Local versus nonlocal information in quantum-information theory: Formalism and phenomena

Michał Horodecki,[1] Paweł Horodecki,[2] Ryszard Horodecki,[1] Jonathan Oppenheim,[1,3,4] Aditi Sen(De),[1,5] Ujjwal Sen,[1,5] and Barbara Synak-Radtke[1]

[1]*Institute of Theoretical Physics and Astrophysics, University of Gdańsk, 80-952 Gdańsk, Poland*
[2]*Faculty of Applied Physics and Mathematics, Technical University of Gdańsk, 80-952 Gdańsk, Poland*
[3]*Dept. of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, United Kingdom*
[4]*Racah Institute of Theoretical Physics, Hebrew University of Jerusalem, Givat Ram, Jerusalem 91904, Israel*
[5]*Institut für Theoretische Physik, Universität Hannover, D-30167 Hannover, Germany*

(Received 12 October 2004; published 9 June 2005)

In spite of many results in quantum information theory, the complex nature of compound systems is far from clear. In general the information is a mixture of local and nonlocal ("quantum") information. It is important from both pragmatic and theoretical points of view to know the relationships between the two components. To make this point more clear, we develop and investigate the quantum-information processing paradigm in which parties sharing a multipartite state distill *local* information. The amount of information which is lost because the parties must use a classical communication channel is the *deficit*. This scheme can be viewed as *complementary* to the notion of distilling entanglement. After reviewing the paradigm in detail, we show that the upper bound for the deficit is given by the relative entropy distance to so-called pseudoclassically correlated states; the lower bound is the relative entropy of entanglement. This implies, in particular, that *any entangled state is informationally nonlocal*—i.e., has nonzero deficit. We also apply the paradigm to defining the thermodynamical cost of erasing entanglement. We show the cost is bounded from below by relative entropy of entanglement. We demonstrate the existence of several other nonlocal phenomena which can be found using the paradigm of local information. For example, we prove the existence of a form of nonlocality without entanglement and with distinguishability. We analyze the deficit for several classes of multipartite pure states and obtain that in contrast to the GHZ state, the Aharonov state is extremely nonlocal. We also show that there do not exist states for which the deficit is strictly equal to the whole informational content (*bound local information*). We discuss the relation of the paradigm with measures of classical correlations introduced earlier. It is also proved that in the one-way scenario, the deficit is additive for Bell diagonal states. We then discuss complementary features of information in distributed quantum systems. Finally we discuss the physical and theoretical meaning of the results and pose many open questions.

DOI: 10.1103/PhysRevA.71.062307 PACS number(s): 03.67.−a

I. INTRODUCTION

"Quantum information" is emerging as a primitive notion in physics following an essential extension of classical Shannon information theory [1] into the quantum domain. Quantum information cannot be defined precisely, but it is necessary to understand the role of this mysterious and "unspeakable" information [2] in newly discovered quantum phenomena such as teleportation [3] or cryptography [4,5]. These phenomena suggest that quantum states represent quantum information—reality we process in the laboratory, but which cannot be described as a sequence of classical symbols on a Turing tape [6,7]. Recently the no-deleting and no-cloning theorems have been connected with the principle of conservation of quantum information [8]. Like physical quantities such as energy, quantum information has different forms and one of them is entanglement—an exotic resource extraordinarily sensitive to the environment. One finds a loss of entanglement in the transition from a pure entangled state to a noisy entangled state, yet remarkably this process can be partially reversed within the distant laboratories paradigm. Namely from a large number of noisy bipartite states shared between two distant parties one can distill a number of entanglement bits (*e*-bits) at the optimal conversion rate using local operations and classical communications (LOCC) [9].

Despite a plethora of measures which can be used to quantify entanglement, we are still far from properly understanding it. Part of the difficulty is that measures of a quantity are not enough to understand the quantity—one needs to understand entanglement in relation to something else. You cannot understand entanglement in relation to entanglement. In the above context, basic questions arise: (i) *Does entanglement exhaust all aspects of quantum information?* (ii) *Are there resources other than entanglement in the distant laboratory paradigm?* (iii) *Does quantum information involve a nonlocality which goes beyond Bell's theorem?*

The above questions have been recently considered [10–20]. In particular, a new quantum-information processing paradigm has been introduced, where we proposed the idea of attributing cost to local resources such as pure local qubits [14,15]. Instead of asking how much entanglement can be distilled from a state shared between two parties, one can ask how many local pure qubits I_l can be drawn from it. This gives a quantity (called *localizable information*) which can then be used to get insight into the double nature of quantum information. Namely, it was shown that local information can be thought of as being complementary to entanglement [16], thereby allowing one, in particular, to understand entanglement in relation to I_l.

At first glance, the idea of considering local pure states to be a resource may seem curious. In traditional entanglement theory, one thinks of local pure states as being a free resource. Each party can use as many pure-state ancillas as desired. Furthermore, one can obtain pure states from a mixed state simply by performing a measurement on the state. Note, however, that the second law of thermodynamics tells us that purity is indeed a resource. One can never decrease the entropy of a closed system; entropy only increases. The reason a measurement appears to produce pure states is that we ignore the fact that the measuring apparatus must have initially been set in some pure state, and after the measurement, the apparatus will be in a mixture of all the different measurement outcomes. In other words, in a closed system which includes the state, the measurement apparatus, and the observer, the total number of pure qubits can never increase. We must therefore be careful how we define the allowable class of operations in order to account for all pure states which might be introduced by various parties from the outside. We will discuss such a useful class, called *closed operations*, which can properly be used to account for pure states.

By considering pure states as a resource, one is immediately connecting quantum-information theory with thermodynamics. In fact, it was the early foundational work on reversible computation [21] where the entropic cost of computation was considered [22]. The relationship between information and physical tasks such as performing work also has a long history beginning with Szilard [23]. In fact, as shown in [15,24] the information function is exactly equal to the number of pure qubits one can extract from a state while having many copies of the state. We will therefore talk of extracting *information I* from a state. One can think of this as extracting pure states from more mixed states. From Szilard, we also know that the information I is closely related to the amount of work W one can extract from a single heat bath (see [25] for a rigorous derivation). Thus we sometimes talk of extracting work, or information, or purity from states. These connections will be discussed in Sec. II where we review the basic concepts.

The rough essence of the approach is that if separated individuals extract local pure states (i.e., information) from a shared state, using only local operations and classical communication, then they will in general be able to extract less information than if they were together. If the amount of information they can extract when they are together from a state ϱ is $I(\varrho)$ and the optimal [26] amount they can extract when separated is $I_l(\varrho)$, then the difference (called the *deficit*) $\Delta(\varrho) \equiv I(\varrho) - I_l(\varrho)$ feels some nonclassical correlations in the state ϱ.

Note that the quantity Δ is not an entanglement measure, at least in the regime of finite copies of a state ϱ. It feels not only entanglement, but also so-called *nonlocality without entanglement* [10]. We say that it quantifies *the quantumness of correlations* rather than entanglement (first attempts to formally quantify such features for quantum states are due to [11] and for ensembles in [10]). The state which has nonzero deficit we will call "informationally nonlocal." The term *nonlocality* means here that distant parties can do worse than parties that are together, despite the fact that they can communicate classically [10]. Thus it is a different notion than the nonlocality understood as a violation of local realism (we have discussed the relations in [27]).

In this work we review some of the results of [14–17,24] and provide more detail. We then give a number of new, essential results within the paradigm of distillation of local information. In particular we provide a lower bound for the deficit: it is bounded from below by the relative entropy of entanglement [28,29]. We also find that the closed LOCC (CLOCC) paradigm allows one to define the thermodynamical cost of erasure of entanglement. The cost is also bounded from below by the relative entropy of entanglement. We also analyze the deficit for multiparty pure states such as the Aharonov state [30], Greenberger-Horne-Zeilinger (GHZ) state [31], and W state. We obtain that, according to the deficit, the Aharonov state exhibits the greatest quantum correlations, while the GHZ state, the least. We show that in the finite regime (i.e., where Alice and Bob deal with a single copy of a state), any entangled state is *informationally nonlocal*; i.e., it has nonzero deficit. Moreover, we provide states which exhibit informational nonlocality even though they are separable and have an eigenbasis of distinguishable states—call it *nonlocality without entanglement but with distinguishability* (on the level of ensembles, it has its counterpart in [10]). We also provide many other interesting results, including the impossibility of catalysis with local pure states and the nonexistence of states whose entire informational contents is nonlocalizable.

The paper is organized as follows. In Sec. II an operational meaning of information is briefly recalled in terms of transition rates and basic laws of thermodynamics. In Sec. III, the idea of information as a resource in the distant laboratory paradigm is presented. Here the central notion of the present formalism—i.e., the *quantum-information deficit*—is defined. In Sec. IV the various aspects of the information deficit and its dual notion *localizable information* are discussed and an interpretation of the deficit in the context of quantum nonlocality is provided. Section V presents the deficit as the entropy production needed to reach the set of pseudoclassically correlated states. The concept is then generalized to an arbitrary set, including a set of separable states, and the cost of erasure of entanglement is defined. Section VI provides upper and lower bounds for the deficit in terms of the relative entropy distance and an upper bound for the entanglement erasure cost.

We next turn to exploring new phenomena which can be discovered using our methods. In Sec. VII, the main implications of the results of previous section are provided including the key conclusion that any entangled state is *informationally nonlocal* in a well-defined, natural sense. We also prove the existence of separable states which have a locally distinguishable eigenbasis, yet contain nonlocalizable information. Section VIII is devoted to a generalization to a multipartite case. Some of these results were briefly noted in [14]. Here the information deficit is calculated and the asymptotic behavior is analyzed for special examples of pure multipartite states: the GHZ state, W state, and Aharonov state. We find that the Aharonov state can be considered to be the most nonlocal. Section IX contains an exhaustive analysis of Bell states. In Sec. X we prove that (as opposed to pure nondistillable entanglement—i.e., the bound entanglement phenomenon) pure unlocalizable information does not exist. Section XI includes an analysis of the proportions of quantum and classical correlations in quantum states, addressing the question, can the first component exceed the second? In

Sec. XII zero-way and one-way subclasses of informational deficit are presented. It is shown that in the asymptotic regime, the one-way deficit is nonzero for separable (disentangled) states, stressing that quantum correlations are more than quantum entanglement. Section XIII discusses the relation of our measure to other measures of the quantumness of correlations; i.e., *one-way and two-way quantum discord* is discussed. Section XIV contains discussion of the result in the context of classical correlations measure introduced by other authors including the Henderson-Vedral measure. A discussion of complementarity between information quantities in distributed quantum systems is provided in Sec. XV. The paper closes with a general discussion of the results and a list of open questions in Sec. XVI.

II. INFORMATION: AN OPERATIONAL MEANING

Before turning to the case of parties who are in distant laboratories, it will prove worthwhile to discuss the notion of information from a more general perspective. Although we often talk about information as an abstract concept, here we use it as a term of art which refers to a specific function

$$I(\varrho) = \log_2 d - S(\varrho), \tag{1}$$

where $S(\varrho) = -\mathrm{tr}\varrho \log \varrho$ is the von Neumann entropy of ϱ acting on a Hilbert space of dimension d. We will usually work with qubits, in which case $\log d = N$ is an integer. As we will see in the next section, the information function has an operational meaning: it is the number of pure qubits one can draw from many copies of the state.

Let us now shortly discuss the information function (1) in the context of the more common Shannon picture. In the latter approach 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 which is known by the sender, but is not known by the receiver. The receiver treats the message as the information, if she did not know it. However, one can also consider an objective picture; a system represents information if it is in a pure state (zero entropy). We *know* what state it is in. The state is itself the information.

We obtain a picture where two kinds of information are dual. Shannon's entropy represents the information one *can get to know* about the system, while the information of Eq. (1) represents the information one *knows* about the system. Together they add up to a constant, which characterizes the system only (not its particular state):

$$I(\text{total}) = \log d = S(\rho) + I(\rho). \tag{2}$$

Note that the "objective" picture is more natural in the context of thermodynamics. There, a heat bath is highly entropic, and we are ignorant of exactly what state it is in. On the other hand, it is known that using pure states, one can draw work from a single heat bath using a Szilard heat engine [23]. The pure state represents information needed to order the energy of the heat bath. Knowing which side of a box the molecules of gas are in allows one to draw work by

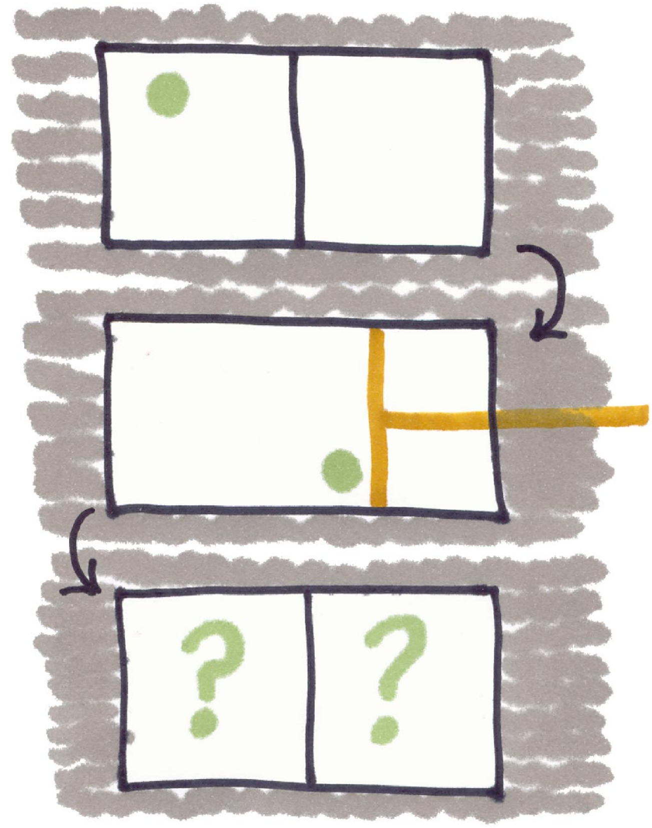

FIG. 1. Drawing work from a single heat bath using knowledge about the position of the molecule (the Szilard engine). In the first stage the molecule is known to be on the right-hand side. Next, a piston is inserted, and the molecule pushes it out, thus performing kT bits of work. After this stage, the position of the molecule is unknown, and we cannot use it to perform more work.

having the molecules push out a piston (see Fig. 1). High entropy of the gas implies ignorance of the molecule positions and an inability to draw work from the system. In general from a single heat bath of temperature T by use of a system in state ρ, one can draw amount of work (cf. [32])

$$W = kTI. \tag{3}$$

The process does not violate the second law because the information is depleted as entropy from the heat bath accumulates in the engine, and one cannot run a perpetual mobil. Thus a quantum system in a nonmaximally mixed state can be thought of as a type of fuel or resource. In fact, originally, our motivation for considering the function 1 in [14] was to understand entanglement in a thermodynamical context. We thus interchangeably speak of work W or information (purity) I.

A. Information and transition rates

In [15,24] it was shown that the function I has operational meaning in the asymptotic regime of many identical copies. It gives the number of pure states that one can obtain from a state ϱ under a certain class of operations we call noisy operations (NO's): operations that consist of (i) unitary transformations (ii) partial trace and (iii) adding ancillas in maximally mixed state. The motivation for considering such a class is that if we want to measure purity (i.e., information), as a resource to be counted, then we should restrict the class of operations so as not to allow pure states being added for free. For example, if our class of operations allowed the creation of pure states out of nothing, it would be impossible to measure how much information (purity) could be extracted from a state, because an infinite amount could always be created. Thus, for example, we only allow adding maximally mixed states with maximal entropy, since with these alone, pure states cannot be created.

Having defined the class of operations one can show that it is the unique function (up to constants) that is not increasing under the class of NO's. One then shows that I determines the optimal rate of transitions between states under NO's. Let us now discuss two special cases.

First, given n copies of state ϱ one can obtain $nI(\varrho)$ qubits in a pure state. This is done essentially by quantum data compression [33] (cf. [34,35]). In data compression, one keeps the signal and discards the qubits which are in the pure state. Here we do the opposite. We discard the "signal," treating it as noise, and keep instead the redundancies (which are in pure state). Thus we obtain pure states. This is essentially like cooling [36]. The protocol does not require using noisy ancillas (e.g., maximally mixed states).

A second protocol of interest is that one can take $n[N-S(\varrho)]$ pure qubits and produce n copies of ϱ. The protocol, described in [24], takes pure states and dilutes them using ancillas in the maximally mixed state (noise). The existence of such dual protocols is similar to entanglement concentration and dilution [37]. And similarly as in [38,39], this can be used to prove that there is a unique function that does not increase under the NO class of operations.

Note that for K pure qubits, the information I is equal to K. For the maximally mixed state $I=0$. As mentioned, I is monotonically decreasing under partial trace and adding ancillas in the maximally mixed state. It is of course constant under unitary operations. The property that makes it a unique measure of information in the asymptotic regime is *asymptotic continuity* (see [39–41]) which means that if two states are close to each other, then so is their informations per qubit. It is important to remember that I is not expansible; i.e., if we embed the state into larger Hilbert space, then it changes (because the number of qubits increases). The reason is obvious even within the classical framework: if there are two possible states of the system, knowledge of the state represents less information than knowledge of the state in the case of, say, three possible configurations. It is in contrast with entanglement theory where a pure state of Schmidt rank 2 means always the same thing, independently of how large the system is. Also the entropy of the state depends only on nonzero eigenvalues: e.g., the entropy of a pure state is zero, independently of how large the system is. However, in the present case, the Hilbert space and its dimension are important elements in our considerations.

B. Information in the context of "closed operations"

In the previous section we argued that the information function gives transition rates from the mixed state to pure and backwards, and that it gives uniqueness of information in the context of NO's. For the rest of this paper we will not treat additional mixed states as a free resource. Thus let us now discuss the meaning of information in the context of a class that is compatible with the class of operations which we will use in the case of distributed systems further in this paper. Namely, we can consider *closed operations* (CO's). They are arbitrary compositions of the following two basic operations: (i) unitary transformations and (ii) dephasing $\rho \to \Sigma_i P_i \rho P_i$ where $\Sigma_i P_i = I$, and P_i are projectors not necessarily of rank one.

We call the class closed, though it is not actually fully closed. Information cannot go in, but can go out (via dephasing). The name closed is motivated by the fact that the number of qubits is the same, and the qubits cannot be exchanged between the system of interest and environment. The only allowed contact with the environment is decoherence caused by operation (ii). As with NO operations, the CO class is motivated by wanting to quantify purity (i.e., information and entropy). Namely, just as in thermodynamics, if we want to consider entropy, then we must isolate our system. Thus the class does not allow one to bring in pure states for free and thus allows us to count the amount of purity which is contained in a given system. In the next section we will introduce this "closed" paradigm to the distant laboratory scenario, by use of which we will define the quantum deficit.

Now, let us ask about drawing pure qubits out of a given state by the present class of operations. The operations do not change the size of the system, so that when we start, e.g., with many copies of the state ρ, we cannot end up with a smaller system in almost a pure state. However, this is not a big problem. Imagine for a while that in addition we can apply a partial trace (which is not allowed in CO's). Then the process of drawing pure qubits can be divided into two stages: (1) some CO operations aiming to concentrate the pure part into some number of qubits and (2) partial trace of the remaining qubits.

Since we do not allow for partial trace, one can simply stop before tracing out. The obtained state will have a form of (approximate) product of qubits in a pure state and the rest of the system—some garbage. Thus the process of dividing a system into a pure part and garbage we can treat as extraction pure qubits.

Now, let us ask how many pure qubits can be drawn from a state by closed operations in the above sense? Actually, the process of drawing qubits by NO's did not use maximally mixed states. It was just a unitary operation, plus partial trace. Thus we can apply this operation (unitaries are allowed in CO's) and get again I qubits per input states. Thus also within the "closed picture" information has the same interpretation of a maximal amount of pure qubits that can be obtained from a state per input copy by closed operations.

III. RESTRICTING THE CLASS OF OPERATIONS IN THE DISTANT LABORATORY PARADIGM: CLOCC AND THE INFORMATION DEFICIT

In the preceding section, we discussed the notion of information from the perspective of being able to reversibly distill pure states from a given state ϱ. Now, one can ask about how things change when the allowable class of operations one can perform is somehow restricted. This is a rather general question, but since here we are interested in understanding entanglement and nonlocality, we will examine the restricted class of operations which occurs when various parties hold some joint state, but are in distant laboratories. One then imagines that Alice and Bob wish to distill as many locally pure states as possible—i.e., product pure states such as $|0^{\otimes m_A}\rangle_A \otimes |0^{\otimes m_B}\rangle_B$. The amount of local information which is distillable we call I_l.

In the ordinary approach to the distant laboratories paradigm, one imagines that two parties (Alice and Bob) are in distant laboratories and can only perform local operations and classical communication (LOCC). However, as we noted, this class of operations is not suitable to deal with the questions of concentration of information to local form. That is because under LOCC, one does not count the information that gets added to the systems through ancillas, measuring devices, etc. We thus have to state the paradigm more precisely. Since we are interested in *local information*, we must treat it as a resource, assuming it cannot be created, but only manipulated. Once we have a compound state, the task is to localize the information by using classical channel between Alice and Bob.

The new paradigm was introduced in Ref. [14] where one essentially looks at a closed system as one does in thermodynamics when calculating changes in entropy. One imagines that Alice and Bob are in some closed box, which does not allow them to import additional quantum states, except for ones which we specifically keep track of and account for.

In defining a class of operations, the crucial point is that here, unlike in usual LOCC (local operations and classical communication) schemes, one must explicitly account for all entropy transferred to measuring devices or ancillas. So in defining the class of allowable operations one must ensure that no information loss is being hidden when operations are being carried out. Moreover, the operations should be general enough to represent faithfully the ultimate possibilities of Alice and Bob to concentrate information. In other words, we would not like to introduce any limitation apart from two basic ones: (i) there is a classical channel between Alice and Bob and (ii) local information is a resource (cannot be increased).

We consider a state ϱ_{AB} acting on Hilbert space $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$. Let us first define the elementary allowable elements of closed LOCC operations.

Definition 1. By CLOCC operations on bipartite system of n_{AB} qubits we mean all operations that can be composed out of (i) local unitary transformations and (ii) sending subsystems down a completely decohering (dephasing) channel.

The latter channel is of the form

$$\varrho_{in} \to \varrho_{out} = \sum_i P_i \varrho P_i, \tag{4}$$

where P_i are one-dimensional projectors. For a qubit system, it acts as

$$\varrho_{in} = \begin{bmatrix} \varrho_{11} & \varrho_{12} \\ \varrho_{21} & \varrho_{22} \end{bmatrix} \to \varrho_{out} = \begin{bmatrix} \varrho_{11} & 0 \\ 0 & \varrho_{22} \end{bmatrix}. \tag{5}$$

It is understood that ϱ_{in} is at the sender's site, while ϱ_{out} is at the receiver's site. The operation (ii) accounts for both local measurements and sending the results down a classical channel. It can be disassembled into two parts: (a) local dephasing (at, say, the sender site) and (b) sending a qubit intact (through a noiseless quantum channel) to the receiver. Thus suppose that Alice and Bob share a state $\varrho_{AB} \equiv \varrho_{A'A''B}$, and Alice decided to send subsystem A'' to Bob, down the dephasing channel. The following action will have the same effect. Alice dephases locally the subsystem A'':

$$\varrho_{A'A''B} \to \sum_i P_i^{A''} \otimes I_{A'B} \varrho_{A'A''B} P_i^{A''} \otimes I_{A'B}. \tag{6}$$

The state is now of the form

$$\varrho^{out}_{A'A''B} = \sum_i p_i P_i^{A''} \otimes \varrho_i^{A'B}. \tag{7}$$

Thus part A'' is classically correlated with the rest of the system (it is stronger than to say that the state is separable with respect to $A'':A'B$). Now Alice sends system A'' to Bob through an ideal channel. Thus the final state differs from the state $\varrho^{out}_{A'A''B}$ only in that system A'' is at the Bob site. It follows that operation 1 can be replaced by the following two operations: (iia) local dephasing and (iib) sending a completely dephased subsystem.

Note that operations (i) and (iib) are reversible. Only operation (iia) can, in general, be irreversible. Actually it is irreversible if only it *changes* the state—i.e., in all nontrivial cases. Note also that the operations do not change the dimension of the total Hilbert space or, equivalently, the number of qubits of the total system, even though the particular qubits can be reallocated; for example, at the end all qubits can be at Alice's site.

Let us finally note that it may happen that after the protocol, one of the parties will be left without any system at all, as everything has been sent to the other parties. It is only the total number of particles which is conserved.

Comparison with other classes of operations

For the purpose of the present paper, we will use solely CLOCC operations. Yet we have also found it useful to use another class of operations; thus, we will describe the other class and compare it with CLOCC.

Let us first present the other (likely equivalent) class of operations, called noisy LOCC (NLOCC). The relation between NLOCC and CLOCC will be similar to the relations between NO and CO: the elementary operations will be the same as in CLOCC, plus tracing out local systems and adding maximally mixed ancillas.

Definition 2. By NLOCC operations on bipartite system of n_{AB} qubits we mean all operations that can be composed out of (i) local unitary transformations, (ii) sending the subsystem down the completely decohering (dephasing) channel, (iii) adding ancilla in the maximally mixed state, and (iv) discarding the local subsystem.

As in CLOCC we can decompose (ii) into (iia) and (iib). CLOCC operations are more basic than NLOCC. Namely, the latter can be treated as CLOCC with an additional resource: an unlimited supply of maximally mixed states (which have zero informational content). Indeed, similarly as in Sec. II B one can argue that the operation of the local partial trace is not essential.

IV. LOCALIZABLE INFORMATION AND INFORMATION DEFICIT

In this section we define the central quantity: *the information deficit*. To this end we will first introduce the notion of localizable information. To be more precise, we will first deal with the single-copy case and define basic quantities on this level. Then we will discuss the asymptotic regime, which will require regularization of the quantities.

Definition 3. The localizable information $I_l(\varrho_{AB})$ of a state ϱ_{AB} on Hilbert space $C^{d_A}\otimes C^{d_B}$ is the maximal amount of local information that can be obtained by CLOCC operations. More formally,

$$I_l(\varrho_{AB}) = \sup_{\Lambda\in CLOCC}[I(\varrho'_A) + I(\varrho'_B)], \qquad (8)$$

where $\varrho'_{AB}=\Lambda(\varrho_{AB})$, I is the information function $I(\varrho'_X) = N'_X - S(\varrho'_X)$; $N'_A=\log d'_A$, $N'_B=\log d'_B$ are the number of qubits of subsystems of the output state. When one of the numbers of qubits is zero (null subsystem) we apply the convention that information is zero.

Alternatively, we have the formula

$$I_l(\varrho_{AB}) = N - \inf_{\Lambda\in CLOCC}[S(\varrho'_A) + S(\varrho'_B)], \qquad (9)$$

where N is the total number of qubits. Again, if it happens that all particles are with one party (i.e., the output dimension is equal to 1) so that the subsystem of the other party is null, then we apply the convention that the entropy of such a subsystem is zero. Further states on the system with one subsystem null we will call null-subsystem states.

It is important here, that "to obtain local information" does not mean as usual getting some outcomes of local measurements. Rather it means to apply such an operation, after which information, as a function of states of subsystems, will be maximal. Thus, we only deal with *state changes* and calculate some function (information function) on the states.

Actually it is not localizable information which will be the most important quantity. Rather, the central quantity is a closely connected one, which we call the *quantum-information deficit* (in short quantum deficit). It is defined as a difference between the information that can be localized by means of CLOCC operations and total information of the state.

Definition 4. The quantum deficit $\Delta(\varrho_{AB})$ of a state ϱ_{AB} on Hilbert space $C^{d_A}\otimes C^{d_B}$ is given by

$$\Delta(\varrho_{AB}) = I(\varrho_{AB}) - I_l(\varrho_{AB}). \qquad (10)$$

Using the definition of localizable information I_l, we get an alternative formula for the quantum deficit:

$$\Delta = \inf_{\Lambda\in CLOCC}[S(\varrho'_A) + S(\varrho'_B)] - S(\varrho_{AB}), \qquad (11)$$

where $\varrho'_{AB}=\Lambda(\varrho_{AB})$.

It is important to note that both quantities are functions not only of a state but also the dimension of the Hilbert space. This is because CLOCC operations are defined for a fixed Hilbert space. That I_l depends on the dimension of the Hilbert space is even more obvious, because the latter is explicitly written in the formula. However, in the formula for the deficit as written in Eq. (11), the dimension does not appear explicitly, so it could happen that there is no dependence on dimension. Actually, it is rather important that Δ does not actually depend on dimension; i.e., when one locally increases Hilbert space by, e.g., adding a qubit in a pure state, Δ should not change. This is because, as we will see later, the deficit will be interpreted as a measure of the quantumness of correlations, which should not change upon adding local ancilla. We will discuss this issue later in more detail. In particular in Sec. X we will show that regularization of the deficit *does not change* upon adding local ancilla in a pure state.

A. Interpretation of the quantum deficit: Measure of "informational nonlocality"

A nonzero deficit means that Alice and Bob are not able to localize all the information contained within the state. This, however, means that part of the information is necessarily destroyed in the process of localizing by use of classical communication. This part of information cannot survive traveling classical channels. It implies that it must be somehow quantum. In addition, this part of information must come from correlations, since information that is not in correlations is already local and need not be localized. We could say that the quantum deficit quantifies quantum correlations. However, we will see that in the regime of single copies, the quantum deficit can be (and often is) nonzero for separable states, which can be generated by local quantum actions and solely classical communication. It is not clear then if we can talk here about quantum correlations—can quantum correlations be created by only classical communication between the parties? However, the quantum deficit being nonzero indicates that there is something quantum in correlations of the state. One can say that these are *classical* correlations of *quantum* properties. We will then propose to interpret the quantum deficit as the amount of "quantumness of correlations." Whether it represents also quantum correlations when regularized is still open.

Let us now discuss the issue in the context of a notion of nonlocality considered by [10]. The authors exhibited ensembles of product states which are fully distinguishable if globally accessed, but cannot be perfectly distinguished by distant parties that can communicate only via classical channels. Then they called this effect *nonlocality without entanglement*. The reason for using term "nonlocality" was the following: one can do better if the system is accessible as a whole, rather than when it is accessible by local operations and classical communication.

In our case, the situation is similar: Alice and Bob can do better in distilling local information if they have two subsystems at the same place rather than shared in distant laboratories. Thus, we have a similar kind of nonlocality, and the quantum deficit is a measure of such nonlocality, which we can call "informational," as it concerns the difference in access to informational contents. Thus, any state with a nonzero deficit will be called informationally nonlocal (or nonlocal, when the context is obvious).

B. Classical information deficit of quantum states

It is important to investigate not only the "quantumness" of compound quantum states, but also the relationships between their "classical" and "quantum" parts. To this end consider the quantity I_{LO}—the information that is local from the very beginning—i.e.,

$$I_{LO} = N - S(\rho_A) - S(\rho_B). \tag{12}$$

We will call it local information. It represents how much information each party can extract if only local operations are performed. We can now define an analogous quantity to the quantum deficit, on a "lower" level.

Definition 5. The classical deficit [16] of a quantum state is the difference between local information and the information that can be obtained by CLOCC (i.e., by localizable information):

$$\Delta_c = I_l - I_{LO}. \tag{13}$$

This tells us how much more information can be obtained from the state by exploiting additional correlations in the state ρ_{AB}. Since these correlations are exploitable using a classical channel, this quantity tells us something about classical correlations. We will refer to Δ_c as the *classical deficit*. Also, as we will see later, the quantity can be used in the context of the quantifying of classical correlations (though it is not immediate; see [42]).

C. Restricting resources: Zero-way and one-way subclasses

Additional measures of the quantumness of correlations which arise when one restricts the communications between Alice and Bob are as follows.

One can define the one-way (Alice to Bob) deficit ($\Delta^{\rightarrow}$) and one-way (Bob to Alice) information deficit ($\Delta^{\leftarrow}$) by restricting the classical communication to only be in one direction. Furthermore, one has also a zero-way deficit ($\Delta^{\emptyset}$). The name *zero way* is perhaps confusing. It refers to the situation where no communication is allowed between Alice and Bob until after they have completely dephased (or performed measurements of) their systems. After they have done this, they may then communicate in order to exploit the (what are now) purely classical correlations in order to localize the information. These restricted deficits correspond to locally accessible information $I_l^{\rightarrow}$, $I_l^{\leftarrow}$, and $I_l^{\emptyset}$.

D. Asymptotic regime: Distillation of local information as a dual picture to entanglement distillation

In this section we will argue that the idea of localization of information, though at a first glance exotic, can be recast in terms typical for quantum information theory, where of central importance are manipulations over resources. Even more, our present formulation will be analogous to the scheme which is a basis for entanglement theory: entanglement distillation. We will use the interpretation of the information function as the amount of pure qubits one can draw from a state in the limit of many copies.

Instead of singlets our precious resource will be a pure local qubit. The aim of Alice and Bob is, given many copies of state ϱ_{AB} to distill the maximal amount of local pure qubits by means of CLOCC operations (in entanglement theory, we had LOCC operations; however, here we need CLOCC; otherwise, one could add for free states, and the maximal distillable amount of pure local qubits would be infinite). One way of doing that is the following: Alice and Bob take state ϱ_{AB}, apply the CLOCC protocol that optimizes the formula for localizable information—i.e., they obtain state ϱ'_{AB} which has maximal local information I'_A and I'_B. They apply such a protocol to every copy of the state they share. As a result they obtain many copies of state ϱ'_{AB}. Now, Alice in her laboratory, can apply a protocol of drawing pure qubits out of her state $\varrho_A'^{\otimes n}$, obtaining I'_A pure qubits. The same does Bob. Finally, they possess $I'_A + I'_B$ pure local qubits which is equal just to localizable information, and actually it is the best they can do, when acting first on single copies using communication, and only locally performing collective actions on many copies.

Alice and Bob could do better when they act collectively from the very beginning. In this way we get that the optimal amount of local pure qubits that can be distilled by CLOCC is equal to regularization of localizable information:

$$I_l^{\infty} = \lim_n \frac{I_l(\rho^{\otimes n})}{n}. \tag{14}$$

Similarly we can define the regularized quantum and classical deficits

$$\Delta^{\infty} = \lim_n \frac{\Delta(\rho^{\otimes n})}{n}, \quad \Delta_c^{\infty} = \lim_n \frac{\Delta_c(\rho^{\otimes n})}{n}. \tag{15}$$

Thus we conclude that regularizations of our quantities have operational meaning connected to the amount of pure local qubits which can be distilled out of a large number of copies of the input state by means of different resources (global operations, CLOCC, local operations). Let us emphasize here that when Alice and Bob are given a single copy of state, they usually cannot distill pure qubits. When they are given many copies, the ultimate amount of distillable pure qubits is described by regularized I_l. Thus the nonregularized quantity does not represent the amount of pure qubits that can be drawn either from single copy or from many copies. However, since in the definition of I_l there is an information function that has operational asymptotic meaning, then I_l also *has* some *asymptotic interpretation*, representing the amount of pure local qubits that can be drawn when at the stage of communication, Alice and Bob operate on single copies, and only after that stage operates collectively.

In entanglement theory, there is a similar situation with entanglement of formation and entanglement cost. The first is not the ultimate cost of producing a state out of singlets, though it already contains "some asymptotics" by definition—the von Neumann entropy, which is the asymptotic cost of producing pure states out of singlets. The ultimate cost of producing states out of singlets is the regularization of entanglement of formation.

Finally, one can also consider the amount of local information that can be distilled by means of one-way classical communication. It is equal to the regularized one-way quantum deficit $\Delta^{\rightarrow}$. In a similar vein we can consider regularizations of other quantities based on restricted resources, such as $\Delta_c^{\rightarrow}$, $\Delta^{\emptyset}$, $\Delta_{cl}^{\emptyset}$, etc. Again, all those regularizations have operational meaning.

E. Additional local resources

One of the basic features of the paradigm is that adding local ancillas is not for free. The reason is that, otherwise, all the quantities would become trivial. However, there are two kinds of local resources that still can be taken into account.

First of all, we can allow adding for free local ancillas in a maximally mixed state. Thus given a state ϱ_{AB} we can ask, what about the quantities of interest for the state $\varrho_{AB} \otimes I_{A'}/d$? Note here that this would mean that I_l^{∞} does not change if we use the NLOCC class instead of CLOCC. Indeed, as have already mentioned, the only difference between two classes for the problem of distillation of local information may appear when adding local maximal noise could help. In general, upon adding such local noise, localizable information could only go up. However, it is more likely that it will not change. In fact, Devetak has shown [20] that the one-way deficit does not change upon adding noise. We were not able to show the same in the case of two-way communication, though we believe it is also the case.

The second possibility is *borrowing* local pure qubits. This would be the most welcome, as it would mean that the deficit does not depend on the dimension of the Hilbert space as discussed in the introduction of Sec. IV. We actually show that it is the case for the regularized deficit in Sec. X. For the one-way case it is shown also in the asymptotic regime in [20].

There is a more general possibility: borrowing local ancilla in any mixed state. However, in the asymptotic limit, this is actually equivalent to borrowing noise and pure qubits, as in that regime any state can be reversibly composed out of noise and pure qubits [15,24].

F. An example: Pure states

As we have mentioned, in our definition of quantum correlations, we do not speak about entanglement at all. We do not work in the established paradigm of the optimal rate of transformation to or from maximally entangled states [43]. We consider distillation of pure product states. Thus, it was perhaps surprising to find [14] that for pure states, this definition of the quantumness of correlations is just equal to the unique asymptotic entanglement for pure states [37,43].

We shall now see that by taking as an example the Bell state

$$|\psi^-\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle). \tag{16}$$

It is a two-qubit state of zero entropy, so its informational content, as given by Eq. (1), is $I=2$. We will now see that $I_l=1$. Clearly, without communicating, neither party can draw any information from the state, since locally, the state is maximally mixed. It turns out that the best protocol is for Alice to send her qubit down the dephasing channel. After she has done this, Bob will hold the classically correlated state

$$\rho_{CC} = \frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 11|) \tag{17}$$

from which one can extract 1 bit of information by performing a CNOT gate to extract one pure state $|0\rangle$. We thus have that $\Delta=1$. One can actually view this process in terms of measurements and classical communication, as long as we keep track of the measuring device. Alice performs a measurement on the state to find out if she has a $|0\rangle$ or $|1\rangle$. She then tells Bob the result. Bob now holds a known state, without having to perform any measurement. Alice, on the other hand, had to perform a measurement to learn her state. The informational cost of the measurement is 1 bit since a measuring apparatus is initially in a pure state and must have two possible outcomes. After the measurement, the measuring device needs to be reset. The classical state-correlated state ρ_{CC}, if held between two parties, has $\Delta=0$. That the process is optimal for the singlet state is obvious, as this is actually the only thing which Alice and Bob can do given a single copy. However, it is highly nontrivial to show that the regularization of a I_l is still the same. The optimality of this protocol also in many copy case was shown in [15]. It also follows from the general theorem we give in this paper, which connects the deficit with the relative entropy distance from some set of states.

In general, it is not hard to see that for an arbitrary pure state, the same protocol can be used with Alice first performing local compression on her state. For any pure state $|\psi\rangle_{AB}$, the two-way Δ is given by [14,15]

$$\Delta(|\psi\rangle) = S(\mathrm{tr}_A(|\psi\rangle\langle\psi|)).$$

Thus, for pure bipartite states, the quantum deficit is equal to entanglement. It is quite interesting that we have obtained entanglement by *destroying* entanglement.

V. DEFICIT AS THE PRODUCTION OF ENTROPY NECESSARY TO REACH PSEUDOCLASSICALLY CORRELATED STATES

In this section we will show that the quantum deficit can be interpreted as the amount of entropy one has to produce in the process of transforming a given state into a so-called pseudoclassically correlated state [14]. This expression of deficit makes it possible to define the entropy production connected with a given subset of states. For example, we can then speak about the entropy production needed to reach the set of separable states. In this way our paradigm provides a consistent definition of the *thermodynamical cost of erasure of entanglement*, while the original deficit can be called the *thermodynamical cost of erasing quantum correlations*.

A. Important classes of states

Let us first define sets of states which are important for our analysis. Notice that in place of a simple dichotomy between separable and entangled states [44], one can have a whole hierarchy of levels of quantumness [45]. Already Werner recognized [44] that within entangled states there might be ones that do not violate Bell's inequalities (cf. [46,47]). One may also go in a *converse* direction and within separable states find a subclass which is most classical and wider classes which are still somehow classical, though in some sense to a lesser degree (cf. [11]).

First, let us consider a set if states which we choose to call *properly classically correlated* or, shortly, classically correlated. These are states of the form

$$\varrho = \sum_{ij} p_{ij}|i\rangle\langle i| \otimes |j\rangle\langle j|, \tag{18}$$

where $\{|i\rangle\}$ and $\{|j\rangle\}$ are local bases. Thus any such state is the classical joint probability distribution naturally embedded into a quantum state. Note that the set of classically correlated states is invariant under local unitary operations. The states are diagonal in a special product basis, which can be called the *biproduct basis*.

Now let us define the set of states of our central interest. We will call them *pseudoclassically correlated states* and denote them by $\mathcal{PC}$. These are the states that can be *reversibly* transformed into classically correlated ones by CLOCC. "Reversibly" means that no entropy is produced during the protocol. This implies that no dephasing is needed in transformations: Alice and Bob use only unitaries and sending subsystems such that dephasing does not change the total state. Thus they can send only such subsystems X, which are in the following state with the rest R: $\rho_{XR}=\Sigma_i p_i |i\rangle_X\rangle i| \otimes \rho_i^R$. The states that can be in such a way transformed into classical ones can be also described as the set of states which Alice and Bob can create under the allowed class of operations (CLOCC) out of classical states. The eigenbasis of these states was called an *implementable product basis* (IPB) in [15], since it is the eigenbasis that Alice and Bob are able to dephase in.

Let us note that one can have an intermediate class, *one-way classically correlated* states, which are of the form

$$\varrho = \sum_{ii} p_{ii}|i\rangle\langle i| \otimes \varrho_i. \tag{19}$$

These are states which can be produced out of classically correlated states by one-way reversible CLOCC. They are diagonal in basis which is of the form $\{|i\rangle|\psi_k^{(i)}\rangle\}$ where $\{|\psi_k^{(i)}\rangle\}$ are bases themselves.

The above sets are *proper* subsets of separable states, and all the inclusions between them are proper too.

B. Formula for the quantum deficit in terms of pseudoclassically correlated states

Any protocol of attaining the information deficit looks as follows: Alice chooses a subsystem of her system, dephases it, and then sends it to Bob. Bob then chooses a subsystem from his system (which now includes his original system and the system sent by Alice). He dephases his chosen part and sends it to Alice. They can send the states using an ideal channel, as the sent subsystems are already dephased. Thus sending is here only reallocating subsystems, nothing more. Alice and Bob continue such a process as long as they wish. When they decide to stop, the final step is ρ' and the obtained local information is equal to $N-S(\rho_A')-S(\rho_B')$ while the initial total information was $I=N-S(\rho_{AB})$. Thus the deficit obtained in a particular protocol $\mathcal{P}$ is $\Delta_{\mathcal{P}}=S(\rho_A')+S(\rho_B')-S(\rho_{AB})$. Alice and Bob wish this quantity to be minimal. Suppose then that they preformed an optimal protocol, for which indeed this value is minimal.

There are two cases: (i) one of subsystems is null (all particles with the other party) or (ii) both parties have subsystems that are not null. Note that in the second case the system must be in a product state. Suppose it is not. Then, Alice and Bob can dephase the state in the eigenbasis of states of local subsystems. This will not change local entropies, but will transform the state into a classically correlated one. Then Alice can send her part to Bob, so that the information contents of the total state will be unchanged. However, if only the state was nonproduct, the total information was greater than the sum of local information. This means that the protocol was not optimal, so that we have contradiction.

Thus we conclude that the optimal protocol ends up with either product state or state of a system, which one of the subsystems is null (all particles either with Bob or with Alice). Even more, when a state is a product, one of the subsystems can be sent to the other party, so that the whole system is with one party. This is compatible with the philosophy of "localizing" of information.

However, it turns out that we can divide the total process of localizing of information into two stages:

(i) Irreversible stage: transforming input state ϱ into some pseudoclassically correlated one ϱ'.

(ii) Reversible stage: localizing information of the state ϱ'.

In the first stage Alice and Bob try to produce the least entropy. The amount of information that they are able to localize is determined by this stage. In second stage, the entropy is not produced, and the information is constant.

We have the following proposition.

Proposition 1. The quantum deficit is of the form

$$\Delta = \inf_{\mathcal{P}}[S(\rho') - S(\rho)], \tag{20}$$

where the infimum is taken over all CLOCC protocols that transform initial state ρ into pseudoclassically correlated state ρ'.

Proof. The proof actually reduces to noting that pseudoclassically correlated states can be reversibly created from states with one null system. Simply, by definition

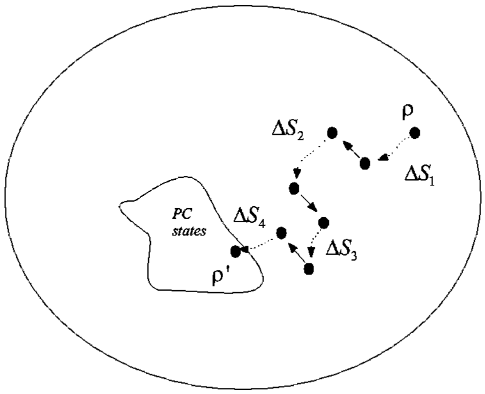

FIG. 2. CLOCC protocol of concentration of information to local form is a series of actions aiming to reach the set of pseudoclassically correlated states. The solid lines denote reversible actions: sending dephased qubits or local unitary transformations. The dotted lines denote dephasings. The goal is to make the total entropy increase $\Delta S=\Delta S_1+\Delta S_2+\cdots$ minimal. Then the deficit is given by $\Delta=\Delta S$, because once the state is pseudoclassically correlated, its full information content can be localized.

pseudoclassically correlated states can be reversibly produced out of classically correlated states. The latter, in turn, can be reversibly produced out of one-subsystem states. Thus, consider an optimal protocol for drawing local information. As we have argued, it can end up with a one-subsystem state. Out of the state we can reversibly create a classically correlated state which is a special case of pseudoclassically correlated states. Conversely, suppose that we have a protocol that ends up with a pseudoclassically correlated state. Then one can reversibly transform it into a one-subsystem state. □

Thus the quantum deficit is equal to *minimum entropy production* during the process of making the state be pseudoclassically correlated by CLOCC operations. In other words, to draw an optimal amount of local information from a given state, one should try to make it a pseudoclassically correlated state in the most gentle way—i.e., producing the least possible amount of entropy. Once the state is pseudoclassically correlated, the further process of the localization of entropy is trivial. The first stage is illustrated in Fig. 2.

C. Defining the cost of erasing entanglement

The above formulation of the deficit allows one to generalize the idea of the thermodynamical cost to other situations. Namely, instead of the set of pseudoclassically correlated states one can take any other set and ask the same question: how much entropy must be produced, while reaching this set by use of CLOCC. Thus our concept of localizing information allows us to ascribe thermodynamical costs to other tasks than localizing information. With any chosen set we can associate a suitable deficit Δ_{Set}. An important application of this concept is to take a set of separable states. Then the associated deficit Δ_{sep} has an interpretation of the *thermodynamical cost of erasing entanglement*. As such it is a good candidate for an entanglement measure. In this paper we will show that it is bounded from below by the relative entropy of entanglement. Since the set of separable states is a superset of pseudoclassically correlated states, we have

$$\Delta_{sep} \leqslant \Delta, \tag{21}$$

so that the cost of erasing all quantum correlations is no smaller than the cost of erasing entanglement. For the sake of further proofs, let us put here a formal definition of Δ_{sep}.

Definition 6. The thermodynamical cost of erasing entanglement Δ_{sep} is given by

$$\Delta_{sep}(\rho) = \inf_{\mathcal{P}}[S(\rho') - S(\rho)], \tag{22}$$

where the infimum runs over all CLOCC protocols $\mathcal{P}$ which transform initial state ρ into a separable output state ρ'

VI. RELATIONS BETWEEN THE DEFICIT AND RELATIVE ENTROPY DISTANCE

In this section we will present the proof of the theorem relating the deficit to the relative entropy distance obtained in [15].

Theorem 1. The information deficit is bounded from above by the relative entropy distance from the set of pseudoclassically correlated states:

$$\Delta(\varrho_{AB}) \leqslant \inf_{\sigma\in\mathcal{PC}} S(\varrho_{AB}|\sigma) \equiv E_r^{\mathcal{PC}}, \tag{23}$$

where $S(\varrho\,|\,\sigma)=\mathrm{tr}\varrho\log\varrho-\mathrm{tr}\varrho\log\sigma$.

Let us first prove the proposition.

Proposition 2. Localizable information and the deficit satisfy the following bounds:

$$I_l(\varrho) \geqslant N - \inf_{\mathcal{B}\in IPB} H(\varrho,\mathcal{B}), \tag{24}$$

$$\Delta(\varrho) \leqslant \inf_{\mathcal{B}\in IPB} H(\varrho,\mathcal{B}) - S(\varrho), \tag{25}$$

where $H(\rho,\mathcal{B})$ denotes the entropy of diagonal entries of state ρ in basis $\mathcal{B}$,

$$H(\rho,\mathcal{B}) = -\sum_i p_i \log p_i, \tag{26}$$

with $p_i=\langle\psi_i|\rho|\psi_i\rangle$, with $\psi_i\in\mathcal{B}$.

Proof. We will exhibit a simple protocol to achieve a reasonable (and perhaps optimal) amount of local information. Namely, Alice and Bob choose some implementable basis $\mathcal{B}$ and dephase a state in such a basis. They can do this, as, by definition, an IPB is a basis in which Alice and Bob can dephase by use of CLOCC. The final state has entropy

$$S(\rho') = H(\rho,\mathcal{B}). \tag{27}$$

Alice and Bob can now choose the basis that will produce the smallest possible entropy $H(\rho,\mathcal{B})$. In this way we obtain the following bound for Δ:

$$\Delta \leqslant \inf_{\mathcal{B}\in IPB} H(\rho,\mathcal{B}) - S(\rho). \tag{28}$$

This ends the proof of proposition. □

Let us now express this bound in terms of the relative entropy distance. This is done by the following lemma.

Lemma 1. Given a state ϱ,

$$H(\varrho,\mathcal{B}) = \inf_{\sigma\in\mathcal{S}_{\mathcal{B}}} S(\varrho|\sigma) + S(\varrho), \tag{29}$$

where $H(\varrho,\mathcal{B})$ is the Shannon entropy of the probability distribution of the outcomes when ϱ is measured in a given basis $\mathcal{B}$ and $\mathcal{S}_{\mathcal{B}}$ is the set of all states with eigenbasis $\mathcal{B}$.

Proof. We have

$$\begin{aligned}\inf_{\sigma\in\mathcal{S}_{\mathcal{B}}} S(\varrho|\sigma) + S(\varrho) &= \inf_{\sigma\in\mathcal{S}_{\mathcal{B}}}[-\operatorname{tr}(\varrho\log_2\sigma)] \\ &= -\operatorname{tr}(\varrho_{\mathcal{B}}\log_2\varrho_{\mathcal{B}}) + \operatorname{tr}(\varrho_{\mathcal{B}}\log_2\varrho_{\mathcal{B}}) \\ &\quad + \inf_{\sigma\in\mathcal{S}_{\mathcal{B}}}[-\operatorname{tr}(\varrho_{\mathcal{B}}\log_2\sigma)] \\ &= S(\varrho_{\mathcal{B}}) + \inf_{\sigma\in\mathcal{S}_{\mathcal{B}}} S(\varrho_{\mathcal{B}}|\sigma) = H(\varrho,\mathcal{B}).\end{aligned} \tag{30}$$

Here $\varrho_{\mathcal{B}}$ is the state ϱ dephased in the basis $\mathcal{B}$. In the second equality, we have used the fact that $\operatorname{tr}(\varrho\log_2\sigma) = \operatorname{tr}(\varrho_{\mathcal{B}}\log_2\sigma)$, because σ is diagonal in the basis $\mathcal{B}$. In the fourth equality, we have used that $\varrho_{\mathcal{B}}$ belongs to the set $\mathcal{S}_{\mathcal{B}}$ so that $\inf_{\sigma\in\mathcal{S}_{\mathcal{B}}} S(\varrho_{\mathcal{B}}|\sigma)=0$ and also that $S(\varrho_{\mathcal{B}})\equiv H(\varrho,\mathcal{B})$. This ends the proof of the lemma. □

Now combining the lemma with the proposition we obtain the above theorem. We have not been able to prove equality, and in Sec. VI C we discuss the origin of the difficulties.

A. Deficit, cost of erasure of entanglement, and relative entropy of entanglement

In the previous section we have reproduced the result of [15] which provided an upper bound for the deficit in terms of the relative entropy distance from pseudoclassically correlated states. In this section we will prove a new result, providing a *lower bound for the deficit* in terms of an entanglement measure—the relative entropy of entanglement.

Theorem 2. For any bipartite state ρ the quantum deficit is bounded from below by the relative entropy of entanglement:

$$\Delta(\rho) \geqslant E_r(\rho). \tag{31}$$

To prove the above theorem it is enough to show that Δ_{sep}—the cost of erasing entanglement—is lower bounded by E_r, which is the contents of the next theorem. Indeed, by definition of Δ_{sep} and by the proposition 1 the deficit is no smaller than Δ_{sep}.

Theorem 3. For any bipartite state ρ the cost of erasing entanglement is bounded from below by the relative entropy of entanglement:

$$\Delta_{sep}(\rho) \geqslant E_r(\rho). \tag{32}$$

To prove this theorem we will need the following lemma.

Lemma 2. Consider any subset S of states, invariant under product unitary transformations. Then the relative entropy distance from this set E_r^S given by

$$E_r^S = \inf_{\sigma\in\mathcal{S}} S(\rho|\sigma) \tag{33}$$

decreases no more than the entropy increases under local dephasing—that is,

$$E_r^S(\rho) - E_r^S(\Lambda(\rho)) \leqslant S(\Lambda(\rho)) - S(\rho), \tag{34}$$

where Λ is local dephasing.

Proof. Note first that local dephasing can be represented as a mixture of local unitaries:

$$\Lambda(\rho) = \sum_i p_i U_A^i \otimes I_B \rho U_A^{i\,\dagger} \otimes I_B. \tag{35}$$

Indeed, consider any set of projectors $\{P_j\}_1^k$. The suitable unitaries are given by

$$U(s_1,\ldots,s_k) = \sum_{j=1}^k s_j P_j, \tag{36}$$

where $s_j=\pm1$ are chosen at random. Thus p_i's are equal, but this is irrelevant for our purpose.

Now, let us rewrite the inequality (34) as follows:

$$E_r^S(\rho) + S(\rho) \leqslant S(\Lambda(\rho)) + E_r^S(\Lambda(\rho)). \tag{37}$$

Thus we have to prove that the function $f(\rho)=E_r^S(\rho)+S(\rho)$ is nondecreasing under dephasing. This is a somehow parallel result to the result of [48] where it was proved that the above function does not decrease under (global) mixing. The proof is directly inspired by [49].

We have

$$\begin{aligned} f(\Lambda(\rho)) &= \inf_{\sigma\in S} -\operatorname{tr}\Lambda(\rho)\log\sigma = \inf_{\sigma\in S}\sum_i p_i \operatorname{tr}\rho_i\log\sigma \\ &\geqslant \sum_i p_i \inf_{\sigma\in S}\operatorname{tr}\rho_i\log\sigma = \sum_i p_i \inf_{\sigma\in S}\operatorname{tr}\rho\log\sigma_i \\ &= \sum_i p_i \inf_{\sigma\in S}\operatorname{tr}\rho\log\sigma = f(\rho),\end{aligned} \tag{38}$$

where $\rho_i=U_A^i\otimes I_B\rho U_A^{i\,\dagger}\otimes I_B$ and $\sigma_i=U_A^{i\,\dagger}\otimes I_B\sigma U_A^i\otimes I_B$. The inequality comes from the properties of the infimum; the last but one equality comes from the fact that the set S is invariant under product unitary operations. This ends the proof of the lemma. □

Proof of theorem 3. The basic ingredient of the proof is the monotonicity of the function $f(\rho)=E_r(\rho)+S(\rho)$ under CLOCC. (In entanglement theory important functions are the ones that cannot increase under a suitable class of operations, while here we need a function that does not *decrease* under our class of operations. This once more shows that our approach is in a sense dual to the usual entanglement theory.) As we have already discussed, any CLOCC operation can be decomposed into basic ones: (i) local unitary transformation, (ii) local dephasing, and (iii) noiseless sending of dephased qubits. Of course the local unitary operation does not change either the entropy or E_r, so that the function f remains constant. The lemma we have just proved tells us that local dephasing can only increase the function f. Consider now the last component—sending dephased qubits. Clearly the entropy again does not change during such operations. It remains to show that E_r does not change under sending dephased qubits. Consider the state $\rho_{ABB'}$ with one dephased qubit B' on Bob's site. Consider the closest separable state to the state $\sigma_{ABB'}$. Since the relative entropy of entanglement is in particular monotone under dephasings, we can choose this state to have the qubit B' dephased too. Consider then the state $\rho_{AA'B}$, where A' qubit is the B' qubit after being sent by Bob. We now apply the procedure of sending qubit B' to the state $\sigma_{ABB'}$ and obtain a new separable state $\sigma_{AA'B}$. By construction we have $S(\rho_{ABB'}|\sigma_{ABB'})=S(\rho_{AA'B}|\sigma_{AA'B})$. Thus E_r could only go down. However, we can repeat the reasoning with the qubit sent in the converse direction and conclude that E_r does not change.

In this way we have shown that the function f cannot decrease under CLOCC operations. This means that for any protocol that brings the initial state ρ to a final separable state ρ' we have

$$f(\rho') \geqslant f(\rho). \tag{39}$$

However, the target state is separable; hence, it has $E_r=0$. We obtain

$$S(\rho') - S(\rho) \geqslant E_r(\rho), \tag{40}$$

which tells us that in any protocol that ends up with a separable state, the increase of entropy is no smaller than the relative entropy of entanglement. This ends the proof. □

B. Connection with bounds obtained via semidefinite programming

In [19] semidefinite programming techniques were used to obtain lower bounds on the regularized deficit. The following general bound was obtained:

$$\Delta^{\infty}(\rho) \geqslant \sup_{\sigma}[-\log_2 \lambda_{\max}(|\sigma^{\Gamma}|) - S(\varrho) - S(\varrho|\sigma)], \tag{41}$$

where $\lambda_{\max}$ denotes the greatest eigenvalue and Γ is the partial transposition of the matrix. The value of the bound has been calculated for Werner states and isotropic states. It turned out that for those states it is exactly equal to the regularized relative entropy of entanglement. This is compatible with theorem 3. It is interesting, what is the general relation of the bound (41) with regularized E_r.

C. Discussion of the problem of "noncommuting choice"

We have proved that the deficit satisfies the inequality

$$E_r^{\mathcal{PC}} \geqslant \Delta \geqslant E_r. \tag{42}$$

Yet we have not been able to prove that $\Delta=E_r^{\mathcal{PC}}$. Let us discuss the main obstacles which we encountered. The question is actually as follows: Can there be a better protocol than dephasing in an optimal IPB basis? The latter protocol has some fundamental features. Namely, in the series of subsequent local dephasings, each dephasing is compatible with the previous one in the sense that they commute with each other. In other words, each dephasing is in some sense ultimate: it divides the total Hilbert space into blocks, so that all subsequent dephasings are performed within blocks and in a basis that is compatible with the blocks. Another way of viewing it is to say that what was sent from Alice to Bob or vice versa will remain classical—that is, diagonal in a fixed distinguished basis. The main open question is now the following: Is it enough for Alice and Bob to follow this restriction, or should they violate this rule to draw more information?

We can formulate this fundamental problem in a more tractable way if we look through the proof of theorem 3 and find where the proof fails if instead of separable states one takes pseudoclassically correlated states. Almost the entire proof can be carried forward without alteration, apart from one small item: the invariance of $E_r^{\mathcal{PC}}$ under sending dephased qubits. E_r was invariant mainly because we could choose the closest separable state to be also dephased on that qubit. This is because the set of separable states is closed under local dephasings. However, the set of pseudoclassically correlated states is not. It does not rule out the possibility that indeed the closest pseudoclassically correlated state has the qubit dephased. However, we were not able to prove it or disprove. We will formulate here the problem in a formal way.

Problem. Consider a bipartite state that can be written in the following form:

$$\rho_{AB} = p_1\rho_{AB}^1 + p_2\rho_{AB}^2, \tag{43}$$

where ρ_{AB}^1 and ρ_{AB}^2 are orthogonal on subsystem A; i.e., the reduced states ρ_A^i have disjoint support. Can the closest pseudoclassically correlated state in the relative entropy distance be written in this form?

D. Deficit and relative entropy distance for one-way and zero-way scenarios

Finally let us note that the needed results can be obtained easily for one-way and zero-way scenarios. The problem with the two-way scenario is that Alice and Bob could draw more information than they obtain by measuring in an optimal IPB basis. The source of the difficulty was that in a many rounds protocol, Alice and Bob could make dephasings that would not commute with dephasings they made in a previous step. In the case of the one-way scenario there is no such danger, as there is only one round. The zero-way situation is simplest. The only thing Alice and Bob can do is to dephase the subsystems in some bases, and the only problem is to find the optimal basis (so that they will produce the smallest amount of entropy). The versions of lemma 1 in the one-way and zero-way cases can be proved in the same way. Thus in those cases the deficits are equal to the relative entropy distance to the two sets of states—classically correlated states and one-way classically correlated states (19).

E. Multipartite states

We can define a set of pseudoclassically correlated states also in the case of multipartite states. Then one can formulate a version of theorem 1 in the latter case. Since the arguments we have used did not depend on the number of parties, theorem 1 is then true also in the multipartite case. Similarly theorems 2 and 3 also hold in the multipartite case.

VII. BASIC IMPLICATIONS OF THE THEOREM (INFORMATIONAL NONLOCALITY)

The theorems from the previous section allow us to obtain the following results for both bipartite as well as multipartite states.

(i) Δ is no smaller than distillable entanglement E_D:

$$\Delta \geqslant E_D. \tag{44}$$

Indeed, the latter is bounded from above by the relative entropy of entanglement [50].

(ii) Moreover, theorem 3 implies that the quantum deficit is no smaller than coherent information:

$$\Delta(\rho) \geqslant S(\rho_X) - S(\rho), \tag{45}$$

where $X=A,B,C,\ldots$. or

$$I(\rho) \leqslant N - S(\rho_X). \tag{46}$$

This is because it was proved that in the bipartite case [51], the relative entropy of entanglement is bounded from below by coherent information $S_X - S$. For multipartite states, one gets it by noting that the multipartite relative entropy of entanglement is no smaller than the one versus some bipartite cut. Then one applies the mentioned bipartite result.

(iii) *Any entangled state is informationally nonlocal*; i.e., it has a nonzero deficit:

$$\Delta(\varrho_{entangled}) > 0. \tag{47}$$

This follows from the fact that when a state is entangled, then it has a nonzero relative entropy of entanglement.

Note, however, that there exist *separable* states which are informationally nonlocal:

$$\Delta(\varrho_{separable}) > 0, \tag{48}$$

for some separable states. We will now discuss an example of such a state and relate it to so-called "nonlocality without entanglement." Whether such an effect survives in the asymptotic limit of many copies is unclear.

(iv) Theorem 3 allows for easy proof that for pure bipartite states the deficit is equal to entanglement. Indeed, from the theorem we have that the deficit is no greater than entanglement. On the other hand, a simple protocol of dephasing Alice's state in the eigenbasis of the state of her subsystem and sending it to Bob gives the amount of information $2 \log d - S(\rho_A)$. Thus the deficit is also no greater than the entropy of the subsystem. However, the latter is equal to the relative entropy of entanglement (this is a reflection of the fact that in the asymptotic regime there is only one measure of entanglement for pure states). For multipartite pure states there does not exist a unique entanglement measure. We have the following open question: *For multipartite pure states, is the deficit equal to the relative entropy of entanglement?* That is,

$$E_r(\psi) \overset{?}{=} \Delta(\psi). \tag{49}$$

If so, the deficit would be an entanglement measure for all pure states. And since the deficit is an operational quantity, we would have an *operational* interpretation for the relative entropy of entanglement for pure states.

Note here that in general the deficit is not a monotone under LOCC and even under CLOCC. In contrast, I_l is a monotone under CLOCC.

(v) From the above reasoning and theorem 3 it follows that the *thermodynamical cost of erasure of entanglement of pure states is equal to their entanglement* (cf. [14,15]).

A. Nonlocality without entanglement and with distinguishability

One form of nonlocality we are familiar with is entanglement. Another form of nonlocality was introduced in [10]: the so-called *nonlocality without entanglement*. There, it was shown that there are ensembles of states, which, although product, cannot be distinguished from each other under LOCC with certainty. Ensembles of product states can have a form of nonlocality. Other ensembles were exhibited which were distinguishable, but distinguishing was thermodynamically irreversible. This can be thought of as *nonlocality without entanglement but with distinguishability*. All those results were done for ensembles.

Here we report a similar kind of nonlocality for states. Namely, we will exhibit states which are separable and which can be created out of ensembles of distinguishable states but which contain unlocalizable information such that $\Delta \neq 0$ (at least for single copies). In fact, one can find such states which have an eigenbasis where each eigenket is perfectly distinguishable.

An example is the state given by

$$\rho = \frac{1}{4}|00\rangle\langle 00| + \frac{1}{4}|11\rangle\langle 11| + \frac{1}{2}|\psi^-\rangle\langle\psi^-|. \tag{50}$$

It is a separable state, which can be seen either by construction or because it has a positive partial transpose which is a sufficient condition for dimension $2\otimes 2$. Its eigenkets $|00\rangle$, $|11\rangle$, $|\psi^-\rangle$ are clearly perfectly distinguishable under LOCC, since Alice and Bob just need to measure in the computation basis and compare results to know which of the three basis states they have. Nonetheless, it clearly has nonlocalizable information. To localize all the information, one would need to dephase it in the basis $|00\rangle$, $|11\rangle$, $|\psi^-\rangle$, but this cannot be done under CLOCC, since one cannot dephase using a projector on $|\psi^-\rangle$. The proof follows from theorem 2—we know that the optimal protocol is for Alice to dephase her side in some basis and then send the state to Bob. Indeed, for two qubits, all implementable product bases are one-way implementable; i.e., they are of the form $\{|i\rangle|\psi_k^{(i)}\rangle\}$ where $\{|\psi_k^{(i)}\rangle\}$ are bases themselves. Thus for the one-copy case, which we consider here, the optimal protocol is a one-way protocol. Since the state is symmetric, then it does not matter which way (from Alice to Bob or vice versa).

A direct calculation shows that the optimal basis is $|0\pm 1\rangle$ at one of the sites. This yields $I_l = 3/4 \log 3 - 1$, while $I=1/2$, giving a value of $\Delta = 0.1887$. There are thus separable states which exhibit nonlocality in that all the information cannot be localized even though all the basis elements of the state are perfectly distinguishable.

VIII. INFORMATIONAL NONLOCALITY OF MULTIPARTITE STATES

The approach considered here turns out to be quite valuable in the case of multipartite states. One of the reasons for this is that one cannot only quantify the quantumness of correlations along various splittings, as is commonly done, but one can also look at the total amount of localizable information that a given state possesses if all parties cooperate. In other words, in addition to the various *vector measures* defined for a particular splitting of the state—e.g. $AB|CD$—one also has a *scalar measure* which is defined for the state as a whole. One can calculate Δ for various bipartite splittings by grouping parties together, or one can calculate Δ for the entire state. In fact, one can consider all possible groupings, such as $AB|CD|EF$, etc. This allows one to explore multipartite correlations in more detail and also allows one to ascribe a single quantity to a particular state in order to rank various states in terms of their total quantum correlations.

By considering a family of states for a number of parties, N, one can calculate the information deficit per party $\Delta(\rho_N)/N$, and we find that it goes to zero for the generalized GHZ and to 1 for the Aharonov state, as N goes to infinity. Of the states we consider, we shall thus find that the GHZ state is the least informationally nonlocal, while the so-called Aharonov state is the most informationally nonlocal.

A. Schmidt decomposable states

The information deficit for the N-party GHZ state,

$$|\psi_{NGHZ}\rangle = |111\ldots 1\rangle + |222\ldots 2\rangle + \cdots + |NNN\ldots N\rangle, \tag{51}$$

where we depart slightly from convention by taking the dimension of each party state to also scale like N. This state is thus more entangled than if one were to give each party a qubit, and we do so in order to fairly compare our results with other entangled states. The deficit for the GHZ was calculated in [14] where it was found to be $\Delta(\psi_{NGHZ}) = \log N$. Essentially, once one party makes a measurement, all the other parties can learn which state they have without performing a measurement, and thus $I_l = (N-1)\log N$, while the total state is of dimension N^N, and hence $I = N\log N$. Therefore,

$$\lim_{N\to\infty} \Delta(\psi_{NGHZ})/N = 0. \tag{52}$$

This is in keeping with the notion that the GHZ state is rather fragile, since if only one of the qubits becomes dephased, the entire state becomes classical.

One can generalize this to any multipartite state which can be written in a Schmidt basis; i.e.,

$$|\psi_{NS}\rangle = \sum_i c_i \prod_{n=1}^{N} |\phi_{Ni}\rangle. \tag{53}$$

In that case, one finds $\Delta(\psi_{NS}) = S(\rho_A)$ where ρ_A is any of the subsystem entropies (they are all equal). This follows directly from inequality (46) and it holds in the asymptotic regime of many copies.

B. Example of a non-Schmidt decomposable state: The *W* state of three qubits

A more complicated example is the "*W* state" [52]

$$|\psi_W\rangle_{ABC} = \frac{1}{\sqrt{3}}(|100\rangle + |010\rangle + |001\rangle).$$

and we ask the question of how much localizable information I_l can be extracted under one-way CLOCC by using it as a shared state. Since each party only has one qubit, we can use theorem 2 to calculate it. This is because if each party only holds a single qubit, the optimal protocol will only need one-way communication and will be equivalent to having one-party measure, and then tell her results to the other parties who will then hold a pure state between them.

Let Alice measure her part of the state in basis $\{|e_i\rangle\}$ and send the result to Bob and Charlie. After the measurement,

$$|\psi_W\rangle_{ABC} \rightarrow \varrho_{ABC} = \sum_i p_i |e_i\rangle\langle e_i| \otimes \varrho_{BC}^i. \tag{54}$$

Then Alice obtains the ensemble $\{p_i, |e_i\rangle\}$. Bob and Charlie obtain the ensemble $\{p_i, \varrho_{BC}^i\}$. ϱ_{BC}^i are of course pure states. Bob and Charlie know which of the states $\{\varrho_{BC}^i\}$ they have, because they have obtained information about the result of the measurement by Alice. Therefore, the total amount of information that can be extracted from $|\psi_W\rangle_{ABC}$ locally by such a protocol is given by

$$I(\varrho_A) + p_1 I_l(\varrho_{BC}^1) + p_2 I_l(\varrho_{BC}^2), \tag{55}$$

where $\varrho_A = \Sigma_i p_i |e_i\rangle\langle e_i|$ so that

$$I(\varrho_A) = 1 - H(\{p_i\}) \tag{56}$$

and where (since ϱ_B^i are pure)

$$I_l(\varrho_{BC}^i) = 2 - S(\varrho_B^i), \tag{57}$$

with ϱ_B^i being the reduced density matrix of ϱ_{BC}^i. So for an arbitrary von Neumann measurement, we have that for the W state, I_l is given by

$$I_l(\psi_W) = 3 - H(\{p_i\}) - \sum_1^2 p_i S(\varrho_B^i) = 3 - H\left(\frac{1+|x|^2}{3}\right) - \frac{1+|x|^2}{3} H\left(\frac{1}{2} + \frac{\sqrt{-3|x|^4 + 2|x|^2 + 1}}{2+2|x|^2}\right) - \frac{2-|x|^2}{3} H\left(\frac{1}{2} + \frac{\sqrt{4|x|^2 - 3|x|^4}}{4-2|x|^2}\right),$$

where the measurement is performed in the basis $\{|e_i\rangle\}$ given by

$$|e_1\rangle = x|0\rangle + y|1\rangle,$$

$$|e_2\rangle = y^*|0\rangle - x^*|1\rangle.$$

One can check that for von Neumann measurements, the largest amount of local information extractable is 1.450 26. It

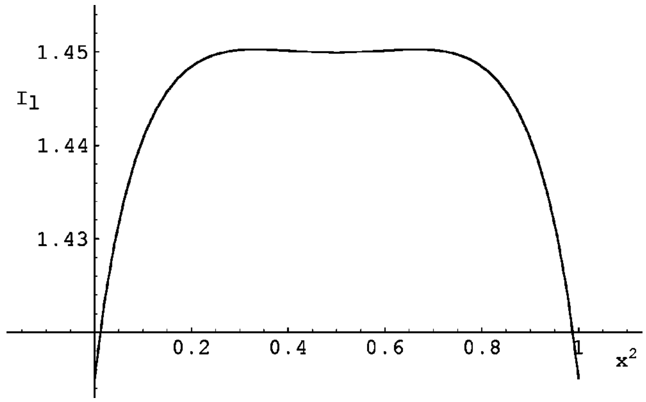

FIG. 3. Plot of I_l^x versus x^2 for measurement in basis (61) for the W state. The optimal basis for maximizing I_l^x is for Alice to dephase (or measure) with $x^2=1/3$ or $2/3$. The basis $|\pm\rangle$ ($x^2=1/2$) is not optimal.

is achieved for measurements in the basis $\{|e_i\rangle\}$, where either $x^2=1/3$ or $x=2/3$ (see Fig. 3). Contrary to naive expectations, dephasing in the computational basis is the worst choice. Also the basis $|\pm\rangle$ ($x=1$) is not optimal. It is interesting that optimal bases are not incidental. Rather these are those bases for which the probabilities of a transition into $|0\rangle$, $|1\rangle$ states are the same as the probabilities of getting those states by Alice measuring the W state in basis $|0\rangle$, $|1\rangle$. In the regime of single copies, this protocol is optimal by theorem 2; therefore, for the W state, $I_l=1.450\,26$. This is less than the amount of localizable information for the corresponding GHZ state $|\psi_{GHZ}\rangle=(1/\sqrt{2})(|000\rangle+|111\rangle)$; thus, we would argue that the W state exhibits more nonlocal correlations.

C. Aharonov state and quasiunlocalizable information

We next consider the so-called Aharonov "diamond" state. It is essentially given by antisymmetrizing N N-dimensional states. For three parties, the unnormalized state is

$$|\psi_{3A}\rangle = |012\rangle - |021\rangle + |120\rangle - |102\rangle + |201\rangle - |210\rangle \tag{58}$$

and in general it is

$$|\psi_{NA}\rangle = \frac{1}{\sqrt{N!}}\sum_{permutations}\epsilon^{a_1\cdots a_N}|a_1\cdots a_N\rangle, \tag{59}$$

where $\epsilon^{a_1\cdots a_N}$ is the permutation symbol (Levi-Cività density).

It has the property that if one party measures their state in any basis and tells their result to the rest of the parties, they will then still hold another Aharonov state of dimension $N-1$. Since this is a pure state of dimension N^N, the total amount of information is $I=N\log N$. On the other hand, under the protocol where the parties take turns measuring, it is easy to see that after each measurement, the other parties will still be left with a locally maximally mixed state. However, the maximally mixed state will reside in a dimension lower than N. Finally, there will be two parties left, and they will share a singlet. One of the parties can convert her pair into $\log N-1$ bits of information, while the other can get $\log N$. The kth party can get $\log N-\log k$. The amount of localizable information is therefore $I_l=\log N^N/N!$. This is optimal by theorem 2 for single copies. We thus have that $\Delta(\psi_{NA})/N=\log N!/N$ which goes to 1 in the limit of $N\to\infty$. Compared to the GHZ state of equivalent dimension, the Aharonov state has far more unlocalizable information. Related behavior has been found independently [53,54]. One might wonder if one can make the localizable information strictly zero, as is the case for entanglement with bound entangled states. We will soon show that this is not the case.

D. General pure three-qubit states

In Sec. VIII A, we considered the localizable information of Schmidt decomposable states, and in Sec. VIII B, we considered the W state, an example of a non-Schmidt-decomposable state.

Let us here consider the general three qubit *pure* state, which can be written in the form [55,56]

$$|\psi\rangle_{ABC} = a|000\rangle + b|010\rangle + c|100\rangle + d|001\rangle + e|111\rangle, \tag{60}$$

where only a need be complex, while the rest of the coefficients are real. Of course we have $|a|^2+b^2+c^2+d^2+e^2=1$.

We again can use theorem 2 to obtain the amount of localizable information. Let us suppose that Alice (A) measures in the basis

$$|e_1\rangle = x|0\rangle + y|1\rangle,$$

$$|e_2\rangle = y^*|0\rangle - x^*|1\rangle, \tag{61}$$

and sends the measurement outcome to Bob (B) and Charlie (C).

Depending on the measurement outcome, Bob and Charlie share the state

$$|\psi_{e_1}\rangle = \frac{1}{\sqrt{p}}((x^*a+y^*c)|00\rangle + x^*d|01\rangle + x^*b|10\rangle + y^*e|11\rangle)$$

or

$$|\psi_{e_2}\rangle = \frac{1}{\sqrt{1-p}}((ya-xc)|00\rangle + yd|01\rangle + yb|10\rangle - xe|11\rangle),$$

corresponding to the outcome $|e_1\rangle$ or $|e_2\rangle$ at Alice, where

$$p = |(x^*a+y^*c)|^2 + |x|^2d^2 + |x|^2b^2 + |y|^2e^2$$

is the probability that $|e_1\rangle$ is obtained by Alice.

For such a protocol, the localizable information amounts to

$$I_l = \sup_{x,y}[3 - H(p) - pS(\mathrm{tr}_A|\psi_{e_1}\rangle\langle\psi_{e_1}|) - (1-p)S(\mathrm{tr}_A|\psi_{e_2}\rangle\langle\psi_{e_2}|)], \tag{62}$$

where we maximize over x and y to obtain the highest localizable information. This is an optimal protocol, and thus we

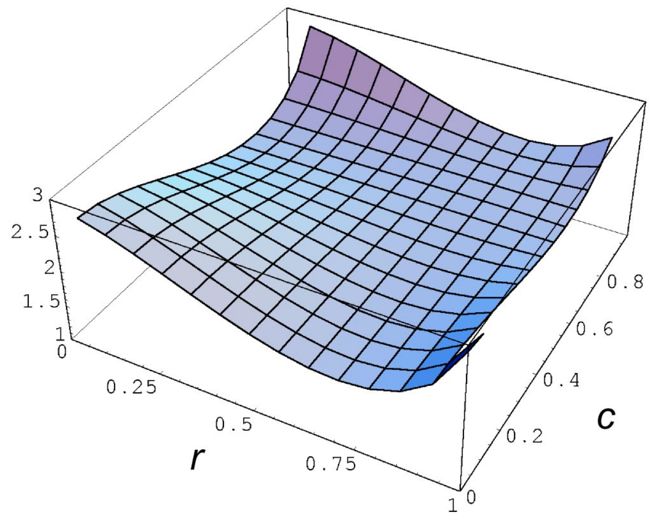

FIG. 4. Plot of the function I_l^{xy} [Eq. (62)] for the three-qubit state in Eq. (60) for the case when $a=e=0$, $b=0.1$, in the (c,r) plane. Here $x=r$, $y=\sqrt{1-r^2}$, and $r>0$. The value of localizable information I_l for a given c is the supremum of I_l^{xy} for that value of c.

obtain I_l. Let us denote the quantity in square brackets as I_l^{xy}.

Let us now choose an exemplary one-parameter subclass from the class in Eq. (60):

$$a = e = 0, \; b = 0.1.$$

For this class, we plot the localizable information I_l using real values of x and y. Taking $x=r>0$ and $y=\sqrt{1-r^2}$, I_l^{xy} is plotted (in Fig. 4) as a function of r and c. For a given c (which then fixes the state), the value of I_l can be read from the figure.

IX. BELL MIXTURES

The state of Eq. (50) is a particular example of a mixture of Bell states:

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

$$|\psi^{\pm}\rangle = \frac{1}{\sqrt{2}}(|01\rangle \pm |10\rangle),$$

Here, for completeness, we calculate Δ for all states of this for so-called Bell-diagonal states. Up to local unitaries, this includes all $2\otimes 2$ states with local density matrices that are maximally mixed. optimization could way classical Due to theorem 2, we only need consider optimizing are over projection measurements (without adding any ancilla locally) at one of the parties—say, Alice. Consider therefore the mixture

$$\varrho_{Bm} = p_1 P_{\phi^+} + p_2 P_{\phi^-} + p_3 P_{\psi^+} + p_4 P_{\psi^-} \tag{63}$$

of the four Bell states in $2\otimes 2$.

After an arbitrary projection-valued (PV) measurement on Alice's side, projecting in the basis

$$\{|\overline{0}\rangle = a|0\rangle + b|1\rangle, \; |\overline{1}\rangle = \overline{b}|0\rangle - \overline{a}|1\rangle\},$$

let the global state be projected, respectively, to

$$P_{|\overline{0}\rangle} \otimes \varrho_0, \; P_{|\overline{1}\rangle} \otimes \varrho_1. \tag{64}$$

At this stage, the whole state is essentially on Bob's side. This is because we allow dephasing as one of our allowed operations. Consequently, the locally extractable information after this set of operations is the von Neumann entropy of

$$pP_{|\overline{0}\rangle} \otimes \varrho_0 + (1-p)P_{|\overline{1}\rangle} \otimes \varrho_1,$$

where p is the probability of Alice obtaining the state $|\overline{0}\rangle$. The optimization yields the value

$$\Delta = 1 + H(p_1 + p_2) - S(\varrho_{Bm}), \tag{65}$$

where p_1 and p_2 are the two highest coefficients of the Bell mixture ϱ_{Bm}.

If we consider only von Neumann measurements (without addition of ancilla) and if Alice and Bob are not allowed to make any communication before they perform their measurements, then the zero-way information deficit $\Delta^{\emptyset}$ for the Bell mixtures (63) is given by

$$1 + H(p_{\max}),$$

where

$$p_{\max} = \frac{1}{2}(1 + |\max\{t_{11}, t_{22}, t_{33}\}|),$$

with $t_{ii}=\mathrm{tr}(\sigma_i\otimes\sigma_i\varrho_{Bm})$. Note however that in this case, we are unable to show whether one can do better by positive-operator-valued measures (POVM's) or whether more copies are useful.

Consider, however, the isotropic $d\otimes d$ state

$$\varrho_{iso} = \lambda|\phi_{\max}\rangle + (1-\lambda)\frac{I}{d^2} \tag{66}$$

in $d\otimes d$, where $\phi_{\max}$ is the maximally entangled state in $d\otimes d$ which is invariant under $U\otimes U^*$ for any unitary U. The one-way information deficit $\Delta^{\rightarrow}$ (as well as $\Delta^{\emptyset}$) is given by

$$\Delta^{\emptyset} = \Delta^{\rightarrow} = \left(\lambda + \frac{1-\lambda}{d}\right)\log_2\left(1 + \frac{1-\lambda}{d}\right) + (d-1)\frac{1-\lambda}{d}\log_2\frac{1-\lambda}{d} - \log_2 d + S(\varrho_{iso}), \tag{67}$$

where

$$S(\varrho_{iso}) = -\left(\lambda + \frac{1-\lambda}{d}\right)\log\left(\lambda + \frac{1-\lambda}{d}\right) - \frac{d^2-1}{d^2}(1-\lambda)\log\frac{1-\lambda}{d^2}. \tag{68}$$

For the isotropic state, it is possible to prove, along the same lines as for Bell mixtures, that POVM's as well as more than one copy cannot help.

A. Asymptotic regime

For two qubits we easily evaluated the deficit, because one-way and two-way deficits are equal in this case and because Alice's first measurement leaves no room for other measurements. So the only thing she should do is to communicate the results to Bob, and communication from Bob is not needed. In other words, the set of pseudoclassically correlated states is equal to the one-way classically correlated states of the form (19). Thus it was enough to evaluate only the one-way deficit. However, if we turn to regularization, this equivalence is no longer valid. This is because, to calculate regularization, one needs to evaluate the deficit for many copies. Thus the dimension of the system is high, and there is room for many rounds. We are not able to regularize the two-way deficit.

Concerning the one-way deficit, one can argue that it is additive for Bell diagonal states. Moreover, borrowing qubits does not help (it has been independently shown that, in general, in the one-way case, borrowing pure local qubits does not help [20]). We will provide the arguments in Sec. XIV A.

X. PURELY NONLOCALIZABLE INFORMATION DOES NOT EXIST

One important aspect of entanglement theory is the existence of bound entangled states. These are states which are entangled in that they require entanglement to create, yet no entanglement can be drawn from them. In Sec. VIII C we saw that in the multipartite case, there were states for which the amount of localizable information per party was small as the number of parties increases. One can ask whether there is a strict analogy to bound entanglement: are there states which have positive I, but which $I_l=0$. It turns out that the answer is no; the only state which has $I_l=0$ is the maximally mixed state. Here we prove this in the following lemma for the case of two parties. The generalization to many parties is straightforward.

Lemma 3. From any state other than the maximally mixed state we can draw local information.

Proof. Consider a state $\varrho \sim C^d \otimes C^d$ such that $\varrho \neq \varrho_{mmix} = I/d^2$; then, there exists an observable for which the mean value in state ϱ has a different value than ϱ_{mmix}. Every nonlocal observable can be decomposed into local operators, so we can always find such an observable of the form $A \otimes B$ for which

$$\mathrm{Tr}(A \otimes B)\varrho \neq \mathrm{Tr}(A \otimes B)\frac{I}{d^2}. \tag{69}$$

Then

$$\mathrm{Tr}(A \otimes B)\varrho \neq \frac{1}{d^2}\mathrm{Tr}A\mathrm{Tr}B, \tag{70}$$

$$\sum_{ij} p_{ij}\lambda_i\lambda_j \neq \sum_{ij} \frac{1}{d^2}\lambda_i\lambda_j. \tag{71}$$

Notice that distribution of probability for ϱ in Eq. (71) is classical. We know that we can obtain a nonzero amount of local information from any classical state besides the maximally mixed one. We can see that we are able to find such a local operation that transforms every state which agrees with the assumptions of lemma 3 into a state from which we can draw local information. □

There is an open question whether there exist states for which localizable information is entirely equal to *local* information content, but which nevertheless are not product. In such a case, one would not be able to draw information from correlations at all. The classical deficit Δ_c would be zero, even though the state would be nonproduct. It is rather unlikely that such states exist, yet we have not been able to solve this question.

We now prove a related theorem which follows from the above lemma and which will be useful for the following section. Namely, we show that using pure states as a resource cannot help when distilling local information. One can think of such a process as *catalysis* where one uses pure states to produce more pure states from some shared state.

Theorem 4. Local pure ancillas do not help in the process of distilling local information.

Proof. Assume that catalysis can help in drawing local information. Consider a state ϱ, which is not the maximally mixed state, and the optimal protocol of distilling local information P_1, which does not use ancillas. Consider also another protocol P_2, in which we distill information from some of the copies of state ϱ. Using P_1 and then using the distilled pure states to do catalytic distillation on the rest of the copies. Notice that we can do this, because we know from lemma 3 that we can distill local information and thus also pure states from it. If catalysis is helpful, that means that using P_2 we are able to obtain more local information than in the previous protocol. Protocol P_2 does not use ancillas and is better than P_1, which is optimal. This leads to the required contradiction.

We showed that catalysis is useless for a state with nonzero distillable information. It could help only in the case of states with pure unlocalizable information, but we know from lemma 3 that such states do not exist. This ends the proof.

Remark. We know that to do catalytic distillation we need pure ancillas. One can notice that states which we want to use in protocol P_2 to do catalysis are not exactly pure. But these states come from distillation, so they are equal in the limit of many copies to $|0\rangle^{\otimes rn}$ (r is the rate of distillation of local information and n is the amount of copies). This fact assures us that in the asymptotic regime of many copies we are able to do catalysis.

XI. CAN CORRELATIONS BE MORE QUANTUM THAN CLASSICAL?

The total amount of correlations contained in a bipartite state is given by the mutual information

$$I_M = S(\rho_A) + S(\rho_B) - S(\rho_{AB}). \tag{72}$$

One can easily see that our quantities for dividing correlations into ones which behave quantumly (Δ) and classically (Δ_c) satisfy

$$I_M = \Delta + \Delta_c. \tag{73}$$

In other words, the total amount of correlations (given by I_M) can be divided into classical and quantum components [16]. Now one can ask whether the total correlations I_M can be divided arbitrarily. Certainly for pure states this is not the case. For pure states, correlations which behave quantumly cannot exceed $I/2$. For pure states ψ, we showed that $\Delta = S(\rho_A)$, and thus it is always the case that $\Delta(\psi) = I_M/2$. For pure states, the quantumness of correlations can never exceed the classicalness of correlations.

Now one can ask, can it be that one has states for which

$$\Delta(\varrho_{AB}) > I/2? \tag{74}$$

If so, one could think of these states as having *supersaturated* quantum correlations, in that for a given amount of mutual information I_M they have a greater proportion of correlations which behave quantumly. In this sense, one can think of such states as being more nonlocal than maximally entangled states.

One way of approach to the above problem is to work with the relative entropy of entanglement. We know that both the relative entropy of entanglement (E_r), with the distance taken from separable states, and the von Neumann entropy (S_{AB}) are not greater than $\log_2 d$ for $d\otimes d$ states. Consequently, one has $E_r + S_{AB} \leqslant 2 \log_2 d$. Can we have the following stronger inequality:

$$2E_r + S_{AB} \overset{?}{\leqslant} 2 \log_2 d. \tag{75}$$

This is tight for maximally entangled states. Because the deficit is no smaller than the relative entropy of entanglement, it follows that if the inequality is violated, then for some states inequality (74) is true, and we would have this curious phenomenon. On the other side, when the inequality is satisfied for all states, we would obtain a nice trade-off between entanglement and noise.

In a recent work, Wei *et al.* [57] calculated (for two-qubit states) the maximal possible relative entropy of entanglement E_r (as well as other entanglement measures) for a given amount of mixedness (quantified by the von Neumann entropy). Note that the inequality (75) would generically hold for two-qubit states if it is satisfied by these optimal values. Indeed examining the curves of the above paper, one finds that for any two-qubit state the inequality is satisfied.

One can also find that for Werner states and maximally correlated states, the inequality is satisfied too for the regularized relative entropy of entanglement. To see this, the asymptotic relative entropy of entanglement ($E^{\infty}_{r(PPT)}$) [with distance taken from states with positive partial transpose (PPT)] is known for Werner states (mixture of projectors on symmetric and antisymmetric spaces) in $d\otimes d$ [58]. One may check that the relation

$$2E^{\infty}_{R(PPT)} + S_{AB} \leqslant 2 \log_2 d. \tag{76}$$

is satisfied for all Werner states in arbitrary dimensions. However, note here that the relative entropy of entanglement (from PPT states) is not additive for Werner states.

For the maximally correlated states, the relative entropy of entanglement (from PPT states) is known to be additive. Its value is also explicitly known for all such states in $d\otimes d$. Via additivity, this would exactly be equal to its asymptotic relative entropy of entanglement (from PPT states). Precisely, for any state of the form

$$\varrho_{mc} = \sum_{ij} a_{ij} |ii\rangle\langle jj|,$$

we have

$$E^{\infty}_{R(PPT)} = E_{R(PPT)} = \sum_i a_{ii} \log_2 a_{ii} - S(\varrho_{mc}).$$

It is easy to check that the relation (76) is satisfied by any ϱ_{mc} in $d\otimes d$.

Thus we have not found states for which the inequality would be violated for the regularized relative entropy of entanglement. It remains an open question whether the trade-off between noise and entanglement represented by inequality (75) is universally true or whether there exist states for which there is more quantum than classical correlations.

XII. ZERO-WAY AND ONE-WAY SUBCLASSES

We now turn to additional measures of the quantumness of correlations which arise when one restricts the communications between Alice and Bob. In Secs. IX and VIII such restrictions were useful for evaluations of perhaps more basic two-way quantities. However, they are more than just for ease of calculation—we shall also see that the restricted measures allow one to explore other aspects of nonlocality. Additionally, there appear to be strong connections between the deficit and distillation of randomness from shared states. For example, it has just been shown in [20] that the one-way deficit is equal to the mutual information minus the one-way distillable randomness [59].

As before, the optimal protocols by which the corresponding local informations are obtained amounts to producing "classical-like" states of least entropy by the respective operations. As mentioned in Sec. V B the theorems proved there apply equally well in these restricted scenarios with suitable modification.

In any protocol of concentrating information to local form, the parties can stop at states of the form

$$\varrho'_{AB} = \sum_{ij} p_{ij} |i\rangle\langle i| \otimes |j\rangle\langle j|. \tag{77}$$

However, for the two-way scenario, we have argued that one can stop already at pseudoclassically correlated states. When one-way protocols are allowed, it is sufficient for the parties to stop at states of the form

$$\varrho'_{AB} = \sum_i p_i |i\rangle\langle i| \otimes \varrho_i. \tag{78}$$

Finally, for zero-way protocols, one has to achieve classical states (77). Consider, for example, the zero-way protocol for a state ϱ_{AB} by which $I_l^{\emptyset}$ is attained. Without any classical communication (just by dephasing via an environment), Alice and Bob change the state ϱ_{AB} into a classical-like state ϱ'_{AB} [of the form given in Eq. (77)], so that $S(\varrho'_A)+S(\varrho'_B)$ is minimized, where ϱ'_A and ϱ'_B are the local density matrices of ϱ'_{AB}. Note that the parties must concentrate information using classical communication. But this is only after they have performed all their dephasings. The situation is therefore like in a Bell-type experiment.

Let us now show that $\Delta^{\emptyset}$ is an independently useful candidate for quantum correlations and can capture interesting aspects of nonlocality. The states that contain no quantum correlations would be then the ones with $\Delta^{\emptyset}=0$. Consider, for example, the states with eigenbasis (without normalization)

$$|0\rangle_A|0\rangle_B, |0\rangle_A|1\rangle_B, |1\rangle_A(|0\rangle+|1\rangle)_B, |1\rangle_A(|0\rangle-|1\rangle)_B, \quad (79)$$

where $|0\rangle$ and $|1\rangle$ are the eigenvectors of the Pauli matrix σ_z. Such states are the ones used in the Bennett-Brassard 1984 (BB84) quantum cryptography protocol [4]. This set of orthogonal states is distinguishable locally. But it is *not* distinguishable by zero-way communication. Bob must wait for Alice's measurement result (in the σ_z basis) to decide whether to perform a measurement in the σ_z basis or in the σ_x basis. Therefore a mixture of the states in Eq. (79), where the mixing probabilities are all different from each other (so that the spectrum of the resulting state is nondegenerate), would have nonvanishing $\Delta^{\emptyset}$. This is because an arbitrary dephasing by Bob on such a mixture, before obtaining Alice's result, would result in no information being extracted from the state (by Bob). Consequently there would be an information deficit when trying to extract information locally, because globally of course all the information is extractable from such a state. All the information is also extractable by one-way or two-way communication. This is in contrast to states which have an eigenbasis

$$|0\rangle_A|0\rangle_B, |0\rangle_A|1\rangle_B, |1\rangle_A|0\rangle_B, |1\rangle_A|1\rangle_B,$$

for which all the information is extractable from the state locally, by measurement by both the parties in the σ_z basis, without any communication.

We therefore see that the quantum behavior of correlations could result from the distinctly quantum but "local" property of nonorthogonality. Here we call nonorthogonality a local property, as it does not *a priori* require a tensor product structure to manifest itself. It is this nonorthogonality that manifests itself in a more complex form in the examples of LOCC-indistinguishable orthogonal product bases [10,60,61]. More generally, it may be the reason for any case of LOCC indistinguishability of orthogonal states [62–66].

An interesting issue is the relation between $\Delta^{\emptyset}$ and mutual information. In Sec. XI we have asked a question whether there exist states for which Δ would be more than half of the mutual information. The same question can be asked in the case of one-way and zero-way deficits. Pankowski [73] has performed numerical simulations to evaluate $\Delta^{\emptyset}$ versus mutual information. The results are presented in Fig. 5. Surprisingly, there are states for which the deficit is almost equal to the mutual information. Thus the measurement destroys almost all correlations. The quantum correlations do not imply classical correlations (see [67] in this context).

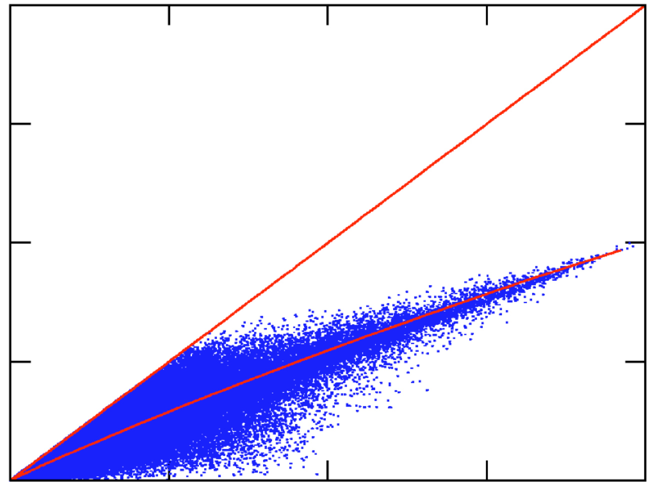

FIG. 5. Zero-way deficit is plotted versus mutual information for 100 000 random two-qubit states. The upper line stands for $\Delta^{\emptyset}=I_M$. The lower line denotes isotropic states of Eq. (66). Two regimes are evident: in the first regime, there are states for which the deficit is almost equal to mutual information. In the second region, the deficit tends to half the mutual information.

With respect to the pure states considered in Sec. IV F, it is easy to see that Δ is also equal to $\Delta^{\rightarrow}$. This is also true for single copies of single-qubit states, due to theorem 2.

A. Expression for one-way Δ

In this subsection we consider the expression for the one-way deficit. In the case when only one-way communication is allowed between the parties, the only thing that Alice and Bob can do is that Alice dephases her part in some basis and then sends her part to Bob. Dephasing transforms the state as

$$\varrho_{AB} \rightarrow \varrho'_{AB} = \sum_i P_i \otimes I \varrho_{AB} P_i \otimes I = \sum_i p_i |i\rangle\langle i| \otimes \varrho_B^i,$$

where $\{P_i=|i\rangle\langle i|\}$ forms a set of orthogonal one-dimensional projectors on the Hilbert space of Alice's part of ϱ_{AB} and p_i are probabilities of the corresponding outcomes which Alice would obtain if she performed measurements with the same P_i's rather than dephasing, while ϱ_B^i is the state that Bob would obtain conditionally on measurement outcome $|i\rangle$. Thus

$$p_i = \mathrm{tr}(\varrho_{AB} P_i \otimes I),$$

$$\varrho_B^i = \frac{1}{p_i} \mathrm{tr}_A(P_i \otimes I \varrho_{AB} P_i \otimes I). \quad (80)$$

The process of sending does not change the form of the state, so that the entropy of the final state at Bob is

$$S(\varrho'_{AB}) = S(\varrho'_A) + \sum_i p_i S(\varrho_B^i),$$

where $\varrho'_A=\Sigma_i p_i |i\rangle\langle i|$ is the reduced density matrix of the A part of ϱ'_{AB}. So finally $I_l^{\rightarrow}$ takes the form

$$I_l^{\to} = n_{AB} - \inf_{\{P_i\}}\left(S(\varrho_A') + \sum_i p_i S(\varrho_B^i)\right)$$

and correspondingly

$$\Delta_l^{\to} = \inf_{\{P_i\}}\left(S(\varrho_A') + \sum_i p_i S(\varrho_B^i)\right) - S(\varrho_{AB}). \qquad (81)$$

Just as we showed that Δ was equal to the relative entropy distance to pseudoclassically correlated states, one can also write $\Delta^{\to}$ and $\Delta^{\emptyset}$ as the minimum relative entropy distance to the set of states $\mathcal{S}^{\to}$ and $\mathcal{S}^{\emptyset}$ which can be created reversibly under the one-way and zero-way classes of operations.

XIII. RELATIONSHIP WITH OTHER MEASURES OF THE QUANTUMNESS OF CORRELATIONS

Let us now compare the deficit with other measures of the quantumness of correlations, in particular the quantum discord [11,12]. The latter is defined, formally, as the difference of two classically equivalent expressions for the mutual information, applied to quantum systems (taken to be a measuring apparatus and system). It was defined with respect to a measurement $A_{\mathcal{M}}$ (either a projective one or a POVM performed on the apparatus A. One then defines the *discord* $\delta(A_{\mathcal{M}}|B)$ with respect to this measurement that results with probabilities p_i in joint states $\varrho_{AB}'^{i} = (|\psi_i\rangle\langle\psi_i|)^A \otimes \varrho_i^B$. The discord is defined as

$$\delta(A_{\mathcal{M}}|B) = H(\{p_i\}) + \sum_i p_i S(\varrho_i) - S(\varrho_{AB}). \qquad (82)$$

The relationship between $\delta(A_{\mathcal{M}}|B)$ and $\Delta^{\to}$ (defined on single copies) was recently shown in [68] where it was shown that the discord also has the interpretation of the extraction of work by a demon, if one minimizes $\delta(A_{\mathcal{M}}|B)$ over all possible measurements $A_{\mathcal{M}}$. Care, however, must be taken, since with the definition of discord there is no cost associated with pure states which are used in a POVM. Therefore, we note here that the relationship between the discord and $\Delta^{\to}$ only applies if one optimizes the discord over von Neumann measurements and disallows POVM's.

Finally, let us provide two explicit examples of cases where two-way communication is more powerful than one-way communication. For example, one has the strict inequality $\Delta^{\leftrightarrow} > \Delta^{\to} = \inf_{A_{\mathcal{M}} \in PVmeas} \delta(A_{\mathcal{M}}|B)$.

To this aim consider the basis related to the sausage states of [10] which has been analyzed in [16]:

$$\begin{aligned}
&\qquad\;\; A \qquad\;\; B\\
\psi_1 &= |0+1\rangle\ |2\rangle,\\
\psi_2 &= |0-1\rangle\ |2\rangle,\\
\psi_3 &= |0\rangle \qquad |0+1\rangle,\\
\psi_4 &= |0\rangle \qquad |0-1\rangle,\\
\psi_5 &= |1+2\rangle\ |0\rangle,\\
\psi_6 &= |1-2\rangle\ |0\rangle,\\
\psi_7 &= |1\rangle \qquad |1\rangle,\\
\psi_8 &= |2\rangle \qquad |2\rangle,\\
\psi_9 &= |2\rangle \qquad |1\rangle. \qquad (83)
\end{aligned}$$

Consider now any bipartite $3\otimes 3$ state $\varrho_{two\text{-}way}$ that is diagonal in the above basis, but has a nondegenerate spectrum. It is relatively easy to provide a two-way protocol that distinguishes vectors (83) without destroying them (see [16]). Hence $\Delta^{\leftrightarrow}$ vanishes. Evidently $\varrho_{two\text{-}way}$ is not of the form $\sum_{i=1}^{3}|\phi_i\rangle\langle\phi_i|\otimes\varrho_i$ with orthogonal ϕ_i, since there are no three eigenvectors among Eq. (83) that have the same component on Alice's side. So both $\Delta^{\to}$ and discord are strictly positive for this state. Thus Maxwell's demon which communicates in both directions is more powerful than a demon who can only communicate in one direction.

Another simple example is to take states which have zero optimized discord or one-way deficit,

$$\rho_{\to} = \sum_i p_i |i\rangle_A\langle i|_A \otimes \rho_B^i, \quad \rho_{\leftarrow} = \sum_j p_j \rho_A^j \otimes |j\rangle_B\langle j|_B, \qquad (84)$$

but in different directions of communication. Then take them each to be on orthogonal Hilbert spaces and mix. Such a state will have $\Delta^{\leftrightarrow}=0$ since both parties can just project onto the two orthogonal Hilbert spaces to determine whether they hold $\rho_{\to}$ or $\rho_{\leftarrow}$ and then the appropriate party can send her state down the channel. On the other hand, one-way communication will be sufficient to completely localize one of the states but not always both.

XIV. RELATION WITH MEASURES OF CLASSICAL CORRELATION

In this section we shall analyze the relation of the classical deficit [16] to already known measures of classical correlations. It happens that both zero-way and one-way deficits have their "counterparts" in such measures. There is no known analog, however, for the two-way deficit.

Let us recall that the quantum deficit was defined as

$$\Delta = I - I_l.$$

One can think of it as describing how much better Alice and Bob can do under CO's if they are given a quantum channel instead of the classical channel. Because it feels the difference between the quantum and classical channels, it tells us about the quantumness of correlations. Likewise, the classical deficit is given by

$$\Delta_c = I_l - I_{LO}.$$

It tells us how much better two parties can do at localizing information if, instead of having no access to a channel—i.e., closed *local* operations—they have access to a classical channel. Because the added resource is a classical channel, it shows how much better the parties can do by exploiting a classical channel.

One can verify that Δ_c and Δ add up to the quantum mutual information $I_M(\varrho_{AB}) = S(\varrho_A) + S(\varrho_A) - S(\varrho_{AB})$. Thus

$$\Delta_{cl} = I_M - \Delta.$$

More explicitly we have [cf. Eq. (11)]

$$\Delta_{cl}(\varrho_{AB}) = S(\varrho_A) + S(\varrho_B) - \inf_{CLOCC}[S(\varrho'_A) + S(\varrho'_B)]; \quad (85)$$

i.e., Δ_{cl} is the optimal decrease of local entropies by means of CLOCC.

A. One-way measures

Corresponding to the measure of quantumness of correlation under one-way classical communication (from Alice to Bob) ($\Delta^{\rightarrow}$), given by Eq. (81), we could have the following formula for classical correlation:

$$\Delta_{cl}{}^{\rightarrow}(\varrho_{AB}) = \sup_{P_i}\left[\{S(\varrho_A) - S(\varrho'_A)\} + \left\{S(\varrho_B) - \sum_i p_i S(\varrho_B^i)\right\}\right]$$
$$\equiv \sup_{P_i}[\delta S(A) + \delta S(B|A)]. \quad (86)$$

Note that the supremum is taken over all local dephasings on Alice's side. Although we optimize over projection measurements, one can effectively include POVM's by including all the required ancillas from the start. Remarkably, it has been shown [20] that POVM's need not be considered when one goes to to the limit of many copies.

In Eq. (86), we have distinguished two terms. The second term

$$\delta S(B|A) = S(\varrho_B) - \sum_i p_i S(\varrho_B^i)$$

shows the decrease of Bob's entropy after Alice's measurement. The first one

$$\delta S(A) = S(\varrho_A) - S(\varrho'_A)$$

denotes the cost of this process on Alice's side and is nonpositive. It is zero only if Alice measures in the eigenbasis of her local density matrix ϱ_A.

The expression for $\Delta_{cl}^{\rightarrow}$ is very similar to the measure of classical correlation introduced by Henderson and Vedral [13]:

$$C_{HV} = \sup_{P_i}\left(S(\varrho_B) - \sum_i p_i S(\varrho_B^i)\right). \quad (87)$$

Originally the supremum was taken over by POVM's, but as mentioned we take the state acting already on a suitably larger Hilbert space, unless stated otherwise explicitly.

The difference between the Henderson-Vedral classical correlation measure and the one given in Eq. (86) is that the former does not include Alice's entropic cost $\delta S(A)$ of performing dephasing. Hence, in general,

$$\Delta_{cl}^{\rightarrow} \leqslant \Delta_{HV}.$$

In the asymptotic limit of many copies, one has equality [20]. Actually in [20] it was shown that the regularized one-way classical deficit is equal to another operational measures of classical correlations: *distillable common randomness* introduced in [59]. The latter is in turn equal to the regularized Henderson-Vedral measure. It is interesting that $\Delta_{cl}^{\rightarrow}$ without regularization, although it seems to be an important characteristic of classical correlations, does not meet a basic requirement for being a measure of classical correlations: it is not monotonous under local operations [42]. Thus regularization plays here a role of monotonization. There is an interesting question: what happens with the two-way classical deficit after regularization?

B. Additivity of the one-way quantum and classical deficits for Bell diagonal states

Here we will prove the fact mentioned in Sec. IX A: that the one-way deficits are additive and that borrowing pure qubits does not help for Bell diagonal states. First of all in [69] it was shown that a measure of classical correlations C_{HV} is additive for Bell diagonal states. Let us recall the argument, as it will be useful for making a connection with the classical deficit. For a Bell diagonal state ρ, consider a related channel Λ [i.e., such a channel that $(I\otimes\Lambda)(|\phi\rangle^{+}\langle\phi^{+}|)=\varrho$]. The maximum output Holevo function over all input ensembles, denoted by $\chi^*(\Lambda)$, is, for general channels, no smaller than C_{HV}. They are equal if the density matrix of ensemble attaining χ^* is equal to ϱ_A. In the case of Bell diagonal states, we have $\varrho_A=I/2$, and it turns out that the optimal ensemble for corresponding channels consists of two orthogonal states and hence gives rise to the same matrix. King [70] has shown that χ^* is additive for channels coming from the Bell diagonal states. From this and from the fact that, in general, $\chi^*\geqslant C_{HV}$ one gets that for Bell diagonal states C_{HV} for many copies is also equal to χ^* for many copies of corresponding channels. This proves that C_{HV} must be additive.

Now, let us make a connection with the classical deficit. As discussed in [42], if χ^* is attained on such an ensemble that its density matrix is equal to ϱ_A, then by looking at the ensemble maximizing χ, one can tell something about measurements that attain C_{HV}. Namely, when the ensemble is orthogonal, then one attains C_{HV} by measurements in the eigenbasis of ϱ_A. Now, it is obvious from Eq. (86) and the discussion thereafter that in the latter case C_{HV} is actually *equal* to the classical deficit, as they differ from one another only by entropy production during Alice's measurement, which vanishes, if it is done in the eigenbasis. Since C_{HV} is additive, then for many copies it is again attained by measurements in the orthogonal basis that is an eigenbasis of Alice's subsystem. Thus the classical deficit for many copies is also not less than C_{HV}, and it by Eq. (86) cannot be greater.

Thus for Bell diagonal states the deficit is equal to C_{HV} and it is additive. Moreover, since the measurement was a von Neumann one, the deficit is attained without using POVM's. This means that additional pure ancillas do not help.

So far we have talked about the classical deficit. Now, since the quantum and classical deficits add up to mutual information which is additive, it follows that the quantum deficit is additive too. Also, since borrowing local qubits does not increase the classical deficit, it cannot decrease the quantum deficit.

C. Zero-way measures

Let us now consider measures of classical correlations under no classical communication, $\Delta_{cl}^{\emptyset}$. Again, this is taken to mean that the parties are not allowed to communicate before making measurements, but can do so afterward in order to concentrate on the classical records. The information deficit under no classical communication, $\Delta^{\emptyset}$, is given by

$$\Delta^{\emptyset} = S(\varrho_{AB}) - S(\varrho'_{AB}),$$

where $S(\varrho'_{AB})$ is the von Neumann entropy of the optimal final state ϱ'_{AB} (which is classical like) and was obtained by local complete measurements, without classical communication. We then obtain

$$\Delta_{cl}^{\emptyset} = S(\varrho_A) + S(\varrho_B) - S(\varrho'_{AB}) = \{S(\varrho_A) - S(\varrho'_A)\} + \{S(\varrho_B) - S(\varrho'_B)\} + \{S(\varrho'_A) + S(\varrho'_B) - S(\varrho'_{AB})\} \equiv \delta S(A) + \delta S(B) + I_M(\rho').$$

We have three terms here: the last one

$$I_M(\rho') = S(\varrho'_A) + S(\varrho'_B) - S(\varrho'_{AB})$$

is the classical mutual information of the final state, while the first two $\delta S(A)=S(\varrho_A)-S(\varrho'_A)$ and $\delta S(B)=S(\varrho_B)-S(\varrho'_B)$ denote, respectively, the local entropic costs of the process at the respective sides. We therefore have a trade-off similar to that in the one-way case. And again there was defined a classical correlation measure [69] which consists only of the last term of our quantity

$$C^{\emptyset} = \sup_{P_i} I_M(\varrho'), \tag{88}$$

where ϱ' is obtained out of ϱ by local complete measurements. Again the original definition of $C^{\emptyset}$ involved POVM's, but as we have suitably increased our Hilbert space from the very beginning, we need not do so.

XV. COMPLEMENTARITY FEATURES OF INFORMATION IN DISTRIBUTED QUANTUM SYSTEMS

Bohr was the first who recognized a fundamental feature of quantum formalism: complementarity between incompatible observables. Complementarity was not explicitly related to entanglement, now regarded as an important quantum-information resource. Namely, Bohr's complementarity concerned mutually exclusive quantum phenomena associated with a *single* system and observed under different experimental arrangements.

Let us comment on complementarity in the case of composite systems and Bohr complementarity. Roughly speaking, the latter says that one cannot access the properties of the systems necessary to describe it by one measurement. The rule is formulated for single-quantum systems and is a consequence of noncommutativity.

On the other hand, we know that one can also divide the properties of the system into local and nonlocal ones, and they are complementary with each other too [16]. For example, one can perform measurement in Bell basis or in standard product basis. However, one cannot perform those measurements simultaneously. In other words, one cannot access global and local properties of the system (see also [71] in this context).

The latter phenomenon is not merely a consequence of Bohr's complementarity. Indeed, if the only allowable states of composite systems were the classically correlated states

$$\rho = \sum_{ij} p_{ij} |e_i\rangle |f_j\rangle \langle e_i| \langle f_j|, \tag{89}$$

then maximal information about the total system would be available through measurements on subsystems. Global measurements would not access any further knowledge about the properties of the system. On the other hand, Bohr complementarity would still hold, in the sense that one cannot access all properties of the system in one measurement.

Thus we see that the local-nonlocal complementarity [16] is a consequence of two distinct phenomena: *noncommutativity* and *the existence of entanglement (or quantum correlations)*. So not only is there noncommutativity, but there is too much of it, so that it affects also relations between local and nonlocal informational contents.

In distributed systems one usually imposes constraints by allowing operations that can be done solely by classical communication and local operations. It turns out that in such a situation there also arises an interesting complementarity. Namely, in [16] we considered two tasks: localizing information (which we have presented in this paper) and sending quantum information (e.g., teleportation), performed simultaneously. It was shown that for a fixed protocol $\mathcal{P}$, the rates of those two tasks obey the relation

$$I_l(\mathcal{P},\rho) + Q(\mathcal{P},\rho) \leqslant I_l(\rho), \tag{90}$$

where $I_l(\mathcal{P},\rho)$ is the amount of information localized by the protocol $\mathcal{P}$ and $Q(\mathcal{P},\rho)$ is the amount of qubits transmitted by the protocol.

For example, for the singlet state, the total informational content is equal to the total correlation content and amount to two bits. The right-hand side of the inequality is equal to 1. This number 2 in light of the above complementarity we can interpret as follows: 2 is equal not to 1 plus 1 but it is equal to 1 *or* 1. One can either draw 1 bit of local information (classical correlations) or teleport 1 qubit (quantum correlations); however, we cannot access both bits.

One can see that this phenomenon is connected with the above-mentioned Bohr complementarity for distributed systems: for the task of teleportation, Alice makes a Bell measurement on her part of the singlet and unknown state to be sent, while to localize information, she measures only the half of the singlets. Interestingly, as far as those two exclusive measurements are concerned, the "local versus nonlocal" complementarity occurs within Alice's laboratory, while it results in complementarity between tasks that refer to local-nonlocal properties of systems belonging to Alice and Bob.

The above inequality suggests an interesting problem: to find the trade-off curves for performances of teleportation and localizing information of a given state. In particular, an interesting question is whether there exist states for which if we teleport the amount of qubits equal to distillable entanglement, one not only would not localize any information, but would need to spend some additional pure states (see [17] in this context).

XVI. DISCUSSION AND OPEN QUESTIONS

In conclusion we have developed a quantum-information processing paradigm which involves local information as a natural resource in the context class of CLOCC operations. We have presented proof that the central quantity of the paradigm, the quantum-information deficit, is bounded from above by the relative entropy distance from the set of pseudoclassically correlated states. We showed how the paradigm allows one to define the thermodynamical cost of erasure of entanglement: entropy production necessary to make the states separable by CLOCC operations. We proved that the cost is no smaller than the relative entropy of entanglement. Since the cost is no greater than the deficit, we have obtained that the deficit is no smaller than the relative entropy of entanglement. This in turn implies that *every entangled state* exhibits informational nonlocality.

We have also found that the paradigm offers a new method of analysis of correlations of multipartite states. The most nonlocal state from this point of view (we call it informationally nonlocal) would be the one for which one has to produce the largest entropy while converting it into classical states. It turned out that according to such a criterion, the Aharonov state is much more nonlocal than the GHZ one. The nonlocality that can be probed by our methods is one that is not caught by Bell's inequalities, since we have found that also separable states can exhibit a nonzero deficit. Rather, it has much in common with nonlocality without entanglement, which was found for ensembles of states [10]. Thus our nonlocality is not identical to entanglement. As a matter of fact, it is a broader notion.

The information deficit has then some peculiar properties. Since it is not an entanglement measure, it can increase under local operations. It is not unreasonable: Local operations may destroy a local property and make it impossible to carry out some action by separated parties, while when the parties meet, the action may still be achievable. This curious behavior of quantum states may be attributed to the fact that even for separable states, when they are mixtures of nonorthogonal states, we cannot ascribe to the subsystems local properties (this may have some connection with the Kochen-Specker theorem).

The paradigm developed in this paper opens many important questions. Here are some of them.

(i) Are *"noncommuting-choice protocols" better in localizing information?* This is the major problem in the paradigm of localizing information by CLOCC operations.

(ii) *Is the quantum deficit equal to the relative entropy distance to pseudoclassically correlated states?* This question would be answered positively, if the noncommuting-choice protocols do not help.

(iii) *Is the regularized deficit still nonzero for all entangled states?* For the regularized deficit we have a lower bound given by the regularized relative entropy distance. However, we do not know if for any entangled state the latter is nonzero.

(iv) *Is the deficit for multiparty pure states equal to the relative entropy of entanglement?* For bipartite states it was proved that the deficit is equal to entanglement. For the multiparty case it is also true for Schmidt decomposable states. It is an open problem whether it is true in general. The same question can be asked for the regularized deficit. Is it equal to regularized E_r for multipartite pure states?

(v) *Is the two-way classical deficit a legitimate measure of classical correlations?* The classical deficit definitely is an important quantity describing some aspects of classical correlations. However, there is a question whether it can be used to quantify them. To this end, it should not increase under local operations [13]. For the one-way case, the classical deficit is not monotonous under local operations as shown in [42]. Yet it turns out that after regularization, the monotonicity is regained [20], because the regularized one-way classical deficit is equal to the one-way distillable common randomness of [59]. Can the two-way classical deficit be also monotonous after regularization? This is connected to the next question.

(vi) *Is the classical two-way deficit equal to the two-way distillable common randomness [59]?*

(vii) *Is the relative entropy of entanglement the thermodynamical cost of erasure of entanglement?* We have shown that the cost is bounded from below by the relative entropy of entanglement. If there is equality, the relative entropy of entanglement would acquire operational status: it would be interpreted as the thermodynamical cost of *erasure of entanglement*.

(viii) *What is the relation between the deficit and mutual information?* We have shown that if a trade-off inequality for E_r Eq. (75), would be violated, then the quantum deficit would be more than the classical deficit for some states. We have also touched on this question by analysis of the zero-way deficit versus mutual information. Preliminary results suggest that there is a very interesting phenomenon while going from quantum to classical states via local measurements: for some states before measurement there are large correlations quantified by mutual information, while after measurement, the remaining amount of information is equal almost exclusively to the initial local information. This means that for some states, even an optimal measurement may destroy most of the information contained in correlations. The question can be recast in the following way: how small can the classical deficit be versus mutual information?

In [67] the measure of classical correlations (88) closely related to the zero-way deficit was compared with mutual information. The authors showed that when this measure is smaller than ϵ, then mutual information is smaller than $\epsilon\mathrm{poly}(d)$ where d is a dimension of the Hilbert space. They were, however, unable to improve the factor to be of order of $\log d$. This means that most probably there is place for a dramatic divergence between the two measures of correlations. Since the deficit can be only smaller from the measure of Eq. (88), the effect can be even stronger. All that suggests that there may be a large gap between the classical and quantum.

(ix) A fundamental open problem, or rather program, is to analyze complementarity between drawing local information and distilling singlets initiated in [16]. In the latter paper, the two tasks—drawing local information and teleporting qubits—were treated as complementary ones. One obtains trade-offs if one wants to perform those tasks simultaneously. An open question is whether distilling singlets can lead to a negative amount of local information gained—i.e., whether in the process of distillation we have to use up local pure qubits rather than we gain them [17]. Moreover, one can define the following quantity: the maximal amount of pure qubits one can draw by CLOCC from a given state [19]. Note that here we do not speak about local qubits. Thus, for example, the singlet is already pure and needs no action. Due to reversibility in entanglement transformations for pure bipartite states [37], the question is in fact reduced to the problem of drawing simultaneously singlets and local pure qubits.

(x) An interesting question arises in the context of [72]. There the authors probe correlations by applying random local unitaries to transform the state to product or separable form, using the smallest number of unitaries. This method allows one to define not only quantum correlations but also total correlations in terms of entropy production while reaching some set of states. It differs from our approach in that the authors do not use classical communication in an essential way (it cannot help). Therefore a natural application of their method is to probe total correlations. This allows them to give a fresh, operational meaning to the quantum mutual information—it is the entropy production needed to bring a quantum state into product form. Our method could be applied in a similar way—one tries to bring a state into product form using CLOCC but without the classical communication (i.e., CLO). Then one finds that the entropy production (i.e., deficit to product states $\Delta_{\text{prod}}^{\text{CLO}}$) is equal to $I(\rho_{AB})$. This can be seen simply from the fact that the optimal protocol is for one party to locally compress her state and then to dephase in the eigenbasis of the compressed state. She then dephases in a basis complementary to the eigenbasis. The latter measurement completely destroys all correlations between A and B. Since the initial entropy was $S(\rho_{AB})$ and the final entropy is $S(A)+S(B)$, the deficit and, hence, entropy production are $I(\rho_{AB})$. Just as the relative entropy distance to some set of states (pseudoclassically correlated and separable states) played a crucial role in the case of Δ and Δ_{sep}, here the relative entropy distance to product states plays the crucial role and is equal to the quantum mutual information.

It is rather amusing that this gives the same answer as the method used in [72], since in our cases, Alice performs her measurement without any knowledge of the density matrix of Bob, while in [72], she must use this information. Furthermore, the number of unitaries which would be needed to perform the dephasing in our case is $S(A)^2$, far greater than the optimal number found in [72]. Understanding in greater detail why these two methods give the same answer might be an interested avenue of further research. It is also interesting to compare how one divides the total correlations into quantum and classical ones. For example, in the case of the singlet, the authors of [72] interpret the two bits of mutual information as requiring one bit of noise to destroy the entanglement and one bit of noise required to destroy the secret correlations. In [16] we interpreted the two bits in terms of one use of a quantum channel *or* one bit of local information.

In the case of destroying correlations due to entanglement, our method uses classical communication in an essential way; therefore, on the surface, it appears to naturally encode the notion of entanglement whose definition relies on the class of LOCC. For pure states the authors of [72] also obtain entanglement, as in this case communication is not needed to reach the set of separable states. It is interesting then to compare what both approaches would produce as far as the entropic cost of erasing entanglement is concerned. One could expect that our method will show less cost in the case of erasing entanglement.

Finally we strongly believe that the present, paradigm analyzed and developed here will be helpful as a rigorous tool in searching for a border or rather a way of coexistence between quantumness and classicality in physical states. It may also enrich our understanding of quantum-information processing and its relation to other branches of physics like thermodynamics and statistics.

ACKNOWLEDGMENTS

The results of this paper were presented at numerous conferences between 2002 and 2004, and put into final form during the Quantum Information Science program at the Isaac Newton Institute. We are grateful to the Institute for their hospitality. We thank Łukasz Pankowski for the performing simulations of Fig. 5, and Sandu Popescu and Andreas Winter for discussions on their draft of [72]. We also thank Igor Devetak for discussions on [20] and Piotr Badziag for pointing out an error in an earlier version of this work. This paper has been supported by the Polish Ministry of Scientific, Research and Information Technology under (solicited) Grant No. PBZ-MIN-008/P03/2003 and by EC grants RESQ, Contract No. IST-2001-37559 and QUPRODIS, Contract No. IST-2001-38877. J.O. also acknowledges the support of the Lady Davis Trust and ISF Grant No. 129/00-1 as well as support from the Cambridge-MIT Institute. A.S. and U.S. also acknowledge support from the Alexander von Humboldt Foundation.

[1] R. S. Ingarden, Rep. Math. Phys. **10**, 43 (1976).
[2] A. Peres, IBM J. Res. Dev. **48**, 63 (2004).
[3] C. H. Bennett, G. Brassard, C. Crepeau, R. Jozsa, A. Peres, and W. K. Wootters, Phys. Rev. Lett. **70**, 1895 (1983).
[4] C. H. Bennett and G. Brassard, in *Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing, Bangalore, India* (IEEE Computer Society Press, New York, 1984), pp. 175–179.
[5] A. K. Ekert, Phys. Rev. Lett. **67**, 661 (1991).
[6] M. Horodecki, P. Horodecki, and R. Horodecki, IBM J. Res.

Dev. **48**, 139 (2004).
[7] R. Jozsa, IBM J. Res. Dev. **48**, 79 (2004).
[8] M. Horodecki, R. Horodecki, A. Sen(De), and U. Sen, e-print quant-ph/0407038.
[9] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. Smolin, and W. K. Wootters, Phys. Rev. Lett. **78**, 2031 (1997).
[10] C. H. Bennett, D. P. DiVincenzo, C. A. Fuchs, T. Mor, E. Rains, P. W. Shor, J. A. Smolin, and W. K. Wootters, Phys. Rev. A **59**, 1070 (1999).
[11] W. Zurek, Ann. Phys. **9**, 853 (2000).
[12] H. Ollivier and W. H. Zurek, Phys. Rev. Lett. **88**, 017901 (2002).
[13] L. Henderson and V. Vedral, J. Phys. A **34**, 6899 (2001).
[14] J. Oppenheim, M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. Lett. **89**, 180402 (2002).
[15] M. Horodecki, K. Horodecki, P. Horodecki, R. Horodecki, J. Oppenheim, A. Sen(De), and U. Sen, Phys. Rev. Lett. **90**, 100402 (2003).
[16] J. Oppenheim, K. Horodecki, M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. A **68**, 022307 (2003).
[17] J. Oppenheim, M. Horodecki, and R. Horodecki, Phys. Rev. Lett. **90**, 010404 (2003).
[18] K. Maruyama, F. Morikoshi, and V. Vedral, Phys. Rev. A **71**, 012108 (2005).
[19] B. Synak, K. Horodecki, and M. Horodecki, e-print quant-ph/0405149.
[20] I. Devetak, e-print quant-ph/0406234.
[21] C. H. Bennett, Int. J. Theor. Phys. **21**, 905 (1982).
[22] R. Landauer, IBM J. Res. Dev. **5**, 183 (1961).
[23] L. Szilard, Z. Phys. **53**, 840 (1929).
[24] M. Horodecki, P. Horodecki, and J. Oppenheim, Phys. Rev. A **67**, 062104 (2003).
[25] R. Alicki, M. Horodecki, P. Horodecki, and R. Horodecki, Open Sys. Inf. Dyn. **11**, 205 (2004).
[26] An interesting twist to this paradigm has recently been added in [18]. There, the authors considered I_l which is *not* optimized over Alice and Bob's measurements. These considerations resulted in entropic inequalities related to Bell inequalities. In general, unoptimized drawing of work (or information) is better using entangled states because the correlations are stronger.
[27] R. Horodecki, M. Horodecki, and P. Horodecki, Phys. Rev. A **60**, 4144 (1999).
[28] V. Vedral, M. B. Plenio, M. A. Rippin, and P. L. Knight, Phys. Rev. Lett. **78**, 2275 (1997).
[29] V. Vedral and M. B. Plenio, Phys. Rev. A **57**, 1619 (1998).
[30] Y. Aharanov (unpublished).
[31] D. M. Greenberger, M. A. Horne, and A. Zeilinger, *Bell's Theorem, Quantum Theory, and Conceptions of the Universe*, edited by M. Kafatos (Kluwer, Dordrecht, 1989).
[32] L. Levitin, in *Proceedings of the Workshop of Physics and Computation, Physics and Computation '92* (IEEE Computer Society Press, Silver Spring, MD, 1993), pp. 223–226.
[33] B. Schumacher, Phys. Rev. A **51**, 2738 (1995).
[34] D. Petz and M. Mosonyi, J. Math. Phys. **42**, 4857 (2001).
[35] F. Hiai and D. Petz, Commun. Math. Phys. **143**, 99 (1991).
[36] L. J. Schulman and U. Vazirani e-print quant-ph/9804060.
[37] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, Phys. Rev. A **53**, 2046 (1996).
[38] S. Popescu and D. Rohrlich, Phys. Rev. A **56**, R3319 (1997).
[39] M. Donald, M. Horodecki, and O. Rudolph, J. Math. Phys. **43**, 4252 (2002).
[40] G. Vidal, J. Mod. Opt. **47**, 355 (2000).
[41] M. Horodecki, QIC, **1**, 3 (2001).
[42] B. Synak and M. Horodecki, J. Phys. A **37**, 11465 (2004).
[43] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Phys. Rev. A **54**, 3824 (1996).
[44] R. F. Werner, Phys. Rev. A **40**, 4277 (1989).
[45] R. Horodecki, *Correlation and Information-theoretic Aspects of Quantum Non-separability of Mixed States* (Uniwersytet Gdnski, Gdansk, 1996).
[46] S. Popescu, Phys. Rev. Lett. **74**, 2619 (1995).
[47] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. Lett. **78**, 574 (1997).
[48] N. Linden, S. Popescu, B. Schumacher, and M. Westmoreland, e-print quant-ph/9912039.
[49] K. Horodecki, M. Horodecki, P. Horodecki, and J. Oppenheim, e-print quant-ph/0404096.
[50] E. Rains, e-print quant-ph/0008047.
[51] M. Plenio, S. Virmani, and P. Papadopoulos, J. Phys. A **33**, L193 (2000).
[52] W. Dür, G. Vidal, and J. I. Cirac, Phys. Rev. A **62**, 062314 (2000).
[53] Y. Aharonov and S. Popescu (private communication).
[54] S. Bravyi, Phys. Rev. A **67**, 012313 (2003).
[55] A. Acin, A. Andrianov, L. Costa, E. Jane, J. I. Latorre, and R. Tarrach, Phys. Rev. Lett. **85**, 1560 (2000).
[56] H. Carteret, A. Higuchi, and A. Sudbery, J. Math. Phys. **41**, 7932 (2000).
[57] T.-C. Wei, K. Nemoto, P. M. Goldbart, P. G. Kwiat, W. J. Munro, and F. Verstraete, Phys. Rev. A **67**, 022110 (2003).
[58] K. Audenaert, J. Eisert, E. Jane, M. Plenio, S. Virmani, and B. D. Moor, Phys. Rev. Lett. **87**, 217902 (2001).
[59] I. Devetak and A. Winter, e-print quant-ph/0304196.
[60] C. H. Bennett, D. P. DiVincenzo, T. Mor, P. W. Shor, J. A. Smolin, and B. M. Terhal, Phys. Rev. Lett. **82**, 5385 (1999).
[61] D. DiVincenzo, T. Mor, P. Shor, J. Smolin, and B. Terhal, Commun. Math. Phys. **238**, 379 (2003).
[62] J. Walgate and L. Hardy, Phys. Rev. Lett. **89**, 147901 (2002).
[63] S. Ghosh, G. Kar, A. Roy, A. Sen(De), and U. Sen, Phys. Rev. Lett. **87**, 277902 (2001).
[64] S. Ghosh, G. Kar, A. Roy, D. Sarkar, A. Sen(De), and U. Sen, Phys. Rev. A **65**, 062307 (2002).
[65] M. Horodecki, A. Sen(De), U. Sen, and K. Horodecki, Phys. Rev. Lett. **90**, 047902 (2003).
[66] P. X. Chen and C. Z. Li, Phys. Rev. A **68**, 062107 (2003).
[67] D. P. DiVincenzo, M. Horodecki, D. W. Leung, J. A. Smolin, and B. M. Terhal, Phys. Rev. Lett. **92**, 067902 (2004).
[68] W. H. Zurek, Phys. Rev. A **67**, 012320 (2003).
[69] B. M. Terhal, M. Horodecki, D. P. DiVincenzo, and D. Leung, J. Math. Phys. **43**, 4286 (2002).
[70] C. King, e-print quant-ph/0103156.
[71] M. Jakob and J. Bergou, e-print quant-ph/0302075.
[72] B. Groisman, S. Popescu, and A. Winter, e-print quant-ph/0410091.
[73] L. Pankowski (private communication).