Entanglement measures and purification procedures

V. Vedral and M. B. Plenio
Optics Section, Blackett Laboratory, Imperial College London, London SW7 2BZ, England

(Received 17 July 1997; revised manuscript received 19 August 1997)

We improve previously proposed conditions each measure of entanglement has to satisfy. We present a class of entanglement measures that satisfy these conditions and show that the quantum relative entropy and Bures metric generate two measures of this class. We calculate the measures of entanglement for a number of mixed two spin-1/2 systems using the quantum relative entropy, and provide an efficient numerical method to obtain the measures of entanglement in this case. In addition, we prove a number of properties of our entanglement measure that have important physical implications. We briefly explain the statistical basis of our measure of entanglement in the case of the quantum relative entropy. We then argue that our entanglement measure determines an upper bound to the number of singlets that can be obtained by any purification procedure. [S1050-2947(98)03202-8]

PACS number(s): 03.67.−a, 03.65.Bz

I. INTRODUCTION

It was thought until recently that Bell's inequalities provided a good criterion for separating quantum correlations (entanglement) from classical ones in a given quantum state. While it is true that a violation of Bell's inequalities is a signature of quantum correlations (nonlocality), not all entangled states violate Bell's inequalities [1]. So, in order to completely separate quantum from classical correlations a new criterion was needed. This also initiated the search into the related question of the amount of entanglement contained in a given quantum state. There are a number of ''good'' measures of the amount of entanglement for two quantum systems in a pure state (see [2] for an extensive presentation). A ''good'' measure of entanglement for mixed states is, however, very hard to find. In an important work Bennett *et al.* [3] have recently proposed three measures of entanglement (we will discuss the entanglement of formation and distillation in more detail later in this paper). Their measures are based on concrete physical ideas and are intuitively easy to understand. They investigated many properties of these measures and calculated the entanglement of formation for a number of states. More recently, Hill and Wootters have proposed a closed form for the entanglement of formation for two spin-1/2 particles [4]. Uhlmann's recent work implies that the entanglement of formation can also be calculated numerically in an efficient way for those cases that are not analytically known [5].

We have recently shown how to construct a whole class of measures of entanglement [6,7], and also imposed conditions that any candidate for such a measure has to satisfy [6]. In short, we consider the disentangled states that form a convex subset of the set of all quantum states. Entanglement is then defined as a distance (not necessarily in the mathematical sense) from a given state to this subset of disentangled states (see Fig. 1). An attractive feature of our measure is that it is independent of the number of systems and their dimensionality, and is therefore completely general [6,7]. We present here two candidates for measuring distances on our set of states and prove that they satisfy improved conditions for a measure of entanglement (the third condition presented here is an improvement over the one given in [6]).

It should be noted that in much the same way we can calculate the amount of classical correlations in a state. One would then define another subset, namely, that of all product states that do not contain any classical correlations. Given a disentangled state one would then look for the closest uncorrelated state. The distance could be interpreted as a measure of classical correlations. In addition to many analytical results we also explain how to calculate efficiently using numerical methods our measure of entanglement of two spin-1/2 particles. We present a number of examples and prove several properties of our measure that have important physical consequences. To illuminate the physical meaning behind the above ideas we present a statistical view of our entanglement measure in the case of quantum relative entropy [7]. We then relate our measure to a purification procedure and use it to define a reversible purification. This reversible purification is then linked to the notion of entanglement through the idea of distinguishing two classes of quantum states. We also argue that the measure of entanglement generated by the quantum relative entropy that we propose gives

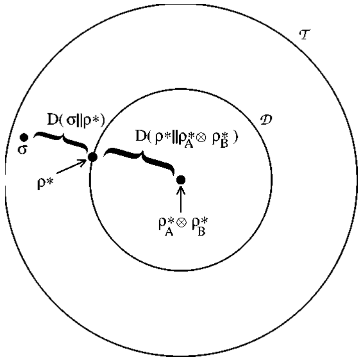
$\mathcal{T}$
$\mathcal{D}$
σ
$D(\sigma\|\rho^*)$
$D(\rho^*\|\rho_A^*\otimes\rho_B^*)$
ρ^*
$\rho_A^*\otimes\rho_B^*$

FIG. 1. The set of all density matrices $\mathcal{T}$ is represented by the outer circle. Its subset, a set of disentangled states $\mathcal{D}$, is represented by the inner circle. A state σ belongs to the entangled states, and ρ^* is the disentangled state that minimizes the distance $D(\sigma||\rho)$, thus representing the amount of quantum correlations in σ. State $\rho_A^*\otimes\rho_B^*$ is obtained by tracing ρ^* over A and B. $D(\rho^*||\rho_A^*\otimes\rho_B^*)$ represent the classical part of the correlations in the state σ.

an upper bound for the number of singlet states that can be distilled from a given state. We find that in general the distillable entanglement is smaller than the entanglement of creation. This result was independently proven by Rains for Bell diagonal states using completely different methods [8].

The rest of the paper is organized as follows. Section II introduces the basis of purification procedures, conditions for a measure of entanglement and our suggestion for a measure of entanglement. We also prove that the quantum relative entropy and the Bures metric satisfy the imposed conditions and can therefore be used as generators of measures of entanglement. We compute our measure explicitly for some examples. In Sec. III we introduce a simple numerical method to compute our measure of entanglement numerically and we apply it to the case of two spin-1/2 systems. We present a number of examples of entanglement computations using the quantum relative entropy. In Sec. IV we present a statistical basis for the quantum relative entropy as a measure of distinguishability between quantum states and hence of amount of entanglement. Based on this, in Sec. V we derive an upper bound to the efficiency (number of maximally entangled pairs distilled) of any purification procedure. We also show how to extend our measure to more than two subsystems.

II. THEORETICAL BACKGROUND

A. Purification procedures

There are three different ingredients involved in procedures aiming at distilling locally a subensemble of highly entangled states from an original ensemble of less entangled states.

(1) *Local general measurements* (LGM): these are performed by the two parties A and B separately and are described by two sets of operators satisfying the completeness relations $\Sigma_i A_i^\dagger A_i = \mathbb{1}$ and $\Sigma_j B_j^\dagger B_j = \mathbb{1}$. The joint action of the two is described by $\Sigma_{ij} A_i \otimes B_j = \Sigma_i A_i \otimes \Sigma_j B_j$, which is again a complete general measurement, and obviously local.

(2) *Classical communication* (CC): this means that the actions of A and B can be correlated. This can be described by a *complete measurement* on the whole space $A+B$ and is not necessarily decomposable into a sum of direct products of individual operators (as in LGM). If ρ_{AB} describes the initial state shared between A and B then the transformation involving ''LGM+CC'' would look like

$$\Phi(\rho_{AB}) = \sum_i A_i \otimes B_i \rho_{AB} A_i^\dagger \otimes B_i^\dagger, \tag{1}$$

where $\Sigma_i A_i^\dagger A_i B_i^\dagger B_i = \mathbb{1}$ i.e., the actions of A and B are ''correlated.''

(3) *Postselection* (PS) is performed on the *final ensemble* according to the above two procedures. Mathematically this amounts to the general measurement not being complete, i.e., we leave out some operations. The density matrix describing the newly obtained ensemble (the subensemble of the original one) has to be renormalized accordingly. Suppose that we kept only the pairs where we had an outcome corresponding to the operators A_i and B_j, then the state of the chosen subensemble would be

$$\rho_{AB} \to \frac{A_i \otimes B_i \rho_{AB} A_i^\dagger \otimes B_i^\dagger}{\mathrm{Tr}(A_i \otimes B_i \rho_{AB} A_i^\dagger \otimes B_i^\dagger)}, \tag{2}$$

where the denominator provides the necessary normalization.

A manipulation involving any of the above three elements or their combination we shall henceforth call a *purification procedure*. It should be noted that the three operations described above are local. This implies that the entanglement of the total ensemble cannot increase under these operations. However, classical correlations between the two subsystems *can* be increased, even for the whole ensemble, if we allow classical communication. A simple example confirms this. Suppose that the initial ensemble contains states $|0_A\rangle \otimes (|0_B\rangle + |1_B\rangle)/\sqrt{2}$. The correlations (measured by, e.g., von Neumann's mutual information [2,6]) between A and B are zero. Suppose that B performs measurement of his particles in the standard 0, 1 basis. If 1 is obtained, B communicates this to A who then ''rotates'' his qubit to the state $|1_A\rangle$. Otherwise they do nothing. The final state will therefore be

$$\rho = \tfrac{1}{2}(|0_A\rangle\langle 0_A| \otimes |0_B\rangle\langle 0_B| + |1_A\rangle\langle 1_A| \otimes |1_B\rangle\langle 1_B|), \tag{3}$$

where the correlations are now ln2 (i.e., nonzero). So, the classical content of correlations can be increased by performing local general measurements and classically communicating.

An important result was proved for pairs of spin-1/2 systems in [9]: all states that are not of the form $\rho_{AB} = \Sigma_i p_i \rho_A^i \otimes \rho_B^i$, where $\Sigma_i p_i = 1$ and $p_i \geqslant 0$ for all i, can be distilled to a subensemble of maximally entangled states using only operations 1, 2, and 3. (The states of the above form obviously remain of the same form under any purification procedure). The local nature of the above three operations implies that we define a disentangled state of two quantum systems A and B as a state from which by means of local operations no subensemble of entangled states can be distilled. It should be noted that these states are sometimes called separable in the existing literature. We also note that it is not proven in general that if the state is not of this form then it can be purified.

Definition 1. A state ρ_{AB} is disentangled iff

$$\rho_{AB} = \sum_i p_i \rho_A^i \otimes \rho_B^i, \tag{4}$$

where, as before, $\Sigma_i p_i = 1$ and $p_i \geqslant 0$ for all i. Otherwise it is said to be entangled. Note that all the states in the above expansion can be taken to be pure. This is because each ρ^i can be expanded in terms of its eigenvectors. So, in the above sum we can in addition require that $(\rho_A^i)^2 = \rho_A^i$ and $(\rho_B^i)^2 = \rho_B^i$ for all i. This fact will be used later in this section and will be formalized further in Sec. III.

B. Quantification of entanglement

In the previous section we have indicated that out of certain states it is possible to distill by means of LGM+CC+PS a subensemble of maximally entangled states (we call these states entangled). The question remains open about how much entanglement a certain state contains. Of course, this question is not entirely well defined unless we state what physical circumstances characterize the amount of entanglement. This suggests that there is no unique measure of entanglement. Before we define three different measures of entanglement we state three conditions that every measure of entanglement has to satisfy. The third condition represents a generalization of the corresponding one in [6].

(E1) $E(\sigma)=0$ iff σ is separable.

(E2) Local unitary operations leave $E(\sigma)$ invariant, i.e., $E(\sigma)=E(U_A \otimes U_B \sigma U_A^\dagger \otimes U_B^\dagger)$.

(E3) The expected entanglement cannot increase under LGM+CC+PS given by $\Sigma V_i^\dagger V_i = \mathbb{1}$, i.e.,

$$\sum \operatorname{tr}(\sigma_i) E(\sigma_i/\operatorname{tr}(\sigma_i)) \leqslant E(\sigma), \tag{5}$$

where $\sigma_i = V_i \sigma V_i^\dagger$.

Condition (E1) ensures that disentangled and only disentangled states have a zero value of entanglement. Condition (E2) ensures that a local change of basis has no effect on the amount of entanglement. Condition (E3) is intended to remove the possibility of increasing entanglement by performing local measurements aided by classical communication. It is an improvement over the condition (3) in [6], which required that $E(\Sigma_i V_i \sigma V_i^\dagger) \leqslant E(\sigma)$. This condition (E3) is physically more appropriate than that in [6] as it takes into account the fact that we have some knowledge of the final state. Namely, when we start with n systems all in the state σ we know exactly which $m_i = n \times \operatorname{tr}(\sigma_i)$ pairs will end up in the state σ_i after performing a purification procedure. Therefore we can separately access the entanglement in each of the possible subensembles described by σ_i. Clearly the total expected entanglement at the end should not exceed the original entanglement, which is stated in (E3). This, of course, does not exclude the possibility that we can select a subensemble whose entanglement per pair is higher than the original entanglement per pair. We emphasize that if we assume that $E(\sigma)$ is also convex (as it, indeed, is in the case of the quantum relative entropy presented later in the paper) then (E3) immediately implies that $E(\Sigma_i V_i \sigma V_i^\dagger) \leqslant E(\sigma)$. On the other hand, convexity of $E(\sigma)$ and $E(\Sigma_i V_i \sigma V_i^\dagger) \leqslant E(\sigma)$ do not imply (E3), which also provides a reason for requiring (E3) rather than the condition in [6]. We now introduce three different measures of entanglement that obey (E1)–(E3).

First we discuss the entanglement of creation [3]. Bennett *et al.* [3] define the entanglement of creation of a state ρ by

$$E_c(\rho) := \min \sum_i p_i S(\rho_A^i), \tag{6}$$

where $S(\rho_A) = -\operatorname{tr}\rho_A \ln \rho_A$ is the von Neumann entropy and the minimum is taken over all the possible realizations of the state, $\rho_{AB} = \Sigma_j p_j |\psi_j\rangle\langle\psi_j|$ with $\rho_A^i = \operatorname{tr}_B(|\psi_i\rangle\langle\psi_i|)$. The entanglement of creation satisfies all three conditions (E1)–(E3) [3]. The physical basis of this measure presents the number of singlets needed to be shared in order to create a given entangled state by local operations. We will discuss this in greater detail in Sec. IV. It should also be added that progress has been made recently in finding a closed form of the entanglement of creation [4].

Related to this measure is the entanglement of distillation [3]. It defines the amount of entanglement of a state σ as the proportion of singlets that can be distilled using a purification procedure (Bennett *et al.* distinguish one- and two-way communication which give rise to two different measures, but we will not go into that much detail; we assume the most general two-way communication). As such, it is dependent on the efficiency of a particular purification procedure and can be made more general only by introducing some sort of universal purification procedure or asking for the best state-dependent purification procedure. We investigate this in Sec. V. We now introduce our suggestion for a measure of an amount of entanglement. It is seen in Sec. V that this measure is intimately related to the entanglement of distillation by providing an upper bound for it.

If $\mathcal{D}$ is the set of all disentangled states, the measure of entanglement for a state σ is then defined as

$$E(\sigma) := \min_{\rho \in \mathcal{D}} \quad D(\sigma || \rho), \tag{7}$$

where D is any measure of *distance* (not necessarily a metric) between the two density matrices ρ and σ such that $E(\sigma)$ satisfies the above three conditions (E1)–(E3) (see Fig. 1).

Now the central question is what condition a candidate for $D(\sigma||\rho)$ has to satisfy in order for (E1)–(E3) to hold for the entanglement measure? We present here a set of sufficient conditions.

(F1) $D(\sigma||\rho) \geqslant 0$ with the equality saturated iff $\sigma = \rho$.

(F2) Unitary operations leave $D(\sigma||\rho)$ invariant, i.e., $D(\sigma||\rho) = D(U\sigma U^\dagger || U\rho U^\dagger)$.

(F3) $D(\operatorname{tr}_p \sigma || \operatorname{tr}_p \rho) \leqslant D(\sigma||\rho)$, where tr_p is a partial trace.

(F4) $\Sigma p_i D(\sigma_i/p_i || \rho_i/q_i) \leqslant \Sigma D(\sigma_i||\rho_i)$, where $p_i = \operatorname{tr}(\sigma_i)$, $q_i = \operatorname{tr}(\rho_i)$, and $\sigma_i = V_i \sigma V_i^\dagger$ and $\rho_i = V_i \rho V_i^\dagger$ (note that V_i's are not necessarily local).

(F5a) $D(\Sigma_i P_i \sigma P_i || \Sigma_i P_i \rho P_i) = \Sigma_i D(P_i \sigma P_i || P_i \rho P_i)$, where P_i is any set of orthogonal projectors such that $P_i P_j = \delta_{ij} P_i$.

(F5b) $D(\sigma \otimes P_\alpha || \rho \otimes P_\alpha) = D(\sigma||\rho)$ where P_α is any projector.

Conditions (F1) and (F2) ensure that (E1) and (E2) hold; (F2), (F3), (F4), and (F5) ensure that (E3) is satisfied. The argument for the former is trivial, while for the latter it is more lengthy and will be presented in the remainder of this section.

C. Proofs

We claim that (F2), (F3), (F4), and (F5) are sufficient for (E3) to be satisfied and hence need to prove that (F2)−(F5)⇒(E3). If (F2), (F3), and (F5b) hold, then we can prove the following statement.

Theorem 1. For any completely positive, trace preserving map Φ, given by $\Phi\sigma=\Sigma V_i\sigma V_i^\dagger$ and $\Sigma V_i^\dagger V_i=\mathbb{1}$, we have that $D(\Phi\sigma||\Phi\rho)\leqslant D(\sigma||\rho)$.[1]

Proof. It is well known that a complete measurement can *always* be represented as a unitary operation+partial tracing on an extended Hilbert Space $\mathcal{H}\otimes\mathcal{H}_n$, where $\dim\mathcal{H}_n=n$ [10–12]. Let $\{|i\rangle\}$ be an orthonormal basis in $\mathcal{H}_n$ and $|\alpha\rangle$ be a unit vector. So we define

$$W=\sum_i V_i\otimes|i\rangle\langle\alpha|. \tag{8}$$

Then, $W^\dagger W=\mathbb{1}\otimes P_\alpha$, where $P_\alpha=|\alpha\rangle\langle\alpha|$, and there is a unitary operator U in $\mathcal{H}\otimes\mathcal{H}_n$ such that $W=U(\mathbb{1}\otimes P_\alpha)$ [10]. Consequently,

$$U(A\otimes P_\alpha)U^\dagger=\sum_{ij} V_iAV_j^\dagger\otimes|i\rangle\langle j|, \tag{9}$$

so that

$$\mathrm{tr}_2\{U(A\otimes P_\alpha)U^\dagger\}=\sum_i V_iAV_i^\dagger . \tag{10}$$

Now using (F3), then (F2), and finally (F5b) we find the following:

$$D(\mathrm{tr}_2\{U(\sigma\otimes P_\alpha)U^\dagger\}||\mathrm{tr}_2\{U(\rho\otimes P_\alpha)U^\dagger\}) \tag{11}$$

$$\leqslant D(U(\sigma\otimes P_\alpha)U^\dagger||U(\rho\otimes P_\alpha)U^\dagger) \tag{12}$$

$$=D(\sigma\otimes P_\alpha||\rho\otimes P_\alpha) \tag{13}$$

$$=D(\sigma||\rho). \tag{14}$$

This proves Theorem 1.

Corollary. Since for a complete set of orthonormal projectors P, $\Sigma_iP_i\sigma P_i$ is a complete positive trace preserving map, then

$$\sum_i D(P_i\sigma P_i||P_i\rho P_i)\leqslant D(\sigma||\rho). \tag{15}$$

[The sum can be taken outside as (F5a) requires that $D(\Sigma_iP_i\sigma P_i||\Sigma_iP_i\rho P_i)=\Sigma_iD(P_i\sigma P_i||P_i\rho P_i)$.] Now from (F2), (F3), (F5b), and Eq. (15) we have the following.

Theorem 2. If $\sigma_i=V_i\sigma V_i^\dagger$ then $\Sigma D(\sigma_i||\rho_i)\leqslant D(\sigma||\rho)$.

Proof. Equations (8) and (9) are introduced as in the previous proof. From Eq. (9) we have that

$$\mathrm{tr}_2\{\mathbb{1}\otimes P_iU(A\otimes P_\alpha)U^\dagger\mathbb{1}\otimes P_i\}=V_iAV_i^\dagger , \tag{16}$$

where $P_i=|i\rangle\langle i|$. Now, from (F3), the corollary, and (F5b) it follows that

$$\sum_i D(\mathrm{tr}_2\{\mathbb{1}\otimes P_iU(\sigma\otimes P_\alpha)U^\dagger\mathbb{1}\otimes P_i\}|| \times\mathrm{tr}_2\{\mathbb{1}\otimes P_iU(\rho\otimes P_\alpha)U^\dagger\mathbb{1}\otimes P_i\}) \tag{17}$$

$$\leqslant\sum_i D(\mathbb{1}\otimes P_iU(\sigma\otimes P_\alpha)U^\dagger\mathbb{1}\otimes P_i||\mathbb{1} \otimes P_iU(\rho\otimes P_\alpha)U^\dagger\mathbb{1}\otimes P_i) \tag{18}$$

$$\leqslant D(U(\sigma\otimes P_\alpha)U^\dagger||U(\rho\otimes P_\alpha)U^\dagger) \tag{19}$$

$$=D(\sigma\otimes P_\alpha||\rho\otimes P_\alpha) \tag{20}$$

$$=D(\sigma||\rho). \tag{21}$$

This proves Theorem 2.

From Theorem 2 and (F4) we have

$$\sum p_iD\left(\frac{\sigma_i}{p_i}\middle|\middle|\frac{\rho_i}{q_i}\right)\leqslant D(\sigma||\rho). \tag{22}$$

Now let $E(\sigma)=D(\sigma||\rho^*)$, i.e., let the minimum of $D(\sigma||\rho)$ over all $\rho\in\mathcal{D}$ be attained at ρ^*. Then from Eq. (22),

$$E(\sigma):=D(\sigma||\rho^*)\geqslant\sum p_iD\left(\frac{\sigma_i}{p_i}\middle|\middle|\frac{V_i^\dagger\rho^*V_i}{q_i}\right)$$

$$\geqslant\sum p_iE(\sigma_i/p_i) \tag{23}$$

and (E3) is satisfied. Note that in all the proofs for $D(\sigma||\rho)$ we never use the fact that the completely positive, trace preserving map Φ is local. This is only used in the last inequality of Eq. (23) where LGM (+CC+PS) maps disentangled states onto disentangled states. This ensures that ρ_i^* is disentangled and therefore $D(\sigma_i/p_i||\rho_i^*/q_i)\geqslant E(\sigma_i/p_i)$. So, the need for local Φ arises only in Eq. (23); otherwise all the other proofs hold for a general Φ. Note also that one can prove, by the same methods, a slightly more general condition:

(E3*) The expected entanglement of the initial state $\sigma^n=\sigma_1\otimes\cdots\otimes\sigma_n$ cannot increase under LGM+CC+PS given by $\Sigma V_i^\dagger V_i=\mathbb{1}$, i.e.,

$$E(\sigma^n)\equiv E(\sigma_1\otimes\cdots\otimes\sigma_n)$$

$$\geqslant\sum \mathrm{tr}(V_i\sigma^nV_i^\dagger)E(V_i\sigma^nV_i^\dagger/\mathrm{tr}(V_i\sigma^nV_i^\dagger)). \tag{24}$$

However, in the following we will not make use of this generalization.

D. Two realizations of $D(\sigma,\rho)$

In this section we show that (F1)–(F5) hold for the quantum relative entropy and for the Bures metric, which as we have seen immediately renders them generators of a good measure of entanglement.

[1] We frequently interchange the Φ and $\Sigma V^\dagger V$ notations for one another throughout this section.

1. Quantum relative entropy

We first prove (F1)–(F5) for the quantum relative entropy, i.e., when $D(\sigma||\rho)=S(\sigma||\rho):=\text{Tr}\{\sigma(\ln\sigma-\ln\rho)\}$. (Note that the quantum relative entropy is not a true metric, as it is not symmetric and does not satisfy the triangle inequality. In the next section the reasons for this will become clear. For further properties of the quantum relative entropy see [13–15].) Properties (F1) and (F2) are satisfied [16]. (F3) follows from the strong subadditivity property of the von Neumann Entropy [11,16,17]. Since $\Sigma S(\sigma_i||\rho_i)=\Sigma p_i S(\sigma_i/p_i||\rho_i/q_i)+\Sigma p_i \ln p_i/q_i$ and $\Sigma p_i \ln p_i/q_i \geqslant 0$ (see [18] for proof) (F4) is also satisfied. Property (F5) can be proved to hold by inspection [11]. Now, a question arises as to why the entanglement is not defined as $E(\sigma)=\min_{\rho\in\mathcal{D}}S(\rho||\sigma)$. Since the quantum relative entropy is asymmetric this gives a different result from the original definition. However, the major problem with this convention is that for all pure states this measure is infinite. Although this does have a sound statistical interpretation (see the next section) it is hard to relate it to any physically reasonable scheme (e.g., a purification procedure) and, in addition, it fails to distinguish between different entangled pure states. This is the prime reason for excluding this convention from any further considerations. The measure of entanglement generated by the quantum relative entropy will hereafter be referred to as the relative entropy of entanglement.

Properties of the relative entropy of entanglement. For pure, maximally entangled states we showed that the relative entropy of entanglement reduces to the von Neumann reduced entropy [6]. We also conjectured [6] that for a general pure state this would be true. Now we present a proof of this conjecture. In short, our proof goes as follows: we already have a guess as to what the minimum for a pure state σ should be—say it is a disentangled state ρ^*. Then we show that the gradient $(d/dx)S(\sigma||(1-x)\rho^*+x\rho)$ for any $\rho\in\mathcal{D}$ is non-negative. However, if ρ^* was not a minimum the above gradient would be strictly negative, which is a contradiction. Now we present a more formal proof [19] that applies to arbitrary dimensions of the two subsystems. An alternative proof that also applies to arbitrary dimensions will be given in Sec. III. In the Appendix we present a third proof that is restricted to two spin-1/2 systems but that can be generalized to arbitrary dimensions.

Theorem 3. For pure states $\sigma=\Sigma_{n_1n_2}\sqrt{p_{n_1}p_{n_2}}|\phi_{n_1}\psi_{n_1}\rangle\langle\phi_{n_2}\psi_{n_2}|$ the relative entropy of entanglement is equal to the von Neumann reduced entropy, i.e., $E(\sigma)=-\Sigma_n p_n \ln p_n$.

Proof. For $a>0$, $\ln a=\int_0^\infty[(at-1)/(a+t)]\,dt/(1+t^2)$, and thus, for any positive operator A, $\ln A=\int_0^\infty[(At-1)/(A+t)\,dt/(1+t^2)$. Let $f(x,\rho)=S(\sigma||(1-x)\rho^*+x\rho)$. Then

$$\begin{aligned}\frac{\partial f}{\partial x}(0,\rho)&=-\lim_{x\to 0}\text{tr}\left\{\frac{\sigma(\ln[(1-x)\rho^*+x\rho]-\ln\rho^*\}}{x}\right\}\\ &=\text{tr}\left(\sigma\int_0^\infty(\rho^*+t)^{-1}(\rho^*-\rho)(\rho^*+t)^{-1}dt\right)\\ &=1-\int_0^\infty\text{tr}[\sigma(\rho^*+t)^{-1}\rho(\rho^*+t)^{-1}]dt\\ &=1-\int_0^\infty\text{tr}[(\rho^*+t)^{-1}\sigma(\rho^*+t)^{-1}\rho]dt.\end{aligned}\tag{25}$$

Take $\rho^*=\Sigma_n p_n|\phi_n\psi_n\rangle\langle\phi_n\psi_n|$ (this is our guess for the minimum). Then

$$\begin{aligned}(\rho^*+t)^{-1}\sigma(\rho^*+t)^{-1}=&\sum_{n_1,n_2,n_3,n_4}(p_{n_1}+t)^{-1}|\phi_{n_1}\psi_{n_1}\rangle\langle\phi_{n_1}\psi_{n_1}|\sqrt{p_{n_2}p_{n_3}}|\phi_{n_2}\psi_{n_2}\rangle\langle\phi_{n_3}\psi_{n_3}|(p_{n_4}+t)^{-1}|\phi_{n_4}\psi_{n_4}\rangle\\ &\times\langle\phi_{n_4}\psi_{n_4}|\\ =&\sum_{n,n'}(p_n+t)^{-1}\sqrt{p_np_{n'}}(p_{n'}+t)^{-1}|\phi_n\psi_n\rangle\langle\phi_{n'}\psi_{n'}|.\end{aligned}\tag{26}$$

Set $g(p,q)=\int_0^\infty(p+t)^{-1}\sqrt{pq}(q+t)^{-1}dt$. Then it follows that $g(p,p)=1$ and, for $p<q$,

$$g(p,q)=\sqrt{pq}\int_0^\infty\left(\frac{1}{p+t}-\frac{1}{q+t}\right)\frac{1}{q-p}dt\tag{27}$$

$$=\frac{\sqrt{pq}}{q-p}\ln\frac{q}{p}.\tag{28}$$

Lemma: $0\leqslant g(p,q)\leqslant 1$ for all $p,q\in[0,1]$.

Proof. We know that $g(p,q)=\sqrt{pq}\int_0^\infty(p+t)^{-1}(q+t)^{-1}dt$. But,

$$(p+t)(q+t)=pq+t(p+q)+t^2\geqslant pq+2t\sqrt{pq}+t^2=(\sqrt{pq}+t)^2,\tag{29}$$

and so

$$g(p,q)\leqslant\sqrt{pq}\int_0^\infty(\sqrt{pq}+t)^{-2}dt=1. \tag{30}$$

Let $\rho=|\alpha\rangle\langle\alpha|\otimes|\beta\rangle\langle\beta|$ where $|\alpha\rangle=\Sigma_n a_n|\phi_n\rangle$ and $\beta=\Sigma_n b_n\psi_n$ are normalized vectors. Then

$$\begin{aligned}\frac{\partial f}{\partial x}(0,\rho)-1&=-\operatorname{tr}\left(\int_0^\infty(\rho^*+t)^{-1}\sigma(\rho^*+t)^{-1}dt\rho\right)\\&=-\operatorname{tr}\left(\sum_{n_1,n_2,n_3,n_4,n_5,n_6}g(p_{n_1},p_{n_2})|\phi_{n_1}\psi_{n_1}\langle\phi_{n_2}\psi_{n_2}|a_{n_3}b_{n_4}\overline{a}_{n_5}\overline{b}_{n_6}|\phi_{n_3}\psi_{n_4}\rangle\langle\phi_{n_5}\psi_{n_6}|\right)\\&=-\sum_{n_1,n_2}g(p_{n_1},p_{n_2})a_{n_2}b_{n_2}\overline{a}_{n_1}\overline{b}_{n_1}\end{aligned} \tag{31}$$

and

$$\left|\frac{\partial f}{\partial x}(0,\rho)-1\right|\leqslant\sum_{n_1,n_2}|a_{n_1}||b_{n_1}||a_{n_2}||b_{n_2}|=\left(\sum_n|a_n||b_n|\right)^2\leqslant\sum_n|a_n|^2\sum_n|b_n|^2=1. \tag{32}$$

Thus it follows that $(\partial f/\partial x)(0,|\alpha\beta\rangle\langle\alpha\beta|)\geqslant 0$.

But any $\rho\in\mathcal{D}$ can be written in the form $\rho=\Sigma_i r_i|\alpha^i\beta^i\rangle\langle\alpha^i\beta^i|$ and so $(\partial f/\partial x)(0,\rho)=\Sigma_i r_i(\partial f/\partial x)(0,|\alpha^i\beta^i\rangle\langle\alpha^i\beta^i|)\geqslant 0$.

Proposition: Let $\Phi\in H$ have Schmidt decomposition [20]

$$|\Phi\rangle=\sum_n\sqrt{p_n}|\varphi_n\psi_n\rangle \tag{33}$$

and set $\sigma=|\Phi\rangle\langle\Phi|$. Then $E(\sigma)=-\Sigma_n p_n\ln p_n$.

Proof. $S(\sigma||\rho^*)=-\Sigma_n p_n\ln p_n$ so it is sufficient to prove that $S(\sigma||\rho)\geqslant S(\sigma||\rho^*)$ for all $\rho\in D$. Suppose that $S(\sigma||\rho)<S(\sigma||\rho^*)$ for some $\rho\in\mathcal{D}$. Then, for $0<x\leqslant 1$,

$$\begin{aligned}f(x,\rho)&=S(\sigma||(1-x)\rho^*+x\rho)\leqslant(1-x)S(\sigma||\rho^*)+xS(\sigma||\rho)\\&=(1-x)f(0,\rho)+xf(1,\rho).\end{aligned} \tag{34}$$

This implies

$$\frac{f(x,\rho)-f(0,\rho)}{x}\leqslant f(1,\rho)-f(0,\rho)<0. \tag{35}$$

This is impossible since $(\partial f/\partial x)(0,\rho)=\lim_{x\to 0}[f(x,\rho)-f(0,\rho]/x\geqslant 0$. This therefore proves the above proposition.

Therefore we have shown that for arbitrary dimensions of the subsystems the entropy of entanglement reduces to the entropy of entanglement for pure states. This is, in fact, a very desirable property, as the entropy of entanglement is known to be a good measure of entanglement for pure states. In fact one might want to elevate Theorem 3 to a condition for any good measure of entanglement, i.e.:

(E4): For pure states the measure of entanglement reduces to the entropy of entanglement, i.e.,

$$E(\sigma)=-\operatorname{tr}\{\sigma_A\ln\sigma_A\}, \tag{36}$$

with $\sigma_A=\operatorname{tr}_B\{\sigma\}$ being the reduced density operator of one subsystem of the entangled pair.

However, in Sec. II D 2 we will see that measures that do not satisfy (E4) can nevertheless contain useful information. We will discuss this point later in this paper.

We would like to point out another property of the relative entropy of entanglement that helps us find the amount of entanglement. It gives us a method to construct from a density operator σ with known entanglement a new density operator σ' with known entanglement.

Theorem 4. If ρ^* minimizes $S(\sigma||\rho^*)$ over $\rho\in\mathcal{D}$ then ρ^* is also a minimum for any state of the form $\sigma_x=(1-x)\sigma+x\rho^*$.

Proof. Consider

$$\begin{aligned}S(\sigma_x||\rho)-S(\sigma_x||\rho^*)=&\operatorname{tr}\{\sigma_x\ln\rho^*-\sigma_x\ln\rho\}=-x\operatorname{tr}(\sigma\ln\rho)\\&-(1-x)\operatorname{tr}(\rho^*\ln\rho)+x\operatorname{tr}(\sigma\ln\rho^*)\\&+(1-x)\operatorname{tr}(\rho^*\ln\rho^*)\\&=x\{S(\sigma||\rho)-S(\sigma||\rho^*)\}+(1-x)\\&\times S(\rho^*||\rho)\geqslant 0.\end{aligned} \tag{37}$$

This is true for any ρ. Thus ρ^* is indeed a minimum of σ_x. For completeness we now prove here that $E(\sigma)$ is convex:

Theorem 5. $E(x_1\sigma_1+x_2\sigma_2)\leqslant x_1E(\sigma_1)+x_2E(\sigma_2)$, where $x_1+x_2=1$.

Proof. This property follows from the convexity of the quantum relative entropy in both arguments [15]

$$S(x_1\sigma_1+x_2\sigma_2||x_1\rho_1+x_2\rho_2)\leqslant x_1S(\sigma_1||\rho_1)+x_2S(\sigma_2||\rho_2). \tag{38}$$

Now,

$$\begin{aligned}E(x_1\sigma_1+x_2\sigma_2)&\leqslant S(x_1\sigma_1+x_2\sigma_2||x_1\rho_1^*+x_2\rho_2^*)\\&\leqslant x_1S(\sigma_1||\rho_1^*)+x_2S(\sigma_2||\rho_2^*)\\&=x_1E(\sigma_1)+x_2E(\sigma_2),\end{aligned} \tag{39}$$

which completes our proof of convexity. This is physically a very satisfying property of an entanglement measure. It says that when we mix two states having a certain amount of entanglement we cannot get a more entangled state, i.e., succinctly stated, ''mixing does not increase entanglement.'' This is what is indeed expected from a measure of entanglement to predict.

As a last property we state that the entanglement of creation E_c is never smaller than the relative entropy of entanglement E. We will show later that this property has the important implication that the amount of entanglement that we have to invest to create a given quantum state is usually larger than the entanglement that you can recover using quantum state distillation methods.

Theorem 6. $E_c(\sigma)\geqslant E(\sigma)=\min_{\rho\in\mathcal{D}}S(\sigma||\rho)$.

Proof. Given a state σ then by definition of the entanglement of creation there is a convex decomposition $\sigma=\Sigma p_i\sigma_i$ with pure states σ_i such that

$$E_c(\sigma)=\sum p_iE_c(\sigma_i). \tag{40}$$

As the entanglement of creation coincides with our entanglement for pure states and as our entanglement is convex it follows that

$$E_c(\sigma)=\sum p_iE_c(\sigma_i)=\sum p_iE(\sigma_i)\geqslant E\left(\sum p_i\sigma_i\right)=E(\sigma), \tag{41}$$

and the proof is completed.

The physical explanation of the above result lies in the fact that a certain amount of additional knowledge is involved in the entanglement of formation, which gives it a higher value to the relative entropy of entanglement. This will be explained in full detail in Sec. V. We add that the relative entropy of entanglement $E(\sigma)$ can be calculated easily for Bell diagonal states [6]. Comparing the result to those for the entanglement of creation [3] one finds that, in fact, strict inequality holds. In general, we have unfortunately found no ''closed form'' for the relative entropy of entanglement and a computer search is necessary to find the minimum ρ^*, for each given σ. However, we can numerically find the amount of entanglement for two spin-1/2 subsystems very efficiently using general methods independent of the dimensionality and the number of subsystems involved which are described in the next section.

2. Bures metric

Another distance measure that leads to a measure of entanglement that satisfies the conditions (E1)–(E3) is induced by the Bures metric. However, it will turn out that it does not satisfy condition (E4) and is therefore a less useful measure. In fact some people would say it is not a measure of entanglement at all, however, we believe that this very much depends on the questions one asks.

We now prove (F1)–(F5) for the Bures metric, i.e., when $D(\sigma||\rho)=D_B(\sigma||\rho):=2-2\sqrt{F(\sigma,\rho)}$, where $F(\sigma,\rho):=[\mathrm{tr}\{\sqrt{\rho}\sigma\sqrt{\rho}\}^{1/2}]^2$ is the so-called fidelity (or Uhlmann's transition probability). Property (F1) follows from the fact that the Bures metric is a true metric and (F2) is obvious. (F3) is a consequence of the fact that D_B does not increase under a complete positive trace preserving map [21]. We can also easily check that $p_iq_iF(\sigma_i/p_i,\rho_i/q_i)=F(\sigma_i,\rho_i)$, from where (F4) immediately follows as $q_i\in[0,1]$. (F5) is seen to be true by inspection. As conditions (F1)–(F5) are satisfied, it immediately follows that conditions (E1)–(E3) are satisfied too.

In the following we present some properties of the Bures measure of entanglement $E_B(\sigma)$. First we show that for pure states we do not recover the entropy of entanglement.

Theorem 7: For a pure state $|\psi\rangle=\alpha|00\rangle+\beta|11\rangle$ one has

$$E_B(|\psi\rangle\langle\psi|)=4\alpha^2(1-\alpha^2). \tag{42}$$

Proof. To prove Theorem 7 we have to show that the closest disentangled state to $\sigma=|\psi\rangle\langle\psi|$ under the Bures metric is given by $\rho^*=\alpha^2|00\rangle\langle00|+\beta^2|11\rangle\langle11|$. To this end we consider a slight variation around ρ^* of the form $\rho_\lambda=(1-\lambda)\rho^*+\lambda\rho$ where $\rho\in\mathcal{D}$. Now we need to calculate

$$\frac{d}{d\lambda}D_B(\sigma||\rho_\lambda)|_{\lambda=0}=\frac{d}{d\lambda}\mathrm{tr}\{\sqrt{\sqrt{\sigma}\rho_\lambda\sqrt{\sigma}}\}\leqslant0. \tag{43}$$

Using the fact that $\sqrt{\sigma}=\sigma$ as σ is pure we obtain

$$\frac{d}{d\lambda}D_B(\sigma||\rho_\lambda)|_{\lambda=0}=\frac{d}{d\lambda}\sqrt{\alpha^4+\beta^4+\lambda(\langle\psi|\rho|\psi\rangle-1)}|_{\lambda=0}\leqslant0. \tag{44}$$

Using the closest state ρ^* one then obtains Eq. (42). To obtain the entanglement of an arbitrary pure state one first has to calculate the Schmidt decomposition [20] and then by local unitary transformation transform the state to the form $|\psi\rangle=\alpha|00\rangle+\beta|11\rangle$. As local unitary transformations do not change the entanglement, we have therefore shown that the Bures measure of entanglement does not reduce to the entropy of entanglement for pure states. The proof presented here can be generalized to many-dimensional systems but we do not state this generalization.

In fact, it is now easy to see the following.

Corollary. The Bures measure of entanglement for pure states is smaller than the entropy of entanglement, i.e., for any pure state σ,

$$E_B(\sigma)\leqslant-\{\sigma_A\ln\sigma_A\}. \tag{45}$$

Proof. One can see quickly that for $\alpha\in[0,1]$

$$4\alpha^2(1-\alpha^2)\leqslant-\alpha^2\ln\alpha^2-(1-\alpha^2)\ln(1-\alpha^2) \tag{46}$$

from which the corollary follows.

As the Bures measure of entanglement does not satisfy condition (E4), i.e., does not reduce to the entropy of entanglement for pure states, one might argue that it does not provide a sensible measure of entanglement. However, it should be noted that the Bures metric immediately gives an upper bound on the following very special purification procedure. Assume that Alice and Bob are given EPR pairs, but one pair at a time. Then they are allowed to perform any local operations they like, and then decide whether we keep the pair or discard it. Then, they are given the next EPR pair. The question is, how many pure singlet states they can possibly distill out of such a purification procedure. The answer is immediately obvious from condition (E3). The best that Alice and Bob can do is to have one subensemble with pure singlets and all other subensembles with disentangled states. Then the probability to obtain a singlet is simply given by the Bures measure of entanglement for the initial ensemble. As this is smaller than the entropy of entanglement we have found the nontrivial, though not very surprising, result that this restricted purification procedure is strictly less efficient than entanglement concentration described in [27].

3. Other candidates

A reasonable candidate to generate a measure of entanglement is the Hilbert-Schmidt metric. Here we have that $D(A||B)=||A-B||^2:=\mathrm{tr}(A-B)^2$. (F1) follows from the fact that $||A-B||$ is a true metric, and (F2) is obvious. (F3) and (F4) remain to be shown to hold. We also believe that there are numerous other nontrivial choices for $D(A||B)$ (by nontrivial we mean that the choice is not a simple scale transformation of the above candidates). Each of those generators would arise from a different physical procedure involving measurements conducted on σ and ρ^*. None of the choices could be said to be more important than any other *a priori*, but the significance of each generator would have to be seen through physical assumptions. To illustrate this point further, let us take an extreme example. Define

$$D(A||B)=\begin{cases}1, & A\neq B,\\ 0, & A=B.\end{cases}$$

If entanglement is calculated using this distance, then

$$E(\sigma)=\begin{cases}1, & \sigma\notin\mathcal{D},\\ 0, & \sigma\in\mathcal{D}.\end{cases}$$

This measure therefore tells us if a given state σ is entangled, i.e., when $E(\sigma)=1$, or disentangled, i.e., when $E(\sigma)=0$. We can call it the ‘‘indicator measure’’ of entanglement. It should be noted that this measure trivially satisfies conditions (E1)–(E3). This shows that there are numerous different choices for $D(A||B)$ and each is related to different physical considerations. We explain the statistical basis of the relative entropy of entanglement in Sec. IV. The relative entropy of entanglement is then seen to be linked very naturally to the notion of a purification procedure. First, however, we present an efficient numerical method to obtain entanglement for arbitrary particles.

III. NUMERICS FOR TWO SPIN-1/2 PARTICLES

In order to understand how our program for calculating the amount of entanglement works, we first need to introduce one basic definition and one important result from convex analysis [22]. From this point onwards we concentrate on the quantum relative entropy as a measure of entanglement although most of the considerations are of a more general nature.

Definition 2. The convex hull $[\mathrm{co}(A)]$ of a set A is the set of all points that can be expressed as (finite) convex combinations of points in A. In other words, $x\in\mathrm{co}(A)$ if and only if x has an expression of the form $x=\Sigma_{k=1}^{K}p_k a_k$, where K is finite, $\Sigma_{k=1}^{K}p_k=1$, and, for $k=1,\dots,K$, $p_k>0$ and $a_k\in A$.

We immediately see that the set of disentangled states $\mathcal{D}$ is a convex hull of its pure states. This means that any state in $\mathcal{D}$ can be written as a convex combination of the form $\Sigma p_n|\phi_n\psi_n\rangle\langle\phi_n\psi_n|$. However, there is now a problem in the numerical determination of the measure of entanglement. We have to perform a search over the set of disentangled states in order to find that disentangled state that is closest to the state σ of which we want to know the entanglement. But how can we parametrize the disentangled states? We know that the disentangled states are of the form given by Definition 1. However, there the number of states in the convex combination is not limited. Therefore one could think that we have to look over all convex combinations with one state, then two states, then 1000 states, and so forth. The next theorem, however, shows that one can put an upper limit to the number of states that are required in the convex combination. This is crucial for our minimization problem as it shows that we do not have to have an infinite number of parameters to search over.

Caratheodory's theorem. Let $A\subset\mathbf{R}^N$. Then any $x\in\mathrm{co}(A)$ has an expression of the form $x=\Sigma_{n=1}^{N+1}p_n a_n$ where $\Sigma_{n=1}^{N+1}p_n=1$, and, for $n=1,\dots,N+1$, $p_n\geqslant 0$ and $a_n\in A$.

A direct consequence of Caratheodory's theorem is that any state in $\mathcal{D}$ can be decomposed into a sum of at most $[\dim(H_1)\times\dim(H_2)]^2$ products of pure states. So, for two spin-1/2 particles there are at most 16 terms in the expansion of any disentangled state. In addition, each pure state can be described using two real numbers, so that there are altogether at most $15+16\times 4=79$ real parameters needed to completely characterize a disentangled state in this case.

A random search over the 79 real parameters would still be very inefficient. However, we can now make use of another useful property of the relative entropy, which is the fact that it is convex. This means that we have to minimize a convex function over the convex set of disentangled states. It can easily be shown that any local minimum must also be a global minimum. Therefore we can perform a gradient search for the minimum (basically we calculate the gradient and then perform a step in the opposite direction and repeat this procedure until we hit the minimum). As soon as we have found any relative minimum we can stop the search, since this is also a global minimum. To make the gradient search efficient we have to choose a suitable parametrization. The parametrization that we use has the advantage that it also provides us with another proof of Theorem 3, which states that for pure states the relative entropy of entanglement reduces to the von Neumann reduced entropy. We first explain the parametrization and then state the alternative proof for Theorem 3. The following results can easily be extended to two subsystems of arbitrary dimensions but for clarity we restrict ourselves to two spin-1/2 systems.

Our aim is to find the amount of entanglement of a state σ of two spin-1/2 states, i.e., we have to minimize $\mathrm{tr}\{\sigma\ln\sigma-\sigma\ln\rho\}$ for all $\rho\in\mathcal{D}$. From Caratheodory's theorem we know that we only need convex combinations of at most 16 pure states ρ_k^i to represent $\rho\in\mathcal{D}$, i.e.,

$$\rho=\sum_{i=1}^{16}p_i^2\rho_1^i\otimes\rho_2^i. \tag{47}$$

(Notice that we use p_i^2 instead of p_i for convenience, so that here we require that $\Sigma_{i=1}^{16}p_i^2=1$.) The parametrization we chose is now given by

$$p_i=\sin\phi_{i-1}\prod_{j=i}^{15}\cos\phi_j \quad \text{with} \quad \phi_0=\frac{\pi}{2} \tag{48}$$

and

$$\begin{aligned}\rho_k^i&=|\psi_k^i\rangle\langle\psi_k^i|,\\ |\psi_1^i\rangle&=\cos\alpha_i|0\rangle+\sin\alpha_i e^{i\eta_i}|1\rangle,\\ |\psi_2^i\rangle&=\cos\beta_i|0\rangle+\sin\beta_i e^{i\mu_i}|1\rangle.\end{aligned} \tag{49}$$

All angles $\alpha_i,\beta_i,\phi_i,\eta_i,\mu_i$ can have arbitrary values, but due to the periodicity only the interval $[0,2\pi]$ is really relevant. Numerically this has the advantage that our parameter space has no edges at which problems might occur. The program for the search of the minimum is now quite straightforward. The idea is that given σ we start from a random ρ, i.e., we generate 79 random numbers. Then we compute $S(\sigma||\rho)$, as well as small variations of the 79 parameters of ρ, to obtain the approximate gradient of $S(\sigma||\rho)$ at the point ρ. We then move opposite to the gradient to obtain the next ρ. We continue this until we reach the minimum. As explained before, a convex function over a convex set can only have a global minimum, so that the minimum value we end up with is the one and only. The method outlined above immediately generalizes to two subsystems of arbitrary dimension, however, the number of parameters rises quickly to large values, which slows down the program considerably.

Before we state some numerical results we now indicate an alternative proof of Theorem 3 using Caratheodory's theorem and the parametrization given in Eqs. (47)–(49). For this proof we use the fact that we can represent the logarithm of an operator ρ by

$$\ln\rho=\frac{1}{2\pi i}\oint \ln z\frac{1}{z\mathbb{1}-\rho}, \tag{50}$$

where the path of integration encloses all eigenvalues of ρ. We can now take the partial derivative of $\ln\rho$ with respect to a parameter ϕ on which ρ might depend.

$$\frac{\partial\ln\rho}{\partial\phi}=\frac{1}{2\pi i}\oint \ln z\frac{1}{z\mathbb{1}-\rho}\frac{\partial\rho}{\partial\phi}\frac{1}{z\mathbb{1}-\rho}. \tag{51}$$

Now, we have a given pure state

$$\begin{aligned}\sigma=\alpha^2|00\rangle\langle00|+\alpha\sqrt{1-\alpha^2}(|00\rangle\langle11|+|11\rangle\langle00|)\\ +(1-\alpha^2)|11\rangle\langle11|.\end{aligned} \tag{52}$$

The suspected closest approximation to σ within the disentangled states is given by

$$\rho_{\min}=\alpha^2|00\rangle\langle00|+(1-\alpha^2)|11\rangle\langle11|. \tag{53}$$

If we want to represent $\rho_{\min}$ using the parametrization given in Eqs. (47)–(49) then we find for these parameters $\cos^2\phi_1=\alpha^2$; $\alpha_2=\beta_2=\pi/2$ and zero for all other parameters. Using Eq. (51) one can now calculate all the partial derivatives of the relative entropy around the point $\rho_{\min}$. It is easy, but rather lengthy, to check that these derivatives vanish and that therefore $\rho_{\min}$ is a relative minimum. This concludes the proof as a relative minimum of a convex function on a convex set is also a global minimum.

After this additional proof of Theorem 3 we now state some results that we have obtained or confirmed with the program that implements the gradient search. We present four nontrivial states σ for which we can find the closest disentangled state ρ that minimize the quantum relative entropy thereby giving the relative entropy of entanglement. Using the same ideas as for the proof of Theorem 3 in Eq. (50)–(53) one can then prove that these are indeed the closest disentangled states.

Example 1:

$$\sigma_1=\lambda|\Phi^+\rangle\langle\Phi^+|+(1-\lambda)|01\rangle\langle01|, \tag{54}$$

$$\begin{aligned}\rho_1=\frac{\lambda}{2}\left(1-\frac{\lambda}{2}\right)|00\rangle\langle00|+\frac{\lambda}{2}\left(1-\frac{\lambda}{2}\right)\{|00\rangle\langle11|+\text{ H.c.}\}\\ +\left(1-\frac{\lambda}{2}\right)^2|01\rangle\langle01|+\frac{\lambda^2}{4}|10\rangle\langle10|\\ +\frac{\lambda}{2}\left(1-\frac{\lambda}{2}\right)|11\rangle\langle11|,\end{aligned} \tag{55}$$

$$E(\sigma_1)=(\lambda-2)\ln\left(1-\frac{\lambda}{2}\right)+(1-\lambda)\ln(1-\lambda). \tag{56}$$

Here $|\Phi^+\rangle$ is one of the four Bell states defined by

$$|\Phi^\pm\rangle=\frac{1}{\sqrt{2}}(|00\rangle\pm|11\rangle), \tag{57}$$

$$|\Psi^\pm\rangle=\frac{1}{\sqrt{2}}(|01\rangle\pm|10\rangle). \tag{58}$$

Example 2:

$$\sigma_2=\lambda|\Phi^+\rangle\langle\Phi^+|+(1-\lambda)|00\rangle\langle00|, \tag{59}$$

$$\rho_2=\left(1-\frac{\lambda}{2}\right)|00\rangle\langle00|+\frac{\lambda}{2}|11\rangle\langle11|, \tag{60}$$

$$\begin{aligned}E(\sigma_2)=s_+\ln s_++s_-\ln s_--\left(1-\frac{\lambda}{2}\right)\ln\left(1-\frac{\lambda}{2}\right)\\ -\left(1-\frac{\lambda}{2}\right)\ln\left(1-\frac{\lambda}{2}\right),\end{aligned} \tag{61}$$

where

$$s_\pm=\frac{1\pm\sqrt{1-2\lambda(1-\lambda/2)}}{2} \tag{62}$$

are the eigenvalues of σ_2. One could argue that in the above two cases the following reasoning can be applied: $\sigma_{1(2)}$ is a mixture of a maximally entangled state (for which the amount of entanglement is given by ln2) and a completely disentangled state ($E=0$). Thus one would expect a total amount of entanglement of λ ln2. It is curious that this reasoning does not work for either of the two states, since, in fact, $E(\sigma_{1(2)})\leqslant\lambda\ \ln 2$. Now, we show how to use Theorem 4 to generate more states and their minima. For pure states $\sigma^2=\sigma$ we know the minimum ρ. Now, the state that is a convex sum of σ and ρ should also have the same minimum ρ. So we have the following.

Example 3:

$$\sigma_3=A|00\rangle\langle 00|+B|00\rangle\langle 11|+B^*|11\rangle\langle 00|+(1-A)|11\rangle\langle 11|, \tag{63}$$

$$\rho_3=A|00\rangle\langle 00|+(1-A)|11\rangle\langle 11|, \tag{64}$$

$$E(\sigma_3)=e_+\ln e_+ + e_-\ln e_- - A\ln A-(1-A)\ln(1-A), \tag{65}$$

where

$$e_{\pm}=\frac{1\pm\sqrt{1-4A(1-A)-|B|^2}}{2}. \tag{66}$$

Using Theorem 4, the amount of entanglement can be found for a number of other spin-1/2 states. Our program can also help us infer the entanglement of some other nontrivial states as the last example shows.

Example 4:

$$\sigma_4=A|00\rangle\langle 00|+B|00\rangle\langle 11|+B^*|11\rangle\langle 00|+(1-2A)|01\rangle \times\langle 01|+A|11\rangle\langle 11|, \tag{67}$$

$$\rho_4=C|00\rangle\langle 00|+D|00\rangle\langle 11|+D^*|11\rangle\langle 00|+E|01\rangle\langle 01| \tag{68}$$

$$+(1-2C-E)|10\rangle\langle 10|+C|11\rangle\langle 11|, \tag{69}$$

where

$$E=\frac{(1-2A)(1-A)^2}{(1-A)^2-B^2}, \tag{70}$$

$$C=1-A-E, \tag{71}$$

$$D=\sqrt{E(1-E-2C)}=\frac{(1-2A)(1-A)}{(1-A)^2-B^2}B. \tag{72}$$

It is now easy to compute the amount of entanglement from the above information.

In addition to the above described methods there is a simple way of obtaining a lower bound for the amount of entanglement for any two spin-1/2 system. Suppose that we have a certain state σ. We first find the *maximally* entangled state $|\psi\rangle$ such that the fidelity $F=\langle\psi|\sigma|\psi\rangle$ is maximized. Then we apply local unitary transformations to σ, which transform $|\psi\rangle$ into the singlet state (this is, of course, always possible). Now, we apply local random rotations [3] to both particles. These will transform σ into a Werner state, where the singlet state will have a weight F (since it is invariant under rotations) and all the other three Bell states will have equal weights of $(1-F)/3$ (since they are randomized). Since these operations are local they cannot increase the amount of entanglement, and we have that for any σ

$$E(\sigma)\geqslant E(W_F)=F\ \ln F+(1-F)\ln(1-F)+\ln 2, \tag{73}$$

where W_F is the above-described Werner state (the relative entropy of entanglement for a general Bell diagonal state is calculated in [6]).

We note that this efficient computer search provides an alternative criterion for deciding when a given state σ of two spin-1/2 systems is disentangled, i.e., of the form given in Eq. (4). The already existing criterion is the one given by Peres and Horodecki *et al.* (see second and third references in [1]), which states that a state is disentangled iff its partial trace over either of the subsystems is a non-negative operator. This criterion is only valid for two spin-1/2, or one spin-1/2 and one spin-1 systems. In the absence of a more general analytical criterion our computational method provides a way of deciding this question. In addition we would like to point out that the program is also able to provide us with the convex decomposition of a disentangled state ρ.

At the end of this section we mention *additivity* as an important property desired from a measure of entanglement, i.e., we would like to have

$$E(\sigma_{12}\otimes\sigma_{34})=E(\sigma_{12})+E(\sigma_{34}), \tag{74}$$

where systems $1+2$ and systems $3+4$ are entangled separately from each other. The exact definition of the left-hand side is

$$E(\sigma_{12}\otimes\sigma_{34})=\min_{p_i,\rho_{13},\rho_{24}} S\left(\sigma_{12}\otimes\sigma_{34}\middle\|\sum_i p_i\rho_{13}^i\otimes\rho_{24}^i\right). \tag{75}$$

Why this form? One would originally assume that $\sigma_{12}\otimes\sigma_{34}$ should be minimized by the states of the form $(\Sigma_i p_i\rho_1^i\otimes\rho_2^i)\otimes(\Sigma_j p_j\rho_3^j\otimes\rho_4^j)$. However, Alice, who holds systems 1 and 3, and Bob, who holds systems 2 and 4, can also perform arbitrary unitary operation on their subsystems (i.e., locally). This obviously leads to the creation of entanglement between 1 and 3 and between 2 and 4 and hence the form given in Eq. (75). Additivity is, of course, already true for the pure states, as can be seen from the proof above, when our measure reduces to the von Neumann entropy. For more general cases we were unable to provide an analytical proof, so that the above additivity property remains a conjecture. However, for two spin-1/2 systems, our program did not find any counterexample. It should be noted that it is easy to see that we have

$$E(\sigma_{12}\otimes\sigma_{34})\leqslant E(\sigma_{12})+E(\sigma_{34}). \tag{76}$$

In the following we will assume that Eq. (74) holds and use it in Sec. V to derive certain limits to the efficiency of purification procedures.

IV. STATISTICAL BASIS OF ENTANGLEMENT MEASURE

Let us see how we can interpret our entanglement measure in the light of experiments, i.e., statistically. This was presented in [7] in greater detail. Here we present a summary which is sufficient to understand the following section. Our interpretation relies on the result concerning the asymptotics of the quantum relative entropy first proved in [14], and here presented under the name of quantum Sanov's theorem. We first show how the notion of relative entropy arises in classical information theory as a measure of distinguishability of two probability distributions. We then generalize this idea to the quantum case, i.e., to distinguishing between two quantum states (for a discussion of distinguishability of pure quantum states see e.g., [23]). We will see that this naturally leads to the notion of the quantum relative entropy. It is then straightforward to extend this concept to explain the relative entropy of entanglement. Suppose we would like to check if a given coin is ''fair,'' i.e., if it generates a ''head-tail'' distribution of $f=(1/2,1/2)$. If the coin is biased then it will produce some other distribution, say $uf=(1/3,2/3)$. So, our question of the coin fairness boils down to how well we can differentiate between two given probability distributions given a finite, n, number of experiments to perform on one of the two distributions. In the case of a coin we would toss it n times and record the number of 0's and 1's. From simple statistics we know that if the coin is fair than the number of 0's $N(0)$ will be roughly $n/2-\sqrt{n}\leqslant N(0)\leqslant n/2+\sqrt{n}$, for large n and the same for the number of 1's. So if our experimentally determined values do not fall within the above limits the coin is not fair. We can look at this from another point of view; namely, what is the probability that a fair coin will be mistaken for an unfair one with the distribution of $(1/3,2/3)$ given n trials on the fair coin? For large n the answer is [7,18]

$$p(\text{fair}\rightarrow\text{ unfair})=e^{-nS_{\text{cl}}(uf||f)},\tag{77}$$

where $S_{\text{cl}}(uf||f)=1/3\ \ln1/3+2/3\ \ln2/3-1/3\ \ln1/2-2/3\ \ln1/2$ is the classical relative entropy for the two distributions. So,

$$p(\text{fair}\rightarrow\text{ unfair})=3^n2^{-(5/3)n},\tag{78}$$

which tends exponentially to zero with $n\rightarrow\infty$. In fact we see that already after ~20 trials the probability of mistaking the two distributions is vanishingly small, $\leqslant10^{-10}$.

This result is true, in general, for any two distributions. Asymptotically the probability of not distinguishing the distributions $P(x)$ and $Q(x)$ after n trials is $e^{-nS_{\text{cl}}(P(x)||Q(x))}$, where

$$S_{\text{cl}}(P(x)||Q(x))=\sum_i p_i\ln p_i-p_i\ln q_i\tag{79}$$

(this statement is sometimes called Sanov's theorem [18]). To generalize this to quantum theory, we need a means of generating probability distributions from two quantum states σ and ρ. This is accomplished by introducing a general measurement $E_i^\dagger\Sigma_iE_i=\mathbb{1}$. So, the probabilities are given by

$$\begin{aligned}p_i&=\text{tr}(E_i^\dagger E_i\rho),\\ q_i&=\text{tr}(E_i^\dagger E_i\sigma).\end{aligned}\tag{80}$$

Now, we can use Eq. (79) to distinguish between σ and ρ. The above is not the most general measurement that we can make, however. In general we have N copies of σ and ρ in the state

$$\sigma^N=\underbrace{\sigma\otimes\sigma...\otimes\sigma}_{\text{total of N terms}}\tag{81}$$

$$\rho^N=\underbrace{\rho\otimes\rho...\otimes\rho}_{\text{total of N terms}}\tag{82}$$

We may now apply a POVM $\Sigma_iA_i=\mathbb{1}$ acting on σ^N and ρ^N. Consequently, we define a new type of relative entropy

$$S_N(\sigma||\rho):=\sup_{A's}\left\{\frac{1}{N}\sum_i\ \text{tr}A_i\sigma^N\ln\ \text{tr}A_i\sigma^N-\text{tr}A_i\sigma^N\ln\ \text{tr}A_i\rho^N\right\}.\tag{83}$$

Now it can be shown that [15]

$$S(\sigma||\rho)\geqslant S_N,\tag{84}$$

where, as before,

$$S(\sigma||\rho):=\text{tr}(\sigma\ \ln\sigma-\sigma\ \ln\rho)\tag{85}$$

is the quantum relative entropy [6,7,11,12,15,16] (for the summary of the properties of the quantum relative entropy see [13]). Equality is achieved in Eq. (84) iff σ and ρ commute [24]. However, for any σ and ρ it is true that [14]

$$S(\sigma||\rho)=\lim_{N\rightarrow\infty}S_N.$$

In fact, this limit can be achieved by projective measurements, which are independent of σ [25]. It is known that if Eq. (79) is maximized over all general measurements E, the upper bound is given by the quantum relative entropy (see, e.g., [15]). In quantum theory we therefore state a law analogous to Sanov's theorem (see also [7]),

Theorem 8 (or quantum Sanov's theorem). The probability of **not** distinguishing two quantum states (i.e., density matrices) σ and ρ after n measurements is

$$p(\rho\rightarrow\sigma)=e^{-nS(\sigma||\rho)}.\tag{86}$$

In fact, as explained before, this bound is reached asymptotically [14], and the measurements achieving this are global projectors independent of the state σ [25]. We note that the quantum Sanov theorem was presented by Donald in [26] as a definition justified by properties uniquely characterizing the quantity $e^{-nS(\sigma||\rho)}$. The underlying intuition in the above measurement approach and Donald's approach are basically the same. Now the interpretation of the relative entropy of entanglement becomes immediately transparent [7]. The probability of mistaking an entangled state σ for a closest, disentangled state, ρ, is $e^{-n\min_{\rho\in\mathcal{D}}S(\sigma,\rho)}=e^{-nE(\sigma)}$. If the amount of entanglement of σ is greater, then it takes fewer measurements to distinguish it from a disentangled state (or, fixing n, there is a smaller probability of confusing it with some disentangled state). Let us give an example. Consider a state $(|00\rangle+|11\rangle)/\sqrt{2}$, known to be a maximally entangled state. The closest to it is the disentangled state $(|00\rangle\langle 00|+|11\rangle\langle 11|)/2$ [6]. To distinguish these states it is enough to perform projections onto $(|00\rangle+|11\rangle)/\sqrt{2}$. If the state that we are measuring is the above mixture, then the sequence of results (1 for a successful projection, and 0 for an unsuccessful projection) will contain on average an equal number of 0's and 1's. For this to be mistaken for the above pure state the sequence has to contain all n 1's. The probability for that is 2^{-n}, which also comes from using Eq. (86). If, on the other hand, we performed projections onto the pure state itself, we would then never confuse it with a mixture, and from Eq. (86) the probability is seen to be $e^{-\infty}=0$. We next apply this simple idea to obtaining an upper bound to the efficiency of any purification procedure.

V. THERMODYNAMICS OF ENTANGLEMENT: PURIFICATION PROCEDURES

There are two ways to produce an upper bound to the efficiency of any purification procedure. Using condition (E3) and the fact that the relative entropy of entanglement is additive, we can immediately derive this bound. However, this bound can be derived in an entirely different way. In this section we now abandon conditions (E1)–(E3) and use only methods of the previous section to put an upper bound to the efficiency of purification procedures. In particular, we show that the entanglement of creation is in general larger than the entanglement of distillation. This is in contrast with the situation for pure states where both quantities coincide. The quantum relative entropy is seen to play a distinctive role here, and is singled out as a ''good'' generator of a measure of entanglement from among other suggested candidates.

A. Distinguishability and purification procedures

In the previous section we presented a statistical basis to the relative entropy of entanglement by considering distinguishability of two (or more) quantum states encapsulated in the form of the quantum Sanov theorem. We now use this quantum Sanov theorem to put an upper bound on the amount of entanglement that can be distilled using any purification procedure. This line of reasoning follows from the fact that any purification scheme can be viewed as a measurement to distinguish entangled and disentangled quantum states. Suppose that there exists a purification procedure with the following property: Initially there are n copies of the state σ. If σ is entangled, then the end product is $0<m\leqslant n$ singlets and $n-m$ states in $\rho\in\mathcal{D}$. Otherwise, the final state does not contain any entanglement, i.e., $m=0$ (in fact, there is nothing special about singlets: the final state can be any other known, maximally entangled state because these can be converted into singlets by applying local unitary operations).

Note that we can allow the complete knowledge of the state σ. We also allow that purification procedures differ for different states σ. Perhaps there is a ''universal'' purification procedure independent of the initial state. However, in reality, this property is hard to fulfill [9]. At present the best that can be done is to purify a certain class of entangled states (see, e.g., [27–29]). The above is therefore an idealization that might never be achieved. Now, by calculating the upper bound on the efficiency of a procedure described above we present an absolute bound for any particular procedure. We ask: ''What is the largest number of singlets that can be produced (distilled) from n pairs in state σ''? Suppose that we produce m pairs. We now project them *nonlocally* onto the singlet state. The procedure will yield positive outcomes (1) with certainty so long as the state we measure indeed is a singlet. Suppose that after performing singlet projections onto all m particles we get a string of m 1's. From this we conclude that the final state is a singlet (and therefore the initial state σ was entangled). However, we could have made a mistake. But with what probability? The answer is as follows: the largest probability of making a wrong inference is $2^{-m}=e^{-m\ln 2}$ (if the state that we were measuring had an overlap with a singlet state of 1/2). On the other hand, if we were measuring σ from the very beginning (without performing the purification first), then the probability (i.e., the lower bound) of the wrong inference would be $e^{-nE(\sigma)}$. But, purification procedure might waste some information (i.e., it is just a particular way of distinguishing entangled from disentangled states, not necessarily the best one), so that the following has to hold

$$e^{-nE(\sigma)}\leqslant e^{-m\ln 2}, \tag{87}$$

which implies that

$$nE(\sigma)\geqslant m, \tag{88}$$

i.e., we cannot obtain more entanglement than is originally present. This, of course, is also directly guaranteed by our condition (E3). The above, however, was a deliberate exercise in deriving the same result from a different perspective, abandoning conditions (E1)–(E3). Therefore the measure of entanglement given in Eq. (7), when $D(\sigma||\rho)=S(\sigma||\rho)$, can be used to provide an upper bound on the efficiency of any purification procedure. For Bell diagonal states, Rains [8] found an upper bound on distillable entanglement using completely different methods. It turns out that the bound that he obtains in this case is identical to the one provided by the relative entropy of entanglement.

Actually, in the above considerations we implicitly assumed that the entanglement of n pairs, equivalently prepared in the state σ, is the same as $n\times E(\sigma)$. We already indicated that this is a conjecture with a strongly supported basis in the case of the quantum relative entropy. Based on the upper bound considerations we can introduce the following definition.

Definition 3. A purification procedure given by a local complete positive trace preserving map $\sigma\rightarrow\Sigma V_i\sigma V_i^\dagger$ is defined to be *ideal* in terms of efficiency iff

$$\sum \operatorname{tr}(\sigma_i)E(\sigma_i/\operatorname{tr}(\sigma_i))=E(\sigma), \tag{89}$$

where, as usual, $\sigma_i=V_i\sigma V_i^\dagger$ and $p_i=\operatorname{tr}(V_i\sigma V_i^\dagger)$ (i.e., a the ideal purification is the one where (E3) is an equality rather than an inequality). Notice an apparent formal analogy between a purification procedure and the Carnot cycle in thermodynamics. The Carnot cycle is the most efficient cycle in thermodynamics (i.e., it yields the greatest ''useful work to heat'' ratio), since it is reversible (i.e., it conserves the thermodynamical entropy). We would now like to claim that the

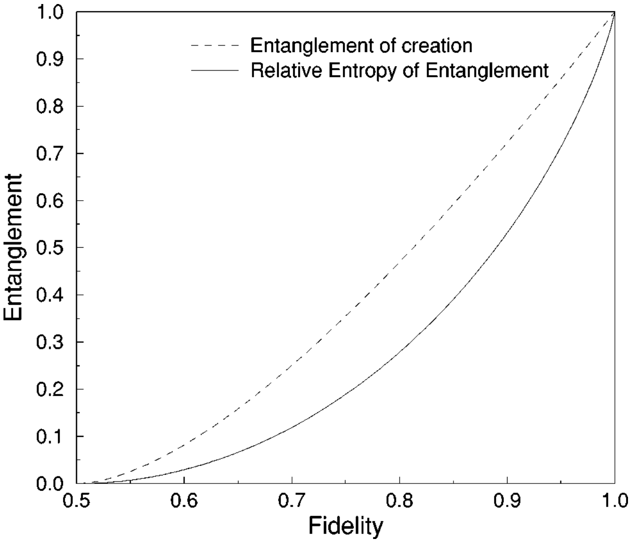

FIG. 2. Comparison of the entanglement of creation and the relative entropy of entanglement for the Werner states (these are Bell diagonal states of the form $W=\text{diag}(F,(1-F)/3,(1-F)/3,(1-F)/3)$. One clearly sees that the entanglement of creation is strictly larger than the relative entropy of entanglement for $0<F<1$.

ideal purification procedure is the most efficient purification procedure (i.e., it yields the greatest number of singlets for a given input state), since it is reversible (i.e., it conserves entanglement, measured by the minimum of the quantum relative entropy over all disentangled states). Unfortunately this analogy between the Carnot cycle and purification procedures is not exact (it is only strictly true for the pure states). This is seen when we compare the entanglement of creation with the relative entropy of entanglement. In Theorem 6 we have, in fact, shown that the entanglement of creation is never smaller than the relative entropy of Entanglement. As an example one can consider Bell diagonal states for which we can exactly calculate both the entanglement of creation [6] and the relative entropy of entanglement [3]. It turns out that the entanglement of creation is always strictly larger than the relative entropy of entanglement except for the limiting cases of maximally entangled Bell states or of disentangled Bell diagonal states (see Fig. 2 for Werner states). This result leads to the following.

Implication. In general, the amount of entanglement that was initially invested in creation of σ cannot all be recovered (''distilled'') by local purification procedures.

Therefore, the ideal purification procedure, though most efficient, is nevertheless irreversible, and some of the invested entanglement is lost in the purification process itself. The solution to this irreversibility lies in the loss of certain information as can easily be seen from the following analysis. Suppose we start with an ensemble of N singlets and we want to locally create any mixed state σ. Now σ can always be written as a mixture of pure states $\Psi_1,\Psi_2,\ldots$ with the corresponding probabilities $p_1,p_2,\ldots$. We now use Bennett *et al.*'s (de)purification procedure [27] for pure states (whose efficiency is governed by the von Neumann entropy). We convert the first $p_1\times N$ singlets into the state Ψ_1, the second $p_2\times N$ singlets into the state Ψ_2, and so on. In this way, the whole ensemble is in the state σ. But, we have additional information: we know exactly that the first $p_1\times N$ pairs are in the state Ψ_1, the second $p_2\times N$ states are in the state Ψ_2, and so on. This is not the same as being given an initial ensemble of identically prepared pairs in the state sigma without any additional information. In this, second, case we do not have the additional information of knowing exactly the state of each of the pairs. This is why the purification without this knowledge is less efficient, and hence one expects that the relative entropy of entanglement is smaller than the entanglement of formation.

An open question remains as to whether we can use some other generator, such as the Bures metric, to give an even more stringent bound on the amount of distillable entanglement.

B. More than two subsystems

We see that the above treatment does not refer to the number (or indeed dimensionality) of the entangled systems. This is a desired property as it makes our measure of entanglement universal. However, in order to perform minimization in Eq. (7) we need to be able to define what we mean by a disentangled state of say N particles. As pointed out in [7] we believe that this can be done inductively. Namely, for two quantum systems, A_1 and A_2, we define a disentangled state as one that can be written as a convex sum of disentangled states of A_1 and A_2 as follows [6,7]:

$$\rho_{12}=\sum_i p_i\rho_i^{A_1}\otimes\rho_i^{A_2}, \tag{90}$$

where $\Sigma_i p_i=1$ and the p's are all positive. Now, for N entangled systems $A_1,A_2,\ldots,A_N$, the disentangled state is

$$\rho_{12\cdots N}=\sum_{\text{perm}\{i_1 i_2\cdots i_N\}} r_{i_1 i_2\cdots i_N}\rho^{A_{i_1}A_{i_2}\cdots A_{i_n}}\otimes\rho^{A_{i_{n+1}}A_{i_{n+2}}\cdots A_{i_N}}, \tag{91}$$

where $\Sigma_{\text{perm}\{i_1 i_2\cdots i_N\}} r_{i_1 i_2\cdots i_N}=1$, all r's are positive and where $\Sigma_{\text{perm}\{i_1 i_2\cdots i_N\}}$ is a sum over all possible permutations of the set of indices $\{1,2,\ldots,N\}$. To clarify this let us see how this looks for 4 systems:

$$\begin{aligned}\rho_{1234}=\sum_i\ & p_i\rho_i^{A_1A_2A_3}\otimes\rho_i^{A_4}+q_i\rho_i^{A_1A_2A_4}\otimes\rho_i^{A_3}+r_i\rho_i^{A_1A_3A_4}\\ &\otimes\rho_i^{A_2}+s_i\rho_i^{A_2A_3A_4}\otimes\rho_i^{A_1}+t_i\rho_i^{A_1A_2}\otimes\rho_i^{A_3A_4}+u_i\rho_i^{A_1A_3}\\ &\otimes\rho_i^{A_2A_4}+v_i\rho_i^{A_1A_4}\otimes\rho_i^{A_2A_3}\end{aligned} \tag{92}$$

where, as usual, all the probabilities $p_i,q_i,\ldots,v_i$ are positive and add up to unity. The above two equations, at least in principle, define the disentangled states for any number of entangled systems. Note that this form describes a different situation from the one given in Eq. (75), which refers to a number of pairs shared by Alice and Bob only. The above definition of a disentangled state is justified by extending the idea that local actions cannot increase the entanglement between two quantum systems [3,6,7]. In the case of N particles we have N parties (Alice, Bob, Charlie,..., Wayne) all acting locally on their systems. The general action that also includes communications can be written as [7]

$$\rho\rightarrow\sum_{i_1,i_2,\ldots,I_N} A_{i_1}\otimes B_{i_2}\otimes\cdots\otimes W_{i_N}\rho A_{i_1}^{\dagger}\otimes B_{i_2}^{\dagger}\otimes\cdots\otimes W_{i_N}^{\dagger} \tag{93}$$

and it can be easily seen that this action does not alter the form of a disentangled state in Eqs. (91) and (92). In fact, Eq. (91) is the most general state invariant *in form* under the transformation given by Eq. (93). This can be suggested as a definition of a disentangled state for $N\geqslant 3$, i.e., it is the most general state invariant in form under local POVM and classical communications. Of course, an alternative to defining a disentangled state would be

$$\rho_{12\cdots N}=\sum_i r_i\rho_i^{A_1}\otimes\rho_i^{A_2}\cdots\otimes\rho_i^{A_N}, \tag{94}$$

which means that we do not allow any entanglement in any subset of the N states. This would be a disentangled state based on some local hidden variable model. Again we repeat that the particular choice of a form of disentangled states will depend on the physical background in our model and there is no absolute sense in which we can resolve this dichotomy. It should be stressed that for two particles this free choice does not exist as both pictures coincide.

VI. CONCLUSIONS

We can look at the entanglement from two different perspectives. One insists that local actions cannot increase entanglement and do not change it if they are unitary. The other one looks at the way we can distinguish an entangled state from a disentangled one. In particular, the following question is asked: what is the probability of confusing an entangled state with a disentangled one after performing a certain number of measurements? These two, at first sight different approaches, lead to the same measure of entanglement. This results in the fact that a purification procedure can be regarded as a protocol of distinguishing an entangled state from a disentangled set of states. From this premise we derived the upper bound on the efficiency of any purification procedure. It turns out that distillable entanglement is in general smaller than the entanglement of creation. Our entanglement measure is independent of the number of systems and their dimensionality. This suggests applying it to more than two entangled systems in order to understand multiparticle entanglement. We have shown how to compute entanglement efficiently for two spin-1/2 subsystems using computational methods. However, a closed form for the expression of this entanglement measure is desirable. However, a closed form for the entanglement of formation has been proposed for two spin-1/2 particles in [4]. An interesting problem is to specify all the states that have the same amount of entanglement. We know that all the states that are equivalent up to a local unitary transformation have the same amount of entanglement [by definition (E2)]. However, there are states with the same amount of entanglement but that are not equivalent up to a local unitary transformation (for example, one state is pure and the other one is mixed). A question for further research is whether they are linked by a local complete measurement. Our work in addition suggest a question of finding a general local map that preserves the entanglement of a given entangled state.

ACKNOWLEDGMENTS

We thank A. Ekert, C. A. Fuchs and P. L. Knight for useful discussions and comments on the subject of this paper. We are very grateful to M. J. Donald for communicating an alternative proof of Theorem 3 to us and for very useful comments on the manuscript. This work was supported by the European Community, the United Kingdom Engineering and Physical Sciences Research Council, by the Alexander von Humboldt Foundation, and by the Knight Trust. Part of this work was completed during the 1997 Elsag-Bailey–I.S.I. Foundation research meeting on quantum computation.

APPENDIX A: ANOTHER PROOF FOR THE PURE STATE ENTANGLEMENT

In the following we present a third proof for the value of the relative entropy of entanglement for pure states. As in the second proof we use the representation of the logarithm of a density operator in terms of a complex integral as in Eqs. (50) and (51). We would like to know the value of the relative entropy of entanglement for a pure state $\sigma=|\psi\rangle\langle\psi|$ with $|\psi\rangle=\alpha|00\rangle+\beta|11\rangle$. We assume that $\rho=\alpha^2|00\rangle\langle 00|+\beta^2|11\rangle\langle 11|$ is the closest disentangled state to σ. Therefore we would have that

$$E(\sigma)=S(\sigma||\rho). \tag{A1}$$

Assume that we change ρ a little bit, i.e., we have

$$\rho_\lambda=(1-\lambda)\rho+\lambda\rho^* \tag{A2}$$

with a small λ such that ρ_λ and ρ^* are disentangled. For ρ to be the closest disentangled state to σ we have to have that

$$\frac{d}{d\lambda}S(\sigma||(1-\lambda)\rho+\lambda\rho^*)|_{\lambda=0}\geqslant 0. \tag{A3}$$

Using the complex representation of Eq. (51) for the derivative of the logarithm we quickly find

$$\begin{aligned}
&\frac{d}{d\lambda}S(\sigma||(1-\lambda)\rho+\lambda\rho^*)|_{\lambda=0}\\
&\quad=-\frac{d}{d\lambda}\mathrm{tr}\{\sigma\ \ln[(1-\lambda)\rho+\lambda\rho^*]\}|_{\lambda=0}\\
&\quad=-\frac{d}{d\lambda}\frac{1}{2\pi i}\oint dz\ \mathrm{tr}\left\{\sigma\frac{1}{z\mathbb{1}-\rho_\lambda}\right\}\mathrm{ln}z|_{\lambda=0}\\
&\quad=-\frac{1}{2\pi i}\oint dz\ \mathrm{tr}\{(\rho^*-\rho)(z\mathbb{1}-\rho)^{-1}\\
&\qquad\times\sigma(z\mathbb{1}-\rho)^{-1}\}\mathrm{ln}z\\
&\quad=1-\mathrm{tr}\{\rho^*(|00\rangle\langle 00|+|11\rangle\langle 11|+x|00\rangle\\
&\qquad\times\langle 11|+x|11\rangle\langle 00|),
\end{aligned} \tag{A4}$$

where $x=\alpha\beta(\ln\alpha^2-\ln\beta^2)/(\alpha^2-\beta^2)$ and we have used the explicit form of σ and ρ together with Cauchy's theorem [30]. Now we have to show that Eq. (A4) is always positive. One easily checks that

$$x=\alpha\beta(\ln\alpha^2-\ln\beta^2)/(\alpha^2-\beta^2)\leqslant 1, \qquad \text{(A5)}$$

where the maximum is achieved for $\alpha^2=1/2$. The right-hand side of Eq. (A4) can become smallest for $x=1$. For Eq. (A3) to be positive we therefore need to show that

$$\mathrm{tr}\{\rho^*(|00\rangle\langle 00|+|11\rangle\langle 11|+|00\rangle\langle 11|+|11\rangle\langle 00|)\}\leqslant 1. \qquad \text{(A6)}$$

Using $|\phi^+\rangle=(|00\rangle+|11\rangle)/\sqrt{2}$ this follows easily as ρ^* is not entangled and therefore $\langle\phi^+|\rho^*|\phi^+\rangle\leqslant 1/2$, which immediately confirms Eq. (A6). Therefore ρ indeed represents the closest disentangled state to σ and our proof is complete.

This proof can easily be extended to arbitrary dimensional subsystems where the maximally entangled states have the form $\Sigma_n\alpha|nn\rangle$. In that case the proof becomes more similar to the one presented in Sec. II.

[1] N. Gisin, Phys. Lett. A **210**, 151 (1996), and references therein; A. Peres, Phys. Rev. A **54**, 2685 (1996); M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Lett. A **223**, 1 (1996).

[2] A. K. Ekert, Ph.D. thesis, Clarendon Laboratory, Oxford, 1991.

[3] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Phys. Rev. A **54**, 3824 (1996).

[4] S. Hill and W. K. Wootters, Phys. Rev. Lett. **78**, 5022 (1997).

[5] A. Uhlmann, lanl e-print quant-ph/9704017.

[6] V. Vedral, M. B. Plenio, M. A. Rippin, and P. L. Knight, Phys. Rev. Lett. **78**, 2275 (1997).

[7] V. Vedral, M. B. Plenio, K. Jacobs, and P. L. Knight, Phys. Rev. A **56**, 4452 (1997).

[8] E. Rains, lanl e-print quant-ph/9707002.

[9] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. Lett. **78**, 574 (1997).

[10] M. Reed and B. Simon, *Methods of Modern Mathematical Physics–Functional Analysis* (Academic Press, New York, 1980).

[11] G. Lindblad, Commun. Math. Phys. **40**, 147 (1975).

[12] G. Lindblad, Commun. Math. Phys. **39**, 111 (1974).

[13] M. Ohya, Rep. Math. Phys. **27**, 19 (1989).

[14] F. Hiai and D. Petz, Commun. Math. Phys. **143**, 99 (1991).

[15] M. J. Donald, Commun. Math. Phys. **105**, 13 (1986); M. J. Donald, Math. Proc. Camb. Philos. Soc. **101**, 363 (1987).

[16] A. Wehrl, Rev. Mod. Phys. **50**, 221 (1978).

[17] E. Lieb and M. B. Ruskai, Phys. Rev. Lett. **30**, 434 (1973).

[18] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley-Interscience, New York, 1991).

[19] M. J. Donald (private communication).

[20] The original reference is E. Schmidt, Math. Ann. **63**, 433 (1907); in the context of quantum theory see H. Everett III, in *The Many-World Interpretation of Quantum Mechanics*, edited by B. S. DeWitt and N. Graham (Princeton University Press, Princeton, 1973), p. 3; and Rev. Mod. Phys. **29**, 454 (1957). A graduate level textbook by A. Peres, *Quantum Theory: Concepts and Methods* (Kluwer, Dordrecht, 1993), Chap. 5 includes a brief description of the Schmidt decomposition.

[21] H. Barnum, C. M. Caves, C. A. Fuchs, R. Jozsa, and B. Schumacher, Phys. Rev. Lett. **76**, 2818 (1996).

[22] T. Rockafeller, *Convex Analysis* (Princeton University Press, Princeton, 1970).

[23] W. K. Wootters, Phys. Rev. D **23**, 357 (1981).

[24] C. A. Fuchs, Ph.D. thesis, The University of New Mexico, Albuquerque, NM, 1996 (lanl e-print quant-ph/9601020).

[25] M. Hayashi, lanl e-print quant-ph/9704040 1997.

[26] M. J. Donald, Found. Phys. **22**, 1111 (1992).

[27] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, Phys. Rev. A **53**, 2046 (1996).

[28] D. Deutsch, A. Ekert, R. Jozsa, C. Macchiavello, S. Popescu, and A. Sanpera, Phys. Rev. Lett. **77**, 2818 (1996).

[29] V. Vedral, M. A. Rippin, and M. B. Plenio, J. Mod. Opt. **44**, 2185 (1997).

[30] R. V. Churchill and J. W. Brown, *Complex Variables and Applications* (McGraw-Hill, New York, 1990).