Strathprints Institutional Repository

Maximum observable correlation for a bipartite quantum system

Michael J. W. Hall

Theoretical Physics, IAS, Australian National University, Canberra ACT 0200, Australia

Erika Andersson and Thomas Brougham

Department of Physics, University of Strathclyde,

Glasgow G4 0NG, United Kingdom

Abstract

The maximum observable correlation between the two components of a bipartite quantum system is a property of the joint density operator, and is achieved by making particular measurements on the respective components. For pure states it corresponds to making measurements diagonal in a corresponding Schmidt basis. More generally, it is shown that the maximum correlation may be characterised in terms of a 'correlation basis' for the joint density operator, which defines the corresponding (nondegenerate) optimal measurements. The maximum coincidence rate for spin measurements on two-qubit systems is determined to be $(1 + s)/2$, where s is the spectral norm of the spin correlation matrix, and upper bounds are obtained for n-valued measurements on general bipartite systems. It is shown that the maximum coincidence rate is never greater than the computable cross norm measure of entanglement, and a much tighter upper bound is conjectured. Connections with optimal state discrimination and entanglement bounds are briefly discussed.

PACS numbers: 03.65.Ta, 03.67.-a

I. INTRODUCTION

Suppose that two observers, Alice and Bob, have access to the respective components of a bipartite quantum system. If the observers make measurements of observables A and B respectively, the correlation between the measurement outcomes will clearly depend on A and B. It is therefore of interest to ask *what choice of A and B will give the maximum possible correlation*. The answer would allow bipartite states to be ranked in terms of their joint-correlation properties. It is also relevant to the efficient generation of secure keys in quantum cryptography where, all other things being equal, Alice and Bob should aim to compare measurement outcomes which are maximally correlated for a given shared state [1, 2].

It is important to make a distinction here between trivial and non-trivial correlations. For example, if Alice and Bob each simply measure the unit operator, their results will of course be perfectly (but trivially) correlated. Hence the answer to the above question is only of interest if it can be ensured that the measurement outcomes for each component have some useful degree of randomness. This is critical, for example, if Alice and Bob wish to generate a secure cryptographic key [2]. As will be shown, a natural approach is to require that the measured observables are 'maximally informative' or 'nondegenerate'. This is equivalent to requiring the observables to be described by *maximal* probability operator measures (POMs), i.e., $A \equiv \{|a_j\rangle\langle a_j|\}$, $B \equiv \{|b_k\rangle\langle b_k|\}$. It turns out that this requirement is in fact naturally built into some measures of correlation (eg, the mutual information), while it must be imposed explicitly for others (eg, the coincidence rate).

For the case of a pure bipartite state, $|\psi\rangle\langle\psi|$, there is an intuitively obvious answer to the above question: Alice and Bob should choose A and B such that the kets $\{|a_j\rangle\}$ and $\{|b_j\rangle\}$ correspond to a Schmidt decomposition of $|\psi\rangle$, i.e., such that

$$|\psi\rangle = \sum_j \sqrt{p_j}\,|a_j\rangle \otimes |b_j\rangle. \tag{1}$$

Thus, each possible measurement outcome $A = a_j$ will be perfectly correlated with the corresponding measurement outcome $B = b_j$. Note for this case that $\langle a_j|a_k\rangle = \delta_{jk} = \langle b_j|b_k\rangle$. Hence, the optimal observables are described by *orthogonal* POMs, and can be equivalently represented by the Hermitian operators $\hat{A} = \sum_j a_j|a_j\rangle\langle a_j|$ and $\hat{B} = \sum_j b_j|b_j\rangle\langle b_j|$ acting on the respective Hilbert space components [3, 4].

More generally, when the bipartite state is described by some density operator ρ, finding the maximal POMs $A \equiv \{|a_j\rangle\langle a_j|\}$, $B \equiv \{|b_k\rangle\langle b_k|\}$ that maximise a given measure of correlation is quite difficult. Such a pair of maximally-correlated observables will determine a corresponding basis set, $\{|a_j\rangle \otimes |b_k\rangle\}$, for the bipartite system. This basis set generalises the notion of the Schmidt basis for pure states, and may be called a *correlation basis* for ρ. Unlike the Schmidt basis, the correlation basis need not always be orthonormal.

Mutual information and coincidence rate, as measures of correlation, are briefly discussed in Sec. II. Formal equations for the correlation basis are given in Sec. III, for the case of coincidence rate, and illustrated with examples in Sec. III C, including connections with the problem of optimal state discrimination. It is conjectured that at least one of the optimal observables A and B can always be chosen to correspond to an orthogonal POM. In Sec. IV, the maximum coincidence rate for two-valued measurements on pairs of qubits is explicitly determined as a simple function of the spectral norm of the 3×3 spin correlation matrix. This result is generalised in Sec. V, where general upper bounds for coincidence rate are obtained for n-valued measurements, based on a singular value decomposition of the Fano form of the density matrix [5, 6]. These bounds are related to the computable cross norm [7], and are generalised in Sec. VI to connect other linear correlation bounds (such as spin covariance) with entanglement properties.

II. MUTUAL INFORMATION VS COINCIDENCE RATE

To find the optimally correlated observables for a given bipartite system, it is necessary to first quantify joint correlation in some manner. Now, the statistics of any two observables A and B, measured on the respective components of the system, can always be described by corresponding probability operator measures (POMs) $\{A_j\}$ and $\{B_k\}$ (i.e., sets of positive operators which sum to the unit operator [3, 4]), with the joint probability of measurement outcomes $A = a_j$ and $B = b_k$ for a bipartite density operator ρ being given by

$$p_{jk} = \mathrm{tr}[\rho A_j \otimes B_k].$$

Any measure of correlation will be some function of the probability distribution p_{jk}, and two well known examples are discussed in the following.

First, the *mutual information* is defined by [8]

$$I(A, B|\rho) := \sum_{jk} p_{jk} \log_2 \frac{p_j q_k}{p_{jk}}, \tag{2}$$

where p_j and q_k denote the marginal distributions for A and B respectively. This quantity vanishes for uncorrelated observables; is invariant under relabellings of measurement outcomes; and has a simple physical interpretation: if A and B are each measured for a large number of copies of ρ, then $I(A, B|\rho)$ is the average amount of data gained per measurement outcome of A, about the corresponding sequence of measurement outcomes of B (as quantified by the number of bits required to represent the data), and vice versa [8].

The convexity of mutual information implies that the *maximum* mutual information for a given state ρ (also called the accessible information), can always be achieved via observables described by *maximal* POMs [9], i.e., with $A_j \equiv |a_j\rangle\langle a_j|$, $B_k \equiv |b_k\rangle\langle b_k|$. Thus,

$$I_{max}(\rho) := \max_{A,B} I(A, B|\rho) = \max_{A,B \text{ maximal}} I(A, B|\rho). \tag{3}$$

While it is very difficult to determine the optimal observables A and B in Eq. (3), a useful upper bound follows from application of the Holevo bound to the ensemble of states induced on one component of the bipartite system by a measurement on the other component (see Eqs. (12) of Ref. [10]):

$$I_{max}(\rho) \leq \min\{S(\rho_1), S(\rho_2)\}. \tag{4}$$

Here $S(\cdot)$ denotes the von Neumann entropy, and ρ_1 and ρ_2 are the reduced operators for the first and second components of the bipartite system. This bound is sufficiently strong to obtain the maximum mutual information for any mixture $\rho = \sum_\alpha \lambda_\alpha |\psi_\alpha\rangle\langle\psi_\alpha|$ of pure states sharing a common Schmidt basis up to trivial phase factors, i.e., with

$$|\psi_\alpha\rangle = \sum_j \sqrt{p_j^{(\alpha)}} \exp[i\phi_j^{(\alpha)}] \, |a_j\rangle \otimes |b_j\rangle.$$

In particular, Eq. (4) is saturated by choosing A and B to be the maximal POMs generated by this Schmidt basis, yielding

$$I_{max}(\sum_\alpha \lambda_\alpha |\psi_\alpha\rangle\langle\psi_\alpha|) = -\sum_j P_j \log_2 P_j, \tag{5}$$

where $P_j := \sum_\alpha \lambda_\alpha p_j^{(\alpha)}$. Note that for a *pure* state, $|\psi\rangle\langle\psi|$, the maximally correlated observables are therefore those corresponding to a Schmidt basis for $|\psi\rangle$, justifying the intuitive answer given in the Introduction.

Second, the *coincidence rate* measure of correlation is defined by

$$C(A,B|\rho) := \sum_j p_{jj}. \tag{6}$$

This quantity is simply the probability of the observers obtaining matched outcomes, and reaches a maximum of unity only when the outcomes are perfectly correlated (i.e., $p_{jk} = p_j\delta_{jk}$). It is also a little more tractable than mutual information, and will therefore be the focus of this paper. Note that coincidence rate (unlike mutual information) has no clear meaning for continuously-valued outcomes: the quantity $C = \int dx\, p_{xx}$ is not invariant under relabellings of the outcomes (eg, for $x \to \lambda x$ one has $C \to C/\lambda$). Hence, only discretely-valued POMs will be considered in what follows.

Unlike mutual information, the coincidence rate does *not* intrinsically distinguish between trivial and non-trivial correlations. For example, if Alice and Bob each merely measure the unit operator, they will obtain the maximum possible value of coincidence rate (unity), but the minimum possible value of mutual information (zero). Hence, as discussed in the Introduction, it is only of interest to maximise coincidence rate subject to some constraint that ensures a useful degree of randomness for the individual measurement outcomes. One reasonable constraint is the requirement that *the measured observables are maximal POMs*. This constraint is consistent with Eq. (3) for mutual information; does not allow the observers to remove potential information about correlations by merging measurement outcomes; and automatically rules out trivial correlations. The relevant problem of interest is then the determination of observables A and B which achieve the maximum value

$$C_{max}(\rho) := \max_{A,B \text{ maximal}} C(A,B|\rho) = \max_{A,B \text{ maximal}} \sum_j \langle a_j, b_j|\rho|a_j, b_j\rangle. \tag{7}$$

In analogy to Eq. (5) for mutual information, one finds

$$C_{max}(\sum_\alpha \lambda_\alpha |\psi_\alpha\rangle\langle\psi_\alpha|) = \sum_j P_j = 1,$$

for mixtures of states sharing a common Schmidt basis, including all pure states. From Eq. (7) one also obtains the general convexity property

$$C_{max}\left(\lambda\rho + (1-\lambda)\sigma\right) \leq \lambda C_{max}(\rho) + (1-\lambda)C_{max}(\sigma),$$

Hence, if two given observables A and B maximise the coincidence rate for some set of states, S_{AB}, then this set is convex. Eq. (7) further implies that the maximum coincidence rate for

any member of S_{AB} is bounded above by the largest eigenvalue of the 'coincidence operator' $K_{AB} := \sum_j |a_j\rangle\langle a_j| \otimes |b_j\rangle\langle b_j|$, and hence that S_{AB} contains a pure state if and only if this largest eigenvalue is unity.

Finally, it may be recalled that mutual information and coincidence rate are both not only useful measures of correlation per se, but may also be used to differentiate 'classical' from 'quantum' correlations, via corresponding Bell inequalities. For example, if Alice can measure either of A and $\overline{A}$, and Bob can measure either of B and $\overline{B}$, and it is assumed that the statistics of these four observables can be generated by some classical joint probability distribution, then from Eq. (6.5) of Ref. [11] one has

$$I(A,B|\rho) + I(A,\overline{B}|\rho) + I(\overline{A},B|\rho) - I(\overline{A},\overline{B}|\rho) \leq H(A) + H(B),$$

where $H(\cdot)$ denotes the Shannon entropy, while from Eq. (8) of Ref. [12] one has

$$C(A,B|\rho) + C(A,\overline{B}|\rho) + C(\overline{A},B|\rho) - C(\overline{A},\overline{B}|\rho) \leq 2.$$

Each of these inequalities is violated, for example, by suitable spin measurements on a singlet state. The use of correlation measures to characterise the minimum degree of entanglement present has been recently discussed in Ref. [13]. Connections between correlation and entanglement bounds are obtained in Secs. V and VI below.

III. MAXIMISING COINCIDENCE RATE

A. Conditions for extrema

The linearity of coincidence rate with respect to A and B makes it straightforward to characterise the extremal observables, as per the following proposition. The conditions for such observables to *maximise* coincidence rate are less straightforward, however, and are left to the next subsection.

Proposition 1: *Necessary and sufficient conditions for maximal POMs* $A \equiv \{|a_j\rangle\langle a_j|\}$ *and* $B \equiv \{|b_j\rangle\langle b_j|\}$ *to attain an extremal value of coincidence rate, for bipartite state* ρ*, are*

$$\langle a_k, b_l|\rho|a_l, b_l\rangle = \langle a_k, b_k|\rho|a_l, b_k\rangle, \qquad \langle a_k, b_l|\rho|a_k, b_k\rangle = \langle a_l, b_l|\rho|a_l, b_k\rangle \tag{8}$$

for all $k \neq l$*. Moreover, these conditions are equivalent to the existence of Hermitian operators* V *and* W*, acting on the first and second components respectively, satisfying*

$$\left(V - \langle b_j|\rho|b_j\rangle\right)|a_j\rangle = 0, \qquad \left(W - \langle a_j|\rho|a_j\rangle\right)|b_j\rangle = 0 \tag{9}$$

for all j. The corresponding extremal value of coincidence rate is given by

$$C(A,B|\rho) = \mathrm{tr}_1[V] = \mathrm{tr}_2[W]. \tag{10}$$

Proof. Consider the variational quantity

$$J := \sum_j \langle a_j, b_j|\rho|a_j, b_j\rangle - \mathrm{tr}_1[V\,(\sum_j |a_j\rangle\langle a_j| - \hat{1}_1\,)\,] - \mathrm{tr}_2[W\,(\sum_j |b_j\rangle\langle b_j| - \hat{1}_2\,)\,],$$

defined for arbitrary sets of kets $\{|a_j\rangle\}$ and $\{|b_j\rangle\}$ of the same cardinality, where V and W are Hermitian operators that act as Lagrange multipliers for enforcing the completeness constraints

$$\sum_j |a_j\rangle\langle a_j| = \hat{1}_1, \qquad \sum_j |b_j\rangle\langle b_j| = \hat{1}_2. \tag{11}$$

Clearly, $C_{max}(\rho)$ in Eq. (7) corresponds to the global maximum of J under these constraints. Letting $J(\epsilon)$ denote J evaluated under the variations $|a_j\rangle \to |a_j\rangle + \epsilon|m_j\rangle$, $|b_j\rangle \to |b_j\rangle + \epsilon|n_j\rangle$, the extremal points of J correspond to the solutions of $J'(0) = 0$, i.e.,

$$\sum_j \mathrm{tr}_1\left[(|m_j\rangle\langle a_j| + h.c.)(\langle b_j|\rho|b_j\rangle - V)\right] + \sum_j \mathrm{tr}_2\left[(|n_j\rangle\langle b_j| + h.c.)(\langle a_j|\rho|a_j\rangle - W)\right] = 0.$$

Choosing at most one element of the $\{|m_j\rangle, |n_j\rangle\}$ to be non-vanishing (and arbitrary) then yields Eq. (9). Multiplying the latter on the left by $\langle a_k|$ and $\langle b_k|$ further yields

$$\langle a_k, b_j|\rho|a_j, b_j\rangle = \langle a_k|V|a_j\rangle, \qquad \langle a_j, b_k|\rho|a_j, b_j\rangle = \langle b_k|W|b_j\rangle,$$

and Eq. (8) immediately follows from the requirement that V and W are Hermitian. Multiplying on the right of Eq. (9) by $\langle a_j|$ and $\langle b_j|$, and summing over j, yields

$$V = \sum_j \langle b_j|\rho|b_j\rangle\, |a_j\rangle\langle a_j|, \qquad W = \sum_j \langle a_j|\rho|a_j\rangle\, |b_j\rangle\langle b_j|. \tag{12}$$

Taking these as *defining* relations conversely yields Eq. (9) from Eq. (8). The trace of Eq. (12) yields Eq. (10). □

Proposition 1 has a formal connection to the well known problem of distinguishing between members of a given statistical ensemble. In particular, let $\{\rho_j; \lambda_j\}$ denote the ensemble containing state ρ_j with probability λ_j. It is known that necessary and sufficient conditions for a POM $\{\Pi_j\}$ to optimally discriminate between members of this ensemble are [3]

$$(\Upsilon - \lambda_j\rho_j)\Pi_j = 0, \qquad \Upsilon \geq \lambda_j\rho_j \tag{13}$$

for all j, for some Hermitian operator Υ. The first of these conditions is equivalent to Eq. (9) of Proposition 1, for the ensembles $\{\sigma_j; p_j\}$ and $\{\tau_j; q_j\}$ defined by $p_j\sigma_j := \langle b_j|\rho|b_j\rangle$ and $q_j\tau_j := \langle b_j|\rho|b_j\rangle$. Further, summing this first condition over j yields $\Upsilon = \sum_j p_j\sigma_j\Pi_j$, corresponding to Eq. (12).

However, there is no simple analogue of the *second* condition in Eq. (13) - in particular, while the conditions

$$V \geq \langle b_j|\rho|b_j\rangle, \qquad W \geq \langle a_j|\rho|a_j\rangle \tag{14}$$

would immediately imply that A optimally discriminates between members of the ensemble $\{\sigma_j; p_j\}$, and that B optimally discriminates between members of the ensemble $\{\tau_j; q_j\}$, these conditions are *not* sufficient to ensure a maximum for the coincidence rate, as will be shown by explicit example in Sec. III C. Indeed, it is not clear that these conditions are even *necessary*.

Finally, some general properties of extremal observables are worth nothing. First, for *pure* states, the matrix coefficients in Eq. (8) vanish identically for $k \neq l$, for the case where observables A and B correspond to the Schmidt basis decomposition in Eq. (1), and hence these observables are extremal as expected. Second, Eq. (8) implies that if A and B are extremal for two density operators ρ and ρ', then they are extremal for any mixture of ρ and ρ'. Third, if ρ is invariant under some local unitary transformation, i.e., $\rho = U_1 \otimes U_2 \rho U_1^\dagger \otimes U_2^\dagger$, then, for a given solution A and B of Eq. (8), there will be a second solution $\overline{A}$ and $\overline{B}$, with $|\overline{a}_j\rangle = U_1^\dagger|a_j\rangle$ and $|\overline{b}_j\rangle = U_2^\dagger|b_j\rangle$. A similar symmetry holds when ρ is invariant under the interchange of the two component systems.

B. Maxima and n-valued measurements

The second-order variation of the quantity $J(\epsilon)$ appearing in the proof of Proposition 1 immediately yields the condition $J''(0) \leq 0$ for two extremal observables A and B to correspond to a local maximum of coincidence rate. This condition is required to hold only for all kets $|m_j\rangle$ and $|n_j\rangle$ satisfying

$$\sum_j \left(|m_j\rangle\langle a_j| + |a_j\rangle\langle m_j|\right) = 0 = \sum_j \left(|n_j\rangle\langle b_j| + |b_j\rangle\langle n_j|\right) \tag{15}$$

(corresponding to the completeness constraints in Eq. (11), to first order in ϵ). However, the set of such kets is not straightforward to characterise explicitly, and is dependent on the

particular POMs A and B in question, making the condition difficult to verify in practice. In contrast, an explicit and generic condition for $C(A, B|\rho)$ to be a local maximum is obtained in Proposition 2 below, based on the Naimark extension theorem. The restricted problem of maximising coincidence rate over n-valued measurements is also discussed.

Attention will be limited to the case where ρ has finite support. In particular, if H_1 and H_2 are defined to be the Hilbert spaces spanned by the eigenstates of the reduced density operators $\rho_1 := \mathrm{tr}_2[\rho]$, $\rho_2 := \mathrm{tr}_1[\rho]$, then it is assumed that these Hilbert spaces are finite-dimensional, i.e.,

$$d_1 := \dim(H_1) < \infty, \qquad d_2 := \dim(H_2) < \infty. \tag{16}$$

Now, consider a maximal POM $A \equiv \{|a_j\rangle\langle a_j|\}$ on a d-dimensional Hilbert space H, having less than or equal to n non-zero elements (hence, from Eq. (11), $n \geq d$). The Naimark extension theorem then implies there is an n-dimensional Hilbert space H_n containing H as a subspace, and a maximal *orthogonal* POM $X \equiv \{|x_j\rangle\langle x_j|\}$ on H_n (i.e., with $\langle x_j|x_k\rangle = \delta_{jk}$), such that $|a_j\rangle = E|x_j\rangle$, where E denotes the d-dimensional projection operator from H_n to H [4, 14]. The converse result trivially holds: any maximal orthogonal POM on H_n, with 'eigenstates' $\{|x_j\rangle\}$, generates a maximal POM A on H with at most n non-zero elements, defined via $|a_j\rangle := E|x_j\rangle$. Since all d-dimensional subspaces of H_n are unitarily equivalent, this establishes the following Lemma:

Lemma (Naimark extension theorem for maximal POMs): *For a d-dimensional Hilbert space, H, the set of maximal POMs on H having at most $n \geq d$ non-zero elements is characterised by the set of maximal orthogonal POMs on any n-dimensional Hilbert space H_n that contains H as a subspace.*

It follows immediately, taking the limit $n \to \infty$, that the class of *all* maximal POMs on H can be represented by the class of maximal orthogonal POMs on H_∞. Thus, the joint measurement of any two maximal POMs A and B, on the respective components of the tensor product $H_1 \otimes H_2$ spanned by ρ, can be represented by the measurement of two maximal orthogonal POMs X and Y on the respective components of the tensor product $H_\infty \otimes H_\infty$, with

$$|a_j\rangle = E|x_j\rangle, \qquad |b_k\rangle = F|y_k\rangle, \qquad (E \otimes F)\rho = \rho = \rho(E \otimes F), \tag{17}$$

where E and F denote the d_1 and d_2-dimensional projections onto H_1 and H_2 respectively.

In particular, one has

$$C(A,B|\rho) \equiv C(X,Y|\rho) := \sum_{j=1}^{\infty} \langle x_j, y_j|\rho|x_j, y_j\rangle. \tag{18}$$

The advantage of this representation is that maximal orthogonal POMs on H_∞ are connected by unitary transformations. This allows one to explicitly write down the necessary and sufficient conditions for an extremal value of coincidence rate to be a local maximum, as per the following proposition.

Proposition 2: *Two maximal orthogonal POMs X and Y on H_∞, and hence the corresponding maximal POMs A and B defined via Eq. (17), generate a local maximum of coincidence rate if and only if*

$$\langle x_k, y_l|\rho|x_l, y_l\rangle = \langle x_k, y_k|\rho|x_l, y_k\rangle, \qquad \langle x_k, y_l|\rho|x_k, y_k\rangle = \langle x_l, y_l|\rho|x_l, y_k\rangle \tag{19}$$

for all $k \neq l$, and

$$\sum_j \{\langle x_j|M(V - \langle b_j|\rho|b_j\rangle)M|x_j\rangle + \langle y_j|N(W - \langle a_j|\rho|a_j\rangle)N|y_j\rangle + \mathrm{tr}\left(\rho\,[M, |x_j\rangle\langle x_j|] \otimes [N, |y_j\rangle\langle y_j|]\right)\} \geq 0, \tag{20}$$

for all Hermitian operators M and N on H_∞, where V and W are defined as per Eq. (12).

The proof is given in the Appendix. Note that the first condition is equivalent to Eq. (8) of Proposition 1 (and hence to Eq. (9) also), as an immediate consequence of Eq. (17). Further, the second condition is equivalent to the condition $J''(0) \leq 0$ discussed above, if one defines $|m_j\rangle := iEM|x_j\rangle$ and $|n_j\rangle := iFN|y_j\rangle$ (the constraints in Eq. (15) follow from the anti-Hermiticity of the operators $iEME$ and $iFNF$). Note that the presence of the last term in Eq. (20) implies that the conditions in Eq. (14) are *not* sufficient to ensure a local maximum. Examples will be given in Sec. III C below.

Proposition 2 applies to observables having an arbitrary number of possible outcomes. However, it is also of interest to consider the case where A and B are restricted to have a maximum of n possible outcomes, i.e., where the corresponding POMs have at most n non-zero elements. The completeness constraints in Eq. (11) imply that $n \geq d_1, d_2$. The maximum of the coincidence rate over such observables, for a given density operator ρ, will be denoted by $C^{(n)}_{max}(\rho)$. Noting the above Lemma, one has

$$C^{(n)}_{max}(\rho) = \max_{X_n, Y_n} C(X_n, Y_n|\rho) = \max_{X_n, Y_n} \sum_{j,k=1}^{n} \langle x_j, y_j|\rho|x_j, y_j\rangle, \tag{21}$$

where the maximum is over all maximal orthogonal POMs X_n and Y_n on H_n. Clearly, $C^{(n)}_{max}(\rho)$ is a non-decreasing function of n, and converges to $C_{max}(\rho)$, i.e., defining $d := \max\{d_1, d_2\}$,

$$C^{(d)}_{max}(\rho) \leq C^{(n)}_{max}(\rho) \leq C^{(\infty)}_{max}(\rho) = C_{max}(\rho). \tag{22}$$

An explicit expression for $C^{(2)}_{max}(\rho)$ is given in Sec. IV, and general upper bounds for $C^{(n)}_{max}(\rho)$ are obtained in Sec. V.

Now, a maximal POM A with n elements may trivially be extended to an infinite number of elements by defining $|a_j\rangle := 0$ for $j > n$. Hence, such n-valued POMs may be thought of as lying on the 'boundary' of the set of *all* maximal POMs. It would be of interest to show that $C^{(n)}_{max}(\rho)$, corresponding to the maximum of coincidence rate over a restricted portion of this boundary, is also (at the least) a *local* maximum of coincidence rate with respect to the *full set* of maximal POMs. The following corollary to Proposition 2 shows that the conditions in Eq. (14) are sufficient for this to be the case.

Corollary: *If the Hermitian operators V and W defined in Eq. (12) satisfy $V \geq \langle b_j|\rho|b_j\rangle$ and $W \geq \langle a_j|\rho|a_j\rangle$ for all j, for maximal POMs $A^{(n)}$ and $B^{(n)}$ achieving $C^{(n)}_{max}(\rho)$, then $C^{(n)}_{max}(\rho)$ is a local maximum of coincidence rate with respect to the set of* all *maximal POMs.*

Proof. By the above Lemma, maximal POMs with at most n non-zero elements can be represented by the set of maximal orthogonal POMs on H_n. Further, since the group of unitary transformations $U(n) \times U(n)$ is compact, the global maximum of coincidence rate over such orthogonal POMs must be actually be achievable, by two orthogonal POMs X_n and Y_n on H_n, having eigenstates $|x_1\rangle, \ldots, |x_n\rangle$ and $|y_1\rangle, \ldots, |y_n\rangle$ respectively. It may be shown, just as per the proof of Proposition 2, that these eigenstates must satisfy Eqs. (19) and (20) with the ranges of j, k, l restricted $1, 2, \ldots, n$ (and with M and N restricted to H_n). Further, any extension of X_n and Y_n to orthogonal POMs X and Y on H_∞ must satisfy $E|x_j\rangle = 0 = F|x_j\rangle$ for all $j > n$. It follows for such X and Y that (i) Eq. (19) is trivially satisfied (implying the corresponding POMs A and B are extremal); (ii) the first and second terms of Eq. (20) are the same as for X_n and Y_n when $j \leq n$, and nonnegative when $j > n$ (as a consequence of the premise of the Corollary); and (iii) the third term in Eq. (20) is the same as for X_n and Y_n when $j \leq n$, and vanishes when $j > n$. Hence, from Proposition 2, $C^{(n)}_{max}(\rho)$ is a local maximum of coincidence rate with respect to POMs having an *arbitrary* number of elements. □

Note from the above proof that the maximal POMs $A^{(n)}$ and $B^{(n)}$ achieving $C^{(n)}_{max}(\rho)$ must satisfy Eq. (8), with k and l restricted to the range $1, 2, \ldots, n$. It may be checked (noting that ρ is Hermitian) that this places $2n(n-1)$ real constraints on the elements of $A^{(n)}$ and $B^{(n)}$, which are invariant under the $n!$ permutations of the elements that preserve the condition $k \neq l$. On the other hand, to specify two arbitrary maximal POMs, each having no more than n non-zero elements, requires $2n(n-1)$ real parameters (corresponding to specifying the unitary transformations $|x_j\rangle = U_X|z_j\rangle$, $|y_j\rangle = U_Y|z_j\rangle$ on H_n relative to some fixed orthonormal basis $\{|z\rangle\}$, up to arbitary phases), with $(n!)^2$ possible orderings of the elements (i.e., $n!$ orderings for each POM). It is therefore expected, for a generic density operator ρ, that there are $n!$ pairs of extremal candidates for $A^{(n)}$ and $B^{(n)}$ (for density operators having particular symmetries, there will be further extrema, as per the last paragraph of Sec. III A). However, it is conjectured in the next subsection that $C^{(n)}_{max}(\rho)$ is in fact *independent* of n, which is equivalent to equality throughout in Eq. (22). If true, this means that no more than $d!$ candidates for the optimal observables need be checked in the generic case.

C. Two examples and one conjecture

As a first example, we will consider the case of 'trine' measurements on a two-qubit system, for which the measurement on each qubit optimally distinguishes between the states of the ensemble prepared by the measurement on the other qubit, and vice versa. Surprisingly, these measurements do *not* generate a global (or even a local) maximum of coincidence rate.

In particular, let $\{|1\rangle, |2\rangle\}$ be a basis set for either qubit, and consider the 3-valued 'trine' observables $A \equiv B \equiv \{\frac{2}{3}|\phi_j\rangle\langle\phi_j|\}$, where the normalised kets

$$|\phi_1\rangle := |1\rangle, \quad |\phi_2\rangle := \frac{1}{2}\left(|1\rangle + \sqrt{3}\,|2\rangle\right), \quad |\phi_2\rangle := \frac{1}{2}\left(|1\rangle - \sqrt{3}\,|2\rangle\right)$$

form the vertices of an equilateral triangle in the Bloch representation. For the pure bipartite state $\rho = |\psi\rangle\langle\psi|$, with

$$|\psi\rangle := \frac{1}{\sqrt{2}}\left(|1\rangle \otimes |1\rangle + |2\rangle \otimes |2\rangle\right),$$

it is then easily checked that

$$\langle a_j|\rho|a_j\rangle = \frac{1}{3}|\phi_j\rangle\langle\phi_j| = \langle b_j|\rho|b_j\rangle$$

on the respective components. It follows that the operators V and W defined in Eq. (12) are each equal to $\frac{1}{3}\hat{1}$, implying from Proposition 1 that A and B generate an extremal value

of coincidence rate, given by

$$C(A, B|\rho) = \mathrm{tr}[V] = 2/3.$$

Further, the conditions in Eqs. (14) are trivially satisfied for this example, implying via Eq. (13) that A optimally distinguishes between members of the ensemble of states $\{|\phi_j\rangle\langle\phi_j|; \frac{1}{3}\}$ prepared by measurement of B, and vice versa (see also Sec. IV.1(a) of Ref. [3]).

However, A and B above do *not* generate a global maximum of coincidence rate for state ρ, as the maximum possible value of unity may be achieved by instead choosing POMs with elements diagonal with respect to any Schmidt decomposition of $|\psi\rangle$. Indeed, A and B above do not even generate a *local* maximum of coincidence rate - the extremal value of $2/3$ in fact corresponds to a saddle point. To see this, note first that the optimal distinguishing property implies that varying either A or B (while keeping the other fixed) must *decrease* the coincidence rate. Hence, the extremal value of $2/3$ represents a *maximum* with respect to such variations. On the other hand, consider the one-parameter 'mirror-symmetric' family of observables $A^{(\alpha)} \equiv B^{(\alpha)} \equiv \{f_j(\alpha)|\phi_j^{(\alpha)}\rangle\langle\phi_j^{(\alpha)}|\}$, with $0 \leq \alpha \leq 1$, $f_1(\alpha) = 1 - \alpha$, $f_{2,3}(\alpha) = (1 + \alpha)/2$, and [15]

$$|\phi_1^{(\alpha)}\rangle := |1\rangle, \qquad |\phi_{2,3}^{(\alpha)}\rangle = (1+\alpha)^{-1/2}\left(\sqrt{\alpha}|1\rangle \pm |2\rangle\right).$$

Choosing $\alpha = 1/3$ corresponds to the trine observables. It is straightforward to calculate

$$C(A^{(\alpha)}, B^{(\alpha)}|\rho) = \frac{2}{3} + \frac{3}{4}\left(\alpha - \frac{1}{3}\right)^2,$$

and hence the extremal value of $2/3$ represents a *minimum* of coincidence rate with respect to the variation of α [16].

As an example of Proposition 2, consider now a separable state of the form

$$\rho = \sum_{j=1}^{d} \lambda_j \, |\psi_j\rangle\langle\psi_j| \otimes |\chi_j\rangle\langle\chi_j|, \tag{23}$$

where the mutual orthogonality property $\langle\psi_j|\psi_k\rangle = \delta_{jk}$ is satisfied, and each $|\chi_j\rangle$ is arbitrary. Let A be the maximal orthogonal POM defined by $|a_j\rangle := |\psi_j\rangle$, i.e., the optimal POM for distinguishing members of the ensemble $\{|\psi_j\rangle\langle\psi_j|; \lambda_j\}$; and let B be the maximal POM which optimally distinguishes between members of the pure-state ensemble $\{|\chi_j\rangle\langle\chi_j|; \lambda_j\}$ (the existence of such a maximal POM B follows from Theorem 2 of Ref. [17]). Thus, from Eqs. (12) and (13),

$$(V - \lambda_j|\psi_j\rangle\langle\psi_j|)\,|a_j\rangle = 0 = (W - \lambda_j|\chi_j\rangle\langle\chi_j|)\,|b_j\rangle, \qquad V \geq \lambda_j|\psi_j\rangle\langle\psi_j|, \;\; W \geq \lambda_j|\chi_j\rangle\langle\chi_j|.$$

It is then straightforward to check that both conditions of Proposition 2 are satisfied by any X and Y corresponding to A and B respectively (in particular, the third term in Eq. (20) vanishes identically, since the orthogonality of the elements of A implies that ρ and $|x_j\rangle\langle x_j|$ must commute). Hence this choice of A and B generates a local maximum of coincidence rate. Indeed, since a measurement outcome $A = a_j$ for the first component is perfectly correlated with preparation of state $|\chi_j\rangle$ for the second component, and since B is the best possible measurement for distinguishing between such prepared states, the above choice of A and B is intuitively expected to generate a *global* maximum of coincidence rate.

Note that if $d = 2$ in Eq. (23), then B is the orthogonal POM generated by the eigenstates of [3]

$$\eta := \lambda_1|\chi_1\rangle\langle\chi_1| - \lambda_2|\chi_2\rangle\langle\chi_2|.$$

The corresponding maximum value of coincidence rate follows as (cf. Eq. (2.34) in Chap. IV of Ref. [3])

$$C(A, B|\rho) = \frac{1}{2}\left(1 + \mathrm{tr}[|\eta|]\right) = \frac{1}{2}\left[1 + \left(1 - 4\lambda_1\lambda_2|\langle\chi_1|\chi_2\rangle|^2\right)^{1/2}\right]. \tag{24}$$

This result is significantly generalised in Sec. IV.

Finally, note that in the above example that A is an *orthogonal* POM, having the minimum possible number, $n = d$, of non-zero elements. We conjecture that this may be an instance of a general rule. As motivation, observe that if Alice and Bob each measure observables having $n \geq d$ possible outcomes, then the outcomes will typically have a greater degree of randomness when $n > d$. For example, the entropy $H(A)$ of a maximal POM A is bounded below by

$$H(A) = -\sum_j p_j \log_2 p_j \geq -\log_2 \max_j p_j \geq -\log_2 \max_j \langle a_j|a_j\rangle,$$

which is nontrivial if A is non-orthogonal (i.e., if $n > d$). Similarly, the joint entropy of A and B is bounded below by

$$H(AB) \geq -\log_2 \max_{j,k} \langle a_j|a_j\rangle\,\langle b_k|b_k\rangle.$$

Further, the more random a distribution is, the more spread out it is over the set of possible outcomes. Hence, the sum over the diagonal elements of the joint distribution p_{jk} (i.e., the coincidence rate), will typically be smaller. It follows that choosing $n > d$ is typically expected to have a decreasing effect on the maximum achievable coincidence rate:

Conjecture: *The global maximum of coincidence rate, for a bipartite density operator with finite support, can always be achieved by observables A and B having at most $d = \max\{d_1, d_2\}$ possible outcomes, where d_1 and d_2 are the Hilbert space dimensions defined in Eq. (16).*

Note that the conjecture implies at least one of A and B corresponds to an orthogonal POM, depending on whether $d = d_1$ and/or $d = d_2$. Note further that the conjecture corresponds to the case of equality throughout in Eq. (22), i.e., to the condition

$$C_{max}(\rho) \equiv C^{(d)}_{max}(\rho). \tag{25}$$

This conjecture is consistent with the convexity properties discussed in Sec. II and, if true, would greatly simplify the numerical determination of the maximum coincidence rate, as only POMs with d elements would need to be considered. Partial numerical support has been found for the conjecture, for the case of two-qubit systems. In particular, the evaluation of coincidence rate for $\approx 10^{11}$ pairs of maximal POMs having no more than 3 non-zero elements, for each member of a random sample of 1200 bipartite density operators, indicates that $C^{(2)}_{max} \equiv C^{(3)}_{max}$.

IV. MAXIMUM SPIN CORRELATION FOR TWO QUBITS

An exact result for two-qubit systems is derived here, which also introduces the basic method used in the following section to derive general upper bounds for the coincidence rate.

A system of two qubits is described by a density operator ρ on $H_2 \otimes H_2$, so that $d_1 = d_2 = d = 2$. Consider the problem of finding the maximal *two*-valued POMs A and B which maximise the coincidence rate. Such POMs are necessarily orthogonal, corresponding to the measurement of spin in some direction, and hence, noting Eq. (21), the corresponding coincidence rate can be written as

$$C^{(2)}_{max}(\rho) = \max_{a,b} C(\sigma^{(1)} \cdot a, \sigma^{(2)} \cdot b|\rho), \tag{26}$$

where a and b are unit directions. Note that $C^{(2)}_{max}(\rho)$ is in fact equal to the global maximum of coincidence rate, $C_{max}(\rho)$, if the conjecture in Eq. (25) is correct.

To determine $C^{(2)}_{max}(\rho)$, let $|m\rangle$ denote the $+1$ eigenstate of $\sigma \cdot m$ for unit direction m, so that $|m\rangle\langle m| = (1 + \sigma \cdot m)/2$. Hence, $A \equiv \{|a\rangle\langle a|, |-a\rangle\langle -a|\}$, $B \equiv \{|b\rangle\langle b|, |-b\rangle\langle -b|\}$, and

the coincidence rate follows via Eq. (6) as

$$C(A,B|\rho) = \frac{1}{2}\mathrm{tr}[\rho(1+\sigma^{(1)}\cdot a\otimes\sigma^{(2)}\cdot b)] = \frac{1}{2}(1+a^T S b),$$

where S is the 3×3 'spin correlation' matrix defined by

$$S_{jk} := \langle \sigma_j^{(1)}\otimes\sigma_k^{(2)}\rangle. \tag{27}$$

Note that S is real, but in general is not symmetric.

Now, the singular value decomposition theorem [18] states that any real $p\times q$ matrix S can be put in the form

$$S = R_1 D R_2^T, \tag{28}$$

where R_1 and R_2 are real orthogonal matrices (of dimensions $p\times p$ and $q\times q$ respectively), and D is a real $p\times q$ matrix of the form

$$D_{jk} = s_j\delta_{jk}, \qquad s_1\geq s_2\geq\ldots\geq 0.$$

The numbers s_j are called the singular values of S, and are just the square roots of the eigenvalues of each of S^TS and SS^T, while R_1 and R_2 are formed by the respective eigenvectors of S^TS and SS^T [18]. The largest singular value, s_1, is also known as the *spectral norm* of S.

It follows in particular, defining $u = R_1^T a$ and $v = R_2 b$, and using the Schwarz inequality, that for unit vectors a and b one has

$$\begin{aligned}
\max_{a,b} a^T S b &= \max_{u,v}|u^T D v| \\
&= \max_{u,v}\left|\sum_j(\sqrt{s_j}u_j)(\sqrt{s_j}v_j)\right| \\
&\leq \max_{u,v}[\sum_j s_j(u_j)^2]^{1/2}\ [\sum_k s_k(v_k)^2]^{1/2} \\
&= \max_u \sum_j s_j(u_j)^2 \leq (\max_j s_j)\sum_j (u_j)^2 = s_1,
\end{aligned}$$

with equality obtained for the choice $u = v = x := (1,0,0)$. Thus,

$$C^{(2)}_{max}(\rho) = \frac{1}{2}(1+s_1), \tag{29}$$

where s_1 is the spectral norm of the spin-correlation matrix S defined in Eq. (27) (hence one must have $s_1 = 1$ for all pure states), with this maximum coincidence rate being achieved via spin measurements in the directions

$$a = R_1 x, \qquad b = R_2 x.$$

The case of spin measurements on two qubits is thus completely solved.

As a simple example, consider the separable state

$$\rho = \lambda_1 |z\rangle\langle z| \otimes \tau_1 + \lambda_2 |-z\rangle\langle -z| \otimes \tau_2$$

for arbitary qubit density operators τ_1 and τ_2. One finds that all elements of the spin-correlation matrix vanish other than the third row, which is given by the 3-vector r with components

$$r_k := \lambda_1 \mathrm{tr}[\tau_1 \sigma_k^{(2)}] - \lambda_2 \mathrm{tr}[\tau_2 \sigma_k^{(2)}].$$

It follows that only the 33-component of SS^T is non-zero, and equal to $r \cdot r$, yielding

$$C^{(2)}_{max}(\rho) = (1 + |r|)/2.$$

This result generalises Eq. (24) of the previous section, and greatly simplifies calculation of the corresponding coincidence rate, as it does not require explicit diagonalisation of the operator η. Note that the coincidence rate is equal to the average probability for optimally discriminating between members of the ensemble $\{\tau_j; \lambda_j\}$ [3].

As a second example, consider the isotropic state [19]

$$\rho_w = w|\Psi^-\rangle\langle\Psi^-| + \frac{1-w}{3}\hat{1}_T,$$

where $|\Psi^-\rangle$ denotes the singlet state, $\hat{1}_T = \hat{1} - |\Psi^-\rangle\langle\Psi^-|$ denotes the unit operator on the triplet subspace, and $0 \leq w \leq 1$. This state is rotationally-invariant, and the spin-correlation matrix is easily calculated to be $S_{jk} = -(4w-1)\delta_{jk}/3$. It follows immediately that

$$C^{(2)}_{max}(\rho_w) = \frac{1}{2}\left(1 + \frac{|4w-1|}{3}\right),$$

with the maximum coincidence rate being achieved by the choice $a = b$ for $0 \leq w \leq 1/4$, and $a = -b$ for $1/4 \leq w \leq 1$.

Finally, for a general factorisable state $\rho = \rho_1 \otimes \rho_2$, with $\rho_1 = (1 + m \cdot \sigma)/2$ and $\rho_2 = (1 + n \cdot \sigma)/2$, the spin correlation matrix is just the outer product $S = m\, n^T$, so that $SS^T = (n \cdot n) m m^T$, with eigenvalues $(m \cdot m)(n \cdot n)$, 0, and 0. It follows immediately from Eq. (29) that the maximum possible coincidence rate for two uncorrelated qubits is given by

$$C^{(2)}_{max}(\rho_1 \otimes \rho_2) = \frac{1}{2}\left(1 + |m|\,|n|\right),$$

achieved by the choice of the measurement directions $a = m$ and $b = n$.

V. BOUNDS FOR COINCIDENCE RATE

A. A general upper bound

Here an upper bound is given for $C^{(n)}_{max}(\rho)$ in Eq. (21), i.e., for the maximum achievable coincidence rate when Alice and Bob are restricted to measurements of n-valued observables. This bound is tight for the case $n = 2$, reducing to Eq. (29) above. Conversely, taking the limit $n \to \infty$ gives a global upper bound for $C_{max}(\rho)$, which turns out to be equal to the computable cross norm of ρ [7]. Note that if the conjecture in Eq. (25) is correct, then taking $n = d$ will give a much tighter bound in general for $C_{max}(\rho)$.

First, it is well known that the traceless Hermitian operators on an n-dimensional Hilbert space H_n form a real vector space of dimension n^2-1, with inner product $(M, N) := \mathrm{tr}[MN]$ [20]. Hence, if $\{K_p\}$ and $\{L_q\}$ denote two orthonormal basis sets for this vector space, then

$$\mathrm{tr}[K_p K_q] = \delta_{ij} = \mathrm{tr}[L_p L_q], \qquad K_p = \sum_q R_{pq} L_q, \tag{30}$$

for some orthogonal matrix R (i.e., $RR^T = I$). It follows that the trace-free part of any operator Z on H_n can be written as

$$Z - \frac{\mathrm{tr}[Z]}{n}\hat{1} = \sum_p \mathrm{tr}[ZK_p]\, K_p = \sum_q \mathrm{tr}[ZL_q]\, L_q, \tag{31}$$

and that any bipartite density operator ρ on $H_n \otimes H_n$ can be expressed as

$$\rho = \frac{1}{n^2}\hat{1}\otimes\hat{1} + \sum_p u_p\, K_p \otimes \hat{1} + \sum_q v_q\, \hat{1}\otimes L_q + \sum_{p,q} T_{pq}\, K_p \otimes L_q, \tag{32}$$

where

$$u_p := \langle K_p \otimes \hat{1}\rangle/n, \quad v_q := \langle \hat{1}\otimes L_q\rangle/n, \quad T_{pq} := \langle K_p \otimes L_q\rangle. \tag{33}$$

This is referred to as a Fano form for ρ [5, 6].

Now, using Eqs. (30)-(33), the coincidence rate for two maximal orthogonal POMs $X_n \equiv \{|x_j\rangle\langle x_j|\}$ and $Y_n \equiv \{|y_j\rangle\langle y_j|\}$ on H_n simplifies to

$$C(X_n, Y_n|\rho) = 1/n + \mathrm{Tr}[TRW],$$

where Tr denotes the *matrix* trace, and

$$W_{pq} := \sum_j \langle x_j|L_p|x_j\rangle\, \langle y_j|L_q|y_j\rangle. \tag{34}$$

Further, Eq. (7.4.14) of Ref. [18] implies, for any two real matrices T and W and orthogonal matrix R, that

$$|\mathrm{Tr}[TRW]| \leq \sum_k s_k(T) s_k(W),$$

where $s_1(P) \geq s_2(P) \geq \ldots$ denote the singular values of matrix P (see Sec. IV). Hence, noting Eq. (21), one has

$$C^{(n)}_{max}(\rho) \leq 1/n + \sum_k s_k(T) s_k(W). \tag{35}$$

To simplify this upper bound, note first that W can be written, in terms of the vectors

$$\boldsymbol{f}^{(j)}_p := \langle x_j | L_p | x_j \rangle, \quad \boldsymbol{g}^{(j)}_p := \langle y_j | L_p | y_j \rangle,$$

as the sum of outer products $W = \sum_j \boldsymbol{f}^{(j)} (\boldsymbol{g}^{(j)})^T$. Using Eq. (31) one finds

$$\boldsymbol{f}^{(j)} \cdot \boldsymbol{f}^{(k)} = \delta_{jk} - 1/n = \boldsymbol{g}^{(j)} \cdot \boldsymbol{g}^{(k)},$$

implying that

$$W^T W = \sum_j \boldsymbol{g}^{(j)} (\boldsymbol{g}^{(j)})^T = (W^T W)^2, \quad \mathrm{Tr}[W^T W] = n - 1.$$

Thus, $W^T W$ is an $(n-1)$-dimensional projection matrix, implying that the non-zero singular values of W consist of precisely $n-1$ 1s. The above upper bound therefore reduces to

$$C^{(n)}_{max}(\rho) \leq 1/n + \sum_{k=1}^{n-1} s_k(T), \tag{36}$$

where the matrix T is defined in Eq. (33).

The bound can be further simplified, via a judicious choice of the basis sets $\{K_p\}$ and $\{L_q\}$. In particular, recall that ρ only has support on the subspace $H_1 \otimes H_2$ of $H_n \otimes H_n$ (see Sec. III B). The first $(d_1)^2 - 1$ elements of $\{K_1, K_2, \ldots\}$ can therefore be chosen to form a basis set for the traceless operators on H_1, and the first $(d_2)^2 - 1$ elements of $\{L_1, L_2, \ldots\}$ can similarly be chosen to form a basis set for the traceless operators on H_2. Two further basis elements, relabelled as K_0 and L_0 for convenience, will be chosen to have the forms

$$K_0 := \alpha_1 E - \beta_1 (\hat{1} - E), \quad L_0 := \alpha_2 F - \beta_2 (\hat{1} - F),$$

where E and F denote the projections from H_n to H_1 and H_2. The requirements $\mathrm{tr}[K_0] = \mathrm{tr}[L_0] = 0$ and $\mathrm{tr}[(K_0)^2] = \mathrm{tr}[(L_0)^2] = 1$ imply that

$$\alpha_1 = (1/d_1 - 1/n)^{1/2}, \quad \beta_1 = \alpha_1 d_1/(n - d_1),$$

$$\alpha_2 = (1/d_2 - 1/n)^{1/2}\,, \quad \beta_2 = \alpha_2 d_2/(n-d_2).$$

Since the remaining basis elements must be orthogonal to the above basis elements, they cannot contribute to the Fano form of ρ in Eq. (32). Hence, using Eq. (33), the only nonzero rows and columns of the matrix T are given by the $(d_1)^2 \times (d_2)^2$-submatrix

$$T^{(n)} := \begin{pmatrix} \langle K_0 \otimes L_0\rangle & \langle K_0 \otimes L_q\rangle \\ \langle K_p \otimes L_0\rangle & \langle K_p \otimes L_q\rangle \end{pmatrix} = \begin{pmatrix} \alpha_1\alpha_2 & \alpha_1\langle \hat{1}_1 \otimes L_q\rangle \\ \alpha_2\langle K_p \otimes \hat{1}_2\rangle & \langle K_p \otimes L_q\rangle \end{pmatrix}, \tag{37}$$

where $\hat{1}_1$ and $\hat{1}_2$ denote the identity operators on H_1 and H_2 respectively, and $1 \le p \le (d_1)^2 - 1$, $1 \le q \le (d_2)^2 - 1$. Substitution into Eq. (36) yields the main result of this section:

Theorem: *The maximum coincidence rate obtainable for a bipartite state with finite support, via maximal POMs having no more than n nonzero elements, is bounded above by*

$$C^{(n)}_{max}(\rho) \le 1/n + \sum_{k=1}^{\min\{n-1,\delta^2\}} s_k(T^{(n)}) \tag{38}$$

where $\delta := \min\{d_1, d_2\}$*, and the matrix* $T^{(n)}$ *is defined in Eq. (37).*

Since local unitary transformations correspond to left and right multiplication of $T^{(n)}$ by orthogonal matrices, which leave the singular values unchanged [18], this upper bound is invariant under such transformations.

For the case of two qubits, with $n = d_1 = d_2 = 2$, one may choose $K_p = \sigma^{(1)}_p/\sqrt{2}$ and $L_q = \sigma^{(1)}_q/\sqrt{2}$. The 'zeroth' row and column of $T^{(n)}$ vanish for this case, since $\alpha_1 = \alpha_2 = 0$, leaving a 3×3-submatrix equal to one-half of the spin-correlation matrix S in Eq. (27). Thus, for this case, the upper bound of the theorem reduces to $(1 + s_1(S))/2$, which can in fact always be achieved, as per Eq. (29) of the previous section. However, for $n > 3$ the upper bound in Eq. (38) cannot always be attained, essentially because the set of orthogonal matrices R in Eq. (30) is larger than the set of unitary transformations on H_n [20].

B. Examples

Note first that taking the limit $n \to \infty$ in Eq. (38) yields a *global* upper bound for the coincidence rate, independent of the possible number of measurement outcomes:

$$C_{max}(\rho) \le \sum_{k=1}^{\delta^2} s_k(T^{(\infty)}) = \mathrm{Tr}\left[\sqrt{(T^{(\infty)})^T T^{(\infty)}}\right] = \left\|T^{(\infty)}\right\|_{\mathrm{Tr}}. \tag{39}$$

Thus, the upper bound is just the trace norm of $T^{(\infty)}$. Noting that $\alpha_1 \to 1/\sqrt{d_1}$ and $\alpha_2 \to 1/\sqrt{d_2}$ in this limit, it follows that the coefficients of $T^{(\infty)}$ yield a Fano form for ρ on $H_1 \otimes H_2$, via

$$\rho = T_{00}^{(\infty)} \hat{1}_1 \otimes \hat{1}_2/\sqrt{d_1 d_2} + \sum_{p\geq 1} T_{p0}^{(\infty)}\, K_p \otimes \hat{1}_2/\sqrt{d_2} + \sum_{q\geq 1} T_{0q}^{(\infty)}\, \hat{1}_1 \otimes L_q/\sqrt{d_1} + \sum_{p,q\geq 1} T_{pq}^{(\infty)}\, K_p \otimes L_q. \tag{40}$$

The trace norm of $T^{(\infty)}$ may therefore be recognised as the 'computable cross norm' measure of quantum entanglement [6, 7], i.e., *the maximum possible coincidence rate, $C_{max}(\rho)$, is bounded above by the computable cross norm.*

The computable cross norm cannot be greater than unity for any separable states [6, 7], and hence the upper bound in Eq. (39) is always nontrivial for separable states (and for a large proportion of entangled states). However, a stronger bound is postulated further below.

Second, it is of interest to consider measurements restricted to the *minimum* number of possible measurement outcomes, i.e., with $n = d$. For this case $\alpha_1 = \alpha_2 = 0$, implying that the only nonvanishing part of $T^{(d)}$ is the $(d_1^2 - 1) \times (d_2^2 - 1)$ submatrix $\tilde{T}_{pq} = \langle K_p \otimes L_q \rangle$ with $p, q \geq 1$, yielding

$$C_{max}^{(d)}(\rho) \leq 1/d + \sum_{k=1}^{\min\{d-1, \delta^2-1\}} s_k(\tilde{T}). \tag{41}$$

As noted above, this bound is tight for the case $d_1 = d_2 = 2$. For two-qudit systems it bounds the coincidence rate for the case of measurements described by orthogonal POMs. It may also be noted, in analogy to the computable cross norm above, that $\tilde{T}$ has similarly been used in partial characterisations of entanglement [21, 22]. For example, the trace norm of $\tilde{T}$ is never greater than $[(1 - 1/d_1)(1 - 1/d_2)]^{1/2}$ for any separable state [21]. These general underlying connections, between bounds for correlations and measures of entanglement, would be an interesting subject for further investigation (see also Sec. VI).

Third, a simple yet general example of the Theorem is provided by the Werner state for two qudits [19], which has the Fano form [21]

$$\rho_x := \frac{1}{d^2} \hat{1} \otimes \hat{1} + \frac{x - 1/d}{d^2 - 1} \sum_{p\geq 1} K_p \otimes K_p,$$

with $-1 \leq x \leq 1$. It follows via Eqs. (32), (33) and (37) that $T^{(n)}$ is diagonal, and so, noting

that $\alpha_1\alpha_2 = 1/d - 1/n$ for this case, the Theorem yields

$$C^{(n)}_{max}(\rho_x) \leq \frac{1}{n} + (D-1)\frac{|x-1/d|}{d^2-1} + \max\left\{\frac{1}{d} - \frac{1}{n}, \frac{|x-1/d|}{d^2-1}\right\},$$

where $D := \min\{n-1, d^2\}$. Note that, in the limit $n \to \infty$, the righthand side approaches the computable cross norm for Werner states, $1/d + |x - 1/d|$, as expected [7]. It is also straightforward to verify via direct calculation that this bound is *tight* for the case $n = d$ and $x \geq 1/d$, i.e.,

$$C^{(d)}_{max}(\rho_x) = 1/d + |x - 1/d|/(d+1) \tag{42}$$

for $x \geq 1/d$, achieved by the choice $A = B$. For $x < 1/d$, a modification of the Theorem for negative definite $\tilde{T}$ gives the tight upper bound $C^{(d)}_{max}(\rho_x) = 1/d + |x - 1/d|/(d^2-1)$, achieved by maximal orthogonal POMs satisfying $|a_j\rangle = |b_{P(j)}\rangle$ for any permutation P of $1, 2, \ldots, d$ with $P(j) \neq j$ for all j. This example may be regarded as a generalisation of the $d = 2$ isotropic example in Sec. IV, where one identifies x with $1 - 2w$.

Note finally that if the conjecture in Eq. (25) is correct, then the bound in Eq. (41) is in fact an upper bound for $C_{max}(\rho)$, which is generally much tighter than the computable cross norm bound in Eq. (39). For example, consider any state for which the reduced density operators are maximally random, i.e., where $\rho_1 = \hat{1}_1/d_1$ and $\rho_2 = \hat{1}_2/d_2$ (eg, the Werner state ρ_x considered above). It then follows trivially via Eq. (37) that

$$\left\|T^{(\infty)}\right\|_{\rm Tr} = (d_1 d_2)^{-1/2} + \left\|\tilde{T}\right\|_{\rm Tr} \geq 1/d + \left\|\tilde{T}\right\|_{\rm Tr}.$$

Thus, for such states, the bound in Eq. (41) is never greater than that in Eq. (39), and is generally smaller whenever $d \leq \delta^2$.

VI. CONCLUSIONS

It is well known that determining the maximum mutual information between the components of a given bipartite system is a difficult problem [9]. The results of this paper indicate that it is similarly not a straightforward matter to maximise the coincidence rate, despite (i) its linearity with respect to the density operator, and (ii) formal similarities with the well known problem of optimal state discrimination. A notable exception is the case of spin measurements on two-qubit systems, which has been fully solved in Sec. IV. More generally,

one only has available the formal equations for the correlation basis derived in Propositions 1 and 2 of Sec. III, and the upper bounds for n-valued measurements derived in the Theorem of Sec. IV. These general results could be substantially strengthened if the Conjecture of Sec. III C could be verified. It would further be of interest to determine whether or not the 'optimal discrimination' conditions in Eq. (14) must be satisfied by observables corresponding to a global maximum of coincidence rate.

It is worth mentioning here some generalisations of the results in Secs. IV and V, to other linear measures of correlation. For example, note that the spin correlation matrix S in Eq. (27) is closely related to the *spin covariance* matrix $\overline{S}$ defined by

$$\overline{S}_{jk} := \langle \sigma_j^{(1)} \otimes \sigma_k^{(2)} \rangle - \langle \sigma_j^{(1)} \rangle \langle \sigma_k^{(2)} \rangle.$$

In particular, explicitly indicating dependence on the density operator, one has $\overline{S}(\rho) = S(\rho) - S(\rho_1 \otimes \rho_2)$. This covariance matrix has been of recent interest in the characterisation of entanglement [13, 22]. For example, the main result in Sec. IV of Ref. [22] may be simplified to

$$\mathrm{Tr}[\overline{S}^T \overline{S}] = 4\,\mathrm{tr}[(\rho - \rho_1 \otimes \rho_2)^2] \leq 1, \tag{43}$$

for all separable two-qubit states, i.e., *a separable state ρ can lie at distance of at most $1/2$ from $\rho_1 \otimes \rho_2$, as measured by the Hilbert-Schmidt metric.*

Now, the covariance of two arbitrary spin observables, corresponding to directions a and b, may be written as

$$\mathrm{Cov}(A, B|\rho) = a^T \overline{S} b.$$

The methods of Sec. IV then immediately lead to the upper bound

$$\max_{a,b} \mathrm{Cov}(A, B|\rho) = s_1(\overline{S}) \tag{44}$$

analogous to Eq. (29), i.e., *the maximum possible spin covariance for state ρ is given by the spectral norm of the spin covariance matrix.* Note that this bound is invariant under local unitary transformations. It follows, for example, that Theorem 1 of Ref. [13] may be strengthened to the observable-independent statement that

$$\mathrm{tr}[\rho^2] + \frac{1}{2} s_1(\overline{S}) \leq 1 \tag{45}$$

for all separable states of two-qubit systems. Noting Eq. (44), this inequality is also valid if $\mathrm{Cov}(A, B|\rho)$ is substituted for $s_1(\overline{S})$, for *any* spin observables A and B. Similarly, Eq. (24)

of Ref. [13] may be strengthened, using the methods of Sec. IV, to the entanglement bound

$$E_N(\rho) \geq \max\{0, \log_2[s_1(S) + s_2(S)]\} \tag{46}$$

for the logarithmic negativity of a two-qubit system. Thus, as in Sec. V, correlation and entanglement bounds are seen to be closely related.

Finally, consider some general linear measure of correlation, of the form

$$G(A, B|\rho) := \sum_{j,k} p_{jk} g_{jk} = \sum_{jk} g_{jk} \langle a_j, b_k|\rho|a_j, b_k\rangle.$$

Coincidence rate corresponds to the choice $g_{jk} = \delta_{jk}$. The related 'covariance' measure

$$\overline{G}(A, B|\rho) := G(A, B|\rho) - G(A, B|\rho_1 \otimes \rho_2)$$

then has the desirable property of automatically vanishing for uncorrelated states. The methods of Sec. V A may then be applied to $\overline{G}$, with ρ replaced by $\rho - \rho_1 \otimes \rho_2$, to yield the corresponding upper bound

$$\left|\overline{G}(A^{(n)}, B^{(n)}|\rho)\right| = \left|\mathrm{Tr}[\overline{T}RW]\right| \leq \sum_k s_k(\overline{T})\, s_k(W^G), \tag{47}$$

analogous to Eq. (35), for maximal POMs $A^{(n)}$ and $B^{(n)}$ having n elements each. Here

$$\overline{T}_{pq} := \langle K_p \otimes L_q\rangle - \langle K_p\rangle \langle L_q\rangle,$$

and the definition of W in Eq. (34) is generalised to

$$W^G_{pq} := \sum_{j,k} g_{jk} \langle x_j|L_p|x_j\rangle \langle y_k|L_q|y_k\rangle.$$

This bound is tight for spin measurements on two-qubit systems. For the choice $g_{jk} = \delta_{jk}$ the bound simplifies to

$$\left|\mathrm{Corr}(A^{(n)}, B^{(n)}|\rho)\right| \leq \sum_{k=1}^{\min\{n-1, \delta^2-1\}} s_k(\overline{T}) \tag{48}$$

for the 'correlation' $\sum_j (p_{jj} - p_j q_j)$ of any two n-valued maximal POMs (see Sec. II), generalising Eq. (44) above. Note that one may simplify the calculation of the above bounds by choosing the basis elements as in Sec. V, allowing one to replace $\overline{T}$ by the submatrix corresponding to $1 \leq p \leq (d_1)^2 - 1$ and $1 \leq q \leq (d_2)^2 - 1$.

APPENDIX A: PROOF OF PROPOSITION 2

To prove Proposition 2 in Sec. III B, note first that all infinitesimal variations of the orthogonal POMs X and Y in Eq. (18) are generated by infinitesimal unitary transformations on H_∞, and hence are of the form

$$|x_j\rangle \to \exp(i\epsilon M)|x_j\rangle, \qquad |y_j\rangle \to \exp(i\epsilon N)|y_j\rangle$$

for arbitrary Hermitian operators M and N on H_∞, where ϵ is a infinitesimal real parameter. Note from Eq. (18) that these variations are equivalent to keeping X and Y fixed and instead varying the density operator, viz.

$$\rho \to \rho_\epsilon := \exp(-i\epsilon K)\rho \exp(i\epsilon K),$$

with $K := M \otimes 1 + 1 \otimes N$. Expanding in powers of ϵ gives

$$\rho_\epsilon = \rho - i[K,\rho] - (1/2)\epsilon^2[K,[K,\rho]] + \ldots$$

and hence the corresponding variation in coincidence rate is

$$\delta C = -i\epsilon \sum_j \langle x_j, y_j|[K,\rho]|x_j, y_j\rangle - (1/2)\epsilon^2 \sum_j \langle x_j, y_j|[K,[K,\rho]]|x_j, y_j\rangle + \ldots.$$

Requiring the first-order variation to vanish yields

$$0 = \mathrm{tr}_1(M \sum_j [X_j, \langle y_j|\rho|y_j\rangle]\,) + \mathrm{tr}_2(N \sum_j [Y_j, \langle x_j|\rho|x_j\rangle]\,)$$

for arbitrary M and N, where X_j and Y_j denote $|x_j\rangle\langle x_j|$ and $|y_j\rangle\langle y_j|$ respectively. Hence, each operator sum must vanish identically, and Eq. (19) follows as the matrix components of these sums, with respect to the X and Y basis sets respectively.

Requiring the second-order variation to be no greater than zero, as is required for a local maximum, is equivalent to

$$\sum_j \{\mathrm{tr}_1\left([M,[M,X_j]\langle y_j|\rho|y_j\rangle\right) + \mathrm{tr}_2\left([N,[N,Y_j]\langle x_j|\rho|x_j\rangle\right) + 2\,\mathrm{tr}\left(\rho\,[M,X_j|] \otimes [N,Y_j|]\right)\} \geq 0.$$

Now, defining the Hermitian operators

$$\tilde{V} := \sum_j \langle y_j|\rho|y_j\rangle\, |x_j\rangle\langle x_j|, \qquad \tilde{W} := \sum_j \langle x_j|\rho|x_j\rangle\, |y_j\rangle\langle y_j|,$$

and using Eq. (17) and $\sum_j X_j = 1$, the summation over the first term may be simplified to give

$$\begin{aligned}\sum_j \mathrm{tr}_1\left([M,[M,X_j]\langle y_j|\rho|y_j\rangle\right) &= \sum_j \mathrm{tr}_1\left(MX_j\langle y_j|\rho|y_j\rangle M + h.c. - 2X_jM\langle y_j|\rho|y_j\rangle M\right)\\ &= \sum_j \mathrm{tr}_1\left(X_jM[\tilde{V}+\tilde{V}^\dagger]M - 2X_jM\langle y_j|\rho|y_j\rangle M\right)\\ &= 2\sum_j \langle x_j|M(\tilde{V}-\langle b_j|\rho|b_j\rangle)M|x_j\rangle.\end{aligned}$$

The summation over the second term may be similarly simplified in terms of $\tilde{W}$. Equation (20) then immediately follows if it can be shown that $\tilde{V} = V$ and $\tilde{W} = W$.

To do so, note first from Eqs. (12) and (17) that $\tilde{V}E = V$ and $\tilde{W}F = W$. Together with their conjugates, these equations imply $[\tilde{V}, E] = 0 = [\tilde{W}, F]$, and hence that

$$\tilde{V} = V + (1-E)\tilde{V}(1-E), \qquad \tilde{W} = W + (1-F)\tilde{W}(1-F).$$

Substitution into

$$(\tilde{V} - \langle y_j|\rho|y_j\rangle)\,|x_j\rangle = 0, \qquad (\tilde{W} - \langle x_j|\rho|x_j\rangle)|y_j\rangle = 0,$$

(which is equivalent to Eq. (19) precisely as per the equivalence of Eqs. (8) and (9) in Proposition 1), and using Eqs. (8) and (17), then gives

$$(1-E)\tilde{V}(1-E)|x_j\rangle = 0 = (1-F)\tilde{W}(1-F)|y_j\rangle$$

for all j. But $\{|x_j\rangle\}$ and $\{|y_j\rangle\}$ are basis sets for H_∞, implying the operators must vanish identically, and the desired result immediately follows.

---

[1] C.H. Bennett and P.W. Shor, Science **284**, 747 (1999)

[2] B. Schumacher and M.D. Westmoreland, e-print quant-ph/0604207.

[3] C.W. Helstrom, *Quantum Detection and Estimation Theory* (Academic, New York, 1976), Chaps. III and IV.

[4] A.S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982), Chap. II.

[5] U. Fano, Rev. Mod. Phys. **55**, 855 (1983).

[6] I. Bengtsson and K. Zyczkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement* (Cambridge University Press, UK, 2006), Sec. 15.4.

[7] O. Rudolph, e-print quant-ph/0202121; K. Chen and L.A. Wu, Quant. Inf. Comp. **3**, 193 (2003); O. Gühne, M Mechler, G. Tóth, and P. Adam, Phys. Rev. A 74, 010301(R) (2006).

[8] C.E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948).

[9] E.B. Davies, IEEE Trans. Inf. Theory, **24**, 596 (1978).

[10] M.J.W. Hall, Phys. Rev. A **55**, 100 (1997).

[11] S.L. Braunstein and C.M. Caves, Ann. Phys. (N.Y.) **202**, 22 (1990).

[12] M.J.W. Hall, in *Quantum Communication, Computing, and Measurement*, edited by O. Hirota, A.S.Holevo and C.M. Caves (Plenum Press, New York, 1997), pp. 53-61.

[13] K. Audenaert and M.B. Plenio, e-print quant-ph/0608067.

[14] K.R. Parthasararthy, Infin. Dimens. Anal. Quantum Probab. Rel. Top. **2**, 557 (1999).

[15] E. Andersson, S.M. Barnett, C.R. Gilson and K. Hunter, Phys. Rev. A **65**, 052308 (2002).

[16] Note that the coincidence rate reaches the maximum possible value of unity for $\alpha = 1$, with $A^{(1)}$ and $B^{(1)}$ corresponding to orthogonal POMs generated by the Schmidt basis $\{|\phi_2^{(1)}\rangle \otimes |\phi_2^{(1)}\rangle, |\phi_3^{(1)}\rangle \otimes |\phi_3^{(1)}\rangle\}$ for $|\psi\rangle$.

[17] Y.C. Eldar, A. Megretski and G.C. Verghese, IEEE Trans. Inf. Theory **49**, 1007 (2003).

[18] R.A. Horn and C.R. Johnson, *Matrix analysis* (Cambridge University Press, UK, 1985).

[19] R.F. Werner, Phys. Rev. A **40**, 4277 (1989).

[20] J.E. Harriman, Phys. Rev. A **17**, 1249 (1978); K. Lendi, J. Phys. A **20**, 15 (1987).

[21] J.L. Vicente, e-print quant-ph/0607195.

[22] C. Kothe and G. Björk, e-print quant-ph/0608041.