Uncertainty, Entropy, and the Statistical Mechanics of Microscopic Systems

Richard Blankenbecler
Stanford Linear Accelerator Center, Stanford University, Stanford, California, 94305

and

M. Hossein Partovi
Department of Physics, California State University, Sacramento, California 95819,[(a)] and Stanford Linear Accelerator Center, Stanford University, Stanford, California 94305
(Received 7 September 1984)

Noting that quantum measurements are in general incomplete, we develop, starting from a recent entropic formulation of uncertainty, a *maximum uncertainty principle* to define the statistical mechanics of microscopic systems. The resulting ensemble entropy coincides with the expression of von Neumann, thus providing a unified, quantum basis for statistical physics of all systems. Examples involving momentum and position measurements are discussed.

PACS numbers: 05.30.Ch, 03.65.Bz, 05.20.−y

Quantum measurements[1] are as a rule incomplete in that they fail to provide an exhaustive specification of the state of a system. A quantum-mechanical system is, in general, described by a density matrix,[2] $\hat{\rho}$, whose complete specification requires the measurement of N^2-1 independent elements, N being the dimensionality of the relevant Hilbert space. Clearly, except for those cases where the density-matrix description is used for only a finite subspace (e.g., spin substates) of the full Hilbert space, N will be infinite and $\hat{\rho}$ inaccessible to a complete measurement. A pure state is therefore an idealization, since otherwise it would constitute an example of a completely measured system in an infinite Hilbert space.[3] Hence, the realizable (or preparable) states of a system must be considered members of an ensemble, and so there arises the basic problem of the assignment of $\hat{\rho}$ on the basis of the incomplete information obtained from the preparation process. While this is a problem of quantum statistics at a very basic level and therefore fundamental to both quantum theory and statistical mechanics, there is a distinct lack of a systematic treatment of it in the literature,[4] principally because of the customary restriction to idealized measurements. The object of this note is to present such a treatment, and to demonstrate the fundamental nature of the results that follow from it. The underlying principle will be the equality of *a priori* probabilities[5] (EAP), implemented on the basis of a recently formulated definition of the entropy of a quantum measurement.[6] This formulation will in particular imply the standard expression for the entropy of an ensemble, thus providing a unified, quantum-mechanical basis for microscopic and macroscopic, equilibrium and nonequilibrium statistical mechanics.

Let us recall the formulation of *measurement entropy* in Ref. 6. In general, the measurement of an observable $\hat{A}$, in the state $\hat{\rho}$, is accomplished by means of a measuring device D^A which will include a partitioning of the spectrum of $\hat{A}$ into a number of subsets α_i^A called "bins." [7] The result of the measurements are summarized in a set of probabilities, $\mathscr{P}_i^A$, of finding the outcome of the measurement to be within the given bin α_i^A. This partitioning of the spectrum of $\hat{A}$ generates a corresponding one of the Hilbert space into (orthogonal) subspaces $\mathscr{M}_i^A$ together with a set of projections $\hat{\pi}_i^A$. The entropy (in units of the Boltzmann constant) of such a measurement was defined in Ref. 6 as

$$S(\hat{\rho}|D^A) = -\sum_i \mathrm{tr}(\hat{\rho}\hat{\pi}_i^A)\ln[\mathrm{tr}(\hat{\rho}\hat{\pi}_i^A)], \qquad (1)$$

and shown there to be a suitable measure of quantum-mechanical uncertainty. Note that measurement entropy is a joint property of the *system and the measuring device*.

The problem to be solved is this: A quantum system, measured (or prepared) to have the set of probabilities $\{\mathscr{P}_i^\nu\}$, where ν labels the measured observables $\hat{A}^\nu$ of the system, is to be assigned a density matrix according to EAP. As usual, this requires the identification of an *ensemble entropy, S*, as a measure of uncertainty (or lack of information) about the system.[8] The desired $\hat{\rho}$ would then be so determined as to maximize this uncertainty/entropy. According to (1), then, we must identify two ingredients, an observable (a self-adjoint operator), $\hat{W}$, and an associated measuring device, D^W, the two of which would jointly serve to define the ensemble entropy as $S=S(\hat{\rho}|D^W)$. It is immediately clear that lack of information must be gauged against the most accurate measuring device available. This requires the device to be as finely binned as possible; such an *idealized* device will be denoted by $D^W_{\max}$.[9] The observable $\hat{W}$, on the other hand, should be identified as the operator which embodies the greatest amount of information, hence the least uncertainty/entropy, about the system. This proposition, that $\hat{\rho}$ is to be determined so that the least uncertain observable of the system has the maximum possible measurement entropy, will be referred to as the *maximum uncertainty principle.* It is the quantum expression of the preposition that the most-probable state is that which embodies what is known and none else. In symbols, $S=\inf_{\hat{W}} S(\hat{\rho}|D^W_{\max})$. Now it can be shown mathematically[10] that the above infimum is obtained for $\hat{W}=\hat{\rho}$, so that $S=S(\hat{\rho}|D^\rho_{\max})$. In other words, our measure of entropy has singled out $\hat{\rho}$ as the best determined observable of the system, as indeed it must. That the measurement entropy of the density matrix is to be designated as the measure of our lack of information about the system is in perfect accord with its significance as the depository of all available information about the system.

Since $\hat{\rho}$ has a strictly discrete spectrum (recall that $\mathrm{tr}\hat{\rho}=1$ and $\mathrm{tr}\hat{\rho}^2\leq 1$), the projection $\hat{\pi}_i^\rho$ corresponding to $D^\rho_{\max}$ is simply $|i\rangle\langle i|$, $|i\rangle$ being the ith eigenvector of $\hat{\rho}$,[6] implying therefore that

$$S=S(\hat{\rho}|D^\rho_{\max}) = -\sum_i \mathrm{tr}(\hat{\rho}\hat{\pi}_i^\rho)\ln[\mathrm{tr}(\hat{\rho}\hat{\pi}_i^\rho)] = -\mathrm{tr}(\hat{\rho}\ln\hat{\rho}). \tag{2}$$

This is the familiar expression introduced by von Neumann.[2] It is well known that the standard results of statistical mechanics immediately follow upon maximizing S (subject to the constraints imposed by the known data). Here, we have achieved a unified basis for microscopic and macroscopic statistics by adopting a unified measure of uncertainty/entropy. In retrospect, the uniqueness and universality properties of the entropy function, introduced into physics by Boltzmann with new uses advocated by Jaynes,[5] uniquely qualify it for such a role.

We now have the solution to the problem posed above: The ensemble entropy S is maximized subject to the constraints $\mathscr{P}_i^\nu=\mathrm{tr}(\hat{\pi}_i^\nu\hat{\rho})$, yielding

$$\hat{\rho}=Z^{-1}\exp\left(-\sum_{\nu,i}\lambda_i^\nu\hat{\pi}_i^\nu\right), \tag{3}$$

where the partition function Z and the Lagrange multipliers λ_i^ν are determined from

$$\mathrm{tr}\hat{\rho}=1,\quad \mathscr{P}_i^\nu=-(\partial/\partial\lambda_i^\nu)\ln Z. \tag{4}$$

While the similarity of (3) to the standard distribution functions is evident, a basic difference between the microscopic and macroscopic cases must be noted: Whereas the known data $\{\mathscr{P}_i^\nu\}$ for the former pertain to individual, microscopic systems (i.e., momentum of a particle in a beam), those in the latter correspond to bulk properties of large aggregates (e.g., volume of a gas, total magnetization of a spin system). Furthermore, the latter is almost exclusively concerned with stationary ensembles (for which $\hat{\rho}$ commutes with the Hamiltonian), whereas there is no such stipulation for microscopic ensembles and, *a fortiori,* no question of time averages or ergodicity. It goes without saying that the maximum uncertainty principle as formulated above, even though it has wider implications than those already tested within statistical mechanics, is one that is implicitly accepted and routinely applied by physicists. Indeed, in practice, any systematic deviation away from its predictions would be attributed to an unaccounted "bias" in the preparation procedure and searched for. Furthermore, insofar as it specifies a definite measure of uncertainty/entropy for any physical system, it resolves, in principle at least, the ambiguity which arises in the implementation of EAP for continuous distributions.[5,8]

An important aspect of the above formulation is the mutual harmony of its mathematical rigor and its physical sense. Consider, for example, a state prepared by the measurement of one observable $\hat{A}$. One finds that $\hat{\rho}=\sum_i[\mathscr{P}_i^A\hat{\pi}_i^A/\mathrm{tr}(\hat{\pi}_i^A)]$. Clearly, unless every $\hat{\pi}_i^A$ has finite trace, this $\hat{\rho}$ will not be normalizable. For example, if $\hat{A}$ is the momentum operator $\hat{p}$, $\mathrm{tr}\hat{\pi}_i^p=\infty$ for $\hat{\pi}_i^p$ corresponding to any nonempty bin (or interval). Therefore, preparation by momentum measurement *only,* no matter how accurate, leads to a mathematically unacceptable $\hat{\rho}$. However, such a measurement is also self-contradictory physically, since any procedure allowing for momentum measurements requires the presence, in the confines of the laboratory, of the particle being measured. The latter of course implies a constraint on the possible values of the position x, which in turn constitutes information on $\hat{x}$, contrary to the initial assumption. Once the possible values of $\hat{x}$ are thus constrained, the spectrum of $\hat{p}$ becomes discrete, $\mathrm{tr}(\hat{\pi}_i^p)$ finite, and $\hat{\rho}$ normalizable. In fact any preparation procedure implies such a constraint on the possible values of position, a fact which when overlooked will lead to an unacceptable $\hat{\rho}$. It is perhaps worth emphasizing that this is not merely a matter of mathematical purism, but rather an integral feature of the present formalism. Indeed the very starting point, the definition of measurement entropy in Eq. (1), would in general be meaningless if the imperfect resolving power of actual devices was not accounted for.

We shall now turn to examples involving momentum and position measurements, say on the particles of a beam, considered in one dimension for convenience. The first example is modeled after the celebrated thought experiments (such as that of the Heisenberg microscope) often discussed in connection with the uncertainty principle. It involves the measurement of the probabilities $\mathscr{P}^x$ and $\mathscr{P}^p$ that the particle has a position, respectively momentum, in the interval $\alpha^x=(-\frac{1}{2}\Delta x,\frac{1}{2}\Delta x)$, respectively $\alpha^p=(-\frac{1}{2}\Delta p,+\frac{1}{2}\Delta p)$, with $\hat{\pi}^{x,p}$ being the projection operators onto $\alpha^{x,p}$. Additionally, we have the above-mentioned spatial constraint on the position, say $|x|<\frac{1}{2}L$, corresponding to

$\alpha_L^x = (-\frac{1}{2}L, \frac{1}{2}L)$, $\mathscr{P}_L^x = 1$, expressing the information that the particles are certainly to be found with the laboratory (whose "size" is L).

A direct application of Eqs. (3) and (4) gives

$$\hat{\rho} = Z^{-1}\exp(-\lambda^x\hat{\pi}^x - \lambda^p\hat{\pi}^p - \lambda_L^x\hat{\pi}_L^x),$$

$$Z = 2\exp(-\lambda_+)\sum_n \cosh[(1-\mu_n^2)\lambda_-^2 + \mu_n^2\lambda_+^2]^{1/2},$$

where $2\lambda_\pm = \lambda^x \pm \lambda^p$, and where $\{\mu_n^2\}$ is the (finite) set of the eigenvalues of the operator $\hat{B} = \hat{\pi}^x\hat{\pi}^p\hat{\pi}^x$ acting on the Hilbert space $L^2(\alpha_L^x)$. The operator $\hat{B}$ is of finite rank (and of the Hilbert-Schmidt class in the limit $L \to \infty$), positive definite, and possesses a bounded spectrum $0 < \mu_n^2 \leq \mu_{\max}^2 < 1$. Moreover, one can show that for the physically interesting case of $(\Delta x/L) << 1$, and with

$$(\Delta x)(\Delta p)/2\pi = k, \quad \mathrm{tr}\hat{B} = \sum_n \mu_n^2 = k,$$

$$\mu_{\max}^2 = k[1-(\pi k/6)^2 + O(k^4)] \quad \text{for } k \to 0,$$

and

$$\mu_{\max}^2 \to 1 \quad \text{for } k \to \infty.$$

The uncertainty principle is now manifested in that the joint probability $\mathscr{P} = \mathscr{P}^x\mathscr{P}^p$ of finding the particle in the intervals Δx and Δp has an upper bound $\mathscr{P}_{\max}$ which is less than unity and which decreases with decreasing k. Indeed considering the symmetric case of $\mathscr{P}^x = \mathscr{P}^p$, which corresponds to $\lambda^x = \lambda^p$ and which for a given value of $\mathscr{P}^x + \mathscr{P}^p$ maximizes $\mathscr{P}$, we find from the solution given above that

$$\mathscr{P}^{1/2} = \frac{1}{2} - \frac{1}{2}\frac{\sum_n \mu_n \sinh(\mu_n\lambda_+)}{\sum_n \cosh(\mu_n\lambda_+)}.$$

It is immediately clear from this result that $1-\mu_{\max} \leq 2\mathscr{P}^{1/2} \leq 1+\mu_{\max}$ (recall that $0 < \mu_{\max} < 1$). The two extreme limits above correspond, not unexpectedly, to infinitely "hot" and "cold" ensembles corresponding to $\lambda_+ \to \pm\infty$. One finds that at these limits $\hat{\rho}$ reduces to pure states $|H\rangle\langle H|$ and $|C\rangle\langle C|$, respectively, where $|H/C\rangle = (\mu_{\max}\hat{\pi}^x \mp \hat{\pi}^p)|\max\rangle$, with $\hat{B}|\max\rangle = \mu_{\max}|\max\rangle$ and $\mathscr{P}_{H/C}^{1/2} = \frac{1}{2} \pm \frac{1}{2}\mu_{\max}$. Given the above restriction to the symmetric case $\mathscr{P}^x = \mathscr{P}^p$, $|H\rangle$ represents the worst, and $|C\rangle$ the best possible x-p definition, while the mixed state corresponding to $\lambda_+ = 0$ represents an intermediate case with $\mathscr{P}^{1/2} = \frac{1}{2}$. The latter, in fact, describes an ensemble for which the particle is as likely as not to be found in the interval Δx, respectively, Δp, and can be realized by an equal mixture of two beams, one filtered through α^x and the other through α^p.

As a measure of how well $|C\rangle$ fares in realizing a beam with a well-defined position and momentum, we shall compare it with a Gaussian (pure) state $|G\rangle$ whose $\hat{x}$ and $\hat{p}$ variances are matched to the bins of our measuring device so that $\delta x = \frac{1}{2}\Delta x$ and $\delta p = \frac{1}{2}\Delta p$, with $(\delta x)(\delta p) = \frac{1}{2}$. Since $k = 1/\pi$ for these values, we can use the small k expansion for $\mu_{\max}^2$ given above to find that $\mathscr{P}_C^{1/2} = 0.79$, whereas the corresponding probability for the Gaussian state is $\mathscr{P}_G^{1/2} = 0.68$. As expected, $\mathscr{P}_C > \mathscr{P}_G$ since $|C\rangle$ is optimal in this respect. This large 79% probability notwithstanding, a simple calculation shows that the conventional variance measure of uncertainty for $|C\rangle$ is essentially infinite simply because its x- and p-space wave functions have power law [e.g., $O(1/p)$], albeit very small, tails. This is another instance of the limitations of the conventional variance measure of uncertainty.[6] In summary then, the uncertainty condition for the typical two-bin measurement of position and momentum may be expressed by

$$\mathscr{P} = \mathscr{P}^x\mathscr{P}^p \leq \tfrac{1}{4}\{1 + [(\Delta x)(\Delta p)/2\pi]^{1/2}\}^2, \qquad (\Delta x)(\Delta p) \lesssim 1. \tag{5}$$

As the above analysis shows, this is an experimentally more meaningful statement than the conventional one involving variances.

Given that the conventional minimum uncertainty state is a pure state not preparable by means of an actual measuring apparatus, there arises the question of how closely it can be approximated given the "resolutions" Δx and Δp of the apparatus. It is clear that there must exist a universal bound $U_{\inf}$, a function of k on dimensional grounds, which is the infimum of $U_\rho = (\delta x)_\rho(\delta p)_\rho$ for all $\hat{\rho}$ preparable by the given device. It is immediately clear that $U_{\inf}(k) \to \frac{1}{2}$ for $k \to 0$ and $U_{\inf}(k) \to (\pi/6)k$ for $k \to \infty$, the latter representing purely classical probability distributions. While the determination of $U_{\inf}(k)$ represents a problem of immense complexity, the following result (for small k) provides useful information on the question.

Given a device with resolutions Δx and Δp,[7] the upper bound to what can be achieved is represented by the (ideal) covering of the entire range of the values of x and p by means of bins of size Δx and Δp, labeled by

$$\alpha_s^x = [(s-\tfrac{1}{2})\Delta x, (s+\tfrac{1}{2})\Delta x],$$

$$\alpha_s^p = [(s-\tfrac{1}{2})\Delta p, (s+\tfrac{1}{2})\Delta p], \quad s = 0, \pm 1, \ldots$$

The general solution corresponding to a measurement by means of this device is given by $\hat{\rho} = Z^{-1} \times \exp(-\sum_s \lambda_s^x\hat{\pi}_s^x + \lambda_s^p\hat{\pi}_s^p)$, where, as usual, Z is determined by $\mathrm{tr}\hat{\rho} = 1$ and the λ's by $(-\partial/\partial\lambda_s^{x,p})\ln Z = \mathscr{P}_s^{x,p}$, where $\mathscr{P}_s^{x,p}$ are the measured probabilities.[11] Consider now the case where these probabilities match those of a Gaussian state with $(\delta x)/(\delta p) = (\Delta x)/(\Delta p)$.[12] A straightforward calculation then gives the following expansions in k:

$$\mathscr{P}_s^x = \mathscr{P}_s^p = (2k)^{1/2}\exp(-2\pi k s^2)[1+\tfrac{1}{3}(2\pi k s)^2+\dots],$$

$$S(\hat{\rho}|D^x)+S(\hat{\rho}|D^p)-2(\delta x)_\rho(\delta p)_\rho = -\ln(2k)-\tfrac{1}{3}\pi k+\dots.$$

Now an inequality derived in Ref. 13 states that

$$S(\hat{\rho}|D^x)+S(\hat{\rho}|D^p)\geq 1-\ln(2k),$$

so that the above equation can be restated as

$$U_\rho=(\delta x)_\rho(\delta p)_\rho \gtrsim \tfrac{1}{2}+\tfrac{1}{12}(\Delta x)(\Delta p), \quad k \ll 1, \qquad (6)$$

where this lower bound is in fact approached as $(\Delta x)(\Delta p)\to 0$.[13] Physically, this result indicates that for $k \ll 1$, the uncertainties resulting from finite resolutions are additive corrections to the intrinsic quantum mechanical ones. We conclude by noting that the ensemble entropy calculated for this state is $\frac{1}{12}(\Delta x)(\Delta p)\ln[12/(\Delta x)(\Delta p)]$ for small k. This is a measure of the uncertainty and impurity of the state determined by the measured probabilities $\mathscr{P}_s^{x,p}$. This impurity is characteristic of, and will persist for, any actual measuring device.

One of us (M. H. P.) wishes to thank Professor S. Drell for hospitality and support at the Stanford Linear Accelerator Center, where most of this work was carried out. This work was supported by the U. S. Department of Energy under Contract No. DE-AC03-76SF00515.

[a]Permanent address.

[1]In this Letter, "measurement" refers to a process that *prepares* a state, and it entails the production of a sufficient number of copies of the system, a fraction of which is subjected to interaction with measuring devices, thereby serving to measure/prepare *reproducibly* the remaining copies.

[2]The first mention of the density matrix seems to be in L. Landau, Z. Phys. **45**, 430 (1927). The corresponding formalism was developed by J. von Neumann; see *Mathematical Foundations of Quantum Mechanics* (Princeton Univ. Press, Princeton, 1955). A standard reference is U. Fano, Rev. Mod. Phys. **29**, 74 (1957).

[3]This idealization, often employed for position-momentum degrees of freedom, is usually an adequate approximation.

[4]The familiar treatment of partially polarized systems, an essentially trivial case from the present point of view, is an exception to this statement.

[5]E. T. Jaynes has forcefully and effectively advocated the use of this principle in a variety of contexts; Phys. Rev. **106**, 620 (1957); Found. Phys. **3**, 477 (1973). For a proof of the consistency of this principle for "reproducible" experiments, see Y. Tikochinsky, N. Z. Tishby, and R. D. Levine, Phys. Rev. Lett. **52**, 1357 (1984).

[6]M. H. Partovi, Phys. Rev. Lett. **50**, 1883 (1983). This work was in turn inspired by D. Deutsch, Phys. Rev. Lett. **50**, 631 (1983).

[7]The more realistic case of nonuniform acceptances and/or overlapping bins may be treated by straightforward generalizations.

[8]This is essentially equivalent to identifying the appropriate set of independent events in implementing equal probabilities, the *bete noire* of EAP; see Ref. 5.

[9]If the spectrum of $\hat{W}$ is not purely discrete, there will not exist a $D_{\max}^W$, only an interminable sequence of ever finer devices; see Ref. 6. However, the sought-after operator will have a point spectrum only.

[10]This is proved in von Neumann (Ref. 2), p. 381.

[11]The problem of determining $U_{\inf}$ mentioned in the text is now seen as one of minimizing U_ρ with respect to the λ's.

[12]It is important to realize that this does *not* imply that the state being measured is a Gaussian; such a result would follow only if $\Delta x = \Delta p = 0$.

[13]I. Bialynicki-Birula, to be published.