Distinguishability of states and von Neumann entropy

Richard Jozsa
Department of Computer Science, Merchant Venturers Building, University of Bristol, Woodland Road, Bristol BS8 1UB, United Kingdom

Jürgen Schlienz*
School of Mathematics and Statistics, University of Plymouth, Plymouth, Devon PL4 8AA, United Kingdom
(Received 9 November 1999; published 5 June 2000)

Let $\{|\psi_1\rangle, \ldots, |\psi_n\rangle; p_1, \ldots, p_n\}$ be an ensemble of pure quantum states. We show that it is possible to increase all of the pairwise overlaps $|\langle\psi_i|\psi_j\rangle|$, i.e., make each constituent pair of the states more parallel (while keeping the prior probabilities the same), in such a way that the von Neumann entropy S is increased, and dually, make all pairs more orthogonal while decreasing S. We show that this phenomenon cannot occur for ensembles in two dimensions but that it is a feature of almost all ensembles of three states in three dimensions. It is known that the von Neumann entropy characterizes the classical and quantum information capacities of the ensemble and we argue that information capacity, in turn, is a manifestation of the distinguishability of the signal states. Hence, our result shows that the notion of distinguishability within an ensemble is a global property that cannot be reduced to considering distinguishability of each constituent pair of states.

PACS number(s): 03.67.−a

I. INTRODUCTION

The interpretation and physical significance of nonorthogonality is one of the fundamental enigmas in the foundations of quantum theory. Let $|\phi\rangle$ and $|\chi\rangle$ be two nonorthogonal states of a quantum system. We may decompose $|\phi\rangle$ as a superposition of components parallel and perpendicular to $|\chi\rangle$

$$|\phi\rangle = a|\chi\rangle + b|\chi^\perp\rangle, \tag{1}$$

where $\langle\chi^\perp|\chi\rangle = 0$ and $a = \langle\chi|\phi\rangle$. Since any time evolution in quantum mechanics is unitary (when we include also the state of any ambient environment) $|\chi\rangle$ and $|\chi^\perp\rangle$ will evolve as though independent, remaining orthogonal, and the decomposition in Eq. (1) is preserved. Thus, we see that the overlap $|\langle\chi|\phi\rangle|$ measures the extent to which the state $|\phi\rangle$ behaves as though it were actually *equal* to the state $|\chi\rangle$. This view is further formalized in the many worlds interpretation of quantum theory according to which Eq. (1) may be thought of as a ''splitting'' into two ''worlds.'' In one of these worlds the state $|\phi\rangle$ is indeed actually precisely *equal* to $|\chi\rangle$.

The overlap (by which we will always mean the absolute value of the inner product) is also a fundamental ingredient in the question of (non)distinguishability of quantum states. In standard quantum measurement theory $|\langle\chi|\phi\rangle|^2$ is the probability that $|\phi\rangle$ passes the test of ''being the state $|\chi\rangle$.'' Although $|\phi\rangle$ and $|\chi\rangle$ are *distinct* states in the mathematical formalism of quantum theory, there is no physical process that can distinguish them with certainty and indeed the overlap provides a quantitative measure of the extent to which the states cannot be distinguished (as for example, in the Peres measurement [1] for optimal distinguishability of nonorthogonal states. This is in contrast to distinct states in classical physics, which are always perfectly distinguishable in principle.

The purpose of this paper is to describe a situation that appears to contradict the above intuitions. We will describe a situation in which quantum states actually become more distinguishable (in a certain natural sense) when they are made more parallel, i.e., when their overlap increases. Our notion of distinguishability will be based on information–theoretic considerations and will rest on the concept of von Neumann entropy. Recent work in quantum information theory [3–6] has shown that this alternative quantification of distinguishability is very natural and compelling. Indeed if we view quantum states as carriers of information then their capacity for embodying information is a very natural measure of distinguishability i.e. a set of states can communicate more information if and only if the states are made more ''distinguishable'' (all else, such as prior probabilities remaining the same). Quantum states may be used to carry two different kinds of information, classical and quantum information, and we will first briefly outline the essential results characterizing the respective information capacities that form the basis of our quantification of distinguishability in terms of von Neumann entropy.

Consider first the case of quantum information. Suppose that Alice has a source that emits an unending sequence of qubit signal states $|\psi_1\rangle = |0\rangle$ and $|\psi_2\rangle = 1/\sqrt{2}(|0\rangle + |1\rangle)$. Each emission is chosen to be $|\psi_1\rangle$ or $|\psi_2\rangle$ with an equal prior probability a half. Let $\rho = \frac{1}{2}|\psi_1\rangle\langle\psi_1| + \frac{1}{2}|\psi_2\rangle\langle\psi_2|$ be the density matrix of the source and let $S = S(\rho) = -\mathrm{Tr}\,\rho \log_2 \rho$ be its von Neumann entropy. Alice wishes to communicate the sequence of states to Bob. Clearly this may be achieved by transmitting one qubit per emitted state but according to the quantum source coding theorem [3,4,6] she can communicate the quantum information (with arbitrarily

*Present address: Test and Measurement Division, Rohde & Schwarz GmbH & Co. KG, Mühldorfstrasse 15, P.O. Box 801469, D-81671 Munich, Federal Republic of Germany.

high fidelity) using (asymptotically) only S qubits per state. Furthermore, this compression is optimal: no fewer number of qubits per signal can achieve this task. In our example a direct calculation gives that $S=0.601$ qubit per signal.

Now consider the analogous situation in classical physics: if we have two classical signals with equal prior probabilities of a half then no compression beyond 1 bit per signal is possible (by Shannon's source coding theorem [2]) and one may ask what is the origin of the extra nonclassical compression in the quantum case? Clearly it is related to the overlap: if $\theta=\cos^{-1}|\langle\psi_1|\psi_2\rangle|$ is the angle between the two signal states then S depends only on θ and increases monotonically from 0 to 1 as θ increases from 0 to $\pi/2$ (corresponding to the classical situation). In view of Eq. (1) and the discussion following it, one is tempted to think of $|\langle\psi_1|\psi_2\rangle|$ as representing a redundancy or overlap of quantum information between $|\psi_1\rangle$ and $|\psi_2\rangle$, which may be ''compressed out,'' i.e., to some extent $|\psi_1\rangle$ and $|\psi_2\rangle$ are the ''same'' and this common quantum information in every signal, already known to Bob, need not be sent.

More generally, if we have an ensemble of signal states $\{|\psi_1\rangle,\ldots,|\psi_n\rangle\}$ with prior probabilities $p_1,\ldots,p_n$ then the quantum source coding theorem asserts that the quantum information may be compressed to S qubits per signal where S is the von Neumann entropy of $\rho=\Sigma_i p_i|\psi_i\rangle\langle\psi_i|$ and that this compression is optimal. Note that the von Neumann entropy $S(\rho)$ is always less than or equal to the Shannon entropy $H(p_1,\ldots,p_n)$ [7,8] and we might think of the extra quantum compression to S qubits beyond the classical limit of H bits per signal as being due to the overlap of the quantum information represented by the constituent states as expressed in Eq. (1). Our results below will imply that this interpretation is incorrect. Hence the origin of the extra quantum compression is evidently more subtle.

Let $\{|\phi_1\rangle,\ldots,|\phi_n\rangle;p_1,\ldots,p_n\}$ denote the ensemble of quantum states $|\phi_i\rangle$ taken with prior probabilities p_i respectively. Let

$$\mathcal{E}=\{|\psi_1\rangle,\ldots,|\psi_n\rangle;p_1,\ldots,p_n\}$$

and

$$\tilde{\mathcal{E}}=\{|\psi_1\rangle,\ldots,\widetilde{|\psi_n\rangle};p_1,\ldots,p_n\}$$

be two ensembles with the same number of states and with the same corresponding prior probabilities. Let

$$\rho=\sum_i p_i|\psi_i\rangle\langle\psi_i|,\quad \tilde{\rho}=\sum_i p_i|\tilde{\psi}_i\rangle\langle\tilde{\psi}_i|$$

be the respective density matrices and let S and $\tilde{S}$ be the von Neumann entropies. We will show that it is possible to have the following situation: the states of $\mathcal{E}$ are all pairwise more parallel (i.e., have greater overlap) than the corresponding states of $\tilde{\mathcal{E}}$ yet the von Neumann entropy of $\mathcal{E}$ is greater than that of $\tilde{\mathcal{E}}$, i.e., we simultaneously have

$$\text{(ENS1):}\quad |\langle\tilde{\psi}_i|\tilde{\psi}_j\rangle|\leq|\langle\psi_i|\psi_j\rangle|\quad\text{for all } i,j$$

$$\text{(ENS2):}\quad \tilde{S}<S.$$

This is in contradiction to our intuitive discussion above. Each *pair* of states of $\mathcal{E}$ has a greater overlap of quantum information than the corresponding states of $\tilde{\mathcal{E}}$ in the sense of Eq. (1) yet as a *totality* they embody *more* quantum information.

We will say that an ensemble $\mathcal{E}_2$ is a deformation of $\mathcal{E}_1$ if they have the same number of states and the two ensembles have the same list of prior probabilities (in the given order), i.e., the states of $\mathcal{E}_2$ are being thought of as obtained by ''deforming'' the corresponding states of $\mathcal{E}_1$ while keeping the probabilities fixed.

We will show that (ENS1) with (ENS2) can never be satisfied for any pair of ensembles $\mathcal{E}$ and $\tilde{\mathcal{E}}$ in *two* dimensions (regardless of the number of states) but that for almost all ensembles $\mathcal{E}$ of three states in three dimensions, there is an ensemble $\tilde{\mathcal{E}}$ satisfying (ENS1) and (ENS2).

The phenomenon in (ENS1) and (ENS2) shows curiously, that information capacity (or distinguishability) is a *global* property of a set of states and not an accumulative local property of pairs of constituent states. Indeed for ensembles $\{|\psi_1\rangle,|\psi_2\rangle;p_1,p_2\}$ of just two states, the von Neumann entropy is a monotonically decreasing function of the overlap $|\langle\psi_1|\psi_2\rangle|$. Thus the overlap conditions in (ENS1) imply that each constituent pair of signals has a *diminished* capacity for information as we pass from $\tilde{\mathcal{E}}$ to $\mathcal{E}$ yet (ENS2) states that the ensemble as a whole develops an *increased* capacity.

Von Neumann entropy also characterises the *classical* information capacity of an ensemble of quantum states. Suppose that Alice is constrained to use the states $|\psi_i\rangle$ with prior probabilities p_i and she wishes to communicate classical information to Bob. On receiving a string of states, Bob is allowed to perform any joint measurement on a signal block of any length in order to maximise his acquired mutual information about the identity of the states. Then it may be shown [5] that the von Neumann entropy S of the signal ensemble gives the maximum amount of information per signal that Alice is able to reliably transmit to Bob under the above constraints. Now *classical* information capacity is very closely related to the concept of distinguishability, which, by any definition, is itself a form of classical information about the identity of the states. Then (ENS1) with (ENS2) shows that the members of an ensemble of quantum states can become pairwise less distinguishable yet as a whole the ensemble becomes more distinguishable, i.e., has more classical information capacity.

II. GRAM MATRIX FORMULATION

We now describe a formalism for studying the conditions represented by (ENS1) and (ENS2) and give a method for generating realizations in dimension $d>2$.

Consider an ensemble $\{|\psi_1\rangle,\ldots,|\psi_n\rangle;p_1,\ldots,p_n\}$ of n states in d dimensions. The density matrix is

$$\rho=\sum_{i=1}^{n} p_i|\psi_i\rangle\langle\psi_i|. \tag{2}$$

We introduce the Gram matrix **G** defined as the $n\times n$ matrix of rescaled inner products

$$G_{ij}=\sqrt{p_i p_j}\langle\psi_i|\psi_j\rangle. \tag{3}$$

The Gram matrix enjoys the following two fundamental properties:

(G1): The nonzero eigenvalues of **G** are the same as the nonzero eigenvalues of ρ (and their respective multiplicities are also the same). Note that in general $d\neq n$ so the mismatch in the numbers of eigenvalues is made up by zero eigenvalues. It follows that ρ and **G** also have the same von Neumann entropy.

To see this, introduce n orthogonal vectors $|e_i\rangle$ in an auxiliary Hilbert space and consider the pure state

$$|\phi\rangle=\sum_{i=1}^{n} \sqrt{p_i}|\psi_i\rangle|e_i\rangle. \tag{4}$$

Then ρ and **G** are just the two reduced states obtained by partial trace of $|\phi\rangle\langle\phi|$ over the second and first components, respectively. Hence, they must have the same nonzero eigenvalues (c.f. appendix of Ref. [7]).□

(G2): **G** is always a positive matrix and $\mathrm{Tr}\,\mathbf{G}=1$. Conversely, if **A** is any $m\times m$ positive matrix with $\mathrm{Tr}\,\mathbf{A}=1$ then **A** is the Gram matrix of an ensemble of m states in m dimensions [9].

The first part follows immediately from Eq. (4) where **G** is identified as a density matrix itself. For the converse statement note that if **A** is positive we can write $\mathbf{A}=\mathbf{B}^2$ where **B** is Hermitian so $\mathbf{A}=\mathbf{B}^\dagger\mathbf{B}$. Let $\hat{b}_i$ be the normalized ith column of **B** and let t_i be its squared length. Then $\mathbf{A}=\mathbf{B}^\dagger\mathbf{B}$ expresses precisely the fact that **A** is the Gram matrix of the ensemble of m-dimensional states $\{\hat{b}_1,\ldots,\hat{b}_m;t_1,\ldots,t_m\}$. The probabilities t_i are just the diagonal entries of **A**.□

The Gram matrix, expressed in terms of the inner products rather than the states themselves, provides a natural vehicle for studying the conditions (ENS1) and (ENS2). Indeed, we are generally not interested in the actual positions of the ensemble states but only in their relative positions, i.e., inner products. The following theorem shows that the Gram matrix encodes this information while eliminating the superfluous data of overall unitary repositionings:

Lemma 1: Two ensembles

$$\mathcal{E}_1=\{|\alpha_1\rangle,\ldots,|\alpha_m\rangle;p_1,\ldots,p_m\}\ \text{ on }\ \mathcal{H}_1$$

and

$$\mathcal{E}_2=\{|\beta_1\rangle,\ldots,|\beta_n\rangle;q_1,\ldots,q_n\}\ \text{ on }\ \mathcal{H}_2$$

have equal Gram matrices $G_1=G_2$ if and only if $m=n$, $p_i=q_i$ for $i=1,\ldots,m$ and there is a unitary transformation U on $\mathcal{H}_1\oplus\mathcal{H}_2$ with $|\beta_i\rangle=U|\alpha_i\rangle$ for all i.□

We give the proof in the appendix.

Let **G** and $\tilde{\mathbf{G}}$ be respectively the Gram matrices of the ensembles $\mathcal{E}$ and $\tilde{\mathcal{E}}$, which have the same prior state probabilities (i.e., **G** and $\tilde{\mathbf{G}}$ have the same diagonals). Then (ENS1) is equivalent to $|\tilde{G}_{ij}|\leqslant|G_{ij}|$ for each i and j, i.e.

$$\tilde{G}_{ij}=r_{ij}G_{ij}\quad r_{ij}\in\mathcal{C}\text{ and }|r_{ij}|\leqslant 1\text{ for each }i,j. \tag{5}$$

The matrix $\mathbf{r}=[r_{ij}]$ satisfies the following properties:

(R1) **r** is Hermitian (without loss of generality)

(R2) the diagonal entries r_{ii} are all equal to 1 and $|r_{ij}|\leqslant 1$ for all i and j.

However, for given **G**, **r** cannot be chosen arbitrarily subject to (R1) and (R2) because $\tilde{\mathbf{G}}$ is required to be a positive matrix [by (G2)]. Also, we wish to choose **r** so that Eq. (5) induces a decrease in the entropy of the Gram matrix **G**.

The componentwise product of **G** and **r** in Eq. (5) is known as the Hadamard product of matrices. We denote it by

$$\tilde{\mathbf{G}}=\mathbf{r}*\mathbf{G} \tag{6}$$

to distinguish it from the usual matrix product. The Schur product theorem [9] asserts that the Hadamard product of any positive matrices is again a positive matrix. Hence, if **r** is chosen to be positive then by (G2) $\tilde{\mathbf{G}}$ will again correspond to an ensemble of states. However, in this case (ENS2) can never be satisfied, i.e., the entropy is nondecreasing. To see this, let **G** be the Gram matrix of an ensemble $\mathcal{E}$ comprising n states $|\psi_i\rangle$ with probabilities p_i. If **r** satisfying (R1) and (R2) is *positive* then (G2) implies that $\mathbf{r}/n$ is also a Gram matrix of some collection of states, $|\xi_1\rangle,\ldots,|\xi_n\rangle$ say, taken with equal prior probabilities $1/n$. Thus,

$$r_{ij}=\langle\xi_i|\xi_j\rangle.$$

Hence, Eq. (5) asserts that $\tilde{\mathbf{G}}$ is the Gram matrix of the ensemble $\mathcal{E}(\xi)$ comprising the states $|\psi_i\rangle\otimes|\xi_i\rangle$ with probabilities p_i. Thus, $\mathcal{E}(\xi)$ is an ''extension'' of $\mathcal{E}$ obtained by simply adjoining the states $|\xi_i\rangle$ to the corresponding $|\psi_i\rangle$'s. As such, the entropy $S(\xi)$ of $\mathcal{E}(\xi)$ can never be smaller than the entropy S of $\mathcal{E}$. We give three brief proofs of this fact, each invoking a different (substantial) theorem. First, if $S(\xi)<S$ then Alice could reliably communicate the quantum information of $\mathcal{E}$ to Bob using less than S qubits per signal. She simply adjoins the states $|\xi_i\rangle$, compresses to $S(\xi)$ qubits per signal and on reception and decompression, Bob just discards the extensions. This contradicts the quantum noiseless coding theorem [3,4,6]. Secondly, in a similar way, $S(\xi)<S$ contradicts the classical information capacity theorem [5] (which asserts that the von Neumann entropy is the classical information capacity): the extended ensemble $\mathcal{E}(\xi)$ cannot have a smaller information capacity since Alice and Bob can always just ignore the presence of the extensions for the purposes of classical communication. Thirdly, the passage from $|\psi_i\rangle|\xi_i\rangle$ to $|\psi_i\rangle$ (i.e. discarding the second state) is a physically realizable operation and hence a completely positive (CP) map. Then Uhlmann's monotonicity theorem [10,11] (asserting that relative entropy can never increase under any CP map) immediately implies that the entropy of $\mathcal{E}$

cannot exceed the entropy of $\mathcal{E}(\xi)$ [since the entropy of any ensemble of pure states $\rho_i = |\phi_i\rangle\langle\phi_i|$ is just the average relative entropy $\Sigma_i p_i S(\rho||\rho_i)$ where $\rho = \Sigma_i p_i \rho_i$ is the ensemble density matrix].

Hence, we have

Lemma 2: If we wish to satisfy (ENS2) together with (ENS1) it is necessary that the matrix **r** of multipliers in Eq. (5) have at least one negative eigenvalue.□

An example described later (having $\tilde{\mathbf{G}}$ positive) will show that this necessary condition is unfortunately not also sufficient.

III. ENSEMBLES IN TWO DIMENSIONS

We begin by proving

Lemma 3: (ENS1) and (ENS2) can never be simultaneously satisfied for any ensembles in two dimensions.□

For the case of two states ($n=2$) we have already noted that lemma 3 follows readily from the explicit formula for von Neumann entropy which, for any p_1 and p_2, depends monotonically on the overlap of the two states. Alternatively in this case we note that **r** is a 2 by 2 matrix which by (R1) and (R2) must take the form

$$\mathbf{r} = \begin{pmatrix} 1 & e^{i\beta}\cos\alpha \\ e^{-i\beta}\cos\alpha & 1 \end{pmatrix}.$$

This is always a positive matrix (as the eigenvalues $1 \pm \cos\alpha$ are both non-negative) and we then apply lemma 2.

For general values of n we introduce the linearized entropy S_{lin} defined by

$$S_{\text{lin}} = \text{Tr}(\rho - \rho^2).$$

Substituting Eq. (2) we get

$$S_{\text{lin}} = \left(1 - \sum_{i=1}^{n} p_i^2\right) - 2\sum_{i<j} p_i p_j |\langle\psi_i|\psi_j\rangle|^2. \tag{7}$$

Hence, S_{lin} is a monotonically decreasing function of each of the overlaps $|\langle\psi_i|\psi_j\rangle|$. Next, we note that for $d=2$, the von Neumann entropy $S(\rho)$ is also a monotonically increasing function of S_{lin}, giving our claimed result that $S(\rho)$ is a monotonically decreasing function of each overlap. To see that S is monotonically increasing with S_{lin} for $d=2$ let the eigenvalues of ρ be λ and $1-\lambda$. Then,

$$S = -\lambda\log\lambda - (1-\lambda)\log(1-\lambda)$$

and

$$S_{\text{lin}} = \lambda + (1-\lambda) - \lambda^2 - (1-\lambda)^2.$$

Computing $dS/d\lambda$ and $dS_{\text{lin}}/d\lambda$ shows that

$$\frac{dS}{dS_{\text{lin}}} = \frac{dS}{d\lambda} \Big/ \frac{dS_{\text{lin}}}{d\lambda} > 0.$$

IV. ENSEMBLES IN THREE DIMENSIONS

We will focus on the case that the inequalities in (ENS1) are all equalities

$$|\langle\tilde{\psi}_i|\tilde{\psi}_j\rangle| = |\langle\psi_i|\psi_j\rangle| \quad \text{for all } i,j. \tag{8}$$

This will readily imply our basic result *viz* that the entropy can be increased by increasing the overlap of each pair of states. Indeed suppose that $\mathcal{E}$ and $\tilde{\mathcal{E}}$ are two different ensembles (with the same prior probabilities) satisfying Eq. (8) but with $S \neq \tilde{S}$. Without loss of generality suppose that $\tilde{S} < S$. Now consider a small deformation of the states of $\mathcal{E}$, which slightly increases all the pairwise overlaps, giving an ensemble $\mathcal{E}^*$ with entropy S^*. Then either $S^* > S$ (so $S^* > \tilde{S}$ giving a direct example of our result) or $S^* \leqslant S$. In the latter case, we will always have $S^* > \tilde{S}$ if the deformation is sufficiently small, again giving an example of our result (with ensembles $\tilde{\mathcal{E}}$ and $\mathcal{E}^*$).

We will show that for almost any ensemble $\mathcal{E}$ of three linearly independent states (i.e., $n=3$ and $d=3$) it is possible to deform $\mathcal{E}$, in two ways such that

(D1) all pairwise overlaps are increased and the entropy increases,

(D2) all pairwise overlaps are decreased and the entropy decreases.

For clarity in this section, it is useful to distinguish (normalized) state vectors, written as kets $|\psi\rangle$ from physical states, which we will denote with square brackets as $[\psi]$. A physical state is a full set of all normalized vectors that differ only in overall phase:

$$[\psi] = \{e^{i\phi}|\psi\rangle : 0 \leqslant \phi < 2\pi\},$$

i.e., state vectors are elements of the Hilbert space whereas physical states are elements of the projective Hilbert space. The vectors $e^{i\phi}|\psi\rangle$ in $[\psi]$ are called phase representatives of the state $[\psi]$.

Consider three normalized vectors $|\psi_1\rangle, |\psi_2\rangle, |\psi_3\rangle$ in $\mathcal{H}_3$. We have the overlaps (non-negative real numbers)

$$a_{12} = |\langle\psi_1|\psi_2\rangle|, \quad a_{23} = |\langle\psi_2|\psi_3\rangle|, \quad a_{31} = |\langle\psi_3|\psi_1\rangle|,$$

and the triple quantity denoted Υ_{123} (generally complex) defined as

$$\Upsilon_{123} = \langle\psi_1|\psi_2\rangle\langle\psi_2|\psi_3\rangle\langle\psi_3|\psi_1\rangle = a_{12}a_{23}a_{31}e^{i\xi}, \quad \text{for some phase } \xi.$$

Note that the real numbers $a_{12}, a_{23}, a_{31}, \xi$ are well defined on physical states rather than just on the vectors, i.e., if we arbitrarily change the phases of the vectors then these four numbers remain invariant. Also for any prior probabilities, the density matrix and entropy of the ensemble $\{|\psi_1\rangle, |\psi_2\rangle, |\psi_3\rangle; p_1, p_2, p_3\}$ is a function of the corresponding physical states. Note also that unitary transformations are well defined on physical states and leave the quantities $a_{12}, a_{23}, a_{31}, \xi$ invariant.

From the point of view of physics, we are primarily interested in ensembles of physical states rather than ensembles of state vectors. In contrast to the overlaps a_{ij} and ξ, the Gram matrix depends on the choices of phase representatives. For any ensemble, given any choice of phase representative of $[\psi_1]$ we may always choose representatives of $[\psi_2]$ and $[\psi_3]$ to make $\langle\psi_1|\psi_2\rangle$ and $\langle\psi_1|\psi_3\rangle$ real and positive so the Gram matrix has the form

$$\mathbf{G}=\begin{pmatrix} p_1 & \sqrt{p_1p_2}a_{12} & \sqrt{p_1p_3}a_{31} \\ \sqrt{p_1p_2}a_{12} & p_2 & \sqrt{p_2p_3}a_{23}e^{i\xi} \\ \sqrt{p_1p_3}a_{31} & \sqrt{p_2p_3}a_{23}e^{-i\xi} & p_3 \end{pmatrix}. \quad (9)$$

To any ensemble of physical states we will associate a Gram matrix of this form depending only on the invariants a_{ij}, ξ and the prior probabilities. Then, if $\mathcal{E}$ and $\tilde{\mathcal{E}}$ have the same overlaps [as in Eq. (8)], the Gram matrices must be related as in Eq. (6) by a matrix of multipliers $\mathbf{r}$ of the form

$$\mathbf{r}(\phi)=\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & e^{i\phi} \\ 1 & e^{-i\phi} & 1 \end{pmatrix}. \quad (10)$$

To study deformations preserving overlaps we begin by giving a characterization of the set of all possible triples of physical states $\{[\psi_1],[\psi_2],[\psi_3]\}$ compatible with a given prescribed set a_{12},a_{23},a_{31} of overlaps. We clearly have any overall unitary transformation of any allowed triple but we are especially interested in triples that are not unitarily related. A complete characterization is given in the following theorem, whose proof is given in the appendix.

Theorem 1: Suppose we are given real numbers

$$0\leqslant a_{12}\leqslant 1, \quad 0\leqslant a_{23}\leqslant 1, \quad 0\leqslant a_{31}\leqslant 1.$$

(a) If $\{[\psi_1],[\psi_2],[\psi_3]\}$ is any set of physical states having a_{12},a_{23},a_{31} as overlaps then the phase ξ of Υ_{123} satisfies

$$1+2a_{12}a_{23}a_{31}\cos\xi\geqslant a_{12}^2+a_{23}^2+a_{31}^2. \quad (11)$$

Conversely, for any solution ξ of Eq. (11) there is a set of states having overlaps a_{12},a_{23},a_{31} and triple quantity $\Upsilon_{123}=a_{12}a_{23}a_{31}e^{i\xi}$. Thus the existence of a solution of Eq. (11) is a necessary and sufficient condition for a set of real numbers a_{12},a_{23},a_{31} to be realisable as a set of overlaps.

(b) The solutions ξ of Eq. (11) give a one-to-one parameterization of all sets of states up to unitary equivalence, that have the prescribed overlaps. We always take ξ to be in the interval $[-\pi,\pi]$ so if Eq. (11) has solutions they are always of the form $-\xi_{max}\leqslant\xi\leqslant\xi_{max}$ with $\xi_{max}\leqslant\pi$ (and we identify the values $\pm\pi$ if $\xi_{max}=\pi$).

(c) (Explicit formulas up to unitary equivalence). Suppose that $\{[\psi_1],[\psi_2],[\psi_3]\}$ is any set of physical states having a_{12},a_{23},a_{31} as overlaps. Then there is an orthonormal basis $\{|0\rangle,|1\rangle,|2\rangle\}$ of $\mathcal{H}_3$ such that phase representatives of the states are given by

$$|\psi_1(\xi)\rangle=|0\rangle,$$

$$|\psi_2(\xi)\rangle=a_{12}|0\rangle+\sqrt{1-a_{12}^2}|1\rangle, \quad (12)$$

$$|\psi_3(\xi)\rangle=a_{13}|0\rangle+\frac{a_{23}e^{i\xi}-a_{23}a_{31}}{\sqrt{1-a_{12}^2}}|1\rangle + \frac{\sqrt{1-a_{12}^2-a_{23}^2-a_{31}^2+2a_{12}a_{23}a_{31}\cos\xi}}{\sqrt{1-a_{12}^2}}|2\rangle.$$

Here, ξ is the phase of Υ_{123}.□

To illustrate theorem 1 and its significance for further developments we describe an example.

Example 1. Consider the ensemble $\mathcal{E}=\{|\alpha_1\rangle,|\alpha_2\rangle,|\alpha_3\rangle;\frac{1}{3},\frac{1}{3},\frac{1}{3}\}$ where

$$|\alpha_1\rangle=|0\rangle,$$

$$|\alpha_2\rangle=\frac{1}{\sqrt{2}}(|0\rangle+|1\rangle), \quad (13)$$

$$|\alpha_3\rangle=\frac{1}{\sqrt{3}}(|0\rangle+|1\rangle+|2\rangle).$$

The Gram matrix has eigenvalues 0.053, 0.145, and 0.802 with entropy $S=0.613$ (where natural logarithms have been used). According to theorem 1, up to unitary equivalence any ensemble with the same overlaps has the form $\mathcal{E}(\xi)=\{|\alpha_1(\xi)\rangle,|\alpha_2(\xi)\rangle,|\alpha_3(\xi)\rangle;\frac{1}{3},\frac{1}{3},\frac{1}{3}\}$ where

$$|\alpha_1(\xi)\rangle=|0\rangle,$$

$$|\alpha_2(\xi)\rangle=\frac{1}{\sqrt{2}}(|0\rangle+|1\rangle), \quad (14)$$

$$|\alpha_3(\xi)\rangle=\frac{1}{\sqrt{3}}|0\rangle+\frac{2e^{i\xi}-1}{\sqrt{3}}|1\rangle+\sqrt{\frac{4}{3}\cos\xi-1}|2\rangle.$$

Here, the parameter ξ is constrained by Eq. (11) giving

$$\cos\xi\geqslant\frac{3}{4} \quad \text{i.e. } \xi_{max}=0.72\,\text{rad}.$$

$\mathbf{G}(\xi)$, the Gram matrix of $\mathcal{E}(\xi)$, is positive so long as $|\xi|\leqslant\arccos\frac{3}{4}$. At the maximum value of ξ the amplitude of $|2\rangle$ becomes zero and the states become linearly dependent.

The von Neumann entropy of $\mathcal{E}(\xi)$ may be computed from the eigenvalues of $\mathbf{G}(\xi)$. This is shown in Fig. 1 and we see that the entropy falls monotonically with ξ for the

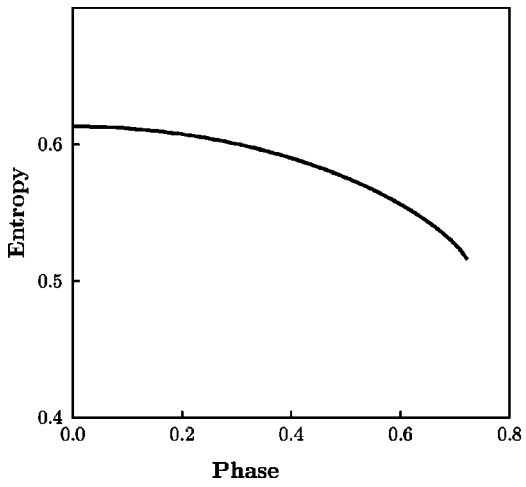

FIG. 1. Variation of von Neumann entropy for the family of ensembles in Eq. (14) with unchanging overlaps, as a function of ξ, the phase of Υ_{123}. ξ is measured in radians and the entropy is calculated using natural logarithms. The maximum value of ξ is $\arccos\frac{3}{4}\approx 0.72$ radians.

ensembles $\mathcal{E}(\xi)$ which have constant overlaps. (For negative values of ξ the graph is reflected in the vertical axis). Thus, for any $0<\xi<\arccos\frac{3}{4}$ we may deform $\mathcal{E}(\xi)$ to obtain an ensemble $\mathcal{E}^*$ with the same overlaps but having strictly lower (respectively higher) entropy. Then by slightly deforming $\mathcal{E}^*$ to make all overlaps lower (respectively higher) we obtain a deformation of $\mathcal{E}(\xi)$ satisfying **(D2)** [respectively **(D1)**]. Note that if $\xi=0$ (i.e. Υ_{123} is real) then we only get **(D2)** [and not **(D1)**] by *this* method. Also for $\xi=\xi_{max}=\arccos\frac{3}{4}$ we get only **(D1)** (by *this* method) as ξ can only be decreased.□

We now return to general ensembles of three states and study the variation of von Neumann entropy with deformations that preserve the overlaps. According to Theorem 1 these deformations, up to unitary equivalence, are parametrized by ξ. We will find that the behavior exhibited in example 1 is generic–the entropy $S(\xi)$ falls monotonically as ξ increases from 0 to ξ_{max} just as in Fig. 1. Then the same deformation arguments as in example 1 give the following theorem:

Theorem 2: Suppose the ensemble $\mathcal{E}=\{|\psi_1\rangle,|\psi_2\rangle,|\psi_3\rangle;p_1,p_2,p_3\}$ has rank 3 (i.e. the unnormalized states $\sqrt{p_i}|\psi_i\rangle$ are linearly independent) with overlaps a_{12},a_{23},a_{31} and phase ξ_0 of Υ_{123}.

(a) If Υ_{123} is not real (i.e., $\xi_0\neq 0,\pi$) then $\mathcal{E}$ can be deformed according to both **(D1)** and **(D2)**.

(b) If Υ_{123} is real positive (i.e., $\xi_0=0$) then $\mathcal{E}$ can be deformed according to **(D2)**. If Υ_{123} is real negative (i.e., $\xi_0=\pi$) then $\mathcal{E}$ can be deformed according to **(D1)**.□

The proof of theorem 2 is given in the appendix.

Note that in theorem 2(b) we have not ruled out the possibility of a **(D1)** deformation [respectively **(D2)** deformation] when $\mathcal{E}$ has Υ_{123} real positive (respectively negative). We have shown only that such deformations cannot be achieved by the particular method of first altering the entropy while keeping the overlaps constant and then slightly increasing or decreasing the overlaps. Indeed consider the ensemble $\mathcal{F}$ of states

$$|\psi_1\rangle=|0\rangle$$

$$|\psi_2\rangle=\frac{1}{\sqrt{2}}(|0\rangle+|1\rangle)$$

$$|\psi_3\rangle=\frac{\sqrt{2}-1}{6}|0\rangle+\frac{\sqrt{2}+1}{6}|1\rangle$$

taken with equal probabilities. Here, we have Υ_{123} real positive so our method cannot give a **(D1)** deformation. However such deformations do exist, for example the ensemble $\mathcal{F}'$ of states:

$$|\psi_1\rangle=|0\rangle$$

$$|\psi_2\rangle=\frac{1}{\sqrt{2}}(|0\rangle+|1\rangle)$$

$$|\psi_3\rangle=\frac{2}{3\sqrt{2}}|0\rangle+\frac{4}{3\sqrt{3}}|1\rangle+\frac{\sqrt{7}}{3\sqrt{3}}|2\rangle.$$

$\mathcal{F}'$ has overlaps greater than or equal to those of $\mathcal{F}$ yet its entropy $S'=0.91$ is greater than the entropy $S=0.85$ of $\mathcal{F}$, i.e., $\mathcal{F}'$ is a deformation of $\mathcal{F}$ of type **(D1)**.

This leads us to conjecture that the conclusions in theorem 2(a) also hold for the ensembles in (b) and more generally that any ensemble containing a subset of three states that has no parallel or orthogonal pairs, admits deformations of both types **(D1)** and **(D2)**. However, it is possible to show from our results that if such deformations always exist, they are *not* always realisable as *continuous* deformations of the appropriate type starting with the given ensemble. Indeed let $\mathcal{E}$ be any ensemble with Υ_{123} complex, consisting of three linearly dependent states, lying in a plane P [so that from theorem 1(c) the phase ξ of Υ_{123} must take the maximum value allowed by the overlaps]. From the result of Sec. III we know that there are no deformations (continuous or not) of type **(D1)** or **(D2)** lying within the plane P. Since $\mathcal{E}$ is planar, its density matrix has a zero eigenvalue, say $\lambda_3=0$. The entropy function $S(\lambda_1,\lambda_2,\lambda_3)$ has infinite slope at $\lambda_3=0$ in the λ_3 direction (and finite slopes in the λ_1 and λ_2 directions, for $\lambda_1,\lambda_2>0$) so any continuous deformation of $\mathcal{E}$ out of the plane P must begin to *increase* S, regardless of the overlaps. Hence, no *continuous* deformations, by any method whatever, of type **(D2)** can exist, yet there may still exist a ''distant'' ensemble $\mathcal{E}'$ with decreased overlaps and decreased entropy [i.e., a **(D2)** deformation] which cannot be connected to $\mathcal{E}$ by a continuous family of **(D2)** deformations.

In relation to theorem 1 it is interesting to note that in Ref. [17] Gisin and Popescu have described another method for (discontinuous) deformation of a particular class of ensembles that preserves all overlaps. Let $|n\rangle$ denote the qubit

state corresponding to the unit vector n on the Bloch sphere. Let $\mathcal{E}=\{|n_1\rangle|n_1\rangle, \ldots, |n_k\rangle|n_k\rangle; p_1 \ldots, p_k\}$ be any ensemble of two qubit states where each constituent state comprises two copies of a single qubit state. We now replace each second component $|n_i\rangle$ by the orthogonal state $|-n_i\rangle$ to get $\mathcal{E}^*=\{|n_1\rangle|-n_1\rangle, \ldots, |n_k\rangle|-n_k\rangle; p_1 \ldots, p_k\}$. Clearly, $\mathcal{E}$ and $\mathcal{E}^*$ have the same pairwise overlaps but in general the von Neumann entropies will be different. Indeed $\mathcal{E}$ always lies within the three dimensional subspace of symmetric states whereas $\mathcal{E}^*$ generally spans the full four dimensions of two qubits. In Ref. [17] the ensemble $\mathcal{E}$ uniformly distributed over the whole Bloch sphere was considered. It was shown that $\mathcal{E}^*$ allows one to guess the identity of the direction n with greater average fidelity than is possible with the states from $\mathcal{E}$. This provides another manifestation of the idea that the distinguishability in an ensemble can vary while keeping all pairwise overlaps fixed.

V. DISCUSSION

One of the initial motivations for this work was the problem of determining the optimal compression of quantum information in mixed state signals [12,13]. For pure states the optimal compression is given by the von Neumann entropy of the source (and Schumacher compression [3,4] provides an explicit asymptotically optimal compression protocol) but for mixed states the value is unknown. One method of compressing the ensemble $\{\rho_1, \ldots, \rho_n; p_1, \ldots, p_n\}$ of mixed states (where the compresser knows the identity of the signals) is to first prepare purifications $|\psi_i\rangle$ of the states ρ_i and then apply Schumacher compression to the pure state ensemble $\{|\psi_1\rangle, \ldots, |\psi_n\rangle; p_1, \ldots, p_n\}$. Thus, we wish to construct the ensemble of purifications that has the least von Neumann entropy. For $n=2$ the solution is given by the purifications with the largest overlap and the corresponding minimum entropy can be readily calculated from Uhlmann's transition probability formula [14,15]. However, for three or more states the results in this paper show that the maximum overlap condition is no longer correct and the problem of minimizing the entropy is evidently more subtle. It has been shown in Ref. [16] that the problem of optimal mixed state compression, in full generality, may be translated into a problem of minimizing the entropy of a suitable ensemble of purifications (of blocks of signal states).

From a mathematical viewpoint our results amount to an investigation of the von Neumann entropy S not as a function of a density matrix, but as a function of variables defining an ensemble:

$$S=S(\rho)=S(\Sigma p_i|\psi_i\rangle\langle\psi_i|)=S(|\psi_1\rangle, \ldots, |\psi_n\rangle; p_1, \ldots, p_n).$$

In particular, we have considered the behavior of S when the states $|\psi_i\rangle$ are varied while the probabilities p_i are held fixed. In this context, many further interesting questions arise. For example, given an ensemble, what is the largest entropy that can be attained by deformations that make the states more parallel? Conversely given a value S_0 of von Neumann entropy, what is the maximum possible (average) overlap of any ensemble that has entropy S_0? Are there ensembles with the property that every continuous deformation that increases overlaps, begins to increase the entropy? We may also consider varying the probabilities. Suppose we have two ensembles $\mathcal{E}=\{|\psi_1\rangle, |\psi_2\rangle, |\psi_3\rangle; p_1, p_2, p_3\}$ and $\mathcal{E}^*=\{|\phi_1\rangle, |\phi_2\rangle, |\phi_3\rangle; q_1, q_2, q_3\}$ where the states of $\mathcal{E}^*$ have greater pairwise overlap than the corresponding states of $\mathcal{E}$. Let cap $(\mathcal{E})$ be the maximum entropy attainable from the states in $\mathcal{E}$ by varying the prior probabilities. Is it possible to have $\text{cap}(\mathcal{E}^*) > \text{cap}(\mathcal{E})$? This question is of particular interest since $\text{cap}(\mathcal{E})$ is the classical information capacity of a quantum channel using the states of $\mathcal{E}$ as basic signal states [5].

In the introduction we argued that von Neumann entropy quantifies a notion of distinguishability for the constituent states of an ensemble and pointed out that it is surprising that distinguishability (as well as the lower limit for compressibility) may be increased by increasing all pairwise overlaps (for ensembles of three or more states). Thus, the notion of distinguishability and the redundancy involved in quantum information compression depend not only on the overlaps but also on the relative phases of the amplitudes $\langle\psi_i|\psi_j\rangle$. For ensembles with just two states there is only one such phase and it is rendered physically irrelevant by the overall phase freedom in a physical quantum state. For three or more states the relative phases cannot be eliminated and they provide further parameters for issues of compressibility and distinguishability. For general ensembles we can consider the quantities:

$$\Upsilon(i_1, i_2, \ldots, i_k)=\langle\psi_{i_1}|\psi_{i_2}\rangle\langle\psi_{i_2}|\psi_{i_3}\rangle\cdots\langle\psi_{i_{k-1}}|\psi_{i_k}\rangle \times\langle\psi_{i_k}|\psi_{i_1}\rangle \tag{15}$$

associated to each subset $[\psi_{i_1}], \ldots, [\psi_{i_k}]$ of physical states (noting the cyclic chain of indices $i_1, \ldots, i_k, i_1$ returning to i_1 to complete the cycle.) These quantities are all unitary invariants and independent of the choice of phase representatives. Chains of length 1 define the normalization while chains of length 2 are just (the squares of) the pairwise overlaps. The modulus of any chain of length k is a product of overlaps but its phase is a new quantity. For example for three states, the phase of any chain of length 3 is (up to a sign) the parameter ξ considered previously. It would be interesting to understand the physical bearing of these parameters on issues of distinguishability and compressibility.

We have seen that a family of ensembles with fixed overlaps can exhibit a variation of information content. In particular for ensembles of three states we have the extra quantum mechanical phase parameter ξ. It is interesting to note that an analogous phenomenon may occur in a purely classical context [19]. Suppose Alice wishes to communicate information to Bob using three signals *A, B, C* with equal prior probabilities. The signals themselves are probability distributions on three values $\{1,2,3\}$. *A* is the uniform distribution on $\{1,2\}$, *B* on $\{2,3\}$, and *C* on $\{1,3\}$. To send a signal, Alice samples the corresponding distribution and sends the result, e.g., to send *B* she tosses a fair coin labeled by 2 and 3, and sends the outcome. (Alternatively we may attribute the probabilistic nature of the signals to noise in the channel). Bob then reads the received value. We may readily

calculate the probability $p(y|X)$ that Bob reads y (1, 2, or 3) given that Alice sent X (A, B, or C) and the mutual information $I(X:Y)$ between Alice and Bob. In classical information theory, $I(X:Y)$ quantifies the amount of information that Bob gets about Alice's message, i.e., the information capacity of the communication protocol. Consider now the same scenario with three new signals A', B', and C', which are probability distributions on four values $\{1,2,3,4\}$. A' is the uniform distribution on $\{1,4\}$, B' on $\{2,4\}$, and C' on $\{3,4\}$. The mutual information is now different but in any reasonable sense, the signals A', B', and C' have the same pairwise overlaps as the corresponding signals from the original set A, B, C—in every case the two distributions coincide on one value and are disjoint on the other value. Thus, in this purely classical context we again have the phenomenon that the global information content (in the sense of mutual information here) differs even though the pairwise overlaps (and pairwise information contents) are the same.

There are various other possible natural concepts of distinguishability, which one can associate to an ensemble $\{|\psi_1\rangle, \ldots, |\psi_n\rangle; p_1, \ldots, p_n\}$ of pure quantum states. Two such concepts are the accessible information and the minimum error probability [18]. For the latter, we attempt to identify the state by applying a measurement $\mathcal{M}$ with outcomes $1,2,\ldots,n$. Let $p(j|i)$ be the probability of obtaining outcome j for the input state $|\psi_i\rangle$. The error probability is defined by $P_{\text{error}}=\Sigma_i p_i(1-p(i|i))$. We choose $\mathcal{M}$ to minimize P_{error} and use that minimum value as a measure of distinguishability for the ensemble (where a smaller value indicates greater distinguishability). The accessible information of an ensemble is the maximum Shannon mutual information of any POVM measurement on the ensemble states. For $n=2$ both of these measures are monotonic functions of the overlap $|\langle\psi_1|\psi_2\rangle|$ but for $n\geqslant 3$, like the von Neumann entropy, they are not functions of the overlaps alone. Thus it seems plausible that they too—like the von Neumann entropy—will exhibit increased distinguishability with suitable deformations of the ensemble that decreases all pairwise overlaps. But unlike the von Neumann entropy this is difficult to study analytically: the computation of minimum error probability or accessible information involves difficult optimisations and is generally intractible analytically for all but the simplest ensembles.

ACKNOWLEDGMENTS

This work was initiated at the Elsag-Bailey ISI workshop on quantum computation held in Torino and we are grateful for the support and opportunity of collaboration provided by that meeting. We are especially grateful to Armin Uhlmann for originally suggesting the Gram matrix approach adopted in this paper and for other helpful comments. Much of the work was supported by the European TMR Research Network ERB-FMRX-CT96-0087. RJ is also supported by the UK Engineering and Physical Sciences Research Council.

APPENDIX

Proof of Lemma 1

Suppose that the two ensembles are unitarily related. Then it is easy to see that they have the same Gram matrices (as the inner products are unitary invariants).

Conversely, suppose that $\mathbf{G}_1=\mathbf{G}_2$. Then the number of states and corresponding probabilities (being the diagonal of the Gram matrix) must be the same. So write

$$\mathcal{E}_1=\{|\alpha_1\rangle, \ldots, |\alpha_k\rangle; p_1, \ldots, p_k\}$$

and

$$\mathcal{E}_2=\{|\beta_1\rangle, \ldots, |\beta_k\rangle; p_1, \ldots, p_k\}.$$

We will use proof by induction on k to show that $|\alpha_i\rangle = U|\beta_i\rangle$ for some unitary U. The result is clearly true for $k=1$, i.e., ensembles with only one state. Assume (the induction hypothesis) that it is true for all ensembles of k states Consider two ensembles of $k+1$ states:

$$\mathcal{E}_1=\{|\alpha_1\rangle, \ldots, |\alpha_k\rangle, |\alpha_{k+1}\rangle; p_1, \ldots, p_k, p_{k+1}\}$$

$$\mathcal{E}_2=\{|\beta_1\rangle, \ldots, |\beta_k\rangle, |\beta_{k+1}\rangle; p_1, \ldots, p_k, p_{k+1}\}$$

with the same Gram matrices. Then the subensembles of just the first k states [with probabilities rescaled by $1/(1-p_{k+1})$] will have the same Gram matrices so by the induction hypothesis there is a unitary U with

$$|\beta_i\rangle = U|\alpha_i\rangle \quad \text{for } i=1,\ldots,k\,.$$

Now compare the ensembles $U(\mathcal{E}_1)$ and $\mathcal{E}_2$. They differ only in their $(k+1)$th states that are respectively $U|\alpha_{k+1}\rangle$ and $|\beta_{k+1}\rangle$. Let $B=\text{span}(|\beta_1\rangle, \ldots, |\beta_k\rangle)$. Consider the parallel and perpendicular projections with respect to B

$$U|\alpha_{k+1}\rangle = (U\alpha)_{\parallel} + (U\alpha)_{\perp}$$

$$|\beta_{k+1}\rangle = (\beta)_{\parallel} + (\beta)_{\perp}\,.$$

Since $\mathcal{E}_1$ and $\mathcal{E}_2$ [and hence also $U(\mathcal{E}_1)$] have equal Gram matrices, $U|\alpha_{k+1}\rangle$ and $|\beta_{k+1}\rangle$ have equal inner products with $|\beta_1\rangle, \ldots, |\beta_k\rangle$ and hence with a basis of B. Thus, the parallel projections $(U\alpha)_{\parallel}$ and $(\beta)_{\parallel}$ are equal and so the perpendicular projections have the same length. If this length is zero then $\mathcal{E}_2=U\mathcal{E}_1$. If it is not zero, let $B^{\perp}$ be the two-dimensional space spanned by $(U\alpha)_{\perp}$ and $(\beta)_{\perp}$ and let V be a unitary transformation in $B^{\perp}$ with $V(U\alpha)_{\perp}=(\beta)_{\perp}$. Finally, let U' be the unitary transformation which is the identity in B and V in $B^{\perp}$. Then $\mathcal{E}_2=U'U\mathcal{E}_1$, completing the proof of the lemma.□

Proof of Theorem 1

For any set of states $\{[\psi_1],[\psi_2],[\psi_3]\}$ we can always choose phase representatives making $\langle\psi_1|\psi_2\rangle$ and $\langle\psi_1|\psi_3\rangle$ real non-negative. Hence, $\{[\psi_1],[\psi_2],[\psi_3]\}$ will have overlaps a_{12}, a_{23}, a_{31}

iff there are representatives with

$$\langle\psi_1|\psi_2\rangle = a_{12} \quad \langle\psi_1|\psi_3\rangle = a_{31}$$

and

$$\langle\psi_2|\psi_3\rangle = a_{23}e^{i\xi} \quad \text{for some } \xi \tag{A1}$$

iff

$$\mathbf{G}(\xi) \text{ defined by } \frac{1}{3}\begin{pmatrix} 1 & a_{12} & a_{31} \\ a_{12} & 1 & a_{23}e^{i\xi} \\ a_{31} & a_{23}e^{-i\xi} & 1 \end{pmatrix}$$

is a Gram matrix (A2)

iff $3\mathbf{G}$ is a positive matrix [where $\mathbf{G}$ is the matrix in Eq. (A2)]

iff the eigenvalues $\lambda_1,\lambda_2,\lambda_3$ of $3\mathbf{G}$ are all non-negative.

Next, we claim (and prove in the next paragraph) that the last condition holds **iff** $\det 3\mathbf{G}\geqslant 0$. Then computing $\det 3\mathbf{G}$ directly from $\mathbf{G}$ above, gives the required condition (11). Note also from Eq. (A1) that ξ is the phase of the triple quantity Υ_{123} and this completes the proof of (a).

To justify the claim above, we show that $3\mathbf{G}$ cannot have exactly two negative eigenvalues. Consider $\beta=\lambda_1\lambda_2+\lambda_2\lambda_3+\lambda_3\lambda_1$ and suppose that there are two negative eigenvalues λ_1 and λ_2. Then, $\mathrm{Tr}\,3\mathbf{G}=3=\lambda_1+\lambda_2+\lambda_3$ gives

$$\begin{aligned}\beta&=\lambda_1\lambda_2+(\lambda_1+\lambda_2)[3-(\lambda_1+\lambda_2)]\\&=-\lambda_1^2-\lambda_2^2-\lambda_1\lambda_2+3(\lambda_1+\lambda_2).\end{aligned}$$

Thus, $\beta<0$ as all terms are negative. But β is the linear coefficient in the characteristic equation of $3\mathbf{G}$ and a direct calculation of $\det(\lambda I-3\mathbf{G})$ gives

$$\beta=3-a_{12}^2-a_{23}^2-a_{31}^2$$

so that $\beta\geqslant 0$. Hence, $3\mathbf{G}$ can never have exactly two negative eigenvalues and the non-negativity of $\det 3\mathbf{G}=\lambda_1\lambda_2\lambda_3$ is equivalent to the non-negativity of all three eigenvalues.

To prove (b) we recall that Υ_{123} and hence ξ, is an invariant of any overall unitary transformation. Hence, for fixed a_{12}, a_{23}, and a_{31}, different ξ's correspond to unitarily inequivalent sets of states.

To prove (c) suppose that $\{[\psi_1],[\psi_2],[\psi_3]\}$ is any set of states with the given overlaps. We choose an orthonormal basis $\{|0\rangle,|1\rangle,|2\rangle\}$ of $\mathcal{H}_3$ and phase representatives $|\psi_i\rangle$ as follows. Choose an arbitrary phase representative $|\psi_1\rangle$ of $[\psi_1]$ and set

$$|0\rangle=|\psi_1\rangle.$$

Choose $|1\rangle$ orthogonal to $|0\rangle$ in the plane of $[\psi_1]$ and $[\psi_2]$ (with phase of $|1\rangle$ to be fixed later). Then any representative $|\psi_2\rangle$ of $[\psi_2]$ has the form

$$|\psi_2\rangle=a_{12}e^{i\alpha}|0\rangle+\beta|1\rangle.$$

Choose the overall phase of $|\psi_2\rangle$ to make $\alpha=0$ and choose the phase of $|1\rangle$ to make β real and non-negative, so

$$|\psi_2\rangle=a_{12}|0\rangle+\sqrt{1-a_{12}^2}|1\rangle.$$

Then choose $|2\rangle$ orthonormal to $|0\rangle$ and $|1\rangle$ (with phase to be fixed later) so any representative of $[\psi_3]$ has the form

$$|\psi_3\rangle=\omega|0\rangle+\eta|1\rangle+\zeta|2\rangle.$$

Choose the overall phase of $|\psi_3\rangle$ to make ω real non-negative so $\omega=a_{31}$. Choose the phase of $|2\rangle$ to make ζ real non-negative. Then,

$$|\psi_3\rangle=a_{31}|0\rangle+\eta|1\rangle+z|2\rangle \quad \text{with } z \text{ real}\geqslant 0,\ \eta \text{ complex.} \tag{A3}$$

We have two further conditions

$$|\langle\psi_2|\psi_3\rangle|=a_{23}=|a_{31}a_{12}+\sqrt{1-a_{12}^2}\eta| \tag{A4}$$

and normalization

$$z^2=1-|\eta|^2-a_{31}^2. \tag{A5}$$

From Eq. (A4) we introduce ξ so that

$$a_{31}a_{12}+\sqrt{1-a_{12}^2}\eta=a_{23}e^{i\xi}.$$

This gives η parameterized by ξ

$$\eta=\frac{a_{23}e^{i\xi}-a_{31}a_{12}}{\sqrt{1-a_{12}^2}}. \tag{A6}$$

Then Eq. (A5) gives

$$\begin{aligned}z^2&=\frac{1-a_{12}^2-a_{23}^2-a_{31}^2+2a_{12}a_{23}a_{31}\cos\xi}{1-a_{12}^2}\\&=\frac{\det\begin{pmatrix}1 & a_{12} & a_{31}\\ a_{12} & 1 & e^{i\xi}a_{23}\\ a_{31} & e^{-i\xi}a_{23} & 1\end{pmatrix}}{1-a_{12}^2}\end{aligned} \tag{A7}$$

Substituting Eqs. (A6) and (A7) into Eq. (A3) gives the required form of $|\psi_3\rangle$. Note that the condition $z^2\geqslant 0$ in Eq. (A7) reproduces the condition (11) in part (a) of the theorem. □

Proof of Theorem 2

Given the ensemble $\mathcal{E}=\{|\psi_1\rangle,|\psi_2\rangle,|\psi_3\rangle;p_1,p_2,p_3\}$ let $\mathcal{E}(\xi)$ denote the family of ensembles (up to unitary equivalence) which have the same overlaps as $\mathcal{E}$. Here, ξ is the phase of Υ_{123}. We study the variation of S with ξ through a sequence of lemmas (with proofs given at the end):

Lemma A1: The ensembles $\mathcal{E}(\xi)$ have constant $\mathrm{Tr}\,\rho$ and constant $\mathrm{Tr}\,\rho^2$. Hence, S may be viewed as a function of $\mathrm{Tr}\,\rho^3=\mathrm{Tr}\,\mathbf{G}^3$.

Lemma A2: If $\mathrm{Tr}\,\rho$ and $\mathrm{Tr}\,\rho^2$ are held constant then S is a monotonically increasing function of $\mathrm{Tr}\,\rho^3$.

Lemma A3: For any ensemble $\mathcal{E}$, $\mathrm{Tr}\,\rho^3$ has the form

$$\begin{aligned}\mathrm{Tr}\,\rho^3&=C+6p_1p_2p_3\,\mathrm{Re}\,\Upsilon_{123}\\&=C+6p_1p_2p_3a_{12}a_{23}a_{31}\cos\xi,\end{aligned} \tag{A8}$$

where C is independent of ξ.

In view of these lemmas we have

Proof of Theorem 2: We consider the family of ensembles $\mathcal{E}(\xi)$ with the same overlaps as $\mathcal{E}$, parametrized by ξ, the phase of Υ_{123}.

(a) If Υ is not real and $\xi_0 \neq \pm \xi_{max}$ (as $\mathcal{E}$ has rank 3) then we can perturb ξ_0 slightly in both the positive and negative directions inducing either an increase or decrease in $\mathrm{Tr}\,\rho^3$ (by Lemma A3). Hence, by Lemmas A1 and A2 we may either increase or decrease S while keeping the overlaps the same, giving the possibilities **(D1)** and **(D2)**.

(b) If $\xi_0 = 0$ then $\cos\xi_0$ is at its maximum value. Thus, any perturbation of ξ_0 leads to a decrease in $\mathrm{Tr}\,\rho^3$ and hence a decrease in S (while keeping all overlaps constant). Thus, we get **(D2)** only. At $\xi_0 = \pi$, $\cos\xi_0$ is minimum so similarly, we get only **(D1)**.□

Proof of Lemma A1: We have $\mathrm{Tr}\,\rho = 1$ for all ξ. Also $\mathrm{Tr}\,\rho^2 = \mathrm{Tr}\,\mathbf{G}^2 = \Sigma_{ij} G_{ij} G_{ji} = \Sigma |G_{ij}|^2$ as $\mathbf{G}$ is Hermitian. Since $\mathcal{E}(\xi)$ have constant overlaps, the Gram matrices are related by Hadamard product with $\mathbf{r}$ of the form in Eq. (10) with $|r_{ij}| = 1$ for all i,j. Hence, $\mathrm{Tr}\,\rho^2$ remains constant. The three eigenvalues of ρ are uniquely determined by the values of $\mathrm{Tr}\,\rho$, $\mathrm{Tr}\,\rho^2$, and $\mathrm{Tr}\,\rho^3$ so S may be viewed as a function of $\mathrm{Tr}\,\rho^3$. □

Proof of Lemma A2: Note first that S is unitarily invariant so we may assume without loss of generality that ρ is diagonal, and work on the classical probability simplex

$$\mathcal{P}_3 = \{(\lambda_1,\lambda_2,\lambda_3): \lambda_1 \geqslant 0, \lambda_2 \geqslant 0, \lambda_3 \geqslant 0 \quad \text{and} \quad \lambda_1 + \lambda_2 + \lambda_3 = 1\}.$$

To represent the constraint $\mathrm{Tr}\,\rho^2 = \text{const.}$ it is convenient to introduce a polar coordinate system (r,θ) in $\mathcal{P}_3$ as shown in Fig. 2. r is measured from the center $M = (\frac{1}{3},\frac{1}{3},\frac{1}{3})$ of $\mathcal{P}_3$ and θ is measured anticlockwise from the line joining M to the vertex (1,0,0).

A direct calculation gives the coordinates of a general state as

$$\lambda_1 = \frac{1}{3} + \sqrt{\frac{2}{3}} r \cos\theta,$$

$$\lambda_2 = \frac{1}{3} + \sqrt{\frac{2}{3}} r \cos\left(\theta + \frac{2\pi}{3}\right), \tag{A9}$$

$$\lambda_3 = \frac{1}{3} + \sqrt{\frac{2}{3}} r \cos\left(\theta + \frac{4\pi}{3}\right).$$

The constraint

$$\mathrm{Tr}\,\rho^2 = \sum \lambda_i^2 = A^2 = \text{const.} \tag{A10}$$

corresponds to the intersection of the simplex XYZ with a sphere of radius A centred at (0,0,0). This gives a circle or part of a circle (as shown in Fig. 2) within the simplex. In terms of polar coordinates we get

$$\sum \lambda_i^2 = \frac{1}{3} + r^2. \tag{A11}$$

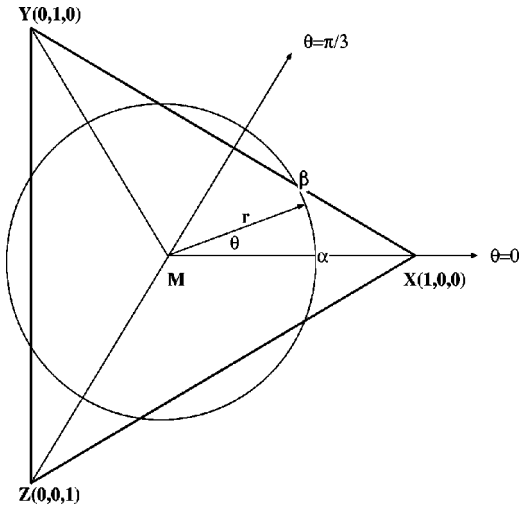

FIG. 2. The triangle **XYZ** is the probability simplex $\mathcal{P}_3$ [or space of diagonal density matrices $\mathrm{diag}(\lambda_1,\lambda_2,\lambda_3)$] in $d=3$ dimensions. It is obtained by intersecting the plane $\lambda_1+\lambda_2+\lambda_3=1$ with the positive octant in the space of all triples $(\lambda_1,\lambda_2,\lambda_3)$. The three vertices are the three pure states and the edges are rank 2 diagonal states. The central point $M = \mathrm{diag}(\frac{1}{3},\frac{1}{3},\frac{1}{3})$ is the maximally mixed state. Along lines MX, MY, and MZ two eigenvalues are equal. We introduce polar coordinates on this simplex with r being the distance from M and the polar angle θ is measured anticlockwise from the line MX. The coordinates for a general state are given in Eq. (A9). The arc $\alpha\beta$ includes all diagonal states ρ (up to a permutation of the diagonal entries) at a constant radius r with a fixed $\mathrm{Tr}\,\rho^2 = \frac{1}{3} + r^2$ [cf Eq. (A11) *et seq.*]

In Fig. 2 the lines MX, MY, and MZ divide the circle into six symmetrical parts corresponding to the six possible permutations of the eigenvalues. Since we will be interested only in the (unordered) set of values of the λ's it suffices to consider just one of these regions. Thus, without loss of generality we may take

$$0 \leqslant \theta \leqslant \theta_{max} \leqslant \frac{\pi}{3}.$$

For some values of constant $\mathrm{Tr}\,\rho^2$ or r, the angle θ will have a maximum value θ_{max} smaller than $\pi/3$, e.g., for the r shown in Fig. 2 we take only the arc $\alpha\beta$. At the point β one of the eigenvalues has become zero.

From Eq. (A11) we see that the constraint $\mathrm{Tr}\,\rho^2 = \text{const}$ corresponds to r being constant. Using Eq. (A9) we get

$$\mathrm{Tr}\,\rho^3 = \sum \lambda_i^3 = \frac{1}{9} + r^2 + \frac{1}{\sqrt{6}} r^3 \cos 3\theta.$$

Since θ lies in the range $[0,\pi/3]$ we see that $\mathrm{Tr}\,\rho^3$ is a monotonically decreasing function of θ.

Next we show that S is also monotonically decreasing with θ. To study the variation of entropy

$$S = -\sum \lambda_i \log \lambda_i$$

with θ when $\mathrm{Tr}\,\rho^2$ is held constant, note that $\Sigma\lambda_i=1$ so we get

$$\sum \frac{\partial\lambda_i}{\partial\theta}=0 \tag{A12}$$

and so

$$\frac{\partial S}{\partial\theta}=-\sum_i \frac{\partial\lambda_i}{\partial\theta}\log\lambda_i .$$

Since $0\leqslant\theta\leqslant\pi/3$, we have from Eq. (A9)

$$\frac{\partial\lambda_1}{\partial\theta},\frac{\partial\lambda_2}{\partial\theta}\leqslant 0 \qquad \frac{\partial\lambda_3}{\partial\theta}\geqslant 0.$$

Hence, using Eq. (A12) we get

$$\frac{\partial S}{\partial\theta}=\left|\frac{\partial\lambda_1}{\partial\theta}\right|\log\frac{\lambda_1}{\lambda_3}+\left|\frac{\partial\lambda_2}{\partial\theta}\right|\log\frac{\lambda_2}{\lambda_3}.$$

Using the inequality $\log x\leqslant x-1$ we get directly

$$\frac{\partial S}{\partial\theta}\leqslant 0$$

with equality possible only for $\theta=0$ or $\pi/3$.

To summarize, any given ρ corresponds to a unique point (r,θ) with $0\leqslant\theta\leqslant\pi/3$. If $\mathrm{Tr}\,\rho^2=\frac{1}{3}+r^2$ is held constant then both $\mathrm{Tr}\,\rho^3$ and S are monotonically decreasing functions of θ. Hence, S is a monotonically increasing function of $\mathrm{Tr}\,\rho^3$, which completes the proof of Lemma A2. □

Proof of Lemma A3: We have

$$\mathrm{Tr}\,\rho^3=\mathrm{Tr}\,\mathbf{G}^3=\sum_{ijk} G_{ij}G_{jk}G_{ki}. \tag{A13}$$

Since ρ is invariant under choices of phase representatives we may assume without loss of generality that **G** has the form given in Eq. (9). Then in Eq. (A13) the ξ dependence arises only through three terms (via cyclic permutation of the subscripts) of the form $G_{23}G_{31}G_{12}=p_1p_2p_3\Upsilon_{123}$ and the three corresponding complex conjugate terms $G_{32}G_{21}G_{13}=p_1p_2p_3\overline{\Upsilon}_{123}$. This gives Eq. (A8).□

[1] A. Peres, Phys. Lett. A **128**, 19 (1988).
[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).
[3] B. Schumacher, Phys. Rev. A **51**, 2738 (1995).
[4] R. Jozsa and B. Schumacher, J. Mod. Opt. **41**, 2343 (1994).
[5] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, and W. Wootters, Phys. Rev. A **54**, 1869 (1996).
[6] H. Barnum, Ch. Fuchs, R. Jozsa, and B. Schumacher, Phys. Rev. A **54**, 4707 (1996).
[7] L. Hughston, R. Jozsa, and W. Wootters, Phys. Lett. A **183**, 14 (1993).
[8] A. Wehrl, Rev. Mod. Phys. **50**, 221 (1978).
[9] R. Horn and C. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, England, 1985).
[10] A. Uhlmann, Commun. Math. Phys. **54**, 21 (1977).
[11] G. Lindblad, Commun. Math. Phys. **40**, 147 (1975).
[12] H. Barnum, C. Caves, Ch. Fuchs, R. Jozsa, and B. Schumacher (unpublished).
[13] M. Horodecki, Phys. Rev. A **57**, 3364 (1998).
[14] A. Uhlmann, Rep. Math. Phys. **9**, 273 (1976).
[15] R. Jozsa, J. Mod. Opt. **41**, 2315 (1994).
[16] M. Horodecki, e-print quant-ph/9905058.
[17] N. Gisin and S. Popescu, e-print quant-ph/9901072.
[18] C. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976), Chap. 4.
[19] Peter Shor (private communication).