Stronger subadditivity of entropy

Elliott H. Lieb* and Robert Seiringer†
Department of Physics, Jadwin Hall, Princeton University, P.O. Box 708, Princeton, New Jersey 08544, USA
(Received 5 March 2005; published 23 June 2005)

The strong subadditivity of entropy plays a key role in several areas of physics and mathematics. It states that the entropy $S[\varrho]=-\mathrm{Tr}(\varrho \ln \varrho)$ of a density matrix ϱ_{123} on the product of three Hilbert spaces satisfies $S[\varrho_{123}]-S[\varrho_{12}]\leqslant S[\varrho_{23}]-S[\varrho_{2}]$. We strengthen this to $S[\varrho_{123}]-S[\varrho_{12}]\leqslant \Sigma_\alpha n^\alpha(S[\varrho_{23}^\alpha]-S[\varrho_{2}^\alpha])$, where the n^α are weights and the ϱ_{23}^α are partitions of ϱ_{23}. Correspondingly, there is a strengthening of the theorem that the map $A\mapsto \mathrm{Tr}\exp[L+\ln A]$ is concave. As applications we prove some monotonicity and convexity properties of the Wehrl coherent state entropy and entropy inequalities for quantum gases.

DOI: 10.1103/PhysRevA.71.062329 PACS number(s): 03.67.−a, 05.30.−d

I. INTRODUCTION

The strong subadditivity of entropy (SSA), whose proof in the noncommutative case was given by Lieb and Ruskai [1,2], is one of the main ingredients in various fields of mathematics and physics in which the von Neumann/Shannon entropy plays a role. Over the years other proofs have appeared [3–5]. SSA is an inequality among various entropies that can be formed from one density matrix on the product of three Hilbert spaces $\mathcal{H}_{123}=\mathcal{H}_1\otimes\mathcal{H}_2\otimes\mathcal{H}_3$ and states that

$$S[\varrho_{123}]-S[\varrho_{12}]\leqslant S[\varrho_{23}]-S[\varrho_{2}]. \quad (1)$$

Here, ϱ_{123} is a density matrix (i.e., a positive semidefinite operator whose trace is 1) on the tensor product space $\mathcal{H}_{123}$, and ϱ_{12} is the reduced density matrix on $\mathcal{H}_{12}=\mathcal{H}_1\otimes\mathcal{H}_2$, formed by taking the partial trace of ϱ_{123} over $\mathcal{H}_3$ (i.e., $\varrho_{12}=\mathrm{Tr}_{\mathcal{H}_3}\varrho_{123}$), and so forth. The entropy $S[\varrho]$ on a Hilbert space $\mathcal{H}$ is given by the von Neumann/Shannon formula

$$S[\varrho]=-\mathrm{Tr}_{\mathcal{H}}(\varrho \ln \varrho). \quad (2)$$

(Henceforth, the Hilbert space notation $\mathcal{H}$ on the trace, Tr, will be omitted if it is not needed, or we may simply write Tr_1 to denote $\mathrm{Tr}_{\mathcal{H}_1}$, etc.; likewise, S_{123} will denote $S[\varrho_{123}]$, etc., when the meaning is clear. As a technical matter, all Hilbert spaces in this paper are assumed to be separable in order that a trace can be defined.)

Inequality (1) appears to be straightforward enough that it seems unlikely that it can be improved, i.e., that one can insert a quantity between the left and right sides that preserves the inequality. That, however, is what we do in this paper [cf. Eq. (9)]. Admittedly, our theorems can be derived from SSA [or, equivalently [6], from the monotonicity of relative entropy under completely positive trace preserving (CPT) maps] and thus, when viewed from a sufficiently remote perspective, there is little new here. From the point of view of applications and of understanding the potential of SSA, however, our results and proof techniques may merit attention, especially our applications to the theory of quantum gases in Corollary 4 of Sec. III C.

Inequality (1) is written in a slightly unusual way. Instead of the usual $S_{123}+S_2\leqslant S_{12}+S_{23}$, (1) displays the inequality as the decrease of the conditional entropy $S_{23}-S_2$ when 2 is replaced by 12, i.e., information about the state on $\mathcal{H}_1$ is added. Our focus will be on the conditional entropy.

In Sec. III we give some examples of the utility of the improved version of inequality (1), Eq. (8). As one example, we show that the "mutual information" $S_1+S_2-S_{12}$ is decreased if the density matrix is replaced by Wehrl's corresponding classical phase-space function (whose definition will be recalled later). Wehrl had shown [7] that his entropy is always greater than the true entropy, but the monotonicity of the difference $S_{12}-S_1-S_2$ is new. This is a special case of Corollary 2 below. We also show that the difference between the Wehrl and the true entropy is a convex function of the density matrix.

Originally, we had proved the monotonicity of $S_{12}-S_1$, and we are grateful to Ruskai for suggesting the stronger version to us; her argument, which uses the theory of CPT maps, is briefly sketched in Appendix C [8]. We also acknowledge other helpful correspondence about this paper.

In another direction, it will be recalled that one of the ways to prove SSA is by means of the theorem [9] (for one Hilbert space) that the map

$$A\mapsto \mathrm{Tr}\exp(L+\ln A) \quad (3)$$

for positive definite operators A is concave for each fixed self-adjoint L. This, too, will be improved, and its improvement will lead to the improved version of SSA.

Our main result is the following.

Theorem 1 (Stronger subadditivity). *Let $\mathcal{H}_i$, $i=1,2,3$, be Hilbert spaces, and let ϱ_{123} be a density matrix on $\mathcal{H}_{123}=\mathcal{H}_1\otimes\mathcal{H}_2\otimes\mathcal{H}_3$ with finite entropy. Let Ω be a measure space, with elements labeled by α, and let $d\mu(\alpha)$ be a measure on Ω. Let K^α be bounded operators on $\mathcal{H}_{12}=\mathcal{H}_1\otimes\mathcal{H}_2$ that are weakly measurable, and satisfy (with $K^{\alpha*}$ the adjoint of K^α)*

$$\int_\Omega d\mu(\alpha)K^{\alpha*}K^\alpha=\mathbb{I}_{\mathcal{H}_{12}}. \quad (4)$$

With the usual notational abuse $K^\alpha\leftrightarrow K^\alpha\otimes\mathbb{I}_{\mathcal{H}_3}$, let

*Electronic address: lieb@princeton.edu
†Electronic address: rseiring@princeton.edu

$$n^{\alpha} = \mathrm{Tr}_{123}\, K^{\alpha}\varrho_{123}K^{\alpha *} \tag{5}$$

and, in case $n^{\alpha}>0$, *let*

$$\varrho_{23}^{\alpha} = \mathrm{Tr}_1\, K^{\alpha}\varrho_{123}K^{\alpha *}/n^{\alpha}, \tag{6}$$

$$\varrho_{2}^{\alpha} = \mathrm{Tr}_3\, \varrho_{23}^{\alpha} = \mathrm{Tr}_{13}\, K^{\alpha}\varrho_{123}K^{\alpha *}/n^{\alpha}. \tag{7}$$

Then

$$S[\varrho_{123}] - S[\varrho_{12}] \leqslant \int_{\Omega} d\mu(\alpha) n^{\alpha}(S[\varrho_{23}^{\alpha}] - S[\varrho_{2}^{\alpha}]). \tag{8}$$

Remarks.

(1) We recall that weakly measurable means that $\langle\phi|K^{\alpha}|\psi\rangle$ is measurable for any vectors $|\phi\rangle$ and $|\psi\rangle$ in the Hilbert space. (This is implied, via polarization, by the seemingly weaker condition of measurability for all $|\phi\rangle=|\psi\rangle$.) The integrals then have to be interpreted in the weak sense, e.g., (4) means that $\int_{\Omega}d\mu(\alpha)\langle\phi|K^{\alpha *}K^{\alpha}|\psi\rangle=\langle\phi|\psi\rangle$ for all $|\phi\rangle$ and $|\psi\rangle$.

(2) Because of cyclicity of the trace, $n^{\alpha}=\mathrm{Tr}\, K^{\alpha *}K^{\alpha}\varrho_{123}$, and hence (4) implies that $\int_{\Omega}d\mu(\alpha)n^{\alpha}=\mathrm{Tr}\,\varrho_{123}=1$.

(3) Both sides of the inequality (8) are homogeneous of order 1 in ϱ_{123}. Hence this inequality holds also without the normalization condition $\mathrm{Tr}\varrho_{123}=1$, i.e., it holds for all positive trace class operators.

(4) It is no restriction to assume the K^{α} to be bounded. Because of (4), they must be bounded almost everywhere, and hence one can absorb their norm into the measure $d\mu(\alpha)$.

(5) In case all the K^{α} act nontrivially only on $\mathcal{H}_1$ [i.e., $K^{\alpha}=k^{\alpha}\otimes\mathbb{I}_2$ and $\int_{\Omega}d\mu(\alpha)k^{\alpha *}k^{\alpha}=\mathbb{I}_{\mathcal{H}_1}$] we have that $\int_{\Omega}d\mu(\alpha)n^{\alpha}\varrho_{23}^{\alpha}=\varrho_{23}$. Since the map $\varrho_{23}\mapsto S[\varrho_{23}]-S[\varrho_2]$ is concave, as shown in [1], the right side of (8) is bounded above by $S[\varrho_{23}]-S[\varrho_2]$ in this special case. Theorem 1 is, therefore, stronger than the usual strong subadditivity of entropy because we have

$$\begin{aligned} S[\varrho_{123}] - S[\varrho_{12}] &\leqslant \int_{\Omega} d\mu(\alpha) n^{\alpha}(S[\varrho_{23}^{\alpha}] - S[\varrho_{2}^{\alpha}]) \\ &\leqslant S[\varrho_{23}] - S[\varrho_2] \end{aligned} \tag{9}$$

in this case.

(6) Everything remains true in the classical case as well. That is, instead of Hilbert spaces $\mathcal{H}_i$ we have measure spaces (X_i,Σ_i,ν_i) (which are not necessarily related in any way with the measure μ on the space Ω), the density matrix ϱ is replaced by a non-negative measurable function on the product of the three measure spaces, and the trace is replaced by an integral. The K^{α} are then functions on the product of the measure spaces (X_1,Σ_1,ν_1) and (X_2,Σ_2,ν_2). Note that in the limit that $K^{\alpha *}K^{\alpha}$ is just a δ function supported at a point in this product measure space, labeled by α, inequality (8) is actually an *equality* in the classical case.

A. The special case of matrices and sums

To keep things simple we shall first deal with the finite dimensional case, when $\mathcal{H}_i=\mathbb{C}^{n_i}$ for finite n_i, and with the case where the integral in (4) is just a finite sum. In this special case, Theorem 1 is then just Theorem 2 below. We will first prove Theorem 2. The extension to the case of a general measure space in (4) is given in Appendix A, and the extension to the infinite dimensional case is given in Appendix B.

Theorem 2 (Stronger subadditivity, matrix case). *Let ϱ_{123} be a density matrix on a finite dimensional Hilbert space $\mathcal{H}_{123}=\mathcal{H}_1\otimes\mathcal{H}_2\otimes\mathcal{H}_3$. Let K^{α}, $1\leqslant\alpha\leqslant M$, be a finite set of operators on $\mathcal{H}_{12}=\mathcal{H}_1\otimes\mathcal{H}_2$, that satisfy*

$$\sum_{\alpha} K^{\alpha *}K^{\alpha} = \mathbb{I}_{\mathcal{H}_{12}}. \tag{10}$$

With the usual notational abuse $K^{\alpha}\leftrightarrow K^{\alpha}\otimes\mathbb{I}_{\mathcal{H}_3}$, let

$$n^{\alpha} = \mathrm{Tr}_{123}K^{\alpha}\varrho_{123}K^{\alpha *} \tag{11}$$

and, in case $n^{\alpha}>0$, *let*

$$\varrho_{23}^{\alpha} = \mathrm{Tr}_1 K^{\alpha}\varrho_{123}K^{\alpha *}/n^{\alpha}, \tag{12}$$

$$\varrho_{2}^{\alpha} = \mathrm{Tr}_3\varrho_{23}^{\alpha} = \mathrm{Tr}_{13}K^{\alpha}\varrho_{123}K^{\alpha *}/n^{\alpha}. \tag{13}$$

Then

$$S[\varrho_{123}] - S[\varrho_{12}] \leqslant \sum_{\alpha} n^{\alpha}(S[\varrho_{23}^{\alpha}] - S[\varrho_{2}^{\alpha}]). \tag{14}$$

We will give two independent proofs of Theorem 2 in the next section. The first one uses Theorem 3 below, which states the generalization of the concavity in (3) mentioned above. The second proof uses the theory of completely positive maps.

Theorem 3 (A concave map). *Let L be a self-adjoint operator on a finite dimensional Hilbert space $\mathcal{H}$. For $1\leqslant\alpha\leqslant M$, let A^{α} be positive operators on $\mathcal{H}$ and let K^{α} be operators such that $\Sigma_{\alpha}K^{\alpha *}K^{\alpha}\leqslant\mathbb{I}_{\mathcal{H}}$. Then the map*

$$(A^1,\ldots,A^M) \mapsto \mathrm{Tr}_{\mathcal{H}}\exp\left(L + \sum_{\alpha=1}^{M} K^{\alpha *}(\ln A^{\alpha})K^{\alpha}\right) \tag{15}$$

is jointly concave.

Remarks.

(1) This theorem was proved in Theorem 6 of [9] for one A and $K^{\alpha}=\mathbb{I}$, and it was generalized there, in Corollary 6.1, to $A^1,\ldots,A^M$, but only in the case that the K^{α} are non-negative numbers $\sqrt{p^{\alpha}}$ with $\Sigma_{\alpha}p^{\alpha}\leqslant 1$.

(2) Theorem 3 can be extended to the infinite dimensional case as well, using the methods of Sec. 4 in [9]. It is also possible to generalize to the case of continuous variables α in some measure space Ω; in this case, A^{α} is a measurable function on Ω with values in the positive operators, and the sum over α in (15) is replaced by the integral $\int_{\Omega}d\mu(\alpha)K^{\alpha *}(\ln A^{\alpha})K^{\alpha}$, with K^{α} satisfying $\int_{\Omega}d\mu(\alpha)K^{\alpha *}K^{\alpha}\leqslant\mathbb{I}$. For simplicity we will not give this generalization here since we will not need it for the proof of our main Theorem 1.

(3) Note the switching of K^{α} and $K^{\alpha *}$ between (11)–(13) and (15). Note also that only the inequality $\Sigma_{\alpha}K^{\alpha *}K^{\alpha}\leqslant\mathbb{I}$ is required for Theorem 3, whereas *equality* is necessary in Theorem 2.

II. PROOF OF THEOREMS 2 AND 3

Proof of Theorem 3. We start with two preliminary remarks. (a) It is clearly enough to assume that $\Sigma_\alpha K^{\alpha *}K^\alpha = \mathbb{I}$, for otherwise we can add one more $K^{M+1} = (\mathbb{I} - \Sigma_\alpha K^{\alpha *}K^\alpha)^{1/2}$ and take $A^{M+1} = \mathbb{I}$. (b) We can also assume that all K^α are invertible; the general case follows by continuity.

Let $\mathcal{K} = \mathcal{H} \otimes \mathbb{C}^M$. Every operator in $\mathcal{B}(\mathcal{K})$ can be thought of as an $M \times M$ matrix (indexed by α, β) with entries in $\mathcal{B}(\mathcal{H})$. Define $\hat{L}$, $\hat{A}$, $\hat{P} \in \mathcal{B}(\mathcal{K})$ by

$$\hat{L}_{\alpha\beta} = \frac{1}{M}\delta_{\alpha\beta}\frac{1}{K^{\alpha *}}L\frac{1}{K^\alpha}, \tag{16}$$

$$\hat{A}_{\alpha\beta} = \delta_{\alpha\beta}A^\alpha, \tag{17}$$

and

$$\hat{P}_{\alpha\beta} = K^\alpha K^{\beta *}. \tag{18}$$

Note that $\hat{P} = \hat{P}^*$ and (since $\Sigma_\alpha K^{\alpha *}K^\alpha = \mathbb{I}$) $\hat{P}^2 = \hat{P}$, so $\hat{P}$ is an orthogonal projection. We know from Theorem 6 of [9] that the map

$$(A^1, \ldots, A^M) \mapsto \mathrm{Tr}_{\mathcal{K}} \exp[-\lambda(\mathbb{I} - \hat{P}) + \hat{L} + \ln \hat{A}] \tag{19}$$

is concave, for every $\lambda \in \mathbb{R}$. The concavity property survives in the limit $\lambda \to \infty$, in which limit the operator in the exponent of (19) is $-\infty$ on the orthogonal complement of $\hat{P}\mathcal{K}$. Therefore, in the $\lambda \to \infty$ limit the concavity in (19) becomes the statement that

$$(A^1, \ldots, A^M) \mapsto \mathrm{Tr}_{\hat{P}\mathcal{K}} \exp(\hat{P}[\hat{L} + \ln \hat{A}]\hat{P}) \tag{20}$$

is concave. We shall show that

$$\mathrm{Tr}_{\hat{P}\mathcal{K}} \exp(\hat{P}[\hat{L} + \ln \hat{A}]\hat{P}) = \mathrm{Tr}_{\mathcal{H}} \exp\left(L + \sum_\alpha K^{\alpha *}(\ln A^\alpha)K^\alpha\right), \tag{21}$$

which finishes the proof.

Equation (21) can be proved as follows. The trace on the left side is over $\hat{P}\mathcal{K}$, which is isomorphic to $\mathcal{H}$. In fact, the map $\mathcal{U}: \mathcal{H} \to \hat{P}\mathcal{K}$, defined by

$$(\mathcal{U}\Psi)^\alpha = K^\alpha \Psi, \tag{22}$$

is clearly onto since every vector in $\hat{P}\mathcal{K}$ has the form $K^\alpha\Psi$. Moreover, since $\Sigma_\alpha K^{\alpha *}K^\alpha = \mathbb{I}$, $\mathcal{U}$ preserves norms and hence $\mathcal{U}$ is a unitary. A simple calculation shows that

$$\mathcal{U}^* \hat{P}[\hat{L} + \ln \hat{A}]\hat{P}\mathcal{U} = L + \sum_\alpha K^{\alpha *}(\ln A^\alpha)K^\alpha. \tag{23}$$

□

Proof of Theorem 2 using Theorem 3. We need to show that

$$\mathrm{Tr}_{123}\, \varrho_{123}\left(-\ln \varrho_{123} + \ln \varrho_{12} - \sum_\alpha K^{\alpha *}(\ln\varrho_2^\alpha - \ln\varrho_{23}^\alpha)K^\alpha\right) \leq 0. \tag{24}$$

Using the Peierls-Bogoliubov inequality [10], we see that (24) holds if we can show that

$$\mathrm{Tr}_{123} \exp\left(\ln\varrho_{12} - \sum_\alpha K^{\alpha *}(\ln\varrho_2^\alpha - \ln\varrho_{23}^\alpha)K^\alpha\right) \leq 1. \tag{25}$$

We now use an idea of Uhlmann [11]. Let U_3 be a unitary operator on $\mathcal{H}_3$, and let dU_3 denote the corresponding normalized Haar measure. Since the trace is invariant under unitary transformations, and since the K^α commute with U_3, we see that the left side of (25) equals

$$\int \mathrm{Tr}_{123}\, U_3^* \exp\left(\ln\varrho_{12} - \sum_\alpha K^{\alpha *}(\ln\varrho_2^\alpha - \ln\varrho_{23}^\alpha)K^\alpha\right)U_3 dU_3$$

$$= \int \mathrm{Tr}_{123} \exp\left(\ln\varrho_{12} - \sum_\alpha K^{\alpha *}(\ln\varrho_2^\alpha)K^\alpha + \sum_\alpha K^{\alpha *}(\ln U_3^*\varrho_{23}^\alpha U_3)K^\alpha\right)dU_3. \tag{26}$$

Now $\int[U_3^*\varrho_{23}^\alpha U_3]dU_3 = d^{-1}\mathrm{Tr}_3\, \varrho_{23}^\alpha = d^{-1}\varrho_2^\alpha$, where d denotes the dimension of $\mathcal{H}_3$. Using the concavity result of Theorem 3, we see that

$$(26) \leq \mathrm{Tr}_{123}\exp\left(\ln\varrho_{12} - \sum_\alpha K^{\alpha *}K^\alpha(\ln d)\right) = 1. \tag{27}$$

The last equality follows from $\Sigma_\alpha K^{\alpha *}K^\alpha = \mathbb{I}$ and $\mathrm{Tr}_{123}\, \varrho_{12} = d$. □

Proof of Theorem 2 using CPT theory. Consider the map $\Phi: \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3 \mapsto \mathbb{C}^M \otimes \mathcal{H}_2 \otimes \mathcal{H}_3$, given for a general density matrix ϱ_{123} by

$$\Phi(\varrho_{123}) = \bigoplus_\alpha n^\alpha \varrho_{23}^\alpha. \tag{28}$$

This map is trace preserving and completely positive (CPT) [12]. It is known that the relative entropy, $H(\varrho, \gamma) = \mathrm{Tr}\, \varrho(\ln\varrho - \ln \gamma)$ decreases under such maps [6,13,14], and hence

$$H(\varrho_{123}, \varrho_{12} \otimes \varrho_3) \geq H(\Phi(\varrho_{123}), \Phi(\varrho_{12} \otimes \varrho_3)). \tag{29}$$

The left side of this inequality equals $S[\varrho_{12}] + S[\varrho_3] - S[\varrho_{123}]$. To compute the right side, note that

$$\Phi(\varrho_{12} \otimes \varrho_3) = \bigoplus_\alpha n^\alpha \varrho_2^\alpha \otimes \varrho_3. \tag{30}$$

It is then easy to see that the right side of (29) equals $\Sigma_\alpha n^\alpha(S[\varrho_2^\alpha] - S[\varrho_{23}^\alpha] + S[\varrho_3])$. Thus (29) is the same statement as (14). □

III. COROLLARIES AND APPLICATIONS

Taking $\mathcal{H}_2 = \mathbb{C}$, we get as an immediate corollary of Theorem 1 and the concavity of $\varrho \mapsto S[\varrho]$:

Corollary 1 (Improved subadditivity). *Let ϱ_{12} be a density matrix on a Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$. Let P^α be positive,* *bounded and measurable operators on* $\mathcal{H}_1$, *with* $\int_\Omega d\mu(\alpha)P^\alpha=\mathbb{I}_1$. *Let* $n^\alpha=\mathrm{Tr}_{12}P^\alpha\varrho_{12}$ *and, in case* $n^\alpha>0$, *let* $\varrho_2^\alpha=\mathrm{Tr}_1 P^\alpha\varrho_{12}/n^\alpha$. *Then*

$$S[\varrho_{12}]\leq S[\varrho_1]+\int_\Omega d\mu(\alpha)n^\alpha S[\varrho_2^\alpha]\leq S[\varrho_1]+S[\varrho_2]. \tag{31}$$

Remarks.

(1) In the notation of Theorem 1, $P^\alpha=K^{\alpha*}K^\alpha$, but there is no need for this splitting in this case.

(2) One may wonder whether (31) holds if ϱ_1 is also split in a manner similar to ϱ_2. This is not true, in general. As a simple example, consider the case when $\mathcal{H}_1=\mathcal{H}_2$, and $\varrho_{12}=d^{-1}\Sigma_{\alpha=1}^d\Pi^\alpha\otimes\Pi^\alpha$, with Π^α being mutually orthogonal one-dimensional projections. With $P^\alpha=\Pi^\alpha$ we have $S[\varrho_{12}]=\ln d$, whereas $S[\varrho_1^\alpha]=S[\varrho_2^\alpha]=0$ for all α.

A. Classical entropies

Now, suppose we are given a partition of unity of both $\mathcal{H}_1$ and $\mathcal{H}_2$, i.e., a finite set of positive operators P^α and Q^β such that

$$\sum_\alpha P^\alpha=\mathbb{I}_1,\quad \sum_\beta Q^\beta=\mathbb{I}_2. \tag{32}$$

For simplicity, we restrict ourselves to the case of a finite dimensional Hilbert space and discrete sums in this subsection, but, using the methods described in the Appendixes, one can extend the results to the case of infinite dimensional Hilbert spaces and integrals over general measure spaces.

For ϱ_{12} a density matrix on $\mathcal{H}_1\otimes\mathcal{H}_2$, we can define a "classical" entropy as

$$S^{\mathrm{cl}}[\varrho_{12}]=\sum_{\alpha,\beta}-(\mathrm{Tr}_{12}P^\alpha Q^\beta\varrho_{12})\ln(\mathrm{Tr}_{12}P^\alpha Q^\beta\varrho_{12}). \tag{33}$$

Analogously, we can define $S^{\mathrm{cl}}[\varrho_1]$ and $S^{\mathrm{cl}}[\varrho_2]$ for density matrices on $\mathcal{H}_1$ or $\mathcal{H}_2$.

In the case where the P^α and Q^β are one-dimensional projections, this definition agrees with the one in [15]. Note, however, that we define the classical entropy here for any partition of unity. If this partition is trivial, i.e., $P^\alpha=\mathbb{I}$ for $\alpha=1$ and zero otherwise, then $S^{\mathrm{cl}}[\varrho_1]\equiv 0$ identically, and we see from this example that it is not generally true that $S[\varrho_1]\leq S^{\mathrm{cl}}[\varrho_1]$. However, this inequality is true if the partition is such that $\mathrm{Tr}\,P^\alpha\leq 1$ for all α. That this is the *only* condition on P^α needed for $S[\varrho_1]\leq S^{\mathrm{cl}}[\varrho_1]$ to hold follows easily from the concavity of $x\mapsto -x\ln x$:

$$\begin{aligned}S[\varrho_1]&=\sum_\alpha -\mathrm{Tr}\,P^\alpha\varrho_1\ln\varrho_1\\&\leq\sum_\alpha -(\mathrm{Tr}\,P_\alpha\varrho_1)\ln(\mathrm{Tr}\,P_\alpha\varrho_1/\mathrm{Tr}\,P_\alpha)\leq S^{\mathrm{cl}}[\varrho_1].\end{aligned} \tag{34}$$

Corollary 1 above can be used to prove the following inequality for the mutual information $S[\varrho_1]+S[\varrho_2]-S[\varrho_{12}]$.

Corollary 2 (Quantum mutual information bounds classical mutual information). *For any density matrix* ϱ_{12} *on* $\mathcal{H}_1\otimes\mathcal{H}_2$, *and any partition of unity as in (32),*

$$S[\varrho_1]+S[\varrho_2]-S[\varrho_{12}]\geq S^{\mathrm{cl}}[\varrho_1]+S^{\mathrm{cl}}[\varrho_2]-S^{\mathrm{cl}}[\varrho_{12}]. \tag{35}$$

Note that the right side of (35) is just the classical mutual information, since $\mathrm{Tr}_1P^\alpha\varrho_1=\Sigma_\beta\mathrm{Tr}_{12}P^\alpha Q^\beta\varrho_{12}$.

Proof. We learn from Corollary 1 that

$$S[\varrho_{12}]-S[\varrho_1]\leq\sum_\alpha n^\alpha S[\varrho_2^\alpha], \tag{36}$$

where $n^\alpha=\mathrm{Tr}_{12}P^\alpha\varrho_{12}$ and $\varrho_2^\alpha=\mathrm{Tr}_1P^\alpha\varrho_{12}/n^\alpha$. On $\mathbb{C}^M\otimes\mathcal{H}_2$, define a density matrix $\tilde\varrho_{12}$ as

$$\tilde\varrho_{12}=\bigoplus_\alpha n^\alpha\varrho_2^\alpha. \tag{37}$$

Then the right side of (36) can be written as

$$\sum_\alpha n^\alpha S[\varrho_2^\alpha]=S[\tilde\varrho_{12}]-S^{\mathrm{cl}}[\varrho_1]. \tag{38}$$

Note that $\tilde\varrho_2=\mathrm{Tr}_1\tilde\varrho_{12}=\Sigma_\alpha n^\alpha\varrho_2^\alpha=\varrho_2$.

We now apply inequality Corollary 1 again, this time to the expression $S[\tilde\varrho_{12}]-S[\tilde\varrho_2]$. This yields

$$S[\tilde\varrho_{12}]-S[\varrho_2]=S[\tilde\varrho_{12}]-S[\tilde\varrho_2]\leq\sum_\beta m^\beta S[\tilde\varrho_1^\beta], \tag{39}$$

with $m^\beta=\mathrm{Tr}_{12}Q^\beta\tilde\varrho_{12}$ and $\tilde\varrho_1^\beta=\mathrm{Tr}_2Q^\beta\tilde\varrho_{12}/m^\beta$. The right side of this expression equals

$$\sum_\beta m^\beta S[\tilde\varrho_1^\beta]=S^{\mathrm{cl}}[\varrho_{12}]-S^{\mathrm{cl}}[\varrho_2]. \tag{40}$$

In combination (36)–(40) give the desired result. □

Another way to interpret the results above is the following: define a partially classical and partially quantum entropy by

$$S^{\mathrm{cl,Q}}[\varrho_{12}]=-\sum_\alpha\mathrm{Tr}_2(\mathrm{Tr}_1P^\alpha\varrho_{12})\ln(\mathrm{Tr}_1P^\alpha\varrho_{12}). \tag{41}$$

Then Corollary 1 and the proof of Corollary 2 show that

$$\begin{aligned}S[\varrho_{12}]-S[\varrho_1]-S[\varrho_2]&\leq S^{\mathrm{cl,Q}}[\varrho_{12}]-S^{\mathrm{cl}}[\varrho_1]-S[\varrho_2]\\&\leq S^{\mathrm{cl}}[\varrho_{12}]-S^{\mathrm{cl}}[\varrho_1]-S^{\mathrm{cl}}[\varrho_2].\end{aligned} \tag{42}$$

In the same way, Theorem 2, in the special case where the K^α act nontrivially only on $\mathcal{H}_1$, can be interpreted as

$$S[\varrho_{123}]-S[\varrho_{12}]\leq S^{\mathrm{cl,Q,Q}}[\varrho_{123}]-S^{\mathrm{cl,Q}}[\varrho_{12}] \tag{43}$$

for a partition of unity on $\mathcal{H}_1$, with the obvious definition of $S^{\mathrm{cl,Q,Q}}$. We can use this inequality to prove the following.

Corollary 3 (Convexity of classical minus quantum entropy). *The map*

$$\varrho_{12}\mapsto S^{\mathrm{cl,Q}}[\varrho_{12}]-S[\varrho_{12}] \tag{44}$$

is convex.

Proof. Let A_{12} and B_{12} be two density matrices on $\mathcal{H}_1$

$\otimes \mathcal{H}_2$. On $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathbb{C}^2$, consider the density matrix $\varrho_{123} = \frac{1}{2} A_{12} \otimes \Pi + \frac{1}{2} B_{12} \otimes (\mathbb{I} - \Pi)$, where Π is a one-dimensional projection in $\mathbb{C}^2$. Inequality (43) implies that

$$\frac{1}{2}S[A_{12}] + \frac{1}{2}S[B_{12}] - S[\varrho_{12}] \leqslant \frac{1}{2}S^{\text{cl,Q}}[A_{12}] + \frac{1}{2}S^{\text{cl,Q}}[B_{12}] - S^{\text{cl,Q}}[\varrho_{12}]. \tag{45}$$

Since $\varrho_{12} = \frac{1}{2} A_{12} + \frac{1}{2} B_{12}$, this proves the result. □

Remark. In particular, taking $\mathcal{H}_2 = \mathbb{C}$ to be trivial, this corollary shows that the map (for a single Hilbert space)

$$\varrho \mapsto S^{\text{cl}}[\varrho] - S[\varrho] \tag{46}$$

is convex—which is remarkable, given that both entropies are concave functions of ϱ. The inequality implied by convexity of (46) is known as the Holevo bound [16,17].

B. Coherent states and Wehrl entropy

Now, suppose we are given a coherent state decomposition of both $\mathcal{H}_1$ and $\mathcal{H}_2$, i.e., normalized vectors $|\varphi\rangle \in \mathcal{H}_1$, $|\theta\rangle \in \mathcal{H}_2$ and positive measures μ and ν on some measure space (not necessarily the same spaces) such that

$$\int d\mu(\varphi)|\varphi\rangle\langle\varphi| = \mathbb{I}_1, \qquad \int d\nu(\theta)|\theta\rangle\langle\theta| = \mathbb{I}_2. \tag{47}$$

Here, $|\varphi\rangle\langle\varphi|$ is the Dirac notation for the one-dimensional projector onto φ and the integrals are to be interpreted in the weak sense, as explained before.

The classical (Wehrl) entropy [7] for a density matrix ϱ_{12} on $\mathcal{H}_1 \otimes \mathcal{H}_2$ is then defined as

$$S^{\text{W}}[\varrho_{12}] = -\int d\mu(\varphi)d\nu(\theta)\langle\varphi,\theta|\varrho_{12}|\varphi,\theta\rangle \ln\langle\varphi,\theta|\varrho_{12}|\varphi,\theta\rangle, \tag{48}$$

and similarly for density matrices on $\mathcal{H}_1$ or $\mathcal{H}_2$. As Wehrl showed, it follows from concavity of $x \mapsto -x \ln x$ that

$$S[\varrho_{12}] \leqslant S^{\text{W}}[\varrho_{12}]. \tag{49}$$

Corollaries 2 and 3 now also hold for the Wehrl entropy. In particular, it is true that

$$S^{\text{W}}[\varrho_1] + S^{\text{W}}[\varrho_2] - S^{\text{W}}[\varrho_{12}] \leqslant S[\varrho_1] + S[\varrho_2] - S[\varrho_{12}]. \tag{50}$$

Moreover, Corollary 3 implies that the map

$$\varrho \mapsto S^{\text{W}}[\varrho] - S[\varrho] \tag{51}$$

is convex. Note that the infimum of this function is zero. In the finite dimensional case, this infimum is achieved for the totally mixed state $\varrho = d^{-1}\mathbb{I}$, where $d = \dim \mathcal{H}$.

It might be recalled that Wehrl raised the question [7] of evaluating the minimum of his classical entropy and conjectured, in the special case of the Glauber coherent states, that it should be given by the one-dimensional projector onto a coherent state $\varrho = |\theta\rangle\langle\theta|$. (The minimum of the quantum entropy, $-\text{Tr}\, \varrho \ln \varrho$, is always zero.) This particular conjecture was proved in [18], where the (still open) generalized conjecture was made to $SU(2)$ (Bloch) coherent states. Oddly, the minimum of the *difference* of the entropies is a much easier question to answer.

Note that because of convexity the maximum of the function $S^{\text{W}}[\varrho] - S[\varrho]$ is attained for a pure state, where $S[\varrho] = 0$. Hence the question about the maximum of $S^{\text{W}}[\varrho] - S[\varrho]$ is equivalent to *maximizing* $S^{\text{W}}[\varrho]$ over pure states. In the case of the Glauber coherent states, this maximum is infinite [7].

C. Quantum statistical mechanics of point particles

Consider a system of two types of particles, A and B. The state space of the combined system is $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$, where $\mathcal{H}_A$ and $\mathcal{H}_B$ are the spaces of square integrable functions of the particle configurations of particles A and B, respectively. We assume that the configuration space is $\mathbb{R}^{d_A}$ and $\mathbb{R}^{d_B}$, respectively, for some finite d_A and d_B. The usual subadditivity of entropy implies that, for any state ϱ on $\mathcal{H}$,

$$S[\varrho_B] \geqslant S[\varrho] - S[\varrho_A], \tag{52}$$

where ϱ_A and ϱ_B denote the states of the subsystems. However, in applications it can be useful to get a lower bound not only on the entropy of ϱ_B, which is the state averaged over all configurations of the A particles, but rather on the *average entropy* of the state for *fixed* A particles. Such a bound is one of the key ingredients in a rigorous upper bound on the pressure of a dilute Fermi gas at nonzero temperature [19].

More precisely, if X_A and X_B denote particle configurations of the A and B particles, any density matrix on $\mathcal{H}$ will be given by an integral kernel $\varrho(X_A, X_B; X_A', X_B')$. For every fixed configuration of the A particles, X_A, we can then define a density matrix on $\mathcal{H}_B$ by the kernel

$$\varrho_B^{X_A}(X_B, X_B') = n(X_A)^{-1}\varrho(X_A, X_B; X_A, X_B'), \tag{53}$$

where $n(X_A)$ is the normalization factor

$$n(X_A) = \int dX_B \varrho(X_A, X_B; X_A, X_B). \tag{54}$$

Since ϱ is a trace class operator, $\varrho_B^{X_A}$ is well defined by the spectral decomposition of ϱ for almost every X_A, if $n(X_A) \neq 0$. The definition (53) makes sense only if $n(X_A) > 0$; only in this case $\varrho_B^{X_A}$ will be needed below, however.

Note that $n(X_A)$ is the probability density of a configuration of A particles X_A. Moreover, $\int dX_A n(X_A) = 1$, and $\int dX_A n(X_A)\varrho_B^{X_A} = \varrho_B$. Therefore we have, by concavity of $S[\varrho]$,

$$S[\varrho_B] \geqslant \int dX_A n(X_A) S[\varrho_B^{X_A}]. \tag{55}$$

Hence the following is a strengthening of (52).

Corollary 4 (Subadditivity with average entropy instead of entropy of average). *Let ϱ be a density matrix on $\mathcal{H}_A \otimes \mathcal{H}_B$ with finite entropy. With the definitions given above,*

$$\int dX_A n(X_A) S[\varrho_B^{X_A}] \geqslant S[\varrho] - S[\varrho_A]. \tag{56}$$

We remark that it is not necessary to have a fixed particle number for this bound; the integral $\int dX_A$ can as well include a discrete sum over different particle numbers. I.e., our bound also applies to the grand-canonical ensemble, and this is the form that Corollary 4 actually gets used in [19]. For simplicity, we consider only the case of a fixed particle number in the proof below, but the extension is straightforward, using an additional decomposition $\mathbb{I}=\Sigma_{n\geq 0}P_n$, where P_n projects onto the subspace of $\mathcal{H}_A$ with fixed particle number n.

Corollary 4 follows from Corollary 1 by the following limiting argument.

Proof of Corollary 4. For $d\equiv d_A$, let $j:\mathbb{R}^d\mapsto\mathbb{R}$ be a positive and integrable function on the configuration space of the A particles, with $\int dX\, j(X)=1$. For some $\varepsilon>0$ and $Y\in\mathbb{R}^d$, let $j_\varepsilon^Y(X)=\varepsilon^{-d}j((X-Y)/\varepsilon)$. Let

$$P^\varepsilon=\mathbb{I}-\int_{\mathbb{R}^d}dY|j_\varepsilon^Y\rangle\langle j_\varepsilon^Y|. \tag{57}$$

It is not difficult to see that $P^\varepsilon\geq 0$. Hence we can infer from Corollary 1 that, for any density matrix ϱ_{12} on $\mathcal{H}_1\otimes\mathcal{H}_2$ (where we use again the notation 1 and 2 instead of A and B),

$$S[\varrho_{12}]-S[\varrho_1]\leq\int_{\mathbb{R}^d}dY\,n_\varepsilon(Y)S[\varrho_\varepsilon^Y]+n_{P^\varepsilon}S[\varrho_2^{P^\varepsilon}], \tag{58}$$

where we denoted $n_\varepsilon(Y)=\langle j_\varepsilon^Y|\varrho_1|j_\varepsilon^Y\rangle$, $\varrho_\varepsilon^Y=\langle j_\varepsilon^Y|\varrho_{12}|j_\varepsilon^Y\rangle/n_\varepsilon(Y)$, $n_{P^\varepsilon}=\mathrm{Tr}_1P^\varepsilon\varrho_1$, and $\varrho_2^{P^\varepsilon}=\mathrm{Tr}_1P^\varepsilon\varrho_{12}/n_{P^\varepsilon}$. We will show that there exists a sequence ε_j with $\varepsilon_j\to 0$ as $j\to\infty$ such that the right side of (58) converges to the left side of (56) in the limit $j\to\infty$.

We note that, for any square integrable function ϕ on $\mathbb{R}^d$, $\langle j_\varepsilon^Y|\phi\rangle\to\phi(Y)$ strongly in $L^2(\mathbb{R}^d)$ as $\varepsilon\to 0$ ([21], Theorem 2.16). Passing to a subsequence, it is then true that $\langle j_\varepsilon^Y|\phi\rangle\to\phi(Y)$ almost everywhere. Decomposing ϱ_1 into its eigenvalues and eigenfunctions, we see that there is a subsequence such that $\lim_{\varepsilon\to 0}n_\varepsilon(Y)=n(Y)$ for almost every Y. Also, $\varrho_\varepsilon^Y\rightharpoonup\varrho_B^Y$ weakly as $\varepsilon\to 0$ for almost every Y. (Here we used the separability of the Hilbert space to ensure the existence of this subsequence.) Since also the traces converge, this convergence is actually in trace class norm [20].

We first assume that ϱ_{12} has finite rank. Then also ϱ_ε^Y and ϱ_B^Y have finite rank, and hence bounded entropy. It follows that $\lim_{\varepsilon\to 0}S[\varrho_\varepsilon^Y]=S[\varrho_B^Y]$. Moreover, it is easy to see that $P^\varepsilon\rightharpoonup 0$ weakly as $\varepsilon\to 0$. This implies that $\lim_{\varepsilon\to 0}n_{P^\varepsilon}=0$, and hence $\lim_{\varepsilon\to 0}\int dY\,n_\varepsilon(Y)=1$. It then follows from Fatou's lemma that

$$\lim_{\varepsilon\to 0}\int_{\mathbb{R}^d}dY\,n_\varepsilon(Y)S[\varrho_\varepsilon^Y]=\int_{\mathbb{R}^d}dY\,n(Y)S[\varrho_B^Y]. \tag{59}$$

It remains to show that the last term in (58) goes to zero as $\varepsilon\to 0$. As already noted, $\lim_{\varepsilon\to 0}n_{P^\varepsilon}=0$. The entropy $S[\varrho_2^{P^\varepsilon}]$ need not be bounded as $\varepsilon\to 0$, however. Since $P^\varepsilon\leq\mathbb{I}$, $\mathrm{Tr}_1P^\varepsilon\varrho_{12}\leq\varrho_B$. Note that ϱ_{12} has finite entropy by assumption and, without loss of generality, also ϱ_A has finite entropy. This implies that ϱ_B has finite entropy by the triangle inequality for entropies [22]. Hence it follows from dominated convergence ([1], Theorem A3) that $S[\mathrm{Tr}_1P^\varepsilon\varrho_{12}]\to 0$ as $\varepsilon\to 0$, and hence $n_{P^\varepsilon}S[\varrho_2^{P^\varepsilon}]=S[\mathrm{Tr}_1P^\varepsilon\varrho_{12}]+n_{P^\varepsilon}\ln n_{P^\varepsilon}\to 0$ as $\varepsilon\to 0$. This proves (56) in the case that ϱ_{12} has finite rank.

For a general ϱ_{12} on $\mathcal{H}_1\otimes\mathcal{H}_2$, let $\varrho_{12}^j=P^j\varrho_{12}$, where P^j denotes the projection onto the largest j eigenvalues of ϱ_{12}. It is then easy to see that

$$S[\varrho_{12}]=\lim_{j\to\infty}S[\varrho_{12}^j] \tag{60}$$

and

$$S[\varrho_A]=\lim_{j\to\infty}S[\mathrm{Tr}_2\varrho_{12}^j] \tag{61}$$

(cf. the Appendix in [1]). Moreover, with $n^j(Y)$ and $\varrho_{B,j}^Y$ defined as above, for the operator ϱ_{12}^j, we write

$$\int dY\,n^j(Y)S[\varrho_{B,j}^Y]=e\int dY\,n(Y)S[\varrho_{B,j}^Y(n^j(Y)/e\,n(Y))]-\int dY\,n^j(Y)\ln[e\,n(Y)/n^j(Y)]. \tag{62}$$

Note that $n^j(Y)$ is pointwise increasing in j, and also $n^j(Y)\varrho_{B,j}^Y$ is an increasing sequence of operators. Moreover, $\varrho_{B,j}^Y[n^j(Y)/e\,n(Y)]\leq 1/e$. Since $-x\ln x$ is monotone increasing for $0\leq x\leq 1/e$, this implies that the first term on the right side of (62) is bounded from above by

$$e\int dY\,n(Y)S[\varrho_B^Y/e]=\int dY\,n(Y)S[\varrho_B^Y]+\int dY\,n(Y). \tag{63}$$

Moreover, since $\lim_{j\to\infty}n^j(Y)=n(Y)$ for almost every Y, this implies, by monotone convergence,

$$\lim_{j\to\infty}-\int dY\,n^j(Y)\ln[e\,n(Y)/n^j(Y)]=-\int dY\,n(Y). \tag{64}$$

This shows that (56) also holds in the infinite rank case, and finishes the proof of Corollary 4. □

ACKNOWLEDGMENTS

This work was partially supported by U.S. National Science Foundation Grant Nos. PHY 01 39984 (E.H.L.) and PHY 03 53181 (R.S.), as well as by the A.P. Sloan Foundation (R.S.).

APPENDIX A: EXTENSION TO INTEGRALS

In this appendix we extend Theorem 2 in the following way. Again, let $\mathcal{H}_i$ be *finite dimensional* Hilbert spaces, $1\leq i\leq 3$. Let Ω be a measure space, with elements labeled by α, and let $d\mu(\alpha)$ be a measure on Ω. Let K^α be matrices on $\mathcal{H}_1\otimes\mathcal{H}_2$ that are weakly measurable (i.e., all the matrix elements are measurable functions), such that

$$\int_\Omega d\mu(\alpha)K^{\alpha *}K^\alpha=\mathbb{I}_{12}. \tag{A1}$$

The extension of Theorem 2 is the following. With the same definitions as in (11)–(13),

$$S[\varrho_{123}]-S[\varrho_{12}] \leq \int_{\Omega} d\mu(\alpha) n^{\alpha}(S[\varrho_{23}^{\alpha}]-S[\varrho_{2}^{\alpha}]). \quad \text{(A2)}$$

Note that this expression is well defined, since the integrand is measurable and the entropy is bounded.

For the proof of (A2), we may assume that, for each α, $\|K^{\alpha*}K^{\alpha}\| \leq 1$. This is no restriction, since we can always absorb the norm into the measure $d\mu(\alpha)$. Likewise, we may assume that $\operatorname{Tr} K^{\alpha*}K^{\alpha} \geq 1/2$. Taking the trace of (A1), it is then clear that Ω has finite measure.

Pick some $\varepsilon > 0$. By looking at the level sets of the matrix elements of K^{α}, we can divide Ω into *finitely many* disjoint measurable subsets $\mathcal{O}_j$, $1 \leq j \leq M_{\varepsilon}$, with $\|K^{\alpha}-K^{\beta}\| \leq \varepsilon$ if α and β are in the same subset $\mathcal{O}_j$. For each $\alpha \in \Omega$, write $K^{\alpha} = U^{\alpha}(K^{\alpha*}K^{\alpha})^{1/2}$, with U^{α} unitary. For each j, pick some $\alpha_j \in \mathcal{O}_j$, and define

$$L^j = U^{\alpha_j}\left(\int_{\mathcal{O}_j} d\mu(\alpha) K^{\alpha*}K^{\alpha}\right)^{1/2}. \quad \text{(A3)}$$

We then have $\Sigma_j L^{j*}L^j = \mathbb{I}_{12}$, and hence we can apply Theorem 2. That is, we have, for any density matrix ϱ_{123} on $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3$,

$$S[\varrho_{123}]-S[\varrho_{12}] \leq \sum_j n^j(S[\varrho_{23}^j]-S[\varrho_2^j]), \quad \text{(A4)}$$

with $n^j = \operatorname{Tr}_{123} L^{j*}L^j\varrho_{123}$ and $\varrho_{23}^j = \operatorname{Tr}_1 L^j\varrho_{123}L^{j*}/n^j$. We will now show that, as $\varepsilon \to 0$, the right side of (A4) converges to the right side of (A2).

Using the fact that $\|\sqrt{A}-\sqrt{B}\| \leq \|A-B\|^{1/2}$ for any two positive matrices A and B [Eq. (X.2) of [23]], we can estimate

$$\|L^j - |\mathcal{O}_j|^{1/2}K^{\alpha_j}\|^2 \leq \int_{\mathcal{O}_j} d\mu(\alpha)\|K^{\alpha*}K^{\alpha}-K^{\alpha_j*}K^{\alpha_j}\| \leq 2\varepsilon|\mathcal{O}_j|. \quad \text{(A5)}$$

Here, $|\cdot|$ denotes the measure of a subset of Ω, and we used $\|K^{\alpha}\| \leq 1$ and $\|K^{\alpha}-K^{\alpha_j}\| \leq \varepsilon$ in the last step. Using the triangle inequality we thus see that $\|L^j - |\mathcal{O}_j|^{1/2}K^{\alpha}\| \leq |\mathcal{O}_j|^{1/2}(\varepsilon+\sqrt{2\varepsilon})$ for any $\alpha \in \mathcal{O}_j$. This implies that

$$\|L^j\varrho_{123}L^{j*} - |\mathcal{O}_j|K^{\alpha}\varrho_{123}K^{\alpha*}\| \leq |\mathcal{O}_j|2(\varepsilon+\sqrt{2\varepsilon})\|\varrho_{123}\| \quad \text{(A6)}$$

for any $\alpha \in \mathcal{O}_j$, where we again used that $\|K^{\alpha}\| \leq 1$. Note that (A6) implies the same estimate for the partial trace of the operator on the left side, with the right side multiplied by the dimension of the space. Moreover, since in finite dimensions the entropy is Lipschitz continuous, this also implies that, for some constant $C > 0$,

$$\begin{aligned}\sum_j n^j(S[\varrho_{23}^j]-S[\varrho_2^j]) &= \sum_j (S[\operatorname{Tr}_1 L^j\varrho_{123}L^{j*}] \\ &\quad - S[\operatorname{Tr}_{12} L^j\varrho_{123}L^{j*}]) \\ &\leq \sum_j \int_{\mathcal{O}_j} d\mu(\alpha)\{S[\operatorname{Tr}_1 K^{\alpha}\varrho_{123}K^{\alpha*}] \\ &\quad - S[\operatorname{Tr}_{12} K^{\alpha}\varrho_{123}K^{\alpha*}] + C(\varepsilon+\sqrt{2\varepsilon})\} \\ &= \int_{\Omega} d\mu(\alpha)n^{\alpha}(S[\varrho_{23}^{\alpha}]-S[\varrho_2^{\alpha}]) \\ &\quad + C|\Omega|(\varepsilon+\sqrt{2\varepsilon}). \end{aligned} \quad \text{(A7)}$$

Letting $\varepsilon \to 0$ this proves the desired result.

APPENDIX B: EXTENSION TO INFINITE DIMENSIONAL SPACES

We now show that (A2) holds in the case when the $\mathcal{H}_i$ are infinite dimensional Hilbert spaces. This will prove Theorem 1. As stated there, we assume that the K^{α} are bounded, weakly measurable operators, i.e, $\langle\phi|K^{\alpha}|\phi\rangle$ is measurable for all vectors $|\phi\rangle$. (By polarization, this implies that the matrix elements $\langle\phi|K^{\alpha}|\psi\rangle$ are measurable for any vectors $|\phi\rangle$ and $|\psi\rangle$.) We also assume (4) to hold. Again it is then no restriction to assume that $\|K^{\alpha*}K^{\alpha}\| \leq 1$ for each $\alpha \in \Omega$. Note that Ω need not have finite measure in this case, however. We assume that the density matrix ϱ_{123} has finite entropy. We may also assume that $S[\varrho_{12}]$ is finite, otherwise there is nothing to prove. Note that in this case, the triangle inequality for entropy [22] implies that also $S[\varrho_3]$ is finite.

For $m \in \mathbb{N}$, let $P_i^{(m)}$ be rank m projections in $\mathcal{H}_i$, such that $P_i^{(m)} \to \mathbb{I}_{\mathcal{H}_i}$ strongly as $m \to \infty$. Let $\hat{P} = P_1^{(m)} \otimes P_2^{(m)} \otimes P_3^{(m)}$. Then also $\hat{P} \to \mathbb{I}_{123}$ strongly as $m \to \infty$. In the following, we find it convenient to suppress the dependence on m in our notation, but rather put a ^ on all quantities that depend on m.

Given a density matrix ϱ_{123}, we define $\hat{\varrho}_{123} = \hat{P}\varrho_{123}\hat{P}$. Moreover, let $\widehat{K^{\alpha}} = P_1^{(m)} \otimes P_2^{(m)} K^{\alpha} P_1^{(m)} \otimes P_2^{(m)}$ for $\alpha \in \Omega$. Let also

$$\hat{L} = \left[P_1^{(m)} \otimes P_2^{(m)}\left(1 - \int_{\Omega} d\mu(\alpha)K^{\alpha*}P_1^{(m)} \otimes P_2^{(m)}K^{\alpha}\right) \times P_1^{(m)} \otimes P_2^{(m)}\right]^{1/2}. \quad \text{(B1)}$$

If we set $\hat{\mathcal{H}}_i = P_i^{(m)}\mathcal{H}_i$, we then have

$$\int_{\Omega} d\mu(\alpha)\widehat{K^{\alpha*}}\widehat{K^{\alpha}} + \hat{L}^2 = \mathbb{I}_{\hat{\mathcal{H}}_{12}}, \quad \text{(B2)}$$

and hence we can apply the finite dimensional result (A2). I.e.,

$$\begin{aligned}S[\hat{\varrho}_{123}]-S[\hat{\varrho}_{12}] &\leq \int_{\Omega} d\mu(\alpha)\hat{n}^{\alpha}(S[\hat{\varrho}_{23}^{\alpha}]-S[\hat{\varrho}_2^{\alpha}]) \\ &\quad + \hat{n}^L(S[\hat{\varrho}_{23}^L]-S[\hat{\varrho}_2^L]). \end{aligned} \quad \text{(B3)}$$

Here, $\hat{n}^{\alpha}$ and the density matrices $\hat{\varrho}_{23}^{\alpha}$ and $\hat{\varrho}_2^{\alpha}$ are defined as in (11)–(13), with K^{α} replaced by $\widehat{K^{\alpha}}$ and ϱ_{123} replaced by $\hat{\varrho}_{123}$. Moreover, $\hat{\varrho}_{23}^L$, $\hat{\varrho}_2^L$ and $\hat{n}^L$ are defined in the same way, with $\hat{L}$ in place of $\widehat{K^{\alpha}}$. Our goal is to show that we can remove the ^'s in (B3).

Note that $\hat{\varrho}_{123} \to \varrho_{123}$ strongly as $m \to \infty$. Since also the trace of $\hat{\varrho}_{123}$ converges to $\operatorname{Tr} \varrho_{123} = 1$, this implies that the convergence is actually in trace norm, as proved by Wehrl in [20]. Hence also $\hat{\varrho}_{12}\to\varrho_{12}$ in trace norm. Since the eigenvalues of $\hat{\varrho}_{123}$ are smaller than the corresponding eigenvalues of ϱ_{123}, i.e., $\hat{\varrho}_{123}\prec\varrho_{123}$ in Simon's notation in the Appendix of [1], Theorem A2 in [1] implies that

$$\lim_{m\to\infty} S[\hat{\varrho}_{123}] = S[\varrho_{123}]. \tag{B4}$$

By the same reasoning, this also holds for $\hat{\varrho}_{12}$. Taking the limit $m\to\infty$ in (B3), we thus have

$$S[\varrho_{123}] - S[\varrho_{12}] \leq \liminf_{m\to\infty}\int_\Omega d\mu(\alpha)[A_m(\alpha)+B_m(\alpha)] + \liminf_{m\to\infty} \hat{n}^L(S[\hat{\varrho}_{23}^L]-S[\hat{\varrho}_2^L]). \tag{B5}$$

Here we defined the functions $A_m(\alpha)$ and $B_m(\alpha)$ by

$$A_m(\alpha) = \hat{n}^\alpha(S[\hat{\varrho}_{23}^\alpha]-S[\hat{\varrho}_2^\alpha]-S[\hat{\varrho}_3^\alpha]) \tag{B6}$$

and

$$B_m(\alpha) = \hat{n}^\alpha S[\hat{\varrho}_3^\alpha], \tag{B7}$$

with $\hat{\varrho}_3^\alpha = \mathrm{Tr}_2\,\hat{\varrho}_{23}^\alpha = \mathrm{Tr}_{12}\,\widehat{K^{\alpha *}}\widehat{K^\alpha}\hat{\varrho}_{123}/\hat{n}^\alpha$. The reason for the splitting into the two parts (B6) and (B7) is that $A_m(\alpha)$ is negative, which allows for the use of Fatou's lemma, whereas $B_m(\alpha)$ depends on $\widehat{K^\alpha}$ only through $\widehat{K^{\alpha *}}\widehat{K^\alpha}$.

We start by estimating the last term on the right side of (B5). By subadditivity of entropy, $S[\hat{\varrho}_{23}^L]-S[\hat{\varrho}_2^L]\leq S[\hat{\varrho}_3^L]$, with $\hat{\varrho}_3^L=\mathrm{Tr}_2\,\hat{\varrho}_{23}^L$. We claim that $\lim_{m\to\infty}\hat{n}^L=0$. This is true if we can show that

$$\lim_{m\to\infty}\int_\Omega d\mu(\alpha)\mathrm{Tr}_{123}\,\hat{P}K^{\alpha *}\hat{P}K^\alpha\hat{P}\varrho_{123} = 1. \tag{B8}$$

Since $\hat{P}K^{\alpha *}\hat{P}K^\alpha\hat{P}$ converges strongly to $K^{\alpha *}K^\alpha$ for each fixed α (because products of strongly convergent sequences converge strongly), we see that $\mathrm{Tr}_{123}\,\hat{P}K^{\alpha *}\hat{P}K^\alpha\hat{P}\varrho_{123}$ converges to $\mathrm{Tr}_{123}\,K^{\alpha *}K^\alpha\varrho_{123}$ for each fixed α. Hence, using Fatou's lemma, we see that the left side of (B8) is always ≥ 1. On the other hand, estimating the middle $\hat{P}$ in the integrand in (B8) by $\hat{P}\leq 1$ and using (4), we see that the left side of (B8) is bounded above by

$$\lim_{m\to\infty}\mathrm{Tr}_{123}\hat{P}\varrho_{123} = \mathrm{Tr}_{123}\varrho_{123} = 1. \tag{B9}$$

This proves the claim that $\lim_{m\to\infty}\hat{n}^L=0$.

Although $S[\hat{\varrho}_3^L]$ need not be bounded, we claim that $\hat{n}^L S[\hat{\varrho}_3^L]\to 0$ as $m\to\infty$. To see this, write $\hat{n}^L S[\hat{\varrho}_3^L] = S[\mathrm{Tr}_{12}\hat{L}^2\hat{\varrho}_{123}]+\hat{n}^L\ln\hat{n}^L$. Note that the second term goes to zero as $\hat{n}^L\to 0$. Since $\hat{L}^2\leq\mathbb{I}_{12}$, $\mathrm{Tr}_{12}\hat{L}^2\hat{\varrho}_{123}\leq P_3^{(m)}\varrho_3 P_3^{(m)}\prec\varrho_3$. Recall that $S[\varrho_3]$ is finite. Since $\mathrm{Tr}_{12}\hat{L}^2\hat{\varrho}_{123}\to 0$ in trace norm as $m\to\infty$, we can use dominated convergence Theorem A1 of [1] to conclude that $S[\mathrm{Tr}_{12}L^2\hat{\varrho}_{123}]\to 0$ as $m\to\infty$. Hence we have shown that

$$\limsup_{m\to\infty}\hat{n}^L(S[\hat{\varrho}_{23}^L]-S[\hat{\varrho}_2^L])\leq 0. \tag{B10}$$

Next we treat the first term on the right side of (B5), i.e., the integral of $A_m(\alpha)$. Since $A_m(\alpha)\leq 0$ (by subadditivity), we can use Fatou's lemma to estimate

$$\liminf_{m\to\infty}\int_\Omega d\mu(\alpha)A_m(\alpha)\leq\int_\Omega d\mu(\alpha)\limsup_{m\to\infty}A_m(\alpha). \tag{B11}$$

We now claim that $\hat{\varrho}_{23}^\alpha\to\varrho_{23}^\alpha$ in trace norm. It is clear that $\widehat{K^{\alpha *}}\hat{\varrho}_{123}\widehat{K^\alpha}$ converges strongly to $K^{\alpha *}\varrho_{123}K^\alpha$. By the same argument as that after (B8), the trace also converges, and hence the convergence is in trace norm. This implies that the reduced density matrices also converge in trace norm, and hence proves our claim. Moreover, we claim that, for each fixed α,

$$\limsup_{m\to\infty}A_m(\alpha)\leq n^\alpha(S[\varrho_{23}^\alpha]-S[\varrho_2^\alpha]-S[\varrho_3^\alpha]). \tag{B12}$$

This follows from upper semicontinuity of $S[\varrho_{23}]-S[\varrho_2]-S[\varrho_3]$ in ϱ_{23}. This upper semicontinuity, in turn, follows from lower semicontinuity of the relative entropy $H(\varrho,\sigma)=\mathrm{Tr}\,\varrho(\ln\varrho-\ln\sigma)$ [[10], 2.2,22(ii)], since $S[\varrho_{23}]-S[\varrho_2]-S[\varrho_3]=-H(\varrho_{23},\varrho_2\otimes\varrho_3)$. By combining (B11) and (B12) we have thus shown that

$$\liminf_{m\to\infty}\int_\Omega d\mu(\alpha)A_m(\alpha)\leq\int_\Omega d\mu(\alpha)n^\alpha(S[\varrho_{23}^\alpha]-S[\varrho_2^\alpha]-S[\varrho_3^\alpha]). \tag{B13}$$

It remains to show that

$$\liminf_{m\to\infty}\int_\Omega d\mu(\alpha)B_m(\alpha)\leq\int_\Omega d\mu(\alpha)n^\alpha S[\varrho_3^\alpha], \tag{B14}$$

with

$$B_m(\alpha)=\hat{n}^\alpha S[\hat{\varrho}_3^\alpha]=S[\mathrm{Tr}_{12}\widehat{K^{\alpha *}}\widehat{K^\alpha}\hat{\varrho}_{123}]+\hat{n}^\alpha\ln\hat{n}^\alpha. \tag{B15}$$

Note that $\mathrm{Tr}_{12}\widehat{K^{\alpha *}}\widehat{K^\alpha}\hat{\varrho}_{123}\leq P_3^{(m)}\varrho_3P_3^{(m)}\prec\varrho_3$, since $\|K^{\alpha *}K^\alpha\|\leq 1$ and hence $\|\widehat{K^{\alpha *}}\widehat{K^\alpha}\|\leq 1$. By the same argument as above, $\mathrm{Tr}_{12}\widehat{K^{\alpha *}}\widehat{K^\alpha}\hat{\varrho}_{123}$ converges to $\mathrm{Tr}_{12}K^{\alpha *}K^\alpha\varrho_{123}$ in trace norm, and hence, again by Theorem A1 in [1],

$$\lim_{m\to\infty}B_m(\alpha)=n^\alpha S[\varrho_3^\alpha] \tag{B16}$$

for each fixed α. This gives pointwise convergence, but to show convergence of the integral in (B14) we have to use the dominated convergence theorem. We claim that $B_m(\alpha)$ is uniformly bounded, independent of m and α. This is true since $-x\ln x$ is monotone in x for $0\leq x\leq 1/e$. Since $\mathrm{Tr}_{12}\widehat{K^{\alpha *}}\widehat{K^\alpha}\hat{\varrho}_{123}\prec\varrho_3$, the contribution to the entropy from eigenvalues less then $1/e$ is bounded by the corresponding value for ϱ_3. Moreover, since $\mathrm{Tr}\,\varrho_3=1$, there are at most 2 eigenvalues bigger than $1/e$. This gives the claimed uniform bound.

By dominated convergence, we see that, for any subset $\mathcal{O}\subset\Omega$ with finite measure,

$$\lim_{m\to\infty}\int_{\mathcal{O}} d\mu(\alpha)B_m(\alpha) = \int_{\mathcal{O}} d\mu(\alpha)n^\alpha S[\varrho_3^\alpha] \leqslant \int_\Omega d\mu(\alpha)n^\alpha S[\varrho_3^\alpha]. \quad \text{(B17)}$$

Moreover, by concavity of $S[\varrho]$,

$$\int_{\mathcal{O}^c} d\mu(\alpha)B_m(\alpha) \leqslant \left(\int_{\mathcal{O}^c} d\mu(\alpha)\hat{n}^\alpha\right)S[\hat{\varrho}^c], \quad \text{(B18)}$$

with

$$\hat{\varrho}^c = \left[\int_{\mathcal{O}^c} d\mu(\alpha)\hat{n}^\alpha\right]^{-1}\mathrm{Tr}_{12}\int_{\mathcal{O}^c} d\mu(\alpha)\widehat{K^{\alpha*}}\widehat{K^\alpha}\hat{\varrho}_{123}. \quad \text{(B19)}$$

Proceeding as in the proof of (B8), we see that

$$\lim_{m\to\infty}\int_{\mathcal{O}^c} d\mu(\alpha)\hat{n}^\alpha = \int_{\mathcal{O}^c} d\mu(\alpha)n^\alpha. \quad \text{(B20)}$$

Also $\mathrm{Tr}_{12}\int_{\mathcal{O}^c}d\mu(\alpha)\widehat{K^{\alpha*}}\widehat{K^\alpha}\hat{\varrho}_{123}\to\mathrm{Tr}_{12}\int_{\mathcal{O}^c}d\mu(\alpha)K^{\alpha*}K^\alpha\varrho_{123}$ weakly, and thus in trace norm. Since again $\mathrm{Tr}_{12}\int_{\mathcal{O}^c}d\mu(\alpha)\widehat{K^{\alpha*}}\widehat{K^\alpha}\hat{\varrho}_{123}\lessdot\varrho_3$, the same argument as above implies that the right side of (B18) converges, in the limit $m\to\infty$, to the corresponding expression without the ^'s. I.e.,

$$\limsup_{m\to\infty}\int_{\mathcal{O}^c} d\mu(\alpha)B_m(\alpha) \leqslant \left(\int_{\mathcal{O}^c} d\mu(\alpha)n^\alpha\right)S[\varrho^c], \quad \text{(B21)}$$

with ϱ^c given as in (B19), but with all the ^'s removed.

Now as $\mathcal{O}\to\Omega$, $\int_{\mathcal{O}^c}d\mu(\alpha)n^\alpha\to 0$. Using again the dominance by ϱ_3, which follows from the fact that $\int_{\mathcal{O}^c}d\mu(\alpha)K^{\alpha*}K^\alpha\leqslant\mathbb{I}$, we see the right side of (B21) goes to zero as $\mathcal{O}\to\Omega$. Together with (B17), this finishes the proof of (B14), and hence the proof of the theorem.

APPENDIX C: RUSKAI'S PROOF OF COROLLARY 2

In an early version of this paper we had a weaker version of Corollary 2, which read $S[\varrho_{12}]-S[\varrho_1]\leqslant S^{\mathrm{cl}}[\varrho_{12}]-S^{\mathrm{cl}}[\varrho_1]$. In a private correspondence, Ruskai suggested the stronger version for the case of (product) coherent states, and her suggestion motivated us to prove the strengthened version of Corollary 2 using our methods. The following is a sketch of her proof.

(1) The map from a state to its coherent state representation is a completely positive, trace-preserving map (CPT).

(2) The *relative* entropy of two density matrices, $H(\varrho,\mu)=\mathrm{Tr}\varrho(\ln\varrho-\ln\mu)$ is known to decrease under CPT maps.

(3) Apply (2) to $H(\varrho_{12},\varrho_1\otimes\varrho_2)$ using the (product) coherent state map, to obtain (35).

[1] E. H. Lieb and M. B. Ruskai, J. Math. Phys. **14**, 1938 (1973). See also Phys. Rev. Lett. **30**, 434 (1973).
[2] E. H. Lieb, Bull. Am. Math. Soc. **81**, 1 (1975).
[3] H. Narnhofer and W. Thirring, Fizika (Zagreb) **17**, 257 (1985).
[4] D. Petz, Rep. Math. Phys. **23**, 57 (1986).
[5] E. Carlen and E. H. Lieb, Am. Math. Soc. Transl. **189**, 59 (1999).
[6] G. Lindblad, Commun. Math. Phys. **40**, 147 (1975).
[7] A. Wehrl, Rep. Math. Phys. **16**, 353 (1979).
[8] M. B. Ruskai also pointed out that CPT map theory can yield most of the results in Sec. III A and we are grateful for her correspondence on this matter. Her beautiful review article [14] gives a thorough introduction to the subject.
[9] E. H. Lieb, Adv. Math. **11**, 267 (1973).
[10] W. Thirring, *Lehrbuch der Mathematischen Physik 4* (Springer, Berlin, 1980).
[11] A. Uhlmann, Wiss. Z.-Karl-Marx-Univ. Leipzig, Math.-Naturwiss. Reihe **22**, 139 (1973).
[12] E. B. Davies, *Quantum Theory of Open Systems* (Academic, New York, 1976), p. 136.
[13] A. Uhlmann, Commun. Math. Phys. **54**, 21 (1977).
[14] M. B. Ruskai, J. Math. Phys. **43**, 4358 (2002).
[15] A. Wehrl, Rep. Math. Phys. **12**, 385 (1977).
[16] A. S. Holevo, Probl. Inf. Transm. **9**, 31 (1973).
[17] A. S. Holevo, *Statistical Structure of Quantum Theory*, Lecture Notes in Physics, Monographs (Springer, Berlin, 2001).
[18] E. H. Lieb, Commun. Math. Phys. **62**, 35 (1978).
[19] R. Seiringer, Commun. Math. Phys. (to be published).
[20] A. Wehrl, Rep. Math. Phys. **10**, 159 (1976).
[21] E. H. Lieb and M. Loss, *Analysis*, 2nd ed. (American Mathematical Society, Providence, RI, 2001).
[22] H. Araki and E. H. Lieb, Commun. Math. Phys. **18**, 160 (1970).
[23] R. Bhatia, *Matrix Analysis* (Springer, Berlin, 1997).