

ТЕОРЕТИЧЕСКАЯ ХИМИЯ

УДК 539.191

USING CUMULANT ANALYSIS FOR ENTROPIC COMPLEXITY MEASURES

© 2008 A. V. Luzanov

A simplified approach based on the cumulant analysis of the Shannon entropy is proposed for measuring complexity. We report the result of such analysis for some generic problems (mixing r – component ideal gas, the simplest geometric ornament, Schrödinger's cat states, and the logistic map). We argue that the new measures have benefits compared to the currently practiced measures in the Shiner-Davison-Landsberg and Lopes-Ruiz-Mancini-Calbet approaches.

1. Introduction

The quantification of complexity is in conformity with the fundamental science principle "to reduce the problem of nature to the determination of quantities by operations with numbers" (Maxwell, "On Faraday Lines of Force") [1]. There are many approaches for quantifying complexity in various fields of science, and the entropic paradigm is vital to our understanding of physical and chemical issues pertinent to complexity and uncertainties. Here we cite only typical reviews in the field [2-6]. The most sophisticated approach is based on the symbolic dynamics in a form of the so-called computational mechanics which was developed by Crutchfield et al in Ref. [7]. Some applications of the Crutchfield approach and related schemes to molecular dynamics problem are given in Ref. [8]. Furthermore, the simplified schemes for quantifying complexities [9,10] are frequently applied, especially for atomic systems [11]. At last, the purely topological [12] and related approaches involving some notions of differential geometry [13], extended the complexity theory to structural chemistry and chemoinformatics problems.

In this paper we try to simplify measuring complexity by using a conventional cumulant analysis in respect to entropy. This approach automatically leads to additive measures unlike the other frequently applied schemes previously given in [9, 10]. Some examples presented in the paper demonstrate a reasonability of the method proposed.

2. The usual entropy indices

We first consider a discrete random variable. Let P^{discr} be a normalised probability distribution of the form

$$P^{\text{discr}} = \{w_j\}_{1 \leq j \leq r}, \quad (1)$$

where r is a size (or cardinality) of the set P^{discr} . Then the conventional Shannon definition of the statistical entropy is

$$S = \langle -\log w \rangle = - \sum_{1 \leq j \leq r} w_j \log w_j. \quad (2)$$

In the above and in the following, the expectation value of a random variable X is symbolized by $\langle X \rangle$. In the case of continuous variables we must introduce the continuous counterpart of P^{discr} in the form of a normalized probability density function $\rho(x)$ determined in the interval $[a, b]$, so the Shannon entropy is written as

$$S = \langle -\log w \rangle = - \int_a^b \rho(x) \log \rho(x) dx, \quad (3)$$

In Eqs. (1) and (3) logarithm is usually taken to the base 2. The integration in Eq. (3) is performed over generally multidimensional domain.

It is certainly worth emphasizing that for many physical and chemical problems the theoretic-information entropy as being computed by Eqs. (1) and (3) cannot serve to be an appropriate complexity measure. Really, this quantity fails to distinguish correctly between two limiting cases

commonly considered as simplicity patterns. The first one is associated with a full order when only a single nonzero probability $w_1 = 1$, so in the case we naturally have

$$S_{\text{order}} = 0. \quad (4)$$

The second case is referred to a full disorder with the equiprobable distribution $w_j \equiv 1/r$, thus giving the maximal entropy value

$$S_{\text{disorder}} = \log r. \quad (5)$$

We see that the Shannon entropy does not satisfy the important one-humped criterion [2]. This criterion claims a low complexity for *both* ordered and random systems with a maximum of the complexity measure for the intermediate level of order. In other words, the high value in Eq. (4) is not consistent with a low complexity usually attributed to disordered states.

Several complexity measures were devised to obey the one-humped criterion. One of these is the so-called Shiner-Davison-Landsberg (SDL) complexity measure

$$C_{\text{SDL}} = (1 - S/S_{\text{max}})S/S_{\text{max}}, \quad (6)$$

where S_{max} is taken to be the entropy value referred to the equilibrium or relevant disordered state [10]. In this equation, S/S_{max} plays a role of a disorder parameter. An ambiguity of choosing S_{max} and other deficiencies of the SDL measure were discussed in Refs. [14,15]. Moreover, in our opinion the essential drawback of C_{SDL} is a lack of additivity, unlike the initial Shannon entropy.

The other popular complexity measure C_{LMC} is due to Lopes-Ruiz, Mancini, and Calbet [9]. This measure can be defined by the relations

$$C_{\text{LMC}} = S/\bar{r}, \quad (7)$$

$$1/\bar{r} = \sum_{1 \leq j \leq r} (w_j - 1/r)^2, \quad (7')$$

which has no ambiguous quantities. And yet this index also does not obey the necessary additivity since C_{LMC} involves the quantities of a different ‘‘asymptotic’’ behaviour, that is the additive entropy S and the approximately multiplicative quantity \bar{r} (for more details concerning the quantities of type \bar{r} see e.g. Ref. [16]). Nevertheless this complexity measure is rather useful if one is interested in using simple schemes [11]. The situation, however, can be improved by obtaining more consistent measures.

2. Entropy variance and cumulants as complexity measures

Now we turn to the new approach which allow us to produce the certainly additive complexity measures. The key idea is that entropy fluctuations can properly quantify complexity rather than the Shannon entropy itself or the C_{LMC} and C_{SDL} indices. Indeed, examining fluctuation gives an efficient tool in studying dynamic chaos problems as many works demonstrated it (see review [17] and the recent paper [18]). Here we exploit the similar approach to the Shannon entropies (2) and (3).

We recall that for a given random variable X , the statistical cumulants $k_u[X]$ are of the form [19]

$$k_2[X] = \langle (X - \langle X \rangle)^2 \rangle, \quad (8)$$

$$k_3[X] = \langle (X - \langle X \rangle)^3 \rangle, \quad (9)$$

$$k_4[X] = \langle (X - \langle X \rangle)^4 \rangle - 3(k_2[X])^2, \text{ etc.} \quad (10)$$

The cumulant $k_2[X]$ is but the customary variance for the given X . In this work $k_2[X]$ and $k_3[X]$ will be mainly studied. For our purposes these cumulants should be computed for a special choice $X = -\log w$. Therefore, the basic quantities are $\kappa_u \equiv k_u[-\log w]$, that is

$$\kappa_2 = \langle (-\log w - S)^2 \rangle, \quad (11)$$

$$\kappa_3 = \langle (-\log w - S)^3 \rangle. \quad (12)$$

Explicitly we have the typical working expressions as follows:

$$\kappa_2 = \sum_{1 \leq j \leq r} w_j (\log w_j)^2 - S^2 \quad (13)$$

and

$$\kappa_3 = \sum_{1 \leq j \leq r} w_j (-\log w_j - S)^3 \quad (14)$$

for the discrete distribution P^{discr} with the discrete entropy (2), and

$$\kappa_2 = \int_a^b \rho(x) [\log \rho(x)]^2 dx - S^2 \quad (15)$$

and

$$\kappa_3 = \int_a^b \rho(x) [-\log \rho(x) - S]^3 dx \quad (16)$$

for the entropy (3) of the continuous distribution $\rho(x)$.

To recognize why the introduced quantities can serve as admissible complexity measures, let us consider for definiteness the discrete case (1) and as before analyse the two limiting cases. Clearly, for the full order ($w_1 \equiv 1$) all cumulants take zero value. For the full disorder with the equiprobable distribution $w_j \equiv 1/r$, all the cumulants also vanish because $\log w_j \equiv -\log r = -S$. Thereby, the one-humped criterion [2] is satisfied for this principal set of situations. Additional features can be understood by studying specific examples.

3. Some Applications

3.a Mixing r -component ideal gas. This is a very familiar problem which can be found in most text-books on physical chemistry (e.g., see [20], p.240). Let x_j be the mole fraction of the j -th component in a given r -component gas. Then

$$\sum_{1 \leq j \leq r} x_j = 1, \quad (17)$$

so the set $\{x_j\}$ can be viewed as formal counterpart of the discrete distribution (1). More than that, the Gibbs entropy of mixing for ideal gases takes (ignoring prefactors) just the form of Eq.(2) (see Eq. (6.1-15) in Ref. [20]).

For the simplest case $r=2$ the corresponding results are given in Fig. 1. Recall that the value $x=1/2$ with the maximal entropy $S_{\text{max}}=1$ corresponds to the simplest case. We see a clear nontrivial extremum point for all the complexity measures. Evidently, plots in Fig. 1 are symmetric due to a symmetry of the problem under a change $x \rightarrow 1-x$. The numerical estimations give the extrema: $\kappa_2^{\text{max}} = 0.914$ at $x_0 = 0.083$, $\kappa_3^{\text{max}} = 3.451$ at $x_0 = 0.031$, $C_{\text{LMC}}^{\text{max}} = 0.153$ at $x_0 = 0.122$ and $C_{\text{SDL}}^{\text{max}} = 0.249$ at $x_0 = 0.100$.

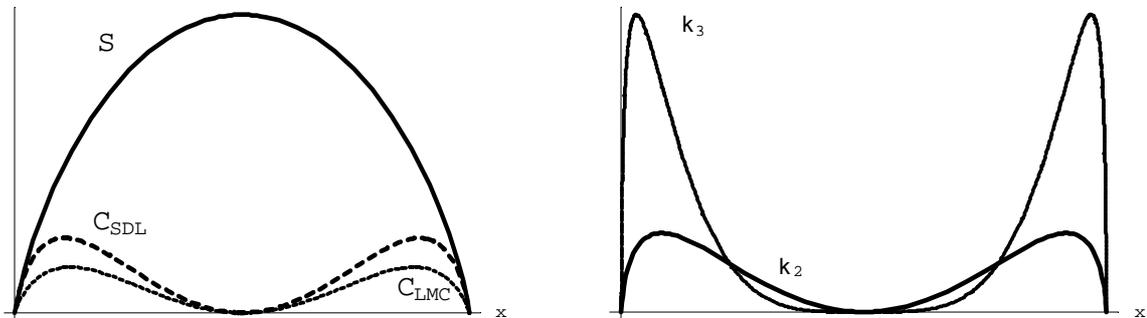


Figure 1. The Shannon entropy S and complexity measures C_{SDL} and C_{LMC} (the left panel) and cumulant complexity measures κ_2 and κ_3 (the right panel) for the 2-component system as a function of mixing parameter x .

For the case $r \geq 3$, due to the normalization condition (17) the geometry of the problem is that of a standard simplex. Therefore, all the maxima are equivalent and they are localized near each vertex of the simplex. Specifically, in the case of $r=3$ we have 3 equivalent maxima at $\{x_0, x_0, 1-2x_0\}$ with $x_0 = 0.062$ and $\kappa_2^{\max} = 1.586$ for Eq. (13), and with $x_0 = 0.023$ and $\kappa_3^{\max} = 6.186$ for Eq. (14). At the centre of simplex $\{1/3, 1/3, 1/3\}$ the entropy takes its maximal value $\log r = \log 3$ whereas the complexity measures naturally vanish at this point. These results are presented graphically in Fig 2.

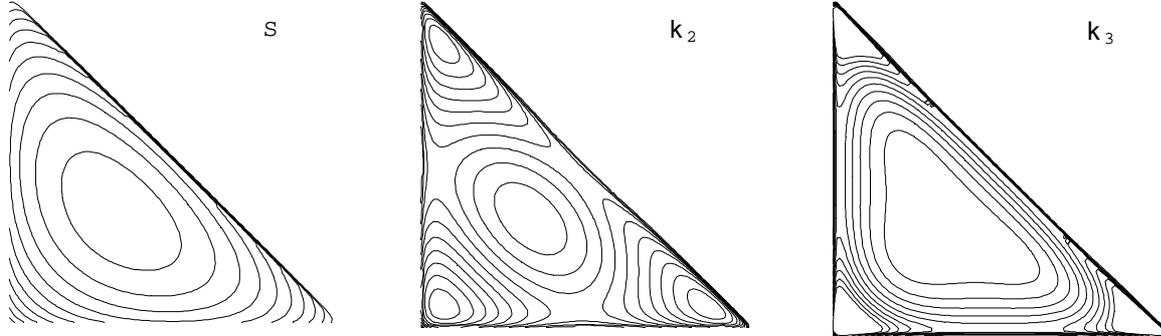


Figure 2. Contour plots for the Shannon entropy S and cumulant complexity measures κ_2 and κ_3 for the 3-component system as a function of mixing parameter in plane $\{x, y\}$ obeying inequalities $x, y \geq 0, x + y \leq 1$.

3.b. The simplest geometric ornament. The simplest two-dimensional ornament, which was previously investigated in Ref. [21], furnishes another example of the applicability of the cumulant complexity measures, at least in the cases of finite size probability distributions. In the model from Ref. [21] the colour of each element of the ornament changes with a definite probability. In so doing, the three-colour ornament is constructed by the following set of probabilities:

$$P_1(t) = P_2(t) = \frac{1}{3}(1 - e^{-P_0 t}), \quad P_3(t) = \frac{1}{3}(1 + 2e^{-P_0 t}) \quad (18)$$

where t takes the role of a temporal parameter, and P_0 is an initial probability. The same value $P_0 = 0.1$ as in Ref. [21] is used in our numerical experiments displayed in Fig. 3. We observe a clear maximum for both cumulant indices κ_2 and κ_3 at comparably low values of $t \approx 2.5$. At the same time, a special and rather involved probabilistic approach in Ref. [21] gives an essentially larger value for the critical t value. At this stage we cannot undoubtedly judge on which approach is more reliable for this problem.

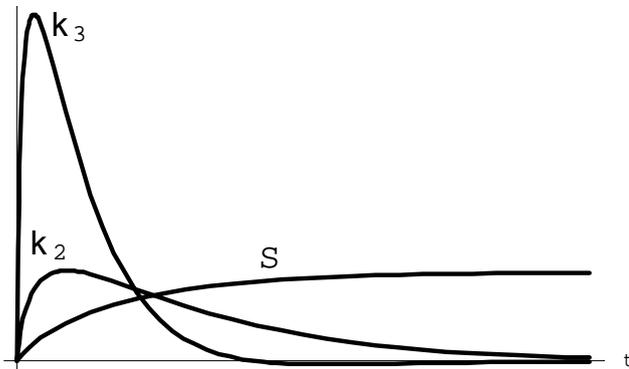


Figure 3. Shannon's entropy S and cumulant complexity measures κ_2 and κ_3 for the 2D-ornament as a function temporal parameter t .

3.c. Schrödinger's cat states. This model is known in quantum optics and quantum information processing theory [22]. In particular, Shannon's entropy for typical Schrödinger cat states is given in Ref. [23] where the Wigner formalism of quantum mechanics is exploited. Here we treat the same problem in the context of complexity theory. We start with the coordinate wave function

$$\psi_{\alpha}^{\text{cat}}(q) = [2\sqrt{\pi}(1 + e^{-2\alpha^2})]^{-1/2} \{e^{-(q-\sqrt{2}\alpha)^2/2} + e^{-(q+\sqrt{2}\alpha)^2/2}\}, \quad (19)$$

for which the momentum wave function can be cast in the form

$$\chi_{\alpha}^{\text{cat}}(p) = [\sqrt{\pi}(1 + e^{-2\alpha^2})]^{-1/2} e^{-p^2/2} \cos(p\alpha\sqrt{2}). \quad (20)$$

The parameter α represents a relative displacement of the two Gaussian functions associated with individual state-vectors. Here we use a simple phase-space description known in the general theory [23]. In this approach the phase space distribution $\rho(x) \equiv \rho(p, q)$ is realized in the form of the product of coordinate and momentum distributions $\rho(p, q) = |\psi(q)|^2 |\chi(p)|^2$. The results of numerical computations are shown in Fig 4. We see that the entropy presents no features of interest, unlike the complexity measures. The behaviour of the κ_3 index is more significant than that of the κ_2 index. It is naturally to expect that such a sensitivity is characteristic of κ_3 index, as Fig. 1 also demonstrates this.

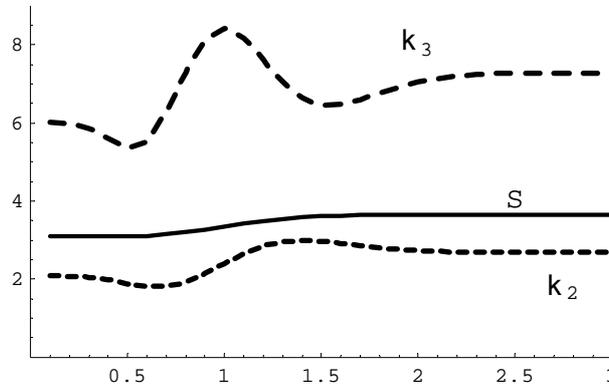


Figure 4. Shannon's entropy S and cumulant complexity measures κ_2 and κ_3 for the Schrödinger cat states as a function of the displacement parameter α .

3.d. The logistic map. The complexity of the logistic map was examined by the C_{SDL} -index in Ref. [10]. We carry out a comparable study of this map with adding the C_{LMC} index to the measures calculated here. The map is normally defined by the following quadratic iterations

$$x_{j+1} = \lambda(1 - x_j)x_j, \quad (21)$$

where λ is the control parameter taken from the interval $[0,4]$ (for more details see e.g. Ref. [25]). The most interesting is the region near the (Feigenbaum) bifurcation accumulation point $\lambda_{\infty} \approx 3.57$ corresponding to the beginning of the developed chaos. We used the same probabilistic scheme as in Ref. [10]. The results of the computations of the Shannon entropy and complexity indices are displayed in Fig. 5 where along with the main indices (13) and (14) the fourth cumulant index κ_4 (defined by Eq. (10)) is also presented. These results deserve a careful analysis because the logistic map takes a fundamental place in the modern theory of dynamic systems. Here we turn an attention to a principal feature of the cumulant complexity indices – they vanish in the region $\lambda < \lambda_{\infty}$ where the system is manifestly deterministic. We also see that the previously introduced indices C_{LMC} and C_{SDL} are nonzero in the same deterministic region.

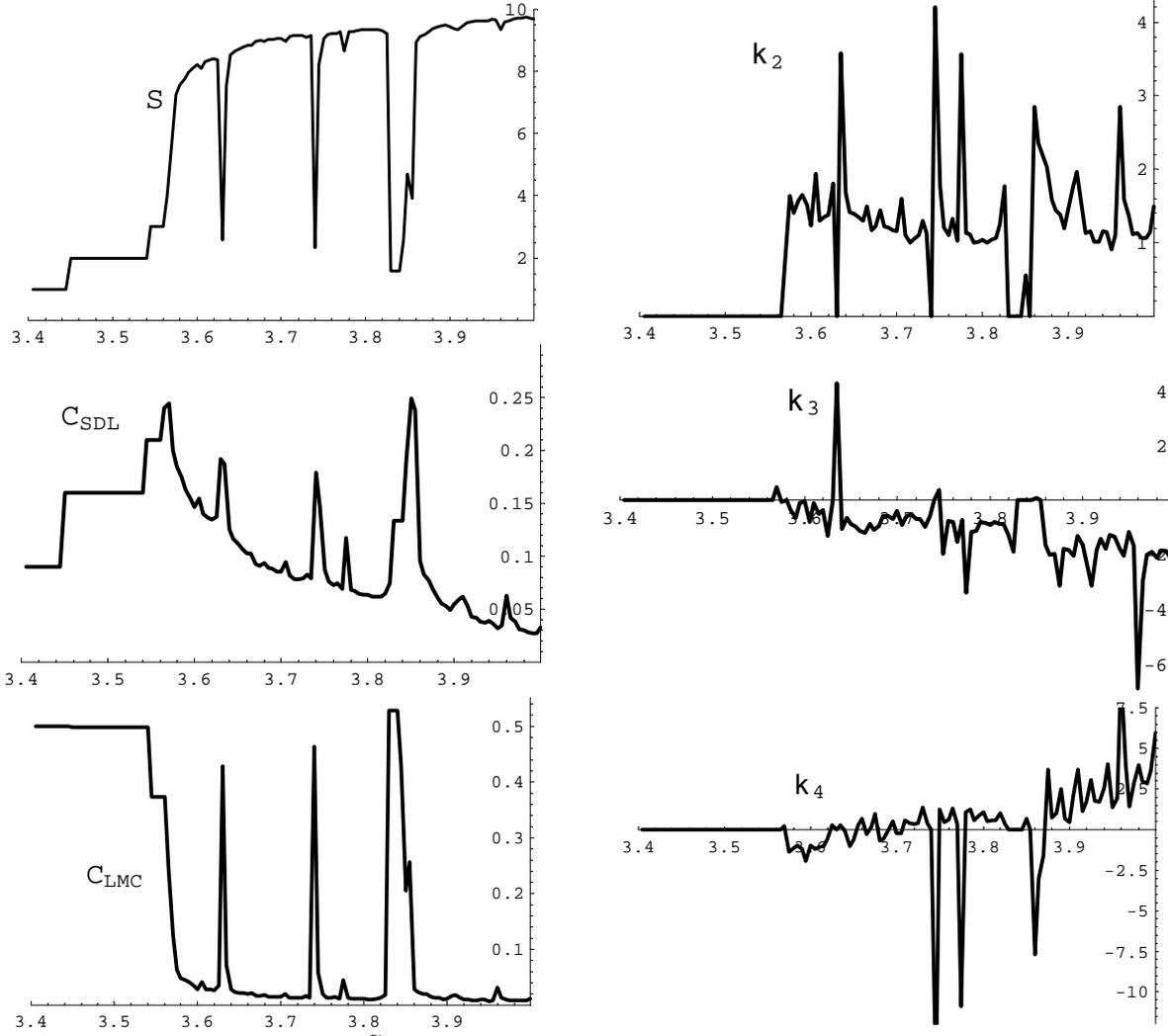


Figure 5. Shannon's entropy S , complexity measures C_{LMC} and C_{SDL} (the left panel) and cumulant complexity measures κ_2 , κ_3 , and κ_4 (the right panel) for the logistic map as a function the parameter λ .

4. Concluding remarks

In this paper we devised a quite simple scheme for quantifying complexity. The working expressions (13) - (16) for the cumulant entropy measures κ_u defined by Eqs. (11) and (12) are of the same feasibility as the Shannon entropy (2) or (3). The results of Section 3 indicate that κ_2 and κ_3 indices can capture the statistical complexity. Let us emphasize once more that the complexity measures (11) and (12) possess additivity what is important for multidimensional problems. On the contrary, the popular C_{SDL} and C_{LMC} measures lack the required additivity.

This study is only preliminary, and further work need to be done to ensure that the cumulant entropy analysis is applied to more involved problems such as time series or pattern recognition problems. In particular, a physical meaning of the cumulant complexity measures requires some clarification. A certain elucidation can be provided by the example of the Gibbs canonical distribution

$$p_j^{\text{Gibbs}} = \{ \exp[-\beta \varepsilon_j] / Z \}, \quad (22)$$

(all notations are standard). The ensuing relation is

$$\kappa_2^{\text{Gibbs}} = \beta^2 \langle (\varepsilon - \langle \varepsilon \rangle)^2 \rangle / (\ln 2)^2, \quad (23)$$

which is in fact the classical identity of heat capacity, C , and mean-square fluctuations $\langle \Delta S^2 \rangle$ corresponding to thermodynamic entropy [26-28]:

$$\langle \Delta S^2 \rangle = k_B C. \quad (24)$$

Here we do not distinguish between the isobaric and isochoric heat capacities as is normally assumed for internal degrees of freedom (e.g., see Ref. [29], Chapt. 2). Notice that some useful applications of the identity of type (24) can be found in the recent investigations of the protein stability and typical folding/unfolding processes [30, 31].

Using the general definition for cumulants [19], the extension of (23) is proved to be of the form

$$\kappa_u^{\text{Gibbs}} = (-\beta \ln 2)^{-u} d^u (\ln Z) / d\beta^u, \quad (23)$$

where, as before, Z is the canonical partition function and β the inverse temperature. Thus, we can expect that in general case, possible unusual behaviour of the κ_u quantities admit a reasonable rationalization in terms of effective capacities or their derivatives and the related phase-like transitions.

To be specific, return to the first example in Section 3.A. Let us treat this example as a two-level problem with $\Delta\varepsilon$ being the corresponding excitation energy. Making a change of variable $x = 1/(1 + \exp[-\beta\Delta\varepsilon])$ and recalculating $(\ln 2)^2 \kappa_2$ as a function of $\beta\Delta\varepsilon$ we just obtain the plot given in Ref. [32] for the heat capacity of the 2-state system (see p. 62 in Ref. [32]). The occurrence of a clear peak in the specific heat capacity is usually referred to as Schottky anomaly. In practice it indicates on the essential strong splitting of the lowest lying energy states (see Ref. [33], p. 60). Similarly, we can reinterpret the well-known sharp peak of C at the critical temperature of the Bose-Einstein condensation in the ideal Bose gas (see Ref. [28], Fig 4.3 on p. 123). The discontinuity of dC/dT and the same behaviour of κ_2^{Gibbs} manifest the crucial reorganization and developing complexity in the Bose gas during the Bose-Einstein transition. Hence, we believe that generally the observed peculiarities in the entropic cumulants reflect possible internal structural transitions, and the such interpretation is rather natural in the framework of the modern complexity theory.

Acknowledgements

The author is indebted to Dr D. Nerukh for the valuable comments.

References

1. Maxwell J. C. Trans. Cambr. Phil Soc. 1855—1856. V. X. [Russian translation: Джемс Клерк Максвелл, «Избранные сочинения по теории электромагнитного поля», М.: Гостехиздат, 1952, 685с.]
2. Adami C. Bioessays. 2002. V.24. P. 1085-1094.
3. Badii R. and Politi A. Complexity: Hierarchical Structures and Scaling in Physics. Cambridge: University Press, England, 1997, 332 p.
4. Entropy, Search, Complexity. Eds.: Csiszar I., Katona G. O. H., Tardos G., New York: Springer, 2007, 264 p.
5. Erdi P. Complexity Explained. Heidelberg: Springer-Verlag, 2008, 397 p.
6. Mainzer K., Thinking in Complexity. The Computational Dynamics of Matter, Mind, and Mankind, Heidelberg: Springer-Verlag, 2007, 482 p.
7. Crutchfield J.P. and Young K. Phys. Rev. Lett. 1989. V. 63. P. 105.-108; Crutchfield J.P. Physica D. 1994. V. 75. P. 11-54; Shalizi C. R. and Crutchfield J. P. J. Stat. Phys. 2001 V. 104. P. 817–879.
8. Nerukh D., Karvounis G. and Glen R. C. J. Chem. Phys. 2002. V. 117. P. 9611-9617; Nerukh D., Karvounis G., and Glen R. C., Complexity 2004. V.10. No 2. P. 40-46; Nerukh D., Ryabov V., and Glen R. C. Phys. Rev. E. 2008. V. 77. P. 036225-1 - 036225-11; Nerukh D. Chem. Phys. Lett. 2008. V.457. P.439-443.
9. Lopez-Ruiz R., Mancini H. L., and Calbet X. Phys. Lett. A. 1995. V. 209. P. 321-326; Catalan R. G., Garay J., and Lopez-Ruiz R. Phys. Rev. E. 2002. V.66. P. 011102-1 - 011102-6.
10. Shiner J. S., Davison. M., and Landsberg P. T. Phys. Rev. E. 1999. V.59. P. 1459-1464.
11. Chatzisavvas K. Ch., Moustakidis Ch. C., and Panos C. P. J. Chem. Phys. 2005. V. 123. P. 1744111-1 - 1744111-10; Panos C. P., Chatzisavvas K. Ch., Moustakidis Ch.C, and Kyrkou E. G. Phys. Lett. A. 2007. V. 363. P. 78-83; Borgoo A., De Proft F., Geerlings P., and Sen K. D., Chem. Phys. Lett. 2007. V. 444. P. 186-191; Angulo J. C. and Antolin J. J. Chem. Phys. 2008. V.8. P. 164109-1 - 164109-6; Angulo J. C., Antolin J., and Send K. D. Phys Lett. A. 2008. V. 372. P. 670–674.

12. Complexity in Chemistry: Introduction and Fundamentals. Taylor and Francis, Bonchev D. and Rouvray D. H. (Eds.), London, 2003. 344 p.; Nicolis S. and Trinajstić N. J. Chem. Inf. Comput. Sci. 2000. V. 40. P. 920-926; Bertz S. H. and Rucker C. Croat. Chem. Acta. 2004. V.77 P. 221-235.
13. Luzanov A. V. and Babich E. J. Mol Struct. (Theochem). 1995. V. 333. P. 279-290; Luzanov A. V. and Nerukh D. Functional Materials. 2005. V. 12. P. 55-64; Luzanov A. V. and Nerukh D. J. Math. Chem. 2007. V. 41. P. 417-435.
14. Crutchfield J. P., Feldman D. P., and Shalizi C. R. Rev. E. 1999. V.62. P. 2996-2997; Feldman D. P., Crutchfield J. P. Phys Lett. A. 2000. V. 238. P. 244-255.
15. Binder P. M. and Perry N. Phys. Rev. E. 2000. V. 62. P. 2998-2999.
16. Luzanov A. V. and Zhikol O. A. I. J. Quant. Chem. 2005. V. 104. P. 167-180.
17. Casetti L., Pettini M., and Cohen E. G. D. Phys. Rep. 2000. V. 337. P. 237 - 341.
18. Luzanov A. V. Kharkov University Bulletin, Chemical Series. 2007 V.770. Issue 11 (34) P. 178-187.
19. Kendall M. G. The Advanced theory of statistics. London: C. Griffin & Co, 1945, 457 p.
20. Mortimer R. G. Physical Chemistry. Third Ed. London: Acad. Press, 2008, 1386 p.
21. Andrienko Yu. A., Brilliantov N. V., and Kurths J. Eur. Phys. J. B. 2000. V. 15. P. 539-546.
22. Schleich W.P. Quantum Optics in Phase Space. Berlin : Wiley-VCH, 2001. 726 p.
23. Luzanov A. V. Kharkov University Bulletin, Chemical Series. 2007. V.770. Issue 11 (34). P. 188-193; Luzanov A. V. J. Chem. Phys. 2008. V.129. P. 094103-1 - 094103-15.
24. Hillary M., O'Connell R. F., Scully M.O., and Wigner E. Phys. Rep. 1984. V. 106 P. 121-167.
25. Tabor. M. Chaos and Integrability in Nonlinear Dynamics: An Introduction. New York: Wiley, 1989. 364 p.
26. Landau L. D. and Lifshitz E. M. Course of Theoretical Physics, vol.5. Statistical Physics, Part 1. Pergamon Press: New York, 1980, 544 c.
27. Problems in Thermodynamics and Statistical Physics. Ed. by Landsberg P. T. London: Pion, 1989, 588 p.
28. ter Haar D. "Elements of Statistical Mechanics", Oxford: Butterworth-Heinemann, 417 p.
29. Hirschfelder J. O., Curtiss C. F., and Bird R. B. Molecular Theory of Gases and Liquids. New York: Wiley; 1954, 1219 p.
30. Biological Systems Under Extreme Conditions: Structure and Function. Eds: Taniguchi Y., Stanley H. E., Ludwig H. Berlin Springer, 2002, 282 p.
31. Prabhu N.V. and Sharp K. A. Annual Review of Physical Chemistry. 2005. V. 56. P. 521-548.
32. Kittel C. and Kroemer H. Thermal Physics. New York: W. H. Freeman & Co, 1980, 473 p.
33. Couture L., Zitoun R. Statistical Thermodynamics and Properties of Matter. CRC Press: 2001, 626 p.

Поступила в редакцию 2 июня 2008 г

Kharkov University Bulletin. 2008. № 820. Chemical Series. Issue 16(39). A. В. Лузанов. Применение кумулянтного анализа для энтропийных мер сложности.

На основе кумулянтного анализа введен ряд упрощенных мер сложности. Предложенным методом численно исследованы несколько типичных задач (смешение в многокомпонентном идеальном газе, простейший геометрический орнамент, квантовые состояния шредингеровского кота и логистическое отображение). Утверждается, что новые меры имеют определенные преимущества по отношению к часто используемым мерам сложности, таким как мера Шинера-Дэвисона-Ландсберга и мера Лопес-Руиза-Манчини-Калбета.