

## MULTISCALE COMPLEXITY/ENTROPY

Y. BAR-YAM

*New England Complex Systems Institute  
24 Mt. Auburn St., Cambridge, MA, 02138, USA  
yaneer@necsi.org*

Received 27 March 2002

Revised 22 August 2003

We discuss the role of scale dependence of entropy/complexity and its relationship to component interdependence. The complexity as a function of scale of observation is expressed in terms of subsystem entropies for a system having a description in terms of variables that have the same *a priori* scale. The sum of the complexity over all scales is the same for any system with the same number of underlying degrees of freedom (variables), even though the complexity at specific scales differs due to the organization/interdependence of these degrees of freedom. This reflects a tradeoff of complexity at different scales of observation. Calculation of this complexity for a simple frustrated system reveals that it is possible for the complexity to be negative. This is consistent with the possibility that observations of a system that include some errors may actually cause, on average, negative knowledge, i.e. incorrect expectations.

*Keywords:* Complexity; entropy; information; variety; scale; negative information.

### 1. Introduction

The study of the relationship of information and physics has a long history. Maxwell's demon has provided much of the impetus, as has the understanding of computation as a physical process [1–9]. More recently, this interest has focused on quantum information and computation. In this paper we will consider the role of the interdependence of the components of a physical system and the relationship of interdependence to microscopic and macroscopic degrees of freedom. Missing from the conventional discussions of information in physics is the recognition that information describing the macroscopic structure and dynamics of non-equilibrium systems, must be related to changes in microscopic entropy.

Not only are non-equilibrium states related to a reduction in microscopic entropy (which follows directly from the Second Law of Thermodynamics) but, more significantly, the degree to which macroscopic behavior occurs must, in some sense, be directly related to the degree of reduction in the entropy. This statement can also be understood to follow from the discussion in Ref. 10 that macroscopically observable degrees of freedom represent aggregates of many microscopic variables. Macroscopically relevant aggregates arise when their internal/relative degrees of

freedom are constrained, creating coherent behaviors that are observable on a large scale. To develop a formalism that can address this topic we go beyond the understanding of pair correlations in physical systems, the usual tool of statistical physics, to consider higher order correlations that are characteristic of coherent behaviors: many particles that are coupled as described by  $k$ -fold correlations, where  $k$  may be a macroscopic number. The use of microscopic correlations to obtain the macroscopic entropy for an equilibrium system was shown by Lindgren [11]. Our objective is to relate the structure of the state space of the system (i.e. measures of the volume of the state space, the entropy, as well as its geometrical structure) to the dependencies between microscopic coordinates. These measures of state space structure are relevant for both equilibrium and non-equilibrium systems. The structure of the state space can be characterized in part by considering its volume at different degrees of resolution (scale). The large-scale structure of the state space is related to macroscopic degrees of freedom (especially for non-equilibrium systems), while the fine scale structure is related to microscopic degrees of freedom.

The logic of this presentation is as follows. We begin by considering the conventional definition of entropy/information as a measure of the state space volume of a system. Then we consider the dependence of this quantity on the scale of observation. Larger scales correspond to reduced precision in observation, i.e. observation only of collective behaviors of multiple components that carry redundant information. Using a detailed consideration of the algebraic properties of the information for multiple variables we provide explicit expressions for the amount of information at each particular level of redundancy (scale). We compare this scale dependent function in its entirety for different systems. This function, “the complexity profile” [10] can be viewed as the scale dependent complexity, synonymous with the scale dependent information, entropy or variety. We arrive at the conclusion that this function can characterize the structure of a system, and in particular it demonstrates that the existence of larger scale parameters (degrees of freedom) directly corresponds to a loss of microscopic degrees of freedom through aggregation. This is shown by deriving a sum rule that constrains the sum over complexity (amount of information) in all degrees of freedom of the system. This fundamental, if simple, observation of a tradeoff between fine and large-scale degrees of freedom has not yet been made in the literature considering the role of information in physics.

In the process of this discussion we develop a formalism based upon information theory that can treat all high order correlations of the components of a system. Indeed, capturing these correlations is what enables us to describe the tradeoff between fine and large-scale degrees of freedom through a sum rule. This sum rule formalizes the tradeoff in state space at different scales. The formalism and the sum rule together constitute a new treatment of the relationship between interdependence of components and measures of state space. We illustrate the formalism and its implications by applying it to several simple models. Detailed applications to models of specific complex systems are deferred to future publications.

The discussion of information in physical systems in this paper can be conceptually linked to many quantitatively oriented discussions of complexity, since they are

often related to information theory. The word “complexity” is used in the literature in two ways. The first is as a quality whose attributes (to be identified) characterize all complex systems. The second is as a quantitative measure, a single number that characterizes a system, thus capturing the notion that some systems (mathematical or physical) are more or less complex than others. The various quantitative approaches that have been taken often relate complexity to mathematical formulations of thermodynamic or statistical entropy and information [12–33]. The notion of descriptive complexity, one of the threads in the discussion of complexity suggests that the complexity of a system is related to the length of its shortest description. Since the length of a description is in some sense the minimal information, the basis of most discussions of descriptive complexity is information theory [13], augmented by algorithmic complexity [16–18], which considers the possibility of using a shorter description and an algorithm for expanding it. The use of algorithmic complexity indicates the importance of computation in the study of the properties of complex systems. Excluding this algorithmic aspect of most discussions, the approach of information theory would be based upon a discussion of state space volume. In considering the behavior of the state space volume at different scales of observation, and the behavior of microscopic and macroscopic degrees of freedom, we are therefore considering how information behaves as a function of scale. Missing from these other treatments is a discussion of the interdependence of components and specifically any formalism that can consider high order correlations between components of the system. As shown below, without a treatment of these high order correlations, an understanding of the relationship of microscopic and macroscopic (or intermediate scale) observations of a system is not possible. Thus in this paper, rather than approach the problem from the point of view of algorithmic complexity, we directly consider how information in physical observations is related to a multiscale approach that treats multiple redundancy, scale and interdependence.

There are two aspects of the concept of complexity that appear to cause the most difficulty in formal discussions. The first problem is that there is a correspondence between information and entropy that is counter to common intuition. The second problem is the existence of observer dependence, suggesting arbitrariness in the definition. The multiscale approach developed in this paper suggests that these issues can be resolved by considering the scale dependence of observation rather than observation at a particular scale, and especially not just the microscopic scale. This leads to the characterization of a system by a scale dependent complexity rather than a single number. This is not a new concept but it has received relatively little attention in recent years and deserves to be reconsidered. In the following, we introduce the two conceptual difficulties as additional motivation for the multiscale formalism.

## 2. Model

The conventional definition of the physical entropy  $S$  of a system with a particular macrostate (e.g. energy, composition, volume,  $(U, N, V)$ ) in statistical physics and

that of information  $H(z)$ , the average information in a set of messages (descriptions) whose probabilities  $P_z(s)$  map onto the ensemble of the microstates of the physical system ( $z$  is a label for the set of possible messages and the probability over this set,  $s$  is a particular value from the set; see the discussion of notation in Ref. 10, Sec. 1.8.3) can be related by

$$H(z) = \frac{S(U, N, V)}{k \ln(2)} = - \sum_s P_z(s) \log_2 P_z(s). \quad (1)$$

The sum over  $s$  indicates a sum over the finite or infinite set of messages, where the messages are in one-to-one correspondence with the states of the system and therefore can be considered as labels or descriptions of each state of the system. Equation (1) allows the entropy to be generalized to treat non-equilibrium systems, as long as we assume a well-defined ensemble probability distribution,  $P_z(s)$ . We will further assume that the messages can be written using binary variables  $s = \{s_i\}_{i=1, \dots, n}$ ,  $s_i = \pm 1$ , that these binary variables correspond to the components of the system, and are all of the same scale, although they may be interdependent. The interdependence of these components is captured by the probability of the system state  $P_z(s) = P_z(s_1, \dots, s_n)$ , which does not generally factor into probabilities of the individual binary variables. The variables  $s_i$  considered individually have the probability  $P_{z_i}(s_i)$ , which we can for simplicity assume to be equal for the two possible values so that the information  $H(z_i)$  is 1.

The conceptual difficulty with the relationship given by Eq. (1) is that equilibrium systems, which maximize entropy, are not the systems we normally identify as most complex. The equilibrium form of a human being is a murky liquid — it is the most disordered/random system without visible structure or dynamics (even through its microscopic description requires the most information). This difficulty has led to definitions of complexity that give a low value for completely disordered systems, and a higher value for systems away from equilibrium. Probably the earliest and most discussed variant on information which can be used to characterize the non-equilibrium character of a system is Schrödinger's concept of negentropy [12] which, while originally defined as the negative of the entropy, can be better understood as the difference between the equilibrium entropy and the entropy of the system when it is in a non-equilibrium but well-defined ensemble  $P_z(s)$ . Negentropy, therefore, measures the entropic deviation of the system from the equilibrium state. Since this modified definition amounts to the addition of a constant offset (the equilibrium entropy) for all possible non-equilibrium macrostates of a particular system, it changes only in a non-essential way the concept defined by Schrödinger. The negentropy defined in this way is high for systems that are out of equilibrium. Its relationship with the multiscale formalism will be discussed below.

The second conceptual problem arises from a concern with the apparent dependence of description length on the observer. In contrast to the unique definition of the entropy of a physical system, the observer dependence of physical quantities

(e.g. kinematic ones) is well-known and the objective of theory is relating what different observers report. In light of this perspective, it should not be surprising that the complexity of a system may be observer dependent. We will argue that rather than a single number for characterizing the complexity, at the very least, a scale dependent complexity is needed, i.e. the complexity as reported by a sequence of observers whose observations of the system are at different scales. This scale dependence reflects the observer's ability to distinguish details of the system. All observers observe the same scope, i.e. the system in its entirety. Larger scale observers, however, see only redundant information. The concept of a scale dependent entropy dates back to the  $\epsilon$ -entropy of Kolmogorov [15], which is similar to Shannon noisy-channel information theory [13]. The study of the  $\epsilon$ -entropy has been largely limited to the fine scale limit to describe the Kolmogorov–Sinai dynamical systems entropy rate [15, 34]. There are several more recent works that have explored the scale dependence of entropy in specific cases [10, 35–38].

The expressions we will give for the scale dependent entropy in terms of component interdependence can be related to an intuitive notion of complexity without either of the conceptual difficulties discussed above. Underlying the resolution of the conceptual problems is the sum rule we will prove that reflects a tradeoff between fine-scale and large-scale complexity for systems with different probability distributions over the same underlying degrees of freedom. This sum rule has widespread relevance to our understanding of complex systems.

The multiscale representation is written by considering the level of redundancy of specification of the information by the elements. The amount of information at each scale is not an integer, but the scale is integral and ranges from 1 to  $n$ . The amount of information that is shared by at least  $k$  variables is the complexity  $C(k)$ , we also define  $D(k)$ , the information that has a redundancy of  $k$ , but not higher.  $D(k)$  partitions the information disjointly and

$$C(k) = \sum_{k'=k}^n D(k'). \quad (2)$$

The conventional entropy is the fine-scale complexity  $H(z) = C(1)$ .

We can obtain an explicit expression for multiscale complexity in terms of the probability distribution over the state space from a consideration of the algebraic properties of the information. After an involved derivation, we obtain an explicit expression for the scale dependent information in terms of the sum over the information in all subsets of a particular size. The derivation requires ensuring that a proper count is made of the different levels of redundancy of information. In general, if we consider the information present in a particular subset of the system, this will overlap the information of other subsets of the system. The degree to which this overlap occurs (number of redundancies) depends on the multiplicity of the underlying dependencies of the variables. In order to obtain just the information in the system that has at least a particular redundancy and not that with less, we determine the necessary combination of the possible sums over the information in

subsets of the system. A careful treatment of the combinatorics of  $k$ -fold redundancy for all  $k$  is necessary. Rather than provide this derivation here we will prove below that our derived form is the only possible expression. The result is

$$C(k) = \sum_{j=0}^{k-1} (-1)^{k-j+1} \binom{n-j-1}{k-j-1} Q(n-j) \quad (3)$$

$$D(k) = \sum_{j=0}^k (-1)^{k-j+1} \binom{n-j}{k-j} Q(n-j), \quad (4)$$

where

$$\begin{aligned} Q(n-j) &= \sum_{\{i_1, \dots, i_j\}} H(z - \{z_{i_1}, \dots, z_{i_j}\}) \\ &= - \sum_{\{i_1, \dots, i_j\}} \sum_s P_z(s_1, \dots, s_n) \log_2 \sum_{\{s'_{i_1}, \dots, s'_{i_j}\}} P_z(s_1, \dots, s'_{i_\alpha}, \dots, s_n) \end{aligned} \quad (5)$$

is the sum over subspace entropies, with  $n-j$  components, i.e. the information excluding sets of  $j$  binary variables from the  $n$  total number of binary variables. In the notation used the subtraction  $z - \{z_{i_1}, \dots, z_{i_j}\}$  indicates set subtraction, i.e. omission from the set of variables considered. The sum over  $\{i_1, \dots, i_j\}$  implies a sum over all distinct subsets of  $j$  variables out of the entire set of  $n$  variables. We note that these are new expressions that have not appeared in the literature and provide an explicit form for the information in  $k$ -fold dependencies for any system described by a probability distribution over the ensemble [39].

These expressions can be concretized by considering the behavior of a system that partitions into subsets  $\Theta_\alpha$  each of which consists of  $|\Theta_\alpha|$  completely dependent variables, i.e. there are only two possible states of each subset, a binary degree of freedom. The redundancy of information inherent in the dependency of the variables implies that the scale of the degree of freedom is equal to the number of variables in the subset. Specifically, we consider the partition

$$n = \sum_\alpha |\Theta_\alpha| = \sum_q n_q q, \quad (6)$$

where  $n_q$  is the number of subsets consisting of  $q$  completely interdependent variables, i.e.

$$P_z(s_1, \dots, s_n) = \prod_{i,j \in \Theta_\alpha} \delta(s_i, s_j) \quad (7)$$

(while this expression indicates variables of a subset all adopt the same orientation, it is only necessary that the choice of one variable fixes the orientation of all other variables of the set) and for this simple case we have the result (by direct substitution):

$$D(k) = n_k, \quad (8)$$

this means that  $D(k)$  counts the number of distinct degrees of freedom that are  $k$ -fold redundant, which in this case is the same as the number of partitions with  $k$  elements.

To prove the results in Eqs. (3) and (4) we note the following must hold for the correct expressions: (a) the expressions must be symmetric in all subsets of the variables of the same size (this includes the case of single variables), (b) they must satisfy Eq. (8) for the example chosen. This proves the correctness of the formulas given in Eqs. (3) and (4), since by direct algebraic inversion of Eq. (4).<sup>a</sup>  $D(k)$  is the only function of average subsystem entropies with property (b).

For the case specified by Eq. (8), there is a simple relationship of the multiscale complexity with robustness to noise. In the presence of noise that is independent for each variable, the probability of observer error is given by the likelihood that more than half of the redundant variables are incorrect, which for  $k$  completely correlated variables is

$$\begin{aligned}
 p_{\text{err}}(k) &= \sum_{m=0}^{k/2} \left(1 - \frac{1}{2} \delta(m, k/2)\right) \binom{k}{m} (1-r)^m r^{k-m} \\
 &\sim \frac{1}{2} \left(1 - \text{erf}\left(\sqrt{2k}(1/2 - r)\right)\right),
 \end{aligned}
 \tag{9}$$

where  $r$  is the probability of a single variable error, and the asymptotic expression is for large  $k$ .

We can obtain additional intuitive results that generalize Eq. (8) by considering a system that partitions into subsets. First consider partitions that are independent of each other but are not as simply constrained within. For such a system that partitions into several subsets (clusters), each of which is interdependent, but with no dependency between them, the complexities of subgroups add. If we have subsets of variables indexed by  $\alpha$ , we have

$$C(k) = \sum_{\alpha} C_{\alpha}(k),
 \tag{10}$$

where  $C_{\alpha}(k)$  is the  $k$ th entropy of the  $\alpha$ th subset. This generalizes the extensivity of the entropy and is a first step toward considering the topic of subdivision and hierarchy [10, Chap. 2, 41–43]. We can also consider the possibility of subdivisions that are exact replicas (i.e. multiple copies that have the same state) in which case it can be seen that:

$$C(k) = C_1(\lfloor k/g \rfloor),
 \tag{11}$$

where  $C_1$  is the complexity of one subdivision, and  $g$  is the number of copies. This means that the scale is multiplied, while keeping the same value of the complexity.

<sup>a</sup>Using the identity

$$\delta_{k,k'} = \sum_{j=1}^n (-1)^{k-j} \binom{n-k'}{n-j} \binom{n-j}{k-j}.$$

We see that for the case of completely dependent or independent variables, the definition of the complexity as a function of scale provides intuitive results.

It is now possible to describe the relationship between interdependence and the scale of degrees of freedom. This relationship is captured by a sum rule. Several qualitative and quantitative applications of such a sum rule have been discussed previously [10, 38], however this formal statement is presented here for the first time. The sum rule is proven by direct substitution of the expressions in Eqs. (3) or (4), and can be written as

$$\sum_k C(k) = \sum_k kD(k) = Q(1) = \sum_i H(z_i), \quad (12)$$

where the final expression is equal to  $n$  for equal probabilities for the case of the binary variables. A conceptual illustration of the sum rule is provided in Fig. 1. For a class of systems that have the same degrees of freedom, but different correlations between them, the sum rule reflects a tradeoff between complexity at different scales. When parts of a system are independent, they do not provide redundant information and cannot be visible at a large-scale. Independence indicates a large amount of information at a fine scale, while dependence indicates information at a larger scale. This tradeoff alleviates the conceptual difficulty with the conventional definition of entropy and complexity.

It is useful to compare the scale dependent complexity with the concept of negentropy. Adding to the usual definition of negentropy, the entropy of the equilibrium system (as discussed previously), and assuming that the equilibrium entropy is

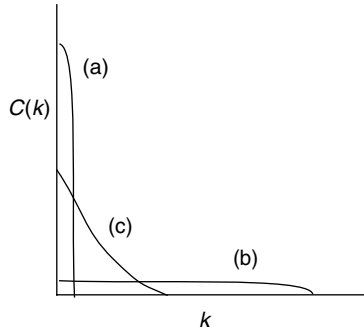


Fig. 1. Schematic illustration of the sum rule for entropy/complexity  $C(k)$  (vertical axis) as a function of scale,  $k$  (horizontal axis, increasing to the right). A system with the highest possible fine scale complexity corresponds to a system with independent/random degrees of freedom (curve (a)). When all degrees of freedom are coherent the system has a behavior at the largest scale, but the same low value of complexity at all finer scales (curve (b)). Intuitively, complex non-equilibrium systems (curve (c)) have a much higher complexity than either random or coherent systems, except at the finest scale for random systems and except a small but slightly higher complexity for coherent systems at the largest scales. The area under all curves is the same for systems with the same underlying degrees of freedom.



given by independent binary variables, the negentropy can be seen to be

$$\sum_i H(z_i) - H(z) = \sum_{k=2}^n C(k), \tag{13}$$

which includes all of the larger scale complexity without distinguishing its scale. A reason to consider the scale dependent complexity in its entirety is that we can have correlations between pairs of variables of the system (microscopic correlations) that give a high negentropy, without any large-scale behavior, i.e. observable degrees of freedom.

The above expressions can be made more intuitive by considering the case of three variables. In this case we have

$$\begin{aligned} C(1) &= H(z), \\ C(2) &= \sum_{\{i,j\}} H(z_i, z_j) - 2H(z), \\ C(3) &= \sum H(z_i) - \sum_{\{i,j\}} H(z_i, z_j) + H(z). \end{aligned} \tag{14}$$

We can illustrate the multiscale complexity for three elements using a Venn diagram (Fig. 2), which shows the information overlap between these elements. The area of the entire diagram is the total entropy  $H(z) = H(z_1, z_2, z_3)$ . The area of each circle is  $H(z_i)$ . The intersection of a pair of circles is the mutual information  $I(z_i; z_j) = H(z_i, z_j) - H(z_i) - H(z_j)$ , which corresponds to the information that can be learned from  $z_i$  about  $z_j$  or *vice versa*, or the information that is available to an observer from either  $z_i$  or  $z_j$ . The three pairwise overlapping regions also overlap in the region labeled  $D(3)$  in the figure. This might be given the notation  $I(z_1; z_2; z_3)$  and called the mutual information of three variables (with caveats given below).

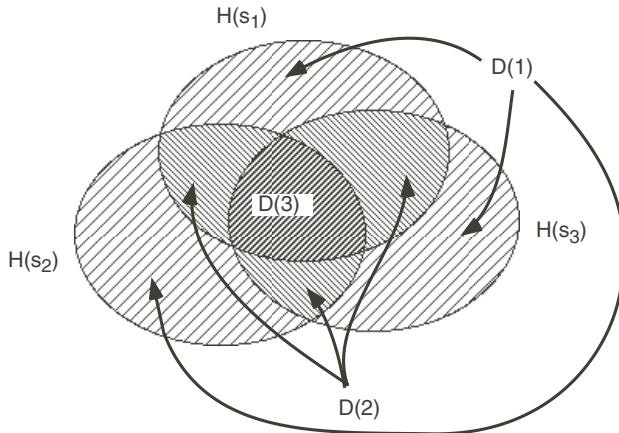


Fig. 2. Venn-type diagram illustrating the information and redundancy of the information in a system with three variables.

We identify the regions  $D(1)$ ,  $D(2)$ , and  $D(3)$  as the singly, double, and triply redundant regions, respectively. Expressions in Eq. (8) and the sum rule can be verified graphically, or the analogous expressions from Eq. (5) for  $D$ .

To work out a particular example, we consider the three binary variables as binary (Ising model) spins interacting through a magnetic interaction  $J$  to give the probability:

$$P(s) = e^{J \sum_{i,j} s_i s_j} / \sum_s e^{J \sum_{i,j} s_i s_j}. \quad (15)$$

We evaluate

$$\begin{aligned} H(z) &= 6J(e^{-J} - e^{3J})/Z \ln(2) + \log_2(Z), \\ \sum_{\{i,j\}} H(z_i, z_j) &= 6(2e^{-J} \log_2(2e^{-J}) \\ &\quad + (e^{-J} + e^{3J}) \log_2(e^{-J} + e^{3J}))/z + 3 \log_2(Z), \\ \sum_i H(z_i) &= 3, \end{aligned} \quad (16)$$

where

$$Z = 2(3e^{-J} + e^{3J}). \quad (17)$$

The results for  $C(k)$  from Eq. (3) are plotted in Fig. 3 as a function of  $J$ . For positive  $J$  the spin alignment causes the increase in  $C(2)$  and  $C(3)$  with the information overlap reaching 1. For negative  $J$ , the spins are frustrated in trying to antialign leading ultimately to six possible states  $C(1) = \log_2(6)$ .

What is interesting is the behavior of  $C(3)$ , which becomes negative for negative  $J$ , with a limiting value of

$$C(3) \rightarrow 3 - \log_2(54) + \log_2(6) \approx -0.169925. \quad (18)$$

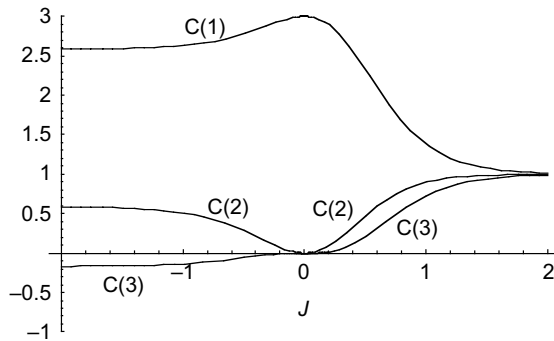


Fig. 3. Plot of the amount of information  $C(k)$  observable at each scale  $k$  as a function of the interaction strength  $J$  between three spins.

This is not consistent with our normal understanding of the Venn diagram in Fig. 2. However, the Venn diagram is only a visual aid, and is not a formally valid representation of information theory. Indeed, the negative value of  $C(3)$  a counter example. We note that the average mutual information of two variables must always be positive, but the mutual information that results from a single message may be negative. The case we have here is different because it is the average information rather than the information in a single message that is negative, this can occur when we consider the mutual information of three variables [43]. The meaning of a negative mutual information can be understood from the example we are considering here. Its origin is the incompatibility of the interactions that try to anti-align any pair of the three spins, technically known as frustration. Specifically, if we ask what does the state of spin 1 inform us about the state of spin 3, we learn that it is likely to be opposite. If, however, we ask, what does spin 1 inform us about the state of spin 2, and what does the state of spin 2 tell us about the state of spin 3, we will discover that the state of spin 3 is likely to be the same as spin 1. This conflicting information leads to the negative value of  $C(3)$ . More generally, the negative large-scale entropy reflects the possibility that partial information gained about a system from messages that contain errors may be misleading, i.e. this is disinformation resulting from partial knowledge. The negative value of  $C(3)$ , also implies that the multiscale complexity is not always a monotonically decreasing function, as has been stated elsewhere [10].

As a step towards discussing dynamics, consider a Markov chain with transition matrix  $T$  where the possible states at a particular time are assumed to be of a single scale (but higher scales arise from interplay between times)

$$p_{z^{j-1}|z^j}(s^j) = \sum_{s^{j-1}} T(s^j|s^{j-1})p_{z^{j-1}}(s^{j-1}). \quad (19)$$

We obtain expressions for  $C(k)$  per time step in terms of the conditional information along the chain ( $H(z'|z)$  is the information in  $s'$  given  $s$  taken from the probability distribution  $P_z(s)$ ):

$$C(k) = H(z^k|z^0) - H(z^{k-1}|z^0), \quad (20)$$

$$H(z^k|z^0) = - \sum_{s^0, s^k} p_{z^0}(s^0) T^k(s^k|s^0) \log_2(T^k(s^k|s^0)). \quad (21)$$

The expression for  $C(1)$  reduces to the standard entropy of a Markov chain, and  $C(k)$  is always positive. The sum rule is given by:

$$\sum_k C(k) = \lim_{k \rightarrow \infty} H(z^k|z^0) = H(z^0), \quad (22)$$

consistent with equation (12). For the more specific case of a single binary variable with transition matrix:

$$\begin{pmatrix} (1+f)/2 & (1-f)/2 \\ (1-f)/2 & (1+f)/2 \end{pmatrix}, \quad (23)$$

we have

$$H(z^k|z^0) = h((1 + f^k)/2), \quad (24)$$

where

$$h(p) = -p \log_2 p - (1 - p) \log_2 (1 - p) \quad (25)$$

showing the role of the eigenvalues of the transition matrix  $(1, f)$  in determining the multiscale complexity. If the information transferred along the chain was assumed to be retained from step to step in a random way, the result would be  $C(k) = h(p) (1 - h(p))^{k-1}$ , so the difference between this expression and the equations above reflects “correlations” in the propagation of information along the Markov chain.

The above treatment considered binary variables, but can be generalized. To generalize the treatment to apply also to continuum variables, such as position  $x$  of a particle, we begin using a simplified approach that considers the information associated with a single variable, even if that variable is a real number, to be at a single scale. In this case, the concept of scale becomes solely associated with numbers of particles that are behaviorally related, rather than the distance a particular particle travels. While this violates some aspects of intuition, it enables a direct analysis of a variety of physical systems in which particles are the underlying physical entity. In particular, all of the above expressions for  $C(k)$  and  $D(k)$  apply to multiparticle systems with the interpretation of  $s_i$  as the particle coordinate(s) and  $H(z_i)$  as the information needed to specify each coordinate. The simplest non-trivial case is  $C(k)$  for an ideal gas of indistinguishable particles. While it might seem that the independence of the particles would lead to a non-zero  $C(k)$  only for  $k = 0$ , indistinguishability is an implicit interdependence. Since the entropy of any subset of  $j$  particles has the form  $S(j) = j k_B \ln(v/j)$ , where  $v$  is temperature dependent and proportional to the volume, we have

$$C(k) = \sum_{j=0}^{k-1} (-1)^{k-j+1} \binom{n-j-1}{k-j-1} \binom{n}{j} (n-j) \log_2(v/n-j). \quad (26)$$

$C(k)$  for  $k > 0$  depends only on  $n$ . A plot is shown in Fig. 4. A more complete treatment of continuum variables would also consider that the information about a single real number does not have a single scale, but rather has multiple scales associated with smaller and larger variations in the position variable  $x$ . This separation of scales is, in a sense, orthogonal to the issue of distinguishability described above. The total amount of information at all scales, appears in the formalism as  $C(1)$  which can be inferred from the usual result  $C(1) = \log(L/d)$  where  $L$  is the maximum variation and  $d$  is the finest observable precision. A more complete discussion of the scale dependence of continuum variables will be given elsewhere.

The technical results reported thus far may not provide sufficient intuitive understanding of the nature of this formal framework. The central idea is to consider the collective motions of a system and the relationship of such collective motions to the dependencies that exist between smaller scale degrees of freedom. As an aid

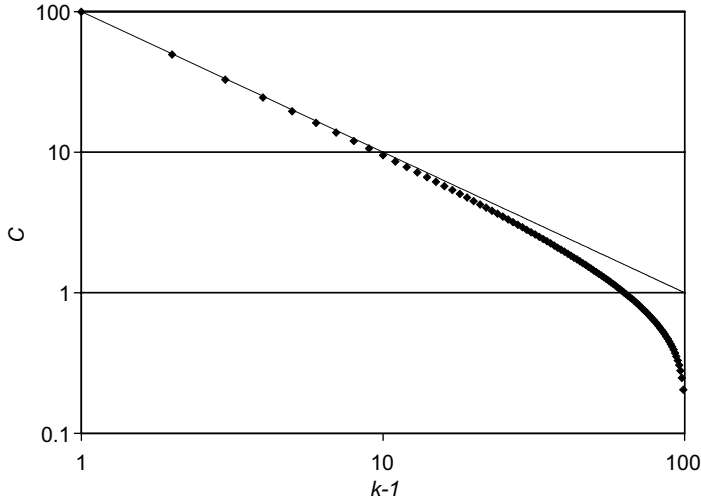


Fig. 4. Plot of  $C(k)$  for a gas of  $N = 100$  indistinguishable particles. The line is the curve  $N/(k - 1)$ .

to communicating the broad relevance of this concept, it may be helpful to apply this framework to an example of a complex system. Consider the comparison of the multiscale complexity of two systems, the first is a human being, the second is a collection of biological cells. We consider the human being to be confined to a single large room, the collection of cells, e.g. microbes, to be confined to a pond of the same volume as the room. As a first approximation, we consider as variables describing the system the set of possible locations of the cells of the person, and the set of possible locations of the microbes. The multiscale complexity of the locations of the microbes might be first approximated as randomly located at any location in the volume available. The cells of the person are similarly randomly located anywhere in the room. The primary difference being that the location of the cells of the person must all be near the same location at any particular time, while those of the microbes need not be. This statement might be described by correlations between positions of the cells, the pairwise distribution, but such a representation does not capture the many possible dependencies that might exist between the variables. In the formalism developed in this paper, we use the  $N$ -fold correlation of all of the positions of the cells to capture their mutual dependency. Indeed, the multiscale complexity of these two systems can be simply characterized. For the human being at the simplest level of approximation it is  $C(k) = H(x)$  for  $k \leq N$ , and zero otherwise; and  $D(k) = H(x)$  for  $k = N$ , and zero otherwise. For the microbes it is  $C(k) = NH(x)$  for  $k = 1$ , and zero otherwise; and in this case  $D(k)$  is equal to  $C(k)$ . If we relax the simplest approximations used in this analysis and consider the possibility of interactions between microbes, or freedom of motion of parts of the human being with respect to other parts, this would give rise to corrections to the multiscale complexity. All such corrections

reflect the dependence/independence of the degrees of freedom of the system. The formal result contained in Eqs. (3) and (11) provides proof that a tradeoff between fine-scale and large-scale degrees of freedom is preserved no matter how realistic a model is used for the discussion of biological (or other complex) systems.

Often, complex systems are considered to be systems that have behaviors on multiple scales. This is frequently captured conceptually by self-similar geometrical fractals and more general scaling behaviors [44]. For the purpose of information theory, a specific geometric fractal has no information aside from its generating algorithm. However, we can consider various stochastic fractal models that have increasing information at finer scales. For simplicity, we can consider stochastic generalizations of conventional geometric fractals such as Cantor's set and the Sierpinski gasket. For Cantor's set, recursively defined as the omission of the middle third of remaining line segments starting from the segment  $[0, 1]$ , we define a stochastic fractal by defining  $s(x)$ , a binary variable located at a finite but large number  $N$  of evenly spaced locations  $x$  along the unit interval  $[0, 1]$ , through assigning recursively the middle third of the remaining intervals with a probability of  $1/2$  to the value 1, and otherwise zero. The resulting multiscale complexity is given by  $D(N/3m) = 2m - 1$ , for  $m = 1, 2, 3, \dots$ , so that  $C(k) = 2^{\lfloor 1 + \log_3(N/k) \rfloor} - 1$ . This gives the approximate expression  $\log C(k) \approx d(\log(N) - \log(k))$ , where  $d$  is the box-counting dimension  $d = \log(2)/\log(3)$ . Similarly, for the Sierpinski gasket, we define the value of  $s(x, y)$ , a binary variable located at one of a finite but large number  $A$  of locations uniformly distributed in a triangle, recursively with a probability of  $1/2$  to the value 1 in the middle (upside-down) triangle of the remaining unselected triangles partitioned into four triangles. This gives the value  $D(A/4m) = 3m - 1$ . This gives the result  $C(k) = 3^{\lfloor 1 + \log_4(N/k) \rfloor} - 1$ , or

$$\log C(k) \approx (d/d_0)(\log(N) - \log(k)), \quad (27)$$

where  $d$  is the fractal dimension  $d = \log(3)/\log(2)$ , and  $d_0$  is the embedding dimension  $d_0 = 2$ . The power-law result in Eq. (20) is quite general as can be seen by considering a  $d_0$  dimensional body and recursively partitioning it into  $W$  parts, of which  $X$  are assigned the value 0 or 1 with probability  $1/2$  and the remainder are subject to the next level of recursion. The box-counting dimension is  $d = d_0 \log(X)/\log(W)$  and the above expression gives  $C(k)$ . The power law behavior of the complexity of stochastic fractals is consistent with the idea that they represent systems with behaviors at all scales, and according to the multiscale complexity they are more complex than either fully disordered or fully ordered systems with the same number of degrees of freedom at all scales except the smallest and largest.

Finally, to relate the multiscale complexity to the usual study of large-scale behavior through the order parameters of an equilibrium thermodynamic system, we label the entropy by the size of the system,  $C_N(k)$ , and suggest that the thermodynamic limit is given by  $C_\infty(\alpha) = \lim_{N \rightarrow \infty} C_N(\alpha N)$ . We note that the order parameter reflects the highly redundant information that is available about the system in the thermodynamic limit. For example, we can consider an interacting Ising system with  $N$  interacting spins. Assuming that the system is ferromagnetic,

regardless of the geometry of interactions (long range or spatial), the high temperature limit corresponds to independent spins and the low temperature limit to having all spins aligned. The multiscale complexity of such a system in these two limits is readily understood. In the high temperature limit each of the microscopic variables is independent and the complexity is given by  $C(k) = N$  for  $k = 1$ , and zero otherwise. In the low temperature limit  $C(k) = 1$  for  $k \leq N$ . The same result applies for an anti-ferromagnet on a bipartite lattice. We can also consider generalizations of the Ising model, such as the Mattis model [46]. The conventional treatment of the thermodynamic limit requires explicitly recognizing the order parameter to study its properties. By contrast, the multiscale complexity, because it evaluates all  $k$ -fold correlations, allows for all possible orderings. This can be readily proven for the Mattis model, which, if the order parameter is known, can be transformed into the conventional infinite range Ferromagnetic Ising model by redefining the sign of a subset of the spin variables. The proof follows from the invariance of the multiscale complexity under redefinition of the variables. Similar considerations apply to more complex models such as spin glass [45] models, where conventional treatments require identifying “by-hand” the order parameter. Thus at low temperatures, the information in the macroscopic order will be present in  $C_N(k)$  at macroscopic values of  $k$ . Application of this methodology to thermodynamic systems to obtain the characteristic behavior through the phase transitions will be further discussed elsewhere [47].

### 3. Conclusion

In summary, we have studied the multiscale properties of the structure of the state space of a system in terms of the interdependence of components, treating all possible ( $k$ -fold) correlations. This treatment made use of the formal properties of the information of subsets of the system. We derived a specific expression for the information that is present in  $k$  or more fold redundancy in the system,  $C(k)$  and for precisely  $k$ -fold redundancy,  $D(k)$ . We proved the expressions by showing that they are the only ones that are correct for the case of a system partitioned into subsets that are completely interdependent. We further proved a sum rule demonstrating a tradeoff, for different systems with the same underlying variables, of information needed to describe systems at different scales. This multiscale treatment provided a resolution of two conceptual difficulties facing formal studies of quantitative complexity. Furthermore, we illustrated the utility of the formalism and its conceptual framework by applying it to a variety of models of simple and complex systems. We (a) calculated the multiscale complexity of simple interdependent systems revealing some interesting results, particularly that negative values for the  $k$ -fold information are possible; (b) calculated the multiscale complexity of a simple Markov chain; (c) made an initial step toward extending the formalism to continuum variables and indistinguishability; (d) showed the broader conceptual relevance of the formalism by considering the difference between the multiscale characterization of independent biological cellular microorganisms, and cells forming a

human being; (e) considered a model of stochastic fractals illustrating the idea that complex systems have behaviors on multiple scales; and (f) considered the application of the formalism to traditional studies of thermodynamic phase transitions. These results for models of simple and complex systems suggest that the multiscale complexity provides a widely applicable quantitative and conceptual tool for characterizing complex systems and their behaviors on multiple scales. In separate work we address the relevance of the multiscale complexity to the response of the system to external forces and the functional capabilities of complex biological and social systems, and particularly the importance of the sum rule in capturing a trade-off between behaviors/capabilities of a system at different scales [48–50].

## References

- [1] Maxwell, J. C., *Theory of Heat* (Appleton, London, 1871).
- [2] Szilard, L., ber die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen, *Z. Physik* **53**, 840–856 (1929).
- [3] Brillouin, L., *Science and Information Theory* (Academic Press, New York, 1956).
- [4] Landauer, R., Irreversibility and heat generation in the computing process, *IBM J. Res. Dev.* **3**, 183–191 (1961).
- [5] Bennett, C. H., Logical reversibility of computation, *IBM J. Res. Dev.* **17**, 525–532 (1973).
- [6] Bennett, C. H., Thermodynamics of computation — A review, *Int. J. Theor. Phys.* **21**, 905–940 (1982).
- [7] Fredkin, E. and Toffoli, T., Conservative logic, *Int. J. of Theor. Phys.* **21**, 219–253 (1982).
- [8] Zurek, W. H., Thermodynamic cost of computation, algorithmic complexity and the information metric, *Nature* **341**, 119–124 (1989).
- [9] Landauer, R., Information is physical, *Phys. Today* **44**, 23–29 (1991).
- [10] Bar-Yam, Y., *Dynamics of Complex Systems* (Perseus, 1997), Chaps. 8 and 9.
- [11] Lindgren, K., Microscopic and macroscopic entropy, *Phys. Rev.* **E38**, 4794 (1988).
- [12] Schrödinger, E., *What is Life?* (Cambridge University Press, Cambridge, 1944).
- [13] Shannon, C. E., A mathematical theory of communication, *Bell Systems Technical Journal*, July and October 1948; reprinted in C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, 1963).
- [14] Jaynes, E. T., Information theory and statistical mechanics, *Phys. Rev.* **106**, 620 (1957).
- [15] Kolmogorov, A. N., *Dokl. Akad. Nauk SSSR* **119**, 861 (1958); *ibid.* **124**, 754 (1959), Sinai, Ya. G., *Dokl. Akad. Nauk SSSR* **124**, 768 (1959).
- [16] Solomonoff, R. J., *Information and Control* **7**, 1 (1964); *ibid.* 224 (1964).
- [17] Shiriyayev, A. N. ed., *Selected Works of A. N. Kolmogorov, Volume III: Information Theory and the Theory of Algorithms (Mathematics and Its Applications)*, (Kluwer, Dordrecht, 1987).
- [18] Chaitin, G. J., *Information, Randomness & Incompleteness*, 2nd Edn (World Scientific, Singapore, 1990); *Algorithmic Information Theory* (Cambridge University Press, Cambridge, 1987).
- [19] Carnap, R., *Two Essays on Entropy* (University of California Press, Berkeley, 1977).
- [20] Bennett, C. H., *Int. J. of Theoretical Physics* **21**, 905 (1982).
- [21] Denbigh, K. G. and Denbigh, J. S., *Entropy in Relation to Incomplete Knowledge* (Cambridge University Press, Cambridge, 1985).



- [22] Lloyd, S. and Pagels, H., Complexity as Thermodynamic Depth, *Annals of Physics* **188**, 186–213 (1988).
- [23] Zurek, W. H. ed., *Complexity, Entropy and the Physics of Information* (Addison-Wesley, 1990).
- [24] Tononi, G., Sporns, O. and Edelman, G. M., *Proc. Nat. Acad. Sci.* **91**, 5033 (1994).
- [25] Gell-Mann, M., *Complexity* **1**, 16–19 (1995).
- [26] Toffoli, T., Biafore, M. and Leo, J., eds., *Proc. 4th Workshop on Physics and Computation* (PhysComp96) New England Complex Systems Institute (1996).
- [27] Gell-Mann, M. and Lloyd, S., *Complexity* **2**, 44–52 (1996).
- [28] Lindgren, K., Moore, C. and Nordahl, M., Complexity of two-dimensional patterns, *J. of Stat. Phys.* **91**, 909–951 (1998).
- [29] Wolpert, D. H. and Macready, W. G., Self-dissimilarity: An empirically observable complexity measure, in *Unifying Themes in Complex Systems*, ed. Y. Bar-Yam, (Perseus Press, 2000), pp. 626–643.
- [30] Andrienko, Yu. A., Brilliantov, N. V. and Kurths, J., Complexity of two-dimensional patterns, *Eur. Phys. J.* **B15**, 539–546 (2000).
- [31] Standish, R. K., On complexity and emergence, *Complexity International* **9** (2001).
- [32] Edmonds, B., Hypertext Bibliography of Measures of Complexity, <http://bruce.edmonds.name/combib/>
- [33] NECSI Discussion group, <http://necsi.org:8100/Lists/complex-science/List.html>
- [34] Latora, V. and Baranger, M., *Phys. Rev. Lett.* **82**, 520 (1999).
- [35] Gaspard, P. and Wang, X.-J., *Phys. Rep.* **235**, 291 (1993).
- [36] Aurell, E., Bofetta, G., Crisanti, A., Paladin, G. and Vulpiani, A., *Phys. Rev. Lett.* **77**, 1262 (1996); Abel, M., Biferale, L., Cencini, M., Falcioni, M., Vergni, D. and Vulpiani, A., *Physica* **D147**, 12 (2000).
- [37] Binder, P.-M. and Plazas, J. A., *Phys. Rev.* **E63**, 065203(R) (2001).
- [38] Bar-Yam, Y., General features of complex systems, *UNESCO Encyclopedia of Life Support Systems* (in press).
- [39] Wolf, D. R., Information and correlation in statistical mechanical systems, Ph.D. Dissertation, U. Texas, Austin, Texas, 1996; a quantity discussed here appears to be the same as  $D(k)$  though no explicit form was given.
- [40] Simon, H. A., *The Sciences of the Artificial*, 3rd Edn. (MIT Press, Cambridge, 1998).
- [41] Havel, I. M., Scale dimensions in nature, *Int. J. General Systems* **24**, 295 (1996).
- [42] Auger, P., *Dynamics and Thermodynamics in Hierarchically Organized Systems* (Pergamon, Oxford, 1990). Auger considers a system subdivided into components and equates the total entropy to a sum of subsystem entropies and entropies of the collective subsystem variables.
- [43] Cover, T. M. and Thomas, J. A., *Elements of Information Theory* (Wiley, NY, 1991), p. 12.
- [44] Mandelbrot, B., *The Fractal Geometry of Nature* (W. H. Freeman & Co., 1982).
- [45] Sherrington, D. and Kirkpatrick, S., *Phys. Rev. Lett.* **35**, 1792 (1975).
- [46] Mattis, D. C., *Phys. Lett.* **56A**, 421 (1976).
- [47] Gheorghiu-Svirschewski, S. and Bar-Yam, Y., *Phys. Rev. E* (submitted).
- [48] Bar-Yam, Y., Complexity rising: From human beings to human civilization, A complexity profile, in *Encyclopedia of Life Support Systems (EOLSS)*, UNESCO (EOLSS Publishers, Oxford, UK, 2004).
- [49] Bar-Yam, Y., Complexity of military conflict: Multiscale complex systems analysis of littoral warfare, Report to Chief of Naval Operations Strategic Studies Group, 2003.
- [50] Bar-Yam, Y., *Complexity* **9**, 37–45 (2004).