PHYSICAL REVIEW
LETTERS

VOLUME 73 5 DECEMBER 1994 NUMBER 23

Ensemble-Dependent Bounds for Accessible Information in Quantum Mechanics

Christopher A. Fuchs and Carlton M. Caves
Center for Advanced Studies, Department of Physics and Astronomy, University of New Mexico, Albuquerque, New Mexico 87131-1156
(Received 3 August 1994)

We simplify the derivation of the Holevo upper bound on the maximum information extractable from a quantum communication channel. This simplification leads to upper and lower bounds for binary channels, both of which depend explicitly on the message ensemble.

PACS numbers: 03.65.Bz, 02.50.–r, 89.70.+c

A quantum communication channel is defined by the action of sending one of n possible messages, with prior probabilities $p_1,\ldots,p_n$, to a specified receiver in the form of one of n distinct (possibly mixed) density operators $\hat{\rho}_1,\ldots,\hat{\rho}_n$ on an N-dimensional Hilbert space; the message states $\hat{\rho}_i$, together with their probabilities p_i, constitute the *message ensemble.* The receiver can perform any *generalized* quantum measurement, described by a positive-operator-valued measure (POVM), in an attempt to discern which message was sent. The fundamental question of quantum communication theory is this [1]: Which measurements maximize the Shannon mutual information about the actual message, and just how much information is that maximal amount $I_{\rm acc}$? Previous results on $I_{\rm acc}$ include an upper bound due to Holevo [2] and a lower bound recently exhibited by Jozsa, Robb, and Wootters [3]. Both are the best bounds that can be expressed solely in terms of the total density operator $\hat{p}=\sum p_i\hat{\rho}_i$, whenever all the $\hat{\rho}_i$ are pure states, but for just this reason both bounds are fairly loose for many message ensembles. We simplify Holevo's original derivation, making it accessible to most physicists, and along the way derive upper and lower bounds for binary channels, both of which *always* depend explicitly on the message ensemble.

Since Holevo's original derivation, various improved versions have appeared [4,5], but the improvement has been directed at proving the upper bound for more general situations by removing finiteness assumptions on the Hilbert-space dimensionality, the number of messages, and the number of measurement outcomes. In contrast, we retain the finiteness assumptions, our aim being a deeper understanding through a more accessible derivation of the bound. Specific payoffs of this approach are the ensemble-dependent bounds reported here.

The quantum communication problem [1] is formalized with the help of the most general quantum mechanical measurements, described by so-called POVMs [6]. A POVM is a set of non-negative, Hermitian operators $\hat{E}_b$ such that $\sum_b \hat{E}_b=\hat{\mathbb{1}}=$ (N-dimensional unit operator). The subscript b indexes the outcomes of the measurement. The Shannon mutual information [7] with respect to a measurement $\{\hat{E}_b\}$ is defined by

$$I=H(\hat{\rho})-\sum_{i=1}^{n}p_iH(\hat{\rho}_i), \tag{1}$$

where $H(\hat{\rho})=-\sum_b \mathrm{tr}(\hat{\rho}\hat{E}_b)\ln[\mathrm{tr}(\hat{\rho}\hat{E}_b)]$ is the average information gain upon finding outcome b, when the density operator is assumed to be $\hat{\rho}$.

The *accessible information* $I_{\rm acc}$ is defined to be the maximum of the mutual information I over all measurements $\{\hat{E}_b\}$. The Holevo upper bound to $I_{\rm acc}$ is

$$I_{\rm acc}\le S(\hat{\rho})-\sum_{i=1}^{n}p_iS(\hat{\rho}_i), \tag{2}$$

where $S(\hat{\rho})=-\mathrm{tr}(\hat{\rho}\ln\hat{\rho})=-\sum\lambda_j\ln\lambda_j$ is the von Neumann entropy of the density operator $\hat{\rho}$, whose eigenvalues are λ_j. Taking $0\ln 0=0$, one sees that $S(\hat{\rho}_i)=0$ whenever $\hat{\rho}_i$ is a pure state; thus, when *all* the input states $\hat{\rho}_i$ are pure, the upper bound reduces to $S(\hat{\rho})$ and does not depend explicitly on the message ensemble.

0031-9007/94/73(23)/3047(4)$06.00 © 1994 The American Physical Society

Holevo's derivation of inequality (2) can be summarized as follows. Trivially the mutual information (1) can be written as a sum of binary-channel mutual informations,

$$I = \sum_{k=2}^{n} s_k \left(H(\hat{\tau}_k) - \frac{p_k}{s_k} H(\hat{\rho}_k) - \frac{s_{k-1}}{s_k} H(\hat{\tau}_{k-1}) \right), \quad (3)$$

where $s_k = \sum_{i=1}^{k} p_i$ and $\hat{\tau}_k = s_k^{-1} \sum_{i=1}^{k} p_i \hat{\rho}_i$. (All the terms in this sum containing a $\hat{\tau}_k$ cancel, except the contributions from $\hat{\tau}_n = \hat{\rho}$ and $\hat{\tau}_1 = \hat{\rho}_1$.) Thus, if one can find an upper bound for the mutual information of a *binary* channel, one can immediately build an upper bound for the general case. The bound (2) can, in fact, be built in just this way, thereby allowing the derivation to focus on the binary case.

For a binary channel specified by density operators $\hat{\rho}_0$ and $\hat{\rho}_1$ with probabilities $1 - t$ and t, the mutual information $I \equiv I(t)$ can be considered a function of the parameter t. The Holevo bound takes the form $I_{\rm acc}(t) \le S(\hat{\rho}) - (1 - t)S(\hat{\rho}_0) - tS(\hat{\rho}_1) \equiv S(t)$, where $\hat{\rho} = (1 - t)\hat{\rho}_0 + t\hat{\rho}_1 = \hat{\rho}_0 + t\hat{\Delta} = \hat{\rho}_1 - (1 - t)\hat{\Delta}$ and $\hat{\Delta} = \hat{\rho}_1 - \hat{\rho}_0$. The derivation relies on properties of $I(t)$ and $S(t)$ as functions of t. Note first that, trivially, $I(0) = I(1) = S(0) = S(1) = 0$. Moreover, both $I(t)$ and $S(t)$ are downwardly convex functions of t, as can be seen from their second derivatives. For $I(t)$, that is,

$$I''(t) = -\sum_b \frac{[\mathrm{tr}(\hat{\Delta}\hat{E}_b)]^2}{\mathrm{tr}(\hat{\rho}\hat{E}_b)}. \quad (4)$$

The expression (4) for $-I''(t)$ is well known as the Fisher information [8], a quantity of use in parameter estimation. The second derivative of $S(t)$ is most easily found by representing $S(\hat{\rho})$ as a contour integral [9],

$$S(\hat{\rho}) = -\frac{1}{2\pi i} \oint z \ln z \, \mathrm{tr}[(z\hat{1} - \hat{\rho})^{-1}] \, dz, \quad (5)$$

where the contour encloses the poles at all the nonzero eigenvalues of $\hat{\rho}$, but does not enclose $z = 0$. Differentiating within the integral and using the operator identity $(\hat{A}^{-1})' = -\hat{A}^{-1}\hat{A}'\hat{A}^{-1}$, one finds that

$$S''(t) = -\sum_{\{j,k|\lambda_j+\lambda_k\neq 0\}} \Phi(\lambda_j, \lambda_k) \, |\Delta_{jk}|^2, \quad (6)$$

where $\Phi(x, y) = (\ln x - \ln y)/(x - y)$ if $x \neq y$, $\Phi(x, x) = 1/x$, $\Delta_{jk} = \langle j|\hat{\Delta}|k\rangle$, and $|j\rangle$ is the eigenvector of $\hat{\rho}$ with eigenvalue λ_j. Expressions (4) and (6) are both nonpositive.

The key to deriving the Holevo bound is the following: $S(t)$ is an upper bound to $I_{\rm acc}(t)$ for any t if and only if, when plotted versus t, the curve for $S(t)$ has a more negative curvature than the curve for $I(t)$, regardless of which measurement is used, i.e.,

$$S''(t) \le I''(t) \le 0 \quad \text{for any POVM } \{\hat{E}_b\}. \quad (7)$$

The meat of the derivation is in showing this inequality. Holevo does this by demonstrating the existence of a function $L''(t)$, independent of $\{\hat{E}_b\}$, such that $S''(t) \le L''(t) \le I''(t)$ for any POVM $\{\hat{E}_b\}$. Upon enforcing the boundary conditions $L(0) = L(1) = 0$, it follows that $I_{\rm acc}(t) \le L(t) \le S(t)$ [though in [2] $L(t)$ is not computed explicitly].

At this juncture quite a dramatic simplification can be made to the original proof. The easiest way to get at such a function $L''(t)$ is to minimize $I''(t)$ over all POVMs and thereafter to show that $S''(t) \le L''(t)$. [This is more tractable than maximizing the mutual information itself because no logarithms appear in $I''(t)$.] This approach generates the same function $L''(t)$ as used by Holevo, though the two derivations appear to have little to do with each other. In addition to simplication, the advantage of this approach is that it pinpoints a measurement that minimizes $I''(t)$, thus revealing the significance of the upper bound $L(t)$; moreover, this measurement, though it generally does not maximize $I(t)$, nonetheless provides a *lower* bound $M(t)$ to the accessible information $I_{\rm acc}(t)$ in the binary-channel case.

An efficient way to minimize $I''(t)$ is through a clever use of the Schwarz inequality, as in the formally identical problem considered in [10]. The steps are as follows. The operator inner product $\mathrm{tr}(\hat{A}^\dagger \hat{B})$ obeys a Schwarz inequality $|\mathrm{tr}(\hat{A}^\dagger \hat{B})|^2 \le \mathrm{tr}(\hat{A}^\dagger \hat{A})\mathrm{tr}(\hat{B}^\dagger \hat{B})$, where equality holds if and only if $\hat{A} = \mu \hat{B}$ for some constant μ. The idea is to think of the numerator within the sum (4) as the analog of the left-hand side of the Schwarz inequality and to use that inequality in such a way that the $\mathrm{tr}(\hat{\rho}\hat{E}_b)$ term in the denominator is cancelled. This leaves an expression linear in $\hat{E}_b$; summing over the index b then eliminates the dependence on the measurement. This requires introduction of the "lowering" superoperator $\mathcal{L}_{\hat{\rho}}$, whose action on the operator $\hat{A}$ is defined by

$$\tfrac{1}{2}[\hat{\rho}\,\mathcal{L}_{\hat{\rho}}(\hat{A}) + \mathcal{L}_{\hat{\rho}}(\hat{A})\hat{\rho}] = \hat{A}. \quad (8)$$

In a basis $|j\rangle$ that diagonalizes $\hat{\rho}$, $\mathcal{L}_{\hat{\rho}}(\hat{\Delta})$ becomes

$$\mathcal{L}_{\hat{\rho}}(\hat{\Delta}) = \sum_{\{j,k|\lambda_j+\lambda_k\neq 0\}} \frac{2}{\lambda_j + \lambda_k} \Delta_{jk} |j\rangle\langle k|, \quad (9)$$

which depends on the fact that $\Delta_{jk} = 0$ if $\lambda_j + \lambda_k = 0$. [For further discussion of why Eq. (9) is a suitable extension of $\mathcal{L}_{\hat{\rho}}(\hat{A})$ to the zero-eigenvalue subspaces of $\hat{\rho}$, see [10]; there $\mathcal{L}_{\hat{\rho}}$ is denoted by $\mathcal{R}_{\hat{\rho}}^{-1}$.] Using Eq. (8), one easily derives the identity that for Hermitian $\hat{A}$,

$$\mathrm{tr}(\hat{A}\hat{\Delta}) = \mathrm{Re}\{\mathrm{tr}[\hat{\rho}\hat{A}\mathcal{L}_{\hat{\rho}}(\hat{\Delta})]\}. \quad (10)$$

The desired optimization now follows in short order:

$$I''(t) = -\sum_b (\mathrm{Re}\{\mathrm{tr}[\hat{\rho}\hat{E}_b \mathcal{L}_{\hat{\rho}}(\hat{\Delta})]\})^2 [\mathrm{tr}(\hat{\rho}\hat{E}_b)]^{-1}$$
$$\ge -\sum_b |\mathrm{tr}[\hat{\rho}\hat{E}_b \mathcal{L}_{\hat{\rho}}(\hat{\Delta})]^2 [\mathrm{tr}(\hat{\rho}\hat{E}_b)]^{-1} \quad (A)$$

$$= -\sum_b \frac{|\mathrm{tr}\{(\sqrt{\hat{E}_b}\sqrt{\hat{\rho}})^\dagger [\sqrt{\hat{E}_b}\mathcal{L}_{\hat{\rho}}(\hat{\Delta})\sqrt{\hat{\rho}}]\}|^2}{\mathrm{tr}(\hat{\rho}\hat{E}_b)}$$

$$\geq -\sum_b \mathrm{tr}[\hat{E}_b \mathcal{L}_{\hat{\rho}}(\hat{\Delta})\hat{\rho}\,\mathcal{L}_{\hat{\rho}}(\hat{\Delta})] \quad \text{(B)}$$

$$= -\mathrm{tr}[\mathcal{L}_{\hat{\rho}}(\hat{\Delta})\hat{\rho}\,\mathcal{L}_{\hat{\rho}}(\hat{\Delta})] = -\mathrm{tr}[\hat{\Delta}\mathcal{L}_{\hat{\rho}}(\hat{\Delta})]. \quad (11)$$

The conditions for equality in Eq. (11)—i.e., for achieving the lower bound—arise from steps (A) and (B): $\mathrm{Im}\{\mathrm{tr}[\hat{\rho}\hat{E}_b\mathcal{L}_{\hat{\rho}}(\hat{\Delta})]\} = 0$ for all b and

$$\sqrt{\hat{E}_b}[\hat{\mathbb{1}} - \mu_b \mathcal{L}_{\hat{\rho}}(\hat{\Delta})]\sqrt{\hat{\rho}} = 0 \quad \text{for all } b. \quad (12)$$

These conditions can be met [10] by choosing the operators $\hat{E}_b$ to be one-dimensional projectors onto a basis that diagonalizes $\mathcal{L}_{\hat{\rho}}(\hat{\Delta})$ and choosing the constants μ_b to be the inverse eigenvalues of $\mathcal{L}_{\hat{\rho}}(\hat{\Delta})$.

The function $L''(t)$ can now be defined as

$$L''(t) = -\mathrm{tr}[\hat{\Delta}\mathcal{L}_{\hat{\rho}}(\hat{\Delta})] = -\sum_{\{j,k|\lambda_j+\lambda_k\neq 0\}} \frac{2}{\lambda_j + \lambda_k}|\Delta_{jk}|^2. \quad (13)$$

The remainder of the derivation of Eq. (2), to show that $S''(t) \leq L''(t)$, consists of demonstrating the inequality $\Phi(x,y) \geq 2/(x+y)$ (see [2]). This completes our discussion of the Holevo upper bound (2).

Now we focus on deriving explicit expressions for the binary-channel ensemble-dependent upper and lower bounds $L(t)$ and $M(t)$. The lower bound $M(t)$, in particular, can be written in quite a simple form. We start with the generic formula for the binary-channel mutual information expressed in a slightly different guise. By defining $\alpha_b = \mathrm{tr}(\hat{\rho}_0\hat{E}_b)/\mathrm{tr}(\hat{\rho}\hat{E}_b)$ and $\beta_b = \mathrm{tr}(\hat{\rho}_1\hat{E}_b)/\mathrm{tr}(\hat{\rho}\hat{E}_b)$, we write

$$I(t) = (1-t)\sum_b \mathrm{tr}(\hat{\rho}_0\hat{E}_b)\ln\alpha_b + t\sum_b \mathrm{tr}(\hat{\rho}_1\hat{E}_b)\ln\beta_b = \mathrm{tr}\left((1-t)\hat{\rho}_0\sum_b(\ln\alpha_b)\hat{E}_b + t\hat{\rho}_1\sum_b(\ln\beta_b)\hat{E}_b\right). \quad (14)$$

The lower bound $M(t)$ arises from using a measurement described by projectors $\hat{E}_b$ onto a basis that diagonalizes $\mathcal{L}_{\hat{\rho}}(\hat{\Delta})$. Note that $\mathcal{L}_{\hat{\rho}}(\hat{\Delta})$, $\mathcal{L}_{\hat{\rho}}(\hat{\rho}_0)$, and $\mathcal{L}_{\hat{\rho}}(\hat{\rho}_1)$ all commute and are thus simultaneously diagonalizable. This follows from the linearity of the $\mathcal{L}_{\hat{\rho}}$ superoperator: $\mathcal{L}_{\hat{\rho}}(\hat{\rho}_0) = \mathcal{L}_{\hat{\rho}}(\hat{\rho} - t\hat{\Delta}) = \hat{\mathbb{1}} - t\mathcal{L}_{\hat{\rho}}(\hat{\Delta})$ and $\mathcal{L}_{\hat{\rho}}(\hat{\rho}_1) = \mathcal{L}_{\hat{\rho}}[\hat{\rho} + (1-t)\hat{\Delta}] = \hat{\mathbb{1}} + (1-t)\mathcal{L}_{\hat{\rho}}(\hat{\Delta})$. One can then show, using Eq. (8), that α_b and β_b are the eigenvalues of $\mathcal{L}_{\hat{\rho}}(\hat{\rho}_0)$ and $\mathcal{L}_{\hat{\rho}}(\hat{\rho}_1)$ corresponding to the projector $\hat{E}_b$. Hence $M(t)$ takes the form

$$M(t) = \mathrm{tr}\{(1-t)\hat{\rho}_0\ln[\mathcal{L}_{\hat{\rho}}(\hat{\rho}_0)] + t\hat{\rho}_1\ln[\mathcal{L}_{\hat{\rho}}(\hat{\rho}_1)]\}. \quad (15)$$

The upper bound $L(t)$ has not so far yielded such a simple form. In principle all that need be done is to integrate Eq. (13) twice, applying the boundary conditions $L(0) = L(1) = 0$; the problem lies in finding a tractable representation for $\mathcal{L}_{\hat{\rho}}(\hat{\Delta})$. Here we note that *when $\hat{\rho}$ has no zero eigenvalues*, $\mathcal{L}_{\hat{\rho}}(\hat{A})$ can be written as a contour integral,

$$\mathcal{L}_{\hat{\rho}}(\hat{A}) = \frac{2}{2\pi i}\oint (z\hat{\mathbb{1}} - \hat{\rho})^{-1}\hat{A}(z\hat{\mathbb{1}} + \hat{\rho})^{-1}\,dz, \quad (16)$$

where the contour contains the pole at $z = \lambda_j$ for all eigenvalues λ_j of $\hat{\rho}$, but does not contain the pole at $z = -\lambda_j$ for any j. This contour representation leads to a Fourier series expansion for $L(t)$. It is not difficult, again using the operator-inverse differentiation formula, to work out that

$$\hat{\Delta}\frac{d^n}{dt^n}\mathcal{L}_{\hat{\rho}}(\hat{\Delta}) = \frac{2(n!)}{2\pi i}\sum_{k=0}^{n}(-1)^k D_{\hat{\rho}}(n;k), \quad (17)$$

where

$$D_{\hat{\rho}}(n;k) = \oint [\hat{\Delta}(z\hat{\mathbb{1}} - \hat{\rho})^{-1}]^{n+1-k}[\hat{\Delta}(z\hat{\mathbb{1}} + \hat{\rho})^{-1}]^{k+1}\,dz.$$

With this, one can derive a Taylor series expansion for $L''(t)$ and then use the standard algorithm for Fourier expansions to obtain

$$L(t) = \sum_{m=1}^{\infty} b_m \sin(m\pi t), \quad (18)$$

where

$$b_m = \frac{1}{m^3\pi^4 i}\sum_{n=0}^{\infty} n!\,B(n;m)\sum_{k=0}^{n}(-1)^k\mathrm{tr}[D_{\hat{\rho}_0}(n;k)], \quad (19)$$

$$B(n;m) = b(n;m) - (-1)^m\sum_{j=0}^{n}\frac{1}{(n-j)!}b(j;m), \quad (20)$$

and $b(j;m) = (-1)^{j/2}[1 + (-1)^j](m\pi)^{-j}$. The advantages of this representation are that it automatically satisfies the boundary conditions and only the first few terms in Eq. (18) are significant.

Finally, we consider a special case of some practical interest—binary communication channels on two-dimensional Hilbert spaces. Here the bounds $L(t)$ and $M(t)$ are expressible in terms of elementary functions; moreover, the optimal orthogonal projection-valued measurement can be found via a variational calculation. Let the message states $\hat{\rho}_0$ and $\hat{\rho}_1$ be represented by vectors within the Bloch sphere, i.e., $\hat{\rho}_0 = \frac{1}{2}(\hat{\mathbb{1}} + \boldsymbol{a}\cdot\boldsymbol{\sigma})$ and $\hat{\rho}_1 = \frac{1}{2}(\hat{\mathbb{1}} + \boldsymbol{b}\cdot\boldsymbol{\sigma})$, where $a = |\boldsymbol{a}| \leq 1, b = |\boldsymbol{b}| \leq 1$, and $\boldsymbol{\sigma}$ is the Pauli spin vector. (A state is pure if its Bloch vector has unit modulus.) The total density operator for the channel can be written as $\hat{\rho} = \frac{1}{2}(\hat{\mathbb{1}} + \boldsymbol{c}\cdot\boldsymbol{\sigma})$, where $\boldsymbol{c} = (1-t)\boldsymbol{a} + t\boldsymbol{b} = \boldsymbol{a} + t\boldsymbol{d} = \boldsymbol{b} - (1-t)\boldsymbol{d}$ and $\boldsymbol{d} = \boldsymbol{b} - \boldsymbol{a}$.

An orthogonal projection-valued measurement is defined by two projectors, specified by unit Bloch vectors $\boldsymbol{n}$ and $-\boldsymbol{n}$. For such a measurement the mutual information $I(t)$ takes the form

$$I(t) = \frac{1}{2}[(1-t)K_0 + tK_1]. \quad (21)$$

Here $K_0 = A_+\ln(A_+/C_+) + A_-\ln(A_-/C_-)$ and $K_1 = B_+\ln(B_+/C_+) + B_-\ln(B_-/C_-)$, where $C_\pm = 1 \pm \boldsymbol{c}\cdot\boldsymbol{n}$ and similarly for $A_\pm$ and $B_\pm$. The optimal projectors can

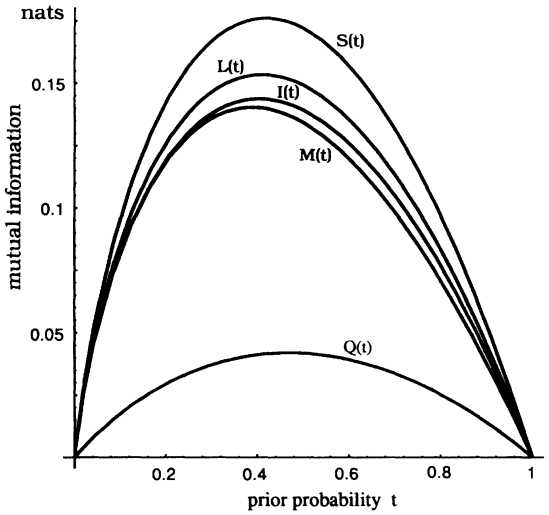

FIG. 1. Holevo upper bound $S(t)$, ensemble-dependent upper bound $L(t)$, information $I(t)$ extractable by optimal orthogonal projection-valued measurement, ensemble-dependent lower bound $M(t)$, and Jozsa-Robb-Wootters lower bound $Q(t)$, all for the case that $\hat{\rho}_0$ is pure ($a = 1$), $\hat{\rho}_1$ is mixed with $b = 2/3$, and the angle between the two Bloch vectors is $\pi/3$.

be found by varying Eq. (21) over all unit vectors $\boldsymbol{n}$; the resulting equation for the optimal $\boldsymbol{n}$ is

$$0 = (1 - t)\ln\left(\frac{C_+A_-}{C_-A_+}\right)\boldsymbol{a}_\perp + t\ln\left(\frac{C_+B_-}{C_-B_+}\right)\boldsymbol{b}_\perp , \qquad (22)$$

where $\boldsymbol{a}_\perp = \boldsymbol{a} - (\boldsymbol{a}\cdot\boldsymbol{n})\boldsymbol{n}$ and $\boldsymbol{b}_\perp = \boldsymbol{b} - (\boldsymbol{b}\cdot\boldsymbol{n})\boldsymbol{n}$. Though this transcendental equation generally has no explicit solution, we can obtain solutions in four situations: (i) a classical channel, where $\hat{\rho}_0$ and $\hat{\rho}_1$ commute ($\boldsymbol{a}$ and $\boldsymbol{b}$ parallel); (ii) $\hat{\rho}_0$ and $\hat{\rho}_1$ are pure states ($a = b = 1$); (iii) $a = b$ and $t = \frac{1}{2}$; (iv) $t = [1 + \sqrt{(1 - b^2)/(1 - a^2)}]^{-1}$. Cases (ii)–(iv) are a consequence of setting $(1 - t)\boldsymbol{a}_\perp = t\boldsymbol{b}_\perp$ and requiring that the arguments of the logarithms be multiplicative inverses. Cases (ii) and (iii), reported previously by Levitin [11], are limits of case (iv).

In case (iv) the optimal $\boldsymbol{n}$ (unnormalized) is given by

$$\boldsymbol{n} \propto \frac{\boldsymbol{a}\cdot(\boldsymbol{a} - \boldsymbol{c})}{\boldsymbol{a}\cdot(\boldsymbol{a} - \boldsymbol{b})}\boldsymbol{b} - \frac{\boldsymbol{b}\cdot(\boldsymbol{b} - \boldsymbol{c})}{\boldsymbol{b}\cdot(\boldsymbol{b} - \boldsymbol{a})}\boldsymbol{a} . \qquad (23)$$

In cases (i)–(iii) the optimal $\boldsymbol{n}$ is

$$\boldsymbol{n} = \frac{(1 - \boldsymbol{a}\cdot\boldsymbol{c})\boldsymbol{b} - (1 - \boldsymbol{b}\cdot\boldsymbol{c})\boldsymbol{a}}{D} = \frac{\boldsymbol{d} + \boldsymbol{c}\times(\boldsymbol{c}\times\boldsymbol{d})}{D} , \qquad (24)$$

where $\boldsymbol{c}\cdot\boldsymbol{n} = \boldsymbol{c}\cdot\boldsymbol{d}/D$ and

$$D = \sqrt{\delta^2(1 - c^2) + (\boldsymbol{c}\cdot\boldsymbol{d})^2} , \qquad (25)$$

$$\delta = \sqrt{D\boldsymbol{d}\cdot\boldsymbol{n}} = \sqrt{d^2 - |\boldsymbol{c}\times\boldsymbol{d}|^2} = \sqrt{d^2 - |\boldsymbol{a}\times\boldsymbol{b}|^2} . \qquad (26)$$

Case (ii) is of special interest because two pure states in a Hilbert space of any dimension span only a two-dimensional subspace; hence Eq. (24) gives the optimal orthogonal projection-valued measurement for a pure-state binary channel in all dimensions.

It turns out that $I''(t)$ is minimized by the measurement (24). The lower bound $M(t)$ follows from substituting this measurement into the expression for $I(t)$ in Eq. (21). The upper bound $L(t)$, found by substituting the measurement (24) into $I''(t)$ and integrating twice, is given by

$$L(t) = \frac{\delta}{2d^2}[\,-(\delta - \boldsymbol{c}\cdot\boldsymbol{d})\ln(\delta - \boldsymbol{c}\cdot\boldsymbol{d}) - (\delta + \boldsymbol{c}\cdot\boldsymbol{d})\ln(\delta + \boldsymbol{c}\cdot\boldsymbol{d}) + \beta_1 t + \beta_2\,] , \qquad (27)$$

where β_1 and β_2 are determined by the boundary conditions $L(0) = L(1) = 0$. How tight these bounds are relative to the Holevo upper and Jozsa-Robb-Wootters lower bounds is illustrated in a typical case in Fig. 1.

Much remains to be done. Although the Holevo upper bound (2) for more than two message states can be built from binary-channel bounds using Eq. (3), the lower bound $M(t)$ cannot be so generalized. The upper bound $L(t)$ *can* be generalized, but we speculate that it becomes increasingly loose as it is generalized to many message states. A route to tight upper and lower bounds for many message states, we conjecture, is to deal directly with the matrix of second derivatives of the mutual information (1), a matrix known to parameter-estimation theory as the Fisher information matrix.

This work was supported in part by the Office of Naval Research Grant No. N00014-93-1-0116. We thank S. L. Braunstein and C. Chandler for helpful discussions.

[1] C. M. Caves and P. D. Drummond, Rev. Mod. Phys. **66**, 481 (1994).
[2] A. S. Holevo (Kholevo), Probl. Peredachi Inf. **9**(3), 3 (1973) [Prob. Inf. Transm. **9**, 177 (1973)].
[3] R. Jozsa, D. Robb, and W. K. Wootters, Phys. Rev. A **49**, 668 (1994).
[4] H. P. Yuen and M. Ozawa, Phys. Rev. Lett. **70**, 363 (1993).
[5] M. J. W. Hall and M. J. O'Rourke, Quantum Opt. **5**, 161 (1993).
[6] K. Kraus, *States, Effects, and Operations: Fundamental Notions of Quantum Theory* (Springer, Berlin, 1983).
[7] C. E. Shannon, Bell Syst. Tech. J. **27**, 379 623 (1948).
[8] R. A. Fisher, Proc. R. Soc. Edinburgh **42**, 321 (1922).
[9] H. Poincaré, Trans. Cambridge Philos. Soc. **18**, 220 (1899).
[10] S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. **72**, 3439 (1994).
[11] L. B. Levitin, in *Workshop on Physics and Computation: PhysComp '92*, edited by D. Matzke (IEEE Computer Society Press, Los Alamitos, CA, 1993).