PHYSICAL REVIEW LETTERS

VOLUME 70 25 JANUARY 1993 NUMBER 4

Ultimate Information Carrying Limit of Quantum Systems

Horace P. Yuen
Department of Electrical Engineering and Computer Science, Department of Physics and Astronomy, Northwestern University, Evanston, Illinois 60208

Masanao Ozawa
Department of Mathematics, College of General Education, Nagoya University, Nagoya 464, Japan
(Received 28 August 1992)

The possible amount of information transfer between any source and any user via a quantum system is bounded through the quantum entropy function. In contrast to the classical case, this shows that infinite information transfer implies infinite entropy. The entropy bound is also applied to obtain the ultimate quantum information transmission capacity of the free electromagnetic field under a power and a bandwidth constraint.

PACS numbers: 03.65.Bz, 05.90.+m, 42.50.Dv

All data transmissions, in computer and communications technology as well as in other human endeavors, have to be accomplished by physical processes subject to the laws of physics. A fundamental issue arises as to the maximum amount of information one can transfer with a physical system described by quantum mechanics. By *information* we mean the standard measure of quantity of data in statistical communication theory, namely, the Shannon entropy [1] of a data source. If a source generates one of M possible messages with equal probability during each use, and successive generations are statistically independent, the information rate is $\log_2 M$ bits per use, or $r \log_2 M$ bits per second for r uses per second. Thus, if a real variable over any interval is selected, the source information rate is infinite per use. If such a variable can be transmitted without error to a user, the resulting information transfer between the source and the user is also infinite per use. As we first discuss in the following, this situation prevails when the physical system for information transfer is described by classical physics, even in the absence of any compensating infinite magnitude such as infinite power, infinite entropy, or infinite space. Only infinite (ideal) resolution is required. In this paper we show that in quantum physics infinite information transfer requires infinite quantum entropy, which in turn would require some other physical magnitude to become infinite. We will demonstrate that the underlying cause of this phenomenon of infinite resolution requiring an infinite magnitude is a manifestation of the uncertainty principle. In this regard we would suggest that quantum physics is *de facto* more realistic than classical physics. Toward our end we will give a rigorous proof of an entropy inequality that has been widely conjectured before, but never proved in any useful generality. As another application of this inequality in quantum communication theory, we will also establish the ultimate information carrying capacity of a free boson field.

The entropy of a discrete random variable with probability distribution p_i is $-\sum_i p_i \log p_i$, which is always non-negative. The (*differential*) *entropy* of a continuous random variable $\mathbf{x}$ with probability density $p(x)$ is

$$S(\mathbf{x}) \equiv -\int p(x) \log p(x) dx. \tag{1}$$

As in statistical physics, $S(\mathbf{x})$ has no absolute significance—it is defined up to an arbitrary additive constant depending on the scale one measures $\mathbf{x}$. However, the difference between two differential entropies does have an absolute significance, as in the case of the *(average) mutual information* between two continuous random variables $\mathbf{x}$ and $\mathbf{y}$,

$$I(\mathbf{x};\mathbf{y}) \equiv S(\mathbf{x}) - S(\mathbf{x}|\mathbf{y}), \tag{2}$$

where $S(\mathbf{x}|\mathbf{y}) \equiv -\int dy\, p(y) \int p(x|y) \log p(x|y)\, dx$ is the *conditional (differential) entropy.* The definition of mutual information $I(\mathbf{x};\mathbf{y})$ is extended to *any* pair of random variables $\mathbf{x}$ and $\mathbf{y}$ by using Radon-Nikodym derivatives as discussed later.

The significance of $I(\mathbf{x},\mathbf{y})$ lies in Shannon's coding theorem and its converse, which [1] states that error-free communication via $\mathbf{x}$ and $\mathbf{y}$ is possible if and only if the source rate per use is less than $I(\mathbf{x};\mathbf{y})$. In the noise-free limit of ideal resolution referred to above, $I(\mathbf{x};\mathbf{y})$ indeed becomes infinite since $p(x|y)=\delta(x-y)$ which makes the conditional entropy $S(\mathbf{x}|\mathbf{y})$ negative infinite. However, $S(\mathbf{x})$ can remain finite and it often does, as in the case of a Gaussian random variable.

Usually, $\mathbf{x}$ is taken to be the input of a "channel," and $\mathbf{y}$ the channel output, with the *channel* specified by the conditional probability density $p(y|x)$. The mutual information $I(\mathbf{x};\mathbf{y})$ is maximized with respect to $p(x)$ to yield the so-called *capacity* of the channel.

Information transmission between a source and a user via a quantum system can be formulated as follows. Let Θ be the system input alphabet *one picks*, which is either a discrete set or a subset of $\mathbf{R}^n$ consisting of n-dimensional real vectors or a mixed discrete-continuous set. In general a source coder may be employed to convert the data from any given data source into elements from the alphabet Θ before transmission; this has no relevance for our purpose and will not be discussed. Each $\theta\in\Theta$ is to be modulated into a quantum system with Hilbert space $\mathcal{H}$ of pure states, so that the system state as presented to the user is given by a density operator ρ_θ on $\mathcal{H}$. That is, the mapping $\theta\mapsto\rho_\theta$ includes all effects of modulation, transmission, and noise. It represents the "quantum channel" or the quantum channel plus modulation. The user determines which θ was transmitted through the result of a quantum measurement on the system.

The most general description of a quantum measurement that can be performed on a system is given by the mathematical concept of a completely positive instrument [2] on the system state space. It can be easily shown that for extracting information, it suffices to concentrate on the measurement probability without the need of successive measurements on the already measured system. The most general description of quantum measurement probability is given by the mathematical concept of a *probability-operator-valued measure* (POM) [2, 3] on the system state space. To facilitate comprehension, we describe a POM with the strictly speaking illegitimate mathematical notation $X(x)$ instead of $X(dx)$ that involves the precise mathematical definition of a POM. [They are related by the *formal* decomposition $X(dx)=X(x)\,dx$.] In this notation a POM X corresponding to a measurement with output variable $\mathbf{x}$, with values $x\in\mathbf{R}^n$, is an operator-valued generalized function $X(x)$ such that for each x, $X(x)$ is a positive (formal) self-adjoint operator and all the $X(x)$ sum to the identity operator, i.e.,

$$X(x)\geq 0, \quad \int X(x)dx = I. \tag{3}$$

When $X(x)=|x\rangle\langle x|$ with orthogonal $|x\rangle$'s, the POM X corresponds to a unique self-adjoint operator $\hat{X}\equiv\int xX(x)dx$ obeying the function calculus $f(\hat{X})=\int f(x)X(x)dx$, whereas this is *not* satisfied for an arbitrary POM. This POM description, by now fairly well known, generalizes the usual textbook description of quantum measurements by self-adjoint operators in that it includes the case when some "apparatus" is adjointed to the system and commuting self-adjoint operators on the system plus apparatus are measured to produce an outcome referring to the system alone [2, 3]. When the POM X is measured on a system in state σ, the *output probability distribution* $\mu[\sigma]$ is the probability distribution of $\mathbf{x}$ given by

$$\mu[\sigma](dx)=\mathrm{Tr}[X(dx)\sigma]. \tag{4}$$

Now suppose that we are given the input variable $\boldsymbol{\theta}$ with probability distribution $P(d\theta)$ on the alphabet Θ, the parametrized states ρ_θ ($\theta\in\Theta$), and the measurement described by POM X with output variable $\mathbf{x}$. Then, the conditional probability distribution of the output $\mathbf{x}$ given input θ is obtained by $P(dx|\theta)=\mu[\rho_\theta](dx)$, and hence we have a standard information theoretical channel with conditional probability distribution $\mu[\rho_\theta](dx)$, for which we can talk about the mutual information. The amount of information transfer between the input alphabet Θ and the user that measures POM X is described by the mutual information $I(\boldsymbol{\theta};\mathbf{x})$ between the random variables $\boldsymbol{\theta}$ and $\mathbf{x}$. Taking account of the symmetry property $I(\boldsymbol{\theta};\mathbf{x})=I(\mathbf{x};\boldsymbol{\theta})$, an expression for $I(\boldsymbol{\theta};\mathbf{x})$ is obtained as follows. Consider the *mixture of parametrized states* ρ_θ,

$$\rho\equiv\int\rho_\theta P(d\theta). \tag{5}$$

The (unconditional) probability distribution of $\mathbf{x}$ is then $P(dx)=\mu[\rho](dx)$. The classical relative entropy of $\mu[\rho_\theta]$ with respect to $\mu[\rho]$ is given by

$$S(\mu[\rho_\theta]/\mu[\rho])\equiv\int\mu[\rho_\theta](dx)\log\frac{d\mu[\rho_\theta]}{d\mu[\rho]}(x), \tag{6}$$

where $d\mu[\rho_\theta]/d\mu[\rho]$ is the Radon-Nikodym derivative of $\mu[\rho_\theta]$ with respect to $\mu[\rho]$. Then the mutual information is given by

$$I(\boldsymbol{\theta};\mathbf{x})=\int S(\mu[\rho_\theta]/\mu[\rho])P(d\theta). \tag{7}$$

Note that the maximum $I(\boldsymbol{\theta};\mathbf{x})$ obtained by optimization over $P(d\theta)$ is the capacity of this information channel. When Θ, ρ_θ ($\theta\in\Theta$), and X are allowed to vary, the resulting maximum $I(\boldsymbol{\theta};\mathbf{x})$ obtained for a given quantum system can be referred to as the *ultimate quantum capacity* of the system. The search for such maximum quantum information transfer may be greatly facilitated by the following entropy bound.

The entropy of a quantum state σ is $S(\sigma)\equiv$

$-\mathrm{Tr}[\sigma \log \sigma]$. For arbitrary parameter $\boldsymbol{\theta}$ with probability distribution $P(d\theta)$ on Θ, parametrized states ρ_θ ($\theta \in \Theta$) with mixture ρ, and POM X, the following inequality holds.

Theorem (the entropy bound).—

$$I(\boldsymbol{\theta};\mathbf{x}) \le S(\rho) - \int S(\rho_\theta)P(d\theta). \tag{8}$$

We will prove this theorem via the following general inequality.

The Uhlmann inequality: Let $\Phi : U_1 \to U_2$ be a unit-preserving completely positive map from a C*-algebra U_1 to a C*-algebra $U_2, \Phi^* : U_2^* \to U_1^*$ the dual map, and $S(\cdot/\cdot)$ the relative entropy between two states of a C*-algebra. Then for any two states σ_1 and σ_2 of U_2,

$$S(\Phi^*(\sigma_1)/\Phi^*(\sigma_2)) \le S(\sigma_1/\sigma_2). \tag{9}$$

This inequality (9) is proved by Uhlmann [4], generalizing the results by Araki [5], Lindblad [6], and Umegaki [7]. Its importance in quantum information theory has also been noted by Ohya [8].

To prove (8) we write

$$S(\rho) - \int S(\rho_\theta)P(d\theta) = \int S(\rho_\theta/\rho)P(d\theta), \tag{10}$$

where the relative entropy between two density operators is given by

$$S(\sigma_1/\sigma_2) \equiv \mathrm{Tr}[\sigma_1 \log \sigma_1 - \sigma_1 \log \sigma_2]. \tag{11}$$

We will show that for each θ,

$$S(\mu[\rho_\theta]/\mu[\rho]) \le S(\rho_\theta/\rho). \tag{12}$$

From (7), (10), and (12) our bound (8) follows. To obtain (12) from (9), let $U_1 = C(\overline{\mathbf{R}^n})$, the algebra of complex-valued continuous functions on $\overline{\mathbf{R}^n}$, the one-point compactification of $\mathbf{R}^n$, and $U_2 = \mathcal{L}(\mathcal{H})$, the algebra of bounded linear operators on $\mathcal{H}$. Then U_1^* is the space of countably additive complex-valued finite Borel measures, and U_2^* *includes* the space of trace class operators on $\mathcal{H}$. Let $\Phi : U_1 \to U_2$ be the map such that $\Phi(f) = \int f(x)X(dx)$. Then the dual $\Phi^* : U_2^* \to U_1^*$ maps a density operator σ to the probability measure $\mathrm{Tr}[X(dx)\sigma]$, i.e., $\Phi^*(\sigma) = \mu[\sigma]$. Under this Φ, which is trivially completely positive, the inequality (9) becomes (12) and the proof is completed.

The entropy bound has a long and complicated history which we can only briefly outline here. Forney [9] first gave an interesting discussion but an incorrect proof in an unpublished master thesis. In the published literature, it was first explicitly conjectured by Gordon [10]. Since then there have been a few more unpublished incomplete or incorrect proofs. All of these appeared before POM found its way into quantum mechanics. The one correct proof in print by Holevo [11] is restricted under the conditions of a finite set Θ and POM defined over a finite-dimensional operator *-algebra; the latter essentially restricts its applicability to a finite-dimensional state space $\mathcal{H}$. It is clear that either one of these restrictions drastically reduces the significance of the entropy bound, and we will see later that under such restrictions one cannot establish the ultimate capacity limit of a quantum channel which one can do with (8) in its full generality.

The entropy bound is, of course, not a surprising result as it was conjectured and thought to be established many times before. Indeed, it is a very intuitively satisfying result. It implies in particular, since $S(\rho_\theta) \ge 0$,

$$I(\boldsymbol{\theta};\mathbf{x}) \le S(\rho), \tag{13}$$

i.e., independently of the quantum measurement one may make, the information transfer is bounded by the quantum entropy of the mixed state presented to the user as averaged over the input alphabet [12]. Thus, whenever there is an infinite amount of information transfer between the source and the user, the entropy $S(\rho)$ would become infinite too. This is in marked contrast to the situation in classical physics discussed above, where infinite information transfer can occur without the (differential) entropy becoming infinite itself. The absolute character of the quantum entropy thus enforces the requirement that some "physical magnitude" would become infinite for infinite information transfer. In the following, we will examine how this infinite entropy gives rise to an infinite physical magnitude in some examples which are of basic interest in their own right.

First consider the maximum possible amount of information transfer with a single electromagnetic field mode subject to the constraint of a given level of average photons available, i.e., we wish to pick a set Θ, a probability distribution $P(d\theta)$ on Θ, a mapping $\theta \mapsto \rho_\theta$, and a measurement X, such that $I(\boldsymbol{\theta};\mathbf{x})$ is as large as possible subject to the constraint

$$\int \mathrm{Tr}[a^\dagger a \rho_\theta]P(d\theta) \le N, \tag{14}$$

where a is the photon annihilation operator of the field mode. Complicated as it may seem, this problem is easily solved from the entropy bound (8) or (13) since it is well known that the entropy of a boson field mode is maximized, under an average energy constraint, by the number of eigenstates with the exponential distribution—the canonical ensemble of quantum statistical mechanics. Thus, by optimizing the right-hand side of (8) and achieving the optimum, we see that $I(\boldsymbol{\theta};\mathbf{x})$ is maximized by using $\Theta = \mathcal{N}$, the set of natural numbers, together with $n \mapsto \rho_n = |n\rangle\langle n|$ ($n \in \mathcal{N}$), $P(\{n\}) = N^n(1+N)^{-(n+1)}$, and $\hat{X} = a^\dagger a$. It yields the result, often conjectured or alleged but *never proved before*, that the ultimate quantum capacity is given by the following entropy in nats per use per mode [1 nat $= (\log 2)^{-1}$ bits from logarithm with base e]:

$$\max I(\boldsymbol{\theta};\mathbf{x}) = C(N) \equiv (N+1)\log(N+1) - N\log N. \tag{15}$$

It is immediate from (15) that infinite information transfer in this case requires infinite N. It should be emphasized that the entropy bound in its full generality is required to establish this result (15). This is because the state subspace satisfying $\mathrm{Tr}[a^\dagger a\rho] \leq N$ is infinite dimensional, and, more importantly, there is *no a priori reason why a continuous* Θ *with matching measurement would not yield a higher* $I(\boldsymbol{\theta};\mathbf{x})$. Certainly it would classically, as we discussed above. More specifically, one would wonder why the use of two-photon coherent states [13] and homodyne detection, or some other continuous observable and states which involve a continuous $\mathbf{x}$, may not do better. It is here that the energy limit (14) plays a crucial role; a highly squeezed state requires too many photons wasted in the large noise quadrature, a requirement from the uncertainty principle. Similarly, phase eigenstates also have infinite number of photons [14] and the phase is not as good as the energy for information transmission which was first argued by Gabor [15]. One may *not*, however, conclude that it is always the energy that is placing the limitation, as the following second example shows.

Consider a single quantum degree of freedom with the Hamiltonian

$$H = P^2, \tag{16}$$

where P has a spectrum equal to the real line and can be interpreted, for example, as the momentum of an unbounded fermion. It is clear that under an average H constraint,

$$\int \mathrm{Tr}[H\rho_\theta]P(d\theta) \leq N, \tag{17}$$

an infinite information transfer is possible with, say, $\Theta = [0,1]$, $\rho_\theta = |p_\theta\rangle\langle p_\theta|$, $\hat{X} = P$, where each $|p_\theta\rangle$ is an eigenstate of P, $p_\theta \neq p_{\theta'}$ for $\theta \neq \theta'$, such that (17) is satisfied for whatever $P(d\theta)$ we pick, e.g., $|p_\theta| \leq N^{1/2}$. The only important point here is that for an observable with a continuous spectrum, error-free transmission is possible via the generalized eigenstates, and the energy constraint need not be violated when the continuous observable is itself the Hamiltonian. It would not matter if the state is strictly confined to the Hilbert space $\mathcal{H}$; that would merely replace $I(\boldsymbol{\theta};\mathbf{x}) = \infty$ by $I(\boldsymbol{\theta};\mathbf{x}) \to \infty$; i.e., $I(\boldsymbol{\theta};\mathbf{x})$ can be made larger than any specified number, say, by approaching a Dirac delta function with Gaussians of increasingly narrow widths. This result does not contradict the entropy bound because $S(\rho) \to \infty$ in such limits, too [16]. The catch in this case is, of course, that the observable Q conjugate to P would take on infinite values as a consequence of the uncertainty principle.

We conclude by pointing out that the number-state photon-counting capacity for the free electromagnetic field is indeed the ultimate quantum capacity under an average power constraint and a frequency bandwidth constraint. With the assumption that number states and photon counting are used, it has been shown [17] that the capacity in nats per second, converted from per mode per use with a frequency band $\mathcal{W}$ (modes per second), is given by

$$C = \int_{\mathcal{W}} C\Big(\frac{1}{e^{\lambda hf}-1}\Big)df, \tag{18}$$

where $C(\cdot)$ is given by (15) and λ is determined from the power constraint

$$\mathcal{P} = \int_{\mathcal{W}} \frac{hf}{e^{\lambda hf}-1}df. \tag{19}$$

The optimality of number states and photon counting among all possible states and measurements can be proved from (8) similar to the derivation of (15). We have a collection of statistically independent frequency modes, each one collectively subject to the energy constraint $\mathcal{P}T$ in any time interval T. Thus, with $S(\rho_j) = 0$ for a pure state, the entropy is maximized by number states. The limit $T \to \infty$ can be properly taken with the help of Teoplitz distribution theorems, and the problem reduces to obtaining the frequency distribution of power that maximizes the total entropy, which is the Bose-Einstein distribution given in (19). In the limit $\mathcal{W} \to \infty$, we have the infinite-bandwidth ultimate capacity [17, 18]

$$C \to \pi\sqrt{\frac{2\mathcal{P}}{3h}}. \tag{20}$$

This work was supported by the National Science Foundation and the Office of Naval Research. M.O. is supported in part by the Okawa Institute of Information and Telecommunication.

[1] C.E. Shannon, Bell System Tech. J. **27**, 379 (1948); **27**, 623 (1948); Proc. IRE **37**, 10 (1949).
[2] M. Ozawa, J. Math. Phys. **25**, 79 (1984).
[3] H.P. Yuen, Phys. Lett. **91A**, 101 (1982).
[4] A. Uhlmann, Commun. Math. Phys. **54**, 21 (1977).
[5] H. Araki, Publ. RIMS Kyoto Univ. **11**, 809 (1976).
[6] G. Lindblad, Commun. Math. Phys. **39**, 111 (1974).
[7] H. Umegaki, Kodai Math. Sem. Rep. **14**, 59 (1962).
[8] M. Ohya, Rep. Math. Phys. **27**, 19 (1989).
[9] G.D. Forney, Jr., S.M. thesis, Massachusetts Institute of Technology, 1963 (unpublished).
[10] J.P. Gordon, in *Quantum Electronics and Coherent Light*, Proceedings of the International School of Physics "Enrico Fermi," Course XXXI, edited by P.A. Miles (Academic, New York, 1964), p. 156.
[11] A.S. Holevo, Probl. Inf. Transm. **9**, 177 (1973).
[12] This bound (13) might be regarded as a rigorous version of Brillouin's "negentropy principle of information;" see L. Brillouin, *Scientific Uncertainty and Information* (Academic, New York, 1964). The detailed discussion on interpretation and difference will be given elsewhere.
[13] H.P. Yuen, Phys. Rev. A **13**, 2226 (1976).
[14] J.P. Shapiro, S.R. Shepard, and N.C. Wong, Phys. Rev. Lett. **62**, 2377 (1989).
[15] D. Gabor, Philos. Mag. **41**, 1161 (1950).
[16] Note that, for example, $\mathrm{Tr}(\exp -\beta P^2)$ is infinite, which is also revealed by a careful use of Dirac notation.
[17] J.I. Bowen, IEEE Trans. Inf. Theory **13**, 230 (1967).
[18] D.S. Lebedev and L.B. Levetin, Inform. Control **9**, 1 (1966).