Quantum, classical, and total amount of correlations in a quantum state

Berry Groisman,[1,*] Sandu Popescu,[1,2,†] and Andreas Winter[3,‡]
[1]H. H. Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol BS8 1TL, United Kingdom
[2]Hewlett-Packard Laboratories, Stoke Gifford, Bristol BS12 6QZ, United Kingdom
[3]Department of Mathematics, University of Bristol, Bristol BS8 1TW, United Kingdom
(Received 1 February 2005; published 13 September 2005)

We give an operational definition of the quantum, classical, and total amounts of correlations in a bipartite quantum state. We argue that these quantities can be defined via the amount of work (noise) that is required to erase (destroy) the correlations: for the total correlation, we have to erase completely, for the quantum correlation we have to erase until a separable state is obtained, and the classical correlation is the maximal correlation left after erasing the quantum correlations. In particular, we show that the total amount of correlations is equal to the quantum mutual information, thus providing it with a direct operational interpretation. As a by-product, we obtain a direct, operational, and elementary proof of strong subadditivity of quantum entropy.

DOI: 10.1103/PhysRevA.72.032317 PACS number(s): 03.67.Mn, 03.65.Ud, 03.65.Yz

I. INTRODUCTION

Landauer [1], in analyzing the physical nature of (classical) information, showed that the amount of information stored, say, in a computer's memory, is proportional to the work required to erase the memory (reset to zero all the bits). These ideas were further developed by other researchers (most prominently Bennett) into a deep connection of classical information and thermodynamics (see [2] for a recent survey). Here we follow Landauer's idea in analyzing quantum information: we want to measure correlation by the (thermodynamical) effort required to erase (destroy) it.

The main idea of our paper can be understood on a simple example. Consider a maximally entangled state of two qubits (equivalent to a singlet)

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A \otimes |0\rangle_B + |1\rangle_A \otimes |1\rangle_B). \quad (1)$$

Usually this state is seen as containing 1 ebit, i.e., one bit of entanglement, based on the asymptotic theory of pure-state entanglement [3]. The temptation is to think that it contains 1 bit of correlation, and that this correlation is in pure quantum form (which can be used either in a quantum way—e.g., for teleportation—or to obtain one perfectly correlated classical bit).

We will argue however that this state contains in fact 2 bits of correlation—1 bit of entanglement and 1 bit of (secret) classical correlations, as follows.

Suppose that Alice wants to erase the entanglement between her bit and Bob's. She can do this by applying 1 bit of *randomness*: she applies to her qubit one of two unitary transformations $\mathbb{1}$ or σ_z with equal probability. By this the pure state in Eq. (1) becomes a mixture

$$\rho = \frac{1}{2}|\Phi^+\rangle\langle\Phi^+| + \frac{1}{2}|\Phi^-\rangle\langle\Phi^-|,$$

where

$$|\Phi^-\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A \otimes |0\rangle_B - |1\rangle_A \otimes |1\rangle_B).$$

This mixed state is disentangled because it is identical with a mixture of two direct product states

$$\rho = \frac{1}{2}|0\rangle\langle 0|_A \otimes |0\rangle\langle 0|_B + \frac{1}{2}|1\rangle\langle 1|_A \otimes |1\rangle\langle 1|_B.$$

But although the entanglement is now gone, Alice's and Bob's qubits are still correlated. Indeed, ρ contains now 1 bit of purely classical correlations; furthermore, these correlations are secret since, given the procedure by which they were obtained, they are not correlated with any third party, such as an eavesdropper.

To also erase these classical correlations Alice has to "work" more. She can do this by randomly applying a "bit flip" to ρ, that is, applying either $\mathbb{1}$ or σ_x at random, with equal probability . This brings the state to

$$\rho' = \frac{1}{2}\mathbb{1}_A \otimes \frac{1}{2}\mathbb{1}_B,$$

where qubit A is completely independent of qubit B.

To summarize, two bits of erasure (or, depending on the point of view, "bits of noise," or "error"), are required to completely erase the correlations in the singlet. The first bit erases the entanglement and the second erases the classical secret correlations that are left after the entanglement is gone. We then say that the singlet contains 1 bit of pure entanglement, *and* 1 bit of secret classical correlation. The total amount of correlation is 2 bits. We emphasize, however, that this total amount only makes sense in the above operational description; obviously, Alice and Bob cannot make use of both bits of correlation simultaneously.

A couple of remarks concerning the connection to Landauer's theory of information erasure. Just as Landauer did for

*Electronic address: b.groisman@bris.ac.uk
†Electronic address: s.popescu@bris.ac.uk
‡Electronic address: a.j.winter@bris.ac.uk

information (entropy), our approach quantifies correlations via their robustness against destruction. However, there seems to be a contradiction: whereas Landauer considers resetting the memory to a standard state (and we take for granted that one can generalize his argument to quantum memory), effectively exporting—"dissipating"—the entropy of the system, we *inject* entropy into it. This is actually only an apparent contradiction, as can be seen easily once we realize that in the above example we tacitly assumed that Alice forgets which Pauli operator she has applied. Indeed, we can present what she does in more detail as follows. She has a reservoir of random bits, which she uses to apply one of the Pauli operators as above in a reversible way (by a quantum-controlled unitary). This step does not affect the correlations between A and B. Only when she decides to erase (forget) the random bits, then the correlations are affected, as we have shown above. Now it is evident that the entropy pumped into the state is equal to the Landauer erasure cost of the random bits.

In this paper we develop these ideas, as follows.

For an arbitrary bipartite quantum state ρ_{AB} the quantum mutual information is defined as

$$I(A:B) = S(\rho_A) + S(\rho_B) - S(\rho_{AB}).$$

(The name is taken from Cerf and Adami [4], but Stratonovich [5] considered this quantity already in the mid-1960s.)

While this definition is formally very simple, an operational interpretation for it was hitherto missing [6] (at least not for the quantity itself with given state; however, it plays a crucial role in the formula for the entanglement-assisted capacity of a quantum channel [7]). We show here that the total amount of correlation, as measured by the minimal rate of randomness that is required to completely erase all the correlations in ρ_{AB} (in a many-copy scenario), is equal to the quantum mutual information. This is the main result of Sec. II. As an important consequence of this result we shall demonstrate that it leads to the *strong subadditivity* of von Neumann entropy.

In our above example this amount of total correlation divides neatly into the amount required to obliterate the quantum correlations (1 bit), and the amount to take the resulting separable state to a product state (1 bit). We will follow on this in our discussion contained in Sec. III, where we use this approach to define the quantum and the classical correlations in a state, and conjecture how they compare with the total correlations.

Section IV contains some observations how the total, quantum, and classical correlations as defined here relate to other such measures.

Then, in Sec. V, we extend our considerations to correlations (quantum and classical) of more than two players, after which we conclude.

An Appendix quotes the technical results about typical subspaces and our main tool, an operator version of the classical Chernoff bound, which are used repeatedly, as well as miscellaneous proofs.

II. TOTAL BIPARTITE CORRELATIONS

As explained in the Introduction, we want to add randomness to a state $\rho = \rho_{AB}$ of a bipartite system AB (with local Hilbert space dimensions $d_A, d_B < \infty$) in such a way as to make it into a product state. In fact, we shall consider $n \to \infty$ many copies of ρ, and be content with achieving decorrelation (product state) approximately (but arbitrarily well in the asymptotic limit).

In detail, the randomization will be engineered by an ensemble of local unitaries $\{p_i, U_i \otimes V_i\}_{i=1}^N$, to which is associated the randomizing map

$$R: \tau \mapsto \sum_{i=1}^N p_i (U_i \otimes V_i) \tau (U_i \otimes V_i)^\dagger. \qquad (2)$$

We call the class of such completely positive and trace preserving (CPTP) maps on AB "coordinated local unitary randomizing" (COLUR) maps. Considering that our object is to study the correlation between A and B, it may seem a bit suspicious to allow coordinated application of U_i and V_i at the two sites. Hence we define A-LUR to be those maps where all $V_i = \mathbb{1}$, and B-LUR those where all $U_i = \mathbb{1}$—because they can be implemented by application of noise strictly locally at A or B alone, respectively. The combination of an A-LUR with a B-LUR map (i.e., independent local noise at either side) we call simply "local unitary randomizing" (LUR).

We say that R *ϵ-decorrelates* a state ρ if there is a product state $\omega_A \otimes \omega_B$ such that

$$\|R(\rho) - \omega_A \otimes \omega_B\|_1 \leq \epsilon, \qquad (3)$$

where $\|\cdot\|_1$ is the trace norm of an operator, i.e., the sum of the absolute values of the eigenvalues. For technical reasons, when we study the asymptotics of such transformations (i.e., acting on n copies of the state ρ), we will demand that the output of the map R (and similar maps studied below) is supported on a space of dimension d^n, for all n, with some finite d.

How shall we account for the amount of noise introduced? From the point of view of the ensemble of unitaries, the most conservative option will be to take $\log N$ (all logarithms in this paper are taken to base 2), the space required to identify the element i uniquely. A smaller, and in the many-copy asymptotic meaningful, quantity would be $H(p) = -\sum_i p_i \log p_i$. Note, however, that they are not uniquely associated with the randomizing map R. However, Schumacher [8], and earlier Lindblad [9], have proposed a measure of the entropy a CPTP map T injects into the system P on which it acts: for this purpose, one has to introduce an environment E, which is initially in a pure state, and to fix a reference system Z, which purifies ρ_P to $|\psi\rangle_{ZP}$—note that all such purifications are related via unitaries on Z. Then, the entropy exchange is defined as

$$S_e(T, \rho_P) \coloneqq S((\mathrm{id}_Z \otimes T_P)|\psi\rangle\langle\psi|).$$

It is the entropy the environment (initially in a pure state) acquires in a unitary dilation of the CPTP map. In this paper, P will be a bipartite system AB.

Based on elementary properties of the von Neumann entropy, one can see that for every randomizing map R as above, and every state ρ,

$$\log N \geqslant H(p) \geqslant S_e(R,\rho). \tag{4}$$

Proposition 1. Consider any COLUR map on the bipartite system A^nB^n,

$$R:\tau \mapsto \sum_{i=1}^{N} p_i (U_i \otimes V_i)\tau(U_i \otimes V_i)^\dagger,$$

which ϵ-decorrelates $\rho^{\otimes n}$. Then the entropy exchange of R relative to $\rho^{\otimes n}$ is lower bounded

$$S_e(R,\rho^{\otimes n}) \geqslant n[I(A:B) - 3\epsilon \log d - \eta(3\epsilon)], \tag{5}$$

where

$$\eta(x) := \begin{cases} -x\log x & \text{for } x \leqslant \dfrac{1}{e}, \\ \dfrac{1}{e}\log e & \text{for } x \geqslant \dfrac{1}{e}. \end{cases}$$

In particular, the right-hand side is also a lower bound on $H(p)$, and even more so on $\log N$.

Proof. First of all, because R acts locally,

$$R_A := \mathrm{Tr}_B R(\rho^{\otimes n}) = \sum_{i=1}^{N} p_i U_i \rho_A^{\otimes n} U_i^\dagger,$$

and similarly for $R_B := \mathrm{Tr}_A R(\rho^{\otimes n})$. Hence we have (using the concavity of the von Neumann entropy)

$$S(R_A) \geqslant nS(\rho_A), \quad S(R_B) \geqslant nS(\rho_B). \tag{6}$$

On the other hand, we can argue that $R(\rho^{\otimes n})$ is very close to $R_A \otimes R_B$. Indeed, from Eq. (3) it follows that

$$\|R_A - \omega_A\|_1 \leqslant \|R(\rho^{\otimes n}) - \omega_A \otimes \omega_B\|_1 \leqslant \epsilon.$$

Similarly,

$$\|R_B - \omega_B\|_1 \leqslant \epsilon.$$

Thus, by the triangle inequality,

$$\|R_A \otimes R_B - \omega_A \otimes \omega_B\|_1 \leqslant 2\epsilon,$$

and we get

$$\|R(\rho^{\otimes n}) - R_A \otimes R_B\|_1 \leqslant 3\epsilon. \tag{7}$$

Hence, by the Fannes inequality [10],

$$S(R_A) + S(R_B) - S(R(\rho^{\otimes n})) \leqslant 3\epsilon \log d^n + \eta(3\epsilon). \tag{8}$$

Taking into account Eq. (6) we obtain

$$S(R(\rho^{\otimes n})) \geqslant n[S(\rho_A) + S(\rho_B) - 3\epsilon \log d - \eta(3\epsilon)]. \tag{9}$$

Here we use the fact that multiplying the last term in Eq. (8) by n will only weaken the inequality. Now, introduce a purifying reference system Z for our state, $\rho = \mathrm{Tr}_Z \psi$, with a pure state $\psi = |\psi\rangle\langle\psi|$ on ZAB. Then the randomizing map acts on A^nB^n, producing the state

$$\Omega = (\mathrm{id}_Z^{\otimes n} \otimes R)(\psi^{\otimes n})$$

on $Z^nA^nB^n$. So, by definition of the entropy exchange,

$$\begin{aligned} S_e(R,\rho^{\otimes n}) &= S(\Omega_{Z^nA^nB^n}) \geqslant S(\Omega_{A^nB^n}) - S(\Omega_{Z^n}) \\ &= S(R(\rho^{\otimes n})) - S(\rho^{\otimes n}) \\ &\geqslant n[S(\rho_A) + S(\rho_B) - S(\rho) - 3\epsilon \log d - \eta(3\epsilon)], \end{aligned}$$

where in the first line we have used the Araki-Lieb (or triangle) inequality [11], and in the second line the fact that R acted only on A^nB^n, i.e., initially $S(\rho_{Z^n}) = S(\rho_{A^nB^n})$; in the last line we have inserted Eq. (9). ■

On the other hand, we have the following proposition.

Proposition 2. For any state ρ and $\epsilon > 0$ there exists, for all sufficiently large n, an A-LUR map

$$R:\tau \mapsto \frac{1}{N}\sum_{i=1}^{N} (U_i \otimes \mathbb{1})\tau(U_i \otimes \mathbb{1})^\dagger$$

on A^nB^n, which ϵ-decorrelates $\rho^{\otimes n}$, and with

$$\log N \leqslant n[I(A:B) + \epsilon].$$

Proof. For large n, we change the state $\rho^{\otimes n}$ very little by restricting it to its typical subspace, with projector Π (see the Appendix), and even restricting the systems A^n (B^n) to the local typical subspaces of $\rho_A^{\otimes n}$ ($\rho_B^{\otimes n}$), with projector Π_A (Π_B):

$$\hat{\rho} := (\Pi_A \otimes \Pi_B)\Pi\rho^{\otimes n}\Pi(\Pi_A \otimes \Pi_B). \tag{10}$$

By definition of the typical subspace projectors,

$$\|\hat{\rho} - \rho^{\otimes n}\|_1 \leqslant \epsilon + \sqrt{8(2\epsilon)} \leqslant 5\sqrt{\epsilon},$$

using the "gentle measurement lemma" 2.

From the properties of the typical projectors (see again the Appendix) we obtain that $\hat{\rho}$ is an operator of trace $\geqslant 1 - 3\epsilon$ supported on a tensor product of (typical sub)spaces of dimensions $D_A \leqslant 2^{n(S(\rho_A)+\epsilon)}$ and $D_B \leqslant 2^{n(S(\rho_B)+\epsilon)}$, and such that

$$\hat{\rho} \leqslant \frac{1}{D}\Pi_A \otimes \Pi_B,$$

where $D = 2^{n(S(\rho)-\epsilon)}$. It is for this latter property that we needed to put the global typical projector Π in the definition of $\hat{\rho}$, Eq. (10).

For the following argument we will also need a lower bound on the reduced state on B, which we engineer by a further reduction. Define the projection Π_B' on the subspace where $\mathrm{Tr}_A \hat{\rho} \geqslant \epsilon / D_B$, and let

$$\tilde{\rho} := (\mathbb{1}_A \otimes \Pi_B')\hat{\rho}(\mathbb{1}_A \otimes \Pi_B').$$

Then it is immediate that $\mathrm{Tr}\tilde{\rho} \geqslant \mathrm{Tr}\hat{\rho} - \epsilon \geqslant 1 - 4\epsilon$; hence by the gentle measurement lemma 2

$$\|\tilde{\rho} - \hat{\rho}\|_1 \leqslant \sqrt{8\epsilon},$$

and we can keep for later reference the approximation

$$\|\tilde{\rho} - \rho^{\otimes n}\|_1 \leqslant 8\sqrt{\epsilon}. \tag{11}$$

Observe that we have defined all these projections in such a way that

$$\omega_B' \coloneqq \mathrm{Tr}_A\tilde{\rho} \geq \frac{\epsilon}{D_B}\Pi_B'.$$

Now take any ensemble of unitaries $\{p(dU),U\}$ such that for all states φ from the typical subspace of $\rho_A^{\otimes n}$,

$$\int_U p(dU)U\varphi U^\dagger = \frac{1}{D_A}\Pi_A \eqqcolon \omega_A$$

(a *private quantum channel* in the terminology of [12]), for example, the discrete Weyl operators on the typical subspace of $\rho_A^{\otimes n}$, but all unitaries on that subspace with corresponding Haar measure are good, as well. (The unitaries can behave in any way outside the subspace.) By elementary linear algebra,

$$\int_U p(dU)(U\otimes \mathbb{1})\tilde{\rho}(U^\dagger \otimes \mathbb{1}) = \omega_A \otimes \omega_B'.$$

Now we show, using the "operator Chernoff bound," Lemma 3 in the Appendix, that we can select a small subensemble of these unitaries doing the same job to sufficient approximation (this is an argument like those used in [13]). To this end, we understand Alice's local unitary U as a random variable with distribution $p(dU)$, and define the operator-valued random variable

$$X \coloneqq D(U\otimes \mathbb{1})\tilde{\rho}(U^\dagger\otimes \mathbb{1}).$$

By the above, $0 \leq X \leq \mathbb{1}$ and

$$\mathbb{E}X = D\omega_A \otimes \omega_B' \geq \epsilon 2^{-n[I(A:B)+3\epsilon]}\Pi_A\otimes\Pi_B'.$$

Thus, if $X_1,\ldots,X_N$ are independent realizations of X, Lemma 3 yields

$$\Pr\left(\frac{1}{N}\sum_{i=1}^N X_i \notin [(1-\epsilon)\mathbb{E}X;(1+\epsilon)\mathbb{E}X]\right) \leq 2d_A^n d_B^n \exp(-N\epsilon 2^{-n[I(A:B)+3\epsilon]}\epsilon^2/2)$$

where the factor 2 on the right-hand side follows from adding the two probability bounds of Lemma 3. For $N = 2^{n[I(A:B)+4\epsilon]}$ or larger (and sufficiently large n) this is smaller than 1, and we can conclude that there exist $U_1,\ldots,U_N$ from the *a priori* ensemble such that

$$(1-\epsilon)\omega_A\otimes\omega_B' \leq \frac{1}{N}\sum_{i=1}^N (U_i\otimes\mathbb{1})\tilde{\rho}(U_i\otimes\mathbb{1})^\dagger \leq (1+\epsilon)\omega_A\otimes\omega_B'.$$

Note that it is enough to show that this probability is just smaller than 1, i.e., that at least one such set of unitaries exists.

Putting this together with Eq. (11), we get

$$\left\|\frac{1}{N}\sum_{i=1}^N (U_i\otimes\mathbb{1})\rho^{\otimes n}(U_i\otimes\mathbb{1})^\dagger - \omega_A\otimes\omega_B'\right\|_1 \leq \epsilon + 8\sqrt{\epsilon};$$

hence for the state $\omega_B \coloneqq \omega_B'/\mathrm{Tr}\,\omega_B'$,

$$\left\|\frac{1}{N}\sum_{i=1}^N (U_i\otimes\mathbb{1})\rho^{\otimes n}(U_i\otimes\mathbb{1})^\dagger - \omega_A\otimes\omega_B\right\|_1 \leq 5\epsilon + 8\sqrt{\epsilon}.$$

The last inequality shows that the map R we have constructed does indeed $(5\epsilon+8\sqrt{\epsilon})$-decorrelate $\rho^{\otimes n}$. ■

Putting Eq. (4) and Propositions 1 and 2 together, we obtain the (robust) asymptotic measure of total correlation in a quantum state.

Theorem 1. The total correlations in a bipartite state ρ_{AB}, as measured by the asymptotically minimal amount of local noise one has to add to turn it into a product [let us denote this $C_{\mathrm{er}}(\rho)$, the correlation of erasure of ρ], is $I(A:B) = S(\rho_A)+S(\rho_B)-S(\rho_{AB})$. Mathematically,

$$\sup_{\epsilon>0}\lim_{n\to\infty}\inf\frac{1}{n}\min\{S_e(R,\rho^{\otimes n}):R\ \epsilon\text{-decorr. COLUR}\}$$

$$= \sup_{\epsilon>0}\lim_{n\to\infty}\sup\frac{1}{n}\min\{\log N:R\ \epsilon\text{-decorr. } A\text{-LUR}\} = I(A:B).$$

So, whether we allow general LUR ensembles or ones restricted to A (or B), whether we count conservatively the size of the ensemble, $\log N$, or are lax and charge only the entropy exchange, and whether we define the best rate optimistically or pessimistically, it all comes down to the quantum mutual information as the optimal noise (erasure) rate to remove the total correlation. ■

In passing we note that this implies the perhaps surprising result that the three ways of measuring the noise in Eq. (4) are asymptotically equivalent, as expressed in Propositions 1 and 2. In [14] the authors argue that the entropy exchange is a way of measuring the noise of a CPTP map based on compressibility—it seems to us that the connection to that work is the following: while one can always change the basis of the environment to interpret the entropy exchange as the "entropy of Kraus operators acting," this change of basis will turn our initially unitary Kraus operators into something else. We instead want to modify the CPTP map so as to preserve the entropy exchange *and* unitarity of the Kraus operators.

We now want to present a line of thought intended to reconcile our earlier doubts whether allowing coordinated LUR would be a well-behaved concept. This is based on the realization that providing the players with the perfectly correlated data i (with probability p_i) is effectively giving them another state $\gamma = \sum_i p_i |i\rangle\langle i|_A \otimes |i\rangle\langle i|_B$. This gives us the idea of regarding the situation as a kind of catalysis; the task, for given (general) γ, is to decorrelate $\rho\otimes\gamma$, but we will have to discount the overhead $C_{\mathrm{er}}(\gamma)$ of just erasing the correlations in γ.

So we really want to consider the infimum (over all γ) of the erasure cost of $\rho\otimes\gamma$ minus the cost of γ. Of course, in the light of our Theorem 1, this is $I(A:B)$ (which means that allowing catalysis does not change the content of our theorem). Conceptually, however, we gain an insight: supposing we allow only LUR in the randomization, then giving the parties a perfect correlation γ allows them the following strategy: they use the perfect correlation to implement a general COLUR map to erase the correlations in ρ and after this the one in γ. We do not need to know how much the latter costs because we subtract the same cost anyway.

Thus, even though we may be restricted to LUR at first, the availability of appropriate γ in a catalytic scenario effectively motivates consideration of general COLUR maps. It is a nice observation, though, that in Theorem 1 we can locally restrict to A-LUR without the need to resort to catalysts.

Remark 1. It may be worth noting that our lower bound in Proposition 1 is valid for an even larger class of operations, namely, "local unital" (LUN) CPTP maps: these are compositions of unital (i.e., identity preserving) maps locally at A and at B. This is because all we need for the argument is that the local entropies of Alice and Bob can only increase under the map, which is exactly the property of unital CPTP maps; the rest of the proof is the same (observe in particular that entropy exchange makes sense for whatever CPTP map we have, not just mixtures of unitaries). Clearly LUR is a subset of LUN, and we can even emulate COLUR maps by including catalysis in the sense of the previous remarks.

We can interpret this result intuitively using our explanation of our approach in terms of (reversible) local unitaries and Landauer erasure, as given in the Introduction; namely, it is well known that unital maps T are exactly those that admit a dilation

$$T(\varphi) = \mathrm{Tr}_E\left[U\left(\varphi \otimes \frac{1}{d_E}\mathbb{1}_E\right)U^\dagger\right].$$

Hence, the local unital maps of Alice and Bob can be understood as reversibly interacting their registers with local noise, with subsequent erasure of that noise. The cost of the latter is bounded by the entropy exchange.

Corollary 1 (strong subadditivity). For any tripartite state ρ_{ABC},

$$I(A:C|B) = S(\rho_{AB}) + S(\rho_{BC}) - S(\rho_{ABC}) - S(\rho_B) \geqslant 0.$$

Proof. The strong subadditivity inequality as expressed above is equivalent to

$$I(A:BC) \geqslant I(A:B).$$

However, by Theorem 1 above, the left-hand side is the minimum local noise necessary and sufficient to asymptotically decorrelate A from BC, and we may consider an A-LUR for this, i.e., randomization acting only on A. Since a map which ϵ-decorrelates $\rho_{A|BC}$ surely also ϵ-decorrelates ρ_{AB}, this minimum noise is larger than or equal to the minimum noise to decorrelate the latter state, which is the right-hand side, once more by Theorem 1.

Observe that the proof of Theorem 1 did not invoke strong subadditivity: in the lower bound, Proposition 1, we have only used concavity (Schur convexity) and subadditivity of the entropy; in the upper bound, Proposition 2, only typical subspaces and random coding were employed. ■

Remark 2. While it is worth noting that in our noise model we have not allowed communication between the parties, and that indeed (and unsurprisingly) communication can decrease as well as increase the total correlation, our result shows that the total correlation $C(\rho)$ is indeed monotonic under local operations and classical communication (LOCC) in the following sense.

Every LOCC is a succession of steps of the form that Alice (Bob) performs a quantum instrument [15] locally, transforming the state ρ into an ensemble $\{p_i,\rho_i\}$, of which she (he) communicates i to the other party. In general, such a local quantum instrument can be characterized by adding an ancillary system A' on, say, Alice's side and letting A' interact with an original subsystem A. Thus, the transformation

$$\rho_{AB} \otimes \rho_{A'} \mapsto \sum_i p_i|i\rangle\langle i|_{A'} \otimes (\rho_i)_{AB} =: \sigma_{AA'B}$$

is implemented locally by a CPTP map. By adding a local ancilla Alice cannot change the quantum mutual information between her and Bob, i.e., initially

$$I(AA':B)_{\rho_{AB}\otimes\rho_{A'}} = I(A:B)_{\rho_{AB}}. \tag{12}$$

On the other hand,

$$\begin{aligned} I(AA':B)_{\rho_{AB}\otimes\rho_{A'}} &\geqslant I(AA':B)_{\sigma_{AA'B}} = I(A':B)_\sigma + I(A:B|A')_\sigma \\ &\geqslant I(A:B|A')_\sigma = \sum_i p_i I(A:B)_{\rho_i}, \end{aligned} \tag{13}$$

where in the first line we used monotonicity of I under local operations and the formal "quantum conditional mutual information," and in the second line we used standard properties of the von Neumann entropy. Combining Eqs. (12) and (13) we obtain

$$I(A:B)_{\rho_{AB}} \geqslant \sum_i p_i I(A:B)_{\rho_i}.$$

The expression on the right is the average of the total correlations after the instrument. We can interpret this as the correlation between Alice and Bob conditional on an eavesdropper who monitors the classical communication between them; in this way the common knowledge of the classical message i does not count as correlation between Alice and Bob.

III. BIPARTITE ENTANGLEMENT AND CLASSICAL CORRELATIONS

A. Quantum correlations

Now we use the same method of randomization to define an entanglement measure. It will be the minimum noise one has to add locally to a state ρ to make it a separable state σ. Of course, as in Sec. II we will adopt an asymptotic and approximate point of view.

To the disentanglement process we associate the randomizing map R as in Eq. (2). We say that R *ϵ-disentangles* a state ρ if there is a separable state $\sigma=\sum_\mu q_\mu \sigma_A^\mu \otimes \sigma_B^\mu$ such that

$$\|R(\rho) - \sigma\|_1 \leqslant \epsilon. \tag{14}$$

As in Sec. II we can (and will) restrict ourselves to LUR, keeping in mind that the appropriate γ in a catalytic scenario will easily motivate a generalization to COLUR maps.

In the previous section there was an undercurrent message that the minimum noise we have to add is the (minimal) entropy difference between the state and the target class. There it was product states achievable by LUR; here we will aim at separable states achievable by LUR (up to ϵ approximations). In detail, we can prove the following.

Proposition 3. Let T be an ϵ-disentangling map for $\rho^{\otimes n}$. Then,

$$\log N \geq H(p) \geq S_e(T,\rho^{\otimes n}) \geq \inf_{\|\sigma - R(\rho^{\otimes n})\|_1 \leq \epsilon} [S(\sigma) - nS(\rho) - n\epsilon \log d - \eta(\epsilon)],$$

where the infimum is over all COLUR maps R and separable states σ with $\|\sigma - R(\rho^{\otimes n})\|_1 \leq \epsilon$.

Proof. Just as in the proof of Proposition 1, we introduce a purification ψ of ρ on the extended system ZAB; the randomizing map acts on A^nB^n, resulting in the state

$$\Omega = (\mathrm{id}_Z^{\otimes n} \otimes T)(\psi^{\otimes n}).$$

As before, by the definition of the entropy exchange,

$$\begin{aligned} S_e(T,\rho^{\otimes n}) &= S(\Omega_{Z^nA^nB^n}) \geq S(\Omega_{A^nB^n}) - S(\Omega_{Z^n}) \\ &= S(T(\rho^{\otimes n})) - S(\rho^{\otimes n}) \\ &\geq S(\sigma) - nS(\rho) - n\epsilon \log d - \eta(\epsilon), \end{aligned}$$

where in the first line we have used the triangle inequality [11], and in the second line the fact that R acted only on A^nB^n; in the last line we have substituted the separable state σ with $\|\sigma - T(\rho^{\otimes n})\|_1 \leq \epsilon$, which exists by assumption, and have used the Fannes inequality. ■

Proposition 4. Let $k > 0$ and T be a COLUR map such that $\sigma := T(\rho^{\otimes k})$ is separable. Then for all ϵ and sufficiently large n there exists an ϵ-disentangling COLUR map R as in Eq. (2), with

$$\log N \leq n[S(\sigma) - kS(\rho) + \epsilon].$$

Proof. We assume the form of Eq. (2) for the map T. To begin with, we have for all n,

$$T^{\otimes n}(\rho^{\otimes kn}) = \sigma^{\otimes n}, \tag{15}$$

which is separable. Our goal will be to construct a COLUR map with the desired properties, which approximates $T^{\otimes n}$.

To this end, we use a typical projector Π_1 of $\rho^{\otimes kn}$ and a typical projector Π_2 of $\sigma^{\otimes n}$: for sufficiently large n, the right-hand side is changed by not more than ϵ if we sandwich the state between Π_2, and the left-hand side is changed by not more than ϵ if we replace $\rho^{\otimes kn}$ by $\tilde{\rho} := \Pi_1 \rho^{\otimes kn} \Pi_1$. [This has the effect of making $\tilde{\rho} \leq (1/D_1)\Pi_1$, with $D_1 \geq 2^{kn[S(\rho)-\epsilon]}$.] Hence,

$$\hat{\sigma} := \Pi_2 (T^{\otimes n}(\tilde{\rho}))\Pi_2$$

satisfies $\|\hat{\sigma} - \sigma^{\otimes n}\|_1 \leq 2\epsilon$.

Since $\hat{\sigma}$ is supported on a subspace of dimension $D_2 = \mathrm{Tr}\,\Pi_2 \leq 2^{n[S(\sigma)+\epsilon]}$, we alter it again only by not more than ϵ if we restrict it to the subspace where it is $\geq \epsilon/D$; denote the corresponding projector Π_3 and let $\tilde{\sigma} := \Pi_3 \hat{\sigma} \Pi_3$.

Now we are in a position to use the operator Chernoff bound once more: we understand the ensemble of unitaries defining $T^{\otimes n}$,

$$W = U_I \otimes V_I = (U_{i_1} \otimes \cdots \otimes U_{i_n}) \otimes (V_{i_1} \otimes \cdots \otimes V_{i_n})$$

as a random variable with probability density $p(W) = p_I = p_{i_1} \cdots p_{i_n}$. Now we can define random operators

$$X := D_1 \Pi_3 \Pi_2 W(\tilde{\rho}) W^\dagger \Pi_2 \Pi_3,$$

which by the above obey $0 \leq X \leq \mathbb{1}$, and

$$\mathbb{E}X = D_1 \tilde{\rho} \geq \epsilon \frac{D_1}{D_2} \Pi_3 \geq \epsilon 2^{-n[S(\sigma)-kS(\rho)+2\epsilon]} \Pi_3.$$

Hence, for independent realizations $X_1, \ldots, X_N$ of X, Lemma 3 gives

$$\Pr\left(\frac{1}{N} \sum_{j=1}^{N} X_j \notin [(1 \pm \epsilon)\mathbb{E}X] \right) \leq 2d^n \exp(-N\epsilon 2^{-n[S(\sigma)-kS(\rho)+2\epsilon]}\epsilon^2/2).$$

Hence, for $N = 2^{n[S(\sigma)-kS(\rho)+3\epsilon]}$ (and sufficiently large n), this probability is less than 1; this means that there are unitaries $W_1, \ldots, W_N$ from the original ensemble of product unitaries, such that

$$(1 - \epsilon)\tilde{\sigma} \leq \frac{1}{N} \sum_{j=1}^{N} W_j \tilde{\rho} W_j^\dagger \leq (1 + \epsilon)\tilde{\sigma}.$$

This statement, however, yields

$$\left\| \frac{1}{N} \sum_{j=1}^{N} W_j \rho^{\otimes kn} W_j^\dagger - \sigma^{\otimes n} \right\|_1 \leq 4\epsilon,$$

and we are done. ■

Remark 3. By the same proof technique as in Propositions 2 and 4 one can show that for many independent copies of a COLUR map T (acting on as many copies of a state ρ), the entropy exchange has, in the asymptotic limit, the actual character of a classical entropy rate, in the following sense: the action of the map $T^{\otimes n}$ on a purification of $\rho^{\otimes n}$ is approximated by a different COLUR map with N terms, where

$$\log N \leq n[S_e(T,\rho) + \epsilon].$$

These two propositions can be summarized in the following theorem. Let us define, for a given state ρ, integer n, and $\epsilon > 0$, $N(n,\epsilon)$ as the smallest N such that there exists an ϵ-disentangling COLUR map as in Eq. (2). Then, the *entanglement erasure* of ρ is defined as the minimal asymptotic noise rate needed to turn ρ into a separable state:

$$E_{\mathrm{er}}(\rho) := \sup_{\epsilon>0} \limsup_{n\to\infty} \frac{1}{n} \log N(n,\epsilon).$$

As usual in information theory, we also define the *optimistic entanglement erasure* by replacing the lim sup by the lim inf in the previous formula:

$$\underline{E}_{\mathrm{er}}(\rho) := \sup_{\epsilon>0} \liminf_{n\to\infty} \frac{1}{n} \log N(n,\epsilon).$$

Theorem 2. For all bipartite states $\rho = \rho_{AB}$,

$$\underline{E}_{\mathrm{er}}(\rho) \geq \sup_{\epsilon>0} \limsup_{n\to\infty} \inf_{\|\sigma - R(\rho^{\otimes n})\|_1 \leq \epsilon} \frac{1}{n} S(\sigma) - S(\rho),$$

where the infimum is over all COLUR maps R and separable states σ; and

$$E_{\mathrm{er}}(\rho) \leq \liminf_{n\to\infty} \inf_{\sigma=R(\rho^{\otimes n})} \frac{1}{n} S(\sigma) - S(\rho),$$

with the infimum is again over all COLUR maps R and separable states σ. ■

We conjecture (without proof, at the moment) that the two limits on the right-hand side coincide. Note that the main difference (apart from the uses of lim inf and lim sup) is that in the one we consider maps taking the original state to perfectly separable states, while in the other we still allow ϵ approximations (which is why we need to include the ϵ in the formula). If this conjecture turns out to be true we have warranted our intuition from the beginning of this section that the entanglement erasure is the minimal entropy one has to "add" to the state to make it separable.

It remains as a major open problem to prove this conjecture, and perhaps to find a single-copy optimization formula for the entanglement erasure E_{er}.

B. Classical correlations

Now we want to use the same approach to define and study the *classical correlation* content of a quantum state. The intuitive idea here is that what is left of the correlations after erasing the quantum part ought to be addressed as the classical correlations. In particular, a separable state has no quantum correlations, so its total correlation (quantum mutual information) should be addressed as classical correlation.

This motivates not one but two definitions of classical correlations. In the one we consider separable states σ such that there exists a LUR map R such that

(a) $\|R(\rho^{\otimes n}) - \sigma\|_1 \leq \epsilon$;

in the other, any local CPTP map $T = T_A \otimes T_B$ with

(b) $\|T(\rho^{\otimes n}) - \sigma\|_1 \leq \epsilon$.

Then let

$$C\ell_{\mathrm{er}}(\rho) := \sup_{\epsilon>0} \limsup_{n\to\infty} \sup_{\sigma \text{ s.t.(a)}} \frac{1}{n} I(A:B)_\sigma,$$

$$C\ell^*_{\mathrm{er}}(\rho) := \sup_{\epsilon>0} \limsup_{n\to\infty} \sup_{\sigma \text{s.t.(b)}} \frac{1}{n} I(A:B)_\sigma.$$

In words, $C\ell_{\mathrm{er}}(\rho)$ is the largest asymptotic total erasure cost of (near-)separable states accessible from many copies of ρ by LUR, while $C\ell^*_{\mathrm{er}}(\rho)$ extends the maximization over all states accessible by arbitrary local operations (but, as in LUR, no communication or correlation).

Of course, we use the quantum mutual information to measure the total correlations of the resulting near-separable state, because of Theorem 1. There are also "optimistic" versions of these definitions, denoted $\underline{C\ell}_{\mathrm{er}}$ and $\underline{C\ell}^*_{\mathrm{er}}$, by replacing the lim sup by lim inf; but here we will not talk about these variants.

C. The pure-state case

For a bipartite pure state $\psi = |\psi\rangle\langle\psi|$, $|\psi\rangle = \Sigma_i \sqrt{\lambda_i} |i\rangle|i\rangle$ in Schmidt form, the total correlation is $I(A:B) = 2S(\psi_A) = 2E(\psi) = 2H(\lambda)$ (with $\psi_A = \mathrm{Tr}_B \psi$), i.e., twice the entropy of entanglement. We will show that both the quantum and the classical correlations are equal to $E(\psi) = H(\lambda)$, the entropy of entanglement. This is to be expected in the light of our introductory example and from the fact of entanglement concentration [3]: indeed, for many copies of ψ, both Alice and Bob can, without much distortion of the state, restrict themselves to their respective typical subspaces, and share a state that is pretty much maximally entangled, at which point the reasoning of the Introduction should hold.

In rigorous detail, both Alice and Bob have typical subspace projectors Π_A and Π_B for their reduced states $\psi_A^{\otimes n}$ and $\psi_B^{\otimes n}$, respectively, according to Lemma 1 in the Appendix. Because of that result, we have that $\mathrm{Tr}(\psi^{\otimes n} \Pi_A \otimes \Pi_B) \geq 1 - \epsilon$ for large enough n, and the state $|\Phi\rangle := \Pi_A \otimes \Pi_B |\psi\rangle^{\otimes n}$ has Schmidt rank $D \leq 2^{n[S(\psi_A)+\epsilon]}$. On the other hand, by the gentle measurement Lemma 2, $\|\Phi - \psi^{\otimes n}\|_1 \leq \sqrt{8\epsilon} =: \delta$.

Now a pure state of Schmidt rank D can always be disentangled by a local phase randomization using D equiprobable unitaries: if $|\Phi\rangle = \Sigma_j \sqrt{f_j} |j\rangle_A |j\rangle_B$, $\Phi = |\Phi\rangle\langle\Phi|$, we let $U_k := \Sigma_j e^{2\pi i jk/D} |j\rangle\langle j|$, and have

$$\frac{1}{D} \sum_{k=1}^{D} (U_k \otimes \mathbb{1}) \Phi (U_k \otimes \mathbb{1})^\dagger = \sum_j f_j |j\rangle\langle j|_A \otimes |j\rangle\langle j|_B.$$

Hence, applying this same randomization map to $\psi^{\otimes n}$ will δ-disentangle this state.

On the other hand, let an ϵ-disentangling map R for $\psi^{\otimes n}$ be given. Then, just as in the proof of Proposition 1,

$$\begin{aligned}
\log N &\geq H(p) \geq S_e(R, \psi^{\otimes n}) \geq S(R(\psi^{\otimes n})) - S(\psi^{\otimes n}) \\
&\geq S(\sigma) - n\epsilon \log d - \eta(\epsilon) - 0 \\
&\geq S(\sigma_A) - n\epsilon \log d - \eta(\epsilon) \\
&\geq S(R(\psi^{\otimes n})_A) - 2n\epsilon \log d - 2\eta(\epsilon) \\
&\geq S(\psi_A^{\otimes n}) - 2n\epsilon \log d - 2\eta(\epsilon) \\
&\geq n[S(\psi_A) - 2\epsilon \log d - 2\eta(\epsilon)],
\end{aligned}$$

using, in this order, the triangle inequality in the first line, then the Fannes inequality [with the separable state σ which we assume to exist ϵ-close to $R(\psi^{\otimes n})$], then the inequality $S(\sigma_{AB}) \geq S(\sigma_A)$ for separable states (this is implied by the majorization result of [16]), then the Fannes inequality once more, and finally the fact that the local entropy can only increase since we use a locally unital map.

Letting $n \to \infty$ and $\epsilon \to 0$, these considerations prove that $E_{\mathrm{er}}(\psi) = \underline{E}_{\mathrm{er}}(\psi) = E(\psi) = S(\psi_A)$.

By a similarly simple consideration, we can also calculate the classical correlation of ψ (up to one only conjectured entropic inequality).

First, by simply locally dephasing the state $\psi^{\otimes n}$ in its Schmidt basis, we can obtain a separable, perfectly correlated state $\sigma^{\otimes n}$, which has as its quantum mutual information

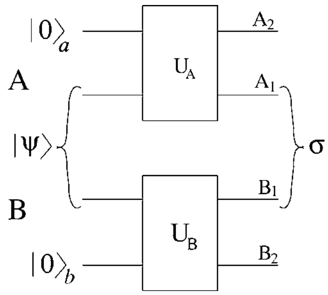

FIG. 1. One can locally implement the CPTP maps T_A and T_B using ancillas and unitaries. These unitaries rotate the initial pure state $|\psi\rangle$ to a pure state $|\zeta\rangle = (U_A \otimes U_B)(|0\rangle_a |\psi\rangle |0\rangle_b)$, which hence has the same entanglement as $|\psi\rangle$. The conjecture is thus a statement about the pure state ζ: relative to ζ, it states that $I(A_1{:}B_1) \leq S(A_1 A_2)$.

$$I(A{:}B)_{\sigma^{\otimes n}} = S(\psi_A^{\otimes n}) = nE(\psi).$$

On the other hand, to show that this is (asymptotically) optimal, we need to consider local operations T_A and T_B (now completely general, in the spirit of the definition of $C\ell_{\mathrm{er}}^*$), such that $\tau = (T_A \otimes T_B)(\psi^{\otimes n})$ is close to a separable state σ: $\|\tau - \sigma\|_1 \leq \epsilon$.

By implementing the local operations as local unitaries U_A, U_B, with ancillas which we keep for reference (compare Fig. 1), we preserve the purity of the overall state: the output state $|\vartheta\rangle = (U_A \otimes U_B)(|0\rangle_a |\psi\rangle |0\rangle_b)$ is the purification of τ. Hence (by Uhlmann's theorem) there is a purification $|\zeta\rangle$ of σ such that $\|\vartheta - \zeta\|_1 \leq \epsilon'$, with an ϵ' universally dependent on ϵ [17]. Now, invoking the Fannes inequality a couple of times,

$$\begin{aligned} nE(\psi) = E(\vartheta) &\geq E(\zeta) - n\epsilon' \log d - \eta(\epsilon') \\ &\overset{*}{\geq} I(A_1{:}B_1)_\sigma - n\epsilon' \log d - \eta(\epsilon') \\ &\geq I(A_1{:}B_1)_\tau - n(3\epsilon + \epsilon') \log d - 3\eta(\epsilon) - \eta(\epsilon'). \end{aligned}$$

Division by n, and letting $n \to \infty$ (such that $\epsilon, \epsilon' \to 0$), yields the claim that the mutual information rate can asymptotically not exceed $E(\psi)$. The inequality marked $*$ in the third line we were not able to prove rigorously (it is easily seen to be true in a great number of cases)—it is codified in the following conjecture, which we think is very plausible.

Conjecture 1. For a pure entangled state $\psi = \psi_{AB}$, and local operations T_A, T_B, such that $\sigma = (T_A \otimes T_B)(\psi)$ is separable,

$$I(A{:}B)_\sigma \leq E(\psi).$$

The major difficulty of proof stems from the fact that Alice and Bob may use quite general local operations if their goal is to maximize the classical correlation, e.g., they may apply local unitaries involving ancillas, i.e., enlarge the Hilbert space (see Fig. 1). If they do not do this, let us say for example that Alice acts only on her typical subspace, then she cannot increase her local entropy above $n[S(\psi_A) + \epsilon]$, which also is an upper bound for the mutual information of the separable state. In general, of course, we would like to be able to avoid such an assumption, and indeed the feeling is that going out of the typical subspace is suboptimal anyway.

IV. GENERAL PROPERTIES OF QUANTUM AND CLASSICAL CORRELATION; COMPARISON WITH OTHER ENTANGLEMENT MEASURES

A. Total correlation

About the total correlation $C_{\mathrm{er}}(\rho) = I(A{:}B)_\rho$ of a state we know most, primarily so because we have a usable formula. For example, because of strong subadditivity, it is monotonic under local operations, and in Remark 2 we have already argued that monotonicity extends to local operations and classical communication.

Again because of its coinciding with the quantum mutual information, we can easily relate the total correlation to distillability measures of quantum states, namely, total distillable correlation, distillable secret key, and distillable entanglement (which are decreasing in this order)

$$I(A{:}B) \geq \mathrm{CR}(\rho) \geq K(\rho) \geq E_D(\rho).$$

[For the second quantity, the common randomness $\mathrm{CR}(\rho)$ in a state, see [18]; for the third and fourth, the distillable key $K(\rho)$ and the distillable entanglement $E_D(\rho)$, see the recent results in [19].]

B. Quantum and classical correlations

Our Theorem 2 narrows down the entanglement erasure up to the regularization and getting rid of ϵ. This is not good enough to decide any of the properties we would like an entanglement measure to have—in the first place, monotonicity under local operations and classical communication. Similarly, we do not know how to prove or disprove convexity of E_{er} (a situation much in contrast to the total correlations).

On the other hand, these properties are easily seen for the second variant of our classical correlation quantity, $C\ell_{\mathrm{er}}^*$: it is monotonic under local operations (no communication allowed, of course), and it is convex.

Once more, we have at present little to offer in terms of comparing the erasure (quantum and classical) correlation measures to other quantifications of entanglement and classical correlation; clearly, we would like E_{er} to be an upper bound on the distillable entanglement, and some version of the classical correlation to be an upper bound on the distillable secret key. It has been suggested [20,21] that the (regularized) relative entropy of entanglement should relate to the entanglement erasure—while this would be a most interesting result, we see no clear evidence either way.

An interesting question arises when we return to the pure-state example of the Introduction, where the total correlations could be erased neatly in two steps: first by adding the minimal noise to dephase the state, and then going on from there by adding noise to classically decorrelate it. We have seen that for pure states this is so generally, even for the asymptotic cost. But *a priori*, the definitions of quantum correlations E_{er} and classical correlations $C\ell_{\mathrm{er}}$ require us to tar-

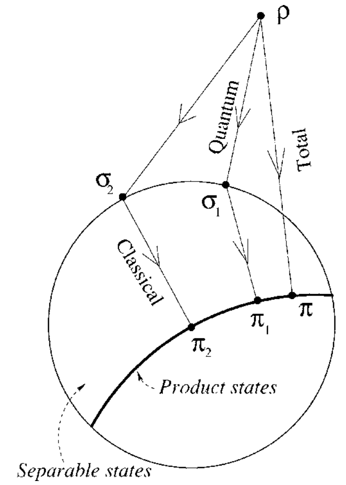

FIG. 2. Starting from ρ, this figure illustrates the different objectives one has when considering (i) the total correlations, (ii) the quantum correlations, and (iii) the classical correlations. For this purpose we have ignored the subtleties of the asymptotics, and symbolize the noise required to go from one point in state space to another by their distance. Then for (i) we seek the shortest way (minimal noise) from ρ to the manifold of product states (and we expect the target π to be $\approx\rho_A \otimes \rho_B$); for (ii) we seek instead the shortest way from ρ to the convex set of separable states, and going to the optimal point σ_1 and from there on to a product state π_1 may in total yield a suboptimal erasure procedure. Finally, for (iii), we want to go from ρ to a separable state σ_2 of maximal correlation (=distance from product states). Even if the transition from ρ to σ_2 is done by a local randomizing map, it could be that the noise cost is significantly larger than that of going from ρ to σ_1. For the pure state $\rho=\psi$ we have argued in Sec. III C that all three optimal paths coincide, and that in fact $E_{\text{er}}(\psi)=C\ell^{*}_{\text{er}}(\psi)=(1/2)C_{\text{er}}(\psi)=E(\psi)$.

get quite different separable states; Fig. 2 illustrates this point.

Heuristically, for an initial state ρ, we could have a strategy of adding local noise to turn it into a separable state σ (or close to). Theorem 2 indicates that the cost will asymptotically be the entropic gap between σ and ρ: $S(\sigma)-S(\rho)$. In the spirit of the introductory example, we then go further and completely decorrelate σ; according to Theorem 1 this will cost $I(A:B)_\sigma=S(\sigma_A)+S(\sigma_B)-S(\sigma)$ bits of noise. Hence the total cost of this two-step process will be

$$S(\sigma_A)+S(\sigma_B)-S(\rho),$$

whereas if we had destroyed the correlations in one go, we would have spent noise amounting to

$$I(A:B)_\rho = S(\rho_A)+S(\rho_B)-S(\rho),$$

which is in general smaller. We have equality if the optimal disentangling map does not increase the local entropies (or, in the asymptotic picture, only by a sublinear amount). While this seems reasonable to expect, we have no argument in favor of this expectation.

Finally, is it true that the quantum correlation, measured by the entanglement erasure E_{er}, is always smaller or equal to the classical correlation? Our and perhaps the reader's intuition would answer yes, but to prove this from our definitions seems not obvious.

V. MULTIPARTITE CORRELATIONS

By obvious generalizations of the approaches presented in the previous two sections one can also easily define total correlation and entanglement measures for more than two parties in the many-copy limit.

We do not want to go into too much detail here but discuss an aspect of the total correlation measure $C_{\text{er}}(\rho)$ of a state $\rho_{A_1\ldots A_p}$ of p parties.

By easy generalizations of Propositions 1 and 2 (and Remark 1), one obtains that

$$C_{\text{er}}(\rho)=\sum_{i=1}^{p} S(A_i)-S(A_1,\ldots,A_p). \qquad (16)$$

As before, this asymptotic measure does not depend on the details of definition, and we find a generalization of the fact that the randomization can be performed by one party alone in the bipartite case: the parties can decorrelate themselves locally one by one from the rest, and the individual costs add up to C_{er} of Eq. (16). In detail, let A_1 decorrelate herself from $A_2,\ldots,A_p$ using $I(A_1:A_2,\ldots,A_p)$ bits of randomness (by Theorem 1); then let A_2 decorrelate himself from $A_3,\ldots,A_p$ using $I(A_2:A_3,\ldots,A_p)$; etc. Then adding up these quantities yields obviously the right-hand side of Eq. (16).

VI. DISCUSSION

In this paper we have addressed the problem of an operational definition of the total, quantum, and classical amounts of correlation in a bipartite quantum state. We have shown that the above quantities can be defined via the amount of noise that is required to destroy the correlations.

We have proved that the total correlation in a bipartite quantum state, measured by the asymptotically minimal amount of noise needed to erase the correlation, equals the quantum mutual information $I(A:B)$. Thus, our approach gives the first clear operational definition of $I(A:B)$ for any given state. This even leads to an operational proof of strong subadditivity; it is an interesting question whether the equality conditions derived recently [22] can be derived in this way, too.

Then we extended our approach to definitions of the quantum (entanglement) and classical correlation content: after definitions of these quantities in the spirit of erasure, by the noise needed to destroy the entanglement, and the maximum correlation left after destroying the entanglement, we proved partial results on these quantities, and related them to other entanglement and correlation measures. In that context, we also put forward the conjecture that the amount of quantum correlations is always at most as large as the amount of classical correlations. For pure states we have verified, up to a plausible conjectured information inequality for separable states, that the proposed quantum and classical correlation measures coincide with the entropy of entanglement. In general, we had to leave open the questions of LO(CC) monotonicity and convexity of E_{er} and $C\ell_{\mathrm{er}}$. (That $C\ell^*_{\mathrm{er}}$ is monotonic under local operations is, however, trivial from the definition.)

The reader who is acquainted with the work of Horodecci and co-workers [23] will sense that there is a relation between their "thermodynamical" approach to correlations via extractable work (=purity), and ours, even though superficially we seem to go in opposite directions: we consider the entropy increase necessary to destroy correlations—and this directly gives a correlation measure; in the approach of [23] the purity content decreases as one restricts the set of allowed operations, and the "total correlation" appears as a *deficit* between global operations and local operations. If one allows also communication, the deficit is a quantum correlation measure. Recently, however, these authors have been able to relate this latter deficit to the entropy production when turning the given state into a product via certain LOCC maps [20]. Via Landauer erasure, this now looks a lot more like our model, and indeed it seems to be the case that by their including classical communication, [20] allows for a wider class of destructive operations, and consequently, the resulting entanglement measure is no larger than our E_{er}. This makes the lower bound from [20] applicable, yielding that the entanglement erasure is lower bounded by the relative entropy of entanglement (with respect to the separable set). It remains to be investigated whether there is indeed a gap between them or whether the difference is washed out in the asymptotic limit involved in both definitions.

ACKNOWLEDGMENTS

The authors acknowledge support from the U.K. Engineering and Physical Sciences Research Council (IRC "Quantum Information Processing") and the EU under European Commission project RESQ (Contract No. IST-2001-37559). S.P. and A.W., and also B.G., gratefully acknowledge the hospitality of the Isaac Newton Institute for Mathematical Sciences, during the program "Quantum Information Science," where part of the present work was done. Thanks to C. H. Bennett, I. Devetak, and N. Linden for interesting conversations on the subject of this paper, and special thanks to M. Horodecki, P. Horodecki, R. Horodecki, and J. Oppenheim for making available their paper [20] prior to publication and for discussions comparing their and our approaches.

APPENDIX A: TYPICALITY, OPERATOR CHERNOFF BOUND, AND MISCELLANEOUS RESULTS

From [24] we cite the following definitions and properties of *typical subspaces*.

For the state density operator ρ choose a diagonalization $\rho=\Sigma_i p_i|i\rangle\langle i|$ [such that $S(\rho)=H(p)$]. Then, with $I=i_1,\ldots,i_n$ and

$$p_I = p_{i_1}\cdots p_{i_n},$$

$$|I\rangle\langle I| = |i_1\rangle\langle i_1| \otimes \cdots \otimes |i_n\rangle\langle i_n|,$$

$\rho^{\otimes n}=\Sigma_I p_I|I\rangle\langle I|$. We call (with $\epsilon>0$ fixed implicitly) a state $|I\rangle$ *typical* if

$$|-\log p_I - nS(\rho)| < \epsilon n.$$

We define the *ϵ-typical subspace* to be the subspace spanned by all typical states, and Π to be the orthogonal projector onto the typical subspace (n and ϵ as before implicit).

The following theorem states the properties of the typical subspace and its projector Π (which can easily be proved by the definitions and the law of large numbers).

Lemma 1 (typical subspace theorem). For any state ρ, integer n, and $\epsilon>0$ let Π be the typical subspace projector. Then we have the following properties.

(1) For all $\delta>0$ and sufficiently large n,

$$\mathrm{Tr}(\rho^{\otimes n}\Pi) \geqslant 1-\delta.$$

In other words, by enlarging n the probability of ρ to be found in the typical subspace can be made as close to 1 as desired.

(2) For sufficiently large n, the dimension of the typical subspace equals $\mathrm{Tr}\,\Pi$, and satisfies

$$2^{n(S(\rho)-\epsilon)} \leqslant \mathrm{Tr}\,\Pi \leqslant 2^{n(S(\rho)+\epsilon)}.$$

Indeed, for all n,

$$2^{n(S(\rho)-\epsilon)}\Pi \leqslant \Pi\rho^{\otimes n}\Pi \leqslant 2^{n(S(\rho)+\epsilon)}\Pi.$$

Lemma 2 (gentle measurement [25]). Let ρ be a density operator with $\mathrm{Tr}\,\rho\leqslant 1$, and X an operator with $0\leqslant X\leqslant \mathbb{1}$, such that $\mathrm{Tr}\,\rho X \geqslant \mathrm{Tr}\,\rho-\delta$, then

$$\|\rho - \sqrt{X}\rho\sqrt{X}\|_1 \leqslant \sqrt{8\delta}.$$

(The factor 8 can be improved to 4: see [26].) Here the operator order is defined by saying that $X\geqslant Y$ if and only if $X-Y$ is positive semidefinite. This is a partial order. The interval $[A;B]$ is defined as the set of all operators X such that $A\leqslant X$ and $X\leqslant B$.

Furthermore, we shall make use of the following result.

Lemma 3 (operator Chernoff bound [27]). Let $X_1,\ldots,X_N$ be i.i.d. (independent identically distributed) random variables taking values in the operator interval $[0;\mathbb{1}]\subset\mathcal{B}(\mathbb{C}^d)$ and with expectation $\mathbb{E}X_i=M\geqslant\mu\mathbb{1}$. Then, for $0\leqslant\epsilon\leqslant 1$, and denoting $\bar{X}=(1/N)\Sigma_{i=1}^N X_i$,

$$\Pr[\bar{X}\not\leqslant(1+\epsilon)M] \leqslant d\exp\left(-N\frac{\mu\epsilon^2}{2}\right),$$

$$\Pr[\bar{X}\not\geqslant(1-\epsilon)M] \leqslant d\exp\left(-N\frac{\mu\epsilon^2}{2}\right).$$

[1] R. Landauer, IBM J. Res. Dev. **5**, 183 (1961).
[2] C. H. Bennett, Stud. Hist. Philos. Mod. Phys. **34**, 501 (2003); IBM J. Res. Dev. **32**, 16 (1988).
[3] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, Phys. Rev. A **53**, 2046 (1996).
[4] C. Adami and N. J. Cerf, Phys. Rev. A **56**, 3470 (1997).
[5] R. L. Stratonovich, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **8**, 116 (1965) [Probl. Inf. Transm. **2**, 35 (1966)].
[6] Quantum mutual information is a straightforward formal generalization of a classical concept, which has clear operational definition: it quantifies the rate of information transmission via a classical noisy channel.
[7] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, IEEE Trans. Inf. Theory **48**, 2637 (2002).
[8] B. Schumacher, Phys. Rev. A **54**, 2614 (1996); B. Schumacher and M. A. Nielsen, *ibid.* **54**, 2629 (1996).
[9] G. Lindblad, in *Quantum Aspects of Optical Communications, Proceedings of a Workshop Held at the CNRS, Paris, France, 1990*, Lecture Notes in Physics Vol. 378 (Springer-Verlag, Berlin, 1991), p. 71.
[10] M. Fannes, Commun. Math. Phys. **31**, 291 (1973).
[11] H. Araki and E. H. Lieb, Commun. Math. Phys. **18**, 160 (1970).
[12] A. Ambainis, M. Mosca, A. Tapp, and R. de Wolf, in *Proceedings of the 41st FOCS*, 2000 (unpublished), p. 547.
[13] P. Hayden, D. Leung, P. Shor, and A. Winter, Commun. Math. Phys. **250**, 371 (2004).
[14] J. Oppenheim and B. Reznik, Phys. Rev. A **70**, 022312 (2004).
[15] E. B. Davies and J. Lewis, Commun. Math. Phys. **17**, 239 (1970).
[16] M. A. Nielsen and J. Kempe, Phys. Rev. Lett. **86**, 5184 (2001).
[17] Uhlmann's theorem is usually formulated in terms of fidelity F. Here we use the fact that fidelity and trace norm are two equivalent distance measures for density operators, obeying the following relations: $F(\theta,\zeta)^2=1-\|\theta-\zeta\|_1^2$ for pure states and $\|\sigma-\tau\|_1\geqslant 1-F(\sigma,\tau)$ for mixed states.
[18] I. Devetak and A. Winter, IEEE Trans. Inf. Theory **50**, 3138 (2004).
[19] I. Devetak and A. Winter, Proc. R. Soc. London, Ser. A **461**, 207 (2005).
[20] M. Horodecki, P. Horodecki, R. Horodecki, J. Oppenheim, A. Sen(De), U. Sen, and B. Synak, Phys. Rev. A **71**, 062307 (2005).
[21] V. Vedral, M. B. Plenio, M. A. Rippin, and P. L. Knight, Phys. Rev. Lett. **78**, 2275 (1997); V. Vedral and M. B. Plenio, Phys. Rev. A **57**, 1619 (1998).
[22] P. Hayden, R. Jozsa, D. Petz, and A. Winter, Commun. Math. Phys. **246**, 359 (2004).
[23] J. Oppenheim, M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. Lett. **89**, 180402 (2002); M. Horodecki, P. Horodecki, and J. Oppenheim, Phys. Rev. A **67**, 062104 (2003); J. Oppenheim, M. Horodecki, and R. Horodecki, Phys. Rev. Lett. **90**, 010404 (2003); M. Horodecki, K. Horodecki, P. Horodecki, R. Horodecki, J. Oppenheim, A. Sen(De), and U. Sen, *ibid.* **90**, 100402 (2003).
[24] B. Schumacher, Phys. Rev. A **51**, 2738 (1995); R. Jozsa and B. Schumacher, J. Mod. Opt. **41**, 2343 (1994).
[25] A. Winter, IEEE Trans. Inf. Theory **45**, 2481 (1999).
[26] T. Ogawa and H. Nagaoka, e-print quant-ph/0208139.
[27] R. Ahlswede and A. Winter, IEEE Trans. Inf. Theory **48**, 569 (2002); **49**, 346 (2003).