Collision experiments with partial resolution of final states: Maximum entropy procedure and surprisal analysis

Y. Alhassid and R. D. Levine

Department of Theoretical Physics and Department of Physical Chemistry and Institute for Advanced Studies, The Hebrew University, Jerusalem, Israel

(Received 30 April 1979)

The application of the maximal entropy approach as a dynamical procedure to collisions when there is incomplete resolution of the final states is discussed. The major objective is to provide a collision theoretic foundation for the phenomenological procedure of surprisal analysis which is applied to heavy-ion induced reactions in a companion paper. A reduced description where during the entire time evolution the degree of detail is commensurate with that available for the analysis of the final state is introduced. An exact equation of motion, in the form of a continuity equation, is derived for the reduced description. The most notable feature of the equation is the inherent presence of a dissipative term. Because of the finite duration of the collision, the final reduced distribution need not, however, be fully relaxed. It is shown it is possible to characterize the deviance from the statistical limit in terms of time-dependent constants of the motion. The post-collision measure of this deviance is the surprisal. The significance of the (time-independent) constraints, identified by surprisal analysis of experimental results, is discussed and the use of sum rules as a practical route to the identification of such constraints is noted. The formulation of the approach as a sophisticated statistical theory capable of representing direct reactions is discussed in the appendix. The variational character of the procedure of maximal entropy is also noted therein as a route to approximate descriptions.

[NUCLEAR REACTIONS Statistical theories of heavy-ion reactions, surprisal analysis of the deviation of the observed distribution from the statistical limit, dynamical derivation of functional form of the surprisal, identification of constraints using sum rules, procedure of maximal entropy, equation of motion for reduced description.]

I. INTRODUCTION

Collision experiments using composite projectiles often fail to fully resolve the possible final states. Such is particularly the case at higher energies where many internal states are accessible. It is then appropriate to adopt a reduced level of description where a coarser analysis suffices. It is not necessary to account for the cross sections into specific final quantum states but only for the averaged behavior. Despite this loss of detail a purely statistical theory is not necessarily sufficient as a reduced description. Yet a conventional [e.g., distorted wave Born approximation (DWBA)] dynamical approach is both prohibitive (because of the enormous number of final states that need to be summed over) and wasteful (because of the considerable detail which gets averaged over). The purpose of this article is to argue that under such circumstances it makes sense to provide a reduced characterization by the procedure of maximal entropy.[1-7]

The method can be applied in two modes. One is analysis, where the procedure retains the simplicity of a statistical theory, yet is applicable to direct processes where there is often considerable specificity in the distribution of final states. The essential idea here is that in a purely statistical theory all accessible final quantum states are equally probable. Such a distribution can be characterized as one of maximal entropy (subject only to the conservation of the additive constants of the motion, e.g., energy, as these define the range of accessible states). In a direct reaction there are additional constraints that prevent the final phase space from being uniformly populated. The analysis assumes that the correct description of the actual observed distribution in phase space is one of maximal entropy subject to these additional constraints. It seeks to identify such constraints by an analysis of the data. The method has been extensively employed for molecular collisions,[6,8-10] where it is often the case that very few constraints (one or two) of obvious physical significance suffice to account for highly nonstatistical reduced distributions of direct reactions. Moreover, similar reactions are found to have similar constraints.[7] The analysis thus seeks to codify, compact, and correlate the data. Applications to specific heavy-ion nuclear reactions are reported in a companion paper[11] and elsewhere.[12,13]

The second mode of application and a central topic of this paper is that of synthesis. Here one attempts to address the two questions raised by the

success of the analysis: (a) Why does the procedure of maximal entropy subject to constraints work?, (b) As the constraints identified by the analysis are not time-independent constants of the motion (of the kind made familiar by statistical theories or, in general, by equilibrium statistical mechanics[14]), what then is the physical significance of these constraints and why is it that few constraints often suffice to account for the major features of the data? Ultimately, the purpose of the synthesis is to provide a predictive route, where one computes directly the reduced distribution of interest.

This paper is roughly divided into two parts. Sections II and III discuss the concept of a reduced distribution and the procedure of maximal entropy with special reference to reduced descriptions. Section IV is the transition between the two parts. It motivates the selection of the initial state before the collision as one of maximal entropy subject to constraints. It is then noted[7,15] that such an initial state evolves into a state that is of maximal entropy throughout the collision. In other words, it is not necessary to invoke the procedure at every time t of interest. Having invoked it once, it remains valid throughout the collision. The constraints during the collision are identified[7,15] as time-dependent constants of the motion. These are functions which depend explicitly on time, but owing to their dependence on additional variables have time-independent expectation values. The one concession to simplicity is the use of classical mechanics. The intended application is, however, to collisions where the density of internal states is high and the kinetic energy is plentiful so that the restriction is not severe. Previously we have discussed[7,15] the synthesis for the full description of the collision using both a quantal and a classical description of the internal states and noted that the essential difference is not in the identification of the constraints. Rather it was that the quantal constraints, being operators, may fail to commute.

The application to collisions where a reduced description is sufficient is discussed in the second part. The two central technical results (Sec. V) are (a) the derivation of an exact yet innocent looking equation of motion for the reduced description in the form of a continuity equation, and (b) the proof that even for the reduced description the deviance from the statistical limit remains a time-dependent constant of the motion. It is this result that points out the special role of the surprisal. Section VI considers both phenomenological and theoretical routes to the identification of the constraints. The most important practical tool is the use of sum rules, since these are often available from diverse sources, including models.[16,17]

There are two appendices. We specifically recommend an examination of Appendix A. It is argued there that the compactness of the maximum entropy description is achieved by its failing to compute such details which get averaged out in the reduction. From the very start, the formalism centers attention only on those features which are relevant at the reduced level of description. Borrowing the terminology of the statistical theories,[18] it is convenient to introduce the idea of a fluctuation term as follows. Two densities in phase space which upon averaging yield the same reduced density differ only in their fluctuating parts. It is then shown that their common average part is given by the maximum entropy formalism. Finally, it is shown that the solution of the equation of motion can be cast in a maximal entropy form. One can also show[15] that the variational procedure of maximal entropy converges to the exact solution of the equation of motion. This complementary point of view is suggested as a practical route to approximations motivated by physical considerations. Appendix B demonstrates how the irrelevant details (which get averaged out upon reduction) can be removed by the use of a projection operator.

II. THE REDUCED DESCRIPTION

Consider for simplicity a classical system. A complete description of a single isolated classical collision is given by the trajectory in phase space, where a point along the trajectory is specified by the coordinates and momenta of the colliding partners. A single trajectory evolves into a particular final classical state. In the sense of the correspondence principle, a quantal system can be represented as an entire set of trajectories and this is true even when the system is initially in a pure quantum state. At any time, the collision is specified by a density function $\rho(p,q)$ in phase space. The time evolution of ρ is given by Liouville's theorem[19] $d\rho/dt = 0$ or

$$\partial\rho/\partial t + \sum_i [(\partial\rho/\partial q_i)\dot{q}_i + (\partial\rho/\partial p_i)\dot{p}_i] = 0 . \qquad (2.1)$$

Here $\partial\rho/\partial t$ is evaluated at a given point in phase space and the dots denote the time derivative. The density in phase space is thus a constant of the motion despite the fact that it is explicitly time dependent. It follows from Liouville's theorem that any function of ρ, and in particular the entropy $S(\rho)$, where

$$S(\rho) = -\mathrm{tr}(\rho \ln \rho) , \qquad (2.2)$$

are constants of the motion. The notation trace, as in (2.2), denotes integration over the accessible regions in phase space. Thus, for a system with a sharp value of the energy, the integration is restricted to the energy shell. In general, we shall use the term "energy shell" for the range of states that are allowed by conservation laws. We also stipulate that the area element $dpdq$ in phase space is measured in units of h, so that the density ρ is dimensionless.

The reduced description which is of direct experimental interest centers attention on the values of one, say A, or a few, $A_1, A_2, \ldots, A_n$, dynamical variables. These do not suffice to identify a point in phase space as there is an entire shell of points $A(p, q) = a$ which correspond to A having the given value a. The measured distribution is $P(a)$ [or $P(\vec{a}) = P(a_1, \ldots, a_n)$] such that $P(a)da$ is the frequency for A to have a value in the range a to $a + da$.

The reduced distribution at the time t can be computed[20] from $\rho(p, q; t)$ by summing over all points which are on the shell $A(p, q) = a$,

$$p(a; t) = \mathrm{tr}[\rho(p, q; t)\delta(A(p, q) - a)]. \tag{2.3}$$

If there is more than one variable A, there will be a product of δ functions in (2.3), one for each A_r, which can be written in the compact notation

$$p(\vec{a}; t) = \mathrm{tr}[\rho(p, q; t)\delta(\vec{A}(p, q) - \vec{a})]. \tag{2.4}$$

$p(a; t)$ provides a complete reduced description. Average values can be computed directly from $p(a; t)$ as follows:

$$\begin{aligned} \langle A \rangle &= \mathrm{tr}[A(p, q)\rho(p, q; t)], \\ &= \int da\, a\, \mathrm{tr}[\rho(p, q; t)\delta(A(p, q) - a)], \\ &= \int da\, a p(a; t), \end{aligned} \tag{2.5}$$

and similarly for functions of A.

The purpose of this paper is to characterize directly the reduced distribution $p(a; t)$. The motivation is that such a characterization may in some sense be simpler for $p(a; t)$ than for $\rho(p, q; t)$ because the latter carries considerably more detail which is not really relevant as it only gets integrated over during the reduction. A more technical discussion of this point of view is given in Appendix A. In this connection, it is useful to point out a complementary interpretation of (2.3). As it now stands, the final reduced distribution is determined by coarse graining the final detailed density in phase space over all the irrelevant details. One can, however, transform (2.3) to a form where the reduction is over all the details of the initial state which are irrelevant. This transformation corresponds to the change from the Schrödinger to the Heisenberg pictures in quantum mechanics,

$$p(a; t) = \mathrm{tr}[\rho(p, q)\delta(A(p, q; t) - a)]. \tag{2.6}$$

Here $\rho(p, q)$ is the initial distribution in phase space and $A(p, q; t)$ is the solution of the equation of motion for phase functions, which reflects the evolution of p and q with time

$$\partial A/\partial t = \sum_i [(\partial A/\partial q_i)\dot{q}_i + (\partial A/\partial p_i)\dot{p}_i]. \tag{2.7}$$

III. PROCEDURE OF MAXIMAL ENTROPY

The concept of entropy can be introduced in an axiomatic fashion[21] as a measure of "uncertainty" or "missing information," when what is known is only a probability distribution. Adopting this point of view, one is no longer restricted to systems with very many degrees of freedom, nor to systems at equilibrium. In a collision of composite projectiles there is typically more than just one accessible final state. When many independent binary collisions are considered, there will be observed a distribution of final states and one can consider the entropy of that distribution. Indeed, one can even consider a particular state and talk of its "surprisal," which is the amount of information provided when the probability of that final state is determined.[22] The only question then is whether it is useful to do so. That is, whether one can establish links between these abstract concepts and quantities which are not defined within the axiomatic structure but which are relevant to the dynamics of the collision.

As an example, consider the initial state before the collision. On practical grounds one requires that it does not change in the absence of collisions, so that any modifications in the distribution of states reflect the interaction during the collision. One can thus characterize the initial state by the values of those quantities which are conserved in the absence of collisions. If only a finite number of such quantities are invoked, there may be other states which lead to identical values for the conserved quantities. One can, however, show (for both classical[23] and quantal[15] systems) that among all those states, the stationary state is the one of maximal entropy. The purpose of this paper is to discuss and interpret a similar characterization of the final, post-collision state. As an inter-

mediate stage we shall demonstrate the same type of characterization throughout the collision.

In phase space, on the energy shell, the entropy can be interpreted as the effective area covered by the distribution. The higher the entropy, the less localized (or more "chaotic") is the distribution. It is shown below to be maximal for the uniform distribution (on the energy shell). Any other normalized distribution has a lower entropy. The uniform distribution $\sigma^0(p,q)=1/\mathrm{tr}(1)$ is not, however, a very interesting initial state for collision problems. The reason is that for any Hamiltonian such an initial state evolves into a final, post-collision state which is also uniform.[24] Scattering experiments using such an initial state do not serve to probe the dynamics. A more useful initial state is one where the entropy is maximized not over all distributions but only over some more limited class. We prove below (Sec. IV B) that such an initial state will also evolve into a final state of maximal entropy, but one which does provide insight into the dynamics. Before that, we need to consider the class of distributions over which entropy is to be maximized. We have previously[7,15] discussed the general case, but here we make a specific requirement: All the information used to specify the class of initial states should be in the A space. The degree of control exercised in selecting the initial state should be comparable to that available for the analysis of the final state. If that is not the case, one should use the more general formalism.[7]

A. Constraints

The purpose of the maximal entropy procedure is to select a unique particular distribution, say σ, from an entire class of distributions $\{\rho\}$, where all members are consistent with a given reduced data. An example is to choose σ from all the densities which are consistent with a given reduced initial density $p(a)$. The unique solution is (cf. Sec. III C)

$$\sigma(p,q)=p(a)/\Omega(a)\,, \tag{3.1}$$

so that all points where $a=A(p,q)$ have the same value of σ and $\Omega(a)$,

$$\Omega(a)=\mathrm{tr}[\delta(A(p,q)-a)]\,, \tag{3.2}$$

is the density of states on the $a=A(p,q)$ shell. For every value of a (i.e., on the A shell), σ, as given by (3.1), is a uniform distribution, but its value differs for different shells.

The density $p(a)$ is the most detailed possible reduced initial input. Less detailed reduced inputs are those where the expectation values $\langle C_r\rangle$ of M functions, $C_r(a)$ are specified

$$\langle C_r\rangle=\int da\, C_r(a)p(a)\,,\quad r=0,1,\ldots,M\,. \tag{3.3}$$

By taking $C_0(a)=1$, the condition that the density is normalized is included in (3.3). These $M+1$ constraining conditions can be rewritten, using (2.3), as conditions on the density in phase space

$$\langle C_r\rangle=\mathrm{tr}[\rho(p,q)C_r(A)]\,,\quad r=0,1,\ldots,M \tag{3.4}$$

where $A\equiv A(p,q)$.

If all that is specified is that the distribution is normalized, then $M=0$. As M increases, the range of densities $p(a)$ which satisfy (3.3) is more and more limited. In the limit $M\to\infty$ [say, by using for $C_r(a)$ a complete basis set] these conditions specify $p(a)$ uniquely.

B. Distribution of maximal entropy

The constraints (3.3) provide, therefore, the means for stating the available data on the A space. It is from the class of functions consistent with the constraints that one selects σ as the one of maximal entropy. The result is[1-7] (cf. Sec. III C)

$$\sigma(p,q)=\exp\Big[-\sum_{r=0}^{M}\lambda_r C_r(A)\Big]\,, \tag{3.5}$$

where, as usual, $A\equiv A(p,q)$. The $M+1$ (Lagrange) parameters in (3.5) are to be determined by the $M+1$ implicit equations (3.4). An efficient computer program has been described[25] and is available upon request.

The reduced distribution $p(a)$ corresponding to $\sigma(p,q)$ is

$$\begin{aligned}p(a)&=\mathrm{tr}[\sigma(p,q)\delta(A(p,q)-a)]\\&=\Omega(a)\exp\Big[-\sum_{r=0}^{M}\lambda_r C_r(a)\Big]\,.\end{aligned} \tag{3.6}$$

The reduced distribution corresponding to $\sigma^0(p,q)$, the uniform distribution (which is subject only to the normalization constraint), is $p^0(a)$,

$$\begin{aligned}p^0(a)&=\Omega(a)\exp(-\lambda_0)\\&=\Omega(a)\Big/\int da\,\Omega(a)\,.\end{aligned} \tag{3.7}$$

The constraints are reflected in the deviance of $p(a)$ from $p^0(a)$.

C. Entropy and entropy deficiency

Given two normalized distributions, it is convenient to consider the following measure:

$$DS[\rho|\sigma]=\mathrm{tr}(\rho\ln\rho-\rho\ln\sigma)\,. \tag{3.8}$$

It is non-negative and equals zero if and only if $\rho=\sigma$. If ρ and σ are consistent with the same set of constraints [e.g., both satisfy the $M+1$ conditions (3.4)] and σ is of maximal entropy, i.e.,

of the form (3.5),

$$\mathrm{tr}(\rho \ln\sigma) = \mathrm{tr}(\sigma \ln\sigma) \tag{3.9}$$

and then, using (3.9) in (3.8),

$$DS[\rho|\sigma] = S[\sigma] - S[\rho] \geq 0 . \tag{3.10}$$

For any ρ which satisfies the constraints, but is not equal to σ, DS is positive and hence σ is the unique state of maximal entropy.

Since all densities of interest are normalized, $DS[\rho|\sigma^0]$ is always the difference between the actual and unconstrained maximal values of the entropy

$$DS[\rho|\sigma^0] = S[\sigma^0] - S[\rho] . \tag{3.11}$$

It is therefore referred to[6,8-10] as the entropy deficiency.

When the constraints are only on the A space, (σ/σ^0) will be only a function of a and hence the entropy deficiency can be expressed in terms of $p(a)$ itself:

$$\begin{aligned} DS[\sigma|\sigma^0] &= \int da\, p(a) \ln[p(a)/p^0(a)] \\ &= S[p^0] - S[p] \\ &= DS[p|p^0] , \end{aligned} \tag{3.12}$$

where

$$S[p] = -\int da\, p(a) \ln[p(a)/\Omega(a)] , \tag{3.13}$$

$$\begin{aligned} S[p^0] &= -\int da\, p^0(a) \ln[p^0(a)/\Omega(a)] \\ &= \ln\left[\int da\, \Omega(a)\right] . \end{aligned} \tag{3.14}$$

The result (3.13) identifies the correct expression for the entropy in the A space. With this expression one can derive (3.6) directly as the distribution of maximal entropy subject to the constraints (3.3), or identify $p^0(a) \propto \Omega(a)$ as the distribution of maximal entropy subject only to normalization.

IV. COLLISION PROCESSES

In the absence of constraints, the distribution in phase space after the collision will be uniform (on the energy shell). The corresponding reduced distribution is $p^0(a)$. It need not be uniform and may depend on a, but such dependence is that implied by the density of states [cf. (3.2)] $\Omega(a)$. An observed (or computed) reduced final state distribution $q(a)$ need not agree with $p^0(a)$. The entropy deficiency $DS[q|p^0]$ provides an integral measure of the deviance. The local measure is provided by the surprisal[26] $I(a)$, where

$$I(a) = -\ln[q(a)/p^0(a)] . \tag{4.1}$$

In information-theoretic terms,[22] $-I(a)$ is the information received from a message which stated that the event a has the probability $q(a)$, if before the message the probability of a was taken to be $p^0(a)$. As expected, the total information in such a message, given by $DS[q|p^0]$ [cf. (3.12)], is nonnegative. In dynamical terms, the surprisal will be characterized by the constraints. In other words, our purpose is to identify a set of functions $C_r(a)$ such that $q(a)$ is given as a distribution of maximal entropy subject to these functions as constraints

$$-\ln[q(a)/p^0(a)] = \sum_{r=0} \lambda_r C_r(a) . \tag{4.2}$$

This section examines briefly the surprisal of $\rho(p, q; t)$ both during and after the collision. The applications to the reduced distributions are made in Secs. V and VI.

A. Time-dependent constants of motion

To show that an initial state of a maximal entropy retains this characterization throughout the collision, it is convenient to employ the concept of a time-dependent constant of the motion.[27,28] A phase function which depends explicitly on time, but whose average value (computed over the density at time t) is independent of time,

$$d\{\mathrm{tr}[\rho(p, q; t) I_r(p, q; t)]\}/dt = 0 , \tag{4.3}$$

is a time-dependent constant of the motion. It follows from the Liouville equation (2.1) that (4.3) can be rewritten as

$$\begin{aligned} &\mathrm{tr}\{\rho(p, q; t)[(\partial I_r/\partial q)\dot{q} + (\partial I_r/\partial p)\dot{p} + \partial I_r/\partial t)]\} \\ &\quad = \mathrm{tr}[\rho(p, q; t)(dI_r/dt)] = 0 . \end{aligned} \tag{4.4}$$

Since the definition (4.3) should hold for any density function, a time-dependent constant of the motion is formally defined by

$$dI_r(p, q; t)/dt = 0 . \tag{4.5}$$

$I_r(p, q; t)$ satisfies the same equation as the density $\rho(p, q; t)$ itself. (This, incidentally, will no longer be true in the reduced description.) I_r is a conserved quantity because its p and q dependence is such that it balances its explicit time dependence, as is evident in (4.4).

Given the initial p and q dependence of I_r, e.g., $C_r(p, q)$, then

$$\mathrm{tr}[\sigma(p, q) C_r(p, q)] = \mathrm{tr}[\sigma(p, q; t) I_r(p, q; t)] , \tag{4.6}$$

and the equation of motion (4.5) determines $I_r(p, q; t)$ uniquely.

That constants of motion with the stipulated properties do exist follows from the following sim-

ple consideration[27]: The equations of motion define uniquely the present values p and q in terms of their initial values $p^0=p(t=0)$ and q^0 as follows:

$$p=P(p^0,q^0;t)\,,\qquad q=Q(p^0,q^0;t)\,. \tag{4.7}$$

Since the time evolution is reversible, one can take the present values as initial values and recover p^0 and q^0. Explicitly

$$p^0=P(p,q;-t)\,,\qquad q^0=Q(p,q;-t)\,. \tag{4.8}$$

For a system with N degrees of freedom, the $2N$ functions P and Q are a set of constants of the motion, with the initial values

$$p=P(p,q;0)\,,\qquad q=Q(p,q;0)\,. \tag{4.9}$$

Another time-dependent constant of the motion, such as $I_r(p,q;t)$, can be expressed in terms of P and Q. Since the initial value of I_r is $I_r(p,q;t=0)=C_r(p,q)$, it follows that

$$I_r(p,q;t)=C_r[P(p,q;-t),Q(p,q;-t)]\,. \tag{4.10}$$

To prove (4.10), note that at $t=0$ it satisfies the boundary condition [cf. (4.6)], and that since it is a function of constants of the motion, it satisfies the equation of motion $dI_r/dt=0$.

The construction that led to the constants of motion (4.8) is valid in general. To see this, consider the phase function $C_r(p,q;t)$, whose initial value is $C_r(p,q)$. Its time evolution is given by (2.7). On comparing (2.7) and (4.4) it is evident that an alternative expression for I_r is $I_r(p,q;t)=C_r(p,q;-t)$.

B. Once is enough

For a given collision process it is sufficient to apply the procedure of maximal entropy once.[7,15] Say the initial state $\sigma(p,q)$ has been so specified and is of the form (3.5). It follows from the definition of $I_r(p,q;t)$ as a time-dependent constant of the motion that the exact solution of the Liouville equation for the given initial state is

$$\sigma(p,q;t)=\exp\Big[-\sum_{r=0}^{M}\lambda_r I_r(p,q;t)\Big]\,. \tag{4.11}$$

The reasoning is immediate. Any function of a constant of motion is itself a constant of motion. Hence, $\sigma(p,q;t)$ is a solution of the Liouville equation. It also satisfies the boundary conditions [i.e., for $t\to 0$ it reduces to $\sigma(p,q)$] and is therefore the required solution.

Among all possible densities $\rho(p,q;t)$ the particular result (4.11) is the one of maximal entropy. The proof is based on showing that all the alternative densities satisfy the same constraints as $\sigma(p,q;t)$. Let $\rho(p,q)$ be the initial state that evolved to $\rho(p,q;t)$. Such an initial state must be consistent with the same constraints as $\sigma(p,q)$. Then, using (4.6) and (3.4),

$$\begin{aligned}\langle C_r\rangle&=\mathrm{tr}\{\sigma(p,q)C_r(p,q)\}=\mathrm{tr}\{\rho(p,q)C_r(p,q)\}\\&=\mathrm{tr}\{\sigma(p,q;t)I_r(p,q;t)\}\\&=\mathrm{tr}\{\rho(p,q;t)I_r(p,q;t)\}\,.\end{aligned} \tag{4.12}$$

The functional form (4.11) is that of maximal entropy subject to the $M+1$ constraints $I_r(p,q;t)$. But any alternative density $\rho(p,q;t)$ satisfies the very same constraints. Hence, $\sigma(p,q;t)$ has a higher entropy than any other density that one might consider.

V. THE EQUATION OF CONTINUITY

An equation of motion for the reduced density $p(a;t)$ is derived and discussed in this section. Like the Liouville equation for the detailed density $\rho(p,q;t)$, here too the equation of motion has the interpretation of a continuity equation (in A space). The formal advantage, of course, is that while ρ is defined in a $2N$ dimensional phase space, where N is the number of degrees of freedom (for an isolated collision), the dimension of A space can be much lower (e.g., one, if A is a single phase function). The conceptual advantage is that all irrelevant details have been eliminated from the very start. This elimination of detail is reflected in the observation that, as opposed to the full Liouville equation, the entropy of $p(a;t)$ is not a constant of the motion.

Transport equations for reduced distributions have been previously discussed.[29-32] However, the present equation is exact and, because of its form as a continuity equation, is easily interpreted and manipulated. The entire complexity of the dynamics is contained in the velocities (as defined below). One can show that the leading approximation to the equation of continuity is the Fokker-Planck equation, which is extensively used as an equation of motion for reduced descriptions.

A. The velocity field

The reduced distribution $p(a;t)$ is the average, over the initial density, of $G(a;t)=\delta[A(t)-a]$, where $A(t)=A(p,q;t)$ [cf (2.6)]. Using a dot to denote time derivative, $\dot A(t)$ is given by (2.7) and

$$\begin{aligned}\partial G(a;t)/\partial t&=\frac{d}{dA(t)}\delta(A(t)-a)\dot A(t)\\&=-\frac{d}{da}\delta(A(t)-a)\dot A(t)\,.\end{aligned} \tag{5.1}$$

$\partial p(a;t)/\partial t$ is obtained by averaging (5.1) over the initial state. To write the final result in a compact fashion, it is useful to introduce a velocity field in A space, as follows. From the basic definitions of the reduction, the average rate of change of $A(t)$ on the $A(p,q;t)=a$ shell is given by $v(a;t)$, where

$$v(a;t)=\frac{\mathrm{tr}[\rho(p,q)\delta(A(t)-a)\dot{A}(t)]}{\mathrm{tr}[\rho(p,q)\delta(A(t)-a)]}$$
$$=\mathrm{tr}[\rho(p,q)\delta(A(t)-a)\dot{A}(t)]/p(a;t)\,. \quad (5.2)$$

Using (2.6), (5.1), and (5.2)

$$\partial p(a;t)/\partial t=-\partial[v(a;t)p(a;t)]/\partial a\,. \quad (5.3)$$

and

$$\langle\dot{A}\rangle=\int da\,a\,\partial p(a;t)/\partial t$$
$$=\int p(a;t)v(a;t)da=\langle v\rangle\,. \quad (5.4)$$

The dimension of the A space is determined by the number of different variables A_j, $j=1,\ldots,n$ that are of interest. The variable a is thus, in general, an n dimensional vector (cf. Sec. II). To explicitly exhibit this point, one can rewrite the equation of motion (5.3) as

$$\partial p(\vec{a};t)/\partial t=-\mathrm{div}[\vec{v}p(\vec{a};t)]$$
$$=-\sum_{j=1}^{n}\partial[v_j p(\vec{a};t)]/\partial a_j\,. \quad (5.5)$$

Here $v_j(\vec{a};t)$ is given by (5.2) with $\dot{A}_j(t)$ replacing $\dot{A}(t)$.

As consistency checks, note the following: (a) If as the n variables A_j we take the $2N$ variables p and q, then $p(\vec{a};t)$ is the actual detailed density $\rho(p,q;t)$. It is readily verified that in this case (5.5) is just the Liouville equation

$$\partial\rho/\partial t=-\sum_{j=1}^{2N}[\partial(\dot{q}_j\rho)/\partial q_j+\partial(\dot{p}_j\rho)/\partial p_j]\,. \quad (5.6)$$

(b) Say that for some particular j, $j=k$, $v_k(\vec{a};t)=0$. $p(a_k,t)$ is given, as always, by integrating $p(\vec{a};t)$ over all the a_j's for $j\neq k$. Then it follows from (5.5) (by integration by parts) that if $v_k=0$, then $p(a_k;t)$ does not change with time. Hence, there is no reason why time-independent constants of the motion (i.e., such observables that $\dot{A}_k=0$) should not be included in the set $\{A_j\}$. It simply means that the corresponding term will be missing in the right-hand side of (5.5).

B. Time-dependent constants of the reduced motion

A time-dependent constant of the motion in A space is a phase function $I_r(a;t)$ such that its average value over $p(a;t)$ is time independent:

$$d\left[\int da\,p(a;t)I_r(a;t)\right]\Big/dt=0\,. \quad (5.7)$$

Using the equation of motion (5.5) and integration of (5.7) by parts, we obtain the equation of motion of $I_r(a;t)$,

$$\partial I_r(\vec{a};t)/\partial t=-\vec{v}\,\mathrm{grad}I_r(\vec{a};t)$$
$$=-\sum_{j=1}^{n}v_j\partial I_r(\vec{a};t)/\partial a_j\,, \quad (5.8)$$

or $dI_r(\vec{a};t)/dt=0$.

The equation of motion (5.8) is to be contrasted with the equation of continuity (5.5), which can be written as

$$dp(\vec{a};t)/dt=-p(\vec{a};t)\,\mathrm{div}\vec{v} \quad (5.9)$$

or

$$d\ln p(\vec{a};t)/dt=-\mathrm{div}\vec{v}\,. \quad (5.10)$$

When the $\vec{A}$ space is the entire phase space, it follows from Hamilton's equations that $\mathrm{div}\vec{v}=0$, so that the density is a time-dependent constant of the motion, and the equation of continuity represents the motion of an incompressible "fluid." In the reduced description, $p(\vec{a};t)$ is not necessarily a constant of the motion. In a thermodynamic-like language, an isolated system cannot "relax" because its entropy is a constant of the motion, but a subsystem can. Its entropy is not a constant of the motion, and the coupling between the A space and the other degrees of freedom may cause $p(\vec{a};t)$ to approach a statistical distribution but may also drive it further away from equilibrium.[33,34]

C. Dissipation

That the equation of motion (5.5) for the reduced distribution does allow for a dissipative time evolution also follows from considering a volume element

$$d\vec{a}=\prod_{j=1}^{n}da_j \quad (5.11)$$

In A space. At the time $t'=t+\tau$, where τ is infinitesimal, $a_j'=a_j+\tau v_j$ and hence the volume element is

$$d\vec{a}'=(1+\tau\,\mathrm{div}\vec{v})\,d\vec{a}\,. \quad (5.12)$$

When $\mathrm{div}\vec{v}\neq 0$, a volume element in A space is no longer invariant.

Our primary interest is in the $t\to\infty$ limit of $p(\vec{a};t)$ and so we do not consider in detail the dissipative evolution of $p(\vec{a};t)$ in this article. It is of interest, however, to note that despite the dissipation one can still employ the concepts of invariants. To see this, consider the "velocity" vector

u of $n+1$ components such that $u_0=1$ and the other n components are the v_j's. Then $p(\vec{a};t)$ satisfies the equation

$$\mathrm{div}(\vec{u}p)=0\,, \tag{5.13}$$

where $a_0=t$. The motion of such a fluid is shown schematically in Fig. 1.

D. The surprisal

The prior distribution is uniform in phase space (on the energy shell). It represents the limit of a fully relaxed distribution, where conservation of energy (and other good quantum numbers) is the only constraint. Owing to the finite duration of the collision, the final reduced distribution need not, however, accord with the prior reduced distribution $p^0(a)$. The surprisal is the (logarithmic) measure of the deviance between the two. In phase space the prior distribution is uniform and the surprisal is just the asymptotic limit of $-\ln[\rho(p,q;t)]$ and hence can be expressed in terms of the time-dependent constants of the motion (Sec. IV B).

To interpret the surprisal for the reduced distribution we factor out of $p(\vec{a};t)$ the term that

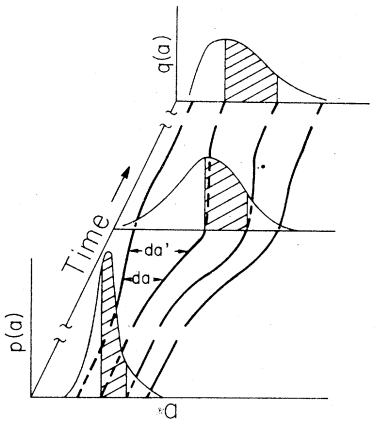

FIG. 1. The evolution of the initial $p(a)$ to the final $q(a)$ reduced distribution as a motion of an incompressible fluid, $\mathrm{div}[\vec{u}p(\vec{a};t)]=0$, where time is one of the coordinates. Shown are the streamlines, where the velocity vector $\vec{u}=(1,\vec{v})$ is the tangent to the line. Before (or after) the collision $\vec{u}=(1,0)$, the streamlines are parallel to the time axis, and there is no dissipation. The nonconservation of the volume element da, Eq. (5.11) is also shown. Note, however, that $\int p(\vec{a};t)\,d\vec{a}=\mathrm{const}$ where the integral is between two streamlines. (The conservation of this integral is shown as the invariance of the dashed area.) The possibility that the entropy first decreases and then increases during the collision (reflected by the narrowing and widening of the cross section between streamlines) has been noted in both analytic solutions (Ref. 33) and computational studies (Ref. 34) of model systems.

evolves to the prior distribution

$$p(\vec{a};t)=p_e(\vec{a};t)\omega(\vec{a};t)\,. \tag{5.14}$$

Here $p_e(\vec{a};t)$ satisfies the same equation of motion as $p(\vec{a};t)$,

$$\partial p_e(\vec{a};t)/\partial t+\mathrm{div}(\vec{v}p_e)=0\,, \tag{5.15}$$

but with the boundary condition that after the collision it equals the prior distribution $p^0(\vec{a})$. The surprisal is thus the asymptotic limit of $I(\vec{a};t)=-\ln\omega(\vec{a};t)$.

The equation of motion for $\omega(\vec{a};t)$ is obtained by inserting the resolution (5.14) in the equation of motion (5.5) of $p(\vec{a};t)$ and using (5.15) so that

$$\partial\omega(\vec{a};t)/\partial t+\vec{v}\,\mathrm{grad}(\omega)=0\,, \tag{5.16}$$

$$\partial I(\vec{a};t)/\partial t+\vec{v}\,\mathrm{grad}(I)=0\,. \tag{5.17}$$

$\omega(\vec{a};t)$ and $I(\vec{a};t)$ are therefore time-dependent constants of the reduced motion (cf. Sec. V B). Despite the dissipation, the surprisal identifies the (asymptotic limit of the) time-dependent constants of the motion.

VI. THE CONSTRAINTS OF THE REDUCED DISTRIBUTION

A complete formal account of the time evolution in a space is now available: The distribution at any time t (and also after the collision) is of the form

$$p(a;t)=p_e(a;t)\exp[-I(a;t)]\,. \tag{6.1}$$

Here $p_e(a;t)$ is the distribution at the time t which becomes, after the collision, the prior distribution $p^0(a)$ (a uniform distribution in phase space or, in quantum mechanics, a uniform distribution over all accessible final quantum states).

The surprisal is the post-collision a dependence of the time-dependent constant of the motion $I(\vec{a};t)$. The average value of $I(\vec{a};t)$ is constant throughout the collision. If after the collision we expand the surprisal

$$I(\vec{a})=\sum_{r=0}\lambda_r C_r(\vec{a}) \tag{6.2}$$

and recall that the Lagrange parameters λ_r are themselves functions of the average values $\langle C_r(\vec{a})\rangle$, it follows that the values of the averages $\langle C_r(\vec{a})\rangle$ constrain the distribution from being fully relaxed. This dynamical interpretation is, of course, the same as that provided directly by the procedure of maximal entropy. Indeed the average value of $I(\vec{a})$ is just the entropy deficiency [cf. (3.12)] which vanishes if (and only if) $p(\vec{a})=p^0(\vec{a})$. This section examines some dynamic considerations which guide the determination of the functions $C_r(a)$ in the expansion (6.2) for the surprisal. The infor-

mation-theoretic approach (which can be derived from a variational principle[15]) is discussed in Sec. C of Appendix A.

One possible query may arise concerning the functional form (6.1). After all, one can set $p(a;t)=p_e(a;t)\omega(a;t)$, where $\omega(a;t)$ is also [cf. (5.16)] a time-dependent constant of the motion. Why not analyze $\omega(a;t)$ instead of $I(a;t)$. Well, one can, and it is shown in Appendix B that this is strictly equivalent to the analysis of $I(a;t)$, except that the structure of $I(a;t)$ is simpler.

A. Sum rules

The most practical route to the identification of the constraints [the functions $C_r(a)$ in (6.2)] that govern the final-state distribution is the use of so-called "sum rules." This route has been extensively used,[6,10,16] on the basis of information-theoretic considerations. Here we consider the dynamical interpretation. The application to heavy-ion reactions[17] is discussed in the companion paper.[11]

Let $\langle C_r\rangle$ be the initial mean value of a function $C_r(a)$,

$$\langle C_r\rangle = \int da C_r(a)p(a)\,, \tag{6.3}$$

and let $\Delta\langle C_r\rangle$ be the change in the mean value of $C_r(a)$ between the initial and the final state. It is often the case that the following "sum rule" is satisfied:

$$\Delta\langle C_r\rangle = \sum_s (F_{sr}-\delta_{sr})\langle C_s\rangle\,. \tag{6.4}$$

That is, that the change in $\langle C_r\rangle$ is a linear combination of the initial values of several functions $C_s(a)$. One can rewrite (6.4) as

$$\langle C_r\rangle' = \sum_s F_{sr}\langle C_s\rangle\,, \tag{6.5}$$

where the prime refers to the final value. If the numerical coefficients F_{sr} are independent of the initial state, it follows that there is a time-dependent constant of motion whose initial form (i.e., for $t\to-\infty$) is

$$I_{\rm in}(a) = \sum_s F_{sr}C_s(a)\,, \tag{6.6}$$

and whose final form is

$$I_{\rm out}(a) = C_r(a)\,. \tag{6.7}$$

Taking the expectation value of $I_{\rm in}(a)$ over the initial state and equating it to the expectation value of $I_{\rm out}(a)$ in the final state [remember (5.7)], we recover (6.5).

An observable that satisfies a sum rule of the form (6.5) with coefficients F that are independent of the particular initial state is thus guaranteed to be an asymptotic value of a constant of motion and hence is to be used as a constraint.

Sum rules need not be purely of phenomenological origin. In Sec. VIC we shall derive exact sum rules for such systems where the Hamiltonian is known. A very important practical source of sum rules is models. It is often possible to summarize a model in the form of a sum rule such as (6.5), and an actual case is discussed in the companion paper.[11] Sum rules which derive from models identify, of course, the constraint implied by the model, which need not be the completely correct constraint. Even so, if the model is any good, the constraint is likely to be quite reasonable. Indeed, one can argue[6] that models should be used primarily for the identification of the constants of the motion and that the actual final reduced distribution $q(a)$ be obtained using the procedure of maximal entropy as

$$q(a) = p^0(a)\exp\left[-\sum_{r=0}\lambda_r C_r(a)\right], \tag{6.8}$$

where the values of the λ_r's are determined from the sum rules. The reason is, of course, that to obtain the distribution the model builder often needs to introduce additional assumptions which are not germane to the central idea. A more objective route is to limit the model to relations between average values and let the maximum entropy procedure provide the least biased route from these averages to the actual distribution.

B. The time-dependent constraints

The traditional route to the solution of a partial differential equation such as (5.17) is the separation of variables. Hence, to solve for $I(a;t)$ we try the ansatz

$$I(a;t) = \sum_{r=0}^{M}\lambda_r(t)C_r(a)\,, \tag{6.9}$$

where the linearly independent functions $C_r(a)$ do not depend on time, and the expansion (6.9) is to be valid for all times throughout the collision. Inserting (6.9) in the equation of motion (5.17), we have

$$\sum_{s=0}^{M} C_s(\vec{a})\partial\lambda_s(t)/\partial t = -\sum_{r=0}^{M}\lambda_r(t)\vec{v}\cdot\mathrm{grad}C_r(\vec{a})\,. \tag{6.10}$$

Since the functions $C_s(\vec{a})$ are linearly independent, one can equate the coefficients of $C_s(\vec{a})$ on both sides of (6.10). It follows that (6.9) is an exact expansion if (and only if) the functions $C_s(\vec{a})$ are closed under the operation $\vec{v}\cdot$ grad. Explicitly,

$$-\vec{v}\cdot \mathrm{grad} C_r(\vec{a}) = \sum_s g_{sr} C_s(\vec{a}) . \tag{6.11}$$

Using (6.11) in (6.10) we obtain the equation of motion for the expansion coefficients

$$\partial \lambda_s(t)/\partial t = \sum_{r=0}^{M} g_{sr}\lambda_r(t) . \tag{6.12}$$

A closed set of functions that satisfies (6.11) obtains, for example, when the velocities are linear in a,

$$v_j(\vec{a};t) = \sum_{k=1}^{n} g_{kj}(t) a_k . \tag{6.13}$$

In this case, all the moments, $\vec{r} = r_1, r_2, \ldots,$

$$C_{\vec{r}}(\vec{a}) = \prod_i a_i^{r_i} , \tag{6.14}$$

up to any given order $\sum r_i = R$, form a closed set. If the A space is one dimensional, then $C_r(a) = a^r$ are the ordinary moments.

The closure condition (6.11) is, of course, closely related to the closure condition discussed in connection with the detailed solution in phase space.[7] There one required a set which is closed under the operation $\{H, \}$, i.e., that of taking the Poisson bracket with the Hamiltonian. Indeed, the form (6.13) does obtain if the A's themselves are closed,[35]

$$\{H, A_j\} = \sum g_{kj} A_k . \tag{6.15}$$

Recalling that $\dot{A}_j = -\{H, A_j\}$ one obtains (6.13) from (6.15) upon averaging on the $\vec{A}(p,q) = \vec{a}$ shell. It is also possible[15] to extend such considerations to rearrangement collisions, where different sets of constraints may be useful in different arrangement channels.

C. Sum rules for time-independent constraints

When a set of functions $C_r(a)$, closed in the sense of (6.11), can be determined, then one has an expansion for the surprisal with a fixed set of time-independent constraints, which is valid throughout the collision. Such a set will also satisfy sum rules (Sec. VIA) at any time t and also after the collision. To show this, consider the equation of motion for $\langle C_r \rangle(t)$:

$$\begin{aligned} \partial \langle C_r \rangle(t)/\partial t &= \int da\, C_r(a) \partial p(a;t)/\partial t \\ &= -\int da\, C_r(a)\, \mathrm{div}[\vec{v} p(a;t)] \\ &= \int da\, p(a;t) \vec{v}\, \mathrm{grad} C_r(a) \\ &= -\sum_s g_{sr} \langle C_s \rangle(t) . \end{aligned} \tag{6.16}$$

The equation of motion (6.16) for the mean values of the constraints is the counterpart of the equation of motion (6.12) for the Lagrange parameters. Either equation can be formally integrated by the introduction of a matrix $\vec{F}(t)$,

$$-\partial \vec{F}/\partial t = \vec{F}\cdot \vec{g} , \tag{6.17}$$

with $\vec{F}_{\mathrm{in}} = \vec{I}$. Then

$$\langle C_r \rangle(t) = \sum_s F_{sr}(t) \langle C_s \rangle , \tag{6.18}$$

and the matrix $\vec{F}$ of (6.5) is simply $\vec{F}_{\mathrm{out}}$, the asymptotic limit of $\vec{F}(t)$.

Consider, finally, different collision processes which do, however, satisfy the same type of sum rules, i.e.,

$$\Delta \langle C_r \rangle = \sum_s (F_{sr} - \delta_{sr}) \langle C_s \rangle , \tag{6.19}$$

where, of course, the elements of the matrix $\vec{F}$ may have different values for different processes. Here too, the functional form of the surprisal will be the same for all the different processes, provided that the initial states are specified by the same set of constraints. It is indeed possible[7] to argue that similar reactions have similar constraints.

VII. CONCLUDING REMARKS

Surprisal analysis has typically been applied to only partially resolved final state distributions. A statistical dynamics formulation which provides the theoretical foundation for such phenomenological applications has been presented. The central result, in this respect, is the proof that even for a reduced description the surprisal remains a time-dependent constant of the motion. This enables one to expand the surprisal, both during and after the collision, in a set of time-independent constraints. During the collision the expansion coefficients will be time dependent. The identification of these constraints can be made on theoretical grounds. A much more practical method is the use of sum rules which state that a change in the mean value of some observable (e.g., the kinetic energy) during the collision is linearly dependent on the initial values of one or more observables. Observables related in this fashion are the required constraints.

ACKNOWLEDGMENTS

We thank Professor S. Alexander, Professor Y. Shaham, and Professor S. G. Steadman for their comments. This work was supported by the Office of Scientific Research, U. S. A. F. under Grant No. AFOSR 77-3135.

APPENDIX A: PROCEDURE OF MAXIMAL ENTROPY IN A SPACE

Critics of the maximal entropy point of view tend to argue that it violates the principle of conservation of effort. The proper answer is that it does not; all that is avoided is the unnecessary effort. This is achieved by restricting attention to those variables that are relevant (i.e., that constrain) the density of interest. The description of the collision in the reduced A space can be used as an explicit illustration of this claim. In particular, we show how such a point of view can be naturally accommodated in the framework of statistical theories.[18] The removal of irrelevant details by projection operator techniques is discussed in Appendix. B.

A. The detailed vs the reduced solution

It is indeed possible to make an effort to "balance" and identify the origin of the savings. A person who is only concerned with a reduced description will not be able to use his results once scattering experiments at higher resolution are performed. This is not the case for the person with a complete solution. To see this, consider first a fully detailed solution of the scattering problem. This means the ability to compute the final state for any initial state of interest. In classical mechanics this requires computing classical trajectories for a dense mesh of initial values of p and q. Every such trajectory specifies the density that evolved from an initial δ function, i.e.,

$$G(p,q;t)=\delta(p(t)-p)\delta(q(t)-q)\,. \tag{A1}$$

The density that corresponds to any desired initial state is obtained by averaging (A1) over the initial distribution of p and q as in (2.6).

For the person of more practical nature, who is only able to specify the initial state in A space, the general solution (A1) is sufficient but not really necessary. The reason is that he is unable to resolve different initial states which correspond to the same reduced initial distribution $p(a)$. He can, of course, introduce some ansatz for his initial $\rho(p,q)$ density and use (2.6). In so doing he will have to average over all those initial p,q values that are on the $A(p,q)=a$ shell. All the effort that went into calculating the different trajectories that originate from distinct p,q points on the A shell is thereby lost. For example, if the initial state is specified by the M constraints (3.3), then it is, in principle, sufficient to solve the equation of motion [Eq. (5.8)] M times. The effort is considerably reduced, and the price is that such a description would only suffice for an initial state specified by the M constraints $C_r(a), r=1,2,\ldots,M$. (Note, however, that changes in the values $\langle C_r\rangle$ of the constraints can be accommodated without solving again the equations of motion.) However, when the initial state is changed in such a manner that it can no longer be specified by the M constraints, one needs to redetermine new constants of the motion. This is not so for the person with the fully detailed solution. If his tape with all the trajectories has not been erased in the meantime, he can accommodate any modifications in the initial state.

Finally, one reaches the limit of a uniform initial state. There is only one constraint ($M=0$), namely, that the initial density is normalized. The integration of the equation of motion is immediate: A normalized initial state remains normalized. The final state density is thus available (i.e., it is uniform) without any numerical integrations of trajectories. To be sure, if one did determine the trajectories and used those to compute the final state for a uniform initial state, one would reach the same conclusion. However, there is clearly no objection to using the intertwining theorem[24] in order to bypass this laborious task and to proceed directly to the conclusion that a uniform initial state is invariant. The maximum entropy procedure seeks to point out that this line of reasoning can be usefully extended, that there are additional invariants that can be used, at least in principle, to determine the functional form of the final state for more detailed initial states.

B. Statistical theories and the fluctuation term

Statistical theories[18] often seek to resolve the cross section (and the S matrix) into a "direct" and a "fluctuation" term. This is typically done by considering an "average" over initial states, such that the direct contribution is the average value and the fluctuation is the deviance from the average. Except when the average is over the initial energy, the precise nature of the averaging is seldom specified. Here we point out that the procedure of maximal entropy provides a well defined framework for the introduction and implementation of these ideas.

The average is here defined by specifying which quantities are held constant. All other quantities are allowed to vary and one seeks the greatest possible latitude for such variations. In other words, the result should be maximally noncommitted with respect to the unconstrained variables. This desideratum is precisely provided by the procedure of maximal entropy. One determines the density which is consistent with the values of the specified quantities and is otherwise least biased. In a quantal treatment it can be applied to the S matrix itself.

The simplest example of such averaging is when only the time-independent constants of the motion are constrained to their values. Then the average density is the uniform one (on the energy shell) and every deviance represents a fluctuation. Another example is the person with resolution only in A space. For him, any density in phase space can be resolved as

$$\rho(p,q;t)=\sigma(p,q;t)+\delta\rho(p,q;t)\,. \tag{A2}$$

Here $\sigma(p,q;t)$ is the density of maximal entropy as discussed in Sec. IV [i.e., (4.11)]. The fluctuation term $\delta\rho(p,q;t)$ cannot be observed by such a person, for, by construction, both ρ and σ yield identical values for those variables that such a person can measure. Only by refining the measurement procedure, i.e., by resolving different states on the $A(p,q)=a$ shell can one discern the fluctuation term.

The considerations above also illustrate once more the sense in which the maximal entropy procedure foregoes the computation of irrelevant details; e.g., why should one determine $\delta\rho(p,q;t)$, if it has no observable consequences on the level of detail which is required? An explicit form for the average part is derived in Appendix B, where a projection operator technique is employed to project out the fluctuating part. There we also discuss a similar resolution for observables.

C. Surprisal synthesis as an inductive inference

Sections V and VI provided a dynamical interpretation of the surprisal of the exact reduced density and a determination of a set of constraints which specifies an exact solution of the equation of motion. Once those objectives are realized, it is possible to formulate approximations to these exact results. It is useful, however, to note that the point of view discussed in the introduction to this appendix provides an alternative method. The difference is not one of technical detail, but one of motivation. If a set of variables is deemed particularly relevant to a given system, then why not determine the final distribution by maximizing the entropy subject to the expectation values of these variables? The result will not necessarily be exact, but will be the most conservative inference that can be made from the given data. Synthesis is thus regarded as an inductive inference[3,4] in a manner first discussed in connection with statistical mechanics.[3,14] In such an approach the constraints have the same interpretation as those variables which are relevant to the particular dynamics. Sum rules (Sec. VI A) are very useful in this respect. Any irrelevant constraints are identified as such by the procedure, since their Lagrange parameters are found to equal zero[25] (or are very small). Moreover, the procedure is backed by a variational principle.[15] By including more and more constraints, one is sure to converge to the exact dynamical results. In practice, when the constraints are sensibly chosen, the inclusion of one[36] or two[37] constraints results in very close accord with exact numerical solutions of the Schrödinger equation. It follows from the discussion in Sec. III C that the measure of convergence is[10] $DS[\rho|\sigma]$, where ρ is the exact result and σ is the maximum entropy inference. When σ is a good approximation, $DS[\rho|\sigma]$ is numerically equal to the square of the fractional error in the distribution function.

APPENIDX B: PROJECTING ON THE MAXIMUM ENTROPY PART

The use of a projection operator technique[5] for separating a density into a term given by the procedure of maximum entropy and a fluctuating part is discussed. The equivalence of a linear expansion and the maximum entropy form is demonstrated. As discussed in Appendix A, the fluctuation part corresponds to the irrelevant details which get averaged out in a specified level of description. The same details may of course be quite relevant when a finer resolution is available.

A. The reciprocal set

The discussion is based on the introduction of a set of observables that is reciprocal [as defined by (B3) below] to the set of constraints. Such a set will offer a linear expansion for the density of maximum entropy. Similarly, the constraints themselves will be shown to offer a linear expansion for the relevant part of an observable [as defined in (B16)].

To simplify the notation, a scalar product of two phase functions $(X\cdot Y)$ is defined by

$$(X\cdot Y)=\mathrm{tr}(X\cdot Y)\,. \tag{B1}$$

Here tr is, as before, an integration over phase space, the point being that the proofs obtain equally well for the quantal case where X and Y are operators.

Let σ be a distribution in phase space of maximal entropy subject to the constraints $\langle C_r\rangle$, $r=0,1,\ldots,M$. Then the set of phase functions

$$B_r=\partial\sigma/\partial\langle C_r\rangle,\quad r=0,1,\ldots,M \tag{B2}$$

is reciprocal (i.e., biorthogonal) to the set of constraints

$$\begin{aligned}(C_r\cdot B_s)&=\mathrm{tr}(C_r\partial\sigma/\partial\langle C_s\rangle)\\&=\partial\,\mathrm{tr}(C_r\sigma)/\partial\langle C_s\rangle\\&=\partial\langle C_r\rangle/\partial\langle C_s\rangle=\delta_{r,s}\,.\end{aligned} \tag{B3}$$

The last equality follows from the linear independence[25] of the constraints used to specify σ. If σ is time dependent, then so will be the observables B_r of the reciprocal set.

B. The projectors

The set of constraints $\{C_r\}$ and the reciprocal set $\{B_r\}$ are thus biorthogonal. Either set can be used as a basis for a linear expansion, and the expansion coefficients are the scalar products with the corresponding members of the other set. For example,

$$X = \sum_{r=0}^{M} \mu_r B_r + \cdots \tag{B4}$$

with $\mu_r = (C_r \cdot X)$. One can thus define two types of projection operators. The first is a Robertson-type[5] projector

$$PX = \sum_{r=0}^{M} (C_r \cdot X) B_r \,. \tag{B5}$$

Here the basis functions are time dependent (if σ is). Using the general relation

$$(PX \cdot Y) = (X \cdot P^\dagger Y) \,, \tag{B6}$$

we obtain from (B5)

$$P^\dagger X = \sum_{r=0}^{M} (B_r \cdot X) C_r \,. \tag{B7}$$

Here the basis functions are not necessarily time dependent. The second (or Mori[38]-type) projector is the adjoint of the first.

The two special properties of the projectors are (a) σ is invariant under P,

$$\sigma = P\sigma = \sum_{r=0}^{M} (C_r \cdot \sigma) B_r \,. \tag{B8}$$

This property is the promised linear expansion of σ. One can either expand $\ln\sigma$ in the C_r's or σ in their reciprocal B_r's, and the two expansions are strictly equivalent. To prove (B8) note [e.g., using (4.11)] that $\sigma = -\partial\sigma/\partial\lambda_0$ and hence that $\partial\langle C_r\rangle/\partial\lambda_0 = \partial(C_r \cdot \sigma)/\partial\lambda_0 = -(C_r \cdot \sigma)$. Thus,

$$\sigma = -\partial\sigma/\partial\lambda_0 = -\sum_{r=0}^{M} (\partial\sigma/\partial\langle C_r\rangle)(\partial\langle C_r\rangle/\partial\lambda_0)$$
$$= \sum_{r=0}^{M} (C_r \cdot \sigma) B_r \,. \tag{B9}$$

(b) The constraints are invariant under $P^\dagger$:

$$P^\dagger C_r = C_r \,. \tag{B10}$$

The proof is immediate, using the definition (B7) and the biorthogonality property (B3).

C. Projecting out the irrelevant part

Consider a distribution ρ in phase space and let σ be the distribution of maximal entropy which is consistent with ρ on the values of the constraints,

$$(C_r \cdot \rho) = (C_r \cdot \sigma) \,. \tag{B11}$$

Then, using (B11), (B8), and (B5),

$$P\rho = \sum_{r=0}^{M} (C_r \cdot \rho) B_r$$
$$= \sum_{r=0}^{M} (C_r \cdot \sigma) B_r = P\sigma = \sigma \,. \tag{B12}$$

P projects the average or direct part out of any density in phase space. The fluctuating part is $(1-P)\rho$. It can also be shown that, for any small variation $\delta\rho$ in ρ, P projects out its average part,

$$\delta\sigma = P\delta\rho \,. \tag{B13}$$

The proof is immediate, using (B12) and the chain rule.

D. The fluctuating term for observables

The resolution into the average and fluctuating terms can also be introduced for observables. For any observable X, the adjoint operator $P^\dagger$ projects onto its average part. To see this, note that

$$\langle X\rangle = \mathrm{tr}[X\rho(t)]$$
$$= \mathrm{tr}[XP\rho(t)] + \mathrm{tr}[X(1-P)\rho(t)]$$
$$= \mathrm{tr}[X\sigma(t)] + \mathrm{tr}[X\delta\rho(t)] \,. \tag{B14}$$

Using the definition (B6) of the adjoint, we have shown in (B14) that X can be resolved as

$$X = X_a(t) + \delta X(t) \,, \tag{B15}$$

with $X_a(t) = P^\dagger X$. The identification of $X_a(t)$ as the average part follows since

$$\mathrm{tr}[X_a(t)\rho(t)] = \mathrm{tr}[X\sigma(t)] = \mathrm{tr}[X_a(t)\sigma(t)] \,, \tag{B16}$$

whereas the fluctuating part $\delta X(t) = (1-P^\dagger)X$ has a vanishing average over $\sigma(t)$,

$$\mathrm{tr}[\delta X(t)\sigma(t)] = 0 \,. \tag{B17}$$

If X is one of the constraints, it follows from (B10) that it has no fluctuating part.

---

[1]For a current status report, see *The Maximum Entropy Formalism*, edited by R. D. Levine and M. Tribus (MIT Press, Cambridge, 1978).
[2]W. M. Elsasser, Phys. Rev. 52, 987 (1937).
[3]E. T. Jaynes, Phys. Rev. 106, 620 (1957) and in Ref. 1.
[4]E. H. Wichmann, J. Math. Phys. 4, 884 (1963).
[5]B. Robertson, Phys. Rev. 144, 151 (1966) and in Ref. 1.

[6]R. D. Levine, Ann. Rev. Phys. Chem. 29, 59 (1978) and in Ref. 1.
[7]Y. Alhassid and R. D. Levine, Phys. Rev. A 18, 89 (1978).
[8]R. D. Levine and R. B. Bernstein, *Dynamics of Molecular Collisions*, edited by W. H. Miller (Plenum, New York, 1976), p. 323.
[9]R. D. Levine and A. Ben-Shaul, *Chemical and Biochemical Applications of Lasers*, edited by C. B. Moore (Academic, New York, 1977), Vol. II, p. 145.
[10]For a user's handbook, see R. D. Levine and J. L. Kinsey, *Atom Molecule Collision Theory—A Guide for the Experimentalist*, edited by R. B. Bernstein (Plenum, New York, 1979), Chap. 22.
[11]Y. Alhassid, R. D. Levine, J. S. Karp, and S. G. Steadman, following paper, Phys. Rev. C 20, 1789 (1979).
[12]R. D. Levine, S. G. Steadman, J. S. Karp, and Y. Alhassid, Phys. Rev. Lett. 41, 1537 (1978).
[13]Y. Alhassid, Ph.D. thesis, The Hebrew University, Jerusalem, 1978 (unpublished).
[14]A. Katz, *Principles of Statistical Mechanics—The Information Theory Approach* (Freeman, San Francisco, 1967).
[15]Y. Alhassid and R. D. Levine, J. Chem. Phys. 67, 4321 (1977).
[16]See, for example, R. Kosloff, A. Kafri, and R. D. Levine, Astrophys. J. 215, 497 (1977) and A. Kafri, E. Pollak, R. Kosloff, and R. D. Levine, Chem. Phys. Lett. 33, 201 (1975) for two different applications.
[17]For sum rules in nuclear heavy-ion reactions, see, for example: P. J. Siemens, J. P. Bondorf, D. H. E. Gross, and F. Dickmann, Phys. Lett. 36B, 24 (1971); D. M. Brink, *ibid.* 40B, 37 (1972); T. Mikumo, T. Kohno, K. Katori, T. Motobayashi, S. Nakajima, M. Yoshie, and H. Kamitsubo, Phys. Rev. C 14, 1458 (1976); C. K. Gelbke, C. Olmer, M. Buenerd, D. L. Hendrie, J. Mahoney, M. C. Mermaz, and D. K. Scott, Phys. Rep. 42, 311 (1978); T. M. Cormier, P. Braun-Munzinger, P. M. Cormier, J. W. Harris, and L. L. Lee, Jr., Phys. Rev. C 16, 215 (1977).
[18]G. E. Brown, Rev. Mod. Phys. 31, 893 (1959); H. Feshbach, Ann. Phys. (N.Y.) 19, 287 (1962); P. A. Moldauer, Phys. Rev. 177, 1841 (1969); M. Kawai, A. K. Kerman, and K. W. McVoy, Ann. Phys. (N.Y.) 75, 156 (1973); H. M. Hufmann, J. Richert, J. W. Tepel, and H. A. Weidenmüller, *ibid.* 90, 403 (1975).
[19]R. C. Tolman, *The Principles of Statistical Mechanics* (Clarendon, Oxford, 1938).
[20]For a similar point of view in statistical mechanics, see R. Zwanzig, Phys. Rev. 124, 983 (1961).
[21]C. E. Shannon, Bell Syst. Tech. J. 27, 379 (1948); A. I. Khinchin, *Mathematical Foundation of Information Theory* (Dover, New York, 1957); M. Tribus, *Rational Descriptions, Decisions and Design* (Pergamon, New York, 1969).
[22]N. Wiener, *Cybernetics* (Wiley, New York, 1948); S. Goldman, *Information Theory* (Dover, New York, 1953).
[23]J. W. Gibbs, *Elementary Principles in Statistical Mechanics* (Yale University Press, New Haven, 1902),
[24]This property is known as the intertwining theorem in quantal collision theory. See, for example, R. D. Levine, *Quantum Mechanics of Molecular Rate Processes* (Clarendon, Oxford, 1969).
[25]Y. Alhassid, N. Agmon, and R. D. Levine, Chem. Phys. Lett. 53, 22 (1978) and in Ref. 1.
[26]Even without the information-theoretic (Ref. 22) interpretation, examining the deviance of the actual final state distribution from $p^0(a)$ is a sensible approach. The reason is that $p^0(a)\,da$ is [cf. (3.7)] essentially the number of accessible final states when A is in the range a, $a+da$. Hence $q(a)/p^0(a)$ is the average magnitude of the transition rate into a final quantum state (i.e., the average of the squared element of the T matrix). The division by $p^0(a)$ eliminates, therefore, the phase space bias, which is due to the fact that $p^0(a)$ may be strongly varying with a. Any nonuniformity of $q(a)/p^0(a)$ reflects a bias due to the dynamics. Another result of the elimination of the phase space bias is that $q(a)/p^0(a)$ has a simple behavior under time reversal. See J. L. Kinsey, J. Chem. Phys. 54, 1206 (1971) and Ref. 10.
[27]J. Dothan, Phys. Rev. D 2, 2944 (1970).
[28]R. L. Anderson, S. Kumei, and C. E. Wulfman, Rev. Mex. Fis. 21, 1 (1972); E. B. Aronson, I. A. Malkin, and V. I. Man'ko, Fiz. Elem. Chastits At. Yadra 5, 122 (1974) [Sov. J. Part. Nucl. 5, 47 (1974)].
[29]For a recent statistical mechanics discussion, see K. S. J. Nordholm and R. Zwanzig, J. Stat. Phys. 13, 347 (1975).
[30]Applications to nuclear reactions are discussed, for example, by W. Nörenberg, Phys. Lett. 53B, 289 (1974); S. Ayik, B. Schurmann, and W. Nörenberg, Z. Phys. A279, 299 (1976); J. S. Sventek and L. G. Moretto, Phys. Lett. 65B, 326 (1976); H. Hofmann and F. P. Siemens, Nucl. Phys. A275, 464 (1977); D. Agassi, C. M. Ko, and H. A. Weidenmüller, Ann. Phys. (N.Y.) 107, 140 (1977).
[31]For applications to molecular collisions, see, for example, S. D. Augustin and H. A. Rabitz, J. Chem. Phys. 67, 2082 (1977); G. C. Schatz, *ibid.* 66, 5220 (1977); B. C. Eu, Chem. Phys. 27, 301 (1978).
[32]For an information-theoretic derivation and analysis of transport equations, see, for example, R. D. Levine, J. Chem. Phys. 65, 3302 (1976); R. D. Levine and M. Berrondo, Chem. Phys. Lett. 47, 399 (1977). I. Oppenheim and R. D. Levine, Physica (to be published) and Ref. 5.
[33]G. L. Hofacker and R. D. Levine, Chem. Phys. Lett. 15, 165 (1972).
[34]See, for example, J. N. L. Connor, W. Jakubetz, and J. Manz, Chem. Phys. Lett. 44, 516 (1976); R. K. Nesbet, *ibid.* 42, 197 (1976); R. E. Wyatt, *State-to-State Chemistry*, edited by P. R. Brooks and E. F. Hayes (A. C. S., Washington, 1976), p. 185.
[35]It is important to note that if H is time independent then $\mathrm{tr}\,g=0$, so that when (6.15) obtains $\mathrm{div}\,\vec{v}=0$ and the motion is dissipationless. It is, however, possible that (6.13) obtains even if (6.15) does not, as will be discussed in a separate note on dissipative time evolution. In this latter case $\mathrm{tr}\,g\neq 0$.
[36]For inelastic collisions, see, for example, I. Procaccia and R. D. Levine, J. Chem. Phys. 64, 796 (1976). For rearrangement collisions, see R. K. Nesbet and D. C. Clary, Chem. Phys. Lett. 59, 437 (1978).
[37]U. Halavee and R. D. Levine, Chem. Phys. Lett. 46, 35 (1977).
[38]H. Mori, Prog. Theor. Phys. 33, 423 (1965).