Increase-in-entropy law

J. L. del Río-Correa*
Departamento de Física, Escuela Superior de Física y Matemáticas, Instituto Politécnico Nacional, Edificio 9, Unidad Profesional Zacatenco, Apartado Postal 75-702 México, Distrito Federal 07738, México

L. S. García-Colín†
Departamento de Física, Universidad Autónoma Metropolitana-Iztapalapa, Avenida Michoacán y Purísima s/n, Apartado Postal 55-524, 09340 México 13, Distrito Federal, México

(Received 14 May 1992; revised manuscript received 3 March 1993)

An increase-in-entropy law is derived from two very simple and direct premises: First, the probability distribution function describing the states of the system depends only on the macroscopic information that an observer can obtain by a measurement of the constraints imposed on the system. Second, each time a measurement is performed the observation generates information that depends only on the present state of the system, a fact which is explicitly determined by the mathematical properties of Ehrenfest's coarse-graining procedure, shown here to be identical to the action of Zwanzig's projection operator on the true N-body distribution function. Further, the act of observation is restricted to take place in time intervals that cannot be smaller than a collision time nor larger than the relaxation time to assure the full meaning of a statistical average. Under these conditions the process governing the time-dependent statistical phenomena is shown to be Markovian, although not necessarily linear in the variables associated with the observables of the system. The results are also placed within the context of previous work done along these lines.

PACS number(s): 05.20.−y, 05.70.Ln

I. INTRODUCTION

In a recent paper [1] we have studied the problem of defining the entropy function for nonequilibrium states of arbitrary closed systems. This was done using the condition requiring that the entropy, regarded as a function of state, depends only on information available to the observer who characterizes the nonequilibrium state he wishes to study. This task is accomplished by constructing a probability density function whose values are given not by all the microscopic information pertinent to the system but only through that information provided by an observation. As we showed in paper I, the mathematical characteristics of such a distribution function are defined by the action of Zwanzig's projection operator on the exact N-body distribution function. Moreover, the entropy associated with the states of the system is defined in terms of the Shannon-Jaynes information entropy $S_I(t)$ through the standard variational procedure which we need not repeat here.

The main property of $S_I(t)$ which we were able to exhibit in paper I is that for every instant of time t following the initial time from which the evolution of the system takes place, this entropy can never be smaller than the corresponding Gibbs entropy at that time. Furthermore we gave some evidence, not completely conclusive, that if repeated observations are performed on such system at discrete times $t_r > t_{r-1} > \cdots > t_0$ then

$$S_I(t_r) > S_I(t_{r-1}) > \cdots > S_I(t_0) , \tag{1}$$

which is a discrete law of increasing entropy such that for $t_r \to \infty$, $S_I \to S_G(\text{eq})$, the Gibbs equilibrium entropy. Nevertheless, Eq. (1) is still far from providing a general irreversibility criterion in the form of the standard H theorem of Boltzmann.

In this paper we wish to address ourselves to these aspects of the problem. First we want to show that indeed the inequality expressed in Eq. (1) holds true provided that $\tau_c < \|t_r - t_{r-1}\| < \tau_{\text{rel}}$ for arbitrary r, τ_c being the time of duration of a collision. This means that a macroscopic observation must await the occurrence of times much larger than a collision time before an averaged microscopic variable has a macroscopic significance. Of course it must be also smaller than τ_{rel}, the time it takes the variable to achieve its equilibrium value (or zero). This physical constraint imposed on measurements has a strong implication, namely, that the time-evolution operator governing the dynamics of the variables within this time scale is shown to satisfy the Chapman-Kolmogoroff equation. This means that although it may be nonlocal in the space spanned by the numerical values of the appropriate Γ space functions (**a** space in paper I) it is instantaneous in time. It is needless to emphasize the fact that for those favoring the ideal of regarding such variables as random variables, this statement simply reads that the process is Markovian. The consequences of this result and the relationship to other previous work done along these lines is also underlined.

The structure of the paper is as follows: In Sec. II we shall summarize the results of paper I rederived in a slightly different language more suitable to our purposes here. In Sec. III we give a more physical derivation of the inequality expressed in Eq. (1) by using the concept of

an observation on the physical system. Section IV is devoted to a study of the necessary and sufficient conditions that the time evolution operator $M(t,t')$, introduced in Eq. (31) has to obey in order that the process is a Markovian one. Since these conditions are basically dependent on characteristic time scales we undertake a careful discussion of their significance in Sec. V. Finally Sec. VI contains an entirely independent discussion by means of which we attempt to explain in terms of the concepts of information theory why the operator $M(t,t')$ will obey a semigroup property for Markovian processes. Emphasis is laid on the fact that this property when viewed in this context is intimately related to an inference principle, here called the observational compatibility principle fully discussed in the text.

II. GIBBS-EHRENFEST COARSE GRAINING

To make this paper self-contained we shall establish the relationship between the Gibbs-Ehrenfest concept of coarse graining and the projection-operator technique in a language which is advantageous for future purposes. Recall that the Gibbs-Ehrenfest entropy [2] is constructed by dividing the Γ space in cells of volume Ω_i and introducing the average value of the N-body distribution function $\rho(\Gamma,t)$ over the cell, as

$$P_i(t)=\frac{1}{\Omega_i}\left[\int_{\Omega_i} d\Gamma\,\rho(\Gamma,t)\right] . \tag{2}$$

The Gibbs-Ehrenfest entropy is then defined as

$$S_{\mathrm{GE}}(t)=-k_B\sum_i \Omega_i P_i(t)\ln P_i(t) , \tag{3}$$

where k_B is Boltzmann's constant. Denoting by $\Xi(\Gamma,t)$ the coarse-grained distribution function, namely,

$$\Xi(\Gamma,t)=\sum_i P_i(t)\Delta_i(\Gamma) , \tag{4}$$

where $\Delta_i(\Gamma)$ is the characteristic function of the ith cell, then [1,2]

$$S_{\mathrm{GE}}(t)=-k_B\int \Xi(\Gamma,t)\ln\Xi(\Gamma,t)d\Gamma . \tag{5}$$

Next, we choose to select the Ω_i cells in Γ space following van Kampen [3], if $\boldsymbol{\alpha}$ is a vector of macroscopic variables associated to a set of phase-space functions $\mathbf{A}(\Gamma,t)$,

$$\boldsymbol{\alpha}(\mathbf{t})=\int d\Gamma\,\rho(\Gamma,t)\,\mathbf{A}(\Gamma,0) , \tag{6}$$

a "measurement" is meant as the process whereby one can assign numerical values $\mathbf{a}$ to the set $\{\mathbf{A}(\Gamma,t)\}$ with a certain accuracy $\Delta\mathbf{a}$. Hence the Ω_i cell is given by

$$\Omega_i=\{\Gamma|\mathbf{a}^{(i)}\le \mathbf{A}(\Gamma,0)\le\mathbf{a}^{(i)}+\Delta\mathbf{a}\}$$

and therefore, with this selection

$$\Xi(\Gamma,t;\Delta\mathbf{a})=\sum_i P(\mathbf{a}^{(i)},t)D\left[\frac{\mathbf{A}(\Gamma,0)-\mathbf{a}^{(i)}}{\Delta\mathbf{a}}\right] , \tag{7}$$

where $D(x)$ is the characteristic function of region Ω_i [4]. Clearly $\Xi(\Gamma,t;\Delta\mathbf{a})$ depends on the accuracy of the measurement $\Delta\mathbf{a}$. Letting $\Delta\mathbf{a}\to\mathbf{0}$ which corresponds physically to an exact measurement of $\mathbf{A}(\Gamma,0)$, using the fact that

$$\lim_{\Delta\mathbf{a}\to 0}\frac{1}{\Delta\mathbf{a}}D\left[\frac{\mathbf{A}(\Gamma,0)-\mathbf{a}^{(i)}}{\Delta\mathbf{a}}\right]=\delta(\mathbf{A}(\Gamma,0)-\mathbf{a}^{(i)}) \tag{8}$$

and that $\mathbf{a}$ takes a continuous set of values when $\Delta\mathbf{a}\to 0$, we get that

$$\Xi(\Gamma,t)=\lim_{\Delta\mathbf{a}\to 0}\Xi(\Gamma,t;\Delta\mathbf{a})=[\Omega(\mathbf{A}(\Gamma,0))]^{-1}\int d\Gamma'\delta(\mathbf{A}(\Gamma',0)-\mathbf{A}(\Gamma,0))\rho(\Gamma',t) , \tag{9}$$

where $\Omega(\mathbf{a})$ denotes the extension of the $\mathbf{a}$ cell, namely,

$$\Omega(\mathbf{a})=\int d\Gamma\,\delta(\mathbf{A}(\Gamma,0)-\mathbf{a})) . \tag{10}$$

Using the ordinary scalar product for two arbitrary functions $A(\Gamma)$ and $B(\Gamma)$, namely,

$$(A,B)=\int d\Gamma\,A(\Gamma)B^*(\Gamma) , \tag{11}$$

and defining

$$G(\mathbf{a},t)=\delta(\mathbf{A}(\Gamma,t)-\mathbf{a}) \tag{12}$$

the coarse-grained distribution function given by Eq. (9) may be readily written as

$$\Xi(\Gamma,t)=\int d\mathbf{b}\frac{(\rho(\Gamma,t),G(\mathbf{b},0))}{(1,G(\mathbf{b},0))}G(\mathbf{b},0) . \tag{13}$$

Since Zwanzig's projection operator [15] is defined as

$$P_zF(\Gamma)=\int d\mathbf{b}\frac{(F(\Gamma),G(\mathbf{b},0))}{(1,G(\mathbf{b},0))}G(\mathbf{b},0) , \tag{14}$$

we see that

$$\Xi(\Gamma,t)=P_z\rho(\Gamma,t) \tag{15}$$

or, in other words, the mathematical operation which gives rise to the Gibbs-Ehrenfest coarse-grained distribution function is provided by P_z. Since the physical significance of P_z as the generalization of the microcanonical ensembles for nonequilibrium states has been discussed in many papers [5–7] we avoid unnecessary repetition.

The derivation of Eq. (15) as outlined above is based solely on the basic concepts of statistical mechanics. However, as we emphasized in paper I it may also be derived by specifying the information that the observer has on the system, namely, the numerical variables $\mathbf{a}$ associated to the set of observables represented by $\mathbf{A}=(A_1=H,A_2,\ldots)$. These observations are expressed in terms of the nonequilibrium distribution $g_{\mathrm{ne}}(\mathbf{a},t)d\mathbf{a}$ in terms of which, as shown in paper I,

$$\hat{\rho}(\Gamma,t)=\int d\mathbf{a}[\Omega(\mathbf{a})]^{-1}g_{\rm ne}(\mathbf{a},t)G(\mathbf{a},0) \tag{16}$$

is the distribution function which maximizes the Gibbs-Shannon information entropy $S_I(t)$,

$$S_I(t)=-k_B\int\hat{\rho}(\Gamma,t)\ln\hat{\rho}(\Gamma,t)d\Gamma \tag{17}$$

subject to the restrictions that

$$\begin{aligned}&\int\hat{\rho}(\Gamma,t)d\Gamma=1\ ,\\&\int\hat{\rho}(\Gamma,t)\delta(\mathbf{A}(\Gamma,0)-\mathbf{a})d\Gamma=g_{\rm ne}(\mathbf{a},t)\ .\end{aligned} \tag{18}$$

Therefore,

$$\Xi(\Gamma,t)=\hat{\rho}(\Gamma,t)=P_z\rho(\Gamma,t)\ , \tag{19}$$

which establishes the equality between the Gibbs-Ehrenfest coarse-grained distribution function and the Jaynes-Shannon probability distribution function, both being uniquely determined by Zwanzig's projection operator. This result will have a strong bearing on the ideas to be presented in the following sections.

III. THE LAW OF INCREASE OF ENTROPY

In this section we wish to address ourselves to one of the unanswered questions raised in paper I, namely, the properties that the information entropy defined in Eq. (17) exhibits as a function of time. Using Eq. (19) we immediately see that

$$S_I(t)=-k_B\int\rho(\Gamma,t)\ln P_z\rho(\Gamma,t)d\Gamma\ , \tag{20}$$

whereas the Gibbs entropy $S_G(t)$ is given by

$$S_G(t)=-k_B\int\rho(\Gamma,t)\ln\rho(\Gamma,t)d\Gamma\ . \tag{21}$$

As is well known, $dS_G(t)/dt=0$, a property which follows at once from the fact that $\rho(\Gamma,t)$ satisfies Liouville's equation.

If $f_1(\Gamma)$ and $f_2(\Gamma)$ are two arbitrary distribution functions we define the relative entropy as

$$S(f_2|f_1)=k_B\int d\Gamma f_1(\Gamma)\ln[f_1(\Gamma)/f_2(\Gamma)]\ . \tag{22}$$

Using the fact that for $x>0$, $\ln x\geq 1-x^{-1}$ and taking $x=[f_1/f_2]$ it is then clear that $S(f_2|f_1)\geq 0$, the equality being satisfied if $f_1=f_2$. It follows then at once that if we conveniently choose $f_1=\rho(\Gamma,t)$ since it is the natural choice for the weighting function and $f_2=\hat{\rho}(\Gamma,t)$, Eqs. (20), (21), and (22) imply that for every time $t>0$

$$S_I(t)>S_G(t)\ , \tag{23}$$

provided $\rho(\Gamma,t)\neq\hat{\rho}(\Gamma,t)$. This result already derived in paper I, simply states that the information entropy can never be less than the Gibbs entropy. Emphasis should be placed on the fact that this does not imply that $S_I(t)$ is an ever-increasing function in time ($S_G=$const), simply that $S_G(t)$ is a lower bound to the possible values the function $S_I(t)$ may take at any arbitrary instant of time.

We now show that $S_I(t)$ does in fact obey an increase-in-time property provided we understand clearly the physical meaning of observing the macrostate of the system at any time and are also clear about the fact that two consecutive observations cannot take place within arbitrarily small intervals of time. We again place emphasis on the fact that the dynamical variables characterizing the state of the system undergoing an irreversible process, which we have labeled $\{A_i(\Gamma)\}$, generates a corresponding set of numerical values $\{a_i+da_i\}$ upon measurement, characterized by the distribution function $g_{\rm ne}(\mathbf{a},t)d\mathbf{a}$. Thus, "observing" the states of the system at a finite sequence of times $t_1<t_2<\ldots<t_r$ implies that one establishes the corresponding sequence

$$g_{\rm ne}(\mathbf{a},t_1);g_{\rm ne}(\mathbf{a},t_2);\ldots;g_{\rm ne}(\mathbf{a},t_r)\ldots\ .$$

The finiteness of time intervals between two successive observations arises from the intrinsic concept of a macrostate, enough collisions among the particles must take place before the time averaging of a dynamical quantity is macroscopically meaningful. Also the interval cannot be arbitrarily large allowing the variables to relax to their equilibrium value (or zero). This means that if τ_c and $t_{\rm rel}$ are the collision and relaxation times, respectively,

$$\tau_c<\|t_r-t_{r-1}\|<\tau_{\rm rel}\quad\text{for all } r\ .$$

Under such conditions, for each $g_{\rm ne}(\mathbf{a},t_r)$ of the above sequence one can construct the corresponding coarse-grained density $\hat{\rho}(\Gamma,t_r)$ which, following the methodology described in the previous section, is given by

$$\hat{\rho}(\Gamma,t_r)=\int d\mathbf{a}[\Omega(\mathbf{a})]^{-1}g_{\rm ne}(\mathbf{a},t_r)G(\mathbf{a},0)\ . \tag{24}$$

From Eq. (19) it is clear that for all observations times t_r, we have that P_z leaves any $\hat{\rho}$ invariant, i.e.,

$$\hat{\rho}(\Gamma,t_r)=P_z\hat{\rho}(\Gamma,t_r)\ . \tag{25}$$

To establish the relationship between the different values of $\hat{\rho}$ obtained from the time sequence, we proceed as follows: Let $\rho_0(\Gamma,t)$ be the probability distribution function obtained from the solution to Liouville's equation when $\rho_0(\Gamma,t_0)$ is taken as the initial condition, so that

$$\rho_0(\Gamma,t)=U(t,t_0)\rho_0(\Gamma,t_0)\ , \tag{26}$$

where $U(t,t_0)=\exp[-iL(t-t_0)]$ and L is Liouville's operator (iL is defined by the Poisson bracket of the Hamiltonian and any phase function, i.e., $\{H,\ \}$). From Eq. (19), we know that at any time $t>t_0$, the distribution function which maximizes $S_I(t)$ is the projected part of the distribution given by Eq. (26). Calling this function $\hat{\rho}_0(\Gamma,t)$ we have that

$$\hat{\rho}_0(\Gamma,t)=P_z\rho_0(\Gamma,t)=P_zU(t,t_0)\rho_0(\Gamma,t_0)\ . \tag{27}$$

Moreover, this distribution function maximizes $S_I(t)$ subject to the constraints given by Eq. (18) and as pointed out is unique since it depends only on $g_{\rm ne}(\mathbf{a},t)$, the value for the distribution obtained when an observation is performed at time t. Therefore, $\hat{\rho}_0(\Gamma,t)(t>t_0)$ is the same for any initial condition imposed on Liouville's equation.

Now we establish an inequality between the information entropy and Gibbs entropy which will be very useful to establish the increase of entropy. Because of the fact

that $\hat{\rho}_0(\Gamma,t)\neq\rho_0(\Gamma,t)$, Eq. (23) implies the inequality $S_I(t)>S_G(t)$ and as $\rho_0(\Gamma,t_0)$ evolves according to Liouville's equation, we have that $S_G(t)=S_G(t_0)$, a fact implying that

$$S_I(t)>S_G(t_0)\ . \tag{28}$$

Up to this point we have not yet specified the way to select the initial distribution function, nor the time t at which an observation is performed on the system. We select any two times t_r and t_q in the measurement sequence, such that $t_r>t_q$, and choose $\rho_0(\Gamma,t_0)=\hat{\rho}(\Gamma,t_q)$ and $t=t_r$. With this selection, Eq. (27) simply states that

$$\begin{aligned}\hat{\rho}(\Gamma,t_r)&=P_z U(t_r,t_q)\hat{\rho}(\Gamma,t_q)\\&=P_z U(t_r,t_q)P_z\hat{\rho}(\Gamma,t_q)\ ,\end{aligned} \tag{29}$$

where we have used Eq. (25) in the second equality. Clearly Eq. (29) may be rewritten in the form

$$\hat{\rho}(\Gamma,t_r)=M(t_r,t_q)\hat{\rho}(\Gamma,t_q)\ , \tag{30}$$

where the operator M is defined by

$$M(t,t')=P_z U(t,t')P_z=M(t-t',0)\ , \tag{31}$$

as follows from the definition of U. Equation (30) provides a relation which clearly exhibits the fact that the dynamics governing the time evolution of $\hat{\rho}$ during the time interval that elapses between two observations of the system is no longer governed by Liouville's time-reversible operator $U(t)$.

Using now the fact that $t_0=t_q$ has been taken as the initial (first-measurement) time so that $\rho_0(\Gamma,t_0)=\hat{\rho}(\Gamma,t_q)$ and that S_G the Gibbs entropy is constant in time, from Eqs. (21) and (17) we have that

$$S_G(t_0)=S_G(t_q)=S_I(t_q)\ . \tag{32}$$

Since the second measurement is performed at time $t=t_r$, from Eqs. (28) and (32) we obtain that

$$S_I(t_r)>S_I(t_q)\ . \tag{33}$$

Since t_r and t_q represent two arbitrary times in the observational sequence, such that $t_r>t_q$, it follows at once that

$$S_I(t_r)>S_I(t_{r-1})>\ldots>S_I(t_1)>S_I(t_0)\ , \tag{34}$$

which is the sought after increasing-entropy law. Emphasis should be placed on the fact that Eq. (34) is by no means equivalent to the inequality $dS/dt>0$ nor to a demonstration of the second law of thermodynamics. It merely implies that if a sequence of measurements are performed on the nonequilibrium macrostates of an arbitrary closed system, characterized by a reduced set of dynamical variable $\{a_i\}$, $i=1,2,\ldots,r\ll N$, within time intervals obeying the restriction imposed by the inequality $\tau_c<\|t_r-t_{r-1}\|<\tau_{\rm rel}$, then the corresponding values of the information entropy as defined by Eq. (20) cannot decrease. What occurs to the behavior of $S_I(t)$ within each interval and even the question of whether or not we can define an entropy for every time t is unknown. This result is not surprising; Jaynes [8] more than 20 years ago illustrated this point by showing that for a real gas under specific conditions the inequality $dH/dt\leq 0$ can be openly violated without entering into a conflict either with the irreversibility concept or with the second law of thermodynamics. In the following section we will reinforce this view by showing that the sequence established in Eq. (34) is not general but characteristic only of processes which are local in time although they may be nonlocal in the macroscopic **a** variables.

IV. DYNAMICS OF THE COARSE-GRAINED DISTRIBUTION FUNCTION

In the study of the relationship between deterministic dynamics and probabilistic description of physical processes, we have previously obtained the increase-of-entropy law for slow processes [9]. In that paper we analyzed two types of slow processes; the first one corresponds to a stochastic process in which the state vector $\mathbf{A}(\Gamma,t)=\mathbf{a}(t)$ is nonstationary. In this case we showed that it obeys a couple of equations which resembles the Chapman-Kolmogoroff (CK) equation and that it is also quasi-Markovian; the second one corresponds to the case in which the state vector $\mathbf{A}(\Gamma,t)=\mathbf{a}(t)$ is stationary and thus obeys the Chapman-Kolmogoroff equation, where the process is Markovian. For these two cases it is possible to obtain generalized H theorems and applying these theorems to the Gibbs-Ehrenfest nonequilibrium entropy we obtained the increase-in-entropy law for slow process.

Notice the fact that a previous effort in this direction was made by Misra, Prigogine, and Courbage [10] who establishes the following theorem: If an operator W_t has the semigroup property

$$W_t W_s=W_{t+s}\quad \text{for } t,s\geq 0\ ,$$

and also satisfies that (a) W_t preserves positivity, (b) $W_t\rho_{\rm eq}=\rho_{\rm eq}$, (c) $W_t^\dagger$, its Hermitian conjugate is such that $W_t^\dagger\rho_{\rm eq}=\rho_{\rm eq}$, where $\rho_{\rm eq}$ is the microcanonical distribution, then W_t defines a Markov process, in such a form that if $\rho_{\rm in}$ denotes the initial distribution function, the time evolution of ρ under the Markov process is governed by $W_t^\dagger$, or

$$\rho_t=W_t^\dagger\rho_{\rm in}\ .$$

Indeed, when $W_t^\dagger$ satisfies that $\|W_t^\dagger(\rho-\rho_{\rm eq})\|^2=\|W_t^\dagger\rho-\rho_{\rm eq}\|^2\to 0$ for all normalized and non-negative distribution functions ρ, W_t defines then a strong Markov semigroup, and the associated stochastic process displays the irreversibility expressed in the increasing-entropy law.

Here we will follow a different approach. Indeed, in the previous section we showed, using information-theory arguments, that the increase-in-entropy law is obtained with the coarse-grained distribution $\hat{\rho}(\Gamma,t)$ which is governed by the $M(t,t')$ operator. Therefore in view of the results found in Ref. [9] it is natural to raise the question of whether $M(t,t')$ is a quasi-Markovian operator, a Markovian operator or an operator of another type. We will show in this section the conditions that must prevail in order that $M(t,t')$ generates a Markov process.

In order to study this question we need to introduce a

number of well-known concepts briefly summarized in what follows. We recall that the nonequilibrium binary distribution $g^{(2)}_{ne}(\mathbf{a},t;\mathbf{b},s)$ is defined as [11,12]

$$g^{(2)}_{ne}(\mathbf{a},t;\mathbf{b},s)=\int d\Gamma\,\rho(\Gamma,s)G(\mathbf{a},t-s)G(\mathbf{b},0)\ , \tag{35}$$

with $G(\mathbf{a},t)$ given by Eq. (12). Now we choose that $t>s$ and at these times the system was subject to observation so that from information theory we know from Eq. (24) that

$$\rho(\Gamma,s)=\hat{\rho}(\Gamma,s)=\int d\mathbf{c}\,g_{ne}(\mathbf{c},s)\rho(\mathbf{c},\Gamma,0)\ , \tag{36}$$

where $\rho(\mathbf{c},\Gamma,0)=[\Omega(\mathbf{c})]^{-1}G(\mathbf{c},0)$ is the microcanonical distribution in the $\mathbf{c}$ cell. From Eqs. (35) and (36) we get that

$$g^{(2)}_{ne}(\mathbf{a},t;\mathbf{b},s)=g_{ne}(\mathbf{b},s)[\Omega(\mathbf{b})]^{-1}\times(G(\mathbf{a},t-s),G(\mathbf{b},0))\ , \tag{37}$$

where use has been made of Eq. (11). The conditional probability $P_{ne}(\mathbf{a},t|\mathbf{b},s)$ for the two events, event $\mathbf{b}$ occurring at time s and event $\mathbf{a}$ occurring at time t, is defined as

$$P_{ne}(\mathbf{a},t|\mathbf{b},s)=[g_{ne}(\mathbf{b},s)]^{-1}g^{(2)}_{ne}(\mathbf{a},t;\mathbf{b},s)\ , \tag{38}$$

which according to Eq. (37) may be rewritten as

$$P_{ne}(\mathbf{a},t|\mathbf{b},s)=[\Omega(\mathbf{b})]^{-1}(G(\mathbf{a},t-s),G(\mathbf{b},0))=(G(\mathbf{a},t-s),\rho(\mathbf{b},\Gamma,0))\ , \tag{39}$$

the last equality following from the definition of the microcanonical distribution in the $\mathbf{b}$ cell. When we select $s=0$, from Eq. (39) we have that

$$P_{ne}(\mathbf{a},t|\mathbf{b},0)=(G(\mathbf{a},t),\rho(\mathbf{b},\Gamma,0))\ . \tag{40}$$

Using this fact in Eq. (39) we have that

$$P_{ne}(\mathbf{a},t|\mathbf{b},s)=P_{ne}(\mathbf{a},t-s|\mathbf{b},0)\ . \tag{41}$$

This equation shows that when the coarse-grained density is obtained from information-theory arguments, the stochastic processes $\mathbf{A}(\Gamma,t)=\mathbf{a}(t)$ is stationary, a result which is a direct consequence that $\hat{\rho}(\Gamma,s)=P_z\rho(\Gamma,s)$. Now we proceed to prove a central theorem about the M operator.

Theorem. The necessary and sufficient condition for the dynamics to generate a Markov process is that M has the semigroup property

$$M(\Delta t_1+\Delta t_2;0)=M(\Delta t_1,0)M(\Delta t_2,0)\ , \tag{42}$$

where the time intervals involved in Eq. (42) namely, Δt_1 and Δt_2, when they are different from zero, need to be larger compared with some time τ_c characteristic of the system. We begin showing that when the M operator has the property given by Eq. (42), we obtain the Chapman-Kolmogoroff equation, which defines a Markov process.

Taking into account that $G(\mathbf{a},\Delta t)=\exp[iL\Delta t]G(\mathbf{a},0)$, that $P_zG(\mathbf{a},0)=G(\mathbf{a},0)$ and that P_z and L are Hermitian operators, we can rewrite Eq. (40) in the form

$$P_{ne}(\mathbf{a},\Delta t|\mathbf{b},0)=(G(\mathbf{a},0),M(\Delta t,0)\rho(\mathbf{b},\Gamma,0))\ . \tag{43}$$

As by assumption, M satisfies, the semigroup property, we have that

$$P_{ne}(\mathbf{a},\Delta t|\mathbf{b},0)=(G(\mathbf{a},0),M(\Delta t_1,0)M(\Delta t_2,0)\rho(\mathbf{b},\Gamma,0))\ , \tag{44}$$

where $\Delta t_1+\Delta t_2=\Delta t$.

In the Appendix we prove that from the definition of the M operator given in Eq. (31), it follows that

$$M(\Delta t,0)\rho(\mathbf{b},\Gamma,0)=\int d\mathbf{c}\,P_{ne}(\mathbf{c},\Delta t|\mathbf{b},0)\rho(\mathbf{c},\Gamma,0)\ . \tag{45}$$

Substituting Eq. (45) twice in the right-hand side of Eq. (44) we find that

$$P_{ne}(\mathbf{a},\Delta t|\mathbf{b},0)=\int d\mathbf{c}\,P_{ne}(\mathbf{c},\Delta t_2|\mathbf{b},0)(G(\mathbf{a},0),M(\Delta t_1,0)\rho(\mathbf{c},\Gamma,0))=\int d\mathbf{c}\int d\mathbf{c}'P_{ne}(\mathbf{c},\Delta t_2|\mathbf{b},0)P_{ne}(\mathbf{c}',\Delta t_1|\mathbf{c},0)(G(\mathbf{a},0),\rho(\mathbf{c}',\Gamma,0))\ ,$$

which, using the fact that $(G(\mathbf{a},0),\rho(\mathbf{c}',\Gamma,0))=\delta(\mathbf{a}-\mathbf{c}')$, leads to the result

$$P_{ne}(\mathbf{a},\Delta t|\mathbf{b},0)=\int d\mathbf{c}\,P_{ne}(\mathbf{a},\Delta t_1|\mathbf{c},0)\times P_{ne}(\mathbf{c},\Delta t_2|\mathbf{b},0)\ . \tag{46a}$$

As $\Delta t=t_2-t_1=t_2-t_3+t_3-t_1$, where t_3 is an arbitrary intermediate time, selecting $\Delta t_1=t_2-t_3$, $\Delta t_2=t_3-t_1$, and using the stationarity property given by Eq. (41), we obtain that

$$P_{ne}(\mathbf{a},t_2|\mathbf{b},t_1)=\int d\mathbf{c}\,P_{ne}(\mathbf{a},t_2|\mathbf{c},t_3)\times P_{ne}(\mathbf{c},t_3|\mathbf{b},t_1)\ , \tag{46b}$$

which is the Chapman-Kolmogoroff equation.

We now proceed to prove that when the stochastic process satisfies the Chapman-Kolmogoroff equations, then the M operator satisfies the semigroup property.

Using the idempotent property of the projector, P_z, namely, that $P_z^2=P_z$, we have from Eq. (31) that $M(\Delta t,0)=M(\Delta t,0)P_z$. From Eq. (14) and the definition of $\rho(\mathbf{a},\Gamma,0)$, we obtain that the operator $M(\Delta t,0)$ is given by

$$M(\Delta t,0)=\int d\mathbf{a}(\ ,G(\mathbf{a},0))M(\Delta t,0)\rho(\mathbf{a},\Gamma,0)\ ,$$

where $(\ ,G(\mathbf{a},0))$ indicates that P_z acts upon any function defined in Γ space. Use of Eq. (45) into this equation leads to

$$M(\Delta t,0)=\int d\mathbf{b}\int d\mathbf{a}\,\rho(\mathbf{b},\Gamma,0)P_{\rm ne}(\mathbf{b},\Delta t|\mathbf{a},0)(\ ,G(\mathbf{a},0))\ . \tag{47}$$

Using the result, twice, we proceed to evaluate the product $M(\Delta t_1,0)M(\Delta t_2,0)$. Thus

$$M(\Delta t_1,0)M(\Delta t_2,0)=M(\Delta t_1,0)\int d\mathbf{b}\int d\mathbf{a}\,\rho(\mathbf{b},\Gamma,0)P_{\rm ne}(\mathbf{b},\Delta t_2|\mathbf{a},0)(\ ,G(\mathbf{a},0))\ ,$$

and using Eq. (47) again, we have that

$$M(\Delta t_1,0)M(\Delta t_2,0)=\int d\mathbf{b}\int d\mathbf{a}\left\{\int d\mathbf{c}\,P_{\rm ne}(\mathbf{c},\Delta t_1|\mathbf{b},0)\rho(\mathbf{c},\Gamma,0)\right\}P_{\rm ne}(\mathbf{b},\Delta t_2|\mathbf{a},0)(\ ,G(\mathbf{a},0))\ .$$

Interchanging the first and third integrals, we obtain

$$M(\Delta t_1,0)M(\Delta t_2,0)=\int d\mathbf{c}\int d\mathbf{a}\left[\int d\mathbf{b}\,P_{\rm ne}(\mathbf{c},\Delta t_1|\mathbf{b},0)P_{\rm ne}(\mathbf{b},\Delta t_2|\mathbf{a},0)\right]\rho(\mathbf{c},\Gamma,0)(\ ,G(\mathbf{a},0))\ . \tag{48}$$

Up to this point, the result is exact, namely, Eq. (48) is only a convenient form of rewriting the product $M(\Delta t_1,0)M(\Delta t_2,0)$. However, if we assume that the stochastic process is Markovian, the term inside the square brackets is equal to $P_{\rm ne}(\mathbf{c},\Delta t|\mathbf{a},0)$ with $\Delta t=\Delta t_1+\Delta t_2$, as follows from Eq. (46a), so that further using Eq. (47), we obtain that for a Markov process,

$$M(\Delta t_1,0)M(\Delta t_2,0)=M(\Delta t,0)\ ,$$

which is the semigroup property, as required.

Apparently, we never use in the demonstration the fact that the times involved need to be greater than some time τ_c, but this fact is implicitly assumed in the second part of the demonstration, because it is well known that to obtain the Chapman-Kolmogoroff equation from the Liouville equation it is necessary to require that the process is slow, a fact first noticed by Green in 1952 [13]. This implies that the time intervals which appear in the CK equation are greater than τ_c, where τ_c is the correlation time of the memory kernel which appears in the exact kinetic equation. Because of the importance of this feature of the work we shall discuss it in detail in the following section.

V. TIME SCALES FOR MARKOV PROCESSES

We discuss in this section the time scale in which the semigroup property of the M operator is valid; with this purpose in mind we briefly remind the reader some results about Markov processes.

In Ref. [9], it was shown that to obtain a Markov process from Liouville's equation, it is necessary to make three assumptions, namely, first the form of the initial phase-distribution function, second that the stochastic process $\mathbf{A}(\Gamma,t)=\mathbf{a}(t)$ is a slow one, which is valid only with a certain time scale, and third that the stochastic process is stationary in this time scale.

The first assumption concerning the form of the initial phase-distribution function is that it has the form

$$\rho(\Gamma,0)=\rho_{\rm eq}(\Gamma)\int d\mathbf{a}\frac{g_{\rm ne}(\mathbf{a},0)}{g_{\rm eq}(\mathbf{a})}G(\mathbf{a},0)\ . \tag{49}$$

In paper I we showed that the coarse-grained density $\hat{\rho}(\Gamma,t)$ obtained from information-theory arguments can be expressed as [see Eq. (11) in Ref. [1]]

$$\hat{\rho}(\Gamma,0)=\rho_{\rm eq}(\Gamma)\int d\mathbf{a}\frac{g_{\rm ne}(\mathbf{a},0)}{g_{\rm eq}(\mathbf{a})}G(\mathbf{a},0)\ . \tag{50}$$

It is then clear that $\hat{\rho}(\Gamma,0)$ is the initial distribution function given by Eq. (49), and therefore the first assumption about the initial distribution function can be justified using information-theory arguments.

From the solution of Liouville's equations with the initial conditions given by Eq. (49), the kinetic equation for $P_{\rm ne}(\mathbf{a},t|\mathbf{a}_0,0)$ can be obtained using the Zwanzig-Mori projection operator technique [5,7]. It is well known that one obtains two equivalent forms for the kinetic equation, the forward and backward equations [11], which have the following form:

$$\frac{\partial P_{\rm ne}(\mathbf{a},t|\mathbf{b},0)}{\partial t}=Z(\mathbf{a},t)P_{\rm ne}(\mathbf{a},t|\mathbf{b},0) \tag{51}$$

$$=\Lambda(\mathbf{b},t)P_{\rm ne}(\mathbf{a},t|\mathbf{b},0)\ , \tag{52}$$

where the Zwanzig-Mori operators are given by

$$Z(\mathbf{a},t)f(\mathbf{a},t)=\int d\mathbf{c}\,i\Omega(\mathbf{a},\mathbf{c})f(\mathbf{c},t)-\int_0^t ds\int d\mathbf{c}\,K(\mathbf{a},\mathbf{c},s)f(\mathbf{c},t-s)\ , \tag{53}$$

$$\Lambda(\mathbf{b},t)h(\mathbf{b},t)=\int d\mathbf{c}\,h(\mathbf{c},t)i\Omega(\mathbf{c},\mathbf{b})-\int_0^t ds\int d\mathbf{c}\,h(\mathbf{c},t-s)K(\mathbf{c},\mathbf{b},s)\ , \tag{54}$$

respectively. The explicit form to the quantities $i\Omega$ and K are irrelevant for this work, but are explicitly given in Eq. (33) and (34) in Ref. [11].

The second assumption to obtain the Markovian processes is that the stochastic process $\mathbf{A}(\Gamma,t)=\mathbf{a}(t)$ is a slow one. This property requires that the times in which we are interested in observing the system are sufficiently long compared with τ_c, the correlation time of the kernel

$K(\mathbf{a},\mathbf{c},t)$ which appears in the forward and backward equations (51) and (52). For long times compared with τ_c the kernel can be expressed in the form

$$K(\mathbf{a},\mathbf{c},t)=2K(\mathbf{a},\mathbf{c})\delta(t)\ , \tag{55}$$

where $K(\mathbf{a},\mathbf{c})=\int_0^\infty K(\mathbf{a},\mathbf{c},t)dt$.

When we introduce the slowness condition equation (55) into the kinetic equations (51) and (52) they take the form

$$\frac{\partial P_{\rm ne}(\mathbf{a},t|\mathbf{b},0)}{\partial t}=F(\mathbf{a})P_{\rm ne}(\mathbf{a},t|\mathbf{b},0) \tag{56}$$

$$=\Lambda(\mathbf{b})P_{\rm ne}(\mathbf{a},t|\mathbf{b},0)\ , \tag{57}$$

where $F(\mathbf{a})$ and $\Lambda(\mathbf{a})$ are operators such that one is the transpose of the other. Thus

$$\int d\mathbf{a}\,h(\mathbf{a},t)F(\mathbf{a})f(\mathbf{a},t)=\int d\mathbf{a}\,f(\mathbf{a},t)\Lambda(\mathbf{a})h(\mathbf{a},t) \tag{58}$$

and the $F(\mathbf{a})$ operator is given by

$$F(\mathbf{a})f(\mathbf{a},t)=\int d\mathbf{c}[i\Omega(\mathbf{a},\mathbf{c})-K(\mathbf{a},\mathbf{c})]f(\mathbf{c},t)\ . \tag{59}$$

When the process is slow, the solution to the forward equation (56) and to the backward equation (57) is given by

$$P_{\rm ne}(\mathbf{a},\Delta t|\mathbf{b},0)=e^{F(\mathbf{a})t}\delta(\mathbf{a}-\mathbf{b})\ , \tag{60}$$

$$P_{\rm ne}(\mathbf{a},\Delta t|\mathbf{b},0)=e^{\Lambda(\mathbf{b})t}\delta(\mathbf{a}-\mathbf{b})\ , \tag{61}$$

these equations being valid only for $t>\tau_c>0$. Using Eq. (60) we obtain that

$$P_{\rm ne}(\mathbf{a},\Delta t|\mathbf{b},0)=e^{F(\mathbf{a})\Delta t_1}P_{\rm ne}(\mathbf{a},\Delta t_2|\mathbf{b},0)\ , \tag{62}$$

where $\Delta t>\tau_c>0$, $\Delta t_1>\tau_c>0$, $\Delta t_2>\tau_c>0$. If we now use the identity

$$P_{\rm ne}(\mathbf{a},\Delta t|\mathbf{b},0)=\int d\mathbf{c}\,\delta(\mathbf{c}-\mathbf{a})P_{\rm ne}(\mathbf{c},\Delta t|\mathbf{b},0) \tag{63}$$

and introduce Eq. (62) into Eq. (63), use the fact that according to Eq. (58) Λ is the transpose of F, and that the solution of the backward equation is given by Eq. (61) we arrive at Eq. (46a), namely,

$$P_{\rm ne}(\mathbf{a},\Delta t|\mathbf{b},0)=\int d\mathbf{c}\,P_{\rm ne}(\mathbf{a},\Delta t_1|\mathbf{c},0)\times P_{\rm ne}(\mathbf{c},\Delta t_2|\mathbf{b},0)\ . \tag{46a}$$

It is now clear that the time intervals which appear in this equation are not arbitrary, they need to satisfy the requirements of being greater than the memory kernel's correlation time τ_c.

To obtain the CK equation from Eq. (46a) we need further to assume that the stochastic process is stationary. Nevertheless this fact is a consequence of the information-theory approach, because as we showed in the last section, due to the fact that the coarse-grained distribution function satisfies that $\hat\rho(\Gamma,s)=P_z\hat\rho(\Gamma,s)$, the conditional probability has the stationary property given by Eq. (41). Using this property it is clear that Eq. (46a) reduces to the CK equation.

The fact that the CK equation is valid only for long time intervals, implies for the theorem showed in the last section, that the semigroup property of the M operator is valid under the restriction that the times involved are not only positive as usual, but a more restrictive condition, namely, that the times need to be greater than τ_c.

The question concerning the justification of the property given by Eq. (55) from first principles, is outside our treatment. However, in this aspect the interested reader can refer to Ref. [14], where the spectral properties of Liouville's operator are analyzed associating the problem with the nonintegrability of the great majority of the physical systems. The answer to the question about which class of physical systems have the property that for long times exhibit a slow behavior as used in this work remains an open one.

Finally we wish to point out that as shown in this section in order to obtain the CK equation it is necessary to introduce only one dynamical argument to justify the appearance of a slow process. The other two assumptions can be completely justified from the information-theory approach. Thus we may ask if it is possible to obtain the semigroup property of the M operator such that it guarantees the existence of a Markovian behavior, using only information-theory arguments. This possibility is the subject of the following section.

VI. OBSERVATIONAL COMPATIBILITY PRINCIPLE

In this section we will argue that it is possible to establish the semigroup property of the M operator using only an information-theory approach. Accordingly the distribution function at some time t is determined when we maximize the Gibbs-Shannon information entropy $S_I(t)$, and it is expressed in terms of the information that we have about the state of the system at such time, namely, $g_{\rm ne}(\mathbf{a},t)$. Thus $\hat\rho(\Gamma,t)$ can be expressed in terms of the values of the observables at this time. When we perform measurements at different times, we need to introduce another criterion which involves the appearance of several times. Clearly it is therefore somewhat compulsory that such a criterion must arise as an inference principle. To justify this point of view we introduce the following inference principle: The distribution function $\hat\rho(\Gamma,t)$, determined using the maximum-entropy principle from the observations made on the system at a given time t, correctly predicts the values of the system's observables at any observation time greater than time t in which $\hat\rho(\Gamma,t)$ was determined.

We call this assertion the observational compatibility principle (OCP), and study its consequences. In fact we will show that the semigroup property of the M operator can be justified using this principle. Let us assume that we have the sequence of observations at times $t_1<t_2<\ldots t_r<\ldots$ so that the corresponding distribution functions are $g_{\rm ne}(\mathbf{a},t_1)$, $g_{\rm ne}(\mathbf{a},t_2)$, $\ldots$, $g_{\rm ne}(\mathbf{a},t_r)$. We begin by considering a pair of observation times, say t_p, t_q, such that $t_p>t_q$. The knowledge of $g_{\rm ne}(\mathbf{a},t_q)$ is used to construct $\hat\rho(\Gamma,t_p)$, so that by Eq. (24) we have,

$$\hat\rho(\Gamma,t_q)=\int d\mathbf{a}[\Omega(\mathbf{a})]^{-1}g_{n-c}(\mathbf{a},t_q)G(\mathbf{a},0)\ .$$

To obtain the distribution function at any other time

$t > t_q$, we use Liouville's equation with the initial condition given by $\hat{\rho}(\Gamma, t_q)$. Then,

$$\rho_q(\Gamma, t) = U(t, t_q)\hat{\rho}(\Gamma, t_q) ,$$

which in turn predicts that

$$g_{\rm ne}(\mathbf{a}, t) = \int d\Gamma \rho_q(\Gamma, t) G(\mathbf{a}, 0) .$$

The OCP asserts that if t_r is an observation time greater than time t_q, the distribution function $\rho_q(\Gamma, t)$ gives at time t_r the observed function $g_{\rm ne}(\mathbf{a}, t_r)$, so that

$$g_{\rm ne}(\mathbf{a}, t_r) = \int d\Gamma \rho_q(\Gamma, t_r) G(\mathbf{a}, 0) . \tag{64}$$

At time t_r we construct $\hat{\rho}(\Gamma, t_r)$ using Eq. (24) and introducing Eq. (64), which is a consequence of the OCP, we obtain that $\hat{\rho}(\Gamma, t_r)$ is given by

$$\hat{\rho}(\Gamma, t_r) = \int d\mathbf{a} \frac{G(\mathbf{a}, 0)}{\Omega(\mathbf{a})} \times \left[\int d\Gamma \{ U(t_r, t_q)\hat{\rho}(\Gamma, t_q) \} G(\mathbf{a}, 0) \right] . \tag{65}$$

Using now the definition of the P_z operator given by Eq. (14) we have that

$$\begin{aligned} \hat{\rho}(\Gamma, t_r) &= P_z U(t_r, t_q)\hat{\rho}(\Gamma, t_q) \\ &= M(t_r, t_q)\hat{\rho}(\Gamma, t_q) , \end{aligned} \tag{66}$$

which yields the relationship between the coarse-grained distribution function at times t_r and t_q, with $t_r > t_q$. The relationship between the coarse-grained distribution functions at two times is a central result in order to obtain the increase-in-entropy law, as we saw in Sec. III.

We now analyze the consequences of the OCP when we consider three observation times $t_r > t_q > t_p$. When we use the OCP for the pair of times (t_q, t_p), we have that the coarse-grained distribution functions are related by Eq. (66),

$$\hat{\rho}(\Gamma, t_q) = M(t_q, t_p)\hat{\rho}(\Gamma, t_p) \tag{67a}$$

and similarly, when we do focus our attention to the pair of times (t_r, t_p) we have that

$$\hat{\rho}(\Gamma, t_r) = M(t_r, t_p)\hat{\rho}(\Gamma, t_p) . \tag{67b}$$

Also we must have observational compatibility between the observables at times t_r and t_q, thus the OCP implies that

$$\hat{\rho}(\Gamma, t_r) = M(t_r, t_q)\hat{\rho}(\Gamma, t_q) , \tag{67c}$$

therefore, the observational compatibility principle applied to the observations performed at times $t_r > t_p > t_q$ implies that the relationships given by Eqs. (67a), (67b), and (67c) are simultaneously satisfied. Thus using Eqs. (67b) and (67c) we have

$$\hat{\rho}(\Gamma, t_r) = M(t_r, t_p)\hat{\rho}(\Gamma, t_p) = M(t_r, t_q)\hat{\rho}(\Gamma, t_q)$$

and using (67a) in the last equality, we obtain

$$\begin{aligned} \hat{\rho}(\Gamma, t_r) &= M(t_r, t_p)\hat{\rho}(\Gamma, t_p) \\ &= M(t_r, t_q)M(t_q, t_p)\hat{\rho}(\Gamma, t_p) . \end{aligned} \tag{68}$$

Therefore the OCP implies that the M operator has the property

$$M(t_r, t_p) = M(t_r, t_q)M(t_q, t_p) \tag{69}$$

for the observation times $t_r > t_q > t_p$. We wish to underline here that we obtain the semigroup property for the M operator not as a consequence of the dynamical properties of the system, but from an inference principle, namely, OCP. One consequence of this fact is that we cannot quantify the time scale in which the semigroup property is valid. Indeed it is only feasible to give a qualitative argument about the times involved asserting that they need to be large because a measurement of the system implies some time average of the dynamical quantities. However the fact that Eq. (69) is, as we showed in Sec. V only true for slow processes, defines the time scale at which the OCP holds true, namely, each measurement has to be performed in a time interval necessarily greater than time τ_c.

VII. CONCLUDING REMARKS

The three important results derived in this paper are expressed in Eqs. (34), (46a), and (69), respectively. The former one which gives the sought increase in entropy law is not really new. Several workers in this field have in the past reached similar if not identical conclusions using the concepts of information theory as first proposed by Jaynes in his well-known papers of 1957 [15]. A restricted version of Eq. (34) in a way resembling Eq. (23) was derived by Robertson in 1966 [16]. In that same year Hobson [17] wrote a paper on the subject which has undeservedly remained unnoticed. In that paper he uses practically the same ideas of this work to derive an increase-in-entropy law for single- and two-time measurements of a given system. In the latter case his results differ from ours because he uses δ functions of the system's observables for each measurement instead of the coarse-grained property expressed in Eq. (15). Later on Lewis in 1967 [18] proposed a unifying principle to derive equations of motion for state functions in statistical mechanics which contains a nondecreasing time property for the entropy. It is relevant to point out the Lewis principle can be shown to stem from this present formalism but we shall leave the underlying details for a future publication. Also, specific theorems dealing with the question of how the amount of information gathered by a system influences the behavior of its nonequilibrium states through the entropy were derived by Lloyd a few years ago [19]. Similar work on this context can also be found in several review articles on the subject so the interested reader is referred to them [20–23].

The second result summarized in Eq. (46a) has to our knowledge not been previously published. What we have achieved is the proof of the semigroup property that the time evolution operator $M(t, t')$ satisfies under the condition that the process is a slow one. The reader should be

very careful about this statement. The property stated in Eq. (46a) is not valid for any arbitrary dynamical process, only for Markovian ones. That slow processes are Markovian and therefore described by a nonequilibrium conditional probability which satisfies the Chapman-Kolmogoroff equation was first pointed out by Green over 40 years ago [13] $M(t,t')$ is no longer a time-reversible operator in contrast with similar operators that have been used previously in the literature [24,25].

Finally our third result expressed by Eq. (69) simply makes explicit what was said in the previous paragraph. The difference is that the argument used to reach this result is based on an inference principle stated here as the OCP. This places emphasis once more on the fact that the semigroup property exhibited by the operator $M(t,t')$ is not general at all. Nevertheless it can be understood with the aid of arguments based exclusively on the precepts of information theory without resorting to the detailed dynamics obeyed by the system [26]. As emphasized in the text the domain of validity of the OCP is not arbitrary; measurements have to be performed within time intervals longer than the correlation time of the system.

APPENDIX

We proceed to prove Eq. (45) which gives the effect of the M operator on the microcanonical distribution function $\rho(\mathbf{b},\Gamma,0)$ associated with the $\mathbf{b}$ cell, defined by

$$\rho(\mathbf{b},\Gamma,0)=[\Omega(\mathbf{b})]^{-1}G(\mathbf{b},0)\ . \tag{A1}$$

Because of the fact that $P_z G(\mathbf{b},0)=G(\mathbf{b},0)$, the distribution function $\rho(\mathbf{b},\Gamma,0)$ is invariant under P_z, so that $P_z\rho(\mathbf{b},\Gamma,0)=\rho(\mathbf{b},\Gamma,0)$. Applying the M operator to $\rho(\mathbf{b},\Gamma,0)$, and using Eq. (31) we have that

$$\begin{aligned}M(t,t')\rho(\mathbf{b},\Gamma,0)&=P_z U(t-t',0)P_z\rho(\mathbf{b},\Gamma,0)\\&=P_z U(t-t',0)\rho(\mathbf{b},\Gamma,0)\ ,\end{aligned} \tag{A2}$$

where in the last step we used the invariance of $\rho(\mathbf{b},\Gamma,0)$ under the projection operator. Substituting the explicit form of Zwanzig's projector given by Eq. (14) into Eq. (A2), we obtain

$$M(t,t')\rho(\mathbf{b},\Gamma,0)=\int d\mathbf{c}[\Omega(\mathbf{c})]^{-1}(U(t,t')\rho(\mathbf{b},\Gamma,0),G(\mathbf{c},0))G(\mathbf{c},0)\ . \tag{A3}$$

Taking into account that $U(t-t')=\exp\{-iL(t-t')\}$, the fact that L is a Hermitian operator, and Eq. (A1), Eq. (A3) reduces to

$$\begin{aligned}M(t,t')\rho(\mathbf{b},\Gamma,0)=\int d\mathbf{c}(&\rho(\mathbf{b},\Gamma,0),G(\mathbf{c},t-t'))\\&\times\rho(\mathbf{c},\Gamma,0)\ ,\end{aligned} \tag{A4}$$

where we use that $G(\mathbf{c},t)=[\exp\{iLt\}]G(\mathbf{c},0)$. Recalling that for two arbitrary phase functions A and B, $(A,B)=(B,A)^*$, Eq. (A4) can be rewritten as

$$\begin{aligned}M(t,t')\rho(\mathbf{b},\Gamma,0)=\int d\mathbf{c}(&G(\mathbf{c},t-t'),\rho(\mathbf{b},\Gamma,0))^*\\&\times\rho(\mathbf{c},\Gamma,0)\ .\end{aligned} \tag{A5}$$

Finally, using Eq. (40) and the fact that the conditional probability is real, we arrive at

$$M(t,t')\rho(\mathbf{b},\Gamma,0)=\int d\mathbf{c}\,P_{\rm ne}(\mathbf{c},t-t'|\mathbf{b},0)\rho(\mathbf{c},\Gamma,0)\ , \tag{A6}$$

which is Eq. (45) in the text.

*On leave from UAM-Iztapalapa.
†Also at El Colegio Nacional, Luis González Obregón 23, 06020 México, D.F., Mexico.

[1] J. L. Del Río-Correa and L. S. García-Colín, Phys. Rev. A **43**, 6657 (1991) (hereafter referred to as paper I).
[2] J. L. del Río-Correa and L. S. García-Colín, Phys. Lett. **85A**, 201 (1981), and references cited therein.
[3] N. G. Van Kampen, in *Fundamental Problems in Statistical Mechanics,* edited by E. G. D. Cohen (North-Holland, Amsterdam, 1962), p. 173.
[4] For the explicit definition of the function $D(x)$ see the Appendix of paper I.
[5] R. W. Zwanzig, Phys. Rev. **24**, 983 (1961).
[6] L. S. García-Colín and J. L. del Río-Correa, Physica A **96**, 606 (1979).
[7] H. Graebert, *The Projection Operator in Non-Equilibrium Statistical Mechanics* (Springer-Verlag, Berlin, 1982).
[8] E. T. Jaynes, Phys Rev. A **4**, 747 (1971).
[9] J. L. del Río-Correa, Physica A **131**, 329 (1985).
[10] B. Misra, I. Prigogine, and M. Courbage, Proc. Natl. Acad. Sci. USA **76**, 3607 (1979); B. Misrá, I. Prigogine, and M. Courbage, Physica A **98**, 1 (1979).
[11] L. S. García-Colín and J. L. del Río-Correa, J. Stat. Phys. **16**, 235 (1977); **17**, 109 (1978).
[12] See Ref. 6, Sec. 3.
[13] M. S. Green, J. Chem. Phys. **20**, 1281 (1952).
[14] T. Petrosky and I. Prigogine, Physica A **175**, 146 (1991).
[15] E. T. Jaynes, Phys. Rev. **106**, 620 (1957); **108**, 171 (1957); *Statistical Physics,* 1962 Brandeis Lectures (Benjamin, New York, 1963), Vol. 3, p. 181.
[16] B. Robertson, Phys. Rev. **144**, 151 (1966).
[17] A. Hobson, J. Chem. Phys. **45**, 1352 (1966). See also his book, *Concepts in Statistical Mechanics* (Gordon and Breach, New York, 1971).
[18] R. W. Lewis, J. Math. Phys. **8**, 1448 (1967).
[19] S. Lloyd, Phys. Rev. **39**, 5378 (1989).
[20] A. Wehrl, Rev. Mod. Phys. **50**, 221 (1978).
[21] F. Schlogl, Phys. Rep. C **62**, 267 (1980).
[22] W. T. Grandy, *Foundations of Statistical Mechanics* (Reidel, Dordrecht, 1988), Vols. I and II.

[23] M. C. Mackey, Rev. Mod. Phys. **61**, 981 (1989). See also his book, *Time's Arrow: The Origins of Thermodynamical Behavior* (Springer-Verlag, New York, 1992).
[24] L. P. Horwitz, J. A. LaVita, and J. P. Marchand, J. Math. Phys. **12**, 2537 (1971).
[25] D. H. Williams, Commun. Math. Phys. **21**, 314 (1971).
[26] J. L. del Rio-Correa and L. S. García-Colín (unpublished).