Reversible transformations from pure to mixed states and the unique measure of information

Michał Horodecki,[1] Paweł Horodecki,[1] and Jonathan Oppenheim[1,2]

[1]*Institute of Theoretical Physics and Astrophysics, University of Gdańsk, Poland*
[2]*Racah Institute of Theoretical Physics, Hebrew University of Jerusalem, Givat Ram, Jerusalem 91904, Israel*
(Received 6 December 2002; published 20 June 2003)

Transformations from pure to mixed states are usually associated with information loss and irreversibility. Here, a protocol is demonstrated allowing one to make these transformations reversible. The pure states are diluted with a random noise source. Using this protocol one can study optimal transformations between states, and from this derive the unique measure of information. This is compared with irreversible transformations where one does not have access to noise. The ideas presented here shed some light on attempts to understand entanglement manipulations and the inevitable irreversibility encountered there where one finds that mixed states can contain "bound entanglement."

I. INTRODUCTION

There are two opposing pictures of *information*. In the first picture, a source produces a large amount of information if it has large entropy. Thus information can be associated with entropy. This is because the receiver is being informed only if he is "surprised." In such an approach the information has a subjective meaning: something that is known by the sender, but is not known by receiver. The receiver treats the message as the information, if she did not know it.

One can consider a different approach to information—an objective one where a system represents information if it is in pure state (zero entropy). The state is itself the information. This view is more natural in the context of thermodynamics. There, "knowledge is power" in the sense that one can draw work from a single heat bath by use of systems in known pure states [1]. On the other hand, the heat bath is represented by a maximally entropic state, hence it is the less informative one. The pure state represents information needed to order the energy of the heat bath.

There can be many candidates for functions to measure information. However, Shannon recognized that there is a unique function that shares some natural properties to describe information. Shannon derived his unique measure based on the subjective picture of information. Therefore his information function (Shannon entropy) increases as the dispersion of the probability distribution increases. The same is true for the generalization of Shannon's entropy to the quantum case which is the von Neumann's entropy $S(\varrho) = -\text{Tr}\varrho \ln \varrho$.

One can consider a measure of objective information, that has the converse tendency: namely, $I = \log_2 d - H$ where $\log_2 d$ is the maximal entropy of the system (i.e., the system has d states). In the quantum case it would be $\log_2 d - S(\varrho)$. Such a function was naturally interpreted as the information contents of the state as introduced by Brillouin.

One can ask the question: can this function be derived independently of the notion of entropy, so that it is not just a subtraction of two known terms, but rather has an autonomous meaning?

It turns out that there it is such a possibility and it is offered by quantum information theory: in Ref. [2] we have derived the function I as the unique one, which does not increase under some class of operations. The motivation came from considering information as a resource in distributed systems [3]. The main aim of the present paper is to present the full rigorous version of that derivation. In the process, we give a protocol for reversible transformations between states using a random source of noise. We also discuss these results in the context of the issue of reversibility and entanglement theory.

It is quantum information theory (QIT) that provides us with a suitable perspective to attack the problem. Indeed, one of the central themes of QIT is the idea of *optimal transitions between states under a restricted class of operations*. This originates from attempts to describe entanglement of quantum states. Although it was difficult to say what exactly entanglement was, it was clear that it could not increase under the class of operations made up of *local operations and classical communication* (LOCC) [4,5]. These operations allow one to use any amount of separable states for free, but do not allow one to create entangled states. One can take the converse point of view: one starts with a given class of operations (LOCC operations), and treat the states that are not free as containing a resource, which can be called entanglement (cf. Ref. [6]). The basic question of entanglement theory is: can state ϱ be transformed into σ by LOCC? What is the optimal rate of such a transition?

In entanglement theory, this allowed one to define a number of measures of entanglement, since essentially, any function that does not increase under LOCC is a measure. However, thus far, no one has found a unique measure. The essential difficulty (as will become clearer) is that operations under LOCC are not reversible. However, if one has a restricted class of operations for which transitions are reversible, then we will see that the rate of transitions gives one a unique measure. This is similar to pure bipartite state entanglement where we have reversibility, and there is unique measure of entanglement (entropy of subsystem) [7,8].

In the present work, we consider a restricted class of operations we shall call noisy operations (NO) [9] and use this to develop a unique measure for information. Essentially, we consider operations where one is allowed to use random noise as a free resource. Perhaps counter intuitively, random-

ness allows one to make the transformations reversible: the number of pure states and noise which is needed to form the state, is the same as the amount that can be obtained from the state. The usual interpretation of mixed states is that their creation involves irreversibly destroying information. Here we see that if one has access to noise as a resource, then there is no irreversibility.

This has interesting consequences concerning entanglement theory, since there, the irreversibility is often associated with the fact that one is dealing with mixed states. Here, we see that transitions into mixed states need not involve irreversibility. In fact, the axiomatic structure of the paradigm presented here involving mixed states is very similar to pure-state entanglement manipulation. This shows that *a priori*, mixed-state entanglement manipulation need not involve irreversibility, leaving open the question of why entanglement manipulation involves inevitable irreversibilities.

Other restricted classes of operations may lead one to find unique measures for other quantities. Here, we consider the optimal transitions between states by means of NO. In the asymptotic limit of many identical copies, we will obtain that there is only one function that does not increase under NO. We will establish that the optimal ratio of conversion between a state ϱ of a N qubit system and a state σ of a N' qubit system is equal to $N-S(\varrho)/N'-S(\sigma)$. The transitions are reversible, even though mixed states are involved. Finally we will consider operations without free noisy ancillas. Then the mixed states have to be created from pure states by partial trace, which introduces irreversibility. We discuss the implications of our results on understanding entanglement transformations, especially bound entanglement.

The work is organized as follows. In Sec. II we introduce the class of noisy operations. Then in Sec. III we show how one can transform a given state into another state, under NO, provided certain conditions are met. In Sec. IV we go to the asymptotic regime, and show that these transition rates are optimal. This will allow us to find the unique measure of information in Sec. V. In Sec. VI we discuss the case of transitions without access to noise, and give the transition rates in this case. We discuss this in terms of understanding the source of irreversibility in transitions, and relate it to attempts to understand entanglement in Sec. VII. We conclude with some open questions in Sec. VIII.

II. NOISY OPERATIONS

Perhaps the most important restricted class of operations that has been considered in quantum information theory is LOCC, which was introduced in the context of understanding entanglement in shared systems. One is then interested in questions as how many maximally entangled states can a particular state be transformed into (i.e., the *rate* of distilling singlet). However, analyzing LOCC operations proved rather difficult. Therefore, to facilitate the investigation of entanglement, a larger class of operations was analyzed—the so-called PPT operations [10–12], which are superoperators that preserve the positivity of partial transpose.

One can also consider other restricted classes of operations, and consider various versions of the state transforma-

tion problem. On the extreme end, one allows all operations, and adding any ancilla. Then any state can be created for free, so that there is no resources to be manipulated, and the theory becomes trivial.

As one knows, any operation can be composed out of a unitary operation, adding an ancilla in some state, and removing an ancilla. Suppose that we want to make the theory nontrivial, while keeping all unitaries in our class of allowable operations. The only way is then to restrict the state of the free ancilla, or somehow restrict removing ancillas. In the present work, we consider only restrictions to the free ancilla. While one could instead consider restrictions on removing ancillas, we believe that this would give identical results [13].

Thus we will restrict to choosing states that can be added for free by means of ancillas. Remarkably, the choice of which ancillas to allow is forced on us. It turns out that the only choice that does not make the theory trivial is that the free ancilla must be in maximally mixed state. Essentially, we will see in Sec. VII that if one allows any other ancilla, then all transition rates become infinite. This fixes the class of operations we will call NO. The class NO is therefore very natural, as it is the only one that gives nontrivial transition rates.

In entanglement theory, an entangled state of Schmidt rank 2 represents the same resource whether it acts on a Hilbert space $C^2 \otimes C^2$ or on a larger space $C^d \otimes C^d$. This is because embedding a state into a larger Hilbert space is equivalent to adding local ancillas in a pure state. In our case, a state acting on a Hilbert space C^2 is not the same resource as the one acting on C^d. This is because adding ancilla in a pure state is adding a new resource.

III. OPTIMAL TRANSITIONS UNDER NOISY OPERATIONS: SINGLE-COPY CASE

In this section we will present a protocol to transform single copies of states into each other by diluting them with noise. We will show that the transition from a single copy of ϱ to a single copy of state σ is possible if and only if the latter is *more mixed* than the former. This is provided the Hilbert space is the same for both states, i.e., they occupy the same number of qubits. We will also consider the transitions between systems of different number of qubits. One then has to add maximally mixed ancillas to one of the systems (or to both), so that the number of qubits become equal. Then we can apply the above criterion. The term "more mixed" [14] has the following meaning: for states ϱ and ϱ' on the Hilbert space $\mathcal{H}=C^d$, we say that ϱ is more mixed than ϱ' ($\varrho > \varrho'$) if their eigenvalues in decreasing order satisfy $\sum_{i=1}^k \lambda_k \leq \sum_{i=1}^k \lambda_k'$, for all $k \leq \dim\mathcal{H}$. (In the same way, one can say that some probability distribution is more mixed than another one.) If the state is more mixed, its eigendistribution is more spread. The order introduced by the relation ">" has a largest element—the maximally mixed state. It is easy to see that it is more mixed than any other state.

Let us now prove the main result of this section. *Proposition 1.* For states ϱ and σ of d-level systems the transition $\varrho \to \sigma$ by NO is possible if and only if $\varrho > \sigma$.

Proof. "$\Rightarrow$" follows from the fact [15] that $\sigma > \varrho$ iff there exists a bistochastic map [16] that maps ϱ into σ. Since noisy operations (for equal input and output dimensions) are bistochastic, then $\varrho \rightarrow \sigma$ implies $\sigma > \varrho$. To prove "$\Leftarrow$" we cannot use the result of Ref. [15], because we do not know if the existing map can be taken to be noisy operations. Instead we will construct the map explicitly. Let us then assume that $\sigma > \varrho$. First we can always rotate ϱ unitarily, so that it commutes with σ. Thus we can assume without loss of generality that the states commute. We can now use the fact [17] that if probability distribution $\{q_i\}$ is more mixed than $\{p_i\}$, then the former can be obtained from the latter via a mixture of permutations, i.e.,

$$q_i = \sum_j \alpha_j p_{\sigma_j(i)}, \tag{1}$$

where $\Sigma_j \alpha_j = 1$, while σ_j are permutations of indices of the probability distribution. Let then p_i be the eigenvalues of ϱ and q_i the eigenvalues of σ. We will consider state $\varrho \otimes \tau_N$ (where τ_N is an added maximally mixed state of dimension N) and construct some permutation of eigenvalues of the latter density matrix. After such permutation, and removing the ancilla, the state will approach σ for large d. For simplicity we will assume that there are only two permutations σ_1 and σ_2, so that $q_i = \alpha p_{\sigma_1(i)} + (1 - \alpha) p_{\sigma_2(i)}$.

The state $\varrho \otimes \tau_N$ consists of blocks, of dimensions N,

$$\frac{1}{N} (\underbrace{p_1, \ldots, p_1}_{N}, \ldots, \underbrace{p_d, \ldots, p_d}_{N}). \tag{2}$$

We will divide each block into two groups of entries: N_1 first entries and the rest $N_2 = N - N_1$ entries. Now we will apply permutation σ_1 to the first entries of each block. Similarly, we apply it to the second set of entries, and so on, in the first group. The second group is subjected to permutation σ_2 in a similar way. The resulting density matrix is

$$\frac{1}{N} (\underbrace{p_{\sigma_1(1)}, \ldots, p_{\sigma_1(1)}}_{N_1}, \underbrace{p_{\sigma_2(1)}, \ldots, p_{\sigma_2(1)}}_{N_2}, \ldots, \underbrace{p_{\sigma_1(N)}, \ldots, p_{\sigma_1(N)}}_{N_1}, \underbrace{p_{\sigma_2(N)}, \ldots, p_{\sigma_2(N)}}_{N_2}). \tag{3}$$

Now we trace out the ancilla. This means that we sum all elements of each block, and instead of the block, take the resulting number. The obtained eigendistribution is given by

$$\tilde{q}_i = \frac{N_1}{N} p_{\sigma_1(i)} + \frac{N_2}{N} p_{\sigma_2(i)}. \tag{4}$$

Choosing large N and suitable N_1, N_2 one can approach α and $1 - \alpha$ with arbitrarily high accuracy. This ends the proof of the proposition.

IV. OPTIMAL TRANSITIONS UNDER NOISY OPERATIONS: ASYMPTOTIC REGIME

Here we will consider asymptotic transitions of type

$$\varrho^{\otimes n} \rightarrow \sigma^{\otimes m_n}. \tag{5}$$

Usually it is not possible to obtain a perfect state $\sigma^{\otimes m_n}$ from $\varrho^{\otimes n}$ even if an arbitrarily large amount of copies can be used. It is, however, possible to obtain the state σ_n that will asymptotically converge to $\sigma^{\otimes m_n}$,

$$\varrho^{\otimes n} \rightarrow \sigma_n \approx \sigma^{\otimes m_n}. \tag{6}$$

Thus we allow for inaccuracy, provided it vanishes in the limit of large n. The fidelity can be measured by the trace norm, i.e., one requires that

$$\|\sigma_n - \sigma^{\otimes m_n}\| \rightarrow 0 \quad \text{for} \quad n \rightarrow \infty. \tag{7}$$

The rate of given protocol of asymptotic $\varrho \rightarrow \sigma$ transition is given by the asymptotic ratio $\lim_n (m_n / n)$. The optimal transition rate denoted by $R(\varrho \rightarrow \sigma)$ is given by supremum over rates attainable by protocols that satisfy the asymptotic accuracy condition (7).

A. Conversion from mixed to pure states

We will now consider the optimal rate for transition to the one qubit pure state π, i.e., $\varrho \rightarrow \pi$. We will show that if ϱ is a state of d-level system then

$$R(\varrho \rightarrow \pi) = I(\varrho), \tag{8}$$

where $I = N - S(\varrho)$ with $N = \log_2 d$ being the amount of qubits occupied by the state ϱ. In other words, the transformation from pure states to mixed states is reversible, in the sense that the number of pure states which is needed, or which can be obtained, is the same. The proof could be just use of Schumacher compression [18], however, with a different interpretation (similar to that in Ref. [19]). We will also show that conversely, the amount of copies in state ϱ that can be obtained under NO per input pure qubit is also equal to I. The proofs will be similar to the reasoning of Nielsen in Ref. [20] where he derived asymptotic rates of pure-state entanglement manipulations from single copies based on majorization.

We will use law of large numbers [18,21], which implies that there exists a subset of eigenvalues of $\varrho^{\otimes n}$ called the typical set T with useful properties. More precisely, given $\epsilon, \delta > 0$, there exists large enough n, and the set T of eigenvalues such that

$$\sum_{p_i \in T} p_i \geq 1 - \epsilon, \tag{9}$$

$$2^{-n(S+\delta)} \leq p_i \leq 2^{-n(S-\delta)} \quad \text{for} \quad p_i \in T. \tag{10}$$

These are thus the eigenvalues that carry almost the whole weight and they are more or less uniform. One can consider two states ϱ_{typ} and ϱ_{atyp}, given by

$$\varrho_{typ} = \frac{1}{c} \sum_{p_i \in T} p_i |i\rangle\langle i|, \quad \varrho_{atyp} = \frac{1}{1-c} \sum_{p_i \notin T} p_i |i\rangle\langle i|, \tag{11}$$

where $|i\rangle$ are eigenvectors corresponding to p_i, and $c = \sum_{p_i \in T} p_i$ is a normalization constant. Clearly $\varrho^{\otimes n}$ is a mixture of those states

$$\varrho^{\otimes n} = c \varrho_{typ} + (1-c) \varrho_{atyp}. \tag{12}$$

Since $c \geq 1 - \epsilon$ one finds that ϱ_{typ} is close to $\varrho^{\otimes n}$,

$$\|\varrho_{typ} - \varrho^{\otimes n}\| \leq 2\epsilon. \tag{13}$$

Thus it suffices to use ϱ_{typ} instead of $\varrho^{\otimes n}$. Let us first show that one can convert ϱ_{typ} into approximately $n(N-S)$ copies of pure qubits. To this end, note that the eigenvalues of ϱ_{typ} satisfy

$$\lambda_i \equiv \frac{p_i}{c} \geq \frac{1}{c} 2^{-n(S+\delta)}. \tag{14}$$

Thus ϱ_{typ} is less mixed than the state ϱ_{out} with eigenvalues

$$\left\{ \underbrace{\frac{1}{D}, \dots, \frac{1}{D}}_{D}, \underbrace{0, \dots, 0}_{d^n - D} \right\}, \tag{15}$$

where D is given by

$$D = \left\lceil \frac{1}{\frac{1}{c} 2^{-n(S+\delta)}} \right\rceil \tag{16}$$

(the eigenvectors of ϱ_{out} are irrelevant, as we can perform any unitary transformation for free). Both of the states act on a d^n dimensional space, so that we can apply our Proposition 1. Thus it is possible to go from ϱ_{typ} to ϱ_{out} via noisy operations. If we choose D to be larger than in Eq. (16), namely, so that it is a power of 2, the transition is still possible. The smallest such D satisfies $\log_2 D = \lceil n(S+\delta) \rceil \leq n(S+\delta) + 1$. Then the state ϱ_{out} represents exactly the tensor product of $\log_2 D$ qubits in maximally mixed state and $n \log_2 d - \log_2 D \geq n(\log_2 d - S - \delta) - 1$ qubits in pure states. Thus one can remove the mixed qubits, and keep the obtained pure qubits. Call the obtained state π_{out}. The rate of the transition is the number of obtained pure qubits divided by n. For large n this tends to $\log_2 d - S - \delta$. Since δ can be chosen arbitrarily small, we obtain the optimal asymptotic rate equal to $\log_2 d - S$.

One could think that we obtain the pure qubits exactly. However, we used Proposition 1, where the transition is not exact, though arbitrarily precise.

Yet we have not transformed $\varrho^{\otimes n}$ but ϱ_{typ}. We now take instead of ϱ_{typ}, the state $\varrho^{\otimes n}$ and apply the same action, which transformed ϱ_{typ} into the required amount of pure qubits (called the action Λ). It is now easy to see that $\Lambda(\varrho^{\otimes n})$ is close to a final state of pure qubits π_{out}. Indeed, we have

$$\|\Lambda(\varrho^{\otimes n}) - \pi_{out}\| = \|\Lambda(\varrho^{\otimes n}) - \Lambda(\varrho_{typ})\| \leq \|\varrho^{\otimes n} - \varrho_{typ}\| \leq \epsilon, \tag{17}$$

where the second last inequality comes from the fact that completely positive trace-preserving maps are contractions on Hermitian operators in trace norms, i.e., $\|\Lambda(A)\| \leq \|A\|$ for Hermitian A [22].

Now we should show that the converse is possible, i.e., to create a state $\varrho^{\otimes n}$ it is sufficient to start with $\log_2 d - S$ pure qubits per output copy of ϱ. However, the proof is similar to the above. The only difference is that we now use the other part of Eq. (10). Namely, we note that ϱ_{typ} is *more* mixed than the state with eigenvalues

$$\left\{ \underbrace{\frac{1}{D'}, \dots, \frac{1}{D'}}_{D'}, \underbrace{0, \dots, 0}_{d^n - D'} \right\}, \tag{18}$$

where D' is given by

$$D' = \left\lfloor \frac{1}{\frac{1}{c} 2^{-n(S-\delta)}} \right\rfloor. \tag{19}$$

Again, due to Proposition 1 we can turn ϱ_{typ} into the latter state. Changing D' into a suitable power of 2 (so that it is smaller than D' of the above equation hence passing from ϱ_{typ} is still possible) one gets that the latter state is a tensor product of $\log_2 D'$ qubits in maximally mixed states and approximately $n(\log_2 d - S)$ qubits in pure states.

Thus starting with $n(\log_2 d - S)$ qubits in a pure state, one has to add $\log_2 D'$ qubits in the maximally mixed state, and pass to the state ϱ_{typ} which can be made arbitrarily close to $\varrho^{\otimes n}$ by choosing small ϵ.

B. Optimality of $\log_2 d - S$ transition rates and optimal mixed-mixed transition rates

We will now show that the obtained rates are optimal. We will follow Ref. [7] invoking standard thermodynamical reasoning concerning Carnot efficiency (cf. Ref. [8]). Essentially, we will show that $I = N - S$ cannot increase under NO maps, and then show that if our transitions are not optimal, one could increase I under NO. We will use the reversibility of our protocol, and also the asymptotic continuity property of von Neumann entropy.

We will prove optimality by contradiction. Suppose that for the transition to pure qubits $\varrho \to \pi$ one can obtain a better rate than $R(\varrho \to \pi) = N - S$ (where $N = \log_2 d$, ϱ acts on C^d).

Then one can run the following transition

$$\pi \rightarrow \varrho \rightarrow \pi, \tag{20}$$

and obtain a rate of such transition which is more than 1. In other words, employing $n < m(N-S)$ pure qubits, according to the assumption, one gets m pairs in state ϱ. Then one can apply the protocol of the preceding section to the m pairs of ϱ, to obtain $m(N-S)$ pure qubits. Thus one would be able to *increase* the number of pure qubits from n to m. Repeating the procedure one can obtain an arbitrary number of pure qubits.

Now, we have to show that this is impossible. This follows from the fact that $N-S$ cannot increase under NO maps. Indeed, unitary maps do not change the quantity. Partial trace of one qubit decreases N by 1, and can increase entropy at most by 1. Finally, adding a system in maximally mixed state, increases N by 1, but also increases entropy by 1. Now, for m pure qubits, $N-S=m$, while for n qubits we have $N-S=n<m$, thus the function $N-S$ must increase.

This is yet not the full proof, as we have made an implicit assumption, that the final qubits are exactly pure states. In fact it is not true, as all our conversions are only asymptotically true. However, the von Neumann entropy is asymptotically continuous, namely, for N qubit states ϱ and σ we have [23]

$$|S(\varrho) - S(\sigma)| \leq N \|\varrho - \sigma\| + O(1). \tag{21}$$

In our case we take $\varrho = \pi^{\otimes m}$ and σ_m being the actual final state. We know then that $S(\varrho) = 0$ and that $\|\sigma_m - \varrho\|$ tends to zero as m goes to infinity. Thus $|S(\sigma_m)|/m \rightarrow 0$ for large m. Thus the *density* of the function I tends to 1 for the state σ_m. This density is also 1 for the initial state $\pi^{\otimes n}$. Thus we can write that in our process $I_{out} = m_n - o(m_n)$; on the other hand $I_{in} = n$. We will show that for large n (which also implies that m_n is large) $I_{in} < I_{out}$. Indeed that latter inequality is equivalent to the following set of equivalent inequalities:

$$m_n - o(m_n) > n,$$

$$\frac{m_n}{n} - \frac{o(m_n)}{n} > 1, \tag{22}$$

$$\frac{m_n}{n}\left(1 - \frac{o(m_n)}{m_n}\right) > 1.$$

The quantity inside the bracket tends to 1, while in our protocol m_n/n goes to a number greater than 1. Thus the inequality holds, which is impossible. Therefore our assumption that our rate is not optimal is incorrect. In a similar way one can show that one cannot obtain a better rate than

$$R(\pi \rightarrow \varrho) = \frac{1}{I} \tag{23}$$

while going from pure states to mixed ones.

Clearly since the transitions from mixed to pure states are reversible and optimal, one can use these protocols to go from one mixed state to another in a reversible and optimal way by just distilling pure states and then creating another mixed state. This gives that the optimal ratio of conversion between state ϱ of a N qubit system and state σ of a N' qubit system is equal to

$$R(\varrho \rightarrow \sigma) = N - S(\varrho)N' - S(\sigma). \tag{24}$$

V. INFORMATION MONOTONES AND THE UNIQUE MEASURE OF INFORMATION

Here we will derive the unique measure of information I, with virtually no assumptions. The derivation will be mostly operational. We will actually assume two properties. The first will concern the intuition of what information is—namely, noisy operations should not increase it. Indeed, information, whatever it is, should not be increased by unitary operations, by adding a qubit in maximally mixed state (supposed to be informationless) and discarding qubit (rather obvious requirement). Thus we postulate the following.

Postulate 1. I should be monotonic under noisy operations.

We will actually see in the following section that this postulate is rigid, in the sense that if instead of noisy operations, we had chosen operations with a free resource other than maximally mixed states, the theory would be trivial, and all rates would be infinite.

The second assumption will not be connected with the expected properties of information. Rather it will display the properties any function used in the asymptotic regime (limit of many copies) should possess.

Postulate 2. I is asymptotically continuous.

By asymptotically continuous, one means that for the state ϱ_N and σ_N of N qubits, such that $\|\varrho_N - \sigma_N\| \rightarrow 0$ for $N \rightarrow \infty$. One would then require

$$|f(\varrho_N) - f(\sigma_N)| \rightarrow 0. \tag{25}$$

We then say that f is *asymptotically continuous*. The motivation for this is that in the asymptotic regime, one identifies the states that asymptotically converge to each other. Thus the only relevant functions of states are those that also somehow identify those states. Of course in the asymptotic limit, the interesting functions become infinite, so that one has to pass to intensive quantities and divide by the number of copies to obtain *densities*. The relevant functions would be those whose densities converge on convergent sequences. Note that this not merely a technical requirement. Rather this follows from the basic assumption of the asymptotic regime—that similar states should be identified. The latter assumption is necessary, and physically natural—it is simply impossible to obtain exact transitions.

Let us now prove that there is a unique function that satisfies these two postulates.

The proof can be obtained from Refs. [24,25]. According to [25] the following inequality is true:

$$R(\varrho \rightarrow \sigma) \leq \frac{f^{\infty}(\varrho)}{f^{\infty}(\sigma)}, \tag{26}$$

where R denotes the rate of transition under any given class of operations, and f is an asymptotically continuous function nonincreasing under the class. The symbol ∞ stands for *regularization*. The regularization of function $f(\varrho)$ is $M^{\infty}(\varrho) = \lim_{n \to \infty} (1/n) M(\varrho^{\otimes n})$.

Choosing as σ the one qubit pure state π and exchanging the roles of ϱ and σ we obtain

$$R(\varrho \to \pi) \leqslant \frac{f^{\infty}(\varrho)}{f^{\infty}(\pi)}, \tag{27}$$

$$R(\pi \to \varrho) \leqslant \frac{f^{\infty}(\pi)}{f^{\infty}(\varrho)}.$$

Denoting $1/f^{\infty}(\pi) = a$ we obtain

$$R(\varrho \to \pi) \leqslant a f^{\infty}(\varrho) \leqslant \frac{1}{R(\pi \to \varrho)}. \tag{28}$$

However, we have explicit protocols which show that $R(\varrho \to \pi) \geqslant I$ and $1/R(\pi \to \varrho) \leqslant I$. Thus up to the constant a we obtain that $f^{\infty} = I$. In this sense I is the *unique* measure of information.

It is interesting to see how other measures of information are removed in the asymptotic limit. Suppose that we consider measures of information that only satisfying the first postulate. Since we see that everything is very similar to the problem of pure-state entanglement, one is not surprised that all monotones under NO are so called Shur concave functions of the density matrix. In particular there is a set of information measures (or "monotones") which is enough to determine if a transition is possible. These are the so called Ky Fan k norms, i.e., sums of the first k largest eigenvalues. By definition of the "more mixed" condition, we have $\varrho \succ \sigma$ iff for all k norms, $||\varrho||_k \leqslant ||\sigma||_k$. Thus the process $\sigma \to \varrho$ is possible iff in the process no monotone increases.

One might get the feeling that there is some contradiction here. Namely, in asymptotic transitions, the only restriction for the rate is the monotone I. Thus there are allowed transitions for which other monotones increase. Indeed, we say that $\varrho^{\otimes n} \to \sigma^{\otimes m}$ is possible, though it is clear that some of the monotones will increase. The solution is that, in fact we are not talking about exact transitions. Thus in the actual transition, the final state obeys the nonincreasing of monotones. For that state, all monotones are not greater than for the initial state. The monotones are, however, not asymptotically continuous, and they see differences between that actual state, and the required state $\sigma^{\otimes m}$. The only monotone that does not see the difference is I. Therefore only this function survives in the asymptotic limit.

The choice of free resource is unique

One could think that the way we have obtained the information measure is not fully operational, as we assumed, somewhat arbitrarily that the free resource is the maximally mixed state. Here we will show that this is the only reasonable choice, if we want to allow ancillas at all, and if the theory is to be nontrivial, i.e., the transition rates are finite,

and therefore, not all states can be obtained for free. We thus assume that our operations include unitary transformations, and partial trace, and will try to play with third component—adding ancillas.

Suppose that instead of maximally mixed states τ, we chose any other state ϱ_0 as a free resource. This means that we can use arbitrarily many copies of this state. From $\varrho_0^{\otimes n}$ we can produce *without use of noise* pure states by Schumacher compression [18,19] (in this paper we have not described this—we always used noise). Thus we have pure states for free. From pure states we can produce noise by entangling two qubits in a maximally entangled state and rejecting one qubit. The remaining one will be in a maximally mixed state. This is not very efficient: we spend two qubits in a pure state to get one qubit of noise. However, pure states are for free, hence this method is sufficiently good in our situation. Now we have both noise and pure states for free, hence via the protocol described in the preceding section, we can create any state. The theory becomes trivial—all states are for free; all rates are infinite. Thus if we allow adding systems for free at all, we can only add ones in maximally mixed state. We thus see that Postulate 1 is rather rigid, in the sense that changing it to a class of operations which allows any other ancilla, will result in a trivial theory.

VI. REVERSIBILITY AND IRREVERSIBILITY

Note that we have a kind of reversibility: the amount of pure qubits that can be drawn from a given state is equal to the amount we need to create the state. Let us consider another situation, where we count everything (no free resource). We then see that there is basic irreversibility: transition from almost any state ϱ to any other state is irreversible. For example, one can draw I pure qubits from ϱ, but to create ϱ, one needs many more pure qubits. There are two reasons for this. The first reason is trivial—to get N qubits in state ϱ one needs N qubits anyway. This is 1 qubit per output qubits, which is already more than $I = N - S(\varrho)$. Now, however, even more pure qubits are needed. Namely, the output state has nonzero entropy. However, the only way of producing entropy out of pure states is rather wasteful: one entangles two qubits, and removes one of them (as already described in the preceding section). Indeed, previously, we had a free source of entropy—maximally mixed states, now we have only pure states to our disposal, and we count them.

Interestingly, in the classical world there is no way to produce entropy at all. Therefore in classical statistical mechanics, one has to assume mixed state from the very beginning. Quantum mechanics allows one to produce mixed states out of pure ones. This may lead one to prefer Bayes concept of probability.

Proposition 2. $N + S$ pure qubits are necessary and sufficient to produce ϱ if one does not have access to noise.

That this is sufficient can be seen by noting that ϱ can be created by *purification* of ϱ_{typ}. We thus consider a pure state of two systems A and B. Subsystem A has N qubits, and its state is ϱ_{typ}. The state of subsystem B (the purification) is also ϱ_{typ}, but we do not need it to be an N qubit system, but

rather want it to use the smallest possible amount of qubits. The latter is equal to S qubits. Thus $N+S$ qubits in pure state are needed to prepare ϱ_{typ} (preparation is discarding the system B). That this number of qubits are necessary simply stems from the fact that we start from an initially pure state, so to get a mixed state we must trace out part of the initial system, and the "garbage" that gets traced out must have at least S qubits (since the number of qubits of garbage cannot be less than its entropy, and the garbage must have entropy S since the system is initially pure). We must also have at least N qubits left over to form the state. So, in general, to create the N qubit state ϱ we need $N+S$ pure qubits, but we can draw only $N-S$ qubits. The "information of preparation" is much greater than "information of distillation." During the transition

$$\psi \rightarrow \varrho \rightarrow \psi \qquad (29)$$

we lose $2S$ pure qubits.

Proposition 3. To produce the mixed-mixed transition $\varrho \rightarrow \sigma$, without access to noise, $\Delta N + \Delta S$ qubits are necessary and sufficient where $\Delta N \equiv N(\sigma) - N(\varrho)$ and $\Delta S \equiv S(\sigma) - S(\varrho)$.

To see that these resources are necessary, we note that a general protocol involves an initial state $\varrho \otimes |\psi\rangle\langle\psi|$, where $|\psi\rangle$ is some initial pure state. One then performs unitaries to give a state ϱ', and then one traces out the garbage g to leave the state σ. We can then use the triangle inequality

$$|S(\sigma) - S(g)| \leqslant S(\varrho') = S(\varrho) \qquad (30)$$

to see that the number of garbage bits traced out $N(g)$ satisfies $N(g) \geqslant S(g) \geqslant \Delta S$ [if $S(g) \geqslant S(\sigma)$ then trivially $S(g) \geqslant \Delta S$]. So, we need a minimum of $N(\sigma) + \Delta S$ pure qubits to create σ, but we already had $N(\rho)$ bits to start with, so the minimum amount of additional qubits needed is $\Delta N + \Delta S$.

The protocol that realizes this bound is to reversibly distill ϱ into $N(\varrho) - S(\varrho)$ pure qubits and $S(\varrho)$ bits of noise in a manner that we shall shortly describe. We then add in an additional $\Delta N - \Delta S$ pure qubits. However, we also need ΔS bits of noise, which costs $2\Delta S$ pure states (this is the only part of the protocol which is irreversible). We then create σ reversibly as described in the preceding section, using the $\Delta N + \Delta S$ additional qubits.

The distillation procedure can be realized using a scheme similar to quantum data compression [18] and to the concentration of entanglement scheme of Ref. [8] (here, however, the procedure is applied to the entire state). The protocol is essentially a projective measurement onto blocks proportional to the identity. On average, the size of the Hilbert space that the state is projected onto will be of size $S(\varrho)$, and so, the state can then be unitarily rotated to leave $N(\varrho) - S(\varrho)$ pure states. We will explicitly give the protocol for n qubits, i.e., $N(\varrho) = 1$ but the extension to higher dimensional states is straightforward.

We can write the state in the eigenbasis which we label as 0 and 1, i.e., $\varrho = a|0\rangle + b|1\rangle$. We have have n copies, i.e., we operate on the state $\varrho^{\otimes n}$, and then we measure how many zeros this state has. This is a measurement with $n+1$ out-

comes and it will yield a result $k = 0, \ldots, n$ telling us how many zeros there are. This projects us onto a state that has $d_k = \binom{n}{k}$ basis vectors, all with equal coefficients. That is, it is proportional to the identity. The probability of finding a particular outcome k is $p_k = \binom{n}{k} a^{2k} b^{2(n-k)}$ and since it does not, in general, span the entire Hilbert space, it can be unitarily transformed to yield $I_k = n - \log_2 d_k$ pure states.

Each process $\varrho \rightarrow \{p_k, \rho_k\}$ after which I_k pure states are extracted from ρ_k with probability p_k, provides

$$N_o = \sum_k p_k I_k - H(\{p\}) \qquad (31)$$

total pure states. The Shannon entropy $H(\{p\})$ of distribution $\{p_k\}$ equals the cost of the erasure of information which allows us to work with an ensemble of ρ_k's [26]. Thus we need $I_{er} = H(\{p\})$ bits of erasure to pay for the next part of the scheme, in which they draw $\Sigma_k p_k I_k$ pure states. This quantity, which is of order $\log_2 n$ is negligible in the large n limit. We can divide the above equation by n to obtain the amount of extractable pure states per qubit.

$$N_o/n = 1 - S(\varrho), \qquad (32)$$

where the erasure cost has been neglected since it is of order $\log_2 n/n$. This completes our proof of the proposition.

This allows one to think of states in the following way: the mixed state consists of $N-S$ bits of information and S bits of noise. Thus to produce it one needs $N-S$ qubits in pure states, to account for information, and $2S$ qubits to produce noise. Indeed, one bit of noise costs two pure qubits—since noise is produced by rejecting part of entangled system.

It is interesting that one needs to add a free resource (noise) in order to achieve efficient transitions from pure to mixed states which are much less "useful" than mixed-to-pure transitions. Indeed, the latter is a task that can be associated with such actions as cooling, error correction, increasing signal, etc. This useful task can be performed *without* the help of an additional resource at the optimal rate. Only the converse direction, which is not useful (who wants to have mixed states instead of pure ones?) needs noise, and is much less efficient without noise.

There are other cases where reversibility needs noise. For example, according to the Shannon second theorem, one can simulate one use of noiseless channel by $1/C$ uses of a noisy channel of capacity C. However, one cannot do the converse, i.e., simulate noisy channels by a noiseless one, without sharing random correlated data [27]. Again, the useful task does not need any additional resource, while the useless task needs one. This is clear, if one realizes that in both situations we deal with dilution of some valuable resource into noise. Similarly in thermodynamics, the thermodynamical system with difference of temperature can be thought as being "pure energy" (such as mechanical energy) diluted into "pure heat." To draw work out of it one does not need any additional resource. However, to create the system of heat baths efficiently, one needs a heat reservoir at the beginning. Oth-

erwise, one has to spend work to produce heat, exactly as we needed to spend pure states to produce noise.

VII. DISCUSSION: COMPARISON WITH ENTANGLEMENT TRANSFORMATIONS

The paradigm discussed in this paper may be useful to understand the problems of entanglement theory. As one knows there is a basic irreversibility in entanglement transformations. We deal there with bipartite systems, shared by distant parties. One is interested in how many pure singlets are needed to form a state ρ_{AB} (the *entanglement cost*, and also, how many singlets can be obtained from the state (the *distillable entanglement*). If ρ_{AB} is pure, then the entanglement cost is equal to the distillable entanglement in the limit of many copies of ρ_{AB} [8]. The tranformations are reversible.

However, it is known that for a number of mixed states, the distillable entanglement is not equal to the entanglement cost [28]. One has irreversibility. It has generally been assumed that this is because one is making transformations between pure states (in this case, singlets), and the mixed state ρ_{AB}. One therefore expects some information loss. However, as we have seen here, one can make transformations between pure and mixed states completely reversible, provided one has access to noise. And indeed, in the paradigm of entanglement theory, there is no reason why two distant parties could not share some initial noisy resource. There is no special *a priori* reason for irreversibility in entanglement theory. It is therefore interesting to compare the situation discussed here with that of entanglement theory. This comparison is summerized in the following table, and described below.

Paradigm	Class of operations	Free resource	Expensive resource	Reversible
Information	NO	Maximally mixed states	Pure states	Yes
Pure-state entanglement	LOCC	Separable states	Singlets	Yes
Mixed-state entanglement	LOCC	Separable states	Singlets	No
Thermodynamics	Adiabatic processes	Heat [29]	Work	Yes
TOE [30,31]	LOCC + PPT states	PPT states	Singlets	No (?)
PPT [32]	PPT operations	PPT states	Singlets	In some cases

Instead of NO, in entanglement theory we have there LOCC, which means that (1) arbitrary *local* unitary operations can be performed, (2) any *local* ancilla can be added, (3) any *local* partial trace can be performed, and (4) qubits can be communicated between distant parties only via a *dephasing channel*. The role of noise is played by separable states—all the states that can be produced for free within the allowed class of operations are free resources. The role of pure states is played by pure entangled states.

One could imagine that like with "local information theory," in entanglement theory, any state is a reversible mixture of two phases: pure entanglement and a separable noisy phase. One should be able to draw the same amount of pure entanglement from a given state as is needed to produce it. Creation of mixed states would be reversible *dilution* of pure entanglement into mixed, separable states.

In this simple picture we would have only two kinds of basic elements in entanglement theory: pure entangled systems and disentangled systems. One is useful, the other is useless. A state that is neither pure entangled nor disentangled, consists of those two basic elements. This is in parallel to the paradigm presented in this paper, where the useful elements were pure states, the useless maximally mixed ones.

As noted, such a situation exists for pure states, where we can reversibly concentrate and dilute entanglement. However, such a situation does not exist with mixed states in entanglement theory. What is the basic difference between mixed-state entanglement and the paradigms (I) of pure entanglement, and (II) the present NO one?

In both (I) and (II) we have the following common point. We define states that can be added for free, and then the class of operations. Then in both cases it turns out that the free states remain the only nontrivial set of states closed under the class of operations. Now in mixed-state entanglement we may have another basic element-bound entangled ones. One cannot obtain them from separable states, but also one cannot obtain any pure entanglement from them. Thus the set of states closed under the class of operations is greater than it would seem from the construction of the paradigm. Thus in situations (I) and (II) we have only two elements: useful and useless. In paradigm (II) the useful element is information, the useless one being noise. In paradigm (II) the useful element is entanglement, the useless one being separability. Here, entanglement *itself* is divided into at least two phases: bound and pure. From bound entanglement we cannot make pure states, so we call it useless as well. Thus we can have states that have entanglement, but are useless. This is different from (I) and (II), but similar to thermodynamics: we have there two forms of energy, useful and useless. In Ref. [31] we have asked a question—is it possible that mixed-state entanglement is like thermodynamics. There would be three basic elements: separable states (no entanglement), bound entanglement, and pure entanglement; similarly as in thermodynamics there are states without energy, with disordered energy (single heat bath), and with ordered energy (mechani-

cal energy). All three kinds could be reversibly mixed.

In Ref. [31] it was shown that such a picture can be treated as a sort of "first-order approximation" rather than full description of asymptotic bipartite entanglement. Related questions were studied in Ref. [32] where reversibility for some states holds, if the so-called PPT superoparators are allowed [10]. The relation between the latter result and the "thermodynamic" approach of Ref. [31] goes beyond the scope of this paper and is explained in Ref. [31] itself.

VIII. CONCLUSION

Contrary to what might be imagined, we have shown that mixed states do not necessarily impose irreversibility. One can reversibly transform pure states into mixed ones, provided one has access to random noise. This defines a class of operations (NO) which can then be used to explore the transition rates between various states. It is found that the infor-

mation measure $I = N - S$ cannot decrease under NO, and is therefore the unique asymptotically continuous measure of information. It would be extremely interesting to explore other restricted classes of operations in addition to NO, to see whether there are other nontrivial theories. Exploring the connection between this, and the LOCC paradigm of entanglement theory, would be extremely useful in understanding entanglement in distributed quantum systems. Perhaps ideas along the lines of Ref. [31] may prove fruitful.

[1] L. Szilard, Z. Phys. **53**, 840 (1929).

[2] M. Horodecki, K. Horodecki, P. Horodecki, R. Horodecki, J. Oppenheim, A. Sen(De), and U. Sen, Phys. Rev. Lett. **90**, 100402 (2002).

[3] J. Oppenheim, M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. Lett. **89**, 180402 (2002).

[4] C.H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. Smolin, and W.K. Wootters, Phys. Rev. Lett. **76**, 722 (1996).

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

[6] G. Vidal, J. Mod. Opt. **47**, 355 (2000).

[7] S. Popescu and D. Rohrlich, Phys. Rev. A **56**, R3319 (1997).

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

[9] Such maps were considered from a formal point of view in K. Życzkowski (unpublished).

[10] E. Rains, Phys. Rev. A **60**, 179 (1999).

[11] E. Rains, Phys. Rev. A **63**, 019902 (2001).

[12] V. Vedral and M.B. Plenio, Phys. Rev. A **57**, 1619 (1998).

[13] See in this context Ref. [3], where we introduced closed local operations and considered the case where one is not allowed to remove ancillas.

[14] A. Uhlmann, Wiss. Z.-Karl-Marx-Univ. Leipzig, Math.-Naturwiss. Reihe **20**, 633 (1971).

[15] A. Chefles, e-print quant-ph/0109060.

[16] A linear completely positive map is called bistochastic if it

preserves both trace and identity.

[17] R. Bhatia, *Matrix Analysis*, Graduate Texts in Mathematics, Vol. 169 (Springer, New York, 1997).

[18] B. Schumacher, Phys. Rev. A **51**, 2738 (1995).

[19] L.J. Schulmann and U. Vazirani, e-print quant-ph/9804060.

[20] M.A. Nielsen, Phys. Rev. Lett. **83**, 436 (1999).

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

[22] M.-B. Ruskai, Rep. Math. Phys. **6**, 1147 (1994).

[23] M. Fannes, Commun. Math. Phys. **31**, 291 (1973).

[24] M. Donald, M. Horodecki, and O. Rudolph, e-print quant-ph/0106052.

[25] M. Horodecki, QIC **1**, 3 (2001).

[26] R. Landauer, IBM J. Res. Dev. **5**, 183 (1961).

[27] C.H. Bennett, P.W. Shor, J.A. Smolin, and A.V. Thapliyal, e-print quant-ph/0106052.

[28] G. Vidal and J.I. Cirac, Phys. Rev. Lett. **86**, 5803 (2001).

[29] Here, heat is understood as a single heat bath; two baths at difference temperatures are not "pure" heat in the sense that they can be used to perform work.

[30] Here, TOE stands for "Thermodynamics of Entanglement."

[31] M. Horodecki, J. Oppenheim, and R. Horodecki, Phys. Rev. Lett. **89**, 240403 (2002).

[32] K. Adenaert, M. Plenio, and J. Eisert, Phys. Rev. Lett. **90**, 027901 (2003).