ARTICLES

Lower bound for accessible information in quantum mechanics

Richard Jozsa,[1,*] Daniel Robb,[2,3] and William K. Wootters[3]

[1] *Département d'Informatique et de Recherche Opérationelle, Université de Montréal, Case Postale 6128, Succursale "A," Montréal, Québec, Canada H3C 3J7*

[2] *Department of Physics, The University of Texas at Austin, Austin, Texas 78712*

[3] *Department of Physics, Williams College, Williamstown, Massachusetts 01267*

(Received 16 August 1993)

It has long been known that the von Neumann entropy S is an upper bound on the information one can extract from a quantum system in an unknown pure state. In this paper we define the "subentropy" Q, which we prove to be a *lower* bound on this information. Moreover, just as the von Neumann entropy is the *best* upper bound that depends only on the density matrix, we show that Q is the best lower bound that depends only on the density matrix. Other parallels between S and Q are also demonstrated.

I. INTRODUCTION

Suppose one is presented with a single quantum system which is known to be in one of several possible pure states, each having a certain *a priori* probability. For example, the system may be an electromagnetic pulse used in a quantum communication scheme, and the allowed states might be nonclassical states that are not necessarily orthogonal to each other. Given this *a priori* description, one wishes to perform a measurement on the system that will provide as much information about the state as possible, where "information" is defined in the sense of Shannon [1]. Although this problem has been solved for a few special cases [2,3], no simple and general method has yet been discovered for determining either an optimal measurement or the amount of information one can expect to gain. However, a 20-year-old theorem due to Kholevo does provide a general upper bound on the accessible information [4]. The theorem states that the amount of information one can extract from a quantum system is no greater than the von Neumann entropy S of the system. In the present paper we prove what might be thought of as the mirror image of Kholevo's theorem: the accessible information is *no less than* a certain quantity Q, which we call the "subentropy" of the system. As we will see, there are a number of parallels between the von Neumann entropy and the subentropy [5].

The present work belongs to the general field of quantum information theory, which is the study of information carried by quantum systems. The potential practical applications of quantum information theory include quantum cryptography, in which one intentionally uses nonorthogonal states to prevent eavesdropping [6], and quantum computation, in which the basic information storing and processing elements are objects that must be treated quantum mechanically, such as localized electrons in an array of quantum dots [7]. Aside from these potential applications, the study of information in quantum mechanics also has value for the foundations of physics. It has already contributed to our understanding of thermodynamic entropy [3, 4, 8], and it gives us new ways of thinking about quantum theory as well [9].

To define our problem precisely, let $\mathcal{E}$ be an ensemble of possible states, specified by a set of vectors $|\psi_1\rangle, \ldots, |\psi_m\rangle$ in an n-dimensional Hilbert space, together with a set of corresponding probabilities $p_1, \ldots, p_m$. The amount of information one initially lacks about the state is given by Shannon's entropy formula [1]

$$H(\mathcal{E}) = -\sum_{i=1}^{m} p_i \ln p_i. \tag{1}$$

$H(\mathcal{E})$ is called the mixing entropy [10] to distinguish it from the von Neumann entropy. Note that $H(\mathcal{E})$ depends only on the probabilities of the states $|\psi_i\rangle$ and not on the states themselves. We use the natural logarithm in our definition, so that the entropy given by Eq. (1) is measured in "nats," but for ease of interpretation we will sometimes quote specific values of entropy in bits, where 1 bit $= \ln 2$ nats.

If the states $|\psi_i\rangle$ are not all mutually orthogonal, then there is no measurement that can distinguish them from each other perfectly, but one can at least reduce the mixing entropy and in that sense gain information. A general quantum measurement is described by a probability-operator-valued measure (POM), which is a set of positive operators A_j, $j = 1, \ldots, r$, such that $\sum_{j=1}^{r} A_j = I$, where I is the identity [11]. Each operator A_j corresponds to a possible outcome of the measurement and

*Permanent address: Department of Mathematics and Statistics, University of Plymouth, Plymouth, Devon PL4 8AA, United Kingdom.

the probability of the jth outcome when the system being measured is in the state $|\psi\rangle$ is $\langle\psi|A_j|\psi\rangle$. (A special case is an orthogonal measurement, for which the operators A_j are orthogonal projection operators.) It is helpful to define the symbol p_{ij}, which represents the joint probability that the system is initially in the ith state *and* that the jth outcome of the measurement occurs. Thus $p_{ij} = p_i\langle\psi_i|A_j|\psi_i\rangle$. After one has performed the measurement and obtained the jth outcome, one's knowledge of the original state of the system has changed and is now described by the *a posteriori* probabilities $p_{i|j}$, computed via Bayes's formula:

$$p_{i|j} = p_{ij} \Big/ \sum_{i=1}^{m} p_{ij}. \tag{2}$$

Here $p_{i|j}$ is the probability that the system was in the ith state, given that the jth outcome of the measurement has occurred. Note that the system will typically change its state as a result of the measurement, but this is not our concern here. We are interested in inferring as well as possible the *original* state of the system.

The post-measurement mixing entropy is given again by Eq. (1), but with p_i replaced by $p_{i|j}$. This final entropy will typically depend on the outcome of the measurement. When choosing a measurement, one does not know in advance which outcome will occur, so to assess the potential usefulness of a measurement one averages the final entropy over all the possible outcomes. The average final entropy is

$$H(\mathcal{E}|A) = -\sum_{j=1}^{r} P_j \sum_{i=1}^{m} p_{i|j} \ln p_{i|j}, \tag{3}$$

where $P_j = \sum_{i=1}^{m} p_{ij}$ is the probability of getting the jth outcome. The *mutual information* between the measurement A and the ensemble $\mathcal{E}$ is defined as

$$I(\mathcal{E}:A) = H(\mathcal{E}) - H(\mathcal{E}|A). \tag{4}$$

That is, the mutual information is the average amount by which the mixing entropy is reduced as a result of the measurement. From Eqs. (1)–(4) we obtain the following more symmetric expression for $I(\mathcal{E}:A)$:

$$I(\mathcal{E}:A) = -\sum_i \left(\sum_j p_{ij}\right) \ln\left(\sum_j p_{ij}\right) - \sum_j \left(\sum_i p_{ij}\right) \ln\left(\sum_i p_{ij}\right) + \sum_{ij} p_{ij} \ln p_{ij}. \tag{5}$$

Some measurements are better than others at providing information, and it is interesting to ask how well one can do when one is free to use any measurement allowed by the laws of quantum mechanics, that is, any POM. Let $I(\mathcal{E})$ be the maximum value of $I(\mathcal{E}:A)$ over all POMs A. This maximum information, called the *accessible information* [12], is the quantity on which we focus our attention. In a communication context, the accessible information has a very practical meaning: it determines the number of signals necessary to transmit a long message faithfully. For example, if for each photon sent along an optical fiber the accessible information is 1/2 bit, then it will take approximately 200 photons to convey a message consisting of 100 random binary digits.

Kholevo's upper bound on the accessible information is based on the density matrix of the ensemble $\mathcal{E}$,

$$\rho = \sum_{i=1}^{m} p_i |\psi_i\rangle\langle\psi_i|. \tag{6}$$

As was mentioned above, the Kholevo bound is equal to the von Neumann entropy of ρ, which is defined as $S(\rho) = -\mathrm{tr}\,\rho\ln\rho$, or, in terms of the eigenvalues λ_k of ρ,

$$S(\rho) = -\sum_{k=1}^{n} \lambda_k \ln\lambda_k, \tag{7}$$

n again being the dimension of the state space [13].

Our new *lower* bound, the subentropy, is also a function only of the density matrix, and is defined by the equation

$$Q(\rho) = -\sum_{k=1}^{n} \left(\prod_{l\neq k} \frac{\lambda_k}{\lambda_k - \lambda_l}\right) \lambda_k \ln\lambda_k. \tag{8}$$

If two or more of the λ's are equal, one takes the limit as they become equal, and one finds that $Q(\rho)$ is finite for all ρ. The origin of this unusual formula will become clear in Sec. II.

It is important to note that a given density matrix is consistent with many different ensembles [14], and for some of these ensembles the accessible information will be greater than for others. For example, in a two-dimensional state space, the density matrix $\rho = I/2$ could represent the ensemble consisting of the two states $(1,0)$ and $(0,1)$ with equal probabilities, or it could represent the ensemble consisting of the three states $(1,0)$, $(1/2,\sqrt{3}/2)$, and $(1/2,-\sqrt{3}/2)$ with equal probabilities. In the former case, the accessible information is 1 bit, whereas in the latter case it is 0.585 bits (see the end of Sec. II). Thus the accessible information is not a function of the density matrix alone but depends on the specific ensemble.

It is therefore reasonable to ask, for a given density matrix ρ, how large or small the accessible information can get. The upper limit is the one given by Kholevo, $S(\rho)$. This limit is achieved by the "eigenensemble" of ρ, which consists of the eigenvectors of the density matrix with *a priori* probabilities λ_i. For this ensemble one can always determine the state perfectly (by making an orthogonal measurement along the eigenvectors) and thereby gain information $S(\rho)$. Thus the Kholevo bound is the best upper bound that depends only on the density matrix.

Similarly, we show in Sec. III that the smallest value the accessible information can have for a given density matrix ρ is $Q(\rho)$. For an arbitrary ρ, we construct explicitly an ensemble of states consistent with ρ for which the accessible information is exactly $Q(\rho)$. Thus Q is the best lower bound that depends only on the density matrix. This is one sense in which the subentropy is the natural complement of the von Neumann entropy.

Before presenting the proofs, it is interesting to give

an alternative pair of formulas for S and Q that emphasizes their relatedness. The von Neumann entropy can be written as the contour integral

$$S(\rho) = -\frac{1}{2\pi i}\oint (\ln z)\ \mathrm{tr}\,(I-\rho/z)^{-1}dz, \tag{9}$$

where I is the identity matrix and the contour encloses all the nonzero eigenvalues of ρ. It turns out, somewhat remarkably, that the subentropy is given by the same formula except that the trace is replaced by the determinant:

$$Q(\rho) = -\frac{1}{2\pi i}\oint (\ln z)\ \det\,(I-\rho/z)^{-1}dz. \tag{10}$$

It is not difficult to show that Eqs. (9) and (10) are equivalent to Eqs. (7) and (8) [noting that the eigenvalues of $(I-\rho/z)^{-1}$ are $z/(z-\lambda_k)$].

II. PROOF THAT Q IS A LOWER BOUND

To show that Q is a lower bound on the accessible information, we average the mutual information $I(\mathcal{E}:A)$ over all complete orthogonal measurements A. (A measurement is complete if every POM element A_j is proportional to a one-dimensional projection. For a complete *orthogonal* measurement these projections are associated with the elements of an orthonormal basis.) It will turn out that this average is equal to $Q(\rho)$, which will imply that there must be at least one measurement for which the mutual information is at least $Q(\rho)$, thus proving that Q is a lower bound on the accessible information.

In order to carry out the average, it is helpful to write the mutual information $I(\mathcal{E}:A)$ in another form. Equation (5) shows that $I(\mathcal{E}:A)$ is more symmetric in $\mathcal{E}$ and A than our original definition (4) suggests. In fact $I(\mathcal{E}:A)$ can also be written as

$$I(\mathcal{E}:A) = H(A) - H(A|\mathcal{E}), \tag{11}$$

where $H(A)$ is the *a priori* Shannon entropy of the outcome of the measurement A and $H(A|\mathcal{E})$ is the average Shannon entropy of the outcome given a knowledge of the system's state. For a complete orthogonal measurement, the *a priori* probability of the jth outcome is $\langle\alpha_j|\rho|\alpha_j\rangle$ and the probability of the jth outcome given that the system is in the state $|\psi_i\rangle$ is $|\langle\alpha_j|\psi_i\rangle|^2$, where $|\alpha_j\rangle$ is the jth eigenstate of the measurement. Thus the mutual information can be written as

$$\begin{aligned} I(\mathcal{E}:A) = &-\sum_j \langle\alpha_j|\rho|\alpha_j\rangle \ln\langle\alpha_j|\rho|\alpha_j\rangle \\ &+\sum_i p_i \sum_j |\langle\alpha_j|\psi_i\rangle|^2 \ln|\langle\alpha_j|\psi_i\rangle|^2. \end{aligned} \tag{12}$$

We now want to average this expression over all complete orthogonal measurements, that is, all orthogonal bases $\{|\alpha_j\rangle\}$. There is a natural measure on the set of such bases which is invariant under all unitary transformations, and this is the measure we use in our average.

Notice first that for each of the sums over j in Eq. (12), every term in the sum has the same average, because each basis element $|\alpha_j\rangle$ covers the same range—the unit sphere in $\mathcal{H}_n$—with the same uniform distribution. Thus we can replace each $|\alpha_j\rangle$ by a generic unit vector $|\alpha\rangle$. In order to find the average of the first sum—the one containing ρ—it is convenient to express $|\alpha\rangle$ in terms of its components a_k along the eigenvectors of ρ. The summand depends only on the squared magnitudes of the a's and not on their phases, so we can can convert the average over the unit sphere into an average over the $(n-1)$-dimensional probability simplex whose points are labeled by $(x_1,\ldots,x_n)$, where $x_k = |a_k|^2$ and $\sum_{k=1}^n x_k = 1$. Under this mapping, Sykora [15] has shown that the uniform measure on the unit sphere in $\mathcal{H}_n$ is taken to the uniform measure on the probability simplex, that is, the measure $k dx_1\cdots dx_{n-1}$, where k is a normalizing constant. In a similar way, we can also reduce the other sum in Eq. (12) to an integral over the probability simplex: each $|\psi_i\rangle$ can be replaced by a generic $|\psi\rangle$, and we can let x_1 be the squared magnitude of the component of $|\alpha\rangle$ along $|\psi\rangle$. We thereby find that the average of $I(\mathcal{E}:A)$ is

$$\begin{aligned} \langle I\rangle = &-n\int (\lambda_1 x_1+\cdots+\lambda_n x_n) \\ &\times \ln(\lambda_1 x_1+\cdots+\lambda_n x_n)\,dx + n\int x_1\ln x_1 dx. \end{aligned} \tag{13}$$

Each integral in this equation is over the $(n-1)$-dimensional probability simplex, with the uniform measure normalized so that $\int dx = 1$. As before, $\lambda_1,\ldots,\lambda_n$ are the eigenvalues of the density matrix.

The evaluation of the two integrals is discussed in Appendix A. One finds that the first term is equal to

$$-\sum_{k=1}^n \left((\lambda_k\ln\lambda_k)\prod_{l\neq k}\frac{\lambda_k}{\lambda_k-\lambda_l}\right)+\left(\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}\right) \tag{14}$$

and the second is

$$-\left(\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}\right). \tag{15}$$

Thus the mutual information averaged over all complete orthogonal measurements, being the sum of the expressions (14) and (15), is

$$\langle I\rangle = -\sum_{k=1}^n \left((\lambda_k\ln\lambda_k)\prod_{l\neq k}\frac{\lambda_k}{\lambda_k-\lambda_l}\right) = Q(\rho). \tag{16}$$

This result shows that Q is a lower bound on the accessible information. Note that there is no need to consider nonorthogonal POMs here; we have shown that there is an orthogonal measurement that provides an amount of information equaling or exceeding $Q(\rho)$, and this is sufficient.

As an example of the way Q functions as a lower bound, consider a single photon that could be in any of the following three linear polarization states with equal probability: vertical, 60° to the right of vertical, and 60° to the left of vertical. One can show that the optimal measurement in this case is a nonorthogonal POM with three outcomes, each of which is associated with a state orthogonal to one of the given states [16]. This measurement has the effect of eliminating one of the three possible

states and leaving the other two equally likely. Thus the accessible information is

$$I = H(\mathcal{E}) - H(\mathcal{E}|A) = \ln 3 - \ln 2 = 0.405 \text{ nats.} \quad (17)$$

The density matrix in this example has eigenvalues $\lambda_1 = \frac{1}{2}$ and $\lambda_2 = \frac{1}{2}$. For these eigenvalues, the upper and lower bounds S and Q take the values $S = \ln 2 = 0.693$ and $Q = \ln 2 - \frac{1}{2} = 0.193$. In this case the accessible information happens to be somewhat closer to the subentropy than to the von Neumann entropy, but in other cases it often happens that S is the closer bound.

III. ATTAINMENT OF Q

As was mentioned in the Introduction, for any density matrix ρ there is a special ensemble, the eigenensemble, for which the accessible information equals the Kholevo upper bound $S(\rho)$. We now show that the lower bound $Q(\rho)$ is similarly attained as the accessible information for a special ensemble called the "Scrooge ensemble." This is an ensemble for which *every* complete measurement yields the same mutual information $Q(\rho)$. The designation "Scrooge" derives from the fact that the ensemble is particularly stingy with its information.

The Scrooge ensemble will be defined in terms of the following construction called "ρ distortion." Suppose that we have an ensemble in an n-dimensional Hilbert space given by (normalized) states $|\psi_1\rangle, \ldots, |\psi_m\rangle$ taken with probabilities $p_1, \ldots, p_m$ and having density matrix $\frac{1}{n}I$:

$$\sum_{i=1}^{m} p_i |\psi_i\rangle\langle\psi_i| = \frac{1}{n} I.$$

It is often more convenient to give ensembles simply as a list of the corresponding states *unnormalized* so that the squared lengths are the probabilities. The above ensemble is then represented as $\{|\tilde{\psi}_1\rangle, \ldots, |\tilde{\psi}_m\rangle\}$ where

$$|\tilde{\psi}_i\rangle = \sqrt{p_i}\,|\psi_i\rangle,$$

so that $p_i = \langle\tilde{\psi}_i|\tilde{\psi}_i\rangle$ and

$$\sum_{i=1}^{m} |\tilde{\psi}_i\rangle\langle\tilde{\psi}_i| = \frac{1}{n} I. \quad (18)$$

Clearly the normalized states and probabilities can be easily recovered from the data $\{|\tilde{\psi}_1\rangle, \ldots, |\tilde{\psi}_m\rangle\}$.

Now for any other density matrix ρ let

$$|\tilde{\phi}_i\rangle = \sqrt{n\rho}\,|\tilde{\psi}_i\rangle.$$

Then $\{|\tilde{\phi}_1\rangle, \ldots, |\tilde{\phi}_m\rangle\}$ is an ensemble with density matrix ρ since multiplying each side of Eq. (18) on the left and right by the Hermitian operator $\sqrt{n\rho}$ gives

$$\sum_{i=1}^{m} |\tilde{\phi}_i\rangle\langle\tilde{\phi}_i| = \rho.$$

The ensemble $\{|\tilde{\phi}_1\rangle, \ldots, |\tilde{\phi}_m\rangle\}$ is said to be obtained from $\{|\tilde{\psi}_1\rangle, \ldots, |\tilde{\psi}_m\rangle\}$ by ρ distortion. The new probabilities p'_i for the $|\phi_i\rangle$ are

$$p'_i = \langle\tilde{\phi}_i|\tilde{\phi}_i\rangle = n\,p_i\,\langle\psi_i|\rho|\psi_i\rangle. \quad (19)$$

ρ distortion provides a one-to-one correspondence between the set of all ensembles with density matrix $\frac{1}{n}I$ and the set of all ensembles with density matrix ρ. (A generalization of this construction applying to any pair of density matrices is described in Ref. [14].)

ρ distortion can also be applied to continuous ensembles, which will be our main interest. Let x parametrize the state space and suppose that the distribution of (normalized) states $|x\rangle$ taken with probability density $p(x)$ has density matrix $\frac{1}{n}I$:

$$\int p(x)\,|x\rangle\langle x|\,dx = \frac{1}{n} I.$$

ρ distortion induces a map on normalized states

$$|x\rangle \mapsto |x'\rangle = \sqrt{n\rho}\,|x\rangle / \parallel \sqrt{n\rho}|x\rangle \parallel, \quad (20)$$

which defines x' as a function of x:

$$x' = f(x).$$

The new probability density $p'(x')$ is defined by an expression analogous to Eq. (19) but with p_i replaced by $p(x)dx$:

$$p'(x')dx' = n\,p(x)\,\langle x|\rho|x\rangle\,dx. \quad (21)$$

So explicitly we get

$$p'(x') = n\,p(x)\,\langle x|\rho|x\rangle\,Jac(x/x') \quad (22)$$

with $x = f^{-1}(x')$ on the right-hand side. Then the continuous distribution of states $|x'\rangle$ taken with probability density $p'(x')$ has density matrix ρ [as can be directly verified from Eqs. (20) and (21)].

We can now define the Scrooge ensemble (or "Scrooge distribution," since it is a continuous distribution). We start with the uniform distribution on the unit sphere in state space, that is, the unique distribution that is invariant under all unitary transformations. This uniform distribution has density matrix $\frac{1}{n}I$. The Scrooge distribution for any given density matrix ρ is simply the ρ distortion of the uniform distribution.

To represent continuous distributions explicitly we label (normalized) states by their components with respect to some fixed chosen orthonormal basis $|e_1\rangle, \ldots, |e_n\rangle$:

$$|\xi_1, \ldots, \xi_n\rangle = \xi_1|e_1\rangle + \cdots + \xi_n|e_n\rangle.$$

All state distributions that we consider here will be independent of the phases of the ξ's so that they may be represented as distributions over the $(n-1)$-dimensional probability simplex labeled by $(x_1, \ldots, x_n)$, where $x_i = |\xi_i|^2$. For example, the uniform distribution in state space is represented by the uniform distribution $k dx_1 \cdots dx_{n-1}$ on the probability simplex. Here k is the normalizing constant, which happens to have the value $(n-1)!$.

It is interesting in particular to get an explicit formula for the Scrooge distribution. For this purpose let us take our basis vectors $|e_1\rangle, \ldots, |e_n\rangle$ to be the eigenvectors of ρ. Then the Scrooge distribution can be written explicitly as follows, as is shown in Appendix B:

$$\frac{kn}{\lambda_1 \cdots \lambda_n} \frac{dx_1 \cdots dx_{n-1}}{(x_1/\lambda_1 + \cdots + x_n/\lambda_n)^{n+1}}. \quad (23)$$

As before, the numbers λ_k are the eigenvalues of ρ.

We now show that the amount of mutual information obtained by performing a complete measurement on the Scrooge ensemble is constant, independent of the choice of measurement. This constant must be $Q(\rho)$, since we have already seen that the average over all complete orthogonal measurements is $Q(\rho)$. By a theorem of Davies [2] the maximal information can always be obtained using a complete measurement, so we will get our desired result that the accessible information in the Scrooge ensemble is $Q(\rho)$.

To see that the mutual information for the Scrooge ensemble is independent of the choice of complete measurement, we write the equations first for the case of a discrete ensemble, which makes the argument more transparent, and indicate afterwards the modifications necessary for continuous ensembles (as Scrooge ensembles are always continuous).

Let $\mathcal{E}$ denote an ensemble $\{|\tilde{\phi}_1\rangle, \ldots, |\tilde{\phi}_m\rangle\}$, and let the POM A represent a complete measurement with r possible outcomes. The elements of A can then be written as

$$A_j = |\tilde{\alpha}_j\rangle\langle\tilde{\alpha}_j|, \tag{24}$$

where the $|\tilde{\alpha}_j\rangle$ are unnormalized vectors satisfying

$$\sum_{j=1}^{r} |\tilde{\alpha}_j\rangle\langle\tilde{\alpha}_j| = I.$$

If the measurement A is performed on $\mathcal{E}$ then the resulting probability matrix has entries

$$p_{ij} = \text{Prob}(|\phi_i\rangle \text{ and outcome } A_j) = |\langle\tilde{\alpha}_j|\tilde{\phi}_i\rangle|^2.$$

Now suppose that $\mathcal{E}$ is obtained by ρ distortion of an ensemble $\{|\tilde{\sigma}_1\rangle, \ldots, |\tilde{\sigma}_m\rangle\}$. Then

$$|\tilde{\phi}_i\rangle = \sqrt{n\rho}\,|\tilde{\sigma}_i\rangle, \qquad i = 1, \ldots, m$$

and

$$\langle\tilde{\alpha}_j|\tilde{\phi}_i\rangle = \langle\tilde{\alpha}_j|\sqrt{n\rho}|\tilde{\sigma}_i\rangle = \langle\tilde{\beta}_j|\tilde{\sigma}_i\rangle,$$

where we have set

$$|\tilde{\beta}_j\rangle = \sqrt{n\rho}|\tilde{\alpha}_j\rangle.$$

The mutual information is given by substituting these expressions into Eq. (5), which results in

$$I(\mathcal{E}:A) = -\sum_i |\langle\tilde{\phi}_i|\tilde{\phi}_i\rangle| \ln |\langle\tilde{\phi}_i|\tilde{\phi}_i\rangle| \tag{25}$$

$$-\sum_j \left(\sum_i |\langle\tilde{\beta}_j|\tilde{\sigma}_i\rangle|^2\right) \ln\left(\sum_i |\langle\tilde{\beta}_j|\tilde{\sigma}_i\rangle|^2\right) \tag{26}$$

$$+\sum_{ij} |\langle\tilde{\beta}_j|\tilde{\sigma}_i\rangle|^2 \ln(|\langle\tilde{\beta}_j|\tilde{\sigma}_i\rangle|^2). \tag{27}$$

The first term (25) is clearly independent of A. For the terms (26) and (27) we separate out the magnitude of each $|\tilde{\beta}_j\rangle$, writing

$$b_j^2 = \langle\tilde{\beta}_j|\tilde{\beta}_j\rangle \quad |\beta_j\rangle = |\tilde{\beta}_j\rangle/b_j.$$

Note that $\sum_j |\tilde{\beta}_j\rangle\langle\tilde{\beta}_j| = \rho n$, so

$$\sum_j b_j^2 = n.$$

Also the quantity

$$B = \sum_i |\langle\beta_j|\tilde{\sigma}_i\rangle|^2$$

will be independent of $|\beta_j\rangle$ if the $|\tilde{\sigma}_i\rangle$'s are exactly uniformly distributed in state space (as would be the case in the continuous limit if $\mathcal{E}$ were a Scrooge distribution). The term (26) of $I(\mathcal{E}:A)$ becomes

$$-B \ln B - B \sum_j b_j^2 \ln(b_j^2) \tag{28}$$

and the term (27) becomes

$$B\sum_j b_j^2 \ln(b_j^2) + \sum_j b_j^2 \sum_i |\langle\beta_j|\tilde{\sigma}_i\rangle|^2 \ln(|\langle\beta_j|\tilde{\sigma}_i\rangle|^2). \tag{29}$$

In the expression (29) the term $\sum_i |\langle\beta_j|\tilde{\sigma}_i\rangle|^2 \ln(|\langle\beta_j|\tilde{\sigma}_i\rangle|^2)$ will be independent of $|\beta_j\rangle$ if the $|\tilde{\sigma}_i\rangle$'s are exactly uniformly distributed. Hence the combination of (28) plus (29) will be independent of $|\beta_j\rangle$ and b_j, i.e., of the POM A.

The above argument is only approximate in that the *discrete* ensemble $\{|\tilde{\sigma}_1\rangle, \ldots, |\tilde{\sigma}_m\rangle\}$ cannot be exactly uniformly distributed over the whole state space (but can provide an arbitrarily good approximation). However, the same cancellations occur if the above calculation is carried out starting with a continuous ensemble $\mathcal{E}$ which is the ρ distortion of the exactly uniform distribution (i.e., if $\mathcal{E}$ is a Scrooge ensemble.) The formulas are given in Appendix C.

Thus for any given density matrix, $Q(\rho)$ is the amount of information one can obtain about the system's state in the worst case, when the ensemble is as stingy with information as it could possibly be (given that it is composed of *pure* states).

IV. PROPERTIES OF Q

We now compare and contrast some properties of $Q(\rho)$ and $S(\rho)$. If $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of ρ listed in decreasing order, then both S and Q have the form

$$-\sum_{i=1}^{n} c_i \ln \lambda_i,$$

where the coefficients c_i satisfy

$$\sum_{i=1}^{n} c_i = 1.$$

For S the coefficients are just the eigenvalues of ρ, whereas for Q

$$c_i = \lambda_i^n \Big/ \prod_{j\neq i} (\lambda_i - \lambda_j),$$

which alternate in sign and can become arbitrarily large in magnitude.

Since $Q(\rho)$ is a lower bound for accessible information it follows from Kholevo's theorem that

$$Q(\rho) \leq S(\rho) \qquad \text{for all } \rho.$$

Alternatively, we can avoid an appeal to Kholevo's weighty result by considering the eigenensemble of ρ. In this case $S(\rho)$ is clearly the maximum mutual information, since the ensemble is orthogonal, whereas $Q(\rho)$ is the average mutual information over complete orthogonal measurements and hence is less than or equal to $S(\rho)$.

For a fixed dimension n, S is a maximum when $\lambda_1 = \cdots = \lambda_n = 1/n$, in which case

$$S(1/n, \ldots, 1/n) = \ln n,$$

which is monotonically increasing with n and unbounded. We now show that for a fixed n, Q is also a maximum when $\lambda_1 = \cdots = \lambda_n = 1/n$. We begin with

$$Q(\lambda_1, \ldots, \lambda_n) = G(\lambda_1, \ldots, \lambda_n) - \left(\frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right), \tag{30}$$

where

$$G(\lambda_1, \ldots, \lambda_n) = -n \int (\lambda_1 x_1 + \cdots + \lambda_n x_n) \times \ln(\lambda_1 x_1 + \cdots + \lambda_n x_n)\, dx \tag{31}$$

and the integral is over the probability simplex [cf. Eq. (13)]. G is a strictly convex function of $(\lambda_1, \ldots, \lambda_n)$ and hence has a *unique* maximum on its convex domain (the probability simplex). Also G is symmetrical in its arguments so that the maximum can occur only where the λ's are all equal. Then Eqs. (30) and (31) give the maximum value of Q as

$$Q(1/n, \ldots, 1/n) = \ln n - \left(\frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right).$$

This is again monotonically increasing with n but now bounded above by

$$\lim_{n\to\infty} \left[\ln n - \left(\frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right)\right] = 1 - \gamma \simeq 0.422\,78,$$

where γ is Euler's constant. Thus for any ρ, $Q(\rho)$ never exceeds 0.422 78 nats (or 0.609 95 bits) whereas $S(\rho)$ may be arbitrarily large.

V. EXTENSION TO MIXED STATES

So far we have considered only the problem of trying to distinguish among a set of *pure* states. Imagine now a communication scenario in which each of the messages is represented by a *mixed* state. Such a scenario is actually more realistic than the one with pure states, because any pure-state signal will almost certainly be degraded to some extent during its passage through the channel.

Let the ensemble $\mathcal{E}$, then, consist of a set of mixed states $\rho_1, \ldots, \rho_m$, with corresponding probabilities $p_1, \ldots, p_m$. There is a general form of Kholevo's theorem that applies to this case—it is the form in which he actually stated the theorem—which places the following upper bound on the accessible information [4]:

$$I(\mathcal{E}) \leq S(\rho) - \sum_{i=1}^{m} p_i S(\rho_i). \tag{32}$$

Here $\rho = \sum_i p_i \rho_i$ is the density matrix of the whole ensemble. To put it in words, the accessible information is no greater than the von Neumann entropy of the whole ensemble minus the average von Neumann entropy of the individual signals.

We now prove the following analogous theorem which gives a *lower* bound on the accessible information for general ensembles:

$$I(\mathcal{E}) \geq Q(\rho) - \sum_{i=1}^{m} p_i Q(\rho_i). \tag{33}$$

As in Sec. II, we prove the theorem by averaging the mutual information $I(\mathcal{E}: A)$ over all complete orthogonal measurements A. The mutual information is

$$\begin{aligned} I(\mathcal{E}: A) &= H(A) - H(A|\mathcal{E}) \\ &= -\sum_j \langle\alpha_j|\rho|\alpha_j\rangle \ln\langle\alpha_j|\rho|\alpha_j\rangle \\ &\quad + \sum_i p_i \left(\sum_j \langle\alpha_j|\rho_i|\alpha_j\rangle \ln\langle\alpha_j|\rho_i|\alpha_j\rangle\right), \end{aligned} \tag{34}$$

where $|\alpha_1\rangle, \ldots, |\alpha_n\rangle$ are the eigenstates of the complete orthogonal measurement A. The first sum has appeared before, in Eq. (12), and we have already computed its average which has the value [cf. Eq. (14)]

$$Q(\rho) + \left(\frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right). \tag{35}$$

The remaining sums over j—one for each value of i—are all of the same form as the one we have just evaluated, and their averages are therefore also of the same form, with $Q(\rho)$ replaced by $Q(\rho_i)$. Thus the average mutual information is

$$\langle I\rangle = Q(\rho) - \sum_i p_i Q(\rho_i). \tag{36}$$

The right-hand side of Eq. (36) is therefore a lower bound on the accessible information, which is what we wanted to prove.

VI. DISCUSSION

Thus there are two natural entropic quantities that one can associate with a density matrix: the von Neumann entropy $S(\rho)$ and the subentropy $Q(\rho)$. These are the upper and lower limits on the amount of information one can extract from an ensemble of pure states with the given density matrix, and they also have significance for ensembles of mixed states, as we have just seen in Sec. V.

There is another, rather different, context in which

both S and Q appear, and to the best of our knowledge it is the only other context in which Q has been mentioned previously in the literature [17, 18]. Rather than trying to infer the initial state of a quantum system as we have been imagining here, suppose that one knows the initial state, which may be mixed or pure, and one is trying to predict the *outcome of a complete measurement* on the system. The complete measurement whose outcome can be predicted best, that is, whose outcome entropy $H(A)$ is least, is an orthogonal measurement along the eigenstates of the density matrix. For that particular measurement, the outcome entropy is $S(\rho)$. (If the state happens to be pure, then the outcome entropy for this measurement is zero.) On the other hand, suppose one does not consider this "most predictable" measurement but considers instead a complete orthogonal measurement chosen at random. The outcome entropy, averaged over all such measurements, is typically rather large if the state space is large—nearly $\ln n$—and it is slightly larger for a mixed state than for a pure state. The *amount* by which it is larger, for a given mixed state ρ, turns out to be $Q(\rho)$ [17, 18]. Thus both S and Q quantify the loss of predictability of the outcome of a measurement owing to the fact that the system in question may be in a mixed state rather than a pure state. S applies to the case where the measurement being made is the most predictable measurement, while Q applies to a randomly chosen measurement.

Evidently there are a number of parallels between S and Q, and this fact raises an interesting question. The von Neumann entropy S, in addition to having the information-theoretic interpretations mentioned above, has an additional and more famous significance: it is *the* entropy of thermodynamics. As such it plays a role in such practical matters as the determination of the amount of work that can be extracted from a thermodynamic system. One wonders whether Q, the subentropy, has a significance that is at all parallel to this meaning of S. One context where Q might conceivably have such a role—but this is only speculation—is in quantum chaos, since that is an area where the dynamics can effect a kind of average over the set of pure states [19]. Are there, for example, common situations in which the time-averaged density matrix of a "chaotic" quantum system maximizes the subentropy (subject to certain constraints), just as a system in thermodynamic equilibrium maximizes the von Neumann entropy? We offer no answer here but only raise the question as something that might be interesting to consider.

It is worth noting in this context that every thermodynamic system has a subentropy as well as a von Neumann entropy, but for ordinary objects the former is miniscule compared to the latter. For example, the von Neumann entropy of a glass of water at room temperature is around 10^{26} bits, whereas the subentropy of the same system is just over half a bit.

The formula (8) for Q may in fact be written down for *any* probability distribution, even in a classical setting. However, we have not been able to find an interpretation of this expression in classical information theory. It may be that Q is an essentially quantum mechanical quantity, like the correspondingly small nonzero ground state energy of a harmonic oscillator. Indeed, the problem we have addressed in this paper appears to have no interesting classical analog, in that it depends in an essential way on two features peculiar to quantum mechanics: (i) the existence of incompatible measurements and (ii) the fact that even a complete measurement generally yields only partial information about the identity of the input state. In contrast, the state of a *classical* system can in principle be exactly determined by measurement, and thus the classical analog of the Scrooge distribution—if one thinks of it as a distribution on probability space—is simply a δ-function from which no information could be extracted. The "classical subentropy" is in this sense always zero.

A point we particularly want to emphasize is that the von Neumann entropy is not the only interesting entropy-like quantity that one can associate with a quantum system in an unknown state. The subentropy of a system, small though it may be by macroscopic standards, plays a role in quantum information theory that is quite parallel to that of the von Neumann entropy. It remains to be seen what significance, if any, the subentropy has for the rest of physics.

ACKNOWLEDGMENTS

We are grateful for the opportunity for collaboration provided by the Quantum Computation Workshop, conducted with the support of ELSAG-Bailey, Genova. We would also like to thank Ben Schumacher for helpful discussions, and David Park for suggesting the integration method used in Appendix A.

APPENDIX A: EVALUATING THE AVERAGE MUTUAL INFORMATION

The integrals in Eq. (13) have been evaluated by Jones using Riemann-Liouville fractional integration [18]. Here we present an alternative method that may be valuable in its own right.

The first term in (13), which we have called G, can be written more fully as follows:

$$G = -n! \int_0^1 \cdots \int_0^1 [s(x) \ln s(x)]\, \delta\left(1 - \sum_i x_i\right) \times dx_1 \cdots dx_n \tag{A1}$$

where $s(x) = \lambda_1 x_1 + \cdots + \lambda_n x_n$. The factor $n!$ comes from combining the n that appears in Eq. (13) with the factor $(n-1)!$ that normalizes the measure:

$$(n-1)! \int_0^1 \cdots \int_0^1 \delta\left(1 - \sum_i x_i\right) dx_1 \cdots dx_n = 1.$$

Our strategy for evaluating the integral G is first to replace the quantity $s(x) \ln s(x)$ by $s(x)^t$. We will later take the derivative with respect to t and evaluate it at $t = 1$ to recover the original integral.

Thus we need to evaluate the integral

$$F(t) \equiv \int_0^1 \cdots \int_0^1 s(x)^t \, \delta\left(1 - \sum_i x_i\right) dx_1 \cdots dx_n. \tag{A2}$$

We now make the following two substitutions:

$$\delta\left(1 - \sum_i x_i\right) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left[i\alpha\left(1 - \sum_i x_i\right)\right] d\alpha\,; \tag{A3}$$

$$s(x)^t = \frac{1}{\Gamma(-t)} \int_0^{\infty} e^{-us(x)} \frac{du}{u^{t+1}}\,. \tag{A4}$$

With these substitutions, the integrals over the x_i's factorize and can thus be done separately. Once those integrals are done, one is left with the following integral over α and u:

$$F(t) = \frac{1}{2\pi\Gamma(-t)} \int_0^{\infty} \left\{ \int_{-\infty}^{\infty} \left[\prod_{i=1}^{n} \left(\frac{1 - e^{-i\alpha - u\lambda_i}}{i\alpha + u\lambda_i} \right) \right] e^{i\alpha} \, d\alpha \right\} \frac{du}{u^{t+1}}. \tag{A5}$$

Consider now the numerator of the product in square brackets, that is,

$$\prod_{i=1}^{n} (1 - e^{-i\alpha - u\lambda_i}),$$

and imagine expanding it in a power series in the quantity $e^{-i\alpha}$. For the first term in this series, the integral over α can be made into a contour integral by closing it in the upper half plane. For all the other terms in the series, the contour can be closed in the lower half plane. Now the integrand has poles only at $\alpha = iu\lambda_i$, which lie in the upper half plane, so the only term in the series that gives a nonzero contribution to the integral is the first one. In other words, the numerator in the product can be replaced by the number 1. We now perform the contour integral over α, under the assumption that the λ's are all distinct. [When two or more of the λ's are equal, one can evaluate G by taking the limit as they become equal. This follows from the continuity of the integrand in Eq. (A1) and the fact that the integral itself is a bounded function of the λ's.] The result of the contour integration is

$$F(t) = \frac{1}{\Gamma(-t)} \int_0^{\infty} \sum_i \left[e^{-u\lambda_i} \prod_{j \neq i} \frac{1}{\lambda_j - \lambda_i} \right] \frac{du}{u^{t+n}}\,. \tag{A6}$$

We can do the integral over u to obtain

$$F(t) = \frac{t!}{(t+n-1)!} \sum_i \left[\lambda_i^{t+n-1} \prod_{j \neq i} \frac{1}{\lambda_i - \lambda_j} \right]. \tag{A7}$$

It is now a straightforward matter to take the derivative of $F(t)$ with respect to t and to set t equal to 1. We thus obtain

$$\begin{aligned} G &= -n! \left[\frac{dF(t)}{dt} \right]_{t=1} \\ &= -\left(\frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right) \sum_i \left[\lambda_i \prod_{j \neq i} \frac{\lambda_i}{\lambda_i - \lambda_j} \right] \\ &\quad + \sum_i \left[\lambda_i \ln \lambda_i \prod_{j \neq i} \frac{\lambda_i}{\lambda_i - \lambda_j} \right]. \end{aligned} \tag{A8}$$

Finally, we note that the first sum over i appearing in Eq. (A8) has the value unity, so that we have obtained the expression given in Eq. (14).

There is yet another way in which we could have evaluated the integral G, based on a theorem used in numerical analysis [20]. If $f(\lambda)$ is a function whose value is known at only n points $\lambda_1, \ldots, \lambda_n$, then there is a unique polynomial of degree $n-1$, the Lagrange interpolating polynomial, that agrees with the function at those points. The coefficient of λ^{n-1} in this polynomial is called the Newton divided difference and one can show that this quantity is equal to the integral over the probability simplex of $f^{(n-1)}(\lambda_1 x_1 + \cdots + \lambda_n x_n)$, where $f^{(n-1)}$ is the $(n-1)$th derivative of f. (This is the Hermite-Gennochi theorem.) If we take $f(\lambda)$ to be $-\lambda^n \ln \lambda$, then this integral is essentially the one we wanted. One can then use the standard explicit form of the Lagrange polynomial in terms of λ_i and $f(\lambda_i)$ to arrive immediately at Eq. (A8).

The second term in Eq. (13) is

$$n! \int_0^1 \cdots \int_0^1 [x_1 \ln x_1] \, \delta\left(1 - \sum_i x_i\right) dx_1 \cdots dx_n. \tag{A9}$$

This integral can be evaluated by elementary methods (e.g., by parts), and one obtains the result given in Eq. (15). Alternatively, once one has obtained the formula for G, one can take the limit of this formula as all but one of the λ's approach zero.

APPENDIX B: DERIVING THE SCROOGE DISTRIBUTION FORMULA

For any density matrix ρ with eigenvalues $\lambda_1, \ldots, \lambda_n$ let $(x_1, \ldots, x_n)$ be the coordinates on the probability simplex induced by an orthonormal basis of eigenstates of ρ in the state space. We use $(x_1, \ldots, x_{n-1})$ as independent variables setting $x_n = 1 - x_1 - \cdots - x_{n-1}$. The ρ-distortion map on normalized states is given by

$$(\sqrt{x_1}, \ldots, \sqrt{x_n}) \mapsto (\sqrt{\lambda_1 x_1}, \ldots, \sqrt{\lambda_n x_n}) / \sqrt{\lambda_1 x_1 + \cdots + \lambda_n x_n},$$

inducing a map on the probability simplex

$$x_i' = \frac{\lambda_i x_i}{\lambda_1 x_1 + \cdots \lambda_n x_n}\,, \qquad i = 1, \ldots, n.$$

The inverse map is

$$x_i = \frac{x'_i/\lambda_i}{(x'_1/\lambda_1 + \cdots + x'_n/\lambda_n)}.$$

To evaluate the Jacobian $\mathcal{J}(x_1, \ldots, x_{n-1}/x'_1, \ldots, x'_{n-1})$ introduce

$$A = (x'_1/\lambda_1 + \cdots + x'_n/\lambda_n), \qquad a_i = 1/\lambda_i - 1/\lambda_n.$$

Then direct calculation of the partial derivatives gives

$$\mathcal{J}(x/x') = \frac{1}{\lambda_1 \cdots \lambda_{n-1} A^{2n-2}} \begin{vmatrix} A - a_1 x'_1 & -a_2 x'_1 & \cdots & -a_{n-1} x'_1 \\ -a_1 x'_2 & A - a_2 x'_2 & \cdots & -a_{n-1} x'_2 \\ \vdots & \vdots & \vdots & \vdots \\ -a_1 x'_{n-1} & -a_2 x'_{n-1} & \cdots & A - a_{n-1} x'_{n-1} \end{vmatrix}.$$

The remaining determinant can be evaluated by factoring $x'_1, \ldots, x'_{n-1}$ out of the rows and multiplying them back into the columns. Then subtracting the last row from each of rows 1 to $n-2$ followed by addition of each column to the last column results in the form $\mathrm{diag}(A, \ldots, A, 1/\lambda_n)$. Hence $\mathcal{J}(x/x') = 1/(\lambda_1 \cdots \lambda_n A^n)$. Substituting into Eq. (22) with $p(x) = k$ and

$$\langle x|\rho|x\rangle = \lambda_1 x_1 + \cdots + \lambda_n x_n = 1/(x'_1/\lambda_1 + \cdots + x'_n/\lambda_n)$$

gives the stated formula (23) for the Scrooge distribution.

APPENDIX C: PROOF OF THE A-INDEPENDENCE OF $I(\mathcal{E}: A)$ FOR THE SCROOGE DISTRIBUTION

Suppose that $\mathcal{E}$ is the continuously distributed ensemble of states $|x'\rangle$ with probability density $p'(x')dx'$. Then for the POM (24) the mutual information is

$$\begin{aligned} I(\mathcal{E}: A) = & -\int p' \ln p' \, dx' \\ & - \sum_j \left(\int |\langle \alpha_j | x' \rangle|^2 \, p' \, dx' \right) \\ & \times \ln \left(\int |\langle \alpha_j | x' \rangle|^2 \, p' \, dx' \right) \\ & + \sum_j \int |\langle \alpha_j | x' \rangle|^2 p' \, \ln \left(|\langle \alpha_j | x' \rangle|^2 p' \right) dx' . \end{aligned}$$

Suppose now that $\mathcal{E}$ is a Scrooge ensemble obtained from the uniform distribution by ρ distortion. Let x represent the coordinates $(x_1, \ldots, x_{n-1})$ on the probability simplex. Then from Eqs. (20) and (21) we have

$$|x'\rangle \, p'(x') \, \langle x'| \, dx' = \sqrt{n\rho}|x\rangle \, p(x) \, \langle x|\sqrt{n\rho} \, dx$$

where $p(x)dx = k \, dx_1 \cdots dx_{n-1}$ is the uniform distribution. With $|\tilde{\beta}_j\rangle, |\beta_j\rangle, b_j$ as before,

$$I(\mathcal{E}: A) = -\int p' \ln p' \, dx' \tag{C1}$$

$$- \sum_j \left(\int |\langle \tilde{\beta}_j | x \rangle|^2 dx \right) \ln \left(\int |\langle \tilde{\beta}_j | x \rangle|^2 \, dx \right) \tag{C2}$$

$$+ \sum_j \int |\langle \tilde{\beta}_j | x \rangle|^2 \ln \left[\mathcal{J}(x/x') \, |\langle \tilde{\beta}_j | x \rangle|^2 \right] dx . \tag{C3}$$

Writing

$$B = \int |\langle \beta_j | x \rangle|^2 \, dx ,$$

which is now exactly independent of $|\beta_j\rangle$, the terms (C2) and (C3) become

$$-B \ln B - B \sum_j b_j^2 \ln(b_j^2)$$

and

$$B \sum_j b_j^2 \ln(b_j^2) + \sum_j b_j^2 \int |\langle \beta_j | x \rangle|^2 \ln[\mathcal{J}(x/x') \, |\langle \beta_j | x \rangle|^2] \, dx .$$

The integral in the final term is independent of $|\beta_j\rangle$ so that altogether $I(\mathcal{E}: A)$ is independent of the POM A.

[1] C. E. Shannon, Bell Syst. Tech. J. **27**, 379 623 (1948).

[2] E. B. Davies, IEEE Trans. Inform. Theory **IT24**, 596 (1978).

[3] L. B. Levitin, in *Information, Complexity, and Control in Quantum Physics*, edited by A. Blaquière, S. Diner, and G. Lochak (Springer, Vienna, 1987).

[4] A. S. Kholevo, Probl. Inf. Transm. (USSR) **9**, 177 (1973) [Probl. Peredachi Inf. **9** (3), (1973)]; H. P. Yuen and M. Ozawa, Phys. Rev. Lett. **70**, 363 (1993).

[5] Some of the results presented in this paper were conjectured but not proved in W. K. Wootters, *Proceedings of the Workshop on the Physics of Computation: PhysComp '92* (IEEE Press, Dallas, 1993).

[6] S. Wiesner, Sigact News **15:1**, 78 (1983); C. H. Bennett and G. Brassard, *Proceedings of IEEE International Conference on Computers, Systems, and Signal Processing, Bangalore, India* (IEEE, New York, 1984), pp. 175–179; A. K. Ekert, Phys. Rev. Lett. **67**, 661 (1991); C. H. Bennett, G. Brassard, and N. D. Mermin, *ibid.* **68**, 557 (1992); C. H. Bennett, *ibid.* **68**, 3121 (1992); A. K. Ekert, J. G. Rarity, P. R. Tapster, and G. M. Palma, *ibid.*, **69**, 1293 (1992); C. H. Bennett, G. Brassard,

C. Crépeau, and M.-H. Skubiszewska, in *Advances in Cryptology-Crypto '91 Proceedings* (Springer, New York, 1992), p. 351; G. Brassard and C. Crépeau, in *Advances in Cryptology-Crypto '90 Proceedings* (Springer, New York, 1991), p. 49.

[7] D. Deutsch, Proc. Roy. Soc. Lond. **A 400**, 97 (1985); D. Deutsch and R. Jozsa, *ibid.* **439**, 553 (1992); A. Berthiaume and G. Brassard, *Proc. 7th Annual IEEE Conf. on Structure in Complexity Theory, Boston, 1992* (IEEE, New York, 1993), pp. 132–137; E. Bernstein and U. Vazirani, in *Proceedings of the 25th Annual ACM Symposium on the Theory of Computation, San Diego, 1993* (ACM, New York, 1993), pp. 11–20; S. Lloyd, Science **261**, 1569 (1993).

[8] B. Schumacher, Phys. Rev. A (to be published); P. Hausladen, B.A. thesis, Williams College, 1993.

[9] D. Deutsch, Phys. Rev. Lett. **50**, 631 (1983); H. Maassen and J. B. M. Uffink, *ibid.* **60**, 1103 (1988); S. L. Braunstein and C. M. Caves, *ibid.* **61**, 662 (1988); B. Schumacher, Phys. Rev. A **44**, 7047 (1991).

[10] A. Wehrl, Rev. Mod. Phys. **50**, 221 (1978).

[11] A. Peres, *Quantum Theory: Concepts and Methods* (Kluwer Academic, Dordrecht, 1993), Chap. 9; A. Peres, Found. Phys. **20**, 1441 (1990); C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976), pp. 74–83.

[12] B. Schumacher, in *Complexity, Entropy, and the Physics of Information*, edited by W. H. Zurek (Addison-Wesley, Redwood City, CA, 1990).

[13] J. von Neumann, *Mathematische Grundlagen der Quantenmechanik* (Springer, Berlin, 1932), [translated by E. T. Beyer *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, 1955)].

[14] This example was given in A. S. Kholevo, Probl. Inf. Trans. **9**, 110 (1973) [Probl. Peredachi Inf. **9** (2), 31 (1973)]. The determination of the optimal measurement is based on the methods of Davies [2].

[15] S. Sykora, J. Stat. Phys. **11**, 17 (1974).

[16] L. P. Hughston, R. Jozsa, and W. K. Wootters, Phys. Lett. A **183**, 14 (1993).

[17] W. K. Wootters, Found. Phys. **20**, 1365 (1990).

[18] K. R. W. Jones, J. Phys. A **24**, 1237 (1991).

[19] R. Schack and C. M. Caves, Phys. Rev. Lett. **71**, 525 (1993).

[20] K. E. Atkinson, *An Introduction to Numerical Analysis* (Wiley, New York, 1978), pp. 107–123.