

Irreversibility and typicality: a simple analytical result for the Ehrenfest model

Marco Baldovin^{a,1,*}, Lorenzo Caprini^{b,1}, Angelo Vulpiani^{c,1,1}

^a*Dipartimento di Fisica, Università di Roma “Sapienza”, P.le A.Moro 5 I-00185, Rome, Italy*

^b*Gran Sasso Science Institute (GSSI), Via F.Crispi 7, I-67100 L’Aquila, Italy*

^c*Dipartimento di Fisica, Università di Roma “Sapienza”, P.le A.Moro 5 I-00185, Rome, Italy
and Centro Interdisciplinare “B.Segre”, Accademia dei Lincei, Rome, Italy*

Abstract

With the aid of simple analytical computations for the Ehrenfest model, we clarify some basic features of macroscopic irreversibility. The stochastic character of the model allows us to give a non-ambiguous interpretation of the general idea that irreversibility is a typical property: for the vast majority of the realizations of the stochastic process, a single trajectory of a macroscopic observable behaves irreversibly, remaining “very close” to the deterministic evolution of its ensemble average, which can be computed using probability theory. The validity of the above scenario is checked through simple numerical simulations and a rigorous proof of the typicality is provided in the thermodynamic limit.

Keywords: Macroscopic irreversibility, Typicality, Ehrenfest model

1. Introduction

Understanding the irreversibility from first principles is an old and noble problem of Physics. The technical reason of its difficulty is rather clear: on the one hand, the microscopic world is ruled by laws (Hamilton equations) which are invariant under the transformation of time reversal ($t \rightarrow -t$, $\mathbf{q} \rightarrow \mathbf{q}$, $\mathbf{p} \rightarrow -\mathbf{p}$, being \mathbf{q} and \mathbf{p} the positions and momenta of the system); the macroscopic world, on the other hand, is described by irreversible equations, e.g. the Fick equation for the diffusion [1, 2, 3, 4].

How is it possible to conciliate the two above facts? On this topic there is an aged debate which started with the celebrated Boltzmann’s H theorem and the well known criticisms by Loschmidt (about reversibility) and Zermelo

*Corresponding author

Email addresses: marco.baldovin@roma1.infn.it (Marco Baldovin), lorenzo.caprini@gssi.it (Lorenzo Caprini), angelo.vulpiani@roma1.infn.it (Angelo Vulpiani)

(about recurrency). Of course we cannot enter in a detailed discussion about this fascinating chapter of statistical mechanics [2, 5]. Already Boltzmann and Smoluchowski understood that the criticism by Zermelo is not a real serious problem as long as macroscopic systems are considered: basically, because of the Kac's lemma, in macroscopic systems the recurrence time is so large that it cannot be observed [2, 6, 7]. We can summarise the conclusions of Boltzmann by saying that the irreversibility describes an empirical regularity of macroscopic objects which is valid for a "vast majority" of the possible initial conditions. Often such validity for the "vast majority" of initial conditions is called *typicality*. According to Lebowitz [8] (as well as many others) *a certain behaviour is typical if the set of microscopic states for which it occurs comprises a region whose volume fraction goes to one as the number of molecules N grows*. We can state that irreversibility is an emergent property [9, 6, 5] which appears as the number of degrees of freedom becomes (sufficiently) large; in such a limit, a single observation of the system is enough to determine its macroscopic properties.

Several mathematical results, as well as detailed numerical simulations, support the coherence of the scenario proposed by Boltzmann [2]. Among the others, Lanford's work about the rarified gases is particularly important: he was able to prove, in a rigorous fashion, the validity of the Boltzmann equation for short times (of the order of the collision time) in the so-called Boltzmann-Grad limit [10].

In spite of the above mentioned results, irreversibility still remains a somehow misinterpreted and controversial issue. The reader may appreciate the diversity of opinions from the comments [11] to a well known paper by Lebowitz on Boltzmann's approach to the irreversibility. For instance, Prigogine and his school claim that irreversibility is either true on all levels or on none: it cannot emerge as if out of nothing, on going from one level to another [12, 6, 11]. For others, irreversibility either results from (microscopic) chaotic dynamics or it is a mere consequence of the interaction with the external environment.

One source of the controversy about the Boltzmann point of view, in particular among philosophers of science, is how to interpret typicality [13].

This article aims at supplementing, mainly for pedagogical purposes, the basic aspects of Boltzmann's explanation of macroscopic irreversibility. In order to present a clear non ambiguous analysis we treat a stochastic system, i.e the celebrated Ehrenfest model, which is nothing but a Markov chain. A simple and neat analytical computation for the model shows, in a precise way, that for $N \gg 1$ each realization is very close, at any time, to the average value (which can be easily computed).

The paper is organized as follows: in Sec. 2 we introduce the problem of typicality together with some remarks about ensembles and entropies. Sec. 3 is devoted to the Ehrenfest model, and to some numerical results. Then, we rigorously prove the typicality of a trajectory for such a model in Sec. 4. Finally we summarize the results in Sec. 5.

2. Remarks about ensembles, entropies and typicality

Traditionally, entropy has an important relevance in the treatment of irreversibility; it seems to us that this central role is mainly based on historical grounds. In the present paper we do not discuss irreversibility in terms of entropy, for two main reasons. First, the word entropy can be source of confusion: for instance the entropy S_G , defined in terms of the probability distribution function in the Γ - space, has a completely different behaviour from S_B , i.e the entropy obtained from the probability density of a single particle (μ -space); for a discussion on this point see [4, 14, 8]. Second, at a practical, as well as at a conceptual level, for understanding of irreversibility it is enough to observe that, if the system starts from a typical far-from-equilibrium initial state, the macroscopic observables stay close to their mean values during the evolution, and therefore they approach their equilibrium values. In Sec. 4 we will discuss this point in a precise mathematical way.

Even if probability theory has a great relevance for statistical mechanics, it is necessary to avoid mixing mathematics and physics. It is true that the building of the standard formulation of statistical mechanics is based on statistical ensembles, but this approach can be seen as a mere stratagem, and it is ultimately unconvincing in the following sense: in experiments, as well as in numerical computations, we are forced to treat a unique system, and we have not access to a collection of identical systems [15, 16]. At a physical level the relevant problem is: what is the link between the probabilistic computations (i.e. the averages over an ensemble) and the actual results obtained by looking at a single realization (or sample) of the system under investigation?

In particular one should be careful to avoid the confusion between irreversibility and relaxation of the phase space probability distribution [14, 8]. In presence of “good chaotic properties” (mixing systems) one has that the probability density, $\rho(\mathbf{X}, t)$, relaxes (in a suitable technical sense) to the invariant distribution for large times, i.e. $\rho(\mathbf{X}, t) \rightarrow \rho_{inv}(\mathbf{X})$ as $t \rightarrow \infty$. This property is remarkable, and rather important in the dynamical systems context; still, it cannot be considered physical irreversibility. Actually, from a physical point of view, the true question is to show that a *single macroscopic* system shows an irreversible behaviour, for a “generic” initial state. In crude terms the interesting point is to understand the cooling of a single (initially hot) pot and not the behaviour of an ensemble of pots.

In deterministic systems a delicate point is how to intend typicality i.e. the precise mathematical meaning of “vast majority”. As already mentioned, for many scientists active in statistical mechanics “vast majority” means with probability close to 1 with respect to the Lebesgue measure, in physical terms microcanonical distribution [17, 8, 16]. Such an interpretation has been considered not convincing by some authors who criticised the privileged status of the Lebesgue measure [13]. Although, in our opinion, there are very good reasons to privilege the microcanonical distribution, we do not insist in such controversial aspect. In the following we will consider a well known stochastic model [18, 19] where there is no ambiguity about the meaning of “vast majority”. An analysis

of simplified stochastic models of this kind can help very much, in particular at pedagogical level, for understanding irreversibility (see for instance [20] for a discussion about Kac ring model).

3. The Ehrenfest model: heuristic results

3.1. Description and basic properties

For the sake of self-consistency, we briefly recall the Ehrenfest model [18]. We consider N particles, labeled with an index $i = 1, \dots, N$, and two boxes, A and B: at the beginning, each particle can be placed either in box A or in box B. At every time step we randomly choose an integer number between 1 and N , with a uniform distribution, and we move the corresponding particle from its box to the other one. The “macroscopic” state of the model at time t is identified by n_t , the number of particles in A, while the corresponding “microscopic” configuration is defined by the (labeled) particles which actually are in box A at that time. The Markovian evolution for the macroscopic state is ruled by the transition probabilities for $n_t = n$ to become $n_{t+1} = n \pm 1$:

$$\begin{cases} P_{n \rightarrow n-1} = \frac{n}{N} \\ P_{n \rightarrow n+1} = 1 - \frac{n}{N}. \end{cases} \quad (1)$$

As a consequence, for any starting value n_0 , during t steps the macroscopic state can realize 2^t different trajectories. Let us notice that in the Ehrenfest model the detailed balance holds: this property in the stochastic context is somehow equivalent to the time reversibility.

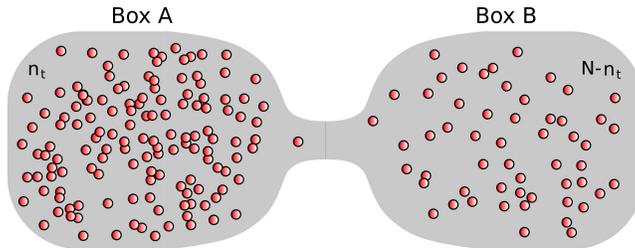


Figure 1: Physical interpretation of the Ehrenfest model: box A (containing n_t particles) is connected to the box B (containing $N - n_t$ particles) by a narrow channel.

This model can be seen as a crude description of a system with N particles in two vessels (A and B), connected by a narrow pipe, as shown in Fig. 1. Of course in this case the true dynamics is deterministic, while in the Ehrenfest model the evolution is stochastic. As a link between the original, Hamiltonian system and its Markovian counterpart, we can imagine to associate to each realization of the Ehrenfest model a set of initial conditions in the deterministic system.

The simplicity of the model allows us to study the statistical features of the evolution of an ensemble of (microscopic) initial conditions in the same

(macroscopic) state, n_0 , by computing the evolution of the first and the second conditional momenta of the state n_t , namely $\langle n_t \rangle$ and $\sigma_t^2 = \langle n_t^2 \rangle - \langle n_t \rangle^2$; for the sake of simplicity, we omit the conditional argument in the average $\langle \cdot | n_0 \rangle$. It is easy, see Appendix A, to show that:

$$\langle n_t \rangle = \frac{N}{2} + \left(1 - \frac{2}{N}\right)^t \Delta_0, \quad (2)$$

$$\sigma_t^2 = \frac{N}{4} + \left(1 - \frac{4}{N}\right)^t \left(\Delta_0^2 - \frac{N}{4}\right) + \left(1 - \frac{2}{N}\right)^{2t} \Delta_0^2, \quad (3)$$

where $\Delta_0 = n_0 - N/2$. From Eq.(2) is clear that $\langle n_t \rangle$ relaxes monotonically to the equilibrium (mean) value $n_{eq} = N/2$, with an exponential decay ruled by the characteristic time $\tau_c = -1/\ln(1 - 2/N) \simeq N/2$. In a similar way the conditional standard deviation σ_t tends to its equilibrium value $\sqrt{N}/2$ with a characteristic time $O(N)$.

3.2. First numerical clues of typicality

Let us note that Eqs.(2) and (3) provide a description of the process only at an average level, giving no information about the single realization. We will see that under the assumption $N \gg 1$ (somehow equivalent to the *thermodynamic limit* in real physical systems) almost all actual realizations are arbitrarily close to the average evolution at any time, i.e. almost all trajectories are “typical”.

Simple numerical computations suggest that a single trajectory is typical in the sense discussed above, i.e. that behaviors very different from the average one are extremely rare. In Fig.2 we show n_t/N in function of time, panels (a) and (b), for several single realizations of n_t for the same n_0 , where the exponential behavior clearly emerges for each trajectory. Panels (c) and (d), on the other hand, display $n_t - \langle n_t \rangle$ vs t for the same trajectories; we superimpose a confidence interval (light blue region) obtained by considering a stripe $\pm 3\sigma_t$ around the conditional average value, $\langle n_t \rangle$, as computed in Eqs. (2) and (3). Each trajectory is contained in this stripe for almost all times: since the trajectories are closer to their mean value as N increases, Fig.2 can be seen as a first, rough numerical clue of the emergence of typicality in the limit $N \gg 1$.

3.3. Maximal deviations from the average

In the next Section, using just simple mathematical methods of the probability theory, we will show that starting from a far-from-equilibrium initial condition (e.g. $|n_0 - N/2| \gg \sqrt{N}$), n_t exhibits an irreversible behavior in the limit $N \gg 1$, namely it remains close to $\langle n_t \rangle$, which exponentially decays to its equilibrium value $N/2$. More precisely, defining the quantity

$$\delta_T = \sup_{0 \leq t \leq T} |n_t - \langle n_t \rangle|, \quad (4)$$

i.e. the maximal deviation of n_t from its average value along a trajectory of T time-steps, we will show that

$$\text{Prob} \left(\frac{\delta_T}{N} < c_N \right) \geq 1 - \zeta_N, \quad (5)$$

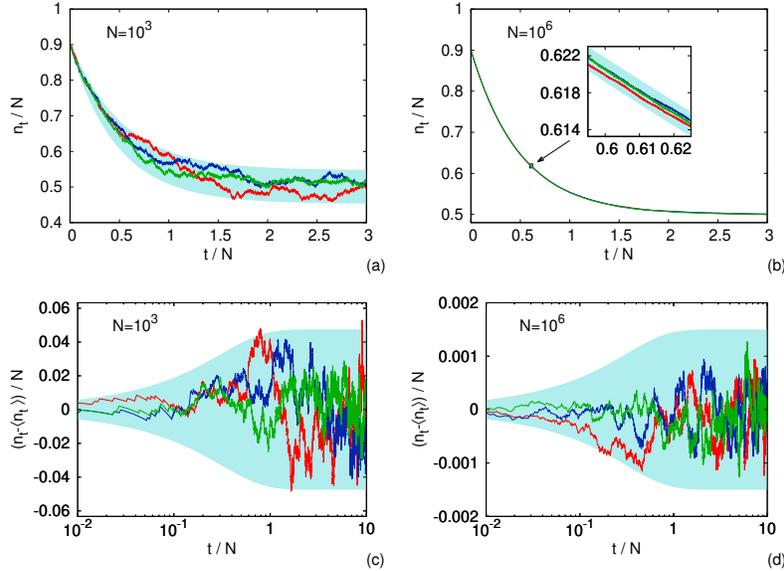


Figure 2: Single trajectories of the model. Above: n_t/N as a function of the time t , for $N = 10^3$ (panel (a)) and $N = 10^6$ (panel (b)); three different realizations are shown for each case. The inset in panel (b) is a magnified detail of the plot. Below: $(n_t - \langle n_t \rangle)/N$ vs t for the same trajectories ($N = 10^3$ in panel (c) and $N = 10^6$ in panel (d)). Light-blue regions are stripes of size $\pm 3\sigma_t$ around the mean value $\langle n_t \rangle$. In all considered cases, $n_0 = 0.9N$.

where T is $O(N)$, and both the constants c_N and ζ_N tend to zero in the limit $N \rightarrow \infty$.

Before we exhibit a mathematical proof of relation (5), let us provide a numerical evidence of its validity. In Fig. 3 we show $\delta_{max}(M)$, defined as the maximal value of δ_T along M different trajectories of length T , in function of M . The scaling of such a quantity with M is particularly interesting, since the number of realizations we need to observe a certain maximal deviation δ_{max} from the average is an indication of its probability.

Using Kac's lemma for recurrence times [21, 19], in Appendix B we derive the following scaling:

$$\frac{\delta_{max}}{\sqrt{N}} \sim \sqrt{\ln M}, \quad (6)$$

under the assumption that δ_T , for large values, is distributed as a Gaussian variable. Let us remark that detailed statistical features of extremal events can be established in the framework of Gumbels theory [22]: however, since we are only interested in the scaling law of δ_{max} vs M , we can avoid the use of the complete theory by a direct application of Kac's lemma. Let us remark that with a different distribution for δ_T , still exponentially decaying to zero, only minor changes occur in the result. We observe that Eq.(6) is in fair agreement with the numerical data presented in Fig.3.

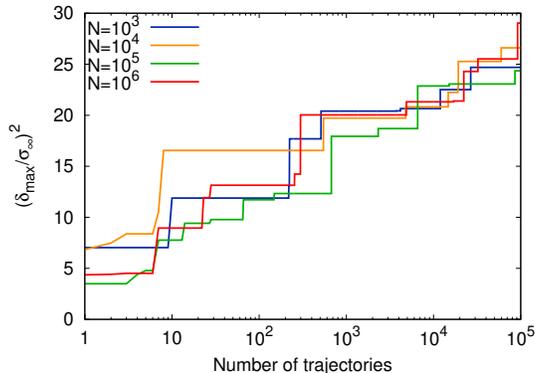


Figure 3: Scaling law between the maximal deviation δ_{max} from average as function of the number of trials M , for $N = 10^3, 10^4, 10^5, 10^6$ and $T = 2N$. As in Fig.2, we set $n_0 = 0.9N$ in all cases.

Roughly speaking, Eq. (6) means that the maximal observed value of δ_T/\sqrt{N} grows very slowly with the number of trials M , implying that significant deviations from the average are extremely rare.

4. A simple analytical result

In order to understand how much a single realization of n_t deviates from its average $\langle n_t \rangle$, i.e. how much a single trajectory is “typical” in the sense discussed in the previous sections, we focus on the probability:

$$\mathcal{P}_{N,T} = \text{Prob}\left(|n_t - \langle n_t \rangle| < \varepsilon_N \quad \forall t : 0 < t \leq T\right) \quad (7)$$

where ε_N is a constant (it only depends on N) and by $\langle n_t \rangle$ we mean $\langle n_t | n_0 \rangle$. Our goal is to prove that

$$\mathcal{P}_{N,T} \rightarrow 1 \quad \text{as } N \rightarrow \infty, \quad \varepsilon_N/N \rightarrow 0, \quad T/N \rightarrow \text{const.} \quad (8)$$

In other words, we aim at showing that a single realization is almost surely contained in a stripe $\pm\varepsilon_N$, at least up to times $\sim O(N)$, being ε_N a quantity that grows slower than N when N tends to infinity.

First of all let us define the following sets:

$$\begin{aligned} \Omega_\tau &\equiv \{t = k\tau : k = 1, 2, \dots, \lfloor T/\tau \rfloor\} \\ \bar{\Omega}_\tau &\equiv \{t \in \mathbb{Z}^+ : t \leq T, t \notin \Omega_\tau\}, \end{aligned} \quad (9)$$

where $\tau \in \mathbb{Z}$ and $\lfloor x \rfloor$ is the integer part of the real number x ; Ω_τ is the set of the equidistant discrete times $t \leq T$ separated by τ , while the remaining times of the trajectory form the set $\bar{\Omega}_\tau$. Defining \mathcal{A}_t as the event that $\delta n_t \equiv |n_t - \langle n_t \rangle| < \alpha_N$, where $\alpha_N < \varepsilon_N$ is a constant, we can write the following lower bound for $\mathcal{P}_{N,T}$:

$$\mathcal{P}_{N,T} \geq \text{Prob}\left[(\delta n_t < \varepsilon_N, \forall t \in \bar{\Omega}_\tau) \text{ and } (\mathcal{A}_t, \forall t \in \Omega_\tau)\right] \quad (10)$$

that can be decomposed, using the definition of the conditional probability P , as

$$\mathcal{P}_{N,T} \geq P(\delta n_t < \varepsilon_N, \forall t \in \overline{\Omega}_\tau | \mathcal{A}_t, \forall t \in \Omega_\tau) \cdot \text{Prob}(\mathcal{A}_t, \forall t \in \Omega_\tau). \quad (11)$$

Because of the markovian character of the process, the above inequality can be written as:

$$\mathcal{P}_{N,T} \geq P(\delta n_t < \varepsilon_N, \forall t \in \overline{\Omega}_\tau | \mathcal{A}_t, \forall t \in \Omega_\tau) \prod_{t \in \Omega_\tau} P(\mathcal{A}_t | \mathcal{A}_{t-\tau}). \quad (12)$$

We will see that, with a suitable choice of τ , such decomposition allows us to prove our statement (8). Now we need to study the two factors of the right hand side of Eq. (12): the first one can be estimated using the transition rules of the Markov chain (1), whereas for the second one we will apply the Chebyshev's inequality.

Since at every time step the value of n_t can only increase (or decrease) by 1, and the same holds for $\langle n_t \rangle$, we can write the following inequality:

$$\delta n_t \leq \delta n_{k\tau} + 2\tau, \quad \forall t : k\tau < t < (k+1)\tau, \quad (13)$$

i.e. during τ time steps, the distance between a particular trajectory and the average cannot spread more than 2τ . We could consider even stronger bounds on δn_t , but inequality (13) is enough to prove our result. In particular, if $\alpha_N + 2\tau < \varepsilon_N$ relation (13) implies

$$P(\delta n_t < \varepsilon_N, \forall t \in \overline{\Omega}_\tau | \mathcal{A}_t, \forall t \in \Omega_\tau) = 1. \quad (14)$$

Let us notice that if we consider:

$$\alpha_N = N^a, \quad \varepsilon_N = N^b, \quad \tau = \lfloor \varepsilon_N / 3 \rfloor \quad (15)$$

if N is large enough, equation (14) holds as soon as

$$0 < a < b < 1. \quad (16)$$

In order to evaluate the product in the left hand side of Eq.(12) we use a different strategy: denoting with \mathcal{A}_t^c the complementary event of \mathcal{A}_t , for every t Chebyshev's inequality [23] ensures that

$$P(\mathcal{A}_t^c | \mathcal{A}_0) \leq \frac{\sigma_t^2}{\alpha_N^2} \leq \frac{N^{1-2a}}{4}, \quad (17)$$

where we have used the bound $\sigma_t^2 \leq N/4$ (Eq. (A.9) of Appendix A). Noting that

$$\begin{aligned} P(\mathcal{A}_t | \mathcal{A}_0) &= P(\mathcal{A}_t | \mathcal{A}_{t-\tau}) P(\mathcal{A}_{t-\tau} | \mathcal{A}_0) + P(\mathcal{A}_t | \mathcal{A}_{t-\tau}^c) P(\mathcal{A}_{t-\tau}^c | \mathcal{A}_0) \\ &\leq P(\mathcal{A}_t | \mathcal{A}_{t-\tau}) + P(\mathcal{A}_{t-\tau}^c | \mathcal{A}_0), \end{aligned} \quad (18)$$

from (17) we easily get

$$\begin{aligned} P(\mathcal{A}_t|\mathcal{A}_{t-\tau}) &\geq P(\mathcal{A}_t|\mathcal{A}_0) - P(\mathcal{A}_{t-\tau}^c|\mathcal{A}_0) \\ &= 1 - [P(\mathcal{A}_{t-\tau}^c|\mathcal{A}_0) + P(\mathcal{A}_t^c|\mathcal{A}_0)] \geq 1 - \frac{N^{1-2a}}{2}. \end{aligned} \quad (19)$$

Finally, using relation (19) to estimate the product in Eq.(12), we find:

$$\mathcal{P}_{N,T} \geq \prod_{t \in \Omega_\tau} P(\mathcal{A}_t|\mathcal{A}_{t-\tau}) \geq \left(1 - \frac{N^{1-2a}}{2}\right)^{\lfloor T/\tau \rfloor}, \quad (20)$$

where $\lfloor T/\tau \rfloor$ is the cardinality of Ω_τ . If we choose T to be proportional to N , we can always find a constant β , independent of N , such that

$$\lfloor T/\tau \rfloor \leq 2\beta N^{1-b}. \quad (21)$$

It is easy to show that, if the constraint (16) holds and, in addition, we choose a such that:

$$a > 1 - \frac{b}{2}, \quad (22)$$

the right hand side of Eq. (20) approaches to one as $\exp(-\beta N^{2-2a-b})$ when $N \rightarrow \infty$ (note indeed that from inequalities (16) and (22) one has $1 - 2a < 0$). This completes the proof, since Eq. (15) ensures that $\varepsilon_N/N \rightarrow 0$ in such limit.

We remark that the proof can be achieved choosing a and b of order $O(1)$. For instance, the couple $b = 0.8$ and $a = 0.7$ satisfies the relations (16) and (22), giving $\mathcal{P}_{N,T} \geq \exp(-\beta N^{-0.2})$.

5. Conclusions

The time irreversibility is an experimental fact whose validity must be accepted as a (quite obvious) empirical property of macroscopic systems. On the other hand it is not easy at all to build a coherent theory that conciliates such macroscopic behaviour with the reversible nature of the laws at the microscopic level (i.e. Newton's equations). The difficult point is to give a mathematical dignity to the great visionary conjecture of Boltzmann, stating that in macroscopic systems an irreversible behaviour occurs for the overwhelming majority of the initial conditions.

One of the most important steps in the ambitious program of formalizing the idea that irreversibility is a typical property is due to Lanford: he had been able to show the validity of the conjecture of Boltzmann for rarified gases in a suitable limit. Lanford proved his result only for a short time (of the order of one collision time); in addition, some authors claim that the use of the Lebesgue measure in the formulation of the idea of typicality is questionable.

In the present paper we give an additional contribution, mainly at a pedagogical level, to support the Boltzmann's conjecture. For the Ehrenfest model we show a result which shares the same philosophy of the Lanford's work: in

the limit $N \gg 1$, a single trajectory of n_t is very close to its mean value $\langle n_t \rangle$, with probability close to 1.

Due to the stochastic nature of our system, it is possible to obtain analytical results up to large times, and there is no ambiguity about the possible interpretations of typicality.

Appendix A. Derivation of Eqs. (2) and (3)

Let us start with the derivation of Eq.(2). Defining the variable Δ_t as a random quantity which takes values ± 1 with probabilities $1 - n_t/N$ and n_t/N , respectively, we have the recurrence relation:

$$n_{t+1} = n_t + \Delta_t. \quad (\text{A.1})$$

Taking the conditional average of the state n_{t+1} for a given n_t , using Eq.(A.1), we get:

$$\langle n_{t+1} | n_t \rangle = n_t + \langle \Delta_t | n_t \rangle = n_t + \left(1 - \frac{n_t}{N}\right) - \frac{n_t}{N} = \left(1 - \frac{2}{N}\right) n_t + 1 \quad (\text{A.2})$$

where we have just applied the definition of Δ_t . Defining the variable $\delta_t = \langle n_t | n_{t-1} \rangle - N/2$, we can replace Eq. (A.2) with a recursive relation for δ_t :

$$\delta_{t+1} = \left(1 - \frac{2}{N}\right) \delta_t. \quad (\text{A.3})$$

Fixing the initial state n_0 , corresponding to some δ_0 , we finally get:

$$\delta_t = \left(1 - \frac{2}{N}\right)^t \delta_0, \quad (\text{A.4})$$

which, in terms of $\langle n_t | n_0 \rangle$ and the initial state n_0 , reads:

$$\langle n_t | n_0 \rangle = \frac{N}{2} + \left(1 - \frac{2}{N}\right)^t \left(n_0 - \frac{N}{2}\right). \quad (\text{A.5})$$

Analogue calculations allows us to compute Eq.(3), using the same strategy of Eq.(A.2), we can obtain a relation for $\langle n_{t+1}^2 | n_t \rangle$:

$$\langle n_{t+1}^2 | n_t \rangle = 1 + \frac{N-4}{N} n_t^2 + 2n_t. \quad (\text{A.6})$$

Applying recursively this equation and using that $\langle n_0^2 | n_0 \rangle = n_0^2$, we get:

$$\begin{aligned} \langle n_t^2 | n_0 \rangle &= \frac{N(N+1)}{4} + N \left(n_0 - \frac{N}{2}\right) \left(1 - \frac{2}{N}\right)^t \\ &+ \left(n_0^2 + \frac{N}{4} (N - 4n_0 - 1)\right) \left(1 - \frac{4}{N}\right)^t. \end{aligned} \quad (\text{A.7})$$

Using Eq.(A.5) it is straightforward to derive the conditional variance:

$$\begin{aligned}\sigma_t^2 &= \langle (n_t^2 - \langle n_t^2 \rangle) \rangle \\ &= \frac{N}{4} + \left(1 - \frac{4}{N}\right)^t \left[\left(n_0 - \frac{N}{2}\right)^2 - \frac{N}{4} \right] + \left(1 - \frac{2}{N}\right)^{2t} \left(n_0 - \frac{N}{2}\right)^2, \quad (\text{A.8})\end{aligned}$$

which leads to Eq.(3). As a remark, we note that Eq.(A.8) is bounded by:

$$\sigma_t^2 \leq \sigma_\infty^2 = \frac{N}{4}, \quad (\text{A.9})$$

the stationary variance of the Ehrenfest model.

Appendix B. Scaling law for $\delta_{max}(M)$

Let us consider the Markov chain (1) starting with the initial condition $n = n_0$. For each trajectory, Eq. (4) defines the largest deviation δ_T from the average (within a time T), and we have indicated as $\delta_{max}(M)$ the maximal occurrence of this quantity in M independent realizations. Our goal is to find an (approximate) relation between δ_{max} and M . We assume an asymptotic behavior of the probability density function, $p(x)$, of the random variable $x = \delta_T/\sigma_\infty$ for a single trajectory. For instance, we can assume that such distribution decays with a gaussian tail,

$$p(x) \simeq \alpha e^{-cx^2} \quad \text{for } x \gg 1. \quad (\text{B.1})$$

The (average) number of independent attempts M that are needed in order to observe a value of x larger than X is simply given by Kac's lemma [19]:

$$\langle M \rangle = \left(\int_X^\infty dx p(x) \right)^{-1}. \quad (\text{B.2})$$

With the assumption (B.1), in the limit $X \gg 1$, the above relation can be written as

$$\ln \langle M \rangle \simeq -\ln \left(\text{const.} \frac{e^{-cX^2}}{2X} \right) \simeq cX^2. \quad (\text{B.3})$$

Inserting $X = \delta_{max}/\sigma_\infty \sim \delta_{max}/\sqrt{N}$ into the above relation, we finally get the scaling law in Eq. (6). It is easy to understand that the above argument for the logarithmic dependence of X as a function of $\langle M \rangle$ is rather robust and does not depend on the details of the distribution: assuming an asymptotic shape $p(x) \sim \exp(-cx^A)$, one obtains

$$\ln \langle M \rangle \sim x^A. \quad (\text{B.4})$$

Acknowledgments

We thank M. Falcioni and R. Figari for a critical reading of the manuscript.

- [1] Ludwig Boltzmann. *Lectures on Gas Theory*. Dover Publications, New York, reprint edition, 2003.
- [2] Carlo Cercignani. *Ludwig Boltzmann: The Man Who Trusted Atoms*. Oxford University Press, Oxford, 1998.
- [3] Gerard G. Emch and Chuang Liu. *The Logic of Thermostatistical Physics*. Springer-Verlag, Berlin Heidelberg, 2002.
- [4] Patrizia Castiglione, Massimo Falcioni, Annick Lesne, and Angelo Vulpiani. *Chaos and Coarse Graining in Statistical Mechanics*. Cambridge University Press, Cambridge, UK, 2008.
- [5] Nino Zanghì. I fondamenti concettuali dell’approccio statistico in fisica. In Valia Allori, Mauro Dorato, Federico Laudisa, and Nino Zanghì, editors, *La natura delle cose: Introduzione ai fondamenti e alla filosofia della fisica*, pages 202–247. Carocci Ed., Roma, 2005.
- [6] Sergio Chibbaro, Lamberto Rondoni, and Angelo Vulpiani. *Reductionism, Emergence and Levels of Reality: The Importance of Being Borderline*. Springer International Publishing, 2014.
- [7] Dustin Lazarovici and Paula Reichert. Typicality, Irreversibility and the Status of Macroscopic Laws. *Erkenntnis*, 80:689–716, 2015.
- [8] Joel L. Lebowitz. Boltzmann’s Entropy and Time’s Arrow. *Physics Today*, 46:32, 1993. doi: 10.1063/1.881363.
- [9] Michael V. Berry. Asymptotics, singularities and the reduction of theories. *Logic, Methodology and Philosophy of Science*, IX:597, 1994.
- [10] Oscar E. Lanford. The hard sphere gas in the Boltzmann-Grad limit. *Physica A*, 106:70 – 76, 1981. doi: [https://doi.org/10.1016/0378-4371\(81\)90207-7](https://doi.org/10.1016/0378-4371(81)90207-7).
- [11] Howard Barnum, Carlton M. Caves, Christopher Fuchs, Rüdiger Schack, Dean J. Driebe, William G. Hoover, Harald Posch, Brad Lee Holian, Rudolf Peierls, and Joel L. Lebowitz. Is Boltzmann Entropy Time’s Arrow’s Archer? *Physics Today*, 47:11, 1994. doi: 10.1063/1.2808690.
- [12] J. Bricmont. Science of chaos or chaos in science? *Annals of the New York Academy of Sciences*, 775:131–175, 1995. doi: 10.1111/j.1749-6632.1996.tb23135.x.
- [13] Roman Frigg. Typicality and the Approach to Equilibrium in Boltzmannian Statistical Mechanics. *Philosophy of Science*, 76:997–1008, 2009.

- [14] L. Cerino, F. Cecconi, M. Cencini, and A. Vulpiani. The role of the number of degrees of freedom and chaos in macroscopic irreversibility. *Physica A*, 442:486–497, 2016. doi: 10.1016/j.physa.2015.09.036.
- [15] Sergio Caprara and Angelo Vulpiani. Law Without Law or Just Limit Theorems? *Foundations of Physics*, 48:1112–1127, 2018.
- [16] Wojciech H. Zurek. Eliminating ensembles from equilibrium statistical physics: Maxwell’s demon, Szilard’s engine, and thermodynamics via entanglement. *Physics Reports*, 755:1 – 21, 2018.
- [17] Sheldon Goldstein. Typicality and Notions of Probability in Physics. In Yemima Ben-Menahem and Meir Hemmo, editors, *Probability in Physics*, pages 59–71. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.
- [18] Paul Ehrenfest and Tatiana Ehrenfest. *The Conceptual Foundations of the Statistical Approach in Mechanics*. Dover Publications, New York, reprint edition, 2015.
- [19] Mark Kac. *Probability and Related Topics in Physical Sciences*. Amer Mathematical Society, Providence, RI, 1957.
- [20] Georg A. Gottwald and Marcel Oliver. Boltzmann’s Dilemma: An Introduction to Statistical Mechanics via the Kac Ring. *SIAM Rev.*, 51:613–635, August 2009. doi: 10.1137/070705799.
- [21] Mark Kac. On the notion of recurrence in discrete stochastic processes. *Bull. Amer. Math.Soc.*, 53:10021010, 1947.
- [22] E.J. Gumbel. *Statistics of Extremes*. Columbia University Press, New York, 1958.
- [23] Boris V. Gnedenko. *Probability Theory*. Taylor & Francis Ltd, sixth edition, 1998.