

The Evolution of Complex Systems

G. L. Hofacker

Technische Universität München, Theoretische Chemie, D-8046 Garching,
W-Germany, Lichtenbergstr. 4

R. D. Levine

The Fritz Haber Research Center for Molecular Dynamics, Hebrew University,
Jerusalem 91904, Israel

Z. Naturforsch. **43a**, 73–77 (1988); received July 11, 1987

A principle of evolution of highly complex systems is proposed. It is based on extremal properties of the information $I(X, Y)$ characterizing two states X and Y with respect to each other,

$$I(X, Y) = H(Y) - H(Y/X),$$

where $H(Y)$ is the entropy of state Y , $H(Y/X)$ the entropy in state Y given the probability distribution $P(X)$ and transition probabilities $P(Y/X)$.

As $I(X, Y)$ is maximal in $P(Y)$ but minimal in $P(Y/X)$, the extremal properties of $I(X, Y)$ constitute a principle superior to the maximum entropy principle while containing the latter as a special case. The principle applies to complex systems evolving with time where fundamental equations are unknown or too difficult to solve. For the case of a system evolving from X to Y it is shown that the principle predicts a canonic distribution for a state Y with a fixed average energy $\langle E \rangle$.

Key words: Mutual information, Channel capacity, Maximum entropy principle, Extremal properties of information.

I. Introduction

There is a need for principles which govern evolution, that is the change with time in the state of a system. Such changes can be brought about by both internal processes and external perturbations. We are concerned here with such systems where the fundamental processes are too numerous, too intricate or not even known. Nor is it necessary to provide a description of the system on a fundamental level since such a description would be far too detailed.

For a system in equilibrium our considerations would imply that a statistical description is the method of choice. However, we are interested in change and hence one possibility is to combine statistics with equations of motion [1–3]. Here we explore an approach suitable for such systems where the fundamental processes are far less well characterized. The approach is thus complementary to the progress currently being made [4] using ‘evolution equations’ as input.

We do borrow from statistical mechanics [5] the notion of using the maximum entropy principle [6] as a means of characterizing a complex system in terms of fewer relevant parameters. We fully recognize however that the principle, as it stands, is ‘stationary’ in character. It assigns probabilities to the different possible states of the system at a given moment in time. To discuss the process of change one requires either to add the equations of motion as a separate input (which can be done [7]) or an extension of the principle. Specifically, one requires a principle which makes explicit reference to the possibility of change in the state of the system.

Towards this goal we propose to use the central concept of information theory [8–10] namely that of information itself. Interestingly enough, this central concept has hardly had any impact outside of information theory proper. Most papers in the sciences use ‘an information theoretic approach’ in the sense of ‘one which uses entropy’. The technical concept of information has been left largely undisturbed. In previous work on the evolution of protein amino acid sequences [11] we had occasion to appeal to this concept. In this paper we intend to further develop the germ of same ideas which were formulated then [11, 12].

Reprint Requests to Prof. Dr. G. L. Hofacker, Technische Universität München, Theoretische Chemie, Lichtenbergstraße 4, 8046 Garching.

0932-0784 / 88 / 0100-0073 \$ 01.30/0. – Please order a reprint rather than making your own copy.



Dieses Werk wurde im Jahr 2013 vom Verlag Zeitschrift für Naturforschung in Zusammenarbeit mit der Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V. digitalisiert und unter folgender Lizenz veröffentlicht: Creative Commons Namensnennung-Keine Bearbeitung 3.0 Deutschland Lizenz.

Zum 01.01.2015 ist eine Anpassung der Lizenzbedingungen (Entfall der Creative Commons Lizenzbedingung „Keine Bearbeitung“) beabsichtigt, um eine Nachnutzung auch im Rahmen zukünftiger wissenschaftlicher Nutzungsformen zu ermöglichen.

This work has been digitalized and published in 2013 by Verlag Zeitschrift für Naturforschung in cooperation with the Max Planck Society for the Advancement of Science under a Creative Commons Attribution-NoDerivs 3.0 Germany License.

On 01.01.2015 it is planned to change the License Conditions (the removal of the Creative Commons License condition “no derivative works”). This is to allow reuse in the area of future scientific usage.

II. Information

The technical concept of information requires the explicit introduction of a change in the distribution over the states of the system. If the previous probability of the state i was $P_0(i)$ and is now $P(i)$ then the information provided is

$$I(i) = \log [P(i)/P_0(i)]. \quad (1)$$

We shall often need a special case of this result when the observation is a transition from state i to state j . Then we need to revise our estimate of the probability of the state i from the previous value, $P_0(i)$, to the new value

$$P(i|j) = P(j|i) P_0(i) / \sum_{i'} P(j|i') P_0(i'). \quad (2)$$

Here $P(j|i)$ is the transition probability from state i to j and (2) is just Bayes' theorem. (The change in the probability by additional evidence). The information provided is then

$$\log [P(i|j)/P_0(i)] = \log [P(j|i)/P(j)], \quad (3)$$

where

$$P(j) = \sum_{i'} P(j|i') P_0(i') \quad (4)$$

and the right hand side of (3) is expressed in terms of the physical transition probabilities. (Note that the state i precedes the state j in time). Hence $P(i|j)$ is not a physical transition probability while $P(j|i)$ is. Further details can be found in standard sources [8–10].

By its very definition, information satisfies what can be called a 'generalized H-theorem'. Loosely stated it claims that information degrades or at least does not increase. The theorem can be formulated using the schematic diagram

$$X \rightarrow Y \rightarrow Z. \quad (5)$$

Here X , Y and Z are three consecutive stages in the evolution of the system. The transition probabilities into the different states of Z can, in principle, depend on the situation at both X and Y . Say, however, the evolution from Y to Z is independent of the previous history. Then [9, 10]

$$I(Z, X) \leq I(Z, Y), \quad (6)$$

$$I(Z, X) \leq I(Y, X). \quad (7)$$

Here $I(A, B)$ is the information provided by the distribution at B about that at A (or vice versa).

Information tends inherently to degrade. At stage Z we have less information about X than at the earlier stage Y .

The information provided between two stages in the evolution of the system depends on two sets of probabilities. (i) The input probabilities (these are the probabilities of the different states at the earlier stage) and (ii) The physical transition probabilities. These two sets are taken by us to be independent of one another and so can be separately varied. Our proposed variational principle is based on the mathematical results that the information has an extremal value with respect to either one of these variations [10]. But please note an essential difference. The extremum is a minimum when the transition probabilities are being varied and it is a maximum when it is the input probabilities that are subject to variation. The latter property is much better known [9], but both results are correct [10, 13].

III. A Principle of Evolution

We propose to characterize the evolution of complex systems by the following principle:

Between any two stages in the evolution of the system the information provided is extremal.

It should also be recognized that implicit in our principle is that the extremum may be subject to constraints. In that respect our principle is similar to that of maximal entropy and we shall argue that it reduces to the latter principle in the suitable limit. The essential difference between the two principles is that the one proposed above can determine the transition probabilities as well, if these are unknown. This is the technical sense of our earlier statement that we are dealing with processes of such complexity that the fundamental mechanism of change is not known or too detailed to be handled.

To see the motivation behind the principle consider an evolution from stage X to stage Y . One readily shows that the information between them can be written as [9, 10]

$$I(X, Y) = H(Y) - H(Y|X). \quad (8)$$

Here $H(Y)$ is the entropy of the distribution of states at the stage Y and is the quantity which is made extremal in the usual principle of maximum entropy. The new feature here is the second term, $H(Y|X)$. It is

the entropy of the distribution at stage Y given that it originated by evolution from stage X . $H(Y|X)$ is non negative and by making I an extremum one strikes a balance between making $H(Y)$ as large as possible and making $H(Y|X)$ as small as possible. Why? Examine the motivation for the principle of maximum entropy, [5]. We are to tell (i) the truth, (ii) the whole truth and (iii) nothing but the truth. We tell the truth by imposing all the constraints that are relevant and we tell nothing but the truth by maximising our uncertainty (the entropy) so as to obtain the least biased distribution. Where, however, have we told the *whole* truth about the distribution? We should make the uncertainty in Y *conditional upon what we know* as low as possible. That is precisely what is achieved by (8). It makes the uncertainty at stage Y minimal with respect to what we do know and maximal with respect to what we do not know.

The practitioners of maximal entropy have been aware for some time about the missing ‘whole truth’ aspect of the principle. What is then often done in practice is to seek the minimal value of the maximal entropy by varying over all possible constraints. Our principle includes that practice as a special case.

Our principle is meant to apply to systems in the physical world. Within information theory proper, seeking the extremum of I is an established procedure for the selection of codes and for the design of (noisy) information channels. In information theoretic terminology what our principle does is to match the code (the states of the system) and the channel (the transition probabilities). The maximal value of the information (for given transition probabilities) is termed ‘the capacity’ of the channel. Hence one can also state our principle that the evolution channel operates at capacity. We have previously discussed this result for the special case of the evolution of amino acid sequences in proteins [11]. Another aspect related to a match between a physical code and a physical channel was discussed elsewhere [12]. The subject was the information in an indirect experiment where observable B is measured in order to learn about observable A . Our present considerations are, in a sense, a special case where B is observable A , but at a later time.

Textbooks necessarily deal with the simpler cases. It is therefore worthwhile to emphasize that the concept of information is by no means limited to the type of evolution which is described as a Markov process. Nor must the process be linear in the input probabilities. The required concepts (the ‘sum’ and ‘product’

channels) have already been introduced by Shannon [14].

IV. Implications

The first requirement from a new principle is that it subsumes the presently known results as special cases. We have already mentioned that a generalized H-theorem is inherently satisfied for systems without memory. Indeed the result is somewhat stronger. Consider the scheme (5). Then

$$I(X \& Y, Z) \geq I(Y, Z) \quad (9)$$

with equality if *and only if* the transition probabilities into the different states at stage Z depend only on the states at stage Y . If they do also depend on stage X (i.e. the system does have memory) then more information is provided.

Already it was also noted that the principle of maximal entropy is a special case and thus such results that have previously been derived from it are included here as well.

Where do we go beyond the available principles? First and foremost is that we do not have to know the transition probabilities. They can be determined, in terms of the constraints, by seeking the extremum of I . Since I is a concave function (u) of the transition probabilities our principle will always generate such transitions which will drive the system as much as possible towards equilibrium as is consistent with the constraints. Thus far, the usual route was to assume a reasonable set of transition probabilities, solve the master equation and then examine which features of the solution are robust, i.e. independent of the finer details of the transition probabilities. The proposed procedure assures us the set of least biased (maximally non committal) transition probabilities, from the very start. Of course, we do not have to ‘not to know’ the transition probabilities. If these are given, one can still invoke the principle so as to match the input probabilities to the constraints imposed by the evolution.

V. Example

There are few ‘universal’ results for the evolution of complex systems and fewer still which can be discussed in a short space. Hence the example we chose is a very simple one, yet it suffices to show that the

principle can be implemented and that it yields sensible results.

Consider a physical system which changes from stage X to stage Y . Let i index the states at stage X , $i = 1, \dots, n$ and j index the states at stage Y , $j = 1, \dots, n$. The only constraint we impose on the system at stage Y is that its mean energy $\langle E \rangle$ is given,

$$\langle E \rangle = \sum_j P(j) E_j. \quad (10)$$

Here E_j is the energy of state j . The intuitive expectation is that irrespective of the distribution at stage X , if the only constraint placed on the evolution is the mean energy at stage Y then the system will be found in a (canonical) equilibrium. This is indeed what the principle will predict and hence we do not bother to state constraints on stage X since the system at stage Y will be at equilibrium irrespective of what the constraints on stage X are.

To prove the desired result in only a few steps we start by writing the logarithmic inequality ($\ln x \geq 1 - 1/x$ with equality iff $x = 1$) as

$$L \equiv \sum_i P(i) \sum_j P(j|i) \ln [P(j|i)/p(j)] \geq 0. \quad (11)$$

$p(j)$ in (11) is any normalized distribution of states for stage Y . Equality obtains in (11) if $P(j|i) = p(j)$. For our purpose we take $p(j)$ to be a canonical distribution:

$$p(j) = \exp \{-\beta E_j\}/Q, \quad (12)$$

$$Q = \sum_{j=1}^n \exp \{-\beta E_j\}. \quad (13)$$

The parameter β is adjusted to have that value which insures that the mean energy over the distribution $p(j)$ is the same value $\langle E \rangle$ as in (10):

$$\langle E \rangle = \sum_j E_j p(j). \quad (14)$$

Since $p(j)$ has only one parameter (Q is a function of β given by (13)), (14) can be solved for β . Now the actual distribution, $P(j)$, at stage Y is given in terms of the transition probabilities and the distribution at stage X by

$$P(j) = \sum_i P(j|i) P(i). \quad (15)$$

From (15), (14) and (10)

$$\sum_j E_j p(j) = \sum_j E_j \sum_i P(j|i) P(i). \quad (16)$$

But from (12) and (16), L , which is defined by (11) can be written as

$$\begin{aligned} L &= -H(Y|X) - \sum_i P(i) \sum_j P(j|i) \ln p(j) \\ &= -H(Y|X) + \ln Q + \beta \sum_i P(i) \sum_j P(j|i) E_j \\ &= -H(Y|X) + \ln Q + \beta \langle E \rangle \geq 0. \end{aligned} \quad (17)$$

In (17) one can easily recognize $S(\langle E \rangle)$,

$$S(\langle E \rangle) \equiv \ln Q + \beta \langle E \rangle, \quad (18)$$

as the entropy of the canonical distribution*. We have thus shown that

$$H(Y|X) \leq S(\langle E \rangle) \quad (19)$$

with equality iff the transition probabilities insure that the system reaches equilibrium. In other words, equality in (19) obtains if

$$P(j|i) = \exp \{-\beta E_j\}/Q, \quad \text{all } i. \quad (20)$$

The upper bound (19) on $H(Y|X)$ provides a lower bound for $I(X, Y)$. Using (8)

$$I(X, Y) \geq H(Y) - S(\langle E \rangle). \quad (21)$$

On the other hand

$$H(Y) \leq S(\langle E \rangle). \quad (22)$$

The reason for the inequality (22) being that $\langle E \rangle$ is the mean energy of the distribution at stage Y . But $S(\langle E \rangle)$ is, by construction, the maximal possible value of the entropy for all possible distributions whose mean energy $\langle E \rangle$ is given. Indeed, using the logarithmic inequality once more:

$$\sum_j P(j) \ln [P(j)/p(j)] \geq 0 \quad (23)$$

with equality iff $P(j) = p(j)$. We readily verify that equality obtains in (22) iff

$$P(j) = \exp \{-\beta E_j\}/Q. \quad (24)$$

Since $I \geq 0$ the two rigorous bounds (21) and (22) can coexist only with an equality sign in both. Since the conditions for equality are both necessary and sufficient we have proven (20) and (24).

* We use the symbol S rather than H since S is an extremal value of the entropy whereas H is an entropy functional. S depends on $\langle E \rangle$ since $\langle E \rangle$ determines β via (14) and β determines Q via (13).

Like for all variational procedures the result seems to drop out. But it is quite rigorous. In the absence of constraints to the contrary, the variational principle proposed by us 'drove' the system to equilibrium, for all possible initial states. It is important also to note that we have derived not only the distribution of the states of the system but also the transition probabilities which would yield such a state**.

VI. Discussion

The results found in Sect. V are just what one would expect. What needs to be done next is to study the

** At the risk of stating the obvious we note that all transition probabilities which satisfy detailed balance would *eventually* drive the system to equilibrium. What is unique about (20) is that in *one stage* the system reaches equilibrium, irrespective of the initial state.

more interesting cases where the constraints suffice to prevent the system from reaching equilibrium. One may also examine additional cases where the answers are either known or are intuitively reasonable. A further important direction is the study of the minimal constraints required to insure various types of generic behaviour, like non-equilibrium phase transitions, bifurcations, instabilities, chemical lasers [15] or the transition to turbulence oscillation and chaos in pumped, interlocking chemical reactions.

Acknowledgement

One of the authors (GLH) is indebted to the Fonds der Chemischen Industrie, Frankfurt, and RDL thanks the Stiftung Volkswagenwerk.

- [1] H. Haken, Synergetics, Springer Verlag, Berlin 1977.
- [2] N. G. van Kampen, Stochastic Processes in Physics and Chemistry, North Holland, Amsterdam 1981.
- [3] C. W. Gardiner, Handbook of Stochastic Methods, Springer-Verlag, Berlin 1983.
- [4] H. Haken, Advanced Synergetics, Springer-Verlag, Berlin 1983.
- [5] A. Katz, Principle of Statistical Mechanics – The Information Theory Approach, Freeman, San Francisco 1967.
- [6] R. D. Levine and M. Tribus, editors, The Maximum Entropy Formalism, M. I. T. Press, Cambridge 1979.
- [7] R. D. Levine, J. Chem. Phys. **65**, 3302 (1976); Adv. Chem. Phys. **47**, 239 (1981); R. D. Levine and Y. Alhassid, Phys. Rev. A **18**, 89 (1977); *ibid.* C **20**, 1775 (1979).
- [8] C. E. Shannon, Bell System Tech. J. **27**, 379 (1948).
- [9] R. Ash, Information Theory, Interscience, New York 1961.
- [10] R. G. Gallager, Information Theory and Reliable Communication, Wiley, New York 1968.
- [11] R. Coutelle, G. L. Hofacker, and R. D. Levine, J. Mol. Evol. **13**, 57 (1979).
- [12] G. L. Hofacker and R. D. Levine, Z. Naturforsch. **35a**, 490 (1980).
- [13] C. E. Shannon, Proc. IRE **4**, 142 (1959), also in Information and Decision Processes, R. E. Machol, ed. McGraw Hill, New York 1960.
- [14] C. E. Shannon, Information and Control **1**, 6 (1957).
- [15] A. Ben-Shaul, O. Kafri, and R. D. Levine, Chem. Phys. **10**, 367 (1975).