Quantum information and correlation bounds

Michael J. W. Hall
Department of Theoretical Physics, Institute of Advanced Studies, Australian National University, Canberra, Australian Capital Territory 0200, Australia

(Received 15 July 1996)

This paper is primarily concerned with the development and application of quantum bounds on mutual information, although some of the methods developed can be applied to any figure of merit indicating degree of correlation, such as coincidence rate. Three basic techniques for obtaining bounds are described: mappings between joint-measurement and communication correlation contexts; a duality relation for quantum ensembles and quantum measurements; and an information exclusion principle [M. J. W. Hall, Phys. Rev. Lett. **74**, 3307 (1995)]. Results include a proof of Holevo's communication bound from a joint-measurement inequality; a measurement-dependent dual to Holevo's bound; lower bounds for mutual information under ensemble and measurement constraints; information exclusion relations for measurements described by probability-operator measures; a proof that Glauber coherent states are optimal signal states for quantum communication based on (noisy) optical heterodyne detection; and an information inequality for quantum eavesdropping. Relations between the three techniques are used to further obtain upper bounds for quantum information, and to extend the information exclusion principle to a joint-measurement context. [S1050-2947(97)04701-X]

PACS number(s): 03.65.Bz, 42.50.Ar, 42.50.Dv, 89.70.+c

I. INTRODUCTION

The prediction of correlations between physical systems is a fundamental role of physical theories, and of quantum mechanics in particular. Such correlations may be, for example, between the outputs of two detectors, between the transmitter and receiver of a communication channel, or between a scatterer and a scattered probe signal. It is often of interest to optimize these correlations (e.g., for secure key distribution, maximum information transfer, or efficient estimation of scattering parameters), with respect to some figure of merit such as coincidence rate, mutual information, maximum likelihood, etc. It is also of fundamental interest to explore and understand the effects of quantum mechanics on such correlations.

The natural figure of merit for the degree of correlation between two statistical sources is the (Shannon) mutual information, which quantifies the average amount of error-free data which can be obtained about a sequence of outputs from one source, given the corresponding sequence of outputs from the other source [1,2]. This amount is generally expressed in terms of the number of binary digits (bits) required to represent the data. It is maximized when the sources are perfectly correlated, minimized when they are completely uncorrelated, is (unlike entropy) invariant under continuous reparametrizations of the source outputs, and (unlike coincidence rates and maximum likelihood) does not rely on any mapping connecting the output ranges of the two sources.

A simple classical example is the random placement of a chess piece on some square of a chessboard as the first statistical source, and the subsequent result of measurement of which square is occupied as the (perfectly correlated) second source. Clearly the error-free data gained per output of the second source, concerning the output of the first source, is $\log_2 64=6$ bits of information. Note that such information could be used for communication purposes, where each square represents a symbol or a code word.

For *quantum* systems, complementarity immediately places rather strong constraints on mutual information. A simple semiclassical example is provided by a ''quantum chessboard,'' corresponding to some area A of a two-dimensional phase space. The Heisenberg uncertainty principle implies that a quantum system (''chess piece'') occupies a region (''square'') in A with area of at least $2\hbar$ (so that the product of its position and momentum half-widths is greater than $\hbar/2$). Hence the maximum number of nonoverlapping regions available to the system is $A/2\hbar$, implying that the amount of error-free data which can be gained per measurement about the phase space location of the system is bounded by

$$I_{\max}\approx\log_2(A/2\hbar). \tag{1}$$

For example, the maximum energy constraint

$$p^2/(2m)+V(x)\leq E \tag{2a}$$

implies the semiclassical information bound

$$I_{\max}\approx\log_2\left(\int dx[2m(E-V)]^{1/2}/\hbar\right) \tag{2b}$$

(where integration is over all real values of the integrand). Similar heuristic bounds for bosonic channels are given in [3].

For *classical* systems, where states can occupy arbitrarily small regions of phase space, there is no fundamental finite bound on information analogous to Eq. (1). The derivation of Eq. (1) is, of course, heuristic in nature; however, this inequality has a rigorous counterpart known as Holevo's theorem [3–5] (see also below). Other rigorous demonstrations of differences between quantum and classical information have been given in the areas of entangled states [6], quantum computing [7], quantum cryptography [8,9], quantum measurement [10], and complementary observables [11,12].

This paper is primarily concerned with the development and application of quantum bounds on mutual information; however, the methods developed in Sec. II and III can be applied to *any* figure of merit for the degree of correlation between two sources. The determination of information bounds is not only important for exploring quantum limits on correlations, but also because little progress has been made on exact results in quantum information theory (a notable exception is a paper by Davies [13], which significantly limits the possible measurements which must be considered in determining maximum information gain for a given ensemble of states).

The necessary elements of quantum information theory are briefly reviewed in Sec. II, and a general correspondence between joint-measurement and communication correlation contexts is noted which allows results derived in one context to be mapped to results for the other context. An interesting example is a proof of Holevo's theorem [3–5] (which bounds the information which can be gained by measurement on a given ensemble of quantum states in a communication context), from a joint-measurement inequality. An appendix to this section outlines the context-independent interpretation of mutual information.

In Sec. III a ''source duality'' property of mutual information (first noted in a quantum communication context in [14]), which formally transforms signal states into measurement outcomes and vice versa, is exploited to derive measurement-dependent correlation bounds from measurement-independent bounds. In particular a measurement-dependent dual to Holevo's theorem is obtained and discussed in Sec. III B. Source duality further provides a method for obtaining lower bounds for quantum correlations, which is applied to mutual information in Sec. III C.

In Sec. IV a recent ''information exclusion principle'' [12], which bounds the sum of information gains corresponding to measurements of complementary observables, is heuristically motivated and applied to generalized measurements, single-mode field quadratures, heterodyne detection, and quantum eavesdropping. Upper bounds for mutual information are obtained which are stronger than those provided by Holevo's theorem. Results from Secs. II and III are used to extend the information exclusion principle to a joint-measurement context, and to obtain bounds on the sum of information gains corresponding to a given measurement on two different ensembles.

Results are discussed in Sec. V.

II. QUANTUM INFORMATION

A. Formalism

As mentioned in the Introduction, information theory is concerned with correlations between the output sequences of statistical sources. In a *communication context* these sources comprise a ''transmitter,'' which outputs a sequence of signals, and a ''receiver,'' which outputs a sequence of measurement results (one for each received signal). In a *joint-measurement context* the sources comprise two detectors (e.g., polarization detectors) which generate correlated sequences of measurement results for some fixed sequence of identical input states (e.g., a singlet state) [15].

In either context let P_{ij} denote the joint probability that two sources, X and Y, generate the correlated pair of outputs (x_i, y_j). Defining the marginal probabilities

$$P_i = \sum_j P_{ij}, \quad P_j = \sum_i P_{ij} \tag{3}$$

corresponding, respectively, to output x_i being generated by X and output y_j being generated by Y, the Shannon mutual information is defined by [1,2]

$$I(P_{ij}) = \sum_{i,j} P_{ij} \log_2 P_{ij}/(P_i P_j). \tag{4}$$

As outlined in the Appendix, this quantity is *the average amount of error-free data which can be gained per member of a long sequence of outputs from one source, about the corresponding sequence of outputs from the other source* [1,2]. Note that it reduces to zero in the completely uncorrelated case $P_{ij} = P_i P_j$, and becomes equivalent to the entropy of either source, $-\Sigma_i P_i \log_2 P_i$, in the perfectly correlated case $P_{ij} = P_i \delta_{ij}$. Further, summation in (5) may be replaced by integration in the case of continuously valued outputs, and is invariant under reparametrizations of such outputs [1,16].

The quantum form of P_{ij} differs markedly in the communication and joint-measurement contexts. In the former context each output x_i of source X is associated with some signal state ρ_i on a Hilbert space H, which is transmitted with some prior probability p_i. Further, each output y_j of source Y is associated with a result of some measurement A. In general A can be represented by a probability-operator measure (POM) on H [3,17], i.e., by a set of positive operators $\{A_j\}$ which sum to the identity operator, such that the probability of measurement result y_j for state ρ is given by $\mathrm{tr}[\rho A_j]$. Thus in the communication context one has

$$P_{ij} = p_i \mathrm{tr}[\rho_i A_j], \tag{5a}$$

and the corresponding mutual information will be denoted by

$$I(P_{ij}) \equiv I(A|\mathcal{E}), \tag{5b}$$

where $\mathcal{E}$ denotes the ensemble of signal states $\{\rho_i; p_i\}$.

In contrast, in the joint-measurement context each source X and Y is associated with (compatible) measurements M and N, respectively, with corresponding (commuting) POM's $\{M_i\}$ and $\{N_j\}$, and the measurements are made on a fixed input state ρ. Thus one has in this context

$$P_{ij} = \mathrm{tr}[\rho M_i N_j], \tag{6a}$$

and the corresponding mutual information will be denoted by

$$I(P_{ij}) \equiv I(M,N|\rho). \tag{6b}$$

An interesting example of mutual information in the joint-measurement context arises from noisy quantum cryptography. In particular, if $|\psi\rangle$ denotes the singlet state of two spin-1/2 particles, define the rotationally symmetric ''noisy'' singlet state W_F by [18]

$$W_F=F|\psi\rangle\langle\psi|+[(1-F)/3]1_T, \tag{7a}$$

where $0\leqslant F\leqslant 1$ and 1_T denotes the unit operator on the triplet subspace. If two observers respectively measure the spin of the particles in a fixed direction **a**, the corresponding mutual information may be calculated from Eqs. (3), (4), and (6) as

$$I(\boldsymbol{\sigma}^{(1)}\cdot\boldsymbol{a},\boldsymbol{\sigma}^{(2)}\cdot\boldsymbol{a}|W_F)=\log_2 2-H([1+2F]/3), \tag{7b}$$

where $H(x)$ denotes the entropy function $-x\log_2 x-(1-x)\log_2(1-x)$. This quantity is related to the number of binary digits per joint measurement which can be used to establish a cryptographic key between the observers [8] (see also Appendix). It is maximized for $F=1$ (perfect correlation), corresponding to generating 1 binary digit per measurement, and minimized for $F=1/4$ (zero correlation). The observers should, of course, be able to do better by making measurements on product states $W_F\otimes\cdots\otimes W_F$ (studied in [18] in the context of state purification).

B. Context mappings

Here formal connections between Eqs. (5a) and (6a) will be demonstrated, which allow results obtained in one context to be transformed into results for the other context. In particular, for p_i, ρ_i, and A_j as in Eq. (5a), let $\{|\psi_i\rangle\}$ be a complete orthonormal basis in an ancillary Hilbert space H'. Then the corresponding joint-measurement context defined by

$$M_i=|\psi_i\rangle\langle\psi_i|,\quad N_j=A_j,\qquad \rho=\sum_i p_i|\psi_i\rangle\langle\psi_i|\otimes\rho_i \tag{8}$$

yields, via Eq. (6a), an equivalent joint distribution P_{ij}, and thus $I(M,N|\rho)=I(A|\mathcal{E})$. Conversely, for M_i, N_j, and ρ as in Eq. (6a), the corresponding communication context defined by

$$p_i=\mathrm{tr}[\rho M_i],\quad \rho_i=(M_i)^{1/2}\rho(M_i)^{1/2}/\mathrm{tr}[\rho M_i],\qquad A_j=N_j \tag{9}$$

yields, via Eq. (5a), an equivalent joint distribution P_{ij}, and hence $I(A|\mathcal{E})=I(M,N|\rho)$.

As a useful example of such mappings between contexts, consider the information bound [19,20]

$$I(M,N|\rho)\leqslant S(\rho_1)+S(\rho_2)-S(\rho) \tag{10}$$

in the joint-measurement context, which holds for the case where M and N refer to observables of two subsystems 1 and 2 of a quantum system [21]. Here ρ_1, ρ_2 denote the reduced density operators $\mathrm{tr}_2[\rho]$,$\mathrm{tr}_1[\rho]$ corresponding to subsystems 1 and 2, respectively, and $S(\sigma)$ denotes the quantum entropy $-\mathrm{tr}[\sigma\log_2\sigma]$ of state σ.

Substituting (8) in (10) one finds, noting $\rho_1=\Sigma_i p_i|\psi_i\rangle\langle\psi_i|$ and $\rho_2=\Sigma_i p_i\rho_i$, that

$$I(A|\mathcal{E})\leqslant S\left(\sum_i p_i\rho_i\right)-\sum_i p_i S(\rho_i) \tag{11}$$

in the communication context. This is just Holevo's theorem [3–5] (see also Sec. III), and its (rather simple) derivation here demonstrates a direct connection with the joint-measurement bound (10).

For the case of measurements on two subsystems as above, ρ_i in the correspondence mapping (9) can be replaced by its partial trace over subsystem 1 without destroying the correspondence. Substituting (9) (with this replacement) in (11) gives the result

$$I(M,N|\rho)\leqslant S(\rho_2)-\sum_i \mathrm{tr}[\rho M_i]S(\mathrm{tr}_1[\rho M_i]/\mathrm{tr}[\rho M_i]) \tag{12a}$$

for the joint-measurement context. Interchanging the labeling of the subsystems in (12a) further gives

$$I(M,N|\rho)\leqslant S(\rho_1)-\sum_j \mathrm{tr}[\rho N_j]S(\mathrm{tr}_2[\rho N_j]/\mathrm{tr}[\rho N_j]). \tag{12b}$$

Note that addition of these two inequalities implies immediately that

$$I(M,N|\rho)\leqslant[S(\rho_1)+S(\rho_2)]/2, \tag{13}$$

previously proved for the case of *pure* ρ [19,20].

Application of Eq. (8) to inequality (12a) simply yields Holevo's theorem (11) once more, while the application of Eq. (8) to inequality (12b) gives the trivial result $I(A|\mathcal{E})\leqslant I(A|\mathcal{E})$. Thus Holevo's theorem is equivalent to inequality (12a), and no stronger than inequality (10).

The correspondence mappings (8) and (9) between the two contexts can, of course, be applied to other circumstances, such as information *lower* bounds [22,23], or even to other measures of correlation such as coincidence rate (see Sec. V). Finally, it should be noted that the mapping (8) is only well-defined for countable numbers of signal states, as only in this case do appropriate $|\psi_i\rangle$ exist. It would be of interest to generalize Eq. (8) to arbitrary distributions of signal states.

III. BOUNDS FROM SOURCE DUALITY

A. Source duality

Holevo's theorem (11) provides a finite quantum bound for mutual information in the communication context. The theorem was proved by Holevo in the case of a finite-dimensional Hilbert space and finite numbers of signal states and measurement outcomes [4], and generalized by Yuen and Ozawa [5] (see also [3,23,24]). The upper bound is a function of the signal ensemble $\mathcal{E}$, and in particular is independent of the measurement A made at the receiver. It provides, for example, a rigorous basis for the result that the optimal information rate which can be transmitted per unit time by a wideband bosonic channel with transmission power P is $(\pi P/3\hbar)^{1/2}/\ln 2$ [3,5].

In practice, the class of measurements available at the receiver is likely to be more restricted in scope than the class of signal ensembles which can be generated at the transmitter [25]. It would therefore be useful to have an upper bound, analogous to Eq. (11), which depends upon the actual measurement to be made at the receiver. Such a bound is obtained in Sec. III B below, based on a source-duality property of mutual information. This property is further exploited to obtain lower bounds for mutual information in Sec. III C.

Now, it is clear from definitions (3) and (4) that mutual information is invariant under the interchanging of sources X and Y. This invariance is indeed expected for any measure of correlation between the sources, and for mutual information reflects the property that the data obtainable about X outputs from Y outputs is equivalent to the data obtainable about Y outputs from X outputs (see also Appendix). In the joint-measurement context of Eqs. (6a) and (6b) such an interchange corresponds to swapping the labels M and N of the measured observables, and is formally trivial.

In contrast, for the communication context of Eqs. (5a) and (5b) such an interchange reverses the fundamentally different roles of the signal ensemble $\mathcal{E}$ and the receiver measurement A: signal states in some sense become measurement outcomes, and vice versa. The multiplicative form of P_{ij} in Eq. (5a) immediately suggests a formal modeling of this reversal, via the duality transformation

$$\mathcal{E}\rightarrow\mathcal{E}', \quad A\rightarrow A', \quad p_i\rightarrow p_j', \quad \rho_i\rightarrow\rho_j', \quad A_j\rightarrow A_i', \tag{14}$$

where

$$p_j'=\mathrm{tr}[\rho_{\mathcal{E}}A_j], \tag{15a}$$

$$\rho_j'=\rho_{\mathcal{E}}^{1/2}A_j\rho_{\mathcal{E}}^{1/2}/\mathrm{tr}[\rho_{\mathcal{E}}A_j], \tag{15b}$$

$$A_i'=p_i\rho_{\mathcal{E}}^{-1/2}\rho_i\rho_{\mathcal{E}}^{1/2}, \tag{15c}$$

and where

$$\rho_{\mathcal{E}}=\sum_i p_i\rho_i=\sum_j p_j'\rho_j'=\rho_{\mathcal{E}'}. \tag{16}$$

Relations (15a)–(15c) are chosen to ensure that

$$P_{ji}'=p_j'\mathrm{tr}[\rho_j'A_i']=p_i\mathrm{tr}[\rho_iA_j]=P_{ij}, \tag{17}$$

and thus from Eq. (4) one has in particular

$$I(A'|\mathcal{E}')=I(A|\mathcal{E}). \tag{18}$$

A restricted form of this duality property is noted in Sec. 3.1 of [14]. Note that the duality mapping relation (14) is well-defined even when the measurement A is continuously valued.

It is convenient to define a ‘‘signal subspace,’’ $H_{\mathcal{E}}$, by the linear span of the eigenvectors corresponding to nonzero eigenvalues of $\rho_{\mathcal{E}}$. As the symmetry of (17) suggests, one has $\mathcal{E}''=\mathcal{E}$ and $A''=A$ on $H_{\mathcal{E}}$. Strictly speaking, the POM $\{A_i'\}$ in Eq. (15c) must in general be supplemented by a projection A_∞ onto the subspace orthogonal to $H_{\mathcal{E}}$ (corresponding to the zero eigenvalues of $\rho_{\mathcal{E}}$), to ensure that $\Sigma_iA_i'=1$. Applications of source duality to mutual information are given in Secs. III B and III C below. Noting property (17), analogous applications may, of course, be made to any figure of merit which is a symmetric function of P_{ij} (see Sec. V).

B. Dual to Holevo’s theorem

Substitution of Eqs. (15a), (15b), (16), and (18) into inequality (11) immediately yields the *measurement-dependent* dual to Holevo’s theorem:

$$I(A|\mathcal{E})\leqslant S(\rho_{\mathcal{E}})-\sum_j \mathrm{tr}[\rho_{\mathcal{E}}A_j]S(\rho_{\mathcal{E}}^{1/2}A_j\rho_{\mathcal{E}}^{1/2}/\mathrm{tr}[\rho_{\mathcal{E}}A_j]), \tag{19}$$

where $\rho_{\mathcal{E}}$ is the ensemble density operator in Eq. (16). This may be compared to the measurement-dependent lower and upper bounds for mutual information given by Scutaru and by Schumacher *et al.*, respectively [24], which are also related to Holevo’s theorem.

Unfortunately, while this bound depends upon the receiver measurement as desired, it is not amenable to straightforward exploitation. For example, if the measurement is complete (i.e., $A_j\equiv|\phi_j\rangle\langle\phi_j|$ for suitable kets $\{|\phi_j\rangle\}$), then each ρ_j' in Eq. (15b) is pure and the summation term in relationship (19) vanishes, leaving a weaker bound in general than relationship (11). The bound is therefore directly applicable only to incomplete measurements (e.g., measuring only the sign of a position observable, inefficient photodetection), and in general must be numerically evaluated. The amount of work required depends on the size of the signal subspace $H_{\mathcal{E}}$. For example, if communication is via two pure states then $H_{\mathcal{E}}$ is two-dimensional, and the entropies in relationship (19) can be determined by diagonalizing 2×2 matrices.

It is worth remarking on the cases in which the dual bound (19) can actually be achieved via some suitable signal ensemble $\mathcal{E}$. Now, Holevo’s bound (11) can be achieved only when the signal states all commute [4]. Thus for the dual bound (19) to be achieved the ρ_j' in Eq. (15b) must commute (on the signal subspace $H_{\mathcal{E}}$ spanned by $\rho_{\mathcal{E}}$), i.e.,

$$A_j\rho_{\mathcal{E}}A_k\equiv A_k\rho_{\mathcal{E}}A_j \quad (\text{on } H_{\mathcal{E}}). \tag{20a}$$

Summing over k in this relation further yields

$$[A_j,\rho_{\mathcal{E}}]\equiv0 \quad (\text{on } H_{\mathcal{E}}), \tag{20b}$$

which from Eq. (20a) implies the $\{A_j\}$ must mutually commute on $H_{\mathcal{E}}$. It follows that the bound (19) cannot be achieved (except for the trivial case of pure $\rho_{\mathcal{E}}$) for complete measurements $\{|\phi_j\rangle\langle\phi_j|\}$ with *nonorthogonal* kets, such as canonical phase detection [26], and ideal heterodyne detection (see Sec. IV D).

It is interesting to note that bound (19) cannot be achieved even in some cases where the $\{A_j\}$ *do* commute. For example, suppose that on $H_{\mathcal{E}}$ one has a continuous POM $\{|x\rangle\langle x|\}$ with non-normalizable orthogonal kets $|x\rangle$. Then there is *no* density operator $\rho_{\mathcal{E}}$ satisfying Eq. (20b) above (noting the constraint $\mathrm{tr}[\rho_{\mathcal{E}}]=1$). This result reflects a basic asymmetry between states and measurements in standard quantum mechanics, in that ‘‘state’’ kets must be normalizable whereas as ‘‘measurement’’ kets need not be.

An example of this asymmetry is provided by optical homodyne detection of a single-mode optical field (or equivalently, a position measurement on a harmonic oscillator), for the case where $\rho_{\mathcal{E}}$ spans the whole Hilbert space. Choosing in particular $\rho_{\mathcal{E}}$ to be a thermal state, with number-state expansion

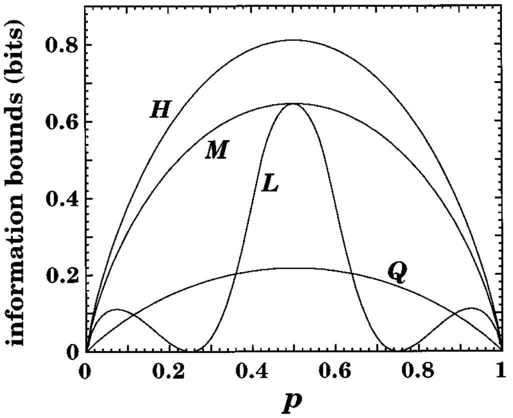

FIG. 1. The ''dual'' lower bound L in Eq. (24), the ''subentropy'' lower bound Q in [22], the ''binary-channel'' lower bound M in [23], and the Holevo upper bound in relation (11); plotted as a function of p for a signal ensemble $\mathcal{E}$ corresponding to transmission of two pure states $|\phi\rangle$, $|\psi\rangle$ with prior probabilities p, $1-p$ and overlap $|\langle\psi|\phi\rangle|=1/2$. These quantities bound the maximum information which may be obtained per signal at the receiver. Equation (26) implies that L and M are equal for any overlap in the case $p=1/2$.

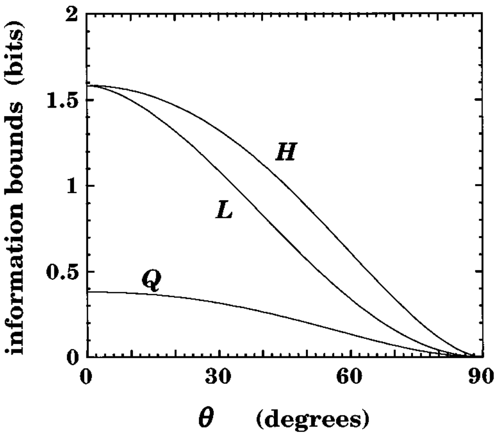

FIG. 2. The ''dual'' lower bound L in Eq. (24), the ''subentropy'' lower bound Q in [22], and the Holevo upper bound H in relation (11); plotted as a function of θ for the signal ensemble corresponding to transmission of the three states $|0\rangle$, $\cos\theta|1\rangle+\sin\theta|0\rangle$, $\cos\theta|-1\rangle-\sin\theta|0\rangle$ with equal prior probabilities of 1/3; where $|0\rangle$, $|1\rangle$, $|-1\rangle$ are mutually orthogonal. Note that L and H constrain the maximum information which can be gained per signal to within a range of at most 0.3 bits.

$$\rho_{\mathcal{E}}=(n_s+1)^{-1}\sum_n (1+1/n_s)^{-n}|n\rangle\langle n| \tag{21}$$

and average photon number $n_s>0$, the upper bound

$$I(X|\mathcal{E})\leqslant S(\rho_{\mathcal{E}})=\log_2(n_s+1)+n_s\log_2(1+1/n_s) \tag{22}$$

following from relationship (19) is *not* achievable. Indeed, as shown in [12] using the information exclusion principle, the stronger upper bound $\log_2(1+2n_s)$ holds for this case.

C. Lower bounds

Source duality does not only provide dual bounds to known bounds. The dual observable A' and dual ensemble $\mathcal{E}'$ can themselves be used as estimates for optimal observables and ensembles, respectively, to obtain useful (although generally nonoptimal) lower bounds. To demonstrate this, let $I_{\max}(\mathcal{E})$ denote the maximum information which can be obtained by measurement on a signal ensemble $\mathcal{E}$, i.e.,

$$I_{\max}(\mathcal{E})=\sup_A\{I(A|\mathcal{E})\}. \tag{23}$$

Then one immediately has the lower bound

$$I_{\max}(\mathcal{E})\geqslant L(\mathcal{E}):=I(A'|\mathcal{E})=\sum_{i,j} p_ip_jX_{ij}\log_2X_{ij} \tag{24}$$

from Eqs. (4), (5), and (15b), where

$$X_{ij}=\mathrm{tr}[\rho_i\rho_{\mathcal{E}}^{-1/2}\rho_j\rho_{\mathcal{E}}^{-1/2}]. \tag{25}$$

The lower bound (24) is in fact optimal in the case that the signal states ρ_i are orthogonal (i.e., $\rho_j\rho_j=0$ for $i\neq j$), as it reduces to the Holevo upper bound (11) (with corresponding mutual information $-\Sigma_ip_i\log_2p_i$). Its degree of usefulness is variable, however, with best results apparently for the case of signal states with equal prior probabilities.

In Fig. 1 the bound $L(\mathcal{E})$ is plotted as a function of p for binary communication based on the transmission of two pure states $|\psi\rangle$, $|\phi\rangle$ with prior probabilities p and $1-p$, respectively, and with a significant overlap $|\langle\psi|\phi\rangle|=1/2$. Also plotted is the corresponding ''subentropy'' lower bound Q from [22], and the ''binary-channel'' lower bound M from [23], as well as the Holevo upper bound H from relationship (11) above. It is seen that L outperforms Q for most values of p, but only reaches M for the case $p=1-p=1/2$.

Algebraic calculation shows that in fact L and M are *always* equal for the case of two pure states $|\psi\rangle$, $|\phi\rangle$ transmitted with equal prior probabilities of 1/2, with

$$L(\mathcal{E})=M(\mathcal{E})=(x_+\log_2x_++x_-\log_2x_-)/2, \tag{26}$$

where $x_\pm=1\pm(1-|\langle\psi|\phi\rangle|^2)^{1/2}$.

In Fig. 2 $L(\mathcal{E})$ is plotted as a function of θ for the transmission of *three* pure states $|0\rangle$, $\cos\theta|1\rangle+\sin\theta|0\rangle$, $\cos\theta|-1\rangle-\sin\theta|0\rangle$ with equal prior probabilities 1/3, 1/3, 1/3, where $|0\rangle$, $|1\rangle$, $|-1\rangle$ are mutually orthogonal. Q and H are again plotted (there is *no* corresponding bound M [23]). It is seen that $L(\mathcal{E})$ places a relatively tight bound on $I_{\max}(\mathcal{E})$, and is optimal when the signal states are orthogonal ($\theta=0$).

As an infinite-dimensional example, consider the case of single-mode optical communication where a Gaussian ensemble of Glauber coherent states is transmitted, i.e., the coherent state

$$|\alpha\rangle=\exp(-|\alpha|^2/2)\sum_n (n!)^{-1/2}\alpha^n|n\rangle \tag{27}$$

is transmitted with probability

$$p(\alpha)=(\pi n_s)^{-1}\exp(-|\alpha|^2/n_s), \tag{28}$$

where n_s denotes the average photon number per signal [3]. In this case $\rho_{\mathcal{E}}$ is given by Eq. (21) above, and $L(\mathcal{E})$ may be calculated from Eqs. (24), (25), (27), and (28) as

$$L(\mathcal{E})=\log_2(n_s+1). \qquad (29)$$

This is close to the *upper* bound $S(\rho_{\mathcal{E}})$ in Eq. (22) [following from relationships (11) and (21)], and thus provides an excellent lower bound. The dual measurement A' in this case is just a rescaled heterodyne measurement (see Sec. IV). Note that the corresponding subentropy lower bound Q cannot exceed 0.609 95 bits [22], and so is far less restrictive than Eq. (29) in general. Equations (26) and (29) and Fig. 2 suggest that L is most useful when the entropy of $\rho_{\mathcal{E}}$ is as large as possible (e.g., equal prior probabilities, orthogonal signal states).

Finally, a *measurement-dependent* lower bound may also be derived from source duality. In particular, let $I^*_{\max}(A,\rho)$ denote the maximum information which may be gained via measurement A, on the class of signal ensembles with ensemble density operator ρ. Thus

$$I^*_{\max}(A,\rho)=\sup_{\mathcal{E}:\rho_{\mathcal{E}}=\rho}\{I(A|\mathcal{E})\}. \qquad (30)$$

One then has, using relationships (4), (5), and (15), the lower bound

$$\begin{aligned} I^*_{\max}(A,\rho)&\geqslant L^*(A,\rho):=I(A|\mathcal{E}')\\ &=-2\sum_j \mathrm{tr}[\rho A_j]\log_2\mathrm{tr}[\rho A_j]+\sum_{i,j} Y_{ij}\log_2 Y_{ij}, \end{aligned} \qquad (31)$$

where Y_{ij} is the joint probability distribution,

$$Y_{ij}=\mathrm{tr}[\rho^{1/2}A_i\rho^{1/2}A_j]. \qquad (32)$$

As an example, let A represent ideal heterodyne detection of a single-mode optical field [3,27], with corresponding POM $\{\pi^{-1}|\alpha\rangle\langle\alpha|\}$, where $|\alpha\rangle$ is the Glauber coherent state in Eq. (27) and α ranges over the complex plane. If ρ is chosen to be the thermal state in Eq. (21), one finds the lower bound

$$L^*(A,\rho)=\log_2(n_s+1) \qquad (33)$$

from Eqs. (31) and (32). This is, in fact, the *optimal* bound [corresponding to transmission of a Gaussian distribution of coherent states as per Eqs. (27) and (28) above], i.e., $L^*(A,\rho)=I^*_{\max}(A,\rho)$, as will be shown in Sec. IV D.

IV. INFORMATION EXCLUSION

A. Motivation

One intuitively expects that the better some quantum observable is at extracting information in a given scenario, the worse a complementary observable will perform in that scenario. In particular, it should be possible to increase the information gain corresponding to measurement of the first observable only at the expense of decreasing the information gain corresponding to the second observable, and vice versa. This idea has recently been formulated as a rigorous ''information exclusion'' principle for quantum observables [12], which will be further generalized and exploited here to obtain strong bounds on mutual information in the communication and joint-measurement contexts.

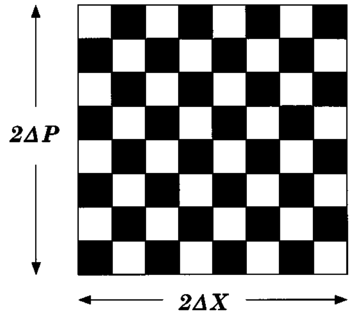

FIG. 3. A semiclassical ''quantum chessboard'' as discussed in Sec. IV A, corresponding to a rectangular region of phase space with position and momentum half-widths ΔX and ΔP, and divided into $N_1=8$ columns and $N_2=8$ rows. The minimum ''square'' size a system can occupy is $2\hbar$ from the uncertainty principle, and hence the sum of the position and momentum information gains, corresponding, respectively, to measurements of which column and row is occupied, is bounded as per Eq. (35) of the text.

As a semiclassical example of information exclusion in the communication context, consider a ''quantum chessboard'' as discussed in the Introduction, corresponding to a rectangular region of a two-dimensional phase space with position and momentum half-widths ΔX and ΔP (see Fig. 3). If the region is divided into N_1 columns and N_2 rows as per Fig. 3, and signals are generated by placing a system at random in one of the N_1N_2 ''squares'' formed thereby, then the information gains corresponding to measurements of position and momentum, respectively, are given by

$$I(X|\mathcal{E})=\log_2 N_1, \quad I(P|\mathcal{E})=\log_2 N_2. \qquad (34)$$

But from the uncertainty principle each signal must occupy an area of at least $2\hbar$, implying the constraint $2\hbar N_1N_2\leqslant 4\Delta X\Delta P$. Combining this constraint with relations (34) gives the *information exclusion relation*

$$I(X|\mathcal{E})+I(P|\mathcal{E})\leqslant\log_2 2\Delta X\Delta P/\hbar. \qquad (35)$$

Thus for a fixed phase space area the position information can be increased only at the expense of the momentum information and vice versa.

Surprisingly, inequality (35) for position and momentum information can in fact be rigorously derived for *all* signal ensembles $\mathcal{E}$, providing that ΔX and ΔP are interpreted as the root-mean-square position and momentum uncertainties of the ensemble density operator $\rho_{\mathcal{E}}$ in Eq. (16) [12] (see also Sec. IV C below). Since information is non-negative the rigorous inequality immediately implies (and hence is stronger than) the Heisenberg uncertainty relation for X and P. Further, it is sufficiently strong to obtain the *tight* bound, $\log_2(1+2n_s)$, on the information which may be gained by homodyne detection on an ensemble of single-mode fields with average photon number n_s [12] (see also Sec. IV C below).

The strategy used in [12] to obtain information exclusion relations from entropic uncertainty relations will be used here also. In particular, if the entropies $S(A|\rho)$, $S(B|\rho)$ of the measurement distributions of A and B for any state ρ satisfy

$$\Lambda \leqslant S(A|\rho)+S(B|\rho)\leqslant U(\rho), \tag{36a}$$

then one has a corresponding exclusion relation

$$I(A|\mathcal{E})+I(B|\mathcal{E})\leqslant U(\rho_{\mathcal{E}})-\Lambda. \tag{36b}$$

In the following subsections the information exclusion principle will be generalized and applied in a number of ways, including a tight bound for the information which can be gained via (noisy) heterodyne detection, and new upper bounds for mutual information. Results will typically be presented in the communication context, in the form

$$I(A|\mathcal{E})+I(B|\mathcal{E})\leqslant J(A,B,\rho_{\mathcal{E}}). \tag{37}$$

However, in Sec. IV F below extensions are given to the joint-measurement context (via context mappings); bounds on the sum of information gains for a single observable and two different ensembles (via source duality); and to an inequality for quantum eavesdropping.

B. Discrete observables

It was shown in [12] that two observables A and B corresponding to (possibly degenerate) *Hermitian operators* on an N-dimensional Hilbert space satisfy the information exclusion relation

$$I(A|\mathcal{E})+I(B|\mathcal{E})\leqslant 2\log_2 Nc, \tag{38}$$

where c denotes the maximum overlap $|\langle a|b\rangle|$ of eigenstates of A and B. For $A=B$ one has $c=1$, and hence that $I(A|\mathcal{E})\leqslant\log_2 N$ [this also follows trivially from the Holevo bound (11), noting $S(\rho_{\mathcal{E}})\leqslant\log_2 N$]. In contrast, for *complementary* A and B (where the distribution of B is uniform for each eigenstate of A and vice versa [28]), one has $c=N^{-1/2}$ and hence the strong bound $I(A|\mathcal{E})+I(B|\mathcal{E})\leqslant\log_2 N$.

Thus $I(A|\mathcal{E})$ can *only* reach its maximum of $\log_2 N$ when the complementary observable B carries *no* information. For nondegenerate A this maximum is obviously attained by transmitting eigenstates of A with equal prior probabilities $1/N$. However, it is not so obvious what the optimal signal states are if A has only non-normalizable eigenkets (e.g., position and momentum), and/or does not correspond to a Hermitian operator (e.g., canonical phase detection and heterodyne detection). The information exclusion principle will be shown sufficiently strong to provide solutions in some instances, in Secs. IV C and IV D.

Relation (38) will be generalized here to

$$I(A|\mathcal{E})+I(B|\mathcal{E})\leqslant\log_2\left(\sum_j d(A_j)\sum_k d(B_k)K_{AB}\right) \tag{39}$$

for observables A and B corresponding to POM's $\{A_j\}$ and $\{B_k\}$, respectively, where $d(E)$ is defined to be the ''degeneracy'' of probability operator E, i.e., the number of nonzero eigenvalues of E, and

$$K_{AB}=\min\{\max_{j,k,r,s}\{|\langle a_{jr}|b_{ks}\rangle|^2\}\}, \tag{40}$$

where the minimum is taken over all $\{|a_{jr}\rangle\}$ and $\{|b_{ks}\rangle\}$ such that $\Sigma_r|a_{jr}\rangle\langle a_{jr}|=A_j$ and $\Sigma_s|b_{ks}\rangle\langle b_{ks}|=B_k$ are orthogonal decompositions of operators A_j and B_k, respectively. Note that for nondegenerate Hermitian operators on an N-dimensional Hilbert space the $\{A_j\}$ and $\{B_k\}$ correspond to sets of one-dimensional orthogonal projections which sum to the identity operator, and hence relation (39) reduces to relation (38), with $\Sigma_j d(A_j)=\Sigma_k d(B_k)=N$ and $K_{AB}=c^2$.

To prove relation (39) suppose first that the POM's $\{A_j\}$ and $\{B_k\}$ are *complete*, i.e., $A_j\equiv|a_j\rangle\langle a_j|$ and $B_k\equiv|b_k\rangle\langle b_k|$. Although $\{|a_j\rangle\}$ and $\{|b_k\rangle\}$ are not necessarily orthonormal sets, the entropic uncertainty relation

$$S(A||\psi\rangle\langle\psi|)+S(B||\psi\rangle\langle\psi|)\geqslant-2\log_2\max_{j,k}\{|\langle a_j|b_k\rangle|\}$$

may be derived from Riesz's theorem exactly as per Eq. (6) of [29] for an arbitrary pure state $|\psi\rangle\langle\psi|$, where $S(A|\rho)$ denotes the entropy of the measurement distribution of A for state ρ. Since $S(A|\rho)$ is concave with respect to ρ [2] one has for general ρ, noting definition (40), that

$$S(A|\rho)+S(B|\rho)\geqslant-\log_2 K_{AB}. \tag{41}$$

Moreover, the maximum entropy of a distribution over M outcomes is trivially bounded by $\log_2 M$. Hence, since $d(A_j)=d(B_k)=1$ for complete observables, one has

$$S(A|\rho)+S(B|\rho)\leqslant\log_2\sum_j d(A_j)+\log_2\sum_k d(B_k). \tag{42}$$

The exclusion relation (39) immediately follows for complete observables from relations (41) and (42), using relations (36a) and (36b).

For the case of *incomplete* observables A and B one may define corresponding complete observables A^* and B^* by the POM's $\{|a_{jr}\rangle\langle a_{jr}|\}$ and $\{|b_{ks}\rangle\langle b_{ks}|\}$ which achieve the minimum in Eq. (40). Thus the probability distributions p_j, q_k of A and B for some state may be obtained by summing over the ''degeneracy'' indices r and s of the distributions p_{jr}, q_{ks} of A^* and B^*, respectively. Hence measurement of A or B cannot yield more information than measurement of A^* or B^*, respectively (Lemma 2 in [13]), and one has

$$\begin{aligned}I(A|\mathcal{E})+I(B|\mathcal{E})&\leqslant I(A^*|\mathcal{E})+I(B^*|\mathcal{E})\\&\leqslant\log_2\Bigg(\sum_{j,r}d(|a_{jr}\rangle\langle a_{jr}|)\\&\qquad\times\sum_{k,s}d(b_{ks}\rangle\langle b_{ks}|)K_{A^*B^*}\Bigg).\end{aligned}$$

But $d(A_j)=\Sigma_r d(|a_{jr}\rangle\langle a_{jr}|)$, $d(B_k)=\Sigma_s d(|b_{ks}\rangle\langle b_{ks}|)$, and $K_{A^*B^*}=K_{AB}$ by construction, and the exclusion relation (39) follows.

It should be possible to improve significantly on the entropic uncertainty relation (41) for the case of non-Hermitian observables, and hence upon the corresponding information exclusion relation (39). For example, consider a ''tetrahedral'' measurement M on a spin-1/2 particle, with POM $\{1/2|\boldsymbol{m}_j\rangle\langle\boldsymbol{m}_j|\}$, where the $|\boldsymbol{m}_j\rangle$ are eigenstates of spin-up in four unit directions $\boldsymbol{m}_j$ which form the vertices of a regular tetrahedron. Choosing $A=B=M$ in relations (39) and (40) yields $K_{MM}=1/4$ and hence the trivial result $I(M|\mathcal{E})\leqslant\log_2 2$. However, a variational calculation over pure states shows that $S(M|\rho)\geqslant\log_2 3$ [thus improving on relation (41) which only gives $S(M|\rho)\geqslant\log_2 2$], and hence via relations (36a), (36b), and (42) one finds the stronger result

$$I(M|\mathcal{E})\leqslant\log_2 4/3. \tag{43}$$

Note that the bound in inequality (43) is optimal, being achieved by transmission of the four signal states $|-\boldsymbol{m}_j\rangle$ with equal prior probabilities [13].

C. Position and momentum

If $\boldsymbol{X}$ and $\boldsymbol{P}$ are n-vectors denoting the position and momentum observables of a quantum system, then one has the corresponding information exclusion relations [12]

$$I(\boldsymbol{X}|\mathcal{E})+I(\boldsymbol{P}|\mathcal{E})\leqslant\tfrac{1}{2}\ \log_2\ \det[\operatorname{cov}(\boldsymbol{X})\operatorname{cov}(\boldsymbol{P})/(\hbar/2)^2] \tag{44a}$$

$$\leqslant\sum_m\ \log_2\ 2(\Delta X_m)_{\mathcal{E}}(\Delta P_m)_{\mathcal{E}}/\hbar, \tag{44b}$$

where

$$\begin{aligned}\operatorname{cov}(\boldsymbol{V}):&=\operatorname{tr}[\rho_{\mathcal{E}}\boldsymbol{V}\boldsymbol{V}^T]-\operatorname{tr}[\rho_{\mathcal{E}}\boldsymbol{V}]\operatorname{tr}[\rho_{\mathcal{E}}\boldsymbol{V}^T]\\ &=\langle(\boldsymbol{V}-\langle\boldsymbol{V}\rangle_{\mathcal{E}})(\boldsymbol{V}^T-\langle\boldsymbol{V}^T\rangle_{\mathcal{E}})\rangle_{\mathcal{E}}\end{aligned} \tag{45}$$

denotes the covariance matrix of n-vector $\boldsymbol{V}$, and $(\Delta V_m)_{\mathcal{E}}$ denotes the root-mean-square uncertainty of component V_m of $\boldsymbol{V}$ with respect to state $\rho_{\mathcal{E}}$. The two middle inequalities in this chain were noted without proof in [12] and will be derived further below.

Inequalities (44a) and (44b) collapse to the exclusion relation (35) for the case $n=1$. It is of interest to note that this case may immediately be generalized to give an exclusion relation for two arbitrary quadratures of a single-mode optical field. In particular, for a harmonic oscillator with annihation operator a define the quadrature operator [3]

$$X_\theta:=(ae^{i\theta}+a^\dagger e^{-i\theta})/2. \tag{46}$$

Then $[X_\theta,X_\phi]=(i/2)\sin(\theta-\phi)$ and so X_θ and X_ϕ are formally related exactly as per position and momentum, identifying $\hbar$ with $|\sin(\theta-\phi)|/2$. Thus the generalized exclusion relation

$$I(X_\theta|\mathcal{E})+I(X_\phi|\mathcal{E})\leqslant\log_2[4(\Delta X_\theta)_{\mathcal{E}}(\Delta X_\phi)_{\mathcal{E}}/|\sin(\theta-\phi)|] \tag{47}$$

follows directly from relation (44b) with $n=1$.

For an average photon number per signal of $n_s=\operatorname{tr}[\rho_{\mathcal{E}}a^\dagger a]$ one has

$$\begin{aligned}(\Delta X_\theta)_{\mathcal{E}}(\Delta X_\phi)_{\mathcal{E}}&\leqslant(\langle X_\theta^2\rangle\langle X_\phi^2\rangle)^{1/2}\leqslant\langle X_\theta^2+X_\phi^2\rangle/2\\ &=\langle X_\alpha^2+X_\beta^2\rangle/2+\cos(\theta-\phi)\langle X_\alpha^2-X_\beta^2\rangle/2\\ &\leqslant[1+|\cos(\theta-\phi)|]\langle X_\alpha^2+X_\beta^2\rangle/2\\ &=[1+|\cos(\theta-\phi)|](2n_s+1)/4,\end{aligned}$$

where $\alpha=(\theta+\phi)/2$ and $\beta=(\theta+\phi+\pi)/2$, and hence from Eq. (47) that

$$I(X_\theta|\mathcal{E})+I(X_\phi|\mathcal{E})\leqslant\log_2\{(2n_s+1)|\cot[(\theta-\phi)/2]|\}. \tag{48}$$

Choosing $\theta-\phi=\pi/2$ in relation (48), one recovers the result that the information gain from homodyne detection has the *strong* bound $\log_2(2n_s+1)$ [12], which may be compared with the weaker bound (22) following from Holevo's theorem. A similar bound for heterodyne detection will be obtained in Sec. IV D from the $n=2$ case.

To prove inequality (44a), note from (45) that $\operatorname{cov}(\boldsymbol{V})$ is a real symmetric positive-semidefinite $n\times n$ matrix, and hence may be diagonalized by some $n\times n$ rotation matrix R with $R^TR=I$ and $\det(R)=1$. Defining the observable $\boldsymbol{W}=R\boldsymbol{V}$ one then has $R\operatorname{cov}(\boldsymbol{V})R^T=\operatorname{cov}(\boldsymbol{W})$ where $\operatorname{cov}(\boldsymbol{W})$ is diagonal, and hence that

$$\det[\operatorname{cov}(\boldsymbol{V})]=\det[\operatorname{cov}(\boldsymbol{W})]=\operatorname{var}_{\mathcal{E}}(W_1)\cdots\operatorname{var}_{\mathcal{E}}(W_n), \tag{49}$$

where $\operatorname{var}_{\mathcal{E}}$ denotes variance with respect to $\rho_{\mathcal{E}}$. Moreover, the entropy of $\boldsymbol{V}$ is invariant under rotation [16], and hence via Eq. (20) of [12] one has

$$\begin{aligned}S(\boldsymbol{V}|\rho_{\mathcal{E}})=S(\boldsymbol{W}|\rho_{\mathcal{E}})&\leqslant(n/2)\log_2 2\pi e+(1/2)\sum_m\ \log_2\operatorname{var}_{\mathcal{E}}(W_m)\\ &=(n/2)\log_2 2\pi e+(1/2)\log_2\det[\operatorname{cov}(\boldsymbol{V})].\end{aligned} \tag{50}$$

Summing the inequalities obtained from (50) with $\boldsymbol{V}=\boldsymbol{X}$ and $\boldsymbol{V}=\boldsymbol{P}$, and noting the entropic uncertainty relation [30]

$$S(\boldsymbol{X}|\rho)+S(\boldsymbol{P}|\rho)\geqslant n\ \log_2\pi e\hbar, \tag{51}$$

the exclusion relation (44a) follows using relations (36a) and (36b).

Finally, to obtain inequality (44b) note that the determinant of a non-negative real symmetric matrix is no greater than the product of its diagonal terms [31]. Hence

$$\det[\operatorname{cov}(\boldsymbol{V})]\leqslant\operatorname{var}_{\mathcal{E}}(V_1)\cdots\operatorname{var}_{\mathcal{E}}(V_n), \tag{52}$$

and relation (44b) follows immediately from relation (44a).

D. Heterodyne detection

It has been shown that under ideal conditions, optical heterodyne detection measures the (commuting) real and imaginary parts of the operator $a+b^\dagger$, where a and b represent the annihilation operators for the signal and image-band fields, respectively [3,27]. Heterodyne detection thus estimates the complex amplitude of the signal field a, subject to image-band noise. The effect of the latter noise is minimized by choosing the image-bound field to be in a vacuum state, and for this case the measurement statistics are given by the Q function of the signal state ρ [3,27],

$$Q(\alpha|\rho)=\pi^{-1}\langle\alpha|\rho|\alpha\rangle, \tag{53}$$

where α ranges over the complex plane and $|\alpha\rangle$ is defined in (27).

To bound the information which may be gained by heterodyne detection, let X_θ, Y_θ denote the quadrature observables corresponding to a and b, respectively [see Eq. (46)], and introduce the observables

$$X = X_0 + Y_0, \quad P = X_{\pi/2} + Y_{\pi/2}, \quad X' = X_{\pi/2} - Y_{\pi/2},$$
$$P' = X_0 - Y_0. \tag{54}$$

Thus $[X,P]=[X',P']=i$, with all other commutators vanishing, and so from relation (44b) one has, with $n=2$ and $\hbar \equiv 1$,

$$I(X,X'|\mathcal{E}) + I(P,P'|\mathcal{E}) \leqslant (1/2)\log_2[4\,\mathrm{var}_{\mathcal{E}}(X)\mathrm{var}_{\mathcal{E}}(P)] + (1/2)\log_2[4\,\mathrm{var}_{\mathcal{E}}(X')\mathrm{var}_{\mathcal{E}}(P')]. \tag{55}$$

For a vacuum image-band field the ensemble density operator has the form $\rho_{\mathcal{E}} \otimes |0\rangle\langle 0|$, which substituted into relation (55) yields

$$I(X,X'|\mathcal{E}) + I(P,P'|\mathcal{E}) \leqslant \log_2\{4[\mathrm{var}_{\mathcal{E}}(X_0)+1/4] \times [\mathrm{var}_{\mathcal{E}}(X_{\pi/2})+1/4]\}.$$

Under the constraint $n_s = \mathrm{tr}[\rho_{\mathcal{E}} a^\dagger a]$ the argument of the logarithm is maximized by the choice $\mathrm{var}_{\mathcal{E}}(X_0) = \mathrm{var}_{\mathcal{E}}(X_{\pi/2}) = (2n_s+1)/4$ [cf. the derivation of relation (48)], and hence

$$I(X,X'|\mathcal{E}) + I(P,P'|\mathcal{E}) \leqslant 2\log_2(n_s+1). \tag{56}$$

Now, from relations (46) and (54), $X+iX' = a+b^\dagger$, and so the first term in relation (56) is in fact the information gained by heterodyne detection, $I(\mathrm{het}|\mathcal{E})$, say. Moreover, if U denotes the antiunitary transformation on the image-band field which maps b to $-b$, then the joint distribution of P and P' for any signal state $\rho \otimes |0\rangle\langle 0|$ must be invariant under $P \to UPU^\dagger = X'$, $P' \to UP'U^\dagger = X$, $\rho \otimes |0\rangle\langle 0| \to U|\rho \otimes |0\rangle\langle 0|U^\dagger = \rho \otimes |0\rangle\langle 0|$. Hence, for a vacuum image-band field, a joint measurement of P and P' is equivalent to a joint measurement of X and X', implying the second term in relation (56) is also equal to $I(\mathrm{het}|\mathcal{E})$. Thus relation (56) reduces to the bound

$$I(\mathrm{het}|\mathcal{E}) \leqslant \log_2(1+n_s). \tag{57}$$

It follows immediately from the lower bound (33) that inequality (57) is strong, i.e., the best possible bound depending on the average photon number per signal. The bound is achieved by transmission of a Gaussian ensemble of Glauber coherent states, as per Eqs. (27) and (28) [3]. In contrast, the Holevo bounds (11) and (19) only yield the weaker bound $S(\rho_{\mathcal{E}})$ in relation (22).

Inequality (57) may be generalized to the strong bound

$$I(\mathrm{het}|\mathcal{E},n_\gamma) \leqslant \log_2[1+n_s/(n_\gamma+1)] \tag{58}$$

when Gaussian noise of variance n_γ is added to the signal states. This bound is also achieved by transmission of Glauber states as per Eqs. (27) and (28) [32]. An analogous bound for noisy homodyne detection is given in [12] (it is not known if the latter is strong for $n_\gamma > 0$).

Gaussian noise models the effects of thermal noise [32], linear amplification and/or attenuation noise [33], and inefficient photodetection [34] on the statistics of heterodyne detection; with corresponding noise variances

$$n_\gamma^{\mathrm{thermal}} = [\exp(\hbar\omega/kT)-1]^{-1}, \tag{59a}$$

$$n_\gamma^{\mathrm{amp./att.}} = \{1-\exp[-2A(N_2-N_1)]\}/(1-N_1/N_2), \tag{59b}$$

$$n_\gamma^{\mathrm{ineff.\ det.}} = (1-\eta)/\eta, \tag{59c}$$

respectively, where k, ω, and T are Boltzmann's constant, frequency, and absolute temperature; N_1 (N_2) the number of excited (unexcited) amplifier atoms, and A is an amplification constant [33]; and η denotes detector inefficiency [34].

Equations (59) are obtained by studying the effects of the respective noise processes on the measurement statistics (53) for heterodyne detection. These effects break down into a rescaling of the statistics, which does not affect information gain [16], and an effective Gaussian noise term as per either Eqs. (13) or (26b) of [32] (general effects of Gaussian noise on quantum optical communication are studied in [32]).

To derive bound (58), suppose that Gaussian noise of variance n_γ is present. The corresponding measurement statistics for state ρ are then given by the Q function [Eq. (26a) of [32]]

$$Q(\alpha|\rho,n_\gamma) = Q(\alpha|\Gamma(\rho)) = \pi^{-1}\langle\alpha|\Gamma(\rho)|\alpha\rangle, \tag{60}$$

where

$$\Gamma(\rho) = (\pi n_\gamma)^{-1}\int d^2\beta\, e^{-|\beta|^2/n_\gamma} e^{\beta a^\dagger - \beta^* a} \rho e^{\beta^* a - \beta a^\dagger} \tag{61}$$

and integration is over the complex plane. Moreover, from relations (36a) and (36b) with $A=B=\mathrm{het}$ one has

$$I(\mathrm{het}|\mathcal{E},n_\gamma) \leqslant S_{\max}(\mathrm{het}|\Gamma(\rho_{\mathcal{E}})) - S_{\min}(\mathrm{het}|\Gamma(\rho)) \tag{62}$$

where the maximum will be evaluated over all signal ensembles with average photon number $n_s = \mathrm{tr}[\rho_{\mathcal{E}} a^\dagger a]$, and the minimum over all states ρ.

Now, from Eq. (60)

$$S(\mathrm{het}|\Gamma(\rho)) = -\int d^2\alpha\, Q(\alpha|\Gamma(\rho))\log_2 Q(\alpha|\Gamma(\rho)), \tag{63}$$

$$\int d^2\alpha |\alpha|^2 Q(\alpha|\Gamma(\rho)) = \mathrm{tr}\left(\Gamma(\rho) a \pi^{-1}\int d^2\alpha |\alpha\rangle\langle\alpha| a^\dagger\right) = \mathrm{tr}[\Gamma(\rho)aa^\dagger] = \mathrm{tr}[\rho a^\dagger a] + n_\gamma + 1, \tag{64}$$

where the relation $a|\alpha\rangle = \alpha|\alpha\rangle$ has been used, and the last equality follows from Eq. (12) of [32]. Hence the first term on the right-hand-side of relation (62) is equivalent to maximizing the entropy (63) of a two-dimensional probability distribution $Q(\alpha|\Gamma(\rho_{\mathcal{E}}))$, subject to the quadratic constraint (64). This is a well known variational problem in classical information theory [2], with solution

$$S_{\max}(\mathrm{het}|\Gamma(\rho_{\mathcal{E}}))=\log_2 \pi e(n_s+n_\gamma+1). \tag{65}$$

Further, to evaluate the last term in (62), note that concavity of the entropy functional [2] implies that only pure states $|\psi\rangle\langle\psi|$ need be considered. Defining $K = S(\mathrm{het}|\Gamma(|\psi\rangle\langle\psi|))+\lambda\langle\psi|\psi\rangle$, where λ is a Lagrange multiplier constraining the normalization of $|\psi\rangle$, and noting from (61) [see Eq. (19c) of [32]] that

$$\langle\alpha|\Gamma(|\psi\rangle\langle\psi|)|\alpha\rangle=\langle\psi|\Gamma(|\alpha\rangle\langle\alpha|)|\psi\rangle,$$

the variational equation $\partial K/\partial\langle\psi|=0$ can be written as

$$\left(\pi^{-1}\int d^2\alpha\Gamma(|\alpha\rangle\langle\alpha|)\log_2 Q(\alpha|\Gamma(|\psi\rangle\langle\psi|)\right)|\psi\rangle=\lambda|\psi\rangle, \tag{66}$$

which is satisfied by the Glauber coherent states in Eq. (27).

In particular, if $|\psi\rangle$ is the vacuum state then from Eqs. (60) and (61) [or Eq. (27) of [32]],

$$Q(\alpha|\Gamma(|0\rangle\langle 0|)=\pi^{-1}(n_\gamma+1)^{-1}\exp[-|\alpha|^2/(n_\gamma+1)], \tag{67}$$

and the integral in Eq. (66) has the form

$$\begin{aligned}\Gamma&\left(\pi^{-1}\int d^2\alpha|\alpha\rangle\langle\alpha|[C_1+C_2|\alpha|^2]\right)\\ &=C_1\Gamma(1)+C_2\Gamma(aa^\dagger)=C_1+C_2(a^\dagger a+n_\gamma+1)\end{aligned} \tag{68}$$

where C_1, C_2 are constants and the last equality follows from Eqs. (12) and (19c) of [32]. Thus Eq. (66) is satisfied, with $\lambda=C_1+C_2(n_\gamma+1)$, and from Eqs. (63) and (67) one finds

$$S_{\min}(\mathrm{het}|\Gamma(\rho))=\log_2 \pi e(n_\gamma+1). \tag{69}$$

The upper bound (58) follows immediately from Eqs. (62), (65), and (69).

E. Upper bounds

It will be now be shown how the information exclusion principle may be combined with source duality to obtain general upper bounds for quantum information, which are related to the lower bounds $L(\mathcal{E})$ and $L^*(A,\rho)$ in Sec. III C.

First, for an exclusion relation of the general form given in relation (37), choose B to be the dual observable A' in Eq. (15c). Then relation (37) may be rewritten in the form

$$I(A|E)\leqslant J(A,A',\rho_{\mathcal{E}})-L(\mathcal{E}), \tag{70}$$

where $L(\mathcal{E})$ is defined by Eqs. (24) and (25). Second, choosing B again to be A', and replacing $\mathcal{E}$ by the dual ensemble $\mathcal{E}'$ defined in Eqs. (15a) and (15b), one has from relation (37) that

$$I(A|\mathcal{E})\leqslant J(A,A',\rho_{\mathcal{E}})-L^*(A,\rho_{\mathcal{E}}), \tag{71}$$

where $L^*(A,\rho)$ is defined by Eqs. (31) and (32), and the duality relations (16) and (18) have been used.

Upper bounds (70) and (71) are typically stronger than those provided by Holevo's theorem (11) and its dual (19). For example, let A correspond to some complete POM $\{|a_j\rangle\langle a_j|\}$ and $\mathcal{E}$ to a mixture $\{|\psi_i\rangle\langle\psi_i|,p_i\}$ of pure states. If $N_{\mathcal{E}}$ and N_A denote the number of signal states and possible measurement outcomes, respectively, then, using relations (24), (25), (31), (32), (39), and (40), the bounds

$$\begin{aligned}I(A|\mathcal{E})\leqslant&\log_2[N_A N_{\mathcal{E}}\max_{i,j}\{p_i|\langle a_j|\rho_{\mathcal{E}}^{-1/2}|\psi_i\rangle|^2\}]\\ &+\sum_{i,i'} p_i p_{i'} h(|\langle\psi_i|\rho_{\mathcal{E}}^{-1/2}|\psi_{i'}\rangle|^2);\end{aligned} \tag{72}$$

$$\begin{aligned}I(A|\mathcal{E})\leqslant&\log_2[N_A N_{\mathcal{E}}\max_{i,j}\{p_i|\langle a_j|\rho_{\mathcal{E}}^{-1/2}|\psi_i\rangle|^2\}]\\ &+\sum_{j,j'} h(|\langle a_j|\rho_{\mathcal{E}}^{1/2}|a_j\rangle|^2)-2\sum_j h(\langle a_j|\rho_{\mathcal{E}}|a_j\rangle)\end{aligned} \tag{73}$$

follow from relations (70) and (71), respectively, where $h(x):=-x\log_2 x$.

For the particular case of an N-dimensional Hilbert space, with A Hermitian (i.e., $\langle a_j|a_k\rangle=\delta_{jk}$); equal prior probabilities (i.e., $p_i=1/N_{\mathcal{E}}$); and $\rho_{\mathcal{E}}=(1/N)\mathbf{1}$; the upper bound (73) reduces to

$$I(A|\mathcal{E})\leqslant\log_2[N\max_{i,j}\{|\langle a_j|\psi_i\rangle|^2\}], \tag{74}$$

reminiscent of inequality (38). This may be compared to the weaker bound $\log_2 N$ following from Holevo's theorem (11) and its dual (19).

F. Other applications

Information exclusion relations for two observables A and B in the communication context have the general form of relation (37). Here it will be briefly indicated how this form can be manipulated to obtain information inequalities for joint measurements; measurement of a given observable on two different ensembles; and quantum eavesdropping.

First, if each of two observables M_1 and M_2 can be measured jointly with some observable N, then the information bound

$$I(M_1,N|\rho)+I(M_2,N|\rho)\leqslant J\left(M_1,M_2,\sum_j N_j^{1/2}\rho N_j^{1/2}\right) \tag{75}$$

for the joint-measurement context follows directly from relation (37) and the relations $I(M_1,N|\rho)=I(M_1|\mathcal{E})$, $I(M_2,N|\rho)=I(M_2|\mathcal{E})$, where $\mathcal{E}$ is the ensemble given by the context mapping

$$p_j=\mathrm{tr}[\rho N_j],\quad \rho_j=(N_j)^{1/2}\rho(N_j)^{1/2}/\mathrm{tr}[\rho N_j] \tag{76}$$

analogous to Eqs. (9). The exclusion relation (75) quantifies the notion that the more strongly an observable M_1 is correlated with N, the less strongly a complementary observable M_2 will be correlated with N.

For example, if M_1 and M_2 correspond to measurements of spin in two orthogonal directions, on one of two spin-1/2 particles, and N corresponds to a spin measurement on the other particle, then from relations (38) and (75) one has

$$I(M_1,N|\rho)+I(M_2,N|\rho)\leqslant\log_2 2. \tag{77}$$

Hence the maximum correlation of $\log_2 2$ between M_1 and N is possible only if there is *zero* correlation between M_2 and N. Noting Eq. (9) of [12] an even stronger inequality holds, with $I(M_3,N|\rho)$ added to the left-hand side of relation (77), where M_3 corresponds to a measurement of spin along the remaining orthogonal axis.

Second, if $\mathcal{E}_1$ and $\mathcal{E}_2$ are two signal ensembles with the same ensemble density operator ρ, then one can use source duality (Sec. III A) and relation (37) to derive an information inequality of the form

$$I(A|\mathcal{E}_1)+I(A|\mathcal{E}_2)\leqslant J(D(\mathcal{E}_1),D(\mathcal{E}_2),\rho) \qquad (78)$$

for any observable A, where $D(\mathcal{E})$ denotes the dual observable for ensemble $\mathcal{E}$ defined in Eq. (15c). Thus the more strongly correlated an observable is with some ensemble $\mathcal{E}_1$, the less strongly correlated it will be with a ‘‘complementary’’ ensemble $\mathcal{E}_2$. The upper bound (71) corresponds to choosing $\mathcal{E}_2$ to be the dual ensemble defined by Eqs. (15a) and (15b), with $\mathcal{E}=\mathcal{E}_1$. An analogous bound can be obtained by choosing $\mathcal{E}_2$ to be the ‘‘Scrooge ensemble’’ defined in [22].

Third, consider a communication channel where an eavesdropper makes a measurement of an observable Z on the signal states before they reach the receiver, thus obtaining $I(Z|\mathcal{E})$ bits of data per signal. Assuming that the eavesdropping process is independent of the particular signal transmitted, then the joint probability P_{ij} of transmitting signal ρ_i and receiving a measurement result corresponding to A_j has the form

$$P_{ij}=p_i\mathrm{tr}[\zeta(\rho_i)A_j]=p_i\mathrm{tr}[\rho_i\zeta^*(A_j)], \qquad (79)$$

where ζ is a completely positive map or ‘‘nonselective operation,’’ with dual ζ^* [17]. Thus from Eqs. (5a) and (5b) the maximum information available at the receiver is given by $I(\zeta^*(A)|\mathcal{E})$, where $\zeta^*(A)$ denotes the observable with POM $\{\zeta^*(A_j)\}$. From relation (37) one then has the exclusion relation

$$I(Z|\mathcal{E})+I(\zeta^*(A)|\mathcal{E})\leqslant J(Z,\zeta^*(A),\rho_{\mathcal{E}}), \qquad (80)$$

which quantifies the notion that the eavesdropper in general can gain information only at the expense of decreasing the information at the receiver [35].

V. DISCUSSION

This paper has in part been an exploration of the symmetries of quantum correlations (context mappings and source duality), and their use as basic tools in obtaining new bounds from old ones. Thus the dual Holevo bound (19) was obtained via source duality from the original Holevo bound (11), which was in turn obtained via a context mapping from the joint-measurement bound (10). Similarly, the new upper bounds (70)–(74) for mutual information were obtained via source duality from the information exclusion relations (37) and (39).

It is worth emphasising that these tools can also be applied to other measures of correlation. For example, if the output ranges of sources X and Y are identical, then the coincidence rate $C(P_{ij})=\Sigma_i P_{ii}$ has the forms

$$C(A|\mathcal{E})=\sum_i p_i\ \mathrm{tr}[\rho_i A_i], \quad C(M,N|\rho)=\sum_i \mathrm{tr}[\rho M_i N_i]$$

in the communication and joint-measurement contexts, respectively. Thus for a bound $C(A|\mathcal{E})\leqslant B(A,\mathcal{E})$ in the former context, the bound $C(M,N|\rho)\leqslant B(A|\mathcal{E})$ in the latter context immediately follows with A and $\mathcal{E}$ given by the context mapping (8). Moreover, the dual bound $C(A|\mathcal{E})\leqslant B(A',\mathcal{E}')$ immediately follows from the source duality relation (17), with A' and $\mathcal{E}'$ given by Eqs. (15a)–(15c). Some results for $C(A|\mathcal{E})$ are given in [36].

The use of source duality to estimate lower bounds (24) and (31) for mutual information under transmitter and receiver constraints may be similarly applied to other measures of correlation. For example, the coincidence rate $C(A|\mathcal{E})$ when maximized with respect to A and $\mathcal{E}$, respectively (with fixed $\rho_{\mathcal{E}}$), has respective lower bounds $C(A'|\mathcal{E})$ and $C(A|\mathcal{E}')$.

An important feature of the information exclusion principle (Sec. IV) is its use for deriving tight bounds for information gain, exemplified for homodyne detection by relation (48) (with $\theta-\phi=90°$) [12], and for heterodyne detection by relations (57) and (58). These bounds cannot be derived from either Holevo's theorem (11) or its dual (19), suggesting that information exclusion relations for *complementary* observables can generally provide stronger bounds than Holevo's theorem. Indeed, from the perspective of the ‘‘quantum chessboard,’’ Eqs. (1) and (35) (corresponding to Holevo's theorem and information exclusion, respectively), yield the same upper bound for both $I(X|\mathcal{E})$ and $I(X|\mathcal{E})+I(P|\mathcal{E})$. This suggests for complementary observables A and B in general that one can find $J(A,B,\rho_{\mathcal{E}})$ as in relation (37) with $J(A,B,\rho_{\mathcal{E}})\approx S(\rho_{\mathcal{E}})$.

While information exclusion is strong enough to yield the optimal signal states for homodyne and (noisy) heterodyne detection, the information exclusion relation (39) is not strong enough to derive tight bounds for non-Hermitian discrete observables such as the ‘‘tetrahedral’’ measurement in relation (43). However, it should be possible to strengthen this relation for such observables by improving the corresponding entropic uncertainty relation (41). This would also improve the related upper bounds (72)–(74) for mutual information.

Finally, it should be noted that the paper has dealt with classical ‘‘bits’’ rather quantum ‘‘qubits’’ [37] or ‘‘ebits’’ [38]. The latter cases involve correlations between sources which generate quantum states as outputs. However, given that the correlations between two sources can be exploited only by (eventually) making measurements on their outputs, this is not regarded as a physically significant restriction on the results of the paper.

Indeed, the qubit and ebit formalisms appear to be more concerned with mapping given states onto particular target states, under various mapping constraints, rather than with source correlations per se. For example, Schumacher [37] considers generation of the product state $\rho_{i1}\otimes\rho_{i2}\otimes\cdots\otimes\rho_{iL}$ at some source X, with respective prior probability $p_{i1}p_{i2}\cdots p_{iL}$ (where $\Sigma_i p_i=1$), and subsequent generation of corresponding states $\phi(\rho_{i1}\otimes\rho_{i2}\otimes\cdots\otimes\rho_{iL})$ at source Y where ϕ represents some physical operation. If ϕ is constrained to include mapping the states into some Hilbert space H_C, and the ρ_i to be pure, then it is shown that for sufficiently large L the ''average fidelity''

$$F = \sum_{i_1, i_2, \ldots, i_L} p_{i1} p_{i2} \cdots p_{iL} \times \mathrm{tr}[\rho_{i1} \otimes \rho_{i2} \otimes \cdots \otimes \rho_{iL} \phi(\rho_{i1} \otimes \rho_{i2} \otimes \cdots \otimes \rho_{iL})]$$

can be made arbitrarily close to unity, providing that $\log_2 \dim(H_C)$ is greater than $LS(\Sigma_i p_i \rho_i)$. The quantity $\log_2 \dim(H_C)$ is referred to as the number of qubits available, and the result shows that $S(\Sigma_i p_i \rho_i)$ qubits per component state ρ_i are necessary for high-fidelity transmission.

Similarly, in [38] products of L two-particle states $\rho \otimes \cdots \otimes \rho$ on Hilbert space $H_1 \otimes H_2$ are generated at a source X, where H_1 (H_2) is the Hilbert space corresponding to the L subsystem 1 (subsystem 2) particles. Operations ϕ are then sought such that the states $\phi(\rho \otimes \cdots \otimes \rho)$ generated at output Y are orthogonal mixtures of maximally entangled pure states. Under the constraints that ρ is pure and ϕ is composed of operations which act on H_1 and H_2 separately, it is shown that for L sufficiently large the average ''entanglement'' of the states in the mixture $\phi(\rho \otimes \cdots \otimes \rho)$ can be made arbitrarily close to $LE(\rho)$, where for any pure state σ on $H_1 \otimes H_2$ one defines the entanglement in ebits to be

$$E(\sigma) = S(\mathrm{tr}_1[\sigma]) = S(\mathrm{tr}_2[\sigma]).$$

Thus an average of up to $E(\rho)$ ebits per two-particle state can be obtained.

Hence, while the qubit and ebit formalisms share some formal similarities with the communication and joint-measurement contexts, respectively [cf. Eqs. (11) and (13), respectively], they are primarily directed at mappings of given states onto target states rather than at correlations between statistical sources as discussed in this paper. Indeed, if two sources X and Y generate the pair of quantum states (ρ_i, σ_j) with associated probability P_{ij} [e.g., with $P_{ij} = p_i \delta_{ij}$ and $\sigma_i = \phi(\rho_i)$], it is difficult to find a general measure of correlation with physical significance. This is perhaps because physics ultimately relies on measurements, and no obvious measurements are available in the general case.

ACKNOWLEDGMENT

I thank Robyn Hall for partial financial support.

APPENDIX

Here the context-independent interpretation of mutual information, and its physical significance as a measure of correlation between statistical sources, are briefly outlined. Three roles of mutual information are distinguished, and the relevance of Shannon's coding theorem [1,2] to the joint-measurement context is discussed.

First, for two statistical sources X and Y as discussed in Sec. II A, consider a sequence of joint outputs of length L:$\{(x_{i_1}, y_{j_1}), (x_{i_2}, y_{j_2}), \ldots, (x_{i_L}, y_{j_L})\}$. It is convenient to denote such a sequence by (α, β), where α and β are the corresponding sequences of outputs $\{x_{i_1}, x_{i_2}, \ldots, x_{i_L}\}$, $\{y_{j_1}, y_{j_2}, \ldots, y_{j_L}\}$ of X and Y, respectively.

For sufficiently large L, the strong law of large numbers implies that nearly all sequences (α, β) are ''typical'' [1,2]; i.e., any given pair (x_i, y_j) will occur approximately LP_{ij} times for most sequences, and the remaining ''nontypical'' sequences may be ignored as having a vanishingly small total probability of occurrence. The probability of a given typical sequence follows as

$$P(\alpha, \beta) \approx \prod_{i,j} (P_{ij})^{LP_{ij}} = \text{const},$$

and hence the total number of typical joint sequences of length L may be estimated as

$$N_{XY} \approx 1/P(\alpha, \beta) = 2^{LH(P_{ij})}, \tag{A1}$$

where $H(P_{ij})$ denotes the entropy of the joint distribution P_{ij}. Similarly, the numbers of typical sequences output by X and Y, respectively, may be estimated as

$$N_X \approx 2^{LH(P_i)}, \quad N_Y \approx 2^{LH(P_j)}, \tag{A2}$$

where $H(P_i)$ and $H(P_j)$ denote the entropies of the marginal distributions defined in Eq. (3).

Now if the sources are completely uncorrelated, with $P_{ij} = P_i P_j$, then $N_{XY} \approx N_X N_Y$ from Eqs. (A1) and (A2). More generally, the correlation between the sources *reduces* the number of typical outputs, from $N_X N_Y$ to N_{XY}. Thus the number of binary digits needed to label the outputs is reduced from $\log_2 N_X N_Y$ to $\log_2 N_{XY}$, i.e., the data residing in the correlation itself is given by

$$\log_2[N_X N_Y / N_{XY}] \approx LI(P_{ij}), \tag{A3}$$

where $I(P_{ij})$ is defined in Eq. (4) of the text.

While Eq. (A3) indicates that mutual information represents an average amount of data per output pair residing in the correlation, and so may be taken as a measure of the strength of the correlation, it has minimal physical content. It does not show, for example, how $I(P_{ij})$ is related to data obtainable about one source from the other source, nor how such data may be exploited.

Second, define $N_{X|\beta}$ to be the number of typical sequences (α, β) with a fixed component β. The total number of sequences can then be written as

$$N_{XY} = \sum_\beta N_{X|\beta}. \tag{A4}$$

Moreover, the number of binary digits required to represent the possible output sequences α of X, given knowledge of output sequence β of Y, is reduced from $\log_2 N_X$ to $\log_2 N_{X|\beta}$. Therefore the amount of data gained about an unknown sequence of outputs α of X, given the corresponding sequence of outputs β of Y, is (measured in binary digits)

$$I(X|\beta) = \log_2 N_X / N_{X|\beta}. \tag{A5}$$

If $N_{X|\beta}$ is a *constant*, then it follows from Eq. (A4) that it has the value N_{XY}/N_Y, and hence from Eqs. (A3) and (A5) that

$$I(X|\beta) \approx LI(P_{ij}), \tag{A6}$$

thus justifying the interpretation of mutual information given in Sec. II A. But $N_{X|\beta}$ is indeed (approximately) constant. In particular, any typical sequence β of outputs of Y may be decomposed into subsequences $\beta_1, \beta_2, \beta_3, \ldots$, where β_j consists of (approximately) LP_j occurrences of output y_j. Each subsequence has some corresponding number N_j of possible typical subsequences α_j of outputs of X, of corresponding length LP_j. By construction N_j is independent of the parent sequence β [indeed one may estimate $N_j \approx 2^{LP_j H(P_{ij}/P_j)}$ in analogy to Eq. (A2)], and thus one has $N_{X|\beta} = \Pi_j N_j = \text{const}$ as required.

Equation (A6) establishes the connection of mutual information with the data obtainable from one source about another source. Shannon's coding theorem [1,2] outlined below further demonstrates that in principle this data is not difficult to exploit, in both the communication and joint-measurement contexts.

Third then, suppose that M typical sequences of outputs of length L from X are chosen at random, $\alpha_1, \alpha_2, \ldots, \alpha_M$, say, to code for M distinct symbols. For an observer at Y to distinguish without error between these sequences at X, and thus obtain $\log_2 M$ bits of error-free data, each possible typical sequence β generated at Y must correspond to at most one coding sequence $\alpha_1, \alpha_2, \ldots, \alpha_M$. But if β does correspond to some coding sequence, then the probability that any particular one of the remaining $N_{X|\beta} - 1$ possible sequences corresponding to β is also a coding sequence is $q = (M-1)/(N_X - 1)$. The probability that *none* of these is also a coding sequence is therefore

$$\text{Prob(no error)} = (1-q)^{N_{X|\beta}-1} \approx (1 - M/N_X)^{N_{X|\beta}}$$
$$= (1 - M/N_X)^{N_{XY}/N_Y}.$$

Finally, if less than $I(P_{ij})$ bits of data per output of Y is to be recovered per coding sequence, i.e., $\log_2 M = L\ [I(P_{ij}) - \delta]$ with $\delta > 0$, then using Eqs. (A1)–(A3) one has

$$\text{Prob(no error)} \approx [1 - 2^{-L\delta}/(N_{XY}/N_Y)]^{N_{XY}/N_Y}$$
$$\to \exp[-2^{-L\delta}] \to 1 \quad \text{as } L \to \infty. \quad \text{(A7)}$$

Thus, for L sufficiently large, there are codings which allow recovery of up to $I(P_{ij})$ bits of data per output per coding sequence with arbitrarily small error. Clearly such a coding may be exploited in the communication context by using the sequences $\alpha_1, \alpha_2, \ldots, \alpha_M$ as an alphabet for messages, and restricting transmission to these sequences.

In the joint-measurement context there is no control over transmission, i.e., over the outputs of X. However if, for example, an observer at X groups the outputs into sequences of length L, and notes the (mutually agreed) coding sequences $\alpha_1, \alpha_2, \ldots, \alpha_M$ as they appear, then the latter generate a random cryptographic key which can be recovered by an observer at Y with an arbitrarily small error. From Eq. (A7) up to $LI(P_{ij})$ bits of key can be generated per coding sequence as M approaches $2^{LI(P_{ij})}$; hence, since the probability of a coding sequence is M/N_X per group of L outputs, the average number of bits generated per output is bounded by

$$R \leqslant I(P_{ij}) 2^{LI(P_{ij})}/N_X \approx I(P_{ij}) 2^{-L[H(P_i) - I(P_{ij})]}, \quad \text{(A8)}$$

where Eq. (A2) has been used. Thus, in general, a balance must be struck between a low error rate (large L) and a reasonable value of R (low L), e.g., via error-correcting codes [2,38].

[1] C. E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948); **27**, 623 (1948), reprinted in *Claude Elwood Shannon: Collected Papers*, edited by N. Sloane and A. Wyner (IEEE, New York, 1993), pp. 5–83.

[2] R. G. Gallagher, *Information Theory and Reliable Communication* (Wiley, New York, 1968).

[3] C. M. Caves and P. D. Drummond, Rev. Mod. Phys. **66**, 481 (1994).

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

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

[6] S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. **61**, 662 (1988).

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

[8] A. Ekert, Phys. Rev. Lett. **67**, 661 (1991).

[9] C. H. Bennett, G. Brassard, and N. D. Mermin, Phys. Rev. Lett. **68**, 557 (1992).

[10] A. S. Holevo, Probl. Inf. Trans. **9**, 110 (1973).

[11] W. K. Wootters and W. H. Zurek, Phys. Rev. D **19**, 473 (1979).

[12] M. J. W. Hall, Phys. Rev. Lett. **74**, 3307 (1995).

[13] E. B. Davies, IEEE Trans. Inf. Theory, **IT-24**, 596.

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

[15] A parameter estimation context could also be defined, where a ‘‘scatterer’’ parametrized by x_i, with prior distribution p_i, scatters a probe state ρ into a state ρ_i, on which an observable A is then measured. This is formally equivalent to the communication context, however.

[16] If continuously valued outputs (x,y) are relabelled by (x',y'), where $x \to x'$ and $y \to y'$ are one-one mappings, then the joint distribution $p'(x',y')$ of x' and y' is given by $p'(x',y')dx'dy' = p(x,y)dx\, dy$, with marginal distributions $p'(x')dx' = p(x)dx$, $p'(y')dy' = p(y)dy$. Hence, replacing summation by integration in Eq. (4), one has $I[p(x,y)] = I[p'(x',y')]$. In contrast, entropy transforms as, e.g., $S[p'(x')] = S[p(x)] + \int dx\ p(x) \log_2 |J_X|$, where J_X denotes the Jacobian of the transformation $x \to x'$.

[17] E. B. Davies, *Quantum Theory of Open Systems* (Academic, New York, 1976); K. Kraus, *States, Effects and Operations*, Lecture Notes in Physics Vol. 190 (Springer, Berlin, 1983).

[18] C. H. Bennett *et al.*, Phys. Rev. Lett. **76**, 722 (1996).

[19] G. Lindblad, in *Quantum Aspects of Quantum Communications*, Lecture Notes in Physics Vol. 378, edited by C. Bendjaballah *et al.* (Springer, Berlin, 1991), pp. 71–80, and references therein; S. J. D. Phoenix and S. M. Barnett, *ibid*, pp. 93–100.

[20] S. M. Barnett and S. J. D. Phoenix, Phys. Rev. A **44**, 535 (1991).

[21] Note that (10) also follows immediately by replacing ρ_θ and ρ in Eq. (12) of Ref. [5] by ρ and $\rho_1 \otimes \rho_2$, respectively.

[22] R. Josza, D. Robb, and W. K. Wootters, Phys. Rev. A **49**, 668 (1994).

[23] C. A. Fuchs and C. M. Caves, Phys. Rev. Lett. **73**, 3047 (1994).

[24] L. B. Levitin, in *Information, Complexity and Control in Quantum Physics*, edited by A. Blacquieve, S. Diner, and G. Lochak (Springer, New York, 1987), pp. 15–47; M. J. W. Hall and M. J. O'Rourke, Quantum Opt. **5**, 161 (1993); H. Scutaru, Phys. Rev. Lett. **75**, 773 (1995); B. Schumacher, M. Westmoreland, and W. K. Wootters, *ibid.* **76**, 3452 (1996).

[25] K. Vogel, V. M. Akulin, and W. P. Schleich, Phys. Rev. Lett. **71**, 1816 (1993).

[26] For mutual information properties of phase detection see M. J. W. Hall, J. Mod. Opt. **40**, 809 (1993); K. R. W. Jones, Phys. Scr. **T48**, 100 (1993); M. J. W. Hall, in *Quantum Communications and Measurement*, edited by V. P. Belavkin *et al.* (Plenum, New York, 1995), pp. 53–59; see also Ref. [12] above.

[27] H. P. Yuen and J. H. Shapiro, IEEE Trans. Inf. Theory **IT-26**, 78 (1980); J. H. Shapiro and S. S. Wagner, IEEE J. Quantum Electron. **QE-20**, 803 (1984).

[28] K. Kraus, Phys. Rev. D **35**, 3070 (1987).

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

[30] I. Bialynicki-Birula and J. Mycielski, Commun. Math. Phys. **44**, 129 (1975).

[31] G. H. Hardy, J. E. Littlewood, and G. Polya, *Inequalities* (Cambridge University Press, London, 1934), Eq. (2.13.7) of Sec. 2.13.

[32] M. J. W. Hall, Phys. Rev. A **50**, 3295 (1994).

[33] S. Friberg and L. Mandel, Opt. Commun. **46**, 141 (1983).

[34] U. Leonhardt and H. Paul, Phys. Rev. A **48**, 4598 (1993).

[35] A. K. Ekert, B. Huttner, G. M. Palma, and A. Peres, Phys. Rev. A **50**, 1047 (1994); C. A. Fuchs and A. Peres, *ibid.* **53**, 2038 (1996); N. Lütkenhaus, *ibid.* **54**, 97 (1996).

[36] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic, New York, 1976).

[37] B. Schumacher, Phys. Rev. A **51**, 2738 (1995).

[38] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, Phys. Rev. A **53**, 2046 (1996).