Séminaire Poincaré XV Le Temps (2010) 141 – 173 **Séminaire Poincaré**

Time's Arrow and Lanford's Theorem

Jos Uffink
Institute for History and Foundations of Science
Utrecht University
PO BOX 80010
3508 TA Utrecht, the Netherlands

Giovanni Valente
Philosophy Department
University of Pittsburgh
1001 Cathedral of Pittsburgh
PA 15260, USA

Abstract. It has been a longstanding problem to show how the irreversible behaviour of macroscopic systems can be reconciled with the time-reversal invariance of these same systems when considered from a microscopic point of view. A result by Lanford (1975, 1976, 1981) shows that, under certain conditions, the famous Boltzmann equation, describing the irreversible behaviour of a dilute gas, can be obtained from the time-reversal invariant Hamiltonian equations of motion for the hard spheres model. Here, we examine how and in what sense Lanford's theorem succeeds in deriving this remarkable result. Many authors have expressed different views on the question which of the ingredients in Lanford's theorem is responsible for the emergence of irreversibility. We claim that the true culprit for the emergence of irreversibility lies in a point that has hitherto not been sufficiently emphasized, i.e. in the choice of incoming, rather than outgoing, configurations for collision points. We argue that this choice ought to be recognized clearly as an explicit assumption in the theorem, and discuss its implications for the question in what sense irreversible behaviour follows from Lanford's theorem.

1 Introduction

The Boltzmann equation is one of the most important tools of statistical physics. It describes the evolution of a dilute gas towards equilibrium, and serves as the key to the derivation of further hydrodynamical equations. A striking aspect of this equation is that it is not invariant under time reversal. Indeed, when Boltzmann (1872, 1875) presented this equation, he immediately derived from it a celebrated theorem, now commonly known as the H-theorem, which shows that a certain quantity H of the gas can only change monotonically in time, so that the gas displays an evolution towards equilibrium.

Despite its long-standing legacy, the status of the H-theorem has remained controversial. The reversibility objection by Loschmidt (1877) questioned the validity of the H-theorem by constructing a counterexample. Essentially, this objection raised the problem of how an irreversible macro-evolution equation can be obtained from the time-reversal invariant micro-evolution equations governing molecular motion. More than twenty years later, Culverwell (1894) posed the same problem and inaugurated a famous debate in *Nature* with a provocative question: "Will anyone say exactly what the H-theorem proves?".

In his responses to the reversibility objection, Boltzmann (1877, 1895) suggested an alternative approach and reading of the H-theorem, which the Ehrenfests (1912) called the "modified formulation of the H-theorem", and which we will refer to as the statistical H-theorem. Yet, the problem of providing a rigorous statistical counterpart of the Boltzmann equation and the H-theorem was left unsolved. It is widely believed that the work by Oscar E. Lanford (1975, 1976, 1981) provides the best available candidate for a rigorous derivation of the Boltzmann equation and the H-theorem from statistical mechanics, in the limiting case of an infinitely diluted gas system described by the hard spheres model, at least for a very brief time. To be sure, these clauses imply that Lanford's result will hardly apply to realistic circumstances. The importance of Lanford's theorem is that it claims to show how the conceptual gap between macroscopic irreversibility and microscopic reversibility can in principle be overcome, at least in simple cases.

However, Lanford's papers suggested various answers to the question how the irreversibility embodied in the Boltzmann equation or the ensuing H-theorem arises in this rigorous statistical mechanical setting. Later commentators on Lanford's theorem (e.g.: Spohn, 1980,1991, 2001; Lebowitz 1983; Cercignani e.a. 1994; Cercignani 2008, Uffink, 2008) have also expressed mutually incompatible views on this particular issue. So, one may well ask: "Will anyone say exactly what Lanford's theorem proves?".

The present paper addresses this question. We analyse the problem of how Lanford's theorem gives rise to the emergence of irreversible behaviour and whether Lanford's result can be interpreted as providing a satisfactory statistical H-theorem. We will argue that the responsibility for the emergence of macroscopic irreversibility in Lanford's theorem is to be sought in an ingredient that has not been sufficiently clearly recognized by previous authors.

1.1 Problems of reduction

Statistical physics is a broad field that comprises many closely related but more strictly circumscribed theories: thermodynamics, hydrodynamics, the kinetic theory of gases, and statistical mechanics, to name the most relevant members. The inter-theoretical relations between them are subtle but it is clear that the Boltzmann equation and the H-theorem are crucial for the precise specification of such relations between all these theories.

For example, the putative reduction of thermodynamics to statistical physics calls for a statistical counterpart of the approach to equilibrium, which, in the kinetic theory of gases, is given by the H-theorem. Similarly, in the analysis of the inter-theoretical relations between kinetic theory and statistical mechanics, regarded as two different theories of statistical physics, it is necessary to spell out what it takes for the original, but untenable, H-theorem to be replaced by a statistical H-theorem. Also, the Boltzmann equation is crucial for the bridge between kinetic theory and hydrodynamics.

For purposes of reduction, it is thus important to establish in what sense the Boltzmann equation and the H-theorem are valid. Since Lanford's theorem aims to address this issue, this theorem is also highly relevant to these foundational issues of reduction in statistical physics. Lanford provided a rigorous derivation of the Boltzmann equation and the H-theorem for the hard spheres gas-model in the so-called Boltzmann-Grad limit. The proof of his theorem is cast in the formalism developed by Bogolyubov, Born, Green, Kirkwood and Yvon (BBGKY). This formalism provides, departing from the Hamiltonian formulation of statistical mechanics, a hierarchy of equations for the time-evolution of macroscopic systems, called the BBGKY hierarchy. On the other hand, the Boltzmann equation itself can also be reformulated in the form of a hierarchy (the Boltzmann hierarchy). Lanford's theorem then shows how the Boltzmann hierarchy can be obtained from the BBGKY hierarchy for the hard spheres model in the Boltzmann-Grad limit under specific assumptions. To be sure, the technical assumptions needed in this rigorous derivation present on several points severe limitations. In particular, the convergence obtained in this Boltzmann-Grad limit holds for a very brief time only, and the Boltzmann-Grad limit itself implies that the density of the gas-model goes to zero, which is incompatible with the usual hydrodynamic limit. Clearly, Lanford's theorem by itself does not provide a full answer to the conceptual issues in the reduction relations mentioned above. However, it provides the most successful approach to these problems yet, and will likely be the guideline for future work towards a more complete coverage of these reduction relations.

We begin by reviewing the Boltzmann equation and the H-theorem in the kinetic theory of gases for the hard spheres model, along with the quest for a statistical H-theorem (Section 2). Section 3 discusses the connection between the BBGKY hierarchy for the hard spheres model and the Boltzmann hierarchy. Lanford's theorem is then stated in section 4. We take up the issue of irreversibility in section 5, and end with our conclusions.

2 The quest for a statistical H-theorem

2.1 The kinetic theory of gases and the Boltzmann equation

In the kinetic theory of gases, one considers a gas as a system consisting of a very large number N of molecules, moving in accordance with the laws of classical mechanics, enclosed in a container Λ with perfectly elastic reflecting and smooth walls. In the hard spheres model, these molecules are further idealized as rigid and impenetrable spheres of diameter a interacting only by collisions. The instantaneous state of the gas system at time t is represented by a *distribution function* $f_t(\vec{q}, \vec{p})$, which represents the relative number of molecules in the gas with positions between $\vec{q}$ and $\vec{q} + d\vec{q}$ inside the container Λ and momenta between $\vec{p}$ and $\vec{p} + d\vec{p}$.

Of course, the question exactly how such a smooth function f_t is meant to represent the distribution of a finite number of particles is somewhat tricky, and we shall come back to it later (section 4.1). Notice that, for each time t, f_t formally defines a probability density on the so-called μ-space $\mu = \Lambda \times \mathbb{R}^3$, i.e.:

$$f_t \geq 0 \quad \text{and} \quad \int_\Lambda d\vec{q} \int_{\mathbb{R}^3} d\vec{p}\, f_t(\vec{q}, \vec{p}) = 1, \tag{1}$$

assigning probabilities to molecular positions and momenta —which thus play the role of stochastic variables. However, in kinetic theory, the distribution function itself is though to represent, in some sense, the relative number of particles over their various possible positions and momenta in the actual microstate of the gas. The distribution function should therefore be sharply distinguished from probability densities as they arise from some probability measure on phase space in statistical mechanics.

In order to describe the evolution of the gas system, one needs to consider how the distribution function $f_t(\vec{q}, \vec{p})$ evolves in time. The crucial assumption to obtain this evolution equation is the *Stoßzahlansatz*, or "assumption about the number of collisions", also often referred to as the Hypothesis of Molecular Chaos, which provides a constraint on the way in which collisions between the particles take place. It can be decomposed in two distinct conditions:

– Factorization
The relative number of pairs of particles, with positions within $d\vec{q}_1$ and momenta within $d\vec{p}_1$, and within $d\vec{q}_2$ and $d\vec{p}_2$, respectively, is given by

$$f_t^{(2)}(\vec{q}_1, \vec{p}_1, \vec{q}_2, \vec{p}_2) d\vec{p}_1\, d\vec{q}_1 d\vec{p}_2 d\vec{q}_2 = f_t(\vec{q}_1, \vec{p}_1)\, f_t(\vec{q}_2, \vec{p}_2)\, d\vec{p}_1 d\vec{q}_1 d\vec{p}_2 d\vec{q}_2 \tag{2}$$

– Pre-collision
The number $N(\vec{p}_1, \vec{p}_2)$ of such pairs of particles which are *about* to collide in a time span dt is proportional to $f_t^{(2)}(\vec{q}, \vec{p}_1, \vec{q}, \vec{p}_2)$ and the volume dV of the "collisions cylinder", i.e. the spatial region around the position q at which the particles are located when colliding, i.e.

$$N(\vec{p}_1, \vec{p}_2) = f_t^{(2)}(\vec{q}, \vec{p}_1, \vec{q}, \vec{p}_2) \cdot dV \tag{3}$$

where

$$dV = a^2 \vec{\omega}_{12} \cdot \left(\frac{\vec{p}_1 - \vec{p}_2}{m} \right) dt d\vec{\omega}_{12} \tag{4}$$

Here, $\vec{\omega}_{12}$ is a unit vector pointing form the center of particle 1 to the center of particle 2 (See Fig. 1). The condition that the particles are "about to collide"can be expressed mathematically by the condition

$$\vec{\omega}_{12} \cdot (\vec{p}_1 - \vec{p}_2) \geq 0. \tag{5}$$

The *Stoßzahlansatz* may be interpreted as saying that particle pairs are uncorrelated in their momenta just before they collide. We note that the literature is somewhat confusing in the terminology here. Many authors use the name "Molecular Chaos" for the factorization condition (2) alone, without including the pre-collision condition.

Whenever a collision occurs, molecular velocities change. If the particles have momenta $\vec{p}_1, \vec{p}_2$ just before the collision, their outgoing momenta will be denoted as $\vec{p}_1{}'$ and $\vec{p}_2{}'$, respectively. In the hard-spheres model, these outtgoing momenta are simple functions of $\vec{p}_1$ and $\vec{p}_2$ and $\vec{\omega}_{12}$. Indeed:

$$\begin{aligned} \vec{p}_1{}' &= \vec{p}_1 - (\vec{\omega}_{12} \cdot (\vec{p}_1 - \vec{p}_2))\, \vec{\omega}_{12} \\ \vec{p}_2{}' &= \vec{p}_2 + (\vec{\omega}_{12} \cdot (\vec{p}_1 - \vec{p}_2))\, \vec{\omega}_{12}, \end{aligned} \tag{6}$$

which can be written more compactly in terms of a linear collision operator $T_{\vec{\omega}_{12}}$, defined by (6):

$$(\vec{p}_1, \vec{p}_2) \longrightarrow (\vec{p}_1{}', \vec{p}_2{}') = T_{\vec{\omega}_{12}}(\vec{p}_1, \vec{p}_2). \tag{7}$$

By standard arguments, one obtains from these assumptions the Boltzmann equation, which the change of the distribution function in the course of time:

$$\begin{aligned} \frac{\partial}{\partial t} f_t(\vec{q}, \vec{p}_1) + \frac{\vec{p}_1}{m} \cdot \frac{\partial}{\partial \vec{q}} f_t(\vec{q}, \vec{p}_1) = N a^2 \int_{\mathbb{R}^3} d\vec{p}_2 \int_{\vec{\omega}_{12} \cdot (\vec{p}_1 - \vec{p}_2) \geq 0} d\vec{\omega}_{12}\, \frac{\vec{p}_1 - \vec{p}_2}{m} \cdot \vec{\omega}_{12} \\ \times \left[f_t(\vec{q}, \vec{p}_1{}') f_t(\vec{q}, \vec{p}_2{}') - f_t(\vec{q}, \vec{p}_1) f_t(\vec{q}, \vec{p}_2) \right] \end{aligned} \tag{8}$$

The second term in the left-hand side of the equation accounts for the change of the distribution function through free motion of particles, whereas the right-hand side is the collision term. Here, the variables $\vec{p}_i{}'$ are to be thought of as implicit functions $\vec{p}_i{}'(\vec{p}_1, \vec{p}_2)$ given by

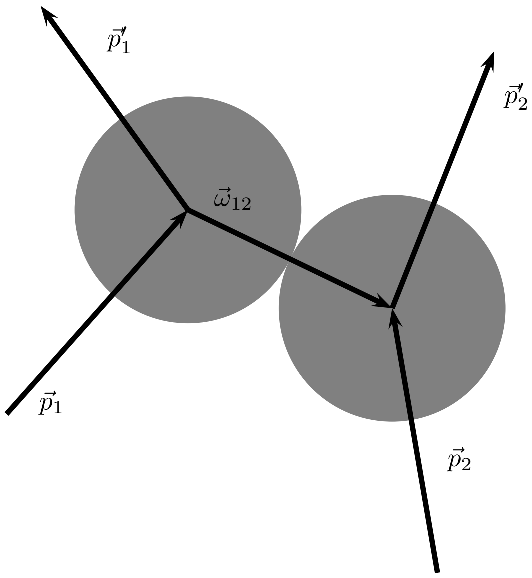

Figure 1: Geometry of a collision between two hard spheres.

(7). Note that the latter term is not linear in f_t. Hence, the overall Boltzmann equation is non-linear, and this is a major obstacle in attempts towards solving the equation. In fact, the question whether the equation does have physically meaningful solutions for all times for some given f_0 as initial condition remains hard even today and has only been answered in special cases.

Boltzmann circumvented this problem by showing that a general theorem could nevertheless be obtained. To derive this result, now commonly known as the H-theorem, he introduced a function of the distribution of state defined as

$$H[f_t] \equiv \int f_t(\vec{q},\vec{p}) \ln f_t(\vec{q},\vec{p})\, d\vec{p}d\vec{q}, \tag{9}$$

and proved that, under the assumption that the Boltzmann equation holds at all times, and f_t is a solution to this equation, then this quantity cannot increase, i.e.

$$\frac{dH[f_t]}{dt} \leq 0 \tag{10}$$

for all t. In the case that the distribution function is and remains spatially uniform, i.e. $f_t(\vec{q},\vec{p}) = f_t(\vec{p})$, equality obtains only for a Maxwell distribution $f(\vec{p}) = Ae^{-\vec{p}^2/B}$, which captures the equilibrium state. If the negative of the H-function is associated with the entropy of the system, the H-theorem means that this entropy increases monotonically through non-equilibrium states until the systems reaches equilibrium and then remains constant.

Boltzmann claimed in 1872 that his result yields a rigorous, analytical proof of the Second Law of thermodynamics. Yet, this is not quite correct. First of all, the Second Law asserts that if an isolated system undergoes an adiabatic process, the entropy associated with the final state cannot be less than the entropy associated with the initial state. But the Boltzmann equation describes a gas evolving in a fixed vessel isolated from its environment and does not refer to general adiabatic processes, e.g when work is performed by moving a piston or stirring the gas. So, the H-theorem does not cover all the relevant thermodynamical processes dealt with in the Second Law. Secondly, the Second Law does not prescribe the monotonic increase of entropy, nor does it allude to non-equilibrium states. Hence, in this sense Boltzmann actually obtained more than what is needed for the purpose of reproducing the Second Law. Therefore, the relationship of the H-theorem and the Second Law of thermodynamics is somewhat indirect (cf. Uffink 2008).

Nevertheless the H-theorem does provide another interesting connection between the kinetic theory of gases and thermodynamics. Indeed, it yields a description of the spontaneous approach to equilibrium for gases. The Second Law does not imply that any isolated gas confined in a finite volume would tend to evolve toward an equilibrium state. This is instead captured by an independent principle, which Brown and Uffink (2001) dubbed the Minus First Law. Therefore, although it does not reduce the Second Law to kinetic gas theory, the H-theorem seems to provide a kinetic underpinning of the Minus First Law, at least for the hard-spheres gas model.

However, the general validity of the H-theorem was called into question soon after its formulation. Loschmidt's reversibility objection, as rephrased by Boltzmann in (1877) goes basically as follows. Take a non-equilibrium initial distribution of state f_0 for which the H-theorem holds and let it evolve for a certain amount of time t, so that $H[f_t] < H[f_0]$. Then, suddenly reverse the velocities of all particles. The particles will now simply retrace all their previous motions back to their original spatial configuration at time $2t$. If at point we reverse their velocities again the distribution of state at time $2t$ will be identical to f_0. But since H, as defined by (9) is invariant under a velocity reversal, this means that under the evolution from t to $2t$, H must have been increasing. In other words, for every dynamically allowed evolution for the particles during which H decreases , one can construct another for which H increases, but also allowed by the dynamics.

This argument relies on the tension between the time-reversal invariance of the dynamics governing the motion of the particles and the explicit time-reversal non-invariance of the H-theorem. In fact, one can trace this time-reversal non-invariance back to the Boltzmann equation from which the H-theorem has been derived. This is shown explicitly in *Proposition* 1 in the Appendix.

The upshot of the reversibility objection is that the irreversible time-evolution of macroscopic systems cannot be a consequence of the laws of Hamiltonian mechanics alone. There must be some additional non-dynamical ingredient in the H-theorem, or indeed in the Boltzmann equation from which it follows, that picks out a preferred direction in time. As we now know, the Stoßzahlansatz is the culprit. The pre-collision condition introduces a time-asymmetric element, since it is assumed to hold only for particle pairs immediately before collisions, but not for pairs immediately after they collided. This is responsible for the failure of the Boltzmann equation to be time-reversal invariant. Indeed, if we had supposed, instead of the pre-collision condition, a similar condition for the momenta immediately after collision, we would, by the same argument, have obtained a version of the Boltzmann equation with an additional minus sign in the collision term, a version commonly called the *anti-Boltzmann equation*[1], and accordingly, we would have derived an anti-H theorem, i.e. $dH[f_t]/dt \geq 0$. Hence the irreversible behaviour in the macro-evolution of non-equilibrium distributions towards equilibrium is due to the preference of this pre-collision rather than a corresponding post-collision condition. But this preference cannot be grounded in the dynamics.

2.2 Boltzmann's statistical reading of the H-theorem

Boltzmann's (1877) response to Loschmidt already argued that one cannot prove that every initial distribution function should always evolve towards the equilibrium distribution function, but rather that there are infinitely many more initial states that do evolve, in a given time, towards equilibrium than do evolve away from equilibrium, and that even these latter states will evolve towards equilibrium after an even longer time . However, Boltzmann did not provide proofs for these claims.

A more detailed argument can be found in Boltzmann (1895, 1897). To any microstate one can associate a curve (the H-curve), representing the behaviour of $H[f_t]$ in the course of time. Boltzmann claimed that, with the exception of certain 'regular' microstates, the curve exhibits the following properties: (i) for most of the time, $H[f_t]$ is very close to its minimal value H_{min}; (ii) occasionally the H-curve rises to a peak well above the minimum value; (iii) higher peaks are extremely less probable than lower ones. If at time $t = 0$ the curve takes on a value $H_[f_0]$ much greater than H_{min}, the function may evolve only in two alternative ways. Either $H[f_0]$ lies in the neighbourhood of a peak, and hence $H[f_t]$ decreases in both directions of time; or it lies on an ascending or descending slope of the curve, and hence $H[f_t]$ would correspondingly decrease or increase. However, statement (iii) entails that the first case is much more probable than the second. One would thus conclude that there is a very high probability that at time $t = 0$ the entropy of the system, associated with the negative of the H-function, would increase for positive time; likewise there is a very high probability that the entropy would increase for negative time.

It is this conclusion that is sometimes called the *statistical H-theorem.* Nevertheless, Boltzmann gave no proof of these claims, nor did he indicate whether or how they might still depend on the *Stoßahlanasatz.* Thus the statistical H-theorem is hardly a theorem. The problem of finding an analogue of Boltzmann's H-theorem in statistical mechanics thus remained unsettled. We now investigate whether Lanford's theorem does provide a mathematically rigorous version of a statistical H-theorem.

3 On the derivation of the Boltzmann equation from Hamiltonian mechanics

3.1 From the Hamiltonian framework to the BBGKY hierarchy

In this section we briefly describe the general form of the BBGKY hierarchy. Again, we consider a classical mechanical system consisting of N particles, each with the same mass m. In order to alleviate a bit the notation of the equations to follow, we will set $m = 1$. The particles are contained in a vessel $\Lambda \subset \mathbb{R}^3$ with a finite volume and smooth wall $\partial\Lambda$. But we now approach this system from statistical mechanics, rather than kinetic theory. Its $6N$-dimensional phase space is given by $\Gamma_N = (\Lambda \times \mathbb{R}^3)^N$ and its evolution is governed

1. As it appears evident from the proof of *Proposition* 1 in the Appendix, the anti-Boltzmann equation can also be obtained by applying a time-reversal transformation to the Boltzmann equation.

by a Hamiltonian

$$H_N(x) = \sum_{i=1}^{N} \vec{p}_i^{\,2} + \sum_{i<j}^{N} \phi(\vec{q}_i - \vec{q}_j) \tag{11}$$

Here, x denotes the microstate $x = (x_1, \ldots, x_N) = (\vec{q}_1, \vec{p}_1, \ldots, \vec{q}_N, \vec{p}_N)$.

Strictly speaking, the Hamiltonian should also contain a term corresponding to the elastic wall potential, describing the interaction when individual particles collide with the boundary $\partial\Lambda$ of the vessel. However, there are ways to suppress this complication. The easiest way is to suppose that each particle i undergoes specular reflection when it hits the wall and identify the values $(\vec{q}_i, \vec{p}_i)$ just before such a collision and the values $(\vec{q}_i, \vec{p}_i')$ immediately after. In this move, the phase-space Γ_N is endowed with the topology of a torus, and the dynamics under wall collisions becomes smooth. Indeed, a collision with the wall becomes indistinguishable from free motion, and consideration of the wall potential becomes redundant.

Now, although we will eventually focus on the hard-sphere model, i.e. the special case when

$$\phi(\vec{q}_i - \vec{q}_j) = \begin{cases} \infty & \text{when } \|(\vec{q}_i - \vec{q}_j)\| \geq a \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

we assume, for now, that ϕ is a smooth function obeying the Lipshitz condition. The virtue of this assumption is that, in this case, the Hamiltonian (11) is known to be integrable, so that there exists a smooth one-parameter group of transformations, $\{T_t, t \in \mathbb{R}\}$ on Γ_N, called the Hamiltonian flow, $T_t : \Gamma_N \longrightarrow \Gamma_N$, $\Gamma_N \ni x \mapsto x_t = T_t(x)$ that characterizes the dynamics.

The statistical state of the system is given by a probability measure μ over Γ_N. We assume that μ is absolutely continuous with respect to Lebesgue measure on Γ, so one can write

$$\mu(A) = \int_A \mu(x)dx \tag{13}$$

in terms of a density function $\mu(x)$ with respect to the Lebesgue measure on Γ.

The evolution of such a statistical state $\mu(x)$ at any instant t is defined by

$$\mu_t(x) = \mu(T_{-t}x) \tag{14}$$

in terms of the Hamiltonian flow or, equivalently, by means of the Liouville equation

$$\frac{\partial \mu}{\partial t} = \{H, \mu\} = \sum_{i=1}^{N} \frac{\partial H}{\partial \vec{q}_i}\frac{\partial \mu}{\partial \vec{p}_i} - \frac{\partial H}{\partial \vec{p}_i}\frac{\partial \mu}{\partial \vec{p}_i} =: \mathcal{H}_N \mu. \tag{15}$$

The BBGKY approach exploits the fact that the above Hamiltonian (11) is invariant under permutation of the particles, and, moreover, the inter-particle potential ϕ only contains pair-interactions. Furthermore, we assume that μ is permutation invariant as well:

$$\mu(x_1, \ldots, x_i \ldots, x_j, \ldots x_N) = \mu(x_1, \ldots, x_j \ldots, x_i, \ldots x_N) \qquad \forall i, j \in \{1, \ldots, N.\} \tag{16}$$

Obviously, for a permutation invariant Hamiltonian such as (11), this property of μ_t will be conserved under the dynamical evolution (15).

With the above symmetry assumptions in place, it is clear that macroscopic quantities of physical interest will only depend on *how many* particles have certain molecular properties, or how many pairs have certain relations to each other, but not on their particle labels. It thus becomes attractive to study the dynamics in terms of reduced probability densities obtained by conveniently integrating out most of the variables. For this purpose, one defines a hierarchy of reduced or marginal probability densities:

$$\begin{aligned} \mu_1(x_1) &:= \int \mu(x_1, \ldots x_N) dx_2 \cdots x_N \\ &\vdots \\ \mu_k(x_1, \ldots, x_k) &:= \int \mu(x) dx_{k+1} \cdots dx_N \\ &\vdots \\ \mu_N(x_1, \ldots, x_N) &:= \mu(x_1, \ldots, x_N). \end{aligned} \tag{17}$$

Here, for instance, μ_k is the probability density that k particles occupy positions $\vec{q}_1, \ldots, \vec{q}_k$ and move with momenta $\vec{p}_1, \ldots, \vec{p}_k$, while the remaining $N-k$ particles possess arbitrary positions and momenta. Note that, although $\mu_1(x)$ is thus a marginal probability density on μ-space, just like the distribution function f discussed in section 2.1, the conceptual status of these two density functions is very different.

With a somewhat different normalization convention, one defines rescaled reduced probability densities:

$$\rho_k(x_1, \ldots, x_k) = \frac{N!}{(N-k)!} \frac{1}{N^k} \mu_k(x_1, \ldots, x_k). \tag{18}$$

It remains, of course, to specify the time evolution of these rescaled reduced probability densities.

Now, the N-particle Liouville operator $\mathcal{H}_N$ in the Liouville equation (15) can be expanded as

$$\mathcal{H}_N = -\sum_{i=1}^{N} \vec{p}_i \cdot \frac{\partial}{\partial \vec{q}_i} + \sum_{i \neq j}^{N} \mathcal{L}_{ij} \tag{19}$$

where

$$\mathcal{L}_{ij} := \frac{\partial \phi(\vec{q}_i - \vec{q}_j)}{\partial \vec{q}_i} \cdot \frac{\partial}{\partial \vec{p}_i}. \tag{20}$$

The evolution of ρ_1 is therefore given by

$$\frac{\partial \rho_{1,t}(x_1)}{\partial t} = \vec{p}_1 \cdot \frac{\partial}{\partial \vec{q}_i} \rho_{1,t}(x_1) + N \int dx_2 \mathcal{L}_{12} \rho_{2,t}(x_1, x_2), \tag{21}$$

and for the higher-order rescaled reduced probability densities:

$$\frac{\partial \rho_{k,t}}{\partial t} = -\sum_{i=1}^{N} \vec{p}_i \cdot \frac{\partial}{\partial \vec{q}_i} \rho_k + \sum_{i \neq j}^{k} \mathcal{L}_{ij} \rho_k + N \sum_{i=1}^{k} \int dx_{k+1} \frac{\partial \phi(\vec{q}_i - \vec{q}_{k+1})}{\partial \vec{q}_i} \cdot \frac{\partial}{\partial \vec{p}_i} \rho_{k+1}. \tag{22}$$

Or, in abbreviated form:

$$\frac{\partial \rho_k}{\partial t} = \mathcal{H}_k \rho_k + \mathcal{C}^{\phi}_{k,k+1} \rho_{k+1}, \quad k = 1, \ldots, N. \tag{23}$$

where the superscript on the operator $\mathcal{C}$ is intended to remind one that it depends on the smooth inter-particle potential ϕ in (11).

These dynamical equations for the rescaled reduced densities of the statistical state μ constitute the BBGKY hierarchy. Note that, taken together, they are strictly equivalent to the Hamiltonian evolution, i.e. nothing else has been assumed yet, except for the rather harmless permutation invariance of ρ and the specific form of the Hamiltonian (11). As one might expect, therefore, solving these equations is just as hard as for the original Hamiltonian equations. Indeed, to find the time-evolution of ρ_1 from (21), we need to know $\rho_{2,t}$. But to solve the dynamical equation for ρ_2, we need to know $\rho_{3,t}$ etc. Moreover, the equations of the BBGKY hierarchy are still perfectly time-reversal invariant.

Nevertheless, the above might already make one hopeful that a counterpart of the Boltzmann equation can be obtained from the exact Hamiltonian dynamics. Indeed, if we tentatively identify Boltzmann's f function with ρ_1, (21) looks somewhat similar to the Boltzmann equation (8). Of course, much work still remains to be done: first of all, the Boltzmann equation pertains to the hard-sphere model, whereas the equation (21) assumes a smooth pair-potential $\phi(\vec{q}_i - \vec{q}_j)$. More importantly, we will have to justify the tentative relationship between ρ_1 and f. These tasks will be tackled in the following subsections.

3.2 From smooth potentials to the hard-spheres gas model

While the BBGKY hierarchy provides a generally useful format for studying the evolution of a statistical state for a system of indistinguishable particles interacting by smooth pair potentials, it is our purpose here to apply it to the hard-spheres potential (12).

There are several caveats when applying Hamiltonian dynamics or the BBKGY hierarchy to the case of a hard-spheres model, in particular, because the potential (12) of this model does not obey the Lipshitz condition. First of all, we have to remove configurations in which particles overlap, i.e. restrict our original phase space Γ_N to:

$$\Gamma^{(a)}_{N,\neq} := \{x \in (\Lambda \times \mathbb{R}^3)^N : \ \|\vec{q}_i - \vec{q}_j\| \geq a, \quad i \neq j, \quad i,j \in \{1, \ldots, N\} \ \}. \tag{24}$$

More importantly, the dynamical evolution of the microstate of a collection of N hard spheres enclosed in a vessel might lead to (i) grazing collisions (ii) more than two particles colliding simultaneously or (iii) an infinite number of collisions (either between the particles mutually or between some particle and the wall) occurring within a finite lapse of time. In all of these cases, the Hamiltonian equations cannot be solved, and the trajectory in phase space cannot be extended for all times. Fortunately, it has been shown by Alexander (1975) that the subset consisting of microstates x showing such anomalous evolutions has a Lebesgue measure zero in $\Gamma^{(a)}_{N,\neq}$. Therefore, if, as we assumed, the statistical state μ is absolutely continuous with respect to the Lebesgue measure, these unwanted microstates make up a set of probability zero, and can be ignored for the purpose of our analysis.

That is to say, we can either delete this unwanted set Δ of measure zero from our phase space $\Gamma^{(a)}_{N,\neq}$, and in doing so guarantee that there is a Hamiltonian flow $\{T_t, t \in \mathbb{R}\}$ defined on the smaller phase space $\Gamma_{N,\neq} \setminus \Delta$, or continue with the original space, with the provision that its Hamiltionian flow is defined only almost everywhere, i.e. outside of the above set Δ.

Thus, the hard-sphere dynamics is such that if we consider any given phase point $x = (x_1, \ldots x_N)$ $(x \notin \Delta)$ and consider how it will move under the flow in the next sufficiently small time increment δt, then either all particles persist in free motion (or perhaps some collide with the wall); or else some pair of particles, say i and j, collide. In the latter case, at the moment of collision, they touch, i.e., their positions obey

$$\vec{q}_j = \vec{q}_i + a\vec{\omega}_{ij} \quad \text{for} \quad \vec{\omega}_{ij} = \vec{q}_j - \vec{q}_i, \tag{25}$$

which implies that the microstate x lies on the boundary of $\Gamma_{N,\neq}$, and in the collision their momenta undergo an instantaneous transition, cf. (7):

$$(\vec{p}_i, \vec{p}_j) \longrightarrow (\vec{p}_i{}', \vec{p}_j{}') = T_{\vec{\omega}_{ij}}(\vec{p}_i, \vec{p}_j). \tag{26}$$

Note that $T_{\vec{\omega}}$ is measure preserving, and an involution, i.e. $T_{\vec{\omega}} \circ T_{\vec{\omega}} = \mathbb{1}$. In other words, whenever the incoming momenta before a collision between particles i, j happen to take values $(\vec{p}_i{}', \vec{p}_j{}')$, they are transformed into $(\vec{p}_i, \vec{p}_j)$:

$$(\vec{p}_i{}', \vec{p}_j{}') \longrightarrow (\vec{p}_i, \vec{p}_j) = T_{\vec{\omega}_{ij}}(\vec{p}_i{}', \vec{p}_j{}'). \tag{27}$$

Now, although this momentum transfer during collision is clearly discontinuous, one can nevertheless maintain the idea that the dynamics is smooth, by mimicking a procedure already applied to deal with collisions with the wall, i.e., by adopting a topology in which the pre-collision coordinates $(\vec{q}_i, \vec{p}_i; \vec{q}_j, \vec{p}_j)$ and the post-collision coordinates $(\vec{q}_i, \vec{p}_i{}'; \vec{q}_j, \vec{p}_j{}')$ are identified. We will discuss this procedure in greater detail in section 5.

With these caveats taken care off, we thus recover a smooth dynamics for the hard spheres model, and indeed one can show that the equations (23) go over in

$$\frac{\partial \rho^{(a)}_{k,t}}{\partial t} = \mathcal{H}_k \rho^{(a)}_{k,t} + \mathcal{C}^{(a)}_{k,k+1} \rho^{(a)}_{k+1,t} \qquad k \in \{1, \ldots, N\} \tag{28}$$

where now

$$\mathcal{C}^{(a)}_{k,k+1}\rho^{(a)}_{k+1} = Na^2 \sum_{i=1}^{k} \int_{\mathbb{R}^3} d\vec{p}_{k+1} \int_{S^2} d\vec{\omega}_{i,k+1} \left(\vec{\omega}_{i,k+1} \cdot (\vec{p}_{k+1} - \vec{p}_i)\right) \times \rho^{(a)}_{k+1}(x_1, \ldots, x_k, \vec{q}_i + a\vec{\omega}_{i,k+1}, \vec{p}_{k+1}) \tag{29}$$

and the superscript (a) is intended to remind one that these operators and the rescaled probability densities refer to a hard spheres model with a sphere diameter $a > 0$. Of course, for each value of k, these rescaled probability densities $\rho^{(a)}_k$ are defined only on the domains

$$\Gamma^a_{k,\neq} = \{(x_1, \ldots x_k) \in (\Lambda \times \mathbb{R}^3)^k : \ \|\vec{q}_i - \vec{q}_j\| \geq a, \quad i \neq j, \quad i, j \in \{1, \ldots k\} \ \} \tag{30}$$

We emphasize that the resulting form (28) of the BBGKY hierarchy for the hard sphere model is still time-reversal invariant. This is indeed the content of Proposition 2 in the Appendix.

However, one more crucial step is needed in order to make the collision term (29) in the BBGKY hierarchy look like a Boltzmann collision term, as appears in (8). First of all, we can split the integral over the unit sphere S^2 into two parts: the hemisphere $\vec{\omega}_{i,k+1} \cdot (\vec{p}_i - \vec{p}_{k+1}) \geq 0$, and the hemisphere $\vec{\omega}_{i,k+1} \cdot (\vec{p}_i - \vec{p}_{k+1}) \leq 0$. In the first hemisphere, the collision configuration $(\vec{q}_i, \vec{p}_i; \vec{q}_i + a\vec{\omega}_{i,k+1}, \vec{p}_{k+1})$ represents a collision between particles i and $k+1$ with incoming momenta $\vec{p}_i, \vec{p}_{k+1}$, and we leave the integrand as it is.

In the other hemisphere, characterized by $\vec{\omega}_{i,k+1} \cdot (\vec{p}_i - \vec{p}_{k+1}) \leq 0$, the momenta in the configuration $(\vec{q}_i, \vec{p}_i; \vec{q}_i + a\vec{\omega}_{i,k+1}, \vec{p}_{k+1})$ appear as outgoing momenta. In this hemisphere, these momenta are replaced by the corresponding incoming momenta, which, according to (27), gives the configuration:

$$(\vec{q}_i, \vec{p}_i{}'; \vec{q}_i + a\vec{\omega}_{i,k+1}, \vec{p}{}'_{k+1}) \quad \text{where} \quad (\vec{p}_i{}', \vec{p}{}'_{k+1}) = T_{\omega_{i,k+1}}(\vec{p}_i, \vec{p}_{k+1}) \tag{31}$$

Also, we replace the integration variable $\vec{\omega}_{i,k+1}$ by $-\vec{\omega}_{i,k+1}$. The result of these operations is that we obtain from (29) the collision term:

$$\hat{C}^{(a)}_{k,k+1}\rho^{(a)}_{k+1,t}(x_1,\ldots,x_k) = Na^2\sum_{i=1}^{k}\int_{\vec{\omega}_{i,k+1}\cdot(\vec{p}_i-\vec{p}_{k+1})\geq 0} d\vec{\omega}_{i,k+1}\ d\vec{p}_{k+1}\ \vec{\omega}_{i,k+1}\cdot(\vec{p}_i-\vec{p}_{k+1})$$
$$\times\left[\rho^{(a)}_{k+1}(x_1,\ldots,\vec{q}_i,\vec{p}_i{}',\ldots x_k,\vec{q}_i-a\vec{\omega_{i,k+1}},\vec{p}{}'_{k+1})-\rho^{(a)}_{k+1}(x_1,\ldots,\vec{q}_i,\vec{p}_i,\ldots x_k,\vec{q}_i+a\vec{\omega}_{i,k+1},\vec{p}_{k+1})\right] \tag{32}$$

Summing up Lanford's argument so far, the general BBGKY hierarchy has been applied to the particular case of the hard-spheres model. An equation (28) for the time-evolution of the relevant reduced probability densities including the details of both collisions and rectilinear motion of the particles is thus obtained. However, in the last passage from (29) to (32) a particular step was made, namely to split integral and rewrite the integrands in terms of pre-collision rather than the post-collision coordinates. But since such a step is accompanied by an argument that, in order to assure that the dynamics is smooth, we can identify these coordinates as representing the same physical phase point, it may seem that this is just a conventional choice of representation, as Lanford himself suggested. Nevertheless, we will argue in section 5 that this step is actually crucial for the emergence of irreversibility in Lanford's theorem.

3.3 From the Boltzmann equation to the Boltzmann hierarchy

In this section, we start from the other side of the bridge that we aim to cross. That is, we take the Boltzmann equation as given, and reformulate it in a mathematically equivalent hierarchy of distribution functions. This idea is captured by the lemma below, which is spelled out by Lanford (1975, p. 88).

First, define a hierarchy of multi-particle distribution functions by

$$f_{k,t}(x_1,\ldots,x_k) = \prod_{i=1}^{k} f_t(x_i) \quad k\in\mathbb{N} \tag{33}$$

where $x_i = (\vec{q}_i,\vec{p}_i)$. Then we have:

Lemma 3.1 *The following two statements are equivalent: (i): f_t is a solution of the Boltzmann equation and (ii) the functions $f_{k,t}$ obey the equations*

$$\frac{\partial f_{k,t}}{\partial t} = \mathcal{H}_k f_{k,t} + \mathcal{C}_{k,k+1}f_{k+1,t} \quad k\in\mathbb{N} \tag{34}$$

where:

$$\mathcal{H}_k := \sum_{i=1}^{k}\mathcal{L}_i := -\sum_{i=1}^{k}\vec{p}_i\cdot\frac{\partial}{\partial\vec{q}_i} \tag{35}$$

and

$$C_{k,k+1}f_{k+1,t}(x_1,\ldots,x_k) = Na^2\sum_{i=1}^{k}\int_{\vec{\omega}_{i,k+1}\cdot(\vec{p}_i-\vec{p}_{k+1})\geq 0} d\vec{\omega}_{i,k+1}d\vec{p}_{k+1}\ \ (\vec{\omega}_{i,k+1}\cdot(\vec{p}_i-\vec{p}_{k+1}))$$
$$\times\left[f_{k+1}(x_1,\ldots,\vec{q}_i,\vec{p}_i{}',\ldots,\vec{q}_i,\vec{p}{}'_{k+1}) - f_{k+1}(x_1,\ldots,\vec{q}_i,\vec{p}_i,\ldots,\vec{q}_i,\vec{p}_{k+1})\right] \tag{36}$$

and

$$(\vec{p}_i{}', \vec{p}_{k+1}{}') = T_\omega(\vec{p}_i, \vec{p}_{k+1}) \tag{37}$$

In other words, the problem of solving the Boltzmann equation for a distribution function f is equivalent to the problem of solving a hierarchy of evolution equations (34), called the *Boltzmann hierarchy*, for the functions $(f_1, f_2, \ldots)$ under the assumption of a factorization condition (33). One can write this hierarchy more compactly by regarding the f_k as components of a vector: $\boldsymbol{f} = (f_1, f_2, \ldots)$. Then we can write (34) schematically as:

$$\frac{\partial}{\partial t}\boldsymbol{f} = \mathcal{H}\boldsymbol{f} + \mathcal{C}\boldsymbol{f}, \tag{38}$$

where $\mathcal{H}$ is a diagonal matrix with diagonal elements $\mathcal{H}_k$ and $\mathcal{C}$ a matrix with elements $\mathcal{C}_{k,k+1}$ and zero elsewhere.

The lemma has two virtues: First, and most important is the point that while the original non-linear Boltzmann equation is notoriously hard to solve, the Boltzmann hierarchy (38) is linear. This contrast arises, of course, because the non-linearity is put, so to say, in the factorization constraint (33). As a consequence, it is easier to write down (at least formal) solutions to the Boltzmann hierarchy. A formal solution to this hierarchy of equations is obtained by writing down an expansion familiar from Dyson's time-dependent perturbation theory:

$$\boldsymbol{f}_t = S(t)\boldsymbol{f}_0 + \sum_{i=1}^{\infty} \int_0^t dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{i-1}} dt_m S(t-t_1)\mathcal{C}S(t_1-t_2)\mathcal{C}\cdots\mathcal{C}S(t_m)\boldsymbol{f}_0 \tag{39}$$

where the operator $\mathcal{S}(t)$ represents the collisionless time evolution, i.e.:

$$\mathcal{S}(t)f_k(x_1, \ldots x_k) := f_k(\vec{q}_1 - t\vec{p}_1,\, \vec{p}_1, \vec{q}_2 - t\vec{p}_2, \vec{p}_2,\, \ldots,\, \vec{q}_N - t\vec{p}_N, \vec{p}_N) \tag{40}$$

Obviously, the above formal way of writing a general solution to the Boltzmann hierarchy does not alleviate the original problems in solving the Boltzmann equation entirely; these problems are merely transposed into a further problem of showing that the series expansion in (39) actually converge.

The second virtue of the lemma is that it brings the Boltzmann equation in a form which is more similar to the results from the BBGKY formalism discussed above, which likewise take the form of a hierarchy, and this alleviates the effort to build a rigorous bridge between them.

As we have remarked above, the factorization condition (33), taken as a generalization of Boltzmann's condition (2), is sometimes called 'molecular chaos'. Interestingly, if the initial data of the Boltzmann hierarchy at time $t = 0$ take the form (33), then this factorization is preserved through time, i.e., it holds for the solution of (34) for all time t, with f_t being a solution of the Boltzmann equation. This important property of the Boltzmann hierarchy is commonly known as 'propagation of chaos'[2]. Note however, that this factorization, and its preservation in time, has nothing to do with the pre-collision condition mentioned in section 2.1 as a crucial ingredient of the molecular chaos hypothesis.

Finally, we stress that the Boltzmann hierarchy, being an equivalent way of expressing the Boltzmann equation, is just as time-reversal non-invariant as the original Boltzmann equation. In fact, by applying a time-reversal transformation to it, one obtains a hierarchy of evolutions equation which has the same form as (34) except for a minus sign in front

2. See (Spohn 1991, p. 45; Cercignani, Illner & Pulvirenti 1994, p. 85) for a more extensive discussion of this property.

of the collision term $C_{k,k+1}$. We refer to the latter as the *anti-Boltzmann hierarchy*. Also, notice that both the collision operators and the distribution functions in (36) resemble those involved in (32), except that they do not depend on the diameter a of the particles. The crucial point in Lanford's theorem is to demonstrate that all relevant terms in the BBGKY hierarchy tend to their counterparts in the Boltzmann hierarchy in the Boltzmann-Grad limit, whereby $a \longrightarrow 0$. That would establish that the Boltzmann hierarchy can be obtained from Hamiltonian mechanics.

4 Lanford's theorem

So far, we have seen how the Hamiltonian dynamics for the hard-spheres model leads, under a relatively harmless assumption of permutation invariance, to a hierarchy of BBGKY equations describing the evolution of reduced density functions of a statistical state. And we have also argued how the Boltzmann equation can be reformulated as a hierarchy of equations in close resemblance to the BBGKY hierarchy. The question still remains how to bridge the gap between these two descriptions. Lanford's theorem establishes the convergence of the BBGKY hierarchy to the Boltzmann hierarchy in the so-called Boltzmann-Grad limit.

This Boltzmann-Grad limit defines a particular limiting regime within the hard spheres model. Not only one lets the number of particles grow to infinity, i.e. $N \to \infty$, but one also requires that their diameter goes to zero, i.e. $a \to 0$ while keeping the volume $|\Lambda|$ of the container fixed. The limit is taken in such a way that the quantity Na^2 remains constant, or at least approaches a finite non-zero value. This guarantees that the collision term in the Boltzmann equation or Boltzmann hierarchy, which is proportional to this quantity, does not vanish. Accordingly, the 'mean free path' $\lambda := \frac{1}{2\pi Na^2}$, which is the typical scale-distance traveled by any particle between two subsequent collisions in an equilibrium state, also remains of order one. The same holds for the 'mean free time', i.e. the typical duration between collisions in equilibrium, which is of the order $\sqrt{(\beta/3)}\pi a^2 N/|\Lambda|$, where β is the inverse temperature. Of course, this limit implies that $Na^3 \to 0$, which means that the gas also becomes infinitely diluted in the Boltzmann-Grad limit.

There is one final technical point we need to mention. Recall that the rescaled probability densities $\rho_k^{(a)}$ of the BBGKY hierarchy have as their domains the sets (28). As one takes the Boltzmann-Grad limit, these sets converge to

$$\Gamma_{k,\neq} := \bigcup_{a>0} \Gamma_{k,\neq}^{(a)} = \{(x_1, \dots x_k) \in (\Lambda \times \mathbb{R}^3)^k : \ \vec{q}_i \neq \vec{q}_j, \quad i \neq j, \quad i,j \in \{1, \dots k\} \ \}. \tag{41}$$

Obviously, we cannot expect the convergence of $\rho_k^{(a)} \longrightarrow f_k$ everywhere in $\Gamma_k := (\Lambda \times \mathbb{R}^3)^k$, but at most on $\Gamma_{k,\neq}$, i.e., away from the hypersurface $\Gamma_{k,=} := \Gamma_k \setminus \Gamma_{k,\neq}$ of phase points for which two particles (now considered as point particles), coincide in space. Actually, we need to be even a little bit more restrictive. Let

$$\Gamma_{k,\neq}(s) := \{(x_1, \dots x_k) \in (\Lambda \times \mathbb{R}^3)^k : \ \vec{q}_i - s\vec{p}_i \neq \vec{q}_j - s\vec{p}_j, \\ i \neq j, \quad i,j \in \{1, \dots k\}, \forall t : 0 \leq t \leq s \ \}. \tag{42}$$

In words, this is the set of k point particle configurations for which no particle pairs collide at time 0, but also have not collided within in a time span $[-s, 0]$.

We are now ready to state the precise version of Lanford's theorem, as given by Spohn (1991, Theorem 4.5). Here, when we write $\lim_{a \longrightarrow 0}$, the Boltzmann-Grad limit is understood, i.e. it is assumed that $N \longrightarrow \infty$ simultaneously, while keeping Na^2 a fixed non-zero constant.

LANFORD'S THEOREM

With the notation introduced in section 3, take $0 < a < a_0$ and let ρ_k^a be a family of functions defined on $\Gamma_{k,\neq}^{(a)}$, and assume that for all such a, the following conditions hold at time $t = 0$.

(i) There exists positive real constants z, β, M, independent of a, such that

$$\rho_{k,0}^{(a)}(x_1, \ldots, x_k) \leq M z^k \prod_{i=1}^{k} h_\beta(\vec{p}_i) \tag{43}$$

for any $k = 1, 2, \ldots$, where $h_\beta(\vec{p}_i)$ denotes the normalized Maxwellian distribution over momenta: $h_\beta(\vec{p}_i) = (\frac{\beta}{2\pi})^{\frac{3}{2}} \cdot e^{-\frac{\beta \vec{p}_i^2}{2}}$ at inverse temperature β, and the spatial distribution is constant inside the vessel Λ with density $z = N/|\Lambda|$.

(ii) There exist continuous functions $f_{k,0}$ on Γ_k, for $k = 1, 2 \ldots$ such that

$$\lim_{a \longrightarrow 0} \operatorname*{ess\,sup}_{(x_1, \ldots x_k) \in K} |\rho_{k,0}^{(a)}(x_1, \ldots, x_k) - f_{k,0}(x_1, \ldots x_k)| = 0 \tag{44}$$

for all compact subsets $K \subset \Gamma_{k,\neq}(s)$ for some $s > 0$.

Then, there exists a strictly positive time τ, such that

$$\lim_{a \longrightarrow 0} \operatorname*{ess\,sup}_{(x_1, \ldots, x_k) \in K} |\rho_{k,t}^{(a)}(x_1, \ldots, x_k) - f_{k,t}(x_1, \ldots, x_k)| = 0 \tag{45}$$

for any $k = 1, 2, \ldots$ and compact subset $K \subset \Gamma_{k,\neq}(s + t)$.

Here, $\rho_{k,t}^{(a)}$ are solutions of the BBGKY hierarchy with initial conditions $\rho_{k,0}^{(a)} = \rho_k^{(a)}$ and $f_{k,t}$ solutions of the Boltzmann hierarchy with initial conditions $f_{k,0} = f_k$ for $t \in [0, \tau]$.

Let us make some comments on the theorem. First, let us put its content in words. Assumption (i) is a regularity condition which admits only initial conditions for the rescaled reduced densities of the BBGKY hierarchy bounded by a product of some constant M times Maxwellian distributions with uniform spatial density z and inverse temperature β . This condition expresses that statistical state of the hard-spheres system should not be "too far away" from equilibrium. By assumption (ii), these initial conditions for the BBGKY hierarchy converge to functions $f_{k,0}$ that serve as initial conditions of the Boltzmann hierarchy. The theorem then states that this convergence is maintained through time, at least for $t \in [0, \tau]$, so that solutions of the BBGKY hierarchy $\rho_{k,t}^{(a)}$ converge to solutions $f_{k,t}$ of the Boltzmann hierarchy as $a \to 0$, except for the phase-points comprised in the set $\Gamma_{k,=}(s+t)$.

Second, if we add the further assumption that the functions $f_{k,0}$ in equation (44) initially factorize:

$$f_{k,0}(x_1, \ldots, x_k) = \prod_{i=1}^{k} f_0(x_i) \tag{46}$$

then we can infer by the Lemma of section 3.3, and the propagation of chaos, that this factorization property is maintained in time:

$$f_{k,t}(x_1, \ldots, x_k) = \prod_{i=1}^{k} f_t(x_i) \tag{47}$$

where f_t is a solution of the Boltzmann equation. In that case, the Lanford theorem not only obtains convergence of $\rho^{(a})_k$ to solutions of the Boltzmann hierarchy, but also obtains the limiting validity of the Boltzmann equation in the sense of (47), for the duration $t \in [0, \tau]$. But note that this factorization condition (46) is not actually needed in the above theorem.

Third, the exceptional sets $\Gamma_{k,=}(t)$ increase in time, i.e. $\Gamma_{k,=}(t) \subset \Gamma_{k,=}(t')$ if $0 < t < t'$. Moreover, they are not invariant under reversal of all velocities. However, being hypersurfaces of codimension one, these sets all have Lebesgue measure zero, and probability zero for any statistical state which is absolutely continuous with respect to the Lebesgue measure.

Finally, an explicit estimate given by (Spohn 1991, p. 62) shows that $0.2\sqrt{(\beta/3)}\pi a^2 z \leq \tau$, i.e. the theorem guarantees convergence for only one-fifth of the mean free time between collisions for the above Maxwellian.

4.1 A measure-theoretic version of Lanford's theorem

While the theorem formulated above deals with the convergence of the solutions of the BBGKY hierarchy to the solutions of the Boltzmann hierarchy, it may seem that these two sets of equations have little to do with each other, each coming from a very different theoretical framework (i.e. Hamiltonian statistical mechanics and kinetic theory, respectively). Thus one may well ask what the relationship between these two frameworks is, in particular how the Boltzmann distribution functions f_k relate to the rescaled probability densities ρ_k in the Hamiltonian framework.

Let us attempt to spell this relationship out more explicitly. To each point $x^N = (\vec{q}_1, \vec{p}_1, \ldots, \vec{q}_N, \vec{p}_N)$ in the N-particle phase space Γ one can associate a normalized (improper) distribution function

$$F^{[x^N]}(\vec{q}, \vec{p}) = \frac{1}{N} \sum_{i=1}^{N} \delta^3(\vec{q}_i - \vec{q})\delta^3(\vec{p}_i - \vec{p}) \tag{48}$$

on the one-particle μ-space. Such a distribution, being a sum of Dirac δ functions it is clearly not a suitable candidate to be fed into the Boltzmann equation (or into the H-theorem, because its H-value would be infinite). However, one might intuitively hope that, for increasingly large N, $F^{[x^N}$ will tend to approximate some continuous density function $f(\vec{q}, \vec{p})$, which can be associated to a solution of the Boltzmann equation. One then needs to express the sense in which the two distributions becomes sufficiently close to each other in the limit of infinitely many N. For this purpose, one can introduce a distance function d: we can then say that $F^{[x^N]}$ converges to f just in case $d(F^{[x^N]}, f) \longrightarrow 0$ when $N \longrightarrow \infty$.

There are obviously various distance functions to choose from (cf. Spohn 1991 p. 26). But a somewhat natural choice would be to stipulate some partition $\mathcal{P}$ of μ-space into disjoint rectangular cells, i.e. $\mathcal{P} = (A_1, \ldots A_n)$, and to say that two (possibly improper) densities f and g on μ-space satisfy $d(f, g) \leq \epsilon$ if and only if

$$\text{for all } A_i \in \mathcal{P} : \left| \int_{A_i} f(\vec{q}, \vec{p}) d\vec{q} d\vec{p} - \int_{A_i} g(\vec{q}, \vec{p}) d\vec{q} d\vec{p} \right| \leq \epsilon \tag{49}$$

Clearly, this distance function is sensitive to the choice of the pair $(\mathcal{P}, \epsilon)$: the larger the number n of cells into which the μ-space is partitioned and the smaller the value of ϵ, the more refined the resulting notion of "closeness" between the functions would be. It is plausible that Boltzmann himself intended some such limiting procedure, whereby a discrete distribution describing N particles would approximate a continuous function.

Let us stress that the limit $N \longrightarrow \infty$ has to be taken with some care. There is no algorithm for how to make the transition from the state x^N in a $6N$-dimensional phase space to the state x^{N+1} in the larger $6(N+1)$-dimensional phase space. Thus, if we consider a given sequence of microstates $x^N = (x_1, \ldots, x_N)$ for $N = 1, 2, \ldots$, it might happen that the convergence described above holds, or it might not hold, depending on how the sequence is chosen, and there is nothing further to say.

However, a measure-theoretic procedure comes to the rescue. Let Γ_N, with $N = 1, 2, \ldots$ be a sequence of N-particle phase spaces, and let

$$\Delta_N = \{x^N \, : \, d(F^{[x^N]}, f) < \epsilon\} \tag{50}$$

be a sequence of sets of points in Γ_N whose corresponding exact distribution function is close to the continuous distribution function f. A sequence μ_N of probability measures for $N = 1, 2, \ldots$, each concentrated on Δ_N, i.e. such that $\mu_N(\Delta_N) = 1$, is said to be an *approximating sequence* for f.

One can thus formulate Lanford's theorem in the measure-theoretic terms introduced above: if μ_N is an approximating sequence for the initial distribution of states f_0, under what conditions is its time-evolution $\mu \circ T_{-t}$ an approximating sequence for the solution f_t of the Boltzmann equation at any subsequent time t?

> LANFORD'S THEOREM (MEASURE-THEORETIC VERSION)
> Suppose μ is an approximating sequence for f_0. If assumptions (i) and (ii) and the factorization condition (46) hold at the initial time $t = 0$, then there exists a finite positive time τ such that $\mu \circ T_{-t}$ is an approximating sequence for the solution f_t of the Boltzmann equation at all $t \in [0, \tau]$.

To put it differently, let us define the conditional probability function $\mu_{\Delta_N} := \mu(\cdot|\Delta_N)$, which assigns probability one to the points in the set Δ_N comprising all microstates x_0 such that the corresponding exact distribution function $F^{[x_0]}$ is close to f_0. Then, for any positive $t \leq \tau$

$$\lim_{N \longrightarrow \infty} \mu_{\Delta_N}(\{x \in \Gamma | d(F^{[x_t^N]}, f_t) < \epsilon\}) = 1 \tag{51}$$

That is, the probability of the set of phase points x_0 whose time-evolved microstates x_t are such that the corresponding exact distribution function $F^{[x_t^N]}$ is close to f_t conditionalized on Δ_N tends to one in the Boltzmann-Grad limit.

Under such a measure-theoretic interpretation of the theorem, the Boltzmann equation seems to give an accurate description of the time development of the overwhelming majority of initial microstates. Some points in phase space whose corresponding exact distribution function is close to f_0 would inevitably have a trajectories which do not agree with the Boltzmann equation. That is a consequence of the fact that the latter, contrary to the Hamiltonian equations of motion, is not invariant under-time reversal. Yet, the micro-dynamics and the macro-dynamics agree for most of the points in the N-particle phase space satisfying the initial conditions.

This means that for the initial microstates belonging to the region Δ_0 of phase space the entropy of the system is very likely to increase. It would thus appear that Lanford's theorem constitutes a statistical H-theorem, hence representing a counterpart of the thermodynamical approach to equilibrium in statistical mechanics. notice that although the time-scale τ of the theorem is very short, it is not too short to make irreversibility unobservable: in a duration of $\frac{1}{5}^{th}$ of the mean collision time, one expects that about 20 % of the particles will have collided, and this can be sufficient for a significant decrease in the H-function.

One might ask what happens if we consider negative times $-\tau < t < 0$. As pointed out by Lanford (1975, p. 109-110), and more explicitly by Lebowitz (1983, p. 9-10), exactly the same conclusions hold in that case too, provided one makes the following changes:

(a.) In condition (ii) we take $s \leq 0$.

(b.) The collision term $\mathcal{C}_{k,k+1}$ in the Boltzmann hierarchy is replaced by $-\mathcal{C}_{k,k+1}$

(c.) We systematically replace the configurations with incoming momenta by the outgoing momenta instead of *vice versa*, as we did in (32).

With this understanding, Lanford's theorem is clearly neutral with respect to time reversal: that is, we obtain convergence to solutions of the Boltzmann equation for positive times, and hence a decrease of H, as the H-theorem requires, but for negative times a convergence towards a solution of the 'anti-Boltzmann' equation, i.e. the Boltzmann equation with the sign of the collision integral reversed, and hence an increase of H . Accordingly, Lanford's theorem would itself be time-reversal invariant.

This is somewhat analogous to Boltzmann's 1897 argument based on the H-curve. Indeed, in this understanding, the theorem proves that for most initial microstates the H-function lies at a local peak of the H-function. So, at the initial time instant $t = 0$, $H[f_t]$ is expected to decrease in both directions of time. This offers a mathematical formalization of Boltzmann's claim that, apart from equilibrium, the most probable case is that the H-function is at a maximum of the curve. Thus, Lanford's result provides a rigorous version of the statistical H-theorem. Notice, however, that this understanding implies that the theorem would yield the wrong retrodictions, thus contradicting everyday experience which would lead one to expect that entropy of an isolated gas system should increase rather than decrease even during the interval $[-\tau, 0]$ (cf. Drory 2008).

5 Time-reversal invariance and Lanford's theorem

Lanford's theorem shows how one can derive the Boltzmann equation from the Hamiltonian equations of motion under precise assumptions. As a statistical version of Boltzmann's H-theorem, it seems to account for the approach to equilibrium for a general class of non-equilibrium initial conditions characterized by the regularity condition (i), at least during the time-interval $[0, \tau]$. The most important question is then how the implied irreversibility of this macro-evolution arises. On this point Lanford and other commentators on his theorem made remarks that are not quite univocal. We first survey and criticize these different views and then present our own argument on the emergence of irreversibility.

5.1 Views on the emergence of irreversibility in the literature

Lanford's first discussion of the issue of irreversibility concerns Boltzmann's heuristic derivation of the H-theorem. Here he writes:

> The inequality $dH/dt \leq 0$ shows that the reversibility of the underlying molecular dynamics has been lost when passing to the Boltzmann equation. The irreversibility must have been introduced in the Hypothesis of Molecular Chaos since the rest of the derivation was straightforward mechanics. Indeed it is not hard to see that the Hypothesis of Molecular Chaos is asymmetric in time; it gives a formula for the number of pairs of particles that are about to collide. If we write down an

> analogous expression for the number of pairs which have just undergone collision and repeat the argument, we obtain the Boltzmann collision term but with the sign reversed. One conclusion is that something more is involved in the Hypothesis of Molecular Chaos than simple statistical independence. [Lanford (1975), p.81]

Here of course we completely agree. [3]

Later, when presenting his own re-examination of the derivation of the Boltzmann equation from the BBGKY hierarchy, i.e. (32) specialized to the case $k = 1$, Lanford discusses a factorization condition which, in our notation, reads $\rho_2^{(a)}(x_1, x_2) = \rho_1^{(a)}(x_1)\rho_1^{(a)}(x_2)$, and comments:

> It must be pointed out that the factorization assumption, like the Hypothesis of Molecular Chaos to which it is evidently related, [4] is more subtle than it may appear. We obtained [the BBGKY hierarchy for the hard spheres model with the collision term expressed by equation (32) in the present paper] by systematically writing collision phase points in their incoming representations. We could have equally well have written them in their outgoing representations; if we then assumed factorization we would have obtained the Boltzmann collision term with its sign reversed. It is thus essential, in order to get the Boltzmann equation, to assume
> $$\rho_2^{(a)}(x_1, x_2) = \rho_1^{(a)}(x_1)\rho_1^{(a)}(x_2) \tag{52}$$
> for *incoming* collision points (x_1, x_2) and not for outgoing ones. [Lanford (1975), p.88]

While this quote is, by and large, consistent with the previous one in pointing out the distinction between an assumption for incoming collision as opposed to the same assumption for outgoing collision as responsible for the sign of the collision term in the Boltzmann equation, and hence for irreversibility, something subtle has changed.

In the intermediate pages, Lanford argued for the *identification* of phase points which differ only by having an incoming collision configuration replaced by the corresponding outgoing collision configuration. It is this demand of identification that makes a subsequent crucial distinction between incoming and outgoing collisions harder to explain. Recall that the appeal to a topology identifying the coordinates $(\vec{q}_i, \vec{p}_i; \vec{q}_j, \vec{p}_j)$ and $(\vec{q}_i, \vec{p}_i{}'; \vec{q}_j, \vec{p}_j{}')$ was introduced in order to assure the technical point that the hard-sphere dynamics becomes smooth. Lanford writes about these as two distinct *representations* of the same phase point. This suggests that the origin of irreversibility, rather than being a question of making either one of two substantially different assumptions, would lie instead in a conventional choice of representation. As Lanford (1975, p. 87) puts it "[T]hese two are really just different representations of the same phase point." Lebowitz (1983, p. 8) argues similarly when he writes about the incoming and outgoing momenta as being "just two different representations of the same phase point." However, we wish to object here that it is not clear at all how physical irreversibility can be due to a mere conventional choice of a representation.

In the final pages of his first paper, Lanford comes back to the issue of irreversibility. There he concludes:

> The Boltzmann hierarchy, like the Boltzmann equation is not invariant under time reversal. That is, irreversibility appears in passing to the limit $a \longrightarrow 0$, not

3. Note that in this quote the phrase "Molecular Chaos" refers the pre-collision condition, and not to the factorization condition alone.

4. Note that, contrary to the previous quote, the envisaged relation to "Molecular Chaos" here refers to the factorization condition rather than the pre-collision condition, as in Lanford's previous quote. (Note added by the present authors.)

> in the assumption that the rescaled correlation functions factorize. (Lanford 1975, p. 110)

This statement seems to put the blame somewhere else entirely. In contrast to the previous quotes, there is no mention here of the distinction between incoming and outgoing collision configurations; rather, the appearance of irreversibility is apparently due to the Boltzmann-Grad limit procedure. Of course there is also a constant feature in all these quotes, namely that Lanford consistently stressed that mere factorization is not in itself the explanation of irreversibility. This claim is indeed fortified by the fact that, as we saw, Lanford's theorem does not require a factorization condition to get convergence towards a solution of the Boltzmann hierarchy, but only to guarantee that the latter becomes equivalent to the Boltzmann equation. However, since irreversible behaviour already appears at the level of the Boltzmann hierarchy, factorization by itself, at least in the version adopted by Lanford, is surely not relevant for the emergence of irreversibility. We shall come back to this at the end of the section.

Still, Lanford's later papers seem to put yet another gloss on the issue. After a discussion of the limit $N \longrightarrow \infty$ (which in the Boltzmann-Grad procedure is equivalent to $a \longrightarrow 0$) he writes in his (1981):

> None of this, however, really implies that irreversible behavior *must* occur in the limiting regime; it merely makes this behavior plausible. For a really compelling argument in favor of irreversibility, it seems to be necessary to rely on some version of Boltzmann's original proof of the H-theorem (Lanford 1981, p.75).

To be sure, the irreversible approach to equilibrium does not follow from taking the limit for $N \longrightarrow \infty$, as it has been argued e.g. by Goldstein (2001). As Lanford (1981) pointed out, a counterexample is given by the Vlasov limit. This is a limiting regime in which the interaction between particles is given by a sum of two-body potentials of the form

$$\phi^{(N)}(\vec{q}_1 - \vec{q}_2) = \frac{1}{N}\phi_0(\vec{q}_1 - \vec{q}_2) \tag{53}$$

where $\phi_0(\vec{q}_1 - \vec{q}_2)$ is a fixed singularity-free potential. The macroscopic distribution function $f(\vec{q}, \vec{p})$ is in this case determined by the microstates of the system in the same way just as in the Boltzmann-Grad limit. However, the time-evolution of f is given by the Vlasov equation, rather than by the Boltzmann equation. In this case one can show that the H-function remains constant through time, and hence the limit does not lead to irreversibility. Yet, in this 1981 paper, Lanford did not specify how "some version of Boltzmann's original proof of the H-theorem" would provide a compelling argument in favor of irreversibility. Be it as it may, there does not seem to be room even for Lanford's claim of plausibility about the emergencence of irreversible behaviour in the limiting regime, even in the case of the hard-sphere model. Indeed, while Lanford's theorem does show that the Boltzmann hierarchy, which is not time-reversal invariant, is somehow derived from the BBGKY hierarchy (28), which is time-reversal invariant, by taking the Boltzmann-Grad limit, our Proposition 3 in the Appendix demonstrates that the BBGKY hierarchy with the collision operator expressed by (32) is *not* time-reversal invariant. Therefore, the time-asymmetric ingredient introducing irreversibility does not depend in any way on the Boltzmann-Grad limit, it must lie hidden in the passage leading to the transformation of the collision term (29) into (32).

An entirely different argument for the emergence of irreversibility is presented by Spohn (1980,1991) and Lebowitz (1983). Spohn (1980, p. 596) devotes a paragraph to the question "how Lanford's theorem escapes the conflict the reversible character and the irreversible character of the Bolztmann hierarchy. (A similar passage is found in (Spohn 1991, p. 66)). He points out how Lanford's theorem will not sustain the construction of a counterexample as in the original reversibility objection (section 2.1). Recall that in that construction we assume an initial distribution function $f_0(x)$ that evolves in accordance with the Boltzmann equation from the initial time 0 to some positive time t, when the distribution function is $f_t(x)$, and then suddenly all the velocities of the particles are reversed. Due to the time-reversal invariance of the microdynamics, $H[f]$ would then have to increase during the interval $[t, 2t]$.

Spohn considers what happens if we try to run this same argument on the basis of Lanford's theorem. The crucial point in his analysis is that the set $\Gamma_{k,=}(t)$ of phase points for which the theorem does not hold increases with time. Hence if we consider the rescaled densities $\rho_{k,t}^{(a)}$ at time t, such that $0 < t < \tau$, and reverse the velocities, the ensuing evolution of these functions will no longer be guaranteed by Lanford's theorem, since $\Gamma_{k,=}(t) \supset \Gamma_{k,=}(0)$. A more elaborate version of the argument is given by Lebowitz (1983).

The analysis by Spohn and Lebowitz is clearly correct, and convincingly shows that the reversibility objection cannot be run against Lanford's theorem in the same way as is it was used by Loschmidt and Culverwell against Boltzmann's original presentation of the H-theorem. However, we believe that it is one thing to show how the reversibility objection is evaded, but it is quite another thing to explain the emergence of irreversibility in Lanford's theorem. And although the Spohn-Lebowitz argument is successful in the first objective, we feel it does little to offer the sought-after explanation. After all, the difference between the measure-zero sets $\Gamma_{k,=}(s)$ and $\Gamma_{k,=}(s + t)$ is also a measure zero set. The difference between these sets is, admittedly, important in the mathematical convergence conditions of the theorem. But in the spirit of the derivation, these measure zero sets in phase-space are not to be held physically significant. Indeed, to explain the emergence of irreversibility, one would like to understand why the overwhelming majority of phase space points that approximate the initial distribution function f_0 will evolve for time $0 < t < \tau$ in accordance with the Boltzmann equation, rather than the anti-Boltzmann equation. The consideration of measure-zero sets like $\Gamma_{k,=}(t)$ will not be helpful in this regard.

Cercignani, Illner & Pulvirenti (1994) and Cercignani (2008) have another analysis. After an informal discussion of the problem, they write:

> In fact, we remark that when giving a justification of the Boltzmann equation in the previous chapter, we used the laws of elastic collisions and the continuity of the probability density at the impact to express the distribution functions corresponding to an after-collision state in terms of the distribution corresponding to the state before the collision, rather than the latter in terms of the former. It is obvious that the first way is the right one to follow if the equations are used to predict the future from the past and *not vice versa*; it is clear, however, that this choice introduced a connection with the everyday concepts of past and future which are extraneous to molecular dynamics and are based on our macroscopic experience. [...] [A] striking consequence of our choice is that the Boltzmann equation describes motions for which the quantity H has a tendency to decrease, while the opposite choice would have led to an equation having a negative sign in front of the collision term and hence describing motions with increasing H (Cercignani, Illner & Pulvirenti 1994, p. 53).

This quote clearly expresses the idea that the preference for the incoming (or ingoing) configurations is responsible for the emergence of irreversibility. At the same time, it states that the motivation for such a preference cannot be based on dynamics, but rather presupposes macroscopic experience, which allows us to distinguish between past and future. Nevertheless, a later passage from the same book squarely contradicts this idea:

> We are compelled to ask whether the representation in terms of ingoing configurations is the right one, i.e. physically meaningful. As we shall later see, in a more careful analysis of the validity problem, the representation in terms of ingoing configurations follows automatically from hard-spheres dynamics and is, indeed, not a matter of an a priori choice (Cercignani, Illner & Pulvirenti 1994, p. 74).

It is unclear to us how to reconcile these conflicting statements. The more so because the "more careful analysis" promised in the above quote refers to a subsequent passage in their book (pp. 77–81) where the choice of the incoming (or ingoing) configuration is assumed from the outset in order for the calculations to go through. We have been unable to find any explicit demonstration that the incoming representation is forced upon us by the hard-spheres dynamics at all.

5.2 The role of the choice of incoming configurations in the theorem

It is rather striking that all of the authors we have considered explicitly recognize the fact that the procedure of choosing incoming momenta configurations above outgoing configurations, as used here in the transition from the collision term as given by(29) to that given by (32), is crucial for obtaining the 'right' sign of the collision term in the Boltzmann equation or hierarchy, and hence for the derivation of an irreversible approach to equilibrium; and that choosing the outgoing configurations instead would lead one to obtain the anti-Boltzmann equation. However, none of the above authors considers this procedure to be crucial to the explanation of the emergence of irreversibility, and indeed, no-one even mentions this choice as a crucial ingredient in Lanford's theorem. Rather, the consensus point of view seems to be that this step from (29) to (32) involves no more than a conventional choice between equivalent representations of the same phase point. The only exception to this consensus is the viewpoint expressed by Cercignani, Illner & Pulveirenti, quoted above, to the effect that only one of these representations is "right", or "physically meaningful".

In contrast, we believe that it is this predilection for writing the BBGKY hierarchy in terms of incoming configurations, i.e. the transition from the collision term (29) to (32) itself, instead of the Boltzmann-Grad limit, as argued by Lanford; the consideration of special measure-zero sets, as argued by Spohn and Lebowitz; or the hard-spheres dynamics, as argued by Cercignani, Illner & Pulvirenti; that is crucial for explaining the emergence of irreversibility in Lanford's theorem. Also, we claim that this predilection is not a matter of a conventional choice between different representations of the same phase-point. Furthermore, we argue that this crucial choice, rather than being derived from the hard-spheres dynamics itself, is in fact, and ought to be recognized as, an independent ingredient of Lanford's theorem. We address these claims in more detail below.

First, let us stress that, as we pointed out in the previous section, the upshot of our Proposition 3 is that the source of irreversibility must lie in the passage leading to the BBGKY hierarchy with the collision term (32). Now, splitting the integral over the unit sphere S^2 in (29) and changing the integration variable $\omega_{i,k+1}$ into $-\omega_{i,k+1}$ are nothing but mere mathematical procedures. Thus, one is left with identifying the adoption of the incoming configurations of collision points as the element introducing the time-asymmetry.

Let us make this point more explicit by considering what happens if we make the 'opposite' choice. In that case, going back to Eqn. (29), we would again split the integral over S^2 into two hemispheres, but on the hemisphere with $\vec{\omega}_{i,k+1} \cdot (\vec{p}_{k+1} - \vec{p}_i) \geq 0$ we would leave the configuration for the pair $(\vec{q}_i\vec{p}_i, \vec{q}_i + a\vec{\omega}_{i,k+1}, \vec{p}_{k+1})$ as it is (i.e. outgoing). On the hemisphere characterized by $\vec{\omega}_{i,k+1} \cdot (\vec{p}_{k+1} - \vec{p}_i) \leq 0$ we replace the coordinates $(\vec{q}_i\vec{p}_i, \vec{q}_i + a\vec{\omega}_{i,k+1}, \vec{p}_{k+1})$ by $(\vec{q}_i\vec{p}_i{}', \vec{q}_i + a\vec{\omega}_{i,k+1}, \vec{p}_{k+1}')$ as well as change the sign of the integration variable $\vec{\omega}_{i,k+1}$ so that the two hemisphere integrals have a common domain. The result is that Eqn. (29) would go over in

$$\check{\mathcal{C}}^{(a)}_{k,k+1}\rho^{(a)}_{k+1,t}(x_1, \ldots, x_k) = Na^2 \sum_{i=1}^{k} \int_{\vec{\omega}_{i,k+1}\cdot(\vec{p}_i - \vec{p}_{k+1})\geq 0} d\vec{\omega}_{i,k+1} d\vec{p}_{k+1}\ \vec{\omega}_{i,k+1} \cdot (\vec{p}_i - \vec{p}_{k+1})$$
$$\times[\rho^{(a)}_{k+1}(x_1, \ldots, \vec{q}_i, \vec{p}_i, \ldots x_k, \vec{q}_i + a\vec{\omega}_{i,k+1}, \vec{p}_{k+1}) - \rho^{(a)}_{k+1}(x_1, \ldots, \vec{q}_i, \vec{p}_i{}', \ldots, \vec{q}_i - a\vec{\omega}_{i,k+1}, \ldots \vec{p}_{k+1}')] \tag{54}$$

as an alternative to (32). *Proposition 4* in the Appendix shows that neither the BBGKY hierarchy with the collision term expressed by (54) is time-reversal invariant. This indicates that the true source of irreversibility lies in the adoption of either one between the incoming and the outgoing configurations of collision points.

Next, we discuss the issue of whether the predilection for the incoming over the outgoing configurations may be due to a mere conventional choice of representation. We argue that this is a misleading line of thought. In more detail, we object that the incoming and the outgoing configurations could be seen as nothing more than different representations of the same physical phase point, as claimed by Lanford and Lebowitz. In fact, the very use of the terminology "representation" appears quite inappropriate in this context.

Recall that the alleged identification between incoming and outgoing collision points arises by a specific choice of topology. This point may perhaps be elucidated by considering what happens if one takes a series of smooth spherically symmetrical pair potentials ϕ in (11) that approaches the hard-spheres model (12). In such a case, whenever a collision between two particles, say i and j, occurs, the momenta of the particles do not change instantaneously from incoming values $(\vec{p}_i, \vec{p}_j)$ to outgoing values $(\vec{p}_i{}', \vec{p}_j{}')$ as given by (6), but by some smooth trajectory in a non-zero time span (cf. Fig. 2.). When we take the hard-sphere limit for such a collision (i.e., if we let the pair potential ϕ approach the hard-sphere potential (12), this interval goes to zero, and the trajectory would jump instantaneously from x_{in} to x_{out}. Now, we can then still regard this hard-sphere collision as a continuous process by adopting a topology on $\Gamma^{(a)}_{\neq}$ in which the holes in this phase space are removed, so that incoming collision coordinates and the outgoing collision coordinates become, as it were, adjacent to each other in phase space, and a trajectory that jumps from $x_{\text{in}} = (x_1, \ldots, x_{i-1}, \vec{q}_i, \vec{p}_i, \ldots, \vec{q}_j, \vec{p}_j, x_{j+1}, \ldots x_n)$ to $x_{\text{out}} = (x_1, \ldots, x_{i-1}, \vec{q}_i, \vec{p}_i{}', \ldots, \vec{q}_j, \vec{p}_j{}', x_{j+1}, \ldots x_n)$ is regarded as continuous (Fig. 3). One may express this fact colloquially as an "identification" of these two points. Indeed, we are free in adopting any topology we like on the boundary of $\Gamma^{(a)}_{\neq}$ (as long as it extends the Euclidean topology on its interior), and in particular we can choose a topology to make an instantaneous transition from x_{in} to x_{out} appear as a smooth trajectory. Such a choice of topology entail that every metric, or distance function, d on $\Gamma^{(a)}_{N\neq}$, compatible with it would have the property that $d(x_{\text{in}}, x_{\text{out}}) = 0$, and hence (by the usual definition of a metric) it would follow that $x_{\text{in}} = x_{\text{out}}$, i.e. those points are identified. But when choosing a topology, we are not forced to introduce a metric. Moreover, even if we identify the incoming and outgoing points x_{in} and x_{out} for the purpose of topological or metrical considerations, it does not follow that they thereby are *physically* identical. Indeed, that would overlook the distinctive relevant fact that the momenta are quite different in these two points! In other words, all that this choice of topology enforces is that a trajectory connecting points like $x_{\rm in}$ and $x_{\rm out}$ becomes smooth, but not that these points are physically one and the same.

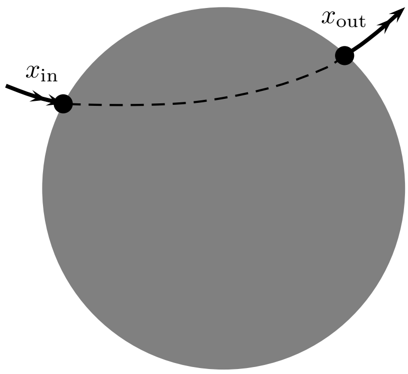

Figure 2: A region of the phase space $\Gamma^a_{N,\neq}$ showing a "hole" (gray area) due to the forbidden overlap of hard spheres i and j. The points $x_{\rm in}$ and $x_{\rm out}$ represent the microstate immediately before and after the collision between particles i and j. The dashed curve between them denotes the continuous trajectory obtained these points in a smooth potential approximation to the hard spheres potential.

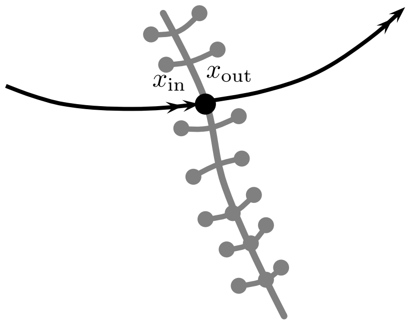

Figure 3: The same region of the phase space $\Gamma^a_{N,\neq}$ with the hole sown up, and the points $x_{\rm pre}$ and $x_{\rm out}$ identified. The phase space trajectory is now smooth even during the hard spheres collision.

Finally, we submit that the choice for the incoming collision configuration is needed as an independent ingredient for the validity of Lanford's theorem. Indeed, it is crucially needed to derive the desired result: If we take the phrase "solutions of the BBGKY hierarchy" in this theorem to refer to the BBGKY hierarchy with the original BBGKY collision term (29), the theorem is false; similarly, if we would write the collision term in the form (54) the theorem would be false too (it would imply convergence to the anti-Boltzmannn hierarchy). The theorem does hold, however, if we adopt the collision term (32) in the BBGKY hierarchy, but as we have argued, the choice of this form of the collision term is not warranted by an appeal to dynamics or topology, and should be explicitly recognized as an independent assumption. A change in this choice will produce radically different conclusions. In particular, if one does not make this choice, one does not succeed in obtaining convergence of $\rho^{(a)}_{k,t}(x_1, \ldots x_k)$ towards the functions $f_{k,t}(x_1, \ldots x_k)$ that solve the Boltzmann hierarchy. As we have noted above, if one chooses the outgoing configurations, one would derive a different collision operator: in particular, $\check{\mathcal{C}}^{(a)} = -\hat{\mathcal{C}}^{(a)}$. Accordingly, one would arrive at the conclusion that if the initial values $\rho^{(a)}_{k,0}$ of the rescaled probability densities converge to $f_{k,0}$, then $\rho^{(a)}_{k,t}$ will converge to solutions of the anti-Boltzmann hierarchy with $f_{k,0}$ as initial conditions. And for the measure-theoretic reformulation of the theorem, one would now have to conclude that "most" phase points which approximate some initial shape of a distribution function f_0 will actually hug closely to a solution of the anti-Boltzmann equation for a time $0 < t \leq \tau$, instead of a solution of the Boltzmann equation. Therefore, as the goal of Lanford's theorem is to obtain the Boltzmann equation, one must opt for the incoming configurations. The more so because for a solution f_t of the anti-Boltzmann equation, the entropy $-H(f_t)$ decreases monotonically in the course of time, and thus the theorem would not agree with our observations at the macroscopic level. This further enforces our claim that the incoming and the outgoing configurations cannot be identified, in that they lead to entirely different predictions.

Since the question of whether we derive $dH/dt \leq 0$ or $dH/dt \geq 0$ is a substantive issue, such a difference cannot be a latter of mere convention. Rather, for Lanford's result to be empirically correct, the choice of the incoming configurations is to be adopted. One is then left with two options: either it is an automatic consequence of the hard-spheres dynamics, as argued by Cercignani e.a., or it is introduced as an independent ingredient in the theorem. However, the hard-spheres dynamics itself is clearly neutral towards the distinction between incoming and outgoing configurations and does not entail preference for one of these. The only option is thus that the choice of the incoming configurations must be regarded as an explicit assumption in the statement of the theorem.

Let us conclude by stressing that the adoption of the incoming configurations as an independent condition in Lanford's theorem plays a similar role as the pre-collision condition in the Stoßzahlansatz in Boltzmann's original H-theorem, in that it is responsible for the emergence of irreversibility. There is, however, a perhaps surprising, crucial difference. If one were to replace the pre-collision condition in the Stoßzahlansatz with a post-collision condition, which applies only to outgoing rather than incoming particles, one would derive the anti-Boltzmann equation. That is equivalent to apply a time-reversal transformation to the Boltzmann equation. To the contrary, *Proposition 5* in our Appendix shows that adopting the outgoing collision configurations for the BBGKY hierarchy is *not* equivalent to adopting the incoming collision configurations for the BBGKY hierarchy followed by applying a time-reversal transformation. Indeed, the resulting hierarchies of evolution equations are different. To be sure, once one takes the Boltzmann-Grad limit, they both reduce to the anti-Boltzmann hierarchy, but the fact remains that the outgoing configurations should not be regarded, so to speck, as the time-reversal counterpart of the incoming configurations. A lesson one may draw from this result is that, contra Lanford and Lebowitz's proposal we reported toward the end of section 4.1, one should not adopt the outgoing configurations for negative times. Instead, one can insist on the incoming configurations, thus consistently obtaining the Boltzmann equation even for $t < 0$.

6 Conclusion

We have investigated the problem of whether and how Lanford's theorem explains the emergence of irreversibility in the macro-evolution equations like the Boltzmann equation or the Boltzmann hierarchy; in particular, we discussed whether the theorem represents a rigorous version of the statistical H-theorem. We have seen that the theorem obtains the approximate validity of the Boltzmann equation from the Hamiltonian equations of motion by providing a link between the Boltzmann hierarchy and the BBGKY hierarchy for the hard spheres model in the Boltzmann-Grad limit. We then criticized the different analysis offered in the literature as to where irreversibility in Lanford's result comes from. We argued, *contra* Spohn and Lebowitz, that it cannot be explained by any measure-theoretic considerations. Instead, the culprit for the emergence of irreversibility lies in the choice of the incoming configurations of collision points, which is introduced when casting the BBGKY hierarchy in a form amenable to derive the Boltzmann hierarchy. However, such a choice is not merely conventional, as Lanford maintains, nor does it follow from the hard-spheres dynamics, as Cercignani e.a. suggest. Rather, the adoption of the incoming collision representation is an independent ingredient which is necessary in order to obtain the sought-after results. In fact, our main claim is that it ought to be included as an explicit assumption of the theorem.

We conclude the paper by pointing out some remarks worth noting about Lanford's theorem.

First of all, there is an issue stressed by Lanford himself and nearly all subsequent commentators. It concerns the validity of his result, that is the fact that the theorem holds for a time length τ of the order of 1/5 the mean free time. Since this time scale, for realistic gas systems under ordinary circumstances, will be of the order of a milliseconds, the theorem will hardly be enough ammunition to provide a justification of the Boltzmann equation through macroscopic time scales, or even the time scale in which equilibration sets in. It is true that Illner and Pulvirenti (1986, 1989), have derived a longer validity but only under much more stringent conditions, i.e. for a gas cloud expanding into a vacuum.. As a matter of fact, this repeated attention to time scale has deluded views from more serious problems. Indeed Lanford already pointed out that there is a simple, if merely technical, "fix" to the above problem: one would only need to require that assumption (i) of the theorem holds for arbitrary times, and not just at $t = 0$, and Lanford's result may be extended to all times. This issue will be taken up somewhere else [5].

However, in our opinion, a more serious drawback to the applicability and physical relevance of Lanford's theorem lies in the usage of the Boltzmann-Grad limit. As we have seen, this limit implies that the density of the gas goes to zero, and hence that the result applies to infinitely diluted gases. And while it seems reasonable to impose this limit in order to give the Boltzmann equation a fighting chance to be valid, it also means that the thus-obtained result can hardly be relevant to real-life gas systems in which the density is not close to zero. The main merit of Lanford's theorem is therefore conceptual, in that it makes a case that, under precise conditions on the initial data, the Boltzmann equation can be derived from Hamiltonian mechanics, although just in rather idealized circumstances.

5. See Valente (2010).

7 Appendix

In this appendix we consider in more detail the issue of time-reversal invariance for the Boltzmann equation and two versions of the BBGKY hierarchy equations and prove the claims concerning this issue in the body of the paper. The commonly accepted criterium for judging time-reversal invariance of such evolution equations, describing a density function (either f or ρ_k) over particle configurations is as follows: the equation determines a class S of allowed solutions, where each solution can be seen as a 'history' of the density function, either $H := \{f_t,\ t \in \mathbb{R}\} \in S$ or $H := \{\rho_{k,t},\ t \in \mathbb{R}\}$, carving out, so to say, a trajectory in their respective spaces of all conceivable density functions.

Now consider a time-reversed version of such a history, defined as $\mathcal{T}H := \{\bar{f}_{-t}\ t \in \mathbb{R}\}$ or $\mathcal{T}H := \{\bar{\rho}_{k,-t}\ t \in \mathbb{R}\}$, where $\bar{f}_t$ and $\bar{\rho}_{k,t}$ are obtained by reversing the momenta in their arguments: $\bar{f}_t(\vec{q}, \vec{p}) = f_t(\vec{q}, -\vec{p})$, and $\bar{\rho}_t(\vec{q}_1, \vec{p}_1; \ldots; \vec{q}_k, \vec{p}_k) = \rho_{k,t}(\vec{q}_1, -\vec{p}_1; \ldots; \vec{q}_k, -\vec{p}_k)$. The question is then whether such a time-reversed history $\mathcal{T}H$ is a solution of the equation too, whenever H is a solution of the equation in question. In other words: Is $\mathcal{T}H \in S$ whenever $T \in S$? If the answer is yes, the equation is time-reversal invariant, otherwise not.

Now in all cases considered below, we are dealing with equations that are first-order differential equations in time. Hence a solution is in principle fixed by choosing an initial value condition f_{t_0} or ρ_{k,t_0} at some instant of time t_0. Thus, it is sufficient for the purpose of determining the time-reversal invariance of these equations to study how they transform under a replaceal of $t \longrightarrow -t$, and $f_{t_0} \longrightarrow \bar{f}_{t_0}$ or $\rho_{k,t_0} \longrightarrow \bar{\rho}_{k,t_0}$.

Moreover all equations we consider below are invariant under time translation. Therefore, it is immaterial which instant of time is taken as the origin of the reversal. That is to say, it does not matter whether one takes $t \longrightarrow -t$ or $(t - t_0) \longrightarrow -(t - t_0)$, for any value of t_0. We will take advantage of this by supposing that the time $t = t_0$ in the equations studied below is the origin of the time reversal, and therefore invariant under the transformation, so that the only temporal change needed in the transformation is $\frac{\partial}{\partial t} \longrightarrow -\frac{\partial}{\partial t}$.

Our strategy will be the same in all three cases. We consider an arbitrary solution of the equation and construct from this a time-reversed history, and derive the equation it obeys. If the resulting equation is equivalent to the original equation we have proved that the original equation is time-reversal invariant. But if it is not, the original equation is not time-reversal invariant.

Proposition 1. The Boltzmann equation (8) is not time-reversal invariant.

Proof. Since nothing interesting happens to the position variables in (8) we will suppress them in the notation below, and also put the mass $m = 1$. Note that $\vec{p}_1$ is the only independent momentum variable in the equation: $\vec{p}_2$ appears only in the right-hand side as a mere integration variable and the outgoing momenta variables $\vec{p}_1{}', \vec{p}_2{}'$ in the collision integral are functions of $\vec{p}_i$: that is, $\vec{p}_i{}' = \vec{p}_i{}'(\vec{p}_1, \vec{p}_2) = T_{\omega_{12}}(\vec{p}_1, \vec{p}_2)$. So, under the transformations $\partial/\partial t \longrightarrow -\partial/\partial t$ and $\vec{p}_1 \longrightarrow -\vec{p}_1$, the left-hand side of (8) transforms into

$$-\frac{\partial}{\partial t} f_t(-\vec{p}_1) - \vec{p}_1 \cdot \frac{\partial}{\partial \vec{q}} f_t(-\vec{p}_1) \quad = \quad -\frac{\partial}{\partial t} \bar{f}_t(\vec{p}_1) - \vec{p}_1 \cdot \frac{\partial}{\partial \vec{q}} \bar{f}_t(\vec{p}_1) \tag{55}$$

while the right-hand side of (8) becomes

$$Na^2 \int d\vec{p}_2 \int_{\vec{\omega}_{12} \cdot (\vec{p}_1 + \vec{p}_2) \leqslant 0} d\vec{\omega}_{12} \, (-\vec{p}_1 - \vec{p}_2) \cdot \vec{\omega}_{12} \; [f_t(\vec{p}_1{}'') f_t(\vec{p}_2{}'') - f_t(-\vec{p}_1) f_t(\vec{p}_2)] \tag{56}$$

Here, the notation $\vec{p_i}''$ is used to indicate that these momenta have be thought of as functions of $(-\vec{p}_1, \vec{p}_2)$: $(\vec{p_1}'', \vec{p_2}'') = T_{\omega_{12}}(-\vec{p}_1, \vec{p}_2)$.

If we now perform an additional (cosmetic) transformation of the integration variables $\vec{p}_2 \longrightarrow -\vec{p}_2$ and $\omega_{12} \longrightarrow -\omega_{12}$, (56) can also be written as

$$Na^2 \int d\vec{p}_2 \int_{\vec{\omega}_{12}\cdot(\vec{p}_1-\vec{p}_2)\geqslant 0} d\vec{\omega}_{12}\, (\vec{p}_1 - \vec{p}_2)\cdot\vec{\omega}_{12}\; [f_t(\vec{p_1}''')f_t(\vec{p_2}''') - f_t(-\vec{p}_1)f_t(-\vec{p}_2)] \tag{57}$$

where $(\vec{p_1}''', \vec{p_2}''') := T_{\omega_{12}}(-\vec{p}_1, -\vec{p}_2)$. But $T_{\omega_{12}}$ is a linear operator, and therefore $\vec{p_i}''' = -\vec{p_i}'$. If we now substitute back $f_t(\vec{p}) = \bar{f}_t(-\vec{p})$ in (57) and equate the transformed left-hand side (55) of (8) to the transformed right-hand side (57) of (8), we find that the reversed solution satisfies the equation

$$-\frac{\partial}{\partial t}\bar{f}_t(\vec{p}_1) - \vec{p}_1\cdot\frac{\partial}{\partial\vec{q}}\bar{f}_t(\vec{p}_1) = \\ Na^2 \int d\vec{p}_2 \int_{\vec{\omega}_{12}\cdot(\vec{p}_1-\vec{p}_2)\geqslant 0} d\vec{\omega}_{12}\, (\vec{p}_1 - \vec{p}_2)\cdot\vec{\omega}_{12}\; \left[\bar{f}_t(\vec{p_1}')\bar{f}_t(\vec{p_2}') - \bar{f}_t(\vec{p}_1)\bar{f}_t(\vec{p}_2)\right] \tag{58}$$

also known as the anti-Boltzmann equation. We conclude: whenever the solution $\{f_t \;, t \in \mathbb{R}\}$ satisfies the Boltzmann equation, the time reversed solution $\{\bar{f}_{-t} \;, t \in \mathbb{R}\}$ solves the (inequivalent) anti-Boltzmann equation, and therefore, the Boltzmann equation is not time-reversal invariant. □

Proposition 2. The BBGKY hierarchy with the collision term expressed by (29) is time-reversal invariant.

Proof. Recall that the BBGKY hierarchy has the form:

$$\frac{\partial\rho^{(a)}_{k,t}(x_1,\ldots,x_k)}{\partial t} - \mathcal{H}_k\rho^{(a)}_{k,t}(x_1,\ldots,x_k) = \left(\mathcal{C}^{(a)}_{k,k+1}\rho^{(a)}_{k+1,t}\right)(x_1,\ldots,x_k) \tag{59}$$

In this equation, we deal with k particles (the momentum of the $k+1$th particle appears in (29) only as an integration variable). If we reverse sign of the momenta $\vec{p}_1, \ldots \vec{p}_k$ and the sign of $\partial/\partial t$ it is easy to see the the left hand side of (65) changes sign. But here, the right-hand side (29) clearly changes sign too when we change sign of all momenta $\vec{p}_1, \ldots, \vec{p}_{k+1}$, due to the fact that the integration over the antisymmetric factor $\big(\vec{\omega}_{i,k+1}\cdot\big(\vec{p}_{k+1} - \vec{p}_i\big)\big)$ in the integrand is extended over the entire unit sphere. More explicitly, if we use the notation $\tilde{x}_i = (\vec{q}_i, -\vec{p}_i)$ along $x_i = (\vec{q}_i, \vec{p}_i)$, the transformed version of the left-hand side of equation (59) is

$$-\frac{\partial}{\partial t}\rho^{(a)}_{k,t}(\tilde{x}_1,\ldots,\tilde{x}_k) + \mathcal{H}_k\rho^{(a)}_{k,t}(\tilde{x}_1,\ldots,\tilde{x}_k) = -\frac{\partial}{\partial t}\bar{\rho}^{(a)}_{k,t}(x_1,\ldots,x_k) + \mathcal{H}_k\bar{\rho}^{(a)}_{k,t}(x_1,\ldots,x_k) \tag{60}$$

while the right-hand side transforms into:

$$\left(\mathcal{C}^{(a)}_{k,k+1}\rho^{(a)}_{k+1,t}\right)(\tilde{x}_1,\ldots\tilde{x}_k) = Na^2\sum_{i=1}^{k}\int_{\mathbb{R}^3} d\vec{p}_{k+1}\int_{S^2} d\vec{\omega}_{i,k+1}\,\big(\vec{\omega}_{i,k+1}\cdot\big(\vec{p}_{k+1}+\vec{p}_i\big)\big) \\ \times\, \rho^{(a)}_{k+1}(\tilde{x}_1,\ldots,\tilde{x}_k, \vec{q}_i + a\vec{\omega}_{i,k+1}, \vec{p}_{k+1}) \tag{61}$$

Hence, if we rewrite the integration variable $\vec{p}_{k+1}$ as $-\vec{p}_{k+1}$, we obtain from (61):

$$\begin{aligned}\left(\mathcal{C}^{(a)}_{k,k+1}\rho^{(a)}_{k+1,t}\right)(\tilde{x}_1,\ldots,\tilde{x}_k) &= -Na^2\sum_{i=1}^{k}\int_{\mathbb{R}^3}d\vec{p}_{k+1}\int_{S^2}d\vec{\omega}_{i,k+1}\left(\vec{\omega}_{i,k+1}\cdot\left(\vec{p}_{k+1}-\vec{p}_i\right)\right)\\ &\quad\times\rho^{(a)}_{k+1}(\tilde{x}_1,\ldots,\tilde{x}_k,\vec{q}_i+a\vec{\omega}_{i,k+1},-\vec{p}_{k+1})\\ &= -Na^2\sum_{i=1}^{k}\int_{\mathbb{R}^3}d\vec{p}_{k+1}\int_{S^2}d\vec{\omega}_{i,k+1}\left(\vec{\omega}_{i,k+1}\cdot\left(\vec{p}_{k+1}-\vec{p}_i\right)\right)\bar{\rho}^{(a)}_{k+1}(x_1,\ldots,x_k,\vec{q}_i+a\vec{\omega}_{i,k+1},\vec{p}_{k+1})\end{aligned} \tag{62}$$

Comparing this with (29), we conclude

$$\left(\mathcal{C}^{(a)}_{k,k+1}\rho^{(a)}_{k+1,t}\right)(\tilde{x}_1,\ldots\tilde{x}_k) = -(\mathcal{C}^{(a)}_{k,k+1}\bar{\rho}^{(a)}_{k+1,t})(x_1,\ldots,x_k) \tag{63}$$

Putting (60) and (63) together, we see that the time-reversed version of an arbitrary solution of (59) obeys the equivalent equation

$$-\frac{\partial}{\partial t}\bar{\rho}^{(a)}_{k,t}+\mathcal{H}_k\bar{\rho}^{(a)}_{k,t} = -\mathcal{C}^{(a)}_{k,k+1}\bar{\rho}^{(a)}_{k+1,t} \tag{64}$$

This shows that if $\{\rho^{(a)}_{k+1,t}\,, t\in\mathbb{R}\}$ solves equation (59), then $\{\bar{\rho}^{(a)}_{k+1,-t}\,, t\in\mathbb{R}\}$ solves the same equation, so that we can conclude that (59) is time-reversal invariant. □

Proposition 3. The BBGKY hierarchy with the collision term expressed by (32) is not time-reversal invariant.

Proof. Recall that, after adopting the incoming representation for collision points, the BBGKY hierarchy takes the form:

$$\frac{\partial\rho^{(a)}_{k,t}(x_1,\ldots,x_k)}{\partial t}-\mathcal{H}_k\rho^{(a)}_{k,t}(x_1,\ldots,x_k) = \left(\hat{\mathcal{C}}^{(a)}_{k,k+1}\rho^{(a)}_{k+1,t}\right)(x_1,\ldots,x_k) \tag{65}$$

where the left-hand side is the same as in (59), but the collision operator is now expressed by (32).

Here, we are again dealing with k particles, but in this case both incoming and outgoing momenta appear in the same formula, just as in the Boltzmann equation. And just as in the Boltzmann equation, one ought to take the outgoing momenta variables here as (implicit) functions of the incoming momenta: $(\vec{p}_i{}', \vec{p}'_{k+1}) = T_{\omega_{i,k+1}}(\vec{p}_i,\vec{p}_{k+1})$.

We now apply a combination of the arguments we used above to judge the time-reversal invariance of the Boltzmann equation and the BBGKY hierarchy in the version (59): we replace $\partial t/\partial t$ by $-\partial t/\partial t$ and $(x_1,\ldots,x_k)$ by $(\tilde{x}_1,\ldots,\tilde{x}_k)$. Since the left-hand side is the same as in (59), we draw the same conclusion: this side transforms into (60). But we have to scrutinize the behaviour of the right-hand side in more detail. This side transforms to:

$$\begin{aligned}\left(\hat{C}^{(a)}_{k,k+1}\rho^{(a)}_{k+1}\right)(\tilde{x}_1,\ldots,\tilde{x}_k) &= -Na^2\sum_{i=1}^{k}\int_{\vec{\omega}_{i,k+1}\cdot(\vec{p}_i+\vec{p}_{k+1})\leqslant 0}d\vec{p}_{k+1}d\vec{\omega}_{i,k+1}\,\vec{\omega}_{i,k+1}\cdot(\vec{p}_i+\vec{p}_{k+1})\\ &\quad\times\big[\rho^{(a)}_{k+1}(\tilde{x}_1,\ldots,\tilde{x}_{i-1},\vec{q}_i,\vec{p}_i{}',\tilde{x}_{i+1},\ldots\tilde{x}_k,\vec{q}_i-a\vec{\omega},\vec{p}'_{k+1})\\ &\quad-\rho^{(a)}_{k+1}(\tilde{x}_1,\ldots,\tilde{x}_{i-1},\vec{q}_i,-\vec{p}_i,\tilde{x}_{i+1},\ldots\tilde{x}_k,\vec{q}_i+a\vec{\omega},-\vec{p}_{k+1})\big]\end{aligned} \tag{66}$$

Where, as before, the variables $(\vec{p}_i{}', \vec{p}'_{k+1})$ are defined as $(\vec{p}_i{}', \vec{p}'_{k+1}) = T_{\omega_{i,k+1}}(-\vec{p}_i, \vec{p}_{k+1})$. Repeating a similar step of our first argument, we perform a conventional transformation on the integration variables $\vec{p}_{k+1} \longrightarrow -\vec{p}_{k+1}$ and $\omega_{i,k+1} \longrightarrow -\omega_{i,k+1}$ and use that the primed momenta transform into $(\vec{p}_i{}', \vec{p}'_{k+1}) = T_{\omega_{i,k+1}}(-\vec{p}_i, -\vec{p}_{k+1}) = (-\vec{p}_i{}', -\vec{p}'_{k+1})$ to rewrite the integral as

$$\left(\hat{C}^{(a)}_{k,k+1}\rho^{(a)}_{k+1}\right)(\tilde{x}_1, \ldots, \tilde{x}_k) = Na^2 \sum_{i=1}^{k} \int_{\mathbb{R}}^{3} d\vec{p}_{k+1} \int_{\vec{\omega}_{i,k+1}\cdot(\vec{p}_i - \vec{p}_{k+1})\geqslant 0} d\vec{\omega}_{i,k+1}\, \vec{\omega}_{i,k+1} \cdot (\vec{p}_i - \vec{p}_{k+1}) \times \left[\rho^{(a)}_{k+1}(\tilde{x}_1, \ldots, \tilde{x}_{i-1}, \vec{q}_i, -\vec{p}_i{}', \tilde{x}_{i+1}, \ldots \tilde{x}_k, \vec{q}_i + a\vec{\omega}, -\vec{p}'_{k+1}) - \rho^{(a)}_{k+1}(\tilde{x}_1, \ldots, \tilde{x}_{i-1}, \vec{q}_i, -\vec{p}_i, \tilde{x}_{i+1}, \ldots \tilde{x}_k, \vec{q}_i - a\vec{\omega}, -\vec{p}_{k+1})\right] \quad (67)$$

For the purpose of comparison of this result with the original equation, we use two further conventions. First, we make the condition $\vec{q}_{k+1} = \vec{q}_i \pm a\vec{\omega}_{i,k+1}$ in the integrand of (67) more explicit by introducing an extra integration over $\vec{q}_{k+1}$, while including a delta function $\delta(\vec{q}_{k+1} - \vec{q}_i \mp a\vec{\omega}_{i,k+1})$ in the respective terms in the integrand. Accordingly,

$$\rho^{(a)}_{k+1}(x_1, \ldots, x_k, \vec{q}_i \pm a\vec{\omega}_{i,k+1}, \vec{p}_{k+1}) = \delta(\vec{q}_{k+1} - \vec{q}_i \mp a\vec{\omega}_{i,k+1})\rho^{(a)}_{k+1}(x_1, \ldots, x_{k+1}) \quad (68)$$

Secondly, we introduce a formal operation on probability density functions that implements the transformation of $(\vec{p}_i, \vec{p}_{k+1}) \longrightarrow (\vec{p}_i{}', \vec{p}'_{k+1})$ in their arguments, for which, with a slight abuse of notation, we use the symbol already in use, $T_{\omega_{i,k+1}}$. Thus, we define:

$$T_{\omega_{i,k+1}}\rho^{(a)}_{k+1}(x_1, \ldots; \vec{q}_i, \vec{p}_i; \ldots x_k; \vec{q}_{k+1}, \vec{p}_{k+1}) = \rho^{(a)}_{k+1}(x_1, \ldots, \vec{q}_i, \vec{p}_i{}', \ldots x_k, \vec{q}_{k+1}, \vec{p}_{k+1}{}') \quad (69)$$

So, we can write

$$T_{\omega_{i,k+1}}\bar{\rho}^{(a)}_{k+1}(x_1, \ldots, x_{k+1}) = \rho^{(a)}_{k+1}(\tilde{x}_1, \ldots, \tilde{x}_{i-1}, \vec{q}_i, -\vec{p}_i{}', \tilde{x}_{i+1}, \ldots \tilde{x}_k, \vec{q}_{k+1}, -\vec{p}'_{k+1}) \quad (70)$$

In this notation, we can rewrite (67) as:

$$\left(\hat{C}^{(a)}_{k,k+1}\rho^{(a)}_{k+1}\right)(\tilde{x}_1, \ldots, \tilde{x}_k) = Na^2 \sum_{i=1}^{k} \int dx_{k+1} \int_{\vec{\omega}_{i,k+1}\cdot(\vec{p}_i - \vec{p}_{k+1})\geq 0} d\vec{\omega}_{i,k+1}\, \vec{\omega}_{i,k+1} \cdot (\vec{p}_i - \vec{p}_{k+1}) \times \left[\delta(\vec{q}_{k+1} - \vec{q}_i - a\vec{\omega}_{i,k+1}) T_{\omega_{i,k+1}}\bar{\rho}^{(a)}_{k+1}(x_1, \ldots, x_{k+1}) - \delta(\vec{q}_{k+1} - \vec{q}_i + a\vec{\omega}_{i,k+1})\bar{\rho}^{(a)}_{k+1}(x_1, \ldots, x_{k+1})\right] \quad (71)$$

while the original collision operator (32), for comparison, takes the form

$$\left(\hat{C}^{(a)}_{k,k+1}\rho^{(a)}_{k+1}\right)(x_1, \ldots, x_k) = Na^2 \sum_{i=1}^{k} \int dx_{k+1} \int_{\vec{\omega}_{i,k+1}\cdot(\vec{p}_i - \vec{p}_{k+1})\geq 0} d\vec{\omega}_{i,k+1}\, \vec{\omega}_{i,k+1} \cdot (\vec{p}_i - \vec{p}_{k+1}) \times \left[\delta(\vec{q}_{k+1} - \vec{q}_i + a\vec{\omega}_{i,k+1}) T_{\omega_{i,k+1}}\rho^{(a)}_{k+1}(x_1, \ldots, x_{k+1}) - \delta(\vec{q}_{k+1} - \vec{q}_i - a\vec{\omega}_{i,k+1})\rho^{(a)}_{k+1}(x_1, \ldots, x_{k+1})\right] \quad (72)$$

Now, in analogy to the previous case (cf. Eqn. (63)), we inquire whether

$$\left(\hat{C}^{(a)}_{k,k+1}\rho^{(a)}\right)(\tilde{x}_1, \ldots, \tilde{x}_k) \overset{?}{=} -\left(\hat{C}^{(a)}_{k,k+1}\bar{\rho}^{(a)}\right)(x_1, \ldots, x_k) \quad (73)$$

holds. But on inspection, we find that $\left(\hat{C}^{(a)}_{k,k+1}\rho^{(a)}\right)(\tilde{x}_1,\ldots,\tilde{x}_k)$ as expressed by (71) is *almost*, but not quite, equal to $\left(\hat{C}^{(a)}_{k,k+1}\bar{\rho}^{(a)}_{k+1}\right)(x_1,\ldots,x_k)$ as obtained from (73) by substituting ρ by $\bar{\rho}$, since the position argument $\vec{q}_{k+1}$ is displaced in different directions in the two terms of the last factor of the integrands in (71) and (73) . (Of course, the distinction between these two expressions will disappear in the limit $a \longrightarrow 0$.)

Nevertheless, for the purpose of testing time reversal invariance, we *can* conclude that in general

$$\left(\hat{C}^{(a)}_{k,k+1}\rho^{(a)}\right)(\tilde{x}_1,\ldots,\tilde{x}_k) \neq -\left(\hat{C}^{(a)}_{k,k+1}\bar{\rho}^{(a)}\right)(x_1,\ldots,x_k) \tag{74}$$

because nothing prevents us from considering special initial conditions for the BBGKY hierarchy for hard spheres, with the property that, for all $1 \leq k \leq N$, $\rho^{(a)}_k(x_1,\ldots x_k)$ is uniform over all allowed position coordinates, (i.e., it takes a constant value, for given $\vec{p}_1,\ldots,\vec{p}_k$, for all $(\vec{q}_1,\ldots,\vec{q}_k)$ compatible with the condition $\|\vec{q}_i - \vec{q}_j\| \geq a$ $(\forall i \neq j)$ in the domain of these probability density functions). For such a special case we do obtain

$$\rho^{(a)}_{k+1}(x_1,\ldots,x_k,\vec{q}_i - a\vec{\omega}_{i,k+1},\vec{p}_{k+1})) = \rho^{(a)}_{k+1}(x_1,\ldots,x_k,\vec{q}_i + a\vec{\omega}_{i,k+1},\vec{p}_{k+1})) \tag{75}$$

, and thus $\delta(\vec{q}_{k+1} - \vec{q}_i + a\vec{\omega}_{i,k+1}) = \delta(\vec{q}_{k+1} - \vec{q}_i - a\vec{\omega}_{i,k+1})$, which implies

$$\left(\hat{C}^{(a)}_{k,k+1}\rho^{(a)}\right)(\tilde{x}_1,\ldots,\tilde{x}_k) = \left(\hat{C}^{(a)}_{k,k+1}\bar{\rho}^{(a)}\right)(x_1,\ldots,x_k). \tag{76}$$

Note that this choice of special initial conditions does not commit us to the thermal equilibrium solution, since no demand had been placed on the dependency on the momentum variables. In fact, the solutions might still deviate in an arbitrary manner from a Maxwellian dependency. As a result, our choice is not trivial, in that both the left-hand side and right-hand side of (76) will in general be different from zero.

So, if we introduce yet one more notational convention, and define a new collision $\breve{C}^{(a)-}_{k,k+1}$ operator to mimick the right-hand side of (71) by

$$\left(\hat{C}^{(a)-}_{k,k+1}\rho^{(a)}\right)(x_1,\ldots,x_k) = Na^2\sum_{i=1}^{k}\int dx_{k+1}\int_{\vec{\omega}_{i,k+1}\cdot(\vec{p}_i-\vec{p}_{k+1})\geq 0} d\vec{\omega}_{i,k+1}\,\vec{\omega}_{i,k+1}\cdot(\vec{p}_i - \vec{p}_{k+1})$$
$$\times\left[\delta(\vec{q}_{k+1} - \vec{q}_i - a\vec{\omega}_{i,k+1})T_{\omega_{i,k+1}}\rho^{(a)}_{k+1}(x_1,\ldots,x_{k+1}) - \delta(\vec{q}_{k+1} - \vec{q}_i + a\vec{\omega}_{i,k+1})\rho^{(a)}_{k+1}(x_1,\ldots,x_{k+1})\right] \tag{77}$$

we conclude that if $\{\rho^{(a)}_{k,t}\ t \in \mathbb{R}\}$ is an arbitrary solution to the equation

$$\frac{\partial\rho^{(a)}_{k,t}(x_1,\ldots,x_k)}{\partial t} - \mathcal{H}_k\rho^{(a)}_{k,t}(x_1,\ldots,x_k) = \left(\hat{\mathcal{C}}^{(a)}_{k,k+1}\rho^{(a)}_{k+1,t}\right)(x_1,\ldots,x_k) \tag{78}$$

the time reversed solution $\{\hat{\rho}^{(a)}_{k,-t}\ t \in \mathbb{R}\}$ will be a solution of the equation

$$-\frac{\partial\bar{\rho}^{(a)}_{k,t}(x_1,\ldots,x_k)}{\partial t} + \mathcal{H}_k\bar{\rho}^{(a)}_{k,t}(x_1,\ldots,x_k) = \left(\hat{\mathcal{C}}^{(a)-}_{k,k+1}\bar{\rho}^{(a)}_{k+1,t}\right)(x_1,\ldots,x_k) \tag{79}$$

but, since it does not hold generally that $\hat{\mathcal{C}}^{(a)-}_{k,k+1} = -\hat{\mathcal{C}}^{(a)}_{k,k+1}$, the equation is not time-reversal invariant. □

Proposition 4. The BBGKY hierarchy with the collision term expressed by (54) is not time-reversal invariant.

Proof. Recall that, after adopting the outgoing representation for collision points, the BBGKY hierarchy takes the form:

$$\frac{\partial \rho^{(a)}_{k,t}(x_1,\ldots,x_k)}{\partial t} - \mathcal{H}_k \rho^{(a)}_{k,t}(x_1,\ldots,x_k) = \left(\check{\mathcal{C}}^{(a)}_{k,k+1}\rho^{(a)}_{k+1,t}\right)(x_1,\ldots,x_k) \tag{80}$$

where the left-hand side is the same as in (59), but the collision operator is now expressed by (54). Exactly the same analysis as in the proof of *Proposition 3* applies here, and thus in analogy to (74) we need to show that

$$\left(\check{C}^{(a)}_{k,k+1}\rho^{(a)}\right)(\tilde{x}_1,\ldots,\tilde{x}_k) \neq -\left(\check{C}^{(a)}_{k,k+1}\bar{\rho}^{(a)}\right)(x_1,\ldots,x_k) \tag{81}$$

Again, the special choice of initial conditions for the BBGKY hierarchy for hard spheres with the property that, for all $1 \leq k \leq N$, $\rho^{(a)}_k(x_1,\ldots x_k)$ is uniform over all allowed position coordinates, yields

$$\left(\check{C}^{(a)}_{k,k+1}\rho^{(a)}\right)(\tilde{x}_1,\ldots,\tilde{x}_k) = \left(\check{C}^{(a)}_{k,k+1}\bar{\rho}^{(a)}\right)(x_1,\ldots,x_k). \tag{82}$$

and so the rest of the proof carries over as in the previous proposition. □

Proposition 5. The BBGKY hierarchy with the collision term expressed by (54) is not equivalent to the time-reversal transformation of the BBGKY hierarchy with the collision term expressed by (32).

Proof. The BBGKY hierarchy with the collision term expressed by (54) is given by eq.(65), whereas the time-reversal transformation of the BBGKY hierarchy with the collision term expressed by (32) is given by eq.(7). Clearly, the two equations have the same form just in case $\hat{\mathcal{C}}^{(a)-}_{k,k+1} = -\check{\mathcal{C}}^{(a)}_{k,k+1}$. Since $\check{\mathcal{C}}^{(a)}_{k,k+1} = -\hat{\mathcal{C}}^{(a)}_{k,k+1}$, we need to check whether $\hat{\mathcal{C}}^{(a)-}_{k,k+1} = \hat{\mathcal{C}}^{(a)}_{k,k+1}$. However, comparing (54) and (77) shows directly that the two collision operators differ in general, in that the displacements $+a\omega_{i,k+1}$ and $-a\omega_{i,k+1}$ for the position variables $\vec{q}_i$ are reversed: specifically, $-a\omega_{i,k+1}$ appears in $\hat{\mathcal{C}}^{(a)}_{k,k+1}$ within the argument of $\bar{\rho}^{(a)}_{k+1}$ with primed momenta $\vec{q}_i\,'$, while it appears in $\hat{\mathcal{C}}^{(a)-}_{k,k+1}$ within the argument of $\rho^{(a)}_{k+1}$ with unprimed momenta $\vec{q}_i$; and vice versa for $+a\omega_{i,k+1}$. □

References

[1] Alexander R.K.: The infinite hard-sphere system. Ph.D. Thesis, University of California at Berkeley (1975).

[2] Boltzmann L.: Weitere Studien über das Wärmeglechgewicht unter Gasmolekülen. Wiener Berichte **66**, 275–370 (1872); in Boltzmann (1909) vol.I. pp. 316–402.

[3] Boltzmann L.: [1875], Über das Wärmegleichgwich von Gasen, auf welche äußere Kräfte wirken. Wiener Berichte **72**, 427–457, in Boltzmann (1909) vol.II, pp. 1–30.

[4] Boltzmann L.: [1877], Bemerkungen über einige Probleme der mechanische Wärmetheorie, Wiener Berichte **66**, 275–370, in Boltzmann (1909) vol.II, pp. 112–148.

[5] Boltzmann L.: [1895], On certain questions of the theory of gases. Nature, **51** 413–415; in Boltzmann (1909) vol.III, pp. 535–544.

[6] Boltzmann L.: [1897], Zu Hrn Zermelos Abhandlung "Über die mechanische Erklärung irreversibler Vorgänge". Wiedemann's Annalen **60**, p.392–398, in Boltzmann (1909), Vol III, paper 119.

[7] Boltzmann L.: [1909], Wissenschaftliche Abhanlungen Vol. I, II, and III. F. Hasenöhrl (ed.) Leipzig. Reissued New York: Chelsea, 1969.

[8] Cercignani C.: 134 years of Boltzmann Equation. In: G. Gallavotti, W.L. Reiter and J. Yngvason (eds.), Boltzmann's Legacy pp. 107–128. Zürich: European Mathematical Society (2008).

[9] Cercignani C., Illner R., and Pulvirenti M.: The mathematical theory of diluted gases. Springer, New York (1994).

[10] Drory A.: Is there a reversibility paradox? Recentering the debate on the thermodynamic time arrow. Studies In History and Philosophy of Modern Physics **39**, 889–913 (2008).

[11] Ehrenfest P. and Ehrenfest-Afanassjewa T.: The Conceptual Foundations of the Statistical approach in Mechanics. New York, Cornell University Press (1912).

[12] Illner R. and Pulvirenti M.: Global validity of the Boltzmann equation for a two-dimensional rare gas in a vacuum. Commun. Math. Phys. **105**, pp. 189–203 (1986); erratum *ibidem* **121** 143–146 (1989).

[13] Lanford O.E.: Time evolution of large classical systems. In: J. Moser (ed.), *Dynamical Systems, Theory and Applications*. Lecture Notes in Theoretical Physics **38**, pp.1–111. Springer, Berlin (1975).

[14] Lanford O.E.: On the derivation of the Boltzmann equation. in Asterisque **40** 117–137 (1976).

[15] Lanford O.E.: The hard sphere gas in the Boltzmann-Grad limit. Physica, **106A**, 70–76 (1981).

[16] Lebowitz J.L.: Microscopic Dynamics and Macroscopic laws. In: C.W. Horton, Jr., L.E. Reichl, and V.G. Szebehely (eds.) *Long-Time Prediction in Dynamics*, pp. 220–233. Wiley, New York (1983).

[17] Loschmidt J.: [1876-1877], Über die Zustand des Wärmegleichgewichtes eines Systems von Körpern mit Rücksicht auf die Schwerkraft. In Wiener Berichte **73**.

[18] Goldstein S.: Boltzmann's approach to statistical mechanics. In: edited by J. Bricmont *et al.* (eds.) *Chance in Physics: Foundations and Perspectives*. Lecture Notes in Physics **574**, pp. 39–54. Springer, Berlin (2001).

[19] Spohn, H.: Kinetic equations from Hamiltonian dynamics: Markovian limits. Reviews of Modern Physics **53**, 569–615 (1980).

[20] Spohn H.: Large Scale Dynamics of Interacting Particles. Springer, Berlin (1991).

[21] Uffink J.: Compendium of the foundations of classical statistical physics. In J. Butterfield and J. Earman (eds.) Handbook for the Philosophy of Physics, pp. 923–1074. Elsevier, Amsterdam (2008).

[22] Valente G.: Can Lanford's theorem be extended in time?, in preparation (2010).