

## Sum Rule for Multiscale Representations of Kinematically Described Systems

Yaneer Bar-Yam,  
New England Complex Systems Institute,  
24 Mt. Auburn St., Cambridge, MA 02138

We derive a sum rule that constrains the scale based decomposition of the trajectories of finite systems of particles. The sum rule reflects a tradeoff between the finer and larger scale collective degrees of freedom. For short duration trajectories, where acceleration is irrelevant, the sum rule can be related to the moment of inertia and the kinetic energy (times a characteristic time squared). Thus, two non-equilibrium systems that have the same kinetic energy and moment of inertia can, when compared to each other, have different scales of behavior, but if one of them has larger scales of behavior than the other, it must compensate by also having smaller scales of behavior. In the context of coherence or correlation, the larger scale of behavior corresponds to the collective motion, while the smaller scales of behavior correspond to the *relative* motion of correlated particles. For longer duration trajectories, the sum rule includes the full effective moment of inertia of the system in space-time with respect to an external frame of reference, providing the possibility of relating the class of systems that can exist in the same space-time domain.

### 1. Introduction

Complex systems are often characterized as systems that have behavior on multiple scales. Existing multiscale description methods with disparate domains of application, including fractals, low-dimensional chaos, scaling, and renormalization group are generally well suited for scale invariant systems where the basic description framework does not change as a function of scale. A more general approach would allow treating systems whose descriptions at different scales vary in the nature of their parameterization (i.e. which parameters and representation framework are used, rather than just changes in the parameter values). Information theory<sup>1-3</sup> may provide aspects of the necessary approach. Standard analogies between information theory and physics focus on the thermodynamic entropy and the information necessary to specify an ensemble microstate. In the context of dynamical systems, the Kolmogorov-Sinai entropy<sup>4</sup> provides additional contact as the microscopic limit of the entropy generation rate and is related to Lyapunov exponents.<sup>5,6</sup> Efforts to further link information theory and physics have taken various forms in recent years. Many of these are concerned with algorithmic aspects of information,<sup>7-10</sup> others with the nonlinear dynamics of nonequilibrium systems and entropy measures.<sup>11-21</sup>

Conventional nonlinear dynamics<sup>11-21</sup> has focused on the *physical mechanism* by which there can appear individual macroscopically observable collective coordinates and their generic behaviors, especially of real-number time series  $x(t)$  associated with a

measurement of nonequilibrium systems. Modern analysis of such time series proceeds in a manner motivated by studies of chaotic low-dimensional deterministic dynamical systems possibly with the addition of noise. While the utility of this approach has been demonstrated, there remain basic issues in understanding the relationship between spatio-temporal behaviors of nonequilibrium systems and the low dimensional dynamical behavior represented by individual time series. An overarching motivation of this paper is the discussion of the *informational* (not physical) *mechanisms* by which the high dimensional microscopic description of a system in terms of particles can be related to macroscopic descriptions either in terms of much lower dimension spatio-temporal field equations or to low dimensional dynamical equations associated with a few macroscopic observables. We assume that the transition between many fine scale parameters and various forms of fewer macroscopic parameters as a function of the *scale of observation* occurs in a way that can be understood from fundamental principles. In this context, it is natural to expect that low dimensional non-equilibrium dynamics is applicable to a system when there are a few large scale real-number parameters that describe the dynamical behavior over many scales and they are well separated from the rest of the fine scale degrees of freedom, and not otherwise. In this context, this paper is a step towards the more general understanding of how descriptions depend on the scale of observation.

Our approach to multiscale representations focuses on a framework to obtain the relevant parameters for describing the behavior of a system at a particular scale of observation, which in general may differ from the relevant parameters at other scales of observation. Initial efforts in this direction by us<sup>22</sup> and others<sup>23</sup> considered the scale dependence of the information necessary to describe the system, and used this scale dependence, "complexity profile", to characterize various systems. This approach builds on Shannon's noisy channel theory for discrete and continuous systems or the similar Kolmogorov - entropy, which has received only limited attention in application to physical systems.<sup>24,25</sup> We have suggested that a fundamental tradeoff exists between fine scale and large scale behaviors. The greater the length of the large scale description of the system, the shorter the length of the fine scale description. Intuitively, the existence of observable behaviors at a large scale implies coherence or correlations of the motions of particles that comprise the system. This, in turn, implies that fine scale motions are not independent and therefore the information necessary to describe the fine scale behavior is significantly reduced. This tradeoff is apparent in the macroscopic microscopic tradeoff of kinetic flow energy and thermal energy found in Bernoulli's law and its diverse applications. The purpose of this article is to describe a sum rule for systems of particles which more generally and explicitly captures this tradeoff. We argue that description lengths at different scales of observation of a system of particles in the same space-time domain are constrained by a quantity, the sum of squares of particle space-time coordinates, that characterizes a class of non-equilibrium systems. The relationship between this quantity and the scale based decomposition of the system's collective degrees of freedom provides a link between system classification and multiscale representations. We first review some aspects of the multiscale approach.

According to the multiscale perspective, we consider a sequence of observers who progressively do not recognize fine scale distinctions between possible states of a system. The aggregation of system states corresponds to less information about the system. This separates the information into that which can be known to the observer, and that which cannot be known. We note that a key to multiscale applications of information theory in the context of physical systems is understanding the relationship between uncertainties in physical system observations and uncertainties in information theory applied to system states. The observations at a particular scale are generally linked to observation of system behaviors which are not obscured by noise. Determining implications of this approach is a key objective of this paper.

Formally, a multiscale representation of a system in space and time can be based upon the use of a probability density of the space of trajectories of the system (i.e. movies),  $P(\{x(t)\})$ . For the case of small duration trajectories this is equivalent to considering the phase space probability and coarse graining the phase space--- acceleration is not significant, and  $P(\{x(t)\})$  can be mapped onto  $P(\{x,p\})$ . The standard information measure of this probability distribution is generally considered to be analogous to the phenomenological entropy of the system.

$$S = - \int P(\{x,p\}) \log P(\{x,p\}) \quad (1)$$

Microscopically distinguishable states have a volume of phase space given by  $h^{3N}$ . Observers unable to distinguish positions within a range  $\Delta x$  and momenta within a range  $\Delta p$  so that  $\Delta x \Delta p \gg h$  do not distinguish all microstates and, in effect, aggregate them together. This corresponds to an observation error and an underlying uncertainty in the observations. Then we have the analog of Shannon's noisy channel theorem

$$S(0) = C(\sigma) + S(\sigma) \quad (2)$$

where  $\sigma$  is a measure of the observation error,  $C(\sigma)$  is the complexity, the information possessed by the observer, and  $S(\sigma)$  is the entropy, the information not accessible to the observer.  $C(\sigma)$  is calculated from the probability distribution of macrostates (aggregated microscopic states), while  $S(\sigma)$  is the information needed to describe the microstate given the macrostate.

The approach of Shannon and Kolmogorov to continuous trajectories in a continuous phase space involves coarse graining the phase space uncertainty (with precision  $\epsilon$ ) and the temporal sampling (with precision  $\tau$ ) so that finite sums can be evaluated to obtain the space-time entropy. Let  $P(\{x(t)\})$  be the probability of trajectories of the system, which is taken to be classical (non quantum mechanical and non relativistic). Then the information (often called Kolmogorov's  $K$ -entropy) is:

$$C(\epsilon, \tau) = - \int P_{\epsilon, \tau}(\{x(t)\}) \log P_{\epsilon, \tau}(\{x(t)\}) \quad (3)$$

where the sum is over probabilities of the trajectories represented as sequences of symbols. Each symbol is a particular spatial configuration of the system. To obtain the Kolmogorov-Sinai entropy the incremental change as a function of trajectory length is evaluated in the microscopic limit. This reflects the usual approach of considering the microscopic limit of all quantities.

This paper will discuss a system consisting of a number of particles that is described by a state space trajectory (history). The treatment is sufficiently general that it applies to any trajectory regardless of dynamical mechanism, or even whether there is a mechanism, deterministic or stochastic. The generality is due to the use only of the descriptive trajectory (kinematics) rather than predictive dynamical equations. Moreover, ultimately the treatment only assumes discrete time observations of the trajectory, so that even continuity is not necessary. Much of the paper will be devoted to a discussion of simpler cases, static particles, and a short time trajectory of particles, in order to make a connection with phase space  $(x,p)$  concepts. We find that here the natural treatment of phase space is somewhat unconventional, i.e. it is found to be  $(\sqrt{m}x, \sqrt{m}v)$ . No apologies are made for this unusual phase space: it is consistent with both classical mechanics or quantum mechanics since the use of the usual conjugate variables  $(x,p)$  is only a matter of convention and  $(\sqrt{m}x, \sqrt{m}v)$  are equally valid. Once these special cases are treated, we consider the full trajectory case. The treatment in this case is quite general and does not rely on assumptions beyond the sufficiency of a trajectory description of the system.

We note that in this paper we do not assume a thermodynamic limit; our approach is most relevant for finite systems. A stronger connection to nonlinear dynamics may be made in the future, but this article does not consider dynamic equations of motion, linear or nonlinear, deterministic or stochastic, and the related attractors, trajectory divergence, dissipation, or many other topics of nonlinear dynamics. Instead we consider only the properties of trajectories as kinematic entities where velocity and acceleration are properties of the trajectory rather than derived from forces or equations of motion.

## **2. Dimensional collapse of the description**

A key issue is relating the behavior of the parts to collective behaviors. In contrast to the --entropy-- which considers a single symbol to represent a point of state space  $\{x(t)\}$ ---we consider the internal structure of state space symbols. Each symbol can be thought of as a message formed of characters where each of the particle positions is a character. However we must consider the distance metric on the positions of particles, the indistinguishability of particles, and the particle statistics under exchange (Fermion or Boson). Another key difference between a description of a physical system and a standard message is the irrelevance of the order of the particles in the description. This irrelevance of order can be inferred from the indistinguishability of particles but does not depend on it---even distinguishable particles (e.g. different types of atoms) can be described (their locations specified) in any order.

The order irrelevance affects the properties of multiscale representations. Specifically, when two particles that are intrinsically indistinguishable and the degree of uncertainty is large enough so that their positions are indistinguishable, then they have the same macroscopic description. Intuitively, the system description can be simplified by specifying it as being composed of two particles with the same description. In relating this to information theory we would say that descriptions of particles are character order-independent messages where identical characters can be grouped together, leading to a smaller number of effective characters but with the additional need to indicate the number of characters of each type. Since each character in the message is selected from a set of possibilities that corresponds to a single dimension of the description, the merging of characters can be considered to be dimensional collapse of the description.

The implications of the irrelevance of order for indistinguishable particles can be studied by considering the case of a separable probability distribution. A separable probability is a product of single particle terms, ignoring correlations/interactions between the particles, e.g. for an ideal gas. This gives

$$P(\{x, p\}) = P(x_i)P(p_i) \quad (4)$$

This description, requires us to "by hand" enter the indistinguishability of the particles under exchanges. The integral of the information should only be done over unique states: only over a region of the phase space which is one of the  $N!$  equivalent sectors of the phase space.

Including the indistinguishability of particles explicitly can be done by symmetrizing this product as:

$$P(\{x, p\}) = \sum_Q P(x_{Q(i)})P(p_{Q(i)}) \quad (5)$$

Where  $Q$  in the sum ranges over all permutations of the indices. The domain of the position and momenta continues to be restricted to a sector of conventional  $6N$  dimensional phase space  $\{x, p\}$ . Because of the symmetry of this function, averages over phase space and the entropy integral Eq. 1 can be extended, allowing the domain of the positions and momenta to be  $3N$  real numbers. Including the  $N!$  redundant sectors of phase space, must be compensated for by dividing the integral by  $N!$ . It should be noted, that this is not the same as extending the domain of the probability and dividing the probability by  $N!$ , because of the nonlinearity of the integrand in Eq. 1. Thus, one must be careful about using  $3N$  real numbers as a representation of the system.

The separable probability distribution corresponds to classical statistics which ignores particle correlations, and is the dilute limit of either the Boson or Fermion system. In the Boson system the correct description is a wavefunction (rather than a probability) which is symmetric under exchange, which corresponds to allowing any number of particles in

the same quantum state. In the Fermion system the correct description is a wavefunction which is antisymmetric under exchange, which corresponds to allowing only zero or one particle in a quantum state. Quantum statistics do not apply to the effective states of a large scale observer, since each such state is a collection of many quantum states. In the dilute limit the particles effectively occupy states independently of each other and Quantum effects due to the overlap of probabilities can be ignored. In this limit we must still address the indistinguishability of the particles and this is our focus.

The conceptual issue involved in the treatment of indistinguishability can be understood when we consider the difference between a physical space picture of the density of a system (the one particle density), and the phase space picture. Contrast two states, the first consists of particles in a box in an essentially uniform distribution, the second consists of the same particles all found in a corner of the box, i.e. uniformly distributed in a volume one-tenth the linear dimension of the entire box. The former is characterized as an equilibrium state, the second as a non-equilibrium state. In conventional physical space the difference between these two cases seems intuitive and evident. The physical space description is the one particle density picture that was just provided (i.e. uniform distribution in the box, or in a corner of the box). In phase space, the usual textbook discussion of these two states would emphasize their similarity. They are simply two points in phase space. The issue of distinguishability combined with uncertainty however, leads to a distinct nature of these two states. In particular, the second point in phase space is much closer to a large number of mirror planes of the phase space. Thus, when we consider an increasing uncertainty / error in determination of the positions of the particles (coarse graining of the phase space) the existence of mirror planes affects the probability distribution of possible observations of the system as a function of increasing uncertainty. In particular, the amplitude of the probability decreases less rapidly for the second case because of the overlap of probabilities across the mirrors (see Appendix A). The same mirror planes exist in the first case but they are farther away. As the uncertainty increases till overlap occurs across these mirror planes, the effect of the mirrors in the first case become more similar to the effect in the second case. The higher probability in the second case is tied to a lower probability away from the phase space point, i.e. the probability of finding particles away from the populated corner is low. The difference between the two cases disappears when the scale of the uncertainty becomes larger than the box itself.

This discussion and the previous discussion about the merging of characters/particles indicates that the description of a system simplifies at characteristic scales which occur at distances to the mirror planes of the system. Our objective is to demonstrate that the scales of these characteristic internal behaviors of a system satisfy a sum rule indicating a tradeoff between fine scale behavior and large scale behavior. The process of obtaining the sum rule makes use of conventional collective coordinates (center of mass and relative coordinates) generated by a physics motivated hierarchical clustering.

### 3. Static particles/points

Many of the concepts relevant to dimensional collapse can be explained by considering just the spatial coordinates of the particles  $\{x_{i0}\}$ . Larger scale observers do not distinguish the positions of nearby particles. For such indistinguishable particles the relative position coordinates are unobservable. As  $\sigma$  increases, the number of bits of precision in the specification of each position decreases and eventually the significant bits of the two positions are the same. The resulting aggregates are assigned collective coordinates, their centers of mass. A center of mass is better defined than the original positions by a factor of the square root of the effective mass. As the scale continues to increase, the aggregates themselves aggregate, until the entire system becomes a single aggregate. The final relevant coordinate is the position of the center of mass.

Clustering is not a complete description of the process of complexity/information loss (entropy increase) implied by the formalism, but the conditions at which clustering occurs are the characteristic scales at which degrees of freedom are lost to the observer. To demonstrate the clustering we start with static particles  $\{x_i\}$  in a one dimensional box containing  $N$  particles. The probability density of the particles in state space is assumed to be written as a product of Gaussian uncertainty:

$$P(\{x_i\}) = \prod_{i=1}^N g(x_{Q(i)} - x_{i0}, \sigma)$$
$$g(x, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} \quad (6)$$

We transform to relative coordinates which characterize the distances between symmetry related points in phase space,  $\{y_i\}$ . For two particles we would have the conventional center of mass and relative coordinates:

$$y_1 = (x_2 - x_1)$$
$$y_2 = (x_2 + x_1)/2 \quad (7)$$

For three particles we would write:

$$y_1 = (x_2 - x_1)$$
$$y_2 = (x_3 - (x_2 + x_1)/2)$$
$$y_3 = (x_1 + x_2 + x_3)/3 \quad (8)$$

There is a symmetry breaking choice in deciding which two of the three particles to include in the first relative coordinate. This is set, quite naturally, by the smallest relative coordinate of the three possible pairwise differences, i.e. the first mirror plane.

In transforming to the relative coordinates  $\{y_i\}$  we must take into account the metric of

these new coordinates. Equivalently, we must recognize that the effective uncertainty in  $\{y_i\}$  is not the same as the uncertainty  $\sigma$  in  $\{x_i\}$ . This can be understood by standard treatment of uncertainty. The relative coordinate  $y_1 = (x_2 - x_1)$  has an uncertainty  $\sigma_1 = \sqrt{2}\sigma$ , while that of  $y_2 = (x_2 + x_1)/2$  is  $\sigma_2 = \sigma/\sqrt{2}$ . The effective uncertainty of three particle coordinates is:

$$\begin{aligned}\sigma_1 &= \sqrt{2}\sigma \\ \sigma_2 &= \sigma/\sqrt{2} \\ \sigma_3 &= \sigma/\sqrt{3}\end{aligned}\tag{9}$$

These corrections to the uncertainty/length can also be understood by considering the relationship between the distance between  $N$  dimensional phase space points and distance between points in physical space which is a projection of the  $N$  dimensional phase space.

For  $N$  particles the general expression of the transformation of coordinates is:

$$\begin{aligned}y_i &= a_{ij}x_j \\ x_i &= a_{ij}^{-1}y_j\end{aligned}\tag{10}$$

where the matrix  $a$  satisfies:

$$a_{ij}^{-1}a_{ik}^{-1} = \mu_j\delta_{jk}\tag{11}$$

where  $\mu_i$  is the effective mass of the  $i$ th coordinate. The probability density in terms of these coordinates is:

$$\begin{aligned}P(\{y_i, v_i\}) &= \prod_{i=1}^N g(y_i - y_{i0}, \sigma_i) \\ \sigma_i &= \sigma/\sqrt{\mu_i}\end{aligned}\tag{12}$$

Explicitly, each of the  $y_i$  except the final one can be written as differences between centers of masses of subsets of the particles (the first subset with mass  $k$  and the second with mass  $l$ ), i.e.

$$y_i = ((x_{i,1} + x_{i,2} + \dots + x_{i,k_i})/k - (x_{i,k_i+1} + x_{i,k_i+2} + \dots + x_{i,k_i+l_i})/l)\tag{13}$$

This relative coordinate has an uncertainty of



$$\sigma_i = \sigma / \sqrt{\mu_i} = \sigma \sqrt{\frac{1}{k_i} + \frac{1}{l_i}} \quad (14)$$

$$\frac{1}{\mu_i} = \frac{1}{k_i} + \frac{1}{l_i}$$

The construction of these coordinates follows an iterative procedure that is known as a hierarchical clustering algorithm.<sup>26</sup> The process of iterative clustering starts by determining the distances between pairs of points and merging the two with the smallest distance into a single “cluster.” This process is continued iteratively. Progressively larger clusters are formed from existing clusters (starting from the trivial clusters of size 1). The iterative clustering is done by merging the closest pair of clusters based on the effective distance between clusters. In the clustering literature, there are various ways of determining the distance between clusters. For this algorithm, the effective distance is the distance between the center of mass of the two clusters divided by the uncertainty of the relative coordinate (i.e.  $\sigma / \sqrt{\mu}$ ). The minimum effective distance corresponds to the coordinate that becomes the next  $y_i$ , and the distance itself is the magnitudes of the transformed coordinate  $y_{i0}$ . There are  $N - 1$  relative coordinates, and the  $N$ th coordinate is the center of mass whose effective mass is the total mass of the system. A simple example is given in Appendix B.

Thus, using the transformed coordinates constructed from a clustering algorithm we can now explain the process of dimensional collapse of the phase space. The process is based upon the restriction of the integral of the phase space to only one of the  $N!$  sectors of the phase space. By construction, each relative coordinate corresponds to an axis which intersects a mirror plane of the phase space. When the uncertainty in observation is larger than the distance to the mirror plane, then there will be overlap of the Gaussians of the state space probability density across the mirror plane.

For scales shorter than the distance to the mirror plane the Gaussians of uncertainty are distinct, while for scales larger than the distance to a mirror plane, the two Gaussians merge into one. In terms of the positions of the particles, we would say that the relative coordinate which is the distance between the particles becomes irrelevant/unobservable to the observer. Specifically for the two particle case the relative coordinate  $y_1 = (x_2 - x_1)$  becomes irrelevant for scales  $\sqrt{2}\sigma > (x_2 - x_1)$ , i.e. the observer does not distinguish the positions of the two particles for  $\sigma > (x_2 - x_1) / \sqrt{2} = \sqrt{\mu_1} y_1$ . More generally for uncertainty  $\sigma > \sqrt{\mu_i} y_i$  the  $i$ th coordinate becomes irrelevant. Thus in relationship to the uncertainty, the natural normalized coordinates are given by  $\sqrt{\mu_i} y_i$ . When mirror planes are well separated in scale, or if they are of similar scale but are orthogonal (i.e. describe well separated pairs of particles, merging two at a time) this is a reasonable description of what happens to the probability distribution for each of the relative coordinates. This is also the circumstance in which clustering is unambiguous in its meaning. When the mirror planes are not so simply organized, then the clustering becomes somewhat ambiguous.

For example, when the particles are evenly spaced and which two are to be clustered becomes ambiguous, the process of assigning collective coordinates becomes ambiguous as does the clustering algorithm. Thus the clustering formalism can be considered an approximation to a more complete multiscale formalism.

The clustering process can be generalized through the multiscale formalism to apply to arbitrary arrangements of particles. One issue is to realize that the shape of the cluster may be a relevant variable, not just the center of mass of the cluster. When the characteristic spatial extent of a cluster is larger than the scale of cluster aggregation additional parameters, e.g. moments of the cluster, should be included in the description of the cluster. This is precisely the circumstance that is known to give difficulties to conventional clustering<sup>26</sup> and is the reason that there are many different kinds of clustering algorithms. These arguments imply that the correct process of clustering must be generalized to include internal coordinates of the cluster (i.e. other descriptions of the probability distribution, by Shannon coding) when simple clustering is not adequate.

Nevertheless, the concept of clustering is helpful in indicating the mechanism by which information is lost as the level of uncertainty increases. Each position coordinate becomes less well known, and at values of uncertainty corresponding to the merging of clusters, the number of coordinates decreases. Because of the role of uncertainty averaging, the effective mass affects the scale of uncertainty because collective coordinates are in general better defined (by a factor of the square root of the effective mass) than the coordinates of individual particles. Equivalently, the scale of collective coordinates is larger than the scale of individual particles by a factor of the square root of the effective mass.

The largest scale coordinate that describes a system of  $N$  particles is its center of mass. The uncertainty in the center of mass coordinate is  $\sigma_N = \sigma/\sqrt{N}$ . The analysis of relevance can be formally extended to include the center of mass coordinate by indicating that it becomes irrelevant for scales  $\sigma > \sqrt{N}y_N = \sqrt{N}x_{cm}$ . This suggests that the center of mass coordinate becomes irrelevant when it is indistinguishable from the origin of the coordinate system. This interpretation must be considered with some care because certain choices of the origin of the coordinate system would result in this scale being smaller than the scale of relative coordinates. Also, we would then expect that at finer scales parts of the system would "aggregate" to the origin of coordinates and become prematurely irrelevant. We can solve this problem by requiring that the origin of coordinates will be located outside of the original system. This would be consistent with an interpretation that the center of mass becomes irrelevant when the system aggregates to another system of higher mass.

#### **4.Short time trajectories / Kinetic Points**

To treat kinematically described systems we will generalize the consideration of positions of particles  $\{x_{i0}\}$  to the conventional  $\{x,p\}$  phase space and then to trajectories. The phase space formalism which we use for this purpose is not quite the standard one. It

might seem natural to treat kinetic points by considering the standard phase space collective position and momentum coordinates, i.e. the center of mass and total momentum. However, we find that the more natural phase space coordinates are not position and momentum  $(r, p)$ , but are instead the position and velocity  $(r, v)$ . The reason has to do with the importance of the mass in the uncertainty of position of an aggregate. The position and the velocity uncertainties are both affected by the mass. The metric on the space  $d((r_1, v_1, m_1), (r_2, v_2, m_2))$  is controlled both by the positions and by the masses of the two aggregates whose relative distance is being evaluated. Scaled distances include the uncertainty that arises from clustering in the form  $(\sqrt{\mu}r, \sqrt{\mu}v)$  where  $\mu$  is the effective mass of the relative coordinate (i.e. this is the scale at which the observer can see a particular coordinate). Since  $\mu$  is proportional to the conventional inertial mass, the distance metric of  $(r, v)$  consists of conjugate coordinates. The product of the uncertainty  $\sqrt{\mu}r\sqrt{\mu}v = rp$  has the appropriate dimensions of state space volume.

The use of velocity rather than momentum coordinates is also conceptually simpler for the purpose of understanding the phase space as a short time approximation to system trajectories / movies. Momentum is an inherently dynamic quantity, i.e. it is related to the application of force and Newton's third law. Inertial mass is defined and measured in this context. Velocity is a kinematic quantity which is directly related to a trajectory. Mass, as defined in this context, is the number of particles in an aggregate. We postpone a more complete discussion of the relationship between trajectories and phase space. However, it is convenient for dimensional arguments to realize that velocity is a surrogate variable for the difference in position of two observations  $(x_i - x_j)$ . Velocity is obtained from this by a-posteriori division by a time interval. Thus the fundamental phase space we are considering is  $((x(t_2) + x(t_1))/2, (x(t_2) - x(t_1))) = (x, v)$  for two observations of position  $x(t_1)$  and  $x(t_2)$  considered to take place at an interval of time  $dt$ . In what follows we should consider  $v$  to actually be  $vdt = (x(t_2) - x(t_1))$  and have the dimension of length. We note that the position coordinate and the velocity coordinate are related to each other as a center of mass and a relative coordinate, thus there is an additional factor of 2 in their relative uncertainty (see above eq. (9)) that can be ignored along with  $dt$  for our present purposes.

The main formal advantage of this remapping of the phase space is that the shape of domains of uncertainty in phase space becomes clearly defined for observers with a particular  $t$ . Quantum mechanics only provides a minimal volume of phase space uncertainty but does not specify the way to determine its shape. The minimum number of states which is observable (i.e. one) is given by the uncertainty relationship  $x p > h$ . When considering an informational description of the system, the product of space and momentum coordinate differences  $x p$  can be understood as a minimum number of possibilities. This assumes that the set of possibilities of  $x$  and those of  $p$  are independent, and the total number of possibilities is given by the product of the two quantities. Thus a region of phase space given by  $x p$  has a certain number of distinguishable possibilities given by  $x p/h$ . However, there is no sense in which a

square region is special, this includes the ambiguity of the correct version of units for  $x$  and  $p$ , i.e. they are not dimensionally equivalent and so the absolute measure of units that can relate  $x$  and  $p$  to each other is not well defined.

Our task is related to the problem of understanding how two distinct regions of phase space, one at  $(x_1, p_1)$  and the other at  $(x_2, p_2)$  (both of size  $x p > h$ ) become indistinguishable from each other for larger scale observers. The region of phase space that encompasses them could be thought to have a volume given by  $(x_2 - x_1)(p_2 - p_1)$ , but this value entirely depends on the shape of the region that becomes indistinguishable and includes  $(x_1, p_1)$  and  $(x_2, p_2)$  (i.e. it assumes a minimal area rectangular region). By using the form of phase space given above  $(\sqrt{\mu}r, \sqrt{\mu}v)$  we have resolved this issue and allowed distances in phase space to be measured using the Euclidean metric. In these units the observer uncertainty grows spherically around points in phase space. Note that this does not imply anything about the underlying quantum uncertainty of the system, since this relates to observer rather than intrinsic uncertainty. The distance between two points in phase space is measured by the norm of this vector, whose square can be seen to be the sum of the moment of inertia and the kinetic energy (multiplied by  $dt^2$ ). Using this definition of phase space we can now proceed in a uniquely defined way with the process of aggregation of particle coordinates, i.e. clustering.

The generalization of the treatment of static points to short time trajectories makes use of our discussion of the effective mass to construct normalized coordinates. The separable probability density is:

$$P(\{x_i, v_i\}) = \prod_{i=1}^N g(x_i - x_{i0}) \prod_{i=1}^N g(v_i - v_{i0}) \quad (15)$$

Using this phase space representation for a collection of points  $\{x_i, v_i\}$ , the previous treatment of clustering can be applied. The natural coordinates to use are the rescaled coordinates given by  $\{\sqrt{\mu}x_i, \sqrt{\mu}v_i\}$  which are also conjugate coordinates. The process of clustering generates the collective coordinates of the system as relative positions and velocities using a vector form of Eq. (6):

$$(y_i, w_i) = ((x_{i,1} + x_{i,2} + \dots + x_{i,k_i})/k - (x_{i,k_i+1} + x_{i,k_i+2} + \dots + x_{i,k_i+l_i})/l), \\ (v_{i,1} + v_{i,2} + \dots + v_{i,k_i})/k - (v_{i,k_i+1} + v_{i,k_i+2} + \dots + v_{i,k_i+l_i})/l) \quad (16)$$

Where  $y$  and  $w$  are the relative coordinates. The objective is to identify the distances to mirror planes for the particles so clustering can be done as in the previous case. The order of clustering the points in phase space is given by the distance between cluster coordinates scaled by the square root of the effective mass,  $\sqrt{\mu} \sqrt{y^2 + w^2}$ . Each of the selected coordinates is the iteratively obtained smallest distance according to this metric.

The scale at which each of these coordinates becomes irrelevant (i.e. the scale of the

mirror planes) are given by

$$\sigma_i = \sqrt{\mu_i} \sqrt{y_{i0}^2 + w_{i0}^2} \quad (17)$$

The sum rule is now proven by adding the squares of the scales of each of the collective degrees of freedom as:

$$\frac{1}{2} \sum_i \sigma_i^2 = I + T \quad (18)$$

where

$$\begin{aligned} I &= \frac{1}{2} \sum_i \mu_i y_{i0}^2 = \frac{1}{2} \sum_i x_{i0}^2 = \frac{1}{2} \sum_i m_i x_{i0}^2 \\ T &= \frac{1}{2} \sum_i \mu_i w_{i0}^2 = \frac{1}{2} \sum_i v_{i0}^2 = \frac{1}{2} \sum_i m_i v_{i0}^2 \end{aligned} \quad (19)$$

are the moment of inertia and the kinetic energy respectively (the mass of individual particles are assumed to be chosen by the units to be one). If the center of mass coordinate of the system is not included in this calculation, the moment of inertia is relative to the center of mass coordinate and the kinetic energy is the internal kinetic energy of the system (including all the relative motion of particles, i.e. including their thermal motion). If the center of mass coordinate is included, then both are relative to the origin of coordinates. While the discussion has been for single spatial and velocity dimensions, the generalization to 3-spatial and velocity dimensions is straightforward.

Interpreting this sum rule, we note that given two systems which have the same kinetic energy and moment of inertia, but a different distribution of velocities and positions of the constituent particles, there is a tradeoff possible between position and velocity coherence and independence of motion of the particles. When particles that are near to each other are also moving in the same direction in patches, then the particles within a patch aggregate together at very fine scales. The patches themselves become identified by the larger scale coordinates, which are few in number. When particles that are near to each other have widely distributed velocities, i.e. they are independent, then the collective coordinates have a distribution of sizes which is more narrow. Thus we see that the sum rule implies a tradeoff between coherence and independence of motion

We can more specifically consider a non-equilibrium fluid where the velocities of particles are correlated rather than the positions which can be uniform, e.g. in an incompressible fluid. Such systems often are considered in discussions of non-equilibrium systems today. The clustering of particles in the phase space would proceed by clustering together first the particles that are adjacent in position and are moving with similar velocities. Since such particles are close in phase space, this would result in many finer scale degrees of freedom than if the velocities were randomly distributed. Such clustering would result in

aggregates representing collective motion. This is true even if the particles are moving partially or mostly at random but with a bias in motion that corresponds to a collective flow. Thus, in the case of correlated motions, the existence of a collective non-zero velocity would survive to larger scales compared to the case of random velocity distributions. At some scale, however, the local aggregates representing correlated motions would cluster, eliminating the effective velocity at a scale that would reflect the scale of the size of local flows as well as the magnitude of local velocities. A simple example is given in Appendix C.

For two incompressible fluids occupying equivalent volumes with the same kinetic energy, but with different flows, this sum rule applies and can be used to compare the decomposition of their scales of behavior. Because of the incompressible nature of the fluid, the sum rule can be thought of as the scale based decomposition of the kinetic energy. The class of systems that can be compared by the sum rule can be expanded by applying it to longer trajectories.

## 5. Trajectories

Implications of the sum rule can be further understood by considering complete trajectories. Consider trajectories of two systems of  $N$  particles in a box. The first system is of particles that correspond to an equilibrium gas. The second is of particles which, at any particular time, are found in a small volume, but over time visit all parts of the box. The trajectory is described by a sequence of spatial coordinates  $\{x_i(t_j)\}$ , where  $i$  is the particle index and  $j$  is the time index. The metric on the space-time for clustering should be taken to be the usual Euclidean metric. For sufficiently frequent time sampling, the individual particle trajectories are apparent through the aggregation of positions at adjacent times. As discussed at the beginning of section 4 the relative coordinate of a particle at different times is equivalent to its velocity multiplied by the time interval. Such details are, however, not necessary for the general application of the sum rule. The key to applying the sum rule in this context is recognizing that the sum rule applies to the entire trajectory rather than to any instant of it. The sum of the coordinates squared analogous to Eq. (18) is (the speed of light is inserted here for dimensional reasons):

$$\sum_{i,j} (x_i(t_j)^2 + c^2 t_j^2) \quad (20)$$

which is the same in the two cases assuming multiple visits to each part of the physical space by the localized particles in the second system. Note that this sum is no longer equal to the instantaneous sum of the kinetic energy and moment of inertia. Instead it is equal to the effective moment of inertia in  $d+1$  dimensions around a single origin over the entire trajectory. According to the sum rule these two systems have the same sum over the square of their scales of aggregate degrees of freedom.

It is helpful to note that in this section, the discussion of trajectories has made no direct

assumptions of the mechanism by which particles are localized in a certain area of physical space, e.g. by external forces or by internal interactions. Thus it is applicable quite generally to systems that are inhomogeneous in space. Much of the literature about non-equilibrium systems is devoted to fluids whose spatial distribution of velocities changes significantly from the equilibrium to a non-equilibrium state, but whose density distribution does not, i.e. an incompressible or weakly compressible fluid. Aggregation of coordinates, and the sum rule implicit in Equation 20 can also be applied in this case.

## 6. Multistate particles / Multicomponent systems

We can extend this treatment to the case of distinguishable particles by considering particles with internal states / structure determined by a (typically discrete) variable  $s$ , which has an associated metric  $d(s,s')$ . This metric determines the distinguishability of the different internal states of the system, i.e. the aggregation of particles in the case where we include both the spatio-temporal distance and the internal state distance. For example, if particles have different spins, then the aggregation of particles with the same spin state is different than the aggregation of particles of different states. There are two interesting limits that compare the intrastate distance with the spatial distance. In the first, whether the spin is different or not does not affect the distinguishability or the aggregation. This leads to a local averaging of spin values during aggregation. In the opposite limit, distinct spins are highly distinguishable, and the two types aggregate separately before collapsing together only at a very large scale. This limit might be more naturally applied to mixed gasses/fluids where components occupy the same volume but are readily distinguishable.

The sum rule in this case follows from a generalized version of eq. 20, i. e. ( $\bar{s}$  is a reference state value),

$$\sum_{i,j} (x_i(t_j)^2 + d(s_i(t_j), \bar{s})^2 + c^2 t_j^2) \quad (21)$$

and aggregation of collective coordinates proceeds by clustering as before. There are many interesting questions that can be addressed using this formalism including the process of aggregation to components, substructure and hierarchy. Specific examples of dynamical and spatio-temporal systems will be discussed in a separate publication.

The equivalence of the sums in eq. 21 for two systems allows us to apply the sum rule to compare the decomposition of the degrees of freedom in systems with trajectories in the same space. With allowances, the discussion of randomly moving particles and coherently moving particles, as well as the more general scale dependent decomposition, can be directly applied to considering other objects that are similar enough to be indistinguishable on a large scale. Thus, even the random or coherent physical movement of people may be considered using this formalism.

## 7. Summary

This paper considers the problem of characterizing the multiscale behavior of non-equilibrium systems. The purpose is to relate descriptions of a system as a function of the scale of observation. We report a sum rule which provides an opportunity to simplify the discussion of systems by classifying them according to the value of the sum. Within a single class, the members can be compared through differences in the scale based decomposition of their behaviors.

In particular, the sum rule relates systems that can exist in the same space-time domain. Within this class of systems the sum rule constrains the scale dependent descriptions so as to ensure a trade-off between the existence of fine scale and large scale behaviors. The tradeoff between scales of behavior characterizes the relationship between dynamic and coherent structures at different scales in non-equilibrium systems. Behaviors that are observable at larger scales reflect the existence of coherent or correlated positions and motions of particles. This implies a reduced randomness / independence of the motion of the particles and thus a reduction in the scale of the finer degrees of freedom. This characterization is more informative than the general concept of a non-equilibrium system as relaxing to an equilibrium system by randomization of the positions or motions of particles, or being formed through the imposition of correlations of motion or position on the particles of an equilibrium system.

## Appendix A: One mirror

Considering a one-dimensional Gaussian at a mirror is a helpful step to considering the multiscale description of a non-equilibrium system. Non-equilibrium systems typically correspond to a state near to a state space boundary. The boundary is composed of mirror planes of symmetries of exchange of the particles that are closer together than they would be in an equilibrium system.

Consider a Gaussian located at  $x = a$  next to a mirror at  $x = 0$ . The domain of  $x$  is always  $x > 0$  The probability is

$$P(x, a, \sigma) = f(x - a, \sigma) + f(x + a, \sigma) \quad (\text{A.1})$$

where

$$f(x, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} \quad (\text{A.2})$$

The entropy as a function of scale is given by an integral:



$$S(a, \sigma) = - \int_0^{\infty} dx P(x, a, \sigma) \log_2(P(x, a, \sigma)) \quad (\text{A.3})$$

Note that we do not have quantum normalization so the entropy can be negative for sufficiently small values of  $\sigma$ .

For a Gaussian located far from the mirror we have the entropy:

$$S(\infty, \sigma) = \frac{1 + \ln[2\pi]}{\ln[4]} + \log_2[\sigma] \approx 2.047 + \log_2[\sigma] \quad (\text{A.4})$$

For a Gaussian located at the mirror we have the probability and entropy:

$$\begin{aligned} P(x, 0, \sigma) &= 2f(x, \sigma) \\ S(0, \sigma) &= S(\infty, \sigma) - 1 \end{aligned} \quad (\text{A.5})$$

Fig. A1 shows the probability distributions for three cases,  $P(x, 0, \sigma)$ ,  $P(x, 1, \sigma)$ ,  $P(x, 5, \sigma)$  with  $\sigma = 1$ . Fig A2 shows  $S(0, \sigma)$ ,  $S(\infty, \sigma)$  and  $S(1, \sigma)$  as a function of  $\sigma$  showing how the entropy increases around  $\sigma = 1$  for the latter case.

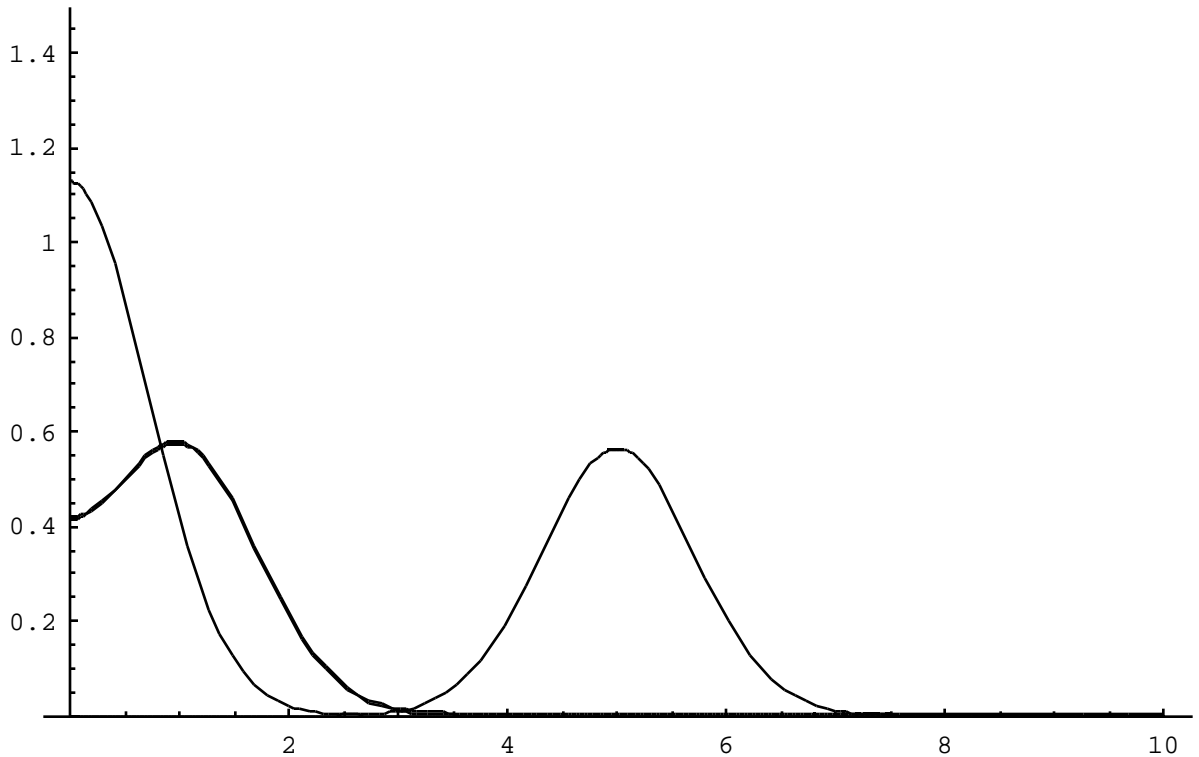


Figure A1

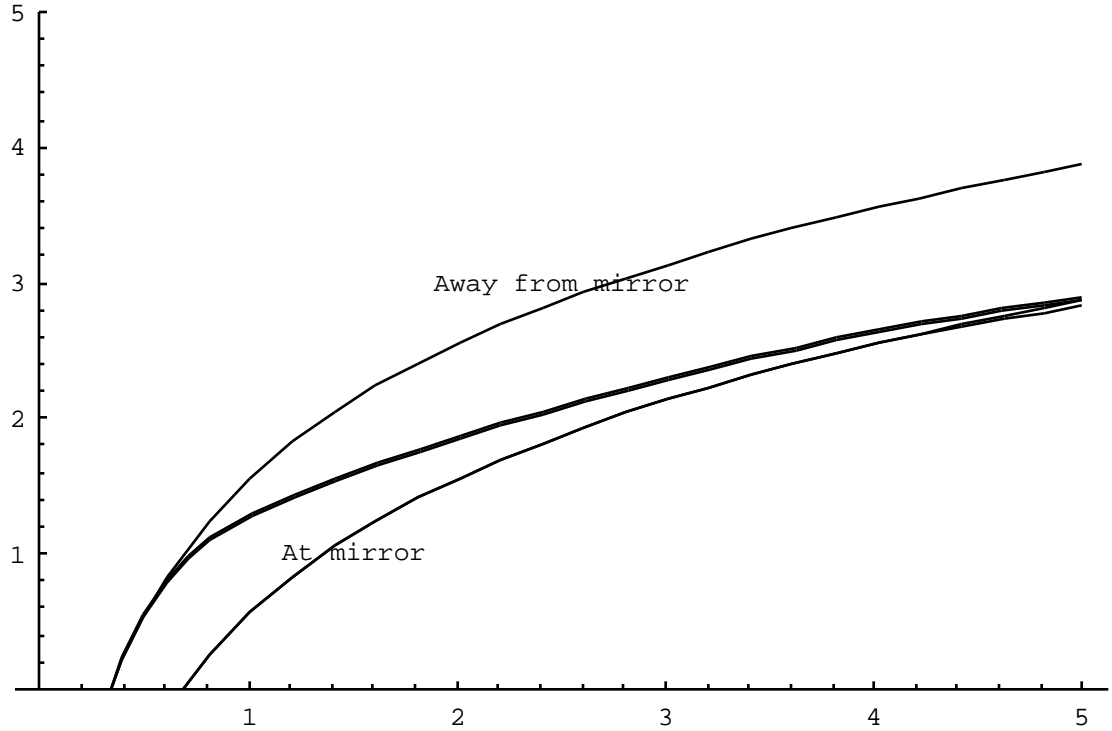


Figure A2

### Appendix B: Example of Sum Rule: Static points.

A first example of the sum rule can be constructed by considering a distribution of static particles/points in the interval  $[0,1]$ . We compare three cases, each with  $n=16$  particles: (a) regularly spaced at intervals of  $1/16$  and centered on the value  $1/2$ , (b) randomly placed and; (c) two clusters symmetrically placed around  $1/2$ . These are illustrated in figure B1 as top, middle and bottom respectively.

In order to apply the sum rule, we must arrange the points so that the moment of inertia of these three sets of points are the same. Using the regularly spaced points as a reference. we can calculate the moment of inertia to be:

$$I = n/3 - 1/12n = 5.328 \tag{B.1}$$

For (b) we generate random points a few times until their moment of inertia approximates this value within 0.02. This does, however, imply that the points are not entirely randomly distributed.

For (c) we position the two clusters at positions:

$$\begin{aligned} x_1 &= 1/4 - w \\ x_2 &= 1/4 + w \end{aligned} \tag{B.2}$$

where

$$w = \frac{-3n + 2\sqrt{3}\sqrt{n^2 - 1}}{12n} = 0.0381 \quad (\text{B.3})$$

The internal and center of mass coordinates can be calculated for (a) and (c) directly and for (b) numerically using a hierarchical clustering algorithm. The results are shown in figure B2. This figure is essentially a histogram, but it indicates only the values at which scales are located, and not all of the zero values. The following three paragraphs discuss each of the three cases, leaving the largest scale for the subsequent paragraph.

For case (a), shown as filled triangles, the pairwise grouping of points gives a smallest scale of  $1/(16\sqrt{2})$ , where the factor of  $\sqrt{2}$  arises from inclusion of the effective mass. This scale has a multiplicity of 8. Progressively higher scales are multiplied by a factor of  $2\sqrt{2}$ , and their multiplicity is halved.

For case (b), shown as filled diamonds, the clustering results are shown, but, consistent with plotting the values as a histogram, all the values below  $1/16$  and those between  $1/16$  and  $1/8$  are combined, each to be plotted as one value with their respective multiplicities. For higher values, the scales are sufficiently separated that they appear as individual points with multiplicity one. Note that one point, at a scale of one, overlaps with that of case (a), the near exact equivalence is a coincidence.

For case (c) there are only two nonzero scales, corresponding to the center of mass discussed below, and the relative coordinate of the two clusters which is given by

$$(\sqrt{n}/2)(.5 + 2w) = 1.152 \quad (\text{B.4})$$

where the first factor is due to the effective mass, and the second is the relative location of the two clusters. All the other degrees of freedom (of which there are 14) are zero.

All three cases have the largest scale as the center of mass coordinate with a value corresponding to the square root of the effective mass ( $n$ ) times the center of mass coordinate. For cases (a) and (c) the center of mass coordinate is exactly 0.5 and the largest scale is 2. For case (b) the center of mass coordinate is not exactly 0.5 and the largest scale differs somewhat from 2.

Comparing the three cases, we see that case (c), the case of two clusters, when compared with the other two cases, illustrates the tradeoff of large scale and fine scale degrees of freedom. The largest relative degree of freedom for the case of two clusters is larger than that for the other two cases (by 15% in this example). This comes at the expense of the disappearance of all the finer scale degrees of freedom. The random and regularly spaced cases are comparatively similar, but can be seen to display more subtle tradeoffs.

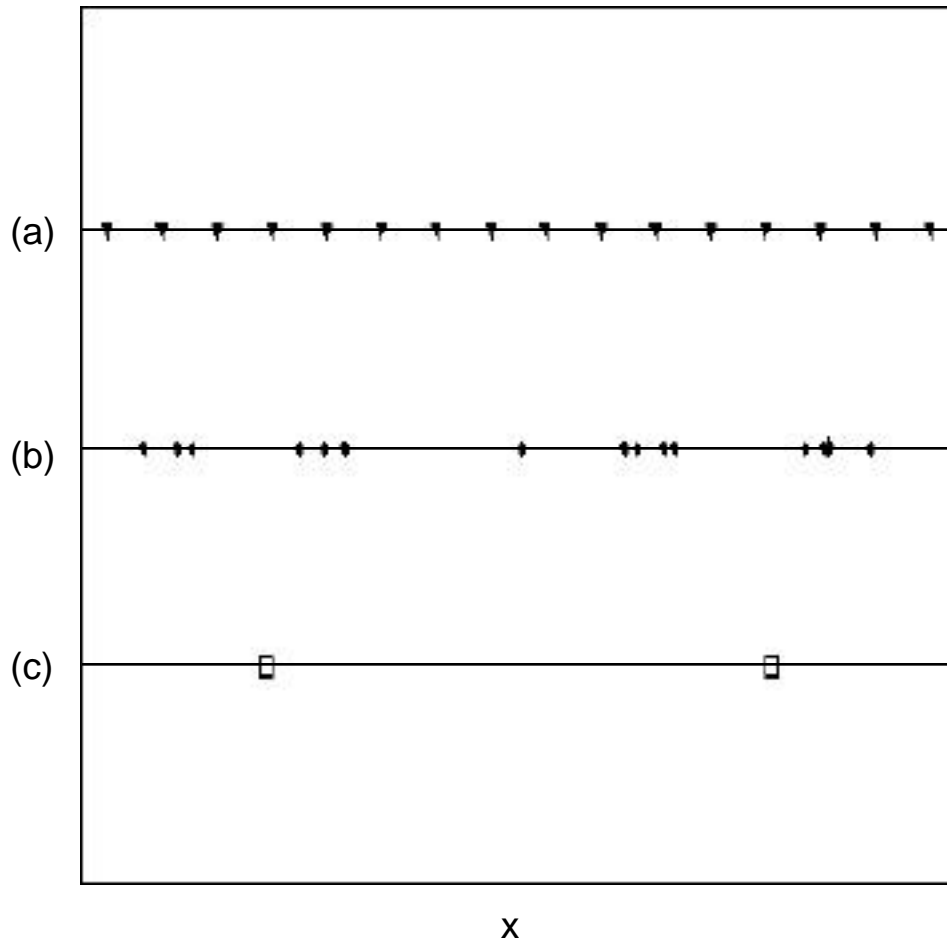


Figure B1: The three scenarios of 16 points in the interval  $[0,1]$  as discussed in the text. The horizontal axis is the interval  $[0,1]$ , the three cases are (a-triangles) evenly distributed, (b-diamonds) randomly distributed subject to selection for having the same squared sum as (a), and (c-rectangles) 8 points are located at two places symmetrically with respect to  $1/2$ .

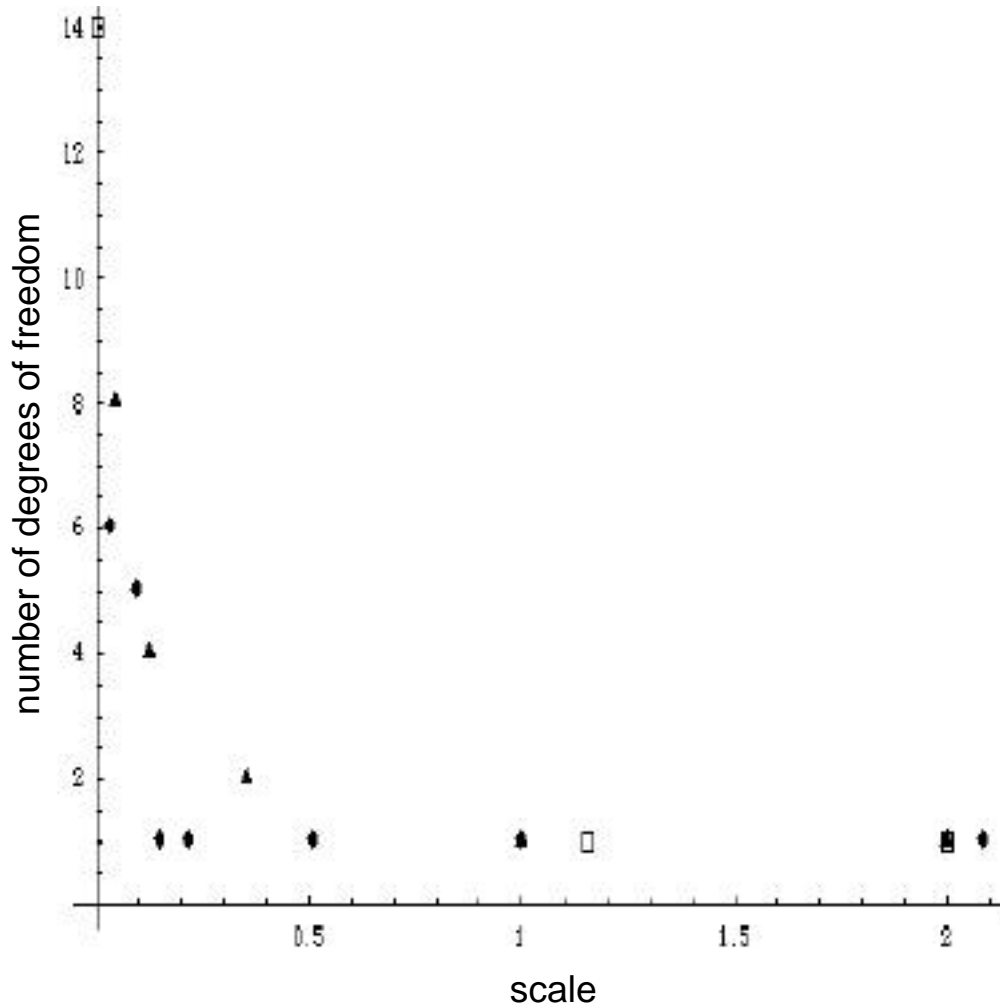


Figure B2: A histogram of the number of degrees of freedom as a function of scale for the three cases illustrated in Figure B1, as indicated by the three symbol types.

### Appendix C: Example of Sum Rule: Dynamic points.

A second example of the sum rule can be constructed by considering points that have positions uniformly distributed in the interval  $[0,1]$  chosen to lie upon a lattice to suggest an incompressible fluid, and different velocity distributions. The first velocity distribution corresponds to a Gaussian distribution with mean 0 and standard deviation 1, corresponding to a toy model of an equilibrium fluid. The second has the velocity distribution:

$$v(x) = \begin{cases} 1 & x > 1/2 \\ -1 & x < 1/2 \end{cases} \quad (\text{C.1})$$

corresponding to a toy model of a non-equilibrium macroscopic flow in a fluid. The latter

case is definite while the former is stochastic and cases for the former were generated till the sum in eq. 18 was the same within one percent (about 20 cases were generated in total), this case was used for results. The two systems are shown in Fig. C1 and the histogram illustrating the tradeoff of larger and finer scale degrees of freedom of their scales is shown in Fig. C2. We see from the figure that the collective flow produces the largest scale behavior and, as required by the sum rule, the finer scale behaviors are smaller than in the equilibrium fluid case.

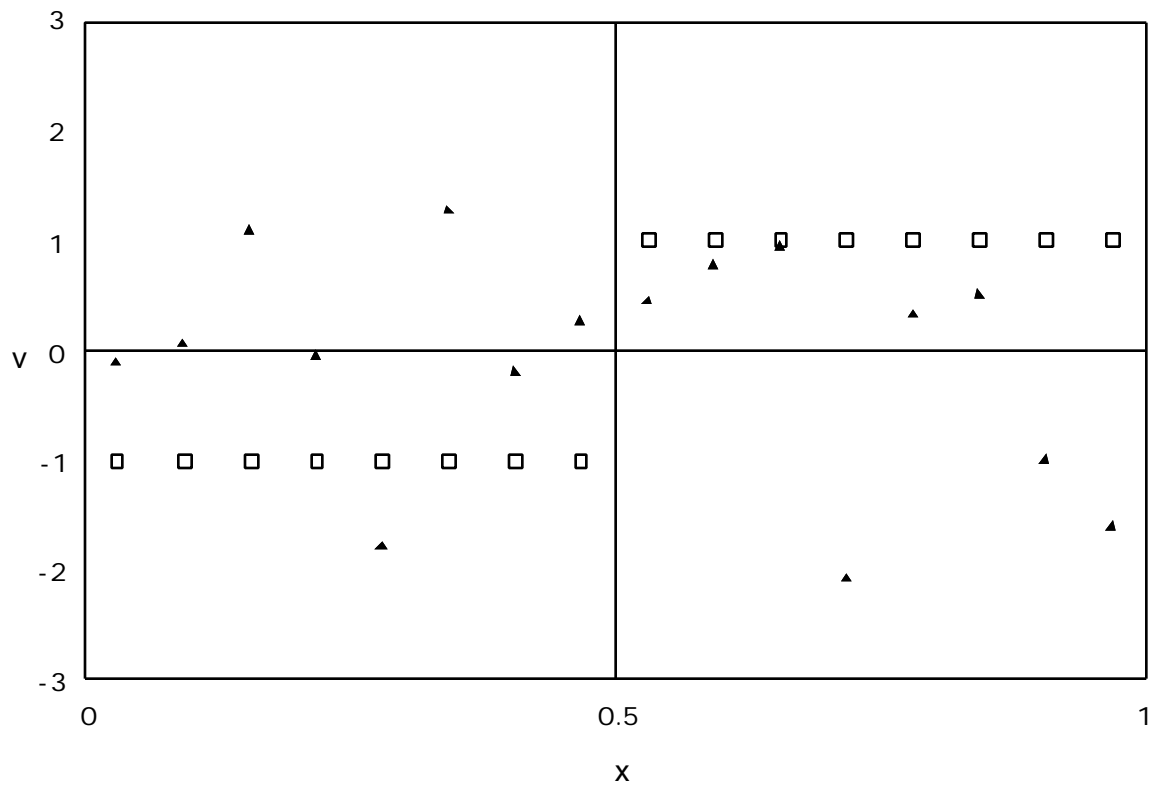


Figure C1: Two scenarios of 16 points in the interval  $[0,1]$  as discussed in the text. The horizontal axis is the interval  $[0,1]$ , the two cases are (a-triangles) uniformly distributed positions with Gaussian random velocities subject to selection for having the same squared sum of scales as (b), (b-rectangles) uniformly distributed positions with velocities  $+1$  for  $x > 0,5$  and  $-1$  for  $x < 0.5$ .

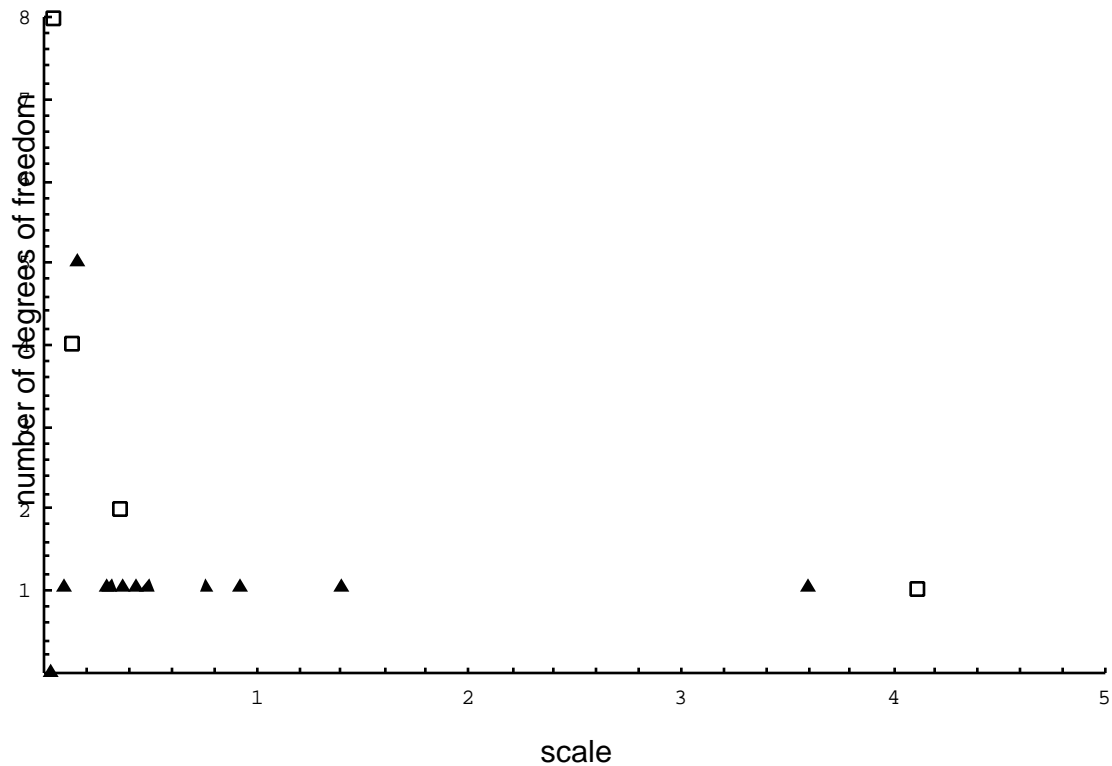


Figure C2: A histogram of the number of degrees of freedom as a function of scale for the two cases illustrated in Figure C1, as indicated by the two symbol types.

Acknowledgments: I would like to thank Michel Baranger, Irving Epstein, Mehran Kardar, David Meyer, Larry Rudolph and Hiroki Sayama for helpful discussions.

References:

1. C. E. Shannon, A Mathematical Theory of Communication, in Bell Systems Technical Journal, July and October 1948; reprinted in C. E. Shannon and W. Weaver, The Mathematical Theory of Communication (Univ. of Illinois Press, Urbana, 1963).
2. Selected Works of A.N. Kolmogorov, Volume III: Information Theory and the Theory of Algorithms (Mathematics and Its Applications), A.N. Shiryayev, ed. (Kluwer, Dordrecht, 1987)
3. G.J. Chaitin, Information, Randomness & Incompleteness, 2nd edition, World Scientific, 1990; Algorithmic Information Theory, Cambridge University Press, 1987.
4. A. N. Kolmogorov, Dokl. Akad. Nauk SSSR 119, 861(1958); 124, 754 (1959), Ya. G. Sinai, Dokl. Akad. Nauk SSSR 124, 768 (1959)
5. V. Latora and M. Baranger, Phys. Rev. Lett. 82(1999)520.

7. C. H. Bennett, *Int. J. of Theoretical Physics* 21, 905 (1982)
8. *Complexity, Entropy and the Physics of Information*, W. H. Zurek, ed., (Addison-Wesley, 1990)
9. *Proceedings of the fourth Workshop on Physics and Computation, PhysComp96*, T. Toffoli, M. Biafore, and J. Leão (eds.), New England Complex Systems Institute (1996)
10. M. Gell-Mann, *Complexity*, 1, 16-19 (1995); M. Gell-Mann and S. Lloyd, *Complexity*, 2, 44-52 (1996); S. Lloyd and H. Pagels, *Annals of Physics* 188 (1988) 186
11. H. Haken: *Synergetics*, (Springer-Verlag, 1978)
12. G. Nicolis and I. Prigogine: *Self-Organization in Nonequilibrium System : From Dissipative Structures to Order Through Fluctuations*, (Wiley, 1977)
13. M. C. Mackey, *Time's Arrow: The Origins of Thermodynamic Behavior*, (Springer-Verlag, 1992)
14. E. Ott, *Chaos in Dynamical Systems* (Cambridge University Press, Cambridge, 1993).
15. S. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*, (Addison-Wesley, 1994)
16. D. K. Kondepudi, I. Prigogine *Modern Thermodynamics: From Heat Engines to Dissipative Structures* (John Wiley & Son, 1998)
17. H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, 1999)
18. K. Kaneko, I. Tsuda, *Complex Systems: Chaos and Beyond, A Constructive Approach with Applications in Life Sciences*, (Springer-Verlag, 2000)
19. A. Winfree, *The Geometry of Biological Time*, (Springer-Verlag, 2001)
20. B. Ermentrout, *Simulating, Analyzing, and Animating Dynamical Systems: A Guide to Xppaut for Researchers and Students (Software, Environments, Tools)* (SIAM, 2002)
21. A. Pikovsky, M. Rosenblum, J. Kurths, *Synchronization: A Universal Concept in Nonlinear Science*, (Cambridge University Press, 2002)
22. Y. Bar-Yam, *Dynamics of Complex Systems* (Perseus, 1997), see especially Chapters 8 and 9.
23. P.-M. Binder and J. A. Plazas, *Phys. Rev. E* 63 (2001)



24. P. Gaspard and X.-J. Wang, Phys. Rep. 235 (1993) 291.

25 E. Aurell, G. Bofetta, A. Crisanti, G. Paladin, and A. Vulpiani, Phys. Rev. Lett. 77, (1996); M. Abel, L. Biferale, M. Cencini, M. Falcioni, D. Vergni, and A. Vulpiani, Physica D 147 12 (2000)

26. see e.g. B. Mirkin, Mathematical Classification and Clustering, (Kluwer, 1996)