Universal geometric approach to uncertainty, entropy, and information

Michael J. W. Hall
Theoretical Physics, IAS, Australian National University, Canberra, Australian Capital Territory 0200, Australia
(Received 14 August 1998)

It is shown that a unique measure of *volume* is associated with any statistical ensemble, which directly quantifies the inherent spread or localization of the ensemble. It is applicable whether the ensemble is classical or quantum, continuous or discrete, and may be derived from a small number of theory-independent geometric postulates. Remarkably, this unique *ensemble volume* is proportional to the exponential of the ensemble entropy, and hence provides an interesting geometric characterization of the latter quantity. Applications include unified volume-based derivations of results in quantum and classical information theory, a precise geometric interpretation of thermodynamic entropy for equilibrium ensembles, a geometric derivation of semi-classical uncertainty relations, a means for defining classical and quantum localization for arbitrary evolution processes, and a proposed definition for the spot size of an optical beam. Advantages of ensemble volume over other measures of localization (root-mean-square deviation, Renyi entropies, and inverse participation ratio) are discussed. [S1050-2947(99)08203-7]

PACS number(s): 03.65.Bz, 03.67.−a, 05.45.−a, 42.60.Jf

I. INTRODUCTION

This paper has two main goals. The first is to demonstrate that, for any ensemble, whether classical, quantum, discrete, or continuous, there is essentially only *one* measure of the ''volume'' occupied by the ensemble which is compatible with basic geometric notions. This *ensemble volume* is thus a preferred and universal choice for characterizing what is variously referred to as the spread, dispersion, uncertainty, or localization of an ensemble.

Remarkably, the derived ''ensemble volume'' turns out to be proportional to the exponential of the entropy of the ensemble. A by-product of the first goal is thus a universal characterization of ensemble entropy, based on geometric notions. Indeed, a number of properties of ensemble entropy turn out to have simple geometric interpretations. The universal nature of the characterization is of particular interest: the only previous *context-independent* interpretation of ensemble entropy to date (and hence applicable in particular to ensembles described by continuous probability distributions) appears to be as a somewhat vague measure of uncertainty or randomness.

The second goal is to apply ''ensemble volume'' to a wide range of contexts in which ensembles appear. The applications demonstrate not only the advantages of ensemble volume over other measures of spread, but also to some extent why it is that ensemble entropy makes a natural appearance in contexts as diverse as statistical mechanics, information theory, chaos, and quantum uncertainty relations. Some results have been briefly reported elsewhere [1]. Here important details and extensions are given, as well as a number of additional results.

The work reported here was originally motivated by several connections between volume and information. Shannon proved an upper bound on information transfer, via classical signals subject to quadratic energy and noise constraints, by considering ratios of spherical volumes in high-dimensional spaces [2]. One can similarly obtain approximate upper bounds on information for *quantum* signals, via semiclassical arguments involving ratios of phase space volumes [3,4], which in some cases turn out to be exact. This raises the question of whether there is some general measure of volume which can be used to derive rigorous information bounds for the general case. This question is answered affirmatively here, and a unified derivation of the classical Shannon and the quantum Holevo information bounds is given, based on simple volume properties.

There are also a number of connections which have been made previously between volume and entropy. For example, derivations in statistical mechanics typically obtain heuristic expressions for thermodynamic entropy by counting ''microstates'' in a phase-space volume of ''small'' thickness containing a constant-energy surface [5]. Ma, in an interesting approach, attempted to *define* the thermodynamic entropy of a system in classical statistical mechanics as proportional to the logarithm of a phase-space volume corresponding to the ''region of motion'' of the system [6], although he could not rigorously define the latter region. A *precise geometric* interpretation of thermodynamic entropy for both classical and quantum equilibrium ensembles will be given here.

Further, Leipnik introduced the exponential of the position entropy of a quantum system as a measure of its ''volume,'' and favorably compared the associated uncertainty relations for position and momentum with the usual Heisenberg uncertainty relations [7] (see also the review in Ref. [8], and Sec. II C below). Generalizations to other measures of ''volume'' were given by Zakai [8,9]. It is demonstrated here that the former measure has a unique geometrical significance, and a geometrical derivation of quantum uncertainty relations is given based on the property that quantum states have a minimum ensemble volume.

Zyckowski [10] and more recently Mirbach and Korsch [11,12], used entropy as a measure of ''localization'' for chaotic quantum and classical systems for various initial states. The results of the present paper show that this measure can be simply related to the spread of ensemble volume for arbitrary evolution processes, and provide support for the use of this measure over all other localization measures.

Rather than going immediately to general postulates for volume, and formal proofs of uniqueness, Sec. II first explores ensemble volume for a familiar class of ensembles: those described by one-dimensional probability distributions. In this case the ensemble volume reduces to a ''length,'' which is calculated for a number of concrete examples and compared with other measures of uncertainty such as root-mean-square deviation. Geometric properties of this ''length'' and an associated quantum uncertainty relation are discussed. Two-dimensional joint probability distributions are also briefly discussed, where the ensemble volume becomes an ''area'' that is geometrically related to the ''lengths'' of the marginal distributions. This ''area'' motivates a definition for the spot size of an optical beam.

In Sec. III and the accompanying Appendix, the derivation of the ensemble volume from universal geometric postulates is given. These postulates depend on theory-independent notions of invariance, projection onto orthogonal axes, and additivity, and in particular are independent of whether the ensemble is classical or quantum. The bonus of a geometrical characteriztion of ensemble entropy is discussed, and a geometrical interpretation of relative entropy is given.

Applications to statistical mechanics, semiclassical quantum mechanics, information theory, chaos and other types of dynamical evolution are given in Sec. IV. Conclusions are presented in Sec. V.

II. ONE- AND TWO-DIMENSIONAL EXAMPLES

Before deriving the unique form of ensemble volume in Sec. III, it is useful first to consider some of its properties and connections to other measures of uncertainty in two familiar settings: continuous distributions on the line and on the plane, for which ''volume'' reduces to the special cases of ''length'' and ''area,'' respectively. These special cases are already sufficient to exemplify a number of general features of ensemble volume, and its advantages as a measure of spread.

A. Length

Consider a one-dimensional probability distribution $p(x)$, corresponding to some random variable X (e.g., position, momentum, or phase). There are then a number of candidates for a direct measure of the ''uncertainty'' or ''spread'' of X, the most well known being the root-mean-square (rms) deviation

$$\Delta X=\left[\int dx\, x^2 p(x)-\left(\int dx\, x p(x)\right)^2\right]^{1/2}. \tag{1}$$

This quantity is a ''direct'' measure in the sense of having the same units as X, and has the virtues of being invariant under translations and reflections, scaling linearly with X ($\Delta Y=\lambda\Delta X$ for $Y=\lambda X$), and vanishing in the limit that X has some definite value x'.

A second candidate is the inverse participation ratio [10,12,13]

$$\xi_X=\left[\int dx\, p(x)^2\right]^{-1}, \tag{2}$$

(which may also be recognized as a monotonic function of the so-called ''linear entropy'' $-\int dx\, p(x)^2$ [14]). This quantity shares all of the above-noted virtues of ΔX. However, it is in fact only a special case of what may be called the ''Renyi length''

$$L_{X,\alpha}=\left[\int dx\, p(x)^{1+\alpha}\right]^{-1/\alpha} \quad (\alpha\geqslant -1) \tag{3}$$

(named for its logarithm—a generalized entropy defined by Renyi [15]). Renyi lengths are directly related to measures of uncertainty considered by Dodonov and Man'ko and Zakai for quantum systems [8,9], and use of their reciprocals as (indirect) measures of uncertainty was extensively investigated in Ref. [16] (see also Ref. [17]). The inverse participation ratio corresponds to $\alpha=1$ in Eq. (3).

The Renyi length $L_{X,\alpha}$ in Eq. (3) satisfies all of the above-noted properties of ΔX [same units as X, translation and reflection invariance, scaling linearly with X, and vanishing as $p(x)$ approaches a δ function]. Equation (3) thus introduces an uncountable infinity of possible candidates for a direct measure of uncertainty. Fortunately, as will be seen in Sec. III, just one of these Renyi lengths may be singled out uniquely over all other possible measures on geometric grounds.

In particular, in this paper special attention will be paid to the case $\alpha\to 0$ in Eq. (3). The corresponding length will simply be denoted by L_X, and is just the exponential of the usual ensemble entropy [18]:

$$L_X=L_{X,0}=\exp\left[-\int dx\, p(x)\ln p(x)\right]. \tag{4}$$

This is a special case of the ''ensemble volume,'' to be derived in Sec. III, and will therefore be referred to as the *ensemble length*.

B. Comparisons

In Table I the rms deviation and ensemble length are calculated for several types of one-dimensional distributions. As noted following Eqs. (1) and (3), both quantities are invariant under translations and scale linearly with X. Hence they can be trivially calculated for distributions of the form $p(x/a-x')/a$ once they have been found for $p(x)$ (by simply multiplying the result for the latter case by a). Table I will be used to highlight a number of differences between ΔX and L_X.

First, it is seen from Table I that the ensemble length exists in cases when the rms deviation does not (for Cauchy-Lorentz and sink-squared distributions in particular). It may further be shown that L_X is finite whenever ΔX is: the well known variational property that ensemble entropy is maximized for a fixed value of ΔX by a Gaussian distribution [19] immediately implies from the scaling property and Table I that

$$L_X\leqslant(2\pi e)^{1/2}\Delta X. \tag{5}$$

TABLE I. Examples of ensemble length and rms deviation.

Distribution	$p(x)$	L_X	ΔX
Uniform	$p_U(x)=1, 0\leqslant x\leqslant 1$	1	$1/(2\sqrt{3})$
Circular	$p_C(x)=2(1-x^2)^{1/2}/\pi, \lvert x\rvert\leqslant 1$	$\pi/\sqrt{e}$	1/2
Gaussian	$p_G(x)=(2\pi)^{-1/2}\exp(-x^2/2)$	$(2\pi e)^{1/2}$	1
Exponential	$p_E(x)=\exp(-x), x\geqslant 0$	e	1
Sink squared	$p_{SS}(x)=\pi^{-1}[\sin(x)/x]^2$	$\pi e^{2(1-C)}$ [a]	
Cauchy-Lorentz	$p_{CL}(x)=\pi^{-1}/(1+x^2)$	4π	
Double uniform	$p_{DU}(x)=\frac{1}{2}, 0\leqslant\lvert x\rvert-a\leqslant 1$	2	$[1/3+a(a+1)]^{1/2}$

[a] $C\approx 0.577\,215\,66$ denotes Euler's constant.

Thus the use of ensemble length as a measure of uncertainty allows a wider quantitative range of applicability than does the rms deviation. This permits, for example, a quantitative discussion of quantum uncertainty relations, expressed in terms of ensemble length, for cases in which the usual Heisenberg uncertainty relations have nothing to say (see Sec. II C).

Second, the calculations for the uniform and circular distributions, p_U and p_C in Table I, respectively, exemplify a maximality property of ensemble length: it is maximized on a given interval by a uniform distribution on the interval, with a maximum value equal to the length of the interval. Thus one may write

$$L_X \leqslant L \tag{6}$$

for a distribution confined to an interval of length L [20]. This property reflects the intuitive notion that $p(x)$ is most spread out or least localized when it is *flat*, having no peaks where probability is concentrated. The rms deviation does not conform to this notion, achieving its maximum possible value in the limit of two maximally separated peaks (a distribution equally concentrated on the endpoints of the interval).

Third, the calculation in Table I for the uniform and double-uniform distributions p_U and p_{DU} illustrates an addivity property of ensemble length: the ensemble length of p_{DU} is twice that of the two nonoverlapping uniform distributions $p_U(x-a)$ and $p_U(-x-a)$, which it comprises in equal mixture. More generally, if $p(x)$ and $q(x)$ denote two nonoverlapping distributions of equal ensemble length L, then any mixture $\lambda p(x)+(1-\lambda)q(x)$ of these distributions satisfies

$$L_X \leqslant 2L, \tag{7}$$

with the upper bound achieved for $\lambda=\frac{1}{2}$ [21]. This property reflects the intuitive notions that such a mixture is least localized (most spread out) when it is not more concentrated in one of the nonoverlapping regions than in the other, and that for this equally weighted case the nonoverlapping lengths simply add. In contrast, the rms deviation of p_{DU} depends strongly on the separation of the peaks, and indeed becomes infinite as this separation increases. This example and the one above emphasize what can be directly seen from Eq. (1): the rms deviation is a measure of *separation* of the region(s) of concentration from a particular point of the distribution (the mean value), rather than a measure of the extent to which the distribution is in fact concentrated.

Fourth, except in cases where the second moment of $p(x)$ has some particular physical meaning, it is difficult to assess the significance of a given value of ΔX without some further information about the distribution. For example, even for single-peaked distributions, the probability that X lies within $\pm\Delta X$ of the mean is highly dependent upon the nature of $p(x)$ [22]. In contrast, as will be seen in Sec. III, the ensemble length L_X has a unique geometrical significance.

Finally, it is of interest to make a quantitative comparison between the degrees to which a given distribution $p(x)$ is concentrated in a region of length L_X on the one hand, and of length $2\Delta X$ on the other hand. To do so, it is natural to define the *maximum confidence* corresponding to a given length L as

$$C(L)=\sup_{\{A:|A|=L\}}\left\{\int_A dx\, p(x)\right\}, \tag{8}$$

where the supremum is over all measurable sets A of total length L. In the case of a distribution symmetric about a single peak this is achieved by choosing A to be the interval of length L centered on the mean value of the distribution.

From Table I one can calculate the values of $C(L_X)$ to be approximately 100%, 99%, 96%, 93%, 91%, and 90% for the uniform, circular, Gaussian, exponential, sink-squared, and Cauchy-Lorentz distributions, respectively. The corresponding values of $C(2\Delta X)$ are 58%, 61%, 68%, and 86% for the first four of the above distributions, with the value being undefined for the last two. It is seen that for these examples $C(L_X)$ varies over a much narrower range than $C(2\Delta X)$, and that L_X typically corresponds to a larger confidence value than $2\Delta X$.

C. Uncertainty relations

The relationship between ensemble length and ensemble entropy in Eq. (4) allows the usual entropic uncertainty relation for the position and momentum of a quantum particle [23] to be equivalently written in the geometric form

$$L_X L_P \geqslant \pi e \hbar, \tag{9}$$

relating the product of the ensemble lengths to a minimum area in phase space. Bounding L_X and L_P from above via Eq. (5) then immediately yields the well-known Heisenberg uncertainty relation

$$\Delta X \Delta P \geqslant \hbar/2. \tag{10}$$

The above two inequalities are similar in form, and have the same broad physical significance: the particle cannot be prepared in a state for which both the position and momentum distributions have arbitrarily small spreads. However, it is seen that the latter inequality is mathematically weaker, as it follows from the former. For example, it follows from Eq. (9) that L_P [and hence, via Eq. (5), ΔP] becomes infinite as $p(x)$ approaches a weighted sum of δ functions. This cannot be concluded from Eq. (10).

Inequality (9) may used to make quantitative evaluations regarding the relative spreads of position and momentum in cases where the Heisenberg inequality (10) yields *no* information. For example, consider a quantum particle confined to an interval of length L, such that the position amplitude is constant over the interval. It follows that the momentum statistics are described by the sink-squared distribution

$$\pi^{-1}(2\hbar/L)(\sin[pL/(2\hbar)]/p)^2. \tag{11}$$

As noted in Table I, the rms deviation ΔP is not defined in this case, and hence the Heisenberg inequality cannot be used to assess the degree to which position and momentum are jointly localized. In contrast, using Eq. (11), Table I, and the scaling property of ensemble length, one finds

$$L_X L_P = 2\pi \exp[2(1-C)]\hbar \approx 15\hbar, \tag{12}$$

where $C \approx 0.577\,215\,66$ denotes Euler's constant. Hence the particle has an associated phase space area close to the lower bound of $\pi e\hbar \approx 9\hbar$ in Eq. (9), i.e., the particle is in fact in an approximate minimum uncertainty state of position and momentum.

A similar example is the case of a particle confined to the positive x axis, with a position amplitude that decays exponentially with x. The position and momentum distributions are then given by exponential and Cauchy-Lorentz distributions of the forms $p_E(x/a)/a$ and $2ap_{\mathrm{CL}}(2ap/\hbar)/\hbar$, respectively, implying via Table I and the scaling property that

$$L_X L_P = 2\pi e\hbar. \tag{13}$$

Hence the state is relatively well localized in position and momentum, with an associated phase-space area only twice that of the minimum in [Eq. (9)]. Again, the Heisenberg uncertainty relation Eq. (10) gives no information about the joint localization in this case.

Finally, it may be mentioned that there is an uncertainty relation relating the Renyi lengths of position and momentum for general α: it follows from Eq. (131) of Ref. [8] that

$$L_{X,\alpha} L_{P,\beta} \geqslant \pi\hbar[1+2\alpha]^{1+1/(2\alpha)}/(1+\alpha) \tag{14}$$

for $\alpha \geqslant -\frac{1}{2}$, where $\beta = -\alpha/(1+2\alpha)$. For $\alpha=\beta=0$ the lower bound is maximum, and the inequality reduces to Eq. (9) above.

D. Area and spot size

This section will be concluded by briefly looking at measures of spread for *two*-dimensional distributions, to highlight a further geometric property of ensemble length of importance in later sections. This property also holds for rms deviation, but not for Renyi lengths in general. A related measure of spot size for optical beams is defined and briefly discussed.

Each of the ''length'' measures in Eqs. (1), (3), and (4) has a natural generalization to a measure of ''area,'' corresponding to the spread or uncertainty of a two-dimensional probability distribution $p(x,y)$ of two random variables X and Y:

$$\Delta A = [\det(\langle \mathbf{x}\mathbf{x}^{\mathbf{T}}\rangle - \langle \mathbf{x}\rangle\langle \mathbf{x}^{\mathbf{T}}\rangle)]^{1/2}, \tag{15}$$

$$A_{XY,\alpha} = \langle p^{\alpha}\rangle^{-1/\alpha}, \tag{16}$$

$$A_{XY} = \exp[\langle -\ln p\rangle] \tag{17}$$

respectively, where $\mathbf{x}$ denotes the column vector (x,y), $\mathbf{x}^{\mathbf{T}}$ its transpose, and $\langle\ \rangle$ the average with respect to p. These areas satisfy properties analogous to to their one-dimensional counterparts, and will be referred to as the rms area, Renyi area, and ensemble area, respectively.

The rms area in Eq. (15) may be recognized as the product of the rms deviations along the principal axes of the distribution in the xy plane, and in general satisfies the inequality [Eq. (2.13.7) of Ref. [24]]

$$\Delta A \leqslant \Delta X \Delta Y, \tag{18}$$

with equality for the case that $p(x,y)$ factorizes into two uncorrelated distributions for X and Y.

This inequality for ''area'' and ''length'' has a simple geometric interpretation, to be generalized in Sec. III. In particular, the marginal distributions $p_1(x)$ and $p_2(y)$ for X and Y are obtained by ''projecting'' the joint distribution $p(x,y)$ onto the two orthogonal x and y axes. The associated rms lengths ΔX and ΔY may be similarly thought of as obtained by ''projecting'' the rms area ΔA onto these axes. However, this is only consistent with Euclidean geometry if inequality (18) holds: the product of the two lengths obtained by projection of an area onto two orthogonal axes can never be less than the original area.

Ensemble area and ensemble length are also consistent with this ''projection'' interpretation: the well-known subadditivity of entropy [19] can be equivalently written via Eqs. (4) and (17), as

$$A_{XY} \leqslant L_X L_Y, \tag{19}$$

in analogy to Eq. (18). The subadditivity of entropy is thus seen to correspond to a projection property of Euclidean geometry. One has the further related property that if $p(x,y)$ is uniform on a rectangular region oriented parallel to the x and y axes, and vanishes outside this region, then equality holds in Eq. (19), with L_X and L_Y corresponding to the lengths of the sides of the rectangle. Thus Eq. (19) reduces in this case to the Euclidean property *area = length × breadth*. In general, the Renyi areas in Eq. (16) are not consistent with the projection property, as will be seen in Sec. III.

Finally, it may be noted that Eq. (17) may be applied to physical distributions other than probability distributions, with corresponding geometrical advantages. For example, let $P(x,y)$ denote the time-averaged power distribution in some plane orthogonal to the direction of propagation of an optical beam. One may then define the ''geometric'' spot size of the beam as the ensemble area of the normalized power distribution $P(x,y)/P_T$, where P_T is the integrated power over the plane:

$$A_{\mathrm{geom}}=P_T\exp\left[-(P_T)^{-1}\int dx\,dy\,P(x,y)\ln P(x,y)\right]. \tag{20}$$

This satisfies desirable properties such as being additive for nonoverlapping identical beams, being invariant with respect to scaling the power up or down, scaling linearly with beam magnification, having a maximum value of A for a beam confined to an area A (attained for a uniform power distribution over that area), and satisfying a ''projection property'' analogous to Eq. (19). It is also invariant under any transformation of coordinates which preserves area in the usual sense (i.e., with unit Jacobian), and so to this extent is independent of the coordinatization of the plane. Alternative definitions based on, for example, Eqs. (15) or (16), are geometrically less satisfying.

III. ENSEMBLE VOLUME

Section II indicates the wide range of possible measures for the spread of one- and two-dimensional probability distributions, and draws attention to a number of geometric and other advantages enjoyed by the ''length'' and ''area'' defined in Eqs. (4) and (17), respectively. As noted in Sec. I, it has often proved useful to employ various notions of ''volume'' for statistical ensembles across a wide variety of contexts, such as information theory, statistical mechanics, uncertainty relations, and chaotic evolution. Other contexts include Ornstein-Uhlenbeck diffusion and semiclassical quantum mechanics (see Ref. [1], and Secs. IV B and IV D below). This raises the question of whether there is in fact some *universal* measure of ''volume'' for classical and quantum ensembles, which may be usefully employed in all of the above contexts and which is not restricted in application or interpretation to various special cases.

Here it will be shown that indeed such a measure exists, which may be uniquely derived from a small number of theory-independent postulates fundamental to the concept of ''volume.'' It generalizes the ensemble length and ensemble area of Sec. II, and will be referred to as the *ensemble volume*. It also leads to geometric characterizations of entropy and relative entropy.

A. Notation

Three generic types of ensemble will be considered here. The first is a classical ensemble described by a continuous probability distribution $p(\mathbf{x})$ on some n-dimensional space X; the second is a classical ensemble described by a discrete probability distribution $\{p_i\}$, where i ranges over some discrete set I; and the third is a quantum ensemble described by a density operator W on some Hilbert space H.

Each of the above types of ensemble shares some universal features. It is essential to abstract a number of these features via a common notation if ''volume'' is to be discussed in a theory-independent manner.

For example, consider the three identities

$$\int_X d^n\mathbf{x}\,p(\mathbf{x})=1,\quad \sum_{i\in I}p_i=1,\quad \mathrm{tr}_H[W]=1. \tag{21}$$

Defining Γ to correspond to the spaces and sets X, I, and H, $\mathrm{Tr}_\Gamma[\ ]$ to correspond to integration over X, summation over I, and the trace over H, and ρ to correspond to the ensembles $p(\mathbf{x})$, $\{p_i\}$, and W, these identities can be subsumed into the generic identity

$$\mathrm{Tr}_\Gamma[\rho]=1. \tag{22}$$

Another universal feature is the notion of *composite* or *joint* ensembles: for a given pair of spaces or sets Γ_1 and Γ_2 of a given type, one can define a composite set or space Γ_{12}, where for classical and quantum ensembles Γ_{12} corresponds to the set product and the tensor product, respectively, of Γ_1 and Γ_2. Further, if ρ describes a composite ensemble on Γ_{12}, one may define two *projected* ensembles ρ_1 and ρ_2 on Γ_1 and Γ_2, respectively, via

$$\rho_1=\mathrm{Tr}_{\Gamma_2}[\rho],\quad \rho_2=\mathrm{Tr}_{\Gamma_1}[\rho]. \tag{23}$$

These projected ensembles correspond to *marginal* distributions and *reduced* density operators for the cases of classical and quantum ensembles, respectively.

Finally, one may define any two ensembles ρ and ρ' of the same type to be nonoverlapping if and only if

$$\mathrm{Tr}_\Gamma[\rho\rho']=0. \tag{24}$$

Note that in general two ensembles are nonoverlapping if and only if they can be distinguished by measurement without error.

B. Postulates for volume

For the three types of ensemble discussed in Sec. III A, it is useful to think of ''volume'' in the following ways. First, for a continuous distribution $p(\mathbf{x})$ on a space X, the volume corresponds to a direct measure of the region(s) of ''spread'' of $p(\mathbf{x})$ in X. Second, for a classical discrete distribution $\{p_i\}$, one may imagine the indices as labeling a set of boxes or bins. In this case ''volume'' corresponds to the spread of the distribution over these bins, i.e., as a continuous measure of the effective number of bins occupied by the distribution. Third, for a quantum ensemble, the volume may be considered as a continous generalization of Hilbert space dimension, corresponding to a measure of the spread of the ensemble in Hilbert space. Consider now a measure of volume, $V(\rho)$, which satisfies the following properties.

(i) *Invariance property:* $V(\rho)$ is invariant under all transformations on Γ which preserve $\mathrm{Tr}_\Gamma[\ ]$ (these are represented by measure-preserving transformations on X for continuous classical ensembles, permutations on I for discrete classical ensembles, and unitary transformations on H for quantum ensembles).

(ii) *Cartesian property:* If ρ describes two *uncorrelated* ensembles ρ_1 and ρ_2 on Γ_1 and Γ_2, respectively, then

$$V(\rho)=V(\rho_1)V(\rho_2) \tag{25}$$

(note that ρ is the product $\rho_1\rho_2$ for classical ensembles, and the tensor product $\rho_1 \otimes \rho_2$ for quantum ensembles).

(iii) *Projection property:* If ρ describes an ensemble of composite systems on Γ_{12}, then

$$V(\rho) \leq V(\rho_1)V(\rho_2), \tag{26}$$

where ρ_1, ρ_2 are the projections of ρ defined in Eq. (23).

(iv) *Additivity property:* An equally weighted mixture of *m nonoverlapping* ensembles $\rho_a, \rho_b, \ldots$, each of equal volume *V*, has a total volume of mV, i.e.,

$$V(m^{-1}[\rho_a + \rho_b + \cdots]) = mV. \tag{27}$$

(v) *Uniformity property:* If ρ is any mixture of m nonoverlapping ensembles of equal volumes *V*, then

$$V(\rho) \leq mV. \tag{28}$$

The above properties are essentially the same as those defined in Ref. [1], where the additivity and uniformity properties were combined in the latter. Their geometrical significance is as follows.

First, the invariance property (i) ensures that the volume $V(\rho)$ is a function of the ensemble alone, independently of a particular coordination, labeling, or measurement basis for Γ. Indeed, the transformations which preserve $\mathrm{Tr}_\Gamma[\;]$ are exactly those which preserve volume, or measure, on Γ in the usual sense. For example, for a classical distribution $p(\mathbf{x})$ on X the measure of a subset $S \subseteq X$ is given by

$$|S| = \int_S d^n\mathbf{x} = \mathrm{Tr}_S[1]. \tag{29}$$

The invariance property then requires that the ensemble volume is invariant under all transformations which preserve the measure of all subsets, i.e., those transformations with a unit Jacobian. For the case of a classical phase space, such transformations include all canonical transformations, and hence $V(\rho)$ will be invariant under Hamiltonian evolution. One may similarly consider the measure $|S| = \mathrm{Tr}_S[1]$ of subsets $S \subseteq I$ and subspaces $S \subseteq H$; in these cases the invariance property again requires that $V(\rho)$ is invariant under measure-preserving transformations, corresponding to permutations and unitary transformations, respectively.

Second, the Cartesian property (ii) is exactly analogous to the geometric property that area equals length times breadth, and more generally that the volume of the Cartesian product of two sets is equal to the product of the volume of the sets. This is illustrated in Fig. 1.

Third, the projection property (iii) is exactly analogous to the geometric property that a volume is less than or equal to the product of the lengths obtained by its projection onto orthogonal axes, and is illustrated in Fig. 2. It is a generalization of the projection property discussed for rms area and ensemble area in Sec. II D.

Fourth, the additivity property (iv) requires the ensemble volume to be additive for a uniform mixture of nonoverlapping ensembles of equal volume. The geometric interpretation of this is self-evident: the total volume of *m* equal nonoverlapping volumes is the sum of the individual volumes.

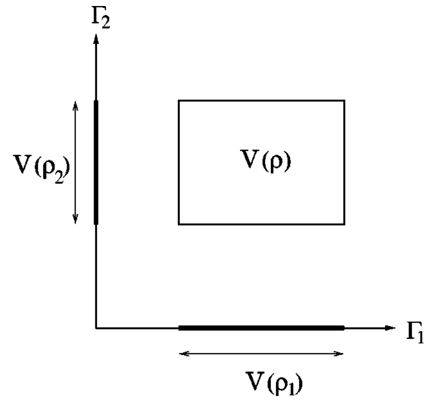

FIG. 1. Two uncorrelated ensembles ρ_1 and ρ_2 on spaces Γ_1 and Γ_2, respectively (shown here compressed to one-dimensional axes), have respective volumes $V(\rho_1)$ and $V(\rho_2)$, as indicated by the darkened axis regions. The *Cartesian property* [Eq. (25)] states that the corresponding joint ensemble ρ has a ''rectangular'' volume $V(\rho) = V(\rho_1)V(\rho_2)$, i.e., $V(\rho)$ corresponds to the Cartesian product of volumes $V(\rho_1)$ and $V(\rho_2)$.

Finally, the uniformity property (v) states that the maximum volume, of a mixture of nonoverlapping ensembles of equal volume, is bounded by the sum of the component volumes. Thus, noting the additivity property, this maximum is achieved for a *uniform* mixture, i.e., one which is not more concentrated on one of the component ensembles than on any other.

C. Derivation

Here the unique, universal measure of volume for ensembles is obtained. It may more generally be applied as a measure of spread for any positive classical or quantum density, such as beam intensity or mass density, by calculating the ''volume'' of the corresponding normalized density. In such cases, where no ensemble is involved, one could alternatively label this quantity as the ''geometric dispersion.'' In particular, one has the following result, first stated in [1], and proved in the Appendix.

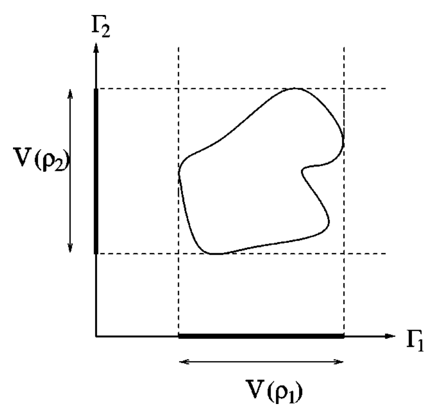

FIG. 2. An ensemble ρ on the product space of Γ_1 and Γ_2 has a volume $V(\rho)$ indicated by the solid closed curve. The corresponding projected ensembles ρ_1 and ρ_2 on Γ_1 and Γ_2, respectively, have projected volumes $V(\rho_1)$ and $V(\rho_2)$, indicated by the darkened axis regions. The *projection property* [Eq. (26)] states that $V(\rho)$ can be no greater than the volume of the rectangular region formed by the dashed lines, i.e., than the product of the projected volumes.

Theorem: Any (continuous) measure of volume satisfying properties (i)–(v) above has the form

$$V(\rho)=K(\Gamma)e^{S(\rho)}, \tag{30}$$

where $S(\rho)$ denotes the ensemble entropy

$$S(\rho)=-\mathrm{Tr}_{\Gamma}[\rho\ln\rho], \tag{31}$$

and $K(\Gamma)$ is a constant which may depend on Γ, and satisfies

$$K(\Gamma_{12})=K(\Gamma_1)K(\Gamma_2). \tag{32}$$

The proof in the Appendix primarily relies on applying properties (i)–(v) to an arbitrarily large number of independent copies of a given ensemble ρ. I believe it may be possible to prove the theorem without the uniformity property (v), but have not been able to do so.

The constant $K(\Gamma)$ in Eq. (30) is a normalization constant, reflecting the notion that only relative volumes are of real interest in comparing different ensembles. For continuous classical ensembles a natural choice is $K(\Gamma)=1$, so that a distribution which is uniform over a set S of measure V, and vanishes outside S, has ensemble volume equal to V.

For discrete classical ensembles the choice $K(\Gamma)=1$ corresponds to measuring the ensemble volume in terms of the number of ''bins'' occupied by the ensemble, with the minimum volume of one bin corresponding to a distribution with $p_i=1$ for some index i. However, if the distribution arises from the discretization of a continuous observable such as position (due to measurement limitations, for example), then it would be natural to choose $K(\Gamma)$ to correspond to the discretization volume. If the index set is finite, with M labels, another possible choice for $K(\Gamma)$ is $1/M$. The ensemble volume then measures the fraction of the total volume occupied by the ensemble.

For quantum ensembles the choice $K(\Gamma)=1$ corresponds to measuring the ensemble volume in terms of the number of Hilbert space dimensions occupied by the ensemble, with pure states occupying the minimum possible of one dimension. However, if the Hilbert space H has a finite dimension M, then one could alternatively take $K(\Gamma)=1/M$, corresponding to a fractional measure of volume in analogy to the classical case. Finally, for quantum systems with classical counterparts, such as spin-zero particles, one may choose $K(\Gamma)$ so that in the classical limit the quantum ensemble volume reduces to the classical ensemble volume. This is explored further in Sec. IV B, and used to obtain semiclassical uncertainty relations.

It should be noted that the assumption of continuity in the statement of the theorem is necessary. For example, one may for a discrete classical ensemble $\{p_i\}$ define the ''support volume'' as the number of nonzero p_i values. This satisfies all of properties (i)–(v), but is not continuous. The simplest counterexample is the discrete probability distribution $\{1-\epsilon,\epsilon\}$ for $\epsilon>0$. As $\epsilon\to 0$ this distribution continuously approaches the distribution $\{1,0\}$, with a support volume of 1; however for all $\epsilon>0$ the support volume is 2.

If one defines the rms volume for an n-dimensional observable $\mathbf{x}$ by generalizing Eq. (15) to arbitrary dimensions [25], it is not difficult to show that the invariance property (restricted to *linear* transformations), the Cartesian property, and the projection property are satisfied. However it does not satisfy the additivity and uniformity properties. Further, the ''Renyi'' volumes

$$V_\alpha(\rho)=(\mathrm{Tr}_{\Gamma}[\rho^{1+\alpha}])^{-1/\alpha}, \tag{33}$$

defined in analogy with the Renyi length and Renyi area in Eqs. (3) and (16), respectively, satisfy properties (i), (ii), (iv), and (v) for all $\alpha\geqslant -1$. However, a counterexample given by Renyi (Theorem 4 of Sec. IX.6 in Ref. [15]) shows that the projection property is *not* satisfied, except for the cases $\alpha=0$ [corresponding to Eq. (30), with $K(\Gamma)=1$], and $\alpha=-1$ (corresponding to the discontinuous case of ''support volume'' discussed above).

D. Geometric characterization of entropy

The appearance of the ensemble entropy in Eq. (30) as a result of geometric postulates (i)–(v) provides an approach to this quantity, which is moreover independent of whether the ensemble is classical or quantum, discrete, or continuous. In particular, *ensemble entropy may be defined* (up to an additive constant) *as the logarithm of the ensemble volume*, where the latter is taken to be the primary quantity. The properties of ensemble entropy may thus be regarded as being geometric in origin. Indeed, it will be seen that its natural appearance in a number of physical contexts can be interpreted as following from its relationship to a ''volume.''

The geometric interpretation of ensemble entropy contrasts markedly with its only other context-independent interpretation as an (indirect) measure of ''uncertainty'' or ''randomness'' [15–17,19,26,27]. Indeed, ensemble volume provides a *direct* measure of uncertainty, which is advantageous when one wishes to compare the spreads of two ensembles of a given type (i.e., with the same Γ). For example, if two ensembles have entropies of 0.5 bits and 1.5 bits, respectively [28], should one compare their ratio or their difference in assessing the degree to which the uncertainty of the second exceeds that of the first? Since entropies are typically only defined up to a multiplicative constant (see below), one might consider the ratio to be the more significant means of comparison. However, the ensemble volume gives an unequivocal answer: the volume of the second ensemble is twice that of the first in this case, and hence has twice the spread.

It is interesting to briefly compare the derivation of ensemble volume from properties (i)–(v) with existing axiomatic derivations of ensemble entropy. Such axiomatic derivations are reviewed in Ref. [29], and are all related to the original derivation given by Shannon [26]. Unlike the theorem of Sec. II C they are limited to *discrete* classical ensembles. Moreover, they lead to an arbitrary multiplicative constant for entropy, whereas the geometric approach leads to an arbitrary *additive* constant for entropy.

To see that the axioms used by Shannon and others are markedly different from properties (i)–(v) used to derive ensemble volume, consider the ''grouping axiom'' of Shannon [26] (see also Sec. 1.2 of Ref. [19]), which may be written in the notation of this paper as

$$S(\lambda\rho+(1-\lambda)\rho')=S(\{\lambda,1-\lambda\})+\lambda S(\rho)+(1-\lambda)S(\rho') \tag{34}$$

for any two nonoverlapping discrete classical ensembles ρ and ρ'. Thus it is assumed that the ''randomness'' $S(\,)$ of a mixture of nonoverlapping distributions is equal to that of the mixing distribution plus the average randomness of the individual ensembles. This axiom, together with a continuity assumption and a symmetry assumption equivalent to the invariance property (i), is sufficient to derive the form $S(\rho) = -C\Sigma_i p_i \ln p_i$ for the entropy of discrete classical ensembles, where C is an arbitrary constant [29].

Equation (34) does *not* translate into a natural axiom for ensemble volume: replacing S by $\ln V$ gives the equivalent constraint

$$V(\lambda\rho+(1-\lambda)\rho')=V(\{\lambda,1-\lambda\})[V(\rho)]^{\lambda}[V(\rho')]^{1-\lambda}, \tag{35}$$

which has no simple geometric interpretation. Conversely, the additivity property Eq. (27), that nonoverlapping equal volumes add, translates under $V\rightarrow\exp S$ into the ''randomness'' constraint

$$S(\rho/2+\rho'/2)=\ln 2+S, \tag{36}$$

which is not a natural property to postulate for a measure of ''randomness.'' The geometric approach to ensemble entropy given here thus differs significantly from former approaches (as is also apparent from comparing the proof in the Appendix with those in Refs. [19,26,29]).

Finally, it is of interest to note that the concavity property of ensemble entropy, $S(\Sigma_i\lambda_i\rho_i)\geqslant\Sigma_i\lambda_i S(\rho_i)$ [19,26], is equivalent to an inequality relating the volume of a mixture to the weighted geometric mean of the volumes of its components:

$$V\left(\sum_i \lambda_i\rho_i\right)\geqslant\prod_i [V(\rho_i)]^{\lambda_i}. \tag{37}$$

This may be regarded as a generalization of the uniformity property [Eq. (28)], as it implies that uniform mixtures have the greatest volumes. Note that the ensemble volume may itself be regarded as a weighted geometric mean [e.g., of the function $p(x)^{-1}$ with respect to $p(\mathbf{x})$ for continous classical ensembles; see Secs. 2.2 and 6.7 of Ref. [24]].

E. Relative entropy

The relative entropy of two ensembles ρ and σ may be defined in a context independent manner by [30]

$$S(\rho|\sigma)=\mathrm{Tr}_\Gamma[\rho(\ln\rho-\ln\sigma)]. \tag{38}$$

It is asymptotically related to the probability of mistaking ensemble ρ for ensemble σ, as was reviewed in Ref. [31]. Here it will briefly be indicated how a geometric interpretation of this quantity can be given.

Consider a compact n-dimensional space X which is divided up into into a set of nonoverlapping bins $\{B_i\}$ (e.g., for measurement purposes). A discrete probability distribution $\{p_i\}$ over the bins (e.g., corresponding to measurement results), may then also be modeled by the *continuous* distribution $p(\mathbf{x})$ on X defined by

$$p(\mathbf{x})=p_i/V_i, \quad \mathbf{x}\in B_i, \tag{39}$$

where $V_i=\int_{B_i}d^n\mathbf{x}$ denotes the measure of bin B_i. Thus $p(\mathbf{x})$ is uniform over each bin, and its integral over bin B_i is equal to p_i. Let ρ_D and ρ_C denote the discrete and continuous ensembles corresponding to $\{p_i\}$ and $p(\mathbf{x})$, respectively.

Now, as discussed earlier, the ensemble volume $V(\rho_D)$ is proportional to the effective number of bins occupied by ρ_D. However, this does not indicate the effective volume or spread of the ensemble relative to X, particularly in the case of varying bin sizes V_i. The latter is given by $V(\rho_C)$, which, making the choice $K(\Gamma)=1$, follows from Eq. (39) as

$$V(\rho_C)=\exp\left[-\sum_i p_i\ln(p_i/V_i)\right]. \tag{40}$$

Note that in the case of *equal* bin sizes $V_i\equiv V$ this reduces to the bin size V multiplied by the effective number of bins occupied, $\exp S(\rho_D)$.

Finally, if X has total measure $\Sigma_i V_i=V_X$, one may define the ''weighting'' ensemble σ_D as corresponding to the discrete probability distribution $\{V_i/V_X\}$. Thus σ_D describes the relative sizes or weightings of the bins. It then follows via Eqs. (38) and (40) that

$$V(\rho_C)/V_X=e^{-S(\rho_D|\sigma_D)}. \tag{41}$$

Hence *the relative entropy $S(\rho|\sigma)$ is directly related to the volume of a discrete ensemble ρ embedded in a continuous space, where σ characterizes the distribution of bin sizes of the embedding.* Note that this geometric interpretation of relative entropy allows its properties to be understood as corresponding to ratios of volumes. For example, the volume of an ensemble on X can never be greater than V_X (corresponding to a uniform distribution on X). Hence the left-hand side of Eq. (41) is never greater than unity, implying that

$$S(\rho|\sigma)\geqslant 0. \tag{42}$$

IV. APPLICATIONS

The results of Sec. II for ensemble length and ensemble area indicate the usefulness of ensemble volume as a direct measure of the spread of an ensemble (and of other positive densities such as optical beam power). Here other applications will be examined, in the contexts of statistical mechanics, semiclassical quantum mechanics, information theory, and quantum chaos. A particular result of note is a unified proof of the classical Shannon information bound and the quantum Holevo information bound based on ratios of ensemble volumes. For the quantum case this proof is conceptually and technically far simpler than previous proofs.

A. Statistical mechanics

First, in the statistical mechanics context, the Gibbs relation $S_{\mathrm{th}}=kS(\rho)$ between thermodynamic entropy and ensemble entropy for equilibrium ensembles can be rewritten, via Eq. (30), as

$$S_{\mathrm{th}}=k\ln[V(\rho)/K(\Gamma)]. \tag{43}$$

Thus the thermodyamic entropy is (up to an additive constant) proportional to the logarithm of the ensemble volume.

From Eq. (43) and the third law of thermodynamics (that thermodynamic entropy vanishes at absolute zero), it follows that one should choose $K(\Gamma)$ to correspond to a minimum ''zero-temperature'' ensemble volume. For quantum ensembles one has from Eqs. (30) and (31) that $V(\rho)=K(\Gamma)$ for pure states, i.e., the *quantum* zero-temperature volume is just that of a *pure* state on Γ. Similarly, for discrete classical ensembles, $K(\Gamma)$ is the volume of the ''pure'' ensemble described by $\{1,0,0,\dots\}$. However, continuous classical ensembles violate the third law [5], and $K(\Gamma)$ remains arbitrary in this case (but see Sec. IV B below).

The geometric expression (43) is very similar to the original Boltzmann relation

$$S_{\rm th}=k\ln W, \tag{44}$$

where W is the number of distinct microstates or ''elementary complexions'' consistent with the thermodynamic description. Indeed, from the above discussion it follows that Eq. (43) provides a *precise geometric* interpretation of the Boltzmann relation for discrete classical and quantum equilibrium ensembles: *thermodynamic entropy is proportional to the logarithm of the number of nonoverlapping zero-temperature volumes contained within the total volume of the ensemble*. Thus the Boltzmann relation and the Gibbs formula for thermodynamic entropy become directly unified in the ensemble volume approach.

Properties of thermodynamic entropy can be reinterpreted in terms of geometric volume. For example, the additivity of thermodynamic entropy for uncorrelated ensembles in thermal equilibrium follows from Eq. (43) and the Cartesian property [Eq. (25)] for uncorrelated ensemble volumes. Note also that irreversible processes correspond geometrically to those which increase the volume of the ensemble.

B. Semiclassical quantum mechanics

Consider now a classical ensemble ρ_C which is the ''classical limit'' of some quantum ensemble ρ_Q, i.e., the physical properties of ρ_C approximate those of ρ_Q. Such ensembles exist, for example, for equilibrium ensembles in the high-temperature limit and for the coherent states of a harmonic oscillator.

For the case of a spinless particle associated with a $2n$-dimensional phase space one can obtain a relationship between the constants $K(\Gamma_C)$ and $K(\Gamma_Q)$ in Eq. (30) by requiring that the ensemble volumes $V(\rho_C)$ and $V(\rho_Q)$ are approximately equal for such ensembles. Since these constants are independent of the dynamics of the ensemble, it suffices to choose an equilibrium ensemble of isotropic oscillators. Equating the calculated values of $V(\rho_C)$ and $V(\rho_Q)$ in the high-temperature limit then yields

$$K(\Gamma_Q)=h^nK(\Gamma_C) \tag{45}$$

for the volume of a pure state, where h is Planck's constant. Thus the Bohr-Sommerfeld quantization rule that a pure quantum state occupies a classical phase-space volume of h^n is recovered [32].

Equation (45) can be used to derive semiclassical uncertainty relations from geometric considerations. For two corresponding ensembles ρ_Q and ρ_C as above the position and momentum entropies S_X and S_P, respectively, must be approximately equivalent for either ensemble. Further,

$$\exp(S_X)\exp(S_P)\geqslant\exp(S(\rho_C)) \tag{46}$$

holds for the classical ensemble from the projection property [Eq. (26)] applied to projections onto the position and momentum axes. Equations (30), (45), and (46) then yield the approximate inequality

$$S_X+S_P-S(\rho_Q)\gtrsim n\ln h \tag{47}$$

for quantum ensembles which have classical limits. I conjecture that *exact* inequality in fact holds for *all* quantum ensembles.

Since the entropy of a quantum ensemble has a minimum value of 0 (corresponding to the existence of a minimum volume for quantum ensembles), it follows from Eq. (47) that one has the semiclassical entropic uncertainty relation

$$S_X+S_P\gtrsim n\ln h, \tag{48}$$

for quantum ensembles with classical limits. As per the derivation of Eq. (10) from Eq. (9), the corresponding semiclassical Heisenberg uncertainty relation

$$\Delta X\Delta P\gtrsim\hbar/e \tag{49}$$

then follows for the $n=1$ case. Equations (48) and (49) are close to the exact results for general quantum ensembles [8,23] [see Eqs. (9) and (10)]. It is seen that geometrically they correspond to application of the projection property [Eq. (26)] to the projections of a pure state of volume h^n onto the position and momentum axes (i.e., replacing Γ_1 and Γ_2 by X and P in Fig. 2).

C. Information bounds

Consider a communication channel where signals represented by ensembles $\rho_1,\rho_2,\dots$ are transmitted with prior probabilities $p_1,p_2,\dots$, respectively [33]. The ensemble of signal states itself corresponds to the mixture

$$\rho=\sum_i p_i\rho_i. \tag{50}$$

For classical ensembles, it was shown by Shannon [26] that the average amount of error-free data I which can be obtained per transmitted signal, measured in terms of the number of binary digits required to represent the data, is bounded above by

$$I\leqslant\left[S(\rho)-\sum_i p_iS(\rho_i)\right]\bigg/\ln 2. \tag{51}$$

The formally equivalent bound for quantum ensembles was proved by Holevo [34], and hence Eq. (51) may be referred to as the Shannon-Holevo information bound.

Proofs given in the literature of Eq. (51) for the quantum case are mathematically rather technical in nature, and quite different in character to proofs for the classical case [34,35]. However, the formal equivalence of the quantum and classical bounds suggests that a unified proof exploiting universal features of statistical ensembles may be possible. Indeed the construction of such a proof, based on simple volume arguments, was recently outlined in Ref. [1], and will be elaborated on here. A second such proof, which reduces the general quantum-classical case to that of discrete classical noiseless channels, will also be pointed out.

First, consider a message consisting of L signals chosen from the set $\{\rho_i\}$. Such a message may be denoted by ρ_α, where $\alpha=(i_1,i_2,\ldots,i_L)$ denotes the labels of the signals comprising the message. In the limit that $L\rightarrow\infty$ the strong law of large numbers implies that the relative frequency of signal ρ_i appearing in the message approaches p_i with probability 1. It follows from the Cartesian property Eq. (25) that the volume of the message satisfies

$$V(\rho_\alpha)\rightarrow V_{\mathrm{mess}}=\prod_i\,[V(\rho_i)]^{p_iL}, \tag{52}$$

as $L\rightarrow\infty$. Moreover, as will be shown below in Eq. (56), the volume of any ensemble of such messages is bounded above by $[V(\rho)]^L$. Hence, using the additivity property Eq. (27), the maximum possible number of nonoverlapping messages of length L, N_L, satisfies

$$N_L\leqslant[V(\rho)]^L/V_{\mathrm{mess}} \tag{53}$$

as $L\rightarrow\infty$. Noting that *error-free* data can only be obtained from distinguishing among a set of *nonoverlapping* messages, and that N_L such messages require at most $1+\log_2 N_L$ binary digits to record, it follows in the limit of infinitely long messages that the average information gained per signal, I, is bounded by

$$I\leqslant\lim_{L\rightarrow\infty}L^{-1}(1+\log_2 N_L)\leqslant\log_2 V(\rho)/\prod_i\,[V(\rho_i)]^{p_i}. \tag{54}$$

Finally, since communication based on finite message lengths cannot transmit more data per signal than communication based on infinite lengths, the bound holds for all signaling schemes, and Eq. (51) follows from Eqs. (30) and (54).

The above proof of the Shannon-Holevo bound is geometrically simple, being based on the ratio of the maximum available volume for an ensemble of messages to the message volume [Eq. (53)]. Note that the argument cannot be used to derive similar bounds based on other invariant volume measures, as all of the defining properties of ensemble volume are required. However, heuristic arguments of the same type for other volume measures can sometimes give excellent results [3,4]. Note that the Shannon-Holevo bound is in fact *tight* for both classical and quantum ensembles [19,26,36], corresponding geometrically to being able to choose a number N_L of messages arbitrarily close to the upper bound in Eq. (53) which can be distinguished with a vanishingly small average error probability as $L\rightarrow\infty$.

To conclude this subsection, it will be shown that the Shannon-Holevo bound may also be proved by considering only messages of finite length, and applying the classical noiseless coding theorem [19,26]. With notation as above, suppose that one chooses a set of codewords C from the set of messages of length L, and that codeword $\rho_\alpha\in C$ is transmitted with probability $q(\alpha)$. Defining $N_i(\alpha)$ as the number of times signal ρ_i appears in codeword ρ_α, and $\overline{\rho_l}=\sum_{\alpha\in C}q(\alpha)\rho_{i_l}$ as the average lth component of the transmitted codewords, consistency requires that

$$p_i=L^{-1}\sum_{\alpha\in C}q(\alpha)N_i(\alpha),\qquad \rho=L^{-1}\sum_{l=1}^{L}\overline{\rho_l}. \tag{55}$$

Using the projection property [Eqs. (26) and (37)], one then has the inequality chain

$$V\left(\sum_\alpha q(\alpha)\rho_\alpha\right)\leqslant V(\overline{\rho_1})\cdots V(\overline{\rho_L})\leqslant\left[V\left(\sum_l L^{-1}\overline{\rho_l}\right)\right]^L = [V(\rho)]^L. \tag{56}$$

To obtain a bound for error-free data, it must be assumed that the codewords are nonoverlapping, so that they can be distinguished without error by measurement. From Eq. (30) and the Cartesian property [Eq. (25)] one may then calculate

$$V\left(\sum_\alpha q(\alpha)\rho_\alpha\right)=e^{S[q]}\prod_{\alpha\in C}[V(\rho_\alpha)]^{q(\alpha)} = e^{S[q]}\prod_{\alpha\in C}\prod_l\,[V(\rho_{i_l})]^{q(\alpha)}, \tag{57}$$

where $S[q]$ denotes the entropy of the discrete distribution $\{q(\alpha)\}$. Combining this with Eqs. (55) and (56) then gives

$$\begin{aligned} S[q]&\leqslant LS(\rho)-\sum_{\alpha\in C}\sum_l q(\alpha)S(\rho_{i_l})\\ &=LS(\rho)-\sum_{\alpha\in C}\sum_i q(\alpha)N_i(\alpha)S(\rho_i)\\ &=L\left[S(\rho)-\sum_i p_iS(\rho_i)\right]. \end{aligned} \tag{58}$$

Finally, from Shannon's classical noiseless coding theorem [19,26] $S[q]/\ln 2$ is the maximum information (measured in binary digits) which can be transmitted on average per codeword, and hence Eq. (51) follows for the average information transmitted per signal.

D. Chaotic and other diffusion processes

Zyckowski [10] and Mirbach and Korsch [11,12] studied connections between quantum and classical chaos via entropies associated with the evolution of coherent states. Here it will be shown that this approach may be simply interpreted in terms of ensemble volume, and considerably generalized.

Consider an ensemble ρ_0, classical or quantum, which evolves in time under some dynamical process D (not necessarily reversible). The ensemble will explore some region of Γ, which may be large for standard diffusion processes, or relatively small for integrable and dissipative systems. The localization of the ensemble in Γ over time is characterized by the time-averaged mixture

$$\overline{\rho}=\lim_{T\to\infty}T^{-1}\int_0^T dt\,\rho_t. \tag{59}$$

This mixture gives greatest weight to regions of Γ where the ensemble spends the most time. Hence its ensemble volume $V(\overline{\rho})$ is a measure of the spread of the region explored by the ensemble as it evolves.

The *localization ratio* for a given initial state and evolution process may now be defined as the ratio of the volumes of $\overline{\rho}$ and ρ_0, i.e.,

$$r=V(\overline{\rho})/V(\rho_0)=\exp[S(\overline{\rho})-S(\rho_0)]. \tag{60}$$

It thus measures the localization of the ensemble under the evolution process, relative to its initial spread. This ratio will be less than or equal to 1 if the ensemble evolves to a fixed point, and greater than or equal to 1 if it diffuses over the whole of Γ. For chaotic systems with integrable regions it will depend strongly on the initial ensemble. The above definition is clearly natural on geometric grounds, and the ensemble entropy appears as a consequence of the uniqueness theorem in Eq. (30).

For classical and quantum systems corresponding to the same evolution process, it is of interest to compare localization properties. This is easily done for the case of initial quantum ensembles ρ_Q which have corresponding classical counterparts ρ_C (such as coherent states). In this case the quantum and classical localization ratios r_Q and r_C can be calculated and compared. Zyckowksi partially carried through this procedure in Ref. [10], where he plotted $S(\overline{\rho})$ for the quantum counterpart of a classically chaotic process, where ρ_Q was chosen to range over a set of coherent states indexed by their corresponding phase-space points. In this case $S(\overline{\rho})$ is just the entropy of the energy distribution of ρ_Q. Noting $S(\rho_Q)=0$ for pure states, it follows from Eq. (60) that this is equivalent to plotting the logarithm of the localization ratio, $\ln r$. However, Zyckowski compared quantum localization features qualitatively with the classical phase space portrait, rather than quantitatively with analogously calculated classical localization ratios.

Mirbach and Korsch extended the approach of Zyckowski by also calculating $S(\overline{\rho})$ for the classical ensembles ρ_C corresponding to the coherent states ρ_Q. For a complete family of such states they then compared the corresponding classical and quantum values of $S(\overline{\rho})$ (Figs. 1 and 3 of Ref. [12]). Since for this case $S(\rho_Q)$ and $S(\rho_C)$ are constants, this amounts to comparing the logarithms of the classical and quantum localization ratios (up to an additive constant).

However, Mirbach and Korsch argued that one should in fact compare *measurement* entropies rather than the direct ensemble entropies, to smear out quantum fluctuations in the latter case [11,12]. This is also easily interpreted in terms of localization ratios. In particular, for a measurement observable A on a classical or quantum ensemble ρ, let $V_A(\rho)$ denote the volume of the measurement distribution of A. The localization ratio of an evolution process with respect to A, for an initial ensemble ρ_0, is then defined in analogy to Eq. (60) as

$$r_A=V_A(\overline{\rho})/V_A(\rho_0). \tag{61}$$

Again one may compare localization ratios for classical and quantum ensembles, where one chooses corresponding observables A_Q and A_C. The logarithm of this quantity (up to an additive constant) is plotted in Figs. 2 and 3 of Ref. [11] for quantum and classical systems, respectively, for a complete set of coherent states, where A_C is chosen to be a phase-space measurement (so that $r_{A_C}=r_C$), and A_Q to be a ‘‘Husimi’’ phase-space measurement corresponding to the complete set of coherent states [37].

V. CONCLUSIONS

In conclusion, an essentially unique measure of volume for classical and quantum ensembles has been found, related to ensemble entropy, which provides a geometric tool for any context in which ensembles appear. This measure is universal in the sense that it may be defined by theory-independent concepts of invariance, uncorrelated ensembles, projection, and nonoverlapping ensembles [properties (i)–(v)].

Its properties as a direct measure of ‘‘spread’’ have been investigated in Sec. II for continuous distributions, and favorably compared with measures based on root-mean-square deviation. Geometric characterizations of ensemble entropy and relative entropy have been discussed in Secs. III D and III E.

Applications include a definition of spot size for optical beams, a precise geometric interpretation of the Boltzmann relation in statistical mechanics, a derivation of semiclassical uncertainty relations based on the existence of a minimum volume for quantum states and a projection property of volumes, a unified derivation of results in classical and quantum information theory based on simple volume ratios, and a universal definition of a localization ratio which measures the time-averaged spreading of an ensemble and underlies entropic measures previously investigated in the context of quantum chaos.

Work is in progress on further applications, particularly to quantum information theory [36], measures of quantum entanglement [31], and information exclusion relations [4,38]. The conjecture suggested following Eq. (47) is also under active investigation, and the (mostly weaker) bound

$$S_X+S_P-S(\rho)\geqslant \ln 2\pi e\hbar-\ln[1+\Delta X\Delta P/(\hbar/2)] \tag{62}$$

has thus far been found for the $n=1$ case.

ACKNOWLEDGMENTS

I am grateful to Professor Wolfgang Schleich for drawing my attention to the inverse participation ratio (thus stimulating the search for the ‘‘best’’ measure of volume), and to Professor Hajo Leschke, Dr. Gernot Alber, and Dr. Bruno Mirbach for useful discussions. This work was carried out at the University of Ulm, and was supported by the Alexander von Humboldt Foundation.

APPENDIX

Here the fundamental theorem stated in Sec. III C is proved, showing essentially that the exponential of the ensemble entropy is the unique measure of the volume of a statistical ensemble. It is convenient to first prove the theorem for discrete classical ensembles, and then extend the arguments to quantum ensembles and to continuous classical ensembles. The notation will be as defined in Sec. III A, and reference will be made to the five assumed properties of the volume measure $V(\rho)$ stated in Sec. III B.

Let ρ denote a classical discrete ensemble $\{p_i\}$, with finite index set $I=\{1,2,\ldots,M\}$. Defining the "pure" ensemble $\rho_j (j\in I)$ as corresponding to the distribution $\{p_i^{(j)}\}$ with $p_i^{(j)}=\delta_{ij}$, one can write ρ as the mixture

$$\rho=\sum_{i\in I} p_i\rho_i . \tag{A1}$$

Note that one has the two basic properties

$$\mathrm{Tr}_\Gamma[\rho_j\rho_k]=0 \quad (j\neq k), \qquad V(\rho_j)=\mathrm{const}=V_I . \tag{A2}$$

The first states that these pure ensembles are nonoverlapping, and the second that they have equal ensemble volumes (this follows from the invariance property, noting that ρ_j map to each other under permutations).

Now consider the ensemble $\rho^L\in\Gamma^L$ corresponding to L uncorrelated copies of ρ. For each $\alpha=(i_1,i_2,\ldots,i_L)$ in I^L, define

$$\rho_\alpha=\rho_{i_1}\rho_{i_2},\ldots,\rho_{i_L}, \quad p(\alpha)=p_{i_1}p_{i_2},\ldots,p_{i_L}. \tag{A3}$$

Thus ρ_α corresponds to the uncorrelated composite ensemble formed by $\rho_{i_1},\rho_{i_2},\ldots,\rho_{i_L}$ (in that order). One can then decompose ρ^L into the mixture

$$\rho^L=\sum_{\alpha\in I^L} p(\alpha)\rho_\alpha . \tag{A4}$$

The proof of the theorem proceeds by finding a suitable set of so-called "typical sequences" $T\subseteq I^L$ [19,26], which allows ρ^L in Eq. (A4) to be approximated by certain mixtures of the ensembles $\{\rho_\alpha\}$ where α is restricted to range over T.

For a given $\alpha\in I^L$ let $N_i(\alpha)$ denote the number of times the index i appears as a component of α, and let $P(\alpha)\in I^L$ correspond to a permutation of the components of α. If $S(\rho)$ denotes the entropy of ρ defined in Eq. (31) of the text, then for any $\epsilon>0$ and L sufficiently large one may choose a set T, with $|T|$ elements, which satisfies

(T1) $C_T=\sum_{\alpha\in T} p(\alpha)>1-\epsilon,$

(T2) $|T|=e^{L[S(\rho)+\delta_L]},$

(T3) $\sum_{i\in I} |L^{-1}N_i(\alpha)-p_i|<\delta_L'$ for all $\alpha\in T,$

(T4) $\alpha\in T$ implies $P(\alpha)\in T$ for all P,

where both δ_L and $\delta_L'\to 0$ as $L\to\infty$. A particular example of such a set is

$$T=\{\alpha:|L^{-1}N_i(\alpha)-p_i|<[Mp_i(1-p_i)/(L\epsilon)]^{1/2}\}. \tag{A5}$$

Properties (T1) and (T2) for this set are proved in Theorem 1.3.1 of Ref. [19]; property (T3) follows noting that $\Sigma_i[p_i(1-p_i)]^{1/2}$ is bounded by $(M-1)^{1/2}$, and hence that one can choose $\delta_L'=M(L\epsilon)^{-1/2}$; and property (T4) is an immediate consequence of $N_i(\alpha)$ being invariant under permutations.

To obtain an upper bound for the volume $V(\rho)$ of ρ, consider now the ensembles defined by the mixtures

$$\rho_L(T)=C_T^{-1}\sum_{\alpha\in T} p(\alpha)\rho_\alpha, \quad \rho_L^*(T)=|T|^{-1}\sum_{\alpha\in T}\rho_\alpha, \tag{A6}$$

where $C_T=\Sigma_{\alpha\in T}p(\alpha)$. From the Cartesian property and Eqs. (A2) and (A3) it follows that $V(\rho_\alpha)=[V_I]^L$ is constant, and further that the ρ_α are nonoverlapping. Hence, from the uniformity and additivity properties, $V(\rho_L(T))\leqslant V(\rho_L^*(T)) =|T|[V_I]^L$. Property (T2) then gives

$$V(\rho_L(T))\leqslant[V_I]^L e^{L[S(\rho)+\delta_L]}. \tag{A7}$$

Further, from property (T1) and Eqs. (A4) and (A6),

$$\begin{aligned}\mathrm{Tr}_{\Gamma^L}[|\rho^L-\rho_L(T)|]&=\sum_{\alpha\in T}|p(\alpha)-p(\alpha)/C_T|+\sum_{\alpha\notin T}p(\alpha)\\ &=(1/C_T-1)C_T+(1-C_T)\leqslant 2\epsilon.\end{aligned}$$

Hence ρ^L can be made arbitrarily close to $\rho_L(T)$ for L sufficiently large, and so from the assumed continuity of $V(\ )$, and noting from the Cartesian property that $V(\rho^L) =[V(\rho)]^L$, one has from Eq. (A7) that

$$V(\rho)=\lim_{L\to\infty}[V(\rho_L(T))]^{1/L}\leqslant V_I\, e^{S(\rho)}. \tag{A8}$$

Thus the exponential of the entropy is an upper bound for the ratio of the volume of ρ to the volume of a "pure" state. Note that only properties (T1) and (T2) of T were needed to obtain this result, and that the projection property has not been used.

To obtain the converse of inequality (A8), note from the projection property that

$$V(\rho_L^*(T))\leqslant\prod_{l=1}^{L} V(\overline{\rho_l}(T)), \tag{A9}$$

where $\overline{\rho_l}(T)$ is the projection of $\rho_L(T)$ onto its lth component, i.e.,

$$\overline{\rho_l}(T)=\sum_{\alpha=(i_1,\ldots,i_L)\in T} p(\alpha)\rho_{i_l}. \tag{A10}$$

From property (T4) of T, $\overline{\rho_l}(T)$ is independent of l and hence may be denoted by $\overline{\rho}$. Equation (A9) then becomes $V(\rho_L^*(T))\leqslant[V(\overline{\rho})]^L$. But, as noted earlier, the volume

$V(\rho_L^*(T))$ follows from the additivity property as $|T|[V_I]^L$, and hence via property (T2) of T Eq. (A9) reduces to

$$V_I e^{S(\rho)+\delta_L} \leqslant V(\overline{\rho}). \tag{A11}$$

Further, from Eqs. (A6) and (A10),

$$\overline{\rho} = L^{-1}\sum_l \overline{\rho_l}(T) = |T|^{-1}\sum_{\alpha\in T}\sum_{i\in I} L^{-1}N_i(\alpha)\rho_i, \tag{A12}$$

and hence, from Eq. (A1) and property (T3) of T,

$$\mathrm{Tr}_\Gamma[|\rho-\overline{\rho}|] = |T|^{-1}\mathrm{Tr}_\Gamma\left[\left|\sum_{\alpha\in T}\sum_{i\in I}(p_i - L^{-1}N_i(\alpha))\rho_i\right|\right]$$

$$\leqslant |T|^{-1}\sum_{\alpha\in T}\sum_{i\in I}|p_i - L^{-1}N_i(\alpha)| \leqslant \delta_L'.$$

Hence $\overline{\rho}$ can be made arbitrarily close to ρ for L sufficiently large, and so, taking the limit $L\to\infty$ in Eq. (A11), the assumed continuity of $V(\ )$ gives

$$V_I e^{S(\rho)} \leqslant V(\rho). \tag{A13}$$

Equations (A8) and (A13) yield the theorem of Sec. III B for classical discrete ensembles with finite index sets [where $K(\Gamma)$ in Eq. (30) is identified with the volume V_I of a pure ensemble $\{p_i=\delta_{ij}\}$ on I, and Eq. (32) for $K(\Gamma)$ follows immediately from the Cartesian property]. The extension to ensembles with infinite index sets is trivial by continuity. The distribution $\{p_i\}$ of such an ensemble ρ can be arbitrarily closely approximated by its (renormalized) first M terms, corresponding to a discrete ensemble ρ_M with a finite index set. Hence, from the assumed continuity of ensemble volume and Eqs. (A8) and (A13), $V(\rho)=V_I\lim_{M\to\infty}\exp[S(\rho_M)]$, where V_I is the volume of a "pure" ensemble with respect to the infinite index set I. Thus $V(\rho)$ is as per the theorem [but becomes infinite in the case that the limit of $S(\rho_M)$ as $M\to\infty$ does not exist].

The extension to quantum ensembles is straightforward. Indeed, for quantum ensembles the above analysis goes through formally unchanged, where the expansion in Eq. (A1) is now identified with an orthogonal decomposition into pure states, and the first product in Eq. (A3) is a tensor product. Thus ρ_i and p_i represent (nonoverlapping) eigenstates and eigenvalues of ρ. The only additional consideration is that V_I, the volume of an eigenstate of ρ, might conceivably depend on the eigenstate basis. However this is ruled out by the invariance property (i): *all* pure states on a given Hilbert space can be connected by unitary transformations, and hence have the same volume.

Finally, the theorem may be extended to continuous classical ensembles as follows. Consider a classical ensemble ρ described by a probability distribution $p(\mathbf{x})$ on an n-dimensional space X. This space may be partitioned into a set $\{S_i\}$ of nonoverlapping sets of equal volume V (i.e., $\int_{S_i}d^n\mathbf{x}=V$ for all i). Define the corresponding "pure" ensembles ρ_i by the associated probability distributions $p^{(i)}(\mathbf{x})=1/V$ for $\mathbf{x}\in S_i$ and $=0$ for $\mathbf{x}\notin S_i$. These pure ensembles can be mapped to each other by measure-preserving transformations, and hence from the invariance property have equal ensemble volumes, say $V_0(V)$. The formal analogs of the properties in Eq. (A2) then hold, and again the above analysis for classical discrete ensembles goes through formally unchanged for mixtures of these pure ensembles, i.e.,

$$V\left(\sum_i p_i\rho_i\right) = V_0(V)\exp\left(-\sum_i p_i\ln p_i\right). \tag{A14}$$

Now consider the particular mixture defined by

$$\rho_V = \sum_i p_i(V)\rho_i, \quad p_i(V)=\int_{S_i}d^n p(\mathbf{x}). \tag{A15}$$

Thus ρ_V is a discrete approximation to ρ, and hence, noting that $\int_X d^n\mathbf{x}\equiv\Sigma_i\int_{S_i}d^n\mathbf{x}$, one has from the Mean Value Theorem that

$$\mathrm{Tr}_\Gamma[|\rho-\rho_V|] = \sum_i \int_{S_i}d^n\mathbf{x}|p(\mathbf{x})-p_i(V)/V| \to 0 \tag{A16}$$

in the continuum limit $V\to 0$. Hence, from Eq. (A14) and the assumed continuity of ensemble volume,

$$V(\rho) = \lim_{V\to 0}V_0(V)\exp(S_V), \tag{A17}$$

where S_V denotes the entropy of $\{p_i(V)\}$. But again approximating an integral by a summation,

$$S(\rho) = \lim_{V\to 0} -V\sum_i [p_i(V)/V]\ln[p_i(V)/V] = \lim_{V\to 0}(S_V + \ln V). \tag{A18}$$

Hence Eq. (A17) can be rewritten as

$$V(\rho) = e^{S(\rho)}\lim_{V\to 0}V_0(V)/V. \tag{A19}$$

Finally, to show that the limit exists in Eq. (A19), note that any set $S\in X$ of measure $\int_S d^n\mathbf{x}=V$ can be partitioned into m nonoverlapping sets of equal measure V/m for any integer m. Moreover, a "pure" ensemble on S, corresponding to a distribution which is uniform over S and vanishing outside S, can trivially be written as an equally weighted mixture of analogously defined ensembles for the members of the partition. Hence from the additivity property one has the relation $V_0(V)=mV_0(V/m)$ for the ensemble volumes of "pure" ensembles. Further, replacing V by nV in this relation for any integer n implies that $V_0(rV)=rV_0(V)$ for any rational number $r=n/m$. This can be extended to all real r from the assumed continuity of ensemble volume, so that $V_0(V)/V=\mathrm{const}=K(\Gamma)$, say, and the theorem follows via Eq. (A19).

[1] M.J.W. Hall, e-print quant-ph/9806013.

[2] C.E. Shannon, Proc. IRE **37**, 160 (1949) [reprinted in *Claude Elwood Shannon: Collected Papers*, edited by N. Sloane and A. Wyner (IEEE, New York, 1993), pp. 160–172].

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

[4] M.J.W. Hall, Phys. Rev. A **55**, 100 (1997).

[5] M. Toda, R. Kubo, and N. Saitô, *Statistical Physics I* (Springer, Berlin, 1983), Sec. 2.1.

[6] S.-K. Ma, *Statistical Mechanics* (World Scientific, Singapore, 1985), Secs. 5.2, 25, and 26.

[7] R. Leipnik, Inf. Control. **2**, 64 (1959).

[8] V.V. Dodonov and V.I. Man'ko, in *Invariants and the Evolution of Nonstationary Quantum Systems*, edited by M.A. Markov (Nova, New York, 1989), Sec. 4, pp. 3–101.

[9] M. Zakai, Inf. Control. **3**, 101 (1960).

[10] K. Zyckowski, J. Phys. A **23**, 4427 (1990).

[11] B. Mirbach and H.J. Korsch, Phys. Rev. Lett. **75**, 362 (1995).

[12] B. Mirbach and H.J. Korsch, Ann. Phys. (N.Y.) **265**, 80 (1998).

[13] E. Heller, Phys. Rev. A **35**, 1360 (1987).

[14] W.H. Zurek, S. Habib, and J.P. Paz, Phys. Rev. Lett. **70**, 1187 (1993).

[15] A. Renyi, *Probability Theory* (North-Holland, Amsterdam, 1970), Sec. IX.6, Theorem 4 and Sec. IX.8, Eq. (31).

[16] J.B.M. Uffink, Ph.D. thesis, University of Utrecht, 1990.

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

[18] To derive Eq. (4), define $f(\alpha)=\ln\int dx\, p(x)^{1+\alpha}$, implying $\lim_{\alpha\to 0}\ln L_{X,\alpha}=\lim_{\alpha\to 0}[f(0)-f(\alpha)]/\alpha=-f'(0)$.

[19] R. Ash, *Information Theory* (Wiley, New York, 1965), Chaps. 1, 2, and 6.

[20] This follows from a straightforward variational calculation, and may be generalized to all Renyi lengths.

[21] For the general case of nonoverlapping $p(x)$ and $q(x)$ with equal Renyi lengths $L_{X,\alpha}=L$, one finds for the mixture $\lambda p(x)+(1-\lambda)q(x)$ that $L_{X,\alpha}/L=[\lambda^{1+\alpha}+(1-\lambda)^{1+\alpha}]^{-1/\alpha}$; differentiating with respect to λ yields a maximum of 2, attained for $\lambda=\frac{1}{2}$.

[22] For an example in the context of optical phase, where the rms deviation gives a rather misleading indication of the spread of the (near optimally phase concentrated) coherent phase states, see M.J.W. Hall, J. Mod. Opt. **40**, 809 (1993).

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

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

[25] If the components of **x** are (possibly noncommuting) quantum observables, the expectation values in Eq. (15) should be symmetrized.

[26] C.E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948); **27**, 623 (1948) [also reprinted in *Claude Elwood Shannon: Collected Papers* (Ref. [2]), pp. 5–83].

[27] Of course, in special contexts, restricted interpretations of ensemble entropy are possible: for example, as the logarithm of the number of microstates in statistical mechanics [5], or as the amount of information obtainable by measurement on discrete classical and quantum ensembles [26,36] [see also L. Brillouin, *Science and Information Theory*, 2nd ed. (Academic, New York, 1962), Chap. 1]. Such interpretations are not applicable more generally, however. For example, for continuous classical ensembles the entropy can take any negative value, which is not the case for numbers of microstates and information gain. In contrast, in the geometric interpretation such negative values merely correspond to the possibility of arbitrarily small ensemble volumes.

[28] Units of bits correspond to dividing $S(\rho)$ in Eq. (31) by $\ln 2$.

[29] M. Behara, *Additive and Nonadditive Measures of Entropy* (Wiley, New Delhi, 1990), Sec. 1.1.

[30] M. Ohya and D. Petz, *Quantum Entropy and Its Use* (Springer-Verlag, Berlin, 1993).

[31] V. Vedral and M.B. Plenio, Phys. Rev. A **57**, 1619 (1998).

[32] A.S. Davydov, *Quantum Mechanics*, 2nd ed. (Pergamon, Oxford, 1976), Sec. III.23.

[33] The signals are, in general, represented by ensembles, to take into account any noise processes in the transmitter and channel medium prior to detection at the receiver.

[34] A.S. Holevo, Probl. Inf. Transm. **9**, 177 (1973).

[35] H.P. Yuen and M. Ozawa, Phys. Rev. Lett. **70**, 363 (1993); C.A. Fuchs and C.M. Caves, *ibid.* **73**, 3047 (1994); B. Schumacher, M. Westmoreland, and W.K. Wootters, *ibid.* **76**, 3452 (1996).

[36] B. Schumacher and M.D. Westmoreland, Phys. Rev. A **56**, 131 (1997); A.S. Holevo, e-print quant-ph/9611023.

[37] A complete set of coherent states $\{|\alpha\rangle\}$ corresponds to a joint position and momentum measurement described by the positive-operator-valued measure $\{\pi^{-1}|\alpha\rangle\langle\alpha|\}$ (see, e.g., Ref. [3]); the statistics of the corresponding observable A_Q for state ρ are then given by the Husimi distribution $\pi^{-1}\langle\alpha|\rho|\alpha\rangle$.

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