Journal of Statistical Physics, Vol. 11, No. 4, 1974

Gibbs vs. Shannon Entropies

Richard L. Liboff[1]

Received June 14, 1974

The Gibbs neg-entropy $\eta_G = \int \Pi \ln \Pi$ is compared to the Shannon neg-entropy $\eta_S = \sum p \ln p$. The coarse-grained density is Π, while $\{p\}$ is a probability sequence. Both objects are defined over partitions of the energy shell within a set-theoretic framework. The dissimilarity of these functionals is exhibited through η_G vs. η_S curves. A positive information interpretation of η_G is given referring it to the maximum information contained in the solution to the Liouville equation. The physical relevance of η_G over η_S in classical physics is argued. In quantum mechanics, the fine-grained Shannon entropy remains relevant to the uncertainty principle, while the coarse-grained densities maintain their properties as in the classical case.

KEY WORDS: Coarse-grained density; density operator; energy shell; ensemble points; fine-grained density; Gibbs–Ehrenfest theorem.

1. INTRODUCTION

Although the information approach to statistical mechanics[1] is currently popular, for the most part authors fail to distinguish sufficiently between the Shannon and Gibbs entropies. In this paper, differences between these entropies are exposed and it is concluded that: (a) The information properties relating to these two entropies are quite dissimilar; (b) the Gibbs entropy is more relevant to physics.

In classical physics, the solution to the Liouville equation D is maximally informative.[2] It contains everything there is to know about the system to

This research was supported by the Physics Branch of the Office of Naval Research under contract N 00014-67-A-0077-0015.

[1] Department of Applied Physics and School of Electrical Engineering, Cornell University, Ithaca, New York.

[2] The most general solution to the Liouville equation contains all the constants of the motion. For further details see Ref. 2.

which it pertains. If A is any dynamical variable, then the average of A at the time t is

$$\langle A \rangle = \int_{\mathbf{E}} D(t) A \, d\mathbf{E}$$

with D normalized to unity. The integral extends over the set $\mathbf{E}$ of phase points which comprise the energy shell in which D is defined. If this energy shell is partitioned, one can construct a coarse-grained distribution Π (after Gibbs[3]) whose value in a cell of the partition is the average of D in that cell. The better Π approximates D, the more information it contains. A measure of this information is given by the Gibbs neg-entropy η_{G},

$$\eta_{\mathrm{G}} \equiv -S_{\mathrm{G}}/\kappa = \int_{\mathbf{E}} \Pi \ln \Pi \, d\mathbf{E} = \sum \Pi_i E_i \ln \Pi_i \tag{1}$$

where κ is Boltzmann's constant and S denotes entropy. The integral extends over the set $\mathbf{E}$ of phase points which comprise the energy shell. The measure (volume) of the ith cell is E_i. The larger is η_{G}, the more informative is Π.

An alternate means of qualifying the information properties of Π is as follows. If D is normalized to unity, then

$$p_i = \Pi_i E_i \tag{2}$$

is the probability of finding the system in a state in the ith cell. The volume of this cell (as noted above) is E_i. Given a sequence of probabilities $\{p_i\}$, with what certainly can one say what state the system is in at a given time? A measure of the related *uncertainty* is given by the Shannon entropy,[4]

$$S_{\mathrm{S}} = -\kappa \sum p_i \ln p_i \tag{3}$$

If $\{p_i\}$ is a uniform sequence, then this uncertainty is maximum. If, on the other hand, say, $p_i = \delta_{i1}$, then this uncertainty is minimum (zero). For this latter case one is certain that measurement finds the system in a state in the first cell.

The Shannon entropy finds important application in communications engineering.[3] In addition to the above interpretation, within communications engineering one also relates the Shannon entropy to the average number of bits per symbol used in encoding messages. Within this scheme, p_i is the probability of the ith element of an alphabet. It is the frequency with which the ith element of the alphabet occurs in a random sampling of messages. In the extreme that, say, the first letter of the alphabet predominates in all messages ($p_i = \delta_{i1}$), messages carry minimal information and $S_{\mathrm{S}} = 0$. In the

[3] For application of the Shannon entropy to communications engineering see Ref. 5.

opposite extreme that all n letters of the alphabet occur equally often, $p_i = 1/n$ and messages contain maximum information; $S_S = \kappa \ln n$.

Of the two interpretations of S_S relevant to communications engineering outlined above, the information-entropy association predominant in physics is the former. Namely, the Shannon neg-entropy, $-S_S$, is a measure of the *certainty* regarding the state of a system, given the sequence $\{p_i\}$.

In the analysis to follow, we investigate the differences between the neg-entropies η_G and η_S ($\eta_S = -S_S/\kappa$). They are examined as functionals of the microdistribution D and the partition on which Π and p are defined. (Here we are denoting $\{p_i\}$ by p. We have already used Π to denote $\{\Pi_i\}$.) The Gibbs neg-entropy varies from zero (least informative about the structure of D) to η_D, where η_D is the fine grained Gibbs neg-entropy,

$$\eta_D = \int D \ln D \, d\mathbf{E} \tag{4}$$

The functional η_G assumes this maximum value on the infinitely refined partition (denoted by M^* below). On this partition $\Pi = D$, and is maximally informative.

The Shannon neg-entropy ranges from $-\infty$, where the aggregate of p_i values is such that one is maximally uncertain (or minimally certain) about the outcome of an experiment on the state of the system, to zero, where this uncertainty is minimum.

Both η_G and η_S are zero for the unit partition. This is the partition consisting of a single cell, i.e., the whole energy shell. These functionals both assume their maximum magnitudes on the infinitely refined partition M^*. The fact that $\eta_S \to -\infty$ in this limit is of small physical relevance. The same is true of the zero value of η_S on the unit partition (one is certain that the system is in a state on the energy shell). On the other hand, the fact that η_G is minimum on the unit partition reveals that Π is least informative on this partition, while the maximum value which η_G assumes on M^* indicates that Π is maximally informative on this partition. Given this latter value of Π (i.e., D) one is most knowledgeable about the properties of the system to which Π pertains.

Both η_S and η_G change monotonically with successive refinement of partition. However, for a given fine-grained distribution D, η_G is not a single-valued function of η_S. The partition on which Π and p are defined enters as a parameter in a plot of η_G vs. η_S. One obtains a single η_S vs. η_G curve for a unique sequence of refinements of increasing order (of the unit partition). These curves are bounded between the axes: (a) $D = \delta(z - z_0)$. For this value of D, η_G varies from 0 to $+\infty$ along the line $\eta_S = 0$. (b) $D = 1$. For this case η_S varies from 0 to $-\infty$ along the line $\eta_G = 0$. (The phase coordinate of the system is z.)

Similar constructions exist in quantum mechanics, where the formalism is centered about the maximally informative density operator $\hat{\rho}$, whose diagonal elements are probabilities. The coarse-grained entropies maintain their classical properties with respect to partitions of the spectrum of energy eigenvalues. Properties of the fine-grained Shannon entropy are discussed which are relevant to the uncertainty principle.

2. ANALYSIS: THE CLASSICAL CASE

2.1. Definitions

Consider the set of points $\mathbf{E}$

$$\mathbf{E} = \{z | \mathscr{E} \leqslant H(z) \leqslant \mathscr{E} + \Delta\mathscr{E}\} \tag{5}$$

where the spread in energy $\Delta\mathscr{E}$ is vanishingly small compared to $\mathscr{E}$. The Hamiltonian of the system is H. For a system with N degrees of freedom, z is a $2N$-dimensional vector in Cartesian Γ space ($\mathbb{R}^{2N}$). In this space, the set $\mathbf{E}$ is called the *energy shell.* Any simply connected subset $\mathbf{E}_n$ of $\mathbf{E}$ has a well-defined Euclidean diameter d_n in Γ space. The measure of the set $\mathbf{E}_n \leqslant \mathbf{E}$ is the normalized volume (Lebesgue) integral[4]

$$E_n = (1/\Omega)\int_{\mathbf{E}_n} d\mathbf{E}; \qquad \Omega \equiv \int_{\mathbf{E}} d\mathbf{E} \tag{6}$$

The dimensions of Ω are those of an element of volume in Γ space, while the measure E_n is dimensionless. In these "units," the volume of the energy shell is unity ($E = 1$).

A partition M of $\mathbf{E}$ is the collection of (simply connected) sets

$$M = \left\{ \mathbf{E}_n{}^M \Big| \bigcup_n \mathbf{E}_n{}^M = \mathbf{E}; \quad \mathbf{E}_n{}^M \cap \mathbf{E}_k{}^M = \varnothing, \quad E_n{}^M > 0 \right\} \tag{7}$$

The *order* of the partition M is the number of sets in M and is written ν_M. A partition M is *finite* if ν_M is greater than one. The empty set is $\varnothing$.

If $D(z)$ is the density of ensemble points in $\mathbf{E}$ ($D \geqslant 0$), then the coarse-grained density Π is a function of M and z and is defined by

$$\Pi(z, M) = E_n^{-1}\int_{\mathbf{E}_n} D\, d\mathbf{E}, \quad z \in \mathbf{E}_n \in M; \qquad \Pi = \Pi_n \quad \text{for} \quad z \in \mathbf{E}_n \tag{8}$$

If

$$\int_{\mathbf{E}} D\, d\mathbf{E} = 1 \tag{9}$$

[4] One can also work with the energy-surface, $E = \{z | \mathscr{E} = H(z)\}$ with the measure $\int_{\mathbf{E}} dz/|\nabla H| = E$. Both measures are preserved under the dynamical motion of the system. The integral (6) is a Poincaré invariant.

then $D\,d\mathbf{E}$ is the probability of finding the system in the state $z \in d\mathbf{E}$ and D may be termed a probability density. (This property of D is assumed in the remainder of the analysis.) For D a probability density,

$$p_n = E_n\Pi_n = \int_{\mathbf{E}_n} D\,d\mathbf{E} \leqslant 1 \tag{10}$$

is the probability of finding the system in a state in $\mathbf{E}_n$. The sequence $\{p_n\}$ satisfies the probability norm,

$$\sum p_n = \sum E_n\Pi_n = \int_{\mathbf{E}} D\,d\mathbf{E} = 1 \tag{11}$$

Next we turn to the definition of the *refinement* of the partition M. If all the elements of the partition M are unions of the elements of the partition N, i.e.,

$$\mathbf{E}_n{}^M = \bigcup_{i=1}^{s} \mathbf{E}_i{}^N, \qquad \forall \mathbf{E}_n{}^M \in M \tag{12}$$

then N is a *refinement* of M and is denoted by

$$N \supset M$$

All partitions are refinements of the unit partition which consists of the whole set $\mathbf{E}$. The unit partition is denoted by I.

If D is constant over the sets of a partition M, it is a *simple* set function, and has the representation

$$D = \sum_{\mathbf{E}_n} D_n \chi_{\mathbf{E}_n}(z) \tag{13}$$

where $\chi_{\mathbf{E}_n}$ is the characteristic set function, and the D_n are constants. If D is simple on a finite partition M, it is clear that

$$D(z) = \Pi(z; M) \tag{14}$$

When D is simple on M, then any refinement of M is said to be *equivalent to M with respect to D.*

The su-partition M^* is defined as

$$M^* = \lim_{\substack{d_M \to 0 \\ \nu_M \to \infty}} M, \qquad \nu_{M^*} = \infty \tag{15}$$

In this expression d_M is the maximum diameter of all sets in M and ν_M is the order of M (defined above).

A sequence of partitions $\{M_i\}$ such that

$$M_1 \subset M_2 \subset M_3 \cdots$$

is called a nested sequence.

2.2. Entropies

With these definitions established we now turn to comparison of the two entropy functions briefly mentioned in the introduction. To repeat, the Gibbs coarse-grained neg-entropy is the functional

$$\eta_G(\Pi, M) = \sum_{i=1}^{\nu_M} \Pi_i E_i \ln \Pi_i \tag{16}$$

It is dependent on the distribution Π as well as the partition M. The Shannon neg-entropy is

$$\eta_S(p, M) = \sum_{i=1}^{\nu_M} p_i \ln p_i \tag{17}$$

Using (2), one obtains

$$\eta_G = \eta_S + \sum p_i \ln E_i^{-1} \equiv \eta_S + \Delta \tag{18}$$

Since $E_i \leqslant 1$, the increment Δ is positive. Microdistributions are easily constructed for which one can construct distinct measure sequences $\{E_i\}$, while maintaining the sequence $\{p_i\}$. In such cases there are many values of Δ corresponding to a single value of η_S, so that in general, η_S is not a single-valued function of η_G.

If the partition M is finite, then

$$-\ln \nu_M \leqslant \eta_S \leqslant 0 \tag{19}$$

For the su-partition M^*, $\ln \nu_{M^*} = +\infty$, and

$$-\infty \leqslant \eta_S \leqslant 0 \tag{20}$$

The right equality is assumed only if $p_i = 1$ for any i. In this case one associates a minimum uncertainty with the sequence $\{p_i\}$. This value is always assumed on the unit partition I.

The bounds on η_G appear as

$$0 \leqslant \eta_G \leqslant \eta_D \tag{21}$$

The right equality is assumed only: (a) on any refinement of M provided D is simple on M; (b) on the partition M^* (since $\Pi = D$ on M^*). The left equality is assumed only: (a) on the partition I; (b) on the partition M provided Π is uniform on M. If Π is uniform on M, then $\Pi = 1$, due to the fact that $\sum \Pi_i E_i = \Pi \sum E_i = \Pi = 1$. This latter condition together with (a) above lead to the Gibbs–Ehrenfest theorem.[3,6] The significance often ascribed to this theorem is as follows. Suppose initially at $t = 0$, D is simple on M. Then $D = \Pi$ on M. Under the natural displacement of the system in time, if D changes, η_G does not increase with respect to the partition M,

$$\eta_G(t) \leqslant \eta_G(0)$$

For this statement to be of any physical relevance, one must show that $\mathbf{E}$ is metrically indecomposable.[2,7] This in turn guarantees that D distorts.

Next we consider the dependence of η_S and η_G on refinement of partition. For a given fine-grained density D, under a refinement of partition, η_S decreases (one is less certain about the outcome of an experiment on the state of the system) and η_G increases (Π better approximates D, whence it is a more informative object). That is, if N is a refinement of M,

$$N \supset M$$

then

$$\eta_S(N) \leqslant \eta_S(M) \tag{22}$$

and

$$\eta_G(N) \geqslant \eta_G(M) \tag{23}$$

Thus η_S and η_G vary monotonically with increasing order of refinement of a given partition. The proof of the first statement is quite well known, and proceeds as follows. Let $N = M$ except on the first cell of M. This cell is divided into two cells so that the p values on $\mathbf{E}_1$ become p_1' and p_2' on $\mathbf{E}_1'$ and $\mathbf{E}_2'$, respectively, where $\mathbf{E}_1 = \mathbf{E}_1' \cup \mathbf{E}_2'$ and $p_1 = p_1' + p_2'$. In this case we can write

$$\eta_S(N) = \eta_S(M) + p_1' \ln(p_1'/p_1) + p_2' \ln(p_2'/p_1) \leqslant \eta_S(M) \tag{24}$$

which is clearly the case since the arguments of both log functions are less than one. The equality holds only if p_1' or p_2' is zero.

Unders similar conditions, to establish (23), we write

$$\begin{aligned} \eta_G(N) &= \eta_G(M) + p_i' \ln(\Pi_1'/\Pi_1) + p_2' \ln(\Pi_2'/\Pi_1) \\ &\equiv \eta_G(M) + p_1 \ln \Lambda \geqslant \eta_G(M) \end{aligned} \tag{25}$$

where we have labeled

$$\Lambda \equiv \frac{w^w}{(1 + w)^{w+1}} \frac{(1 + x)^{w+1}}{x^w}, \qquad w \equiv \frac{p_2'}{p_1'} > 0, \qquad x \equiv \frac{E_2'}{E_1'} \geqslant 0$$

Thus to establish (23) we must show that $\Lambda \geqslant 1$, or equivalently that

$$(1 + x)^{w+1}/x^w \geqslant (1 + w)^{w+1}/w^w \tag{26}$$

To demonstrate the validity of this inequality, we consider the function

$$f(x) = (1 + x)^{w+1}/x^w$$

For (26) to be true, $f(x)$ must be minimum at $x = w$, which is shown by forming the first two derivatives of f:

$$f'(x) = [f(x)/x(x + 1)](x - w)$$

It follows that $f' = 0$ only at $x = w$.

$$f''(x) = [w(w + 1)/x^2(x + 1)^2]f(x)$$

At $x = w$

$$f''(w) = f(w)/w(w + 1)$$

We conclude that

$$f(x) > f(w), \qquad x \neq w \quad (\Pi_1' \neq \Pi_2')$$

and $f(w)$ is the only minimum of $f(x)$, whence (26) is true and (23) is established. The equality in (25) holds only if $\Pi_1' = \Pi_2'$, that is, if Π maintains its uniform structure in $\mathbf{E}_1^M$ under refinement of M. Thus we recapture the Gibbs–Ehrenfest theorem that η_G is minimum for the uniform coarse-grained distribution.[5]

With these properties established we next turn to constructing sketches of η_S vs. η_G curves for given microdistributions D and nested partition sequences of the unit partition. The first two cases serve to determine the η_G–η_S axes.

Case (a). For this case

$$D = \delta(z - z_0) \tag{27}$$

For all partitions M such that $z_0 \in \mathbf{E}_\kappa \in M$ for some κ, one obtains that $p_\kappa = 1$, $p_{\kappa'} = 0$, $\kappa' \neq \kappa$. For such a sequence of p values $\eta_S = 0$. It is maximally certain that measurement finds the state of the system in the κth cell of the partition. However, since η_S is insensitive to volume of partition cells it does not detect that with successive refinement of partition, the κth cell shrinks about the state which the system is in.

For this same sequence of partitions η_G is initially zero on the unit partition, where the value $\Pi = 1$ is least related to D. On the final partition M^*, η_G is infinite and $\Pi = D$ is maximally informative. A sketch of these values is shown in Fig. 1.

Case (b). For this case

$$D = 1 \tag{28}$$

It is uniformly probable to find the system in any state in the energy shell. Since D is simple on all partitions, $D = \Pi$ on all partitions and $\eta_G = \eta_D = 0$. The coarse-grained density is equally informative on all partitions. The Shannon entropy, on the other hand, has the value $\eta_S = 0$ on the unit partition and decreases uniformly with increasing refinement of the unit partition until the value $\eta_S = -\infty$ is attained on M^*. At each level of partition refinement, the Shannon entropy sees the corresponding sequence of

[5] A more direct proof of (25) follows from the Gibbs–Ehrenfest theorem (H. Grad, private communication). However, the proof in the text maintains a parallel structure with the proof of (24) and keeps the paper somewhat self-contained.

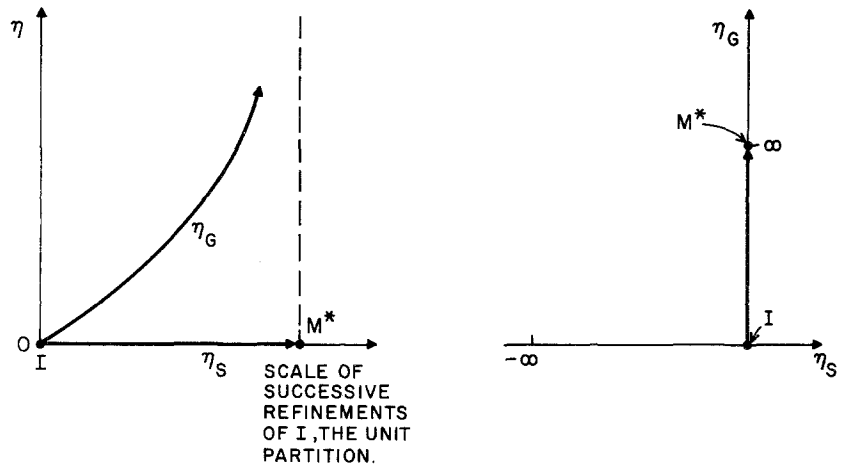

Fig. 1. η_S vs. η_G curves for case (a): $D = \delta(z - z_0)$.

uniform probabilities as carrying a maximum uncertainty which grows logarithmically in magnitude with increasing partition order. A sketch of these values is shown in Fig. 2.

Case (*c*). For this case, D is simple on the finite partition M. To exhibit the difference between the two said entropies for this case, we construct a

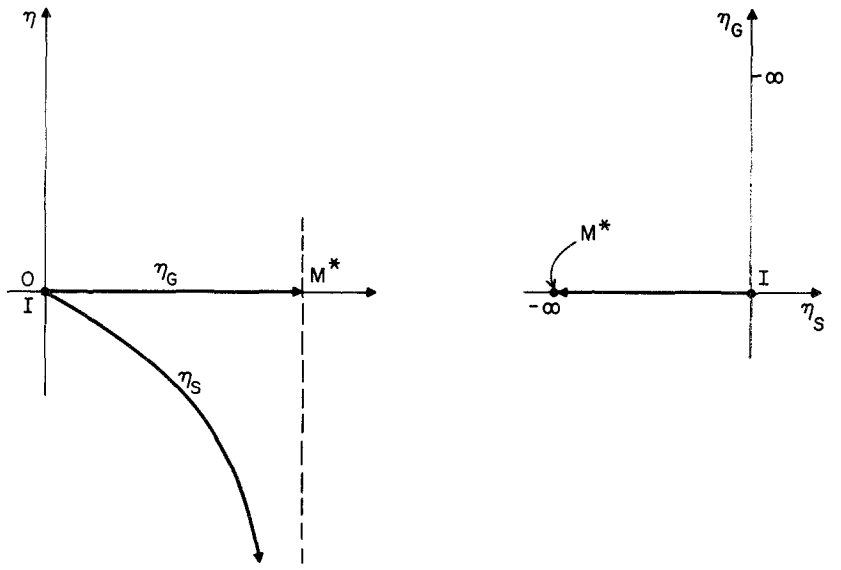

Fig. 2. η_S vs. η_G curves for case (b): $D = 1$.

nested sequence of partitions whose first element is the unit partition and which includes the partition M. On M, $\Pi = D$ and η_G assumes its maximum value η_D. It maintains this value on all successive refinements of M. This property reflects the fact that Π is equally informative on all refinements of M.

On the other hand, the Shannon entropy is relatively insensitive to the fact that D is simple on M. It decreases monotonically with increasing partition order. A sketch of these values is depicted in Fig. 3.

Case (d). This case is offered to illustrate a more peculiar behavior of the η_S vs. η_G curve. It and the case to follow are significant in that they exhibit a D value for which $\eta_G = 0$ is not uniquely attached to $\eta_S = 0$. Let M be a partition of **E** which contains cells of equal measure,

$$E^M = 1/\nu^M$$

where the order ν^M is even. On one-half of these cells D has the value

$$D = 1/(\nu_M/2)E^M = 2 \qquad (p = 2/\nu_M) \tag{29}$$

and on the other half

$$D = 0 \qquad (p = 0)$$

so that D maintains its normalization. These D values form a symmetric array on **E** so that the environment about any cell in which $D > 0$ is the same. Under such conditions a nested sequence of partitions exists such that

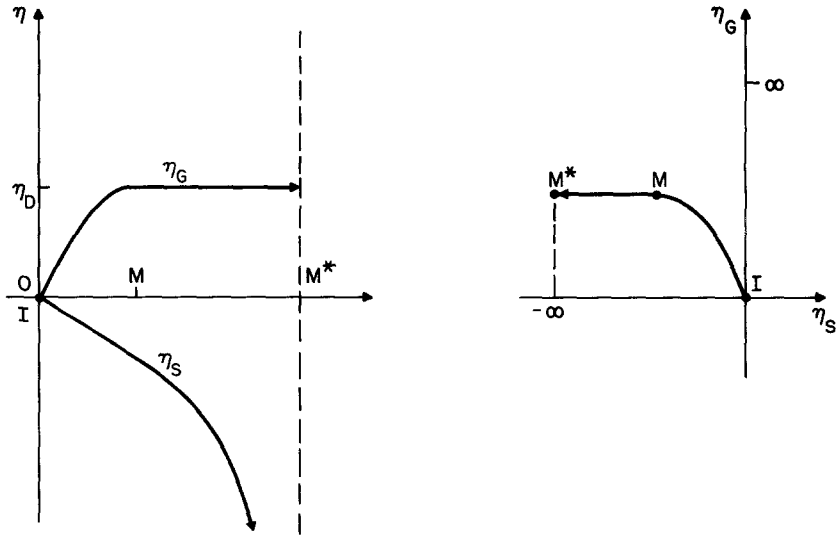

Fig. 3. η_S vs. η_G curves for case (c): D simple on M.

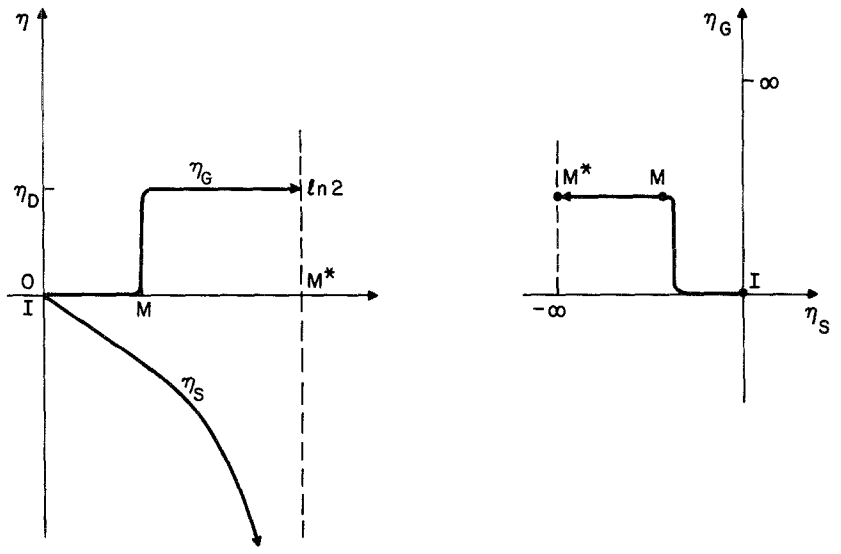

Fig. 4. η_S vs. η_G curves for case (d): $D = 2$, uniformly distributed over half of **E**. $D = 0$ on other half.

on each partition $\Pi = 1$ on all cells and $\eta_G = 0$. This sequence of partitions can be constructed to include M. When this value of M is attained,

$$\eta_G = \eta_D = \ln 2$$

This value of η_G is maintained for all successive refinements.

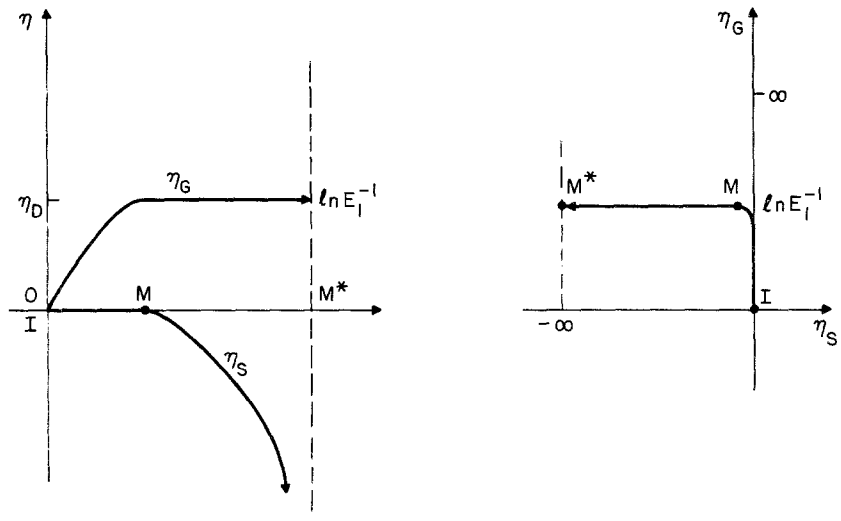

Fig. 5. η_S vs. η_G curves for case (e): $D = E_1^{-1}$ for $z \in \mathbf{E}_1 \subset \mathbf{E}$; $D = 0$ elsewhere on **E**.

The Shannon entropy again merely decreases monotonically with increasing partition order. The corresponding sequence of η_S vs. η_G values is sketched in Fig. 4.

Case (*e*). This is the opposite case to the one just cited. The micro-distribution is nonzero on a simply connected subset $\mathbf{E}_1$ of $\mathbf{E}$ and zero on the complement of $\mathbf{E}_1$, so that

$$\begin{aligned} D &= 1/E_1 \qquad \text{for} \quad z \in \mathbf{E}_1 \\ D &= 0 \qquad\quad \text{elsewhere} \end{aligned} \tag{30}$$

For all partitions M that include an element $\mathbf{E}^M$ such that $\mathbf{E}_1 \subset \mathbf{E}^M$ ($E_1 < E^M$), the Shannon entropy vanishes, while

$$\eta_G = -\ln E^M > 0$$

For partitions that include $\mathbf{E}_1$ as an element, η_G assumes its maximum value,

$$\eta_G = \eta_D = -\ln E_1$$

Thus for a sequence of nested partitions $\{M\}$ with the above cited property and such that in successive refinements

$$\mathbf{E}^M \to \mathbf{E}_1$$

$\eta_S = 0$ until $\mathbf{E}^M = \mathbf{E}_1$, when it begins to decrease with increasing partition order. The Gibbs entropy increases until $\mathbf{E}^M = \mathbf{E}_1$, at which point it assumes its maximum value η_D. These properties are sketched in Fig. 5.

The superposition of these η_S vs. η_G curves is shown in Fig. 6.

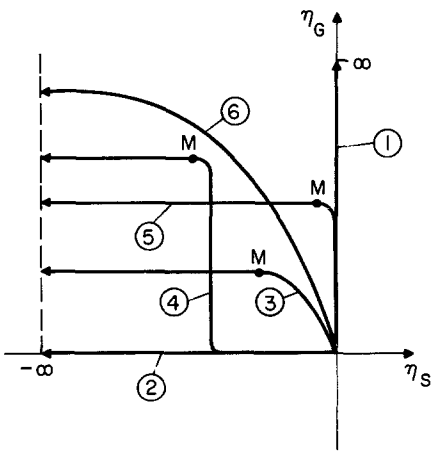

Fig. 6. η_S vs. η_G for the following cases: 1. $D = \delta(z - z_0)$. 2. $D = 1$. 3. D is simple on M. 4. D is uniform on half of $\mathbf{E}$ (see Fig. 4). 5. $D \neq 0$ on one cell of M. 6. The general case.

3. QUANTUM MECHANICS

Quantum statistical mechanics[6] involves the density operator $\hat{\rho}$. This operator is relevant to a system which at a given instant of time is in one of a possible number of states. If the probability that the system is in the kth state is p_k and $|k\rangle$ is the normalized eigenket of this state, then the density operator for this system is

$$\hat{\rho} = \sum_k |k\rangle p_k \langle k| \tag{31}$$

This operator is diagonal in the $\hat{K}$ representation, with eigenvalues p_k. The expectation of an observable A of this system is

$$\langle A \rangle = \mathrm{Tr}\, \hat{\rho}\, \hat{A}, \qquad \mathrm{Tr}\, \hat{\rho} = 1 \tag{32}$$

At this fine-grained level, maximum information about the system is contained in the density operator $\hat{\rho}$. For situations where states are so closely spaced that it is difficult to differentiate between neighboring states, it proves convenient to introduce a coarse-grained density operator (after Pauli[10]). Let the discrete set which is the spectrum of energy eigenvalues for such a system be denoted by $\mathbf{E}$. This set is partitioned into subsets (cells) $\mathbf{E}_i$ such that $\mathbf{E} = \bigcup \mathbf{E}_i$ and $\mathbf{E}_i \cap \mathbf{E}_i = \varnothing$. The "volume" of the set $\mathbf{E}_i$ is the counting measure, i.e., the number of elements in $\mathbf{E}_i$. Over this partition, the coarse-grained density operator has diagonal elements

$$\Pi_i = (1/E_i) \sum_{k \in \mathbf{E}_i} p_k = \langle p \rangle_{i^{\text{th}}\,\text{cell}}$$

The probability of finding the system in the ith cell is

$$P_i = \Pi_i E_i$$

The Shannon and Gibbs neg-entropies appear as

$$\eta_{\mathrm{S}} = \sum P_i \ln P_i; \qquad \eta_{\mathrm{G}} = \sum E_i(\Pi_i \ln \Pi_i) \tag{33}$$

The distinctions between these functionals remain as in the classical case considered above. The Gibbs neg-entropy increases with refinement of partition, rendering Π a more informative distribution.

However, the neg-entropy popular in quantum mechanics is neither of the above functionals but rather that introduced by Pauli (and discussed at length by Tolman[11]). It appears as

$$\eta_{\mathrm{P}} = \sum \Pi_i \ln \Pi_i \tag{34}$$

[6] Good reviews of quantum statistics can be found in Refs. 8 and 9.

This functional as well as η_S decreases with refinement of partition ($\Pi = p$ on M if p is simple on M). The analog of the Gibbs–Ehrenfest theorem in quantum mechanics is referred to as Klein's theorem.[12],[7] With p simple on M at $t = 0$, if p changes, η_P does not increase

$$\eta_P(t) \leqslant \eta_P(0)$$

This property is shared by the Gibbs neg-entropy.

Another entropy encountered in quantum mechanics[11] is the Shannon neg-entropy relevant to the fine-grained distribution $\{p_k\}$,

$$\bar{\eta}_S = \sum p_k \ln p_k = \operatorname{Tr} \hat{\rho} \ln \hat{\rho}$$

(with $\hat{\rho}$ diagonal in the $\hat{K}$ representation). This functional finds useful interpretation with respect to the uncertainty principle. If a system is initially in a pure state of the observable $\hat{K}$, then

$$\hat{\rho} = |k\rangle\langle k|, \qquad \operatorname{Tr} \hat{\rho} \ln \hat{\rho} = 0$$

and $\bar{\eta}_S$ is maximized. One is most certain regarding the outcome of measurement of the observable $\hat{K}$. If measurement is then performed on the observable $\hat{G}$, where $[\hat{G}, \hat{K}] \neq 0$, then subsequent to this measurement, $\bar{\eta}_S$ (in the $\hat{K}$ representation) decreases. On the other hand, if $\hat{G}$ is compatible with $\hat{K}$, then subsequent to the measurement of $\hat{G}$, $\bar{\eta}_S$ is left unchanged (in either $\hat{K}$ or $\hat{G}$ representation, and counting degenerate states only once).

Similarly one finds that if the system is in a mixed state with respect to the degenerate observable $\hat{K}$, then $\bar{\eta}_S$ decreases under a degeneracy-removing perturbation (the perturbation effects a refinement of the partition on which the distribution $\{p_k\}$ is defined), providing the perturbation leaves the system in a mixed state with nondegenerate p_k values maintained.

4. CONCLUSION

Investigation of the properties of the Gibbs and Shannon entropies shows them to be dissimilar in mathematical properties as well as physical relevance. The Shannon entropy is consistently a measure of the uncertainty related to an experiment on the state of a system. This uncertainty increases with increasing refinement of partition. This is so even for the case where the microdistribution is uniform over the whole energy shell, in spite of the fact that coarse-grained distributions over all partitions are equally informative for this case. The Gibbs neg-entropy, on the other hand, is consistently a measure of the information contained in the coarse-grained Π distribution. This is illustrated by way of calculating these entropies as function of succesive refinement of partition for various representative D functions.

[7] For a more inclusive analysis of the properties of the density matrix which includes the coupling of the state of the system to that of the measuring apparatus see Ref. 13.

Properties of the fine-grained Shannon neg-entropy $\bar{\eta}_S$ are shown to be relevant to the uncertainty principle. Examples of successive measurements of commuting and noncommuting observables, respectively, show this entropy to be a consistent measure of the uncertainty produced by such measurement.

ACKNOWLEDGMENTS

Fruitful discussions with T. Fine, A. Rueda, H. Grad, C. Dum, and N. Rostoker are gratefully acknowledged.

REFERENCES

1. A. Katz, *Principles of Statistical Mechanics*, W. H. Freeman, San Francisco (1967); E. A. Jackson, *Equilibrium Statistical Mechanics*, Prentice Hall, Englewood Cliffs, New Jersey (1968); L. Brillouin, *Science and Information Theory*, Academic Press, New York (1956); M. Tribus, *Thermostatics and Thermodynamics*, Von Nostrand, Princeton, New Jersey (1961); D. ter Haar, *Elements of Statistical Mechanics*, Holt, Rinehart and Winston, New York (1954); E. T. Jaynes, *Phys. Rev.* **106**:620 (1957).
2. R. L. Liboff, *Introduction to the Theory of Kinetic Equations*, John Wiley, New York (1969).
3. J. W. Gibbs, *Elementary Principles in Statistical Mechanics* (Vol. II of his collected works), Chapter XII, New Haven (1948).
4. C. E. Shannon, *Bell. Syst. Tech. J.* **27**(I):379 (1948); *Proc. IRE* **37**:10 (1949).
5. H. R. Raemer, *Statistical Communication Theory and Applications*, Prentice Hall, Englewood Cliffs, New Jersey (1969).
6. P. Ehrenfest and T. Ehrenfest, in *Encyklopadie der Mathematischen Wissenshaften IV*, part 32, Leipzig–Berlin (1911); *Physik Z.* **8**:311 (1907).
7. M. Kac, *Probability and Related Topics in Physical Sciences*, Interscience, New York (1959).
8. D. ter Haar, *Rev. Mod. Phys.* **27**:289 (1955).
9. V. Fano, *Rev. Mod. Phys.* **29**:74 (1957).
10. W. Pauli, in *Sommerfeld Festschrift*, Leipzig (1928).
11. R. C. Tolman, *The Principles of Statistical Mechanics*, Oxford, New York (1938).
12. O. Klein, *Z. Physik* **72**:767 (1931).
13. K. Gottfried, *Quantum Mechanics*, Vol. 1, Chapters IV and V, Benjamin, New York (1966).