Noname manuscript No.
(will be inserted by the editor)

On the second law of thermodynamics

Andreas Henriksson

Received: date / Accepted: date

Abstract In this article, it is argued that, given an initial uncertainty in the state of a system, the information possessed about the system, by any given observer, tend to decrease exponentially until there is none left. By linking the subjective, i.e. observer dependent, concepts of information and entropy, the statement of information decrease represent an alternative formulation of the second law of thermodynamics. With this reformulation, the connection between the foundations of statistical mechanics and classical mechanics is clarified. In conclusion, it is argued that concepts such as probability, ergodicity, entropy, as well as the arrow of time, arise naturally as a consequence of the Gibbs-Liouville theorem in combination with the fact that any given observer of a system do not possess infinite knowledge about the initial conditions of the system.

Keywords Gibbs-Liouville theorem · Reversibility · Ergodic theorem · Statistical equilibrium · Second law of thermodynamics · Arrow of time

PACS 05.20.-y · 05.70.-a · 02.50.-r

Contents

Andreas Henriksson
St.Olav school, Jens Zetlitzgt. 33, 4008, Stavanger; Norway
E-mail: andreas.henriksson@skole.rogfk.no
ORCID: 0000-0001-9014-4320

1 Introduction

A key feature of quantum mechanics which distinguish it from classical mechanics is the Heisenberg uncertainty principle [1][2]. It state that there exist a fundamental limit to the precision by which the state of a system on phase space can be determined. For the ultimate purpose of gaining a better understanding and illuminating the key differences between the evolution in time of classical and quantum systems, the concepts and ideas which lay the foundations of statistical mechanics is in this article revisited and reinterpreted.

In the discussion on the foundations of statistical mechanics, it is important to realize that it is a theory which rely both on the fundamental deterministic character of the evolution of classical systems, as characterized mathematically by the Gibbs-Liouville theorem [3][4][5], and the fact that any given observer of such a system do possess some amount of ignorance about the complete set of degrees of freedom which describe the system. It is this ignorance which lead to the appearance of concepts such as uncertainty, probability and entropy. In fact, it is this ignorance which, from the perspective of the observer, lead to the conclusion that classical systems evolve towards a state of equilibrium where the entropy is at a maximum. In other words, the second law of thermodynamics do not have its origin within a fundamental law of Nature but rather from the ignorance possessed by the observer of the system.

2 The Gibbs-Liouville theorem

For classical Hamiltonian systems, the Gibbs-Liouville theorem state that the Hamiltonian flow on phase space is incompressible. A necessary and sufficient condition for incompressibility is that the divergence of the Hamiltonian phase flow velocity vanish, i.e. that

$$\boldsymbol{\nabla} \cdot \mathbf{v} = 0 \tag{1}$$

where

$$\mathbf{v} = (\dot{q}, \dot{p}) = \left(\frac{\partial \mathcal{H}}{\partial p}, -\frac{\partial \mathcal{H}}{\partial q} \right) \tag{2}$$

where $\mathcal{H}$ is the Hamiltonian of the system. The Gibbs-Liouville theorem can be interpreted to represent a mathematical statement on the deterministic evolution of classical Hamiltonian systems, i.e. that distinctions between the possible states of the system, or, equivalently, that the information content within the system, is conserved in time [5].

3 Uncertainty and indistinguishability

Even when ignoring the laws of quantum mechanics, which place a fundamental limit on the precision which can be gained, the dynamical evolution of a system is quite complicated. Most systems of interest contain a vast amount of particles that interact in complicated ways. For such large systems, it is usually very hard to track the individual evolution of each particle as the system evolve in time. Perfect knowledge about the position and velocity, or momenta, of each individual particle is lost. It is lost not because of a fundamental violation of information conservation but merely because of the difficulty for an observer to keep track of all the degrees of freedom. Therefore, from the perspective of the observer, there is an uncertainty Δq associated with the position of a state and an uncertainty Δp associated with the momentum of a state. For this reason, the observer is unable to determine with absolute certainty the state of the system at any given time. The observer can only determine whether or not the system occupy a state which lie within any given region Ω_j on phase space, whose volume V_{Ω_j} is given by the uncertainties Δq and Δp, i.e.

$$V_{\Omega_j} = \Delta q \Delta p \tag{3}$$

The volume V_{Ω_j} is thus a measure of how ignorant the observer is about the details of the system, in the sense that the observer cannot locate an individual state to a greater precision than the size of Ω_j. Due to this lack of precision, the observer is unable to distinguish between states that lie within Ω_j. All states within Ω_j, with their individual sets of degrees of freedom, has, from the perspective of the observer, collapsed into a single state whose single set of degrees of freedom is given by $q + \Delta q$ and $p + \Delta p$. This so-called coarse-grained, or mixed, state is not a fundamental, or pure, state of the system. It is a description that average over all pure states within Ω_j. Put differently, a mixed state ψ_j, $j \in [1, M]$, where M is the number of mixed states on phase space, is a subjective representation, by an ignorant observer, of a collection of pure states $\phi_\alpha, \alpha \in [1, N]$, where N is the number of pure states within Ω_j. As the system evolve in time, the observer is only able to measure the coarse-grained flow, i.e. the jumping from one mixed state ψ_j to a different mixed state $\psi_i, i \neq j$.

It should be noted that due to the lack of perfect knowledge about all the relevant degrees of freedom, the observer is unable to predict a unique evolutionary path on phase space along which the system evolve.

4 Conservation of classical probability

Due to the ignorance of the observer, i.e. the observers inability to distinguish the set of pure states within any given coarse-grained region Ω_j, it is necessary to introduce the notion of probability on phase space. Let P_j be the probability that the system occupy the region Ω_j and let P_α be the probability that the system occupy the pure state ϕ_α within Ω_j. If the observer know with absolute certainty that the system occupy the mixed state ψ_j and not some other state $\psi_i, i \neq j \in [1, M]$, it is given that

$$P_i = 0, \forall i \neq j \in [1, M] \tag{4}$$

$$P_j \equiv \sum_{\alpha=1}^{N} P_\alpha = 1 \tag{5}$$

For continuous systems, the summation is replaced by an integral, i.e.

$$P_j \equiv \int_{\Omega_j} P_\alpha \; dV_\alpha = 1 \tag{6}$$

where $dV_\alpha = dq_\alpha dp_\alpha$ is the phase space volume of the pure state ϕ_α. If the knowledge possessed by the observer about the coarse-grained flow of the system is not lost over time, then the probability P_j is constant in time, i.e.

$$\frac{dP_j}{dt} = 0 \tag{7}$$

In other words, it is assumed that there is no loss of probability from Ω_j to any other coarse-grained region Ω_i, $i \neq j$. Written in terms of the probabilities P_α, the condition of no loss of coarse-grained knowledge become

$$\begin{aligned}\frac{dP_j}{dt} &= \frac{d}{dt}\int_{\Omega_j} P_\alpha \; dV_\alpha \\ &= \int_{\Omega_j}\left(\frac{dP_\alpha}{dt} + P_\alpha \; \nabla\cdot\mathbf{v}\right) dV_\alpha \\ &= 0\end{aligned} \tag{8}$$

Since this should hold independently on the size of Ω_j, the integrand must identically vanish, i.e.

$$\frac{dP_\alpha}{dt} + P_\alpha \; \nabla\cdot\mathbf{v} = 0 \tag{9}$$

This is the continuity equation for probability flow within any given coarse-grained region Ω_j. It is referred to as the Gibbs-Liouville equation for the probability distribution within Ω_j [3][4]. Given that information is conserved within Ω_j it is thus obtained that the probability distribution P_α is conserved, i.e. if $\nabla\cdot\mathbf{v} = 0$ then

$$\frac{dP_\alpha}{dt} = 0 \tag{10}$$

The continuity equation can be rewritten, showing that probability is locally conserved within Ω_j. Using the total time derivative of P_α, i.e.

$$\frac{dP_\alpha}{dt} = \frac{\partial P_\alpha}{\partial t} + \nabla P_\alpha\cdot\mathbf{v} \tag{11}$$

and the product rule

$$\nabla\cdot(P_\alpha \; \mathbf{v}) = \nabla P_\alpha\cdot\mathbf{v} + P_\alpha \; \nabla\cdot\mathbf{v} \tag{12}$$

the continuity equation become

$$\frac{\partial P_\alpha}{\partial t} + \nabla\cdot(P_\alpha \; \mathbf{v}) = 0 \tag{13}$$

The term $\nabla\cdot(P_\alpha \; \mathbf{v})$ represent the difference between the probability outflow and inflow for the pure state ϕ_α. If there is a net probability outflow from ϕ_α to the rest of Ω_j, i.e. if

$$\nabla\cdot(P_\alpha \; \mathbf{v}) > 0 \tag{14}$$

then the continuity equation give that the probability for ϕ_α decrease with time, i.e.

$$\frac{\partial P_\alpha}{\partial t} = -\nabla\cdot(P_\alpha \; \mathbf{v}) < 0 \tag{15}$$

If there is a net probability inflow to ϕ_α from the rest of Ω_j, i.e. if

$$\nabla\cdot(P_\alpha \; \mathbf{v}) < 0 \tag{16}$$

then the continuity equation give that the probability for ϕ_α increase with time, i.e.

$$\frac{\partial P_\alpha}{\partial t} = -\nabla\cdot(P_\alpha \; \mathbf{v}) > 0 \tag{17}$$

In terminology borrowed from quantum mechanics, systems which evolve in such a way that the probability distribution is conserved in time and with a total probability equal to unity are said to exhibit unitary evolution. The assumption of unitary evolution for quantum systems is a key ingredient in the formulation of quantum mechanics. In classical mechanics, the statement of unitary evolution is a direct consequence of the Gibbs-Liouville theorem, i.e that information is conserved.

5 Statistical equilibrium

Consider a system which has been closed for a sufficiently long period of time such that the density of pure states within Ω_j, and hence M, do not change with time. In this situation, the probability distribution P_α has no explicit dependence on time. The continuity equation is then reduced to

$$\nabla\cdot(P_\alpha \; \mathbf{v}) = 0 \tag{18}$$

This is the mathematical condition the system need to satisfy in order for it to be said to exist in statistical equilibrium. In other words, a system is in statistical equilibrium if there is no net probability flow on phase space.

6 The ergodic theorem

The incompressibility of the Hamiltonian flow imply that the time the system spend in any single pure state, before evolving to the next single pure state, is the same for all pure states. If this was not the case, the state points on phase space would lump together which would signify a violation of information conservation. This imply that over the course of a long period of time, the total time spent by the system in any given pure state is expected to be the same for all pure states. This expectation, which is due to a combination of the Gibbs-Liouville theorem and the law of large numbers, is in this article interpreted to be equivalent to the ergodic theorem of statistical mechanics [6][7][8]. Let n_α denote the number of times the system occupy the pure state ϕ_α. The total number of times, n, the system occupy the set of N pure states within Ω_j is then

$$n = \sum_{\alpha=1}^{N} n_\alpha \tag{19}$$

The ergodic theorem then say that over a long period of time, such that n is large, it is expected that the system occupy all pure states within Ω_j an equal number of times, i.e.

$$n_\alpha = n_\beta, \; \forall\beta\neq\alpha\in[1,N] \tag{20}$$

such that

$$n = N\cdot n_\alpha \tag{21}$$

7 Microcanonical probability distribution

It is now possible to define the notion of a probability P_α for the pure state ϕ_α of a closed system from the notion of a relative frequency[1],

$$P_\alpha \equiv \lim_{n\to\infty} \frac{n_\alpha}{n} = \frac{n_\alpha}{N \cdot n_\alpha} = \frac{1}{N} \tag{22}$$

Thus, all the pure states within Ω_j are equally probable. This imply that an observer has lost all information, down to the scale of V_{Ω_j}, about the system, since no distinctions can be made between the possible pure states within Ω_j. The uniform probability distribution given by equation 22 is commonly referred to as the microcanonical [3], or fundamental [9], probability distribution. Thus, given that the system satisfy the Gibbs-Liouville theorem, the microcanonical probability distribution satisfy the condition for statistical equilibrium.

8 Non-uniform probability distributions

There exist also non-uniform probability distributions. The non-uniformity arise due to interactions that the system has, or have had in the not too far distant past, with an environment. In other words, the system is, or was recently, not isolated. Due to the interaction with an environment, the density of states change with time. If the interaction is uniform on phase space, the density change uniformly on phase space. However, in general, this is not the case. An interaction, characterized by a potential energy, do depend on the specific values for the generalized coordinates. In that scenario, the density of states is a local function on phase space. This has the consequence that the total time spent by the system within any given region on phase space is not necessarily the same as within any other equally-sized region. In other words, the ergodic theorem appear to be violated. Thus, not only is the probability distribution non-uniform when there is a non-negligible net interaction with the environment, it can also change over time.

To put it differently, if there exist an interaction between the system and its environment, as seen from the perspective of an observer of the system, this imply that the observer possess knowledge, i.e. information, about the interaction. This information is used by the observer when assigning probabilities for the possible states of the system. The fact that the observer possess some amount of information mean necessarily that the probability distribution is non-uniform. It is only at statistical equilibrium, where all information is lost, that the observer assign a uniform probability distribution.

From the definition of probability in statistical equilibrium it is clear that the probability for any given pure state decrease as the number of pure states N increase, i.e. as the uncertainty volume increase. In non-equilibrium, where probabilities are not equal, it is the average probability which decrease as the uncertainty volume increase.

9 Ergodicity breaking

It should be emphasized that the apparent violation of the ergodic theorem is not of a fundamental character. It is only due to the fact that the degrees of freedom associated with the environment cannot be excluded when defining the degrees of freedom for the system. In other words, the environment should be included in the definition of the system. If that is done then there exist no environment and hence there cannot be any net transfer of energy and particles from, or to, the system. Then, this redefined system, which take into account all degrees of freedom, even those which the experimenter may think belong to an 'environment', do indeed conserve information and ergodicity is not broken. The probability distribution for the states of this redefined system is uniform, i.e. all mixed states for any given system, assuming the system has been defined such that no degrees of freedom are being forgotten, are equally probably. In most practical situations, however, there will always exist an environment to any system under study. The question is to what degree this environment interact with the system. The weaker the interaction, the weaker is the ergodicity breaking and the closer will the system come to a uniform probability distribution.

10 Is information conserved or lost?

At this stage, it is necessary to clarify the notions of information loss and information conservation to avoid confusion. The process of information loss, experienced by an observer, and the notion that information is conserved seem to contradict each other, making it impossible for a given system to reach statistical equilibrium over time unless it started there. The confusion arise due to a key difference between the statistical evolution of the system, as experienced by an ignorant observer, and the deterministic evolution of the system

[1] It must be emphasized that this relative frequency is not possible to obtain from a set of repetitive experimental measurements, since the observer, being ignorant, is not able to distinguish between the set of pure states.

as described by the classical laws of motion. The subtlety which must be emphasized is that the statement of classical determinism, represented mathematically by the Gibbs-Liouville theorem, is a postulate on the fundamental character of classical systems independent on whether there exist any observer or not, whereas the notion of information loss is observer dependent. The statement of information loss tries to capture the tendency of an observer to become more ignorant over time. In conclusion, statistical equilibrium is a statement on the amount of knowledge, or information, an observer of the system possess. In statistical equilibrium, the observer is unable to make any distinctions between the possible states of the system and therefore possess zero information. This do not imply that there are no fundamental distinctions between the states of the system. The observer is simply unaware of them. In fact, if the system do fundamentally conserve information, i.e. the distinctions between the possible states of the system exist for all times, the system is fundamentally never in statistical equilibrium, from the perspective of an observer whose knowledge of the degrees of freedom for the system is perfect and complete.

Thus, the fundamental question of how a system, which is initially not in statistical equilibrium, can evolve into a state of statistical equilibrium, should be modified as follows:

How can an observer of a given classical system, which flow on phase space according to the Hamilton equations, loose information about the system over time?

11 Entropy as a measure of uncertainty

A measure for the amount of ignorance possessed by the observer, i.e. the amount of uncertainty in the determination of the pure state of the system, should depend on the probability distribution $\{P_\alpha\}$. This measure is denoted by $S(\{P_\alpha\})$ and referred to as the entropy of the system. To obtain a specific form for the entropy as a function of the probability distribution, it is noted that this function should satisfy the following conditions.

i The entropy should be zero when the observer has complete knowledge about the evolution of the system. In other words, if the observer know with absolute certainty that the system occupy a specific state ϕ_α, such that $P_\alpha = 1$ and $P_\beta = 0 \ \forall \beta \neq \alpha$, the entropy must vanish.

ii The entropy should always be either zero or a positive number, i.e. $S \geq 0$.

iii The entropy should take a maximum value when the observer is maximally ignorant. This happen when the system is in statistical equilibrium. When all states are equally probable, it imply that the observer possess zero partial knowledge which can be used to distinguish between some of the features of the set of states. Thus,

$$P_\alpha = \frac{1}{N} \ \forall \alpha \in [1, N] \ \rightarrow S(\{P_\alpha\}) = S_{max} \tag{23}$$

iv The entropy should, in statistical equilibrium, be a continuously increasing function of the number of states N. In other words, when N increase, the uncertainty volume V_{Ω_j} increase continuously.

v The entropy should satisfy the following composition law,

$$S(\{P_\alpha\} \cdot \{P_\beta\}) = S(\{P_\alpha\}) + S(\{P_\beta\}) \tag{24}$$

This composition law is understood as follows. Let Ω_j be divided into two subregions Ω_j^α and Ω_j^β such that $V_{\Omega_j} = V_{\Omega_j^\alpha} + V_{\Omega_j^\beta}$. The states $\phi_\alpha, \alpha \in [1, N_\alpha]$, belong to Ω_j^α and the states $\phi_\beta, \beta \in [1, N_\beta]$, belong to Ω_j^β, where $N_\alpha + N_\beta = N$. The corresponding probability distributions, $\{P_\alpha\}_{\alpha=1}^{N_\alpha}$ and $\{P_\beta\}_{\beta=1}^{N_\beta}$, satisfy $\sum_{\alpha=1}^{N_\alpha} P_\alpha + \sum_{\beta=1}^{N_\beta} P_\beta = 1$ and, due to them being independent of each other, their product give the probability distribution associated with the region Ω_j, i.e. $P(\Omega_j) = \{P_\alpha\} \cdot \{P_\beta\}$. The composition law thus state that the total uncertainty within region Ω_j is the sum of the uncertainties associated with the subregions of Ω_j.

Conditions (i) and (v) suggest that the entropy has a logarithmic dependence on the probability distribution. Condition (ii) suggest that it is necessary to include an additional minus sign in the definition of the entropy. This is seen from the general definition of P_α, i.e.

$$\log P_\alpha = \lim_{n\rightarrow\infty} \log \left(\frac{n_\alpha}{n}\right) = \log n_\alpha - \lim_{n\rightarrow\infty} \log n < 0 \tag{25}$$

which, for a system in statistical equilibrium become

$$\log P_\alpha = \log \frac{1}{N} = \log 1 - \log N = -\log N < 0 \tag{26}$$

Since the entropy function should act as a measure for systems both in and out of statistical equilibrium, i.e. for both uniform and non-uniform probability distributions, it is required to take the statistical average of all logarithmic contributions to the entropy, i.e.

$$S(\{P_\alpha\}) \sim -\left(\frac{n_1}{n} \log P_1 + \cdots + \frac{n_N}{n} \log P_N\right) \tag{27}$$

$$\sim -\sum_{\alpha=1}^{N} \frac{n_\alpha}{n} \log P_\alpha \tag{28}$$

$$\sim -\sum_{\alpha=1}^{N} P_\alpha \log P_\alpha \tag{29}$$

This entropy function then satisfy conditions (iii) and (iv). With the proportionality constant identified with the Boltzmann constant k_B, it is referred to as the Gibbs entropy [3] and is, in the information theoretic language, identical to the Shannon entropy [10][11][12].

In conclusion, the entropy of a system measure the amount of uncertainty within the system, and it is given by the Gibbs formula

$$S(\{P_\alpha\}) = -k_B \sum_{\alpha=1}^{N} P_\alpha \log P_\alpha \tag{30}$$

In statistical equilibrium, the Gibbs entropy reduce to the Boltzmann entropy [6][13],

$$S = k_B \log M \tag{31}$$

It is important to emphasize that entropy is not a physical quantity in the same manner as e.g. energy. It is determined by the probability distribution of the states of the system and as such it is a quantity which depend both on the specifics of the system and of the ignorance of the observer.

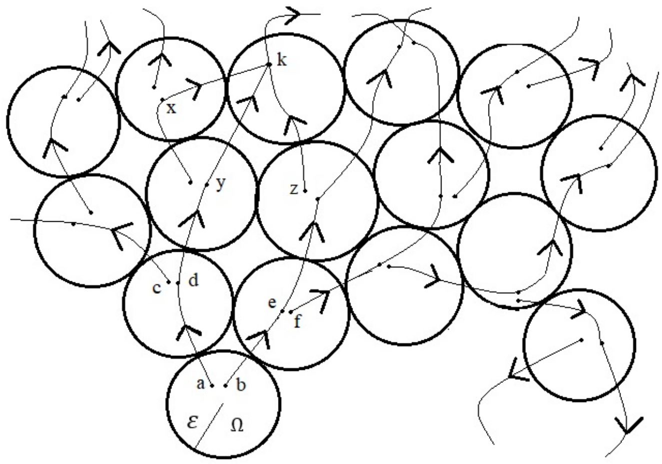

Fig. 1 Irreversible, entropy increasing, flow on phase space.

12 The second law of thermodynamics

If the state of a system is known with infinite precision at some given time, and if the laws of motion are known to infinite precision, then any earlier or later states of the system can be predicted with infinite precision. In such a deterministic situation, information about the system is never lost. However, in practical reality, the experimental precision by which the state can be determined is limited. Instead of knowing the initial conditions with infinite precision they are known to some degree of error, ϵ, on phase space. Therefore, the state of the system is only known to lie within a finite region, Ω, of radius ϵ and volume V_Ω. As the system evolve from the initial conditions it is not possible to predict the exact path on phase space. Any two neighboring states within Ω, e.g. a and b, see figure 1, might evolve differently over time. State a might evolve into either state c or state d. Due to the limited precision, it is impossible to say which state it evolve into. State b, on the other hand, might evolve into state e or state f. This process of diverging paths continue as time unfold. Therefore, the number of states in which the system might exist increase over time. In other words, the amount of uncertainty, i.e. the entropy, increase with time. Alternatively put, over time, any observer will continue to loose information about the system as a consequence of not knowing the initial conditions of the system with infinite precision. Actually, it is also possible for the entropy to decrease over time meaning that the observer has gained information about the system. This correspond to the situation when possible paths converge at some point. For example, the states x, y and z all converge into state k. The uncertainty of the system has thus decreased since there are now fewer possible states in which the system might exist. However, the probability that paths converge to a single state is much lower than the probability that they diverge to separate states. The reason for this is that the state k is merely one possible state out of a large number of possible states within volume V_Ω which x, y and z could have evolved into. Thus, overall, the observer loose information exponentially over time. Eventually, all information has been lost. The observer has become maximally ignorant. The entropy has reached its maximum value. At this stage, the system has reached statistical equilibrium where all states are equally probable since the observer is unable to make any distinctions between them. This tendency, of any given system, as viewed from an observer with limited knowledge of the initial conditions, to increase its entropy and evolve towards statistical equilibrium, is referred to as the second law of thermodynamics. In conclusion, it can be stated as follows:

Any given observer, whose knowledge about the initial conditions of any given system is limited, tend to loose information about the system at an exponential rate until there is none left.

It is important to emphasize that the apparent violation of determinism and reversibility, i.e. violation of the Gibbs-Liouville theorem, is not due to a fundamental character in the dynamical evolution of systems. The apparent irreversibility is only due to the ignorance of the observer.

13 The arrow of time

For an infinitely wise observer, who is able to determine the initial conditions and the laws of motion with infinite precision, the evolution of the system is completely reversible in time. The apparent unique direction in which time flow, i.e. toward the future, is merely a consequence of the fact that any given observer do not possess infinite knowledge. For such an observer, it is overwhelmingly more probable, in fact, exponentially more probable, that the system evolve in such a way that possible paths diverge on phase space. The diverging evolution define the direction, or arrow, of time as seen from the perspective of the observer. In the unlikely scenario that the possible paths converged at a quicker rate than they diverged, such that information on the average was gained, then the system would be observed to evolve backwards in time. Only then could a shattered glass of red wine be enjoyed.

Acknowledgement

The author would like to express gratitude to Anke Olsen for many inspiring discussions in our office at Bacchus.

References

1. W. Heisenberg "Über den anschaulichen inhalt der quantentheoretischen kinematik und mechanik," Zeitschrift für Physik, **43**, No.3-4, p.172-198 (1927).
2. W. Heisenberg "The actual content of quantum theoretical kinematics and mechanics," NASA Technical Memorandum, English translation, 35 pages (1983).
3. J.W. Gibbs "Elementary principles in statistical mechanics," Charles Scribner's sons, New York. (1902).
4. F. Bloch "Fundamentals of statistical mechanics: manuscript and notes of Felix Bloch," Imperial College Press and World Scientific Publishing, 3rd ed. (2000).
5. A. Henriksson "On the Gibbs-Liouville theorem in classical mechanics," arXiv:1905.06185v3 [cond-mat.stat-mech] (2019). https://arxiv.org/abs/1905.06185v3.
6. L.E. Boltzmann "Vorlesungen über Gastheorie, vol II," Verlag von Johann Ambrosius Barth, Leipzig (1898).
7. G.D. Birkhoff "Proof of the ergodic theorem," Proceedings of the National Academy of Sciences (U.S.), **17**, p.656-660 (1931).
8. J.v. Neumann "Physical applications of the ergodic hypothesis," Proceedings of the National Academy of Sciences (U.S.), **18**, p.263-266 (1932).
9. A.Ya. Khintchine "Fundamental laws of probability," Moscow State University, Moscow, 1927; 2nd ed. GTTI, Moscow (1932).
10. C.E. Shannon "A mathematical theory of communication," The Bell System Technical Journal, **27**, p.379-423, 623-656 (1948).
11. E.T. Jaynes "Information theory and statistical mechanics," Physical Review, **106**, No.4, p.620-630 (1957).
12. E.T. Jaynes "Information theory and statistical mechanics II," Physical Review, **108**, No.2, p.171-190 (1957).
13. L.E. Boltzmann "Vorlesungen über Gastheorie, vol I," Verlag von Johann Ambrosius Barth, Leipzig (1896).