Information theoretical analysis of the hydrogen atom

Jye-Chan Chen, Yu-Chuan Yang, Hsi-Wei Jia and Tieh-Sheng Lee

*Department of Chemistry, National Chung-Hsing University Taichung,
Taiwan, ROC*

This is an analysis of the statistical nature of the time-independent Schrödinger equation through the use of the information entropy concept. We first study the Schrödinger equation in a general way and then by actually computing entropies of various states of the hydrogen atom for a re-examination of the problem. It is found that there exists a variational procedure involving maximizing entropy for obtaining all solutions once one solution is known. Based on certain observations of the particular single system, some general conclusions can be deduced. First of all, we can safely say that the Schrödinger equation, among many other interpretations, is but the consequence of a principle of minimum potential energy expectation with certain proper constraints imposed. In addition, the ensemble concept in statistical thermodynamics is also useful in understanding microscopic quantum systems and many quantum mechanical concepts such as energy quantization and wave nodal properties can be discussed in the light of information theory and statistics in general.

1. Introduction

Probably one of the most fascinating concepts in science is the one called entropy. In 1948, besides its importance in thermodynamics, Shannon and Wiener [1,2] recognized the statistical implication of entropy to any set of positive numbers and found applications in the field of communication engineering. Since then physical scientists have re-examined such a mathematical nature and developed a new field known as information theory [3,4]. In particular, Jaynes used this idea to provide a method for constructing the whole of statistical mechanics [5]. Actually, the versatility of this concept even enables its applications to be found in every field of both physical and social sciences as well [6]. Now, a density function in quantum mechanics can be regarded as a continuous set of positive numbers. This naturally renders room for information theory to play a part. In fact, during the last decade, development along this line has been remarkable. Most importantly, Bialynicki–Birula [7,8] discovered an interesting uncertainty relation,

$$S_\rho(N) + S_\gamma(N) \geqslant 3N(1 + \ln \pi) - 2N \ln N$$
$$= N(6.43 - 2 \ln N). \tag{1}$$

Here, $S_\rho(N)$ and $S_\gamma(N)$ are the information entropies in position and momentum spaces, respectively.:

$$S_\rho(N) = -\int \rho(r) \ln \rho(r) \, dr, \tag{2}$$

$$S_\gamma(N) = -\int \gamma(P) \ln \gamma(P) \, dp, \tag{3}$$

where N is the number of electrons. Later, Gadre then computed information entropies for Thomas–Fermi atoms and other atomic systems ranging from H to Xe [9–12]. Of even more importance to the present work are the works of Sears et al. [13] and of Frieden [14,15]. The former authors noticed a close relation between the quantum-mechanical kinetic energy expectation and the Fisher's information measure. They then concluded that the Schrödinger equation formulation may be viewed as a variation principle of minimal Fisher information under other constraints. Frieden later observed that the Fisher information is also the reciprocal of the Cramer–Rao bound, a criterion of noninformativeness in the problem of estimating a probability law $p(x)$. He then showed explicitly a way of obtaining the Schrödinger equation involving a principle of maximum Cramer–Rao bound.

As pointed out by Frieden, on the other hand, one of the useful properties of the maximum entropy approach is its immediate solution. However, for attacking quantum problems, some other constraints are to be found if one insists on using such an approach. Such attempts are known. Gadre and Bendale [11] suggested to maximize the sum of entropies S_ρ and S_γ and studied the possibility of incorporating various constraints such as the $\langle p^2 \rangle$ value and Kato's cusp condition

$$(d\rho/dr)_{r=0} = -2Z\rho(0). \tag{4}$$

The maximum entropy method was also used for obtaining the Compton profile $I(p)$ by Gadre and Sears [16], and by Koga and Morita [17]. Use was also found in constructing approximate quantum wave functions [18,19], as reported by many other authors.

Here, we intend to explore more about the implications of information theory to the Schrödinger equation. Putting any solution in the form of an exponential function, we find that the quantum expectation of the exponent is proportional to the quantum entropy of the system. We show this in section 2. In section 3, we report our computational results on the H atom, the H_2^+ ion, and H_2. In section 4, we re-examine the H atom problem through the maximum entropy approach. Finally in section 5, the statistical basis for certain quantum mechanical concepts and results are discussed.

2. Maximum entropy principle and the Schrödinger equation

Beginning with the classical reduced Hamilton equation for a particle in a conserved field, Schrödinger optimized the integral [20],

$$I^* = \int\int\int \{(K^2/2m)[(\partial\psi/\partial X)^2 + (\partial\psi/\partial Y)^2 + (\partial\psi/\partial Z)^2]$$
$$+ \psi^2(v - E)\} \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z \tag{5}$$

over the whole physical space to obtain his famous equation:

$$(K^2/2m)\nabla^2\psi + (E - v)\Psi = 0. \tag{6}$$

In this formulation, we observe that both the potential expectation and the square of a vector magnitude are minimized under the constraint of total probability conservation. The vector is simply the gradient of the function ψ. For a potential-free system, a minimized I^* yields a probability density $\rho = \psi^*\psi$ everywhere homogeneous as expected. However, it is to be noted that such a purpose may likewise be achieved by maximizing Shannon's information. But then, we have to impose certain necessary constraints, forcing our results to be the same as the solutions of eq. (6) in case there is a potential. In other words, our task here is to find a set of potential-dependent functions $\{f\}$ so as to yield the set of solutions of eq. (6). A detailed comparison is then possible.

For simplicity, consider a general one-dimensional problem. Instead of minimizing I^*, we choose to maximize the integral

$$I = \int_\tau (-\psi^2 \ln \psi^2 + \alpha f(x)\psi^2 + \beta\psi^2) \, \mathrm{d}x \tag{7}$$

to obtain

$$\psi = N \exp(\alpha f(x)), \tag{8}$$

where N and α are constants. Of course, the function set $\{f\}$ is yet unknown. Direct substitution of eq. (8) into eq. (6) gives, after operator factorization,

$$(K^2/2m)(D + \alpha F/2)(D - \alpha F/2)\psi = 0, \tag{9}$$

where $D = \mathrm{d}/\mathrm{d}x$ and $F(x) = \mathrm{d}f(x)/\mathrm{d}x$. It can be shown that $\mathrm{d}F/\mathrm{d}x$ is but the commutator $DF - FD$ so that $F(x) = \int(DF - FD)\mathrm{d}x$ which may be interpreted as the limiting case of a linear combination of infinitely many commutators of D and F over the whole space x. Interestingly, it can be shown that the definite integral $\int_\tau F(x)\mathrm{d}x$ has an expectation which can be shown to be, using eq. (8), proportional to the information entropy of the solution.

3. Information entropy computation

Based on the foregoing discussions, we believe that it is meaningful to compute entropies of some simple systems. We first compute entropies for some states of the

Table 1
Information entropies S_ρ for the Schrödinger hydrogenic orbitals.

Orbital	S_ρ	
	Gadre	This work
1s	4.1447	4.1447
2s	8.1109	8.1112
$2p_z$	7.2649	7.2652
$2p_x$	–	7.2650
$2p_y$	–	7.2651
2p[a]	7.6968	–

[a] Entropy constructed from the function $(p_x + p_y + p_z)/3$.

H atom for the purpose of comparing our results with those obtained by Gadre [11] for confirming our numerical evaluation of integrals. We then proceed to compute entropies for systems such as H_2^+, H_2 and the Dirac hydrogen atom. Results are summarized in tables 1 and 2, figs. 1–3. For the 1s state of the H atom, we have $S_\rho = 4.1447$ and the kinetic energy expectation $\langle T \rangle = 0.5$ hartree. If both S_ρ and $\langle T \rangle$ are considered as functions of β, the exponent of the wave functions, the plots are shown in figs. 1(a) and 1(b), respectively. An increase in β corresponds to charging up the nucleus so that the "volume" of the electron cloud becomes smaller. Entropy is thus decreased just as in thermodynamics. On the other hand, the kinetic energy is expected to increase as the "volume" becomes smaller.

We then come to the 2s orbital. We first point out that although all computations were obtained through the use of machine-numerical integration, they can be evaluated analytically in closed forms by integration by parts. For the 2s state, again if we consider the entropy as a function of N_2, the normalization constant, r', the position of the node, and β, the exponent, we see that the entropy consists of three contributions, namely, the normalization, the node, and the exponent. Although the three may be inter-related, we can still roughly partition the entropy into three terms. Of particular interest to us is the term related to r'. A node in wave function means zero probability density at the node so that the distribution as a whole becomes less homogeneous. Roughly speaking, the introduction of a node gives rise to a negative entropy contribution of -6.2496.

Table 2
Information entropies of the Dirac H atom.

	$m_j = \pm 1/2$	$m_j = \pm 3/2$	m_1[a] $E = 0$
1s	4.1448		4.1448
2s	8.1111		8.1111
$2p_{1/2}$	7.6790		7.6970
$2p_{3/2}$	16.9128	7.5719	

[a] $m_1 = 0$ means the linear combination of states with $m_j = \pm\frac{1}{2}$.
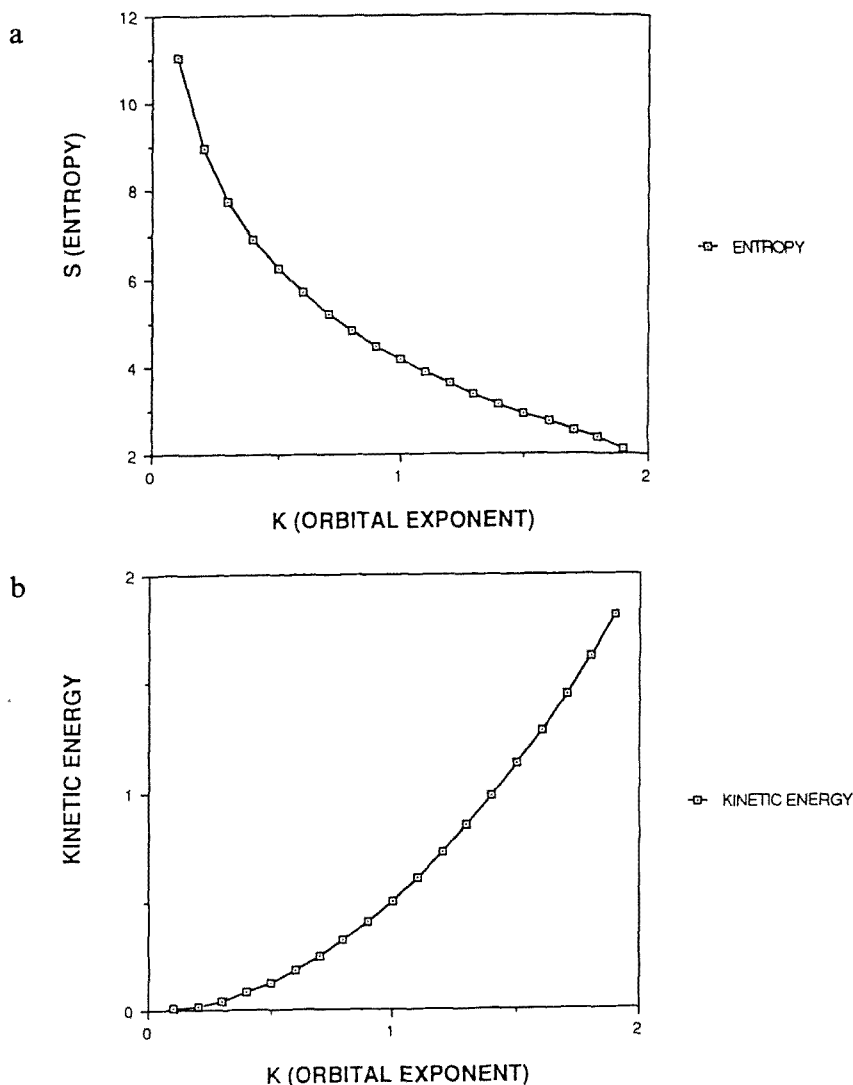
a

b

Fig. 1. (a) 1s information entropy of the H atom vs. K. (b) 1s kinetic energy of H atom vs. K.

As for the Dirac H atom, most results are found almost identical with those of the corresponding Schrödinger atom. This is definitely an indication of the relatively unimportance of the smaller components of any such Dirac wave function, as far as the information entropy is concerned. However, for the $2p_{3/2}$ states with $m_j = \pm 1/2$, entropies are roughly twice as large. No satisfactory explanation can be offered at present. Lastly, not much can be said about the H_2^+ and H_2 systems. For H_2^+, the exact solution [21] has only 11 points corresponding to 11 values of the nuclear separation R (fig. 2(a)). As for the approximate solutions (figs. 2(a) and 2(b)), only very general qualitative behavior shows up in our computation. Further study is needed.
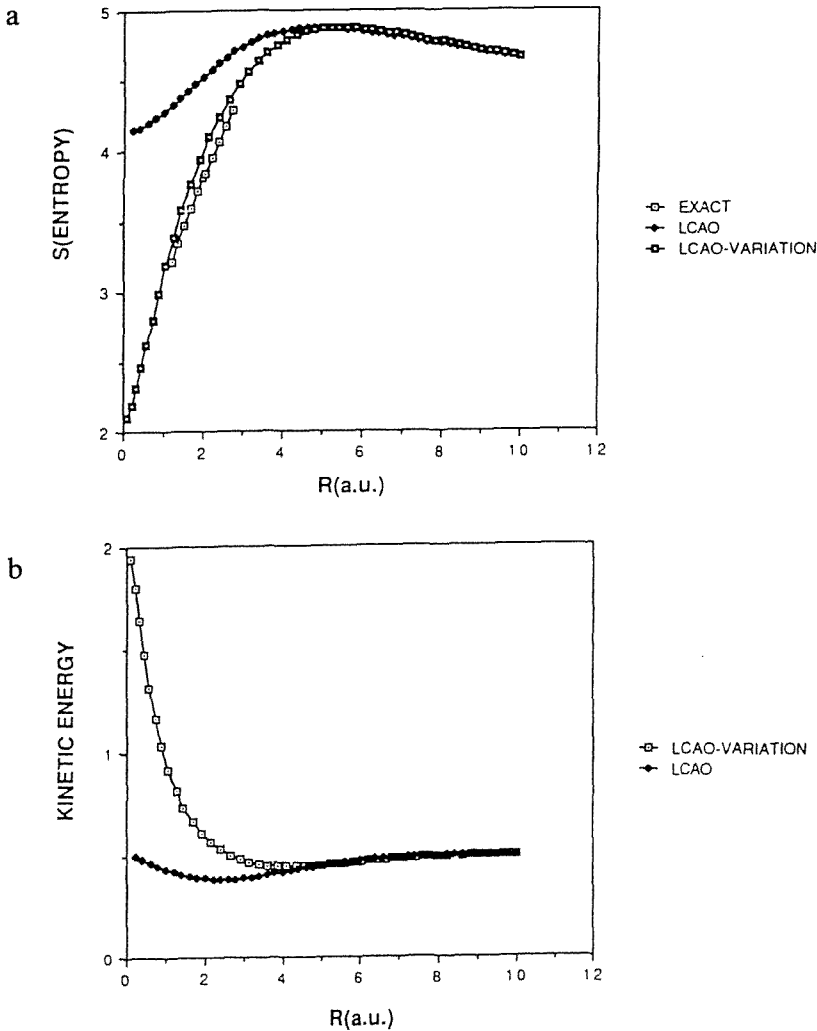
a

b

Fig. 2. (a) Informational entropy of H_2^+ in the ground state [21,24]. (b) Kinetic energy of H_2^+ in its ground state.

4. The hydrogen atom problem

We are now ready for a closer examination of the hydrogen atom problem in the light of information theory. The potential is a function only of the single variable r. Therefore, we expect the ground state probability density function ρ_0 to be a function of r. We have the information entropy of the form

$$I = -\int_\tau \rho_0(r) \ln \rho_0(r) \, d\tau . \tag{10}$$

Here, we do not intend to obtain ρ_0 through the maximum entropy method.
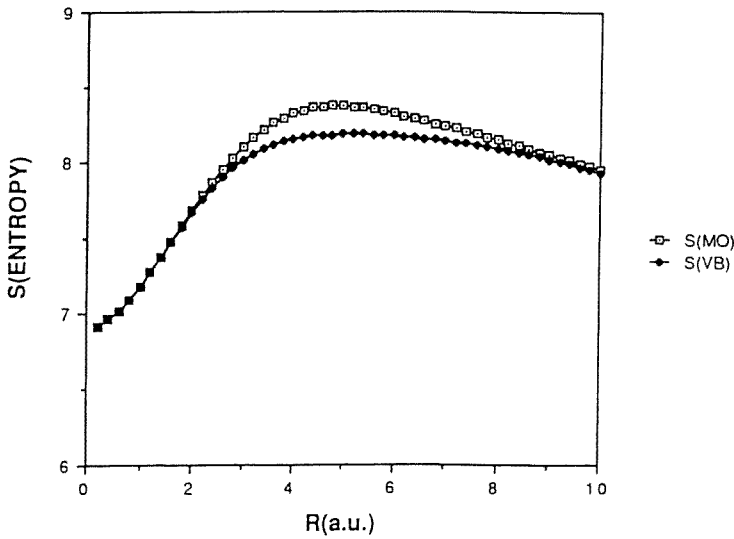
Fig. 3. Informational entropy of H_2 in its ground state [24].

Instead we only try to find the proper variational constraints that may lead us to the correct solution ρ_0 which is obtained by any other means. With a knowledge about those constraints, can we modify them to obtain other solutions? To this end, the 1s orbital is of the form

$$\Psi_{1s} = A e^{-\beta r}, \tag{11}$$

where the constants β and A can be determined by physics and normalization, respectively. If formulated through the maximum entropy approach, in order to obtain the result as in (11), we are involved with the variation

$$\int_\tau \delta\rho_0 (\ln \rho_0 + \alpha + \beta r) r^2 \, dr = 0. \tag{12}$$

As an afterthought, this could be rationalized as the requirement of maximizing the entropy under the constraints of having a finite and fixed radius $\langle r \rangle$ and of a conserved probability, or normalization condition. With the knowledge of (12), we are to obtain the probability distribution function for the 2s state. Our variational scheme then involved the integral equation

$$\int_\tau \delta\rho_0 (\ln \rho_1 + \alpha' + \beta r + \eta h(r)) r^2 \, dr = 0. \tag{13}$$

This is because we are dealing with the same potential so that everything said before remains valid here. However, we are to include an additional term, $\eta h(r)$, for taking into account two criteria, namely, the peak number difference and the orthogonality requirement which will be discussed in more detail in a separate section. From eq. (13), our variational scheme gives

$$\rho_1(r) = G(r) \exp(-\beta' r), \tag{14}$$

where $G(r) = \exp[-\eta h(r)]$ is a function to be determined. For the 2s state, one more peak along the r axis is required as we can see in a later section. Thus, we differentiate eq. (14),

$$d\rho_1(r)/dr = [-\beta' G(r) + dG(r)/dr] \exp(-\beta' r). \tag{15}$$

The function ρ_1 must possess exactly two relative extrema, one corresponding to the point of zero probability density, while the other, to the peak. It will do if $G(r)$ is a quadratic function. Through a little algebraic manipulation, we arrive at

$$\rho_1(r) = N(r - a)^2 \exp(-\beta' r). \tag{16}$$

To determine the constant a, we propose to minimize the potential expectation again for reasons to be explained later. We find that the optimal choice of a is $a = 2/\beta'$. To fix β', we use the orthogonality relation with the 1s state and find $\beta' = \beta/2$. The determination of the normalization constant causes no problem and the derivation for the 2s state distribution function is completed. In fact, this procedure can be generalized to any ns state. We begin right away with the expression

$$\Psi_{ns} = N_n(a_1 + a_2 r + \ldots + a_{n-1} r^{n-2} + r^{n-1}) \exp(-\xi r). \tag{17}$$

The minimization of $\langle v \rangle$ and the orthogonalization relations lead to the matrix equation for the set of coefficients a_i,

$$\begin{pmatrix} 1! & 2! & \ldots & (n-1)! & \xi^{n-2} \\ 2! & 3! & \ldots & n! & \xi^{n-2} \\ . & . & \ldots & & \\ . & . & \ldots & & \\ . & . & \ldots & & \\ (n-1)! & n! & \ldots & (2n-2)! & \xi^{n-2} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ . \\ . \\ . \\ a_{n-1} \end{pmatrix} - \xi^{n-1} \begin{pmatrix} n! \\ (n+1)! \\ . \\ . \\ . \\ (2n-1)! \end{pmatrix} = 0, \tag{18}$$

in which ξ is β/n. Thus, instead of using the recursion formula, using eq. (18) we can obtain the whole set of coefficients all at once. In fact, slight modifications allow us to determine also orbitals of other types such as np and nd, etc.

Moreover, it is found that almost exactly the same procedure can be applied also to the solutions of the Dirac H atom. Auvil and Brown [22] once showed that the Dirac equation

$$(\gamma_\mu \pi^\mu - m)\Psi = 0 \tag{19}$$

can be put in the form

$$(1/\rho^2)(d/d\rho)(\rho^2 [dV^\pm/d\rho]) + [(\lambda/\rho) - (1/4) - (S_\pm(S_\pm + 1)/\rho^2)V^\pm = 0 \tag{20}$$

through some operator algebra manipulation. Here $V^\pm$ is related to Ψ in a definite

way. Now, eq. (20) is exactly identical in form with the nonrelativistic Schrödinger equation. We can therefore likewise obtain all solutions by the same maximum entropy variation approach once one solution is known. We omit the details here. Again, we can obtain all coefficients of any solution without the use of a recursion formula.

5. Discussion and assumption justifications

First of all, we like to point out that when applying the method of Lagrange multipliers to a physical problem, the involved mathematics may allow distinct interpretations corresponding to the same formulation. To see this point, let us digress from the problem of the most probable distribution of N identical particles, a familiar problem in statistical thermodynamics. The usual formulation involves the maximization of entropy under the constraints of fixed number of particles N and fixed total energy E. However, other interpretations which lead to the same mathematics exactly are possible. Actually, we can even follow those interpretations to prepare the system experimentally in three different ways, at least in principle. Namely, we can fix N and E, allowing S to vary; we can likewise fix S and E, allowing N to vary; and lastly, we can also fix S and N, allowing E to vary. Turning back to the hydrogen atom problem, in obtaining solutions other than the 1s state, we have used the criterion of minimizing the potential expectation in addition to maximizing the entropy. The use of such a criterion is thus justified in view of the discussion just presented. Actually, the simultaneous extremalization of both entropy and the potential expectation can be visualized through our computation. The fact that the 2s state has a larger entropy than that of the 1s state is an indication of more importance of potential energy expectation over that of entropy. While in obtaining the distribution of the 2s state, we used the same constraints for the 1s state plus an additional term for satisfying the orthogonality and the peak number criteria. As a result, however, both the potential energy expectation and the entropy are increased. But the latter is a quantity we maximize while the former is one we minimize, implying the former is optimized at a sacrifice of the latter which should have been lower in value because of the introduction of the additional term which is a further constraint. Recall that by the virial theorem, $\langle v \rangle = 2E$ while $\langle T \rangle = -E$ for the hydrogen atom. Therefore the potential plays a more important role as compared with entropy.

In formulating the variational procedure, we have made use of two postulates, namely, the peak number criterion and the orthogonality relationship. By a peak here, we mean a relative maximum in the probability distribution. To be more precise, two relative maxima are to be called a single peak if they are not separated by at least one minimum of the value zero. With this definition, we state our peak number criterion as the following. Two distributions of a quantum system in bound state differ in energy with certainty only if they differ in peak number. Note that

all experimental results are statistical in nature, for even if we have a true micro-
scopic system at our disposal, we could never be certain in specifying the quantum
state of the macroscopic instrument. Uncertainty is thus inevitable. For this rea-
son, in interpreting a set of experimental data, one is forced to watch only the most
remarkable differences and ignore all the minor ones. In other words, in studying
a distribution of a quantum system, all differences in peak shape or in location are
ignored since they can be specified only by using continuous parameters. It means
that all such minor variations are by postulate either ignored, or treated as mix-
tures, i.e., not as eigenstates of the system's Hamiltonian, but as those of the origi-
nal somewhat perturbed in a completely random way. On a statistical basis, all
such random perturbations add up to yield a zero expectation experimentally,
allowing only the peak number change to be singled out for observation. Note that
for a low lying quantum state, the peak number is but a simple integer. Thus, the
argument just presented provides an experimental basis for the energy quantiza-
tion concept. However, the peak number criterion alone is insufficient for establish-
ing energy quantization. Under each statistical hypothesis, one particular
probability distribution which has the representative energy of the whole ensemble
is our only concern. Now, suppose that we have another distribution different in
peak number. Does it have an energy value also the average of another ensemble?
Here, we need the concept of orthogonality. For the statistical implication of this
concept, we consider a discrete set of positive numbers,

$$\{p_i\} = (p_1, p_2, \ldots, p_n)$$ (21)

satisfying the condition

$$\sum_i p_i = 1 .$$ (22)

Statistically more interesting is perhaps the set of square-root numbers, $\{p_i^{1/2}\}$
$= (p_1^{1/2}, p_2^{1/2}, \ldots, p_n^{1/2})$. In addition to the property of order-preserving, $| p_i^{1/2} |$
$\geqslant | p_k^{1/2} |$ whenever $p_i \geqslant p_k$, more interesting is the property that $| p_i^{1/2} | \geqslant p_i$ for all
i, and the difference is the largest for $| p_i^{1/2} | = 0.5$ and the smallest for $p_i = 0$ and
$p_k = 1$. A nice symmetry of this difference set of numbers i to be noted. The use of
the set of square-root values is advantageous in studying quantum effects, for it
magnifies the discrepancy between a boundary and a regular point. Consider now
two square-root functions Ψ_i and Ψ_k. The integral

$$\int (\Psi_i - \Psi_k)^2 \, d\tau = 2 - 2S_{ik}$$ (23)

is maximal if S_{ik} vanishes. Thus, orthogonality guarantees a maximal differences
between two square-root sets corresponding to two different distribution func-
tions. Now, of course, in order to have a vanishing S_{ik}, a sign convention for the Ψ's
is to be introduced. It is done in the following way. A function $\Psi_i(x)$ remains of
one sign as long as it does not cross the point zero. It takes a different sign as we

cross the zero point while changing x continuously. Note that in this way, all functions so constructed do satisfy the Schrödinger equation. In our case, Ψ_{2s} has a zero contamination of Ψ_{1s} due to orthogonality. In addition, the minimization of the potential energy expectation ensures a minimum total energy under necessary constraints. In fact, such arguments can also be extended to handle other quantum systems and further study along this line is under way in this laboratory.

Finally, we like to point out that the concept of information entropy is very versatile and can be used in many ways. For example, Gadre and Bendale [11] suggested to maximize the sum of S_γ and S_ρ for obtaining a ground state solution. They also investigated the upper and lower bounds of atomic information entropies through a study of the relationships between the kinetic energy and entropy [22]. For obtaining approximate ground state solutions of quantum systems, Arrachea et al. [18,19] considered the expansion of the wave function in terms of a suitable basis set $\{|j\rangle, j = 1, \ldots, d\}$. This allows them to define a discrete entropy. The underlying reasoning is that the change for experimentally observing a particular eigenvalue j is assumed a priori the same as for any other value j' in the absence of a potential. Then, one must find different constraints for basis sets if the characteristic potential is given. This is actually an advantage of using entropy concept for a mathematical analysis of quantum systems.

Acknowledgements

References

[1] C.E. Shannon, "A mathematical theory of communication", Bell Syst. Tech. J. 27 (1948) 379–423, 623–656.
[2] N. Wiener, *Cybernetics* (Wiley, New York, 1948).
[3] N. Wiener, "What is information theory?" IRE Trans. Inform. Theory IT-2 (1956) 48.
[4] R. Ash, *Information Theory* (Interscience, New York, 1967).
[5] E.T. Jaynes, Phys. Rev. 106 (1957) 620.
[6] S. Kullback, *Statistics and Information Theory* (Wiley, New York, 1959).
[7] I. Bialynicki-Birula and J. Mycielski, Commum. Math. Phys. 44 (1975) 129.
[8] W. Beckner, Ann. Math. 102 (1975) 159.
[9] S.R. Gadre, Phys. Rev. A30 (1984) 620.
[10] S.R. Gadre, S.J. Chakravorty and R.D. Bendale, Phys. Rev. A32 (1985) 2602.
[11] S.R. Gadre, R.D. Bendale and S.P. Gejji, Chem. Phys. Lett. 117 (1985) 138;
S.R. Gadre and R.D. Bendale, Curr. Sci. 54 (1985) 9970.

[12] G. Maroulis, M. Sana and G. Leroy, Int. J. Quant. Chem. 19 (1981) 43.

[13] S.B. Sears R. Parr and U. Dinur, Israel J. Chem. 19 (1980) 165–173.

[14] B.R. Frieden, J. Mod. Opt. 35 (1988) 1297–1316.

[15] B.R. Frieden, Am. J. Phys. 57 (1989) 1004.

[16] S.R. Gadre and S.B. Sears, J. Chem. Phys. (a) 71 (1979) 4321; 75 (1981) 4626.

[17] T. Koga and M. Morita, J. Chem. Phys. 79 (1983) 1933.

[18] L. Arrachea, N. Canosa, A. Plastino, M. Portesi and R. Rossignoli, *Condensed Matter Theories*, Vol. 7, eds. A.N. Proto and J.L. Aliaga, (Plenum Press, New York, 1992) p. 63.

[19] N. Canosa, R. Rossignoli and A. Plastino, *Condensed Matter Theories*, Vol. 7, eds. A.N. Proto and J.L. Aliaga (Plenum Press, New York, 1992) p. 69.

[20] E. Schrödinger, *Collected Papers on Wave Mechanics* (Blackie, London, 1928; Chelsea, New York, 1982).

[21] V.A. Johnson, Phys. Rev. 60 (1941) 373.

[22] P.R. Auvil and L.M. Brown, Am. J. Phys. 46 (1978) 679.

[23] S.R. Gadre and R.D. Bendale, Phys. Rev. A36 (1987) 1932.

[24] I.N. Levine, *Quantum Chemistry*, 3rd Ed. (Allyn and Bacon, Boston, 1983) pp. 326–329, 354–365.