Probability of Intrinsic Time-Arrow from Information Loss

Lajos Diósi

Research Institute for Particle and Nuclear Physics, P.O. Box 49,
1525 Budapest 114, Hungary

Abstract. Time-arrow $s = \pm$, intrinsic to a concrete physical system, is associated with the direction of information loss ΔI displayed by the random evolution of the given system. When the information loss tends to zero the intrinsic time-arrow becomes uncertain. We propose the heuristic relationship $1/[1+\exp(-s\Delta I)]$ for the probability of the intrinsic time-arrow. The main parts of the present work are trying to confirm this heuristic equation.

The probability of intrinsic time arrow is defined by Bayesian inference from the observed random process. From irreversible thermodynamic systems, the proposed heuristic probabilities follow via the Gallavotti-Cohen relations between time-reversed random processes. In order to explore the underlying microscopic mechanism, a trivial microscopic process is analyzed and an obvious discrepancy is identified. It can be resolved by quantum theory. The corresponding trivial quantum process will exactly confirm the proposed heuristic time-arrow probability.

1 Introduction

Both experiment and theory confirm that physical processes are time-reversal invariant in 'simple' systems. This invariance may eventually be lost if the system is chaotic, singular, of many degrees of freedom, or not isolated [1]. It seems plausible now that time-reversal asymmetry (irreversibility) is always accompanied by some information loss. Yet, little is known quantitatively. The present work discusses an elementary informatic mechanism of irreversibility. It leads to a simple analytic expression for the asymmetric probability of the two possible directions of time.

Suppose we use *reference-time* t to label the order of events but we leave open whether *physical-time* st is passing with increasing or decreasing t, according to the respective *time-arrow* $s = \pm$. We make no a priori (extrinsic) assignment for s. The ambiguity is to be resolved by analyzing irreversible physical processes. We consider informatic irreversibility in a sense that the Shannon information changes by ΔI along the process. We call the resulting a posteriori time-arrow *intrinsic*. It belongs to the given irreversible process. It would not exist in 'empty space' at all. In the spirit of the second law of thermodynamics, the physical entropy production $s\Delta I$ must be positive, hence the intrinsic time-arrow is unique:

$$s = \mathrm{sign}(\Delta I) \ . \tag{1}$$

L. Diósi, Probability of Intrinsic Time-Arrow from Information Loss, Lect. Notes Phys. **633**, 125–135 (2004)
http://www.springerlink.com/

This assignment is only valid if the magnitude $|\Delta I|$ is macroscopic which means that it is much bigger than 1 bit. If, however, the irreversibility is weak then we have to be content with a probabilistic intrinsic time-arrow. The main suggestion of our work is that this probabilistic time-arrow is a relevant concept and, furthermore, the probability $P(s)$ depends on the Shannon information change ΔI under rather general conditions. I will consider the following relationship:

$$P(s) = \frac{1}{1+e^{-s\Delta I}} \ . \tag{2}$$

If we change the sign of the reference time ($t \to -t$) then also the sign of the information loss will change ($\Delta I \to -\Delta I$). Hence the above expression is *covariant* against time-reversal of the reference frame. Asymptotically it yields the unique thermodynamic arrow (1) if the information loss $|\Delta I|$ is much greater than 1 bit. On the contrary, the two time-arrows become equally probable for a reversible process where ΔI is much less than 1 bit [2]. The suggested relationship is heuristic and lacks a general proof. It is intimately related to the fluctuation theorem [3, 4] proved for a particular class of irreversible processes [5]. On the other hand, it intends to reflect a fundamental meaning of the time-arrow in terms of information flow. I am going to prove that the relationship (2) follow from elementary statistical considerations provided we assume some further conditions to fulfill.

Section 2 presents the mathematical steps of Bayesian statistical inference adapted to the estimation of the time-arrow from the observed data. In Sect. 3 we discuss the inference from irreversible thermodynamic process, in Sect. 4 from microscopic process. The time-arrow is derived from quantum irreversibility in Sect. 5. The Appendix offers a short proof of the fluctuation theorem.

2 Bayesian Time-Arrow

Given a statistical system, let X denote a certain random process in a given interval of reference-time t. Let $\tilde{X}$ denote the time-reversal of X. Assume that from the principles of statistical physics we can calculate the probability $\mathcal{P}(X)$ in physical-time! We also introduce the probability distribution $\tilde{\mathcal{P}}(X)$ of the same random process seen from a reference frame with reversed time,

$$\tilde{\mathcal{P}}(X) \equiv \mathcal{P}(\tilde{X}) \ . \tag{3}$$

The conditional probability distribution of X takes the form

$$P(X|s) = \begin{cases} \mathcal{P}(X) & s = + \\ \tilde{\mathcal{P}}(X) & s = - \end{cases} \ , \tag{4}$$

where s is the a priori time-arrow. Prior to the irreversible process X, the distribution of s is symmetric: $P_0(s) = 1/2$. Hence the joint distribution of X and s is the following,

$$P(X, s) = P(X|s)P_0(s) = \frac{1}{2} \begin{cases} \mathcal{P}(X) & s = + \\ \tilde{\mathcal{P}}(X) & s = - \end{cases} \ . \tag{5}$$

According to the Bayes rule, the conditional a posteriori distribution of the time-arrow reads

$$P(s|X) = \frac{1}{\mathcal{P}(X)+\tilde{\mathcal{P}}(X)} \times \begin{cases} \mathcal{P}(X) & s=+ \\ \tilde{\mathcal{P}}(X) & s=- \end{cases} , \tag{6}$$

which can be cast into the following covariant form:

$$P(s|X) = \frac{1}{1+e^{-sD(X)}} , \tag{7}$$

where

$$D(X) = -\log\frac{\tilde{\mathcal{P}}(X)}{\mathcal{P}(X)} . \tag{8}$$

This Bayesian estimate means that if 1) we know the a priori distribution $\mathcal{P}(X)$ of the random process X in physical-time but 2) experimentally we observe either X or $\tilde{X}$ with equal probability since we have no a priori information regarding the relationship of our reference-time to the physical-time then 3) learning X in the reference-time will lead us to the Bayesian probabilistic estimate $P(s)$ of the times-arrow.

Let us calculate the mean fidelity of the estimated time-arrow: from (4-8) we shall obtain the following closed form:

$$F \equiv \sum_X P(+|X)P(X|+) = \left\langle \frac{1}{1+e^{-D(X)}} \right\rangle_{\mathcal{P}} . \tag{9}$$

The expectation value should refer to $\mathcal{P}(X)$ which is the distribution in the physical frame. We can easily derive an ultimate covariant expression of the average Bayesian estimate,

$$P(s) = \left\langle \frac{1}{1+e^{-sD(X)}} \right\rangle , \tag{10}$$

where the average refers already to the observed statistics and the form is valid in time-reversed reference frames as well.

3 Thermodynamic Case

Let X be a coarse-grained macroscopic random process in a given statistical system in the period $[-T,+T] \equiv [t_1,t_2]$ and let $\tilde{X}$ be the same process seen from the time-reversed reference frame,

$$\begin{aligned} X &= \{X(t);\ t_1 \le t \le t_2\} , \\ \tilde{X} &= \{X(-t); t_1 \le t \le t_2\} . \end{aligned} \tag{11}$$

Typically, X can be an irreversible thermodynamic process $X(t)$. Assume that we know the irreversible entropy $\Delta I(X)$ produced by the process X. Obviously, the time-reversed process 'produces' the same entropy with the opposite sign,

$$\Delta I(\tilde{X}) = -\Delta I(X) . \tag{12}$$

Let us introduce the following conditional distributions:

$$P(X|\xi) = \mathcal{P}(X)/P_1(\xi) \ ,$$
$$P(\tilde{X}|\eta) = \tilde{\mathcal{P}}(X)/P_2(\eta) \ , \tag{13}$$

where $P_1(\xi), P_2(\eta)$ are the probability distributions of the extreme values $\xi = X(t_1)$ and $\eta = X(t_2)$, respectively. In the Appendix the reader finds an elementary proof of the fluctuation theorem [3–5] encoding the asymmetry of the time-reversal $X \leftrightarrow \tilde{X}$ into covariant equation,

$$P(\tilde{X}|\eta) = e^{-\Delta I(X)} P(X|\xi) \ . \tag{14}$$

Accordingly, the violation of the time-reversal symmetry is exponentially increasing with the magnitude $|\Delta I|$ of the irreversible entropy. We are going to show that, via the Bayesian statistics of Sect. 2, the relationship (14) reproduces the heuristic probabilities (2) for the thermodynamic time-arrow.

Let us express the r.h.s. of (8) from (13,14),

$$D(X) = \Delta I(X) - \log \tfrac{P_2(\eta)}{P_1(\xi)} \ . \tag{15}$$

For long enough periods, the r.h.s. is dominated by the information loss $\Delta I(X)$, the second (boundary) term can be ignored (cf. [5]). In this limit we can write the covariant Bayesian estimate (7) into this form,

$$P(s|X) = \tfrac{1}{1+e^{-s\Delta I(X)}} \ , \tag{16}$$

which on average leads to the covariant distribution

$$P(s) = \left\langle \tfrac{1}{1+e^{-s\Delta I(X)}} \right\rangle \ . \tag{17}$$

Finally, this yields the heuristic form (2) provided we can ignore the statistical fluctuations of the entropy production around its expectation value $\Delta I = \langle I(X) \rangle$. This is justified for common macroscopically irreversible processes where $|\Delta I| \gg 1$.

4 Microscopic Case

Let us consider a statistical ensemble of $n \gg 1$ independent $d-$state systems characterized by the probability distribution $\rho^i, i = 1, 2, \dots, d$. Let X be an abstract random process as trivial as the transition from an initial microscopic ensemble state ξ into a final one η, the time-reversed process $\tilde{X}$ will be the opposite transition,

$$X = (\xi, \eta) \ ,$$
$$\tilde{X} = (\eta, \xi) \ . \tag{18}$$

Let ρ_1^i and ρ_2^i be the probability distributions of the systems within the ensembles ξ and η, respectively. Then the change of Shannon information along the process X reads:

$$\Delta I \equiv nI_2 - nI_1 = -n\sum_{i=1}^{d}\rho_2^i \log\rho_2^i + n\sum_{i=1}^{d}\rho_1^i \log\rho_1^i \ . \tag{19}$$

The process X is irreversible if $\Delta I \neq 0$ and we should assign the time-arrow s so that $s\Delta I$ be positive (1). The point is that the two samples ξ and η may, by chance, not realize the asymmetry especially when the shapes of their probability distributions ρ_1^i and ρ_2^i do not much differ from each other.

Let us characterize the two constituting configurations of $X = (\xi, \eta)$ by the multiplicities n_1^i and n_2^i,

$$\begin{aligned} \xi &= (n_1^i; i = 1, 2, \ldots, d) \ , \\ \eta &= (n_2^i; i = 1, 2, \ldots, d) \ , \end{aligned} \tag{20}$$

which follow independent multinomial distributions with the respective mean values

$$\begin{aligned} \langle n_1^i \rangle &= n\rho_1^i \ , \\ \langle n_2^i \rangle &= n\rho_2^i \ . \end{aligned} \tag{21}$$

For large n we can approximate the multinomial distributions by Gaussian functions,

$$\begin{aligned} \mathcal{P}(\xi,\eta) &= C\exp\left(-\sum_{i=1}^{d}\tfrac{[n_1^i - n\rho_1^i]^2}{2n\rho_1^i} - \sum_{i=1}^{d}\tfrac{[n_2^i - n\rho_2^i]^2}{2n\rho_2^i}\right) \ , \\ \mathcal{P}(\eta,\xi) &= C\exp\left(-\sum_{i=1}^{d}\tfrac{[n_2^i - n\rho_1^i]^2}{2n\rho_1^i} - \sum_{i=1}^{d}\tfrac{[n_1^i - n\rho_2^i]^2}{2n\rho_2^i}\right) \ . \end{aligned} \tag{22}$$

We substitute these expressions into (8) to calculate $D(\xi, \eta)$, then we calculate the mean value,

$$D = -\tfrac{n}{2}\sum_{i=1}^{d}\left((\rho_2^i)^2 - (\rho_1^i)^2\right)\left(\tfrac{1}{\rho_2^i} - \tfrac{1}{\rho_1^i}\right) \ . \tag{23}$$

Suppose that $D(\xi, \eta)$ is, for very large n, dominated by the mean value D and fluctuations will thus be ignored. Hence the average Bayes estimate (10) reads

$$P(s) = \tfrac{1}{1+e^{-sD}} \ . \tag{24}$$

This could become equivalent with our heuristic proposal provided $D = \Delta I$ which is apparently not true in general. I was looking for further conditions at least to achieve the asymptotic equivalence of D and ΔI. I concluded to the following elementary assumptions. First, the shapes ρ_1^i and ρ_2^i must be close to each other so that the lowest nontrivial order in $\Delta\rho^i = \rho_2^i - \rho_1^i$ will be sufficient. Second, the statistics of *either* ξ *or* η must be totally random. This sets the a priori time-arrow for $s = -$ or $s = +$, respectively. For concreteness, I consider the case $s = +$ and adopt flat distribution for η [6],

$$\rho_2^i = \tfrac{1}{d} \ . \tag{25}$$

This second assumption is a necessary one, otherwise ΔI contains a linear term in $\Delta\rho^i$ while D does not. From (19) and (23) the above two assumptions lead to the following results:

$$\Delta I = \tfrac{nd}{2} \sum_{i=1}^{d} (\Delta\rho^i)^2 \ , \tag{26}$$

and

$$D = nd \sum_{i=1}^{d} (\Delta\rho^i)^2 \ . \tag{27}$$

The result is surprising: D has come out twice the information loss.

Mathematically, D is the Kullback divergence between two neighboring ensembles ξ and η and it should asymptotically coincide with the information loss between them. The reason of the anomalous factor 2 is that we happened to use the Kullback divergence between the composite ensembles (ξ,η) and (η,ξ) instead of ξ and η. This gives a hint how the factor 2 would go away. It is interesting to note that the physical resolution has a typical quantum mechanical motivation. In microphysics it is conceptually impossible to observe the full quantity $X = (\xi,\eta)$. If, e.g., the time-arrow is positive $(s = +)$ then η is testable and ξ is not because its observation would significantly perturb the initial preparation. And vice versa, when $s = -$ then η is testable and ξ is not. Accordingly, we are going to change the concept of experimental data. In the concrete case, we forbid the observation of ξ. In this sense, we have to redefine the distribution of the observed quantities,

$$\begin{aligned} \mathcal{P}(\xi,\eta) &\to \mathcal{P}(\eta) \equiv \sum_{\xi} \mathcal{P}(\xi,\eta) \ , \\ \tilde{\mathcal{P}}(\xi,\eta) &\to \tilde{\mathcal{P}}(\eta) \equiv \sum_{\xi} \tilde{\mathcal{P}}(\xi,\eta) \ . \end{aligned} \tag{28}$$

Repeating the calculation of the Kullback divergence in the leading order, inserting the flat values (25) for ρ_2^i, D turns out to be half of the previous value (26). Thus in the given approximation we have obtained the identity

$$D = \Delta I \ , \tag{29}$$

and confirmed the heuristic relationship (2).

5 Quantum Case

Let us consider the statistical ensemble of $n \gg 1$ independent $d-$state quantum systems where each one has the same density matrix $\hat{\rho}$. Let X be an abstract random process as trivial as the transition from an initial ensemble $\hat{\xi}$ into a final one $\hat{\eta}$, the time-reversed process $\tilde{X}$ will be the opposite transition,

$$\begin{aligned} X &= (\hat{\xi}, \hat{\eta}) \ , \\ \tilde{X} &= (\hat{\eta}, \hat{\xi}) \ , \end{aligned} \tag{30}$$

where

$$\begin{aligned} \hat{\xi} &= \hat{\rho}_1 \otimes \hat{\rho}_1 \otimes \ldots \otimes \hat{\rho}_1 \equiv \hat{\rho}_1^{\otimes n} \ , \\ \hat{\eta} &= \hat{\rho}_2 \otimes \hat{\rho}_2 \otimes \ldots \otimes \hat{\rho}_2 \equiv \hat{\rho}_2^{\otimes n} \ , \end{aligned} \tag{31}$$

if $\hat{\rho}_1$ and $\hat{\rho}_2$ stand for the density matrices of the systems within the ensembles $\hat{\xi}$ and $\hat{\eta}$, respectively. The change of von Neumann information during the process X reads

$$\Delta I \equiv nI_2 - nI_1 = -n\mathrm{Tr}\left(\hat{\rho}_2 \log \hat{\rho}_2\right) + n\mathrm{Tr}\left(\hat{\rho}_1 \log \hat{\rho}_1\right) \ . \tag{32}$$

The process X is irreversible if $\Delta I \neq 0$ and we should assign the time-arrow s so that $s\Delta I$ be positive (1). In order to ΔI have a definite sign the two ensembles $\hat{\xi}$ and $\hat{\eta}$ should display experimentally significant asymmetry.

Quantum theory says that if the reference-time is the physical time ($s = +$) then we cannot test the ensemble $\hat{\xi}$ but the ensemble $\hat{\eta}$. And in the opposite case ($s = -$) the ensemble $\hat{\xi}$ is testable and $\hat{\eta}$ is not. We see that the estimation of the time-arrow s boils down to the statistical decision whether the actually observed ensemble is $\hat{\xi} = \hat{\rho}_1^{\otimes n}$ or $\hat{\eta} = \hat{\rho}_2^{\otimes n}$ whereas both alternatives have equal a priori likelihoods.

We can mechanically follow the Bayes method of the previous chapters. Note, however, the typical quantum informatic arguments: this is the way I approached the issue originally.

The two collective states (31) reside in a Hilbert space of dimension d^n. According to the quantum counterpart of Shannon's code theory [7], in the large n limit such collective states become asymptotically equivalent with totally random states restricted for given subspaces. Our states (31) become random states in subspaces $\hat{E}_1$ and $\hat{E}_2$,

$$\begin{aligned} \hat{\xi} &= \hat{\rho}_1^{\otimes n} \sim e^{-nI_1} \hat{E}_1 \ , \\ \hat{\eta} &= \hat{\rho}_2^{\otimes n} \sim e^{-nI_2} \hat{E}_2 \ , \end{aligned} \tag{33}$$

where $\hat{E}_1$ and $\hat{E}_2$ are Hermitian projectors of dimensions e^{nI_1} and e^{nI_2}, respectively. The dimensions depend on the von Neumann entropies. We are interested in the situations where the experimental distinguishability of the above two ensembles would exclusively depend on the difference $\Delta I = nI_2 - nI_1$ of the informations. This is obviously not true in general because the distinguishability will depend e.g. on the overlap $\mathrm{Tr}(\hat{E}_1\hat{E}_2)$. Nonetheless, the asymptotic forms (33) suggest simple conditions to achieve our goal. Suppose that one of the ensembles, say $\hat{\eta}$, is of minimal information,

$$\hat{\rho}_2 = \frac{\hat{1}}{d} \ , \tag{34}$$

which means that $I_2 = \log d$ and $\hat{E}_2 = \hat{1}^{\otimes n}$. The ensemble $\hat{\eta}$ is totally random over the whole collective Hilbert space of dimension d^n. The overlap between $\hat{E}_1$ and $\hat{E}_2$ becomes trivial. The information loss is always positive and the true time-arrow is thus $s = +$. But we have to find it by deciding whether we have tested the ensemble $\hat{\xi}$ or $\hat{\eta}$ which are of equal a priori likelihoods.

Now the experimental distinguishability of $\hat{\xi}$ and $\hat{\eta}$ is already trivial. All we have to do is to define $\hat{E}_1$ as observable and to observe it! If the tested ensemble is $\hat{\xi}$ itself then we get 1 with certainty since $\mathrm{Tr}(\hat{E}_1\hat{\xi}) = 1$. If the observed ensemble is the fully random $\hat{\eta}$ then we get 1 with probability $\mathrm{Tr}(\hat{E}_1\hat{\eta}) = e^{-\Delta I}$ and we get 0 with the complementary probability. As we see, the complete experimental statistics is determined by the information loss ΔI.

Let us turn to the Bayes method of Sect. 2 to estimate the time-arrow s. As we suggested above, the observed data is the value $E_1 = \{0, 1\}$ of the quantum observable $\hat{E}_1$. The probability $\mathcal{P}(E_1)$ stands for its distribution in the reference time with time-arrow $s = +$ and $\tilde{\mathcal{P}}(E_1)$ stands for its distribution in the reversed time $s = -$. In the preceding paragraph we established their values,

$$\begin{aligned} \mathcal{P}(E_1) &= E_1 e^{-\Delta I} + (1 - E_1)(1 - e^{-\Delta I}) \ , \\ \tilde{\mathcal{P}}(E_1) &= E_1 \ . \end{aligned} \tag{35}$$

Applying the steps of Sect. 2 mechanically, first we write (8) into this form:

$$e^{-D(E_1)} = \frac{\tilde{\mathcal{P}}(E_1)}{\mathcal{P}(E_1)} \ , \tag{36}$$

which is then substituted into the expression (9) of the mean fidelity, yielding

$$F = \left\langle \frac{1}{1+\tilde{\mathcal{P}}(E_1)/\mathcal{P}(E_1)} \right\rangle_{\mathcal{P}} = \frac{1}{1+e^{-\Delta I}} \tag{37}$$

We have used (35) to calculate the average. The result implies exactly the form (2) for the probability of intrinsic time-arrow in function of the information loss.

6 Concluding Remarks

I proposed a heuristic probability distribution (2) for the time-arrow intrinsic to a given irreversible process. The proposed probability is solely a function of the information loss ΔI. The idea itself comes from the phenomenological fluctuation theorem. Indeed, the concrete form of my proposal can easily be confirmed for the intrinsic time-arrow of standard irreversible processes, at least in the limit of macroscopic entropy production $\Delta I \gg 1$. My basic goal, however, was the construction of whatever trivial microscopic process which could underly the proposed dependence on ΔI. I analyze the irreversible process of the simplest possible structure in classical and quantum versions. The quantum version confirmed the proposed probabilistic time-arrow. Let me summarize this central result.

1) Suppose we know that (in physical time) a quantum ensemble $\hat{\xi}$ of $n \gg 1$ identical systems of given (also known to us) state transforms into an ensemble $\hat{\eta}$ of n totally random systems. 2) Suppose we do not know at all whether our reference-time is the physical-time or not, and whether the 'resulting ensemble' of the above process has been $\hat{\eta}$ or $\hat{\xi}$. 3) We test the 'resulting ensemble' and Bayesian inference will give us the time-arrow with fidelity

$$F = \frac{1}{1+e^{-|\Delta I|}} \ .$$

An infite number of conceptual issues could be raised against the presented ideas. I mention and discuss only two. First, the assignment of a non-trivial intrinsic time-arrow to a local irreversible process is a speculation. Nature might retain the same universal time-arrow for the whole Universe independently of the measure or direction of local information flows. Yet, we do not know if Nature is that conservative indeed. We learned from Einstein that Nature delegates the issues of local geometry to local physical systems. I adopted the hypothesis that this happens with time-arrow as well. Second, the proposed confirmation of the time-arrow probability includes Bayesian inference. Many would say that inference is subjective. The obtained probability is also subjective. Nonetheless, famous arguments using inference have been used earlier to confirm objective statistics of quantized fields [11]. It is, furthermore, a common knowledge that the maximum-likelihood inference of the intensive thermodynamic parameters confirms their true equilibrium fluctuations in Gibbs ensembles.

The present work is an attempt to find universal expressions for the hypothetic intrinsic time-arrow. There is a hint of the information loss to play the key role. This does not mean that we can already claim an experimental significance which should, of course, be inevitable after all. But theory of intrinsic time opens a series of natural questions to study in the future and there is apparently a promise of further analytic results.

Acknowledgments

I am grateful to Hans-Thomas Elze and to the sponsors of the workshop "DICE 2002" for the invitation to talk in Piombino and to contribute to this volume. My research was also supported by the Hungarian OTKA Grant No. 032640.

Appendix

Let $X(t)$ denote a thermodynamic variable of equilibrium value $\bar{X}$, where λ is the relaxation rate, and γ is the Onsager kinetic coefficient. The time-dependent fluctuations of $X(t)$ are governed by the phenomenological Langevin equation,

$$\tfrac{dX(t)}{dt} = -\lambda(X(t)-\bar{X}) + \sqrt{2\gamma}\, w(t) \tag{38}$$

with the standard white-noise $w(t)$. The expression

$$\tfrac{dI}{dt} = \tfrac{\lambda}{\gamma}(\bar{X}-X)\tfrac{dX}{dt} \tag{39}$$

will be the local rate of irreversible entropy production (information loss) along the process $X(t)$ (see, e.g., in Landau-Lifshitz [8], or in [9]). According to Onsager and Machlup [10], the conditional probability distribution of the process $X=\{X(t); -T \le t \le T\}$ at fixed initial value $\xi = X(-T)$ and for equilibrium value $\bar{X}$ takes this functional Gaussian form,

$$P(X|\xi;\bar{X}) = \exp\Big(-\frac{1}{4\gamma}\int_{-T}^{T}\Big[\tfrac{dX(t)}{dt} + \lambda(X(t)-\bar{X})\Big]^2 dt\Big)\ . \tag{40}$$

We shall consider *driven* thermodynamic processes which can be described by the (38-40) with time-dependent equilibrium values $\{\bar{X}(t); -T \le t \le T\}$. For convenience of forthcoming calculations let us write down the distribution functional of the driven process,

$$P(X|\xi;\bar{X}) = \exp\Big(-\frac{1}{4\gamma}\int_{-T}^{T}\Big[\tfrac{dX(t)}{dt} + \lambda(X(t)-\bar{X}(t))\Big]^2 dt\Big)\ . \tag{41}$$

Obviously the above equations assume physical time t. Let us express the conditional distribution of the time-reversed process $\tilde{X}$ starting from $\tilde{X}(-T)=\eta$, driven by the time-reversed function $\tilde{\bar{X}}$. Namely, we replace $X, \xi, \bar{X}$ in (41) by $\tilde{X}, \eta, \tilde{\bar{X}}$, respectively,

$$P(\tilde{X}|\eta;\tilde{\bar{X}}) = \exp\Big(-\frac{1}{4\gamma}\int_{-T}^{T}\Big[\tfrac{d\tilde{X}(t)}{dt} + \lambda(\tilde{X}(t)-\tilde{\bar{X}}(t))\Big]^2 dt\Big)\ . \tag{42}$$

Now we change the variable t in the integrand for $-t$ and insert the relations

$$\begin{aligned}\tilde{X}(t) &\equiv X(-t)\ ,\\ \tilde{\bar{X}}(t) &\equiv \bar{X}(-t)\ ,\end{aligned} \tag{43}$$

leading to

$$P(\tilde{X}|\eta;\tilde{\bar{X}}) = \exp\Big(-\frac{1}{4\gamma}\int_{-T}^{T}\Big[\tfrac{dX(t)}{dt} - \lambda(X(t)-\bar{X}(t))\Big]^2 dt\Big)\ . \tag{44}$$

(Recall that this expression would be the conditional distribution of the process had we observed it in the time-reversed frame.) The logarithm of the physical distribution (41) over the time-reversed one (44) will result in a remarkable expression,

$$\log \frac{P(X|\xi;\bar{X})}{P(\tilde{X}|\eta;\tilde{\bar{X}})} = \frac{\lambda}{\gamma} \int_{-T}^{T} (\bar{X}(t) - X(t)) dX(t) \ . \tag{45}$$

It follows from (39) that the r.h.s. is equal to the total entropy production (information-loss) of the driven process,

$$\Delta I(X;\bar{X}) = \frac{\lambda}{\gamma} \int_{-T}^{T} (\bar{X}(t) - X(t)) dX(t) \ . \tag{46}$$

This and the preceding equation yield the fluctuation theorem [3–5],

$$P(\tilde{X}|\eta;\tilde{\bar{X}}) = e^{-\Delta I(X;\bar{X})} P(X|\xi;\bar{X}) \ . \tag{47}$$

References

1. H.D. Zeh: *The Direction of Time* (Springer Verlag, Berlin 1989)
2. Especially in thermodynamic context, the Shannon and von Neumann information (loss) will be deliberately identified with the entropy (production). While information is the fundamental quantity in the present work, I use Boltzmann units of entropy. In Shannon units (bits) the central equation (2) would read:

$$P(s) = \frac{1}{1+2^{-s\Delta I}}$$

3. D.J. Evans, E.G.D. Cohen, and G.P. Morriss: Phys. Rev. Lett. **71**, 2401 (1993)
4. G. Gallavotti and E.G.D. Cohen: Phys. Rev. Lett. **74**, 2694 (1995)
5. C. Maes, F. Redig, and A. Van Moffaert: J. Math. Phys. **41**, 1528 (2000)
6. The choice only means that, for the sake of simplicity, the forthcoming calculations are done in the physical reference time. The covariance for $t \to -t$ will only be restored on the final equations
7. B. Schumacher: Phys. Rev. A **51**, 2738 (1995)
8. L.D. Landau and E.M. Lifshitz: *Statistical Physics* (Clarendon, Oxford 1982)
9. L. Diósi, Katalin Kulacsy, B.Lukács, and A.Rácz: J. Chem. Phys. **105**, 11220 (1996)
10. L. Onsager and S. Machlup: Phys. Rev. **91**, 1505 (1953)
11. N. Bohr and L. Rosenfeld: Kgl. Danske Videnskab S. Mat. Fys. Medd. **12** 1 (1933)