Limit relation for quantum entropy and channel capacity per unit cost

著者	日合　文雄
journal or publication title	Journal of mathematical physics
volume	48
number	9
page range	092102
year	2007
URL	http://hdl.handle.net/10097/46856

doi: 10.1063/1.2779138

JOURNAL OF MATHEMATICAL PHYSICS **48**, 092102 (2007)

Limit relation for quantum entropy and channel capacity per unit cost

Imre Csiszár[a)]
Alfréd Rényi Institute of Mathematics, P.O. Box 127, H-1364 Budapest, Hungary

Fumio Hiai[b)]
Graduate School of Information Sciences, Tohoku University, Aoba-ku, Sendai 980-8579, Japan

Dénes Petz[c)]
Alfréd Rényi Institute of Mathematics, P.O. Box 127, H-1364 Budapest, Hungary

(Received 10 July 2007; accepted 13 August 2007; published online 13 September 2007)

In a thermodynamic model, Diósi *et al.* [Int. J. Quantum Inf. **4**, 99–104 (2006)] arrived at a conjecture stating that certain differences of von Neumann entropies converge to relative entropy as system size goes to infinity. The conjecture is proven in this paper for density matrices. The analytic proof uses the quantum law of large numbers and the inequality between the Belavkin-Staszewski and Umegaki relative entropies. Moreover, the concept of channel capacity per unit cost is introduced for classical-quantum channels. For channels with binary input alphabet, this capacity is shown to equal the relative entropy. The result provides a second proof of the conjecture and a new interpretation. Both approaches lead to generalizations of the conjecture. © *2007 American Institute of Physics.*
[DOI: 10.1063/1.2779138]

I. INTRODUCTION

It was conjectured by Diósi *et al.*[4] that the von Neumann entropy of a quantum state equal to a mixture,

$$R_n := \frac{1}{n}(\sigma \otimes \rho^{\otimes(n-1)} + \rho \otimes \sigma \otimes \rho^{\otimes(n-2)} + \cdots + \rho^{\otimes(n-1)} \otimes \sigma),$$

exceeds the entropy of a component asymptotically by the Umegaki relative entropy $S(\sigma\|\rho)$, that is,

$$S(R_n) - (n-1)S(\rho) - S(\sigma) \to S(\sigma\|\rho), \tag{1}$$

as $n \to \infty$. Here ρ and σ are density matrices acting on a finite dimensional Hilbert space. Recall that $S(\sigma) = -\mathrm{Tr}\, \sigma \log \sigma$ and

$$S(\sigma\|\rho) = \begin{cases} \mathrm{Tr}\, \sigma(\log \sigma - \log \rho) & \text{if supp } \sigma \leq \text{supp } \rho \\ +\infty & \text{otherwise.} \end{cases}$$

Concerning the background of quantum entropy quantities, we refer to Refs. 12 and 14. The set of bounded linear operators on a Hilbert space $\mathcal{H}$ is denoted by $B(\mathcal{H})$. When $\mathcal{H}$ is d dimensional, d finite, $B(\mathcal{H})$ is identified as usual with the set $M_d(\mathbb{C})$ of $d \times d$ matrices with complex entries.

[a)]Electronic mail: csiszar@renyi.hu
[b)]Electronic mail: hiai@math.is.tohoku.ac.jp
[c)]Electronic mail: petz@math.bme.hu

In Ref. 4, a composite system consisting of n molecules has been considered, originally each in a quantum state ρ, and interaction with environment changed the state of one molecule to σ. Irreversibility has been introduced via a completely positive map $\mathcal{M}$ acting as

$$\mathcal{M}(\sigma \otimes \rho^{\otimes(n-1)}) = \frac{1}{n}(\sigma \otimes \rho^{\otimes(n-1)} + \rho \otimes \sigma \otimes \rho^{\otimes(n-2)} + \cdots + \rho^{\otimes(n-1)} \otimes \sigma), \tag{2}$$

interpreted as total randomization over the n subsystems (molecules). A thermodynamical argument showed that the thermodynamical entropy of the system increased by $S(\sigma\|\rho)$. This motivated the conjecture that the increase of the "informatic entropy," given by the left-hand side of Eq. (1), also equals $S(\sigma\|\rho)$, at least in the limit $n\to\infty$.

The quantum formulation includes the case where both ρ and σ are diagonal matrices. This will be referred to as the classical case. If ρ and σ commute, then in an appropriate basis both of them will be diagonal. Apparently, no exact proof of Eq. (1) has been published even for the classical case, although for that case a heuristic proof was offered in Ref. 4.

In this paper, first an analytic proof of Eq. (1) is given, see Theorem 1, using an inequality between the Umegaki and the Belavkin-Staszewski relative entropies and the law of large numbers in the quantum case. The idea is based on the identity

$$R_n = (\rho^{1/2})^{\otimes n}\left(\frac{1}{n}(X \otimes I^{\otimes(n-1)} + I \otimes X \otimes I^{\otimes(n-2)} + \cdots + I^{\otimes(n-1)} \otimes X)\right)(\rho^{1/2})^{\otimes n},$$

where $X=\rho^{-1/2}\sigma\rho^{-1/2}$. The limit of the term in the middle can be computed by the (quantum) law of large numbers. For readers not familiar with the required tools, the arguments are simplified to the classical case, where the ordinary law of large numbers is used, see Theorem 2.

In the second part of this paper, we recognize that $S(R_n)-(n-1)S(\rho)-S(\sigma)$ is a particular Holevo quantity or a classical-quantum mutual information. The Holevo capacity of classical-quantum channels is well understood.[5,8,10] Channel capacity per unit cost has been studied in classical information theory, see primarily Ref. 15, but not in quantum information theory. An indirect approach to capacity per unit cost is possible via the concept of capacity with constrained inputs, which is available for classical-quantum channels.[7] We take a direct approach which—as in the classical case[15]—appears preferable.

We will consider (memoryless) classical-quantum channels with binary input alphabet $\mathcal{X}=\{0,1\}$ which assigns to (classical) input sequences $(x_1,x_2,\ldots,x_n)\in\mathcal{X}^n$ output quantum states $\rho_{x_1}\otimes\rho_{x_2}\otimes\cdots\otimes\rho_{x_n}$, where $\rho_0=\sigma$ and $\rho_1=\rho$, assuming that the cost of an input sequence is the number of characters 0 in it. Considerations similar to Ref. 15 simultaneously provide a proof of Eq. (1), and the result that the capacity per unit cost of the above channel equals $S(\sigma\|\rho)$, see Theorems 3 and 4.

We note that our analytic proof of Eq. (1) requires the assumption that supp $\sigma\leqslant$ supp ρ, while the proof given in Sec. III does not (neither does the simplified version of the analytic proof to the classical case). It is remarkable that the two proofs lead to different generalizations of Eq. (1). The second proof is based on a purely information theoretic interpretation, nevertheless, the result of Theorem 3 admits also a thermodynamical interpretation as in Ref. 4, see the discussion after the proof of Theorem 3.

II. AN ANALYTIC PROOF OF THE CONJECTURE

Assuming that supp $\sigma\leqslant$ supp ρ for the support projections of σ and ρ, one can simply compute

$$S(R_n\|\rho^{\otimes n}) = \mathrm{Tr}(R_n \log R_n - R_n \log \rho^{\otimes n}) = -S(R_n) - (n-1)\mathrm{Tr}\,\rho\log\rho - \mathrm{Tr}\,\sigma\log\rho.$$

Hence, the identity,

$$S(R_n\|\rho^{\otimes n}) = -S(R_n) + (n-1)S(\rho) + S(\sigma\|\rho) + S(\sigma),$$

holds. It follows that conjecture (1) is equivalent to the statement

$$S(R_n\|\rho^{\otimes n}) \to 0 \quad \text{as } n\to\infty,$$

when $\mathrm{supp}\,\sigma \leqslant \mathrm{supp}\,\rho$.

Recall the Belavkin-Staszewski relative entropy

$$S_{\mathrm{BS}}(\omega\|\rho) = \mathrm{Tr}(\omega \log(\omega^{1/2}\rho^{-1}\omega^{1/2})) = -\mathrm{Tr}(\rho\,\eta(\rho^{-1/2}\omega\rho^{-1/2})),$$

if $\mathrm{supp}\,\omega \leqslant \mathrm{supp}\,\rho$, where $\eta(t) := -t\log t$, see Refs. 1 and 12. [The equality of the above two expressions is easily seen from the fact that $Xf(X^*X)=f(XX^*)X$ for a matrix X and for a polynomial f.] It was proved by Hiai and Petz that

$$S(\omega\|\rho) \leqslant S_{\mathrm{BS}}(\omega\|\rho), \tag{3}$$

see Ref. 6 or Proposition 7.11 in Ref. 12.

Theorem 1: *If* $\mathrm{supp}\,\sigma \leqslant \mathrm{supp}\,\rho$, *then* $S(R_n)-(n-1)S(\rho)-S(\sigma)\to S(\sigma\|\rho)$ *as* $n\to\infty$.

Proof: We want to use the quantum law of large numbers, see Proposition 1.17 in Ref. 12. Assume that ρ and σ are $d\times d$ density matrices and we may suppose that ρ is invertible. Due to the GNS construction with respect to the limit φ_∞ of the product states φ_n on the n-fold tensor product $M_d(\mathbb{C})^{\otimes n}$, $n\in\mathbb{N}$, defined by $\varphi_n(A)=\mathrm{Tr}\,\rho^{\otimes n}A$, all finite tensor products $M_d(\mathbb{C})^{\otimes n}$ are embedded into a von Neumann algebra $\mathcal{M}$ acting on a Hilbert space $\mathcal{H}$. If γ denotes the right shift and $X := \rho^{-1/2}\sigma\rho^{-1/2}$, then R_n is written as

$$R_n = (\rho^{1/2})^{\otimes n}\left(\frac{1}{n}\sum_{i=0}^{n-1}\gamma^i(X)\right)(\rho^{1/2})^{\otimes n}.$$

By inequality (3), we get

$$0 \leqslant S(R_n\|\rho^{\otimes n}) \leqslant S_{\mathrm{BS}}(R_n\|\rho^{\otimes n}) = -\mathrm{Tr}(\rho^{\otimes n}\eta((\rho^{-1/2})^{\otimes n}R_n(\rho^{-1/2})^{\otimes n})) = \left\langle \Omega, \eta\left(\frac{1}{n}\sum_{i=0}^{n-1}\gamma^i(X)\right)\Omega\right\rangle, \tag{4}$$

where Ω is the cyclic vector in the GNS construction.

The law of large numbers, see Proposition 1.17 in Eq. (12), gives

$$\frac{1}{n}\sum_{i=0}^{n-1}\gamma^i(X) \to I,$$

in the strong operator topology in $B(\mathcal{H})$, since $\varphi(X)=\mathrm{Tr}\,\rho\rho^{-1/2}\sigma\rho^{-1/2}=1$.

Since the continuous functional calculus preserves the strong convergence (simply due to the approximation by polynomials on a compact set), we obtain

$$\eta\left(\frac{1}{n}\sum_{i=0}^{n-1}\gamma^i(X)\right) \to \eta(I) = 0 \quad \text{strongly}.$$

This shows that upper bound (4) converges to 0 and the proof is complete. □

By the same proof, one can obtain that for

$$R_{\ell,n} := \frac{1}{n-\ell+1}(\sigma^{\otimes \ell}\otimes\rho^{\otimes(n-\ell)} + \rho\otimes\sigma^{\otimes\ell}\otimes\rho^{\otimes(n-\ell-1)} + \cdots + \rho^{\otimes(n-\ell)}\otimes\sigma^{\otimes\ell}),$$

the limit relation

$$S(R_{\ell,n})-(n-\ell)S(\rho)-\ell S(\sigma)\to \ell S(\sigma\|\rho) \tag{5}$$

holds as $n\to\infty$ when ℓ is fixed.

In the next theorem, we treat the classical case in a matrix language.The proof includes the case where supp $\sigma\leqslant$ supp ρ is not true. Those readers who are not familiar with the tools used in the proof of the previous theorem are suggested to follow the arguments below.

Theorem 2: *Assume that* ρ *and* σ *are commuting density matrices. Then*

$$S(R_n)-(n-1)S(\rho)-S(\sigma)\to S(\sigma\|\rho)$$

as $n\to\infty$.

Proof: We may assume that $\rho=\mathrm{Diag}(\mu_1,\ldots,\mu_\ell,0,\ldots,0)$ and $\sigma=\mathrm{Diag}(\lambda_1,\ldots,\lambda_d)$ are $d\times d$ diagonal matrices, $\mu_1,\ldots,\mu_\ell>0$ and $\ell<d$. (We may consider ρ,σ in a matrix algebra of bigger size if ρ is invertible.) If supp $\sigma\leqslant$ supp ρ, then $\lambda_{\ell+1}=\cdots=\lambda_d=0$; this will be called the regular case. When supp $\sigma\leqslant$ supp ρ is not true, we may assume that $\lambda_d>0$ and we refer to the singular case.

The eigenvalues of R_n correspond to elements $(i_1,\ldots,i_n)$ of $\{1,\ldots,d\}^n$,

$$\frac{1}{n}(\lambda_{i_1}\mu_{i_2}\cdots\mu_{i_n}+\mu_{i_1}\lambda_{i_2}\mu_{i_3}\cdots\mu_{i_n}+\cdots+\mu_{i_1}\cdots\mu_{i_{n-1}}\lambda_{i_n}). \tag{6}$$

We divide the eigenvalues in three different groups as follows:

(a) A corresponds to $(i_1,\ldots,i_n)\in\{1,\ldots,d\}^n$ with $1\leqslant i_1,\ldots,i_n\leqslant\ell$;
(b) B corresponds to $(i_1,\ldots,i_n)\in\{1,\ldots,d\}^n$ which contains exactly one d;
(c) C is the rest of the eigenvalues.

If eigenvalue (6) is in group A, then it is

$$\frac{(\lambda_{i_1}/\mu_{i_1})+\cdots+(\lambda_{i_n}/\mu_{i_n})}{n}\mu_{i_1}\mu_{i_2}\cdots\mu_{i_n}.$$

First, we compute

$$\sum_{\kappa\in A}\eta(\kappa)=\sum_{i_1,\ldots,i_n}\eta\left(\frac{(\lambda_{i_1}/\mu_{i_1})+\cdots+(\lambda_{i_n}/\mu_{i_n})}{n}\mu_{i_1}\cdots\mu_{i_n}\right).$$

Below, the summations are over $1\leqslant i_1,\ldots,i_n\leqslant\ell$,

$$\begin{aligned}
&\sum_{i_1,\ldots,i_n}\eta\left(\frac{(\lambda_{i_1}/\mu_{i_1})+\cdots+(\lambda_{i_n}/\mu_{i_n})}{n}\mu_{i_1}\cdots\mu_{i_n}\right)\\
&\quad=-\sum_{i_1,\ldots,i_n}\left(\frac{(\lambda_{i_1}/\mu_{i_1})+\cdots+(\lambda_{i_n}/\mu_{i_n})}{n}\mu_{i_1}\cdots\mu_{i_n}\right)\log(\mu_{i_1}\cdots\mu_{i_n})+Q_n\\
&\quad=-\frac{1}{n}\sum_{k=1}^{n}\left(\sum_{i_1,\ldots,i_n}\lambda_{i_1}\mu_{i_2}\cdots\mu_{i_n}\log\mu_{i_k}+\sum_{i_1,\ldots,i_n}\lambda_{i_2}\mu_{i_1}\cdots\mu_{i_n}\log\mu_{i_k}\right.\\
&\qquad\left.+\cdots+\sum_{i_1,\ldots,i_n}\lambda_{i_n}\mu_{i_1}\cdots\mu_{i_{n-1}}\log\mu_{i_k}\right)+Q_n\\
&\quad=-\frac{1}{n}\sum_{k=1}^{n}\left((n-1)\sum_{i_k}\mu_{i_k}\log\mu_{i_k}+\sum_{i_k}\lambda_{i_k}\log\mu_{i_k}\right)+Q_n=(n-1)S(\rho)-\sum_{i=1}^{\ell}\lambda_i\log\mu_i+Q_n,
\end{aligned}$$

where

$$Q_n := \sum_{i_1,\ldots,i_n} (\mu_{i_1}\cdots\mu_{i_n})\,\eta\left(\frac{(\lambda_{i_1}/\mu_{i_1})+\cdots+(\lambda_{i_n}/\mu_{i_n})}{n}\right).$$

Note that Q_n is equal to the expected value of

$$\eta\left(\frac{X_1+\cdots+X_n}{n}\right)$$

for independent and identically distributed random variables $X_1, X_2, \ldots$ defined on the product probability space $(\{1,2,\ldots,\ell\},(\mu_1,\mu_2,\ldots,\mu_\ell))^{\mathbb{N}}$, where X_n takes the value λ_i/μ_i on a sequence in $\{1,2,\ldots,\ell\}^{\mathbb{N}}$ whose nth component is equal to i.

By the strong law of large numbers,

$$\frac{X_1+\cdots+X_n}{n} \to \mathbb{E}(X_1) = \sum_{i=1}^{\ell}\left(\frac{\lambda_i}{\mu_i}\right)\mu_i = \sum_{i=1}^{\ell}\lambda_i \quad \text{almost surely}.$$

Since $\eta((X_1+\cdots+X_n)/n)$ is uniformly bounded, the Lebesgue bounded convergence theorem implies that

$$Q_n \to \eta\left(\sum_{i=1}^{\ell}\lambda_i\right),$$

as $n\to\infty$.

In the regular case, all nonzero eigenvalues are in group A; hence,

$$S(R_n)-(n-1)S(\rho)-S(\sigma) = -\sum_{i=1}^{\ell}\lambda_i\log\mu_i + \sum_{i=1}^{\ell}\lambda_i\log\lambda_i + Q_n = S(\sigma\|\rho)+Q_n.$$

As $\sum_{i=1}^{\ell}\lambda_i=1$ implies $Q_n\to 0$, the statement follows.

Next, we consider the singular case, when the contributions of the eigenvalues in A is

$$\sum_{\kappa\in A}\eta(\kappa) = (n-1)S(\rho)+O(1),$$

and we turn to eigenvalues in B. If an eigenvalue corresponding to $(i_1,\ldots,i_n)\in\{1,\ldots,d\}^n$ is in group B and $i_1=d$, then the eigenvalue is

$$\frac{1}{n}\lambda_d\mu_{i_2}\cdots\mu_{i_n}.$$

Summation of such eigenvalues gives

$$\begin{aligned}-\sum_{i_2,\ldots,i_n}\left(\frac{\lambda_d\mu_{i_2}\cdots\mu_{i_n}}{n}\right)\log\left(\frac{\lambda_d\mu_{i_2}\cdots\mu_{i_n}}{n}\right) &= -\frac{\lambda_d}{n}\sum_{i_2,\ldots,i_n}(\mu_{i_2}\cdots\mu_{i_n})\log(\mu_{i_2}\cdots\mu_{i_n}) - \frac{\lambda_d}{n}\log\frac{\lambda_d}{n}\\ &= \frac{\lambda_d}{n}(n-1)S(\rho) - \frac{\lambda_d}{n}\log\frac{\lambda_d}{n}.\end{aligned}$$

When $i_j=d$ for some $2\leq j\leq n$, we get the same quantity, so this should be multiplied with n,

$$\sum_{\kappa\in B}\eta(\kappa) = \lambda_d(n-1)S(\rho) - \lambda_d\log\frac{\lambda_d}{n}.$$

It follows that

$$S(R_n)-(n-1)S(\rho)-S(\sigma) \geqslant \sum_{\kappa\in A}\eta(\kappa)+\sum_{\kappa\in B}\eta(\kappa)-(n-1)S(\rho)-S(\sigma) \geqslant \lambda_d(n-1)S(\rho)+\lambda_d \log n + O(1) \to +\infty,$$

as $n\to\infty$. □

III. CHANNEL CAPACITY PER UNIT COST

A classical-quantum channel with classical input alphabet $\mathcal{X}$ transfers the input $x\in\mathcal{X}$ into the output $W(x)\equiv\rho_x$ which is a density matrix acting on a Hilbert space $\mathcal{K}$. We restrict ourselves to the case when $\mathcal{X}$ is finite and $\mathcal{K}$ is finite dimensional.

If a classical random variable X is chosen to be the input, with probability distribution $P=\{p(x):x\in\mathcal{X}\}$, then the corresponding output is the quantum state $\rho_X:=\sum_{x\in\mathcal{X}}p(x)\rho_x$. When a measurement is performed on the output quantum system, it gives rise to an output random variable Y which is jointly distributed with the input X. If a partition of unity $\{F_y:y\in\mathcal{Y}\}$ in $B(\mathcal{K})$ describes the measurement, then

$$\mathrm{Prob}(Y=y|X=x)=\mathrm{Tr}\,\rho_x F_y \quad (x\in\mathcal{X}, y\in\mathcal{Y}). \tag{7}$$

The Holevo bound says that the classical mutual information

$$I(X\wedge Y):=H(Y)-H(Y|X)=H(X)-H(X|Y)$$

(expressed by the classical Shannon entropy H) satisfies[9,10]

$$I(X\wedge Y)\leqslant I(X,W):=S(\rho_X)-\sum_{x\in\mathcal{X}}p(x)S(\rho_x). \tag{8}$$

This bound is a simple consequence of the monotonicity of relative entropy under state transformation[12,13] but has been proved earlier than monotonicity. Here $I(X,W)$ is called a Holevo quantity or a classical-quantum mutual information, and it satisfies the identity

$$\sum_{x\in\mathcal{X}}p(x)S(\rho_x\|\rho)=I(X,W)+S(\rho_X\|\rho), \tag{9}$$

where ρ is an arbitrary density matrix.

When the channel $W:\mathcal{X}\to B(\mathcal{K})$ is used to transfer sequences $\mathbf{x}=(x_1,x_2,\ldots,x_n)\in\mathcal{X}^n$ in a memoryless manner, a sequence $\mathbf{x}=(x_1,x_2,\ldots,x_n)\in\mathcal{X}^n$ is transferred into the quantum state

$$W^{\otimes n}(\mathbf{x})=\rho_{\mathbf{x}}:=\rho_{x_1}\otimes\rho_{x_2}\otimes\cdots\otimes\rho_{x_n}. \tag{10}$$

Formally, this defines a new channel $W^{\otimes n}:\mathcal{X}^n\to B(\mathcal{K}^n)$ called the nth extension of W.

A code of block length n is defined by a subset $A_n\subset\mathcal{X}^n$ called the code word set and by a measurement $\{F_{\mathbf{y}}:\mathbf{y}\in B_n\}$ called the decoder. For technical convenience, the set B_n of possible decoding results may be different from A_n, and only $A_n\subset B_n$ is required. The probability of error is $\mathrm{Prob}(X\neq Y)$, where the input random variable X is uniformly distributed on A_n and the output random variable Y is as in Eq. (7) with x and y replaced by $\mathbf{x}$ and $\mathbf{y}$.

The essential observation is that $S(R_n)-(n-1)S(\rho)-S(\sigma)$ in the conjecture equals the Holevo quantity $I(X,W^{\otimes n})$ for the nth extension of the channel W with input alphabet $\mathcal{X}=\{0,1\}$, $\rho_0=\sigma$, $\rho_1=\rho$ and with X uniformly distributed on those length-n binary sequences that contain exactly one 0. More generally, we shall consider Holevo quantities

$$I(A,\rho_0,\rho_1):=S\left(\frac{1}{|A|}\sum_{\mathbf{x}\in A}\rho_{\mathbf{x}}\right)-\frac{1}{|A|}\sum_{\mathbf{x}\in A}S(\rho_{\mathbf{x}}), \tag{11}$$

defined for any set $A\subset\{0,1\}^n$ of binary sequences of length n.

The concept related to the conjecture we study is the channel capacity per unit cost which is defined next for simplicity only in the case where $\mathcal{X}=\{0,1\}$, the cost of a character $0\in\mathcal{X}$ is 1, while the cost of $1\in\mathcal{X}$ is 0. Given such channel and $\varepsilon>0$, a number $R>0$ is called an ε-achievable rate per unit cost if for every $\delta>0$ and for any sufficiently large T there exists a code of block length $n>T$ with at least $\mathrm{e}^{T(R-\delta)}$ code words such that each of the code words contains at most T 0's and the error probability is at most ε. The largest R which is an ε-achievable rate per unit cost for every $\varepsilon>0$ is the channel capacity per unit cost.

The next theorem is our main result of this section.

Theorem 3: *Let the classical-quantum channel $W:\mathcal{X}=\{0,1\}\to B(\mathcal{K})$ be defined by $W(0)=\rho_0\equiv\sigma$ and $W(1)=\rho_1\equiv\rho$. Let $A_n\subset\{0,1\}^n$, $n=1,2,\ldots$, be sets such that for some natural number ℓ and for some real number $c>0$,*

(a) *each element $\mathbf{x}=(x_1,x_2,\ldots,x_n)\in A_n$ contains at most ℓ copies of 0;*
(b) $\log|A_n|/\log n\to c$ *as* $n\to\infty$;
(c)

$$c(A_n) := \frac{1}{|A_n|}\sum_{\mathbf{x}\in A_n}|\{i:x_i=0\}|\to c \quad \text{as } n\to\infty.$$

Then

$$\lim_{n\to\infty} I(A_n,\sigma,\rho) = cS(\sigma\|\rho).$$

The proof of the theorem is divided into lemmas.

Lemma 1: For an arbitrary $A\subset\{0,1\}^n$,

$$I(A,\rho_0,\rho_1)\leqslant c(A)S(\rho_0\|\rho_1)$$

holds.

Proof: Let $c(\mathbf{x}):=|\{i:x_i=0\}|$ for $\mathbf{x}\in A$. Since $I(A,\rho_0,\rho_1)=I(X,W^{\otimes n})$, we can use identity (9) to get an upper bound

$$\frac{1}{|A|}\sum_{\mathbf{x}\in A}S(\rho_{\mathbf{x}}\|\rho_1^{\otimes n}) = \frac{1}{|A|}\sum_{\mathbf{x}\in A}c(\mathbf{x})S(\rho_0\|\rho_1) = c(A)S(\rho_0\|\rho_1)$$

for $I(A,\rho_0,\rho_1)$. □

Lemma 2: If $A\subset\{0,1\}^n$ is a code word set of a code whose probability of error does not exceed a given $0<\varepsilon<1$, then

$$(1-\varepsilon)\log|A| - \log 2\leqslant I(A,\rho_0,\rho_1).$$

Proof: For the input and output random variables corresponding to the given code, the classical mutual information $I(X\wedge Y)$ is bounded above by $I(X,W^{\otimes n})=I(A,\rho_0,\rho_1)$, see Eq. (8). Since the error probability $\mathrm{Prob}(X\neq Y)$ does not exceed ε, the Fano inequality (see, e.g., Ref. 3) gives

$$H(X|Y)\leqslant \varepsilon\log|A| + \log 2.$$

Therefore,

$$I(X\wedge Y) = H(X) - H(X|Y)\geqslant (1-\varepsilon)\log|A| - \log 2,$$

and the proof is complete. □

We need the direct part of the so-called quantum Stein lemma obtained in Ref. 6, see also Refs. 2, 5, 11, and 14.

Lemma 3: Given arbitrary density matrices ρ and σ in $B(\mathcal{K})$, $\eta>0$, and $0<R<S(\sigma\|\rho)$, if N sufficiently large, there is a projection $E\in B(\mathcal{K}^{\otimes N})$ such that

$$\alpha[E] := \mathrm{Tr}\,\sigma^{\otimes N}(I-E) < \eta$$

and

$$\beta[E] := \mathrm{Tr}\,\rho^{\otimes N}E < \mathrm{e}^{-NR}.$$

Here E [or the measurement $(E, I-E)$] is interpreted as a test of the null hypothesis that the state is $\sigma^{\otimes N}$, against the alternative hypothesis that it is $\rho^{\otimes N}$. This test incorrectly accepts the null hypothesis (error of the first kind) with probability $\alpha[E]$, and incorrectly rejects it (error of the second kind) with probability $\beta[E]$.

Lemma 4: Assume that $\varepsilon > 0$, $0 < R < S(\rho_0 \| \rho_1)$, and ℓ is a positive integer. If n is large enough, then for any set A_n of sequences $\mathbf{x} \in \{0,1\}^n$ that contain at most ℓ copies of 0, there exists a code with error probability smaller than ε whose code words are the N-fold repetitions $\mathbf{x}^N = (\mathbf{x}, \mathbf{x}, \ldots, \mathbf{x})$ of the sequences $\mathbf{x} \in A_n$, where N is the smallest integer satisfying

$$N \geq \frac{1}{R} \log \frac{2n}{\varepsilon}.$$

Proof: We follow the probabilistic construction in Ref. 15. The output states corresponding to input sequences of length nN are density matrices acting on the Hilbert space $\mathcal{K}^{\otimes Nn} \equiv (\mathcal{K}^{\otimes n})^{\otimes N}$. We decompose this Hilbert space into an N-fold tensor product in a different way. For each $1 \leq i \leq n$, let $\mathcal{K}_i$ be the tensor product of the factors $i, i+n, i+2n, \ldots, i+(N-1)n$. So $\mathcal{K}^{\otimes Nn}$ is identified with $\mathcal{K}_1 \otimes \mathcal{K}_2 \otimes \cdots \otimes \mathcal{K}_n$.

We construct a decoder to the code word set in the lemma as follows. For each $1 \leq i \leq n$, we test the null hypothesis that the ith component of the actually chosen $\mathbf{x} \in A_n$ is 0, against the alternative that it is 1, based on the channel outputs at time instances $i, i+n, \ldots, i+(N-1)n$. More exactly, let the projection $E_i \in B(\mathcal{K}_i)$ be a test of the null hypothesis $\sigma^{\otimes n}$ against the alternative $\rho^{\otimes n}$. According to the quantum Stein lemma (Lemma 3), applied with $\eta = \varepsilon/2\ell$ and the given $0 < R < S(\sigma \| \rho)$, for N sufficiently large, there exists a test E_i such that the probability of error of the first kind is smaller than η, while the probability of error of the second kind is smaller than $\mathrm{e}^{-NR} \leq \varepsilon/2n$. The projections E_i and $I-E_i$ form a partition of unity in the Hilbert space $\mathcal{K}_i$, and the n-fold tensor product of these commuting projections will give a partition of unity in $\mathcal{K}^{\otimes Nn}$. For $\mathbf{y} \in \{0,1\}^n$, set $F_{\mathbf{y}} := \otimes_{i=1}^n F_{y_i}$, where $F_{y_i} = E_i$ if $y_i = 0$ and $F_{y_i} = I - E_i$ if $y_i = 1$, and let the decoder be the measurement $\{F_{\mathbf{y}} : \mathbf{y} \in \{0,1\}^n\}$. Thus, the result of the decoding will be an arbitrary 0–1 sequence in $\{0,1\}^n$.

The error probability should be estimated,

$$\begin{aligned}\mathrm{Prob}(Y \neq X | X = \mathbf{x}) &= \sum_{\mathbf{y}:\mathbf{y}\neq\mathbf{x}} \mathrm{Tr}\,\rho_{\mathbf{x}}^{\otimes N} F_{\mathbf{y}} = \sum_{\mathbf{y}:\mathbf{y}\neq\mathbf{x}} \prod_{i=1}^n \mathrm{Tr}\,\rho_{x_i}^{\otimes N} F_{y_i} \leq \sum_{i=1}^n \sum_{\mathbf{y}:y_i \neq x_i} \prod_{j=1}^n \mathrm{Tr}\,\rho_{x_j}^{\otimes N} F_{y_j} \\ &\leq \sum_{i=1}^n \mathrm{Tr}\,\rho_{x_i}^{\otimes N}(I - F_{x_i}).\end{aligned}$$

If $x_i = 0$, then

$$\mathrm{Tr}\,\rho_{x_i}^{\otimes N}(I - F_{x_i}) = \mathrm{Tr}\,\rho_0^{\otimes N}(I - E_i) = \alpha(E_i) \leq \frac{\varepsilon}{2\ell},$$

and if $x_i = 1$,

$$\mathrm{Tr}\,\rho_{x_i}^{\otimes N}(I - F_{x_i}) = \mathrm{Tr}\rho_1^{\otimes N} E_i = \beta(E_i) \leq \mathrm{e}^{-RN} \leq \frac{\varepsilon}{2n}.$$

As $x_i = 0$ holds for at most ℓ indices, it follows that the probability of error of this code is $\mathrm{Prob}(X \neq Y) \leq \varepsilon$. □

Proof of Theorem 3: Since Lemma 1 gives the upper bound,

$$\limsup_{n\to\infty} I(A_n,\rho_0,\rho_1) \leqslant cS(\sigma\|\rho),$$

it remains to prove that

$$\liminf_{n\to\infty} I(A_n,\rho_0,\rho_1) \geqslant cS(\sigma\|\rho).$$

By Lemma 4, the set $\{\mathbf{x}^N:\mathbf{x}\in A_n\}$ with N given there is the code word set of a code with error probability smaller than ε. According to Lemma 2 and Eq. (11), this implies

$$(1-\varepsilon)\log|A_n| - \log 2 \leqslant S(\rho_{X^N}) - \frac{1}{|A_n|}\sum_{\mathbf{x}\in A_n} S(\rho_{\mathbf{x}^N}),$$

where X is uniformly distributed on A_n and X^N denotes its N-fold repetition.

From the subadditivity of the von Neumann entropy, we have

$$S(\rho_{X^N}) \leqslant NS(\rho_X)$$

and

$$S(\rho_{\mathbf{x}^N}) = NS(\rho_{\mathbf{x}})$$

holds due to the additivity for product. It follows that

$$(1-\varepsilon)\frac{\log|A_n}{N} - \frac{1}{N} \leqslant S(\rho_X) - \frac{1}{|A_n|}\sum_{\mathbf{x}\in A_n} S(\rho_{\mathbf{x}}) = I(A_n,\rho_0,\rho_1).$$

From the choice of N in Lemma 4, we have

$$R\frac{\log|A_n}{\log n}\frac{\log n}{\log n + \log 2 - \log \varepsilon} \leqslant \frac{\log|A_n}{N},$$

and the lower bound is arbitrarily close to cR. Since $R < S(\rho_0\|\rho_1)$ was arbitrary, the proof is complete. □

Assume that A_n is the set of all $\mathbf{x}\in\{0,1\}^n$ containing exactly ℓ 0's for a fixed natural number ℓ. Then $c(A_n)=\ell$ and from the Stirling formula, one can easily check $\log|A_n|/\log n \to \ell$. Consequently, Theorem 3 proves that

$$S(R_n(\ell)) - (n-\ell)S(\rho) - \ell S(\sigma) \to \ell S(\sigma\|\rho) \tag{12}$$

holds as $n\to\infty$ when ℓ is fixed and

$$R_n(\ell) := \binom{n}{\ell}^{-1} \sum_{\mathbf{x}\in A_n} \rho_{x_1}\otimes\rho_{x_2}\otimes\cdots\otimes\rho_{x_n} \quad (\rho_0=\sigma,\rho_1=\rho).$$

In particular, when $\ell=1$, conjecture (1) is proven in full generality. We have two generalizations [Eqs. (5) and (12)] of Eq. (1), which are similar but different.

We note that Eq. (12) admits a thermodynamical interpretation, analogous of that of Eq. (1) in Ref. 4, sketched in the Introduction. Indeed, suppose that interaction with the environment changes the state not of 1 but ℓ molecules to σ and irreversibility is introduced again by total randomization. The new state of the system will be $R_n(\ell)$ above and Eq. (12) says that "informatic entropy" increases by ℓ times the relative entropy (in the limit as system size goes to infinity).

Theorem 4: *The capacity per unit cost of the channel with a binary input alphabet and $W(0)=\rho_0$, $W(1)=\rho_1$ is equal to the relative entropy $S(\rho_0\|\rho_1)$.*

Proof: Assume that $R>0$ is an ε-achievable rate per unit cost. For every $\delta>0$ and $T>0$, there is a code $A\subset\{0,1\}^n$ for which we get by Lemmas 1 and 2,

$$TS(\rho_0\|\rho_1) \geqslant c(A)S(\rho_0\|\rho_1) \geqslant I(A,\rho_0,\rho_1) \geqslant (1-\varepsilon)\log|A| - \log 2 \geqslant (1-\varepsilon)T(R-\delta) - \log 2.$$

Since T is arbitrarily large and ε,δ are arbitrarily small, $R\leqslant S(\rho_0\|\rho_1)$ follows.

Let A_n be the set of $\mathbf{x}\in\{0,1\}^n$ containing exactly one 0, and consider the N-times repeated code words given in Lemma 4. Then each of the n code words contains exactly N 0's. For every $R<S(\rho_0\|\rho_1)$ and $\varepsilon,\delta>0$, if N is chosen as in Lemma 4, we have

$$n \geqslant \frac{\varepsilon}{2}e^{NR} = \frac{\varepsilon e^{N\delta}}{2}e^{N(R-\delta)} > e^{N(R-\delta)}$$

as long as n is so large that N satisfies $\varepsilon e^{N\delta}/2>1$. This implies that R is an ε-achievable rate per unit cost for every $\varepsilon>0$. Hence, the result follows. □

ACKNOWLEDGMENTS

The first author (I.C.) was partially supported by the Hungarian Research Grants OTKA T046376 and T068258. The second author (F.H.) was partially supported by Grant-in-Aid for Scientific Research (B)17340043. The third author was partially supported by the Hungarian Research Grant T068258.

[1] Belavkin, V. P., and Staszewski, P., "C*-algebraic generalization of relative entropy and entropy," Ann. Inst. Henri Poincare, Sect. A **37**, 51–58 (1982).
[2] Bjelaković, I., Deuschel, J., Krüger, T., Seiler, R., Siegmund-Schultze, R., and Szkoła, A., "A quantum version of Sanov's theorem," Commun. Math. Phys. **260**, 659–671 (2005).
[3] Cover, T. M., and Thomas, J. A., *Elements of Information Theory*, 2nd ed. (Wiley-Interscience, Hoboken, NJ, 2006).
[4] Diósi, L., Feldmann, T., and Kosloff, R., "On the exact identity between thermodynamic and informatic entropies in a unitary model of friction," Int. J. Quantum Inf. **4**, 99–104 (2006).
[5] Hayashi, M., *Quantum Information: An Introduction* (Springer, New York, 2006).
[6] Hiai, F., and Petz, D., "The proper formula for relative entropy and its asymptotics in quantum probability," Chem. Mater. **143**, 99–114 (1991).
[7] Holevo, A. S., e-print arXiv:quant-ph/9705054.
[8] Holevo, A. S., "Quantum coding theorems," Russ. Math. Surveys **53**, 1295–1331 (1998).
[9] Holevo, A. S., "Some estimates for the amount of information transmittable by a quantum communication channel (in Russian)," Problemy Peredachi Informacii **9**, 3–11 (1973).
[10] Nielsen, M. A., and Chuang, I. L., *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).
[11] Ogawa, T., and Nagaoka, H., "Strong converse and Stein's lemma in quantum hypothesis testing," IEEE Trans. Inf. Theory **46**, 2428–2433 (2000).
[12] Ohya, M., and Petz, D., *Quantum Entropy and its Use* (Springer, New York, 1993).
[13] Ohya, M., Petz, D., and Watanabe, N., "On capacities of quantum channels," Probab. Math. Stat. **17**, 179–196 (1997).
[14] Petz, D., Quantum Information Theory and Quantum Statistics, to be published.
[15] Verdu, S., "On channel capacity per unit cost," IEEE Trans. Inf. Theory **36**, 1019–1030 (1990).