LONG-TIME VALIDITY OF THE GAINLESS HOMOGENEOUS BOLTZMANN EQUATION

KARSTEN MATTHIES AND FLORIAN THEIL

Abstract. This paper introduces a new method to show the validity of a continuum description for the deterministic dynamics of many interacting particles. Here the many particle evolution is analyzed for a hard sphere flow with the addition that after a collision the collided particles are removed from the system. We consider initial conditions, which are Poisson distributed according a spatially homogeneous velocity density $f_0(v)$ that has finite mass and variance (kinetic energy) and does not concentrate mass on lines. Assuming finite energy and no concentration properties on f_0, the homogeneous Boltzmann equation without gain term is derived for arbitrary long times in the Boltzmann-Grad scaling. A key element is a novel description of the many particle flow by a hierarchy of trees which encode the possible collisions. The occurring trees are shown to have favorable properties with a high probability, allowing to restrict the analysis to a finite number of interacting particles, enabling us to extract a single-body distribution. A counter-example is given for a concentrated initial density f_0 even to short-term validity.

The derivation of the continuum models of mathematical physics from atomistic descriptions is a longstanding and fundamental problem, see problem six in [Hil00]. This article is the first of a series of papers ([MT07a], [MT07b], [MT07c]) where we propose and develop a new method that allows us to derive and justify effective continuum limits as scaling limits of large interacting particle systems. In particular we can confront a fundamental challenge in statistical mechanics: The emergence of irreversible macroscopic behavior generated by deterministic reversible Hamiltonian micro-evolution. For earlier work, which was mostly restricted to short times or linear equations, see [Gal70, Lan75, Spo78, BBS83, Spo91, CIP94] and references therein.

We consider the effective Hamiltonian evolution of n hard balls $(u(i,t), v(i,t)) \in \mathbb{T}^d \times \mathbb{R}^d$, $i \in \{1 \dots n\}$ for $d = 2, 3$ with diameter a. We are mostly interested in the effective evolution generated by the kinetic limit where n tends to infinity, the initial values $(u(i, t = 0), v(i, t = 0))$ are iid random variables with law $f_0 \in PM(\mathbb{T}^d \times \mathbb{R}^d)$ where $\mathbb{T}^d$ denotes the d-dimensional unit-torus. The diameter a of the particles is linked to n by the Boltzmann-Grad relation

$$\lim_{n\to\infty} n a^{d-1} = 1. \tag{1}$$

If the particles interact with each other via a hard-core potential it is expected that for every open set $\Omega \subset \mathbb{T}^d \times \mathbb{R}^d$ at every time t the number of particles in Ω divided by the total number of particles converges to $\int_\Omega \mathrm{d}f_t(u, v)$. The time-dependent probability measure f solves the nonlinear Boltzmann equation

$$\partial_t f + v \cdot \partial_u f = Q_+[f, f] + Q_-[f, f], \tag{2}$$

where Q_+ is the gain-term, Q_- is the loss term and $Q = Q_+ + Q_-$ is the collision operator. The collision kernels correspond to a situation of completely independent particles with density $f(u, v, t)$, which collide at position u with a probability depending on the velocities v and v' of the colliding particles. The colliding particles change their velocities from v and v' to v_* and v'_*, so there is a loss in the density at (u, v) and (u, v') and a gain at (u, v_*) and (u, v'_*).

The analysis of (2) is nontrivial and culminated in the celebrated paper by Lions and DiPerna ([DL90]) where existence of renormalized solutions is rigorously established for the first time. The mathematical challenges are a result of the subtle interplay between the transport term $v \cdot \partial_u f$ and the nonlinear gain-term $Q_+[f, f]$. If either of the two terms is not present the analysis of the Boltzmann equation simplifies considerably.
In this paper we will drop both terms and consider the conceptually simplest situation where the motion of each particle moves with constant velocity until it interacts with another particle. After the collision the collided particles are removed from the system. The transport term can be dropped by considering spatially homogenous initial data.
Two of the remaining three scenarios will be treated in [MT07a] and [MT07b]. The analysis of the case where both the transport-term and the gain-term are present requires the development of new compactness results for this type of many-body system.
First we will derive the mean-field theory for a single particle which consists of a homogeneous Boltzmann equation without gain-term. We prove rigorously that the weak-* limit of the empirical densities indeed satisfies the mean-field theory, *provided* that $f_0 \in M_+(\mathbb{R}^d)$ has finite total mass and kinetic energy

$$\int_{\mathbb{R}^d} (1 + |v|)^2 \, \mathrm{d}f_0(v) = K_{\text{ini}} < \infty \tag{3}$$

and does not concentrate mass on single velocity directions, i.e.

$$\int_{\rho(v,\nu)} \mathrm{d}f_0(v') = 0 \text{ for all } v \in \mathbb{R}^d, \nu \in S^{d-1}, \tag{4}$$

where $\rho(v, \nu) = v + \mathbb{R}\,\nu$ is a line. The proof constitutes the core of this article as it is based on new ideas that have not appeared in the literature yet. Instead of extracting the single-body density directly from the complicated n-body evolution we insert an intermediate layer: trees which encode the collision history of the individual particles. Our trees solve in the case of hard-ball dynamics in the Boltzmann-Grad limit many difficulties that haunted previous attempts to answer the question how to extract single-body densities from many-body evolution:

(1) It is not difficult to construct the limiting distribution P of the trees which is obtained by ignoring correlations caused by rare events such as recollisions.
(2) We can extract the single-body density f_t from the distribution of the trees in a relatively simple way. This amounts to distinguishing between the observed degrees of freedom (odof) and the background noise which drives the evolution of the odof.
(3) The convergence of the empirical distribution $\hat{P}$ to the limiting distribution P can be derived on a set of good trees $\mathcal{G}$. The formula for P is so simple that it is not hard to construct a reasonably sharp upper bound $P(\mathcal{G}^c) = o(1)$ as a tends to 0. The combination of these two facts constitutes a rigorous justification of the mean-field theory.

In particular the last point is of crucial importance if one attempts to extract the laws of thermodynamics from deterministic systems with random initial conditions.
For this reason we will first study the tree-equivalent of the Boltzmann-equation: the limiting distribution of trees which is obtained by ignoring correlations. We will show how time plays the role of a parameter which resembles temperature in equilibrium statistical mechanics (equation (20)). Furthermore, the extraction of the single-body distribution will reveal the conceptual link between the Boltzmann equation and the distribution of the trees (Proposition 13).

The hard part of the analysis consists in step (3) where we have to bound the probability of bad trees (Proposition 17).
In Section 3 we will discuss an example which shows that assumption (4) cannot be dropped without losing the approximation property of the Boltzmann equation. We demonstrate that for arbitrarily short but finite times the weak-* limit of the empirical density is not consistent with the mean-field theory. In the last section, we collect some proofs, which are not immediately needed in the understanding and the development of the concepts of this article. An appendix with a list of frequently used notation is included.

1. Main result

For spatially homogeneous initial data the mean-field theory leads to Boltzmann equations without transport and gain term

$$\dot{f} = Q_-[f,f], \quad f_{t=0} = f_0, \tag{5}$$

where $Q_-[f,f](v) = -\int_{\mathbb{R}^d} \mathrm{d}f(v')\, \kappa_d |v-v'|\, f(v)$ is the loss term with κ_d the volume of $d-1$ dimensional unit-ball, in particular $\kappa_2 = 2, \kappa_3 = \pi$.
We prove a somewhat weaker statement than the one stated in the introduction. The number of particles n is a random number, the law of n is a Poisson-distribution with intensity $N = a^{1-d}$. This assumption entails that na^{d-1} is a sequence of random numbers such that $\lim_{a\to 0} na^{d-1} = 1$ almost surely if a assumes only a countable set of values which converges to 0 quickly enough. On the other hand, if the number of particles is determined uniquely by a it is expected that the same result holds but the additional knowledge introduces correlations which are not dealt with in this work.
On the atomistic level we consider n particles with initial values $(u_0(i), v_0(i)) \in \mathbb{T}^d \times \mathbb{R}^d$, $i = 1 \ldots n$, which evolve by Newtonian dynamics

$$\begin{aligned} &u^{(a)}(i, t=0) = u_0(i),\ v^{(a)}(i,t=0) = v_0(i), \\ &\dot{u}^{(a)}(i,t) = v^{(a)}(i,t),\ \dot{v}^{(a)}(i,t) = 0. \end{aligned} \tag{6}$$

For each $t \in [0,\infty)$, $i \in \{1 \ldots n\}$ there exists a unique scattering state $\beta_i^{(a)}(t) \in \{0,1\}$ which satisfies the implicit relation

$$\beta^{(a)}(i,t) = \begin{cases} 1 & \text{if } \operatorname{dist}(z_i, z_{i'}, s) \geq a\beta^{(a)}(i', s) \text{ for all } s \in [0,t),\ i' \neq i, \\ 0 & \text{else} \end{cases} \tag{7}$$

with a modified distance function to ignore initial intersections

$$\operatorname{dist}((u,v),(u',v'),s) = |u - u' + s(v-v')| + a\chi_{[0,a]}(|u-u'|). \tag{8}$$

Definition 1 (Poisson point processes). *Let Ω be a measure space. The random variable $z \in \cup_{n=0}^\infty \Omega^n$ forms a Poisson point process with density $\mu \in M_+(\Omega)$ if*

$$\mathrm{Prob}(z \in \Omega^n) = e^{-\mu(\Omega)} \frac{\mu(\Omega)^n}{n!}, \quad \mathrm{law}(z_i) = \mu/\mu(\Omega),$$

and $z_1, \ldots, z_n$ are independent. The law of the Poisson point process is denoted by $\mathrm{Prob}_{\mathrm{ppp}}$.

Theorem 2. *(Justification of the gainless Boltzmann equation) Let $f_0 \in PM_+(\mathbb{R}^d), d \geq 2$ be a momentum density that satisfies (3, 4) and let for each $N > 0$ the random variable $(u_0, v_0) \in \cup_{n=0}^\infty (\mathbb{T}^d \times \mathbb{R}^d)^n$ be a Poisson point process with intensity $N(\mathbf{1}_{\mathbb{T}^d} \otimes f_0)$. Let n particles with initial values $(u_0(i), v_0(i)) \in \mathbb{T}^d \times \mathbb{R}^d$, $i = 1 \ldots n$ evolve by (6). If N depends on a such that*

$$Na^{d-1} = 1,$$

then for each $t \in [0,\infty)$, $\varepsilon > 0$, measurable $A \subset \mathbb{T}^d \times \mathbb{R}^d$

$$\lim_{a\to 0} \mathrm{Prob}_{\mathrm{ppp}}\Bigg(\bigg|\frac{1}{N}\#\left\{i \in \{1\ldots n\} \mid (u^{(a)}(i,t), v^{(a)}(i,t)) \in A \text{ and } \beta^{(a)}(i,t) = 1\right\} - \int_A \mathrm{d}u\,\mathrm{d}f_t(v)\bigg| > \varepsilon\Bigg) = 0, \tag{9}$$

where $f : [0,\infty) \to M_+(\mathbb{R}^d)$ is the unique solution of (5).

The assumption that $\int_{\mathbb{R}^d} \mathrm{d}f_0(v) = 1$ is not necessary. We make it because it simplifies the notation in the proof which can be found at the end of Section 2.
Stronger results can be obtained, if the rate at which a tends to 0 is controlled.

Corollary 3. *Under the same assumptions as above there exists a subsequence of diameters a_k such that $\lim_{k\to\infty} a_k = 0$ and*

$$\frac{1}{N_k}\sum_{i=1}^{n} \beta_i^{(a_k)}(t)\delta(\cdot - (u^{(a_k)}(i,t), v^{(a_k)}(i,t))) \overset{*}{\rightharpoonup} f_t \tag{10}$$

weak- in $M(\mathbb{T}^d \times \mathbb{R}^d)$ as $k \to \infty$.*

It is not hard to obtain more explicit subsequences such as $a_k = k^{-p}$ e.g. for $p > 1$ if additional regularity assumptions for f_0 are made. The proof of Corollary 3 is given after the proof of the theorem.
Assumption (4) does not exclude the possibility that f_0 is concentrated on lower dimensional subsets, for example the uniform distribution on the sphere S^{d-1} is admissible, i.e. f_0 satisfies

$$\int \varphi(v)\,\mathrm{d}f_0(v) := \frac{1}{\mathcal{H}^{d-1}(S^{d-1})}\int_{S^{d-1}} \varphi(v)\,\mathrm{d}\mathcal{H}^{d-1}(v), \tag{11}$$

for all testfunctions $\varphi \in C_c(\mathbb{T}^d \times \mathbb{R}^d)$, where $\mathcal{H}^d$ is the d-dimensional Hausdorff-measure.
The approach due to Lanford [Lan75] which uses the BBGKY-hierarchy to derive this equation from the Hamiltonian evolution relies heavily on analytic properties in time and high regularity of f.
As a motivation for our analysis, we give an example why the previous is restricted to short times, even for the gainless case. Let us assume that f_0 is given by (11). Solutions f of (5) which satisfy $f_{t=0} = f_0$ can be written as $f_t = \rho(t) f_0$, where ρ satisfies the ordinary differential equation

$$\dot{\rho} = -\gamma\rho^2, \quad \rho(t=0) = 1. \tag{12}$$

The collision rate $\gamma(v) = \int \kappa_d |v - v'|\,\mathrm{d}f_0(v')$ is constant for $v \in S^{d-1}$, the support of f_0, since f_0 is invariant under rotations.
The solution of system (12) is given by

$$\rho(t) = \frac{1}{1+\gamma t}. \tag{13}$$

As the geometric series $\sum_{k=0}^{\infty}(-\gamma t)^k$ diverges if $\gamma|t| > 1$ formula (13) shows that writing f as a power series in t is restricted to small times. Although the solution is a perfectly smooth and bounded function for $t \in [0,\infty)$ the approach is haunted by the singularity at $t = -\frac{1}{\gamma}$. In this particular example, an alternative could be restarting the procedure at small positive time using suitable a-priori estimates.

Never the less such an approach is not extendable to other cases. For this reason we develop a different method to study the coarse-grained many-body dynamics. Furthermore, we will analyze effects due to concentration by a Taylor expansion in time of f_t in Section 3

2. Proof of Theorem 2

2.1. **The hierarchy of evolutions.** Instead of expanding ρ into a power-series in t and matching coefficients in a first step, we replace the initial value problem (5) by an infinite system using general initial distribution without concentrations

$$\dot{f}_k = Q_-[f_{k-1}, f_k], \quad f_{t=0,k} = f_0. \tag{14}$$

Since Q_- is quadratic, for fixed k the integro-differential equation (14) is in fact linear and non-autonomous. We can therefore work with the mathematically much more convenient mild formulation. The differential equation completely decouples in v and the equation for each v is a scalar linear nonautonomous ODE, which can be directly integrated to

$$f_{t,k} = \exp(-\textstyle\int_0^t L[f_{s,k-1}]\,\mathrm{d}s) f_0, \tag{15}$$

where $L[f](v) = \kappa_d \int \mathrm{d}f(v')\,|v - v'|$. We observe that $\mathrm{d}f_{t,k}(v)$ is absolutely continuous with respect to $\mathrm{d}f_0(v)$ due to the decoupling in v.

Lemma 4. *Let $f_0 \in M_{(1+|v|)^2}$ then f_k converges in $C^0_\rho([0,\infty), M_{1+|v|})$ to f for some $\rho > 0$ and $f \in C^1([0,\infty), M_{1+|v|})$ is the unique solution of (5).*

By $M_{1+|v|}$ and $M_{(1+|v|)^2}$ we mean the set of Radon measures with first and second moment, C_ρ denotes the continuous functions which grow not faster than $e^{\rho t}$. The proof of Lemma 4 together with a precise definition of the function spaces can be found in Section 4.

Now we have to translate this idea into the context of deterministic many-body dynamics. To limit the complexity of the notation we will from now on assume that everything except the constants depends on a without displaying the dependency. For every realization of the n-body evolution the random variable $\beta(i,t) \in \{0,1\}$, which encodes the scattering state of particle $i \in \{1 \ldots n\}$ at time $t \in [0,\infty)$ satisfies the implicit relation (7). The computation of β can be simplified by introducing a hierarchy of artificial evolutions indexed by $k \in \mathbb{N}$. We assume that the initial values of the particles at all levels are identical. The particles at level $k = 1$ are simply transported and do not interact with anything. The particles at level $k > 1$ interact only with the particles at level $k-1$, but not with each other. For each $k \in \mathbb{N}$ and $i \in \{1 \ldots n\}$ the scattering state $\beta_k(i,t) \in \{0,1\}$ is defined in the following way

$$\beta_k(i,t) = \begin{cases} 1 & \text{if } \mathrm{dist}(z_i, z_{i'}, s) \geq a\beta_{k-1}(i', s) \text{ for all } s \in [0,t), i' \neq i, \\ 0 & \text{else,} \end{cases} \tag{16}$$

$$\beta_1(i) \equiv 1, \tag{17}$$

with dist as in (8).

Remark 5. *While the determination of the collision-state $\beta(i,t)$ is a complicated problem, the state $\beta_k(i,t)$ emerges via a very simple calculation from $\beta_{k-1}(\cdot,t)$.*

Lemma 6. *For all realizations of the processes of the initial conditions $(u_0, v_0) \in \cup_{n=0}^\infty (\mathbb{T}^d \times \mathbb{R}^d)^n$ both $\beta_k(i,t)$ and $\beta(i,t)$ are well defined and*

$$\lim_{k\to\infty} \beta_k(i,t) = \beta(i,t) \tag{18}$$

pointwise in i and uniformly in t.

Proof. See section 4. ∎

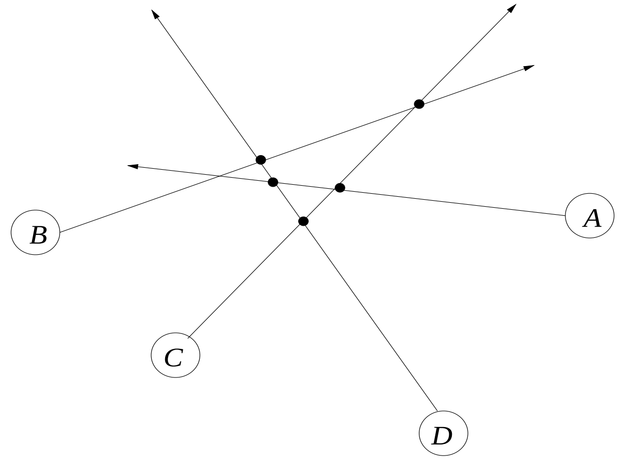

FIGURE 1. Initial positions and velocities of four particles. The bullets indicate the positions where the particles are potentially scattered. The shown configuration is not very likely and consequentially the collision trees are quite complex. Note that not every subset of intersection points of the arrows is a set of potential scattering position. For example, it is not possible to add another bullet at the intersection point of the arrows A and B as it is not possible to assign to each of the six intersections a scattering time which is compatible with the order of the bullets on each ray.

2.2. **The concept of trees.** The translation of the n-body evolution into scattering states β is greatly facilitated by the concept of trees. In the collision tree with root (u, v) we will collect information of collisions and potential collisions up to time t for a particle with initial data u, v.

As an example assume that $n = 4$ and consider the scenario in fig. 1 where the letters A, B, C, D are the labels of the four particles, the empty circles are the initial positions and the arrows are the initial velocities. Consequentially the arrow-tips indicate the positions of the particles at time $t = 1$. To determine whether a certain particle has been scattered before time $t = 1$ it suffices to analyze the associated collision tree which is constructed as follows: The particle of interest is the root with initial data (u, v). The particles which are potentially scattered by the root are added as leaves, i.e. a particle with initial data (u', v') is added, if $|u+sv-(u'+sv')| \leq a$ for some $s \in [0, t]$. This procedure is recursively applied to every leaf but we consider only potential scattering events which are upstream, i.e. before the event which is responsible for adding the leaf. The four collision trees associated to the scenario in fig. 1 are shown in fig. 2. The extraction of the collision trees amounts to a significant reduction of the complexity of the problem. In general, the number of potential scattering events (bullets) is proportional to N but thanks to the Boltzmann-Grad-scaling (1) the number of nodes in the individual trees is a Poissonian random number with an intensity which is asymptotically independent of N and grows exponentially with t, see Lemma 12.

We convert now the example into a general concept.

Definition 7. *Let $\mathbb{N} = \{1, 2, \ldots\}$. The height of a node (or multi-index) $l \in \mathbb{N}^i$ is defined by $|l| := i$, the child node of $l \in \mathbb{N}^i$ is $\bar{l} = (l_1, \ldots, l_{i-1})$. Let $\mathcal{F} = \cup_{i=1}^{\infty} \mathbb{N}^i$ be the set of multi-indices. We say that $m \subset \mathcal{F}$ is a tree skeleton with α roots ($m \in \mathcal{T}^\alpha$), if*

(1) $\#m < \infty$,
(2) $m \cap \mathbb{N} = \{1, \ldots, \alpha\}$,
(3) $\bar{l} \in m$ *for all* $l \in m \setminus \mathbb{N}$,

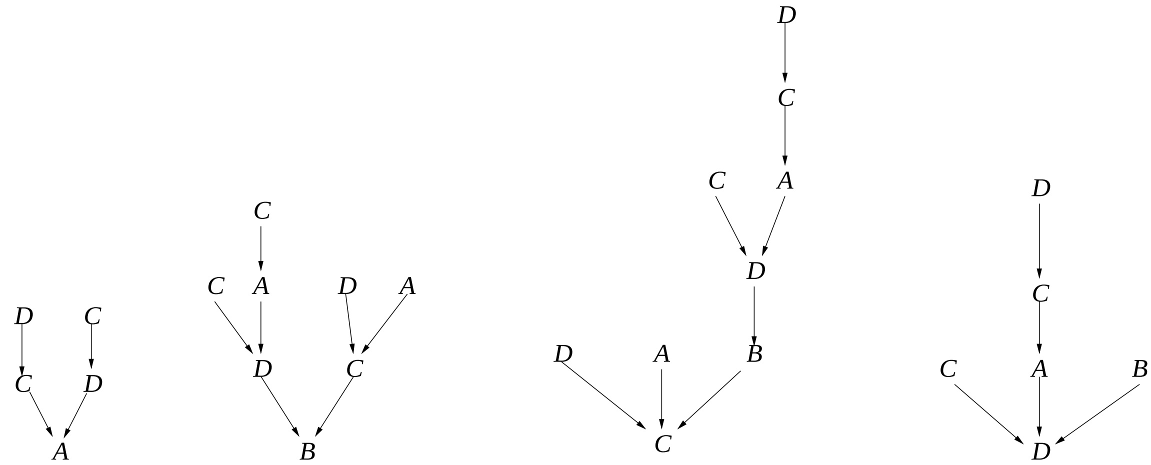

FIGURE 2. Collision trees of the four particles with initial positions and collision structure given in fig. 1. At time $t = 1$ particles C and D have been scattered, particles A and B have not. Note that the labels of the particles which generate the potential scattering events are only included in the picture in order to illustrate the translation of fig. 1 into collision trees. The scattering state of the particle at the root is completely determined by the tree structure, the labels of the tree nodes are irrelevant. For example, the tree of particle B does not contain enough information to decide whether particle A is scattered.

(4) *$l - 1 \in m$ for all $l \in m$ such that $l \neq (*, \dots, *, 1)$,*

where $l - 1 = l - (0, \dots, 0, 1)$. We say that a tree m has at most height k ($m \in \mathcal{T}_k^\alpha$) if $m \cap \mathbb{N}^{k+1} = \emptyset$.
Let $Y = \{(u, v, s, \nu) \in \mathbb{T}^d \times \mathbb{R}^d \times [0, \infty) \times S^{d-1}\}$ be the space of initial values and collision parameters. The set of collision trees is given by

$$\mathcal{T}^\alpha(Y) = \left\{ (m, \phi) \;\middle|\; m \in \mathcal{T}^\alpha, \phi : m \to Y \text{ with the property } s_l \in [s_{l-1}, s_{\bar{l}}] \right.$$
$$\left. \text{and } \nu_l = \frac{1}{a}(u_{\bar{l}} - u_l + s_l(v_{\bar{l}} - v_l)) \text{ for all } l \in m \setminus \mathbb{N} \right\},$$

where $s_{(*,\dots*,0)} = 0$. For each skeleton $m \in \mathcal{T}^\alpha$ we define the set

$$\mathcal{E}(m) = \{(\tilde{m}, \phi) \in \mathcal{T}^\alpha(Y) \mid \tilde{m} = m\}, \tag{19}$$

which contains all trees with skeleton m.

For example, $\{(1), (1,1), (1,2), (1,3), (1,1,1), (1,1,2)\} \in \mathcal{T}_3$, but $\{(1), (2,1)\}$ is not a tree skeleton. The assumption $s_l \in [s_{l-1}, s_{\bar{l}}]$ implies that for all nontrivial permutations $\pi \in S_{\#m} \setminus \mathrm{Id}$ (S_n is the set of permutations of n symbols) and all trees $\Phi = (m, \phi) \in \mathcal{T}^1(Y)$ the permuted tree $\Phi^\pi = (m, \phi^\pi)$ with $\phi_l^\pi = \phi_{\pi(l)}$ is not a tree in the sense of Definition 7. The values ν_l for $l \in \{1, \dots, \alpha\}$ have no relevance. To circumvent this problem we fix a point $\nu^* \in (S^{d-1})^\alpha$, define

$$\mathcal{T}^{\alpha *}(Y) = \{\Phi \in \mathcal{T}^\alpha(Y) \mid \nu_l = \nu_l^* \; \forall l \in m \cap \mathbb{N}\}.$$

and will in future denote $\mathcal{T}^{\alpha *}(Y)$ by $\mathcal{T}^\alpha(Y)$.
It is clear from the definition that for each tree $m \in \mathcal{T}$ there exists a function $r : m \to \mathbb{N} \cup \{0\}$ which counts the number of direct successors, i.e.

$$r_l = \#\{l' \in m \mid \bar{l'} = l\}.$$

We will observe the particles $1,\dots,\alpha$ in the sense that we are interested in evaluating the probability measure on $\mathcal{T}^\alpha(Y)$ which is the joint distribution of the trees generated by α root-particles. In our entire analysis α will be a small natural number, independent of N or a.

Remark 8. *Graph theoretical description of collisions in a hard-sphere gas can lead to many different graphs, which are not necessarily trees. The advantage of our definition is that this graph will always be a tree. Particles might appear several times in a tree, as in fig. 2. This will not destroy the tree structure, as these are due to different collision events. Multiple collisions, which are well-defined in our setting, can lead to identical branches within the tree, but the definition $\mathcal{T}$ will discriminate between these and the graph of collisions is still a tree. The only slight abuse of graph-theoretical language is that elements in $\mathcal{T}^\alpha$ with $\alpha > 1$ are still called trees and not "forests".*

The scattering state $\beta : m \to \{0,1\}$ is determined uniquely by the skeleton, i.e. the labels of the particles are immaterial, but the actual computation is not completely trivial. The most important aspect of the computation of β is that the scattering information flows from the leaves to the root, i.e. the scattering state of a node is completely determined by the state of the nodes above, the nodes below are irrelevant.
We will construct now two families of probability measures $P_{t,k}, \hat{P}_{t,k} \in PM(\mathcal{T}^\alpha(Y))$. The empirical distribution $\hat{P}_{t,k}$ is induced by the many-body dynamics and will be constructed recursively in Section 2.4. The mean-field distribution $P_{t,k}$ is given by an explicit formula (20). The link between $P_{t,k}$ and $\hat{P}_{t,k}$ is provided by the set of good trees $\mathcal{G}(a) \subset \mathcal{T}^\alpha(Y)$ (Definition 15) which has the properties that restriction of $\hat{P}_{t,k}$ on $\mathcal{G}(a)\cap\mathcal{T}^\alpha(Y)$ converges to $P_{t,k}$ and $P_{t,k}(\mathcal{G}(a))$ goes to 1 as a tends to 0 (Proposition 17).
This is the crucial step which eventually yields the justification of the mean-field theory. In other words, the main task consists in analyzing the mean-field measure $P_{t,k}$, the empirical distribution $\hat{P}_{t,k}$ enters only when we prove that $P_{t,k}$ is consistent with $\hat{P}_{t,k}$.

2.3. **The mean-field distribution $P_{t,k}$.** We construct now the mean-field distribution of trees $P_{t,k} \in PM(\mathcal{T}^\alpha(Y))$. Let $\Omega \subset \mathcal{T}^\alpha(Y)$ and $t \in [0,\infty)$. The mean field probability that the observed tree is in Ω is given by

$$P_{t,k}(\Omega) = \sum_{m\in\mathcal{T}^\alpha_k} \int_{\Omega\cap\mathcal{E}(m)} e^{-\sum_{j<k}\Gamma_j(\Phi)}\,\mathrm{d}\lambda^m(\phi) \tag{20}$$

where

$$\Gamma_j(\Phi) = \sum_{l\in m, |l|=j} \gamma_l(\Phi),$$

$$\gamma_l(\Phi) = \int_0^{s_l} L[f_0](v_l)\,\mathrm{d}s' = s_l\, L[f_0](v_l) \geq 0 \text{ is the collision rate of particle } l,$$

$$\lambda^m(\phi) = \prod_{i=1}^{\alpha}[\mu(z_i)\otimes\delta(s_i - t)] \otimes \prod_{l\in m\setminus\mathbb{N}} \left[((v_l - v_{\bar{l}})\cdot\nu_l)_+\, \chi_{[s_{l-1},s_{\bar{l}}]}(s_l)\,\mathrm{d}f_0(v_l)\,\mathrm{d}\nu_l\,\mathrm{d}s_l\right], \tag{21}$$

$$\mu(u,v) = \mathbf{1}_{\mathbb{T}^d}(u)\otimes f_0(v).$$

Remark 9. (1) *Note that the positions u_l are completely determined by $(u_l, v_l)_{l\in m\cap\mathbb{N}}$ and $(v_l, s_l, \nu_l)_{l\in m\setminus\mathbb{N}}$. Since we have assumed that $(\nu_l)_{l\in m\cap\mathbb{N}}$ is fixed, the value of $P_{t,k}(\Omega)$ is well-defined.*
(2) *It is noteworthy that the measures $P_{t,k}$ depend on time only via the parameter t. In other words, time plays the role of a parameter which propagates through the tree and qualifies the local branching structure.*

(3) *For some event* $\Omega \subset \mathcal{T}_k(Y)$ *the probability* $P_{t,k'}(\Omega)$ *is independent of* k' *if* $k' > k$. *Equivalently,* $P_{t,k_1}(\Omega \cap \mathcal{E}(m)) = P_{t,k_2}(\Omega \cap \mathcal{E}(m))$, *if the height of* m *is strictly smaller than* $\min\{k_1, k_2\}$.

We can simplify the measure $P_{t,k}$ by integrating over the collision parameters $\nu_l \in S^{d-1}$, $l \in m$. Let $\hat{Y} = \mathbb{R}^d \times [0, \infty)$ be the reduced set of collision data. For every $\Omega \subset \mathcal{T}^\alpha(\hat{Y})$ we find that when still denoting the collision data as ϕ

$$\bar{P}_{t,k}(\Omega) = \sum_{m \in \mathcal{T}_k^\alpha} \int_{\Omega \cap \mathcal{E}(m)} \mathrm{d}\bar{\lambda}^m(\phi)\, e^{-\sum_{j<k} \Gamma_j(\Phi)} \tag{22}$$

with

$$\bar{\lambda}^m(\phi) = \prod_{i=1}^{\alpha} [f_0(v_i) \otimes \delta(s_i - t)] \otimes \prod_{l \in m \setminus \mathbb{N}} \left[\kappa_k\, |v_l - v_{\bar{l}}|\, \chi_{[s_{l-1}, s_{\bar{l}}]}(s_l)\, \mathrm{d}f_0(v_l)\; \mathrm{d}s_l\right].$$

The measures $P_{t,k}$ have the remarkable property that the expectation of certain random variables can be computed efficiently.

Definition 10. *A random variable* $x : \mathcal{T}^1 \to \mathbb{R}$ *is said to be recursive if there exists a family of functions* $h_b : \mathbb{R}^b \to \mathbb{R}$, $b \in \mathbb{N}$, *which are invariant under permutations of the* b *components in* $\mathbb{R}^b$, *such that for all* $m \in \mathcal{T}$ *the equation*

$$x(m) = h_{r_1}(x(m^{(1)}), \ldots, x(m^{(r_1)}))$$

holds, where

$$m^{(j)} = \{(1, l_3, \ldots, l_{|l|}) \mid l \in m \text{ such that } l_2 = j\} \in \mathcal{T}$$

is the j*-th subtree of* m.

In the same way one can define vector values recursive random variables $x : \mathcal{T}^\alpha \to \mathbb{R}^\alpha$. Examples of recursive random variables which are relevant for our purposes are

$$x^{\#}(m) = \#m \text{ (number of nodes)},$$
$$x^{\beta}(m) = \beta_1(m) \text{ (scattering state of the root)}.$$

It is easy to see that if $m \in \mathcal{T}^1$

$$x^{\#}(m) = 1 + \sum_{j=1}^{r_1} x^{\#}(m_j),$$

$$x^{\beta}(m) = \prod_{j=1}^{r_1} (1 - x^{\beta}(m_j)) \text{ with the convention } \prod_{j=1}^{0} (1 - x^{\beta}(m_j)) = 1,$$

hence the functions h_b are given by

$$h_b^{\#}(x_1, \ldots, x_b) = 1 + \sum_{j=1}^{b} x_j,$$

$$h_b^{\beta}(x_1, \ldots, x_b) = \prod_{j=1}^{b} (1 - x_j)$$

which are clearly invariant under permutations of $x_1, \ldots, x_b$. Similar expressions are also valid for $m \in \mathcal{T}^\alpha$ with $\alpha > 1$. The expectation of recursive random variables with respect to the probability measure $P_{t,k}$ can be computed with a simple recurrence relation.

Lemma 11. *Let $\alpha = 1$ and x be a recursive random variable with recurrence functions h_b. Then*

(23)

$$\int \mathrm{d}\bar{P}_{t,k}(\Phi)\, x(m)$$
$$= \int \mathrm{d}f_0(v)\, e^{-\Gamma_1} \sum_{r=0}^{\infty} \int_0^t \mathrm{d}s_1 \int \mathrm{d}\bar{P}_{s_1,k-1}(\Phi_1)\, \kappa_d |v - v_1| \int_{s_1}^t \mathrm{d}s_2 \int \mathrm{d}\bar{P}_{s_2,k-1}(\Phi_2)\, \kappa_d |v - v_2|$$
$$\ldots \int_{s_{r-1}}^t \mathrm{d}s_r \int \mathrm{d}\bar{P}_{s_r,k-1}(\Phi_r)\, \kappa_d |v - v_r|\, h_r(x(m_1), \ldots, x(m_r))$$
$$= \int \mathrm{d}f_0(v)\, e^{-\Gamma_1} \sum_{r=0}^{\infty} \frac{1}{r!} \int_0^t \mathrm{d}s_1 \int \mathrm{d}\bar{P}_{s_1,k-1}(\Phi_1)\, \kappa_d |v - v_1| \int_0^t \mathrm{d}s_2 \int \mathrm{d}\bar{P}_{s_2,k-1}(\Phi_2)\, \kappa_d |v - v_2|$$
$$\ldots \int_0^t \mathrm{d}s_r \int \mathrm{d}\bar{P}_{s_r,k-1}(\Phi_r)\, \kappa_d |v - v_r|\, h_r(x(m_1), \ldots, x(m_r))$$

where $\Gamma_1 = \kappa_d \int \mathrm{d}f_0(v')\, |v - v'|$.

An analogous formula holds if $\alpha > 1$.

Proof. For each $\Phi \in \mathcal{T}(\hat{Y})$ we define nonnegative Radon measures $\bar{\lambda}_l \in M_+(\mathbb{R}^d \times [0, \infty))$ by

$$\bar{\lambda}_l(v, s) = f_0(v)\, |v_{\bar{l}} - v|\, \chi_{[s_{l-1}, s_{\bar{l}}]}(s). \tag{24}$$

With this notation we find the following formula for the measure $\bar{\lambda}^m$.

$$\bar{\lambda}^m(\phi) = f_0(v_1) \prod_{l \in m \setminus \mathbb{N}} \bar{\lambda}_l(\phi_l).$$

Let now $m \in \mathcal{T}^1$. The definition of $P_{t,k}$ yields

$$\int_{\mathcal{E}(m)} \mathrm{d}\bar{P}_{t,k}(m)\, x(m) = \int_{\mathcal{E}(m)} e^{-\sum_{j<k} \Gamma_j(\Phi)}\, \mathrm{d}f_0(v_1) \prod_{i=1}^{r_1} \left[\mathrm{d}\bar{\lambda}_{1i}(\phi_{1i}) \prod_{\substack{l \in m \setminus (\mathbb{N} \cup \mathbb{N}^2) \\ l_2 = i}} \mathrm{d}\bar{\lambda}_l(\phi_l) \right] x(m).$$

We use now the assumption that x is recursive and find

$$\int_{\mathcal{E}(m)} \mathrm{d}\bar{P}_{t,k}(m)\, x(m)$$
$$= \int \mathrm{d}f_0(v)\, e^{-\Gamma_1} \prod_{i=1}^{b} \left[\int_{\mathcal{E}(m_i)} e^{-\sum_{j<k} \Gamma_j^{(i)}(\Phi)} \prod_{l \in m_j \setminus \mathbb{N}} \mathrm{d}\bar{\lambda}_l(\phi_l) \right] h(x(m^{(1)}), \ldots, x(m^{(j)})),$$

where $\Gamma_j^{(i)}(\Phi) = \sum_{l \in m, |l| = j, l_2 = i} \gamma_l(\phi)$. A simple rearrangement yields that

$$\sum_{m \in \mathcal{T}} \int_{\mathcal{E}(m)} \mathrm{d}\bar{P}_{t,k}(\Phi)\, x(m) = \int \mathrm{d}f_0(v)\, e^{-\Gamma_1} \sum_{r=0}^{\infty} \int_0^t \mathrm{d}s_1 \int \mathrm{d}\bar{P}_{s_1,k-1}(\Phi_1)\, \kappa_d |v - v_1|$$
$$\ldots \int_{s_{r-1}}^t \mathrm{d}s_r \int \mathrm{d}\bar{P}_{s_r,k-1}(\Phi_r)\, \kappa_d |v - v_r|\, h_r(x(m_1), \ldots, x(m_r))\,.$$

This demonstrates the first part of (23), to show the second part we observe that

$$\begin{aligned}&\{(s_1,\dots,s_r)\in[0,t]^r \mid s_j\neq s_i \text{ for } i\neq j\}\\ =&\bigcup_{\pi\in S_r}\{(s_1,\dots,s_r)\in[0,t]^r \mid s_{\pi(1)}<s_{\pi(2)}<\ldots<s_{\pi(r)}\},\end{aligned}$$

where S_r denotes the symmetric group on r elements, such that the union is disjoint. As the set, where $s_j = s_i$ for some $i \neq j$ is of measure zero with respect to Lebesgue measure $\mathrm{d}s_1\ldots\mathrm{d}s_r$, we obtain

$$\int_{[0,t]^r} g(s_1,\dots,s_r)\,\mathrm{d}s_1\ldots\mathrm{d}s_r=\sum_{\pi\in S_r}\int_{0\le s_{\pi(1)}<s_{\pi(2)}<\ldots<s_{\pi(r)}\le t} g(s_1,\dots,s_r)\,\mathrm{d}s_1\ldots\mathrm{d}s_r$$

for any $g\in L^1([0,t]^r)$. Now we define

$$g(s_1,\dots,s_r)=\int \mathrm{d}\bar{P}_{s_1,k-1}(\Phi_1)\,\kappa_d|v-v_1|\ldots\int \mathrm{d}\bar{P}_{s_r,k-1}(\Phi_r)\,\kappa_d|v-v_r|\,h(x(m_1),\ldots x(m_r)).$$

We now observe, that

$$\begin{aligned}&\bar{P}_{s_1,k-1}(\Phi_1)\,\kappa_d|v-v_1|\ldots\bar{P}_{s_r,k-1}(\Phi_r)\,\kappa_d|v-v_r|\\ =&\bar{P}_{s_{\pi(1)},k-1}(\Phi_{\pi(1)})\,\kappa_d|v-v_{\pi(1)}|\ldots\bar{P}_{s_{\pi(r)},k-1}(\Phi_{\pi(r)})\,\kappa_d|v-v_{\pi(r)}|\end{aligned}\tag{25}$$

for all permutations $\pi\in S_r$. Next using (25) and the invariance h under permutations, we obtain

$$\begin{aligned}&\int_{0\le s_1<s_2<\ldots<s_r\le t}\int \mathrm{d}\bar{P}_{s_1,k-1}(\Phi_1)\,\kappa_d|v-v_1|\\ &\ldots\int \mathrm{d}\bar{P}_{s_r,k-1}(\Phi_r)\,\kappa_d|v-v_r|h(x(m_1),\ldots x(m_r))\,\mathrm{d}s_1\ldots\mathrm{d}s_r\\ =&\int_{0\le s_{\pi(1)}<s_{\pi(2)}<\ldots<s_{\pi(r)}\le t}\int \mathrm{d}\bar{P}_{s_{\pi(1)},k-1}(\Phi_{\pi(1)})\,\kappa_d|v-v_{\pi(1)}|\\ &\ldots\int \mathrm{d}\bar{P}_{s_{\pi(r)},k-1}(\Phi_{\pi(r)})\,\kappa_d|v-v_r|h(x(m_{\pi(1)}),\ldots x(m_{\pi(r)}))\,\mathrm{d}s_1\ldots\mathrm{d}s_r.\end{aligned}$$

As there are $r!$ different permutations in S_r we finally obtain

$$\begin{aligned}&\int_{0\le s_1<s_2<\ldots<s_r\le t}\int \mathrm{d}\bar{P}_{s_1,k-1}(\Phi_1)\,\kappa_d|v-v_1|\\ &\ldots\int \mathrm{d}\bar{P}_{s_r,k-1}(\Phi_r)\,\kappa_d|v-v_r|\,h(x(m_1),\ldots x(m_r))\,\mathrm{d}s_1\ldots\mathrm{d}s_r\\ =&\frac{1}{r!}\int_{[0,t]^r}\int \mathrm{d}\bar{P}_{s_1,k-1}(\Phi_1)\,\kappa_d|v-v_1|\\ &\ldots\int \mathrm{d}\bar{P}_{s_r,k-1}(\Phi_r)\,\kappa_d|v-v_r|\,h(x(m_1),\ldots x(m_r))\,\mathrm{d}s_1\ldots\mathrm{d}s_r.\end{aligned}$$

Summing over r and m completes the proof of (23). □

As an application of Lemma 11 we obtain an explicit bound on the expected number of nodes in trees.

Lemma 12. *For a tree $m\in\mathcal{T}^\alpha$ the number of non-root nodes is given by $R(m)=\sum_{r\in m} r_l=\#m-\alpha$. The expected value of R satisfies the estimate uniformly in k*

$$\mathbb{E}(R)\le K_{\text{ini}}\sum_{l=1}^{\alpha}\exp(\kappa_d K_{\text{ini}} t_i),\tag{26}$$

with $K_{\mathrm{ini}} = \int_{\mathbb{R}^d} \mathrm{d}f_0(v)\,(1+|v|)^2$ *as in (3).*

Proof. We will only give a proof of (26) in the case $\alpha = 1$, the general case follows by linearity of the expectation. Let $F_{t,k}(v) = \mathbb{E}(R \mid v_1 = v)$ be the conditional expectation of R if we know that velocity of the root is v and that the tree is in $\mathcal{T}_k^1$. Clearly $\mathbb{E}(R) \le \sup_{k\in\mathbb{N}} \int_{\mathbb{R}^d} \mathrm{d}f_0(v)\, F_{t,k}(v)$. The self-similarity relation (23) implies with $x(m) = R(m)$ and $h_r(R(m_1),\dots,R(m_r)) = r + \sum_{i=1}^r R(m_i)$ and $\gamma_1 = L[f_0](v_1)t = \kappa_d t \int_{\mathbb{R}^d} \mathrm{d}f_0(v')\,|v_1 - v'|$ that

$$
\begin{aligned}
&F_{t,k}(v)\\
=&e^{-\gamma_1} \sum_{r=1}^{\infty} \tfrac{1}{r!} \int_0^t \mathrm{d}s_1 \int \mathrm{d}\bar{P}_{s_1,k-1}(m_1)\,\kappa_d|v-v_1|\\
&\dots \int_0^t \mathrm{d}s_r \int \mathrm{d}\bar{P}_{s_r,k-1}(m_r)\,\kappa_d|v-v_r| \left(r + \sum_{i=1}^r R(m_i)\right)\\
=&e^{-\gamma_1} \sum_{r=1}^{\infty} \left(r\frac{(-\gamma_1)^r}{r!} + \frac{\gamma_1^{r-1}}{r!} \sum_{i=1}^r \int_0^t \mathrm{d}s_i\,\kappa_d \int_{\mathbb{R}^d} \mathrm{d}f_0(v_1^{(i)})\,|v - v_1^{(i)}|\,F_{s_i,k-1}(v_1^{(i)}) \right)\\
=&\gamma_1 + \int_0^t \mathrm{d}s\,\kappa_d \int_{\mathbb{R}^d} \mathrm{d}f_0(v')\,|v_1 - v'|\,F_{s,k-1}(v'),
\end{aligned}
$$

where we used the product structure of the integrals. We define now the norm $\|F\|_1 := \sup_{v\in\mathbb{R}^d} \frac{F(v)}{1+|v|}$ and the integral operator A_{f_0} by

$$
(A_{f_0}F)(v) = \kappa_d \int_{\mathbb{R}^d} \mathrm{d}f_0(v')\,|v-v'|F(v'),
$$

so that

$$
F_{t,k} = t\gamma + \int_0^t \mathrm{d}s\, A_f F_{s,k-1}. \tag{27}
$$

We find the estimates

$$
\|A_{f_0}F\|_1 \le \sup_v \frac{\kappa_d \|F\|_1}{1+|v|} \int_{\mathbb{R}^d} \mathrm{d}f_0(v')\,|v-v'|\,(1+|v'|) \le K\|F\|_1,
$$

and

$$
\|\gamma\|_1 = \sup_v \kappa_d \int_{\mathbb{R}^d} \mathrm{d}f_0(v')\,\tfrac{|v-v'|}{1+|v|} \le \kappa_d \int_{\mathbb{R}^d} \mathrm{d}f_0(v')\,(1+|v'|) \le \kappa_d K_{\mathrm{ini}}.
$$

Furthermore $F_{t,k}(v)$ is monotone in k, as $P_{t,k}$ assigns the probability of trees of height greater than $k+1$ to trees of height k, reducing the number of expected nodes. Hence equation (27) implies that

$$
\|F_{t,k}\|_1 \le \kappa_d K \left(t + \int_0^t \mathrm{d}s\,\|F_{s,k}\|_1 \right).
$$

Gronwall's inequality together with the previous estimate implies that

$$
\|F_{t,k}\|_1 \le e^{\kappa_d K_{\mathrm{ini}} t},
$$

where we used that $F_0 \equiv 0$. Since

$$
\mathbb{E}_k(R) = \int_{\mathbb{R}^d} f_0(v)\,F_{t,k}(v) \le \|F_t\|_1 \int_{\mathbb{R}^d} \mathrm{d}f_0(v)\,(1+|v|) \le K e^{\kappa_d K_{\mathrm{ini}} t}
$$

this implies (26) for $\alpha = 1$ and the proof of the lemma is finished. □

We now turn our attention to the determination of the scattering state of the particle at the root of the tree. For a tree $m \in \mathcal{T}^\alpha$ the scattering state $\beta : m \to \{0,1\}$ is defined recursively by $\beta_l = \prod_{l' \in m, \bar{l}' = l}(1 - \beta_{l'})$. This definition rephrases the original definition of the scattering state in (16), adapting it to the tree structure. It is more convenient in our analysis than the ad-hoc definition, which required already some work to show existence, see Lemma 6.
We define the single-particle density $g_{t,k}(\cdot) \in M_+(\mathbb{R}^d)$ via

$$\int_A \mathrm{d}g_{t,k}(v) = P_{t,k}(\beta_1 = 1 \text{ and } v_1 \in A),$$

for all open $A \subset \mathbb{R}^d$. The density $g_{t,k}$ is closely related to the root marginal of $P_{t,k}$ and provides the link between the Boltzmann equation (5) and the mean-field theory distribution of the trees $P_{t,k}$. Due to the simplicity of the distribution $P_{t,k}$ it is possible to characterize the root-marginal of $P_{t,k}$ explicitly.

Proposition 13. *Let $\alpha \in \mathbb{N}$, $\sigma : \mathcal{T}_1^\alpha \to \{0,1\}$, $A \subset \mathbb{R}^d$, $t \in [0,\infty)$ and $k \in \mathbb{N} \cup \{0\}$. Then the equation*

$$\begin{aligned} &P_{t,k+1}\left(v_l \in A \text{ and } \beta_l = \sigma_l \text{ for all } l \in \{1 \ldots \alpha\}\right) \\ &= \prod_{l=1}^{\alpha} \int_A \mathrm{d}v \; \left[(1-\sigma_l)\;(\mathrm{d}f_0(v) - \mathrm{d}f_{t,k}(v)) + \sigma_l\,\mathrm{d}f_{t,k}(v)\right] \end{aligned} \tag{28}$$

holds, where $f_{t,k}$ is the solution of system (15).

This formula shows that in particular $g_{t,k} = f_{t,k-1}$.

Proof. The proposition is proven using induction over k, the case $k = 0$ is just the definition.
In the induction step it is demonstrated that $P_{t,k+1}$ satisfies formula (28) if $P_{t,k}$ does. Since the collision parameters ν are irrelevant we can integrate them out and work with the simplified version (22) of the measure $P_{t,k}$ instead of (20). We define the set of scattering states that are compatible with σ,

$$\mathcal{A}(\sigma) = \left\{ (m,\sigma') \;\middle|\; m \in \mathcal{T}_2^\alpha, \sigma' : m \to \{0,1\} \text{ such that } \prod_{l' \in m, \bar{l}' = l} (1 - \sigma'_{l'}) = \sigma_l \quad \forall l \right\}, \tag{29}$$

with the standard convention $\prod_{j=1}^0 a_j = 1$ for empty products. The induction assumption and equation (23) implies that

$$\begin{aligned} &P_{t,k+1}(v_l \in A \text{ and } \beta_l = \sigma_l \text{ for all } l = 1,\ldots,\alpha) \\ &= \sum_{(m,\sigma') \in \mathcal{A}(\sigma)} \int_{v \in A} \prod_{l \in m \cap \mathbb{N}} \Bigg(\frac{e^{-\gamma_l}}{r_l!}\,\mathrm{d}f_0(v_l) \prod_{\substack{l' \in m \\ \bar{l}' = l}} \Bigg[(1-\sigma'_{l'}) \int_0^{s_l} \mathrm{d}s \int_{v' \in \mathbb{R}^d} \kappa_d |v_l - v'|\,(\mathrm{d}f_0(v') - \mathrm{d}f_{s,k-1}(v')) \\ &\qquad + \sigma'_{l'} \int_0^{s_l} \mathrm{d}s \int_{v' \in \mathbb{R}^d} \mathrm{d}f_{s,k-1}(v')\,\kappa_d |v_l - v'| \Bigg] \Bigg) \\ &= \int_{v \in A} \sum_{(m,\sigma') \in \mathcal{A}(\sigma)} \prod_{l=1}^{\alpha} \left[\mathrm{d}f_0(v_l)\, I_k(\sigma', v, l)\right] \end{aligned}$$

where

$$\begin{aligned}
&I_k(\sigma', v, l)\\
=&\tfrac{e^{-\gamma_l}}{r_l!}\prod_{l'\in m,\bar{l}'=l}\Big[(1-\sigma'_{l'})\int_0^{t_l}\mathrm{d}s\int_{v'\in\mathbb{R}^d}\kappa_d|v_l-v'|(\mathrm{d}f_0(v')-\mathrm{d}f_{s,k-1}(v'))\\
&\quad+\sigma'_{l'}\int_0^{t_l}\mathrm{d}s\int_{v'\in\mathbb{R}^d}\mathrm{d}f_{s,k-1}(v')\,\kappa_d|v_l-v'|\Big]\\
=&\tfrac{e^{-\gamma_l}}{r_l!}\prod_{l'\in m,\bar{l}'=l}\Big[(1-\sigma'_{l'})\Big(\gamma_l-\int_0^{t_l}\mathrm{d}s\,L[f_{s,k-1}](v_l)\Big)+\sigma'_{l'}\int_0^{t_l}\mathrm{d}s\,L[f_{s,k-1}](v_l)\Big].
\end{aligned}$$

Note that $\mathcal{A}(\sigma)=\prod_{l=1}^{\alpha}\mathcal{A}(\sigma_l)$ if we identify single trees with α roots and α trees with single roots. The algebraic identity $\sum_{i\in I^J}\prod_{j\in J}a(i_j,j)=\prod_{j\in J}\sum_{i\in I}a(i,j)$, where I and J are finite sets and $a: I\times J\to\mathbb{R}$ is a function, implies that $\sum\limits_{\sigma'\in\mathcal{A}(\sigma)}\prod\limits_{l=1}^{\alpha}I_k(\sigma',v,l)=\prod\limits_{l=1}^{\alpha}\sum\limits_{\sigma'\in\mathcal{A}(\sigma_l)}I_k(\sigma',v,1)$. This yields that

(30)

$$P_{t,k+1}(v_l\in A\text{ and }\beta_l=\sigma_l\ \forall l=1\ldots\alpha)=\prod_{l=1}^{\alpha}\int_{v\in A}\mathrm{d}f_0(v)\,[(1-\sigma_l)J_k(0,v)+\sigma_l J_k(1,v)],$$

with $J_k(\sigma,v)=\sum_{\sigma'\in\mathcal{A}(\sigma)}I_k(\sigma',v,1)$. Since by definition $\mathcal{A}(1)=\{(1),(1,(0)),(1,(0,0))\ldots\}$, this shows that

$$J_k(1,v)=\sum_{j=0}^{\infty}\frac{e^{-\gamma}}{j!}\left(\gamma-\int_0^s\mathrm{d}s'\,L[f_{s',k-1}](v)\right)^j=e^{-\int_0^s\mathrm{d}s'\,L[f_{s',k-1}](v)},\tag{31}$$

with $\gamma=sL[f_0](v)$. Clearly $\int_A\mathrm{d}f_0(v)\,J_k(0,v)+\int_A\mathrm{d}f_0(v)\,J_k(1,v)=\int_A\mathrm{d}f_0(v)$ and therefore

(32)

$$\int_{v\in A}\mathrm{d}f_0(v)\,J_k(0,v)=\int_{v\in A}\mathrm{d}f_0(v)\,(1-J_k(1,v))=\int_{v\in A}\mathrm{d}f_0(v)\left(1-e^{-\int_0^s\mathrm{d}s'\,L[f_{s',k-1}]}\right).$$

Plugging the formulas (31) and (32) into equation (30) yields that

$$\begin{aligned}
&P_{t,k+1}(v_l\in A\text{ and }\beta_l=\sigma\ \forall l=1,\ldots,\alpha)\\
=&\prod_{l=1}^{\alpha}\int_{v\in A}\left[(1-\sigma_l)\,\mathrm{d}f_0(v)\,\left(1-e^{-\int_0^{t_l}\mathrm{d}s\,L[f_{s,k-1}]}\right)+\sigma_l\,\mathrm{d}f_0(v)\,e^{-\int_0^{t_l}\mathrm{d}s\,L[f_{s,k-1}]}\right]\\
\overset{(15)}{=}&\prod_{l=1}^{\alpha}\int_{v\in A}\left[(1-\sigma_l)\,(\mathrm{d}f_0(v)-\mathrm{d}f_{t_l,k}(v))+\sigma_l\,\mathrm{d}f_{t_l,k}(v)\right]
\end{aligned}$$

and formula (28) has been established. □

2.4. **The empirical distribution $\hat{P}_{t,k}$.** We return now to the hierarchy of many body evolutions described in Section 2.1. The initial values of the particles form a random set $\omega\subset\mathbb{T}^d\times\mathbb{R}^d$ and it is assumed that the law of ω is the Poisson point process with density $N\mu$, where $\mu=\mathbf{1}_{\mathbb{T}^d}\otimes f_0\in PM(\mathbb{T}^d\times\mathbb{R}^d)$. Hence, the size $n=\#\omega$ is Poissonian random variable with intensity N. As explained in Section 2.2, the family of probability measures $\hat{P}_{t,k}\in PM(\mathcal{T}(Y))$ is the empirical distribution of the tree Φ which is generated by the many-body evolution and has a randomly chosen (tagged) particle as its root. This tree

is only well defined if $n > 0$, i.e. ω is non-empty. For this reason we define $P_{t,k}(\Omega)$ as the conditional probability that the tree is contained in the set Ω, given that $n = \#\omega > 0$.
A particularly simple method of sampling from this conditional distribution consists in drawing a realization of ω according to the unconditioned Poisson point process, and an independent random variable $z \in \mathbb{T}^d \times \mathbb{R}^d$ with law $\mu(z) = \mathbf{1}_{\mathbb{T}^d}(u) \otimes f_0(v)$ which is the initial value of the tagged particle. It can be checked without difficulty that the joint distribution of ω and z is the previously defined conditional distribution.
The trees generated by this procedure are denoted by $\Phi(t,k) = (m(t,k), \phi) \in \mathcal{T}_k(Y)$, where $m(t,k) \in \mathcal{T}_k$ is the skeleton and $\phi : m(t,k) \to Y$ specifies the initial values, the collision times and the impact parameters. The measures $\hat{P}_{t,k}$ are the image measure of $\mathrm{Prob}_{\mathrm{ppp}}$ induced by the many-particle flows so that for each $\Omega \subset \mathcal{T}(Y)$ we obtain

$$\hat{P}_{t,k}(\Omega) := \mathrm{Prob}_{\mathrm{ppp}}((m(t,k), \phi) \in \Omega). \tag{33}$$

The tree measures $\hat{P}_k$ are derived from $\mathrm{Prob}_{\mathrm{ppp}}$, but $\mathrm{Prob}_{\mathrm{ppp}}$ cannot be derived from $\hat{P}_{t,k}$. By construction, for fixed ω the skeleton m is monotonously increasing in t and k, and for fixed $l \in m$ the data ϕ_l does not depend on t or k. This is equivalent to saying that the j-marginal of $\hat{P}_{t,k}$ (trees of hight $j \leq k$) is given by $\hat{P}_{t,j}$, i.e.

$$\hat{P}_{t,k}\left((m(t,k) \cap (\cup_{i=1}^{j} \mathbb{N}^i), (\phi_l)_{|l| \leq j}) \in \Omega\right) = \hat{P}_{t,j}((m(t,j), (\phi_l)_{|l| \leq j}) \in \Omega) \tag{34}$$

for all $\Omega \subset \mathcal{T}_j(Y)$, $k \geq j$.
We will use formula (34) to construct an alternative characterization of $\hat{P}_{t,k}$ which reflects the iterative process that underlies the definition of $m(t,k)$. Using this alternative characterization one can easily establish total-variation bounds for $P_{t,k} - \hat{P}_{t,k}$. Since the time t is arbitrary but fixed we will often write $\hat{P}_k$ instead of $\hat{P}_{t,k}$.
Let $(m', \phi') \in \mathcal{T}_{k-1}(Y)$ and let $\hat{P}_k(\,\cdot\,|\,(m', \phi')) \in PM(\mathcal{T}_k(Y))$ be the conditional distribution of $\hat{P}_k$ in the sense that

$$\hat{P}_k(\Omega \,|\, (m', \phi')) := \hat{P}_k\Big((m(k), \phi) \in \Omega \mid m \cap \mathbb{N}^j = m' \cap \mathbb{N}^j \text{ for all } j \in \{1 \dots k-1\} \\ \text{and } \phi_l = \phi'_l \text{ for all } l \in m \text{ such that } |l| < k\Big).$$

Formula (34), which characterizes the j-marginals of $\hat{P}_{t,k}$, yields the following recurrence relation for $\hat{P}_k$:

$$\hat{P}_k(\Omega) = \int_{\mathcal{T}_{k-1}(Y)} \mathrm{d}\hat{P}_{k-1}(\Phi')\, \hat{P}_k(\Omega \,|\, \Phi'). \tag{35}$$

Repeating this step $k-1$ times we obtain the following iterative representation of $\hat{P}_k$:

(36)

$$\hat{P}_k(\Omega) = \int_{\mathcal{T}_1(Y)} \mathrm{d}P_1(\Phi_1) \int_{\mathcal{T}_2(Y)} \mathrm{d}\hat{P}_2(\Phi_2 \,|\, \Phi_1) \dots \int_{\mathcal{T}_{k-1}(Y)} \mathrm{d}\hat{P}_{k-1}(\Phi_{k-1} \,|\, \Phi_{k-2})\, \hat{P}_k(\Omega \,|\, \Phi_{k-1}),$$

where

$$P_1(z_1 \dots z_\alpha) = \prod_{l=1}^{\alpha} \mu(z_l) \in PM\left((\mathbb{T}^d \times \mathbb{R}^d)^\alpha\right) \tag{37}$$

is the distribution of α initial values.

2.5. **Convergence of $\hat{P}_k$ to P_k.** Having constructed an iterative characterization of $\hat{P}_k$ we will now show that it is very similar to the mean field measure P_k in a precise way. The key is to identify the mechanisms by which the two probability distributions fail to be equal. In this part of the paper we will work with the phase-space representation of the trees: $z_l = (u_l, v_l) \in \mathbb{T}^d \times \mathbb{R}^d$.

Remark 14. *There are only two reasons why $\hat{P}_k$ fails to coincide with P_k in the limit $a \to 0$:*

(1) *The cylinders which are covered by the paths of the particles might contain self-intersections due to the periodic boundary conditions: $v - v' \in R(t,a)$ with*

$$R(t,a) = \left\{ v \in \mathbb{R}^d \mid \min\{|s\,v - \xi| \mid s \in [0,t], \xi \in \mathbb{Z}^d \setminus \{0\}\} \le a \right\}. \tag{38}$$

(2) *Nodes might have more than one child, i.e. the map $z : m \to \mathbb{T}^d \times \mathbb{R}^d$ might be not injective.*

The set $R(t,a)$, which can easily seen to be nonempty, is relevant due to periodic boundary conditions, which will lead to self-intersections of the cylinders. This happens, if $v - v_j$ is sufficiently close to a velocity v^*, where the components of $v_1^*, \ldots, v_d^*$ are rationally dependent, i.e. $\eta \cdot v^* \in \mathbb{Z}$ with $\eta \in \mathbb{Z}^d$, but only if $|\eta| \le t$. The effect is not present in a setting where $(u,v) \in \mathbb{R}^d \times \mathbb{R}^d$.
The second effect is caused by the notorious recollisions. These dependencies disappear as the diameter a tends to zero.
We stipulate now a strict order of the set of nodes m:

$$l < l' \text{ if either } |l| < |l'| \text{ or } (|l| = |l'| \text{ and } \bar{l} < \bar{l}') \text{ or } (\bar{l} = \bar{l}' \text{ and } l_{|l|} < l'_{|l|}) \tag{39}$$

This order is induced by the link between the collision time and the indices $l \in m$ in Definition 7.
Motivated by Remark 14 we define the set of "good" trees.

Definition 15. *For each $a_0 > 0$ the set of "good" trees $\mathcal{G}(a_0) \subset \mathcal{T}(Y)$ consists of those trees $(m, \phi) \in \mathcal{T}(Y)$ with the property that for all $0 < a \le a_0$ and all $l \in m$*

$$v_l - v_{\bar{l}} \in \mathbb{R}^d \setminus R(t,a) \qquad \textit{(all parent-child-pairs are non-resonant),} \tag{40}$$

$$z_l \notin \cup_{\substack{l' < l \\ l' \neq \bar{l}}} C_{l'} \qquad \textit{(no node has more than one child),} \tag{41}$$

where we associate to each node $l \in m$ the set of colliding initial values

$$C_l = \left\{ z' \in \mathbb{T}^d \times \mathbb{R}^d \,\middle|\, \min_{s' \in [0, s_l]} |\mathrm{dist}(z_l, z', s')| \le a \right\},$$

and dist *as in (8) ignores overlap in the initial data.*

Note that $\mathcal{G}(a_0) \subset \mathcal{T}(Y)$ is a family of sets which decreases with a_0. An elementary calculation yields that for all $v' \in \mathbb{R}^d \setminus (v_l + R(t))$

$$N\,\mathcal{H}^{d-1}\left(C_l \cap (\mathbb{T}^d \times \{v'\})\right) = \kappa_d |v_l - v'| s_l. \tag{42}$$

The significance of $\mathcal{G}(a_0)$ is given by the following results

$$\lim_{a_0 \to 0} \inf_k P_k(\mathcal{G}(a_0)) = 1, \tag{43}$$

$$\lim_{a \to 0} \sup_k \left| \hat{P}_k(\Omega) - P_k(\Omega) \right| = 0 \text{ for all } \Omega \subset \mathcal{G}(a_0) \text{ if } a_0 \text{ is fixed,} \tag{44}$$

which are proven in Proposition 17. For the proof we need a more explicit characterization of the distributions $\hat{P}_k(\cdot \mid \Phi_{k-1})$ and $\hat{P}_k(\cdot)$

As an intermediate step we recall a formula which yields the probability of certain complex events with respect to Poisson-point processes. Let $A \subset \cup_{n=0}^{\infty} (\mathbb{T}^d \times \mathbb{R}^d)^n$ be a symmetric set, i.e. $z \in A \cap (\mathbb{T}^d \times \mathbb{R}^d)^n$ if and only if $(z_{\pi(1)}, \ldots, z_{\pi(n)}) \in A \cap (\mathbb{T}^d \times \mathbb{R}^d)^n$ for all permutations $\pi \in S_n$, where S_n is the symmetric group. We use the convention that $(\mathbb{T}^d \times \mathbb{R}^d)^0$ is a single point. For each realization $\omega \subset \mathbb{T}^d \times \mathbb{R}^d$ of the point process we chose an arbitrary enumeration of the elements of ω such that $\omega = \{z_1, \ldots, z_n\}$. We say that $\omega \in A$ if $(z_1, \ldots, z_n) \in A$; the choice of the enumeration is irrelevant since A is symmetric. It can be checked that if ω is a realization of the Poisson-point process with intensity $\mu \in M_+(\mathbb{T}^d \times \mathbb{R}^d)$, then

$$\text{Prob}_{\text{ppp}}(\omega \in A) = e^{-\mu(\mathbb{T}^d \times \mathbb{R}^d)} \sum_{n=0}^{\infty} \frac{1}{n!} \int_{A \cap (\mathbb{T}^d \times \mathbb{R}^d)^n} \mathrm{d}\mu(z_1) \ldots \mathrm{d}\mu(z_n), \tag{45}$$

where the value of integral for $n = 0$ is 1 if $(\mathbb{T}^d \times \mathbb{R}^d)^0 \subset A$ and 0 else. By the definition of Poisson-point processes each set $\mathcal{C} \subset \mathbb{T}^d \times \mathbb{R}^d$ defines a projection denoted by $\mathcal{C} \cap \omega$. We recall the following fundamental independence-principle of Poisson-point processes which asserts that even if we have obtained a certain amount of information over a realization ω of a Poisson-point process it is still possible to use a suitably modified version of formula (45).

Lemma 16. *Let the random set $\omega \subset \mathbb{T}^d \times \mathbb{R}^d$ be distributed according to a Poisson point-process with density μ, $\bar{\mathcal{C}}, \mathcal{C} \subset \mathbb{T}^d \times \mathbb{R}^d$ and $A \subset \cup_{r=0}^{\infty} (\mathcal{C} \setminus \bar{\mathcal{C}})^r$ be symmetric. Then we obtain the following formula for the conditional probability of the event A:*

$$\text{Prob}_{\text{ppp}}\left(\omega \cap \mathcal{C} \in A \;\middle|\; \omega \cap \bar{\mathcal{C}} = \emptyset\right) = \exp\left(-\mu(\mathcal{C} \setminus \bar{\mathcal{C}})\right) \sum_{r=0}^{\infty} \frac{1}{r!} \int_{A \cap \mathcal{C}^r} \mathrm{d}\mu^r(z), \tag{46}$$

where $\mu^r = \underbrace{\mu \otimes \ldots \otimes \mu}_{r \text{ terms}}$

Proof. See section 4. □

To apply Lemma 16 we have to work with the phase space representation of trees. Owing to the decomposition $\Omega = \dot{\cup}_{m \in \mathcal{T}} \mathcal{E}(m) \cap \Omega$ we can assume that $\Omega \subset \mathcal{E}(m)$ for some $m \in \mathcal{T}$. Due to this simplification we can drop the sum in equation (46) since only one term is nontrivial.

Note that for a general tree $\Phi = (m, \phi) \in \mathcal{T}(Y)$ the number of nodes $\#m$ can be bigger than the number of particles involved in the collisions, i.e. it is possible that the map $z : m \to \mathbb{T}^d \times \mathbb{R}^d$ is not injective and $z_l = z_{l'}$ for some pair $l, l' \in m$, $l \neq l'$. This scenario corresponds to a bad tree where one node has two child nodes. For this reason we restrict our attention to sets Ω which are subsets of $\mathcal{G}(a)$. The excluded set has nonzero probability, however we will show that the probability of $\mathcal{T}(Y) \setminus \mathcal{G}(a)$ tends with a to 0. By construction for all trees in Ω the map $l \mapsto z_l$ is injective.

The order defined by (39) induces a representation of the events $\Omega \subset \mathcal{T}(Y)$ in phase-space coordinates:

$$A(\Omega) \subset (\mathbb{T}^d \times \mathbb{R}^d)^{\#m}.$$

In the same spirit one obtains a one-to-one correspondence between the the initial values of particles associated with the tree-nodes at height k and subsets of $(\mathbb{T}^d \times \mathbb{R}^d)^{\#m \cap \mathbb{N}^k}$:

$$Z_k = (z_l)_{|l|=k} \in (\mathbb{T}^d \times \mathbb{R}^d)^{\#m \cap \mathbb{N}^k}.$$

We will also need the conditional events

$$A_k(\Omega, \Phi) = \left\{ Z_k \in (\mathbb{T}^d \times \mathbb{R}^d)^{\#m \cap \mathbb{N}^k} \,|\, (Z_k, \Phi) \in \Omega \right\},$$

where $\Phi \in \mathcal{T}_{k-1}(Y)$ and $(Z_k, \Phi) \in \mathcal{T}_k(Y)$ is the tree obtained by attaching the leaves Z_k to the topmost nodes of Φ.
Recall that the density of the Poisson-point process which generates the initial positions of the particles is given by $N\mu$ where

$$\int \mathrm{d}\mu(z)\,\varphi(z) = \int_{\mathbb{R}^d} \mathrm{d}f_0(v) \int_{\mathbb{T}^d} \mathrm{d}u\,\varphi(u,v)$$

for every testfunction $\varphi \in C_c(\mathbb{T}^d \times \mathbb{R}^d)$.
Before applying Lemma 16 we have to specify the sets $\mathcal{C}$ and $\bar{\mathcal{C}}$. Fix $a_0 > 0$ and let $\Phi \in \mathcal{T}(Y) \cap \mathcal{G}(a_0)$. We are interested in the distribution of those trees which coincide with Φ up to level k. Clearly, the initial positions of the particles at height $k+1$ are contained in the set

$$\mathcal{C}_k(\Phi) := \bigcup_{l \in m_k \cap \mathbb{N}^k} C_l(\phi) \subset \mathbb{T}^d \times \mathbb{R}^d,$$

with $\Phi = (m, \phi)$. In order to apply formula (46) we have to identify the conditioning of the distribution $\omega \cap \mathcal{C}_k(\Phi)$. Define the collection of cylinders

$$\bar{\mathcal{C}}_k(\Phi) := \bigcup_{|l|<k} C_l(\phi) \subset \mathbb{T}^d \times \mathbb{R}^d$$

which contains those initial values that would affect the lower nodes. By construction the information on the point process ω that we have accumulated so far is given by $\omega \cap \bar{\mathcal{C}}_k(\Phi) = \{z_l \mid |l| \leq k\}$. Furthermore, since $\Phi \in \mathcal{G}(a_0)$ we have that $\omega \cap \mathcal{C}_k(\Phi) \cap \bar{\mathcal{C}}_k(\Phi) = \emptyset$. This implies that for each $\Omega \subset \mathcal{T}(Y) \cap \mathcal{G}(a_0)$ and $\Phi \in \mathcal{T}_k(Y) \cap \mathcal{G}(a_0)$ that

$$\hat{P}_{k+1}(\Omega \,|\, \Phi) = \mathrm{Prob}_{\mathrm{ppp}}(\mathcal{C}_k(\Phi) \cap \omega \in \mathrm{sym}(A_k(\Omega, \Phi)) \mid \mathcal{C}_k(\Phi) \cap \bar{\mathcal{C}}_k(\Phi) \cap \omega = \emptyset).$$

where $\mathrm{sym}(A)$ is the symmetrization of the set A, i.e. $(z_1, \dots, z_n) \in \mathrm{sym}(A)$ if there exists a permutation $\pi \in S_n$ such that $(z_{\pi(1)}, \dots, z_{\pi(n)}) \in A$; in particular $A \subset \mathrm{sym}(A)$. This is the crucial step where the complicated dependency on the past of the many-body evolution is reduced to a simple conditional expectation of the Poisson point process. Since $A(\Omega, \Phi) \cap \underbrace{\bar{\mathcal{C}}_k(\Phi) \times \ldots \times \bar{\mathcal{C}}_k(\Phi)}_{r \text{ terms}} = \emptyset$ for each r we can use formula (46) and deduce that

$$P_{k+1}(\Omega \,|\, \Phi) = e^{-\hat{\Gamma}_k(\Phi)} \frac{1}{r!} \int_{\mathrm{sym}(A_{k+1}(\Omega,\Phi))} \mathrm{d}\mu^r(Z_{k+1})$$

where

$$\hat{\Gamma}_k(\Phi) = \mu(\hat{\mathcal{C}}_k(\Phi)) \tag{47}$$

and $\hat{\mathcal{C}}(k) = \mathcal{C}_k(\Phi) \setminus \bar{\mathcal{C}}_k(\Phi)$. Recall the convention that the value of the integral over $(\mathbb{T}^d \times \mathbb{R}^d)^0$ is 1.
Since each permutation of the labels $l \in m$ destroys the tree structure we obtain that if $z_\pi \in A$ and $z \in A$, then necessarily π is the identity transformation, i.e. $z_\pi = z$. This implies that if we replace in the above formula $\mathrm{sym}(A)$ by the non-symmetric set A we have to drop the term $\frac{1}{r!}$.

$$P_{k+1}(\Omega \,|\, \Phi) = e^{-\hat{\Gamma}_k(\Phi)} \int_{A_{k+1}(\Omega,\Phi)} \mathrm{d}\mu^r(Z_{k+1}). \tag{48}$$

Plugging the expression (48) for the conditional expectation $\hat{P}_{k+1}(\cdot\,|\,\Phi)$ into equation (36) yields that

$$\hat{P}_k(\Omega) = \int_{(\mathbb{T}^d\times\mathbb{R}^d)^\alpha} \mathrm{d}P_1(\phi_1(Z_1))\, e^{-\hat{\Gamma}_1(\Phi_1(Z_1))} \int_{(\mathbb{T}^d\times\mathbb{R}^d)^{r_2}} \mu^{r_2}(\Phi_2(Z_2))$$

$$\dots e^{-\hat{\Gamma}_{k-1}(\Phi_{k-1}(Z_1\dots Z_{k-1}))} \int_{A_k(\Omega,\Phi_{k-1}(Z_1\dots Z_{k-1}))} \mathrm{d}\mu^{r_k}(Z_k)$$

$$= \sum_{m\in\mathcal{T}_k} \int_{A(\Omega)} \mathrm{d}\mu^{\#m}(z)\, e^{-\sum_{j<k}\hat{\Gamma}_j(\Phi(z))}. \tag{49}$$

The intermediate step in the computation above relies on the additional assumption that $m \in \mathcal{T}_k\backslash\mathcal{T}_{k-1}$. In general we have to be more careful concerning the domains of integration, but the the final formula is unaffected.
We return now to the collision representation of the trees. This means that the variables $(z_l)_{l\in m}$ are replaced by $(u_1, v_1) \times (s_l, \nu_l, v_l)_{l\in m\backslash\mathbb{N}}$ if $\alpha = 1$ and analogously if $\alpha > 1$. The determinant of the derivative of this transformation is given by

$$\det D_\Phi z(\Phi) = \prod_{l\in m\backslash\mathbb{N}} \left(a^{d-1}[\nu_l\cdot(v_l - v_{\bar{l}})]_+.\right)$$

Thus changing coordinates in the integrals we obtain that for each $m \in \mathcal{T}$

$$\int_{A(\Omega)} e^{-\sum_{j<k}\hat{\Gamma}_j(\Phi(z))}\, \mathrm{d}\mu^{\#m}(z)$$

$$= \int_\Omega \mathrm{d}P_1(z_1\dots z_\alpha)\, e^{-\sum_{j<k}\hat{\Gamma}_j(\Phi)} \prod_{l\in m\backslash\mathbb{N}} \left(N\,\mathrm{d}f_0(v_l)\,\mathrm{d}\nu_l\,\mathrm{d}s_l\,\chi_{[0,s_{\bar{l}}]}(s_l)\, a^{d-1}\,[(v_l - v_{\bar{l}})\cdot\nu_l]_+\right)$$

$$\overset{(1)}{=} \int_\Omega \mathrm{d}P_1(z_1\dots z_\alpha)\, e^{-\sum_{j<k}\hat{\Gamma}_j(\Phi)} \prod_{l\in m\backslash\mathbb{N}} \left(\mathrm{d}f_0(v_l)\,\mathrm{d}\nu_l\,\mathrm{d}s_l\,\chi_{[0,s_{\bar{l}}]}(s_l)\,[(v_l - v_{\bar{l}})\cdot\nu_l)]_+\right)$$

$$= \int_\Omega \mathrm{d}\lambda^m(\phi)\, e^{-\sum_{j<k}\hat{\Gamma}_j(\Phi)},$$

Thus we have shown that for all $\Omega \subset \mathcal{G}(a)$

$$\hat{P}_k(\Omega) = \sum_{m\in\mathcal{T}_k} \int_{\Omega\cap\mathcal{E}(m)} e^{-\sum_{j<k}\hat{\Gamma}_j(\Phi)}\, \mathrm{d}\lambda^m(\phi). \tag{50}$$

and

$$P_k(\Omega) = \hat{P}_k(\Omega) + e_k(\Omega), \tag{51}$$

where the error has the form

$$e_k(\Omega) = \sum_{m\in\mathcal{T}_k} \int_{\Omega\cap\mathcal{E}(m)} \mathrm{d}\lambda^m(\phi) \left(e^{-\sum_{j<k}\hat{\Gamma}_j(\Phi)} - e^{-\sum_{j<k}\Gamma_j(\Phi)}\right). \tag{52}$$

Since $\hat{\Gamma}_j(\Phi) \le \Gamma_j(\Phi)$ the difference $e_k(\cdot)$ is a non-negative measure.
Now we are in a good position to prove that equations (43) and (44) hold.

Proposition 17 (Similarity of $\hat{P}_k$ and P_k)**.** *Let $\mathcal{G}(a)$ the set of good trees from Definition 15, and $\Omega \subset \mathcal{G}(a_0)$. Then equations (43) and (44) hold.*

Proof. For technical reasons we decouple the dependency of $\mathcal{G}$ and $\hat{P}_k$ on the scaling parameter a. We will construct a family of sets of trees $\hat{\mathcal{G}}(a) \subset \mathcal{G}(a)$ with the following two properties

$$\lim_{a_0 \to 0} \inf_k P_k\left(\hat{\mathcal{G}}(a_0)\right) = 1, \tag{53}$$

$$\lim_{a_0 \to 0} \lim_{a \to 0} \sup_k \left| \hat{P}_k\left(\Omega \cap \hat{\mathcal{G}}(a_0)\right) - P_k\left(\Omega \cap \hat{\mathcal{G}}(a_0)\right)\right| = 0 \tag{54}$$

for all $\Omega \subset \mathcal{T}(Y)$. The idea is that the trees in the sets $\hat{\mathcal{G}}(a_0)$ have additional good properties which are controlled by a_0. It is quite clear that for our choice of $\hat{\mathcal{G}}(a_0)$ (see (56)) equation (54) holds even for fixed a_0 but without the limit the proof becomes more complicated.
We show first that (53) and (54) imply (44): Since $\hat{P}_k$ and P_k are probability measures equation (53) implies that

$$\lim_{a_0 \to 0} \lim_{a \to 0} \sup_k \left| \hat{P}_k\left(\mathcal{T}(Y) \setminus \hat{\mathcal{G}}(a_0)\right) - P_k\left(\mathcal{T}(Y) \setminus \hat{\mathcal{G}}(a_0)\right)\right| = 0. \tag{55}$$

Let now $\Omega \subset \hat{\mathcal{G}}(a_0)$ for some $a_0 > 0$ and fix $\varepsilon > 0$. Then

$$\begin{aligned}
\lim_{a_0 \to 0} \lim_{a \to 0} \sup_k |\hat{P}_k(\Omega) - P_k(\Omega)| &\leq \lim_{a_0 \to 0} \lim_{a \to 0} \sup_k \left| \hat{P}_k\left(\Omega \cap \hat{\mathcal{G}}(a_0)\right) - P_k\left(\Omega \cap \hat{\mathcal{G}}(a_0)\right)\right| \\
&\quad + \lim_{a_0 \to 0} \lim_{a \to 0} \sup_k \hat{P}_k\left(\Omega \setminus \hat{\mathcal{G}}(a_0)\right) + \sup_k P_k\left(\Omega \setminus \hat{\mathcal{G}}(a_0)\right) \\
&\overset{(54)}{=} \lim_{a_0 \to 0} \lim_{a \to 0} \sup_k \hat{P}_k\left(\Omega \setminus \hat{\mathcal{G}}(a_0)\right) + \lim_{a_0 \to 0} \sup_k P_k\left(\Omega \setminus \hat{\mathcal{G}}(a_0)\right) \\
&\overset{(55)}{\leq} 2 \lim_{a_0 \to 0} \sup_k P_k\left(\mathcal{T}(Y) \setminus \hat{\mathcal{G}}(a_0)\right) \overset{(53)}{=} 0.
\end{aligned}$$

Equation (43) follows directly from (53) since $\hat{\mathcal{G}}(a) \subset \mathcal{G}(a)$.
Let $\varepsilon(a)$ and $V(a)$ be monotone functions of a such that $\lim_{a \to 0} \varepsilon(a) = 0$ and $\lim_{a \to 0} V(a) = +\infty$. which will be determined later. We define the set

$$\begin{aligned}
\hat{\mathcal{G}}(a_0) = \bigcap_{a < a_0} \Bigg\{ (m, \phi) \in \mathcal{G}(a) \,\Bigg|\, & \min_{\substack{l, l' \in m \\ l' \neq l}} |v_l - v_{\bar{l}}| \geq \varepsilon(a) \text{ and } |v| \leq V(a) \\
& \text{and } \min_{l \in m} \min_{l' < l, l' \neq \bar{l}} \left(1 - \left| \tfrac{v_l - v_{\bar{l}}}{|v_l - v_{\bar{l}}|} \cdot \tfrac{v_{l'} - v_{\bar{l}}}{|v_{l'} - v_{\bar{l}}|}\right|\right) \geq \varepsilon(a) \Bigg\}.
\end{aligned} \tag{56}$$

Due to the monontonicity of $\varepsilon(a)$ and $V(a)$ the set $\hat{\mathcal{G}}(a)$ is increasing with a.
Before proving (53) and (54) we will first estimate the size of the set $R(t, a)$ and demonstrate that

$$\lim_{a \to 0} \int_{R(t,a)} (1 + |v|) \, \mathrm{d}f_0(v) = 0. \tag{57}$$

For each $\xi \in \mathbb{Z}^d \setminus \{0\}$ we define the cone

$$M(\xi, a) = \left\{ v \in \mathbb{R}^d \,|\, (v \cdot \xi)^2 \geq (|\xi|^2 - a^2)|v|^2 \right\}.$$

Let $c(a) := \sup\{\int_{M(\xi,a)} \mathrm{d}f_0(v) \,|\, \xi \in \mathbb{Z}^d \setminus \{0\}\}$ be an upper bound for the volume of this cone. Assumption (4) implies that $c(a) = o(1)$ as $a \to 0$. For each $v \in R(t, a)$ such that $|v| \leq V$ there exists $\xi(v) \in \mathbb{Z}^d \setminus \{0\}$ such that $|\xi(v)| \leq Vt + a$ and $v \in M(\xi(v), a)$, i.e.

each velocity $v \in R(t,a)$ is an element of one of at most $(2tV+2a)^d$ cones. Thus we obtain, using (3),

$$\int_{R(t,a)} (1+|v'|)\,\mathrm{d}f_0(v') \le \int_{R(t,a)\cap\{|v'|\le V\}} (1+|v'|)\,\mathrm{d}f_0(v') + \int_{\{|v'|>V\}} (1+|v'|)\,\mathrm{d}f_0(v')$$
$$\le (1+V)(2tV+2a)^d c(a) + K_{\mathrm{ini}}/V,$$

with $K_{\mathrm{ini}} = \int_{\mathbb{R}^d} \mathrm{d}f_0(v)\,(1+|v|)^2$. So choosing first V large the second term is small. Then choose a so that the first term is small, which completes the proof of the equation (57).

Proof of equation (53).

First we show that we can restrict ourselves to bounded trees. By Lemma 12, the expected value of the number of nodes $\#m$ in a tree m is bounded by $K_{\mathrm{ini}}\exp(\kappa_d K_{\mathrm{ini}} t)$. As $\#m$ is a positive function, this implies immediately the estimate

$$\sum_{\#m-\alpha>r} P_k(\mathcal{E}(m)) < \tfrac{K_{\mathrm{ini}}}{r}\exp(\kappa_d K_{\mathrm{ini}} t). \tag{58}$$

This estimate gives us control over the error which arises if we ignore all trees with more than r nodes:

$$1 = \sum_{m\in\mathcal{T}} P_k(\mathcal{E}(m)) = \sum_{\substack{m\in\mathcal{T}\\ \#m-\alpha\le r}} P_k(\mathcal{E}(m)) + \sum_{\substack{m\in\mathcal{T}\\ \#m-\alpha> r}} P_k(\mathcal{E}(m))$$
$$\le \sum_{\substack{m\in\mathcal{T}\\ \#m-\alpha\le r}} P_k(\mathcal{E}(m)) + \tfrac{K_{\mathrm{ini}}}{r} e^{\kappa_d K_{\mathrm{ini}} t}. \tag{59}$$

In particular, if $r \ge \frac{K_{\mathrm{ini}}}{\delta} e^{\kappa_d K_{\mathrm{ini}} t} + \alpha$, then

$$\sum_{\substack{m\in\mathcal{T}\\ \#m\le r}} P_k(\mathcal{E}(m)) \ge 1-\delta. \tag{60}$$

Recall the ordering of the tree nodes $l\in m$ given by (39) and that $\bar{l}$ denotes the child-node of $l \in m\setminus\mathbb{N}$. Define for each $l\in m$ and each $\Phi\in\mathcal{T}(Y)\cap\mathcal{E}(m)$ the (possibly empty) sets

$$G_l(\Phi,a) = \bigcap_{a<a_0} \Bigg\{ (s,\nu,v)\in[0,s_{\bar{l}}]\times S^{d-1}\times\mathbb{R}^d \;\Bigg|\; \left|\tfrac{v-v_{\bar{l}}}{|v-v_{\bar{l}}|}\cdot\tfrac{v_{l'}-v_{\bar{l}}}{|v_{l'}-v_{\bar{l}}|}\right| \ge \varepsilon(a) \text{ for all } l'<l, l'\neq\bar{l}$$
$$\text{and } v-v_{\bar{l}}\notin R(t,a) \text{ and } (u_l(s,\nu,v),v)\notin \bigcup_{l'<l,l'\neq\bar{l}} C_{l'} \text{ and } |v_l-v_{l'}|\ge\varepsilon\ \forall l'<l \Bigg\},$$

$$B_l(\Phi,a) = [0,s_{\bar{l}}]\times S^{d-1}\times\mathbb{R}^d\setminus G_l(\Phi,a),$$
$$\mathcal{B}_l(a_0,a) = \{\Phi\in\mathcal{T}(Y)\cap\mathcal{E}(m) \mid \phi_l\in B_l(\Phi,a) \text{ and } \phi_{l'}\in G_{l'}(\Phi,a_0) \quad \forall l'<l\}.$$

Note that $\mathcal{B}_1(a_0,a) = B_1(\Phi,a)$. The set $\mathcal{B}_l$ contains those trees which have node l as the first bad node. It is easy to see that $G_l(\Phi,a)$ and $B_l(\Phi,a)$ are monontone in a and that $\mathcal{E}(m) = \left(\hat{\mathcal{G}}(a)\cap\mathcal{E}(m)\right)\dot\cup_{l\in m}\mathcal{B}_l(a,a)$, i.e. the set $\mathcal{E}(m)$ can be written as a disjoint union of good trees and the sets $\mathcal{B}_l$ and we obtain that

$$P_k\left(\hat{\mathcal{G}}(a)\right) = \sum_{m\in\mathcal{T}} \left(P_k(\mathcal{E}(m)) - \sum_{l\in m} P_k(\mathcal{B}_l(a,a)) \right).$$

Another easy consequence is that

$$\sum_{l\in m} P_k(\mathcal{B}_l(a,a)) \le P_k(\mathcal{E}(m)). \tag{61}$$

A simple computation shows that for any sequence $a_0 = a_1 \geq \ldots \geq a_l = a$ the estimate

$$P_k(\mathcal{B}_l(a,a)) \leq \sum_{l' \leq l} P_k(\mathcal{B}_{l'}(a_{l'-1}, a_{l'})),$$

where $l'-1$ is the predecessor of l', and thus

$$1 - P_k\left(\hat{\mathcal{G}}(a)\right) \leq \sum_{\substack{m\in\mathcal{T}\\ \#m\leq r}} \sum_{l\in m, l'\leq l} P_k\left(\mathcal{B}_l(a_{l'-1}, a_{l'})\right) + \delta$$

holds. This inequality is far from being optimal but it suffices for our purposes. We will show that for fixed $a_{l'-1}$ each term in the sum can be made small by choosing $a_{l'}$ small enough. This implies that $\lim_{a\to 0} P_k\left(\hat{\mathcal{G}}(a)\right) = 0$ since we can choose a_{11} first (or a_2 if $\alpha > 1$) depending on a_1 and so on until we reach a_l which serves as an upper bound for a.

We define now the functions $h^m_{l,a_0,a} : \mathcal{E}(m) \to [0,\infty)$ by

$$h^m_{l,a_0,a}(\phi) = e^{-\Gamma(\Phi)}\chi_{\mathcal{B}_l(a_0,a)}(\phi),$$

where $\Gamma(\Phi) = \sum_{|l|<k} \int \mathrm{d}f_0(v)\,|v-v_l|$. Clearly, for fixed $l \in m$, $a_0 > 0$ the number $h^m_{l,a_0,a}(\Phi)$ is nonnegative, monotonously decreasing in a and bounded from above by the function

$$g_m : \mathcal{E}(m) \to [0,\infty): \; g_m(\phi) := e^{-\sum_{j<k}\Gamma_j(\Phi)}.$$

Note that $g_m \in L^1(\lambda^m)$, since $g_m \geq 0$ and $1 = P_k(\mathcal{T}(Y)) = \sum_{m\in\mathcal{T}_k} \int_{\mathcal{E}(m)} \mathrm{d}\lambda^m(\phi)\, g_m(\Phi)$. Thus, it suffices to show that

$$\lim_{a\to 0} \lambda_l\left(B_l(\Phi,a)\right) := N \int_{\mathbb{T}^d} \mathrm{d}u \int_{\mathbb{R}^d} \mathrm{d}v\, \chi_{B_l(\Phi,a)\cap C_{\bar{l}}}(u,v) \tag{62}$$

for all $\Phi \in \mathcal{E}(m) \cap \mathcal{B}_l(a_0,a)$. In order to prove estimate (62) we split $B_l(\Phi,a)$ into five sets, the first four are represented in collision coordinates, the last one is expressed in phase-space coordinates.

$B_{l,1} = \left\{(\nu,\tau,v) \in S^{d-1} \times [0,s_{\bar{l}}] \times \mathbb{R}^d \mid |v - v_{l'}| \leq \varepsilon \text{ for some } l' < l\right\},$

$B_{l,2} = \left\{(\nu,\tau,v) \in S^{d-1} \times [0,s_{\bar{l}}] \times \mathbb{R}^d \;\middle|\; \left|\frac{v-v_{\bar{l}}}{|v-v_{\bar{l}}|}\cdot\frac{v_{l'}-v_{\bar{l}}}{|v_{l'}-v_{\bar{l}}|}\right| \geq 1-\varepsilon \text{ for some } l' < l\right\},$

$B_{l,3} = \left\{(\nu,\tau,v) \in S^{d-1} \times [0,s_{\bar{l}}] \times \mathbb{R}^d \mid v - v_{\bar{l}} \in R(t,a)\right\},$

$B_{l,4} = \left\{(\nu,\tau,v) \in S^{d-1} \times [0,s_{\bar{l}}] \times \mathbb{R}^d \mid |v| > V(a)\right\},$

$$B_{l,5} = \left\{ (u,v) \in \bigcup_{l'<l, l'\neq\bar{l}} C_{l'} \bigcap C_{\bar{l}} \;\middle|\; |v| \leq V(a) \text{ and } \left|\frac{v-v_{\bar{l}}}{|v-v_{\bar{l}}|}\cdot\frac{v_{l'}-v_{\bar{l}}}{|v_{l'}-v_{\bar{l}}|}\right| \leq 1-\varepsilon \text{ for all } l' < l \right\}.$$

We will show now that $\lim_{a\to 0} \lambda_l\left(B_{l,j}(\Phi,a)\right) = 0$ for each $l \in m$, $j \in \{1,\ldots,5\}$ and $\Phi \in \mathcal{B}_l(a_0,a)$ if $a_0 > 0$ is fixed.

$j = 1$**:**

$$\begin{aligned}\lambda_l(B_{l,1}) &= \sum_{\substack{l'<l\\ l'\neq\bar{l}}} \int_{|v-v_{l'}|\leq\varepsilon} \mathrm{d}f_0(v) \int_{S^{d-1}} \mathrm{d}\nu \int_0^{s_{\bar{l}}} \mathrm{d}\tau\, [(v-v_{\bar{l}})\cdot\nu]_+ \\ &\leq \sum_{l'<l} \int_{|v-v_{l'}|\leq\varepsilon} \mathrm{d}f_0(v)\, |v-v_{\bar{l}}|\kappa_d s_{\bar{l}}.\end{aligned}$$

Thanks to (4) the last expression goes to 0 as ε tends to 0. Since a_0 appears nowhere the convergence is uniform in a_0.

$j=2$**:** Let

$$M_\varepsilon = \bigcup_{l'<l, l'\neq \bar{l}} \left\{ v \in \mathbb{R}^d \;\middle|\; \left| \tfrac{v-v_{\bar{l}}}{|v-v_{\bar{l}}|} \cdot \tfrac{v_{l'}-v_{\bar{l}}}{|v_{l'}-v_{\bar{l}}|} \right| > 1-\varepsilon \right\} \subset \mathbb{R}^d.$$

As ε tends to 0 the set M_ε converges to $v_{\bar{l}} + \mathbb{R} \bigcup_{l'<l, l'\neq \bar{l}} (v_{l'} - v_{\bar{l}})$. We obtain that

$$\lambda_l(B_{l,2}) = \int_{M_\varepsilon} \mathrm{d}f_0(v) \int_0^{s_{\bar{l}}} \mathrm{d}\tau \int_{S^{d-1}} \mathrm{d}\nu \, [\nu \cdot (v - v_{\bar{l}})]_+ \leq r \int_{M_\varepsilon} \mathrm{d}f_0(v) \, |v - v_{\bar{l}}| \kappa_d s_{\bar{l}}.$$

Dominated convergence and assumption (4) imply that

$$\lim_{a\to 0} \int_{M_\varepsilon(a)} \mathrm{d}f_0(v) \, (1+|v|)\kappa_d s_{\bar{l}} = 0, \tag{63}$$

hence the last expression goes to 0 as ε tends to 0. The convergence is uniform in a_0.

$j=3$**:**

$$\begin{aligned}\lambda_l(B_{l,3}) =& N \int_{v_{\bar{l}}+R(t)} \mathrm{d}f_0(v) \int_0^{s_{\bar{l}}} \mathrm{d}\tau \int_{S^{d-1}} \mathrm{d}\nu \, [\nu \cdot (v - v_{\bar{l}})]_+ \\ \leq& \int_{v_l + R(t)} \mathrm{d}f_0(v) \, |v - v_{\bar{l}}| \kappa_d s_{\bar{l}} = \int_{R(t)} \mathrm{d}f_0(v) \, |v| \kappa_d s_{\bar{l}}\end{aligned}$$

By equation (57) the last expression converges uniformly in a_0 to 0 as a tends to 0.

$j=4$**:**

$$\begin{aligned}\lambda_l(B_{l,4}) =& N \int_{|v|\geq V(\varepsilon)} \mathrm{d}f_0(v) \int_0^{s_{\bar{l}}} \mathrm{d}\tau \int_{S^{d-1}} \mathrm{d}\nu \, [\nu \cdot (v - v_{\bar{l}})]_+ \\ \leq& \int_{|v|\geq V(\varepsilon)} \mathrm{d}f_0(v) \, |v - v_l| \kappa_d s_{\bar{l}}\end{aligned}$$

By assumption (3) the last expression converges uniformly in a_0 to 0 as ε tends to 0.

$j=5$**:** This is the only case where estimates are not uniform and depend on the constant $\varepsilon(a_0)$. We estimate $\lambda_l(B_{l,5})$ as follows:

$$\lambda\left(B_{l,5}\right) \leq \sum_{\substack{l'<l \\ l'\neq \bar{l}}} N \int_{\mathbb{R}^d} \mathrm{d}f_0(v) \, \mathcal{H}^d\left(C_{\bar{l}}(\Phi) \cap C_{l'}(\Phi) \cap (\mathbb{T}^d \times \{v\})\right).$$

To bound $\mathcal{H}^d\left(C_{\bar{l}}(\Phi) \cap C_{l'}(\Phi) \cap (\mathbb{T}^d \times \{v\})\right)$ we define the number $c(a_0, a, v')$ to be the maximum volume contained within the intersection of two cylinders of diameter a and axes $v - v'$ and $v - v''$ if v, v' and v'' are constrained in a certain geometrical way:

$$\begin{aligned}c(a_0, a, v') = \sup\Big\{ \zeta(u', u'', v, v', v'', a) \;\Big|\; & u', u'' \in \mathbb{T}^d, v, v'' \in \mathbb{R}^d, \; |v' - v''| \geq \varepsilon(a_0) \\ & \text{and } |v|, |v''| \leq V(a) \text{ and } \left| \tfrac{v-v'}{|v-v'|} \cdot \tfrac{v''-v'}{|v''-v'|} \right| \leq 1 - \varepsilon(a_0) \Big\},\end{aligned}$$

where

$$\zeta(u,u',v,v',v'',a) = \mathcal{H}^d\Bigg(\Bigg\{u \in \mathbb{T}^d \;\Bigg|\; \inf_{s\in[0,t]} |u-u'+s(v-v')| \le a \text{ and } \inf_{s\in[0,t]} |u-u''+s(v-v'')|\} \le a\Bigg\}\Bigg).$$

With this notation we obtain that

$$\lambda(B_{l,5}) \le \#m\, N\, c(a_0, a, v_{\bar{l}}).$$

The cylinders can intersect at most $(Vt+1)^2$ times. The volume of each intersection is bounded from above by $(2a)^{d-1}\ell$ where ℓ is the maximal length of a line segment which is parallel to $v-v'$ and is contained in the cylinder with axis parallel to $v''-v'$. A simple geometric consideration yields that $\ell = \frac{2a}{|\sin\psi|}$, where ψ is the angle enclosed by the vectors $v-v'$ and $v-v'$. The law of sines implies that $\sin(\psi) = \frac{|v'-v''|}{|v-v'|}\sin(\psi_0)$, where ψ_0 is the angle enclosed by $v-v''$ and $v'-v''$. Since $\cos(\psi_0) \le 1-\varepsilon$ and $|v'-v''| \ge \varepsilon$ we obtain that $|\sin(\psi)| \ge \frac{1}{|v-v'|}\varepsilon^{\frac{3}{2}}$ and thus the inequality

$$\lambda_l(B_{l,5}) \le \#m\, N\, 2^d\, a^d\, \varepsilon(a_0)^{-\frac{3}{2}} 2V(Vt+1)^2 = 2^d r\, a\, \varepsilon(a_0)^{-\frac{3}{2}} 2V(Vt+1)^2.$$

The right hand side converges to 0 as $a \to 0$ if a_0 is kept fixed. The proof of equation (53) is finished.

Proof of equation (54)**.**

Fix a_0 and let $\Omega \subset \hat{\mathcal{G}}(a_0)$. Like in the first part of the proof we first split off the contribution of the trees with many nodes. For each $r > 0$ one obtains that

$$\begin{aligned}
&\lim_{a\to 0}\sup_k \left|\hat{P}_k(\Omega) - P_k(\Omega)\right| \\
&\le \lim_{a\to 0}\Bigg(\sup_k \sum_{\substack{m\in\mathcal{T}\\ \#m\le r}} \left|\hat{P}_k(\Omega\cap\mathcal{E}(m)) - P_k(\Omega\cap\mathcal{E}(m))\right| \\
&\quad + \sup_k \sum_{\substack{m\in\mathcal{T}\\ \#m> r}} \hat{P}_k\left(\hat{\mathcal{G}}(a)\cap\mathcal{E}(m)\right) + \sup_k \sum_{\substack{m\in\mathcal{T}\\ \#m> r}} P_k\left(\hat{\mathcal{G}}(a)\cap\mathcal{E}(m)\right)\Bigg) \\
&= \lim_{a\to 0}(I_1 + I_2 + I_3).
\end{aligned}$$

We will show that $\lim_{a\to 0} I_1 = 0$ and $\limsup_{a\to 0}(I_2+I_3) = o(1)$ as δ tends to 0 (cf equation (60)).

First we consider I_1. Since there is only a finite number of tree skeletons with fewer than r nodes it suffices to show that

$$\lim_{a\to 0}\sup_k \left|\hat{P}_k(\Omega\cap\mathcal{E}(m)) - P_k(\Omega\cap\mathcal{E}(m))\right| = 0$$

for each $m \in \mathcal{T}$ such that $\#m \le r$. We have seen earlier (formula (51)) that $P_k(\Omega\cap\mathcal{E}(m)) = \hat{P}_k(\Omega\cap\mathcal{E}(m)) + e(\Omega\cap\mathcal{E}(m))$ where

$$0 \le e(\Omega\cap\mathcal{E}(m)) = \int_{\Omega\cap\mathcal{E}(m)} \mathrm{d}\lambda^m(\phi)\left(e^{-\sum_{j<k}\hat{\Gamma}_j(\Phi)} - e^{-\sum_{j<k}\Gamma_j(\Phi)}\right).$$

Since $\Gamma_j(\Phi) \geq \hat{\Gamma}_j(\Phi)$ (cf the remark after (52)) one obtains

$$e(\Omega \cap \mathcal{E}(m)) \leq \int_{\hat{\mathcal{G}}(a)\cap\mathcal{E}(m)} \mathrm{d}\lambda^m(\phi)\, e^{-\sum_{j<k}\Gamma_j(\Phi)} \left(e^{\sum_{j<k}(\Gamma_j(\Phi)-\hat{\Gamma}_j(\Phi))} - 1 \right).$$

We will demonstrate that there is a number $K(a_0, a) > 0$ such that $\lim_{a\to 0} K(a_0, a) = 0$ and for all $\Phi \in \mathcal{E}(m) \cap \hat{\mathcal{G}}(a_0)$ and all $j \in \mathbb{N}$ the estimate

$$0 \leq \Gamma_j(\Phi) - \hat{\Gamma}_j(\Phi) \leq K(a_0, a) \tag{64}$$

holds. Since $\int_{\mathcal{E}(m)} \mathrm{d}\lambda^m(\phi)\, e^{-\sum_{j<k}\Gamma_j(\Phi)} \leq 1$, this yields the bound

$$0 \leq \hat{P}(\Omega \cap \mathcal{E}(m)) - P(\Omega \cap \mathcal{E}(m)) \leq K(a_0, a) e^{K(a_0,a)}. \tag{65}$$

Thus estimate (64) implies $\lim_{a\to 0} I_1 = 0$. To prove (64) we recall that by definition (equation (47))

$$\begin{aligned}
\hat{\Gamma}_j(\Phi) =& N \int_{\mathbb{R}^d} \mathrm{d}f_0(v')\, \mathcal{H}^d \left(\hat{\mathcal{C}}_j(\Phi) \cap (\mathbb{T}^d \times \{v'\}) \right) \\
\geq& N \sum_{|l|=j} \int_{\mathbb{R}^d} \mathrm{d}f_0(v')\, \mathcal{H}^d \left(C_l(\Phi) \cap (\mathbb{T}^d \times \{v'\}) \right) - e_1 \\
=& N \sum_{|l|=j} \int_{\mathbb{R}^d \setminus (v_l + R(t))} \mathrm{d}f_0(v')\, \mathcal{H}^d \left(C_l(\Phi) \cap (\mathbb{T}^d \times \{v'\}) \right) - e_1 + e_2 \\
=& \sum_{|l|=j} \int_{\mathbb{R}^d \setminus (v_l + R(t))} \mathrm{d}f_0(v)\, \kappa_d |v_l - v| - e_1 + e_2 = \Gamma_j(\Phi) - e_1 + e_2 + e_3,
\end{aligned}$$

where the error terms are defined as follows

$$\begin{aligned}
e_1 &= N \int_{\mathbb{R}^d} \mathrm{d}f_0(v')\, \mathcal{H}^d \left((\mathcal{C}_j(\Phi) \setminus \bar{\mathcal{C}}_j(\Phi)) \cap (\mathbb{T}^d \times \{v'\}) \right), \\
e_2 &= N \sum_{|l|=j} \int_{v_l + R(t,a)} \mathrm{d}f_0(v')\, \mathcal{H}^d \left(C_l(\Phi) \cap (\mathbb{T}^d \times \{v'\}) \right), \\
e_3 &= \sum_{|l|=j} \int_{R(t,a)} \mathrm{d}f_0(v)\, \kappa_d |v|.
\end{aligned}$$

We set $K(a_0, a) = -e_1 + e_2 + e_3$ and recycle the estimates from the first part of the proof in order to show that $\lim_{a\to 0} e_j = 0$ for $j = 1, 2, 3$. For all $v' \in \mathbb{R}^d$ one obtains that $N\mathcal{H}^d \left(C_l \cap (\mathbb{T}^d \times \{v'\}) \right) \leq \kappa_d |v_l - v'|$ irrespective whether $v' \in v_l + R(t, a)$ or not. Hence,

$$e_2 + e_3 \leq 2\, \kappa_d\, r \int_{R(t)} |v|\, \mathrm{d}f_0(v)$$

and equation (57) yields that $\lim_{a\to 0}(e_2 + e_3) = 0$.
It remains to estimate e_1. Using the considerations in the case $j = 5$ in the first part of the proof we find that

$$e_1 \leq 2^d\, r^2\, a\, \varepsilon(a_0)^{-\frac{3}{2}} 2V(Vt + 1)^2,$$

and in particular $\lim_{a\to 0} e_1 = 0$. Thus we have shown that $\lim_{a\to 0} K(a_0, a) = 0$ and thereby $\lim_{a\to 0} I_1 = 0$.
We finish the proof by showing that $\lim_{\delta\to 0} \lim_{a_0\to 0} \lim_{a\to 0}(I_2 + I_3) = 0$. Equation (58) yields

$$I_3 = \sup_k \sum_{\substack{m\in\mathcal{T} \\ \#m>r}} P_k(\Omega \cap \mathcal{E}(m)) \leq K_{\mathrm{ini}} \exp(\kappa_d K_{\mathrm{ini}} t)\, \delta. \tag{66}$$

and in a similar way we obtain

$$\lim_{a_0\to 0}\lim_{a\to 0} I_2 = \lim_{a_0\to 0}\lim_{a\to 0}\sup_k \sum_{\substack{m\in\mathcal{T}\\ \#m>r}} \hat{P}_k\left(\Omega\cap\mathcal{E}(m)\right) \le \lim_{a_0\to 0}\lim_{a\to 0}\sup_k \sum_{\substack{m\in\mathcal{T}\\ \#m>r}} \hat{P}_k\left(\hat{\mathcal{G}}(a_0)\cap\mathcal{E}(m)\right)$$

$$= \lim_{a_0\to 0}\lim_{a\to 0}\sup_k \hat{P}_k(\hat{\mathcal{G}}(a_0)) - \lim_{a_0\to 0}\lim_{a\to 0}\inf_k \sum_{\substack{m\in\mathcal{T}\\ \#m\le r}} \hat{P}_k\left(\hat{\mathcal{G}}(a_0)\cap\mathcal{E}(m)\right)$$

$$\overset{(53,65)}{=} 1 - \inf_k \sum_{\substack{m\in\mathcal{T}\\ \#m\le r}} P_k\left(\hat{\mathcal{G}}(a_0)\cap\mathcal{E}(m)\right) \overset{(66)}{\le} \delta\, K_{\text{ini}}\, \exp(\kappa_d K_{\text{ini}} t).$$

Equation (53) yields that the last expression converges to 0 uniformly in a_0 as δ tends to 0.
Thus we have demonstrated that (54) is satisfied and the proof of Proposition 17 is complete. □

Proof of Theorem 2. We first demonstrate that the distribution of a single tagged particle satisfies the Boltzmann equation. Let $A\subset\mathbb{T}^d\times\mathbb{R}^d$ and define $\Omega(A)\subset\mathcal{T}^1(Y)$ by

$$\Omega(A) = \{\Phi\in\mathcal{T}^1(Y)\mid \beta_1(m)=1 \text{ and } z_1\in A\}.$$

With this notation we obtain that for every $a_0>0$

$$\left|\lim_{a\to 0}\lim_{k\to\infty}\hat{P}_{t,k}(\Omega) - \int_A \mathrm{d}u\,\mathrm{d}f_t(v)\right| \overset{\text{Lemma 4}}{=} \lim_{a\to 0}\lim_{k\to\infty}\left|\hat{P}_{t,k}(\Omega) - \int_A \mathrm{d}u\,\mathrm{d}f_{t,k-1}(v)\right|$$

$$\overset{\text{Proposition 13}}{=} \lim_{a\to 0}\lim_{k\to\infty}\left|\hat{P}_{t,k}(\Omega) - P_{t,k}(\Omega)\right|$$

$$= \lim_{a\to 0}\lim_{k\to\infty}\left|\hat{P}_{t,k}(\Omega\cap\mathcal{G}(a_0)) - P_{t,k}(\Omega\cap\mathcal{G}(a_0)) - P_{t,k}(\Omega\setminus\mathcal{G}(a_0)) + \hat{P}_{t,k}(\Omega\setminus\mathcal{G}(a_0))\right|$$

$$\overset{(44)}{\le} \lim_{a\to 0}\lim_{k\to\infty} P_{t,k}(\mathcal{T}(Y)\setminus\mathcal{G}(a_0)) + \lim_{a\to 0}\lim_{k\to\infty}\hat{P}_{t,k}(\mathcal{T}(Y)\setminus\mathcal{G}(a_0))$$

Now using equation (44) again for $\tilde{\Omega} := \mathcal{T}(Y)\cap\mathcal{G}(a_0)$ and that $\hat{P}_{t,k}$ and $P_{t,k}$ are probability measures, we also obtain, that $\lim_{a\to 0}\hat{P}_{t,k}(\mathcal{T}(Y)\setminus\mathcal{G}(a_0)) = P_{t,k}(\mathcal{T}(Y)\setminus\mathcal{G}(a_0))$. Now proceeding

$$\le 2\lim_{k\to\infty} P_{t,k}(\mathcal{T}(Y)\setminus\mathcal{G}(a_0)),$$

we send now a_0 to 0, apply (43) and obtain that $\lim_{a_0\to 0}\lim_{k\to\infty} P_{t,k}(\mathcal{T}(Y)\setminus\mathcal{G}(a_0)) = 0$, hence $\lim_{a\to 0}\lim_{k\to\infty}\hat{P}_{t,k}(\Omega) = \int_A \mathrm{d}u\,\mathrm{d}f_t(v)$.
Next we define the random variables

$$\chi_i(t) = \begin{cases} 1 & \text{if } (u_i^{(a)}(t), v_i^{(a)}(t))\in A \text{ and } \beta_i(t)=1\\ 0 & \text{else.}\end{cases}$$

The previous consideration implies that $\lim_{a\to 0}\langle\chi_i(t)\rangle = \int_A \mathrm{d}u\,\mathrm{d}f_t(v)$. Define now the random variable $s_N = \frac{1}{N}\sum_{i=1}^n \chi_i(t)$. The claim (9) follows if the variance $V_N = \langle (s_N - \langle s_N\rangle)^2\rangle$ converges to 0 as a tends to zero. The standard manipulation yields that

$$V_N \le \frac{\sum_{i=1}^n (\chi_i(t) - \langle\chi_i(t)\rangle)^2}{N^2} + \frac{1}{N^2}\sum_{i\ne j}\langle(\chi_i(t)-\langle\chi_i(t)\rangle)(\chi_j(t)-\langle\chi_j(t)\rangle)\rangle.$$

If we apply the previous reasoning again in the case $\alpha=2$ we find that

$$\lim_{a\to 0}\langle(\chi_i(t)-\langle\chi_i(t)\rangle)(\chi_j(t)-\langle\chi_j(t)\rangle)\rangle = 0$$

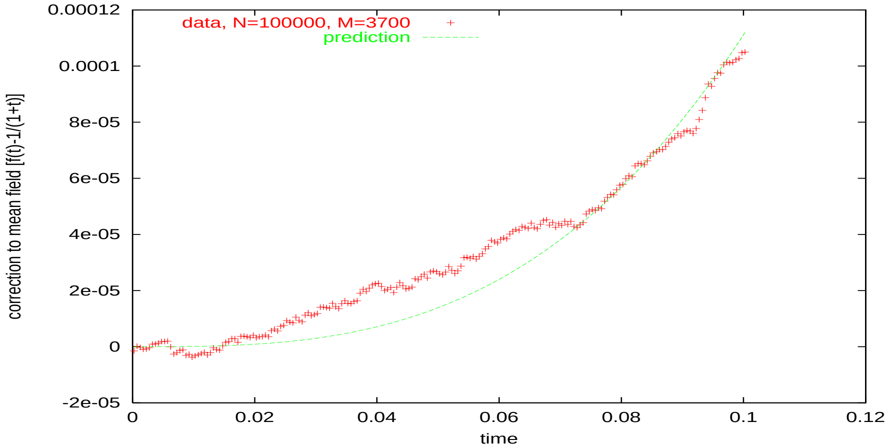

FIGURE 3. Comparison between the empirical probability of colliding and the mean-field prediction. The dashed line is the cubic parabola $t \mapsto \frac{1}{9}t^3$, the signs '+' mark the difference between the number of non-collided particles at time t divided by N and the mean-field prediction $\frac{1}{1+t}$.

uniformly in i and j, and the proof of Theorem 2 is finished. □

Proof of Corollary 3. First we recall a well known principle in probability theory. Let $x_n \in \mathbb{R}$ be a sequence of independent random numbers such that $\mathbb{E}(x_n) = 0$ and let V_n be the variance of x_n. If $\sum_{n=1}^{\infty} V_n < \infty$, then almost surely $\lim_{n\to\infty} x_n = 0$.
Indeed, for every $\varepsilon, N > 0$ Chebyshev's inequality yields the estimate

$$\text{Prob}\left(\sup_{n\geq N} |x_n| \leq \varepsilon\right) \geq \prod_{n=N}^{\infty} \left(1 - \tfrac{V_n}{\varepsilon^2}\right) \geq 1 - \frac{1}{\varepsilon^2}\sum_{n=N}^{\infty} V_n.$$

Consequentially $\lim_{a\to 0} \text{Prob}(\sup_{n\geq N} |x_n| \leq \varepsilon) = 1$, i.e. for each realization and each $\varepsilon > 0$ there exists almost surely a number $N > 0$ such that $\sup_{n\geq N} |x_n| \leq \varepsilon$.
Let s_N be the sum that was defined in the proof of Theorem 2 and V_N be the variance of s_N. Since $\lim_{a\to 0} V_N = 0$ there exists a subsequence V_{N_n} such that $\sum_{n=1}^{\infty} V_{N_n} < \infty$. We apply now the previous consideration to the sequence $x_n = s_{N_n}$. □

3. The effect of concentrations

We illustrate now that the mean field theory does not capture the many-particle dynamics if the initial distribution f_0 exhibits strong concentrations. To simplify the long calculations at the end of the proof we assume that $d = 2$, but similar results are expected to hold in the case $d = 3$.

Theorem 18. *Let $v \in \mathbb{R}^2$ be nonresonant ($\alpha \cdot v \notin \mathbb{Z}$ for all $\alpha \in \mathbb{Z}^d$) such that $|v| = 1$ and set $f_0 = \frac{1}{2}(\delta(\cdot - v) + \delta(\cdot + v))$. If $\hat{Q}(t) = \lim_{a\to 0}\lim_{k\to\infty} \hat{P}_{t,k}(\beta_1 = 1)$ denotes the empirical probability that a tagged particle does not collide, then*

$$\lim_{t\to 0} \frac{1}{t^3}\left(\hat{Q}(t) - \int_{\mathbb{R}^2} \mathrm{d}f_t(v)\right) = \frac{1}{9}, \tag{67}$$

where $f_t = \frac{1}{1+t} f_0$ is the unique solution of the Boltzmann equation (5) which satisfies the initial condition $f_{t=0} = f_0$.

A numerical simulation (fig. 3) confirms the prediction (67).

Proof. It can be assumed without loss of generality that the initial value of the tagged particle is $(0, v)$. We define the set

$$M_\lambda := \left\{ u \in \mathbb{T}^d \;\middle|\; \min_{s\in[0,t]} |2tv - u| \le r \right\},$$

which is basically a cylinder with radius r and centerline given by the particle-trajectory without collisions and contains the initial positions of those particles that might collide with the tagged particle before time t. The parameter r is a function of λ such that $\mathrm{vol}(M_\lambda) = 2at\lambda$.
In this setting the collision rate γ is 1 and we find that the probability that the total number of particles whose initial position is contained in M_λ equals k is given by $e^{-\lambda t}\frac{(\lambda t)^k}{k!}$. Let $p_k(\lambda, t)$ be the probability that the particle does not collide before time t if there are precisely k particles contained in M_λ and there are no particles outside. It is clear from the definition that p_k depends only weakly on t, we will not show the dependency on t in future.
We use the fact that in this system the speed of propagation of information is finite to find explicit approximations for $\hat{Q}$ for short times.

Lemma 19. *Let $n \in \mathbb{N}$ and $\lambda = n + 1$.*

$$\lim_{t\to 0} \frac{1}{t^n} \left| \hat{Q}(t) - e^{-\lambda t} \sum_{k=0}^{n} \tfrac{(\lambda t)^k}{k!} p_k(\lambda) \right| = 0. \tag{68}$$

Proof. Let $\omega = \{u_0(i), \mid i = 1 \dots n\}$ be the set of initial positions and $P_n = \mathrm{Prob}(\#(\omega \cap M_\lambda) > n)$ be the probability that M_λ contains more than n particles. Clearly

$$P_n = e^{-\lambda t} \left(e^{\lambda t} - \sum_{k=0}^{n} \frac{(\lambda t)^k}{k!} \right) \le e^{-\lambda t} t^{n+1} \sup_{s\in[0,t]} \frac{\lambda^{n+1}}{(n+1)!} e^{\lambda s} = \frac{\lambda^{n+1}}{(n+1)!} t^{n+1},$$

where the inequality is due to Taylor's theorem. □

We will only be interested in the case $n = 3$, i.e. $\lambda = 4$.
Let $Q(t) := \frac{1}{1+t}$ be the particle density predicted by the mean-field theory. We are seeking mean-field probabilities $p_k^{\mathrm{mf}}(\lambda) \in [0, 1]$ such that

$$Q(t) = e^{-\lambda t} \sum_{k=0}^{\infty} \tfrac{(\lambda t)^k}{k!} p_k^{\mathrm{mf}}(\lambda). \tag{69}$$

Replacing the exponential function in (69) by the power series one obtains that

$$\sum_{l,m=0}^{\infty} \frac{(-\lambda t)^l}{l!} \frac{(\lambda t)^m}{m!} p_m^{\mathrm{mf}}(\lambda) = \sum_{k=0}^{\infty} (-t)^k. \tag{70}$$

Ordering the left hand side by powers of t and equating coefficients yields the following hierarchical set of equations for the probabilities p_k

$$\sum_{l=0}^{k} \frac{(-1)^l}{l!(k-l)!} p_{k-l}^{\mathrm{mf}} = \left(-\frac{1}{\lambda} \right)^k.$$

We can use the equations above to determine p_k^{mf} recursively and obtain that

$$p_k^{\mathrm{mf}} = (-1)^k \frac{k!}{\lambda^k} - \sum_{l=1}^{k} (-1)^l \binom{k}{l} p_{k-l}^{\mathrm{mf}}.$$

The recurrence relation can be solved explicitly and we obtain

$$p_k^{\mathrm{mf}} = \sum_{l=0}^{k} \frac{k!}{(k-l)!}\left(-\frac{1}{\lambda}\right)^l. \tag{71}$$

Equation (68) and (69) implies that if p_k does not agree with formula (71), then $\hat{Q}(t) \neq Q(t)$ if t is sufficiently small.
If $\lambda > (k+1)^{d-1}$ the probability p_k can be computed explicitly. The reason is that the diameter of the cylinder is so large that the collision probability is not influenced by the initial configuration outside M_λ.
We will show now that for all $\lambda \geq 4$ the values of $p_k(\lambda)$, $k = 0, 1, 2, 3$ are given by $p_0 = 1$, $p_1 = 1 - \frac{1}{\lambda}$, $p_2 = 1 - \frac{2}{\lambda} + \frac{2}{\lambda^2}$, $p_3 = 1 - \frac{3}{\lambda} + \frac{6}{\lambda^2} - \frac{6}{\lambda^3} + \frac{\alpha_2}{\lambda^3}$ with $\alpha_d = \frac{2}{3}$. This implies that

$$\lim_{t \to 0^+} \frac{\hat{Q}(t) - \frac{1}{1+t}}{t^3} = \frac{\alpha_2}{6} = \frac{1}{9}$$

and thus the claim.
Let $k \in \{0, 1, 2, 3\}$ be the number of particles contained in set M_λ. For the sake of simplicity we say that the particles with velocity v are white and the particles with velocity $-v$ are black. One obtains 2^k different color distributions, each of those cases has the same probability of occurring.
We are now in a position to compute an explicit formula for the values of $p_k(\lambda)$. We have to consider several cases, depending on the direction and relative position of the particles in the path of the tagged particle. Particles traveling in the same direction as the tagged particle are denoted by w, particle in the other direction by b. The ordering of the particles in the cylinder is given in the index.

Computation of p_0.
It is clear that $p_0 = 1$ since there is no obstacle in M_λ.

Computation of p_1.
$p_1^w = 1$,
$p_1^b = 1 - \frac{2}{\lambda}$.
We obtain the overall probability $p_1 = \frac{1}{2}(p_1^w + p_1^b) = 1 - \frac{1}{\lambda}$.

Computation of p_2.
$p_2^{ww} = 1$ (No collision possible),
$p_2^{bb} = (1 - \frac{2}{\lambda})^2$ (Probability of avoiding two independent black particles),
$p_2^{bw} = 1 - \frac{2}{\lambda}$ (Probability of avoiding one black particle, the position of the white particle is irrelevant),
$p_2^{wb} = 1 - \frac{2}{\lambda}(1 - \frac{2}{\lambda})$ (Probability of avoiding a black particle which might be removed by a white particle before it comes to a collision.
Adding the probabilities yields that $p_2 = \frac{1}{4}(p_2^{ww} + p_2^{bb} + p_2^{wb} + p_2^{bw}) = 1 - \frac{2}{\lambda} + \frac{2}{\lambda^2}$.

Computation of p_3.
$p_3^{www} = 1$ (No collision possible),
$p_3^{bbb} = (1 - \frac{2}{\lambda})^3$ (Probability of avoiding 3 independent black particles),
$p_3^{bww} = 1 - \frac{2}{\lambda}$ (Probability of avoiding 1 black particle, the white particles are irrelevant),
$p_3^{wbw} = 1 - \frac{2}{\lambda}(1 - \frac{2}{\lambda})$ (Probability of avoiding one black particle which might be removed by one white particle. The second white particle is irrelevant).
$p_3^{wwb} = 1 - \frac{2}{\lambda}(1 - \frac{2}{\lambda})^2$ (Probability of avoiding one black particle which might be removed by two independent white particles).

$p_3^{bbw} = (1-\frac{2}{\lambda})^2$ (Probability of avoiding 2 independent black particles, the white particle is irrelevant)
$p_3^{bwb} = (1-\frac{2}{\lambda})(1-\frac{2}{\lambda}(1-\frac{2}{\lambda}))$ (Probability of avoiding 2 independent black particles, the second black particle might be removed by a white particle).
$p_3^{wbb} = 1-\frac{4}{\lambda}+\frac{12}{\lambda^2}-\frac{24}{\lambda^3}+8\frac{\alpha_2}{\lambda^3}$
To demonstrate that the formula above indeed yields the correct value of p_3^{wbb} we introduce the coordinates perpendicular to v of the three particles $u_i \in \mathbb{R}$, $i=1,2,3$ and consider four mutually exclusive scenarios. In three scenarios the probability of being scattered can be computed analogously to the preceding cases. As these computations are independent of a, we let $a=1$ for notational convenience.

Prob$(|u_2| \geq 1$ and $|u_3| \geq 1) = (1-\frac{2}{\lambda})^2$,
Prob$(|u_2| \geq 3$ and $|u_3| \leq 1$ and $|u_1-u_3| \leq 1) = (1-\frac{6}{\lambda})\frac{4}{\lambda^2}$,
Prob$(|u_2| \leq 1$ and $|u_2-u_1| \leq 1$ and $|u_3| \geq 1) = \frac{4}{\lambda^2}(1-\frac{2}{\lambda})$,
To compute the probability of being scattered in the remaining case where $|u_2| \in [1,3]$, $|u_3| \leq 1$, $|u_1-u_3| \leq 1$ and $|u_1-u_2| \geq 1$ we have to do an explicit integration.

$$\begin{aligned} I_2 = &\int_{-1}^{1} \mathrm{d}u_3 \int_{u_3-1}^{u_3+1} \mathrm{d}u_1 \int_{1}^{3} \mathrm{d}u_2 \,(1-\chi_{[-1,+1]}(u_1-u_2)) \\ &+ \int_{-1}^{1} \mathrm{d}u_3 \int_{u_3-1}^{u_3+1} \mathrm{d}u_1 \int_{-3}^{-1} \mathrm{d}u_2 \,(1-\chi_{[-1,+1]}(u_1-u_2)). \end{aligned}$$

The number I_3 is defined by a similar formula. A simple but lengthy calculation yields that $I_2 = \frac{40}{3}$. The details of this calculation are irrelevant, but for the purpose of checking that this number is indeed correct the detailed calculations are included below. We obtain that

$$p_3^{bww} = (1-\frac{6}{\lambda})(1-\frac{2}{\lambda}) + (1-\frac{6}{\lambda})\frac{4}{\lambda^2} + \frac{4}{\lambda^2}(1-\frac{2}{\lambda}) + \frac{4}{\lambda}(1-\frac{2}{\lambda}) + \frac{I_2}{\lambda^3}.$$

Altogether this yields

$$\begin{aligned} p_3 =& \tfrac{1}{8}(p_3^{www}+p_3^{wwb}+p_3^{wbw}+p_3^{bww}+p_3^{wbb}+p_3^{bwb}+p_3^{bbw}+p_3^{bbb}) \\ =& 1-\tfrac{3}{\lambda}+\tfrac{6}{\lambda^2}-\tfrac{6}{\lambda^3}+\tfrac{I_2-8}{8\lambda^3}, \end{aligned}$$

and therefore $\alpha_2 = \frac{I_2-8}{8} = \frac{2}{3}$.

We calculate now the value of I_2.

$$\begin{aligned}
I_2 = \;& \int_{-1}^1 \mathrm{d}u_3 \int_{u_3-1}^0 \mathrm{d}u_1 \int_1^3 \mathrm{d}u_2 \underbrace{(1-\chi_{[-1,1]}(u_1-u_2))}_{=1} \\
&+ \int_{-1}^1 \mathrm{d}u_3 \int_0^{u_3+1} \mathrm{d}u_1 \int_1^3 \mathrm{d}u_2\,(1-\chi_{[-1,1]}(u_1-u_2)) \\
&+ \int_{-1}^1 \mathrm{d}u_3 \int_{u_3-1}^0 \mathrm{d}u_1 \int_{-3}^{-1} \mathrm{d}u_2\,(1-\chi_{[-1,1]}(u_1-u_2)) \\
&+ \int_{-1}^1 \mathrm{d}u_3 \int_0^{u_3+1} \mathrm{d}u_1 \int_{-3}^{-1} \mathrm{d}u_2 \underbrace{(1-\chi_{[-1,1]}(u_1-u_2))}_{=1} \\
= \;& \underbrace{2\int_{-1}^1 \mathrm{d}u_3\,(1-u_3) + 2\int_{-1}^1 \mathrm{d}u_3\,(1-u_3)}_{=8} \\
&+ \int_{-1}^1 \mathrm{d}u_3 \int_0^{u_3+1} \mathrm{d}u_1 \int_1^3 \mathrm{d}u_2\,(1-\chi_{[-1,1]}(u_1-u_2)) \\
&+ \int_{-1}^1 \mathrm{d}u_3 \int_{u_3-1}^0 \mathrm{d}u_1 \int_{-3}^{-1} \mathrm{d}u_2\,(1-\chi_{[-1,1]}(u_1-u_2)) \\
= \;& 8 + \underbrace{\int_{-1}^1 \mathrm{d}u_3 \int_0^{u_3+1} \mathrm{d}u_1 \int_1^3 \mathrm{d}u_2 + \int_{-1}^1 \mathrm{d}u_3 \int_{u_3-1}^0 \mathrm{d}u_1 \int_{-3}^{-1} \mathrm{d}u_2}_{=8} \\
&- \int_{-1}^1 \mathrm{d}u_3 \int_0^{u_3+1} \mathrm{d}u_1 \int_1^3 \mathrm{d}u_2\,\chi_{[-1,1]}(u_1-u_2) - \int_{-1}^1 \mathrm{d}u_3 \int_0^{1-u_3} \mathrm{d}u_1 \int_1^3 \mathrm{d}u_2\,\chi_{[-1,1]}(u_1-u_2) \\
= \;& 16 - \int_{-1}^1 \mathrm{d}u_3 \int_0^{u_3+1} \mathrm{d}u_1 \int_1^{1+u_1} \mathrm{d}u_2 - \int_{-1}^1 \mathrm{d}u_3 \int_0^{1-u_3} \mathrm{d}u_1 \int_1^{1+u_1} \mathrm{d}u_2 \\
= \;& 16 - \int_{-1}^1 \mathrm{d}u_3 \int_0^{u_3+1} \mathrm{d}u_1\, u_1 - \int_{-1}^1 \mathrm{d}u_3 \int_0^{1-u_3} u_1 \\
= \;& 16 - \int_{-1}^1 \mathrm{d}u_3 \tfrac{1}{2}(u_3+1)^2 - \int_{-1}^1 \mathrm{d}u_3\, \tfrac{1}{2}(u_3-1)^2 \\
= \;& 16 - \tfrac{1}{6}[(u_3+1)^3]_{u_3=-1}^{u_3=1} - \tfrac{1}{6}[(u_3-1)^3]_{u_3=-1}^{u_3=+1} = 16 - \tfrac{8}{3}.
\end{aligned}$$

□

4. Proofs of auxiliary results

This section contains the proofs of Lemmas 4, 6 and 16. These lemmas are not concerned with multi-scale aspects.

We first explain the notation used in Lemma 4. Let $w \in C(\mathbb{R}^d)$, $w \geq 0$ be a weight. For a Radon-measure f we define

$$\|f\|_w := \sup_{\phi \in BC^0(\mathbb{R}^d), \|\phi\| \leq 1} \int |\phi(v) w(v)\, \mathrm{d}f(v)|.$$

Then $M_w = \{f \in (BC^0(\mathbb{R}^d)))^* |\ \|f_0\|_w < \infty\}$ is a Banach space of measures with norm $\|.\|_w$. To control convergence we introduce weighted spaces in time for X-valued functions, for some Banach space X

$$C^0_\rho([0,\infty), X) := \{u \in C^0([0,\infty), X) |\ \sup_{t\in[0,\infty)} (\exp(-\rho t)\|u(t)\|_X < \infty\})\ \text{ with norm}$$
$$\|u\|_\rho := \sup_{t \in [0,\infty)} (\exp(-\rho t)\|u(t)\|_X).$$

Proof of Lemma 4. First, we note that $\|f_{t,k}\|_{(1+|v|)^2}$ is decreasing in t as $0 \leq L[f_{s,k-1}](v) < \infty$. Next we estimate $\exp(-\rho t)\|f_{t,k+1} - f_{t,k}\|_{1+|v|}$ for $0 \leq t < \infty$, with ρ chosen later. Let $\phi \in BC^0(\mathbb{R}^d)$ with $\|\phi\| \leq 1$, then consider

$$\begin{aligned}
&\exp(-\rho t)\left|\int_{\mathbb{R}^d} \phi(v)(1+|v|)\,(\mathrm{d}f_{t,k+1}(v) - \mathrm{d}f_{t,k}(v))\right| \\
&= \int_{\mathbb{R}^d} \phi(v)(1+|v|)\,\mathrm{d}f_0(v)\exp(-\rho t)\left|\exp\left(-\int_0^t L[f_{s,k}](v)\,\mathrm{d}s\right) - \exp\left(-\int_0^t L[f_{s,k-1}](v)\,\mathrm{d}s\right)\right| \\
&\leq \int_{\mathbb{R}^d} \phi(v)(1+|v|)\,\mathrm{d}f_0(v)\exp(-\rho t)\int_0^t |L[f_{s,k}](v) - L[f_{s,k-1}](v)|\ \mathrm{d}s).
\end{aligned}$$

Because of the negativity of L, we obtain a Lipschitz constant of 1 for $\exp(\cdot)$ and we have

$$\leq \int_{\mathbb{R}^d} \phi(v)(1+|v|)\,\mathrm{d}f_0(v)\kappa_d \left(\exp(-\rho t)\int_0^t \int_{\mathbb{R}^d} |\mathrm{d}f_{s,k}(v') - \mathrm{d}f_{s,k-1}(v')|\,|v-v'|\mathrm{d}s\right)$$

$$\leq \int_{\mathbb{R}^d} \phi(v)(1+|v|)\,\mathrm{d}f_0(v)\kappa_d \Big(\int_0^t \exp(-\rho(t-s))\Big[\exp(-\rho s)\|f_{s,k}-f_{s,k-1}\|_{1+|v|}$$
$$+\exp(-\rho s)|v|\|f_{s,k}-f_{s,k-1}\|_1\Big]\mathrm{d}s\Big)$$

$$\leq 2\kappa_d \int_{\mathbb{R}^d} \phi(v)(1+|v|)^2\,\mathrm{d}f_0(v) \sup_{0\leq s<\infty}\big(\exp(-\rho s)\|f_{s,k}-f_{s,k-1}\|_{1+|v|}\big)\int_0^t \exp(-\rho(t-s))\,\mathrm{d}s$$

$$\leq 2\kappa_d \|f_0\|_{(1+|v|)^2}\frac{1}{\rho}(1-\exp(-\rho t))\|f_{t,k}(\cdot)-f_{t,k-1}(\cdot)\|_\rho.$$

Thus for $\rho > 2\kappa_d\|f_0\|_{(1+|v|)^2}$ the sequence $(f_k)_{k\in\mathbb{N}}$ converges in $C^0_\rho([0,\infty), M_{1+|v|})$ by Banach's fixed point theorem and the limit solves $f_t = \exp(-\int_0^t L[f_s](v)\,\mathrm{d}s)f_0$. Hence f is differentiable and solves (5) for $t\in[0,\infty)$. Uniqueness of the solution of the integral equation also follows by the Banach fixed point theorem. On the other hand all solutions of (5) in $C^1([0,T], M_{1+|v|})$ have to satisfy the integrated form too, showing uniqueness of the solutions of (5). □

Proof of Lemma 6. We first show, that the implicit relation $\beta(i,t)$ in Theorem 2 is well-defined. For each particle it indicates whether it has undergone a collision: $\beta(i,t)$ jumps from 1 to 0 at the time of the collision. As the particles are removed after a collision, a collision can only occur when

$$\mathrm{dist}(z_i, z_{i'}, s) = a \text{ for some } i \neq i'.$$

This also takes multiple collisions into account, which lead to an undefined situation in hard-sphere collision dynamics, but as particles are removed here after a collision, the scattering state can be defined.

The distance $\mathrm{dist}(z_i, z_{i'}, s)$ is a continuous piece-wise affine function in s, except possibly a unique point, if there is an initial intersection, but then $\mathrm{dist}(z_i, z_{i'}, s) > a$ near this jump. There are only finitely many different pieces in a finite interval $[0,t]$, because $v(i)-v(i')$ is finite and only a finite number of coverings of the torus $\mathbb{T}^d$ can be visited in a finite time. Hence for every particle i, there are at most $n-1$ possible collision times, i.e. the first time $\tau(i,i')\geq 0$ at which $\mathrm{dist}(z_i, z_{i'}, s) = a$ for each i'. The at most $n(n-1)/2$ possible times for collision of the particles $i = 1,\ldots,n$ can be well-ordered. So by inductively checking at all possible collision times $\tau(i,i')$, there exists a well-defined collision time for each particle i, at which it collides with an unscattered particle ($\beta(i,.)$ has a well-defined jump); or the particle remains unscattered itself for $[0,\infty)$ ($\beta(i)$ is constant), which shows the existence of $\beta(i,t)$.

To prove convergence of $\beta_k(i,t)$ to $\beta(i,k)$ as k tends to ∞ we define $I_1 = \{1\ldots n\}$ and $\tau_0 = 0$. For each $j\geq 1$ let $\tau_j > \tau_{j-1}$ and $C_j, I_{j+1}\subset I_j$ be recursively defined by

$$\begin{aligned}
&\min\{\mathrm{dist}(z_i, z_{i'}, \tau_j)\,|\, i\neq i'\in C_j\} = a \text{ for each } i\in C_j\\
&\mathrm{dist}(z_i, z_{i'}, s) > a \text{ for all } i, i'\in I_j, s\in[\tau_{j-1},\tau_j),\\
&I_{j+1} = I_j\setminus C_j.
\end{aligned}$$

It can be checked that $\beta(i,s) = 1$ if there exists $j\in\mathbb{N}$ such that $i\in I_j$ and $s\in[0,\tau_j]$. For all other choices of i and s we have that $\beta(i,s)=0$. Clearly $\beta(i,\cdot)$ is constant within the intervals $(\tau_{j-1},\tau_j]$. We will show using induction that for each $j\in\{1,2,\ldots\}$ and each

$k \geq j$

(72) $$\beta_k(i,s) = \beta(i,s) \text{ if } s < \tau_j \text{ or } i \in I_1 \setminus I_j.$$

The claim is clear for $j = 1$. Assume now that the claim has been established up to j and let $k \geq j+1$. We will show that

(73) $$\beta_k(i,s) = \beta(i,s) \text{ if } s < \tau_{j+1} \text{ or } i \in I_1 \setminus I_{j+1}.$$

By the induction assumption (73) holds for $s \in [0, \tau_j]$ or $i \in I_1 \setminus I_j$ and we can assume from now that $s > \tau_j$.
Case 1. Let $i \in I_1 \setminus I_{j+1}$.
We have to show that

(74) $$\beta_k(i,s) = \beta(i,s) \text{ for all } k \geq j+1.$$

Since $s > \tau_j$ we have that $\beta(i,s) = 0$. By (72) equation (74) holds if $i \in I_1 \setminus I_j$, hence we can assume that $i \in C_j$. In this case there exists $i' \in C_j$ such that $\text{dist}(z_i, z_{i'}, \tau_j) = a$. The induction assumption (72) implies that $\beta_{k-1}(i', \tau_{j-1}) = 1$ and consequentially

$$\text{dist}(z_i, z_{i'}, \tau_j) = a\beta_{k-1}(i', \tau_j),$$

this implies that $\beta_k(i,s) = 0$.
Case 2: Let $i \in I_{j+1}$ and $s \in (\tau_j, \tau_{j+1})$. We have to show that

(75) $$\beta_k(i,s) = \beta(i,s) = 1 \text{ for all } k \geq j+1.$$

Case 2a. Let $s' \in (\tau_{j-1}, \tau_j)$ and i' be such that $\beta_{k-1}(i', s') = 1$. The induction assumption implies that $\beta_{k-1}(i', s') = \beta(i', s')$ and therefore $i' \in I_j$. Since $i \in I_{j+1}$ and $i' \in I_j$ it is not possible that $\text{dist}(z_i, z_{i'}, \tau_j) = a$.
Case 2b. Let i' be such that $\beta_{k-1}(i', s') = 0$. In this case we find that

$$\text{dist}(z_i, z_{i'}, s') \geq a\beta_{k-1}(i', s') = 0.$$

Cases 2a and 2b together imply that $\beta_k(i,s) = 1$.
Since the number of particles is finite there exists a number K such that $\tau_j = +\infty$ for all $j \geq K$, hence equation (18) is a consequence of (72).

□

Proof of Lemma 16. To simplify the notation we define $\hat{\mathcal{C}} = \mathcal{C} \setminus \bar{\mathcal{C}}$. The assumption $A \subset \cup_{r=0}^{\infty} (\mathcal{C} \setminus \bar{\mathcal{C}})^r$ implies

$$\text{Prob}_{\text{ppp}}\left(\mathcal{C} \cap \omega \in A \mid \bar{\mathcal{C}} \cap \omega = \emptyset\right) = \text{Prob}_{\text{ppp}}\left(\hat{\mathcal{C}} \cap \omega \in A \mid \bar{\mathcal{C}} \cap \omega = \emptyset\right).$$

After a simple rearrangement of the last expression above one obtains

$$\begin{aligned}
&\text{Prob}_{\text{ppp}}\left(\mathcal{C} \cap \omega \in A \mid \bar{\mathcal{C}} \cap \omega = \emptyset\right) \\
=&\text{Prob}_{\text{ppp}}\left(\hat{\mathcal{C}} \cap \omega \in A \text{ and } \bar{\mathcal{C}} \cap \omega = \emptyset\right) \text{Prob}_{\text{ppp}}^{-1}\left(\bar{\mathcal{C}} \cap \omega = \emptyset\right) \\
=&\text{Prob}_{\text{ppp}}\left(\hat{\mathcal{C}} \cap \omega \in A\right)
\end{aligned}$$

since the sets $\hat{\mathcal{C}}$ and $\bar{\mathcal{C}}$ are disjoint. The last expression is an unconditional probability with respect to the Poisson-point process which can be evaluated explicitly using Definition 1:

$$\text{Prob}_{\text{ppp}}\left(\hat{\mathcal{C}} \cap \omega \in A\right) = \sum_{r=0}^{\infty} \text{Prob}_{\text{ppp}}\left(\hat{\mathcal{C}} \cap \omega \in A \cap \mathcal{C}^r\right)$$

$$= \sum_{r=0}^{\infty} e^{-\mu(\hat{\mathcal{C}})} \frac{(\mu(\hat{\mathcal{C}}))^r}{r!} \times (\mu(\hat{\mathcal{C}}))^{-r} \int_{\hat{\mathcal{C}}^r} \mathrm{d}\mu^r(z) \chi_A(z) = e^{-\mu(\hat{\mathcal{C}})} \sum_{r=0}^{\infty} \frac{1}{r!} \int_{A \cap \mathcal{C}^r} \mathrm{d}\mu^r(z).$$

This proves formula (46). □

Appendix A. Notation

Symbol	Meaning
a	diameter of the balls
n	number of particles
N	a^{1-d} intensity
(u,v)	phase space variables in $\mathbb{T}^d \times \mathbb{R}^d$
f_0	initial distribution, element of $PM(\mathbb{R}^d)$
$\mathrm{Prob}_{\mathrm{ppp}}$	probability of the Poisson-point process of the initial data of (6)
f_k	approximated solution of (2) as defined in (15)
$\mathcal{H}^d$	d-dimensional Hausdorff measure
$M_+(\mathbb{T}^d \times \mathbb{R}^d)$	non-negative measures on $\mathbb{T}^d \times \mathbb{R}^d$
$M_w(\mathbb{R}^d)$	measures with weight function w
$\beta^{(a)}(i,t)$	scattering state ($=1$ unscattered, $=0$ scattered) of particle i at time t
$\beta_k^{(a)}(i,t)$	scattering state when restricting to tree of height k
$\mathcal{T}$	$\subset \cup_{i=1}^{\infty} \mathbb{N}^i$ set of tree skeltons
m	$\in \mathcal{T}$ tree (skeleton)
$\#m$	$\in \mathbb{N} \cup \{0\}$ size of m
l	$\in m$ a node in a tree
$\bar{l}$	the child of node l
$\|l\|$	height of a node ($= i$ if $l \in \mathbb{N}^i$)
(u_l, v_l, s_l, ν_l)	$\in \mathbb{T}^d \times \mathbb{R}^d \times [0,\infty) \times S^{d-1}$ data on node l with u_l, v_l initial data
	ν_l collision parameter and s_l collision time
$\mathcal{T}(Y)$	trees with collision data
$\mathcal{E}(m)$	$\subset \mathcal{T}(Y)$ trees with skeleton m
$\Phi = (m, \phi)$	$\in \mathcal{T}(Y)$ tree (with collision data)
$P_{t,k}$	mean field probability, defined in (20)
$\underline{P}_{t,1}$	distribution of root, defined in (37)
$\mathrm{d}\bar{\lambda}_l$	simplified mean field distribution at node l, defined in (24)
$\hat{P}_{t,k}$	empirical distribution, defined in (33)
$R(t,a)$	$\subset \mathbb{R}^d$ resonant initial velocities
$\mathcal{G}(a)$	$\subset \mathcal{T}(Y)$ good trees (Def. 15)
$\hat{\mathcal{G}}(a_0)$	$\subset \mathcal{G}(a_0)$ good trees with additional desirable properties
γ_l	collision rate of particle l (mean-field)
$\Gamma(j)$	joint collision rate of particle of height j (mean-field), (21)
$\hat{\Gamma}(j)$	joint collision rate of particle of height j (empiric), (47)
C_l	colliding initial values of particle at node l, defined in definition 15
$\mathcal{C}(k)$	$:= \bigcup_{l \in m \cap \mathbb{N}^k} C_l \subset \mathbb{T}^d \times \mathbb{R}^d$
$\bar{\mathcal{C}}(k)$	$:= \bigcup_{\|l\|<k} C_l \subset \mathbb{T}^d \times \mathbb{R}^d$
$\hat{\mathcal{C}}(k)$	$:= \mathcal{C}(k) \setminus \bar{\mathcal{C}}(k)$

References

[BBS83] C. Boldrighini, L.A. Bunimovich, Y.G. Sinai. On the Boltzmann equation for the Lorentz gas.*J. Stat. Phys.* **32** 477-501.

[CIP94] C. Cercignani, R. Illner, M. Pulvirenti. *The Mathematical Theory of Dilute Gases.* Applied Mathematical Sciences, Vol **106**, Springer Verlag (1994).

[DL90] R. DiPerna & P.L. Lions. On the Cauchy Problem for Boltzmann Equations: Global Existence and Weak Stability, *Ann. Math.* **130** (1989) 321-366.

[Dur] R. Durrett, *Probability: Theory and Examples.* 3^{rd} ed. Duxbury (2004).

[Gal70] G. Galavotti. Rigorous theory of Boltzmann equation in the Lorentz gas, preprint *Nota interna* **358**, Univ. di Roma (1970).

[Hil00] D. Hilbert, Mathematical problems. Reprinted from Bull. Amer. Math. Soc. 8 (1902), 437–479. *Bull. Amer. Math. Soc. (N.S.)* **37** (2000), 407–436.

[Lan75] O. Lanford. *Time evolution of large classical systems.* ed. by J. Moser pp 1-113. Lecture notes in physics, Vol **38**, Springer Verlag (1975).

[MT07a] K. Matthies, F. Theil. Rigorous justification of the long-time validity of the Boltzmann equation: The gainless, heterogenous case. *In preparation* (2007).

[MT07b] K. Matthies, F. Theil. Rigorous justification of the long-time validity of the Boltzmann equation: The homogeneous case. *In preparation* (2007).

[MT07c] D. Marenduzzo & F. Theil. Non-validity of mean-field theories due to concentration phenomena. *In preparation* (2007).

[Spo78] H. Spohn. The Lorentz process converges to a random flight process. *Comm. Math. Phys.* **60** 277-290 (1978).

[Spo91] H. Spohn. *Large scale dynamics of interacting particles.* Texts and Monographs in Physics, Springer Verlag (1991).

Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, United Kingdom
E-mail address: K.Matthies@maths.bath.ac.uk

Mathematics Institute, University of Warwick, Coventry CV4 7AL, United Kingdom
E-mail address: f.theil@warwick.ac.uk