PHYSICAL REVIEW LETTERS

VOLUME 74 | 24 APRIL 1995 | NUMBER 17

Information Exclusion Principle for Complementary Observables

Michael J. W. Hall

Department of Theoretical Physics, Australian National University, Canberra Australian Capital Territory 0200, Australia

(Received 31 August 1994)

The sum of the information gains corresponding to measurements of position and momentum is bounded by $\log 2\Delta X\Delta P/\hbar$ for a quantum ensemble with position and momentum uncertainties ΔX and ΔP. The bound implies the Heisenberg uncertainty principle and that the gain of position information can be maximized only at the expense of momentum information, and vice versa. This *exclusion* principle for the information contents of complementary observables is extended to finite Hilbert spaces, and to the quadrature, number, and phase observables of bosonic fields degraded by Gaussian noise.

PACS numbers: 03.65.Bz, 42.50.Lc, 89.70.+c

Information theory is a natural and effective tool in communications, parameter estimation, and computation. One may seek to maximize the information which can be gained about, for example, a signal from a space probe, the magnitude of a viral epidemic, or the divisors of a large number. This optimization of information gain, for given prior information, is in principle trivial for classical systems: one can do no better than a complete state determination (e.g., of the electromagnetic field, the medical condition of all population members, or the remainders of all possible divisions). With the exception of Maxwell's demon [1], any physics underlying classical information gain is thus limited to the analysis of particular constraints on measurement (such as sampling rates, research funding, or computer run time).

For quantum systems, the *complementarity* of observables such as position and momentum leads immediately to the impossibility of a complete state determination in general, and hence to a greater physical richness underlying quantum information gain. The most important example is a finite bound on information transfer for any quantum communication scheme [2,3]; other examples arise in the areas of quantum cryptography [4], quantum inference [5], quantum computation [6], and in interpretational issues [7].

Here a very simple and direct signature of complementarity in quantum information theory will be demonstrated, in the form of an *exclusion principle* for the information contents of quantum observables. This principle states that the information gain corresponding to the measurement of some observable can be maximized only at the expense of the information gains corresponding to complementary observables, and is quantified below for a number of cases.

First, let the prior information about a quantum system be represented by a statistical ensemble $\mathcal{E}$ of possible states of the system. If the state of the system is known to be described by density operator ρ_i with probability p_i, then the density operator of the corresponding ensemble $\mathcal{E}$ is given by [8]

$$\rho_{\mathcal{E}} = \sum_i p_i \rho_i \, . \tag{1}$$

Further, if $p(a|\rho)$ denotes the probability distribution of observable A for density operator ρ, then the associated *entropy,* $S(A|\rho)$, is defined by [9]

$$S(A|\rho) = -\int da\, p(a|\rho) \log p(a|\rho) \, . \tag{2}$$

Finally, the (Shannon) mutual information corresponding to a measurement of A on a member of $\mathcal{E}$ is given by ([10], Eq. (3.131))

$$I(A|\mathcal{E}) = S(A|\rho_{\mathcal{E}}) - \sum_i p_i S(A|\rho_i) . \tag{3}$$

Shannon's fundamental theorem ([10], Chap. 12) states that $I(A|\mathcal{E})$ in Eq. (3) is the average amount of error-free data which may be obtained via a measurement of A [11]. For the choice of logarithm base 2 in Eq. (2), this amount is quantified by the number of binary digits (i.e., bits) required to represent the data. A fundamental result of Holevo [2,3] implies the finite bound

$$I(A|\mathcal{E}) \leq -\mathrm{tr}[\rho_{\mathcal{E}} \log \rho_{\mathcal{E}}] \tag{4}$$

for quantum information, and it follows in particular that [3]

$$I_N = \log N, \tag{5}$$

$$I_{\text{boson}}(\overline{n}) = \log(\overline{n} + 1) + \overline{n} \log(1 + \overline{n}^{-1}), \tag{6}$$

where I_N denotes the maximum information obtainable from an ensemble spanning an N-dimensional Hilbert space, and $I_{\text{boson}}(\overline{n})$ the maximum information obtainable from an ensemble of single-mode bosonic fields with an average number $\overline{n}$ of bosons.

The information exclusion principle is exemplified by inequalities of the general form

$$I(A_1|\mathcal{E}) + I(A_2|\mathcal{E}) + \cdots \leq J(A_1, A_2, \ldots, \rho_{\mathcal{E}}) \tag{7}$$

for observables $A_1, A_2, \ldots$, where J is a nontrivial quantum bound. Examples include

$$I(X|\mathcal{E}) + I(P|\mathcal{E}) \leq \log 2(\Delta X)_{\mathcal{E}}(\Delta P)_{\mathcal{E}}/\hbar \tag{8}$$

for position and momentum, and

$$I(\sigma_1|\mathcal{E}) + I(\sigma_2|\mathcal{E}) + I(\sigma_3|\mathcal{E}) \leq \log 2 \tag{9}$$

for orthogonal spin components of a spin-$\frac{1}{2}$ particle. Since a nonideal measurement of an observable A cannot yield more information than a measurement of A itself, one also has an immediate corollary

$$I(\tilde{A}_1|\mathcal{E}) + I(\tilde{A}_2|\mathcal{E}) + \cdots \leq J(A_1, A_2, \ldots, \rho_{\mathcal{E}}), \tag{10}$$

to Eq. (7), where $\tilde{A}_m$ represents an observable corresponding to a (possibly nonideal) measurement of A_m.

The nontrivial bound in Eq. (7) implies that the information content of an observable can be maximized for a given $\rho_{\mathcal{E}}$ only at the expense of the information carried by complementary observables. For example, $I(\sigma_1|\mathcal{E})$ in Eq. (9) can attain its maximum possible value of log2 [see Eq. (5)] *only* if the complementary spin components σ_2, σ_3 carry no information. The exclusion principle is thus an information-theoretic analog of quantum uncertainty principles, in that for both cases certain characteristics of complementary observables cannot simultaneously be optimized. Moreover, Eq. (8) immediately yields (and hence is stronger than) the Heisenberg uncertainty relation for position and momentum, while a close connection between the exclusion principle and entropic uncertainty relations will be demonstrated further below.

A relatively simple strategy will be followed here to obtain the information exclusion relations (8), (9), and others; sharper relations can be obtained via more sophisticated techniques as is indicated at the end of this Letter. The strategy is to search for inequalities of the form

$$L(A_1, \ldots, A_M) \leq \sum_{m=1}^{M} S(A_m|\rho) \leq U(A_1, \ldots, A_M, \rho), \tag{11}$$

which, together with Eq. (3), yield the exclusion relation

$$\sum_{m=1}^{M} I(A_m|\mathcal{E}) \leq U(A_1, \ldots, A_M, \rho_{\mathcal{E}}) - L(A_1, \ldots, A_M). \tag{12}$$

The first inequality in Eq. (11) has the form of an entropic uncertainty relation [12], indicating a strong connection between the entropic-uncertainty and information-exclusion principles. However, the latter has some interpretational advantages, as (i) a primary aim in minimizing uncertainties of various observables (e.g., photon number and quadrature fluctuations) is to increase their information content [13]; (ii) information, in quantifying data, is a quantity of direct physical significance (and unlike entropy is invariant under arbitrary rescaling of observables ([10], Sec. 8.11); (iii) the exclusion relation (7) automatically generalizes to Eq. (10) for nonideal measurements, whereas much effort must be expended to find entropic uncertainty relations for such measurements [14].

The value of the above strategy is that the difficult part, of finding nontrivial lower bounds $L(A_1, \ldots, A_M)$, has already been carried out in many cases [14–18]. For example, let A and B denote two observables of a quantum system with an N-dimensional Hilbert space, and define

$$c = \max_{a,b} |\langle a|b\rangle| \tag{13}$$

as the maximum possible overlap of eigenstates of A and B. For nondegenerate A and B a simple variational calculation yields the upper bound

$$S(A|\rho), S(B|\rho) \leq \log N \tag{14}$$

(attained when $\rho = N^{-1}\hat{1}$), while the entropic uncertainty relation

$$S(A|\rho) + S(B|\rho) \geq -2 \log c \tag{15}$$

follows from [15], Eq. (6). From Eqs. (11), (12), (14), and (15) one obtains the information exclusion relation

$$I(A|\mathcal{E}) + I(B|\mathcal{E}) \leq 2 \log Nc. \tag{16}$$

Note that Eq. (16) is valid for *degenerate* A and B also, as per Eq. (10), since such observables can be regarded as corresponding to nonideal measurements of nondegenerate observables [19].

Two nondegenerate observables A and B as above are defined to be *complementary* if the distribution of A is uniform for any eigenstate of B, and vice versa [20] (thus generalizing the notion of conjugate observables on infinite Hilbert spaces). For complementary observables one has the strongest possible from of Eq. (16),

$$I(A|\mathcal{E}) + I(B|\mathcal{E}) \leq \log N. \tag{17}$$

Note from Eq. (5) that this upper bound is the best possible bound independent of the ensemble density operator.

The orthogonal spin components $\sigma_1, \sigma_2, \sigma_3$ of a spin-$\frac{1}{2}$ particle provide an example of *three* pairwise complementary observables. From Eq. (14) and [16], Eq. (22), one has

$$2\log 2 \leq \sum_{j=1}^{3} S(\sigma_j|\rho) \leq 3\log 2, \tag{18}$$

and the exclusion relation in Eq. (9) follows via Eqs. (11) and (12). Comparison of Eqs. (9) and (17) suggests the strong conjecture

$$\sum_{m=1}^{M} I(A_m|\mathcal{E}) \leq \log N \tag{19}$$

for M pairwise complementary observables $A_1, \ldots, A_M$ on an N-dimensional Hilbert space [21].

Attention will now be turned to exclusion relations for infinite Hilbert spaces. First, let $\mathbf{X}$ and $\mathbf{P}$ be n vectors denoting the position and momentum observables of a quantum system. A simple variational calculation yields the upper bound

$$S(\mathbf{X}|\rho) \leq \frac{n}{2}\log 2\pi e + \frac{1}{2}\sum_{j=1}^{n}\log(\mathrm{Var}X_j) \tag{20}$$

for the position entropy of the system. Combined with a similar bound for $S(\mathbf{P}|\rho)$, and the entropic uncertainty relation [17]

$$S(\mathbf{X}|\rho) + S(\mathbf{P}|\rho) \geq n\log \pi e\hbar, \tag{21}$$

the strategy of Eqs. (11) and (12) may be followed as before to yield the exclusion relation

$$I(\mathbf{X}|\mathcal{E}) + I(\mathbf{P}|\mathcal{E}) \leq \log\prod_j 2(\Delta X_j)_{\mathcal{E}}(\Delta P_j)_{\mathcal{E}}/\hbar, \tag{22}$$

which reduces to Eq. (8) when $n = 1$ [22].

The upper bound in Eq. (22) has a simple semiclassical interpretation. In particular, the argument of the logarithm is just the number of nonoverlapping phase-space cells, of minimum volume $(2\hbar)^n$, available to an ensemble $\mathcal{E}$ restricted to a volume $4^n(\Delta X_1)_{\mathcal{E}}\cdots(\Delta X_n)_{\mathcal{E}}\,(\Delta P_1)_{\mathcal{E}}\cdots(\Delta P_n)_{\mathcal{E}}$ of phase space. More generally, Eqs. (9), (17), and (22) suggest that for mutually complementary observables $A_1, \ldots, A_M$ the upper bound in Eq. (7) is approximately given by $\log\mathcal{N}(\rho_{\mathcal{E}})$, where $\mathcal{N}(\rho_{\mathcal{E}})$ is the maximum possible number of mutually orthogonal states consistent with the ensemble density operator $\rho_{\mathcal{E}}$.

Equations (8) and (22) may be applied to the special case of a one-dimensional harmonic oscillator, with mass m, frequency ω, and annihilation operator a. It is of interest to further include the effects of additive Gaussian noise [18], as degradation of position or momentum information by thermal noise [18], linear amplification [23], and inefficient homodyne and balanced-homodyne detection [24] may be modeled by Gaussian noise of variances

$$n_\gamma^{\text{thermal}} = [\exp(\hbar\omega/kT) - 1]^{-1}, \tag{23}$$

$$n_\gamma^{\text{lin amp}} = \frac{1 - \exp[-2\alpha(N_2 - N_1)]}{1 - N_1/N_2}, \tag{24}$$

$$n_\gamma^{\text{hom}} = (1 - \eta)/2\eta, \tag{25}$$

$$n_\gamma^{\text{bal hom}} = (2 - \eta)/2\eta, \tag{26}$$

respectively, where k is Boltzmann's constant, T absolute temperature, N_1 (N_2) the number of excited (unexcited) amplifier atoms, α is an amplification constant, and η denotes detector efficiency.

For a harmonic oscillator subjected to Gaussian noise of variance n_γ, the position and momentum variances are increased by $\hbar(m\omega)^{-1}n_\gamma$ and $\hbar m\omega n_\gamma$, respectively ([25], Eq. (11)). Further, since the geometric mean never exceeds the arithmetic mean,

$$\begin{aligned}
&[\mathrm{Var}X + \hbar(m\omega)^{-1}n_\gamma](\mathrm{Var}P + \hbar m\omega n_\gamma) \\
&\leq \hbar^2\langle\hbar^{-1}m\omega X^2 + n_\gamma\rangle\langle(\hbar m\omega)^{-1}P^2 + n_\gamma\rangle \\
&\leq \hbar^2(\tfrac{1}{2}\langle\hbar^{-1}m\omega X^2 + n_\gamma\rangle + \tfrac{1}{2}\langle(\hbar m\omega)^{-1}P^2 + n_\gamma\rangle)^2 \\
&= \hbar^2(\overline{n} + n_\gamma + \tfrac{1}{2})^2,
\end{aligned} \tag{27}$$

where $\overline{n}$ denotes the average value of the number operator $a^\dagger a$. Equations (20) and (27), and a generalization of Eq. (21) to include Gaussian noise ([18], Eq. (20)), lead to the noise-dependent exclusion relation

$$I(X|\mathcal{E}, n_\gamma) + I(P|\mathcal{E}, n_\gamma) \leq \log[1 + \overline{n}/(n_\gamma + 1/2)]. \tag{28}$$

Note that the upper bound is *strictly less* than $I_{\text{boson}}(\overline{n})$ in Eq. (6), and decreases as the noise level increases.

To apply Eq. (28) to quantum communication, note that position and momentum measurements on an oscillator are equivalent (up to scale factors) to measurements of the quadratures $(a + a^\dagger)/2, (a - a^\dagger)/2i$ of a single-mode bosonic field, i.e., to homodyne detection of the field [24]. Since information gain is independent of scaling ([10], Sec. 8.11), it follows in particular from Eq. (28) that the information gained from a homodyne measurement is bounded above by $\log(2\overline{n} + 1)$. This bound is achieved for a suitable ensemble $\mathcal{E}$ of *squeezed coherent* states [26], and hence *such states are optimal for single-mode communication based on homodyne detection.* Note as previously that to maximize the information gain in one observable, the complementary observable must carry *no* information, i.e., the optimal signal states are modulated with respect to the measured quadrature [26].

An exclusion relation may also be obtained for the number and phase observables N and Φ of a harmonic oscillator degraded by the Gaussian noise of variance n_γ. For a given average number $\langle N\rangle = \overline{n}$ one has [27]

$$S(N|\rho, n_\gamma) \leq I_{\text{boson}}(\overline{n} + n_\gamma), \tag{29}$$

where $I_{\rm boson}$ is defined in Eq. (6). Further, $S(\Phi|\rho, n_\gamma)$ is bounded above by $\log 2\pi$, while from [18], Eq. (34),

$$S(N|\rho, n_\gamma) + S(\Phi|\rho, n_\gamma) \geq \log 2\pi + I_{\rm boson}(n_\gamma). \quad (30)$$

Applying the strategy of Eqs. (11) and (12) yields the exclusion relation

$$I(N|\mathcal{E}, n_\gamma) + I(\Phi|\mathcal{E}, n_\gamma) \leq I_{\rm boson}(\overline{n} + n_\gamma) - I_{\rm boson}(n_\gamma). \quad (31)$$

This upper bound is also a strong bound for the *maximum* information obtainable from an oscillator degraded by Gaussian noise [25,27].

Finally, while the derivations of exclusion relations (8), (9), (16), (17), (22), (28), and (31) rely on entropic inequalities of the form (11), stronger exclusion relations are possible via different methods. For example, consider an ensemble of equal mixtures of two pure states of a spin-$\frac{1}{2}$ particle. If the angle between the two corresponding spin directions is known to be θ, then the entropic inequalities in Eq. (21) of [18] yield, via Eq. (12),

$$\sum_{j=1}^{3} I(\sigma_j|\mathcal{E}) \leq 3F(3^{-1/2}\cos\theta/2) - 2\log 2, \quad (32)$$

where

$$F(x) = -\tfrac{1}{2}[(1+x)\log(1+x)/2 + (1-x)\log(1-x)/2]. \quad (33)$$

However, direct optimization via standard variational techniques yields the *stronger* bound [28]

$$\sum_{j=1}^{3} I(\sigma_j|\mathcal{E}) \leq \max[J_1(\theta), J_2(\theta)], \quad (34)$$

where

$$J_1(\theta) = \log 2 - F(\sin\theta/2),$$
$$J_2(\theta) = 2F\Big(2^{-1/2}\cos\frac{\theta}{2}\Big) - F\Big(\cos\Big[\frac{\theta}{2} + \frac{\pi}{4}\Big]\Big) - F\Big(\cos\Big[\frac{\theta}{2} - \frac{\pi}{4}\Big]\Big).$$

It would be of interest to determine whether inequalities involving *relative* entropies, as in Eq. (9) of [3], could be used to improve on the exclusion relations derived here.

In conclusion, a new "information exclusion principle" has been presented, which quantifies the notion that the information content of a quantum observable can be increased only at the expense of the information carried by complementary observables. This information-theoretic signature of complementarity has the conceptual advantages of automatic generalization to inexact measurements, invariance under reparametrization of observables, and applicability to both finite and infinite dimensional systems. Technical advantages include measurement-dependent information inequalities (i.e., exclusion relations) which are typically stronger than the measurement-independent Holevo bounds (5) and (6), and direct applicability of results to quantum communication. For example, the exclusion relation (28) for a harmonic oscillator degraded by Gaussian noise immediately implies that squeezed coherent states are optimal for narrow band communication based on homodyne detection. It is hoped to develop the information exclusion principle further, both as a tool for finding optimal signal states and for studying correlations between quantum systems.

I thank Robyn Hall for financial support.

Note added.—A recent result in the literature [29] implies that conjecture (19) cannot hold in general. In particular, the ensemble comprising equal mixtures of the nine states of the form given in Eq. (16) of [29] violates the conjecture for the $N = 3$ case.

[1] The ability of Maxwell's demon to make such complete state determination leads to the physics of information erasure; see, e.g., C. H. Bennett, Int. J. Theor. Phys. **21**, 905 (1982).

[2] A. S. Holevo, Probl. Inf. Trans. **9**, 177 (1973).

[3] H. P. Yuen and M. Ozawa, Phys. Rev. Lett. **70**, 363 (1993).

[4] A. Ekert, Phys. Rev. Lett. **67**, 661 (1991); C. H. Bennett, G. Brassard, and N. D. Mermin, Phys. Rev. Lett. **68**, 557 (1992).

[5] K. R. W. Jones, Ann. Phys. (N.Y.) **207**, 140 (1991).

[6] D. Deutsch and R. Josza, Proc. R. Soc. London A **439**, 553 (1992).

[7] A. S. Holevo, Probl. Inf. Trans. **9**, 110 (1973); W. K. Wootters and W. H. Zurek, Phys. Rev. D **19**, 473 (1979); S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. **61**, 662 (1988).

[8] For example, B. D'Espagnat, *Conceptual Foundations of Quantum Mechanics* (Benjamin, Menlo Park, 1971), Chap. 6.

[9] The logarithm base is left arbitrary throughout, corresponding to a choice of units. The choices of base 2 and base e correspond to units of bits and nats, respectively.

[10] F. M. Reza, *An Introduction to Information Theory* (McGraw-Hill, New York, 1961).

[11] For example, $\mathcal{E}$ may represent the signal states of a communication channel, the possible output states of an optical fiber for a given input state, or a uniform distribution over all possible density operators of the system. $I(A|\mathcal{E})$ then represents, respectively, the average quantity of error-free data transmitted per signal, the information gained about the optical path length, or the information gained about the (completely unknown) state of the system, via a measurement of A.

[12] B. Mamojka, Int. J. Theor. Phys. **11**, 73 (1974); D. Deutsch, Phys. Rev. Lett. **50**, 631 (1983).

[13] MC. Teich and B. E. A. Saleh, Phys. Today **43**, No. 6, 26 (1990).

[14] M. H. Partovi, Phys. Rev. Lett. **50**, 1883 (1983); F. E. Schroeck, J. Math. Phys. **30**, 2078 (1989).

[15] H. Maassen and J. B. M. Uffink, Phys. Rev. Lett. **60**, 1103 (1988).

[16] J. Sánchez, Phys. Lett. A **173**, 233 (1993).
[17] I. Bialynicki-Birula and J. Mycielski, Commun. Math. Phys. **44**, 129 (1975).
[18] M. J. W. Hall, Phys. Rev. A **49**, 42 (1994).
[19] More rigorously, if A and B are degenerate, with maximum overlap c attained for eigenstates $|a\rangle$, $|b\rangle$ of A, B, respectively, let $\overline{A}$ and $\overline{B}$ denote any maximal self-adjoint extensions of A and B which preserve $|a\rangle$ and $|b\rangle$ as eigenstates. Then $\overline{A}$ and $\overline{B}$ are by definition nondegenerate with maximum overlap $\overline{c} = c$ of eigenstates, and hence satisfy Eq. (16). But $I(A|\mathcal{E}) \leq I(\overline{A}|\mathcal{E})$, $I(B|\mathcal{E}) \leq I(\overline{B}|\mathcal{E})$ ([10], Eq. (3.155)), and Eq. (16) follows.
[20] K. Kraus, Phys. Rev. D **35**, 3070 (1987).
[21] A weaker bound, $(N + 1)\log 2N/(N + 1)$, follows via the entropic bound in I. D. Ivanovic, J. Phys. A **25**, L363 (1992).
[22] A tighter bound for Eq. (22) is $\frac{1}{2}\log \times \det[\mathrm{Cov}(\mathbf{X})\mathrm{Cov}(\mathbf{P})/(\hbar/2)^2]$, where $\mathrm{Cov}(\mathbf{V})$ denotes the covariance matrix $\langle \mathbf{V}\mathbf{V}^{\mathbf{T}}\rangle - \langle \mathbf{V}\rangle\langle \mathbf{V}^{\mathbf{T}}\rangle$.
[23] S. Friberg and L. Mandel, Opt. Commun. **46**, 141 (1983).
[24] U. Leonhardt and H. Paul, Phys. Rev. A **48**, 4598 (1993).
[25] M. J. W. Hall, Phys. Rev. A **50**, 3295 (1994).
[26] C. M. Caves and P. D. Drummond, Rev. Mod. Phys. **66**, 481 (1994).
[27] M. J. W. Hall and M. J. O'Rourke, Quantum Opt. **5**, 161 (1993).
[28] The bound in Eq. (34) corresponds to the average spin direction lying halfway between two of the coordinate axes, with the two individual spin directions in the plane parallel (orthogonal) to the remaining axis for J_1 (J_2).
[29] J. Sánchez-Ruiz, J. Phys. A **27**, L843 (1994).