Kac's process and some probabilistic aspects of the Boltzmann equation

Daniel Heydecker

Department of Pure Mathematics and Mathematical Statistics
University of Cambridge
Corpus Christi College

6. July 2021

This dissertation is submitted for
the degree of Doctor of Philosophy

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the introduction (Section 1.4). This is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree, diploma or other qualification at the University of Cambridge or any other university or a similar institution except as declared in the introduction and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the introduction and specified in the text.

Specifications relevant to this declaration are made in Section 1.4.

Kac's Process and some probabilistic aspects of the Boltzmann Equation

D. Heydecker

Abstract

We consider a family of stochastic interacting particle systems introduced by Kac as a model for a spatially homogeneous gas undergoing elastic collisions, corresponding to the spatially homogeneous Boltzmann equation. We consider Kac's problem of showing *propagation of chaos* - that if the velocities of the particles are initially approximately independent, then the same is true at later times - which is equivalent to the convergence of the empirical measures, and which derives the spatially homogeneous Boltzmann equation from the underlying molecular dynamics.

The first two results concern the propagation of chaos for different Kac models. In the first case, we consider the hard spheres kernel, which is appropriate for modelling interactions arising from localised interactions. For this model, we build on previous analyses of the same problem to show that the expected deviation between the Kac process and the Boltzmann equation, measured in expected Wasserstein distance, is of the order $N^{-\alpha}$. Particular care is paid to the time-dependence of the estimates, using the stability properties of the Boltzmann equation: we will show that the expected deviation at a single fixed time is typically of the order $N^{-\alpha}$, uniformly in time, while for the largest deviation on a time interval $[0, t_{\text{fin}}]$, there is a prefactor $(1 + t_{\text{fin}})^{\beta}$, where $\beta > 0$ can be chosen as close to 0 as desired. We also show that similar estimates hold, possibly for smaller α, as soon as the initial data are only assumed to have $2 + \epsilon$ moments.

We next consider the case of non-cutoff hard potentials, which arise from modelling a family of long-range interactions. In this case, the collision kernel is doubly-unbounded, both unbounded as the relative velocity increases, and with a non-integrable angular singularity, so that every particle undergoes infinitely many collisions on any nontrivial time-interval. In this context, we introduce a Tanaka-style coupling for a well-chosen distance function on $\mathbb{R}^d$, which allows us to exploit a negative Povzner-type term. This leads to a proof of propagation of chaos and a new uniqueness and stability result for the corresponding Boltzmann equation, assuming only that the initial measure is a probability measure with finitely many moments, whereas previously established results have required either some additional regularity or an exponential moment. In the case of the corresponding hard-potential Landau equation, the same argument can be further refined

to show uniqueness and stability for probability measures with $2+\epsilon$ moments. We also show that energy-conserving solutions to the Landau equation exist, assuming only that the initial data has finite energy, and we use the new uniqueness result to show that all solutions to the Landau equation immediately admit analytic densities, aside from the degenerate cases where the initial data is a point mass. By contrast, previous results have only shown that such regular solutions exist.

We further study the dynamical large deviations of the N-particle system, either for the case of a cutoff Maxwell molecules kernel, or a caricature of the hard spheres kernel, for a range of initial conditions including equilibrium. We seek large deviation estimates jointly with an auxiliary *flux measure* which records the collision history. We prove a large deviation upper bound, and a lower bound restricted to classes of sufficiently regular paths, with a rate function analagous to those found elsewhere in the literature for similar problems. However, we show by exhibiting a family of counterexamples that this rate function does not capture all possible large deviation behaviours; although the particle system almost surely conserves energy, possible large deviation behaviour includes energy non-conserving solutions to the Boltzmann equation, as found by Lu and Wennberg, but these occur strictly more rarely than predicted by the rate function.

The final section concerns Smoluchowski and Flory coagulation equations with a particular bilinear form, which arise when studying the interaction structure of the Kac process by forming clusters of particles joined by chains of collisions, corresponding to the *cumulant expansion*. Exploiting the bilinear structure and a coupling to random graphs, we are able to give a detailed analysis of the coagulation particle system and limiting Flory equation, including showing the emergence of a unique macroscopic cluster at a finite time $t_{\mathrm{g}} \in (0,\infty)$, which we characterise exactly in the case of the particle system.

To my parents, Deirdre and Benjamin.

Acknowledgements

First and foremost, I would like to thank my supervisor, Professor James Norris. This work would not have been possible without his encouragement, support and many helpful discussions during all the years of my study.

I would also like to thank Dr. Robert Patterson (WIAS) and and Professor Nicolas Fournier (LPSM) for the privilege of working with them. I would also like to thank Dr. Patterson for his encouragement in the study of dynamical large deviations.

I would like to thank the many people at Cambridge who have made my studies - from undergraduate until the present - such an enriching and inspiring experience. Thanks are due to the academic staff of Queens' College, particularly to my director of studies Professor Julia Gog, and to Professor Richard Nickl, whose encouragement sparked my interest in mathematical probability theory. As a graduate, I benefited from the academic and practical support of Corpus Christi College, where I thrived in the stimulating environment of Leckhampton.

This research was made possible by the Cambridge Centre for Analysis (CCA) doctoral training programme, funded by the EPSRC, from which I learned a lot and greatly benefited. I would like to thank the course directors and administrator, as well as my CCA peers.

Finally, thank you to my friends and family for your unwavering support. Thank you to Ben Morely, Jonny Tsang, Jeremy Budd, James Sinclair and Alex Hardwick for all of our stimulating conversations, mathematical and otherwise. My special thanks go to Sidsel Størmer, Gerri Tait and, of course, my parents, to whom this thesis is dedicated.

Contents

Chapter 1

Introduction

1.1 The Boltzmann Equation

The mathematical study of a dilute gas dates back to Maxwell, who modelled a monatomic gas as consisting of a large number N of indistinguishable particles. Each particle has a position x_i and a velocity v_i in a phase space $(x_i, v_i) \in D\times\mathbb{R}^d$, for some spatial domain D in d dimensions, and the particles interact with each other via elastic collisions on a spatial scale $\epsilon_N \ll 1$, chosen so that the binary interactions (involving exactly two particles) contribute nontrivially to the evolution, but that collisions simultaneously involving three or more particles contribute only negligibly, which forces the choice $\varepsilon_N \sim N^{-1/(d-1)}$. Under these conditions, he proposed to study the evolution of a function $F = F(t,x,v) = F_t(x,v) \geq 0$ on $[0,\infty)\times D\times\mathbb{R}^d$ describing the macroscopic density of particles at time t with position x and velocity v. Under these hypotheses, Maxwell proposed an evolution equation, which was subsequently refined by Boltzmann [30, 29] and now bears his name

$$\partial_t F_t + v\cdot\nabla_x F = Q(F,F)(t,x,v) \tag{spBE}$$

where the left-hand side describes the effect of free transport[1], and the right-hand side is the collision operator, acting only on the velocity variables by

$$Q(F,F)(t,v,x) = \int_{\mathbb{R}^d} dv_* \int_{\mathbb{S}^{d-1}} d\sigma\, B(v-v_*,\sigma)(F(t,x,v')F(t,x,v'_*) - F(t,x,v)F(t,x,v_*)).$$

Here, the additional parameter $\sigma\in\mathbb{S}^{d-1}$ parametrises all possible elastic collisions through the direction of separation, since the incoming velocities and the conservation of energy and momentum alone are insufficient to determine the outgoing velocities. With this additional parameter the post-collisional velocities are

$$v' = \frac{v+v_*}{2} + \sigma\frac{|v-v_*|}{2}; \qquad v'_* = \frac{v+v_*}{2} - \sigma\frac{|v-v_*|}{2}. \tag{1.1}$$

[1] The prefix 'sp' refers to spatial inhomogeneity, to distinguish it from the spatially homogeneous Boltzmann equation (BE), which will be our main object of study.

One can check that such collisions preserve the (microscopic) kinetic energy $|v'|^2+|v'_*|^2 = |v|^2+|v_*|^2$ and momentum $v'+v'_* = v+v_*$, and that all such collisions can be written in this way. The factor B appearing in Q is the (effective) cross-section, which depends on the 'microscopic' interactions between the particles. From the Gallilean (translational and rotational) symmetry of the microscopic dynamics, the kernel is usually taken to depend only on $|v-v_*|$ and the *scattering angle* $\theta = \cos^{-1}(\sigma\cdot(v-v_*)/|v-v_*|)$. In particular, $B(v-v_*,\sigma) = B(v'-v'_*, \frac{v-v_*}{|v-v_*|})$, encoding the time-reversibility of the microscopic dynamics. Some natural choices for B arise from modelling particles as hard spheres with an exclusion radius ε_N, at which the particles deflect each other, or from modelling long-range repulsive forces on the same microscopic scale ε_N.

The derivation of (spBE) from the microscopic particle dynamics relies on Boltzmann's *Stoßzahlansatz*, or molecular chaos. For the many-body system, the k-particle marginal $F^{N,k}$ distribution functions satisfy

$$\partial_t F^{N,1}_t + v\cdot\nabla_x F^{N,1}_t = Q^N_2(F^{N,2}_t)$$

where Q^N_2 plays the same rôle as Q, but now integrating the two-particle marginal $F^{N,2}$. While this equation closely resembles the desired mean-field equation (spBE), it does not close, since the evolution of $F^{N,1}$ depends on $F^{N,2}$. In turn, the evolution of $F^{N,2}$ depends on $F^{N,3}$, and in general the evolution of the k-particle marginal depends on the $k+1$-marginal, leading to an infinite hierarchy (BBGKY hierarchy) of differential equations. If one imagines that the positions and velocities of the particles are perfectly independent, then one can eliminate $F^{N,2}_t$ in favour of $F^{N,1}_t$, in which case the equation closes and we find the Boltzmann equation. However, particles cannot be perfectly independent: in the case of hard sphere repulsion, there is the constraint that the positions are separated by at least ε_N, and even if particles are initially sampled independently, conditional to this constraint, the interactions will immediately create dependencies and destroy the independence. Boltzmann instead proposed the *chaoticity property* that

$$F^N_2(t,x,v,x',v') \sim F^N_1(t,x,v)F^N_1(t,x',v'), \qquad N\to\infty$$

for positions and velocities $(x,v),(x',v')$ leading to a collision at scale ε_N; this is then a *low-correlation* assumption on the pre-collisional data. This property certainly holds for natural choices of initial data - for instance, drawing $(x_i(0),v_i(0))$ independently from $F_0(x,v)dxdv$, conditional on $|x_i(0)-x_j(0)|>\varepsilon_N$ for all $i\neq j$ - and the problem is then to show the *propagation of chaos*, so that it remains true at future times. More than one and a half centuries after Boltzmann first proposed his equation (spBE), the propagation of chaos for the limit $N\to\infty, \varepsilon_N\sim N^{-1/(d-1)}\to 0$ (the Boltzmann-Grad limit) remains a significant open problem. Lanford [129, 130] proved that the empirical measures μ^N_t associated to the many-particle system converge to a solution to the Boltzmann equation, at least on a very short time interval $[0,T^*]$ for some $T^*>0$ depending on F_0, typically no larger than one fifth of the mean free time [130], which is required to make a certain 'tree expansion' obtained from the BBGKY hierarchy converge. More recently, a number of works by Bodineau, Saint-Raymond et al. [23, 24] have further studied the error on an even smaller time interval, necessary for a different expansion to converge, and examined the fluctuations in the sense of the central limit theorem and large deviations, and a work by Pulvirenti and Simonella [164] shows that groups of up to N^α particles are asymptotically independent as $N \to \infty$, for $t < T^*$ and for some $\alpha \in (0,1)$.

1.1.1 Kac's Process and the Spatially Homogeneous Boltzmann Equation

In this thesis, we will not consider the spatially inhomogeneous Boltzmann equation, and will restrict ourselves to the less ambitious aim of deriving and studying its spatially inhomogeneous counterpart. As a softer alternative to deriving the Boltzmann equation from a spatially inhomogeneous particle model, Kac [122] proposed to derive the spatially homogeneous Boltzmann equation from a Markov jump process, in which the additional stochasticity of collisions compensates for the lack of a spatial parameter. In this context, we study N-tuples $\mathcal{V}^N_t = (V^1_t, V^2_t, \dots, V^N_t) \in (\mathbb{R}^d)^N$ where, at a rate $2B(V^i_t - V^j_t, \sigma)d\sigma/N$, the velocities V^i_t, V^j_t are updated according to (1.1) to form a new vector $\mathcal{V}^N_{i,j,\sigma}$, so that $\mathcal{V}^N_t$ is a Markov process. In this way, the collisions have a greater degree of independence from each other and from the current state, and the problem of deriving the propagation of chaos is easier. In this case, we can define the collision operator Q for probability measures μ on $\mathbb{R}^d$ by specifying the duality $f : \mathbb{R}^d \to \mathbb{R}$,

$$\langle f, Q(\mu,\nu)\rangle := \int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}} (f(v') + f(v'_*) - f(v) - f(v_*))B(v - v_*, \sigma)\mu(dv)\nu(dv_*)d\sigma.$$

Let us remark that this will make sense throughout either as a signed measure or as a distribution as soon as both μ, ν have bounded second moments; see Remark 1.1 below. We will also write $Q(\mu) := Q(\mu,\mu)$ throughout to simplify notation when the two arguments coincide. We thus reach the (spatially homogeneous) Boltzmann equation, given for measure-valued processes $(\mu_t)_{t\geq 0}$ by

$$\partial_t \mu_t = Q(\mu_t). \tag{BE}$$

For the Kac process, there are two possible ways in which we can derive the Boltzmann equation. As in the discussion of the spatially inhomogeneous case above, one would like the particles to be independent, in which case one would immediately find that the one-particle marginal $\mathrm{Law}(V^1_t) = \mathrm{Law}(V^i_t) = \mu_t$ solves the Boltzmann equation; but as in the inhomogeneous case, there is no hope for perfect independence to be propagated in time. Kac [122] introduces the notion of chaoticity as an *asymptotic independence* property in order to play this same role: he defined a sequence of symmetric probability measures on $(\mathbb{R}^d)^N$ to be μ-chaotic if the k-particle marginals converge weakly to $\mu^{\otimes k}$ as $N \to \infty$ for any fixed k. He proposed to show that, if this property holds for μ_0 at time 0, then it remains true at later times (propagation of chaos). Alternatively, one can prove convergence to μ_t of the empirical measures μ^N_t associated to $\mathcal{V}^N_t$ by

$$\mu^N_t = \theta_N(\mathcal{V}^N_t) = \frac{1}{N}\sum_{i=1}^N \delta_{V^i_t}.$$

Due to the interchangeability of the particles, the empirical measures μ^N_t are themselves a Markov chain, and we call these processes $\mathcal{V}^N_t, \mu^N_t$ *labelled* and *unlabelled* Kac processes respectively, since the map θ_N simply forgets the (unphysical) label $i = 1, 2, \dots, N$ on each particle. Typically, our methods will respect the interchangeability symmetry of the particles and are therefore naturally phrased at the level of the empirical measures μ^N_t, but it will sometimes be helpful to work instead with the labelled processes instead.

1.1.2 Collision Kernels

The mathematical challenges faced in investigating (BE) depend strongly on the choice of collision kernel $B(v,\sigma)$, which encodes the physics of the microscopic interactions. Typically, we will assume that the kernel is of the factorised form

$$B(v,\sigma) = \Psi(|v|)b(\cos\theta)$$

for some $\Psi : (0,\infty) \to [0,\infty)$ and a convex function $b : (-1,1) \to [0,\infty)$; while this restriction could perhaps be relaxed, this would lead to significant additional technicality. We will also often make the transformation $b(\cos\theta) \leftrightarrow \beta(\theta)$ given by

$$\beta(\theta) := b(\cos\theta)\frac{(\sin\theta)^{d-2}}{c_d}; \qquad c_d = \int_0^\pi (\sin\theta)^{d-2}d\theta. \tag{1.2}$$

In this way, specifying b is completely equivalent to specifying β; the additional factor is a Jacobian factor (see Section 2.4) which allows us to move easily between integrals against σ and those against θ via the change of variables

$$\int_{\mathbb{S}^{d-1}} h(\theta)B(u,\sigma)d\sigma = \Psi(|u|)\int_0^\pi h(\theta)\beta(\theta)d\theta. \tag{1.3}$$

We now discuss some possible kernels.

1. Maxwell Molecules The simplest kernel we work with is Grad's kernel, given by

$$\Psi = 1, \qquad b = \frac{1}{2^{d-2}(\sin\theta/2)^{d-2}}. \tag{GMM}$$

Although not physically realistic, this kernel is useful as a toy model of (cutoff) Maxwell molecules, where $b(\cos\theta)$ is integrable and $\Psi = 1$, so that the interaction rate is independent of the relative velocity but may be singular in θ. In dimension $d = 3$, the kernel for 'true' Maxwell molecules is of the form

$$\Psi = 1, \qquad b(\cos\theta) \in L^\infty_{\mathrm{loc}}((0,\pi]), \qquad b(\cos\theta) \sim \theta^{-3/2} \text{ as } \theta \downarrow 0. \tag{tMM}$$

2. Hard Spheres When modelling short-range interactions such as the hard-core interaction described in the inhomogeneous case, where spheres reflect when reaching a distance ε_N, we work with the so-called hard spheres kernel $B(v, \sigma) = |v|$, which fits the factorisation above with

$$\Psi(r) = r, \qquad b = 1. \tag{HS}$$

It will sometimes be useful to have a caricature of this kernel which remains bounded away from 0 when the relative velocity is small:

$$\Psi(r) = 1 + r, \qquad b = \frac{1}{2^{d-2}(\sin\theta/2)^{d-2}}. \tag{rHS}$$

The 'r' here stands for 'regularised', since this choice makes the map $v \mapsto \log \Psi(|v|)$ globally Lipschitz continuous. In both this and Grad's kernel, the angular factor is chosen so that the kernel is a function only of the relative velocity when rewritten in the 'ω-representation', see Section 2.4.

3. Hard Potentials with a Moderate Angular Singularity A more general class of kernels, which combine some of the phenomenology of hard spheres and noncutoff Maxwell molecules, are the noncutoff hard potentials:

$$\begin{cases} \Psi(r) = r^\gamma, & & \gamma \in (0, 1]; \\ \beta(\theta) \in L^\infty_{\text{loc}}((0, \pi]); & \beta(\theta) \sim \theta^{-1-\nu} \text{ as } \theta \downarrow 0, & \nu \in (0, 1). \end{cases} \tag{NCHP}$$

This then has polynomial growth of Ψ at infinity, similar to (HS), as well as an angular singularity at $\theta = 0$, similar to (tMM). The second line can be expressed in terms of b through (1.2). Thanks to the symmetry of collisions, we may assume further that b is supported on $[0, 1)$ and correspondingly that β is supported on $[0, \pi/2]$, since collisions with $\cos\theta < 0$ can be achieved from these collisions by relabelling the particles; see the discussion in Alexandre et al. [8]. As in the case of (tMM) above, let us remark that this is a *moderate* angular singularity in the sense that

$$\int_{\mathbb{S}^{d-2}} B(v, \sigma) d\sigma = \infty; \qquad \int_{\mathbb{S}^{d-2}} \theta B(v, \sigma) d\sigma < \infty.$$

4. Soft Potentials Another form of kernel, which has very different phenomenology from those above, is the case of *soft potentials*, where the kernel has an angular singularity and diverges at small relative velocities instead of large:

$$\begin{cases} \Psi(r) = r^\gamma, & \gamma \in (-d, 0); \\ \beta \in L^\infty_{\text{loc}}((0, \pi]) & \beta(\theta) \sim \theta^{-1-\nu} \text{ as } \theta \downarrow 0, \qquad \nu \in (0, 2). \end{cases} \tag{SP}$$

5. Quadratic Kernels The final family of kernels with which we work are another caricature of cutoff hard spheres, which will be useful when we investigate cluster expansions in Chapter 7. In this case, only the total rate of collision between each pair of particles matters, and we require that this total rate is of the form[2]

$$\int_{\mathbb{S}^{d-1}} B(v,\sigma)d\sigma = a + b|v|^2, \qquad a, b \geq 0 \tag{$\mathrm{Q}_{a,b}$}$$

While this has fairly limited physical relevance, this particular form has some nice algebraic properties allowing some exact calculations, leading to an interesting class of models whose properties can be studied exactly. These kernels interpolate between Maxwellian behaviour when $a > 0, b = 0$ and a sort of 'very hard' behaviour when $a = 0, b > 0$.

Grad's Angular Cutoff In three of the cases (tMM, NCHP, SP) above, the angular part of the kernel is not integrable, reflecting the abundance of *grazing collisions*, that is, collisions with very small θ. In either of the cases (tMM, NCHP), we have

$$\int_{\mathbb{S}^{d-1}} B(v,\sigma)d\sigma = \infty; \qquad \int_{\mathbb{S}^{d-1}} \theta B(v,\sigma)d\sigma < \infty \tag{1.4}$$

and the same for soft potentials (SP) if $\nu \in (0,1)$, whereas with $\nu \in [1,2)$ we have

$$\int_{\mathbb{S}^{d-1}} \theta B(v,\sigma)d\sigma = \infty; \qquad \int_{\mathbb{S}^{d-1}} \theta^2 B(v,\sigma)d\sigma < \infty. \tag{1.5}$$

In all cases, one possible strategy for dealing with the angular singularity is to truncate the collisions at small scattering angles $\theta < \theta_0(K)$, with $\theta_0(K)$ chosen so that $\int_{\mathbb{S}^{d-2}} B_K(u,\sigma)d\sigma = K$ for any unit vector u, called *Grad's angular cutoff*. Since the additional grazing collisions only make very small differences to the particle velocities, one might hope that this preserves, in some meaningful sense, the physics of the system under consideration. Indeed, in the case of (NCHP), we will prove in Chapter 4 that both the particle system and the limit equation are the limits in the weak topology of their cutoff counterparts, uniformly in the particle number in the case of the particle system. On the other hand, other properties which are *not* continuous for the weak topology behave differently, for instance, regularity [8] or the appearance of exponential moments [84].

Let us introduce some notation for this cutoff. In the case (NCHP) we obtain a *family* of such kernels, depending on the truncation parameter K, which we write as

$$B_K(v,\sigma) = B(v,\sigma)\mathbb{1}\{\theta \geq \theta_0(K)\} \tag{CHP_K}$$

where as above θ_0 is defined implicitly by $\int_{\mathbb{S}^{d-1}} B_K(v,\theta)d\sigma = K|v|^\gamma$. We write Q_K for the corresponding collision operator. We will write objects in this way for any $K \in [1,\infty]$, understanding that $K = \infty$ is the original, non-cutoff case.

[2]The meaning of b in this case has nothing to do with the meaning of $b(\cos\theta)$ in the previous cases.

Physical Relevance of Kernels Let us remark on the physics of some of the families of the kernels introduced above. In $d = 3$, if we start with a spatial model in which particles interact at a microscopic scale through a repulsive potential $\mathbb{V}(r) = r^{-s}, s > 0$ on a scale $\varepsilon_N \sim N^{-1/(d-1)}$, then one finds [44, Section II.4], [189, Section 1.4]

$$\gamma = \frac{s-5}{s-1}; \qquad \nu = \frac{2}{s-1}.$$

In this way, we obtain cases of soft potentials (SP) for $s \in (2, 5)$, the edge case Maxwell Molecules (tMM) for $s = 5$, and cases of (noncutoff) hard potentials (NCHP) for $s \in (5, \infty)$. Taking $s \to \infty$, we recover the parameters $\gamma = 1, \nu = 0$ corresponding to (HS), which arises from the spatial model from 'hard core repulsion' forbidding particles from approaching closer than ε_N. The case of quadratic kernels ($Q_{a,b}$) is unphysical, but resembles other toy model with interesting properties [190].

Remark 1.1 (Collision Operator as a Signed Measure or Distribution)**.** *Let us remark on the collision operator Q in light of these kernels. In the cases (GMM, HS, CHP_K), the definition written above makes sense as a convergent integral as soon as f is a bounded function and μ, ν have finite second moments, so that $Q(\mu, \nu)$ defines a signed measure. In the cases (tMM, NCHP), we observe that $|v' - v| \le |v - v_*| \sin\theta$ to get, for Lipschitz f,*

$$|f(v') + f(v'_*) - f(v) - f(v_*)| \le C_f |v - v_*| \sin\theta$$

and this is again integrable as soon as μ, ν have second moments; in this case, the definition above makes $Q(\mu, \nu)$ into a distribution in (say) the negative Sobolev space $W^{-1,\infty}(\mathbb{R}^d)$. In the cases where $\nu \in (1, 2)$, one has to use a 'cancellation of order 2' to see that

$$|f(v') + f(v'_*) - f(v) - f(v_*)| \le C_f |v - v_*|^2 \sin^2\theta$$

so that one is now restricted to f with bounded second derivatives and $Q(\mu, \nu) \in W^{-2,\infty}(\mathbb{R}^d)$. Finally, we cannot allow the edge case $\nu = 2$, as in this case even $\int \sin^2\theta B(u, \sigma) d\sigma = \infty$ diverges; in the physical description above, this corresponds to the case $s = 2$ of Coulomb interaction, see Villani [185] and Alexandre [9].

Let us also remark that the divergence of $\Psi(r) = r^\gamma, \gamma < 0$ causes further problems in the definition of solutions in the case of soft potentials, since $\int_{\mathbb{R}^d \times \mathbb{R}^d} |v - v_*|^\gamma \mu(dv)\mu(dv_*)$ can diverge if μ contains point masses. One further splits into moderately soft ($\gamma \ge -2$) or very soft ($\gamma < -2$). In this thesis, we will only be interested in the other cases (Maxwell molecules, hard spheres and cutoff/noncutoff hard potentials, quadratic kernels), and mention the soft potentials only for contextualisation.

1.1.3 The Landau Equation

The second kind of equation we will study is the spatially homogeneous Landau equation, also called the Landau-Fokker-Planck Equation, which writes

$$\partial_t f_t(v) = \frac{1}{2}\mathrm{div}_v\Big(\int_{\mathbb{R}^d} a(v-v_*)[f_t(v_*)\nabla f_t(v) - f_t(v)\nabla f_t(v_*)]\, dv_*\Big) \tag{LE}$$

where a is the nonnegative, symmetric matrix

$$a(x) = |x|^{2+\gamma}\Pi_{x^\perp}; \quad \Pi_{x^\perp} = I - \frac{xx^*}{|x|^2}.$$

Again, we think of f_t as describing the probability distribution of velocities, which imposes the conditions $f_t \geq 0, \int_{\mathbb{R}^d} f_t dv = 1$. This equation arises naturally [50] for the Coloumb interactions as the limit of Boltzmann equations with b replaced by $b_\varepsilon = (\log\varepsilon)^{-1}\mathbb{1}_{\theta\geq\varepsilon} b$ where b is of the form (SP) with parameter $\nu = 2$ and $\gamma = -3$; the divergent factor $(\log\varepsilon)$ in the denominator ensures that $\int_{\mathbb{S}^{d-1}} \theta^2 B_\varepsilon(v,\sigma)d\sigma$ is bounded, uniformly in ε, and that the contributions from any interval $[\theta_0, \pi), \theta_0 > 0$ vanish in the limit. Moreover, the same equation makes sense for any parameter value $\gamma \in [-3, 1]$, corresponding to $s \in [2, \infty]$; in the cases $\gamma \in (-3, 1]$, this equation is not physically relevant, but shares many of the same phenomena for the corresponding Boltzmann equation. Indeed, in these cases the Landau equation is the *grazing collision limit* of Boltzmann equations, for B_ε having the same exponent γ and b_ε chosen so that

$$\sup_{\theta\geq\theta_0} b_\varepsilon \to 0 \text{ for all } \theta_0 > 0; \qquad \lambda_\varepsilon = \int_{\mathbb{S}^{d-1}} \theta^2 B_\varepsilon(u,\sigma)d\sigma \to \lambda_0 \in (0,\infty). \tag{1.6}$$

See, for example, Arsen'ev and Buryak [13], Desvillettes [52] and Villani [186] for a detailed derivation. We give notice now that, although this equation makes sense in any dimension $d \geq 2$, our main results concerning the Landau equation require $d = 3$, which is the physical dimension in any case, and we will restrict to this case throughout. We use the same terminology as for the Boltzmann equation, and refer to Maxwell Molecules for $\gamma = 0$, hard and soft potentials for $\gamma \in (0, 1], \gamma \in (-3, 0)$ respectively and the Coulomb potential for $\gamma = -3$.

1.1.4 Measure Spaces and Weak Solutions

We next give a definition of weak solutions for the kinetic equations with which we work. Typically, since we will work with the empirical measures of the particle system, the sorts of techniques we will use will not use smoothness or any form of regularity at all. Correspondingly, we do not need any such assumptions on the limiting equation, and so it is natural to consider weak (measure-valued) solutions to (BE, LE).

We begin with some spaces of measures and notation which are in frequent use. We write

$\mathcal{P}(E)$ throughout for the space of Borel probability measures on a topological space E, and for $E = \mathbb{R}^d$, $\mathcal{P}_p(\mathbb{R}^d)$ for those measures where the p^{th} moment is finite. We write $\Lambda_p(\mu) \in [0, \infty]$ for the p^{th} moment of a probability measure $\mu \in \mathcal{P}(\mathbb{R}^d)$, whether or not it is finite, and $\Lambda_p(\mu, \nu) := \max(\Lambda_p(\mu), \Lambda_p(\nu))$. With this notation, we say a family of probability measures $(\mu_t)_{t\geq 0}$ on $\mathbb{R}^d$ belongs to $L^\infty_{\text{loc}}([0,\infty), \mathcal{P}_p(\mathbb{R}^3))$ for $p > 0$ if

$$\sup_{t\leq t_{\text{fin}}} \Lambda_p(\mu_t) < \infty \qquad \text{for all } t_{\text{fin}} > 0$$

and that it belongs to $L^1_{\text{loc}}([0,\infty), \mathcal{P}_p(\mathbb{R}^3))$ if

$$\int_0^{t_{\text{fin}}} \Lambda_p(\mu_t) dt < \infty \quad \text{for all } t_{\text{fin}} > 0.$$

We now give the definitions of weak solutions suitable for our purposes.

Definition 1.1.1 (Weak Solution to the Boltzmann Equation)**.** *Let us fix one of the kernels (GMM, tMM,HS, rHS, NCHP, NCHP, $Q_{a,b}$). We say that a family of probability measures $(\mu_t)_{t\geq 0}$ is a weak solution to (BE) if it belongs to $L^\infty_{loc}([0,\infty), \mathcal{P}_2(\mathbb{R}^d))$ and, for all $f : \mathbb{R}^d \to \mathbb{R}$ bounded and Lipschitz,*

$$\langle f, \mu_t \rangle = \langle f, \mu_0 \rangle + \int_0^t \langle f, Q(\mu_s) \rangle ds$$

for any $t > 0$.

Let us remark that, following the argument sketched in Remark 1.1, the integrand $\langle f, Q(\mu_s)\rangle$ makes sense, since we assume that μ_s has at least two moments. Further, in any of these cases, we find a uniform bound on $\langle f, Q(\mu_s)\rangle$ depending only on the boundedness or Lipschitz constant of f and on the second moment of μ_s, so the integral is well-defined by the assumption that $\mu \in L^\infty_{\text{loc}}([0,\infty), \mathcal{P}_2)$.

We note that *formally*, we have the conservation properties

$$\langle (1, v, |v|^2), Q(\mu) \rangle = 0$$

corresponding to the conservation of energy and momentum at the level of individual collisions, and so solutions to the Boltzmann equation (at least formally) also conserve the integrals of these properties. We will therefore usually normalise to 0 momentum and unit temperature: we define the Boltzmann sphere

$$\mathcal{S} = \mathcal{S}(\mathbb{R}^d) = \left\{ \mu \in \mathcal{P}_2 : \langle v, \mu \rangle = 0, \langle |v|^2, \mu \rangle = 1 \right\}.$$

We write $\mathcal{S}^p$ for $\mathcal{S} \cap \mathcal{P}_p(\mathbb{R}^d)$, and understand $L^1_{\text{loc}}([0,\infty), \mathcal{S}^p), L^\infty_{\text{loc}}([0,\infty), \mathcal{S}^p)$ for the same spaces as above, now additionally requiring that $\mu_t \in \mathcal{S}$ for all $t \geq 0$. By translation and scaling, all energy-conserving solutions to the Boltzmann equation can be captured in this way, aside from the (uninteresting) degenerate case of point masses $\mu_t = \delta_{v_0}$, which are always stationary solutions.

Regarding the Landau equation, we restrict to $d = 3$ as above, and we use the following definition of weak solutions, due to Villani [186] and Goudon [101], which makes sense for the cases at least as hard as Maxwell molecules. For $x \in \mathbb{R}^3$, we recall the definition $a(x) = |x|^{2+\gamma}\Pi_{x^\perp}$ above, and define

$$b(x) = \operatorname{div} a(x) = -2|x|^\gamma x. \tag{1.7}$$

Definition 1.1.2 (Weak Solutions to the Landau Equation)**.** *Let $\gamma \in [0,1]$. We say that $(\mu_t)_{t\ge 0}$ is a weak solution to (LE) if it belongs to $L^\infty_{loc}([0,\infty), \mathcal{P}_2(\mathbb{R}^3)) \cap L^1_{loc}([0,\infty), \mathcal{P}_{2+\gamma}(\mathbb{R}^3))$, if $\Lambda_2(\mu_t) \le \Lambda_2(\mu_0)$ for all $t \ge 0$, and if for all $f \in C^2_b(\mathbb{R}^3)$, all $t \ge 0$,*

$$\langle f, \mu_t\rangle = \langle f, \mu_0\rangle + \int_0^t \langle f, Q_L(\mu_s)\rangle \tag{1.8}$$

where $Q_L(\mu)$ is the distribution given by

$$\langle f, Q_L(\mu)\rangle := \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \mathcal{L}_{\mathrm{L}} f(v, v_*)\mu(dv_*)\mu(dv)ds; \tag{1.9}$$

$$\mathcal{L}_{\mathrm{L}} f(v, v_*) = \frac{1}{2}\sum_{k,\ell=1}^3 a_{k\ell}(v-v_*)\partial^2_{k\ell} f(v) + \sum_{k=1}^3 b_k(v-v_*)\partial_k f(v). \tag{1.10}$$

Since $|\mathcal{L}_{\mathrm{L}} f(v, v_*)| \le C_f(1 + |v| + |v_*|)^{2+\gamma}$ for $f \in C^2_b(\mathbb{R}^3)$, every term makes sense in (1.8) under our assumptions; in the case of Maxwell molecules $\gamma = 0$, the assumption that $\mu \in L^1_{\mathrm{loc}}([0,\infty), \mathcal{P}^{2+\gamma}(\mathbb{R}^3))$ is superfluous, as this is already implied by the other condition, while for soft potentials we would require some extra regularity, taking us back to function-valued solutions. As before, we have the formal conservation properties

$$\langle (1, v, |v|^2), Q_L(\mu)\rangle = 0$$

and the additional assumption that the energy $\Lambda_2(\mu_t)$ is at most its initial value $\Lambda_2(\mu_0)$ is enough to guarantee that such solutions conserve energy, see [58, Theorem 3]; as in the Boltzmann case, we will (almost always) normalise so that solutions take values $\mu_t \in \mathcal{S}$ for all t, in which case the requirements can be succinctly written $\mu \in L^1_{\mathrm{loc}}([0,\infty), \mathcal{S}^{2+\gamma})$. Let us remark on the slight difference that we allow solutions to (BE) which increase energy, whereas this is forbidden for (LE).

1.1.5 Some Formalities of the Kac Process

With the functional spaces introduced above, we can now give a formal definition of the Kac processes described above. We consider fixed, forever, a complete filtered probability space $(\Omega, \mathfrak{F}, (\mathfrak{F}_t)_{t\geq 0}, \mathbb{P})$. The same conservation rules as above apply pathwise almost surely to the Kac dynamics, so that $\mathcal{V}^N_t, \mu^N_t$ almost surely preserve the momentum and energy. We can therefore either allow the (labelled) Kac process to take values in

$$\mathbb{S}_N = \left\{ \mathcal{V} = (v_1, ..., v_N) \in (\mathbb{R}^d)^N : \sum_{i=1}^N v_i = 0, \sum_{i=1}^N |v_i|^2 = N \right\} \tag{1.11}$$

which is the preimage of the Boltzmann sphere $\mathcal{S}$ under the map θ_N associating any N-tuple to its empirical measure, or without normalisation, we allow the state space to be the whole space $(\mathbb{R}^d)^N$. With either state space, and for any choice of the kernels above aside from (SP), we define the generator of the unlabelled process by specifying, for bounded and Lipschitz F,

$$(\mathcal{G}^{\mathrm{L},N}F)(\mathcal{V}^N) = \frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N \int_{\mathbb{S}^{d-1}} \left(F(\mathcal{V}^N_{i,j,\sigma}) - F(\mathcal{V}^N)\right) B(V^i - V^j, \sigma)d\sigma. \tag{1.12}$$

Following Remark 1.1 above, we could allow only bounded F for any of the cutoff cases (GMM, HS, CHP_K), but the restriction to Lipschitz test functions is necessary for the integral to converge in the cases (tMM, NCHP). In the unlabelled case, we obtain a state space of $\mathcal{P}^2_N$ or $\mathcal{S}_N$, of empirical measures on N points (respectively: measures in $\mathcal{S}$ which are empirical measures on N points), and define a generator by specifying, for F Lipschitz with respect to the Wasserstein distance[3] $\mathcal{W}_{1,1}$

$$(\mathcal{G}^{\mathrm{U},N}F)(\mu^N) = N\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}} (F(\mu^{N,v,v_\star,\sigma}) - F(\mu^N))B(v - v_\star, \sigma)\mu^N(dv)\mu^N(dv_\star)d\sigma \tag{1.13}$$

where $\mu^{N,v,v_*,\sigma}$ is the post-collisional measure resulting from a collision between v, v_* with scattering angle σ:

$$\mu^{N,v,v_*,\sigma} = \mu^N + \frac{1}{N}\left(\delta_{v'} + \delta_{v'_*} - \delta_v - \delta_{v_*}\right). \tag{1.14}$$

We include the superscripts L, U on the generators for the labelled and unlabelled cases respectively. We remark that there is a consistency between the generators in the sense that, for any test function F on empirical measures,

$$(\mathcal{G}^{\mathrm{L},N}(F\circ\theta_N))(\mathcal{V}^N) = (\mathcal{G}^{\mathrm{U},N}F)(\theta_N(\mathcal{V}^N))$$

which implies that, if $\mathcal{V}^N_t$ is a labelled Kac process, its empirical measure $\mu^N_t = \theta_N(\mathcal{V}^N_t)$ is an unlabelled Kac process. The converse is also true, in that for any unlabelled Kac process (μ^N_t) defined through this generator, there is a 'lift' $\mathcal{V}^N_t$, which is a labelled Kac process and $\mathrm{Law}((\theta_N(\mathcal{V}^N_t)_{t\geq 0}) = \mathrm{Law}((\mu^N_t)_{t\geq 0})$. In all of the cases we consider except for the noncutoff hard potentials, this is trivially checked using the finiteness of the rates

[3] see Section 2.1 for a definition.

on subsets with bounded energy; we check this assertion for noncutoff hard potentials (NCHP) in the Appendix to Chapter 4. In any case, using this consistency, we can move between labelled and unlabelled Kac processes with no ambiguity, depending on which is more convenient for any given application. We will also write $(\mathfrak{F}_t^N)_{t\geq 0}$ throughout for the natural filtration associated to a Kac process (either labelled or unlabelled) for arguments where this is of interest.

1.2 Topics

With these definitions and notation fixed, we now give an overview of the topics which we will encounter in this thesis and a survey of the relevant literature.

1.2.1 Propagation of Chaos & The Law of Large Numbers

The first topic which will be a recurring theme in this thesis is Kac's original problem [122] of proving the propagation of chaos for the many particle system, which amounts to deriving the Boltzmann equation (BE) from this model; we recall that it is equivalent *either* to prove the convergence of the marginals in the weak topology

$$\mathrm{Law}(V_t^{1,N}, ..., V_t^{l,N}) \to \mu_t^{\otimes l}, \qquad l \geq 1, t > 0$$

or a law of large numbers

$$\mu_t^N \to \mu_t \qquad \text{weakly, in probability.}$$

We refer to Section 2.2 for more details; using metrics which induce the weak topology, either on measures on $(\mathbb{R}^d)^l$ or on $\mathbb{R}^d$, one can ask not only for the qualitative convergence but also about the corresponding rate. Kac himself proposed a combinatorial proof for a one-dimension caricature which preserves energy but not momentum, and required that the collision kernel be bounded. This proof was then applied by McKean [138] to the Kac process as we have defined it here, still requiring the boundedness, which only really allows cutoff Maxwell Molecules (GMM) out of the kernels we have described. Later, Grünbaum developed a functional framework to prove the propagation of chaos for the hard spheres kernel, while Tanaka [177, 178] extended the results to Maxwell molecules without cutoff (tMM). Sznitman [172] also (re)proved the propagation of chaos for the hard spheres model, based on a probabilistic argument which we will use several times in the thesis, proving the tightness of the paths $(\mu_t^N)_{t\geq 0}$ in a suitable space, showing that any subsequential limit $(\mu_t)_{t\geq 0}$ solves the Boltzmann equation, and appealing to a uniqueness result. This then shows the weak convergence of $(\mu_t^N)_{t\geq 0}$ to the corresponding solution to the Boltzmann equation, although this method is inherently non-constructive and cannot be adapted to provide a rate of the convergence. Mischler and Mouhot [142] developed a functional framework for the cases of true Maxwell molecules (tMM) and hard spheres (HS) on which we will build in Chapter 3, and they proved explicit rates for convergence

$$N^{-\alpha_1} \quad \text{(Maxwell Molecules)}; \qquad (\log N)^{-\alpha_2} \quad \text{(Hard Spheres)}, \qquad \alpha_i > 0$$

which are further *uniform in time*. A different probabilistic proof was given for hard spheres by Norris [157] which gives a the optimal N-dependence rate $N^{-1/d}$, uniformly on short time intervals, and a different proof again for the case of Maxwell molecules was given by Cortez and Fontbona [48] with a rate $N^{\varepsilon-1/3}$ in $d=3$ for any $\varepsilon>0$, uniformly in time. As will be discussed in Section 3.1.2, Theorem 1 offers a slight refinement in several senses over the corresponding theorems in the literature [142, 157], since we require fewer moments and have a better rate than the corresponding theorems in the work of Mischler and Mouhot [142].

By comparison, the development of propagation of chaos is significantly less developed for the other kernels (NCHP, SP) where one must deal with both a non-integrable singularity in $b(\cos\theta)$ and the unboundedness of the kinetic factor Ψ at large, respectively small, relative velocities. Fournier and Mischler [87] proved the propagation of chaos for the hard potential Boltzmann equation for a related particle system (the *Nanbu* model) in which only one particle jumps at a time, so that energy is only conserved on average. Recent works by Salem [169] and Xu [196] have considered the same Nanbu model in the case of soft potentials (SP) with various restrictions on the parameters; Salem proved qualitative convergence of chaos, while Xu found a rate $N^{-\alpha}$, but neither treated the (physically more relevant) Kac process. Regarding similar particle models for the Landau equation in place of the Boltzmann, the propagation of chaos was proven for a particle system imitating the Kac process by Fournier and Guérin [88] for the hard potential Landau equation, finding a rate $N^{\varepsilon-1/3}, \varepsilon>0$ in dimension $d=3$, and for a particle system imitating the Nanbu system in the case of the soft potential Landau equation by Fournier and Hauray [89], finding a qualitative result for $\gamma\in(-2,-1]$ and a rate $N^{-\alpha}$ for $\gamma\in(-1,0]$ for some $\alpha>0$. In Theorem 2, we will prove a quantitative rate of propagation of chaos for Kac process in the case (NCHP); to the best of our knowledge, this result (which first appeared in the work [112] by the author) represents the first time that the propagation of chaos has been proven for the true, physical Kac process for either of the cases (NCHP, SP), although our rate is very slow ($(\log N)^{-\alpha}$ on short time intervals).

Time-Dependence of Chaoticity Results We highlighted, in several of the results above, the time-dependence of the rate of convergence of the particle system. This has a natural physical relevance in the validity of the Boltzmann equation (BE) and in particular its spatially inhomogeneous counterpart (spBE). One of the early objections to Boltzmann's equation by Zermelo [198] was that, when the microscopic dynamics are deterministic, Poincaré's recurrence theorem means that, for almost all initial configuration, the particle system will eventually return to a neighbourhood of the initial configuration. On the other hand, Boltzmann's celebrated H-Theorem implies that the macroscopic Boltzmann equation has no such recurrence and converges to equilibrium (see Subsections 1.2.4, 1.2.6 below), and Boltzmann tacitly acknowledged that, for finite N, the spatially inhomogeneous Boltzmann equation (spBE) will fail at some large (N-dependent time), see also the comments in [189, Section 2.5]. Indeed, for a physical system, N is large *but finite*, and so the limit $N \to \infty$ does not make sense; the problem then becomes asking on which timescales the Boltzmann equation is a good approximation to the physical particle system.

Based on this, it is interesting to ask about the time-scales on which the Kac process converges to the spatially homogeneous Boltzmann equation (BE). In this context, we must distinguish between convergence pointwise in time (i.e., estimating, for a single fixed t, $W(\mu_t^N, \mu_t)$ for a suitable metric W) and local uniform estimates, where we consider $\sup_{s\le t} W(\mu_t^N, \mu_t)$. Obtaining estimates of the first type for the cases of Maxwell molecules (tMM) and hard spheres (HS), which are further uniform in the time t, was the subject of the work of Mischler and Mouhot [142]; later, Cortez and Fontbona [48] proved a similar result for Maxwell Molecules with a faster rate in N. In these cases, for fixed N, the Boltzmann equation can be seen as good (pointwise-in-time) approximation to the Kac process for any fixed N. On the other hand, the result of Norris [157] on hard spheres, recalled in Proposition 1.2 below, is an estimate of the second type, but shows only that the Kac process is a good approximation on time scales $t < C\log N$. Even for the large values of N relevant for physical systems (e.g. $N \simeq 10^{23}$), this then fails at times which are small enough to be physically relevant, and it is interesting to try to improve the time dependency. This will be exactly the subject of Theorem 1, which we prove in Chapter 3.

A Fluid Limit Approach to Propagation of Chaos As mentioned above, the propagation of chaos is equivalent to a law of large numbers for the array of (dependent) random variables $\mathcal{V}_t^N = (V_t^{1,N}, \dots, V_t^{N,N})$. Another approach, which is natural in the probabilistic context is the approach of *fluid limits*, see [49]; indeed, we can write the empirical measures μ_t^N as

$$\mu_t^N = \mu_0^N + \int_0^t Q(\mu_s^N)ds + M_t^N \tag{1.15}$$

in the sense that, for any bounded and Lipschitz f,

$$\langle f, \mu_t^N\rangle = \langle f, \mu_0^N\rangle + \int_0^t \langle f, Q(\mu_s^N)\rangle ds + M_t^{N,f} \tag{1.16}$$

for some martingale $M_t^{N,f}$. Further, since the process $\langle f, \mu_t^N\rangle$ makes jumps of order $N^{-1}\sin\theta$ at a rate $\sim Nb(\cos\theta)$, the quadratic variation of the martingale term $M_t^{N,f}$ is of the order $\mathcal{O}(N^{-1})$ with a constant depending on f, see [49, 157]. We can therefore view (1.15) as saying that the empirical measures satisfy a noisy perturbation of the true Boltzmann equation (BE), with vanishing noise in the limit $N \to \infty$. This is exactly the framework of the fluid limit scaling, which has been studied in the probabilistic literature ([49], and references therein [6, 125, 128, 170]), now viewing (BE) as an ODE in the space of measures. This is the approach of Norris [157], whose result we summarise as follows.

Proposition 1.2. *[157, Theorem 10.1] For any $\mu_0 \in \mathcal{S}^p, p > 2$, there is a unique solution in $L^\infty_{loc}([0,\infty), \mathcal{S}^p)$ to the Boltzmann equation (BE), starting from μ_0; we write this solution as $(\phi_t(\mu_0))_{t\geq 0}$.*

Moreover, for any $\epsilon > 0$, $t_{fin} < \infty$, $\lambda < \infty$, there exist constants $C(\epsilon, p) < \infty$ and $\alpha(d,p) > 0$ such that, whenever $(\mu^N_t)_{t\geq 0}$ is a Kac process on $N \geq 1$ particles, with $\Lambda_p(\mu^N_0) \leq \lambda, \Lambda_p(\mu_0) \leq \lambda$, we have

$$\mathbb{P}\left(\sup_{t\leq t_{fin}} W(\mu^N_t, \phi_t(\mu_0)) > e^{C\lambda t_{\text{fin}}}(W(\mu^N_0, \mu_0) + N^{-\alpha})\right) < \epsilon \tag{1.17}$$

where W is a metric of Wasserstein type which is equivalent to the weak topology on $\mathcal{S}$. For $d \geq 3$ and $k > 8$, we can take $\alpha = \frac{1}{d}$.

We will take this same approach in Chapters 3 (in particular) and 6. Following this result, we will often try to obtain local uniform estimates, that is, control on $\sup_{t\leq t_{\text{fin}}} W(\mu^N_t, \mu_t)$ for any $t_{\text{fin}} \geq 0$, as in this result.

1.2.2 Tanaka's Stochastic Interpretation of the Boltzmann Equation & Tanaka Couplings

Tanaka [177, 178, 176] proposed an entirely different probabilistic approach to the Boltzmann equation (BE), which we will also see in the context of the Landau equation (LE). Tanaka argued, for the special case of Maxwell molecules, that the Boltzmann equation can be studied as the Kolmogorov equation for a non-linear stochastic jump differential equation

$$\begin{cases} V_t = V_t + \int_{(0,t]\times\mathbb{R}^d\times\mathbb{S}^{d-1}\times(0,\infty)} (v'(V_{s-}, v_*, \sigma) - V_{s-})1\!\!1[z \leq B(V_{s-} - v_*, \sigma)]\mathcal{N}(ds, dv_*, d\sigma, dz) \\ \mathcal{N} \text{ is a Poisson random measure on } (0,\infty)\times\mathbb{R}^d\times\mathbb{S}^{d-1}\times(0,\infty) \text{ of intensity } dt\mu_t(dv_*)d\sigma dz; \\ \mu_t = \text{Law}(V_t). \end{cases} \tag{stBE}$$

Note that the third condition introduces the nonlinearity, since the intensity of jumps in of $\mathcal{N}$ depends on the distribution $\mu_t = \text{Law}(V_t)$. It follows from the construction that, for all bounded, Lipschitz $f : \mathbb{R}^d \to \mathbb{R}$, the process

$$M^f_t = f(V_t) - f(V_0) - \int_0^t \int_{\mathbb{R}^d} \mathcal{L}_{\text{B}} f(V_s, v_*)\mu_s(dv_*)ds \tag{1.18}$$

is a martingale, where we define the local generator $\mathcal{L}_{\mathrm{B}}$ in analogy to the Landau case

$$\mathcal{L}_{\mathrm{B}} f(v,v_*) := 2\int_{\mathbb{S}^{d-1}} (f(v')-f(v))B(v-v_*,\sigma)d\sigma. \tag{1.19}$$

If we write the symmetrised version

$$\mathcal{L}_{\mathrm{B,s}} f(v,v_*) := \int_{\mathbb{S}^{d-1}} (f(v')+f(v'_*)-f(v)-f(v_*))B(v-v_*,\sigma)d\sigma = \frac{\mathcal{L}_{\mathrm{B}} f(v,v_*)+\mathcal{L}_{\mathrm{B}} f(v_*,v)}{2} \tag{1.20}$$

then for both cases we have the dual representation of $Q(\mu)$ by

$$\langle f, Q(\mu)\rangle) = \int_{\mathbb{R}^d\times\mathbb{R}^d} \mathcal{L}_{\mathrm{B,s}} f(v,v_*)\mu(dv)\mu(dv_*) = \int_{\mathbb{R}^d\times\mathbb{R}^d} \mathcal{L}_{\mathrm{B}} f(v,v_*)\mu(dv)\mu(dv_*)$$

where the first equality is simply the definition of Q, and the second follows by symmetry, since the distribution of (v',v) under $\mu(dv)\mu(dv_*)B(v-v_*,\sigma)d\sigma$ is the same as that of (v'_*,v_*). In particular, it follows that

$$\mathbb{E}\left[\int_{\mathbb{R}^d\times\mathbb{S}^{d-1}} \mathcal{L}_{\mathrm{B}} f(V_t,v_*)\mu_t(dv_*)\right] = \langle f, Q(\mu)\rangle$$

so that taking expectations of (1.18) shows, provided $\mu \in L^\infty_{\mathrm{loc}}([0,\infty),\mathcal{P}_2)$, that $\mu_t = \mathrm{Law}(V_t)$ is a solution to the Boltzmann equation; if we then take V_0 distributed according to any μ_0, independently of $\mathcal{N}$, we then obtain the Boltzmann equation with a specified initial measure. Physically, Tanaka proposed that the process $(V_t)_{t\ge 0}$ should represent the time-dependent behaviour of a 'typical' particle out of the many particles in the cloud of gas. The same idea can be applied to the Landau equation, in which case we write

$$\begin{cases} V_t = V_t + \int_0^t \int_{\mathbb{R}^3} b(V_s-v_\star)\mu_s(dv_\star)ds + \int_0^t\int_{\mathbb{R}^3}\sigma(V_s-v_\star)N(dv_\star,ds); \\ \mu_t = \mathrm{Law}(V_t) \end{cases} \tag{stLE}$$

where N is a 3D-white noise on $\mathbb{R}^3\times[0,\infty)$ with covariance measure $\mu_s(dv_\star)ds$; see Walsh [193]. Again, using the framework (1.8) of weak solutions, it is a straightforward calculation using Itô's formula to see that, if $\mu_t = \mathrm{Law}(V_t)$ has $\mu \in L^1_{\mathrm{loc}}([0,\infty),\mathcal{S}^{2+\gamma})$, then it is a weak solution to (LE).

In either of these two cases, we will say that a process $(V_t)_{t\ge0}$ is a solution to either such equation if there exists a choice of $\mathcal{N}$, respectively N, of the correct distribution which makes the equation true pathwise; we will not insist that V be adapted to the natural filtration for the noise. We will call any such solutions $(V_t)_{t\ge0}$ to (stBE, stLE) Boltzmann, respectively Landau, processes.

Let us review a little the literature on this approach. Tanaka also proposed a probabilistic coupling method for these solutions in the context of the Boltzmann equation, at least in the case of cutoff Maxwell molecules (GMM). Given two solutions $(V_t)_{t\ge0}, (W_t)_{t\ge0}$ driven by Poisson random measures $\mathcal{N}^1,\mathcal{N}^2$ respectively, he built a new solution $(\widetilde{W}_t)_{t\ge0}$ by specifying the associated Poisson random measure $\widetilde{\mathcal{N}}$, with the same law as $\mathcal{N}^2$, in a particular way so that $V_t, \widetilde{W}_t$ both jump at the same times, and such that $\mathbb{E}[|V_t - \widetilde{W}_t|^2]$ decreases in time. The processes $(W_t)_{t\geq 0}, (\widetilde{W}_t)_{t\geq 0}$ have the same law, because $\mathcal{N}^2, \widetilde{\mathcal{N}}$ do, and Tanaka concluded that, *among solutions obtained through (stBE)*, the Boltzmann equation is contractive for the Wasserstein$_2$-distance $\mathcal{W}_2$. Villani and Toscani [180] later extended this conclusion to all solutions.

Beyond the case of cutoff Maxwell molecules, Fournier and Méléard [91] extended this idea to apply to include (NCHP, SP) with $-1 < \gamma < 1$ and without the cutoff assumption; Fournier [83] showed that, if $\gamma \in [0, 1)$, then one can associate such a solution $(V_t)_{t\geq 0}$ to any prescribed solution $(\mu_t)_{t\geq 0}$ to the Boltzmann equation, and for the case $\gamma \in (-1, 0)$ that a solution to (stBE) exists for any given μ_0. A deterministic argument in the same spirit [93] was used to study the stability of the Boltzmann equation in both the cases of hard and soft potentials (NCHP, SP) and, away from the uniqueness and stability, such processes have been used to prove the existence of regular solutions [80] or the finiteness of the entropy[4] for all solutions [83] to the Boltzmann equation using the Malliavin calculus of jump processes. In the Landau case, it was shown in [88, Proposition 10] that all solutions to (LE) with $\Lambda_4(\mu_0) < \infty$ arise as the solution to a stochastic differential equation driven by a Brownian motion which is equivalent to (stLE).

The same 'Tanaka coupling' has also been used in the study of the particle system, both for particle systems leading to the Boltzmann and Landau equations. Rousset [167] developed this coupling for Kac's particle system for Maxwell molecules, showing that the errors are uniform in N, and Cortez and Fontbona [48] used this to prove the uniform in time propagation of chaos. Regarding hard and soft potentials, a similar coupling was used by the works [87, 169, 196] already cited above to prove the convergence of the Nanbu particle system to the Boltzmann equation, and for the Landau equation in [88, 89].

We will use this style of coupling in Chapters 4 - 5, corresponding to Theorems 2-4 for the Boltzmann equation and Landau equation respectively, both in the cases of noncutoff hard potentials. In either case, we present an argument similar to Tanaka's to couple processes $(V_t)_{t\geq 0}$ given by (stBE, stLE) respectively. In either case, we do not find contractivity, as Tanaka did, but the novelty of our work is that the solutions grow apart no faster than a time-dependent multiple of the initial data, measured in a tailor-made cost. In the case of the Boltzmann equation, we also exhibit a Tanaka-coupling of the particle system, with an error uniform in N, which we use to prove propagation of chaos.

[4] see Section 1.2.4 below.

1.2.3 The Cauchy Problem for the Boltzmann and Landau Equations

Another, purely analytic problem on which we have already touched is the theory of well-posedness (existence, uniqueness, continuity in the initial data) of the Boltzmann equation. In the works in the propagation of chaos, uniqueness is already necessary in Sznitman's non-constructive argument; Grünbaum's argument [104] was incomplete for the same reason (see the introduction to [142]), as it assumed this uniqueness in cases for which it is false. Moreover, for the quantitative propagation of chaos, we will typically wish to prove an estimate of $W(\mu^N_t, \mu_t)$ in terms of some sequence $\varepsilon_N \to 0$ depending on N, and on $W(\mu^N_0, \mu_0)$. If we prove such an estimate, then we could apply the same thing to $W(\mu^N_0, \nu_0)$ for any other ν_0 and combine the two to get an estimate on $W(\mu_t, \nu_t)$. In this way, quantitative estimates for the propagation of chaos are at least as hard as proving the uniqueness and (quantitative) stability for the Boltzmann equation, so it will also be necessary to consider the Cauchy problem.

Let us also make a remark on regularity. In the cases without cutoff, μ_t can never be more regular than μ_0; for instance, it can never be function-valued if the initial data μ_0 is not already absolutely continuous with respect to the Lebesgue measure, thanks to the easy estimate

$$\mu_t(dv) \geq \exp\left(-2C\int_0^t \int_{\mathbb{R}^d} \Psi(|v-w|)\mu_s(dw)ds\right)\mu_0(dv), \qquad C = \int_{\mathbb{S}^{d-1}} B(u,\sigma)d\sigma$$

but, assuming only that $\mu_0 \in L^2(\mathbb{R}^d) \cap \mathcal{P}_2(\mathbb{R}^d)$, it holds that μ_t can be decomposed into a regular part and a part which decays exponentially in time [150, Theorem 5.5]. On the other hand, in the cases without cutoff, the (linear) operator $Q(\cdot, \mu)$ behaves in a sense like a fractional diffusion $-(-\Delta)^{\nu/2}, \nu > 0$, see [8], and in particular there is hope for the regularisation of solutions. We will explore this a little in Chapter 5, where we upgrade existing regularity results through a new uniqueness result.

We summarise the state-of-the-art for existence, uniqueness and regularity as follows.

1. Maxwell Molecules For Maxwell molecules, with or without cutoff, the theory of the Cauchy problem is quite advanced. Existence follows from Tanaka's work [177] in which he established solutions to the corresponding stochastic form (stBE), and uniqueness and stability were proven by Toscani and Villani [180]. Regarding regularity in the case (tMM) without cutoff, Fournier [80] proved the existence of a smooth solution in $d = 2$, and more recently Moritmo [146] proved that all solutions immediately admit a smooth density, assuming only finite energy and entropy on the initial data.

2. Hard Spheres Similarly, the Cauchy problem for hard spheres is quite advanced. Arkeryd [12] gave an existence and uniqueness theory for measures with finite entropy; Lu and Mouhot [132] proved the existence of energy-conserving measure-valued solutions for any $\mu_0 \in \mathcal{P}_2$, and that solutions are unique within energy-conserving paths. Norris [157] re-proved, via a different, probabilistic method, existence and uniqueness of energy-conserving solutions[5] with $\mu_0 \in \mathcal{S}^p$, with a quantitative stability estimate in a Wasserstein$_1$-type metric[6]; we will reproduce the proof in Section 3.2 as a stepping stone to sharper results. Currently, no quantitative stability estimates are known with only $p = 2$ moments.

Let us also refer to the work of Mischler and Mouhot [142], who proved a very strong 'twice-differentiability' result for this kernel and for Maxwell molecules, with or without cutoff, in both cases with exponential decay in time. For the case of hard spheres, the distances are measured in weighted total variation norms; the statements and a sketch proof for this case are reproduced in Section 3.2.

3. Non-Cutoff Hard Potentials. The Cauchy theory for the case (NCHP) is, in comparison to the two cases discussed above, comparatively recent and significantly less complete. Villani [186] developed a theory of function-valued solutions, extended to measure-valued solutions by Lu and Mouhot [132, Theorem 1.3], assuming only that $\mu_0 \in \mathcal{P}_2$. Regarding uniqueness and stability, two results have been proven by Desvillettes and Mouhot [56] and by Fournier and Mouhot [93]. For uniqueness among energy conserving solutions, the result of [93], which is recalled in Proposition 4.18, assumes that μ_0 has an exponential moment $\langle e^{\varepsilon|v|^\gamma}, \mu_0\rangle < \infty$ for some $\varepsilon > 0$, and provides a quantitative stability result for μ_0, ν_0 both satisfying this condition, while the result of [56] requires μ_0 to have a density f_0 lying in a weighted $W^{1,1}$ space, and so requires less localisation (by comparison) but much more regularity. Regarding regularity, Alexandre et al. [8] proved in $d = 3$ that if μ_0 admits a density, then μ_t admits a density f_t with $\sqrt{f_t} \in H^{\nu/2}_{\text{loc}}(\mathbb{R}^3)$ for any $t > 0$, which Chen and He [47] improved to the global integrability $(1 + |v|^2)\sqrt{f_t} \in H^{\nu/2}(\mathbb{R}^3)$. Fournier [83] used the Tanaka processes described in the previous subsection to show that, unconditionally, μ_t has a finite entropy $H(\mu_t) < \infty$, and that the density belongs to a certain Besov space. In the case of a regularised hard potentials, where Ψ has the same asymptotic growth but is now smooth, Desvillettes and Wennberg proved the existence of solutions such that μ_t admits a density f_t in the Schwarz space[7], provided only that μ_0 has finite energy and momentum.

[5] The conditions stated in the cited paper are that $\mu \in L^\infty_{\text{loc}}([0, \infty), \mathcal{S}^p)$ for some $p > 2$; this is equivalent to energy conservation plus the finiteness of $\Lambda_p(\mu_0)$ by estimates which we will see in Section 2.5.

[6] See Section 2.1.

[7] The Schwarz space is the space of all smooth functions such that $\sup_v (1 + |v|^s)|D^s f|(v) < \infty$ for all k, s. This is usually denoted $\mathcal{S}(\mathbb{R}^d)$, but this notation is already taken for the Boltzmann sphere.

4. The Landau Equation for Hard Potentials and Maxwell Molecules The Landau equation in the cases $\gamma \in [0,1]$ for hard potentials and Maxwell molecules was studied in detail by Villani [187] and Desvillettes and Villani [58, 59] respectively. For Maxwell molecules, Villani proved the existence of a unique classical solution starting from any μ_0 admitting a density $f_0 \in L^1(\mathbb{R}^3)$, and that f_t is smooth and bounded for $t > 0$. In the case of hard potentials, Desvillettes and Villani proved existence assuming that $\mu_0 \in \mathcal{P}_{2+\epsilon}, \varepsilon > 0$, and uniqueness assuming that the initial data μ_0 has a density $f_0 \in L^1(\mathbb{R}^3) \cap L^2(\mathbb{R}^3, (1+|v|^p)dv)$ for $p > 15+5\gamma$. Regarding regularity, they studied the regularising effects of the diffusion operator $Q_L(\mu)$ and showed the existence of a solution whose density f_t belongs to weighted Sobolev spaces of all orders, provided that μ_0 does not concentrate on a line (see Proposition 5.13). Later, Fournier and Guillin [88] proved a uniqueness and stability result with measure-valued initial data with exponential moments, which is recalled in Proposition 5.8. The regularity results were later extended Morimoto, Pravda-Starov and Xu [145], who showed that any solution μ_t whose density satisfies the conclusions of Desvillettes and Villani further has an analytic density.

5. Soft Potentials. Finally, let us mention the Cauchy theory for soft potentials, either for the Boltzmann or Landau equations. For the Boltzmann equation with $\gamma \in (-1,0), \nu \in (0,1), \gamma+\nu > 0$, Fournier and Mouhot [93, Corollary 2.4] proved existence and and uniqueness in $L^\infty_{\text{loc}}([0,\infty), \mathcal{P}_2) \cap L^1_{\text{loc}}([0,\infty), L^p(\mathbb{R}^d))$ for the case where the initial data has q moments and admits a density $f_0 \in L^1(\mathbb{R}^d) \cap L^p(\mathbb{R}^d)$, for some particular p, q. Xu [196] added a weak-strong uniqueness and stability result so that, under the same hypotheses, the solution starting at μ_0 is unique among all weak solutions $\nu \in L^\infty_{\text{loc}}([0,\infty), \mathcal{P}_2)$, with a stability estimate depending only on the L^p norms of the densities f_t associated to μ_t. In the cases $\gamma \in (-3,0], \nu \in (0,2)$ in dimension $d=3$, Fournier and Guérin [85] proved uniqueness and stability for the classes of solution where

$$\mathcal{J}_\gamma(\mu) = \sup_{v \in \mathbb{R}^3} \int_{\mathbb{R}^3} |v - v_*|^\gamma \mu(dv_*)$$

remains bounded along the solution.

For the Landau equation in the cases of soft and Coulomb potentials, the existence goes back to Villani [187]. For $\gamma > -3$, uniqueness and stability were studied by Fournier and Guérin [86], who found uniqueness and stability, with different hypotheses depending on the value of γ: for $\gamma \in (1-\sqrt{5}, 0)$ finite energy and entropy suffice, for $\gamma \in (-2, 1-\sqrt{5}]$ an additional (sufficiently large) moment is needed, and for $\gamma \in (-3,-2]$ one needs finite energy, and that μ_0 admits a density $f \in L^1(\mathbb{R}^3) \cap L^p(\mathbb{R}^3), p > 3/(3+\gamma)$. For the most physically relevant (and most difficult) Coulomb case $\gamma = -3$, Fournier [82] proved the uniqueness of solutions admitting a bounded density, based on Tanaka's stochastic interpretation, while the only existence result for bounded solutions [14] is for short times, so solutions may become unbounded at finite times.

Energy Non-Conserving Solutions

We insist, in the discussion of hard spheres, and hard potentials above, that uniqueness only holds within the class of *energy conserving* solutions, even though, as remarked above,

$$\langle |v|^2, Q(\mu)\rangle = 0$$

so that one formally expects the energy to be constant. However, Lu and Wennberg [133] showed that in the cases (HS, NCHP, CHP_K), for any initial data μ_0, that there exists a solution with *increasing* energy, and moreover very many such solutions: the energy can be prescribed to have finitely many arbitrary jumps, or to increase in a 'Cantor-like' way[8]; we will see a very similar argument at the level of the Kac process in Chapter 6, where this ill-posedness causes problems for the study of large deviations. Lu and Wennberg also showed that the energy can never decrease, so that these increases are the only way in which the energy fails to be constant.

Let us remark that in the case of the Landau equation, we insist by definition that our solutions have non-increasing energy, and so we cannot have solutions with increasing energy as above, and as already remarked above, Desvillettes and Villani [59, Theorem 3] showed that the energy is constant.

Notation. In each of the cases above, we will often write ϕ_t for the semigroup on suitable spaces of measures $\mathcal{S}' \subset \mathcal{S}$ where the Cauchy problem is well-posed; that is, $(\phi_t(\mu))_{t\ge 0}$ is the unique solution to the Boltzmann equation starting at μ, and the maps ϕ_t take the set $\mathcal{S}'$ (to be specified in each case) to itself. This notation will be a helpful clarification in Chapter 3, where we follow [142] in using properties not only of the Boltzmann equation, but also of the maps ϕ_t, and again in Section 4.9, where we prove uniqueness by constructing *a* solution map ϕ_t with a stability property. In the case of the Landau equation, we write ϕ^{L}_t.

Contributions. Let us briefly discuss the novel contributions of this thesis to the study of the Cauchy problems to these equations in view of the literature above. For (HS), we prove in Theorem 1 a uniform-in-time stability estimate for the (energy-conserving) solution, which is Hölder continuous in a weighted Wasserstein distance and uniform in time, as soon as the initial data have any p moments, $p > 2$. This comes from interpolating between the techniques of Norris [157], which work well in short-time and with only few moments, and the exponential stability in total variation in the long time; we introduce some further ideas to ensure that the dependence on the moments $\Lambda_p(\mu_0, \nu_0)$ of the initial data only appears as a multiplicative factor, and does not change the Hölder exponent.

In the case of non-cutoff hard potentials (NCHP) or the corresponding hard potential

[8] See Theorem 6.5.

Landau equation, one of the main contributions of Theorems 2 - 3 will be a new uniqueness result which applies to measure-valued solutions and only requiring a finite number p moments, which is based on Tanaka's coupling. In the Boltzmann case (NCHP), the required number of moments is potentially large but constructable; in the Landau case, we pay particular attention to this dependence to show that any $p > 2$ is sufficient. We also check, in Theorem 3, that solutions to (LE) exist as soon as the initial data $\mu_0 \in \mathcal{S}(\mathbb{R}^3)$.

1.2.4 Entropy, the H-Theorem and Large Deviations

An important and powerful tool in the study of the Boltzmann equation, both in the inhomogeneous and homogeneous cases (spBE, BE) has been the *entropy* and associated methods, going back as far as Boltzmann himself. Boltzmann proposed the definition of entropy in terms of the logarithm of the volume of accessible microstates, measuring how exceptional a given configuration is; for a probability measure[9] on $\mathbb{R}^d$ one defines the entropy

$$H(\mu) := \begin{cases} \int_{\mathbb{R}^d} f(v) \log f(v)\, dv & \text{if } \mu \text{ has a density } f \text{ with respect to the Lebesgue measure;} \\ \infty & \text{else} \end{cases} \tag{1.21}$$

and similarly the relative entropy with respect to a fixed ν is

$$H(\mu|\nu) := \begin{cases} \int_{\mathbb{R}^d} f(v) \log f(v)\, \nu(dv) & \text{if } \mu \text{ has a density } f \text{ with respect to } \nu; \\ \infty & \text{else.} \end{cases} \tag{1.22}$$

In the latter case, $H(\mu|\nu) \geq 0$ thanks to Jensen's inequality, and $H(\mu|\nu)$ vanishes if and only if $\mu = \nu$. Let us note that, for $\mu \in \mathcal{P}_2$ admitting a density f, if we take γ_μ to be the Gaussian with scalar covariance and the same average momentum $u = \int v\mu(dv) = \int v\gamma_\mu(dv)$ and temperature $T = \frac{1}{d}\int |v-u|^2\mu(dv) = \int |v-u|^2\gamma_\mu(dv)$, then

$$\frac{d\mu}{d\gamma_\mu(v)} = f(v)\frac{e^{-\frac{|v-u|^2}{2T}}}{(2\pi T)^{d/2}}$$

which produces

$$H(\mu|\gamma_\mu) = H(\mu) - \frac{d}{2} - \frac{d}{2}\log(2\pi T) = H(\mu) - H(\gamma_\mu).$$

For $\mu \in \mathcal{S}$, we find the standard Gaussian $\gamma(dv)$ with density

$$\gamma(v) = \frac{1}{(2\pi d^{-1})^{d/2}} e^{-d|v|^2/2}. \tag{1.23}$$

In this context, for solutions $(\mu_t)_{t\geq 0}$ normalised to $\mathcal{S}$, Boltzmann's celebrated H-Theorem writes as follows.

[9] or the same on $D \times \mathbb{R}^d$ in the inhomogeneous case.

Proposition 1.3 (Boltzmann's H-Theorem). *For a solution $(\mu_t)_{t\geq 0} \subset \mathcal{S}$ to (BE), we have*

$$H(\mu_t|\gamma) + \int_0^t D(\mu_s)ds = H(\mu_0|\gamma) \tag{H}$$

where $D(\mu)$ is the entropy dissipation

$$\begin{aligned} D(\mu) &= \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}} (f(v')f(v'_*) - f(v)f(v_*))\log\frac{f(v')f(v'_*)}{f(v)f(v_*)}B(v-v_*,\sigma)dvdv_*d\sigma \\ &= \int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}} f(v)f(v_*)\log\frac{f(v)f(v_*)}{f(v')f(v'_*)}B(v-v_*,\sigma)dvdv_*d\sigma \end{aligned} \tag{1.24}$$

if μ has a density f, in which case we write $D(\mu) = D(f)$, and otherwise $D(\mu) = \infty$. Moreover, $D(\mu) \geq 0$, so that the entropy $t \mapsto H(\mu_t|\gamma)$ is nonincreasing, and $D(\mu) = 0$ if and only if μ is a Maxwellian distribution, so that $t \mapsto H(\mu_t|\gamma)$ is strictly decreasing unless $\mu_t = \gamma$.

It follows that γ is the unique fixed point of the Boltzmann equation (BE) normalised to $\mathcal{S}$, and without the normalisation one finds all Maxwellian distributions $\mu(dv) \propto e^{-|v-u|^2/2T}$, $u \in \mathbb{R}^d, T > 0$, including point masses as the degenerate case $T = 0$. A similar argument holds for the Landau equation [59] for the same entropy, and a suitably defined entropy dissipation. Let us mention a connection here to the stochastic processes which we study; Kac [122, Equation 6.39] proposed that a certain property, called entropic chaoticity in later works [38, 108, 142], is propagated in time, while the relative entropy for the particle system is monotonically decreasing in time. Together, these would lead to a conclusion weaker than (H), which is that $H(\mu_t|\gamma)$ is monotonically decreasing. This is the approach of [142, Theorem 7.1], which derives (H) from the propagation of entropic chaoticity in the cases (GMM, tMM, HS).

Although it is not a topic to which we will contribute, let us mention the study of functional inequalities connected to (H), which has been an important topic in kinetic theory. Boltzmann's H-Theorem, on its own, proves the convergence to equilibrium, but not a rate, and functional inequalities connecting D, H allow 'entropy entropy-dissipation' arguments to estimate the rate of convergence. Cercignani's conjecture [43] posited a linear lower bound on $D(\mu)$, with some constant depending on *a priori* estimtates on μ (entropy, Sobolev norms, moments,...), but successive works of Bobylev [18, 19], Wennberg [194] and Cercignani [21] showed that this linear relation is false in cases of increasing generality, while results of Toscani and Villani showed that a *super*linear lower bound holds under general conditions [181, 182, 190]

$$D(\mu) \geq \lambda_\varepsilon(\mu)H(\mu|\gamma)^{1+\varepsilon}$$

for a suitable constant λ_ε and any $\varepsilon > 0$, leading to convergence to equilibrium with arbitrarily large polynomial exponent:

$$H(\mu_t|\gamma) \leq Ct^{-1/\varepsilon}.$$

Let us refer to the survey work [57] for more details.

A different viewpoint on entropy in the context of the kinetic equations (BE, LE) comes from the theory of gradient flows for measure spaces; Jordan, Kinderlehrer and Otto [120] showed that the heat equation is the gradient flow of the functional H with respect to the Wasserstein$_2$ distance $\mathcal{W}_2$, so that the entropy decreases (in a sense) as efficiently as possible. The same approach was applied to discrete binary reaction networks by Mielke [140] for a particular choice of geometry adapted to each reaction network, and the Boltzmann equation (formally) fits this framework by viewing particle velocities $v \in \mathbb{R}^d$ as a continuum of particle species. Erbar [73] proved that (BE) is the gradient flow of entropy where the kernel is bounded in the 'ω-representation' defined in Section 2.4; out of our kernels, this applies to only (GMM), and where $\mathcal{P}_2$ is equipped with a bespoke metric. The same idea was implemented for the Landau equation in [42].

A probabilistic interpretation, and the one which we will persue in this thesis, of entropy is the concept of large deviations, which gives a precise mathematical meaning to Boltzmann's original idea of entropy in terms of the volume of accessible microstates, see the discussion in [192]. For example, if μ_0^N is the empirical measure from sampling particles independently from $\nu \in \mathcal{P}_2$, then Sanov's Theorem [51] applies to show that the measures μ_0^N satisfy a large deviation principle

$$\mathbb{P}\left(\mu_0^N \approx \mu\right) \asymp e^{-NH(\mu|\nu)} \tag{1.25}$$

which formally means that, for all $A \subset \mathcal{P}$ and $U \subset \mathcal{P}$ closed, respectively open, for the weak topology, we have

$$\limsup_N \frac{1}{N} \log \mathbb{P}\left(\mu_0^N \in A\right) \leq -\inf\left\{H(\mu|\nu) : \mu \in A\right\}; \tag{1.26}$$

$$\liminf_N \frac{1}{N} \log \mathbb{P}\left(\mu_0^N \in U\right) \geq -\inf\left\{H(\mu|\nu) : \mu \in U\right\}. \tag{1.27}$$

Equivalently, for a Wasserstein metric[10] inducing the weak topology

$$H(\mu|\nu) = \limsup_{\varepsilon \to 0} \limsup_{N \to \infty} \left(-\frac{1}{N} \log \mathbb{P}\left(\mathcal{W}_{1,1}(\mu_0^N, \mu_0) < \varepsilon\right)\right) \tag{1.28}$$

which makes the connection to Boltzmann's notion of 'logarithm of volume of microstates' precise. On the probabilistic side, the dynamical large deviations of noisy ODEs have been extensively studied since the work of Freidlin and Wentzell [94], including the seminal work by Feng and Kurtz [77]. In Chapter 6, we will study the dynamical large deviations of the empirical measures μ_t^N on a finite time interval $[0, t_{\text{fin}}]$, which is a Freidlin-Wentzell theory for the Kac process viewed as a stochastic, noisy Boltzmann equation as in (1.15) above. Writing $\bullet$ for processes[11] indexed by $t \in [0, t_{\text{fin}}]$, we seek estimates which are informally

[10] See Section 2.1.

[11] This notation is chosen to try to avoid confusion between functionals of a single measure, for instance $H(\mu|\gamma)$, and those which depend on the whole process.

stated as

$$\mathbb{P}\left(\mu^N_\bullet \approx \mu_\bullet\right) \asymp e^{-N\mathcal{I}(\mu_\bullet)} \tag{1.29}$$

for some function $\mathcal{I}(\mu_\bullet)$ on a suitable space $\mathcal{D}$ of càdlàg paths $\mu_\bullet : [0, t_{\text{fin}}] \to \mathcal{P}_2$.

Auxilliary Flux Following previous works in large deviations [160, 165], it will be useful to consider the Kac process together with an auxiliary *empirical flux*, which records the collision history of the process, on the space $E = (0, t_{\text{fin}}] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{S}^{d-1}$ for the parameter space of collisions[12], and form measures w^N_t on E by setting $w^N_0 = 0$ and changing, at collisions,

$$w^N_t = w^N_{t-} + \frac{1}{N}\delta_{(t,v,v_\star,\sigma)} \tag{1.30}$$

at times t where there is a collision. In this way, the pair (μ^N_t, w^N_t) together form a Markov process, with generator given on bounded functions by

$$\begin{aligned}\mathcal{G}^{N,\mathrm{Fl}}F(\mu^N, w^N) = N\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}} &\left(F(\mu^{N,v,v_\star,\sigma}, w^{N,t,v,v_\star,\sigma}) - F(\mu^N, w^N)\right) \\ &\cdots \times B(v-v_\star,\sigma)\mu^N(dv)\mu^N(dv_\star)d\sigma\end{aligned} \tag{1.31}$$

with

$$w^{N,v,v_\star,\sigma} := w^N + \frac{1}{N}\delta_{(t,v,v_\star,\sigma)}. \tag{1.32}$$

We write $w^N := w^N_{t_{\text{fin}}}$ for the final measure, containing the entire collisional history of the process on the finite time-interval. We then investigate estimates

$$\mathbb{P}\left((\mu^N_\bullet, w^N) \approx (\mu_\bullet, w)\right) \asymp e^{-N\mathcal{I}(\mu_\bullet, w)}. \tag{1.33}$$

Estimates of the form (1.29) can then be derived from these estimates by the contraction principle [77, 147], setting $\mathcal{I}(\mu_\bullet) = \inf_w \mathcal{I}(\mu_\bullet, w)$. Formally, (1.33) means that, for Kac processes $\mu^N_\bullet$ and associated empirical fluxes w^N,

$$\limsup_N \frac{1}{N}\log\mathbb{P}\left((\mu^N_\bullet, w^N) \in \mathcal{A}\right) \le -\inf\left\{\mathcal{I}(\mu_\bullet, w) : (\mu_\bullet, w) \in \mathcal{A}\right\}; \tag{1.34}$$

$$\liminf_N \frac{1}{N}\log\mathbb{P}\left((\mu^N_\bullet, w^N) \in \mathcal{U}\right) \ge -\inf\left\{\mathcal{I}(\mu_\bullet, w) : (\mu_\bullet, w) \in \mathcal{U}\right\} \tag{1.35}$$

for all $\mathcal{A}$ closed and $\mathcal{U}$ open subsets of a suitable topological space, which we will specify in detail in Section 6.1.

1.2.5 Tree Expansion & Interaction Clusters

Our next topic is the study of some combinatorial objects which appear related to the study of the *inhomogeneous* Boltzmann equation (spBE), both in Lanford's work [129] and

[12] although it will be convenient to work with a slightly different parametrisation of collisions with the same parameter space, see Section 2.4.

in a series of more recent works by Bodineau, Gallagher, Saint-Raymond, Simonella and Texier [22, 23, 24, 98]. We probe these objects in the (simpler) case of Kac's interacting system, where an exact derivation of the convergence to some limiting kinetic equations is possible.

We begin with Wild sums [195] and its probabilistic interpretation by McKean [138], see also [191, Section 4.1]. We start by rewriting, for cutoff kernels B, the collision operator Q as $Q = Q^+ - Q^-$, where the gain and loss terms come from integrating the positive, respectively negative, terms:

$$Q^+(\mu,\nu) = \int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}} (\delta_{v'}+\delta_{v'_*})B(v-v_*,\sigma)\mu(dv)\nu(dv_*)d\sigma;$$

and the loss term is

$$Q^+(\mu,\nu) = \int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}} (\delta_{v}+\delta_{v_*})B(v-v_*,\sigma)\mu(dv)\nu(dv_*)d\sigma.$$

Defining the function $A(\mu,v) := \int_{\mathbb{R}^d\times\mathbb{S}^{d-1}} B(v-v_*,\sigma)\mu(dv_*)d\sigma$, we write the loss term as

$$Q^-(\mu,\nu)(dv) = A(v,\nu)\mu(dv) + A(v,\mu)\nu(dv).$$

We now restrict to the case of quadratic kernels ($\mathrm{Q}_{a,b}$), which contains the special case of cutoff Maxwell molecules originally considered by Wild. For the case ($\mathrm{Q}_{a,b}$) and any $\mu \in \mathcal{S}$, we expand the function A as

$$\begin{aligned} A(v,\mu) &= \int_{\mathbb{R}^d} (a + b(|v|^2 - 2v\cdot w + |w|^2))\mu(dw) \\ &= (a+b) + b|v|^2 =: A(v) \end{aligned} \tag{1.36}$$

which simplifies to a function *only* of v, thanks to the specified integrals $\int w\mu(dw) = 0, \int |w|^2\mu(dw) = 1$. The Boltzmann equation, for processes $(\mu_t)_{t\ge 0} \subset \mathcal{S}$, now reads

$$\partial_t \mu_t = -2A\mu_t(dv) + Q^+(\mu_t)$$

where again we supress the repeated argument in $Q^+(\mu) = Q^+(\mu,\mu)$. We now integrate via Duhamel's formula to produce

$$\mu_t = e^{-2At}\mu_0 + \int_0^t e^{-2(t-s)A}Q^+(\mu_s)ds$$

and this process can be repeated, substituting the resulting expression for μ_s into the integrand and using the bilinearity of $Q^+(\cdot,\cdot)$. Iterating this produces more complicated expressions, which fall into a *tree structure*. Let us write $\mathcal{T}_n$ for the set of all rooted, binary trees Γ with n leaves, which we build recursively from the empty tree $\circ$, forming any $\Gamma \in \mathcal{T}_n$ by joining left- and right- trees $\Gamma = (\Gamma(L), \Gamma(R))$ with $\Gamma(L) \in \mathcal{T}_p, \Gamma(R) \in \mathcal{T}_q, p, q \in \{1, \dots, n-1\}, p+q = n$. We then find a recursive expansion of μ_t in terms of this tree structure: we set $\mu_t^\circ = e^{-2tA}\mu_0$, and for $\Gamma = (\Gamma(\mathrm{L}), \Gamma(\mathrm{R})(\in \mathcal{T}_n$, define

$$\mu_t^\Gamma = \int_0^t e^{-2A(t-t_1)}Q^+(\mu_{t_1}^{\Gamma(\mathrm{L})}, \mu_{t_1}^{\Gamma(\mathrm{R})})dt_1.$$

Overall, the iterations of Duhamel's formula produce

$$\mu_t = \sum_{n\geq 1}\sum_{\Gamma\in\mathcal{T}_n} \mu_t^\Gamma. \tag{1.37}$$

In the Maxwell molecule case $b = 0$ (in the notation of ($Q_{a,b}$)), we can simplify this expression a little. In this case, $A = a$ is simply a scalar, and the factors $e^{-2A(t-t_1)}$ can be moved outside the operator Q^+ to find, for $\Gamma \in \mathcal{T}_n$,

$$\mu_t^\Gamma = e^{-2at}(1-e^{-2at})^{n-1}Q_\Gamma^+(\mu_0)$$

where Q_Γ^+ is defined recursively by

$$Q_\Gamma^+(\mu) = Q^+\left(Q^+_{\Gamma(\mathrm{L})}(\mu), Q^+_{\Gamma(\mathrm{R})}(\mu)\right)$$

which exactly reproduces McKean's interpretation [138] of Wild's sum [195]. In the general case of ($Q_{a,b}$), one could find a similar (but more complicated) expansion, since now the weights e^{-2At} have to be kept inside the operator Q^+. Such trees have, in either case, a natural interpretation: for the empty tree $\circ$, μ_t° is the contribution from particles which have undergone no collisions in the time interval $[0,t]$, while $\mu_t^{(\circ,\circ)}$ represents the contribution from particles which have undergone exactly one collision, with a particle which had not collided before, and so in. In general μ_t^Γ represents the contribution from particles with a 'collision history' Γ, that is, particles which previously had a collision history $\Gamma(\mathrm{R})$, and then collided with a particle of collision history $\Gamma(\mathrm{L})$, with $\Gamma(\mathrm{R}), \Gamma(\mathrm{L})$ again the right and left subtrees of Γ. Diagrammatically, these represent particle histories as follows.

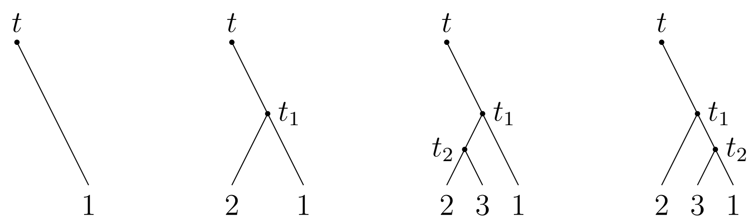

Figure 1.1: Terms in the tree expansion from the first two iterations of Duhamel's formula, from trees $\circ$, $(\circ,\circ)$ and $((\circ,\circ),\circ), (\circ,(\circ,\circ))$ respectively, with labelled leaves.

The contributions to the tree expansion measure μ_t from the first two collision histories given above are, respectively, $e^{-2tA}\mu_0$ and $e^{-2(t-t_1)A}Q^+(e^{-2t_1A}\mu_0, e^{-2t_1A}\mu_0)$, which we integrate over $t_1 \in [0,t]$. For the trees with three leaves, we find the more complicated expressions

$$e^{-2(t-t_1)A}Q^+(e^{-2(t_1-t_2)A}Q^+(e^{-2t_2A}\mu_0, e^{-2t_2A}\mu_0), e^{-2t_1A}\mu_0);$$
$$e^{-2(t-t_1)A}Q^+(e^{-2t_1A}\mu_0, e^{-2(t_1-t_2)A}Q^+(e^{-2t_2A}\mu_0, e^{-2t_2A}\mu_0)).$$

At the level of the particle system, we can similarly think of the (random) interaction history of individual particles. In this case, it is possible that the diagram is no longer a tree; it is possible, for example, that the same two particles collide twice, but these events should contribute negligibly in the limit $N \to \infty$, since the interaction rate between each pair scales as N^{-1}. We also remark that two 'tagged' particles in the particle system can become dependent, even if they do not appear in each others collision histories, if some particle appears in both trees, as in Fig. 1.2 below; in the works [22, 23, 24] investigating the fluctuations of the spatial particle system around (spBE), these are called *recollisions*.

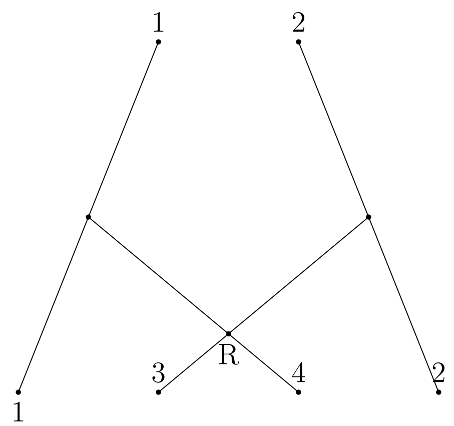

Figure 1.2: Two tagged particles in the Kac process which do not appear in each others collision trees, but whose collision trees overlap, as they both contain the particles labelled $3, 4$. In the language of [22, 23, 24], the event labelled R is a recollision.

Interaction Clusters Bogolyubov [25] introduced the notion of an interaction cluster; at least in short time, the particle system can be partitioned into finite *clusters* whose particles evolve independently of other clusters; the clusters at time t are formed by grouping together any particles which are joined by a chain of collisions. In terms of the previous observation, this amounts exactly to grouping together all particles whose collision trees at time N overlap, or which can be joined by a chain of overlapping collision trees. Following previous works [96, 161], we investigate the statistical properties of these clusters in the limit $N \to \infty$. Let us first show, for the Kac process with kernels ($Q_{a,b}$), that this cluster structure really does capture the structure of dependencies induced by the collisions. If we fix $t > 0$ and a nonrandom $(V_0^1, \dots, V_0^N)$, let us condition on the partition $(\mathcal{C}_1(t), ..., \mathcal{C}_{l_N}(t))$ of the index set $\{1, ..., N\}$ formed by grouping together indexes by collisions before time t. For the kernels ($Q_{a,b}$), we observe that the rate of collisions on $[0,t]$ which this conditioning forbids is

$$\sum_{w,z\le l_N(t), w\neq z}\ \sum_{i\in\mathcal{C}_w(t), j\in\mathcal{C}_z(t)} (a+b|V^i_s|^2-2bV^i_s\cdot V^j_s+b|V^j_s|^2)$$
$$= \sum_{w,z\le l_N(t), w\neq z} (a|\mathcal{C}_w(t)||\mathcal{C}_z(t)|+b|\mathcal{C}_z(t)|\left(\sum_{i\in\mathcal{C}_w(t)}|V^i_0|^2\right)+b|\mathcal{C}_w(t)|\left(\sum_{j\in\mathcal{C}_z(t)}|V^j_0|^2\right)$$
$$\cdots - 2b\left(\sum_{i\in\mathcal{C}_w(t)}V^i_0\right)\cdot\left(\sum_{j\in\mathcal{C}_z(t)}V^j_0\right) \tag{1.38}$$

where we have used the observation that $\sum_{i\in\mathcal{C}_w(t)}(1,V^i_s,|V^i_s|^2)$ is constant over $s\in[0,t]$, since no particles in $\mathcal{C}_w(t)$ interact with any outside it on this time interval. We conclude that the overall rate is constant in time and deterministic (because all V^i_0 are), which ensures that all possible collisions within each part of the partition have the same rates as without the conditioning. In particular, conditional on the partition, for each i,j in different $\mathcal{C}_w(t)\neq\mathcal{C}_z(t)$, the velocities V^i_t, V^j_t are (conditionally) independent, so the clusters really do capture the dependencies of the particles introduced by collisions.

We can also connect this idea to the works [22, 23, 24], which consider the fluctuations of the deterministic dynamics (with random initial condition) about the inhomogeneous Boltzmann equation (spBE) through the *cumulant expansion*, in our notation

$$\Lambda^N_t(h) := N^{-1}\log\mathbb{E}\left[\exp\left(\sum_{i=1}^N h((x_i(s),v_i(s))_{0\le s\le t})\right)\right]$$

for functions h on the space of paths on $[0,t]$, which can be expressed [24, Proposition 2.13] by integrating tensor products of (e^h-1) against a sequence $g^{N,k}_t$ of measures obtained in a one-to-one correspondence with sequence of k-particle marginals $\widetilde{F}^{N,k}=\mathrm{Law}((x_1(s),v_1(s))_{0\le s\le t},\dots,(x_k(s),v_k(s))_{0\le s\le t})$. It is shown [23, Equation 4.5] that $g^{N,k}$ is supported on particle histories of k tagged particles which are completely connected by a sequence of collision events, which is exactly the requirement that the k tagged particles belong to the same interaction cluster. Let us show the same connection for the Kac process; we will show that the cumulant generating function, conditional on the partition $\mathcal{C}_t$, splits up into a sum of the same function over each cluster. As in the argument (1.38) above, let us fix a nonrandom $\mathcal{V}^N_0$ and condition on the partition $\mathcal{C}(t)=(\mathcal{C}_1(t),...,\mathcal{C}_p(t))$ and its length $p=l_N(t)$. For any bounded $f:\mathbb{R}^d\to\mathbb{R}$, we now write

$$\mathbb{E}\left[e^{\sum_{i=1}^N f(V^i_t)}\big|\mathcal{C}(t)\right]=\mathbb{E}\left[\sum_{n\ge0}\frac{1}{n!}\sum_{i_1,..,i_n=1}^N f(V^{i_1}_t)....f(V^{i_n}_t)\big|\mathcal{C}(t)\right].$$

We now group the sum over $i_1, ...i_n$ depending on how many belong to each cluster of the partition, to find

$$\cdots = \mathbb{E}\left[\sum_{n\geq 0}\frac{1}{n!}\sum_{k_1+..+k_p=n}\binom{n}{k_1,k_2,...,k_p}\sum_{i_1,..i_{k_1}\in\mathcal{C}_1(t)}\cdots\sum_{i_{n-k_p+1},..i_n\in\mathcal{C}_p(t)} f(V_t^{i_1})....f(V_t^{i_n})\big|\mathcal{C}(t)\right].$$

We next use the conditional independence above, which generalises to any numbers of particles; the expectation of each term in the sum factors as

$$\mathbb{E}\left[f(V_t^{i_1})...f(V_t^{i_{k_1}})\big|\mathcal{C}(t)\right]\times\cdots\times\mathbb{E}\left[f(V_t^{i_{n-k_p+1}})...f(V_t^{i_n})\big|\mathcal{C}(t)\right].$$

We then cancel the $n!$ from the multinomial coefficient and reverse the order of summation to find a sum over $k_1, ...k_p$ without the constraint, and a factor $(k_1!)^{-1}...(k_p!)^{-1}$, and the overall sum is

$$\cdots = \sum_{k_1,...,k_p\geq 0}\frac{1}{k_1!k_2!\dots k_p!}\sum_{i_1^1,..,i_{k_1}^1\in\mathcal{C}_1(t)}\mathbb{E}\left[f(V_t^{i_1^1})...f(V_t^{i_{k_1}^1})\big|\mathcal{C}(t)\right]\cdots\sum_{i_1^p,..,i_{k_p}^p\in\mathcal{C}_p(t)}\mathbb{E}\left[f(V_t^{i_1^p})...f(V_t^{i_{k_p}^p})\big|\mathcal{C}(t)\right].$$

Performing the sum inside the expectation with indexes belong to each partition element recovers the exponential, so we conclude

$$\mathbb{E}\left[e^{\sum_{i=1}^N f(V_t^i)}\big|\mathcal{C}(t)\right] = \prod_{j=1}^p\mathbb{E}\left[e^{\sum_{i\in\mathcal{C}_j(t)} f(V_t^i)}\big|\mathcal{C}(t)\right].$$

Up to a logarithm, this is exactly the usual cluster expansion of the cumulant [168], conditioned on the (globally random) partition.

Gelation A particular question of interest is the formation of a macroscopic component $\mathcal{C}_1(t)$ of the partition, which we here call *gel* and represents the case when a positive fraction of the particles of the collisional dynamics have become correlated by a chain of interactions. This is already naturally connected to a random graph problem, by placing an edge between any particles which interact. In the special case $a = 1, b = 0$ of ($\mathrm{Q}_{a,b}$), which represents cutoff Maxwell molecules, the associated random graphs are Erdős-Réyni random graphs with parameters $N, 2t/N$, which are well-known [75] to undergo a phase transition at $t = \frac{1}{2}$; for $t \leq \frac{1}{2}$, there is (with high probability) no macroscopic component, whereas for $t > \frac{1}{2}$ there is (with high probability) a connected component containing a positive fraction of the vertexes, see also the comments in [97, Section 11.2]. In general, the rates at which edges appear are inhomogeneous in time and not Markovian, since they depend on previous collisions; we will see in Chapter 7 how these problems can be circumvented in the case ($\mathrm{Q}_{a,b}$), and the time of formation of a giant interaction cluster calculated exactly.

Stochastic Coagulents & Smoluchowski Equation When studying the random dynamics at the level of the clusters, the natural object to study will be the empirical measures λ^N_t associated to the clusters at time t, on a suitably rich metric space S, to be specified so that data $x \in S$ capture all of the particle histories of all the particles involved, from which we can recover either the most recent velocities $\mathbf{v}(x) = (v_1(x), ...v_{m(x)}(x))$, or the tree expansion of each particle by 'pruning' unnecessary data. The data of a cluster are roughly defined in Fig. 1.3 below, and defined formally in Section 7.2. Based on which particles collide in $[0, t]$, we form a partition $\{\mathcal{C}_i(t), i \le l_N(t)\}$ of $\{1, ..., N\}$, and associate clusters $x^1_t, ...x^{l_N(t)}_t \in S$. We study the associated empirical measures $\lambda^N_t = N^{-1}\sum_{i\le l_N(t)} \delta_{x^i_t}$, which we call *stochastic coagulants.*

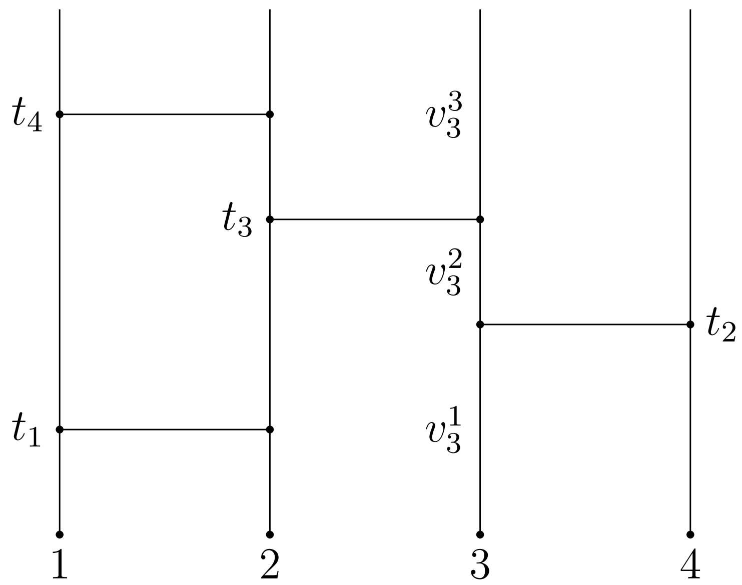

Figure 1.3: A prototypical cluster encoding the collision history of four particles, with the horizontal bars representing collisions at times $t_1, .., t_4$. v^i_3 are the velocities taken by particle 3 before the first collision and after subsequent collisions. The same applies for other particles, but these are omitted from the representation for legibility.

At the level of these coagulants, there are two possible changes when particles collide, depending on whether the two colliding particles already belong to the same part of the partition or not. In the first case, if the $i^{\text{th}}, j^{\text{th}}$ particles of a cluster x collide and scatter with angle σ, the cluster of type x will undergo a transformation to a new cluster of type $U(x, i, j, \sigma)$, with all the other clusters unchanged; this exactly corresponds to recollisions. In the latter case, if the i^{th} particle in a cluster of type x collides with the j^{th} particle of cluster y, both clusters are absorbed into a new cluster of type $M(x, y, i, j, \sigma)$, which contains the same collision history as x, y, as well as recording the new collision. These two types of transitions will be encoded mathematically by transition kernels $J^N \to 0$ on $J^N : S \to \mathcal{M}(S)$ and K on $S \times S \to \mathcal{M}(S)$ so that the rate of internal transition from x to y is $J^N(x, dy)$, and clusters of x, y merge to form a cluster of type z a rate

$2K(x,y,dz)/N$. We are therefore led to a prototypical equation of Smolochowski type in the hydrodynamic limit $N \to \infty$, which we express in the weak formulation

$$\langle f, \lambda_t \rangle = \langle f, \lambda_0 \rangle + \int_0^t \langle f, L(\lambda_s) \rangle ds \tag{Sm}$$

for all compactly supported, continuous $f : S \to \mathbb{R}$, where $L(\lambda)$ is defined by

$$\langle f, L(\lambda) \rangle = \int_{S^3} (f(z) - f(x) - f(y)) K(x,y,dz) \lambda(dx) \lambda(dy). \tag{1.39}$$

We note now that the solutions are now *sub*probability measures, since every coagulation event reduces the number of clusters by 1.

Conserved Quantities & The Flory Equation We already used above the observation that, corresponding to the conservation of particle number, momentum and energy for the collisional Kac dynamics, the quantities π_i given by

$$(\pi_1(x^i_t), \pi_2(x^i_t), \pi_3(x^i_t), \ldots, \pi_{d+2}(x^i_t)) = \sum_{j \in \mathcal{C}_i(t)} (1, |V^j_t|^2, V^j_t) \in [0,\infty)^2 \times \mathbb{R}^d$$

define conserved quantities for the coagulation dynamics: whenever a cluster of type x changes to type y by internal evolution, we have $\pi_i(x) = \pi_i(y)$, and if two clusters x, y merge to form a cluster of type z, then $\pi_i(z) = \pi_i(x) + \pi_i(y)$, for all $i = 1, 2, ..., d+2$. At the level of data in $x \in S$ encoding a particle history of m particles with the most recent velocities $(v_1, ..., v_m)$, π_1 extract the particle number m, π_2 extracts the total (kinetic) energy $\sum |v_i|^2$ and $\pi_3, ..., \pi_{d+2}$ extract the components of the total momentum $\sum_i v_i$. It follows that $\langle \pi_i, \lambda^N_t \rangle$ are pathwise constant for the stochastic coagulants λ^N_t and all i. On the other hand, it does *not* follow that $\langle \pi_i, \lambda_t \rangle$ are constant in time; unlike in the Boltzmann case, where we almost always restrict to solutions where the formally conserved quantity is genuinely conserved, global solutions to equations of the form (Sm) typically lose mass to infinity at a finite time, so that $\langle \pi_i, \lambda_t \rangle$ are decreasing [155, 156]. This corresponds to the formation of a giant cluster at the level of the random coagulation processes λ^N_t. This phenomenon depends strongly on the kernel; in cases where the kernel grows sublinearly in $\pi(x) + \pi(y)$ for a conserved quantity π, there is no gelation for a wide range of initial data [156, Theorem 2.1], while if the coagulation rate $\overline{K}(x,y)$ is bounded below by a product form $\varepsilon\pi(x)\pi(y), \varepsilon > 0$, then there is gelation at a finite time $t_{\mathrm{g}} \in (0,\infty)$ [156, Theorem 2.8]. Indeed, beyond this time, (Sm) is no longer the relevant equation, since it only captures the interaction between clusters in S and makes mass at infinity 'inert', whereas we must still account for the absorption of mass into the giant component. The insight [156] to write down an equation which does take this into account is to return to (1.38) to see that the rate at which two clusters x, y merge is given by the total mass $2\overline{K}(x,y) = 2K(x,y,S)$, which depends only on the conserved quantities $\pi_i(x), \pi_i(y), 1 \le i \le d+2$, and indeed as a bilinear form of the vectors $(\pi_i(x))_{1\le i\le d+2}, (\pi_i(y))_{1\le i\le d+2}$:

$$\overline{K}(x,y) = \sum_{i,j\le d+2} a_{ij}\pi_i(x)\pi_j(y), \qquad a_{ij} \in \mathbb{R}. \tag{1.40}$$

We then identify the mass, energy and momentum of the gel as $g_t^i = \langle \pi_i, \lambda_0 - \lambda_t\rangle$, and using linearity, the rate of absorption of a particle into the gel is $\int_S \overline{K}(x,y)(\lambda_0-\lambda_t)(dy)$. Taking this absorption into account, we modify the coagulation operator to the time-dependent version

$$\langle f, L_{\mathrm{g}}(\lambda_t)\rangle := \langle f, L(\lambda_t)\rangle - \int_S f(x)\overline{K}(x,y)\lambda_t(dx)(\lambda_0-\lambda_t)(dy)$$

and we find a prototypical equation of the type studied by Flory [199]: for all $f \in C_c(S)$,

$$\langle f, \lambda_t\rangle = \langle f, \lambda_0\rangle + \int_0^t \langle f, L_{\mathrm{g}}(\lambda_s)\rangle ds. \tag{Fl}$$

We note that the additional terms only make a difference once $\langle \pi_i, \lambda_t\rangle$ are nonconstant, which corresponds to the emergence of a gel at the level of the limiting equation. We will see in Theorem 5, see also Theorems 7.2 - 7.3 that this equation is globally well-posed, and is the limit of the stochastic coagulants λ_t^N, globally in time.

Previous Literature We will discuss the literature relevant to the study of coagulation problems of the form (Sm, Fl) in detail in Chapter 7. The application to interaction was developped by Patterson, Simonella and Wagner [161, 162] and Gabrielov [96]. In [162], the cases (HS, $Q_{a,b}$) for the underlying collision kernel are considered, and a recursive expansion for the limiting equation is found, as well as upper and lower (not matching) bounds for the gelation times. This paper also introduces the special kernels ($Q_{a,b}$) as a phenomenological proxy to the harder case (HS), and conjectured that gelation occurs strictly before the mean-free time, which we verify for our kernels. We build particularly on the work [156] of Norris, which allows coagulation systems where particles can merge in more than one way, which is relevant for the application to collision clusters.

Characterisation of Gelation Let us remark that we have already encountered two notions of gel and gelation, which could *a priori* differ and give rise to different gelation times. At the level of the particle system, one can study the phase transition where the size of the largest particle goes from size $\ll N$ to a size comparable to N [134]. At the level of the limiting equation, gelation refers to the point where the solution to the Smoluchowski or Flory equation $(\lambda_t)_{t\ge 0}$ fails to conserve the total particle mass, which is known to be related [134, 156] to the divergence of the second moment $\langle \pi_1^2 + \pi_2^2, \lambda_t\rangle$ of the particle masses and energies at the level of the limit equation. We will see that, for our model, these three coincide, so that gelation is equivalently characterised by any of the formation of a giant component in the particle system, the failure of the limiting equation to conserve $\langle \pi_i, \lambda_t\rangle$, or the divergence of the second moment.

Bilinear Coagulation Equations Let us mention our strategy in investigating the equations (Sm, Fl) obtained above in the case of the interaction clusters. We already stressed in (1.40) that the rate of merger between two clusters x, y is given by a bilinear form of the vectors $(\pi_i(x))_{1\le i\le d+2}, (\pi_i(y))_{1\le i\le d+2}$, thanks to the nice algebraic properties of the kernels ($\mathrm{Q}_{a,b}$), and the details of x, y only enter into the kernel through changing the distribution of the new cluster z. This property will allow us to investigate the stochastic coagulants via (Markovian) random graphs, as already mentioned in the context of gelation, and the random graphs are of a kind studied in detail by Bollobás and coauthors [28]. We can therefore play ideas at the level of the coagulation equations (Sm, Fl) from the literature [156] against techniques from random graphs [28]. This connection generalises the case mentioned above (Maxwell molecules $\leftrightarrow$ sparse Erdős-Réyni Graphs $\leftrightarrow$ multiplicative coagulation kernel, see the discussion in Chapter 7). When studying this more general setting in Chapter 7, we will introduce the concept of a bilinear coagulation system, where this property is made an axiom and sufficient conditions are imposed to prevent pathologies, and study results in this generality. All the study of the equations written above can be recovered from this study as a special case.

Recovering the Boltzmann Equation from the Coagulation Equation Having discussed the processes at the level of the coagulation systems, let us conclude by returning to the Boltzmann equation. For any measure λ on S integrating π_1, we can define a measure $\mathcal{F}\lambda$ on $\mathbb{R}^d$ by 'forgetting' the cluster structure and extracting the most recent velocities $\mathbf{v}(x) = (v_1(x), ..., v_{m(x)}(x))$ from each cluster x. Formally, for a bounded function $f : \mathbb{R}^d \to \mathbb{R}$, we set

$$\mathcal{F}^\star f(x) := \sum_{i=1}^{m(x)} f(v_i(x))$$

and define a measure $\mathcal{F}\lambda$ by specifying, for all bounded $f : \mathbb{R}^d \to \mathbb{R}$,

$$\langle f, \mathcal{F}\lambda\rangle = \int_S (\mathcal{F}^\star f)(x)\lambda(dx)$$

which produces a finite measure, since $|(\mathcal{F}^\star f)(x)| \le \|f\|_\infty \pi_1(x)$, which is integrable by hypothesis. To link the coagulations (Sm, Fl) to (BE), let us observe that whenever x, y merge to form z via a collision with velocities v, v_*, we have

$$\mathcal{F}^\star f(z) - \mathcal{F}^\star f(x) - \mathcal{F}^\star f(y) = f(v') + f(v_*') - f(v) - f(v_*)$$

since only the two most recent velocities v, v_* have been replaced by v', v_*'. It follows that, formally,

$$\langle \mathcal{F}^\star f, L(\lambda)\rangle = \langle f, Q(\mathcal{F}\lambda)\rangle.$$

We might therefore hope that, for a solution λ_t to the coagulation equation, we recover the corresponding solution μ_t to (BE) through $\mu_t = \mathcal{F}\lambda_t$. This would then amount to a much finer sum than (1.37), since we break up the sum, when adding a new branch at

$t_1 < t$ not only over the past history (on $[0, t_1)$) but also over the future collision history (up to time t) of the new particle. We will see that this indeed recovers the spatially homogeneous Boltzmann equation (BE), but only up to the gelation time t_{g}.

1.2.6 Relaxation to Equilibrium

Let us finally mention that another topic which we will encounter, but to which this thesis does not contribute, is the speed of relaxation of the kinetic equations and the particle systems to their equilibria. Having quantitative, rather than qualitative estimates, is important here; we refer to [191, Section 2.5]. As mentioned concerning the long-time propagation of chaos, Boltzmann acknowledged that the (spatially inhomogeneous) Boltzmann equation fails at very large times, depending on N. In particular, the behaviour of the Boltzmann equation in the limit $t \to \infty$ is irrelevant for a physical (spatially inhomogeneous) system, and what is interesting is to show the Boltzmann equation is 'reasonably close to' equilibrium on time scales $t \leq t_N$ which are short enough for the Boltzmann equation to remain a good description of the microscopic dynamics for the physical system, where N is large and fixed.

Relaxation of the Boltzmann and Landau Equation We already mentioned, in the context of the H-Theorem (H) above, a long series of works regarding the convergence to equilibrium via functional inequalities which relate the entropy and the entropy dissipation, with positive results obtained by Toscani and Villani [181, 182, 190]; a similar programme was carried out for the Landau equation by Desvillettes and Villani [59]. An entirely different approach is based on finding a spectral gap for the linearised operator $2Q(\cdot, \gamma)$, for $\gamma \in \mathcal{S}$ the standard Gaussian (1.23). This programme goes back to Carlen and co-authors in the case of Maxwell molecules, where they showed that the exponential rate of convergence is determined by the spectral gap in the (much smaller) space $L^2(\mathbb{R}^d, \gamma^{-1}(v)dv)$ in the case of cutoff Maxwell molecules (GMM). For the case of hard spheres and cutoff hard potentials (HS, CHP_K), Grad [102] proved the existence of a spectral gap, and the same for soft potentials by Caflisch [35] and Golse and Poupaud [100]. Arkeryd [12] proved exponential convergence with non-constructive constants, and Mouhot [148] proved a constructive argument, with the rate again determined by the spectral gap in $L^2(\mathbb{R}^d, \gamma^{-1}(v)dv)$. These results were extended for (GMM, tMM, HS) to a differential stability results by Mischler and Mouhot [142], which (in the case (HS)) lie at the heart of our long-time propagation of chaos results in Chapter 3 and are recalled in Section 3.2. In the case of hard potentials, exponential convergence to equilibrium was proven for the Boltzmann equation (BE) by Tristani [183], adapting the ideas of [148] to the noncutoff case, and for the Landau equation by Carrapatoso [41].

Relaxation for the Particle System A different problem related to relaxation is to ask for the relaxation of the many-particle process in $\mathbb{S}_N$ to its equilibrium, which is the uniform distribution (Haussdorf measure) on this space, and further to quantify the dependence of these relaxation rates on the number of particles N, an idea which goes back to Kac's original work [122]. For each fixed N, the chain is dissipative, and Kac proposed that N-uniform relaxation rates for the Kac process could be propagated to obtain relaxation rates for the Boltzmann equation in the limit $N \to \infty$. Janvresse [118] proved the existence of an N-uniform spectral gap in $L^2(\mathbb{S}_N)$ for Kac's caricature in $d = 1$, see also [36, 39, 137] and [40] for the physical (energy and momentum conserving) systems which we consider here; we will recall this result in Proposition 3.27 when discussing the long-time behaviour of the Kac process. However, this itself is not sufficient to get relaxation on N-uniform time scales, since we must start from initial densities F_N which scale like $\|F_N\|_{L^2(\mathbb{S}_N)} \geq C^N$ for chaotic initial data, and so one would still need $t \sim N$ to get good convergence. Carlen and coauthors [38] proposed the use of relative entropy, and functional inequalities for entropy and entropy dissipation at the level of the Markov chains, which have better tensorisation properties, and which were explored in [37], but the best linear inequality between entropy productio and entropy has a constant on the order $\mathcal{O}(N^{-1})$, [190, 67, 68, 69], which matches the absence of a linear inequality in Cercignani's conjecture. Mischler and Mouhot [142] proved the N-uniform relaxation for the cases (GMM, tMM, HS) measured in Wasserstein distances on $\mathcal{P}(\mathbb{S}_N)$ with the correct tensorisation properties, with explicit and N-uniform estimates but a very slow rate in t [142, Theorems 5.2, 6.2], and Rousset [167] proved N-uniform convergence in Wasserstein distance, with arbitrarily fast polynomial rates in t.

Let us reiterate that this thesis makes no contribution to the study of these problems, but that we will encounter the problems of relaxation (both for the particle system and the kinetic equation) in Chapter 3 in deriving the propagation of chaos in large time. We remarked in Subsection 1.2.1 above that one can seek either pointwise-in-time or pathwise, locally-uniform in time estimates for chaoticity; finding good time-dependence for these results will strongly use the differential stability. The proof of the local uniform estimate in particular uses the full strength of exponential convergence to keep a small loss when 'bootstrapping' to make the time-dependence as slow as one likes (see Theorem 1). However, we will also see that the relaxation of the particle system *forbids* improving the local uniform estimate to a uniform estimate, no matter how slow the relaxation.

1.3 Statements of Results

We now summarise our results.

1. Hard Spheres We recall that it is known that there is a unique energy-conserving solution to (BE) starting at any $\mu_0 \in \mathcal{S}$ in the case of hard spheres (HS), which we write as $(\phi_t(\mu_0))_{t\geq 0}$. In this case, our result is as follows.

Theorem 1. *Let $(\mu_t^N)_{t\geq 0} \subset \mathcal{S}$ be a N-particle Kac process for the hard spheres kernel (HS) and $\mu_0 \in \mathcal{S}$, with deterministic initial data μ_0^N having a p^{th} moment bound, $p > 2$. Then, for some metric W_1 on $\mathcal{S}$, which is equivalent to weak convergence, we have the following convergences.*

i). We have convergence for fixed times, uniformly in $t \geq 0$:

$$\sup_{t\geq 0} \mathbb{E}\left[W_1(\mu_t^N, \phi_t(\mu_0))\right] \leq C\left(N^{-\varepsilon_1} + W_1(\mu_0^N, \mu_0)^{\varepsilon_2}\right)$$

where $\varepsilon_1, \varepsilon_2 > 0$, $\varepsilon_1 < \frac{1}{d}$ depend on p, and C depends on the p^{th} moments of μ_0^N, μ_0. By making p large enough, the exponent ε_1 can be made arbitrarily close to the optimal value $\frac{1}{d}$.

ii). We have uniform convergence on compact time intervals in probability: for any time-horizon $t_{\text{fin}} \in [0, \infty)$, we also have

$$\mathbb{E}\left[\sup_{t\leq t_{\text{fin}}} W_1(\mu_t^N, \phi_t(\mu_0))\right] \leq C((1 + t_{\text{fin}})^{\alpha} N^{-\varepsilon_1} + W_1(\mu_0^N, \mu_0)^{\varepsilon_2})$$

where, again, the exponents $\alpha, \varepsilon_1, \varepsilon_2 > 0$ depend only on p, and C depends on p and on the p^{th} moments of the initial data μ_0^N, μ_0. Further, by making p large, α can be made arbitrarily small, while keeping ε_1 bounded away from 0. However, α cannot be taken to be 0, so the result would be false with $t_{\text{fin}} = \infty$.

iii). We have the uniform-in-time Hölder stability of the solutions $\phi_t(\mu)$ in the metric W_1: for any $p > 2$ there exists $\zeta > 0$ such that, whenever μ, ν have finite p^{th} moments, we have

$$\sup_{t\geq 0} W_1\left(\phi_t(\mu), \phi_t(\nu)\right) \leq C W_1\left(\mu, \nu\right)^{\zeta}$$

for some $\zeta > 0$ depending only on p, with C depending on the p^{th} moments of μ, ν.

2. The Boltzmann Equation with Noncutoff Hard Potentials Our second main result concerns the Boltzmann equation in the case of noncutoff hard potentials (NCHP).

Theorem 2. *Let B be a kernel of the form (NCHP). Then for $p<\infty$ sufficiently large, depending only on B, and $K\in[K_0(B,p),\infty]$ sufficiently large, depending only on B,p, and for a semimetric w_p on the space of measures with $p+2$ moments, defined as the optimal transportation cost for a particular function $d_p:\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}$ and equivalent to weak convergence plus convergence of the $(p+2)^{th}$ moment, the following holds.*

i). The energy-conserving solutions $(\mu_t)_{t\geq 0}$ starting at any $\mu_0\in\mathcal{S}$ are exactly $\mu_t=\mathrm{Law}(V_t)$, where $(V_t)_{t\geq 0}$ is a Boltzmann process started from V_0 with $\mathrm{Law}(V_0)=\mu_0$ and such that $\mathrm{Law}(V_t)\in\mathcal{S}$ for all $t\geq 0$.

ii). For any two solutions $(\mu_t)_{t\geq 0},(\nu_t)_{t\geq 0}\subset\mathcal{S}$ to (BE, BE_K) respectively, such that μ_0,ν_0 have $p+2$, $l=p+2+\gamma$ moments, there exists a coupling $(V_t,\tilde{V}_t)_{t\geq 0}$ of Boltzmann processes, with $\mathrm{Law}(V_t)=\mu_t,\mathrm{Law}(\tilde{V}_t)=\nu_t$, such that

$$\mathbb{E}\left[d_p(V_t,\tilde{V}_t)\right]\leq C_1e^{C_1t}\left(w_p(\mu_0,\nu_0)+C_2tK^{-\alpha}\right)$$

for some C_1, depending on B and on the p^{th} moments of μ_0,ν_0, some C_2 depending on the l^{th} moment of ν_0, and some $\alpha>0$ depending only on B. In the case $K=\infty$ of coupling noncutoff solutions, we can relax the requirement on ν_0 to allow only $p+2$ moments.

iii). As a consequence of the previous point, the noncutoff Boltzmann equation has a unique energy conserving solution $(\mu_t)_{t\geq 0}$ as soon as the initial data μ_0 has $p+2$ moments, which is the limit in w_p of the solutions to the cutoff equation (BE_K), with an explicit rate if the initial data have l moments. Moreover, the solution map $\phi_t:\mathcal{S}^{p+2}\to\mathcal{S}^{p+2}$ corresponding to the Boltzmann equation (BE) is Lipschitz-continuous with respect to the semimetric w_p on sets with bounded p^{th} moments.

iv). Concerning the Kac process, we have the following convergence of cutoff processes to a noncutoff process. For some sufficiently large $2+p<q<\infty$ depending on p, if $(\mu^N_t)_{t\geq 0}\subset\mathcal{S}$ and is a (noncutoff) Kac process on N particles and $\mu_0^{N,K}\in\mathcal{S}$ are given, with an almost sure moment bound on the q^{th} moments of both $\mu_0^N,\mu_0^{N,K}$, then $(\mu^N_t)_{t\geq 0}$ can be approximated, uniformly in N and pathwise-uniformly on compact time intervals with respect to the cost w_p, by cutoff Kac processes $(\mu_t^{N,K})_{t\geq 0}$ starting at $\mu_0^{N,K}$, with an error rate, for any $t_{\mathrm{fin}}<\infty$,

$$\mathbb{E}\left[\sup_{t\leq t_{\mathrm{fin}}}w_p(\mu^N_t,\mu^{N,K}_t)\right]\leq Ce^{Ct_{\mathrm{fin}}}\left(w_p(\mu_0^N,\mu_0^{N,K})+K^{-\alpha}+N^{-1/2}\right)$$

where $\alpha>0$ is as above, where C depends on the bound for the q^{th} moments of $\mu_0^N,\mu_0^{N,K}$. The same holds if $K=\infty$, now omitting the second term, so that noncutoff Kac processes can be coupled in the same way.

v). Finally, if $(\mu_t^N)_{t\geq 0}$ *is above and* $\mu_0 \in \mathcal{S}$ *has* $p+2$ *moments, we have the convergence, pathwise-uniformly on compact time intervals in the cost* w_p*,*

$$\mathbb{E}\left[\sup_{t\leq t_{\text{fin}}} w_p(\mu_t^N, \phi_t(\mu_0))\right] \leq Ce^{Ct_{\text{fin}}}\left((\log N)^{-\alpha} + \mathbb{E}\left[w_p(\mu_0^N, \mu_0)\right]\right)$$

where $\phi_t(\mu_0)$ *is the unique energy-conserving solution to (BE), as in point ii), for a new exponent* $\alpha > 0$ *depending on* B*, and where* C *depends on the* q^{th} *moment bound for* μ_0^N *and the* p^{th} *moment bound for* μ_0*.*

3. The Landau Equation with Hard Potentials Our next result adapts points i-iii). of the previous result to the case of the Landau equation with hard potentials in dimension $d = 3$. We recall that a Landau processes is any solution $(V_t)_{t\geq 0}$ to the stochastic differential equation (stLE).

Theorem 3. *Fix* $\gamma \in (0, 1]$*, and consider the corresponding Landau equation (LE) in dimension* $d = 3$*. For all* $p > 2$*, there exists a semimetric* $w_{p,1}$ *on* $\mathcal{S}^p(\mathbb{R}^3)$*, defined as the optimal transport cost for a well-chosen function* $d_{p,1} : \mathbb{R}^3 \times \mathbb{R}^3 \to [0, \infty)$*, and such that* $w_{p,1}$ *is equivalent to weak convergence plus convergence of the* p^{th} *moment, such that the following hold.*

i). If $\mu \in L^1_{loc}([0, \infty), \mathcal{S}^{2+\gamma}(\mathbb{R}^3))$ *is a solution to the Landau equation (LE), then there exists a Landau process* $(V_t)_{t\geq 0}$ *with* $\mu_t = \mathrm{Law}(V_t)$ *for all* $t \geq 0$*. Conversely, if* $(V_t)_{t\geq 0}$ *is a Landau process such that* $\mu_t = \mathrm{Law}(V_t)$ *has* $\mu \in L^1_{loc}([0, \infty), \mathcal{S}^{2+\gamma}(\mathbb{R}^3))$*, then* $(\mu_t)_{t\geq 0}$ *is a solution to the Landau equation (LE).*

ii). If $(\mu_t)_{t\geq 0}, (\nu_t)_{t\geq 0}$ *are two solutions to (LE) such that* μ_0, ν_0 *both have finite* p^{th} *moments,* $p > 2$*, then there exists a coupling* $(V_t, \tilde{V}_t)_{t\geq 0}$ *of Landau processes, with* $\mathrm{Law}(V_t) = \mu_t, \mathrm{Law}(\tilde{V}_t) = \nu_t$ *satisfying, for all* $t \geq 0$*,*

$$\mathbb{E}\left[d_{p,1}(V_t, \tilde{V}_t)\right] \leq e^{C(1+t)} w_{p,1}(\mu_0, \nu_0)$$

for some C *depending on the* p^{th} *moments of* μ_0, ν_0*.*

iii). As a consequence of the previous point, the Landau equation has a unique solution $(\mu_t)_{t\geq 0}$ *as soon as the initial data* μ_0 *has any* $p > 2$ *moments, and the solution map* $\phi_t^{\mathrm{L}} : \mathcal{S}^p(\mathbb{R}^3) \to \mathcal{S}^p(\mathbb{R}^3)$ *corresponding to (LE) is Lipschitz continuous with respect to the semimetric* $w_{p,1}$ *on sets with bounded* p^{th} *moments.*

We also prove the following, which are specific to the Landau case.

iv). For any $\mu_0 \in \mathcal{S}(\mathbb{R}^3)$*, there exists a Landau process* $(V_t)_{t\geq 0}$ *with* $\mu_0 = \mathrm{Law}(V_0)$ *such that* $\mu_t = \mathrm{Law}(V_t)$ *satisfies* $\mu \in L^1_{loc}([0, \infty), \mathcal{S}^{2+\gamma}(\mathbb{R}^3))$*. In particular,* $(\mu_t)_{t\geq 0}$ *is a weak solution to (LE) starting at* μ_0*.*

v). Let $(\mu_t)_{t\geq 0} \subset \mathcal{S}(\mathbb{R}^3)$ be any weak solution to (LE). Then, for all $t>0$, μ_t admits an analytic density f_t with finite entropy, and all weighted Sobolev norms are bounded, uniformly away from $t=0$: for any $k, s \geq 0$ and $t_0>0$,

$$\sup_{t\geq t_0} \int_{\mathbb{R}^3} |D^k f_t|^2(v)(1+|v|^s)dv < \infty$$

which implies that f_t belongs to the class of Schwarz functions: the derivatives of any order decay more quickly than any inverse polynomial.

4. Large Deviations of the Kac Process Chapter 6 is dedicated to the investigation of dynamical large deviations. We summarise as follows.

Theorem 4. *Fix a time horizon $t_{\text{fin}} \in (0,\infty)$, and consider the Kac processes $\mu^N_\bullet = (\mu^N_t)_{0\leq t\leq t_{\text{fin}}}$ with their associated empirical fluxes $w^N = w^N_{t_{\text{fin}}}$, with kernel either regularised hard spheres (rHS) or cutoff Maxwell molecules (GMM), where particles are initially sampled independently from $\mu_0^\star \in \mathcal{S}$. Under general hypotheses on $\mu_0^\star$, which includes the case where $\mu_0^\star = \gamma$ is the equilibrium distribution, we identify a function $\mathcal{I}$, analagous to those found in the literature and which vanishes if, and only if, $\mu_\bullet$ is a solution to the Boltzmann equation, starting at $\mu_0^\star$, and*

$$w(dt, dv, dv_*, d\sigma) = B(v-v_*, \sigma)\mu_t(dv)\mu_t(dv_*)d\sigma.$$

For this function $\mathcal{I}$, and for suitable topological spaces $\mathcal{D}$ on paths and $\mathcal{M}(E)$ on fluxes, we have the following properties.

i). For any closed set $\mathcal{A}$ of $\mathcal{D}\times\mathcal{M}(E)$, we have

$$\limsup_N \frac{1}{N}\log \mathbb{P}\left((\mu^N_\bullet, w^N)\in\mathcal{A}\right) \leq -\inf\left\{\mathcal{I}(\mu_\bullet, w) : (\mu_\bullet, w)\in\mathcal{A}\right\}.$$

ii). For any open set $\mathcal{U}$ of $\mathcal{D}\times\mathcal{M}(E)$, we have

$$\liminf_N \frac{1}{N}\log \mathbb{P}\left((\mu^N_\bullet, w^N)\in\mathcal{U}\right) \geq -\inf\left\{\mathcal{I}(\mu_\bullet, w) : (\mu_\bullet, w)\in\mathcal{U}\cap\mathcal{R}\right\}$$

where $\mathcal{R} = \{(\mu_\bullet, w)\in\mathcal{D}\times\mathcal{M}(E) : \langle 1+|v|^2+|v_*|^2, w\rangle < \infty\}$.

iii). However, the previous item is false without the restriction to $\mathcal{R}$; energy non-conserving solutions to the Boltzmann equation appear as large deviations, but strictly more rarely than predicted by the candidate rate function $\mathcal{I}$.

iv). In the case $\mu_0^\star = \gamma$, the candidate rate function $\mathcal{I}$ is symmetric under a time-reversal on $\mathcal{R}$, from which we recover the first part of the Boltzmann H-Theorem: if $(\mu_t)_{0\leq t\leq t_{\text{fin}}}$ is a solution to the Boltzmann equation with $\mu_\bullet \in L^1([0,t_{\text{fin}}], \mathcal{S}^3)$ such that each μ_t has a nonzero density $f_t>0$, then

$$H(\mu_{t_{\text{fin}}}|\gamma) + \int_0^{t_{\text{fin}}} D(\mu_s)ds = H(\mu_0|\gamma)$$

where $D(\mu_s) = D(f_s) \geq 0$ is the entropy dissipation.

5. Interaction Clusters Our final result concerns the interaction clusters for the Kac process and their limiting behaviour.

Theorem 5. *Consider the stochastic coagulants λ^N_t corresponding to Kac process μ^N_t for quadratic kernels ($Q_{a,b}$). Let us suppose that the initial data μ^N_0 are such that $\mu^N_0 \to \mu_0$ weakly in probability and $\langle |v|^4, \mu^N_0\rangle \to \langle |v|^4, \mu_0\rangle$ in probability, for some $\mu_0 \in \mathcal{S}^6$ which is symmetric under the pushforward by $v \mapsto -v$. We consider (Fl) on a metric space S to be specified, such that $x \in S$ contains all the information of the collision history (see Fig. 1.3) and form λ_0 by pushing μ_0 forward by the map taking v to a single particle with velocity v and no collision history. Then the following hold.*

i). The stochastic coagulants λ^N_t converge to the unique solution $(\lambda_t)_{t\geq 0}$ to the Flory-type equation (Fl) in a metric inducing the weak topology, uniformly in time. Moreover, if we let x^1_t be the largest cluster by particle number in λ^N_t and set $g^N_t = N^{-1}(\pi_i(x^1_t))_{i=1}^{d+2}$ then $g^N_t \to g_t$ in probability, uniformly in time, where $g_t = (M_t, E_t, 0)$ is the mass, energy and momentum of the gel $g_t = (\langle \pi_i, \lambda_0 - \lambda_t\rangle)_{i=1}^{d+2}$.

ii). The limiting equation undergoes a phase transition at the gelation time

$$t_{\mathrm{g}} = \left(a + 2b + \sqrt{(a+2b)^2 + 4b^2(\Lambda_4(\mu_0) - 1)}\right)^{-1}.$$

That is, for $t \leq t_{\mathrm{g}}$, we have $M_t = E_t = 0$, but $M_t > 0, E_t > 0$. Moreover, except in the cases where $b = 0$ (Maxwell molecules) or $\mu_0(|v| = 1) = 1$, we have that the gelation time is strictly smaller than the mean free time, and in the special cases we have equality.

iii). At the level of the particle system, the previous two points imply that $\pi_1(x^1_t)$ is $o(N)$ with high probability if $t \leq t_{\mathrm{g}}$, and on the order N with high probability if $t > t_{\mathrm{g}}$.

iv). The phase transition is continuous and first order. That is, M_t, E_t are continuous at $t = t_{\mathrm{g}}$, but have a strictly positive right-derivative at the critical time. Furthermore, except in the same special cases as item ii), we have $\lim_{t\downarrow t_{\mathrm{g}}} \frac{E_t}{M_t} > 1$, so that the formation of a gel is driven by the fast particles.

v). The second moment

$$\mathcal{E}(t) = \langle (\pi_1 + \pi_2)^2, \lambda_t\rangle$$

is finite and continuous, except for a divergence at the critical time $t = t_{\mathrm{g}}$, and $\mathcal{E}(t) \to \infty$ as $t \to t_{\mathrm{g}}$. In particular, the formation of a giant component exactly coincides with the unique divergence of the second moment.

vi). The limit solution λ_t is supported on $x \in S$ which encode interaction histories without cycles.

vii). Define, for $t \geq 0$, a measure μ_t on $\mathbb{R}^d$ by removing the structure from each cluster:

$$\mu_t = \mathcal{F}\lambda_t.$$

Then on $[0, t_\mathrm{g}]$, μ_t is an energy-conserving solution to the Boltzmann equation, but for $t > t_\mathrm{g}$, the total mass $\mu_t(\mathbb{R}^d) < 1$ and μ_t is not even a probability measure.

Moreover, aside from items vi)-vii), the above properties all generalise to coagulation equations with the bilinear form we derived above.

1.4 Outline of the Thesis

We conclude with an outline of the thesis. We will make the specifications relevant for the declaration here.

ii). Chapter 2 is a 'Technical Toolbox' in which we review some concepts and results of frequent use. We introduce the definitions of the distances on measures which we will use in the theorems above and give a careful study of Kac's notion of chaoticity. We also state and prove several results on moment estimates for both the Kac process and the kinetic equations. Since this is a review chapter, almost none of the content is novel or due to the author (the definitions of some 'tailor made' optimal transport problems $w_p, w_{p,\varepsilon}$, and Proposition 2.10 iv).

iii). Chapter 3 is dedicated to the proof of Theorem 1, and corresponds to the paper [111] by the author. The argument is based on combining some ideas of the work of Mischler and Mouhot [142] with some pathwise techniques of Norris [157].

The idea of using stability estimates in a pathwise sense to study the Kac process was first investigated by the author for the Part III Essay submitted to the University of Cambridge for the Master of Mathematics Degree in 2017, which proved versions of items i-ii) with worse dependence in N and in the time horizon t_fin in item ii), and requiring $\mu_0 = \mu_0^N$. The work presented in this thesis builds on that previous work but was all developed subsequently, which required incorporating some more precise ideas to obtain the (almost sharp) asymptotics in the N, t_fin dependencies presented here, and to include item iii) which allows any μ_0. As a result of the incorporation of these new ideas, there is negligible overlap between the work previously submitted and this current thesis.

iv). Chapter 4 studies the Boltzmann equation and Kac process in the case of noncutoff hard potentials (NCHP) to prove Theorem 2. The key point is the application of a Tanaka coupling in a well-chosen distance produces an advantageous cancellation, either at the level of Boltzmann processes or at the level of the many-particle system.

This chapter corresponds to the paper [112] by the author, but has been refined and extended for this thesis.

v). Chapter 5 continues with the ideas of Chapter 4, now in the context of the hard-potential Landau equation to prove Theorem 3. The key idea remains the same, although we now work only with Landau processes, and substantial further refinement is possible, since we work with a differential operator $\mathcal{L}_\mathrm{L}$ rather than the pseudodifferential operator $\mathcal{L}_\mathrm{B}$, and we obtain some further applications which are specific to the Landau equation. This chapter is based on the work [90] of the author jointly with Prof. Nicolas Fournier, and the parts of this paper corresponding to this chapter were all the result of collaboration between the authors. Some of the presentation has been changed, to fit the overall probabilistic theme of the thesis (for instance, working with stochastic processes, rather than their associated differential equations).

vi). Chapter 6 considers the dynamical large deviations of the Kac process, leading to a precise statement and proof of Theorem 4. This chapter corresponds closely to the paper [113] by the author, although some elements have been added since the paper appeared.

vii). Chapter 7 is dedicated to the study of bilinear coagulations equations in general, which generalise the prototype Flory equation (Fl) above. We carefully define bilinear coagulation equations and the associated stochastic processes, and state and prove results which generalise Theorem 5. We also show how Theorem 5 is then obtained as an application of these results. This work was originally undertaken in collaboration with Robert Patterson [114], to which both of the authors contributed.

As far as possible, we have tried to keep each of Chapters 3 - 7 self-contained, so that these chapters may be read largely independently of each other; results which are (heavily) used in more than one chapter are introduced in sufficient generality in Chapter 2.

1.4.1 Some Approaches of this Thesis

Top-Down vs Bottom Up Let us remark on the two flavours of approaches which run through the works discussed above, and correspondingly through this thesis. When considering results on the Kac process and the Boltzmann equation, it is *either* possible to start with results on the Boltzmann equation in cases where its properties are well-understood, and apply them to the Kac process, *or* to start with properties at the level of the Kac processes $(\mu^N_t)_{t\geq 0}$, and carefully propagate them to the Boltzmann equation, which is appropriate in cases where the desired property for the Boltzmann equation is not well-understood and may be much harder to prove than for the Kac process. We borrow some terminology from the introduction of [142], and call these 'top down' or 'bottom up' approaches respectively. The latter was the approach of Kac [122] in proposing to study relaxation of the Boltzmann equation through the particle system, see the discussion in Section 1.2.6 above, at a time when it was plausible that the (linear, high-dimensional) Markov process would be easier to study than the (nonlinear, infinite-dimensional) kinetic differential equation. To some extent, the opposite has been true; many works were proven on the relaxation of the Boltzmann equation, in particular, before the N-uniform relaxation rates were obtained by Mischler and Mouhot [142]. Let us also mention that the same paper [142] adopted a 'top-down' approach to the propagation of chaos, which used functional properties of the semigroup ϕ_t associated to the Boltzmann equation to obtain propagation of chaos.

We will see both approaches in this thesis. Theorem 1, corresponding to Chapter 3, explores the 'top-down' ideas of [142] in a pathwise framework. However, we will also see some 'bottom up' approaches, notably in Chapter 4 when discussing stability - it is (much) easier to define a coupling at the level of the Kac process than at the level of the Boltzmann equation, and this is the original proof given by the author in [112] - and in Chapter 6, where our analysis of large deviations leads to a rederivation and a probabilistic meaning of (H). In Chapter 7, when we analyse coagulation equations with the same bilinear form as derived for the Kac interaction clusters above, it is particularly profitable to play these two types of idea against each other, which allows us to play ideas from the literature concerning coagulation equations [156] against ideas from the theory of random graphs [28].

Pathwise Analysis Let us also remark that the calculations of this thesis are almost exclusively based on *pathwise analysis*, working fairly directly with $\mu_t^N, \mathcal{V}_t^N$, rather than with their laws $\mathcal{L}_t^N = \text{Law}(\mathcal{V}_t^N)$ and "analytic co-travellers" (transition probabilities, entropy and entropy dissipation of $\mathcal{L}_t^N$, ...). To some extent, the latter can always be recovered, since the law $\text{Law}((\mathcal{V}_t^N)_{t\ge 0})$ contains each $\mathcal{L}_t^N$ as its single-time marginals, but the inverse is definitely false: when we study local uniform estimates

$$\mathbb{P}\left(\sup_{t\le t_{\text{fin}}} \mathcal{W}_{1,1}(\mu_t^N, \mu_t) < \varepsilon\right)$$

in Theorems 1, 2 in the context of propagation of chaos, and in Theorem 4 in the context of large deviations, we cannot recover the desired estimates (just) from knowing all single-time marginals $(\mathcal{L}_t^N)_{t\ge 0}$.

Chapter 2

Technical Toolbox

We will now give an introduction to some technical tools, which will be in frequent use throughout the thesis; except where stated otherwise, none of this is new. Since variants of many of the same ideas appear repeatedly, we will will introduce all of the concepts together to avoid repeated and very similar digressions to cover the different cases in each chapter.

This chapter is structured as follows.

i). First, Section 2.1 introduces some spaces of measures and probability measures, as well as several families of Wasserstein-type distances on probability measures. We will discuss the relationships between the distances, and some (quantitative) equivalences.

ii). In Section 2.2 we carefully formulate Boltzmann's property of chaoticity described in the introduction. We will show that chaoticity follows from the convergence of the empirical measure, which justifies the description of Theorems x, y as proving the propagation of chaos for the respective Kac models.

iii). Section 2.3 introduces a natural scheme for the initial values of the Kac process with normalised energy and momentum.

iv). Section 2.4 introduces two equivalent parameterisations of possible jumps $(v, v_*) \mapsto (v', v'_*)$.

v). In Section 2.5, we review some moment estimates for both the Boltzmann and Landau equations, and for Kac's interacting particle system. These results are (largely) classical, going back as far as Pozvner [163], Elmroth [72] and Desvillettes [53].

2.1 Distances on Probability Measures

For the quantitative results throughout this thesis, we rely on good choices of metrics. In Chapter 3, the choice of metric will be important for the style of proof; in Chapters 4, 5, we will rely crucially on a 'tailor-made' family of optimal transport costs which allow us to exploit some cancellation.

2.1.1 Spaces of Probability Measures

Let us first recall our spaces of measures and probability measures. We define, for any topological space E, the space $\mathcal{P}(E)$ of Borel probability measures on E, and $\mathcal{M}(E)$ for the space of finite, positive Borel measures. In the case $E = \mathbb{R}^d$ and for $p > 0$, we introduce the notation

$$\Lambda_p(\mu) := \int_{\mathbb{R}^3} |v|^p \mu(dv) = \langle |v|^p, \mu \rangle \tag{2.1}$$

and write

$$\mathcal{P}_p(\mathbb{R}^d) = \left\{\mu \in \mathcal{P}(\mathbb{R}^d) : \Lambda_p(\mu) < \infty\right\}. \tag{2.2}$$

In Chapters 3, 4, we will work with spaces of probability measures $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ with prescribed energy and momentum: we define

$$\mathcal{S} := \left\{\mu \in \mathcal{P}_2(\mathbb{R}^d) : \langle v, \mu \rangle = 0,\ \Lambda_2(\mu) = 1\right\} \tag{2.3}$$

and, for $p \geq 2$, we define

$$\mathcal{S}^p := \mathcal{S} \cap \mathcal{P}_p(\mathbb{R}^d). \tag{2.4}$$

We will frequently encounter expressions with moments of two measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$. In this case, we write

$$\Lambda_p(\mu, \nu) := \max\left(\Lambda_p(\mu), \Lambda_p(\nu)\right). \tag{2.5}$$

2.1.2 Metrics Defined by Duality

In Chapter 3, we will work with the following family of metrics defined by duality against Hölder-continuous test functions of quadratic growth. For $f : \mathbb{R}^d \to \mathbb{R}$, we define $\widehat{f}(v) = f(v)/(1 + |v|^2)$, and the γ-Hölder norm

$$\|f\|_{0,\gamma} := \max\left(\sup_v |f|(v),\ \sup_{v \neq w} \frac{|f(v) - f(w)|}{|v - w|^\gamma}\right). \tag{2.6}$$

We write $\mathcal{A}_\gamma$ for the space of weighted γ-Hölder functions:

$$\mathcal{A}_\gamma := \left\{f : \mathbb{R}^d \to \mathbb{R} : \|\hat{f}\|_{0,\gamma} \leq 1\right\} \tag{2.7}$$

and define the weighted Wasserstein metric of type γ by the duality

$$W_\gamma(\mu,\nu) := \sup_{f\in\mathcal{A}_\gamma} |\langle f, \mu-\nu\rangle|. \tag{2.8}$$

In Chapters 6, 7, we will use similarly defined metrics on measures. Given a complete metric space (S,d), we write ρ_1 for the metric on finite measures given by

$$\rho_1(w,w') = \sup\left\{\langle g, w-w'\rangle : \sup_S |g| \le 1, \sup_{p,q\in S, p\neq q} \frac{|g(p)-g(q)|}{d(p,q)} \le 1\right\}. \tag{2.9}$$

Let us remark that this generates the topology of weak convergence on $\mathcal{M}(S)$. Further, if one restricts to subspaces of the form $\{\mu \in \mathcal{M}(S) : \langle \varphi, \mu\rangle \le a\}$ for $a \ge 0$ and some $\varphi : S \to (0,\infty)$ with compact sublevel sets, then the weak topology coincides with the *vague* topology induced by convergence against test functions in $C_c(S)$.

2.1.3 Optimal Transport Costs

Another way of measuring the distances between different probability measures is via optimal transportation problems, which are naturally well-adapted to the coupling arguments we will use in Chapters 4, 5, but are not necessarily metrics. We will use the following family of transportation costs to measure the distance between two solutions. For any $\mu,\nu \in \mathcal{P}(\mathbb{R}^d)$ we write $\Pi(\mu,\nu)$ for the set of couplings

$$\Pi(\mu,\nu) = \left\{\pi \in \mathcal{P}(\mathbb{R}^d\times\mathbb{R}^d) \ : \ \pi \text{ has marginals } \mu \text{ and } \nu\right\}.$$

For $p \ge 0$, and $\varepsilon \ge 0$, we define the functions on $\mathbb{R}^d\times\mathbb{R}^d$

$$d_{p,\varepsilon}(v,\tilde{v}) := (1+|v|^p+|\tilde{v}|^p)\varphi_\varepsilon(|v-\tilde{v}|^2); \qquad \varphi_\varepsilon(r) := \frac{r}{1+\varepsilon r}. \tag{2.10}$$

Let us observe that $\varphi_\varepsilon'(r) = (1+\varepsilon r)^{-2}$ and $\varphi_\varepsilon''(r) = -2\varepsilon(1+\varepsilon r)^{-3}$,

$$r\varphi_\varepsilon'(r) \le \varphi_\varepsilon(r), \quad 0 \le \varphi_\varepsilon'(r) \le 1 \quad \text{and} \quad \varphi_\varepsilon''(r) \le 0. \tag{2.11}$$

When $\varepsilon = 0$, we will drop it from the notation and write $d_p := d_{p,0}$. In this case, d_p has the upper bound growth

$$d_p(v,\tilde{v}) \le C(1+|v|^{p+2}+|\tilde{v}|^{p+2}) \tag{2.12}$$

and we define the corresponding optimal transport cost, for $\mu,\nu \in \mathcal{P}_{p+2}(\mathbb{R}^d)$,

$$w_p(\mu,\nu) = w_{p,0}(\mu,\nu) := \inf\left\{\int_{\mathbb{R}^d\times\mathbb{R}^d} d_p(v,\tilde{v})\pi(dv,d\tilde{v}) : \pi \in \Pi(\mu,\nu)\right\}. \tag{2.13}$$

In the cases $\varepsilon > 0$, we have the upper bound

$$d_{p,\varepsilon}(v,\tilde{v}) \le C_\varepsilon(1+|v|^p+|\tilde{v}|^p) \tag{2.14}$$

and so we can define the optimal transport cost for $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$:

$$w_{p,\varepsilon}(\mu,\nu) := \inf\left\{\int_{\mathbb{R}^3\times\mathbb{R}^3} d_{p,\epsilon}(v,\tilde{v})\pi(dv,d\tilde{v}) : \pi \in \Pi(\mu,\nu)\right\}. \tag{2.15}$$

We will usually work with either $w_{p,0}$ or $w_{p,1}$; for other values of $\varepsilon > 0$, we note that $\min(1,1/\varepsilon)d_{p,1} \le d_{p,\varepsilon} \le \max(1,1/\varepsilon)d_{p,1}$, which gives the bound for the costs

$$\min\left(1,\frac{1}{\varepsilon}\right) w_{p,1}(\mu,\nu) \le w_{p,\varepsilon}(\mu,\nu) \le \max\left(1,\frac{1}{\varepsilon}\right) w_{p,1}(\mu,\nu). \tag{2.16}$$

In either case $\varepsilon = 0, \varepsilon > 0$, it is straightforward to see that there exists a coupling attaining the infimum. Moreover, $w_{p,\varepsilon}$ are always nonnegative, symmetric, and using the existence of a minimiser, $w_{p,\varepsilon}(\mu,\nu) = 0$ if and only if $\mu = \nu$.

However, we remark that the functions $d_{p,\varepsilon}$ are not, in general metrics, and so $w_{p,\varepsilon}$ do not have a triangle inequality. Instead, for the case $\varepsilon = 0$, we note that the function $\delta_p(v,w) = |(1+|v|^p)^{1/2}v - (1+|w|^p)^{1/2}w|$ defines a metric, and that d_p/δ_p^2 is bounded above, and away from 0; in particular, there exists a constant $C = C_p$ such that

$$d_p(v,y) \le C(d_p(v,w) + d_p(w,y)) \tag{2.17}$$

for all $v,w,y \in \mathbb{R}^d$. For the case $\varepsilon > 0$, we observe that $\frac{1}{2}(\varepsilon^{-1}\wedge r) \le \varphi_\varepsilon(r) \le (\varepsilon^{-1}\wedge r)$, and one can prove a bound equivalent to (2.17) for $(1+|v|^p+|\tilde{v}|^p)(|v-\tilde{v}|^2 \wedge \varepsilon^{-1})$ by considering, case-by-case, which of $|v-w|^2$, $|w-y|^2$, $|v-y|^2$ are less than ε^{-1}. All together, we find that, for a constant $C = C_{p,\varepsilon}$,

$$d_{p,\varepsilon}(v,y) \le C(d_{p,\varepsilon}(v,w) + d_{p,\varepsilon}(w,y)) \tag{2.18}$$

for all v,w,y, and integrating, we find that the optimal transportation costs $w_p, w_{p,\varepsilon}$ satisfy *relaxed triangle inequalities*:

$$w_{p,\varepsilon}(\mu,\nu) \le C_{p,e}[w_{p,\varepsilon}(\mu,\lambda) + w_{p,\varepsilon}(\lambda,\nu)] \tag{2.19}$$

for all μ,ν,λ in $\mathcal{P}_{p+2}(\mathbb{R}^d)$ if $\varepsilon = 0$, and for all for all μ,ν,λ in $\mathcal{P}_p(\mathbb{R}^d)$ if $\varepsilon > 0$. The optimal transport costs $w_{p,\varepsilon}$ are therefore *semimetrics*; this failure to be a true metric will not cause any problems in the sequel.

This form of this optimal transportation cost is key to the arguments of Chapters 4, 5; the key calculations rely on a *Povzner effect* of the prefactor $(1+|v|^p+|\tilde{v}|^p)$, which is similar to the one we will see in deriving moment estimates in Section 2.5. In the Landau case in Chapter 5, we will use the cost functions $w_{p,1}$ rather than w_p to minimise the number of moments required, since these are defined with only p rather than $p+2$ moments, and use $w_{p,\varepsilon}$ as an intermediate step.

Let us also mention the Wasserstein$_p$ metrics $\mathcal{W}_p, p \ge 1$, which provide useful benchmarks for the strength of our metrics, and which we will also use in discussing explicit chaoticity estimates. For $p \geq 1$ and $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$, the Wasserstein$_p$-distance is given by

$$\mathcal{W}_p(\mu,\nu) := \inf\left\{\int_{\mathbb{R}^d\times\mathbb{R}^d} |v-\tilde{v}|^p \pi(dv,d\tilde{v}) :\ \pi \in \Pi(\mu,\nu)\right\}^{1/p} \tag{2.20}$$

as well as

$$\mathcal{W}_{1,1}(\mu,\nu) := \inf\left\{\int_{\mathbb{R}^d\times\mathbb{R}^d} 1\wedge|v-\tilde{v}|\pi(dv,d\tilde{v}) : \pi \in \Pi(\mu,\nu)\right\}. \tag{2.21}$$

In the context of chaoticity, we will similarly discuss analogously defined metrics $\mathcal{W}_{1,1,E^l}$ for a general metric space E and $l \in \mathbb{N}$, replacing $|v-\tilde{v}|$ with the metric of E^l.

2.1.4 Relationships Between Distances

Let us now examine some relationships between the different distances. In what follows, C will denote a constant, allowed to change from line to line, depending on p.

1. $\mathcal{W}_1$ and W_γ: Kantorovich-Wasserstein Duality. In the case $p=1$, the metrics $\mathcal{W}_1$ is known as the Monge-Kantorovich-Wasserstein distance, and the well-known Kantorovich-Wasserstein duality [191, Example 5.16] gives

$$\mathcal{W}_1(\mu,\nu) = \sup\left\{\langle f,\mu-\nu\rangle :\ \text{ for all } v,w,\ |f(v)-f(w)| \leq |v-w|\right\}; \tag{2.22}$$

$$\mathcal{W}_{1,1}(\mu,\nu) = \sup\left\{\langle f,\mu-\nu\rangle :\ \text{ for all } v,w,\ |f(v)-f(w)| \leq 1\wedge|v-w|\right\} \tag{2.23}$$

which produces

$$\frac{1}{2}\sup\{\langle f,\mu-\nu\rangle : \|f\|_{0,1}\leq 1\} \leq \mathcal{W}_{1,1}(\mu,\nu) \leq \sup\{\langle f,\mu-\nu\rangle : \|f\|_{0,1}\leq 1\}. \tag{2.24}$$

The supremum occurring here is exactly $W_1(\mu/(1+|v|^2), \nu/(1+|v|^2))$, and this recovers the metric ρ_1 in the case $E=\mathbb{R}^d$ with the Euclidean metric. Without the boundedness condition, the same duality cited above gives

$$\mathcal{W}_1(\mu,\nu) = \sup\left\{\langle f,\mu-\nu\rangle : \sup_{v\neq w}\frac{|f(v)-f(w)|}{|v-w|} \leq 1\right\}. \tag{2.25}$$

Further, $W_\gamma, 0<\gamma\leq 1$ and $\mathcal{W}_1, \mathcal{W}_2$ all generate the topology of weak convergence on $\mathcal{S}$.

Let us mention some quantitative comparisons between and within these classes. For all f, we have the bound $\|f\|_{0,\gamma} \leq 2^{1-\gamma}\|f\|_{0,1}$, which leads to the comparison $W_1 \leq 2^{1-\gamma}W_\gamma$, while approximating $f \in \mathcal{A}_\gamma$ by $f^\epsilon \in c_\epsilon \mathcal{A}_1$ for some $c_\varepsilon \to \infty$ leads to the bound $W_\gamma \leq C_\gamma W_1^\gamma$.

We now compare $\mathcal{W}_{1,1}$ and W_1. On the one hand, if $\|f\|_{0,1} \le 1$, then it is straightforward to see that $\|\widehat{f}\|_{0,1} \le c$ for some absolute constant c, so the duality (2.23) implies that $\mathcal{W}_{1,1} \le cW_1$. In the other direction, if $\|f\|_{0,1} \le 1$ and $\pi \in \Pi(\mu,\nu)$, then we write

$$\langle (1+|v|^2)f, \mu-\nu\rangle = \int_{\mathbb{R}^d\times\mathbb{R}^d} ((1+|v|^2)f(v)-(1+|w|^2)f(w))\pi(dv,dw).$$

By writing the difference as $(|v|^2-|w|^2)f(v)+(1+|w|^2)(f(v)-f(w))$, we find

$$\begin{aligned}|\langle (1+|v|^2)f, \mu-\nu\rangle| \le &\int_{\mathbb{R}^d\times\mathbb{R}^d} (|v-w|)(|v|+|w|)\pi(dv,dw) \\ &+2\int_{\mathbb{R}^d\times\mathbb{R}^d}(1\wedge|v-w|)(1+|w|^2)\pi(dv,dw).\end{aligned} \tag{2.26}$$

If $\mu,\nu \in \mathcal{S}^p$ for some $p>2$, then one can interpolate and optimise over f,π to obtain, for some $C=C(p), \alpha=\alpha(p)>0$,

$$W_1(\mu,\nu) \le C\Lambda_p(\mu,\nu)\mathcal{W}_{1,1}(\mu,\nu)^\alpha. \tag{2.27}$$

2. $w_{p,0}$ and $\mathcal{W}_{p+2}$. We first show interpolation between our optimal transportation costs and the Wasserstein$_p$ metrics in the special case $\varepsilon=0$. For an upper bound, for any $\mu,\nu\in\mathcal{P}_{p+2}(\mathbb{R}^d)$ and $\pi\in\Pi(\mu,\nu)$, we have

$$\begin{aligned}\int_{\mathbb{R}^d\times\mathbb{R}^d} d_p(v,\tilde{v})\pi(dv,d\tilde{v}) &\le \left(\int_{\mathbb{R}^d\times\mathbb{R}^d}(1+|v|^p+|\tilde{v}|^p)^{(p+2)/p}\pi(dv,d\tilde{v})\right)^{p/p+2} \\ &\qquad\cdots\times\left(\int_{\mathbb{R}^d\times\mathbb{R}^d}|v-\tilde{v}|^{p+2}\pi(dv,dw)\right)^{2/p+2} \\ &\le C\Lambda_{p+2}(\mu,\nu)\left(\int_{\mathbb{R}^d\times\mathbb{R}^d}|v-\tilde{v}|^{p+2}\pi(dv,dw)\right)^{2/p+2}.\end{aligned}$$

If we now take π to be a minimiser for the right-hand side, attaining the infimum in the optimal transport problem for $\mathcal{W}_{p+2}(\mu,\nu)$, we conclude that

$$w_p(\mu,\nu) \le C\Lambda_{p+2}(\mu,\nu)\mathcal{W}_{p+2}(\mu,\nu)^{2/p+2}.$$

On the other hand, the easy inequality $|v-\tilde{v}|^{p+2}\le Cd_p(v,\tilde{v})$ produces $\mathcal{W}_{p+2}(\mu,\nu)^{p+2}\le w_p(\mu,\nu)$, and together we conclude

$$C^{-1}\mathcal{W}_{p+2}(\mu,\nu)^{(p+2)} \le w_p(\mu,\nu) \le C\mathcal{W}^2_{p+2}(\mu,\nu)\Lambda_{p+2}(\mu,\nu). \tag{2.28}$$

Further, for any sequence $\mu^n,\mu\in\mathcal{S}^{p+2}$, the convergence $\mathcal{W}_{p+2}(\mu^n,\mu)\to 0$ implies convergence of the moments $\Lambda_{p+2}(\mu^n)\to\Lambda_{p+2}(\mu)$, and so $w_p, \mathcal{W}_{p+2}$ have the same convergent sequences and generate the same topology on $\mathcal{S}^{p+2}$.

Let us also record the elementary interpolation estimate, for any $p'>p+2$ and any $\mu,\nu\in\mathcal{S}^{p'}$,

$$\mathcal{W}_1^2(\mu,\nu) \le w_p(\mu,\nu) \le \mathcal{W}_1(\mu,\nu)^\alpha\Lambda_{p'}(\mu,\nu) \tag{2.29}$$

for some $\alpha=\alpha(p,p')>0$. Indeed, the first inequality follows from $w_p\ge\mathcal{W}_2^2$ and the monotonicity of $\mathcal{W}_p$, and the second follows from an application of the Hölder inequality.

3. w_p and W_γ. Combining the estimates above, if $p \geq 0, p' > p+2$ and $0 < \gamma \leq 1$, we have the equivalence

$$W_\gamma(\mu,\nu) \leq C\Lambda_{p'}(\mu,\nu)w_p(\mu,\nu)^{\gamma/2}; \qquad w_p(\mu,\nu) \leq C\Lambda_{p'}(\mu,\nu)W_\gamma(\mu,\nu)^\alpha \tag{2.30}$$

for some $C = C(p,p'), \alpha = \alpha(p,p')$.

4. $w_{p,1}$ and $\mathcal{W}_p$. Let us now repeat these calculations for $\varepsilon > 0$; thanks to (2.16), it is sufficient to consider the case $\varepsilon = 1$. In this case, we start by noticing that

$$|v-\tilde{v}|^p \leq C(1+|v|^p+|\tilde{v}|^p)\frac{|v-\tilde{v}|^2}{1+|v-\tilde{v}|^2}$$

so that $\mathcal{W}_p^p \leq Cw_{p,1}$. Moreover, it can be checked that if $\mu^n, \mu \in \mathcal{P}_p(\mathbb{R}^d)$ are such that $\mu^n \to \mu$ in the weak topology, and in addition $\Lambda_p(\mu^n) \to \Lambda_p(\mu)$, then $w_{p,1}(\mu^n,\mu) \to 0$; see Lemma 5.11 for a very similar proof. It follows that $w_{p,1}$ and $\mathcal{W}_p$ induce the same topology on $\mathcal{P}_p(\mathbb{R}^d)$

2.2 Chaoticity

We next formulate the concept of chaoticity described in the introduction. As well as important context and a rigorous introduction to this background, this justifies our labelling of Theorems 1, 2 as chaoticity estimates. We follow [142, Section 1.3].

Let us fix a separable and complete metric space (E,d) and, for $N \geq 2$, a probability measure $\mathcal{L}^N$ on E^N. For each such N, the symmetric group $\mathrm{Sym}(N)$ acts on the product space E^N by permuting the indexes

$$\sigma(x) := (x_{\sigma(1)}, x_{\sigma(2)}, ..., x_{\sigma(N)}); \qquad \sigma \in \mathrm{Sym}(N), x = (x_1, \dots, x_N) \in E^N$$

and we say that $\mathcal{L}^N \in \mathcal{P}(E^N)$ is symmetric if it is invariant under the action of all such maps: $\sigma_\#\mathcal{L}^N = \mathcal{L}^N$ for all $\sigma \in \mathrm{Sym}(N)$. We will write $\theta_N : E^N \to \mathcal{P}(E)$ the map associating to any N-tuple its normalised empirical measure

$$(x_1,\dots,x_N) \to \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$$

and we observe that this is preserved by the action of $\mathrm{Sym}(N)$ in the sense that $\theta_N \circ \sigma = \theta_N$ for all $\sigma \in \mathrm{Sym}(N)$.

For any $l \in \mathbb{N}$ and $N \geq l$, we write $M_l : \mathcal{P}(E^N) \to \mathcal{P}(E^l)$ for the marginal distribution on the first l factors; that is, $M_l(\mu)$ is the pushforward of μ by the natural projection $E^N \to E^l, (x_1,\dots,x_N) \to (x_1,\dots,x_l)$. We equip E^l with the metric

$$d_l((v_1,\dots,v_l),(\tilde{v}_1,\dots,\tilde{v}_l)) := \sum_{i=1}^l 1 \wedge d(v_i,\tilde{v}_i)$$

and write $\mathcal{W}_{1,1,E^l}$ for the metric on $\mathcal{P}(E^l)$ defined analagously to (2.21) with this cost function:

$$\mathcal{W}_{1,1,E^l}(\mathcal{L},\mathcal{L}') = \inf\left\{\int_{E^l\times E^l} d_l(v,\tilde{v})\pi(dv,d\tilde{v}) : \pi\in\Pi(\mathcal{L},\mathcal{L}')\right\}.$$

As in the previous section, these are metrics, which are equivalent to the weak convergence of measures on $\mathcal{P}(E^l)$, with the dual formulation

$$\mathcal{W}_{1,1,E^l}(\mathcal{L},\mathcal{L}') = \sup\left\{\int_{E^l} f(V)(\mathcal{L}(V)-\mathcal{L}'(V))\right\}$$

over all those $f: E^l\to\mathbb{R}$ with $|f(V)-f(V')|\le d_l(V,V')$ for all V,V'. With these definitions, we make precise the definition of chaoticity. We begin with the following definition by Kac [122].

Definition 2.2.1 (Finite Dimensional Chaos)**.** *Let $\mathcal{L}^N$, $N\ge 2$ be symmetric laws on E, and $\mu\in\mathcal{P}(E)$. We say that $\mathcal{L}^N$ are μ-chaotic if, for all $l\ge 1$,*

$$\mathcal{W}_{1,1,E^l}\left(M_l[\mathcal{L}^N],\mu^{\otimes l}\right)\to 0 \tag{2.31}$$

or equivalently if $M_l[\mathcal{L}^N]\to\mu^{\otimes l}$ in the weak topology of $\mathcal{P}(E^l)$.

In this context, one can naturally give a quantitative formulation by giving an upper bound $\delta_{N,l}$ for the left-hand side of (2.31). A stronger notion is that of *infinite dimensional* chaos, which allows us to take $N,l\to\infty$ simultaneously.

Definition 2.2.2 (Infinite Dimensional Chaos)**.** *In the setting of the previous definition, we say $\mathcal{L}^N$ are infinite-dimensionally μ-chaotic if*

$$\sup_{l\le N}\left[\frac{1}{l}\mathcal{W}_{1,1,l}\left(M_l[\mathcal{L}^N],\mu^{\otimes l}\right)\right]\to 0. \tag{2.32}$$

As in the previous case, one can make this quantitative by finding an upper bound for the left-hand side. The general result we will use is the following, which is a quantitative version of the well-known equivalence between chaoticity and the convergence of the empirical measure [104, Section 4], [122, 178], [173, Proposition 2.2].

Proposition 2.1. *In the general framework above, suppose that $\mathcal{L}^N$ are symmetric laws such that*

$$\varepsilon_N := \mathbb{E}_{\mathcal{V}^N\sim\mathcal{L}^N}\left[\mathcal{W}_{1,1,E}(\theta_N(\mathcal{V}^N),\mu)\right]\to 0.$$

Then $\mathcal{L}^N$ are chaotic, with an explicit rate

$$\sup_{l\le N}\left[\frac{1}{l}\mathcal{W}_{1,1,E^l}\left(M_l[\mathcal{L}^N],\mu^{\otimes l}\right)\right]\le 2(\varepsilon_N+\delta_N(\mu))\to 0 \tag{2.33}$$

where $\delta_N(\mu)$ is the error with independent samples:

$$\delta_N(\mu) := \mathbb{E}\left[\mathcal{W}_{1,1,E}\left(\mu,\theta_N(\widetilde{\mathcal{V}}^N)\right)\right]; \qquad \widetilde{\mathcal{V}}^N\sim\mu^{\otimes N}$$

which converges to 0 *by the law of large numbers. Conversely, if* $\mathcal{L}^N$ *are infinite-dimensionally* μ*-chaotic, then*

$$\mathbb{E}_{\mathcal{V}^N\sim\mathcal{L}^N}[\mathcal{W}_{1,1,E}(\theta_N(\mathcal{V}^N),\mu)] \le \delta_N(\mu) + \frac{1}{N}\mathcal{W}_{1,1,E^N}(\mathcal{L}^N,\mu^{\otimes N}) \to 0.$$

Remark 2.2 (Rate of convergence of the empirical measures)**.** *In the cases where* E *is a subset of* $\mathbb{R}^d$ *equipped with the Euclidean distance, one can quantify the rate* $\delta_N(\mu)$ *for independent sampling, see works by Talagrand [174, 175], Fournier [87] and Norris [157, Propositions 9.2]. In general, the rate of convergence* $\delta_N \sim N^{-\alpha}, \alpha > 0$*. The best possible rate of converge in this context for general measures is* $N^{-1/d}$*, and it is known that this rate is achieved as soon as* μ *has slightly more than* $\frac{d}{d-1}$ *moments [87, Theorem 1]. We will not sketch these arguments, since we will see arguments very similar to those of [157, Proposition 9.2] in the course of Proposition 3.11 and Theorems 3.1 - 3.2 in Chapter 3.*

Application to the Kac Process Before giving the proof of Proposition 2.1, we briefly remark on how this general framework applies in the case of the Kac process. In this case, we take $E = \mathbb{R}^d$ and are interested in $\mathcal{L}_t^N := \mathrm{Law}(\mathcal{V}_t^N)$ for either $t = 0$, in which case these requirements give a hypothesis on the initial conditions, or $t > 0$, and this will be the conclusion of proving *propagation of chaos*. Regarding symmetry, this property is propagated by the dynamics; if the law of the initial data is symmetric, so is the law of $\mathcal{V}_t^N$ for all $t \ge 0$, reflecting the indistinguishability of the particles. Furthermore, this can always be imposed with no loss of generality; if the law of $\mathcal{V}_0^N$ is not symmetric, form $\widetilde{\mathcal{V}}_0^N$ by randomly permuting the indexes, which does not affect the empirical measure.

Let us also compare the chaoticity results one finds by applying this proposition to Theorems 3.1, 4.5 to the literature, and our techniques to those of [142], on whose work we build in Chapter 3. Applying Proposition 2.1 to Theorem 3.1, we find a uniform-in time chaoticity estimate for (HS) as in [142, Theorem 6.1], but we improve the rate from $(\log N)^{-r}$ to a polynomial estimate $N^{-\alpha}$, mostly thanks to the estimates of the continuity of the Boltzmann flow in W_1 in Theorem 3.6. Let us mention that, although we use many of the same ideas, the method of proof is slightly different, see [142, Theorem 3.1]. In this work, the authors first derive rates for finite-dimensional chaos, and deduce rates for infinite dimensional chaos via a general result [141, Théorème 2.1]. Following the proof of this general result, as we did in the proposition above, one uses the finite-dimensional chaos to deduce the convergence of the empirical measure μ^N, and that this implies infinite dimensional chaos as in the proposition above. Since we prove convergence of the empirical measure directly, this latter part is subsumed into Theorem 3.1, and we can proceed directly to infinite-dimensional chaos. In the case (NCHP) in Chapter 4, we find a rate $e^{Ct}(\log N)^{-r}, r > 0$, which is roughly analagous to the rate of the obtained for the Maxwell Molecule case by Desvillettes and Méléard [55], via a similar method.

We now give the proof of the proposition.

Proof of Proposition 2.1. For the first implication, we follow [141, Théorème 2.1, Steps 3-4]; the proof is simplified since we already start from the convergence of the empirical measure. Let us start from a coupling $(\mathcal{U}^N,\widetilde{\mathcal{U}}^N)$ of $\mathcal{L}^N$ and $\mu^{\otimes N}$ which minimises $\mathbb{E}[\mathcal{W}_{1,1,E}(\theta_N(\mathcal{U}^N),\theta_N(\widetilde{\mathcal{U}}^N))]$ over all such couplings, recalling that θ_N is the map taking an N-tuple to its empirical measure. We now define a coupling $(\mathcal{V}^N,\widetilde{\mathcal{V}}^N)$ by sampling uniformly from the finite set of $2N$-tuples $(v^N,\widetilde{v}^N)$ with $\theta_N(v^N)=\theta_N(\mathcal{U}^N)$, $\theta_N(\widetilde{v}^N)=\theta_N(\widetilde{\mathcal{U}}^N)$ and which attain

$$\frac{1}{N}d_N(v^N,\tilde{v}^N)=\mathcal{W}_{1,1,E}(\theta_N(\mathcal{U}^N),\theta_N(\widetilde{\mathcal{U}}^N))=\mathcal{W}_{1,1,E}(\theta_N(v^N),\theta_N(\widetilde{v}^N))$$

which amounts to choosing reorderings of the indexes of $\mathcal{U}^N$, $\widetilde{\mathcal{U}}^N$, uniformly among pairs attaining the minimum. Since $\mathcal{L}^N$ is already symmetric, it follows that $\mathrm{Law}(\mathcal{V}^N)=\mathcal{L}^N$, and similarly $\mathrm{Law}(\widetilde{\mathcal{V}}^N)=\mu^{\otimes N}$. We now write

$$\mathcal{W}_{1,1,E}(\theta_N(\mathcal{V}^N),\mu)\geq\mathcal{W}_{1,1,E}(\theta_N(\mathcal{V}^N),\theta_N(\widetilde{\mathcal{V}}^N))-\mathcal{W}_{1,1,E}(\theta_N(\widetilde{\mathcal{V}}^N),\mu)$$

by the reverse triangle inequality, so that, taking expectations,

$$\begin{aligned}\mathbb{E}[\mathcal{W}_{1,1,E}(\theta_N(\mathcal{V}^N),\theta_N(\widetilde{\mathcal{V}}^N))]&\leq\mathbb{E}\left[\mathcal{W}_{1,1,E}(\theta_N(\mathcal{V}^N),\mu)\right]+\mathbb{E}\left[\mathcal{W}_{1,1,E}(\theta_N(\widetilde{\mathcal{V}}^N),\mu)\right]\\&=:\varepsilon_N+\delta_N(\mu).\end{aligned}\tag{2.34}$$

On the other hand, by construction of $(\mathcal{V}^N,\widetilde{\mathcal{V}}^N)$, the left-hand side is equal to

$$\mathbb{E}[\mathcal{W}_{1,1,E}(\theta_N(\mathcal{V}^N),\theta_N(\widetilde{\mathcal{V}}^N))]=\mathbb{E}\left[\frac{1}{N}d_N(\mathcal{V}^N,\widetilde{\mathcal{V}}^N)\right]\geq\frac{1}{N}\mathcal{W}_{1,1,E^N}(\mathcal{L}^N,\mu^{\otimes N})$$

where the final bound follows from recalling that $\mathrm{Law}(\mathcal{V}^N,\widetilde{\mathcal{V}}^N)\in\Pi(\mathcal{L}^N,\mu^{\otimes N})$. The claim is now proven in the special case $l=N$.

To obtain the general case with a maximum over $l\leq N$, we fix $l\leq N$ and write $N=ql+r$ with $0\leq r<l$, and observe that $q\geq 1$ since $N\geq l$. We write

$$E^N=E^l\times....\times E^l\times E^r.$$

Let us now choose a coupling $(\mathcal{V}^N,\widetilde{\mathcal{V}})$ attaining the minimum of the optimal transportation problem for $\mathcal{W}_{1,1,E^N}(\mathcal{L}^N,\mu^{\otimes N})$, and write $\mathcal{V}^N=(V_1,...,V_q,V_0)$ with $V_0\in E^r$ and each $V_i\in E^l$, and similarly for the blocks of $\widetilde{\mathcal{V}}$. Since each pair $(V_i,\tilde{V}_i)$ gives a coupling of $(M_l[\mathcal{L}^N],\mu^{\otimes l})$ and using the decomposition

$$d_N(\mathcal{V}^N,\tilde{\mathcal{V}}^N)=\sum_{i=1}^q d_l(V_i,\tilde{V}_i)+d_r(V_0,\tilde{V}_0)$$

from which it follows that

$$\begin{aligned}\frac{1}{N}\mathcal{W}_{1,1,E^N}(\mathcal{L}^N,\mu^{\otimes N})&=\mathbb{E}\left[\frac{1}{N}d_N(\mathcal{V}^N,\widetilde{\mathcal{V}}^N)\right]\\&=\frac{1}{N}\sum_{i=1}^q\mathbb{E}\left[d_l(V_j,V_j))\right]+\frac{1}{N}\mathbb{E}\left[d_r(V_0,\tilde{V}_0)\right]\\&\geq\frac{q}{N}\mathcal{W}_{1,1,E^l}(M_l[\mathcal{L}^N],\mu^{\otimes l}).\end{aligned}\tag{2.35}$$

Now, we note that the prefactor is $\frac{q}{N} = \frac{1}{l}(\frac{ql}{ql+r}) \geq \frac{1}{2l}$ since $ql \geq l > r$, so we conclude that

$$\frac{1}{l}\mathcal{W}_{1,1,E^l}(M_i[\mathcal{L}^N], \mu^{\otimes l}) \leq \frac{2}{N}\mathcal{W}_{1,1,E^N}(\mathcal{L}^N, \mu^{\otimes N})$$

uniformly in $l \leq N$, and the first implication is proven, to control the difference between $\mathcal{L}^l$ and $\mu^{\otimes l}$ in terms of ε_N, δ_N.

On the other hand, for any coupling $(\mathcal{V}^N, \widetilde{\mathcal{V}}^N)$ of $\mathcal{L}^N$ and $\mu^{\otimes N}$, we have

$$\begin{aligned}\mathcal{W}_{1,1,E}(\theta_N(\mathcal{V}^N), \mu) &\leq \mathcal{W}_{1,1,E}(\theta_N(\mathcal{V}^N), \theta_N(\widetilde{\mathcal{V}}^N)) + \mathcal{W}_{1,1,E}(\theta_N(\widetilde{\mathcal{V}}^N), \mu) \\ &\leq \frac{1}{N} d_N(\mathcal{V}^N, \widetilde{\mathcal{V}}^N) + \mathcal{W}_{1,1,E}(\theta_N(\widetilde{\mathcal{V}}^N), \mu).\end{aligned} \tag{2.36}$$

The second term always has expectation $\delta_N(\mu)$, for any coupling $(\mathcal{V}^N, \widetilde{\mathcal{V}}^N)$, and taking the infimum, the first term produces $N^{-1}\mathcal{W}_{1,1,E^N}(\mathcal{L}^N, \mu^{\otimes N})$. We can therefore control the convergence of the empirical measures in terms of $N^{-1}\mathcal{W}_{1,1,E^N}(\mathcal{L}^N, \mu^{\otimes N})$ and δ_N, so we are done. □

2.3 Initial Data with prescribed Energy and Momentum

In Chapters 3 - 4, we will typically work with Kac processes with a prescribed energy and momentum, so that the empirical measures take values in the Boltzmann sphere $\mathcal{S}$. However, in general, empirical measures of independent draws from a base distribution μ_0 do not take values in $\mathcal{S}$, and this can even be an event of probability 0 in the case when μ_0 has a density. We will use the following construction, which produces normalised initial data with a natural modification of the initial values [157, Proposition 9.3].

Let us fix a $\mu_0 \in \mathcal{S}$, and a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$ on which are defined an infinite sequence $U_1, ..U_N, \dots$ from μ_0. For each N, we define

$$\overline{U}_N = \frac{1}{N}\sum_{i=1}^N U_i; \qquad s_N := \frac{1}{N}\sum_{i=1}^N |U_i - \overline{U}_N|^2 \tag{2.37}$$

and construct $\mathcal{V}^N$ by setting, if $s_N > 0$

$$V_i^N = \frac{U_i - \overline{U}_N}{\sqrt{s_N}}; \qquad 1 \leq i \leq N \tag{2.38}$$

or choose $\mathcal{V}^N \in \mathbb{S}_N$ arbitrarily if $s_N = 0$. It follows that $\mathcal{V}^N \in \mathbb{S}_N$, or equivalently the empirical measure $\mu_0^N := N^{-1}\sum_i \delta_{V_i^N} \in \mathcal{S}$ almost surely. The result we will use is as follows.

Proposition 2.3. *Let $\mu_0 \in \mathcal{S}$ and let μ_0^N be given by the construction above. If $\mu_0 \in \mathcal{S}^p$ for some $p > 2$, we have the convergence $\mathbb{E}[\mathcal{W}_1(\mu_0^N, \mu_0)] \to 0$ with a rate $N^{-\alpha}$ for some $\alpha = \alpha(p, d) > 0$, and $\Lambda_q(\mu_0^N) \to \Lambda_q(\mu_0)$ for all $q \in [2, p]$, almost surely.*

The following holds as a corollary, which will be helpful for constructing approximations in the context of the Boltzmann equation in Chapter 4.

Corollary 2.4. *Fix $p \geq 0$, and let $\mu \in \mathcal{S}^q$ for $q > p+2$. Then there exists a sequence $\mu^N \in \mathcal{S}_N$ of discrete approximations to μ such that*

$$w_p(\mu^N, \mu) \to 0; \qquad \Lambda_{p'}(\mu^N) \to \Lambda_{p'}(\mu) \text{ for all } p' \leq q. \tag{2.39}$$

This follows immediately using the interpolation from $\mathcal{W}_1$ to w_p in (2.29); indeed, in the context of the previous proposition, the set of $\omega \in \Omega$ where $\mu^N(\omega)$ satisfies the desired condition (2.39) has probability 1, and in particular is nonempty.

2.4 Parametrisation of Jumps

We next introduce several possible parametrisations of the changes $(v, v_*) \mapsto (v', v_\star')$ which will be useful at various points.

We start with a parametrisation of $\mathbb{S}^{d-2}$. We choose a measurable map $\iota : \mathbb{R}^d \to (\mathbb{R}^d)^{d-1}$ such that, for all j, k and all v, we have

$$\iota_j(v) \cdot \iota_k(v) = |v|^2 1\!\!1_{j=k}$$

and such that $\iota(-v) = -\iota(v)$. It follows that, for all $v \neq 0$, the set

$$\left\{ \frac{v}{|v|}, \frac{\iota_1(v)}{|v|}, ..., \frac{\iota_{d-1}(v)}{|v|} \right\} \tag{2.40}$$

is an orthonormal basis of $\mathbb{R}^d$. With this choice of ι, define $\Gamma : \mathbb{R}^d \times \mathbb{S}^{d-2} \to \mathbb{R}^d$ by

$$\Gamma(v, \varphi) = \sum_{j=1}^{d-1} \varphi_j \iota_j(v). \tag{2.41}$$

With these fixed, for any unit vector u we can write down a parametrisation of $\mathbb{S}^{d-1} \setminus \{\pm u\}$ in terms of $\theta \in (0, \pi), \varphi \in \mathbb{S}^{d-2}$ by

$$F(\theta, \varphi) = u(\cos\theta) + (\sin\theta)\Gamma(u, \varphi).$$

To compute the Jacobian at any (θ, φ), pick an orthonormal basis $\psi_1, ..., \psi_{d-2}$ of the tangent space to $\mathbb{S}^{d-2}$ at φ, and observe that $D_\varphi F[\psi_i] = (\sin\theta)\Gamma(u, \psi_i)$ produces $d-2$ orthogonal vectors of norm $\sin\theta$, and which are all orthogonal to the tangent vector $\partial_\theta F = (-\sin\theta)u + (\cos\theta)\Gamma(u, \varphi)$. The Jacobian is therefore

$$\left| \frac{dF}{d(\theta, \varphi)} \right| = (\sin\theta)^{d-2}.$$

Recalling also that we work with the uniform (probability) measure $d\sigma$ on $\mathbb{S}^{d-1}$, and similarly equipping $\mathbb{S}^{d-2}$ with the uniform measure $d\varphi$, we normalise to write the uniform measure as the pushforward by F of

$$d\sigma = F_\star\left(\frac{(\sin\theta)^{d-2}}{c_d}\right)d\theta d\varphi \tag{2.42}$$

where $c_d = \int_0^\pi (\sin\theta)^{d-2}d\theta$ is the corresponding normalisation constant as in (1.2), and we have proven (1.3). With this parametrisation of σ, the corresponding post-collisional velocities are given by

$$v' = v\left(\frac{1+\cos\theta}{2}\right) + v_*\left(\frac{1-\cos\theta}{2}\right) + \sin\theta\Gamma(v-v_*,\varphi); \tag{2.43}$$

$$v_* = v\left(\frac{1-\cos\theta}{2}\right) + v_*\left(\frac{1+\cos\theta}{2}\right) - \sin\theta\Gamma(v-v_*,\varphi). \tag{2.44}$$

This change of variables is valid for any of the cases of the kernel, as we have made no mention of the kernel in this derivation. In the non-cutoff case[1] (NCHP), we make a further change of variables which is natural for coupling arguments [85, 92]. Recalling the definition of β in (1.2), we define

$$H(\theta) = \int_\theta^{\pi/2}\beta(x)dx, \qquad \theta\in\left(0,\frac{\pi}{2}\right). \tag{2.45}$$

Under the assumption (NCHP), H is now a bijection from $(0,\pi/2)$ to the ray $(0,\infty)$; let us write G for its inverse. We finally define, for distinct $v, v_\star \in \mathbb{R}^d$ and $\varphi\in\mathbb{S}^{d-2}$ and $z>0$,

$$\theta(v,v_\star,z) = G\left(\frac{z}{|v-v_\star|^\gamma}\right) \tag{2.46}$$

and

$$a(v,v_\star,z,\varphi) = -\frac{1-\cos(\theta(v,v_\star,z))}{2}(v-v_\star) + \frac{\sin(\theta(v,v_\star,z))}{2}\Gamma(v-v_\star,\varphi). \tag{2.47}$$

In the case $v=v_\star$, we set $a=0$; we note that, by construction, a is antisymmetric in $v, v_\star$. Some estimates for the function G are established in Section 4.10.1.

Let us now check that this gives a successful parametrisation.

Lemma 2.5. *For any v, v_* and any Lipschitz function $f:\mathbb{R}^d\to\mathbb{R}$, we have*

$$\begin{aligned}2\int_{\mathbb{S}^{d-2}\times(0,\infty)}&(f(v+a(v,v_*,z,\varphi))-f(v))dzd\varphi\\ &= \mathcal{L}_{\mathrm{B}}f(v,v_*) := 2\int_{\mathbb{S}^{d-1}}(f(v')-f(v))B(v-v_*,\sigma)d\sigma.\end{aligned} \tag{2.48}$$

[1] Or indeed for (tMM, SP), although we will not use these cases.

Proof. We fix v, v_*, f. If $v = v_*$, there is nothing to prove, as $v' = v$ for all σ and a is identically 0. Otherwise, we start by observing that

$$v + a(v, v_*, z, \varphi) = \frac{v + v_*}{2} + \frac{|v - v_*|}{2}((\cos\theta)(v, v_*, z)u + \sin\theta(v, v_*, z)\Gamma(u, \varphi)) \quad (2.49)$$

with $u = (v - v_*)/|v - v_*|$ the unit vector parallel to $v - v_*$. The combination in parentheses is a unit vector $\sigma(z, \varphi) \in \mathbb{S}^{d-1}$ and it is straightforward to see that $(z, \varphi) \to \sigma$ defines a bijection from $(0, \infty) \times \mathbb{S}^{d-2} \to \{\sigma \in \mathbb{S}^{d-1} : u \cdot \sigma \geq 0, \sigma \neq u\}$. Thanks to (2.42) and using the chain rule on the composition $(z, \varphi) \mapsto (\theta(v, v_*, z), \varphi) \mapsto \sigma$, the pushforward of the measure $dzd\varphi$ by this map has a density with respect to the uniform measure $d\sigma$ given by

$$\frac{c_d}{(\sin\theta(v, v_*, z))^{d-2}} \left|\frac{d}{dz}\theta(v, v_*, z)\right|^{-1} = \frac{c_d|v - v_*|^\gamma}{(\sin\theta(v, v_*, z))^{d-2}} \left|G'\left(\frac{z}{|v - v_*|^\gamma}\right)\right|^{-1}.$$

Using the construction of G, it follows that $G'(x) = 1/H'(G(x)) = -1/\beta(G(x))$, so all together the density is

$$\frac{c_d|v - v_*|^\gamma}{(\sin\theta(v, v_*, z))^{d-2}}\beta(\cos\theta(v, v_*, z))|v - v_*|^\gamma.$$

We observe that this is precisely $B(v - v_*, \sigma)$ given by the hypothesis (NCHP), since the θ appearing in the rate is exactly $\theta(v, v_*, z)$ and recalling the definition (1.2) of $\beta(\theta)$. We therefore write

$$\begin{aligned} 2\int_{\mathbb{S}^{d-2}\times(0,\infty)} & f(v + a(v, v_*, z, \varphi) - f(v))d\varphi dz \\ &= 2\int_{\{u\cdot\sigma\geq 0\}} (f(v') - f(v))B(v - v_*, \sigma)d\sigma. \end{aligned} \quad (2.50)$$

Meanwhile, the hypothesis in (NCHP) that b is supported on $[0, 1)$ means that $B = 0$ when $u \cdot \sigma < 0$, and we conclude that

$$2\int_{\mathbb{S}^{d-2}\times(0,\infty)} f(v + a(v, v_*, z, \varphi) - f(v))d\varphi dz = 2\int_{\mathbb{S}^{d-1}} (f(v') - f(v))B(v - v_*, \sigma)d\sigma \quad (2.51)$$

as desired. □

Let us briefly observe that, under the hypothesis (NCHP), we find $H(\theta) \sim \theta^{-\nu}$ as $\theta \to 0$, and correspondingly $G(z) \sim (1 + z)^{-1/\nu}$. In particular, it follows that $\int_0^\infty \sin G(z)dz < \infty$. We will analyse G more closely in Section 4.10.1. Finally, in the cases (GMM, rHS), we will rewrite the post-collisional velocities in the 'ω-representation'[2]

$$v' = v - ((v - v_*) \cdot \tau)\tau; \qquad v'_* = v_* - ((v_* - v) \cdot \tau)\tau \quad (2.52)$$

[2] This representation is usually written with ω, but we wish to reserve this for elements of the probability space $(\Omega, \mathfrak{F}, \mathbb{P})$.

for τ running over the unit sphere $\mathbb{S}^{d-1}$. This is equivalent to the usual σ-parametrisation, by taking

$$\sigma = \frac{v - v_* - 2((v-v_*)\cdot\tau)\tau}{|v-v_*|}.$$

Viewed as a function $\sigma = \sigma(\tau)$, the Jacobean is

$$\left|\frac{d\sigma}{d\tau}\right| = 2^{d-1}\sin\left(\frac{\theta}{2}\right)^{d-2}$$

which gives the change of kernel

$$\widetilde{B}(v,\tau) = 2^{d-2}\sin\left(\frac{\theta}{2}\right)^{d-2} B(v,\sigma(\tau))$$

with an extra factor of $\frac{1}{2}$ accounts for the fact that $\tau \to \sigma$ is two-to-one. In particular, the kernels (rHS, GMM) simplify to

$$\widetilde{B}(v,\tau) = 1 + |v|; \qquad \widetilde{B}(v,\tau) = 1$$

respectively.

2.5 Moment Estimates

We next turn to some moment estimates, which will be in frequent use throughout this thesis. Indeed, due to the unboundedness of the kinetic factor $\Psi(|v|)$ in the three cases (HS, rHS, NCHP) in the Boltzmann equation (BE), or $\gamma \in (0,1]$ in the case of (LE), one cannot even make sense of the formal identity $\langle |v|^2, Q(\mu)\rangle = 0$ unless the measure μ already makes $|v|^2\Psi(|v|)$ integrable. Frequently, in our calculations, we will find additional multiplicative factors, due to the unboundedness of these kernels, which depend on the moments of a solution to (BE, LE) at some future time $t \geq 0$, or the same for a particle system. We synthesise results from the literature as follows.

We will mostly be concerned with polynomial moments $\Lambda_p(\cdot), p \in [2,\infty)$. Such moment estimates for the spatially homogeneous Boltzmann and Landau equations are now fairly classical. For the Boltzmann equation, the proof centres on an inequality due to Povzner [163]. Elmroth [72] used this to prove that, for kernels with $\Psi(r) \sim r^\gamma$, that all moments which are initially finite remain finite, globally in time. This result was strengthened to moment *production* by Desvillettes [53], provided that some moment $\Lambda_p(\mu_0)$ is initially finite; this additional requirement was relaxed by Wennberg [194], requiring only finite initial energy $\langle |v|^2, \mu_0\rangle$; see also the work of Lu and Mouhot [133] regarding measure-valued solutions. Very similar arguments hold for the Kac process, which goes back to the works of Mischler and Mouhot [142] and Norris [157]. In this case, an additional martingale term appears, which vanishes again when one considers $\mathbb{E}[\Lambda_p(\mu_t^N)]$; in this context, we also need to argue maximal inequalities for pathwise maxima $\sup_{s\le t}\Lambda_p(\mu_t^N)$ separately. We also contribute a novel 'concentration of moments' result (Proposition 2.10), which allows a different kind of control of the moments of the Kac process; we show that for $p>2$ and sufficiently large b, if the initial data have enough moments, then $\mathbb{P}(\Lambda_p(\mu_t^N)>b)\to 0$ at a rate $\mathcal{O}(N^{-1})$, which cannot be argued solely by bounding $\mathbb{E}[\Lambda_q(\mu_t^N)]$. This will be useful for applications throughout the thesis where we split into events in this way and wish to bound the probability of the 'bad' event where $\Lambda_p(\mu_t^N)$ is large.

We will also use some results on the appearance of *exponential moments* $\langle e^{a|v|^\rho},\mu_t\rangle$ for both the Boltzmann and Landau equations (BE, LE). The study of such moments goes back to Bobylev [20] who proved that, if a moment $\langle e^{a|v|^2},\mu_0\rangle$ is initially finite, then it is propagated in time, at least in the case of cutoff hard potentials (CHP$_K$). The work of Lu and Mouhot [133] proved that exponential moments $\langle e^{\varepsilon|v|^\gamma},\mu_t\rangle$ instantaneously become finite even if they are not initially so, in the cases of cutoff or noncutoff hard potentials, and this has also been investigated by Tasković, Alonso, Gamba and Palović [179], Fournier and Mouhot [93], and Alsono, Cañizo, Gamba and Mouhot [11].

2.5.1 Polynomial Moment Estimates for the Boltzmann Equations

We begin with moment inequalities for the Boltzmann equation. To avoid proving multiple similar propositions for the three kernels (HS, rHS, NCHP) and variants of them, we will prove the moment estimates under the following general assumption.

Proposition 2.6 (Moment Inequalities for the Boltzmann Equation)**.** *Let B be any kernel of the form $B(v,\theta)=\Psi(|v|)b(\cos\theta)$ satisfying, for some $a>0$,*

$$a^{-1}|v|^\gamma\le\Psi(|v|)\le a(1+|v|^\gamma)$$

and

$$\int_{\mathbb{S}^{d-1}}\theta b(\cos\theta)d\sigma\in(0,\infty);\qquad \int_{\mathbb{S}^{d-1}}\mathbb{1}\left\{\frac{\pi}{3}<\theta<\frac{2\pi}{3}\right\}b(\cos\theta)d\sigma>0.$$

Let $(\mu_t)_{t\ge0}$ be any solution to (BE) whose energy $\Lambda_2(\mu_t)=\langle|v|^2,\mu_t\rangle$ is finite and constant (not necessarily normalised to 1). Then the following holds.

i). For all $q\ge p\ge2$ and $t>0$, we have

$$\Lambda_q(\mu_t)\le C_{p,q}(1+t^{(p-q)/\gamma})\Lambda_p(\mu_0). \tag{2.53}$$

Here, $C_{p,q}$ is a constant, depending only on p,q as well as $\gamma,\Lambda_2(\mu_0)$, an upper bound for a, and upper and lower bounds for $\int_{\mathbb{S}^{d-1}}\theta b(\cos\theta)d\sigma$. In particular, all moments $\Lambda_p(\mu_t)$ are finite, uniformly away from $t=0$, and if $\Lambda_p(\mu_0)$ is finite, then $\Lambda_p(\mu_t)$ is bounded, uniformly in time.

ii). For any $p > 2$, if $\Lambda_p(\mu_0)$ is finite, then for some constant C_p with the same dependence as above, for any $t \geq 0$,

$$\int_0^t \Lambda_{p+\gamma}(\mu_u)du \leq C(1+t)\Lambda_p(\mu_0). \tag{2.54}$$

In particular, $\mu \in L^1_{loc}([0,\infty), \mathcal{P}_{p+\gamma}(\mathbb{R}^d))$.

Remark 2.7. *We make the following remarks.*

i) The range $\theta \in (\frac{\pi}{3}, \frac{2\pi}{3})$ in the second item is not essential, and one could equally well replace it with any interval bounded away from $\{0, \pi\}$. The hypothesis as it is written is certainly satisfied in all of the cases of interest.

ii) The second item has to be included separately and cannot be deduced from the first, since we would find a logarithmically diverging integral $\int_0 s^{-1} ds$.

iii) This choice of kernels includes the cases of interest (HS, rHS, NCHP). This general form is also useful when we replace (rHS) with a mild variant (rHS$_\delta$) given by $\Psi(r) = 1 + r\delta$ in Chapter 6.

iv) In the case of Grad's angular cutoff B_K in the case (CHP$_K$), this statement ensures not only that the same estimates apply, but are also uniform as soon as K is bounded away from 0.

v) The following argument follows the approaches of the works [53, 132], incorporating also some elements of [157] to allow any $p \in [2, q]$ in i). In the case of Maxwell Molecules (GMM), Truesdell [115, 184] proved that no moment creation occurs; moments which are initially infinite remain so for all time.

We will use the following as a first step, which gives a qualitative statement of the moment production property and justifies the manipulations.

Proposition 2.8. *Continue in the notation of Proposition 2.6 above. Then for all $t > 0$ and all $p \geq 2$, we have $\mu \in L^\infty_{loc}((0,\infty), \mathcal{P}_p(\mathbb{R}^d))$. Moreover, if $\Lambda_p(\mu_0) < \infty$, then in fact $\mu \in L^\infty_{loc}([0,\infty), \mathcal{P}_p(\mathbb{R}^d))$ and $\Lambda_p(\mu_t) \to \Lambda_p(\mu_0)$ as $t \downarrow 0$.*

For this, we will follow the arguments of Mischler and Wennberg [144, Theorem 1.1', Steps 1-2]; although the uniqueness results of this paper concern only cutoff hard potentials, we will check that the argument holds under our hypotheses, whether the kernel is cutoff or not. Since the argument is a more careful and precise version of the same sort of Povzner inequality we use in the proof of Proposition 2.6, it is more enlightening to first give the proof of Proposition 2.6 taking this for granted, and return to give the justification for this later.

Proof of Proposition 2.6 from Proposition 2.8. Fix, throughout, an energy-conserving solution $(\mu_t)_{t\geq 0}$ to the Boltzmann equation for the kernel B described. If the kinetic energy $\Lambda_2(\mu_t)=0$, then μ_t is always a point mass, and all the conclusions are trivial; otherwise, by rescaling, we assume that $\Lambda_2(\mu_t)=1$ for all $t\geq 0$.

Step 1. Povzner Inequality Let us begin with the Povzner inequality; fix $p\geq 2$ and apply $\mathcal{L}_{\mathrm{B,s}}$ to $f(v)=|v|^p$. For given given $v, v_\star$, we use the parametrisation of jumps in terms of θ,φ given in Section 2.4 and write

$$v' = v\left(\frac{1+\cos\theta}{2}\right)+v_*\left(\frac{1-\cos\theta}{2}\right)+\frac{\sin\theta}{2}\Gamma(v-v_*,\varphi) \tag{2.55}$$

recalling that $|\Gamma(v-v_*,\varphi)|=|v-v_*|$ and $\Gamma(v-v_*,\varphi)\cdot(v-v_*)=0$. From orthogonality, it follows that $v\cdot\Gamma(v-v_*,\varphi)=v_*\cdot\Gamma(v-v_*,\varphi)$, so

$$|v\cdot\Gamma(v-v_*,\varphi)|\leq\min(|v|,|v_*|)|\Gamma(v-v_*,\varphi)|\leq\min(|v|,|v_*|)(|v|+|v_*|)\leq 2|v||v_*|. \tag{2.56}$$

We now return to (2.55) and take the norm of both sides to see that

$$\begin{aligned}|v'|^2&=\left(\frac{1+\cos\theta}{2}\right)|v|^2+\left(\frac{1-\cos\theta}{2}\right)|v_\star|^2+\sin\theta\, v\cdot\Gamma(v-v_*,\varphi)\\ &:=h_1+h_2+h_3.\end{aligned} \tag{2.57}$$

We now raise both sides to the $(p/2)^{\text{th}}$ power, recalling the inequality $(x+y)^{p/2}\leq x^{p/2}+y^{p/2}+C(xy^{p/2-1}+x^{p/2-1}y)$, valid for all $x,y>0$. It is straightforward to see that the cross terms are dominated by

$$h_1^{p/2-1}(h_2+h_3)+h_1(h_2+h_3)^{p/2-1}\leq C(|v|^{p-1}|v_\star|+|v||v_\star|^{p-1})\sin\theta; \tag{2.58}$$

$$h_2^{p/2-1}h_3+h_2h_3^{p/2-1}\leq C(|v|^{p-1}|v_\star|+|v||v_\star|^{p-1})\sin\theta; \tag{2.59}$$

$$h_3^{p/2}\leq C(|v|^{p-1}|v_\star|+|v||v_\star|^{p-1})\sin\theta. \tag{2.60}$$

We thus obtain

$$\begin{aligned}|v'|^p&\leq h_1^{p/2}+h_2^{p/2}+C(|v|^{p-1}|v_\star|+|v||v_\star|^{p-1})\sin\theta\\ &\leq\left(\frac{1+\cos\theta}{2}\right)^{p/2}|v|^p+\left(\frac{1-\cos\theta}{2}\right)^{p/2}|v_*|^p+C(|v|^{p-1}|v_*|+|v||v_*|^{p-1})\sin\theta.\end{aligned} \tag{2.61}$$

From this, and a similar inequality for $|v_*'|^p$, we obtain

$$|v'|^p+|v_*'|^p-|v|^p-|v_*|^p\leq-\lambda(p,\theta)\,(|v|^p+|v_\star|^p)+C_p(|v|^{p-1}|v_\star|+|v_\star|^{p-1}|v|)\sin\theta \tag{2.62}$$

with

$$\lambda(p,\theta) = \left(1 - \left(\frac{1+\cos\theta}{2}\right)^{p/2} - \left(\frac{1-\cos\theta}{2}\right)^{p/2}\right) \geq 0 \tag{2.63}$$

which is nonnegative for all $p \geq 2, \theta \in (0,\pi)$, with strict positivity unless $p = 2$ or $\theta \in \{0,\pi\}$. Integrating against the angular directions and using the hypothesis that $\int \theta b(\cos\theta) d\sigma < \infty$, we find that the function $f(v) = |v|^p$ has

$$\mathcal{L}_{\mathrm{B,s}} f(v,v_*) \leq -\lambda_p \Psi(|v-v_*|)(|v|^p + |v_*|^p) + C\Psi(|v-v_*|)(|v|^{p-1}|v_*| + |v||v_*|^{p-1}) \tag{2.64}$$

with

$$\lambda_p = \lambda_p(B) = \int_{\mathbb{S}^{d-1}} \left(1 - \left(\frac{1+\cos\theta}{2}\right)^{p/2} - \left(\frac{1-\cos\theta}{2}\right)^{p/2}\right) b(\cos\theta) d\sigma > 0 \tag{2.65}$$

and, observing that $\lambda(p,\theta)/\theta$ is bounded and bounded away from 0, we find that λ_p is bounded away from 0, depending only on a lower bound for the mass between $\theta = \frac{\pi}{3}, \theta = \frac{2\pi}{3}$, and similarly C is bounded above depending only on an upper bound for $\int \theta B(u,\sigma) d\sigma$.

We next expand the kinetic factors in (2.64). Using the lower bound on Ψ and that $|v-v_*|^\gamma \geq |v|^\gamma - |v_*|^\gamma$, we bound the first term above by

$$-\lambda_p a^{-1}|v-v_*|^\gamma(|v|^p + |v_*|^p) \leq -\lambda_p a^{-1}(|v|^\gamma - |v_*|^\gamma)|v|^p - \lambda_p a^{-1}(|v_*|^\gamma - |v|^\gamma)|v_*|^p.$$

Similarly, for the positive term, we use the upper bound to see that $\Psi(|v-v_*|) \leq a(1 + |v|^\gamma + |v_*|^\gamma)$. All together, for some λ, C depending now also on a,

$$\mathcal{L}_{\mathrm{B,s}} f(v,v_*) \leq -\lambda(|v|^{p+\gamma} + |v_*|^{p+\gamma}) + Cg(v,v_*) \tag{2.66}$$

where g contains the collected lower order terms

$$\begin{aligned} g(v,v_*) = {} & |v|^p|v_*|^\gamma + |v|^\gamma|v_*|^p + |v|^{p-1}|v_*| + |v||v_*|^{p-1} \\ & + |v|^{p-1+\gamma}|v_*| + |v|^{1+\gamma}|v_*|^{p-1} + |v|^{p-1}|v_*|^{1+\gamma} + |v||v_*|^{p-1+\gamma} \end{aligned} \tag{2.67}$$

which is the (basic) Povzner inequality for the class of kernels in this Proposition.

Step 2. Truncation Argument We next integrate the bound on $\mathcal{L}_{\mathrm{B,s}} f$ found in the previous step, from which we find a bound on $\langle f, Q(\mu_t)\rangle$. We now carefully show that we have $\frac{d}{dt}\Lambda_p(\mu_t) = \frac{d}{dt}\langle f, \mu_t\rangle = \langle f, Q(\mu_t)\rangle$ for $t \in (0,\infty)$, which amounts to taking our choice $f = |v|^p$ in (BE); however, since f does not belong to the class of test functions in the definition of weak solutions (see Definition 1.1.1), this must be shown separately.

We use a trunctation argument, see also [132, Theorem 1.3, Step 1] for a very similar argument. Let us fix $f(v) = |v|^p$ as in the previous step, and a smooth, compactly supported function $\chi : \mathbb{R}^d \to [0,1]$ with $\chi(v) = 1$ when $|v| \le 1$, and $\chi(v) = 0$ when $|v| \ge 2$. We now set, for $R \ge 1$, $f_R(v) := f(v)\chi(v/R)$, so that f_R are Lipschitz and compactly supported, so that we can apply (BE) directly to these functions to find that, for any $0 < s < t$,

$$\langle f_R, \mu_t\rangle - \langle f_R, \mu_s\rangle = \int_s^t \int_{\mathbb{R}^d\times\mathbb{R}^d} \mathcal{L}_{\mathrm{B,s}} f_R(v,v_*)\mu_u(dv)\mu_u(dv_*)du. \tag{2.68}$$

We now take the limit $R \to \infty$. For any fixed v, v_*, we have that $\mathcal{L}_{\mathrm{B,s}} f_R(v,v_*) = \mathcal{L}_{\mathrm{B,s}} f(v,v_*)$ as soon as $R \ge |v| + |v_*|$, so we have pointwise convergence of $\mathcal{L}_{\mathrm{B,s}} f_R \to \mathcal{L}_{\mathrm{B,s}} f$ in the limit $R \to \infty$. Next, we prove that we can bound $\mathcal{L}_{\mathrm{B,s}} f_R$ above, uniformly in R, by a function of polynomial growth. To see this, note that $|v|\nabla\chi(v/R)/R$ is bounded, uniformly in R, leading to the bound

$$|\nabla f_R(v)| = \left| p|v|^{p-1}\left(\frac{v}{|v|}\right)\chi(v/R) + \frac{|v|^p}{R}\nabla\chi(v/R)\right| \le C|v|^{p-1}$$

for some C independent of R. Next, we observe that the line segment $[v, v']$ lies inside the ball of radius $|v| + |v_*|$ and that $|v - v'| = |v - v_*|\sin\theta$ to obtain

$$|f_R(v') - f_R(v)| \le |v - v_*|\sin\theta \sup_{|w|\le|v|+|v_*|} |\nabla f_R(w)| \le C(|v|^p + |v_*|^p)\sin\theta$$

with C again independent of R. Using the same inequality for v_*, v_*' we integrate, recalling that $\int \theta B(u,\sigma)d\sigma = \int \theta b(\cos\theta)d\sigma < \infty$ to find that

$$|\mathcal{L}_{\mathrm{B,s}} f_R(v,v_*)| \le C\Psi(|v - v_*|)(|v|^p + |v_*|^p) \le C(1 + |v|^{p+\gamma} + |v_*|^{p+\gamma})$$

which is a polynomial upper bound as desired. Since we assume validity of Proposition 2.8, for the same choices of s, t in (2.68), the $(p+\gamma)^{\mathrm{th}}$ moment $\Lambda_{p+\gamma}(\mu_u)$ is bounded on $u \in [s,t]$, and we can use dominated convergence to see that

$$\int_s^t \int_{\mathbb{R}^d\times\mathbb{R}^d} \mathcal{L}_{\mathrm{B,s}} f_R(v,v_*)\mu_u(dv)\mu_u(dv_*)du \to \int_s^t \int_{\mathbb{R}^d\times\mathbb{R}^d} \mathcal{L}_{\mathrm{B,s}} f(v,v_*)\mu_u(dv)\mu_u(dv_*)du.$$

On the other hand, $\langle f_R, \mu_t\rangle \to \langle f, \mu_t\rangle \in (0,\infty)$, and similarly for s. We therefore take the limit of (2.68) to obtain

$$\langle f, \mu_t\rangle - \langle f, \mu_s\rangle = \int_s^t \int_{\mathbb{R}^d\times\mathbb{R}^d} \mathcal{L}_{\mathrm{B}} f(v,v_*)\mu_u(dv)\mu_u(dv_*) \tag{2.69}$$

and it follows that $\langle f, \mu_t\rangle = \Lambda_p(\mu_t)$ is weakly differentiable, with derivative

$$\frac{d}{dt}\Lambda_p(\mu_t) = \int_{\mathbb{R}^d\times\mathbb{R}^d} \mathcal{L}_{\mathrm{B,s}} f(v,v_*)\mu_t(dv)\mu_t(dv_*). \tag{2.70}$$

Using a similar truncation argument, using the local boundedness of moments of order strictly higher than $p+\gamma$ and the continuity of $t \mapsto \mu_t$ for the weak topology, it follows that $t \mapsto \int_{\mathbb{R}^d\times\mathbb{R}^d} \mathcal{L}_{\mathrm{B,s}} f(v,v_*)\mu_t(dv)\mu_t(dv_*)$ is continuous on $(0,\infty)$, and so (2.70) holds as a *classical* derivative, and $\Lambda_p(\mu_t) \in C^1(0,\infty)$.

Finally, we integrate the upper bound for $\mathcal{L}_{\mathrm{B,s}} f$ obtained from the Povzner inequality in (2.66). From the normalisation that $\Lambda_2(\mu_t) = 1$ it follows that $q \mapsto \Lambda_q(\mu_t)$ is increasing on $q \in [2,\infty)$, and since $p+\gamma-1 \le p$, the integral of the lower order terms $g(v,v_*)$ is at most

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} g(v,v_*)\mu_t(dv)\mu_t(dv_*) \le C\Lambda_p(\mu_t)$$

and we finally obtain

$$\frac{d}{dt}\Lambda_p(\mu_t) \le -\lambda\Lambda_{p+\gamma}(\mu_t) + C\Lambda_p(\mu_t). \tag{2.71}$$

Step 3. Derivation of Moment Inequalities We now derive the moment inequalities in the Proposition from (2.71), following [157, Proposition 3.1]. For item i), we deal first with the case $q = p$, supposing that $\Lambda_p(\mu_0) < \infty$ and write, for $t \ge 0$ define

$$F_p(t) := \Lambda_p(\mu_t).$$

First, for the case $q = p$, we estimate $F_{p+\gamma}(t) \ge F_p(t)^{1+\gamma/p}$ by Hölder's inequality, and return to (2.71) to find

$$\frac{d}{dt}F_p(t) \le -\lambda F_p(t)^{1+\gamma/p} + CF_p(t). \tag{2.72}$$

We now note that this is a differential inequality valid with $F_p \in C^1((0,\infty)) \cap C([0,\infty))$ thanks to Proposition 2.8 again, and that the right-hand side is negative as soon as $F_p \ge (C/\lambda)^{p/\gamma}$, whence we have the global bound

$$F_p(t) \le \max\left(F_p(0), \left(\frac{C}{\lambda}\right)^{p/\gamma}\right) \le \max\left(1, \left(\frac{C}{\lambda}\right)^{p/\gamma}\right) F_p(0)$$

which proves item i) in the case $q = p$. For $q > p$, we define similarly to F_p the function

$$F_{p,q}(t) := \frac{\Lambda_q(\mu_t)}{\sup_{s\ge 0}\Lambda_p(\mu_s)}$$

where the denominator is finite by the case $q = p$ above. Using Hölder's inequality on the probability measure $|v|^p\mu_t(dv)/F_p(t)$, we find

$$\frac{F_q(t)}{F_p(t)} = \left\langle |v|^{q-p}, \frac{|v|^p\mu_t}{F_p(t)}\right\rangle \le \left\langle |v|^{q-p+\gamma}, \frac{|v|^p\mu_t}{F_p(t)}\right\rangle^{(q-p)/(q-p+\gamma)} = \left(\frac{F_{q+\gamma}(t)}{F_p(t)}\right)^{(q-p)/(q-p+\gamma)} \tag{2.73}$$

which rearranges to

$$F_{q+\gamma}(t) \ge F_p(t)^{-\gamma/(q-p)} F_q(t)^{1+\frac{\gamma}{q-p}}.$$

Returning to (2.71), we find in this case the differential inequality on $(0,\infty)$

$$\frac{d}{dt}F_{p,q}(t) \le -\lambda F_{p,q}(t)^{1+\frac{\gamma}{q-p}} + CF_{p,q}(t) =: h(F_{p,q}(t)). \tag{2.74}$$

Set $u_\star = (C/\lambda)^{(q-p)/\gamma}$, so that $h(r) \le 0$ for $r \in [u_\star, \infty)$ and $h(u_\star) = 0$.

We now fix $t \ge 0$ and let $t_0 \in (0,t)$, and consider the cases $F_{p,q}(t_0) \le u_\star, F_{p,q}(t_0) > u_\star$ separately. In the first case $F_{p,q}(t_0) \le u_\star$, it follows that $F_{p,q}(s) \le u_\star$ for all $s \ge u_\star$, and in particular $F_{p,q}(t) \le u_\star$. Otherwise, in the case $F_{p,q}(t_0) > u_\star$, we observe that the C^1 function $g : (u_\star, \infty) \to \mathbb{R}$ given by

$$g(u) = -\frac{q-p}{\gamma C} \log \left(1 - \frac{C u^{-\gamma/(q-p)}}{\lambda}\right)$$

satisfies

$$\frac{d}{du} g(u) = \frac{1}{Cu - \lambda u^{1+\frac{\gamma}{q-p}}} = -\frac{1}{h(u)} .$$

Now, set $t_1 = \inf\{s > t_0 : F_{p,q}(s) = u_\star\}$. On the interval (t_0, t_1), the right-hand side of (2.74) is negative, and we integrate to find $\frac{d}{ds} g(F_{p,q}(s)) \ge 1$ and hence

$$g(F_{p,q}(s)) \ge g(F_{p,q}(t_0)) + (s - t_0) \ge (s - t_0)$$

on (t_0, t_1), which rearranges to

$$F_{p,q}(s) \le \left(\frac{\lambda}{C}\left(1 - e^{-C\gamma(s-t_0)/(q-p)}\right)\right)^{(p-q)/\gamma} .$$

On the other hand, in the case where $t_1 < \infty$, it holds that $F_{p,q}(s) \le u_\star$ for $s \ge t_1$ for the same reasons as above, so we conclude that for any $s \ge t_0$,

$$\begin{aligned} F_{p,q}(s) &\le \max\left(u_\star, \left(\frac{\lambda}{C}\left(1 - e^{-C\gamma(s-t_0)/(q-p)}\right)\right)^{(p-q)/\gamma}\right) \\ &\le u_\star + \left(\frac{\lambda}{C}\left(1 - e^{-C\gamma(s-t_0)/(q-p)}\right)\right)^{(p-q)/\gamma} \end{aligned}$$

and in particular, this applies at $s = t$. Recalling the definition of $F_{p,q}$, we absorb $u_\star$ into a constant C depending only on the quantities in the statement of the Proposition, and take $t_0 \downarrow 0$ to find that

$$\begin{aligned} \Lambda_q(\mu_t) &\le C\left(1 + \left(\frac{\lambda}{C}\left(1 - e^{-C\gamma t/(q-p)}\right)\right)^{(p-q)/\gamma}\right) \sup_{s\ge 0} \Lambda_p(\mu_s) \\ &\le C\left(\frac{\lambda}{C}\left(1 - e^{-C\gamma t/(q-p)}\right)\right)^{(p-q)/\gamma} \Lambda_p(\mu_0) \end{aligned} \tag{2.75}$$

using, in the final inequality, the bound $\sup_{u\ge 0} \Lambda_p(\mu_u) \le C\Lambda_p(\mu_0)$, for some new choice of C. The general case of point i) now follows by observing that

$$(1 - e^{-C\gamma t/(q-p)})^{(p-q)/\gamma} \sim t^{(p-q)/\gamma}$$

as $t \downarrow 0$, so

$$(1 - e^{-C\gamma(t-s)/(q-p)})^{(p-q)/\gamma} \le C(1 + t^{(p-q)/\gamma}).$$

For item ii), again assuming that $\Lambda_p(\mu_0) < \infty$, we return to (2.71) and integrate over a time interval $[s,t]$ with $s > 0$; this is licit because, thanks to Proposition 2.8, all moments are finite on this time interval. We find

$$\int_s^t \Lambda_{p+\gamma}(\mu_u)du \le \lambda^{-1}\left(\Lambda_p(\mu_t) + C\int_s^t \Lambda_p(\mu_u)du\right).$$

Thanks to item i), we can replace $\Lambda_p(\mu_t), \Lambda_p(\mu_u) \le C\Lambda_p(\mu_0)$ to replace the upper bound by $C(1+t)\Lambda_p(\mu_0)$, uniformly in s, and taking $s \downarrow 0$ gives the result. □

It remains to prove Proposition 2.8, which we used above to justifying the formal manipulations.

Proof. We follow [144, Theorem 1.1', Steps 1-2], which modifies with little alteration to the kernels with which we work, and to allow dimension $d \ge 3$.

Step 1. General Povzner Inequality We start with a general inequality, which generalises the calculations for the functions $f_p(v) = |v|^p$ we found earlier. We fix a convex, Lipschitz function $h : [0,\infty) \to [0,\infty)$ such that h' is Lipschitz continuous; it follows that h' is weakly differentiable, and the (weak) second derivative h'' is nonnegative almost everywhere, and apply $\mathcal{L}_{\mathrm{B,s}}$ to $f_h(v) := h(|v|^2)$.

Let us fix v, v_*, and assume that v is not parallel to $v - v_*$. In this case, we reparametrise the sphere $\mathbb{S}^{d-2}$ by an isometry R, preserving the uniform measure, so that $\Gamma(v - v_*, e_1)$ is parallel to the projection of v onto $v - v_*$. We now recall (2.57) to see that

$$\begin{aligned} |v'|^2 &= \left(\frac{1+\cos\theta}{2}\right)|v|^2 + \left(\frac{1-\cos\theta}{2}\right)|v'|^2 + (\sin\theta)(v\cdot\Gamma(v-v_*,e_1))\varphi_1 \\ &= Y(\theta) + Z(\theta)\varphi_1 \end{aligned} \tag{2.76}$$

where this defines $Y(\theta), Z(\theta)$, and similarly $|v_*'|^2 = Y(\pi-\theta) - Z(\theta)\varphi_1$. The same is true in the case where v is parallel to $v - v_*$, since in this case the last term is identically 0 in both cases. We also observe, recalling (2.56), that $Z(\theta) \le |v||v_*|\sin\theta$ and, since $Y(\theta) + Z(\theta)\varphi_1 \ge 0$ for all φ, we must have $|Z(\theta)| \le Y(\theta)$.

We next fix θ and consider $\int_{\mathbb{S}^{d-2}} f_h(|v'|^2)d\varphi = \int_{\mathbb{S}^{d-2}} h(Y + Z\varphi_1)d\varphi$, surpressing the θ-dependence of $Y(\theta), Z(\theta)$. Away from the poles, we parametrise $\varphi \in \mathbb{S}^{d-2}$ by $\psi \in (0,\pi)$ and $w \in \mathbb{S}^{d-3}$, and recall (2.42) to write

$$\begin{aligned} \int_{\mathbb{S}^{d-2}} h(Y + Z\varphi_1)d\varphi &= h(Y) + \int_0^\pi (h(Y + Z\cos\psi) - h(Y))c_{d-1}^{-1}(\sin\psi)^{d-3}d\psi. \\ &= h(Y) + \int_0^{\pi/2} (h(Y+Z\cos\psi) - 2h(Y) + h(Y - Z\cos\psi))c_{d-1}^{-1}(\sin\psi)^{d-3}d\psi \end{aligned} \tag{2.77}$$

where again $c_{d-1} = \int_0^\pi (\sin\psi)^{d-3}$ is the respective normalising constant[3]. We integrate twice by parts to find

$$\begin{aligned}
\int_{\mathbb{S}^{d-2}} & (h(Y+Z\varphi_1)-h(Y))d\varphi \\
&= \left[\psi(\sin\psi)^{d-3}(h(Y+Z\cos\psi)-2h(Y)+h(Y-Z\cos\psi))\right]_{\psi=0}^{\pi/2} \\
&\qquad + Z\int_0^{\pi/2}\psi(\sin\psi)^{d-2}(h'(Y+Z\cos\psi)-h'(Y-Z\cos\psi))d\psi \\
&= \left[\psi(\sin\psi)^{d-2}(h'(Y+Z\cos\psi)-h'(Y-Z\cos\psi)\right]_{\psi=0}^{\pi/2} \\
&\qquad + Z^2\int_0^{\pi/2} g(\psi)\sin\psi(h''(Y+Z\cos\psi)+h''(Y-Z\cos\psi))d\psi \\
&= Z^2\int_0^\pi g(\psi)\sin\psi h''(Y+Z\cos\psi)d\psi
\end{aligned} \tag{2.78}$$

where we define a bounded function g by $g(\psi) := \int_0^\psi u(\sin u)^{d-2}du$ on $[0,\frac{\pi}{2}]$ and $g(\psi) := g(\pi-\psi)$ on $(\frac{\pi}{2},\pi]$. From this, and the same calculation for $h(|v_*'|^2)$, we find that

$$\begin{aligned}
\int_{\mathbb{S}^{d-2}} & (h(|v'|^2)+h(|v_*'|^2)-h(|v|^2)-h(|v_*|^2))d\varphi \\
&= h(Y(\theta))+h(Y(\pi-\theta))-h(|v|^2)-h(|v_*|^2) \\
&\qquad + Z(\theta)^2\int_0^\pi g(\psi)\sin\psi h''(Y(\theta)+Z(\theta)\cos\psi)d\psi \\
&\qquad + Z(\theta)^2\int_0^\pi g(\psi)\sin\psi h''(Y(\pi-\theta)+Z(\theta)\cos\psi)d\psi.
\end{aligned} \tag{2.79}$$

If we now integrate over the θ direction, we find overall that

$$\frac{\mathcal{L}_{\mathrm{B,s}} f_h(v,v_*)}{\Psi(|v-v_*|)} = P_h(v,v_*) - K_h(v,v_*) \tag{2.80}$$

where $P_h \geq 0$ comes from integrating the last two lines of (2.79), or equivalently the left-hand side of (2.78) with respect to $\beta(\theta)d\theta$, and K_h is given by

$$K_h(v,v_*) = \int_0^\pi (h(|v|^2)+h(|v_*|^2)-h(Y(\theta))-h(Y(\pi-\theta)))\beta(\theta)d\theta. \tag{2.81}$$

Recalling the definition of $Y(\theta)$ and that $Z(\theta) \leq |v||v_*|\sin\theta$, both integrals converge. Recalling also that h is convex, we have

$$h(Y(\theta)) \leq \left(\frac{1+\cos\theta}{2}\right)h(|v|^2) + \left(\frac{1-\cos\theta}{2}\right)h(|v_*|^2)$$

so that $K_h \geq 0$. Moreover, $h \mapsto P_h, K_h$ are linear in h, so that if h', h are two such functions, such that $h'-h$ is nonnegative and convex, we have that

$$P_{h'} = P_h + P_{h'-h} \geq P_h; \qquad K_{h'} = K_h + K_{h'-h} \geq K_h. \tag{2.82}$$

[3] To see this, set $h=1$; the left-hand side is 1, since $d\varphi$ is a probability measure.

2. Application with Approximated Convex Functions We now carefully show, without any integrability hypotheses beyond the conservation of energy, how we find an integral equality for the evolution of $\langle h(|v|^2), \mu_t\rangle$, for any function h satisfying the hypotheses of the previous part. For any n, we let $p_n(x) = h(n) + (x-n)h'(n)$ be the affine function tangent to h at n, and define h_n by

$$h_n(x) = \begin{Bmatrix} h(x) & x \le n \\ p_n(x) & x > n \end{Bmatrix} = \widetilde{h}_n(x) + p_n(x).$$

The function $f_n(v) = \widetilde{h}_n(|v|^2) = h_n(|v|^2) - p_n(|v|^2)$ is compactly supported, and satisfies all of the conditions of the first step, so we can write, for all $t > 0$,

$$\begin{aligned} \langle f_n, \mu_t\rangle + \int_0^t \int_{\mathbb{R}^d\times\mathbb{R}^d} K_{\tilde{h}_n}(v,v_*)\Psi(|v-v_*|)\mu_s(dv)\mu_s(dv_*)ds \\ = \langle f_n, \mu_0\rangle + \int_0^t \int_{\mathbb{R}^d\times\mathbb{R}^d} P_{\tilde{h}_n}(v,v_*)\Psi(|v-v_*|)\mu_s(dv)\mu_s(dv_*)ds. \end{aligned} \tag{2.83}$$

Using the definitions it is immediate to see that the addition of the linear term p_n does not change G, K, so we can replace the integrands with P_{h_n}, K_{h_n}. On ther other hand, $\langle |v|^2, \mu_0\rangle = \langle |v|^2, \mu_t\rangle$, so the same is true for the linear combination $\langle p_n(|v|^2), \mu_t\rangle = \langle p_n(|v|^2), \mu_0\rangle$. Adding this term to both sides, we can replace f_n with $h_n(|v|^2)$ to obtain

$$\begin{aligned} \langle h_n(|v|^2), \mu_t\rangle + \int_0^t \int_{\mathbb{R}^d\times\mathbb{R}^d} K_{h_n}(v,v_*)\Psi(|v-v_*|)\mu_s(dv)\mu_s(dv_*)ds \\ = \langle h_n(|v|^2), \mu_0\rangle + \int_0^t \int_{\mathbb{R}^d\times\mathbb{R}^d} P_{h_n}(v,v_*)\Psi(|v-v_*|)\mu_s(dv)\mu_s(dv_*)ds. \end{aligned} \tag{2.84}$$

We now carefully take the limit $n \to \infty$. We note that $h_{n+1} - h_n \ge 0$ is convex, so by the remark at the end of step 1, $P_{h_{n+1}} \ge P_{h_n} \ge 0, K_{h_{n+1}} \ge K_{h_n} \ge 0$, which increase upwards to $P_h, K_h \ge 0$ respectively, while $h_n(|v|^2) \uparrow h(|v|^2) =: f_h(v)$. We therefore use monotone convergence on both sides to obtain, for all $t > 0$,

$$\begin{aligned} \langle f_h, \mu_t\rangle + \int_0^t \int_{\mathbb{R}^d\times\mathbb{R}^d} K_h(v,v_*)\Psi(|v-v_*|)\mu_s(dv)\mu_s(dv_*)ds \\ = \langle f_h, \mu_0\rangle + \int_0^t \int_{\mathbb{R}^d\times\mathbb{R}^d} P_h(v,v_*)\Psi(|v-v_*|)\mu_s(dv)\mu_s(dv_*)ds \end{aligned} \tag{2.85}$$

possibly understanding this as an equality $\infty = \infty$ if either (and therefore both) side fails to converge.

Step 3. Application with slowly-growing h. We now apply the previous step with a well-chosen h. With a modification of de La Vallée Poussin theorem, we can find a strictly convex, C^2 function $h : [0,\infty) \to [0,\infty)$ with $h'(\infty) = \infty$ such that $\langle h(|v|^2), \mu_0\rangle < \infty$, and by replacing h by a more slowly growing function if necessary, write h in the form

$h(r) = rj(r)$ for a concave, increasing function j with $j(\infty) = \infty$, but increasing slowly enough that $j'(r) \le (1+r)^{-1}$ and, for all $\varepsilon > 0, \alpha \in (0,1)$, there exists some constant $C = C(\alpha, \epsilon)$ such that

$$j(r) - j(\alpha r) \ge C(1+r)^{-\epsilon} \quad \text{for all } r > 1. \tag{2.86}$$

In this case, we bound P_h from the first expression $\int_{\mathbb{S}^{d-2}}(h(Y(\theta) + \varphi_1 Z(\theta)) - h(Y(\theta))d\varphi$ appearing in (2.78); since j is concave, we have

$$j(Y(\theta) + Z(\theta)\varphi_1) \le j(Y) + Z\varphi_1 j'(Y)$$

and so we bound the integrand, suppressing the θ-argument

$$\begin{aligned} h(Y + Z\varphi_1) - h(Y) &\le Yj(Y) + Z\varphi_1(j(Y) + Yj'(Y)) + Z^2\varphi_1^2 j'(Y) - Yj(Y) \\ &= Z\varphi_1(j(Y) + Yj'(Y)) + Z^2\varphi_1^2 j'(Y). \end{aligned}$$

When we integrate over $\varphi \in \mathbb{S}^{d-2}$, the first term on the second line integrates to 0, and the second produces $Z^2(\theta)j'(Y(\theta))/(d-1)$. Recalling that $Z(\theta) \le Y(\theta)$, the hypothesis that $j' \le (1+r)^{-1}$ implies that we can absorb one factor of Z into $Zj'(Y) \le 1$ and obtain, for some constant c,

$$P_h(v, v_*) \le c\int_0^\pi Z(\theta)\beta(\theta)d\theta \le c|v||v_*|.$$

Now using the upper bound on Ψ, we can bound $\Psi(|v - v_*|)P_h(v, v_*)$ by

$$\Psi(|v - v_*|)P_h(v, v_*) \le C(1 + |v|^{1+\gamma})(1 + |v_*|^{1+\gamma}) \le C(1 + |v|^2)(1 + |v_*|^2)$$

since $\gamma \le 1$, and integrating this in time produces

$$\int_0^t \int_{\mathbb{R}^d\times\mathbb{R}^d} P_h(v, v_*)\Psi(|v - v_*|)\mu_s(dv)\mu_s(dv_*) \le Ct(1 + \Lambda_2(\mu_0))^2 < \infty.$$

Returning to (2.85), the right-hand side is now finite for all $t \ge 0$, and indeed bounded on compact time intervals, from which it follows that $\langle f_h, \mu_t\rangle \in L^\infty_{\text{loc}}([0,\infty))$, that $\langle f_h, \mu_t\rangle \to \langle f_h, \mu_0\rangle$ as $t \downarrow 0$, and that, for all $t > 0$,

$$\int_0^t \int_{\mathbb{R}^d\times\mathbb{R}^d} K_h(v, v_*)\Psi(|v - v_*|)\mu_s(dv)\mu_s(dv_*) < \infty. \tag{2.87}$$

We next bound K_h *below*, hence finding a lower bound for the left-hand side of (2.87). A simple calculation shows that

$$\frac{d}{d\theta}(h(Y(\theta)) - h(Y(\pi - \theta))) = -\frac{\sin\theta}{2}(|v|^2 - |v_*|^2)(h'(Y(\theta)) - h'(Y(\pi - \theta)))$$

and recalling that h' is strictly increasing, this vanishes uniquely at $\theta = \pi/2$, which is therefore the unique maximum of $h(|v|^2) + h(|v_*|^2) - h(Y(\theta)) - h(Y(\pi - \theta))$. We bound the integral below by integrating only over $\theta \in (\pi/3, 2\pi/3)$; over this interval, the integrand is minimised at the endpoints $\theta \in \{\frac{\pi}{3}, \frac{2\pi}{3}\}$, which is the value

$$|v|^2 j(|v|^2)+|v_*|^2 j(|v_*|^2)-\left(\frac{3|v|^2+|v_*|^2}{4}\right) j\left(\frac{3|v|^2+|v_*|^2}{4}\right)-\left(\frac{3|v_*|^2+|v|^2}{4}\right) j\left(\frac{3|v_*|^2+|v|^2}{4}\right).$$

If we assume that $|v| \geq 3|v_*|$, we have that $\frac{3|v|^2}{4} \geq \frac{|v|^2}{4} \geq \frac{3|v_*|^2}{4}$, and we bound the arguments of j in the last two terms above by $\frac{7}{9}|v|^2$ to get an overall lower bound, for the same range of θ,

$$|v|^2\left(\left(j(|v|^2)-j\left(\frac{7|v|^2}{9}\right)\right)\right)-|v_*|^2 j\left(\frac{7|v|^2}{9}\right). \tag{2.88}$$

We now integrate over $\theta \in (\pi/3, 2\pi/3)$ to find a bound for K_h, for the same condition $|v_*| \leq 3|v|$. On the first term, we use (2.86) with $\alpha = \frac{7}{9}$, while in the second term we replace $|v_*|^2 \leq |v||v_*|$ to find, for any $\varepsilon > 0$, there exists $C < \infty$ such that, whenever $|v| > 3|v_*|$ and $|v| > 1$,

$$K_h(v, v_*) \geq C|v|^{2-\varepsilon} - |v||v_*|j(|v|^2). \tag{2.89}$$

If instead we have $|v_*| \geq 3|v|, |v_*| \geq 1$, we replace $v \leftrightarrow v_*$, and otherwise $K_h \geq 0$. Overall, up to a new choice of C,

$$K_h(v, v_*) \geq C(|v|^{2-\varepsilon}+|v_*|^{2-\varepsilon})(1-\mathbb{1}(|v|/3 \leq |v_*| \leq 3|v|))-|v||v_*|j(|v|^2)-|v||v_*|j(|v_*|^2)-C \tag{2.90}$$

and so

$$\begin{aligned} K_h(v, v_*)\Psi(|v-v_*|) \geq C(&|v|^{2+\gamma-\varepsilon}+|v_*|^{2+\gamma-\varepsilon})(1-\mathbb{1}(|v|/3 \leq |v_*| \leq 3|v|)) \\ &- C(1+|v|^2 j(|v|^2))(1+|v_*|^2 j(|v_*|^2)). \end{aligned} \tag{2.91}$$

We now return to (2.87). Thanks to the estimate $\langle f_h, \mu_t\rangle \in L^\infty_{\text{loc}}$ we found earlier, the negative terms are integrable, locally uniformly in time, while

$$\begin{aligned} C(|v|^{2+\gamma-\varepsilon}&+|v_*|^{2+\gamma-\varepsilon})(1-\mathbb{1}(|v|/3 \leq |v_*| \leq 3|v|)) \\ &\geq C(|v|^{2+\gamma-\varepsilon}+|v_*|^{2+\gamma-\varepsilon})-c|v|^{1+(\gamma-\varepsilon)/2}|v_*|^{1+(\gamma-\varepsilon)/2}. \end{aligned} \tag{2.92}$$

The second term is integrable, because $1+(\gamma-\varepsilon)/2 \leq 2$ and the second moments of μ_t are integrable, so we finally conclude that, for all $\varepsilon > 0$ and all $t < \infty$,

$$\int_0^t \int_{\mathbb{R}^d} |v|^{2+\gamma-\varepsilon}\mu_s(dv)ds < \infty. \tag{2.93}$$

In particular, $\mu \in L^1_{\text{loc}}([0,\infty), \mathcal{P}_{2+\gamma-\varepsilon}(\mathbb{R}^d))$, and $\Lambda_{2+\gamma-\varepsilon}(\mu_t) < \infty$ for almost all t.

Step 4. Inductive Argument on $(0, \infty)$**.** We now prove the local boundedness of all moments on $(0,\infty)$ by applying the general formulation in step 2 with $h(r) = r^{p/2}$ inductively.

Let us suppose, as our inductive assumption, that $\mu \in L^1_{\text{loc}}((0,\infty),\mathcal{P}_p(\mathbb{R}^d))$; the previous step provides a base case $p = 2+(\gamma/2)$. It follows that, for any $t_2 > t_1 > 0$, we can find $t_0 \in [0,t_1)$ with $\Lambda_p(\mu_{t_1}) < \infty$, and we then apply (2.85) on $(\mu_{t+t_0})_{t\ge 0}$ with $h(r) = r^{p/2}$ to get, for all $t \ge t_0$,

$$\begin{aligned}\Lambda_p(\mu_t) + \int_{t_0}^t K_h(v,v_*)\Psi(|v-v_*|)\mu_s(dv)\mu_s(dv_*)ds \\ = \Lambda_p(\mu_{t_0}) + \int_{t_0}^t P_h(v,v_*)\Psi(|v-v_*|)\mu_s(dv)\mu_s(dv_*)ds\end{aligned} \tag{2.94}$$

as an equality of nonnegative, possibly infinite integrals. We now bound P_h, K_h as we did in the previous step. In this case, we get

$$h' = \frac{p}{2}r^{(p-2)/2}; \qquad h'' = \frac{p(p-2)}{4}r^{(p-4)/2} > 0.$$

We now bound P_h, separating the cases $p \in (2,4], p > 4$, which correspond to the cases where h'' is decreasing or increasing.

4i. Case i $p \in (2,4]$. In this case, we return to (2.79) to write

$$P_h(v,v_*) \le \int_0^\pi \beta(\theta)Z(\theta)^2 Y(\theta)^{(p-4)/2}\int_0^\pi g(\psi)\sin\psi\left(1-\frac{Z(\theta)\cos\psi}{Y(\theta)}\right)^{\frac{p-4}{2}} d\psi d\theta.$$

Since $|Z(\theta)| \le Y(\theta)$ and $g(\psi)\sin\psi$ is bounded, we use the fact that $\frac{p-4}{2} > -1$ to see the integral converges uniformly in θ, and we conclude that

$$P_h(v,v_*) \le C\int_0^\pi Z(\theta)^2 Y(\theta)^{(p-4)/2} \le C|Z(\theta)|^{p/2}\beta(\theta)d\theta$$

where in the final bound, we use again that $|Z| \le Y$ and that $r^{(p-4)/2}$ is noninceasing, thanks to the range of p. Recalling from the definition of $Z(\theta)$ that $|Z(\theta)| \le |v||v_*|\sin\theta$, as remarked below (2.76), we have

$$P_h(v,v_*) \le C|v|^{p/2}|v_*|^{p/2}\int_0^\pi \sin\theta\beta(\theta)d\theta \le C|v|^{p/2}|v_*|^{p/2}$$

and finally, using Young's inequality,

$$P_h(v,v_*) \le C(|v|^{p-\gamma}|v_*|^\gamma + |v_*|^{p-\gamma}|v|^\gamma). \tag{2.95}$$

4ii. Case ii. $p > 4$ In this case, we return to (2.78), and recall that $Z(\theta) \le \sin\theta|v||v_*|$, while $h''(Y(\theta)+Z(\cos\theta)) \le C(|v|^{p-4}+|v_*|^{p-4})$ for some C depending on p. It follows that

$$P_h(v,v_*) \le C(|v|^{p-2}|v_*|^2 + |v_*|^{p-2}) \le C(|v|^{p-1}|v_*| + |v_*|^{p-1}|v|) \tag{2.96}$$

for some C depending on B through $\int_{\mathbb{S}^{d-1}}\theta b(\cos\theta)d\sigma$.

In either case, we use (2.95, 2.96) respectively and the growth of Ψ, recalling that $\gamma \leq 1$. Overall, we get

$$P_h(v, v_*)\Psi(|v - v_*|) \leq C \left((1 + |v_*|^2)(1 + |v|^p) + (1 + |v|^2)(1 + |v_*|^p)\right)$$

so that, using the finiteness of the energy, the corresponding integral is bounded by

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} P_h(v, v_*)\Psi(|v - v_*|)\mu_t(dv)\mu_t(dv_*) \leq C(1 + \Lambda_p(\mu_t)).$$

The upper bound is locally integrable on $[t_0, \infty)$ by the induction hypothesis, and returning to (2.94), we find that

$$\Lambda_p(\mu_t) \leq \Lambda_p(\mu_{t_0}) + \int_{t_0}^{t} C(1 + \Lambda_p(\mu_s))ds. \tag{2.97}$$

It follows that $\Lambda_p(\mu_t)$ is locally bounded on $[t_0, \infty)$ and, using dominated convergence for an upper bound and lower-semicontinuity for a lower bound, $\Lambda_p(\mu_t) \to \Lambda_p(\mu_{t_0})$ as $t \downarrow t_0$. On the other hand, the same argument as leading to (2.62, 2.64) produces

$$K_h \geq \lambda_p(|v|^{p+\gamma} + |v_*|^{p+\gamma}) - C(|v|^{p-1}|v_*| + |v_*|^{p-1}|v|)$$

which implies that

$$K_h(v, v_*)\Psi(|v - v_*|) \geq \lambda_p(|v|^{p+\gamma} + |v_*|^{p+\gamma}) - C(1 + |v|^{p-1+\gamma} + |v_*|^{p-1+\gamma}).$$

The integrals of the negative terms are locally integrable on $[t_0, \infty)$ by the induction hypothesis, since $p - 1 + \gamma \leq p$, so we conclude from (2.94) that, for any $t_2 > t_1 > t_0$, we have

$$\int_{t_1}^{t_2} \Lambda_{p+\gamma}(\mu_s)ds \leq \int_{t_0}^{t_2} \Lambda_{p+\gamma}(\mu_s)ds < \infty.$$

Since $t_2 > t_1 > 0$ were chosen arbitrarily, we conclude that $\mu \in L^1_{\text{loc}}((0, \infty), \mathcal{P}_{p+\gamma}(\mathbb{R}^d))$, and the induction step is complete, allowing us to increase the moment index by γ, so that $\mu \in L^1_{\text{loc}}((0, \infty), \mathcal{P}_p(\mathbb{R}^d))$ for $p = 2 + (2n + 1)\gamma/2, n \in \mathbb{N}$. On the other hand, one trivially has $L^1_{\text{loc}}((0, \infty), \mathcal{P}_q(\mathbb{R}^d)) \subset L^1_{\text{loc}}((0, \infty), \mathcal{P}_p(\mathbb{R}^d))$ for any $q \leq p$, so we conclude that $\mu \in L^1_{\text{loc}}((0, \infty), \mathcal{P}_p(\mathbb{R}^d))$ for all p. This proves the first assertion of Proposition 2.8, since we showed in the course of the induction that if the p^{th} moments are locally integrable on $(0, \infty)$, then they are locally bounded on $(0, \infty)$.

Step 5. Local boundedness near 0 with additional hypotheses. To check the second assertion, we will sketch how to modify the previous proof if the initial data μ_0 has $\Lambda_p(\mu_0) < \infty$. In this case, we choose $\varepsilon > 0$ and $n \geq 1$ such that $p \in [2 - \varepsilon + n\gamma, 2 - \varepsilon + (n + 1)\gamma)$, and set $q_m = 2 - \varepsilon + m\gamma, m = 1, ..., n + 1$, so $p \in [q_n, q_{n+1})$. Using Step 3, we find that $\mu \in L^1_{\text{loc}}([0, \infty), \mathcal{P}_{q_1}(\mathbb{R}^d))$; we then follow step 4 inductively, with indexes $q_1, q_2, \ldots q_n$; since μ_0 has $p \geq q_m$ moments we can choose $t_0 = 0$ each time, to see that

$\Lambda_{q_n}(\mu_t)$ is locally bounded, and $\Lambda_{q_{n+1}}(\mu_t)$ is locally integrable near 0. Since $q_{n+1} \geq p$, it follows that $\Lambda_p(\mu_t)$ is locally integrable near 0, so we can apply the inductive step of Step 4, again with $t_0 = 0$, to obtain that $\Lambda_p(\mu_t)$ is locally bounded near 0, and as remarked below (2.97), that $\Lambda_p(\mu_t) \to \Lambda_0(\mu_0)$ as $t \downarrow 0$. The Proposition is complete. □

As a result of the more general formulation, we also obtain the following, see also [142, Lemma 6.3]

Proposition 2.9. *Let $(\mu_t)_{t\geq 0} \subset \mathcal{S}$ be a weak solution to (BE), for a kernel as in Propositions 2.6-2.8 and $\mu_0 \in \mathcal{S}^6$. Then for all $t \geq 0$ we have the bound*

$$\int_0^t \Lambda_{2+\gamma}(\mu_s)ds \leq C(1+t) + C\log \Lambda_6(\mu_0). \tag{2.98}$$

In particular, for any w, there exists C, p such that

$$\exp\left(w\int_0^t \Lambda_{2+\gamma}(\mu_s)ds\right) \leq e^{C(1+t)}\Lambda_p(\mu_0).$$

Proof. We follow Steps 1-3 of the previous proposition for the convex function $h(r) = r\log(1+r)$ with $j = \log(1+r)$ concave. Steps 1-2 still apply, as does the upper bound on P_h in step 3, so that $\Psi(|v-v_*|)P_h(v,v_*) \leq C(1+|v|^2)(1+|v_*|^2)$. For the lower bound of K_h, we follow the arguments; in the case $|v| \geq 3|v_*|$ we find the same bound (2.88) on the integrand of K_h for $\theta \in (\pi/3, \pi)$, which now simplifies to

$$|v|^2 \log\left(\frac{1+|v|^2}{1+7|v|^2/9}\right) - |v||v_*|\log\left(1+\frac{7|v|^2}{9}\right).$$

We conclude that, for $|v| \geq \max(3|v_*|, 1/2)$, we have

$$K_h(v,v_*) \geq c|v|^2 - C|v||v_*|\log(1+|v|^2)$$

for some c, C, and hence, for the same v, v_*,

$$K_h(v,v_*)\Psi(|v-v_*|) \geq c|v|^{2+\gamma} - C(1+|v|^{1+\gamma})\log(1+|v|^2)(1+|v_*|^{1+\gamma}).$$

Up to a new choice of C, we consider the cases $|v_*| \leq |v| \leq 3|v_*|$ and $|v_*| \leq |v| \leq \frac{1}{2}$ to find, for all v, v_* with $|v| \geq |v_*|$,

$$K_h(v,v_*)\Psi(|v-v_*|) \geq c|v|^{2+\gamma} - C|v|^{1+(\gamma/2)}|v_*|^{1+(\gamma/2)} - C - C(1+|v|^{1+\gamma})(1+\log(1+|v|^2))(1+|v_*|^{1+\gamma})$$

and a symmetric expression holds if $|v| \leq |v_*|$. Grouping everything, we can absorb all lower-order terms into $1+|v|^2$, so, for some new c, C,

$$K_h(v,v_*)\Psi(|v-v_*|) \geq c(|v|^{2+\gamma} + |v_*|^{2+\gamma}) - C(1+|v|^2)(1+|v_*|^2).$$

Integrating, it follows that

$$\begin{aligned}
2c\int_0^t \Lambda_{2+\gamma}(\mu_s)ds &\le \int_0^t \int_{\mathbb{R}^d\times\mathbb{R}^d} (K_h(v,v_*)\Psi(|v-v_*|) + C(1+|v|^2)(1+|v_*|^2))\mu_s(dv)\mu_s(dv_*)ds \\
&\le \langle f_h, \mu_0\rangle + 4Ct + \int_0^t \int_{\mathbb{R}^d\times\mathbb{R}^d} P_h(v,v_*)\Psi(|v-v_*|)\mu_s(dv)\mu_s(dv_*)ds \\
&\le \langle |v|^2\log(1+|v|^2), \mu_0\rangle + Ct
\end{aligned} \tag{2.99}$$

where we changed the value of C in the last line, and changing the value of C again,

$$\int_0^t \Lambda_{2+\gamma}(\mu_t)dt \le Ct + C\langle |v|^2\log(1+|v|^2), \mu_0\rangle.$$

To find the upper bound stated in the proposition, we split the integral on the right-hand side into the regions $1+|v|^2 \le r, 1+|v|^2 > r$ for some $r \ge 1$ to be chosen. The first region contributes at most $(\log r)\langle |v|^2, \mu_0\rangle = \log r$, and the contribution from the second region is bounded by

$$\frac{1}{r}\int_{\mathbb{R}^d} (1+|v|^2)^2\log(1+|v|^2)\mu_0(dv) \le \frac{1}{r}\int_{\mathbb{R}^d}(1+|v|^2)^3\mu_0(dv) \le \frac{16}{r}\Lambda_6(\mu_0).$$

If we now choose $r = \Lambda_6(\mu_0)$, we find the overall integral

$$\langle |v|^2\log(1+|v|^2), \mu_0\rangle \le 2\log\Lambda_6(\mu_0) + 16$$

and the first part of the proposition is proven. The second part follows, changing the value of C and using that $Cw\log\Lambda_6 \le \log\Lambda_{6cW}$ by Jensen's inequality, as soon as $Cw \ge 1$. □

2.5.2 Polynomial Moment Estimates for the Kac Process

We next consider equivalent polynomial moment estimates for the Kac process. In this case, we must also (separately) prove some maximal inequalities, as well as our novel 'concentration of moments' result.

Proposition 2.10. *Let B be a kernel as in Proposition 2.6, and let $(\mu_t^N)_{t\ge0}$ be a Kac process for this kernel, and $a_0 \ge 1$ such that $\Lambda_2(\mu_0^N) \le a_0$ almost surely. Then for all $q \ge p \ge 2$, we have the following bounds.*

i). There exists a constant $C_{p,q} < \infty$ such that, for all $t \ge 0$,

$$\mathbb{E}\left[\Lambda_q(\mu_t^N)\right] \le C_{p,q}(1+t^{(p-q)/\gamma})\mathbb{E}\left[\Lambda_p(\mu_0^N)\right]. \tag{2.100}$$

ii). For some constant C_q, for all $t_{\text{fin}} \ge 0$,

$$\mathbb{E}\left(\sup_{t\le t_{\text{fin}}} \Lambda_q(\mu_t^N)\right) \le (1+Ct_{\text{fin}})\mathbb{E}\left[\Lambda_q(\mu_0^N)\right]. \tag{2.101}$$

iii). For all $p \geq 2$,

$$\mathbb{P}\left(\sup_{t\geq 0} \frac{\Lambda_p(\mu_t^N)}{\Lambda_p(\mu_{t-}^N)} \leq 2^{1+(p/2)}\right) = 1. \tag{2.102}$$

iv). There exist $C_1, C_2 > 0$ such that, for all $\varepsilon > 0$ and $t_{\text{fin}} \geq 0$,

$$\mathbb{P}\left(\sup_{t\leq t_{\text{fin}}} \Lambda_p(\mu_t^N) \geq \max(\Lambda_p(\mu_0^N), C_1) + \epsilon\right) \leq C_2 t_{\text{fin}}\, \mathbb{E}\left[\Lambda_{2p+\gamma}(\mu_0^N)\right]\ N^{-1}\, \epsilon^{-2}. \tag{2.103}$$

Define, for $b \geq 1$,

$$T_b^N = \inf\left\{t \geq 0 : \Lambda_p(\mu_t^N) > \frac{b}{2^{\frac{p}{2}+1}}\right\}. \tag{2.104}$$

As a consequence of the estimate above, there exists $C = C(p)$ such that, if the initial data has the moment estimates $\Lambda_p(\mu_0^N) \leq a$ almost surely, then

$$\mathbb{P}(T_{Ca}^N \leq t_{\text{fin}}) \leq C t_{\text{fin}} \mathbb{E}\left[\Lambda_{2p+\gamma}(\mu_0^N)\right] N^{-1}. \tag{2.105}$$

Throughout, the constants C are allowed to depend on p, q and the same bounds as in Proposition 2.6, and on the almost sure bound a_0 for $\Lambda_2(\mu_0^N) = \Lambda_2(\mu_t^N)$.

Remark 2.11. *i). As in Proposition 2.6, this formulation is general enough to allow the regularised hard spheres kernel (rHS), or to show that the constants are uniform in both K, N in the case of cutoff hard potentials (CHP_K).*

ii). To the best of our knowledge, point iv) is new, and we call this phenomenon concentration of moments. *In some applications we will need a bound with high probability of the form*

$$\mathbb{P}(\Lambda_p(\mu_t^N) \leq b_N) \to 1$$

or the same thing with $\sup_{s\leq t} \Lambda_p(\mu_s^N)$. If we wished to deduce such a bound from item i), we would need to take some sequence $b_N \to \infty$, and the error probability would be on the order $b_N^{-\alpha}$, for some α depending on how many moments we assume on the initial data. However, item iv). instead allows us to achieve this with $b = b_N$ independent of N. The first statement here is somewhat sharper, and may be of independent interest; however, for applications in this thesis, we will mostly use the second form, which absorbs some constants.

Proof. By the conservation of energy, we have $\Lambda_2(\mu_t^N) = \Lambda_2(\mu_0^N) \leq a_0$ for all $t \geq 0$, almost surely, which implies the uniform bound $\text{Supp}(\mu_t^N) \subset \{|v| \leq \sqrt{Na_0}\}$ for all $t \geq 0$, on the same almost sure event.

Step 1. Moment Propagation and Creation For the first point of item i), we argue as in Proposition 2.6. In the case of the Kac process, we write

$$\langle |v|^p, \mu_t^N\rangle = \Lambda_p(\mu_t^N) = \Lambda_p(\mu_0^N) + \int_0^t \langle |v|^p, Q(\mu_s^N)\rangle ds + M_t^N \tag{2.106}$$

where M_t^N is a total variation martingale; this is a noisy version of (2.69). The manipulations here are licit thanks to the time-uniform bound on $\mathrm{Supp}(\mu_t^N)$, for instance replacing $|v|^p$ by a Lipschitz, compactly supported function which agrees with $|v|^p$ on the ball of radius $2\sqrt{N\Lambda_2(\mu_0^N)}$, which contains $\mathrm{supp}(\mu_t^N), \mathrm{supp}(Q(\mu_t^N))$ for all $t\ge 0$, and all expressions are finite by the support bound. We already established in Step 1 of Proposition 2.6 that, for some λ, and allowing C to depend on a_0,

$$\begin{aligned}\langle |v|^p, Q(\mu_t^N)\rangle &\le -\lambda\Lambda_{p+\gamma}(\mu_t^N)+C\Lambda_p(\mu_t^N)\\ &\le -\lambda\Lambda_p(\mu_t^N)^{1+\gamma/p}+C\Lambda_p(\mu_t^N).\end{aligned} \tag{2.107}$$

Taking expectations of (2.106), we find that $F_p^N(t) := \mathbb{E}[\Lambda_p(\mu_t^N)]$ solves a differential inequality

$$\frac{d}{dt}F_p^N(t)\le -\lambda F_{p+\gamma}^N(t)+CF_p^N(t)$$

analagous to (2.71). The desired bound (2.100) now follows by following the arguments of Step 3 of Proposition 2.6.

Step 2. Maximal Inequality We next prove the maximal inequality. In this case, we return to (2.62) to bound the jumps of $\Lambda_q(\mu_t^N)$ at collisions by

$$\Lambda_q(\mu_t^N)-\Lambda_q(\mu_{t-}^N)\le \frac{C}{N}(|v|^{q-1}|v_*|+|v_*|^{q-1}|v|)\,\sin\theta \tag{2.108}$$

when v, v_* are the precollisional velocities, and the deflection angle is θ, since the first term of (2.62) is nonpositive for any value of θ. We now consider the process A_t whose jumps are exactly the right-hand side, so that A_t is increasing and

$$\sup_{s\le t}\Lambda_q(\mu_s^N)\le \Lambda_q(\mu_0^N)+A_t. \tag{2.109}$$

We now estimate

$$\begin{aligned}\mathbb{E}[A_t] &\le C\mathbb{E}\left[\int_0^t\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}}(|v|^{q-1}|v_\star|+|v_\star|^{q-1}|v|)\sin\theta B(v-v_\star,\sigma)\mu_s^N(dv)\mu_s^N(dv_\star)d\sigma\right]\\ &\le C\mathbb{E}\left[\int_0^t\int_{\mathbb{R}^d\times\mathbb{R}^d}(|v|^{q-1}|v_\star|+|v_\star|^{q-1}|v|)|v-v_*|^\gamma\mu_s^N(dv)\mu_s^N(dv_\star)d\sigma\right]\end{aligned} \tag{2.110}$$

thanks to the factor of $\sin\theta$. Simplifying, we see that

$$\mathbb{E}A_t\le C\mathbb{E}\int_0^t\Lambda_{q+\gamma-1}(\mu_s^N)ds\le C\mathbb{E}\int_0^t\Lambda_q(\mu_s^N)ds \tag{2.111}$$

and the conclusion now follows, using the previous point to bound $\mathbb{E}\Lambda_q(\mu_t^N)$.

Step 3. Instantaneous Increase of Moments Almost surely, for any t with $\mu_t^N\neq\mu_{t-}^N$, it follows that μ_t^N is formed by changing velocities $v, v_\star$ in μ_{t-}^N to post-collision velocities $v', v_\star'$. We have the bound

$$|v'|^p\le(|v'|^2+|v_*'|^2)^{\frac{p}{2}}=(|v|^2+|v_*|^2)^{p/2}\le 2^{p/2}(|v|^p+|v_*|^p). \tag{2.112}$$

Using the same bound for $v_\star'$ leads to the claimed result.

Step 4. Concentration of Moments To prove item iii), we start by analysing the martingale M^N in the noisy differential inequality (2.106) which plays the same role for the Kac process that (2.69) does for the Boltzmann equation. From the analysis in [49], we have

$$[M^N]_t = N^{-1}\int_0^t H_p(\mu^N_s)ds; \tag{2.113}$$

where

$$H_p(\mu^N_t) := \int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}} \left(|v'|^p + |v'_*|^p - |v|^p - |v_*|^p\right)^2 B(v-v_*,\sigma)\mu^N_t(dv)\mu^N_t(dv_*)d\sigma. \tag{2.114}$$

To bound the integrand $H_p(\mu^N_t)$, we recall that $|v'-v|\le |v-v_*|\sin\theta$ to obtain

$$\begin{aligned}(|v'|^p-|v|^p)^2 &\le C(p)(1+|v|^{p-1}+|v_*|^{p-1})^2|v-v_*|^2\sin^2\theta\\ &\le C(p)(1+|v|^{2p-2}+|v_*|^{2p-2})|v-v_*|^2\theta\\ &\le C(p)(1+|v|^{2p}+|v_*|^{2p})\theta.\end{aligned} \tag{2.115}$$

By assumption, $\int_{\mathbb{S}^{d-1}}\theta b(\cos\theta)d\sigma = \int_0^\pi \theta\beta(\theta)d\theta<\infty$, and we will have a finite integral when integrating the right-hand side. Using a similar computation for $(|v'_*|^p-|v_*|^p)^2$, we obtain

$$\begin{aligned}H_p(\mu^N_t) &\le C\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}}(1+|v|^{2p+\gamma}+|v_*|^{2p+\gamma})\theta b(\cos\theta)\mu^N_t(dv)\mu^N_t(dv_*)d\sigma\\ &\le C(1+\Lambda_{2p+\gamma}(\mu^N_t)).\end{aligned} \tag{2.116}$$

Returning to (2.113), we conclude that for some $C_2=C_2(p)$,

$$\mathbb{E}\left[|M^N_t|^2\right] \le \frac{C_2}{16N}\mathbb{E}\left[\int_0^t (1+\Lambda_{2p+\gamma}(\mu^N_s))ds\right]. \tag{2.117}$$

By the choice of μ^N_0 and moment propagation results above, the right-hand side is at most $C_2 t_{\text{fin}}(1+\Lambda_{2p+\gamma}(\mu^N_0))/16N$, up to a new choice of C_2. We now return to (2.107) Set $C_1 = (C/\lambda)^{p/\gamma}$, so that the right-hand side of (2.107) is nonpositive as soon as $\langle |v|^p,\mu\rangle \ge C_1$. Define T to be the stopping time

$$T=\inf\left\{t\ge 0: \Lambda_p(\mu^N_t) > \max(C_1,\Lambda_p(\mu^N_0))+\epsilon\right\} \tag{2.118}$$

and on the event $T\le t_{\text{fin}}$, define

$$T' = \sup\left\{t<T: \Lambda_p(\mu^N_t)\le \max(C_1,\Lambda_p(\mu^N_0))\right\}. \tag{2.119}$$

This set is always nonempty, as it includes 0, and we have

$$\limsup_{t\uparrow T'}\langle |v|^p,\mu^N_t\rangle \le \max(C_1,\Lambda_p(\mu^N_0)); \tag{2.120}$$

$$\langle |v|^p,\mu^N_t\rangle > \max(C_1,\Lambda_p(\mu^N_0)) \text{ for all } t\in(T',T]. \tag{2.121}$$

By the choice of C_1, it follows that

$$\int_{(T',T]} \langle |v|^p, Q(\mu_s^N)\rangle ds \le 0 \tag{2.122}$$

and so, from (2.106), we must have $M_T^N - M_{T'-}^N \ge \epsilon$. Therefore, on the event $\{T \le t_{\text{fin}}\}$, we have the lower bound $\sup_{t\le t_{\text{fin}}} |M_t^N| \ge \frac{\epsilon}{2}$. We now use Doob's L^2 inequality to bound $\mathbb{E}[\sup_{t\le t_{\text{fin}}} |M_t^N|^2]$, and Chebychev's inequality to bound the probability

$$\mathbb{P}\left(\sup_{t\le t_{\text{fin}}} |M_t^N| \ge \frac{\epsilon}{2}\right) \le 16\epsilon^{-2}\mathbb{E}\left[|M_{t_{\text{fin}}}^N|^2\right]. \tag{2.123}$$

The second item follows immediately from the first. □

2.5.3 Polynomial Moment Estimates for the Landau Equation

We next turn to the polynomial moment estimates in the case of the Landau equation (LE). In this case, the result analagous to Proposition 2.6 is as follows.

Proposition 2.12. *Let $\gamma \in (0,1]$ and let $(\mu_t)_{t\ge 0} \in L^\infty_{loc}([0,\infty), \mathcal{P}_2(\mathbb{R}^3)) \cap L^1_{loc}([0,\infty), \mathcal{P}_{2+\gamma}(\mathbb{R}^3))$ be a weak solution to (LE). Then the following hold.*

i). For all $q \ge p \ge 2$ and $t > 0$, we have

$$\Lambda_q(\mu_t) \le C_{p,q}(1 + t^{(p-q)/\gamma})\Lambda_p(\mu_0)$$

where $C_{p,q}$ depends (only) on p, q, γ and $\Lambda_2(\mu_0)$. In particular, all moments $\Lambda_p(\mu_t)$ are finite, uniformly away from $t = 0$, and if $\Lambda_p(\mu_0)$ is finite, then $\Lambda_p(\mu_t)$ is bounded, uniformly in time. Moreover, if $\Lambda_p(\mu_0) < \infty$, then the map $t \to \Lambda_p(\mu_t)$ is continuous on $t \in [0,\infty)$.

ii). For any $p > 2$, if $\Lambda_p(\mu_0) < \infty$, then for any $t \ge 0$,

$$\int_0^t \Lambda_{p+\gamma}(\mu_u) du \le C(1+t)\Lambda_p(\mu_0) \tag{2.124}$$

and hence $\mu \in L^1_{loc}([0,\infty), \mathcal{P}_{p+\gamma}(\mathbb{R}^3))$.

The proof of this is very similar to that of Propositions 2.6 - 2.8 above, and so we will omit it. We will see an analogue of (2.71) in the proof of exponential moment creation below (equation (2.128); see also [58, Equations 34-35]) and, having obtained this, the quantitative form given follows by repeating the arguments of Step 3 of Proposition 2.6 verbatim. We remark that in this case, the induction corresponding to Proposition 2.8 is easier to start, since $\mu_t \in \mathcal{P}_{2+\gamma}(\mathbb{R}^3)$ for almost all $t \ge 0$; we refer also to the proof of [58, Theorem 3] for a precise justification of the finiteness of moments. We also did not need to include the conservation of energy as a hypothesis, as the statement that energy does not increase is already part of the definition of weak solutions.

2.5.4 Exponential Moment Estimates for the Boltzmann and Landau Equations

We will also use the some results about the appearance of *exponential* moments for both the Boltzmann equation for the case of noncutoff hard potentials (NCHP) and the Landau equation in Chapters 4, 5, which give us access to previous uniqueness results if we start at any $\mu_s, s>0$. The application will be to show, together with a previous uniqueness result, that if $(\mu_t)_{t\ge 0}, (\nu_t)_{t\ge 0}$ are energy conserving solutions to either (BE, LE) with $\mu_s=\nu_s$ for $s\in[0,\delta)$, for some $\delta>0$, then we have the global uniqueness $\mu_t=\nu_t$ for all $t\ge 0$. .

We will use a pair of recent results by Fournier[4] [84, 90] concerning exponential moments in the cases of the Landau equation (LE) and the Boltzmann equation in the case of non-cutoff hard potentials (NCHP). In the case of non-cutoff hard potentials, exponential moments $\langle e^{a|v|^\rho}, \mu_t\rangle$, $\rho\in(\gamma,2]$ are instantaneously created, which is strictly stronger than the cutoff case, where only exponential moments of order $\rho=\gamma$ are created, see the work of Alonso, Gamba and Taskovic [11]. The result is as follows.

Proposition 2.13. *Fix $\gamma\in(0,1]$ and a kernel B satisfying (NCHP). Then the following hold.*

i). For the Boltzmann case, let $(\mu_t)_{t\ge 0}\subset\mathcal{S}$ be an (energy-conserving) solution to (BE), and let $\rho=\min(2\gamma/(2-\nu),2)$. Then there exists a universal $a=a(B)<\infty$ such that

$$\sup_{t\ge 0}\left\langle e^{a\min(1,t^{\rho/\gamma})|v|^\rho}, \mu_t\right\rangle<\infty.$$

ii). For the Landau case, there are some constants $a>0$ and $C>0$, both depending only on γ, such that all weak solutions $(\mu_t)_{t\ge 0}\subset\mathcal{S}$ to (LE) satisfy

$$\left\langle e^{a|v|^2}, \mu_t\right\rangle<\infty$$

for all $t>0$.

We will give a detailed proof of the Landau case, which is more straightforward than the Boltzmann case, having access to Itô calculus thanks to the locality of $\mathcal{L}_\mathrm{L}$ in v. When investigating $\mathcal{L}_\mathrm{L}f$, $f(v)=|v|^p$, one finds terms of order $p, p-2, p-4$ in v, while by contrast in the Boltzmann case, even when an exact expansion is possible, one finds terms of all orders. A full proof of the Boltzmann case is given in [84].

Proof of Proposition 2.13ii). We follow the argument of Fournier [90]. During the proof, C will denote a constant which may only depend on γ, but may vary from line to line.

[4] The result on Gaussian moments of the Landau equation appeared in a joint work [90] with the author, but was contributed entirely by Prof. Fournier.

Step 1. Povzner Inequality for the Landau Equation We first follow the same calculations as in Step 1 of Proposition 2.6. Let us fix $p > 2$; thanks to Proposition 2.12, we know that for all $q > 0$, all $t_0 > 0$, $\sup_{t\geq t_0} \Lambda_q(\mu_t) < \infty$, and using a similar truncation argument to that of Proposition 2.6, we can apply the weak formulation of the Landau equation with $f(v) = |v|^p$ on $[t_0, \infty)$; as in the Boltzmann case (2.69) we find that $\Lambda_p(\mu_t)$ is of class C^1 on $(0, \infty)$ and get

$$\frac{d}{dt}\Lambda_p(\mu_t) = \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \mathcal{L}_{\mathrm{L}} f(v, v_*)\mu_t(dv_*)\mu_t(dv) \quad \text{for all } t > 0. \tag{2.125}$$

Elementary calculations from the definition of f show that

$$\partial_k f(v) = p|v|^{p-2}v_k \quad \text{and} \quad \partial^2_{k\ell} f(v) = p|v|^{p-2}\mathbb{1}_{\{k=\ell\}} + p(p-2)|v|^{p-4}v_k v_\ell$$

which give

$$\mathcal{L}_{\mathrm{L}} f(v, v_*) = \frac{1}{2}\sum_{k,\ell=1}^{3} a_{kl}(v-v_*)(p|v|^{p-2}\mathbb{1}_{\{k=\ell\}} + p(p-2)|v|^{p-4}v_k v_\ell) + \sum_{k=1}^{3} p|v|^{p-2}b_k(v-v_*)v_k.$$

Let us now examine these terms; to shorten notation, write $x = v - v_*$. First, using symmetry and recalling that $\sigma(x)^2 = a(x)$,

$$\sum_{k,\ell,j=1}^{3} \sigma_{k,j}(x)\sigma_{l,j}(x)\mathbb{1}_{\{k=\ell\}} = \sum_{k,\ell=1}^{3} a_{k\ell}(x)\mathbb{1}_{\{k=\ell\}} = \mathrm{Tr}\ a(x)$$

and

$$\sum_{k,\ell=1}^{3} a_{k\ell}(x)v_k v_\ell = \sum_{k,\ell,j=1}^{3} \sigma_{kj}(x)\sigma_{\ell j}(x)v_k v_\ell = |\sigma(x)v|^2.$$

We can therefore write the expression above concisely as

$$\mathcal{L}_{\mathrm{L}} f(v, v_*) = p|v|^{p-2}v\cdot b(x) + \frac{p}{2}|v|^{p-2}\mathrm{Tr}\ a(x) + \frac{p(p-2)}{2}|v|^{p-4}|\sigma(x)v|^2. \tag{2.126}$$

Recalling the definition $b(x) = -2|x|^\gamma x$ and observing that $\mathrm{Tr}\ a(x) = 2|x|^{\gamma+2}$, we have

$$v\cdot b(x) + \frac{1}{2}\mathrm{Tr}\ a(x)^2 = -2|x|^\gamma(v-v_*)\cdot v + |x|^\gamma(|v|^2 + |v_*|^2 - 2v\cdot v_*) = -|x|^\gamma|v|^2 + |x|^\gamma|v_*|^2.$$

By definition, we have $\sigma(x)x = \sigma(x)v - \sigma(x)v_* = 0$, so we repeat the arguments of (2.56) that

$$|\sigma(x)v| = |\sigma(x)v_*| \leq \min(|v|, |v_*|)||\sigma(x)|| \leq C|x|^{\gamma/2}|v||v_*|.$$

We therefore find that

$$\mathcal{L}_{\mathrm{L}} f(v, v_*) \leq -p|x|^\gamma|v|^p + Cp^2|x|^\gamma|v|^{p-2}|v_*|^2.$$

In the kinetic factor $|x|^\gamma = |v-v_*|^\gamma$, we use as usual the inequalities $|x|^\gamma \geq |v|^\gamma - |v_*|^\gamma$ in the negative term, and $|x|^\gamma \leq |v|^\gamma + |v_*|^\gamma$ in the positive term, to get overall

$$\mathcal{L}_{\mathrm{L}} f(v,v_*) \leq -\,p|v|^{p+\gamma} + p|v|^p|v_*|^\gamma + Cp^2(|v|^{p-2+\gamma}|v_*|^2 + |v|^{p-2}|v_*|^{2+\gamma}). \tag{2.127}$$

Integrating with respect to μ_t in both variables, we obtain from (2.125) that

$$\frac{d}{dt}\Lambda_p(\mu_t) \leq -\,p\Lambda_{p+\gamma}(\mu_t) + p\Lambda_p(\mu_t)\Lambda_\gamma(\mu_t) + Cp^2(\Lambda_{p-2+\gamma}(\mu_t)\Lambda_2(\mu_t) + \Lambda_{p-2}(\mu_t)\Lambda_{2+\gamma}(\mu_t)).$$

Using that $\Lambda_\gamma(\mu_t) \leq [\Lambda_2(\mu_t)]^{\gamma/2} = 1$, we finally find that, for $t \in (0,\infty)$,

$$\frac{d}{dt}\Lambda_p(\mu_t) \leq -p\Lambda_{p+\gamma}(\mu_t) + p\Lambda_p(\mu_t) + Cp^2[\Lambda_{p-2+\gamma}(\mu_t) + \Lambda_{p-2}(\mu_t)\Lambda_{2+\gamma}(\mu_t)]. \tag{2.128}$$

Step 2. A Differential Inequality We now manipulate (2.128) into a closed differential inequality for $\Lambda_p(\mu_t)$ as we did in the Boltzmann case leading to (2.71), but keeping track of lower-order terms, and the dependence of the coefficients on p. First, for any $\alpha \geq \beta \geq 2$, we use the same argument as in (2.73) that, since $|v|^2\mu_t(dv)$ is a probability measure,

$$\Lambda_\alpha(\mu_t) = \int_{\mathbb{R}^3} |v|^{\alpha-2}|v|^2\mu_t(dv) \leq \Big(\int_{\mathbb{R}^3} |v|^{\beta-2}|v|^2\mu_t(dv)\Big)^{(\alpha-2)/(\beta-2)} = [\Lambda_\beta(\mu_t)]^{(\alpha-2)/(\beta-2)}.$$

We deduce that $\Lambda_p(\mu_t) \leq [\Lambda_{p+\gamma}(\mu_t)]^{(p-2)/(p+\gamma-2)}$, whence

$$\Lambda_{p+\gamma}(\mu_t) \geq [\Lambda_p(\mu_t)]^{(p+\gamma-2)/(p-2)} = [\Lambda_p(\mu_t)]^{1+\gamma/(p-2)},$$

that

$$\Lambda_{p-2+\gamma}(\mu_t) \leq [\Lambda_p(\mu_t)]^{(p-4+\gamma)/(p-2)},$$

and that

$$\Lambda_{p-2}(\mu_t)\Lambda_{2+\gamma}(\mu_t) \leq [\Lambda_p(\mu_t)]^{(p-4)/(p-2)+\gamma/(p-2)} = [\Lambda_p(\mu_t)]^{(p-4+\gamma)/(p-2)}.$$

Observing that $(p-4+\gamma)/(p-2) = 1-(2-\gamma)/(p-2)$, we find that, for all $p \geq 4$, we have the differential inequality

$$\frac{d}{dt}\Lambda_p(\mu_t) \leq -p[\Lambda_p(\mu_t)]^{1+\gamma/(p-2)} + p\Lambda_p(\mu_t) + Cp^2[\Lambda_p(\mu_t)]^{1-(2-\gamma)/(p-2)}. \tag{2.129}$$

Step 3. We next analyse general differential inequalities of this form, in the same way we did in the Boltzmann case but again keeping track of lower-order terms. We study general $u : (0,\infty) \to (0,\infty)$ of class C^1 satisfying, for some $a,b,c,\alpha,\beta > 0$, for all $t > 0$,

$$u'(t) \leq -a[u(t)]^{1+\alpha} + bu(t) + c[u(t)]^{1-\beta}.$$

Let us set $h(r) = -ar^{1+\alpha} + br + cr^{1-\beta}$ and we observe that

$$h(r) \leq -\frac{a}{2}r^{1+\alpha} \quad \text{for all } r \geq u_* = \max\{(4b/a)^{1/\alpha}, (4c/a)^{1/(\alpha+\beta)}\}.$$

We now fix $t_0 > 0$. As before, if $u(t_0) \le u_*$, we have $u(t) \le u_*$ for all $t \ge t_0$, because $h(u_*) \le 0$ and $u'(t) \le h(u(t))$.

On the other hand, if we assume instead $u(t_0) > u_*$, set $t_1 = \inf\{t > t_0 : u(t) \le u_*\}$ and observe that for $t \in [t_0, t_1)$,

$$u'(t) \le h(u(t)) \le -\frac{a}{2}[u(t)]^{1+\alpha}.$$

Integrating this inequality, we conclude that, for all $t \in [t_0, t_1)$,

$$u(t) \le \Big[u^{-\alpha}(t_0) + \frac{a\alpha(t-t_0)}{2}\Big]^{-1/\alpha} \le \Big[\frac{2}{a\alpha(t-t_0)}\Big]^{1/\alpha}.$$

This implies that t_1 is finite. Since now $u(t_1) = u_*$ by definition and, as before, it follows that $u(t) \le u_*$ for all $t \ge t_1$.

Hence in any case, for any $t_0 > 0$, any $t > t_0$, $u(t) \le \max\{u_*, [2/(a\alpha(t-t_0))]^{1/\alpha}\}$. Letting $t_0 \to 0$, we deduce that $u(t) \le \max\{u_*, [2/(a\alpha t)]^{1/\alpha}\}$ for all $t > 0$, and using the definitions we conclude that

$$\forall\, t > 0, \quad u(t) \le \Big(\frac{2}{a\alpha t}\Big)^{1/\alpha} + \Big(\frac{4b}{a}\Big)^{1/\alpha} + \Big(\frac{4c}{a}\Big)^{1/(\alpha+\beta)}.$$

We now apply this to (2.129) with $a = p$, $b = p$, $c = Cp^2$, $\alpha = \gamma/(p-2)$ and $\beta = (2-\gamma)/(p-2)$, we find that for all $p \ge 4$, all $t > 0$,

$$\Lambda_p(\mu_t) \le \Big(\frac{2(p-2)}{p\gamma t}\Big)^{(p-2)/\gamma} + 4^{(p-2)/\gamma} + \Big(4Cp\Big)^{(p-2)/2}.$$

Let us remark that this is the same behaviour as in Proposition 2.12 for the exponents $p, 2$, but we have now quantified exactly the coefficient and the lower-order terms. Changing again the value of C, we conclude that for all $p \ge 4$, all $t > 0$,

$$\Lambda_p(\mu_t) \le \Big(1 + \frac{2}{\gamma t}\Big)^{p/\gamma} + (Cp)^{p/2}.$$

Step 4. Conclusion We now conclude. For $a > 0$ and $t > 0$, we write, using that $\Lambda_0(\mu_t) = \Lambda_2(\mu_t) = 1$,

$$\int_{\mathbb{R}^3} e^{a|v|^2}\mu_t(dv) = \sum_{k\ge 0}\frac{a^k\Lambda_{2k}(\mu_t)}{k!} = 1 + a + \sum_{k\ge 2}\frac{a^k\Lambda_{2k}(\mu_t)}{k!}.$$

By Step 4,

$$\int_{\mathbb{R}^3} e^{a|v|^2}\mu_t(dv) \le 1 + a + \sum_{k\ge 2}\frac{1}{k!}\Big[a^k\Big(1+\frac{2}{\gamma t}\Big)^{2k/\gamma} + a^k(2Ck)^k\Big].$$

But $\sum_{k\ge 2}(k!)^{-1}(xk)^k < \infty$ if $x < 1/e$ by the Stirling formula. Hence if $a < 1/(2Ce)$,

$$\int_{\mathbb{R}^3} e^{a|v|^2}\mu_t(dv) \le 1 + a + \exp\Big[a\Big(1+\frac{2}{\gamma t}\Big)^{2/\gamma}\Big] + C.$$

The conclusion follows. □

We also sketch the (much longer) proof of the Boltzmann case.

Sketch Proof of Proposition 2.13i). We sketch the argument of Fournier [84]. First, we return to (2.57), still using the parametrisation of collisions in terms of (θ, φ). In the special case where $p = 2n$ is an even integer, one can find an exact expansion of $|v'|^{2n}$ using Newton's trinomial expansion to find an expansion of $|v'|^{2n}$ in powers of the trigonometric functions $(1 \pm \cos\theta)/2, \sin\theta$ and $v \cdot \Gamma(v - v_*, \varphi)$. The integral over φ produces a Wallis integral, leading to

$$\int_{\mathbb{S}^{d-2}} (v \cdot \Gamma(v - v_*, \varphi))^k d\varphi = \mathbb{1}_{k \in 2\mathbb{N}} \frac{k!}{2^k [(k/2)!]^2} (|v|^2 |v_*|^2 - (v \cdot v_*)^2)$$

to find an exact expression for the average $\int |v'|^{2n} d\varphi$ integrated over φ, in the form

$$\int |v'|^{2n} d\varphi = \sum_{(i,j,k) \in \mathcal{A}_n} \frac{n!}{i! j! ((k/2)!)^2} \left(\frac{1+\cos\theta}{2}\right)^i \left(\frac{1-\cos\theta}{2}\right)^j \left(\frac{\sin\theta}{2}\right)^{k/2} |v|^{2i} |v_*|^{2j} \cdots \times \left(|v|^2 |v_*|^2 - (v \cdot v_*)^2\right)^{k/2}$$

with

$$\mathcal{A}_n := \left\{(i, j, k) \in \mathbb{N}^3 : i + j + k = n, \ k \text{ even}\right\}.$$

Of this sum, the extreme terms $(i, j, k) = (n, 0, 0), (0, n, 0)$ produce the same terms of order $|v|^{2n}, |v_*|^{2n}$ as we found in the more crude expansion (2.62). If one subtracts $|v|^{2n}$ and argues similarly for $|v_*'|^{2n}$, we find the same factor as above reinforcing the negative term

$$\lambda_{2n} := \int_{\mathbb{S}^{d-1}} \left(1 - \left(\frac{1+\cos\theta}{2}\right)^n - \left(\frac{1-\cos\theta}{2}\right)^n\right) b(\cos\theta) d\sigma. \tag{2.130}$$

At this point, one crucially uses the non-cutoff assumption to see that the effect of the Povzner term is stronger for large n. A simple argument shows that, for the case (NCHP), that $\lambda_{2n} \to \infty$; see also Lemma 4.21, whereas in the cutoff case one finds an upper bound in terms of $\int b(\cos\theta) d\sigma$, which is finite in these cases. Fournier [84, Lemma 5] found the lower bound $\lambda_{2n} \geq c n^{\nu/2}$, and with an an analysis of the θ-integrals in the remaining terms, one finds

$$\begin{aligned} \int_{\mathbb{S}^{d-1}} &(|v'|^{2n} + |v_*'|^{2n} - |v|^{2n} - |v_*|^{2n}) b(\cos\theta) d\sigma \\ &\leq -c_1 n^{\nu/2} (|v|^{2n} + |v_*|^{2n}) \\ &+ c_2 \sum_{a=1}^{n-1} \binom{n}{a} \left(\frac{n^{\nu/2}}{(n-a)^{\nu/2+1}} + \frac{1}{a}\right) (|v|^{2a} |v_*|^{2(n-a)} + |v|^{2(n-a)} |v_*|^{2a}). \end{aligned} \tag{2.131}$$

If we multiply by the kinetic factor $|v - v_*|^\gamma$ and integrate to find a differential inequality for Λ_{2n} as before, the negative term produces $-n^{\nu/2} \Lambda_{2n+\gamma}(\mu_t) + 2n^{\nu/2} 2^{2n/\gamma}$, and the positive term leads to the sum

$$S_n(\mu_t) = \sum_{a=1}^{\lfloor n/2 \rfloor} \binom{n}{a} \frac{n^{\nu/2}}{a^{\nu/2}+1} \Lambda_{2a}(\mu_t)\Lambda_{2(n-a)+\gamma}(\mu_t)$$

and we end up with a differential inequality

$$\frac{d}{dt}\Lambda_{2n}(\mu_t) \le -c_1 n^{\nu/2}\Lambda_{2n+\gamma}(\mu_t) + 2c_1 n^{\nu/2} 2^{2n/\gamma} + c_2 S_n(\mu_t). \tag{2.132}$$

We then sum these for the partial sums

$$E_N(t) := \sum_{n=0}^{N} \frac{(at)^{2n/\gamma}\Lambda_{2n}(\mu_t)}{(n!)^\beta}$$

for some $a > 0, \beta \ge 1$ to be chosen. We note that these are continuous at 0, thanks to the asymptotic $\Lambda_{2n}(\mu_t) \le C_n t^{-(2n-2)/\gamma}$ from Proposition 2.6, and $E_N(0) = 1$. Using (2.132) to differentiate, one finds three terms

$$\frac{d}{dt}E_N(t) \le -c_1 F_N(t) + c_2 G_N(t) + \frac{2a}{\gamma} H_N(t) + C \tag{2.133}$$

where F_N gathers the negative terms for $n \ge 2$:

$$F_N(t) := \sum_{n=2}^{N} n^{\nu/2} \frac{(at)^{2n/\gamma}\Lambda_{2n+\gamma}(\mu_t)}{(n!)^\beta}$$

and where G_N comes from summing the positive terms S_n. Finally H_N comes from differentiating the time-dependent coefficient $(at)^{2n/\gamma}$:

$$H_N(t) = \sum_{n=1}^{N} \frac{n(at)^{2n/\gamma-1}\Lambda_{2n}(\mu_t)}{(n!)^\beta}$$

and C is a constant, uniform in N, coming the final term in (2.132). For any $\varepsilon > 0$, the term G_N is controlled by

$$G_N(t) \le \varepsilon E_N(t) F_N(t) + a^{2/\gamma} A_\varepsilon (1 + F_N(t))$$

for some $A = A_\varepsilon$ which does not depend on N. For the term $H_N(t)$, for $\kappa > 0$ to be chosen, we bound

$$\begin{aligned} \frac{n\Lambda_{2n}(\mu_t)}{at} &\le \kappa n^{\nu/2} \left(\Lambda_{2n}(\mu_t)\right)^{1+\gamma/(2n-2)} + \frac{n}{at}\left(\frac{n^{1-\nu/2}}{\kappa a t}\right)^{(2n-2)/\gamma} \\ &\le \kappa n^{\nu/2}\Lambda_{2n+\gamma}(\mu_t) + \frac{n}{at}\left(\frac{n^{1-\nu/2}}{\kappa a t}\right)^{(2n-2)/\gamma} \end{aligned}$$

which leads to

$$H_N(t) \le \kappa F_N(t) + \kappa (at)^{2/\gamma}\Lambda_{2+\gamma}(\mu_t) + (at)^{2/\gamma-1} \sum_{n=1}^{N} \frac{n^{(2-\nu)/\gamma-(2-\nu)/\gamma+1}}{(n!)^\beta \kappa^{(2n-2)/\gamma}}.$$

Using Stirling's formula in the final summand,

$$\frac{n^{1+(1-\nu/2)(2n-2)/\gamma}}{\kappa^{(2n-2)/\gamma}(n!)^\beta} \sim \frac{n^{1-(2-\nu)/\gamma}n^{n((2-\nu)/\gamma)-\beta)}}{(2\pi)^{\beta/2}e^{n\beta}\kappa^{(2n-2)/\gamma}}$$

is summable, provided that we choose $\beta = \max(1, \frac{2-\nu}{\gamma})$ so that the exponent $\frac{2-\nu}{\gamma} - \beta$ of n^n is nonpositive, and if we choose $\kappa = 2e^{\beta\gamma/2}$. We now return to (2.133); provided that we choose a small enough that $2a/\gamma \le \frac{c_1}{2\kappa}$, we can absorb the multiple of F_N coming from H_N into the negative term $-c_1F_N$ to conclude that, for some finite constants c, C,

$$\frac{d}{dt}E_N(t) \le -cF_N(t) + \frac{c}{2}E_N(t)F_N(t) + C.$$

If we then set $T_N = \inf\{t : E_N(t) > 2\}$, then $T_N > 0$ by the continuity of E_N, and since $E_N = 0$, and for $t \le T_N$ we have $\frac{d}{dt}E_N \le C$. It follows that $E_N \ge C^{-1}$, uniformly in N, and for $t \le C^{-1}$ we conclude that

$$\sum_{n\ge 0} \frac{\Lambda_{2n}(\mu_t)(at)^{2n/\gamma}}{(n!)^\beta} \le 2.$$

Using Hölder's inequality, for all n it follows that, still for $t \le C^{-1}$,

$$\frac{\Lambda_{2n/\beta}(\mu_t)(at)^{2n/\gamma\beta}}{n!} \le \left(\frac{\Lambda_{2n}(\mu_t)(at)^{2n/\gamma}}{(n!)^\beta}\right)^{1/\beta} \le 2^{1/\beta}$$

and so

$$\int_{\mathbb{R}^d} \frac{(|v|^{2/\beta}(at/2)^{2/\gamma\beta})^n}{n!}\mu_t(dv) = \frac{\Lambda_{2n/\beta}(\mu_t)(at/2)^{2n/\gamma\beta}}{n!} \le 2^{(\frac{1}{\beta}-\frac{2n}{\gamma\beta})}.$$

The right-hand side is now summable in n, and we observe that $2/\beta$ is the exponent ρ in the theorem, and that the factor $t^{2/\gamma\beta} = t^{\rho/\gamma}$. Summing the last display, we conclude that

$$\left\langle \exp\left(\left(\frac{at}{2}\right)^{\rho/\gamma}|v|^\rho\right), \mu_t\right\rangle \le \frac{2^{\rho/2}}{1-2^{-\rho/\gamma}}$$

for all $t \le C^{-1}$, which proves the instantaneous appearance of such moments. For $t > C^{-1}$, we repeat the same argument on the solution $(\mu_{s+t_0})_{s\ge 0}$ to (BE) with $t_0 = t - C^{-1}$ to conclude that

$$\left\langle \exp\left(\left(\frac{aC^{-1}}{2}\right)^{2/\rho}|v|^\rho\right), \mu_t\right\rangle \le \frac{2^{\rho/2}}{1-2^{-\rho/\gamma}}.$$

The conclusion now follows, with the coefficient a in the statement replaced by

$$(\frac{a}{2})^{\rho/\gamma}\min(1, C^{-1}).$$

□

2.5.5 Some Further Properties Related to Moments

We conclude with some other properties which will be useful, and which follow in the spirit of the results before. First, we have made reference throughout to solutions to the Boltzmann equation for which the energy $\langle |v|^2, \mu_t\rangle$ is constant. Even relaxing this restriction, the energy cannot decrease; this is the content of the following result of Lu and Wennberg [133, Section 3]. This property will be important in Chapter 6, where we are naturally led to consider solutions to (BE) which do not conserve energy.

Proposition 2.14. *Let B be any kernel satisfying the hypotheses of Propositions 2.6-2.8, and $(\mu_t)_{t\ge 0}$ a weak solution to the corresponding Boltzmann equation. Then $\Lambda_2(\mu_t) = \langle |v|^2, \mu_t\rangle$ is nondecreasing.*

We will also make some use of the following 'uniform integrability' property for solutions.

Proposition 2.15. *Let $(\mu_t)_{t\ge 0}$ be a weak, energy-conserving solution to either the Boltzmann equation (BE), for a kernel B as in Proposition 2.6, or (LE) for some $0<\gamma\le 1$ with $d=3$. Let $p\ge 2$ and $\varepsilon>0$. Then there exists $R<\infty$ such that*

$$\sup_{t\ge 0}\int_{\mathbb{R}^d}(1+|v|^p)1\!\!1_{\{|v|>R\}}\mu_t(dv)<\epsilon.$$

Proof. This was argued by the author [90, Lemma 9] with a proof based on the local integrability of the $(p+\gamma)^{\text{th}}$ moment. We will here argue instead using the continuity of the p^{th} moment at 0, which was established in Proposition 2.8, 2.12 for the Boltzmann and Landau cases respectively. Indeed, this absorbs most of the cited proof. Throughout, we work on $\mathbb{R}^d$, understanding $d=3$ in the Landau case.

In either case, we fix $\varepsilon>0$, and let $R_1>0$ be such that $\int_{\mathbb{R}^d}(1+|v|^p)1\!\!1\{|v|\ge R_1\}\mu_0(dv)<\varepsilon$. We now set $\chi_{R_1}:\mathbb{R}^d\to[0,1]$ to be a smooth function with $1\!\!1_{\{|v|\le R_1\}}\le \chi_{R_1}(v)\le 1\!\!1_{\{|v|\le R_1+1\}}$, so that

$$\int_{\mathbb{R}^d}(1-\chi_{R_1}(v))(1+|v|^p)\mu_0(dv)=1+\Lambda_p(\mu_0)-\langle \chi_{R_1}(v)(1+|v|^p),\mu_0\rangle<\varepsilon.$$

Thanks to (BE, LE) and the compact support of $\chi_{R_1}(v)(1+|v|^p)$, the map $t\mapsto \langle\chi_{R_1}(v)(1+|v|^p),\mu_t\rangle$ is continuous, and the map $t\mapsto 1+\Lambda_p(\mu_t)$ is continuous at 0, by Proposition 2.8, 2.12 respectively, and all together the map

$$t\mapsto\int_{\mathbb{R}^d}(1-\chi_{R_1}(v))(1+|v|^p)\mu_t(dv)$$

is continuous at $t=0$. It follows that we can find $t_0>0$ such that, uniformly in $t\in[0,t_0]$,

$$\int_{\mathbb{R}^d}(1+|v|^p)1\!\!1\{|v|\ge R_1+1\}\mu_t(dv)\le\int_{\mathbb{R}^d}(1+|v|^p)(1-\chi_{R_1}(v))\mu_t(dv)<\varepsilon.$$

On the other hand, $\sup_{t\geq t_0}\Lambda_{2p}(\mu_t) < \infty$ by the moment creation property, and for all $R > 0$ a simple Chebychev estimate gives

$$\int_{\mathbb{R}^d}(1+|v|^p)\mathbb{1}\{|v|\geq R\}\mu_t(dv) \leq \frac{1+R^p}{R^{2p}}\Lambda_{2p}(\mu_t).$$

It follows that we can find $R_2 < \infty$ such that the right-hand side is at most ε, uniformly in $t \geq t_0$. The conclusion follows by taking $R := \max(R_1+1, R_2)$. □

Chapter 3

Long-Time Propagation of Chaos for Hard Spheres

3.1 Introduction & Main Results

This chapter is dedicated to Theorem 1, concerning the long-time propagation of chaos for the Kac process in the case of the hard spheres kernel (HS). Throughout this chapter, B will be the hard-spheres kernel, except for some preliminary calculations in Section 3.2.1 where we also allow cutoff hard potentials (CHP_K). Throughout, we will work with the weighted Wasserstein metric W_1 introduced in Section 2.1, which we generalise to W_γ when discussing (CHP_K). In any case, $(\mu_t^N)_{t\geq 0}$ will be a (labelled) Kac process on the relevant kernel, normalised to the state space $\mathcal{S}$.

We now precisely state the main results of this chapter. Our first theorem controls the deviation from the Boltzmann flow at a single, deterministic time $t \geq 0$, which we refer to as a *pointwise* estimate. Moreover, this estimate is *uniform in time.*

Theorem 3.1. *Let $0 < \epsilon < \frac{1}{d}$ and let $a \geq 1$. For sufficiently large p, depending on ϵ, d, let $(\mu_t^N)_{t\geq 0}$ be a Kac process in dimension $d \geq 3$, and let $\mu_0 \in \mathcal{S}^p$, satisfying the moment bounds*

$$\Lambda_p(\mu_0^N) \leq a; \qquad \Lambda_p(\mu_0) \leq a. \tag{3.1}$$

Then for some $C = C(\epsilon, d, p) < \infty$ and $\zeta = \zeta(d) > 0$, we have the uniform bound

$$\sup_{t\geq 0} \left\| W_1\left(\mu_t^N, \phi_t\left(\mu_0\right)\right)\right\|_{L^2(\mathbb{P})} \leq Ca\left(N^{\epsilon-1/d} + W_1\left(\mu_0^N, \mu_0\right)^\zeta\right). \tag{3.2}$$

This generalises, by conditioning, to the case where the initial data μ_0^N is random, provided that $\mathbb{E}\Lambda_p(\mu_0^N) \leq a$.

This result is, to the best of our knowledge, new, although an equivalent result is known for Maxwell molecules [48]. Thanks to Proposition 2.1, we can understand this result as a uniform-in-time-chaoticity result, which gives a power-law rate $N^{-\zeta/d}$, improving on the rates of Mischler and Mouhot [142] for the hard spheres process.

Our second main theorem controls, in $L^q(\mathbb{P})$, the maximum deviation from the Boltzmann flow up to a time t_{fin}, in analogy with Proposition 1.2. We refer to this as a *pathwise, local uniform in time* estimate.

Theorem 3.2. *Let $0<\epsilon<\frac{1}{2d}$, $a\geq 1$ and $q\geq 2$. For sufficiently large $p\geq 0$, depending on ϵ, d, let $(\mu^N_t)_{t\geq 0}$ be a Kac process on $N\geq 2$ particles and let $\mu_0\in\mathcal{S}^p$, with initial moments*

$$\Lambda_{pq}(\mu^N_0)\leq a^q; \qquad \Lambda_p(\mu_0)\leq a. \tag{3.3}$$

For some $\alpha=\alpha(\epsilon,d,q)>0$ and $C=C(\epsilon,d,q,p)<\infty$ and $\zeta=\zeta(d)>0$, we can estimate, for all $t_{fin}\geq 0$,

$$\left\|\sup_{t\leq t_{fin}} W_1\left(\mu^N_t,\phi_t(\mu_0)\right)\right\|_{L^q(\mathbb{P})}\leq Ca\left((1+t_{fin})^{1/q}\,N^{-\alpha}+W_1(\mu^N_0,\mu_0)^\zeta)\right). \tag{3.4}$$

The exponent α is given explicitly by

$$\alpha=\frac{q'}{2d}-\epsilon \tag{3.5}$$

where $1<q'\leq 2$ is the Hölder conjugate to p.

At the end of this section, we will discuss related results, and how they may be compared to this estimate.

An unfortunate feature of these approximation theorems is the dependence on the unknown, and potentially large, moment index p. We will also prove the following variant of the theorems above which allows us to use any moment estimate higher than second.

Theorem 3.3. *[Convergence with few moment estimates] Let $p>2$ and $a\geq 1$. Let (μ^N_t) be an N-particle Kac process, and μ_0 in $\mathcal{S}$ with initial moment estimates*

$$\Lambda_p(\mu^N_0)\leq a; \qquad \Lambda_p(\mu_0)\leq a. \tag{3.6}$$

There exists $\epsilon=\epsilon(d,p)>0$ and a constant $C=C(d,p)$ such that

$$\sup_{t\geq 0}\left\|W_1\left(\mu^N_t,\phi_t(\mu_0)\right)\right\|_{L^1(\mathbb{P})}\leq Ca(N^{-\epsilon}+W_1(\mu^N_0,\mu_0)^\epsilon). \tag{3.7}$$

For a local uniform estimate, if $q\geq 2$, then there exists a constant $C=C(d,p,q)$ and $\epsilon=\epsilon(d,p,q)>0$ such that, for all $t_{\text{fin}}<\infty$,

$$\left\|\sup_{t\leq t_{\text{fin}}} W_1\left(\mu^N_t,\phi_t(\mu_0)\right)\right\|_{L^1(\mathbb{P})}\leq Ca((1+t_{\text{fin}})^{1/q}N^{-\epsilon}+W(\mu^N_0,\mu_0)^\epsilon). \tag{3.8}$$

In the course of proving this result, we will see that the higher moment conditions are only required to obtain the optimal rates on a very short time interval $[0, u_N]$ and, in particular, we can obtain very good time-dependence without higher moment estimates.

We also study the long-time behaviour of the Kac Process. We cannot extend Theorem 3.2 to control the maximum deviations over all times $t \geq 0$, due to the following recurrence features of the Kac process.

Theorem 3.4. *There exists a universal constant $C > 0$ such that, for every N, for every $p > 2$ and $a > 1$, there exists a Kac process $(\mu_t^N)_{t\geq 0} \subset \mathcal{S}$ with initial moment $\Lambda_p(\mu_0^N) \leq a$ but, almost surely,*

$$\limsup_{t\to\infty} W_1\left(\mu_t^N, \phi_t(\mu_0^N)\right) \geq 1 - \frac{C}{\sqrt{N}}. \tag{3.9}$$

Hence we cannot omit the factor of $(1+t_{\mathit{fin}})^{1/q}$ in Theorem 3.2.

In keeping with the terminology of the theorems above, we say that there is no *pathwise, uniform in time* estimate. In the course of proving Theorem 3.4, we will show that the long-time deviation (3.9) is typical for the Kac process. We will show that the Kac process returns, infinitely often, to subsets of $\mathcal{S}_N$ which are far from the Boltzmann flow. However, we make the following remark on the times necessary for such deviations to occur.

Corollary 3.5. *Let $(\mu_t^N)_{t\geq 0, N\geq 2}$ be a family of Kac processes with an initial exponential moment bound $\mathbb{E}\langle e^{z|v|}, \mu_0^N\rangle \leq b$, for some $z > 0$ and $b > 0$ and suppose that $\mathbb{E}[W_1(\mu_0^N, \mu_0)] \leq CN^{-\alpha}$, for some $\alpha > 0$ and some nonrandom μ_0. Define*

$$T_{N,\epsilon} = \inf\left\{t \geq 0 : W_1(\mu_t^N, \phi_t(\mu_0)) > \epsilon\right\} \tag{3.10}$$

and let $t_{N,\epsilon,\delta}$ be the quantile constants of $T_{N,\epsilon}$; that is,

$$t_{N,\varepsilon,\delta} = \inf\left\{t \geq 0 : \mathbb{P}(T_{N,\epsilon} \leq t) \geq \delta\right\}. \tag{3.11}$$

Then, for fixed $\epsilon, \delta > 0$, $t_{N,\epsilon,\delta} \to \infty$, faster than any power of N.

This follows as an immediate consequence of Theorem 3.2. Taken together with Theorem 3.4, we see that macroscopic deviations occur, but typically at times growing faster than any power of N.

In the course of proving Theorems 3.1, 3.2, we will establish the following continuity estimate for the Boltzmann flow ϕ_t measured in the Wasserstein distance W_1, which may be of independent interest.

Theorem 3.6. *There exist constants p, C, w depending only on d such that, whenever $a \geq 1$ and $\mu, \nu \in \mathcal{S}$ have $\Lambda_p(\mu, \nu) \leq a$, we have the estimate*

$$W_1\left(\phi_t(\mu), \phi_t(\nu)\right) \leq Ce^{wt} a W_1(\mu, \nu). \tag{3.12}$$

Moreover, for all $p > 2$, there exist constants $C = C(p, d)$ and $\zeta = \zeta(p, d) > 0$ such that, whenever $\mu, \nu \in \mathcal{S}$ with $\Lambda_p(\mu, \nu) \leq a$, we have the estimate

$$\sup_{t \geq 0} W_1(\phi_t(\mu), \phi_t(\nu)) \leq CaW_1(\mu, \nu)^{\zeta}. \tag{3.13}$$

In the second part of the theorem, and in Theorems 3.1, 3.2 above, the exponent ζ can be taken to be $\lambda_0/(\lambda_0 + 2w)$ by making p large enough, where w is as in the first part of the theorem, and $\lambda_0 = \lambda_0(d) > 0$ is the spectral gap of the linearised Boltzmann operator, see Section 3.2. While it may be possible to obtain better continuity results, with ζ close to 1, we will not explore this here.

Remark 3.7. *Let us remark that all of the analysis in this chapter would apply equally well to the case of cutoff hard potentials (CHP$_K$). However, carefully following the proofs shows that the number of required moments p would diverge as $K \to \infty$. This would be catastrophic for our programme in Chapter 4, where we take $K \to \infty$, uniformly in N and with only finitely many moments, as a step towards the propagation of chaos for the* non*cutoff case (NCHP).*

3.1.1 Plan of the Chapter

Our programme will be as follows:

i. In the remainder of this section, we will then discuss several aspects of our results in view of the literature.

ii. Section 3.2 reviews some elements of the previous works of Mischler and Mouhot [142] and Norris [157] which we will use in this chapter. We cite the analytical *regularity and stability estimates* from [142], and a representation formula in terms of branching processes from [157] with associated estimates. Together, these allow us to prove Theorem 3.6.

iii. In Section 3.3, we use ideas of infinite-dimensional differential calculus, developed by [142] and recalled in Section 3.2, to prove an *interpolation decomposition* of the difference $\mu^N_t - \phi_t(\mu^N_0)$. This is the key identity used for the proofs of Theorems 3.1, 3.2, as all of the terms appearing in our formula can be controlled by the stability estimates.

iv. We then turn to the proof of Theorem 3.1. The main technical aspect is the control of a family of martingales $(M^{N,f}_t)_{f \in \mathcal{A}_1}$, uniformly in f. This is obtained using a quantitative compactness argument similar to that in [157].

v. For a local uniform analysis, we first adopt the ideas of Theorem 3.1 to a local uniform setting, with suitable adaptations, to state a local uniform martingale estimate, and deduce a preliminary, weak version of Theorem 3.2 with worse dependence in t_{fin}. We then use the stability estimates to 'bootstrap' to the improved estimate Theorem 3.2, and finally return to prove the local martingale estimate.

vi. We next prove Theorem 3.3. The strategy here is to use the branching process representation from [157], recalled in Section 3.2, to control behaviour on a very short time interval $[0, u_N]$, and use the previous results, together with the moment production property from Propositions 2.8-2.10, to control behaviour at times larger than u_N.

vii. We prove Theorem 3.4, based on relaxation to equilibrium.

3.1.2 Discussion of Our Results

In this subsection, we will discuss the interpretation of our results, especially in view of the framework of chaoticity set in Chapter 2.

1. Theorems 3.1, 3.2 as a pathwise interpretation of the Boltzmann Equation
In this chapter in particular, our approach is driven by viewing μ_t^N as a noisy perturbation of the Boltzmann equation (BE). This follows the same approach as Norris [157], see also the works [49, 158] on general approximations of ordinary differential equations by Markov processes. The approach of Theorem 3.2 of seeking pathwise local uniform estimates is particularly natural in this context, as this corresponds to typical estimates from the theory of fluid limits, for example [49, Theorem 4.2]. We also write everything in terms of the maps $\phi_t : \mathcal{S} \to \mathcal{S}$, as the analysis is based on exploiting properties of these maps, as in [142, 158].

Another example from kinetic theory in which this philosophy is natural is the case of Vlasov dynamics. In this case, we write $\mu_t^{N,\text{Vl}}$ for the N-particle empirical measure, evolving under (nonrandom) Hamiltonian dynamics; Dobrushin [63] showed that $\mu_t^{N,\text{Vl}}$ is a weak measure solution to the associated mean field Vlasov equation. For the case of Kac dynamics, we may interpret Theorems 3.1, 3.2 as saying that

$$\forall t \geq 0 \qquad \mu_t^N = \phi_t(\mu_0^N) + \mathcal{N}_t^N \tag{3.14}$$

where $\mathcal{N}_t^N$ is a stochastic noise term, which is small in an appropriate sense; compare also to (1.15) in the introduction. This is a general phenomenon in the 'fluid limit' scaling of Markov processes [49, 157, 158]. In this sense, we may interpret the Boltzmann equation in a *pathwise* sense, which is valid for any individual N, and without chaoticity assumptions on the initial data. In estimating $W_1(\phi_t(\mu_0), \mu_0^N)$ using either of the theorems, the assumption of chaoticity of the initial data will enter via the requirement that $W_1(\mu_0, \mu_0^N)$ is small, which is equivalent to chaoticity in the $N \to \infty$ limit by Proposition 2.1.

2. Comparison of Results and Techniques to the literature. The analysis of this chapter follows leading to Theorems 3.1 - 3.2 the ideas of a Mischler and Mouhot [142] in a pathwise framework and incorporating suitable probabilistic ideas from the work of Norris [157], as discussed above, and it is instructive to compare our results to those on which we build.

A first significant difference between Theorems 3.1 - 3.3 to those of Mischler and Mouhot [142, Theorems 6.1-6.2] is that we work pathwise, in the spirit of the fluid limits [49] mentioned above. Results similar to the pointwise estimates (Theorem 3.1 and the first item of Theorem 3.3) could be derived from the estimates on finite marginals similar to [142, Theorems 6.1-6.2], as in Proposition 2.1, but this is now subsumed into our proof, following the techniques of [157]. In the case of the local uniform estimates (Theorem 3.2 and the second item of Theorem 3.3), such estimates cannot even be expressed in terms of the single-time marginals $(\text{Law}(\mu^N_t))_{t\geq 0}$.

Secondly, we obtain a stronger rate of convergence than the results of Mischler and Mouhot, still uniformly in time, and under weaker moment conditions. The conclusions of [142, Theorems 6.1-6.2] have a final rate going as a negative power of $(\log N)$, which is weaker than our rate $N^{-\alpha}$. This is mainly a result of Theorem 3.6, which improves over the analagous estimate in [142, Section 6.8]. Moreover, we achieve these rates with only finitely many moments, allowing a p^{th} moment control, any $p > 2$, in Theorem 3.3 or (constructible, finite) p in Theorems 3.1-3.2, whereas the corresponding theorem in [142] required exponential moments or the compact support of μ_0. Let us also remark that all of the conditions (**A1-5**) from [142] are used our analysis:

i). Assumption (**A1**) corresponds to the moment bounds, which follow from the discussion of moment bounds in Propositions 2.6, 2.10.

ii). Assumption (**A2**i) and (**A5**) concern the continuity of the Boltzmann flow ϕ_t, which is the conclusion of Theorem 3.6, which we prove along the way. Assumption (**A2**ii) concerns the continuity of the collision operator Q, which is discussed in Section 3.2.3.

iii). Assumption (**A3**) is the convergence of the generators. A special case of this is the content of Lemma 3.19, which is used to prove our 'interpolation decomposition' Formula 3.3.1.

iv). Assumption (**A4**) is the differential stability of the Boltzmann flow ϕ_t, recalled in Proposition 3.15, which is crucial to obtaining estimates with good long-time properties.

By comparison to the work of Norris [157], which was the other main inspiration for the content of this chapter, the main novelty is the good behaviour of our estimates in long-time, compared to Proposition 1.2). When using the same branching process representation from [157] recalled in Section 3.2, which is helpful in short-time when only few moments are assumed in Theorem 3.3, or for the Boltzmann equation in Theorem 3.6, we introduce a new 'localisation' idea to control certain bad events on (very) short times before using the moment creation property from Proposition 2.10. This allows us to improve slightly over Proposition 1.2 and have a control in $L^1(\mathbb{P})$ rather than with high probability, see the proof of Theorem 3.3 in Section 3.6.

4. Sharpness of our Results We will now discuss how sharp the main results (Theorems 3.1, 3.2) are, with regards to dependencies in N, and the terminal time t_{fin} in the case of Theorem 3.2.

4a. N-dependence It is instructive to first consider the 'optimal' case of independent particles, for which the empirical measure converges in Wasserstein distance at rate $N^{-1/d}$. More precisely, for $d \geq 3$, let $\mu \in \mathcal{S}^p$ with an estimate $\Lambda_p(\mu) \leq a$ for $p \geq \frac{3d}{d-1}$, and let μ^N be an empirical measure for N independent draws from μ. Then, for some $C = C(a, k, d)$, we have

$$\left\| W_1(\mu^N, \mu) \right\|_{L^2(\mathbb{P})} \leq C N^{-1/d}. \tag{3.15}$$

This is shown in [157, Proposition 9.3]. Moreover, this rate is optimal: if μ is absolutely continuous with respect to the underlying Lebesgue measure, then the optimal approximation in W_1 metric is of the order $N^{-1/d}$, for $d \geq 3$. Results of Talagrand ([174, 175], and discussion in [87]) suggest that this may also be true for higher L^p norms, at least for the simple case of the uniform distribution on $(-1, 1]^d$.

In view of this, we see that the exponent for the pointwise bound is *almost sharp*, in the sense that we obtain exponents $\epsilon - \frac{1}{d}$ which are arbitrarily close to the optimal exponent $-\frac{1}{d}$, but cannot obtain the optimal exponent itself. This appears to be a consequence of using a particular estimate (3.85) from [142], which is 'almost Lipschitz' in a similar sense. For the local uniform estimate Theorem 3.2, we obtain exponent $-\alpha$, where α is given by

$$\alpha = -\epsilon + \frac{q'}{2d}; \qquad \frac{1}{q} + \frac{1}{q'} = 1 \tag{3.16}$$

when considering estimates in $L^q(\mathbb{P})$. In the special case $q = 2$, this produces the almost sharp exponent as discussed above. However, for $q > 2$, the exponents are bounded away from $-\frac{1}{d}$, and so do not appear to be sharp.

We believe that our techniques could be modified to prove an estimate for Theorem 3.1, and Theorem 3.2 in the case $q = 2$, in order to obtain the optimal rate $N^{-1/d}$ discussed above, for instance by using the short-time (Lipschitz) first-order stability estimate found in the proof of Proposition 3.15. However, this would likely come at the cost of poor dependence in time. Since a similar result (Proposition 1.2) is already known, and since this is not the spirit of this work in seeking to optimise time dependence, we will not consider this further.

4b. Time Dependence In light of Theorem 3.4, we see that we cannot exclude the factor $(1+t_{\text{fin}})^{1/q}$ in Theorem 3.2. Hence, this time dependence is sharp *among power laws*. However, we do not know what the *true* sharpest time-dependence is. Similar techniques to those of Graversen and Peskir [103] may be able to provide a sharper bound; we do not explore this here.

We remark that Theorem 3.2 interpolates between almost optimal N dependence at $q = 2$, and almost optimal t_{fin} dependence as $q \to \infty$. Moreover, by taking $q \to \infty$, we sacrifice optimal dependence in N, but the exponent $\alpha(d, q)$ is bounded away from 0, and so we have good convergence, on any polynomial time scale. This is the content of Corollary 3.5.

5. Other Models Since this chapter is based on a pathwise modification of the techniques of Mischler and Mouhot [142], models which satisfy the conditions discussed there may also be amenable to our techniques, see also the work [143].

a). *Other Kac Processes* In the work [142], Mischler and Mouhot show how the assumptions (**A1-5**.) also hold for the Kac process in the case of Maxwell molecules with or without cutoff (GMM, tMM), including the stability estimates analagous to Proposition 3.15 in some different functional spaces. For this case, a result similar to Theorem 3.1 is already known by a different method by Cortez and Fontbona [48, Theorem 2], using the Tanaka coupling discussed in the introduction; see also Chapter 4.

As already commented in Remark 3.7, the same techniques of this chapter would also apply in the case of cutoff hard potentials (CHP_K), although the number of moments p required would also depend on the cutoff K, which is not helpful for our purposes in Chapter 4, see also Remark 3.18. In the case of *non*cutoff hard potentials (NCHP), the study of the Boltzmann equation is significantly less advanced, since one must account for the double-unboundedness (unboundedness in the velocity variable as well as the angular singularity). Indeed, one of the main contributions we make in Chapter 4 is a new uniqueness and stability result (Theorem 4.1 and Corollary 4.2) which applies naturally to the empirical measures μ_t^N of the Kac process, in that it requires neither regularity nor exponential moments. Even with this result in hand, it is not obvious how to apply the techniques of this chapter, since we cannot prove the existence of a functional derivative, nor the 'second-order' estimate in the sense of Proposition 3.15.

b). *McKean-Vlasov Dynamics, and Inelastic Collisions.* Let us also mention the examples of McKean-Vlasov dynamics, and Inelastic Collisions, coupled to a heat bath, which have been studied in the functional framework of [142] by Mischler, Mouhot and Wennberg in the paper [143]. In these cases, the analagous estimates for stability and differentiability, computed in [143], in this case with exponential growth in time (analagous to Step 2 in the proof of Proposition 3.15). As a result, our methods would still apply, but with correspondingly poor time dependence.

For the case of McKean-Vlasov dynamics without confinement potential, this is a fundamental limitation; Malrieu [135] showed that the propagation of chaos is *not* uniform in time. Instead, he proposed to study a *projected* particle system, which satisfies uniform propagation of chaos, and whose limiting flow has exponential convergence to equilibrium [135, Theorem 6.2]. This suggests that it may be possible to use our bootstrap method, used in the proof of Theorem 3.2, to obtain a pathwise estimates with good long-time properties, analagous to Theorem 3.2.

We remark that, in the case of McKean-Vlasov dynamics, the presence of Brownian noise may complicate the derivation of the interpolation decomposition (Formula 3.3.1), which is the key identity required for our argument.

3.2 A review of Previous Analyses

In this section, we will recall in detail some previous analyses of the Kac process and the Boltzmann equation in the cutoff cases (HS, CHP_K) by Norris [157] and Mischler and Mouhot [142]. We will build on these results in this chapter, and we will also use the results in the case of cutoff hard potentials (CHP_K) as a stepping stone when we deal with the noncutoff hard potentials (NCHP) in Chapter 4. The results of Subsections 3.2.1 - 3.2.3 are not novel, and consist of either reproductions or slight modifications of the results of the cited works.

i). Subsection 3.2.1 is dedicated to a branching process representation of either the difference of a Kac process and the Boltzmann Equation, or two solutions to the Boltzmann equation, and some estimates for the objects appearing in this representation. We reproduce the main steps in the proof of [157, Theorem 1.1], which produces estimates in the weighted Wasserstein$_1$-type metrics W_γ assuming only $p > 2$ moments on the initial data. We give as a series of propositions, with the necessary modifications to cover (CHP_K) and to track, in this case, the dependence of the final result on K.

ii). Subsection 3.2.2 gives some consequences of the branching process formula, including a first step towards Theorem 3.6 and a convergence estimate for the Kac process in the case of cutoff hard potentials (CHP_K).

iii). Subsection 3.2.3 reviews the *differential stability* of the Boltzmann semigroup ϕ_t in the case of hard spheres (HS), closely following [142]. We will also review some continuity estimates for the collision operator Q in this case, and prove a technical boundedness property for the functional derivative $\mathcal{D}\phi_t$.

iv). In Subsection 3.2.4, we prove Theorem 3.6, based on comparing the short-time continuity from the branching process representation, and the long-time stability in the differential stability estimates. This contains the only (essential) novel content of this section.

Since we will use the result on the convergence of the Kac process in the case (CHP_K) as a step towards the noncutoff case in Chapter 4, we will present the arguments of Subsections 3.2.1 - 3.2.2 for both cases, understanding $B_K = B$, $K = 1, \gamma = 1$ in the case (HS).

We first note a definition of some spaces of signed measures, which will be useful throughout this section.

Definition 3.2.1. *Consider the space Y of signed measures, given by*

$$Y = \{\xi : \quad \|\xi\|_{\text{TV}} < \infty; \quad \langle 1, \xi \rangle = 0\}. \tag{3.17}$$

We equip Y with the total variation norm $\|\cdot\|_{\mathrm{TV}}$. For real $q \geq 0$, we define the subspace Y_q of measures with finite q^{th} moments:

$$Y_q = \{\xi \in Y : \langle 1 + |v|^q, |\xi| \rangle < \infty\}. \tag{3.18}$$

We define the norm with q-weighting on Y_q by

$$\|\xi\|_{\mathrm{TV}+q} = \langle 1 + |v|^q, |\xi| \rangle. \tag{3.19}$$

The notation $\|\cdot\|_{\mathrm{TV}+q}$ is chosen to emphasise that this is a total variation norm, with additional polynomial weighting of order q, while avoiding potential ambiguity with the L^q norms of random variables.

Remark 3.8. *The total variation norms $\|\cdot\|_{\mathrm{TV}+q}$ appearing in the following analysis are much stronger than the Wasserstein distance W_1 appearing in Theorems 3.1, 3.2, 3.3. We can understand this as follows. Recalling the definitions of $\mathcal{A}_\gamma$ in (2.7), we note that the* TV + 2 *distance is given by a duality*

$$\|\mu - \nu\|_{\mathrm{TV}+2} = \sup_{f \in \mathcal{A}_0} |\langle f, \mu - \nu \rangle| \tag{3.20}$$

and, if we write $\mathcal{A}_\gamma|_r$ for the restriction of functions to $[-r, r]^d$, then the inclusion

$$\mathcal{A}_\gamma|_r \subset \mathcal{A}_0|_r \tag{3.21}$$

is compact in the norm of $\mathcal{A}_0|_r$ for any $\gamma \in (0, 1]$, by the classical theorem of Arzelá-Ascoli. This is at the heart of a quantitative compactness *argument in Proposition 3.11 and Lemmas 3.21, 3.22, which allows us to take the supremum over $f \in \mathcal{A}_\gamma$ inside an expectation.*

3.2.1 Branching Process Representation

We begin by developing a representation formula for both the Kac process and the Boltzmann equation in the cases (HS) and cutoff hard potentials (CHP_K), first developed by Norris [157], which leads to the stability of the Boltzmann flow and the convergence of the Kac process in both cases, albeit not uniformly in time. This framework is particularly useful as a step towards Theorems 3.3 - 3.6, and in the case (CHP_K), because the estimates behave well with any $p > 2$ moments, uniformly in K.

Random Measures Associated to the Kac Process We first introduce the *jump measure* and compensator associated to the cutoff Kac process $(\mu_t^{N,K})_{t\geq 0}$. These definitions are general for Markov processes of finite rate, see [49]. We recall that $\mathcal{S}_N$ is the state space of the unlabelled Kac process normalised to $\mathcal{S}$, given by the set of all measures $\mu^N \in \mathcal{S}$ which are the normalised empirical measure of N points.

Definition 3.2.2. *[Jump Measure and Compensator] Let $(\mu_t^{N,K})_{t\geq 0}$ be a Kac process for the kernel B_K on N particles, either in the case (HS, CHP_K).*

(i.) The jump measure $m^{N,K}$ *is the unnormalised empirical measure on $(0,\infty)\times\mathcal{S}_N$ on all pairs $(t,\mu_t^{N,K})$ such that $\mu_t^{N,K} \neq \mu_{t-}^{N,K}$.*

(ii.) Let $\mathcal{Q}_{N,K}$ be the kernel on $\mathcal{S}_N$ given by

$$\mathcal{Q}_{N,K}(\mu^N, A) = N\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}} \mathbb{1}(\mu^{N,v,v_\star,\sigma}\in A)B_K(v-v_\star,d\sigma)\mu^N(dv)\mu^N(dv_\star). \tag{3.22}$$

The compensator $\overline{m}^{N,K}$ *of the jump measure is the measure on $(0,\infty)\times\mathcal{S}_N$ given by*

$$\overline{m}^{N,K}(dt, A) = \mathcal{Q}_{N,K}(A, d\mu^N)dt. \tag{3.23}$$

Since we are working with the cutoff process, both of these measures are almost surely finite on compact subsets $(0,t]\times\mathcal{S}_N$, for any $t<\infty$. In the case of hard spheres, we omit the sub/superscripts $\cdot^K, \cdot_K$. A calculus of martingales for such processes is recalled in Appendix 3.A.

A Branching Process We next introduce branching processes, which gives a probabilistic representation of the difference between two solutions to (BE) or such a solution and a Kac process.

Definition 3.2.3 (Linearised Kac Process). *Write $V=\mathbb{R}^d$ and V^* for the signed space $V^* = V\times\{\pm 1\} = V^+\sqcup V^-$. We write $\pi: V^*\to V$ as the projection onto the first factor,*

and $\pi_\pm : V^\pm \to V$ for the obvious bijections.
Let $(\rho_t)_{t\geq 0}$ be family of measures on $V = \mathbb{R}^d$ such that[1]

$$\langle 1, \rho_t\rangle = \langle |v|^2, \rho_t\rangle \leq 1 \qquad \textit{for all } t \geq 0; \qquad \Lambda_{2+\gamma}(\rho_t) \in L^1_{loc}([0,\infty)). \tag{3.24}$$

The Linearised Kac Process in environment $(\rho_t)_{t\geq 0}$ *is the branching process on V^* where each particle of type $(v, 1)$, at rate $2B_K(v - v_*, \sigma)\rho(dv_\star)d\sigma$, dies, and is replaced by three particles, of types*

$$(v'(v, v_\star, \sigma), 1); \quad (v'_\star(v, v_\star, \sigma), 1); \quad (v_\star, -1) \tag{3.25}$$

where $v', v'_\star$ are the post-collisional velocities given by (6.1). The dynamics are identical for particles of type $(v, -1)$, with the signs exchanged.

We write Ξ^*_t for the associated process of unnormalised empirical measures on V^*, and define a signed measure Ξ_t on V by including the sign at each particle:

$$\Xi_t = \Xi^+_t - \Xi^-_t; \qquad \Xi^\pm_t = \Xi^\star_t \circ \pi^{-1}_\pm. \tag{3.26}$$

We can also consider the same branching process, started from a time $s \geq 0$ instead. We write E for the expectation over the branching process, which is not the full expectation in the case where ρ is itself random. When we wish to emphasise the initial velocity v and starting time s, we will write $E_{(s,v)}$ when the process is started from $\Xi^*_0 = \delta_{(v,1)}$ at time s, and E_v in the case $s = 0$.

Throughout, we write $|\cdot|$ for the total variation measure associated to the signed measures Ξ_t on $V = \mathbb{R}^d$. Provided that the initial data Ξ_0 has a finite second moment $\langle 1+|v|^2, |\Xi_0|\rangle$, this bound is propagated, and in particular the branching process is almost surely non-explosive:

$$E\langle 1 + |v|^2, |\Xi_t|\rangle \leq \langle 1 + |v|^2, |\Xi_0|\rangle \exp\left[CK\int_0^t \Lambda_{2+\gamma}(\rho_s)ds\right] \tag{3.27}$$

or more generally, if $\rho \in L^1_{\text{loc}}([0,\infty), \mathcal{S}^{p+\gamma})$ for $p \geq 2$, there is a constant C, now also depending on p, such that

$$E\langle 1 + |v|^p, |\Xi_t|\rangle \leq \langle 1 + |v|^p, |\Xi_0|\rangle \exp\left[CK\int_0^t \Lambda_{p+\gamma}(\rho_s)ds\right]. \tag{3.28}$$

In the case where we start from $\Xi_0 = \delta_{(v,1)}$, the first term on the right-hand side is $(1 + |v|^2)$. Recall from Chapter 2 that $\mathcal{A}_\gamma$ is the set of all functions f on $\mathbb{R}^d$, such that $\widehat{f}(v) = (1 + |v|^2)^{-1}f(v)$ satisfies

$$|\widehat{f}(v)| \leq 1; \qquad \frac{|\widehat{f}(v) - \widehat{f}(w)|}{|v - w|^\gamma} \leq 1 \qquad \text{for all } v \neq w. \tag{3.29}$$

[1]This is a mild generalisation of the definition in [157, Section 4], but the proofs work in exactly the same way.

From the bound (3.27), we can now define, for functions of quadratic growth,

$$f_{st}(v_0) = E_{(s,v_0)}\left[\langle f, \Xi_t\rangle\right]. \tag{3.30}$$

When we wish to emphasise the environment, we will write $f_{st}[\rho](v_0)$. We now recall the following estimates from [157]:

Proposition 3.9 (Continuity Estimates for f_{st}). *Fix $t \geq 0$, and for $\beta \in (0,1)$, let $y_\beta(t), z_t$ be given by*

$$y_\beta(t) = z_t \sup_{0\leq s<s'\leq t}\left[(s'-s)^{-\beta}\int_s^{s'}\Lambda_{2+\gamma}(\rho_u)du\right]; \tag{3.31}$$

$$z_t = \exp\left[CK\int_0^t \Lambda_{2+\gamma}(\rho_u)du\right]. \tag{3.32}$$

Then the constant C in the definition of z_t can be chosen, depending only on d, such that, for $f \in \mathcal{A}_\gamma$ and $s \leq t$, we have $f_{st} \in z_t\,\mathcal{A}_\gamma$. Moreover, for the same Cf, for all v and all $0 \leq s \leq s' \leq t$,

$$|f_{st}(v) - f_{s't}(v)| \leq C(1+|v|^{2+\gamma})y_\beta(t)(s'-s)^\beta.$$

This is, in our notation, a reformulation of [157, Propositions 4.3]. We apply this to the Kac process and the Boltzmann equation via the following representation formula, which extends [157, Proposition 4.2].

Proposition 3.10. *[Representation formula for Cutoff Cases] Let us fix $\mu_0, \nu_0 \in \mathcal{S}^{2+\varepsilon}$ for some $\varepsilon > 0$, and corresponding solutions $(\mu^K_t)_{t\geq 0}, (\nu^K_t)_{t\geq 0}$ to the Boltzmann equation (BE), and let $\mu^{N,K}_t \in \mathcal{S}_N$ be a N-paraticle Kac process. Let $m^{N,K}, \overline{m}^{N,K}$ be the jump measure and compensator for the Kac process given by Definition 3.2.2.*

i). (Kac Case) *Consider the branching processes defined above and write $f_{st} = f_{st}[\rho^N]$ with the random environment*

$$\rho^N_t = \frac{\mu^{N,K}_t + \mu^K_t}{2}. \tag{3.33}$$

Then, for all $t \geq 0$, and all functions f of quadratic growth, we have

$$\langle f, \mu^{N,K}_t - \mu^K_t\rangle = \langle f_{0t}[\rho^N], \mu^{N,K}_t - \mu_0\rangle + M^{N,K,f}_t \tag{3.34}$$

where

$$M^{N,K,f}_t = \int_{(0,t]\times\mathcal{S}_N}\langle f_{st}[\rho^N], \mu^N - \mu^{N,K}_{s-}\rangle(m^{N,K} - \overline{m}^{N,K})(ds, d\mu^N). \tag{3.35}$$

ii). (Boltzmann Case) *Consider the deterministic environment*

$$\rho_t = \frac{\mu^K_t + \nu^K_t}{2}. \tag{3.36}$$

Then, for all $t \geq 0$ and all f of quadratic growth,

$$\langle f, \mu^K_t - \nu^K_t\rangle = \langle f_{0t}, \mu_0 - \nu_0\rangle. \tag{3.37}$$

Our last proposition for both cases is the following, which controls a suprema of the stochastic integrals similar to those on the right-hand side of (3.34). The following proposition will give such a control for a related family of martingales, again with a potentially random environment, assuming an almost sure bound on the $(2+\gamma)^{\text{th}}$ moment $\Lambda_{2+\gamma}(\rho_t)$, which diverges no faster than $t^{-1+\varepsilon}$ as $t \downarrow 0$ for some $\varepsilon > 0$.

Proposition 3.11. *Let ρ_t be a potentially random environment such that, for some $\beta \in (0,1)$,*

$$w = \left\| \sup_{t\geq 0} \left(\frac{\Lambda_{2+\gamma}(\rho_t)}{\beta t^{\beta-1}+1} \right) \right\|_{L^\infty(\mathbb{P})} < \infty. \tag{3.38}$$

For $f \in \mathcal{A}_\gamma$ and $0 \leq s \leq t$, let $f_{st} = f_{st}[\rho]$ denote the propagation in this environment, as described above.

Let $p > 2$ and $a \geq 1$, and let $\mu_t^{N,K} \in \mathcal{S}_N$ be a Kac process with initial moment $\mathbb{E}[\Lambda_p(\mu_0^N)] \leq a$, and let $m^{N,K}, \overline{m}^{N,K}$ be the random measures given by Definition 3.2.2. We write

$$\widetilde{M}_t^{N,K,f}[\rho] = \int_{(0,t]\times\mathcal{S}_N} \langle f_{st}[\rho], \mu^N - \mu_{s-}^{N,K}\rangle (m^{N,K} - \overline{m}^{N,K})(ds, d\mu^N). \tag{3.39}$$

In this notation, for any $t_{\text{fin}} \geq 0$, we have the bound

$$\left\| \sup_{t\leq t_{\text{fin}}} \sup_{f\in\mathcal{A}_\gamma} \widetilde{M}_t^{N,K,f} \right\|_{L^1(\mathbb{P})} \leq a e^{CKw(1+t_{\text{fin}})} N^{-\eta} \tag{3.40}$$

for some $C = C(d,p,\beta,\gamma)$ and $\eta = \eta(d,p,\beta,\gamma) > 0$. Here, we emphasise that $\|\cdot\|_{L^1(\mathbb{P})}$ refers to the L^1 norm with simultaneous expectation over μ_t^N and the environment ρ.

Proof of Propositions

We start with the following lemma.

Lemma 3.12. *In either case (HS, CHP$_K$), there exists $C < \infty, \alpha > 0$, depending only on b such that, for all $v, v' \in \mathbb{R}^d$, we have the estimate*

$$\sup_{v_\star\in\mathbb{R}^d} \|B_K(v - v_\star, \cdot) - B_K(v' - v_\star, \cdot)\|_{L^1(d\sigma)} \leq CK^\alpha |v - v'|^\gamma. \tag{3.41}$$

Proof. The case for (HS) is immediate; we check the case (CHP$_K$). Using the convexity of b, it follows that b is differentiable almost everywhere on $(0,1)$, and from the asymptotic of b, one can check the bound

$$b'(x) \leq C(1-x)^{-2-\nu/2} \tag{3.42}$$

for some constant C, which depends only on the singularity of b in (NCHP). From (NCHP), we calculate that the cutoff is at $\theta_0(K) = G(K) \leq CK^{1-1/\nu}$ in the notation of Section 2.4, and so there exists $\alpha > 0, C$ such that

$$\sup_{x\leq\cos\theta_0(K)} (|b(x) + |b'(x)|) \leq CK^\alpha. \tag{3.43}$$

Let us fix $u, u' \in \mathbb{S}^{d-1}$. By splitting the integral into three regions, we find

$$\begin{aligned}\int_{\mathbb{S}^{d-1}} |B_K(u,\sigma) - B_K(u',\sigma)| d\sigma &\le |u'-u| \sup_{\theta \ge \theta_0(K)} (|b'(\cos\theta)| + |b(\cos\theta)|) \\ &\le C|u'-u|\, K^\alpha.\end{aligned} \tag{3.44}$$

Accounting for the change in the kinetic factor, the function $\Psi(r) = r^\gamma$ is Lipschitz continuous on $[1,\infty)$, and so this extends to general v, w of norm at least 1:

$$\|B_K(v,\cdot) - B_K(w,\cdot)\|_{L^1(d\sigma)} \le C|v-w|K^\alpha \qquad \forall v, w : |v|, |w| \ge 1. \tag{3.45}$$

We now consider the function

$$F(v_\star) = \|B_K(e_1 - v_\star, \cdot) - B_K(-e_1 - v_\star, \cdot)\|_{L^1(d\sigma)}. \tag{3.46}$$

If $|v_\star| \le 2$, then we use the bound $\|B_K(\pm e_1 - v_\star, \cdot)\|_{L^1(d\sigma)} \le CK$ to see that $F(v_*) \le CK$. On the other hand, in the region $|v_\star| \ge 2$, it follows from (3.45) that $F \le CK^\alpha$ and, combining, we conclude that

$$\sup_{v_\star \in \mathbb{R}^d} \|B_K(e_1 - v_\star, \cdot) - B_K(-e_1 - v_\star, \cdot)\|_{L^1(d\sigma)} \le CK^\alpha. \tag{3.47}$$

For general $v \ne v'$, there exists a rotation and scaling of $\mathbb{R}^d$ taking $2e_1$ to $v - v'$. Using the invariance under rotation and the scaling property of B_K, we conclude that

$$\sup_{v_\star \in \mathbb{R}^d} \|B_K(v - v_\star, \cdot) - B_K(v' - v_\star, \cdot)\|_{L^1(d\sigma)} \le CK^\alpha |v - v'|^\gamma \tag{3.48}$$

as desired. □

We now turn to Proposition 3.9.

Proof. We mimic the arguments [157, Propositions 4.3, 4.5] with the necessary adaptations for our case.

Step 1. Growth Bound Firstly, the estimate (3.27) already cited above proves the claimed growth condition.

Step 2. Velocity Dependence In order to estimate the difference $f_{0t}(v_0) - f_{0t}(w_0)$, we introduce a coupling of the processes $\Xi^\star_t$ started at the initial data $(v_0, 1), (w_0, 1)$. We consider a branching process on $W = (\mathbb{R}^d \times \mathbb{R}^d) \sqcup \mathbb{R}^d \sqcup \mathbb{R}^d = W_0 \sqcup W_1 \sqcup W_2$, where particles can either be coupled pairs $(v, w) \in (\mathbb{R}^d \times \mathbb{R}^d)$, or uncoupled particles in one of two disjoint copies of $\mathbb{R}^d$, and where each particle is assigned a sign ± 1. The branching rules for uncoupled particles are the same as in Definition 3.2.3, while coupled particles of type (v,w) scatter to remain coupled as far as possible, but undergo decoupling transitions at rate

$$\|B_K(v-v_\star,\cdot)-B_K(w-v_\star,\cdot)\|_{L^1(d\sigma)}\,\rho_t(dv_\star). \tag{3.49}$$

Let $\Gamma^{0\star},\Gamma^{1\star},\Gamma^{2\star}$ be the empirical measures on $W_i\times\{\pm1\}$, $i=0,1,2$, and consider the projection maps $p_i: W_0\times\{\pm1\}\to\mathbb{R}^d\times\pm1$ by projecting onto the i^{th} marginal, $i=1,2$. The empirical measures

$$\Xi^{i\star}_t=\Gamma^{0\star}_t\circ p_i^{-1}+\Gamma^{i\star}_t, \qquad i=1,2 \tag{3.50}$$

are now a coupling of Linearised Kac processes. Using the bound (3.27) on each marginal, we have estimates starting from a coupled pair

$$E_{(0,(v_0,w_0)\in V_0)}\langle 1+|v|^2+|w|^2,\Gamma^{0\star}_t\rangle \le z_t\,(1+|v_0|^2+|w_0|^2) \tag{3.51}$$

or from decoupled particles, for all $0\le s\le t$,

$$E_{(s,v_0\in V_1)}\langle 1+|v|^2,\Gamma^{1\star}_t\rangle \le z_t\,(1+|v_0|^2) \tag{3.52}$$

and similarly for V_2. Let us now run this process starting from a particle of type $(v_0,w_0)\in V_0$. Using the triangle inequality inductively, $\Gamma^{0\star}$ is supported only on coupled pairs $(u,u')\in V_0$ with $|u-u'|\le|v_0-w_0|$, and thanks to Lemma 3.12, the rate of decoupling of such a pair is at most $CK^\alpha|v_0-w_0|^\gamma$. With this modification, the proof of [157, Lemma 4.5] now gives the estimate

$$E_{(0,(v_0,w_0)\in V_0)}\langle 1+|v|^2,\Gamma^{1,\star}_t+\Gamma^{2,\star}_t\rangle \le CK^\alpha|v_0-w_0|^\gamma(1+|v_0|^2+|w_0|^2)\,z_t. \tag{3.53}$$

Let us fix $f\in\mathcal{A}_\gamma$. Since the processes $\Xi^{i,\star}_t$ give a coupling of the linearised Kac processes started at $(v_0,1),(w_0,1)$ respectively, we have

$$f_{0t}(v_0)-f_{0t}(w_0)=E_{(0,(v_0,w_0)\in V_0)}\left\{\langle f\circ p_1-f\circ p_2,\Gamma^0_t\rangle+\langle f,\Gamma^1_t\rangle-\langle f,\Gamma^2_t\rangle\right\}. \tag{3.54}$$

On the support of Γ^0_t, the difference $f\circ p_1-f\circ p_2$ is at most $3(1+|v|^2+|w|^2)|v_0-w_0|^\gamma$, and we can estimate the integral using (3.51). The other terms only gain contributions from decoupled particles, and we can estimate both such terms using (3.53) and recalling that $|f|\le 1+|v|^2$. We therefore put everything together to conclude that

$$|f_{0t}(v_0)-f_{0t}(w_0)|\le CK^\alpha|v_0-w_0|^\gamma(1+|v_0|^2+|w_0|^2)\,z_t. \tag{3.55}$$

which is the regularity desired. Finally, since $K^\alpha\le\exp(CK)$ only appears in the decoupling rate and appears only as a multiplicative factor, rather than in the exponent, it can be absorbed into z_t, by changing the value of C if necessary.

Step 3. Time Dependence We finally deal with the time dependence, as in [157, Proposition 4.4]. To shorten notation, let us fix t and write $f_s := f_{st}$ for $0 \le s \le t$, and let $(\Xi_u)_{u\ge s}, (\Xi'_u)_{u\ge s}$ be independent copies of the branching process, both starting from a single particle of type $(v,1)$, at times s, s' respectively.

Let us set T to be the time of the first branching event in $(\Xi_u)_{u\ge s}$, and $V_\star, V', V'_\star$ the resulting post-collision velocities. We consider separately the cases $T \in (s,s')$ and $T \ge s'$. In the first case, the Markov property gives

$$E(\langle f, \Xi_t - \Xi'_t\rangle | T, V_\star, V', V'_\star) = f_T(V') + f_T(V'_\star) - f_T(V_\star) - f_{s'}(v)$$

which is at most, in absolute value, $z_t(4+|v|^2+|V_\star|^2+|V'|^2+|V'_\star|^2) = z_t(4+2|v|^2+2|V_\star|^2)$, using the growth bound in Step 1 and the conservation of energy $|v|^2+|V_\star|^2 = |V'|^2+|V_\star|^2$. On the other hand, conditional on $T > s'$, the law of $(\Xi_u)_{u\ge s'}$ is exactly that of $(\Xi'_u)_{u\ge s'}$ by the Markov property, so $E(\langle f, \Xi_t - \Xi'_t\rangle | T > s') = 0$. We conclude that

$$\begin{aligned} |f_s(v) - f_{s'}(v)| &= |E\langle f, \Xi_t - \Xi'_t\rangle| = |E\langle f, \Xi_t - \Xi'_t\rangle 1\!\!1_{\{T\in(s,s')\}}| \\ &\le \int_s^{s'} \int_{\mathbb{R}^d\times\mathbb{S}^{d-1}} (4+2|v|^2+2|v_\star|^2) z_t B_K(v-v_\star,\sigma)\rho_u(dv_\star)d\sigma du \\ &\le K z_t \int_s^{s'} (4+2|v|^2+2|v_*|^2)|v-v_*|^\gamma \rho_u(dv_*)du \\ &\le CK(1+|v|^{2+\gamma})z_t \int_s^{s'} \Lambda_{2+\gamma}(\rho_u)du. \end{aligned}$$

The final factor is, by definition of y_β, at most $(s'-s)^\beta y_\beta(t)$, and the prefactor (CK) can be absorbed into z_t, up to a new choice of the constant C in the exponent. □

We next sketch the proof of the representation formula.

Sketch Proof of Proposition 3.10. We follow the argument of [157, Proposition 4.2], and sketch the argument in the Kac case; we discuss the modifications for the Boltzmann case at the end. First, we introduce a signed, random measure $\mathfrak{M}$ on $[0,\infty)\times\mathbb{R}^d$ by specifying, for compactly supported $f : [0,\infty)\times\mathbb{R}^d \to \mathbb{R}$,

$$\int_{(0,\infty)\times\mathbb{R}^d} f(s,v)\mathfrak{M}(ds,dv) := \int_{(0,\infty)\times\mathcal{S}_N} \langle f_s, \mu^N - \mu^{N,K}_{s-}\rangle (m^{N,K} - \overline{m}^{N,K})(ds, d\mu^N)$$

where $m^{N,K}, \overline{m}^{N,K}$ are the jump measure and compensator in Definition 3.2.2. Let us also remark that $\mathfrak{M}$ has compact support in the velocity variable, thanks to the energy bound of μ^N, and so all moments are finite. We now consider the branching process Ξ_t as above, but where particles are intiated randomly according to a Poisson random measure on $[0,\infty)\times V^\star$ of intensity

$$\Theta(dt,dv) = \begin{cases} \mu_0^{N,K}(dv)\delta_0(dt) + \mathfrak{M}^+(dt,dv) & \text{on } V^+; \\ \mu_0(dv)\delta_0(dt) + \mathfrak{M}^-(dt,dv) & \text{on } V^- \end{cases}$$

and, writing E for the expectation over both the branching process and the initialisation, define a measure ξ_t by $\xi_t := E\Xi_t$, which we understand as a Bochner integral in the space Y_2 of signed measured with second moments.

We now derive an evolution equation for ξ_t. For any given (s, v_0), by considering the jumps in both the positive and negative parts, for any bounded, measureable f,

$$\langle f, \Xi_t \rangle - \int_0^t \langle f, 2Q_K(\rho_u, \Xi_u)\rangle du$$

is a martingale for the expectation $E_{(s,v_0)}$. Taking expectations gives

$$E_{(s,v_0)}\langle f, \Xi_t\rangle = f(v_0) + 2\int_s^t E\langle f, Q_K(\Xi_u, \rho_u)\rangle du = f(v_0) + \int_s^t \langle f, Q_K(E_{(s,v_0)}\Xi_u, \rho_u)\rangle du$$

where, in the second equality we use the fact that, in the cutoff cases, $Q_K(\cdot, \rho_u) : (Y_2, \|\cdot\|_{\mathrm{TV}+2}) \to (Y_0, \|\cdot\|_{\mathrm{TV}})$ is a bounded linear map; since the left-hand side is exactly the definition of $f_{st}(v_0)$, we can write

$$f_{st}(v_0) = f(v_0) + \int_s^t \langle f, Q_K(E_{(s,v_0)}\Xi_u, \rho_u)\rangle du. \tag{3.56}$$

We now integrate over (s, v_0) with respect to Θ. On the one hand, we have

$$\begin{aligned}\langle f, \xi_t\rangle = \int_{[0,T]\times V} E_{(s,v)}\langle f, \Xi_t\rangle \Theta(ds, dv) &= \int_{[0,T]\times V} f_{st}(v)\Theta(ds, dv) \\ &= \langle f_{0t}, \mu_0^{N,K} - \mu_0\rangle + \int_{(0,t]\times\mathbb{R}^d} f_{st}(v)\mathfrak{M}(ds, dv)\end{aligned} \tag{3.57}$$

and the second term here is exactly $M_t^{N,K,f}$. On the other hand, integrating the right-hand side of (3.56) produces

$$\langle f, \xi_t\rangle = \langle f, \mu_0^{N,K} - \mu_0\rangle + \int_{(0,t]\times\mathbb{R}^d} f(v)\mathfrak{M}(ds, dv) + 2\int_0^t \langle f, Q_K(\xi_u, \rho_u)\rangle du$$

using Fubini's theorem and the boundedness of the linear map $Q_K(\cdot, \rho_u)$ again, and that $\xi_t = \int_{[0,t]\times V^\star} E_{(s,v_0)}[\Xi_t]\Theta(ds, dv_0)$. Since f was an arbitrary bounded, measurable function, it follows from the previous display that ξ_t satisfies a noisy linear Boltzmann equation

$$\xi_t = \mu_0^{N,K} - \mu_0 + \mathfrak{M}((0,t]\times\cdot) + 2\int_0^t Q_K(\xi_u, \rho_u)du \tag{3.58}$$

which we understand as an equality of signed measures. We now use this evolution formula to show that $\xi_t = \mu_t^{N,K} - \mu_t^K$, and the conclusion of the proposition then follows from (3.57). We first show that $\mu_t^{N,K} - \mu_t^K$ also solves (3.58). For any f as above, we observe that

$$\langle f, \mu_t^{N,K}\rangle = \langle f, \mu_0^{N,K}\rangle + \int_0^t \langle f, Q_K(\mu_u^{N,K})\rangle du + \int_{(0,t]\times\mathbb{R}^d} f(v)\mathfrak{M}(du, dv);$$

$$\langle f, \mu_t^K\rangle = \langle f, \mu_0\rangle + \int_0^t \langle f, Q_K(\mu_u^K)\rangle du$$

where the first equality follows from the definition of $m^{N,K}$ and $\overline{m}^{N,K}$, and where the second equation is simply the (cutoff) Boltzmann equation. Subtracting, $\widetilde{\xi}_t = \mu_t^{N,K} - \mu_t^K$ satisfies

$$\widetilde{\xi}_t = \widetilde{\xi}_0 + \mathfrak{M}((0,t] \times \cdot) + \int_0^t (Q_K(\mu_u^{N,K}) - Q_K(\mu_u^K))du.$$

Using the bilinearity and symmetry of Q_K, and recalling that $Q_K(\mu)$ is shorthand for $Q_K(\mu,\mu)$, we have

$$\begin{aligned} Q_K(\mu_u^{N,K}) - Q_K(\mu_u^K) &= Q_K\left(\mu_u^{N,K} + \mu_u^K, \mu_u^{N,K} - \mu_u^K\right) \\ &= 2Q_K\left(\rho_u, \widetilde{\xi}_u\right) \end{aligned}$$

where, in the final line, we use the definition of ρ in the statement and $\widetilde{\xi}_u := \mu_u^{N,K} - \mu_u^K$, so that $\widetilde{\xi}_t$ also satisfies (3.58). To conclude the claimed equality, the difference $\delta_t := \xi_t - \widetilde{\xi}_t$ satisfies the equation

$$\delta_t = \int_0^t 2Q_K(\delta_u, \rho_u)du$$

so it remains to prove that the only solution to this in $C([0,\infty), Y_2)$ is the the trivial 0 solution. We follow Norris [157, p. 18] using a measure formulation of Di Blasio's L^1 argument [61], see also [142, Lemma 6.3], [76, Lemma 3.2]. For such a solution, there exists a measurable function $f : [0,\infty) \times V \to \{\pm 1, 0\}$ such that $\delta_t = f_t|\delta_t|$ and $|\delta_t| = \int_0^t f_s \nu_s ds, \nu_s = 2Q_K(\delta_s, \rho_s)$. We now define $\check{f}_s(v) := (1+|v|^2)f_s(v)$ and consider

$$\langle \check{f}_s, \nu_s\rangle = 2\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}} 2(\check{f}_s(v') + \check{f}_s(v'_*) - \check{f}(v) - \check{f}(v_*))B(v-v_*,\sigma)\delta_s(dv)\rho_s(dv_*)d\sigma. \tag{3.59}$$

In the terms involving v', v'_*, v_*, we have

$$\check{f}(v') + \check{f}(v'_*) + \check{f}(v_*) \le 3 + |v'|^2 + |v'_*|^2 + |v_*|^2 = 3 + |v|^2 + 2|v_*|^2$$

using the conservation of energy, and we majorise the integrals by replacing δ_s by $|\delta_s|$, while in the term involving $\check{f}(v)$, we keep the negative sign $-\check{f}(v)B(v-v_*,\sigma)\delta_s(dv) = -(1+|v|^2)B(v-v_*,\sigma)|\delta_s|(dv)$. Together,

$$\begin{aligned} \langle \check{f}_s, \nu_s\rangle &\le 2\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}} 2(3+|v|^2+|v_*|^2+|v_*|^2 - (1+|v|^2))B(v-v_*,\sigma)|\delta_s|(dv)\rho_s(dv_*)d\sigma \\ &= 2\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}} 4(1+|v_*|^2)B(v-v_*,\sigma)|\delta_s|(dv)\rho_s(dv_*)d\sigma \\ &\le 4K\Lambda_{2+\gamma}(\rho_s)\langle 1+|v|^2, |\delta_s|\rangle. \end{aligned} \tag{3.60}$$

Now, we use the properties of f again to write

$$\langle 1+|v|^2, |\delta_t|\rangle = \langle \check{f}, \delta_t\rangle = \int_0^t \langle \check{f}, \nu_s\rangle ds \le 4K\int_0^t \langle 1+|v|^2, |\delta_s|\rangle \Lambda_{2+\gamma}(\rho_s)ds.$$

Since $\Lambda_{2+\gamma}(\rho_t) \in L^1_{\text{loc}}([0,\infty))$, we conclude from Grönwall's lemma that $\delta_t \equiv 0$ and the proof of the Kac case is complete. The Boltzmann case is identical, dropping the terms involving $\mathfrak{M}$, in which case we find the linearised Boltzmann equation in place of (3.58):

$$\xi_t = \mu_0 - \nu_0 + 2\int_0^t Q_K(\xi_s, \rho_s)ds; \qquad \rho_t = \frac{\mu_t^K + \nu_t^K}{2}.$$

□

We finally deal with Proposition 3.11, which essentially follows the proof of [157, Theorem 1.1]. Since we will use very similar proofs later in the chapter (Lemma 3.21 and 3.22), we will only sketch the main points of the argument. Thanks to a stochastic calculus for processes of the form $\int_{(0,t]\times\mathcal{S}_N} \langle f, \mu^N - \mu^{N,K}_{s-}\rangle (m^{N,K} - \overline{m}^{N,K})(ds, d\mu^N)$, see Darling and Norris [49], such processes are martingales if f is fixed, with a quadratic variation on the order N^{-1}, and so are small as a function of N. However, unlike in the finite dimensional cases in [49], the problem remains in taking a supremum over f inside the expectation, which corresponds to the estimates degrading as the dimension increases. Instead, we will use the relative compactness discussed in Remark 3.8 to argue that this is an *effectively finite dimensional* problem. More precisely, we show that it can be approximated by a discretised, finite dimensional martingale approximation problem, with the following trade off: that making the truncation error small requires taking a large (finite) dimensional martingale. As in [49, 157], the martingale term is 'small', as a function of N, but will increase as a function of the dimension of the approximation. By optimising over the discretisation, we will be able to balance the two terms to find a useful estimate on the family of processes. This is the same approach as used for an equivalent problem in [157, Theorem 1.1]. We will use variants of this argument repeatedly in similar problems, and this proof may be read as a warm-up for the rather more delicate proofs of Lemmas 3.21, 3.22. In this proposition, we must also deal with the fact that f_{st} may not be adapted if the environment ρ_t is random, since f_{st} also depends on $(\rho_u)_{u=s}^t$.

Sketch Proof of Proposition 3.11. We follow [157], and start with some non-random bounds on the quantities $y_\beta(t), z_t$ as in Proposition 3.9 for this choice of environment. From the definition of z_t we find the almost sure bound, for all $t \le t_{\text{fin}}$,

$$z_t \le \exp\left(CKw\int_0^t (\beta s^{\beta-1}+1)ds\right) \le \exp\left(CKw(1+t_{\text{fin}})\right)$$

up to a new choice of C in the final bound. For $y_\beta(t)$, we bound, almost surely, for all $0 \le s \le s' \le t_{\text{fin}}$,

$$\begin{aligned}\int_s^{s'} \Lambda_{2+\gamma}(\rho_u)du \le w\int_s^{s'} (\beta u^{\beta-1}+1)du &= w\left((s')^\beta - s^\beta + (s'-s)\right) \\ &\le w\left((s'-s)^\beta + (s'-s)\right) \\ &\le w(1+{t_{\text{fin}}}^{1-\beta})(s'-s)^\beta \\ &\le 2w(1+t_{\text{fin}})(s'-s)^\beta.\end{aligned}$$

It follows that, again changing C if necessary, we have the almost sure bounds

$$y_\beta(t), z_t \leq \exp\left(CKw(1+t_{\text{fin}})\right) \qquad \text{for all } t \leq t_{\text{fin}}. \tag{3.61}$$

We now fix $R \geq 1$ and $r \in (0,1]$ such that $t_{\text{fin}}/r, R/r$ are integers. We now set $A = (-R,R]^d$, and let $\mathfrak{P}$ be a partition of $(0,t_{\text{fin}}] \times A$ into $n = (R/r)^d(t_{\text{fin}}/r)$ translates P of $(0,r] \times (-r,r]^d$, and for any $f \in \mathcal{A}_\gamma$, we write

$$f_{(s\wedge t)t}(v) = \sum_{P\in\mathfrak{P}} a_P(f,t)\mathbb{1}_P(s,v)(1+|v|^2) + \varepsilon(f,t)(s,v) \tag{3.62}$$

where $a_P(f,t)$ is the average value of $f_{(s\wedge t)t}(v)/(1+|v|^2)$ over $(s,v)\in P$, and where this defines the remainder function $\varepsilon(f,t)$, and we write

$$\widetilde{M}_t^{N,K,f} = \sum_{P\in\mathfrak{P}} a_P(f,t)M_t^{N,K;P} + Z_t^{N,K,f} \tag{3.63}$$

where

$$M_t^{N,K;P} := \int_{(0,t]\times\mathcal{S}_N} \langle (1+|v|^2)\mathbb{1}_P(s,v), \mu^N - \mu_{s-}^{N,K}\rangle (m^{N,K} - \overline{m}^{N,K})(ds, d\mu^N);$$

$$Z_t^{N,K,f} = \int_{(0,t]\times\mathcal{S}_N} \langle \varepsilon(f,t), \mu^N - \mu_{s-}^{N,K}\rangle (m^{N,K} - \overline{m}^{N,K})(ds, d\mu^N).$$

This is the key decomposition in the proof. Roughly speaking:

- The processes $M^{N,K;P}$ are martingales, which can be controlled by the general theory of Markov chains, *independently of f.*
- The coefficients a_P depend on f and t, but are bounded, uniformly over $f \in \mathcal{A}_\gamma$.
- On A, $\varepsilon(f,t)$ will be small, uniformly in f,t, due to the continuity estimates of f_{st}. This may be viewed as a *relative compactness* argument, as discussed in Remark 3.8: given $\delta > 0$, one could use this construction to produce a finite δ-net for the restriction of functions $\mathcal{A}_\gamma|_A$ in the norm of $\mathcal{A}_0|_A$.
- $|\epsilon(f,t)|$ is bounded by $z_t(1+|v|^2)$ on $\mathbb{R}^d \setminus A$, and the contribution from this region will be controlled by the moment bounds.

Step 1. Control on the martingale sum We start by considering $M_t^{N,K;P}$. The quadratic variation is

$$\left[M^{N,K;P}\right]_t = \int_{(0,t]\times\mathcal{S}_N} \langle (1+|v|^2)\mathbb{1}_P(s,v), \mu^N - \mu_{s-}^{N,K}\rangle^2 \overline{m}^{N,K}(ds, d\mu^N)$$

and by Doob's L^2 inequality,

$$\mathbb{E}\left[\sup_{t\le t_{\text{fin}}}\sum_{P\in\mathfrak{P}}|M_t^{N,K;P}|^2\right] \le 4\mathbb{E}\left[\sum_{P\in\mathfrak{P}}\int_{(0,t_{\text{fin}}]\times\mathcal{S}_N}\langle(1+|v|^2)1\!\!1_P(s,v),\mu^N-\mu^{N,K}_{s-}\rangle^2\overline{m}^{N,K}(ds,d\mu^N)\right]. \tag{3.64}$$

To bound the integrand, if μ^N is obtained from $\mu^{N,K}_{s-}$ by a collision with incoming velocities v,v_*, we note that $\langle(1+|v|^2)1\!\!1_P(s,v),\mu^N-\mu^{N,K}_{s-}\rangle^2$ is at most

$$\begin{aligned}&N^{-2}((1+|v|^2)1\!\!1_P(s,v)+(1+|v_*|^2)1\!\!1_P(s,v_*)+(1+|v'|^2)1\!\!1_P(s,v')+(1+|v_*'|^2)1\!\!1_P(s,v_*'))^2\\ &\quad\le 8N^{-2}((1+|v|^4)1\!\!1_P(s,v)+(1+|v_*|^4)1\!\!1_P(s,v_*)+(1+|v'|^4)1\!\!1_P(s,v')+(1+|v_*'|^4)1\!\!1_P(s,v_*'))\end{aligned}$$

and when we sum over $P\in\mathfrak{P}$ we obtain

$$\begin{aligned}&\sum_{P\in\mathfrak{P}}\langle(1+|v|^2)1\!\!1_P(s,v),\mu^N-\mu^{N,K}_{s-}\rangle^2\\ &\quad\le 8N^{-2}((1+|v|^4)1\!\!1_A(v)+(1+|v_*|^4)1\!\!1_A(v_*)+(1+|v'|^4)1\!\!1_A(v')+(1+|v_*'|^4)1\!\!1_A(v_*'))\\ &\quad\le CN^{-2}R^{(4+\gamma-p)^+}((1+|v|^{p-\gamma})+(1+|v_*|^{p-\gamma})+(1+|v'|^{p-\gamma})+(1+|v_*'|^{p-\gamma}))\\ &\quad\le CN^{-2}R^{(4+\gamma-p)^+}(1+|v|^{p-\gamma}+|v_*|^{p-\gamma}).\end{aligned}$$

We therefore return to (3.64) to find

$$\begin{aligned}&\mathbb{E}\left[\sup_{t\le t_{\text{fin}}}\sum_{P\in\mathfrak{P}}|M_t^{N,K;P}|^2\right]\\ &\le CN^{-1}R^{(4+\gamma-p)^+}\mathbb{E}\left[\int_0^{t_{\text{fin}}}\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}}(1+|v|^{p-\gamma}+|v_*|^{p-\gamma})B(v-v_*,\sigma)\mu^{N,K}_s(dv)\mu^{N,K}_s(dv_*)d\sigma\right]\\ &\qquad\le CN^{-1}R^{(4+\gamma-p)^+}K\mathbb{E}\left[\int_0^{t_{\text{fin}}}\langle 1+|v|^p,\mu^{N,K}_s\rangle ds\right]\\ &\qquad\le CN^{-1}R^{(4+\gamma-p)^+}Kt_{\text{fin}}a\end{aligned}$$

where we use Proposition 2.10, recalling that $\mathbb{E}[\Lambda_p(\mu_0^{N,K})]\le a$ was the moment hypothesis on the initial data, and with C allowed to depend on p,d but not on K,t_{fin} or N or the discretisation parameters R,r.

Step 2. Control of the coefficients & Martingale Sum We next observe that, thanks to the growth bound for $f_{(s\wedge t)t}$ by Proposition 3.9, we have the bound $a_P(f,t)\le z_t\le z_{t_{\text{fin}}}$, uniformly in $f\in\mathcal{A}_\gamma$ and $t\le t_{\text{fin}}$. It follows that we have the pathwise inequality

$$\sum_{P\in\mathfrak{P}}a_P(f,t)M_t^{N,K;P}\le z_t\sqrt{\#\mathfrak{P}}\sqrt{\sup_{s\le t_{\text{fin}}}\sum|M_t^{N,K;P}|^2}$$

valid simultaneously for all $f \in \mathcal{A}_\gamma, t \le t_{\mathrm{fin}}$, which implies the same bound when we replace the left-hand side by its supremum. Using the previous step to take the expectation, we conclude that

$$\begin{aligned}\mathbb{E}\left[\sup_{f\in\mathcal{A}_\gamma, t\le t_{\mathrm{fin}}} \sum_{P\in\mathfrak{P}} a_P(f,t) M_t^{N,K;P}\right] &\le CN^{-1/2}R^{(4+\gamma-p)^+} aKt_{\mathrm{fin}} z_{t_{\mathrm{fin}}} \sqrt{\#\mathfrak{P}} \\ &\le R^{(4+\gamma-p)^+} N^{-1/2} a \exp\left(CKw(1+t_{\mathrm{fin}})\right) \left(\frac{R^d t_{\mathrm{fin}}}{r^{1+d}}\right)^{1/2}\end{aligned} \tag{3.65}$$

using the bound (3.61) and absorbing Kt_{fin} into the exponent, at the cost of a new constant C.

Step 3. Control of the Remainder Term We next bound $\varepsilon(f,t)$, uniformly in $f \in \mathcal{A}_\gamma$. Using Proposition 3.9 and the bounds on z_t, y_β, we find, for any $P \in \mathfrak{P}$, uniformly in $f \in \mathcal{A}_\gamma, t \le t_{\mathrm{fin}}$,

$$|f_{(s\wedge t)t}(v) - a_P(f,t)(1+|v|^2)| \le \exp\left(CK(1+t_{\mathrm{fin}})\right)\left(r^\beta(1+|v|^{2+\gamma}) + r^\gamma(1+|v|^2)\right)$$

on P, and these together cover A. Meanwhile $|\varepsilon(f,t)(v,s)| = |f_{(s\wedge t)t}(v)| \le z_t(1+|v|^2)$ outside of A, so we have the overall bound

$$\begin{aligned}|\varepsilon(f,t)(s,v)| &\le (r^\beta(1+|v|^{2+\gamma})\mathbb{1}_A + r^\gamma(1+|v|^2)\mathbb{1}_A + (1+|v|^2)\mathbb{1}_{A^c})e^{CKw(1+t_{\mathrm{fin}})} \\ &\le ((r^\beta R^{(2+2\gamma-p)^+} + r^\gamma R^{(2+\gamma-p)^+})(1+|v|^{p-\gamma}) + R^{(2-p)/2}(1+|v|^{(p+2)/2}))e^{CKw(1+t_{\mathrm{fin}})} \\ &\le (r^\beta R^{(2+2\gamma-p)^+} + r^\gamma R^{(2+\gamma-p)^+} + R^{(2-p)/2})(1+|v|^{(p+2)/2})e^{CKw(1+t_{\mathrm{fin}})}.\end{aligned}$$

Step 4. Control over Remainder Integrals From the previous step, it follows that

$$\begin{aligned}\sup_{f\in\mathcal{A}_\gamma, t\le t_{\mathrm{fin}}} |Z_t^{N,K,f}| &\le (r^\beta R^{(2+2\gamma-p)^+} + r^\gamma R^{(2+\gamma-p)^+} + R^{(2-p)/2})e^{CKw(1+t_{\mathrm{fin}})} \\ &\cdots \times \int_{(0,t_{\mathrm{fin}}]\times\mathcal{S}_N} |\langle(1+|v|^{(p+2)/2}, \mu^N - \mu_{s-}^{N,K}\rangle|(m^{N,K} + \overline{m}^{N,K})(ds, d\mu^N).\end{aligned} \tag{3.66}$$

Taking expectations, the contributions from the jump measure and compensator contribute equally, since the difference of the two integrals is again a martingale, and we get

$$\begin{aligned}\mathbb{E}\left[\sup_{f\in\mathcal{A}_\gamma, t\le t_{\mathrm{fin}}} |Z_t^{N,K,f}|\right] \le& e^{CKw(1+t_{\mathrm{fin}})}(r^\beta R^{(2+2\gamma-p)^+} + r^\gamma R^{(2+\gamma-p)^+} + R^{(2-p)/2}) \\ &\cdots \times \mathbb{E}\left[\int_{(0,t_{\mathrm{fin}}]\times\mathcal{S}_N} |\langle 1+|v|^{(p+2)/2}, \mu^N - \mu_s^{N,K}\rangle|\overline{m}^{N,K}(ds, d\mu^N)\right].\end{aligned}$$

When μ^N is obtained from $\mu_s^{N,K}$ by a collision in v, v_*, the integrand is at most $C(1 + |v|^{(p+2)/2} + |v_*|^{(p+2)/2})$, so we integrate to find

$$\begin{aligned}\int_{(0,t_{\text{fin}}]\times\mathcal{S}_N} &|\langle 1+|v|^{(p+2)/2}, \mu^N - \mu_s^{N,K}\rangle| \overline{m}^{N,K}(ds, d\mu^N) \\ &\le C\int_{(0,t_{\text{fin}}]\times\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}} (1+|v|^{(p+2)/2}+|v_*|^{(p+2)/2}) B(v-v_*,\sigma)\mu_s^{N,K}(dv)\mu_s^{N,K}(dv_*)ds \\ &\le CK\int_0^{t_{\text{fin}}} \Lambda_{\frac{p+2}{2}+\gamma}(\mu_s^{N,K})ds.\end{aligned} \tag{3.67}$$

Writing q for the moment index in the final line, we have that $q < p + \gamma$ because $p > 2$, so $\frac{p-q}{\gamma} > -1$. In particular, using the moment creation property in Proposition 2.10, we find an integrable time-dependent factor, and

$$\mathbb{E}\left[\int_0^{t_{\text{fin}}} \Lambda_{(p+2)/2+\gamma}(\mu_s^{N,K})ds\right] \le C(1+t_{\text{fin}})\mathbb{E}\left[\Lambda_p(\mu_0^{N,K})\right] = C(1+t_{\text{fin}})a.$$

All together, we conclude that

$$\begin{aligned}\mathbb{E}\left[\sup_{f\in\mathcal{A}_\gamma, t\le t_{\text{fin}}} |Z_t^{N,K,f}|\right] &\le e^{CKw(1+t_{\text{fin}})}(r^\beta R^{(2+2\gamma-p)^+} + r^\gamma R^{(2+\gamma-p)^+} + R^{(2-p)/2})(1+t_{\text{fin}})a \\ &\le e^{CKw(1+t_{\text{fin}})}(r^\beta R^{(2+2\gamma-p)^+} + r^\gamma R^{(2+\gamma-p)^+} + R^{(2-p)/2})a\end{aligned} \tag{3.68}$$

where again we absorbed the time-growing factor into the exponential, at the cost of a new constant C.

Step 5. Conclusion Gathering (3.65, 3.68) and returning to (3.63), we have shown that

$$\begin{aligned}\mathbb{E}\left[\sup_{t\le t_{\text{fin}}}\sup_{f\in\mathcal{A}_\gamma} |\widetilde{M}_t^{N,K,f}|\right] \le e^{CKw(1+t_{\text{fin}})}a\Bigg(&N^{-1/2}R^{(4+\gamma-p)^+}\left(\frac{R^d t_{\text{fin}}}{r^{1+d}}\right)^{1/2} + r^\beta R^{(2+2\gamma-p)^+} \\ &+ r^\gamma R^{(2+\gamma-p)^+} + R^{(2-p)/2}\Bigg)\end{aligned} \tag{3.69}$$

with C independent of $N, R, K, w, t_{\text{fin}}$. We now choose R, r depending on N; in any case, depending on which of the exponents $(4+\gamma-p)^+, (2+2\gamma-p)^+, (2+\gamma-p)^+$ are positive, one can choose $R = \mathcal{O}(N^{\delta_1}), r = \mathcal{O}(N^{-\delta_2})$, still satisfying the conditions above, with $\delta_1, \delta_2 > 0$ small enough that the term in paranetheses has the overall asymptotic $N^{-\eta}, \eta = \eta(p, d, \beta, \gamma) > 0$, as desired. $\square$

3.2.2 Applications of the Branching Process Representation

We now record, for future use, some immediate consequences of the branching process representation developed above. First, we show how these results imply uniqueness and stability for the Boltzmann Equation (BE) in either case (HS, CHP$_K$).

Corollary 3.13. *Continue in the notation above. Let $(\mu_t^K)_{t\geq 0}, (\nu_t^K)_{t\geq 0} \subset \mathcal{S}$ be (energy-conserving) solutions to (BE) with the moment bound $\Lambda_{2+\varepsilon}(\mu_0^K, \nu_0^K) < \infty$ for some $\varepsilon > 0$. Then we have*

$$\begin{aligned} W_\gamma\left(\mu_t^K, \nu_t^K\right) &\leq \exp\left(CK\int_0^t \Lambda_{2+\gamma}(\mu_s^K, \nu_s^K)ds\right) W_\gamma\left(\mu_0^K, \nu_0^K\right) \\ &\leq \exp\left(C_\varepsilon K(1+t)\Lambda_{2+\varepsilon}(\mu_0^K, \nu_0^K)\right) W_\gamma\left(\mu_0^K, \nu_0^K\right) \end{aligned} \tag{3.70}$$

for some constant $C = C(B,d)$ and $C_\varepsilon = C(B,d,\varepsilon)$. In particular, energy-conserving solutions are unique as soon as the initial data have $2+\varepsilon$ moments, and we write $\phi_t^K : \mathcal{S}^{2+\varepsilon} \to \mathcal{S}^{2+\varepsilon}$ for the corresponding solution maps, i.e. $\phi_t^K(\mu_0^K) = \mu_t^K$.

Moreover, there exists a finite $p > 2$, depending on K, such that, whenever $\mu_0^K, \nu_0^K \in \mathcal{S}^p$, we have the estimate

$$W_\gamma\left(\mu_t^K, \nu_t^K\right) \leq e^{CK(1+t)}\Lambda_p(\mu_0^K, \nu_0^K) W_\gamma\left(\mu_0^K, \nu_0^K\right). \tag{3.71}$$

Proof. For (3.70), we use Proposition 3.10 to write, for any $f \in \mathcal{A}_\gamma$,

$$\langle f, \mu_t^K - \nu_t^K \rangle = \langle f_{0t}[\rho], \mu_0^K - \nu_0^K \rangle$$

where $\rho_t = (\mu_t^K + \nu_t^K)/2$ is as in Proposition 3.10. Since $f_{0t}[\rho] \in z_t\mathcal{A}_\gamma$ by Proposition 3.9, we bound the right-hand side to get

$$\langle f, \mu_t^K - \nu_t^K \rangle \leq z_t W_\gamma\left(\mu_0^K, \nu_0^K\right)$$

and, recalling the definition of W_γ, we optimise over $f \in \mathcal{A}_\gamma$ to find

$$W_\gamma\left(\mu_t^K, \nu_t^K\right) \leq z_t W_\gamma\left(\mu_0^K, \nu_0^K\right).$$

This is exactly the first bound of (3.70), using the definition of z_t, and the second line follows using the moment creation and propagation bounds in Proposition 2.6. The second assertion (3.71) follows immediately from the first bound of (3.70) using Proposition 2.9. □

Our second corollary is a convergence result in the transport costs w_p introduced in Section 2.1 for the Kac process in the cases of cutoff hard potentials (CHP$_K$), analagous to Proposition 1.2. Since this will be of use in Chapter 4, where we must already play with a large (finite) number of moments, we will not try to optimise moment dependence.

Corollary 3.14. *Let $p \geq 0$ and $q > \max(4+3\gamma, p+2)$. Then there exists $C = C(G,q,d), \alpha = \alpha(d,p,q) \geq 0$ such that, whenever $a \geq 1$, $\mu_0 \in \mathcal{S}$ and $\mu^{N,K}_t$ is a K-cutoff Kac with $K \geq 1$ and initial moment estimates*

$$\Lambda_q(\mu_0) \leq a, \qquad \mathbb{P}\left(\Lambda_q(\mu^{N,K}_0) \leq a\right) = 1 \tag{3.72}$$

then we have the convergence estimate, for all $t_{\text{fin}} \geq 0$,

$$\mathbb{E}\left[\sup_{t\leq t_{\text{fin}}} w_p\left(\mu^{N,K}_t, \phi^K_t(\mu_0)\right)\right] \leq \left(N^{-\alpha} + \mathbb{E}\left[w_p\left(\mu^{N,K}_0, \mu_0\right)\right]^{\alpha}\right) e^{CaK(1+t_{\text{fin}})}. \tag{3.73}$$

Sketch Proof of Corollary 3.14. Let us first prove the same result with $\mu_0 = \mu^{N,K}_0$. As in Proposition 3.10, we consider the linearised Kac process in the random environment

$$\rho^N_t = \frac{\mu^{N,K}_t + \phi^K_t(\mu^{N,K}_0)}{2} \tag{3.74}$$

and for $b \geq 1$, consider the stopping times T^N_b defined in Proposition 2.10iv). (equation (2.104)) for the $(2+\gamma)^{\text{th}}$ moment. Let us write $\widetilde{M}^{N,K,f,b}_t$ for the stochastic integrals in (3.39) in the environment

$$\rho^{T^N_b}_t : f = \rho_t \mathbb{I}\left\{t < T^N_b\right\}.$$

We consider the events $\{T^N_b \leq t_{\text{fin}}\}, \{T^N_b > t_{\text{fin}}\}$ separately. On the event $\{T^N_b > t_{\text{fin}}\}$, we have the equalities

$$M^{N,K,f}_t = \widetilde{M}^{N,K,f,b}_t \text{ for all } f \in \mathcal{A}_\gamma \text{ and all } t \leq t_{\text{fin}} \tag{3.75}$$

while on $\{T^N_b \leq t_{\text{fin}}\}$ we have the trivial bound

$$\sup_{t\leq t_{\text{fin}}} W_\gamma(\mu^{N,K}_t, \phi^K_t(\mu)) \leq 4 \tag{3.76}$$

since $W_\gamma \leq 4$ on $\mathcal{S}\times\mathcal{S}$. Combining, we have the bound

$$\sup_{t\leq t_{\text{fin}}} W_\gamma\left(\mu^{N,K}_t, \phi^K_t(\mu^{N,K}_0)\right) \leq \sup_{f\in\mathcal{A}_\gamma, t\leq t_{\text{fin}}} \left\{\widetilde{M}^{N,K,f,b}_t\right\} + 4\cdot\mathbb{I}(T^N_b \leq t_{\text{fin}}). \tag{3.77}$$

Since $q > 2+\gamma$, the moment hypothesis on $\mu^{N,K}_0$ implies $\Lambda_{2+\gamma}(\mu^{N,K}_0) \leq a$ almost surely, which is propagated to $\phi^K_t(\mu_0)$ by Proposition 2.6. The first term is therefore controlled by Proposition 3.11, with $w \leq b + Ca$ for some constant C. We now take $b = Ca$, for some large C, depending only on p; by Proposition 2.10iv), C can be chosen so that $\mathbb{P}(T^N_b \leq t_{\text{fin}}) \leq CaN^{-1}t_{\text{fin}}$. Combining, we obtain

$$\mathbb{E}\left[\sup_{t\leq t_{\text{fin}}} W_\gamma\left(\mu^{N,K}_t, \phi^K_t(\mu^{N,K}_0)\right)\right] \leq CaN^{-\eta}\exp(CaK(1+t_{\text{fin}})) + Cat_{\text{fin}}N^{-1} \tag{3.78}$$

and keeping the worse term

$$\mathbb{E}\left[\sup_{t\leq t_{\text{fin}}} W_\gamma\left(\mu^{N,K}_t, \phi^K_t(\mu^{N,K}_0)\right)\right] \leq CaN^{-\eta}\exp(CaK(1+t_{\text{fin}})). \tag{3.79}$$

We now convert this approximation into w_p. Fix $p' \in (p+2, q)$; thanks to the comparisons (2.30) in Section 2.1, for some $\alpha > 0$,

$$\begin{aligned} &\sup_{t \le t_{\text{fin}}} w_p\left(\mu_t^{N,K}, \phi_t^K(\mu_0^{N,K})\right) \\ &\qquad \le \left(\sup_{t \le t_{\text{fin}}} W_\gamma\left(\mu_t^{N,K}, \phi_t^K(\mu_0^{N,K})\right)\right)^\alpha \left(\sup_{t \le t_{\text{fin}}} \Lambda_{p'}\left(\mu_t^{N,K}, \phi_t(\mu_0^{N,K})\right)\right). \end{aligned} \tag{3.80}$$

We now use Hölder's inequality with indexes $\frac{q}{q-p'}, \frac{q}{p'}$ and control the moment term with Propositions 2.6, 2.10 to find that, for some new $\alpha > 0$,

$$\begin{aligned} &\mathbb{E}\left[\sup_{t \le t_{\text{fin}}} w_p\left(\mu_t^{N,K}, \phi_t^K(\mu)\right)\right] \\ &\qquad \le C\,\mathbb{E}\left[\sup_{t \le t_{\text{fin}}} W_\gamma\left(\mu_t^{N,K}, \phi_t^K(\mu_0^{N,K})\right)\right]^\alpha \mathbb{E}\left[\sup_{t \le t_{\text{fin}}} \Lambda_q(\mu_t^{N,K}, \phi_t^K(\mu_0^{N,K}))\right] \\ &\qquad \le CaN^{-\alpha\eta} \exp(CaK(1+t_{\text{fin}})) \cdot Ca(1+t_{\text{fin}}). \end{aligned} \tag{3.81}$$

Absorbing constants and the moment factors into the exponent, we have shown that, for some $\alpha = \alpha(p, q, d) > 0$,

$$\mathbb{E}\left[\sup_{t \le t_{\text{fin}}} w_p\left(\mu_t^{N,K}, \phi_t^K(\mu_0^{N,K})\right)\right] \le N^{-\alpha} \exp(CaK(1+t_{\text{fin}})). \tag{3.82}$$

The conclusion now follows by comparing $\phi_t^K(\mu_0^{N,K})$ and $\phi_t^K(\mu_0)$ using Corollary 3.13 and converting the stability into w_p in the same way, using the other direction of (2.30). □

3.2.3 Stability Estimates

We now turn to some differentiability and stability results for the nonlinear semigroup $(\phi_t)_{t\geq 0}$, which plays a crucial rôle in the time-dependence of Theorems 3.1, 3.2, as in the work of Mischler and Mouhot [142]. We will also use a regularity result for the collision operator Q, which appears in the proof of Lemma 3.22, and *differentiability and stability* results for the flow maps $(\phi_t)_{t\geq 0}$. As in Remark 3.7, from now on we work exclusively with the case of hard spheres in this subsection, so that $K = 1, \gamma = 1$.

Stability Estimates

The key component to our analysis of the Kac process is the *stability* of the limiting Boltzmann equation - that is, that the limit flow suppresses errors, rather than allowing exponential amplification.

We can now state the precise results as they appear in [142, Lemma 6.6]:

Proposition 3.15. *Let $\mu \in \mathcal{S}^p$ for some $p > 2$. Then, for any $\xi \in Y_p$, there exists a unique solution $(\xi_t)_{t\geq 0} \subset Y_p$ to*

$$\partial_t \xi_t = 2Q(\phi_t(\mu), \xi_t); \qquad \xi_0 = \xi \tag{3.83}$$

with the bound

$$\|\xi_t\|_{\mathrm{TV}+p} \leq \exp\left(C\int_0^t \Lambda_{p+1}(\phi_s(\mu))ds\right)\|\xi\|_{\mathrm{TV}+p}. \tag{3.84}$$

Let now $\eta \in (0,1)$. Then there are absolute constants $C \in (0,\infty)$ and $\lambda_0 > 0$ such that, for p large enough (depending only on η), and all $\mu, \nu \in \mathcal{S}^p$, that, if we take $\xi := \nu - \mu$, we have

$$\|\phi_t(\nu) - \phi_t(\mu)\|_{\mathrm{TV}+2} \leq Ce^{-\lambda_0 t/2}\Lambda_p(\mu,\nu)^{\frac{1}{2}}\|\mu-\nu\|^{\eta}_{\mathrm{TV}}; \tag{3.85}$$

$$\|\xi_t\|_{\mathrm{TV}+2} \leq Ce^{-\lambda_0 t/2}\Lambda_p(\mu,\nu)^{\frac{1}{2}}\|\mu-\nu\|^{\eta}_{\mathrm{TV}}; \tag{3.86}$$

$$\|\phi_t(\nu) - \phi_t(\mu) - \xi_t\|_{\mathrm{TV}+2} \leq Ce^{-\lambda_0 t/2}\Lambda_p(\mu,\nu)^{\frac{1}{2}}\|\mu-\nu\|^{1+\eta}_{\mathrm{TV}}. \tag{3.87}$$

This allows us to define a bounded linear map $\mathcal{D}\phi_t(\mu) : Y_p \to Y_p$ by $\mathcal{D}\phi_t(\mu)[\xi_0] := \xi_t$ given by the first assertion, which we understand as a functional derivative thanks to (3.87). This linear map will play the rôle of a functional derivative for the Boltzmann flow ϕ_t in the calculus developed by [142].

In this case, constant λ_0 is given explicitly by the spectral gap of the linearised Boltzmann operator $2Q(\cdot, \gamma)$ at the equilibrium γ, see [142, Theorems 6.5-6.6]. To obtain estimates with the weighted metric W_1, we will use a version of Proposition 3.15 with the difference $\phi_t(\mu) - \phi_t(\nu)$ measured in stronger norms $\|\cdot\|_{\mathrm{TV}+q}$. The following estimate may be obtained by a simple interpolation between Propositions 2.6, 3.15.

Corollary 3.16. *Let $q \geq 2$, $\eta \in (0,1)$ and $\lambda < \lambda_0$. Then for all p large enough, depending on η, λ and q, there exists a constant C such that*

$$\forall \mu, \nu \in \mathcal{S}^p, \qquad \|\phi_t(\mu) - \phi_t(\nu)\|_{\mathrm{TV}+q} \leq Ce^{-\lambda t/2}\Lambda_p(\mu,\nu)^{\frac{1}{2}}\|\mu - \nu\|_{\mathrm{TV}}^{\eta}. \tag{3.88}$$

We emphasise that the rapid decay is the key property that allows us to obtain good long-time behaviour for our estimates. The pointwise estimate Theorem 3.1 and the initial estimate for pathwise local uniform convergence Lemma 3.23 would hold for estimates

$$\|\phi_t(\nu) - \phi_t(\mu)\|_{\mathrm{TV}+5} \leq F(t)\Lambda_p(\mu,\nu)^{\frac{1}{2}}\|\mu - \nu\|_{\mathrm{TV}}^{\eta}; \tag{3.89}$$

$$\|\phi_t(\nu) - \phi_t(\mu) - \xi_t\|_{\mathrm{TV}+2} \leq G(t)\Lambda_p(\mu,\nu)^{\frac{1}{2}}\|\mu - \nu\|_{\mathrm{TV}}^{1+\eta} \tag{3.90}$$

for functions F, G such that

$$\left(\int_0^\infty F^2 dt\right)^{1/2} < \infty; \qquad \int_0^\infty G dt < \infty. \tag{3.91}$$

The full strength of exponential decay is used to 'bootstrap' to the pathwise local uniform estimate Theorem 3.2, which provides better behaviour in the time horizon t_{fin}, with only a logarithmic loss in the number of particles N. Provided that $F \to 0$ as $t \to \infty$, we could use the same 'bootstrap', but with a potentially much larger loss in N.

We now turn to the proof of Proposition 3.15, which is exactly [142, Lemma 6.4]. We will not reproduce all details, but will sketch the most important points for contextualisation and discussion.

Sketch Proof of Proposition 3.15. We reproduce the arguments of [142, Lemmas 6.3-6.4].

Step 1. Existence & Uniqueness, Boundedness Property To start with, we obtain existence for the linearised Boltzmann equation (3.83) using the framework of the previous subsection; as in Proposition 3.10, we run the branching process Ξ_t in the environment $\rho_t = \phi_t(\mu)$, and from initial data Ξ_0 by sampling particles as a Poisson random measure according to $(\xi_0)_\pm$ on $V^\pm$, where $_\pm$ denotes the Hahn-Jordan decomposition of the signed measure. Setting $\xi_t := \mathbb{E}[\Xi_t]$, we argued in the proof of Proposition 3.10 that ξ_t solves (3.83), and that this solution is unique by the same proof. Moreover, the claimed estimate (3.84) follows from (3.28), recalling that $\rho_t = \phi_t(\mu_0) \in L^1_{\mathrm{loc}}([0,\infty), \mathcal{S}^{p+1})$ by Proposition 2.6. Let us remark that we could also allow $p = 2$ here, at the cost of insisting separately that $\phi_t(\mu) \in L^1_{\mathrm{loc}}([0,\infty), \mathcal{S}^3)$.

Step 2. Short Time Bounds We next move to the main case with $\xi_0 := \nu - \mu$, and seek estimates which behave well only on short times. For the difference $\phi_t(\mu) - \phi_t(\nu)$, we use the representation formula again to write, for any continuous f satisfying $|f| \le 1+|v|^2$

$$\langle f, \phi_t(\nu) - \phi_t(\mu)\rangle = \langle f_{0t}[\rho'], \nu - \mu\rangle$$

for the environment $\rho' = (\phi_t(\mu) + \phi_t(\nu))/2$. The right-hand side is at most $z_t[\rho']\|\nu - \mu\|_{\mathrm{TV}+2}$, and optimising over f produces the $\|\cdot\|_{\mathrm{TV}+2}$ on the left-hand side, so we conclude that

$$\|\phi_t(\nu) - \phi_t(\mu)\|_{\mathrm{TV}+2} \le \exp\left(C\int_0^t \Lambda_3(\phi_s(\mu), \phi_s(\nu))ds\right)\|\nu - \mu\|_{\mathrm{TV}+2} \tag{3.92}$$

and from the argument above

$$\|\xi_t\|_{\mathrm{TV}+2} \le \exp\left(C\int_0^t \Lambda_3(\phi_s(\mu))ds\right)\|\nu - \mu\|_{\mathrm{TV}+2}. \tag{3.93}$$

We next estimate $\|\delta_t\|_{\mathrm{TV}+2}$ where δ_t is the signed measure $\delta_t := \phi_t(\nu) - \phi_t(\mu) - \xi_t$. Recalling the notation $\rho_t = \phi_t(\mu)$, $\rho_t' = (\phi_t(\mu) + \phi_t(\nu))/2$ and repeatedly using bilinearity, we have

$$\begin{aligned}\partial_t\delta_t &= Q(\phi_t(\nu)) - Q(\phi_t(\mu)) - 2Q(\xi_t, \phi_t(\mu))\\ &= 2Q\left(\phi_t(\nu) - \phi_t(\mu), \rho_t'\right) - 2Q(\xi_t, \rho_t') - 2Q(\xi_t, \rho_t - \rho_t')\\ &= 2Q(\delta_t, \rho_t') + Q(\xi_t, \phi_t(\nu) - \phi_t(\mu)).\end{aligned} \tag{3.94}$$

By definition, $\delta_0 = 0$, and using the same argument as for uniqueness in Proposition 3.10, we get

$$\|\delta_t\|_{\mathrm{TV}+2} \le \int_0^t (C\|\delta_s\|_{\mathrm{TV}+2}\|\phi_s(\mu) + \phi_s(\nu)\|_{\mathrm{TV}+3} + \|Q(\xi_s, \phi_s(\nu) - \phi_s(\nu))\|_{\mathrm{TV}+2})ds$$

leading to, by Grönwall's Lemma,

$$\|\delta_t\|_{\mathrm{TV}+2} \le \left(\int_0^t \|Q(\xi_s, \phi_s(\nu) - \phi_s(\nu))\|_{\mathrm{TV}+2}ds\right)\exp\left(C\int_0^t \Lambda_3(\phi_t(\mu), \phi_t(\nu))ds\right). \tag{3.95}$$

In the first term, we bound

$$\|Q(\xi_t, \phi_t(\nu) - \phi_t(\nu))\|_{\mathrm{TV}+2} \le \|\xi_t\|_{\mathrm{TV}+3}\|\phi_t(\nu) - \phi_t(\mu)\|_{\mathrm{TV}+2} + \|\xi_t\|_{\mathrm{TV}+2}\|\phi_t(\nu) - \phi_t(\mu)\|_{\mathrm{TV}+3}.$$

The terms in TV + 2 are already controlled, and it remains to estimate the terms in TV + 3 in $L^1_{\mathrm{loc}}([0,\infty))$. For ξ_t, similar arguments to those of Proposition 2.9, using the same argument as for uniqueness above shows that

$$\langle (1+|v|^2)\log(1+|v|^2), |\xi_t| - |\xi_0|\rangle \le \int_0^t (C\|\xi_s\|_{\mathrm{TV}+2}\langle (1+|v|^3)\log(1+|v|^2), \phi_s(\mu)\rangle - C^{-1}\|\xi_s\|_{\mathrm{TV}+3})ds$$

which implies that

$$\begin{aligned}\int_0^t \|\xi_s\|_{\mathrm{TV}+3}ds \le& C\langle (1+|v|^2)\log(1+|v|^2), |\xi_0|\rangle\\ &+ C\left(\sup_{s\le t}\|\xi_s\|_{\mathrm{TV}+2}\right)\left(\int_0^t \langle (1+|v|^3)\log(1+|v|^2), \phi_s(\mu)\rangle ds\right).\end{aligned}$$

The factor $\|\xi_s\|_{\mathrm{TV}+2}$ is bounded by the estimates above, and the same argument again as in Proposition 2.9 gives

$$\int_0^t \langle (1+|v|^3)\log(1+|v|^2), \phi_s(\mu)\rangle ds \le C(t + \langle (1+|v|^2)\log^2(1+|v|^2), \mu\rangle).$$

Absorbing all factors, we find

$$\begin{aligned}\int_0^t \|\xi_s\|_{\mathrm{TV}+3} ds \le \exp\left(C\int_0^t \Lambda_3(\phi_s(\mu))ds\right)(1+\langle (1+|v|^2)\log^2(1+|v|^2), \mu\rangle) \\ \cdots \times \langle (1+|v|^2)\log(1+|v|^2), |\nu-\mu|\rangle.\end{aligned} \tag{3.96}$$

A similar argument holds for $\phi_t(\nu)-\phi_t(\mu)$, and returning to (3.95), we get

$$\begin{aligned}\|\delta_t\|_{\mathrm{TV}+2} \le \exp\left(C\int_0^t \Lambda_3(\phi_s(\mu),\phi_s(\nu))ds\right)(1+\langle (1+|v|^2)\log^2(1+|v|^2), \mu\rangle) \\ \cdots \times \langle (1+|v|^2)\log(1+|v|^2), |\nu-\mu|\rangle \|\nu-\mu\|_{\mathrm{TV}+2}.\end{aligned} \tag{3.97}$$

We now convert everything to a version of what is in the statement, now with time-growing bounds. First, in (3.92, 3.93, 3.97), we use Proposition 2.9 to replace

$$\exp\left(C\int_0^t \Lambda_3(\phi_s(\mu),\phi_s(\nu))ds\right) \le e^{Ct}\Lambda_{p_1}(\mu,\nu)$$

for some finite p_1. For the weighted total variation norm, we use the interpolation

$$\|\xi_0\|_{\mathrm{TV}+2} \le \|\xi_0\|_{\mathrm{TV}}^{\eta}\langle (1+|v|^2)^{1/1-\eta}, |\xi_0|\rangle^{1-\eta} \le 2\Lambda_{2/(1-\eta)}(\mu,\nu)\|\mu-\nu\|_{\mathrm{TV}}^{\eta}.$$

For (3.97), we use the same argument on the last two factors with $\frac{1+\eta}{2}$, to get

$$\|\nu-\mu\|_{\mathrm{TV}+2} \le 2\Lambda_{4/(1-\eta)}(\mu,\nu)\rangle\|\mu-\nu\|_{\mathrm{TV}}^{\frac{1+\eta}{2}};$$

$$\begin{aligned}\langle (1+|v|^2)\log^2(1+|v|^2), |\nu-\mu|\rangle &\le \left\langle ((1+|v|^2)\log^2(1+|v|^2))^{2/(1-\eta)}, \nu+\mu\right\rangle \|\nu-\mu\|^{\frac{1+\eta}{2}} \\ &\le C\Lambda_{5/(1-\eta)}(\mu,\nu)\|\nu-\mu\|_{\mathrm{TV}}^{\frac{1+\eta}{2}}.\end{aligned}$$

All together, for some choice of p depending on η, we have

$$\|\phi_t(\mu)-\phi_t(\nu)\|_{\mathrm{TV}+2} \le e^{Ct}\Lambda_p(\mu,\nu)\|\mu-\nu\|_{\mathrm{TV}}^{\eta}; \tag{3.98}$$

$$\|\xi_t\|_{\mathrm{TV}+2} \le e^{Ct}\Lambda_p(\mu,\nu)\|\mu-\nu\|_{\mathrm{TV}}^{\eta}; \tag{3.99}$$

$$\|\phi_t(\mu)-\phi_t(\nu)-\xi_t\|_{\mathrm{TV}+2} \le e^{Ct}\Lambda_p(\mu,\nu)\|\mu-\nu\|_{\mathrm{TV}}^{1+\eta}. \tag{3.100}$$

Together, these are the conclusions of [142, Lemma 6.3] in our notation.

Step 2. Long-Time Exponential Stability We now show how the previous estimates (3.98 - 3.100) can be extended to the results in the statement, which decay exponentially as $t \to \infty$. Our starting point is the following exponential convergence from [149, Theorem 1.2]. Recalling that $\gamma \in \mathcal{S}$ is the Gaussian equilibrium, we write $\mathcal{L}_\gamma(\mu) = 2Q(\mu, \gamma)$ for the linearised collision operator around γ. Then, for any $z > 0$ and some universal $\lambda > 0$, we can write an equation similar to (3.28) with exponential decay: if $\xi_0 \in Y_2$ and ξ_t^γ is the unique solution to $\partial_t \xi_t^\gamma = \mathcal{L}_\gamma(\xi_t^\gamma), \xi_0^\gamma = \xi_0$, which is (3.83) with $\rho_t = \gamma$, then we have the bound, for some C depending only on z,

$$\langle e^{z|v|}, |\xi_t^\gamma| \rangle \leq C e^{-\lambda t} \langle e^{z|v|}, |\xi_0| \rangle. \tag{3.101}$$

Moreover, for any $\mu \in \mathcal{S}$ with the exponential moment estimate $\langle e^{z|v|}, \mu \rangle \leq a < \infty$, there exists a constant $C = C(B, d, a)$, which depends on μ only through the upper bound a, such that the nonlinear semigroup ϕ_t satisfies

$$\langle e^{z|v|}, |\phi_t(\mu) - \gamma| \rangle \leq C e^{-\lambda t} \tag{3.102}$$

for the same $\lambda > 0$ as above.

Remark 3.17. *Let us remark that the statement, as written in [149], applies only to $\xi_t^\gamma, \phi_t(\mu)$ admitting a density with respect to the Lebesgue measure, but that this can be relaxed to cover all (signed) measures. For instance, the previous points also establish the stability of the Boltzmann flow and linearised Boltzmann flow in Y_2 with respect to the negative Sobolev-type norm $\|\xi\|_{\mathcal{A}_1^\star} := \sup_{f \in \mathcal{A}_1} |\langle f, \xi \rangle|$. For the Boltzmann case, this stability already follows from Corollary 3.13, and for the linearised Boltzmann case, we write $\langle f, \xi_t \rangle = \langle f_{0t}[\gamma], \xi_0 \rangle$, where f_{0t} are the functions defined by the branching process for the environment $\rho_t = \gamma$ or by an equivalent PDE. In any case, once the claimed results (3.101, 3.102) are established with bounds independent of the density, we can approximate μ, ξ in $(Y_2, \|\cdot\|_{\mathcal{A}_1^\star})$ by measures μ^n (resp. signed measures ξ^n) admitting a density. The result then holds for μ^n, ξ^n, uniformly in n, and the measure-valued result holds on taking $n \to \infty$ to revover $\xi_t^\gamma, \phi_t(\mu)$.*

We will also use the appearance of exponential moments for hard spheres, in a similar vein to Proposition 2.13, see [132, Theorem 1.3d)]: there exist some $z > 0, A < \infty$ such that $\langle e^{z|v|}, \phi_t(\mu) \rangle \leq A$ for all $\mu \in \mathcal{S}$ and all $t \geq \frac{1}{2}$. For the linear term ξ_t, we use the same argument as in (3.59) to keep a negeative term so that, for any p and any $t \geq \frac{1}{2}$,

$$\begin{aligned} \langle |v|^p, |\xi_t| - |\xi_{1/2}| \rangle &\leq \int_{1/2}^t \int_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{S}^{d-1}} (|v'|^p + |v'_*|^p + |v_*|^p - |v|^p) B(v - v_*, \sigma) |\xi_s|(dv) \phi_s(\mu)(dv_*) ds \\ &\leq \int_{1/2}^t \left(\langle |v|^p, Q(|\xi_s|, \phi_s(\mu)) \rangle + C \langle 1 + |v|^2, |\xi_s| \rangle \langle 1 + |v|^p, \phi_s(\mu) \rangle \right) ds. \end{aligned} \tag{3.103}$$

Following the arguments of [132, Theorem 1.3d)], the first term produces a sum of weighted total variation norms of $\|\xi_s\|_{\mathrm{TV}+k}, k \leq p + 1$ with a negative term in the highest order term, and the same for $\phi_s(\mu)$, while the second term can be treated as a source term, using the boundedness of $\|\xi_s\|_{\mathrm{TV}+2}$ and the exponential moment of $\phi_s(\mu)$ on $s \geq \frac{1}{2}$. Following the same arguments as the cited result, one finds that the linearised Boltzmann equation also creates exponential moments away from so that overall, for some $z > 0$ and $A < \infty$,

$$\sup_{t\geq 1} \left\langle e^{2z|v|}, \phi_t(\mu) + \phi_t(\nu) + |\xi_t| \right\rangle \leq A \tag{3.104}$$

and, up to choosing a different A,

$$\langle e^{z|v|}, |\phi_t(\mu) - \gamma| + |\phi_t(\nu) - \gamma|\rangle \leq 2Ae^{-\lambda t}. \tag{3.105}$$

With this starting point, we will argue the assertion (3.85) regarding $\varepsilon_t = \phi_t(\mu) - \phi_t(\nu)$; the other cases are similar. We write

$$\begin{aligned} \partial_t \varepsilon_t &= Q(\phi_t(\nu) - \phi_t(\mu), \phi_t(\mu) + \phi_t(\mu)) \\ &= \mathcal{L}_\gamma \varepsilon_t + Q(\varepsilon_t, \phi_t(\nu) - \gamma) + Q(\varepsilon_t, \phi_t(\mu) - \gamma). \end{aligned} \tag{3.106}$$

For any $t \geq t_0 \geq 1$, we use Duhamel's formula to write

$$\varepsilon_t = e^{(t-t_0)\mathcal{L}_\gamma}\varepsilon_{t_0} + \int_{t_0}^t e^{(t-s)\mathcal{L}_\gamma}(Q(\varepsilon_s, \phi_s(\mu) - \gamma) + Q(\varepsilon_s, \phi_s(\nu) - \gamma))ds$$

and taking the norm gives

$$\begin{aligned} \langle e^{z|v|}, |\varepsilon_t|\rangle \leq Ce^{-\lambda(t-t_0)}&\langle e^{z|v|}, |\varepsilon_{t_0}|\rangle \\ &+ C\int_{t_0}^t e^{-\lambda(t-s)}\langle (1+|v|)e^{z|v|}, |\phi_s(\mu) - \gamma| + |\phi_s(\nu) - \gamma|\rangle \\ &\qquad\qquad \cdots \times \langle (1+|v|)e^{z|v|}, |\varepsilon_s|\rangle ds. \end{aligned}$$

Since we have control over the same thing with the weighting $e^{2z|v|} \geq C(1+|v|)e^{z|v|}$, we get, for $u(t) := \langle e^{z|v|}, |\varepsilon_t|\rangle$,

$$u(t) \leq Ce^{-\lambda(t-t_0)}u(t_0) + Ce^{-\lambda t/2}\int_{t_0}^t e^{-\lambda s/2}\langle (1+|v|)e^{z|v|}, |\varepsilon_s|\rangle ds.$$

The integrand is at most

$$\langle (1+|v|)e^{z|v|}, |\varepsilon_s|\rangle \leq Cu(s)\left(1 + (\ln u(s))_-\right)$$

for some C depending on A, z, and where $_-$ denotes the negative part. Let us fix $\delta \in (0,1)$ and choose $t_0 = \max(1, \frac{4}{\lambda}\log\delta)$ so that $e^{-\lambda t/2} \leq \delta e^{-\lambda t/4}$ for all $t \geq t_0$. With this choice of t_0, for all $t \geq t_0$, we get

$$u(t) \leq Ce^{-\lambda(t-t_0)}u(t_0) + \delta e^{-\lambda t/4}\int_{t_0}^t e^{-\lambda s/2}u(s)\left(1 + (\log u(s))_-\right)ds$$

and this integral inequality implies that, for $t \geq t_0$,

$$u(t) \leq Ce^{-\lambda t/4}u(t_0)^{1-\delta}.$$

Recalling that z does not depend on $\mu, \nu \in \mathcal{S}$, we dominate $\|\phi_t(\nu) - \phi_t(\mu)\|_{\mathrm{TV}+2} \le Cu(t)$, and we combine this with the previous step up to time t_0 to get, globally in time,

$$\|\phi_t(\nu) - \phi_t(\mu)\|_{\mathrm{TV}+2} \le Ce^{-\lambda(t-t_0)/4}\Lambda_p(\mu,\nu)\|\nu - \mu\|_{\mathrm{TV}}^{\eta(1-\delta)}$$

by absorbing a factor $e^{\lambda t_0/4}$ into C, and where C, p, t_0 depending on the exponents δ, η. If we choose a new value of η' and $\delta > 0$ so that the exponent $\eta'(1-\delta)$ gives the target exponent η, these values now depend only on η, and using that $\Lambda_p(\mu,\nu) \le \Lambda_{2p}(\mu,\nu)^{1/2}$ and we conclude (3.85) for the moment index $2p$. The other cases are similar. □

Remark 3.18. *We can now justify the comment in Remark 3.7 regarding the applicability of these estimates and the resulting proofs of Theorems 3.1, 3.2 in the cases (CHP_K) where we obtain cutoff hard potentials by truncation of a noncutoff kernel. In this case, one can follow the argument of [149] to show that the exponential rate λ_K is positive, uniformly in K, but the problem in obtaining the result lies in the short-time bounds (Step 2). In the short-time bounds (3.92, 3.93, 3.97), we would find an exponent of the form $CK\int_0^t \Lambda_{2+\gamma}(\phi_s^K(\mu))ds$ and, since the coefficient is proportional to K, the moment index p one would find by applying Proposition 2.9 also depends on K. Even modifying the technique (for instance, keeping a factor $e^{CK\Lambda_{2+\gamma}(\mu_0)}$ and arguing using the concentration of moments Proposition 2.10iv) when applying to the Kac process, as in Corollary 3.14), we still have a factor of the form e^{CKt_0} which would then appear in our final result, and we would not find a final result for the noncutoff case (NCHP) any stronger than we do in Theorem 4.5 in any case.*

Regularity Estimates

For the proof of the local uniform estimate Lemma 3.22, it will be important to control the continuity of Q *after application of the flow maps* ϕ_t; for brevity, we will write the composition as $Q_t = Q \circ \phi_t$. We can exploit the use of the stronger $\|\cdot\|_{\mathrm{TV}+2}-$ norm in the stability estimates Proposition 3.15, to prove a strong notion of continuity for Q_t, including the dependence on t.

It is well known that, for $q \ge 1$, and $\mu, \nu \in \mathcal{S}^{q+1}$, we have the bilinear estimate

$$\|Q(\mu) - Q(\nu)\|_{\mathrm{TV}+q} \le C\Lambda_{q+1}(\mu,\nu)^{\frac{1}{2}}\|\mu - \nu\|_{\mathrm{TV}+(q+1)} \tag{3.107}$$

and, by interpolating, this leads to

$$\|Q(\mu) - Q(\nu)\|_{\mathrm{TV}+q} \le C\Lambda_{3(q+1)}(\mu,\nu)^{\frac{1}{2}}\|\mu - \nu\|_{\mathrm{TV}}^{\frac{1}{2}}. \tag{3.108}$$

Combining this the stability estimate in Corollary 3.16, we deduce the following. For $q \ge 1$, $\eta \in (0,1)$ and $\lambda < \lambda_0$, then there exists p such that, for $\mu, \nu \in \mathcal{S}^p$, we have the estimate

$$\|Q_t(\mu) - Q_t(\nu)\|_{\mathrm{TV}+q} \le Ce^{-\lambda t}\Lambda_p(\mu,\nu)^{\frac{1}{2}}\|\mu - \nu\|_{\mathrm{TV}}^{\eta}. \tag{3.109}$$

Exchange Lemma

We remark, for future reference, the following technical 'Chapman-Kolmogorov' property for the functional derivatives $\mathcal{D}\phi_t$, which we will use in writing a decomposition of $\mu_t^N - \phi_t(\mu_0^N)$ later (Formula 3.3.1). Since the proof relies on the boundedness of the map $\mathcal{D}\phi_t : Y_p \to Y_p$, which we proved via the connection to the branching process, we include the proof here. This property roughly corresponds to (**A3.** - Convergence of the Generators) in the work [142].

Lemma 3.19 (Exchange Lemma)**.** *Let $\mu^N \in \mathcal{S}_N$ and $f \in \mathcal{A}_1$. Then for all times $t \geq 0$, we have the equalities*

$$\begin{aligned} \frac{d}{dt}\langle f, \phi_t(\mu^N)\rangle &= \langle f, \mathcal{D}\phi_t(\mu^N)\left[Q(\mu^N)\right]\rangle \\ &= \int_{\mathbb{R}^d\times\mathbb{R}^d\times S^{d-1}} \langle f, \mathcal{D}\phi_t(\mu^N)[\mu^{N,v,v_\star,\sigma} - \mu^N]\rangle\, N\, B(v-v_*,\sigma)d\sigma\mu^N(dv)\mu^N(dv_\star) \end{aligned} \tag{3.110}$$

where $\mu^{N,v,v_\star,\sigma}$ is the post-collision measure given by (1.14), $\mathcal{Q}_N$ is the transition kernel of the Kac process, as in Definition 3.2.2, and where $\mathcal{D}\phi_t$ is the functional derivative given by Proposition 3.15.

The first equality is familiar from semigroup theory, but is complicated by the non-linearity of the flow maps; we resolve this by using ideas of the infinite dimensional differential calculus developed in [142], and the second equality follows using the boundedness of $\mathcal{D}\phi_t(\mu^N) : Y_p \to Y_p$ for all p. The first step in the following proof corresponds to [142, Lemma 2.11], and the second to [142, Sections 5.5, 6.5].

Proof. Throughout, fix $\mu^N \in \mathcal{S}_N$ and $f \in \mathcal{A}_1$. Recall, for clarity, the notation $Q_t(\mu) = Q(\phi_t(\mu))$.

Step 1. Differentiation in Time Using the boundedness of appropriate moments of $\mu^N \in \mathcal{S}_N$, together with the continuity estimate (3.108), it is straightforward to see that the map $t \mapsto Q_t(\mu^N)$ is Hölder continuous in time, with respect to the weighted norm $\|\cdot\|_{\mathrm{TV}+2}$: for some constant $C_1 = C_1(N)$, we have the estimate

$$\|Q_t(\mu^N) - Q_s(\mu^N)\|_{\mathrm{TV}+2} \leq C_1|t-s|^{\frac{1}{2}}. \tag{3.111}$$

From the definition (BE) of the Boltzmann dynamics, together with a truncation argument (as in Step 2 of Proposition 2.6), we have that

$$\langle f, \phi_t(\mu_0^N)\rangle = \langle f, \mu^N\rangle + \int_0^t \langle f, Q_s(\mu^N)\rangle ds. \tag{3.112}$$

Therefore, the map $t \mapsto \langle f, \phi_t(\mu^N)\rangle$ is continuously differentiable in time, with derivative

$$\frac{d}{dt}\langle f, \phi_t(\mu^N)\rangle = \langle f, Q_t(\mu^N)\rangle \tag{3.113}$$

where, at $t = 0$, this is a one-sided, right derivative. It therefore suffices to show that the first equality of (3.110) holds as a *right* derivative.

Fix $t \geq 0$, and observe that, for $s > 0$ small enough, $\nu_s^N = \mu^N + sQ(\mu^N)$ defines a measure $\nu_s^N \in \mathcal{S}$. From the semigroup property, it follows that $\phi_t(\phi_s(\mu^N)) = \phi_{t+s}(\mu^N)$, and we can therefore expand

$$\begin{aligned}&\left\langle f, \phi_{t+s}(\mu^N) - \phi_t(\mu^N) - s\mathcal{D}\phi_t(\mu^N)[Q(\mu^N)]\right\rangle \\ &= \underbrace{\left\langle f, \phi_t(\phi_s(\mu^N)) - \phi_t(\nu_s^N)\right\rangle}_{:=\mathcal{T}_1(s)} + \underbrace{\left\langle f, \phi_t(\nu_s^N) - \phi_t(\mu^N) - s\mathcal{D}\phi_t(\mu)[Q(\mu^N)]\right\rangle}_{:=\mathcal{T}_2(s)}.\end{aligned} \tag{3.114}$$

We will now show that each of the two terms $\mathcal{T}_1, \mathcal{T}_2$ are of the order $o(s)$, which implies the first equality.

Step 1a. Estimate on $\mathcal{T}_1(s)$ Let $\eta \in (\frac{2}{3}, 1)$, and choose p large enough that the stability estimates (3.85, 3.87) hold with exponent η. As $s \downarrow 0$, the probability measures $\nu_s^N = \mu^N + sQ(\mu^N)$ and $\phi_s(\mu^N)$ are bounded in $\mathcal{S}^p$. Therefore, from (3.85), there exists a constant $C_2 = C_2(N) < \infty$ such that, for all $s > 0$ small enough,

$$\|\phi_t(\phi_s(\mu)) - \phi_t(\nu_s)\|_{\mathrm{TV}+2} \leq C_2\|\phi_s(\mu) - \nu_s\|_{\mathrm{TV}+2}^{\eta}. \tag{3.115}$$

The left-hand side is a bound for $\mathcal{T}_1(s)$. Using the estimate (3.111) above, we estimate the right-hand side, following [142, Lemma 2.11]:

$$\begin{aligned}\|\phi_s(\mu^N) - \nu_s^N\|_{\mathrm{TV}+2} &= \left\|\int_0^s (Q_u(\mu^N) - Q_0(\mu^N))du\right\|_{\mathrm{TV}+2} \\ &\leq \int_0^s \|Q_u(\mu^N) - Q_0(\mu^N)\|_{\mathrm{TV}+2}\, du \\ &\leq C_1(N)\int_0^s u^{\frac{1}{2}}du = \frac{2}{3}C_1(N)s^{\frac{3}{2}}.\end{aligned} \tag{3.116}$$

Combining the estimates (3.115, 3.116), we see that

$$\mathcal{T}_1(s) \leq C_2\left(\frac{2}{3}C_1\right)^{\eta} s^{\frac{3\eta}{2}}. \tag{3.117}$$

Since we chose $\eta > \frac{2}{3}$, this shows that $\mathcal{T}_1$ is $o(s)$ as $s \downarrow 0$.

Step 1b. Estimate on $\mathcal{T}_2$ Let η and p be as above, and recall that in (3.87), ξ_t is the definition of $\mathcal{D}\phi_t(\mu)[\nu - \mu]$. We now apply this estimate to μ^N and ν_s^N, noting that

$\nu_s^N = \mu^N + sQ(\mu^N)$ and $\phi_s(\mu^N)$ are bounded in $\mathcal{S}^p$ as $s \downarrow 0$, and that $\nu_s^N - \mu^N = sQ(\mu^N)$. The bound (3.87) now shows that, for some constants $C_3, C_4 < \infty$,

$$\begin{aligned} \|\phi_t(\nu_s^N) - \phi_t(\mu^N) - s\mathcal{D}\phi_t(\mu^N)[Q(\mu^N)]\|_{\mathrm{TV}+2} &\le C_3\|\nu_s^N - \mu^N\|_{\mathrm{TV}}^{1+\eta} \\ &= C_3\|sQ(\mu^N)\|_{\mathrm{TV}}^{1+\eta} \\ &\le C_4 s^{1+\eta}. \end{aligned} \tag{3.118}$$

The left-hand side is a bound for $\mathcal{T}_2$, which implies that $\mathcal{T}_2$ is $o(s)$, as desired. Together with the previous estimate on $\mathcal{T}_1$, this concludes the proof of the first equality.

Step 2. Exchanging Integration and the Linear Map We now turn to the proof of the second equality in (3.110). Since the fourth moment of μ^N is finite, we have the equality as a Bochner integral in Y_3

$$Q(\mu^N) = \int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}} N(\mu^{N,v,v_*,\sigma} - \mu^N)B(v-v_*,\sigma)d\sigma\mu^N(dv)\mu^N(dv_*). \tag{3.119}$$

Now, from (3.84) and using that the third moments of $\phi_t(\mu^N)$ are bounded by Proposition 2.6, the map $\mathcal{D}\phi_t(\mu^N) : Y_3 \to Y_3$ is a continuous linear map, and using the quadratic growth of f, so is the composition $\xi \mapsto \langle f, \mathcal{D}\phi_t(\mu^N)[\xi]\rangle$. Since (3.119) is a Bochner integral in Y_3, we apply this bounded linear map to both sides to obtain

$$\begin{aligned} &\langle f, \mathcal{D}\phi_t(\mu^N)\left[Q(\mu^N)\right]\rangle \\ &= \int_{\mathbb{R}^d\times\mathbb{R}^d\times S^{d-1}} \langle f, \mathcal{D}\phi_t(\mu^N)[\mu^{N,v,v_\star,\sigma} - \mu^N]\rangle\, N\, B(v-v_*,\sigma)d\sigma\mu^N(dv)\mu^N(dv_\star) \end{aligned} \tag{3.120}$$

which is the second equality in (3.110) as desired. The Lemma is complete. ☐

3.2.4 Proof of Theorem 3.6

As a first application of the stability estimates, we will now prove Theorem 3.6, which extends Corollary 3.13 to a gloabl-in-time continuity result for the Boltzmann flow $(\phi_t)_{t\geq 0}$ with respect to our weighted Wasserstein metric W_1.

Proof of Theorem 3.6. We already established the first claim of the theorem in Corollary 3.13, which we write as, for some p_0, C, w, and all $\mu_0, \nu_0 \in \mathcal{S}^{p_0}$ with a moment estimate $\Lambda_{p_0}(\mu_0, \nu_0) \leq a$, we have

$$W_1\left(\phi_t(\mu), \phi_t(\nu)\right) \leq Ce^{wt} a W_1(\mu, \nu). \tag{3.121}$$

We show the second statement, with uniform-in-time Hölder controls, splitting into the cases first where p is sufficiently large, and then dealing separately with any $p > 2$.

Step 1. Uniform-in-Time control with large p. We first deal with the case where $p_1 \geq p_0$ is large enough that the above holds, and such that the stability estimate Proposition 3.15 holds with Hölder exponent $\eta = \frac{1}{2}$. Fix $\mu, \nu \in \mathcal{S}^{p_1}$ with $\Lambda_{p_1}(\mu, \nu) \leq a$, and assume without loss of generality that $0 < W_1(\mu, \nu) < 1$. From the stability estimate (3.85) we have

$$\|\phi_t(\mu) - \phi_t(\nu)\|_{\mathrm{TV}+2} \leq Ca^{\frac{1}{2}} e^{-\lambda_0 t/2} \tag{3.122}$$

for some constants $C, \lambda_0 > 0$. It is immediate from the definitions that $W_1(\mu, \nu) \leq \|\mu - \nu\|_{\mathrm{TV}+2}$ and so combining with (3.121), we have

$$W_1\left(\phi_t(\mu), \phi_t(\nu)\right) \leq Ca \min\left(e^{-\lambda_0 t/2}, W_1(\mu, \nu)e^{wt}\right). \tag{3.123}$$

The right hand side is maximised when $e^{-\lambda_0 t/2} = W_1(\mu, \nu)e^{wt}$, which occurs when

$$t = -\frac{2}{\lambda_0 + 2w} \log W_1(\mu, \nu). \tag{3.124}$$

Therefore, the maximum value of the right-hand side is

$$\begin{aligned} \sup_{t\geq 0} W_1\left(\phi_t(\mu), \phi_t(\nu)\right) &\leq Ca \exp\left(\frac{\lambda_0}{\lambda_0 + 2w} \log W_1(\mu, \nu)\right) \\ &= CaW_1(\mu, \nu)^{\zeta_1} \end{aligned} \tag{3.125}$$

with

$$\zeta_1 = \frac{\lambda_0}{\lambda_0 + 2w} \tag{3.126}$$

which is the claimed Hölder continuity, for $p = p_1$. Of course, this also holds for any $p \geq p_1$ with the same hypothesis that $\Lambda_p(\mu, \nu) \leq a$.

Step 2. Hölder-Continuity with $p > 2$ Moments. Finally, we deal with the second point for arbitrary $p > 2$. This argument uses a localisation principle to control the moments on a very short initial interval $[0, u]$, and may be read as a warm-up to the more involved arguments in the proof of Theorem 3.3.

As above, let p_1 be the moment index required for the previous step to hold, and let ζ_1 be the resulting exponent. Without loss of generality, let us assume that $p \in (2, 3)$, which is the hardest case; if instead we have an estimate on the $(p')^{\text{th}}$ moments, $p' \geq 3$, then this implies the same estimate holds for the $p^{\text{th}} = (\frac{5}{2})^{\text{th}}$ moments. Let $\beta = \frac{p-2}{2} \in (0, 1)$, let $\mu, \nu \in \mathcal{S}$ have moments $\Lambda_p(\mu, \nu) \leq a$, and let $u \in (0, 1]$ to be chosen later. Set $\rho_t = (\phi_t(\mu) + \phi_t(\nu))/2$ as in Proposition 3.10, and let z_t be the resulting coefficients from Proposition 3.9. Define

$$T = \inf \left\{ t \geq 0 : \Lambda_3(\rho_t) \geq \frac{\beta t^{\beta-1} + 1}{2} \right\}. \tag{3.127}$$

We now deal with the two cases $T > u, T \leq u$ separately.

Case 2i. $T > u$. If $T > u$, then we have the estimate

$$z_u = \exp\left(C \int_0^u \Lambda_3(\rho_s) ds \right) \leq \exp\left(C \int_0^1 \frac{\beta s^{\beta-1} + 1}{2} ds \right) \leq C \tag{3.128}$$

for some new absolute constant C. Using the representation formula Proposition 3.10 as in Corollary 3.13, we therefore obtain

$$\sup_{t \leq u} W_1(\phi_t(\mu), \phi_t(\nu)) \leq C W_1(\mu, \nu). \tag{3.129}$$

Using the previous step on $\phi_u(\mu), \phi_u(\nu)$, and using the moment production property recalled in Proposition 2.6, we have the estimate

$$\sup_{t \geq u} W_1(\phi_t(\mu), \phi_t(\nu)) \leq C \Lambda_{p_1}(\phi_u(\mu), \phi_u(\nu)) W_1(\phi_u(\mu), \phi_u(\nu))^{\zeta_1} \leq C u^{2-p_1} W_1(\mu, \nu)^{\zeta_1}. \tag{3.130}$$

Case 2ii. $T \leq u$. We next deal with the case $T \leq u$. In this case, we first note that the moment production property shows that $\Lambda_3(\phi_t(\mu)) \leq Cat^{p-3}$ for small t, and since $\beta - 1 < p - 3$ by the choice of $\beta = \frac{p-2}{2}$, it follows that $T > 0$, while we assume that $T \leq u \leq 1$. Using the moment production property again, $\Lambda_3(\phi_T(\mu)) \leq CaT^{p-3}$, absorbing the constant term into the (negative) power of $T \in (0, 1]$, and similarly for $\phi_T(\nu)$. Comparing to the definition of T shows that

$$\beta T^{\beta-1} \leq \Lambda_3(\phi_T(\mu)) + \Lambda_3(\phi_T(\nu)) \leq CaT^{p-3}; \qquad T \leq u \tag{3.131}$$

for some new C, and recalling the definition of β, this rearranges to produce the bound $1 \leq Cau^{p/2-1}$. In particular, in this case, we have

$$\sup_{t \geq 0} W_1(\phi_t(\mu), \phi_t(\nu)) \leq 4 \leq Cau^{p/2-1}. \tag{3.132}$$

Combining estimates (3.129, 3.130, 3.132) and keeping the worst terms, we see that in either case,

$$\sup_{t\geq 0} W_1\left(\phi_t(\mu), \phi_t(\nu)\right) \leq Ca\left(u^{2-p_1} W_1\left(\mu,\nu\right)^{\zeta_1} + u^{p/2-1}\right). \tag{3.133}$$

We now optimise by taking $u = \min(1, W_1(\mu,\nu)^{\delta})$ for $\delta = \frac{2\zeta_1}{p+2p_1-6}$ to obtain

$$\sup_{t\geq 0} W_1\left(\phi_t(\mu), \phi_t(\nu)\right) \leq CaW_1(\mu,\nu)^{\zeta} \tag{3.134}$$

for a new exponent $\zeta = \zeta(d,p) = \frac{\zeta_1(p-2)}{(p-2)+2(p_1-2)} > 0$. □

3.3 The Interpolation Decomposition

The goal of this section is to prove the following 'interpolation decomposition' for the difference between Kac's process and the Boltzmann flow, which is the key identity required for the proofs of Theorems 3.1, 3.2. This is based on an idea of Norris [158], which was inspired by [142, Section 3.3]. We recall the definitions of the jump measure m^N, compensator $\overline{m}^N$ and transition kernel $\mathcal{Q}_N$ from Definition 3.2.2.

Formula 3.3.1. *Let μ^N_t be a Kac process on $N \geq 2$ particles, and fix a test function $f \in \mathcal{A}_0$. To ease notation, we write*

$$\Delta(s,t,\mu^N) = \phi_{t-s}(\mu^N) - \phi_{t-s}(\mu^N_{s-}); \quad 0 \leq s \leq t, \;\; \mu^N \in \mathcal{S}_N; \tag{3.135}$$

$$\psi(u,\mu,\nu) = \phi_u(\nu) - \phi_u(\mu) - \mathcal{D}\phi_u(\mu)[\nu-\mu]; \quad u \geq 0, \quad \mu,\nu \in \bigcap_{k>2} \mathcal{S}^k \tag{3.136}$$

where $\mathcal{D}\phi_t$ is the derivative of the Boltzmann flow ϕ_t, defined in Proposition 3.15; this makes sense, provided that all moments of μ,ν are finite. Then we can decompose

$$\langle f, \mu^N_t - \phi_t(\mu^N_0)\rangle = M^{N,f}_t + \int_0^t \langle f, \rho^N(t-s,\mu^N_s)\rangle ds \tag{3.137}$$

where

$$M^{N,f}_t = \int_{(0,t]\times\mathcal{S}_N} \langle f, \Delta(s,t,\mu^N)\rangle (m^N - \overline{m}^N)(ds, d\mu^N_s) \tag{3.138}$$

and where ρ^N is given in terms of the transition kernel $\mathcal{Q}_N$ (3.22) by

$$\langle f, \rho^N(u,\mu^N)\rangle = \int_{\mathcal{S}_N} \langle f, \psi(u,\mu^N,\nu)\rangle \mathcal{Q}_N(\mu^N, d\nu). \tag{3.139}$$

Remark 3.20. *i). This is the key identity needed for Theorems 3.1, 3.2; the remainder of the proofs are to establish suitable controls over each of the two terms.*

ii). This representation formula offers two major advantages over the equivalent representation formula Proposition 3.10 from [157, Proposition 4.2].

- *Firstly, all the quantities appearing in our formula are adapted to the natural filtration $\mathfrak{F}^N_t$ of $(\mu^N_t)_{t\geq 0}$, and so we can use martingale estimates directly. By contrast, Proposition 3.10 contains anticipating terms, as remarked above. This makes it much easier to obtain estimates in $L^p(\mathbb{P}), p \geq 1$, while we had to use different techniques in the case with anticipating terms (for instance, using concentration of moments in Lemma 3.14).*
- *Secondly, all terms appearing in our formula may be controlled by the stability estimates (3.85, 3.87). This allows us to exploit the stability of the limit equation, at the level of* individual realisations of *the empirical particle system μ^N_0.*

Proof of Formula 3.3.1. To begin with, we restrict to bounded, measurable f. Fix $t \geq 0$, and consider the process $\Gamma^{N,f,t}_s = \langle f, \phi_{t-s}(\mu^N_s)\rangle$, for $0 \leq s \leq t$. Then $\Gamma^{N,f,t}$ is càdlàg, and is differentiable on intervals where μ^N_s is constant. On such intervals, Lemma 3.19 tells us that

$$\begin{aligned}\frac{d}{ds}\langle f, \phi_{t-s}(\mu^N_s)\rangle &= -\left.\frac{d}{du}\right|_{u=t-s}\langle f, \phi_u(\mu^N_s)\rangle \\ &= -\int_{\mathbb{R}^d\times\mathbb{R}^d\times S^{d-1}}\langle f, \mathcal{D}\phi_{t-s}(\mu^N_s)[\mu^{N,v,v_\star,\sigma}-\mu^N_s]\rangle B(v-v_*,\sigma)\, N\,\mu^N_s(dv)\mu^N_s(dv_\star)d\sigma \\ &= -\int_{\mathcal{S}_N}\langle f, \mathcal{D}\phi_{t-s}(\mu^N_s)[\mu^N-\mu^N_s]\rangle \mathcal{Q}_N(\mu^N_s, d\mu^N)\end{aligned} \tag{3.140}$$

where the final equality is to rewrite integral in terms of the transition kernel $\mathcal{Q}_N$ of the Kac process, defined in (3.22). Writing $\mathcal{I}_t$ for the (finite) set of jumps $\mathcal{I}_t = \{s \leq t : \mu^N_s \neq \mu^N_{s-}\}$, the contribution to $\Gamma^{N,f,t}_t - \Gamma^{N,f,t}_0$ from drift between jumps is

$$\begin{aligned}&\int_{(0,t]\setminus\mathcal{I}_t}\frac{d}{ds}\langle \phi_{t-s}(\mu^N_s)\rangle\, ds \\ &= -\int_{((0,t]\setminus\mathcal{I}_t)\times\mathcal{S}_N}\langle f, \mathcal{D}\phi_{t-s}(\mu^N_s)[\mu^N-\mu^N_s]\rangle \mathcal{Q}_N(\mu^N_s, d\mu^N)ds.\end{aligned} \tag{3.141}$$

Using the definitions (3.135, 3.136) of ψ and Δ, the integrand can be expressed as

$$\langle f, \mathcal{D}\phi_{t-s}(\mu^N_s)[\mu^N-\mu^N_s]\rangle = \langle f, \Delta(s,t,\mu^N) - \psi(t-s,\mu^N_s,\mu^N)\rangle \tag{3.142}$$

for any $s \notin \mathcal{I}_t$. Since the set $\mathcal{I}_t$ has 0 Lebesgue measure, the set $\mathcal{I}_t\times\mathcal{S}_N$ has 0 measure with respect to $\mathcal{Q}_N(\mu^N_s, d\mu^N)ds$, and so the inclusion of this set does not change the integral. Using the definitions (3.23, 3.139) of $\overline{m}^N$ and ρ^N, we can rewrite the integral as

$$\begin{aligned}&\int_{(0,t]\times\mathcal{S}_N}\langle f, \psi(t-s,\mu^N_s,\mu^N)-\Delta(s,t,\mu^N)\rangle \mathcal{Q}_N(\mu^N_s, d\mu^N)ds \\ &= \int_0^t \langle f, \rho^N(t-s,\mu^N_s)\rangle ds - \int_{(0,t]\times\mathcal{S}_N}\langle f, \Delta(s,t,\mu^N)\rangle \overline{m}^N(ds, d\mu^N).\end{aligned} \tag{3.143}$$

On the other hand, at the times when μ^N_s jumps, we have

$$\Gamma^{N,f,t}_s - \Gamma^{N,f,t}_{s-} = \langle f, \phi_{t-s}(\mu^N_s) - \phi_{t-s}(\mu^N_{s-})\rangle = \langle f, \Delta(s,t,\mu^N_s)\rangle. \tag{3.144}$$

Therefore, the contribution to $\Gamma^{N,f,t}_t - \Gamma^{N,f,t}_0$ from jumps is

$$\begin{aligned}\sum_{s\in\mathcal{I}_t}\Gamma^{N,f,t}_s - \Gamma^{N,f,t}_{s-} &= \int_{(0,t]\times\mathcal{S}_N}\langle f, \Delta(s,t,\mu^N)\rangle\, m^N(ds, d\mu^N) \\ &= M^{N,f}_t + \int_{(0,t]\times\mathcal{S}_N}\langle f, \Delta(s,t,\mu^N)\rangle\, \overline{m}^N(ds, d\mu^N)\end{aligned} \tag{3.145}$$

Combining the contributions (3.143, 3.145), we see that

$$\begin{aligned}\langle f, \mu_t^N - \phi_t(\mu_0^N)\rangle &= \Gamma_t^{N,f,t} - \Gamma_0^{N,f,t} \\ &= \int_{(0,t]\setminus\mathcal{I}_t} \frac{d}{ds}\langle f, \phi_{t-s}(\mu_s^N)\rangle ds + \sum_{s\in\mathcal{I}_t} \Gamma_s^{N,f,t} - \Gamma_{s-}^{N,f,t} \\ &= M_t^{N,f} + \int_0^t \langle f, \rho^N(t-s, \mu_s^N)ds\end{aligned} \tag{3.146}$$

as desired. □

3.4 Proof of Theorem 3.1

We now give the proof of Theorem 3.1 from the representation formula. The main difficulty in obtaining a pathwise statement is taking the supremum of the martingale terms $M^{N,f}_t$ in Formula 3.3.1 inside the expectation, similar to Proposition 3.11, where now

$$M^{N,f}_t = \int_{(0,t]\times\mathcal{S}_N} \Big\langle f, \phi_{t-s}(\mu^N) - \phi_{t-s}(\mu^N_{s-}) \Big\rangle (m^N - \overline{m}^N)(ds, d\mu^N). \tag{3.147}$$

We now take the supremum, inside the expectation, over all those functions $f \in \mathcal{A}_1$, i.e.

$$\forall\, v, v' \in \mathbb{R}^d, \quad |\hat{f}(v)| \leq 1; \qquad |\hat{f}(v) - \hat{f}(v')| \leq |v - v'|. \tag{3.148}$$

We follow the same sort of overall strategy as for the same problem in Proposition 3.11. Finding the best exponents of N we have been able to obtain uses a 'hierarchical decomposition'. This approach was inspired by an equivalent technique used in [157, Proposition 7.1].

Lemma 3.21. *Let $\epsilon > 0$, $a \geq 1$ and $0 < \lambda < \lambda_0$. Let p be large enough that Corollary 3.16 holds with* TV + 4, *exponent λ and Hölder exponent $1 - \epsilon$.*
Let $(\mu^N_t)_{t\geq 0}$ be a Kac process in dimension $d \geq 3$, with initial moment $\Lambda_p(\mu^N_0) \leq a$. Let $M^{N,f}_t$ be the processes given by (3.138). Then we have, uniformly in $t \geq 0$,

$$\left\| \sup_{f\in\mathcal{A}_1} \left| M^{N,f}_t \right| \right\|_{L^2(\mathbb{P})} \leq C a^{1/2}\, N^{\epsilon - 1/d} \tag{3.149}$$

for some C depending on ε, λ, d.

Once we have obtained the control of the martingale term, the remaining proof of Theorem 3.1 is straightforward.

Proof of Theorem 3.1. Take $p = p(\epsilon)$ as in Lemma 3.21, and such that Proposition 3.15 holds with exponent $\max(1 - \epsilon, \frac{1}{2})$.

We first note that it is sufficient to prove the case $\mu_0 = \mu^N_0$. Given this case, we use the continuity established in Theorem 3.6 to estimate the difference

$$W_1\left(\phi_t(\mu^N_0), \phi_t(\mu_0)\right) \leq C a^{1/2} W_1(\mu^N_0, \mu_0)^\zeta \tag{3.150}$$

for some $\zeta = \zeta(d, p)$, which implies the claimed result.

From now on, we assume that $\mu_0 = \mu^N_0$. From the interpolation decomposition Formula 3.3.1, we majorise

$$W_1\left(\mu^N_t, \phi_t\left(\mu^N_0\right)\right) \leq \sup_{f\in\mathcal{A}_1} \left| M^{N,f}_t \right| + \int_0^t \sup_{f\in\mathcal{A}_1} \langle f, \rho^N(t-s, \mu^N_s)\rangle\, ds \tag{3.151}$$

where, as in (3.136, 3.139), the integrand is given by

$$\langle f, \rho^N(t-s,\mu^N_s)\rangle = \int_{\mathcal{S}_N} \langle f, \psi(t-s,\mu^N_s,\nu)\rangle \mathcal{Q}_N(\mu^N, d\nu); \tag{3.152}$$

$$\psi(u,\mu,\nu) = \phi_u(\nu) - \phi_u(\mu) - \mathcal{D}\phi_u(\mu)[\nu-\mu] \tag{3.153}$$

and $\mathcal{Q}_N$ is the transition kernel (3.22) of the Kac process.

The first term of (3.151) is controlled in $L^2(\mathbb{P})$ by Lemma 3.21, and so it remains to bound the second term in $L^2(\mathbb{P})$. Let $s \geq 0$, and let μ^N be a measure obtained from μ^N_s by a collision, as in (1.14). Then, using the estimate (3.87), we bound

$$\begin{aligned} \|\psi(t-s,\mu^N_s,\mu^N)\|_{\mathrm{TV}+2} &= \|\phi_{t-s}(\mu^N) - \phi_{t-s}(\mu^N_s) - \mathcal{D}\phi_{t-s}(\mu^N_s)\|_{\mathrm{TV}+2} \\ &\leq Ce^{-\lambda_0(t-s)/2}\|\mu^N - \mu^N_2\|^{2-\epsilon}_{\mathrm{TV}}\, \Lambda_p(\mu^N,\mu^N_s)^{\frac{1}{2}}. \end{aligned} \tag{3.154}$$

By Proposition 2.10iii), we know that $\Lambda_p(\mu^N) \leq C\Lambda_p(\mu^N_s)$. Moreover, from the form (1.14) of possible μ^N, we know that

$$\|\mu^N - \mu^N_s\|_{\mathrm{TV}} \leq \frac{4}{N} \quad \text{for } \mathcal{Q}_N(\mu^N_s,\cdot)\text{-almost all } \mu^N. \tag{3.155}$$

Therefore, almost surely, for all s and $\mathcal{Q}_N(\mu^N_s,\cdot)$-almost all μ^N, we have the bound

$$\|\psi(t-s,\mu^N_s,\mu^N)\|_{\mathrm{TV}+2} \leq Ce^{-\lambda_0(t-s)/2}N^{\epsilon-2}\, \Lambda_p(\mu^N_s)^{\frac{1}{2}} \tag{3.156}$$

where the implied constants are independent of s, μ^N_s. Integrating with respect to $\mathcal{Q}_N(\mu^N_s, d\mu^N)$, we obtain an upper bound for $\langle f, \rho^N(t-s,\mu^N_s)\rangle$:

$$\begin{aligned} \sup_{f\in\mathcal{A}_1} \langle f, \rho^N(t-s,\mu^N_s)\rangle &\leq \int_{\mathcal{S}_N} \left\|\psi(t-s,\mu^N_s,\mu^N)\right\|_{\mathrm{TV}+2}\, \mathcal{Q}_N(\mu^N_s, d\mu^N) \\ &\leq Ce^{-\lambda_0(t-s)/2}\, N^{\epsilon-1}\, \Lambda_p(\mu^N_s)^{\frac{1}{2}}. \end{aligned} \tag{3.157}$$

We now take the $L^2(\mathbb{P})$ norm of the second term in (3.151). Using Proposition 2.10 to control the moment appearing in the integral, we obtain

$$\begin{aligned} \left\| \int_0^t \sup_{f\in\mathcal{A}_1} \langle f, \rho^N(t-s,\mu^N_s)\rangle\, ds \right\|_{L^2(\mathbb{P})} &\leq \int_0^t \left\| \sup_{f\in\mathcal{A}_1} \langle f, \rho^N(t-s,\mu^N_s)\rangle \right\|_{L^2(\mathbb{P})} ds \\ &\leq C\int_0^t e^{-\lambda(t-s)/2}\, N^{\epsilon-1}\, \left\|\Lambda_p(\mu^N_s)^{\frac{1}{2}}\right\|_{L^2(\mathbb{P})} ds \\ &\leq CN^{\epsilon-1}\, a^{1/2}. \end{aligned} \tag{3.158}$$

Noting that the exponent $\epsilon - 1 < \epsilon - \frac{1}{d}$, we combine this with Lemma 3.21, and keep the worse asymptotics. □

Proof of Lemma 3.21. We begin by reviewing the following estimates for 1−Lipschitz functions from [157]. Following [157], we use angle brackets $\langle f\rangle_P$ to denote the average of a bounded function f over a Borel set P of finite, nonzero measure.
Let f be 1− Lipschitz, and consider $P=[0,2^{-j}]^d$. Then, for some numerical constant C_d, we have

$$\forall v\in P,\ |f(v)-\langle f\rangle_P|\le C_d 2^{-j};\qquad |\langle f\rangle_P-\langle f\rangle_{2P}|\le C_d 2^{-j}. \tag{3.159}$$

We note that both of these bounds are linear in the length scale 2^{-j} of the box. We deal with the case $N\ge 2^{2d}$.

The proof is based on the following 'hierarchical' partition of $\mathbb{R}^d$, given in the proof [157, Proposition 7.1]. For $j\in\mathbb{Z}$, we take $P_j=(-2^j,2^j]$.

Set $A_0=P_0$ and, for $j\ge 1$, $A_j=P_j\setminus P_{j-1}$. For $j\ge 1$ and $l\ge 2$, there is a unique partition $\mathfrak{P}_{j,l}$ of A_j by $2^{ld}-2^{(l-1)d}$ translates of P_{j-l}. Similarly, write $\mathfrak{P}_{0,l}$ for the unique partition of A_0 by 2^{dl} translates of P_{-l}. For $l\ge 3$ and $k\in\mathbb{Z}$, let $P\in\mathfrak{P}_{j,l}$. We write $\pi(P)$ for the unique element of of $\mathfrak{P}_{j,l-1}$ such that $P\subset\pi(P)$.

We deal first with the case $d\ge 3$. Fix discretisation parameters $L,J\ge 1$. Given a test function $f\in\mathcal{A}_1$, we can decompose

$$f=\sum_{j=0}^{J}\sum_{l=2}^{L}\sum_{P\in\mathfrak{P}_{j,l}} a_P(f)(1+|v|^2)\mathbb{1}_P+\beta(f) \tag{3.160}$$

where we define

$$a_P(f)=\begin{cases}\langle\hat f\rangle_P & \text{if } P\in\mathfrak{P}_{j,2},\ \text{for some } j\ge 0\\ \langle\hat f\rangle_P-\langle\hat f\rangle_{\pi(P)} & \text{if } P\in\mathfrak{P}_{j,l},\ \text{for some } j\ge 0, l\ge 3\end{cases} \tag{3.161}$$

and the equation serves to define the remainder[2] term $\beta(f)$. Write $h_P=2^{2j}(1+|v|^2)\mathbb{1}_P$, for $P\in\mathfrak{P}_{j,l}$, and write $M_t^{N;P}=M_t^{N,h_P}$. We can now write

$$M_t^{N,f}=\sum_{j=0}^{J}\sum_{l=2}^{L}\sum_{P\in\mathfrak{P}_{j,l}} 2^{-2j}a_P(f)M_t^{N;P}+R_t^{N,f}; \tag{3.162}$$

$$R_t^{N,f}=\int_{(0,t]\times\mathcal{S}_N}\langle\beta(f),\Delta(s,t,\mu^N)\rangle(m^N-\overline{m}^N)(ds,d\mu^N) \tag{3.163}$$

and where Δ, m^N and $\overline{m}^N$ are defined in Section 3.3. As in Proposition 3.11, this is the key decomposition in the proof. The strategy is very similar:

- The martingales $M^{N;P}$ are controlled by a bound (3.289) from the general theory of Markov chains, *independently of f*.

[2]We called this remainder $\varepsilon(f)$ earlier, but ε is already taken here, as the loss in the exponent.

- The coefficients a_P depend on f, but are bounded, uniformly over $f \in \mathcal{A}_1$.
- On P_J, $\beta(f)$ will be small, uniformly in $f \in \mathcal{A}_1$, due to the Lipschitz bound on f and the estimate (3.159), which is again a sort of quantitative compactness argument, see Remark 3.8.
- The contribution of $|\beta(f)|$ from outside P_J will be controlled by the moment bounds.

Step 1. Control of Martingale Sum We begin by controlling $M_t^{N;P}$ appearing in the decomposition. Let $(M_s^{N;P;t})_{s\le t}$ be the martingale

$$M_s^{N;P;t} = \int_{(0,s]\times\mathcal{S}_N} \langle h_P, \Delta(u,t,\mu^N)\rangle(m^N - \overline{m}^N)(du, d\mu^N). \tag{3.164}$$

We can control the martingale term pointwise in $L^2(\mathbb{P})$ by applying the martingale bound (3.289) at the terminal time t, with $M_t^{N;P} = M_t^{N;P;t}$:

$$\begin{aligned} \left\|M_t^{N;P}\right\|_{L^2(\mathbb{P})}^2 &= \mathbb{E}\int_{(0,t]\times\mathcal{S}_N} \langle (1+|v|^2)2^{2j}\mathbb{1}_P, \Delta(s,t,\mu^N)\rangle^2 \overline{m}^N(ds, d\mu^N) \\ &\le C\mathbb{E}\left[\int_{(0,t]\times\mathcal{S}_N} \langle (1+|v|^4)\mathbb{1}_P, |\Delta(s,t,\mu^N)|\rangle^2 \overline{m}^N(ds, d\mu^N)\right]. \end{aligned} \tag{3.165}$$

Summing over $P \in \mathfrak{P}_{j,l}$ and $j = 0, .., J$, we Minkowski's inequality to move the sum inside the integral against Δ, and note that, $\sum_j \sum_{P\in\mathfrak{P}_{j,l}} h_P \le C(1+|v|^4)$ uniformly in l. This produces the bound

$$\begin{aligned} \sum_{j=0}^J \sum_{P\in\mathfrak{P}_{j,l}} \left\|M_t^{N;P}\right\|_{L^2(\mathbb{P})}^2 &\le C\mathbb{E}\left[\int_{(0,t]\times\mathcal{S}_N} \langle (1+|v|^4), |\Delta(s,t,\mu^N)|\rangle^2 \overline{m}^N(ds, d\mu^N)\right] \\ &= \mathbb{E}\left[\int_{(0,t]\times\mathcal{S}_N} \|\phi_{t-s}(\mu^N) - \phi_{t-s}(\mu^N_{s-})\|^2_{\mathrm{TV}+4}\, \overline{m}^N(ds, d\mu^N)\right] \end{aligned} \tag{3.166}$$

where the second line follows by the definition of Δ in (3.135). Using the stability estimates in Corollary 3.16 with TV + 4, which is licit thanks to the choice of p in the lemma, we find

$$\sum_{j=0}^J \sum_{P\in\mathfrak{P}_{j,l}} \left\|M_t^{N;P}\right\|_{L^2(\mathbb{P})}^2 \le C\mathbb{E}\left[\int_{(0,t]\times\mathcal{S}_N} e^{-\lambda(t-s)}\Lambda_p(\mu^N_s, \mu^N) N^{2(\epsilon-1)}\, \overline{m}^N(ds, d\mu^N)\right]. \tag{3.167}$$

For $\overline{m}^N$-almost all (s, μ^N), we bound $\Lambda_p(\mu_s^N, \mu^N) \le C\Lambda_p(\mu_s^N)$ by Proposition 2.10iii), and $\overline{m}^N(ds, \mathcal{S}_N) \le 2N ds$, to bound the right hand side by

$$\begin{aligned} \sum_{j=0}^J \sum_{P\in\mathfrak{P}_{j,l}} \left\|M_t^{N;P}\right\|_{L^2(\mathbb{P})}^2 &\le C\int_0^t e^{-\lambda(t-s)} N^{2\epsilon-1}\, \mathbb{E}[\Lambda_p(\mu_s^N)]\, ds \\ &\le CN^{2\epsilon-1}a^{\frac{1}{2}} \end{aligned} \tag{3.168}$$

where the second line follows using the moment estimates for the Kac process, established in Proposition 2.10.

Step 2. Control of Coefficients & Martingale Sum We next control the coefficients $a_P(f)$, and hence the sum $\sum a_P(f)M_t^{N;P}$, uniformly in $f \in \mathcal{A}_1$. Observe that for $P \in \mathfrak{P}_{j,l}$, the bound (3.159) gives $2^{-2j}|a_P(f)| \le C2^{-j-l}$, and $\#\mathfrak{P}_{j,l} \le 2^{dl}$. Hence, independently of $f \in \mathcal{A}_1$,

$$\left(\sum_{j=0}^{J}\sum_{P\in\mathfrak{P}_{j,l}} (a_P(f)2^{-2j})^2\right) \le C2^{(d-2)l}. \tag{3.169}$$

Now, by Cauchy-Schwarz,

$$\sup_{f\in\mathcal{A}_1}\left|\sum_{j=0}^{J}\sum_{l=2}^{L}\sum_{P\in\mathfrak{P}_{j,l}} 2^{-2j}a_P(f)M_t^{N;P}\right| \le C\sum_{l=2}^{L}\left(\sum_{j=0}^{J}\sum_{P\in\mathfrak{P}_{j,l}}\left\{M_t^{N;P}\right\}^2\right)^{1/2} 2^{(d/2-1)l}. \tag{3.170}$$

Therefore, (3.168) gives

$$\begin{aligned}\left\|\sup_{f\in\mathcal{A}_1}\left|\sum_{j=0}^{J}\sum_{l=2}^{L}\sum_{P\in\mathfrak{P}_{j,l}} a_P(f)M_t^{N;P}\right|\right\|_{L^2(\mathbb{P})} &\le CN^{\epsilon-1/2}a^{1/2}\sum_{l=2}^{L}2^{(d/2-1)l} \\ &\le CN^{\epsilon-1/2}\ 2^{(d/2-1)L}\ a^{1/2}.\end{aligned} \tag{3.171}$$

Step 3. Control of error term $\beta(f)$. The remaining points are a control on $\beta(f)$, uniformly in $f\in\mathcal{A}_1$, dealing with P_J and $\mathbb{R}^d\setminus P_J$ separately. Fix $f\in\mathcal{A}_1$ and let $P\in\mathfrak{P}_{j,L}$ with $j\le J$. The definition gives $\hat\beta(f) = \hat f - \langle \hat f\rangle_P$ on any P, and so

$$\text{On } P, \quad |\beta(f)| = (1+|v|^2)|\hat f - \langle\hat f\rangle_P| \ \le C\,(1+|v|^2)2^{j-L}. \tag{3.172}$$

Since $|v|\ge 2^{j-1}$ on P, and $P\in\mathfrak{P}_{j,L}$ is arbitrary, we see that

$$\text{On } P_J, \quad |\beta(f)| \le C2^{-L}(1+|v|^4). \tag{3.173}$$

On the other hand, the uniform bound $\|\hat f\|_\infty \le 1$ implies that

$$\text{On } P_J^c, \quad |\beta(f)| \le (1+|v|^2) \le 2^{-2J}(1+|v|^4). \tag{3.174}$$

Combining, we have the global bound for all $f\in\mathcal{A}_1$:

$$\forall v\in\mathbb{R}^d, \quad |\beta(f)| \le C(2^{-2J}+2^{-L})(1+|v|^4). \tag{3.175}$$

Recalling the definition (3.135) of Δ, we use the stability estimate in Corollary 3.16, with TV + 4, and the moment increase bound Proposition 2.10iii), as above to see that almost surely, for $m^N+\overline{m}^N$-almost all (s,μ^N), we have the bound

$$\begin{aligned}\sup_{f\in\mathcal{A}_1}\left|\langle\beta(f), |\Delta(s,t,\mu^N)|\rangle\right| &\le C(2^{-2J}+2^{-L})\ \|\Delta(s,t,\mu^N)\|_{\mathrm{TV}+4} \\ &\le C(2^{-2J}+2^{-L})\ e^{-\lambda(t-s)/2}\ N^{\epsilon-1}\ \Lambda_p(\mu^N_{s-})^{\frac{1}{2}} \\ &=: H_s\end{aligned} \tag{3.176}$$

where we introduced the shorthand H_s for the final expression, for simplicity.

Step 4. Control over Remainder Terms We now integrate the bound found in the previous step. We now start from the trivial observation that

$$\sup_{f\in\mathcal{A}_1}\left|R_t^{N,f}\right| \le \int_{(0,t]\times\mathcal{S}_N}\left\{\sup_{f\in\mathcal{A}_1}\langle|\beta(f)|,|\Delta(s,t,\mu^N)|\rangle\right\}(m^N+\overline{m}^N)(ds,d\mu^N). \quad (3.177)$$

We split the measure $m^N+\overline{m}^N=(m^N-\overline{m}^N)+2\overline{m}^N$ to obtain a uniform bound for the error terms $R_t^{N,f}$ defined in (3.163):

$$\begin{aligned}\left\|\sup_{f\in\mathcal{A}_1} R_t^{N,f}\right\|_{L^2(\mathbb{P})} &\le C\left\|\int_0^t H_s(m^N+\overline{m}^N)(ds,\mathcal{S}_N)\right\|_{L^2(\mathbb{P})} \\ &\le C(2^{-2J}+2^{-L})N^{\epsilon-1}\left[\mathcal{T}_1+\mathcal{T}_2\right]\end{aligned} \quad (3.178)$$

where we have written

$$\mathcal{T}_1=\left\|\int_0^t e^{-\lambda(t-s)/2}\Lambda_p(\mu^N_{s-})^{\frac12}\overline{m}^N(ds,\mathcal{S}_N)\right\|_{L^2(\mathbb{P})} \quad (3.179)$$

$$\mathcal{T}_2=\left\|\int_0^t e^{-\lambda(t-s)/2}\Lambda_p(\mu^N_{s-})^{\frac12}(m^N-\overline{m}^N)(ds,\mathcal{S}_N)\right\|_{L^2(\mathbb{P})}. \quad (3.180)$$

$\mathcal{T}_1$ is controlled by dominating $\overline{m}^N(ds,\mathcal{S}_N)\le 2Nds$ to obtain

$$\begin{aligned}\mathcal{T}_1\le CN\left\|\int_0^t e^{-\lambda(t-s)/2}\Lambda_p(\mu^N_s)^{\frac12}ds\right\|_{L^2(\mathbb{P})} &\le CN\int_0^t e^{-\lambda(t-s)/2}\|\Lambda_p(\mu^N_s)^{\frac12}\|_{L^2(\mathbb{P})}\,ds \\ &\le CNa^{1/2}.\end{aligned} \quad (3.181)$$

We control $\mathcal{T}_2$ by Itô's isometry for $m^N-\overline{m}^N$, which is reviewed in (3.290):

$$\begin{aligned}\mathcal{T}_2^2=\mathbb{E}\left\{\int_0^t e^{-\lambda(t-s)}\Lambda_k(\mu^N_{s-})\overline{m}^N(ds,\mathcal{S}_N)\right\} &\le CN\int_0^t e^{-\lambda(t-s)}\mathbb{E}\left\{\Lambda_p(\mu^N_{s-})\right\}ds \\ &\le CN\,a.\end{aligned} \quad (3.182)$$

Combining (3.178, 3.181, 3.182), we obtain

$$\left\|\sup_{f\in\mathcal{A}_1} R_t^{N,f}\right\|_{L^2(\mathbb{P})} \le C(2^{-2J}+2^{-L})\,N^{\epsilon-1}\,a^{1/2}. \quad (3.183)$$

Step 5. Conclusion. We now conclude. Gathering (3.162, 3.171, 3.183) we have proven that

$$\left\|\sup_{f\in\mathcal{A}_1}\left|M_t^{N,f}\right|\right\|_{L^2(\mathbb{P})} \le C\;N^\epsilon\,a^{1/2}(N^{-1/2}\,2^{(d/2-1)L}+2^{-L}+2^{-2J}). \quad (3.184)$$

Taking $L=\lfloor\log_2(N)/d\rfloor$ and $J\uparrow\infty$ produces the claimed result. For $d=2$, we replace $2^{(d/2-1)L}$ by L in (3.171), and optimise as before, absorbing the factors of $(\log N)$ to make the exponent of N slightly larger. □

3.5 Proof of Theorem 3.2

We now adapt the ideas of Lemma 3.21 to a local uniform setting, and working in L^p, to prove the local uniform approximation result Theorem 3.2. As in the proof above, most of the work is in controlling the martingale term $(M_t^{N,f})_{f\in\mathcal{A}_1}$ defined in (3.138), uniformly in f; for a pathwise local uniform estimate, we wish to control an expression of the form

$$\left\| \sup_{f\in\mathcal{A}_1} \sup_{t\le t_{\text{fin}}} \left|M_t^{N,f}\right| \right\|_{L^p(\mathbb{P})}. \tag{3.185}$$

Since we will frequently encounter suprema of processes on compact time intervals, we introduce notation. For any stochastic process M, we write

$$M_{\star,t} = \sup_{s\le t} |M_t| \tag{3.186}$$

Proving the sharpest asymptotics in the time horizon t_{fin} requires working in $L^q(\mathbb{P})$ instead of $L^2(\mathbb{P})$, for large exponents q. This leads to a weaker exponent in N: we obtain only $N^{\epsilon-q'/2d}$ instead of $N^{\epsilon-1/d}$, where $q'\le 2$ is the Hölder conjugate to p. However, by making q large, we are able to obtain estimates which degrade slowly in the time horizon t_{fin}, with only a factor of $(1+t_{\text{fin}})^{1/q}$. The exponent for t_{fin} can thus be made arbitrarily small, while the resulting exponent for N is bounded away from 0 as we make q large.

The key result required for the local uniform estimate is the following control of the expression (3.185), in analogy to Lemma 3.21.

Lemma 3.22. *Let $\epsilon>0$, $a\ge 1$ and $q\ge 2$, and let $1<q'\le 2$ be the Hölder conjugate to p. Let p be large enough that Corollary 3.16 holds for* TV + 5*, with Hölder exponent $1-\epsilon$, and with some $0<\lambda<\lambda_0$.*
Let $(\mu_t^N)_{t\ge 0}$ be a Kac process on $N\ge 2$ particles, with initial moment $\Lambda_{pq}(\mu_0^N)\le a^q$. Let $M_t^{N,f}$ be the processes given by (3.138), and $M_{\star,t}^{N,f}$ their local suprema, as in (3.186). Then, for any time horizon $t_{fin}\in[0,\infty)$, we have the control

$$\left\| \sup_{f\in\mathcal{A}_1} M_{\star,t_{fin}}^{N,f} \right\|_{L^q(\mathbb{P})} \le Ca^{1/2}\, N^{-\alpha}\, (\log N)^{1/q'}\, (1+t_{fin})^{\frac{3q+1}{2q}} \tag{3.187}$$

where $\alpha = \frac{q'}{2d}-\epsilon$.

The proof of this Lemma follows the same ideas as the proof of the equivalent result, Lemma 3.21, for the pointwise bound. However, in this case, we must modify the argument to work in $L^q(\mathbb{P})$ rather than $L^2(\mathbb{P})$, and also to control all terms uniformly on the compact time interval $[0,t_{\text{fin}}]$. This will be deferred until the end of this section.

Following the argument of the pointwise bound in Theorem 3.1, we can now produce an initial pathwise, local uniform estimate for the case $\mu_0=\mu_0^N$, with worse long-time behaviour. From this, we will 'bootstrap' to the desired long-time behaviour in Theorem 3.2.

Lemma 3.23. *Let $\epsilon > 0$, $a \geq 1$ and $q \geq 2$, with Hölder conjugate $q' \leq 2$. Choose p large enough that Proposition 3.15 holds with exponent $1-\epsilon$, and that Corollary 3.16 holds with exponent $1-\epsilon$ and the norm $\mathrm{TV}+5$. Let $(\mu^N_t)_{t\geq 0}$ be a Kac process on $N \geq 2$ particles, with initial moment $\Lambda_{pq}(\mu^N_0) \leq a^q$. Then, for any time horizon $t_{fin} \geq 0$, we have the control*

$$\left\| \sup_{t\leq t_{fin}} W_1\left(\mu^N_t, \phi_t\left(\mu^N_0\right)\right) \right\|_{L^q(\mathbb{P})} \leq Ca^{1/2}\, N^{\epsilon-\frac{q'}{2d}}\, (\log N)^{1/q'}(1+t_{fin})^{\frac{3q+1}{2q}}. \tag{3.188}$$

Proof of Lemma 3.23. As in Theorem 3.1, it remains to control the supremum of the integral term in Formula 3.3.1

$$\sup_{t\leq t_{\text{fin}}} \int_0^t \sup_{f\in\mathcal{A}_1} \langle f, \rho^N(t-s,\mu^N_s)\rangle ds \tag{3.189}$$

where ρ^N is given by (3.139). Following the previous calculation (3.157), we majorise, for $s \leq t \leq t_{\text{fin}}$,

$$\sup_{f\in\mathcal{A}_1} \langle f, \rho^N(t-s,\mu^N_s)\rangle \leq CN^{\epsilon-1} \sup_{u\leq t_{\text{fin}}} \left\{\Lambda_p(\mu^N_u)^{\frac{1}{2}}\right\} \tag{3.190}$$

from which it follows that

$$\sup_{t\leq t_{\text{fin}}} \int_0^t \sup_{f\in\mathcal{A}_1} \langle f, \rho^N(t-s,\mu^N_s)\rangle ds \leq CN^{\epsilon-1}\, t_{\text{fin}} \sup_{u\leq t_{\text{fin}}} \left\{\Lambda_p(\mu^N_u)^{\frac{1}{2}}\right\}. \tag{3.191}$$

From the local uniform moment bound established in Proposition 2.10ii), and the initial moment bound on μ^N_0,

$$\begin{aligned} \left\| \sup_{u\leq t_{\text{fin}}} \left\{\Lambda_p(\mu^N_u)^{\frac{1}{2}}\right\}\right\|_{L^q(\mathbb{P})} &\leq \left\| \sup_{u\leq t_{\text{fin}}} \left\{\Lambda_p(\mu^N_u)^{\frac{1}{2}}\right\}\right\|_{L^{2q}(\mathbb{P})} \leq \mathbb{E}\left[\sup_{u\leq t_{\text{fin}}} \Lambda_{pq}(\mu^N_u)^{\frac{1}{2}}\right]^{1/2q} \\ &\leq Ca^{1/2}\,(1+t_{\text{fin}})^{1/2q}. \end{aligned} \tag{3.192}$$

Combining the estimates (3.191, 3.192), we see that

$$\left\|\sup_{t\leq t_{\text{fin}}} \int_0^t \sup_{f\in\mathcal{A}_1} \langle f, \rho^N(t-s,\mu^N_s)\rangle ds\right\|_{L^q(\mathbb{P})} \leq CN^{\epsilon-1}\, a^{1/2}\,(1+t_{\text{fin}})^{\frac{2q+1}{2q}}. \tag{3.193}$$

We combine this with Lemma 3.22 and keep the worse asymptotics. □

We will now show how to 'bootstrap' to better dependence of the time horizon t_{fin}. Heuristically, the proof allows us to replace powers of t_{fin} in the initial bound with the same power of $\log N$, and introduce an additional factor of $(1+t_{\text{fin}})^{1/p}$. As was remarked below Proposition 3.15, we could derive Theorem 3.1 and Lemma 3.23 under the milder assumptions

$$\|\phi_t(\nu)-\phi_t(\mu)\|_{\mathrm{TV}+5} \leq F(t)\Lambda_p(\mu,\nu)^{\frac{1}{2}}\|\mu-\nu\|^{\eta}_{\mathrm{TV}}; \tag{3.194}$$

$$\|\phi_t(\nu)-\phi_t(\mu)-\xi_t\|_{\mathrm{TV}+2} \leq G(t)\Lambda_p(\mu,\nu)^{\frac{1}{2}}\|\mu-\nu\|^{1+\eta}_{\mathrm{TV}} \tag{3.195}$$

for functions F, G such that

$$\left(\int_0^\infty F^2 dt\right)^{1/2} < \infty; \qquad \int_0^\infty G dt < \infty. \tag{3.196}$$

If we also assume that $F \to 0$ as $t \to \infty$, we can use an identical bootstrap argument, with $\log N$ replaced by a power of

$$\tau_N := \sup\{t : F(t) > N^{-\alpha}\} \tag{3.197}$$

which produces a potentially larger loss. *Hence, the the full strength of exponential decay in Proposition 3.15 is used to control the asymptotic loss due to the bootstrap.*

Proof of Theorem 3.2. As in the proof of Theorem 3.1, it is sufficient to prove the case $\mu_0^N = \mu_0$. Then, making p larger if necessary, we may use Theorem 3.6 to control $\sup_{t\ge 0} W_1(\phi_t(\mu_0^N), \phi_t(\mu_0))$, which proves the general result.

Let $0 < \epsilon' < \epsilon$, and choose p such that Lemma 3.23 holds for ϵ'. Let $\alpha' > \alpha$ be the exponent of N obtained with ϵ' in place of ϵ. From the stability estimate Proposition 3.15, we have

$$\forall \mu, \nu \in \mathcal{S}^p, \qquad \|\phi_t(\mu) - \phi_t(\nu)\|_{\mathrm{TV}+2} \le C\Lambda_p(\mu,\nu)^{\frac{1}{2}} e^{-\lambda_0 t/2}. \tag{3.198}$$

Define $\tau = \tau_N = -2\lambda_0^{-1}\log(N^{-\alpha'})$ and consider $t_{\text{fin}} > \tau + 1$. Fix a positive integer n, and partition the interval $[0, t_{\text{fin}}]$ as $I_1 \cup I_1 \cup ... \cup I_n$:

$$I_0 = [0,\tau]; \qquad I_r = \left[\tau + (r-1)\frac{t_{\text{fin}} - \tau}{n}, \tau + r\frac{t_{\text{fin}} - \tau}{n}\right] =: [s_r + \tau, t_r]. \tag{3.199}$$

Write also $H_r = [s_r, t_r] \supset I_r$. Since the norm $\|\cdot\|_{\mathrm{TV}+2}$ dominates the Wasserstein distance W_1, we apply (3.198) to bound $W_1(\phi_{t-s_r}(\mu_{s_r}^N), \phi_{t-s_r}(\phi_{s_r}(\mu_0^N)))$ to get bound

$$\sup_{t\in I_r} W_1(\mu_t^N, \phi_t(\mu_0^N)) \le C \sup_{t\in H_r} W_1(\mu_t^N, \phi_{t-s_r}(\mu_{s_r}^N)) + e^{-\lambda\tau}\Lambda_p(\mu_{s_r}^N, \phi_{s_r}(\mu_0^N))^{\frac{1}{2}}. \tag{3.200}$$

We bound the two terms in (3.200) separately. Recalling the notation $(\mathfrak{F}_t^N)_{t\ge 0}$ for the natural filtration of $(\mu_t^N)_{t\ge 0}$, we control the first term by Lemma 3.23, applied to the restarted process $(\mu_t^N)_{t\ge s_r}$: recalling that $t_r - s_r = (1 + \tau + (t-\tau)/n)$, we get

$$\begin{aligned} \left\|\sup_{t\in H_r} W_1(\mu_t^N, \phi_{t-s_r}(\mu_{s_r}^N))\right\|_{L^q(\mathbb{P})}^q &= \mathbb{E}\left\{\mathbb{E}\left(\left[\sup_{s_r\le t\le t_r} W_1(\mu_t^N, \phi_{t-s_r}(\mu_{s_r}^N))\right]^q \middle| \mathfrak{F}_{s_r}^N\right)\right\} \\ &\le C\mathbb{E}\left\{\Lambda_{pq}(\mu_{s_r}^N)^{1/q}\right\}\left(1 + \tau + \frac{t-\tau}{n}\right)^{\frac{3q+1}{2}} N^{-q\alpha'}(\log N)^{\frac{q}{q'}}. \end{aligned} \tag{3.201}$$

We control the moment in the usual way, using Proposition 2.10, to obtain

$$\left\|\sup_{t\in H_r} W_1(\mu_t^N, \phi_{t-s_r}(\mu_{s_r}^N))\right\|_{L^q(\mathbb{P})}^q \le Ca^q\left(1 + \tau + \frac{t-\tau}{n}\right)^{\frac{3q+1}{2}} N^{-q\alpha'}(\log N)^{\frac{q}{q'}}. \tag{3.202}$$

We now turn to the second term in (3.200). Using the definition of τ and the moment estimates in Propositions 2.6, 2.10,

$$\|e^{-\lambda\tau/2}\Lambda_p(\mu^N_{s_r}, \phi_{s_r}(\mu^N_0))^{\frac{1}{2}}\|_{L^q(\mathbb{P})} \leq CN^{-\alpha'}\, a^{1/2}. \tag{3.203}$$

Combining the estimates (3.202, 3.203), and absorbing powers of τ into the powers of $(\log N)$, we obtain

$$\left\| \sup_{t\in I_r} W_1(\mu^N_t, \phi_t(\mu^N_0)) \right\|_{L^q(\mathbb{P})} \leq Ca^{1/2} \left(1 + \frac{t_{\text{fin}}-\tau}{n}\right)^{\frac{3q+1}{2q}} \left(N^{-\alpha'}(\log N)^{\frac{3q+1}{2q}+\frac{1}{q'}}\right). \tag{3.204}$$

To get the supremum over the interval $[\tau, t_{\text{fin}}]$, observe that

$$\left\{\sup_{\tau\leq t\leq t_{\text{fin}}} W_1\left(\mu^N_t, \phi_t(\mu^N_0)\right)\right\}^q \leq \sum_{r=1}^n \left\{\sup_{t\in I_r} W_1\left(\mu^N_t, \phi_t(\mu^N_0)\right)\right\}^q. \tag{3.205}$$

Taking expectations and q^{th} root, we find that

$$\begin{aligned}\left\| \sup_{\tau\leq t\leq t_{\text{fin}}} W_1\left(\mu^N_t, \phi_t(\mu^N_0)\right)\right\|_{L^q(\mathbb{P})} \\ \leq Cn^{\frac{1}{q}}\, a^{1/2} \left(1 + \frac{t_{\text{fin}}-\tau}{n}\right)^{\frac{3q+1}{2q}} \left(N^{-\alpha'}(\log N)^{\frac{3q+1}{2q}+\frac{1}{q'}}\right).\end{aligned} \tag{3.206}$$

We now take $n = \lceil t_{\text{fin}} - \tau\rceil$, and we obtain the estimate

$$\begin{aligned}\left\| \sup_{\tau\leq t\leq t_{\text{fin}}} W_1\left(\mu^N_t, \phi_t(\mu^N_0)\right)\right\|_{L^q(\mathbb{P})} &\leq Ca^{1/2}(t_{\text{fin}}-\tau)^{\frac{1}{q}} \left(N^{-\alpha'}(\log N)^{\frac{3q+1}{2q}+\frac{1}{q'}}\right) \\ &\leq a^{1/2}\, t_{\text{fin}}^{\frac{1}{q}} \left(N^{-\alpha'}(\log N)^{\frac{3q+1}{2q}+\frac{1}{p'}}\right).\end{aligned} \tag{3.207}$$

From Lemma 3.23 applied up to time $\tau = \tau_N$, we have

$$\begin{aligned}\left\| \sup_{0\leq t\leq \tau_N} W_1\left(\mu^N_t, \phi_t(\mu^N_0)\right)\right\|_{L^q(\mathbb{P})} &\leq Ca^{1/2}\, N^{-\alpha'} \left(1 + \frac{2\alpha}{\lambda}\log(N)\right)^{\frac{3q+1}{2q}} (\log N)^{\frac{1}{q'}} \\ &\leq Ca^{1/2} \left(N^{-\alpha}(\log N)^{\frac{3q+1}{2q}+\frac{1}{q'}}\right).\end{aligned} \tag{3.208}$$

Combining (3.207, 3.208), and absorbing the powers of $(\log N)$ into $N^{\epsilon-\epsilon'}$, we have

$$\left\| \sup_{0\leq t\leq t_{\text{fin}}} W_1\left(\mu^N_t, \phi_t(\mu^N_0)\right)\right\|_{L^q(\mathbb{P})} \leq Ca^{1/2}\,(1+t_{\text{fin}})^{\frac{1}{q}}\, N^{-\alpha}. \tag{3.209}$$

The case where $t_{\text{fin}} \leq \tau + 1$ is essentially identical to (3.208). □

Remark 3.24. *We note that this 'bootstrap' argument would produce the same result with* any polynomial *time dependence in Lemma 3.23. As a result, the precise time dependence of Lemmas 3.22, 3.23 is uninteresting, and we do not attempt to optimise it. We also remark that this method produces the same long-time behaviour even starting from an exponential estimate, at the cost of a fractional power of N.*

It remains to prove Lemma 3.22. We draw attention to the fact that $M^{f,N}$ are *not* themselves martingales, despite the general construction (see eq. 3.288), since the integrand $\phi_{t-s}(\mu^N)-\phi_{t-s}(\mu^N_{s-})$ depends on the terminal time t. We address this by computing an associated family of martingales:

Lemma 3.25. *Let $(M^{N,f}_t)_{t\ge 0}$ be the processes defined in Formula 3.3.1. Recalling the notation $Q_t=Q\circ\phi_t$, define*

$$\chi(s,t,\mu^N)=Q_{t-s}(\mu^N)-Q_{t-s}(\mu^N_{s-}). \tag{3.210}$$

Suppose f satisfies a growth condition $|f(v)|\le(1+|v|^q)$, for some $q\ge 0$. Consider the martingales $Z^{N,f}_t$ given by

$$Z^{N,f}_t=\int_{(0,t]\times\mathcal{S}_N}\langle f,\mu^N-\mu^N_{s-}\rangle(m^N-\overline{m}^N)(ds,d\mu^N)\rangle. \tag{3.211}$$

Then we have the equality

$$\begin{aligned}Z^{N,f}_t&=M^{N,f}_t-C^{N,f}_t\\&=M^{N,f}_t-\int_0^t ds\int_{(0,s]\times\mathcal{S}_N}\langle f,\chi(u,s,\mu^N)\rangle(m^N-\overline{m}^N)(du,d\mu^N).\end{aligned} \tag{3.212}$$

Proof. Firstly, we note that $Z^{N,f}_t$ are martingales by standard results from Markov chains, (3.288). Observe that the integrand in the definition of $C^{N,f}_t$ is bounded, since whenever $0\le u\le s$, and μ^N is obtain from μ^N_{u-} by collision, we use the estimate (3.109) with $\eta=\frac{1}{2}$, to obtain for some p

$$\begin{aligned}|\langle f,\chi(u,s,\mu^N)\rangle|&\le\|Q_{s-u}(\mu^N)-Q_{s-u}(\mu^N_{u-})\|_{\mathrm{TV}+q}\\&\le C\Lambda_p(\mu^N,\mu^N_{u-})^{\frac{1}{2}}N^{-\frac{1}{2}}\le CN^{\frac{p-2}{4}}<\infty.\end{aligned} \tag{3.213}$$

Moreover, for initial data $\mu^N\in\mathcal{S}_N$, the Boltzmann flow $(\phi_s(\mu^N))_{s=0}^t$ has uniformly bounded $(q+1)^{\text{th}}$ moments and so, by approximation, the Boltzmann dynamics (BE) extend to f. Now, we apply Fubini to the integral:

$$\begin{aligned}&C^{N,f}_t\\&=\int_{(0,t]\times\mathcal{S}_N}\int_0^t ds\,\langle f,Q_{s-u}(\mu^N)-Q_{s-u}(\mu^N_{u-})\rangle\,\mathbb{1}[u\le s\le t]\,(m^N-\overline{m}^N)(du,d\mu^N)\\&=\int_{(0,t]\times\mathcal{S}_N}\left\{\int_u^t\left(\langle f,Q_{s-u}(\mu^N)\rangle-\langle f,Q_{s-u}(\mu^N_{u-})\rangle\right)ds\right\}(m^N-\overline{m}^N)(du,d\mu^N)\\&=\int_{(0,t]\times\mathcal{S}_N}\{\langle f,\phi_{t-u}(\mu^N)-\phi_{t-u}(\mu^N_{u-})\rangle-\langle f,\mu^N-\mu^N_{u-}\rangle\}\,(m^N-\overline{m}^N)(du,d\mu^N)\\&\qquad=:M^{N,f}_t-Z^{N,f}_t\end{aligned} \tag{3.214}$$

where the third equality is precisely the Boltzmann equation (BE) in the variable $s\in[u,t]$, extended to f as in Proposition 2.6 Step 2. ☐

To prove Lemma 3.22, we return to the decomposition ((3.160, 3.162) used in the proof of Lemma 3.21. We will assume, from now on, the same notation $a_P(f), \beta(f), J, L$. As in the previous proof, our first step is to establish a control on the local uniform sum

$$\mathbb{E}\left[\sum_{j=0}^{J}\sum_{P\in\mathfrak{P}_{j,l}}\left\{M^{N;P}_{\star,t_{\text{fin}}}\right\}^q\right] \tag{3.215}$$

where $\star$ denotes the local supremum (3.186). We will control the local uniform sum so by breaking the supremum into two parts, each of which can be controlled by elementary martingale estimates. Let $(J^{N;P;t}_s)_{0\le s\le t}$ be the process

$$J^{N;P;t}_s = \int_{(0,s]\times\mathcal{S}_N}\langle h_P, Q_{t-u}(\mu^N)-Q_{t-u}(\mu^N_{u-})\rangle(m^N-\overline{m}^N)(du,d\mu^N) \tag{3.216}$$

where, as in the proof of Theorem 3.1,

$$h_P = 2^{2j}(1+|v|^2)\mathbb{1}_P; \qquad P\in\mathfrak{P}_{j,l}. \tag{3.217}$$

Each process $(J^{N;P;t}_s)_{0\le s\le t}$ is a martingale, by standard results for Markov chains (3.288). Writing $Z^{N;P}=Z^{N,h_P}$, Lemma 3.25 gives

$$Z^{N;P}_t = M^{N;P}_t + \int_0^t J^{N;P;s}_s\, ds. \tag{3.218}$$

We achieve the comparison through the following lemma

Lemma 3.26. *Let $q\ge 2$, and let q' be the Hölder conjugate to q. In the notation above, we have the comparison*

$$\mathbb{E}\left[\sum_{j=0}^{J}\sum_{P\in\mathfrak{P}_{j,l}}\left\{\left|M^{N;P}_{\star,t_{fin}}\right|\right\}^q\right]\le C\mathbb{E}\left[\sum_{j=0}^{J}\sum_{P\in\mathfrak{P}_{j,l}}\left\{\left|M^{N;P}_{t_{fin}}\right|^q + t_{fin}^{q/q'}\int_0^{t_{fin}}\left|J^{N;P;t}_t\right|^q dt\right\}\right]. \tag{3.219}$$

for some constant C depending only on q.

Proof. For each partition element P, we observe that

$$\sup_{t\le t_{\text{fin}}}\left|M^{N;P}_t - Z^{N;P}_t\right| \le \int_0^{t_{\text{fin}}}\left|J^{N;P;s}_s\right| ds \tag{3.220}$$

which implies the two bounds

$$M^{N;P}_{\star,t_{\text{fin}}}\le Z^{N;P}_{\star,t_{\text{fin}}}+\int_0^{t_{\text{fin}}}\left|J^{N;P;s}_s\right|ds; \qquad Z^{N;P}_{t_{\text{fin}}}\le M^{N;P}_{t_{\text{fin}}}+\int_0^{t_{\text{fin}}}\left|J^{N;P;s}_s\right|ds. \tag{3.221}$$

By Doob's $L^q(\mathbb{P})$ inequality, we have

$$\left\| Z^{N;P}_{\star,t_{\text{fin}}}\right\|_{L^q(\mathbb{P})}\le q'\left\| Z^{N;P}_{t_{\text{fin}}}\right\|_{L^q(\mathbb{P})}. \tag{3.222}$$

Combining (3.221, 3.222), we obtain

$$\left\| M^{N;P}_{\star,t_{\text{fin}}} \right\|_{L^q(\mathbb{P})} \leq C \left\| M^{N;P}_{t_{\text{fin}}} \right\|_{L^q(\mathbb{P})} + C \left\| \int_0^{t_{\text{fin}}} \left| J^{N;P;s}_s \right| ds \right\|_{L^q(\mathbb{P})}. \tag{3.223}$$

Using Hölder's inequality on the integral,

$$\begin{aligned} \mathbb{E}\left[\left\{ M^{N;P}_{\star,t_{\text{fin}}} \right\}^q \right] &\leq C\mathbb{E}\left[\left| M^{N;P}_{t_{\text{fin}}} \right|^q \right] + C\mathbb{E}\left[\left\{ \int_0^{t_{\text{fin}}} \left| J^{N;P;s}_s \right| \, ds \right\}^q \right] \\ &\leq C\mathbb{E}\left[\left| M^{N;P}_{t_{\text{fin}}} \right|^q \right] + C t_{\text{fin}}^{q/q'} \int_0^{t_{\text{fin}}} \mathbb{E}\left[\left| J^{N;P;t}_t \right|^q \right] \, ds. \end{aligned} \tag{3.224}$$

Summing over $P \in \mathfrak{P}_{j,l}$ and $j = 0, 1, \ldots, J$, we obtain the desired comparison. □

Proof of Lemma 3.22. We follow the same steps as in Lemma 3.21 for the same decomposition (3.162).

Step 1. Control of the Local Uniform Martingale Sum We first control the local uniform sum as in in Lemma 3.26. The quadratic variation of the processes $J^{N;P;t}$ appearing in the upper bound is given by

$$\begin{aligned} \left[J^{N;P;t} \right]_s &= \int_{(0,s]\times\mathcal{S}_N} \langle h_P, \chi(u,t,\mu^N) \rangle^2 m^N(du, d\mu^N) \\ &\leq \int_{(0,s]\times\mathcal{S}_N} \langle h_P, |\chi(u,t,\mu^N)| \rangle^2 m^N(du, d\mu^N) \end{aligned} \tag{3.225}$$

where h_P is as in (3.217) and χ is as in (3.210). Hence, using Burkholder's inequality (See Lemma 3.32) we see that, for all $t \leq t_{\text{fin}}$,

$$\begin{aligned} &\mathbb{E}\left[\sum_{j=0}^J \sum_{P\in\mathfrak{P}_{j,l}} \left\{ \left| J^{N;P;t}_t \right| \right\}^q \right] \\ &\qquad \leq C\mathbb{E}\left[\sum_{j=0}^J \sum_{P\in\mathfrak{P}_{j,l}} \left\{ \int_{(0,t]\times\mathcal{S}_N} \langle h_P, |\chi(u,t,\mu^N)| \rangle^2 m^N(du, d\mu^N) \right\}^{q/2} \right]. \end{aligned} \tag{3.226}$$

We move the double sum inside the parentheses and then inside the integral against $|\chi(u,t,\mu^N)|$, recalling that $\sum_j \sum_{P\in\mathfrak{P}_{j,l}} h_P \leq C(1+|v|^4)$ uniformly in l, we obtain the bound

$$\begin{aligned} &\mathbb{E}\left[\sum_{j=0}^J \sum_{P\in\mathfrak{P}_{j,l}} \left\{ \left| J^{N;P;t}_t \right| \right\}^q \right] \\ &\leq C\mathbb{E}\left[\left\{ \int_{(0,t]\times\mathcal{S}_N} \langle 1+|v|^4, |\chi(u,t,\mu^N)| \rangle^2 m^N(du, d\mu^N) \right\}^{q/2} \right] \\ &\leq C\mathbb{E}\left[\left\{ \int_{(0,t]\times\mathcal{S}_N} \| Q_{t-u}(\mu^N) - Q_{t-u}(\mu^N_{u-}) \|^2_{\text{TV}+4} \, m^N(du, d\mu^N) \right\}^{q/2} \right] \end{aligned} \tag{3.227}$$

where the second equality is the definition of χ given in (3.210).

We now note that with the hypothesis of the lemma, we can use (3.109) with the norm $\|\cdot\|_{\mathrm{TV}+4}$, since we have control over ϕ_t in the norm $\|\cdot\|_{\mathrm{TV}+5}$, and arguing as in the proof of Lemma 3.21, we see that almost surely, for m^N-almost all (u,μ^N), we have

$$\|Q_{t-u}(\mu^N)-Q_{t-u}(\mu^N_{u-})\|_{\mathrm{TV}+4}\leq CN^{\epsilon-1}\Lambda_p(\mu^N_{u-}). \tag{3.228}$$

Therefore, using Cauchy-Schwarz, (3.227) gives the bound

$$\begin{aligned}&\mathbb{E}\left[\sum_{j=0}^{J}\sum_{P\in\mathfrak{P}_{j,l}}\left\{\left|J_t^{N;P;t}\right|\right\}^q\right]\\ &\qquad\leq CN^{q(\epsilon-1)}\,\mathbb{E}\left[\sup_{t\leq t_{\mathrm{fin}}}\Lambda_{pq}(\mu^N_t)\right]^{1/2}\left\|m^N\left((0,t_{\mathrm{fin}}]\times\mathcal{S}_N\right)\right\|^{q/2}_{L^q(\mathbb{P})}.\end{aligned} \tag{3.229}$$

The moment term is controlled by the initial moment bound and Proposition 2.10ii):

$$\mathbb{E}\left[\sup_{t\leq t_{\mathrm{fin}}}\Lambda_{pq}(\mu^N_t)\right]\leq C(1+t_{\mathrm{fin}})\Lambda_{pq}(\mu^N_0)\leq C(1+t_{\mathrm{fin}})a^q. \tag{3.230}$$

For the second factor, the rates of the Kac process are bounded by $2N$, and so we can stochastically dominate $m^N(dt\times\mathcal{S}_N)$ by a Poisson random measure $\mathfrak{m}^N(dt)$ of rate $2N$. By the additive property of Poisson processes, it follows that

$$\|m^N((0,t_{\mathrm{fin}}]\times\mathcal{S}_N)\|_{L^q(\mathbb{P})}\leq\|\mathfrak{m}^N(0,t_{\mathrm{fin}}]\|_{L^q(\mathbb{P})}\leq CN(1+t_{\mathrm{fin}}). \tag{3.231}$$

Combining (3.229, 3.230, 3.231), we have the control of the integrand:

$$\sup_{t\leq t_{\mathrm{fin}}}\ \mathbb{E}\left[\sum_{j=0}^{J}\sum_{P\in\mathfrak{P}_{j,l}}\left\{\left|J_t^{N;P;t}\right|\right\}^q\right]\leq CN^{q(\epsilon-1/2)}\,a^{q/2}(1+t_{\mathrm{fin}})^{\frac{q+1}{2}}. \tag{3.232}$$

This gives the following control of the integral term in Lemma 3.26:

$$t_{\mathrm{fin}}^{q/q'}\ \mathbb{E}\left[\sum_{j=0}^{J}\sum_{P\in\mathfrak{P}_{j,l}}\int_0^{t_{\mathrm{fin}}}\left\{\left|J_t^{N;P;t}\right|\right\}^q dt\right]\leq CN^{q(\epsilon-1/2)}\,a^{q/2}(1+t_{\mathrm{fin}})^{\frac{q+3}{2}+\frac{q}{q'}}. \tag{3.233}$$

Using the definition of q' as the Hölder conjugate to q, it is straightforward to see that the exponent of $(1+t_{\mathrm{fin}})$ is $\frac{3q+1}{2}$.

We now perform a similar analysis for the terms $M^{N;P}_{t_{\mathrm{fin}}}$ in Lemma 3.26. Let $(M^{N;P;t}_s)_{s\leq t}$ be the martingale defined in (3.164). The quadratic variation is

$$\begin{aligned}\left[M^{N;P;t}\right]_s&=\int_{(0,s]\times\mathcal{S}_N}\langle h_P,\phi_{t-u}(\mu^N)-\phi_{t-u}(\mu^N_{u-})\rangle^2\,m^N(du,d\mu^N)\\ &\leq\int_{(0,s]\times\mathcal{S}_N}\langle h_P,|\phi_{t-u}(\mu^N)-\phi_{t-u}(\mu^N_{u-})|\rangle^2\,m^N(du,d\mu^N).\end{aligned} \tag{3.234}$$

Arguing using Burkholder and the stability estimate Corollary 3.16, an identical calculation to the above shows that

$$\sum_{j=0}^{J}\sum_{P\in\mathfrak{P}_{j,l}}\left\|M^{N;P}_{t_{\mathrm{fin}}}\right\|^q_{L^q(\mathbb{P})} \le CN^{q(\epsilon-1/2)}\, a^{q/2}\,(1+t_{\mathrm{fin}})^{\frac{p+1}{2}}. \tag{3.235}$$

Hence, by Lemma 3.26, we obtain

$$\mathbb{E}\left[\sum_{j=0}^{J}\sum_{P\in\mathfrak{P}_{j,l}}\left\{\left|M^{N;P}_{\star,t_{\mathrm{fin}}}\right|\right\}^q\right] \le CN^{q(\epsilon-1/2)}a^{q/2}(1+t_{\mathrm{fin}})^{\frac{3q+1}{2}}. \tag{3.236}$$

Step 2. Control of Coefficients & Martingale Sum We control the coefficients $2^{-2j}a_P(f)$ as in the argument of Lemma 3.21. Repeating the arguments of (3.170) with Hölder's inequality, we obtain

$$\begin{aligned}\left\|\sup_{f\in\mathcal{A}_1}\sup_{t\le t_{\mathrm{fin}}}\left|\sum_{j=0}^{J}\sum_{l=2}^{L}\sum_{P\in\mathfrak{P}_{j,l}}2^{-2j}a_P(f)M^{N;P}_t\right|\right\|_{L^q(\mathbb{P})} &\le C\sum_{l=2}^{L}\left[\mathbb{E}\sum_{j=0}^{J}\sum_{P\in\mathfrak{P}_{j,l}}\left\{M^{N;P}_{\star,t_{\mathrm{fin}}}\right\}^q\right]^{1/q}2^{(d/q'-1)l}\\ &\le C\sum_{l=2}^{L}N^{\epsilon-\frac12}\,a^{1/2}\,(1+t_{\mathrm{fin}})^{\frac{3p+1}{2p}}\,2^{(d/q'-1)l}\\ &\le CN^{\epsilon-\frac12}\,a^{1/2}\,(1+t_{\mathrm{fin}})^{\frac{3q+1}{2q}}\,2^{(d/q'-1)L}.\end{aligned} \tag{3.237}$$

Step 3. Control of the Remainder Integrals Following the argument of Lemma 3.21, we wish to control the error terms $R^{N,f}_t$ given by (3.163), locally uniformly in time. The same bound (3.175) we found for $\beta(f)$ still applies, and we argue as in (3.176), now dropping the time-dependent factor $e^{-\lambda(t-s)/2}\le 1$ in the bound for $\|\phi_{t-s}(\mu^N)-\phi_{t-s}(\mu^N_{s-})\|_{\mathrm{TV}+2}$. We thus end up with, for $m^N+\overline{m}^N$-almost all (s,μ^N),

$$\begin{aligned}\sup_{f\in\mathcal{A}_1}\left|\langle\beta(f),\phi_{t-s}(\mu^N)-\phi_{t-s}(\mu^N_{s-})\rangle\right| &\le C(2^{-2J}+2^{-L})\,N^{\epsilon-1}\,\Lambda_p(\mu^N_{s-})^{\frac12}\\ &=: H'_s.\end{aligned} \tag{3.238}$$

As in (3.177), we may bound everything by considering the two (positive) measures separately, and taking $t=t_{\mathrm{fin}}$:

$$\left\|\sup_{f\in\mathcal{A}_1}\sup_{t\le t_{\mathrm{fin}}}\left|R^{N,f}_t\right|\right\|_{L^q(\mathbb{P})} \le \left\|\int_0^{t_{\mathrm{fin}}}H'_s(m^N+\overline{m}^N)(ds,\mathcal{S}_N)\right\|_{L^q(\mathbb{P})} \le \mathcal{T}_1+\mathcal{T}_2 \tag{3.239}$$

where the two error terms are

$$\mathcal{T}_1 = \left\|\int_0^{t_{\mathrm{fin}}}H'_s\,m^N(ds,\mathcal{S}_N)\right\|_{L^q(\mathbb{P})} \tag{3.240}$$

and

$$\mathcal{T}_2 = \left\| \int_0^{t_{\rm fin}} H'_s \, \overline{m}^N(ds, \mathcal{S}_N) \right\|_{L^q(\mathbb{P})}. \tag{3.241}$$

We now deal with the two terms separately. For $\mathcal{T}_1$, we dominate $\overline{m}^N(ds, \mathcal{S}_N) \le 2N ds$ to see that

$$\int_0^{t_{\rm fin}} H'_s \, \overline{m}^N(ds, \mathcal{S}_N) \le C(2^{-2J} + 2^{-L}) \, N^\epsilon \, t_{\rm fin} \left(\sup_{s \le t_{\rm fin}} \Lambda_p(\mu^N_s)^{\frac{1}{2}} \right). \tag{3.242}$$

Using the monotonicity of $L^q(\mathbb{P})$ norms, and using the moment control in the usual way,

$$\begin{aligned} \mathcal{T}_1 &\le C(2^{-2J} + 2^{-L}) \, N^\epsilon \, t_{\rm fin} \, \mathbb{E}\left[\sup_{s \le t_{\rm fin}} \Lambda_{pq}(\mu^N_s) \right]^{\frac{1}{2q}} \\ &\le C(2^{-2J} + 2^{-L}) \, N^\epsilon \, a^{1/2} \, (1 + t_{\rm fin})^{\frac{2q+1}{2q}}. \end{aligned} \tag{3.243}$$

For $\mathcal{T}_2$, we dominate $m^N(ds, \mathcal{S}_N)$ by a Poisson random measure $\mathfrak{m}^N(ds)$ of rate $2N$, as above. Controlling $\mathfrak{m}^N$ as in (3.231), we obtain

$$\begin{aligned} \mathcal{T}_2 &\le C(2^{-2J} + 2^{-L}) N^{\epsilon - 1} \left\| \int_0^{t_{\rm fin}} \Lambda_p(\mu^N_{s-})^{\frac{1}{2}} \mathfrak{m}^N(ds) \right\|_{L^q(\mathbb{P})} \\ &\le C(2^{-2J} + 2^{-L}) N^{\epsilon - 1} \left\| \left(\sup_{s \le t_{\rm fin}} \Lambda_p(\mu^N_s)^{\frac{1}{2}} \right) \right\|_{L^{2q}(\mathbb{P})} \left\| \mathfrak{m}^N\left((0, t_{\rm fin}]\right) \right\|_{L^{2q}(\mathbb{P})} \\ &\le C(2^{-2J} + 2^{-L}) N^\epsilon a^{1/2} (1 + t_{\rm fin})^{\frac{2q+1}{2q}}. \end{aligned} \tag{3.244}$$

Gathering (3.239, 3.243, 3.244), we have proven that

$$\left\| \sup_{f \in \mathcal{A}_1} \sup_{t \le t_{\rm fin}} \left| R^{N,f}_t \right| \right\|_{L^q(\mathbb{P})} \le C(2^{-2J} + 2^{-L}) N^\varepsilon a^{1/2} (1 + t_{\rm fin})^{\frac{2q+1}{2q}}. \tag{3.245}$$

Step 5. Conclusion Combining the local uniform estimates (3.237, 3.245) of the terms in the decomposition (3.162), we find that

$$\left\| \sup_{f \in \mathcal{A}_1} M^{N,f}_{\star, t_{\rm fin}} \right\|_{L^p(\mathbb{P})} \le C N^\epsilon a^{1/2} \, (1 + t_{\rm fin})^{\frac{3q+1}{2q}} \left(N^{-1/2} \, 2^{(d/q' - 1)L} + 2^{-2J} + 2^{-L} \right).$$

Taking $J \to \infty$ and $L = \lfloor \frac{q'}{2d} \log_2(N) \rfloor$ proves the result claimed. ◻

3.6 Proof of Theorem 3.3

We now turn to the proof of Theorem 3.3, which establishes a convergence estimate in the presence of a p^{th} moment bound, for any $p > 2$. We use the ideas the branching process representation from Section 3.2.1 on a short initial time-interval $[0, u_N]$, for some u_N to be chosen later, and then apply Theorems 3.1, 3.2 on $[u_N, \infty)$ to control the behaviour at times $t \geq u_N$ and the moment production property in Proposition 2.10 to control the moments at time u_N. The argument is similar to the final argument in the proof of the second item of Theorem 3.6 given in Section 3.2, which may be read as a warm-up to this proof.

Throughout, let $p, a, (\mu_t^N), \mu_0$ be as in the statement of the Theorem, and assume without loss of generality that $p \in (2,3)$, and let p_1 be large enough that Theorem 3.1 holds with $\epsilon = \frac{1}{2d}$.

Proof of Theorem 3.3. As in Proposition 3.10i), set $\rho_t = (\mu_t^N + \phi_t(\mu_0))/2$, and write $f_{st} = f_{st}[\rho]$ for the functions obtained from the branching process with this environment. We first introduce a localisation argument, as in the proof of Theorem 3.6, which allows us to guarantee that the conditions (3.38) holds for the environment. Let $\beta = \frac{p-2}{2} \in (0,1)$, and let $u_N \leq 1$ be chosen later. Now, define a new environment

$$\rho_t^T = \rho_t \mathbb{1}\left\{\Lambda_3(\rho_t) \leq \beta t^{\beta-1} + 1\right\} \tag{3.246}$$

and write

$$T_N = \inf\left\{t \geq 0 : \Lambda_3(\rho_t) > \beta t^{\beta-1} + 1\right\} = \inf\left\{t \geq 0 : \rho_t^T \neq \rho_t\right\}. \tag{3.247}$$

We write $f_{st}^T = f_{st}[\rho^T]$ for the functions defined in Section 3.2.1 for this environment, and $M_t^{N,f}, \widetilde{M}_t^{N,f} = \widetilde{M}_t^{N,f}[\rho^T]$ for the stochastic integrals controlled by Proposition 3.11 in the environments ρ, ρ^T respectively, according to (3.35, 3.39). We also remark that, by construction, Proposition 3.11 applies to the modified environment ρ^T with $w = 1$. We will show that a good bound applies on the event $\{T > u_N\}$, and show that u_N may be chosen so that the event $\{T \leq u_N\}$ has a low probability.

Step 1. Control on $\{T_N > u_N\}$. We first control the event with $\{T_N > u\}$. On this event we have the equality $f_{st}^T = f_{st}$ for all $f \in \mathcal{A}_1, s \leq t \leq u_N$, which implies that $M_t^{N,f} = \widetilde{M}_t^{N,f}$ for all $t \leq u_N$. Since the quantities $z_t[\rho^T]$ are bounded in $L^\infty(\mathbb{P})$ on compact time regions, there exists an absolute constant C such that, by the representation formula in Proposition 3.10i), for all $t \leq u_N$,

$$W_1\left(\mu_t^N, \phi_t(\mu_0)\right) \mathbb{1}[T_N > u_N] \leq C \left(W_1(\mu_0^N, \mu_0) + \sup_{f \in \mathcal{A}_1} \widetilde{M}_t^{N,f} \right) \tag{3.248}$$

and we apply Proposition 3.11 to take the expectation of the right-hand side. We thus obtain the estimate

$$\left\|\sup_{t\le u_N} W_1\left(\mu^N_t,\phi_t(\mu_0)\right)\mathbb{1}[T_N>u_N]\right\|_{L^1(\mathbb{P})}\le C\left(W_1(\mu^N_0,\mu_0)+aN^{-\eta}\right) \tag{3.249}$$

where $\eta>0$ is as in Proposition 3.11 for our choice of β. Still on the event $\{T_N>u_N\}$, the stability estimate Theorem 3.6 gives, for some $\zeta_1>0$,

$$\sup_{t\ge u_N} W_1\left(\phi_{t-u_N}(\mu^N_{u_N}),\phi_t(\mu_0)\right)\mathbb{1}[T_N>u_N]\le C\left(W_1\left(\mu^N_{u_N},\phi_{u_N}(\mu_0)\right)\mathbb{1}[T_N>u_N]\right)^{\zeta} \cdots\times\Lambda_{p_1}\left(\mu^N_{u_N},\phi_{u_N}(\mu_0)\right). \tag{3.250}$$

Step 2. Control on the event $\{T_N\le u_N\}$. Using the uniform bound $W_1\le 4$ on $\mathcal{S}\times\mathcal{S}$, we always have the bound

$$\sup_{t\le u_N} W_1\left(\mu^N_t,\phi_t(\mu_0)\right)\mathbb{1}[T_N\le u_N]\le 4\cdot\mathbb{1}[T_N\le u_N] \tag{3.251}$$

and similarly

$$\sup_{t\ge u_N} W_1\left(\phi_{t-u_N}(\mu^N_{u_N},\phi_t(\mu_0)\right)\mathbb{1}[T\le u_N]\le 4\cdot\mathbb{1}[T\le u_N] \tag{3.252}$$

Step 3. Uniform-in-Time Control We recall that we defined p_1 to be large enough that Theorem 3.1 applies with the exponent $-1/2d$. Applying this at time u_N, and using the moment production property, we obtain

$$\begin{aligned}\sup_{t\ge u_N}\left\|W_1(\mu^N_t,\phi_{t-u_N}(\mu^N_{u_N}))\right\|_{L^2(\mathbb{P})}&\le CN^{-1/2d}\,\mathbb{E}\left[\Lambda_{p_1}(\mu^N_{u_N})\right]^{1/2}\\&\le CN^{-1/2d}u_N^{1-p_1/2}.\end{aligned} \tag{3.253}$$

Together with (3.249, 3.251) on the initial interval $[0,u_N]$ and using (3.250, 3.252) to control the difference between $\phi_{t-u_N}(\mu^N_{u_N})$ and $\phi_t(\mu_0)$, we have the overall bound

$$\begin{aligned}\sup_{t\ge 0}\left\|W_1\left(\mu^N_t,\phi_t(\mu_0)\right)\right\|_{L^1(\mathbb{P})}\le C\Bigg(&W_1(\mu^N_0,\mu_0)+aN^{-\eta}+N^{-1/2d}u_N^{1-p_1/2}+\mathbb{P}\left(T_N\le u_N\right)\\&+\mathbb{E}\left[(W_1(\mu^N_{u_N},\phi_{u_N}(\mu_0))\mathbb{1}[T_N>u_N])^{\zeta}\Lambda_{p_1}(\mu^N_{u_N},\phi_{u_N}(\mu_0))\right]\Bigg).\end{aligned} \tag{3.254}$$

To bound the term in the second line, we use Hölder's inequality and (3.249) again to get

$$\begin{aligned}&\mathbb{E}\left[(W_1(\mu^N_{u_N},\phi_{u_N}(\mu_0))\mathbb{1}[T_N>u_N])^{\zeta}\Lambda_{p_1}(\mu^N_{u_N},\phi_{u_N}(\mu_0))\right]\\&\qquad\le C\mathbb{E}\left(W_1(\mu^N_{u_N},\phi_{u_N}(\mu_0))\mathbb{1}[T_N\ge u_N]\right)^{\zeta}\mathbb{E}\left(\Lambda_{p_2}(\mu^N_{u_N},\phi_{u_N}(\mu_0))\right)\\&\qquad\le C(N^{-\eta}+W_1(\mu^N_0,\mu_0))^{\zeta}\,u_N^{2-p_2}\end{aligned} \tag{3.255}$$

where $p_2 = p_1/(1-\zeta)$. Returning to (3.254) and keeping the worst asymptotics, we find that, up to a new choice of $\eta > 0$, we have

$$\sup_{t\geq 0} \left\| W_1\left(\mu_t^N, \phi_t(\mu_0)\right) \right\|_{L^1(\mathbb{P})} \leq Ca\left(\left(W_1(\mu_0^N, \mu_0)^\zeta + N^{-\eta}\right) u_N^{2-p_2} + \mathbb{P}\left(T_N \leq u_N\right)\right). \quad (3.256)$$

Step 4. Estimate of $\mathbb{P}(T_N \leq u_N)$ & Choice of u_N. We now show how u_N may be chosen so that all the terms in (3.256) go to 0 at a rate $N^{-\varepsilon}, W_1(\mu_0^N, \mu_0)^\varepsilon$, and start by bounding $\mathbb{P}(T_N \leq u_N)$. Recalling the definition $\beta = \frac{p-2}{2}$ and that $p \in (2,3)$, let Z_N be given by

$$Z_N = \sum_{l: 2^{-l} \leq u_N} 2^{(\beta-1)l+1-\beta}\beta^{-1} \sup_{t\in[2^{-l}, 2^{1-l}]} \Lambda_3(\rho_t) \quad (3.257)$$

and observe that, for all $t \leq u_N$, we have the bound

$$\Lambda_3(\rho_t) \leq (\beta t^{\beta-1} + 1) Z_N. \quad (3.258)$$

Therefore, recalling the definition of T_N, we have that

$$\mathbb{P}(T_N \leq u_N) \leq \mathbb{P}(Z_N > 1) \leq \mathbb{E}[Z_N]. \quad (3.259)$$

Using the moment production property of the Kac process and Boltzmann equation in Proposition 2.10i-ii), we find, for some C,

$$\mathbb{E}\left[\sup_{t\in[2^{-l}, 2^{1-l}]} \Lambda_3(\mu_t^N)\right] \leq C\mathbb{E}\left[\Lambda_3(\mu_{2^{-l}}^N)\right] \leq C 2^{l(3-p)} a \quad (3.260)$$

and hence, using the same for the Boltzmann equation from Proposition 2.6,

$$\mathbb{E}(Z_N) \leq C \sum_{l: 2^{-l} \leq u_N} 2^{(\beta-1)l} 2^{-l(p-3)} a = C \sum_{l: 2^{-l} \leq u_N} 2^{-(p-2-\beta)l} a. \quad (3.261)$$

The last sum converges, since $p - 2 - \beta = \beta > 0$. Indeed, since the sum ranges over $l \geq \lceil -\log_2 u_N \rceil$, we find, for a new C,

$$\mathbb{P}(T_N \leq u_N) \leq \mathbb{E}[Z_N] \leq C \sum_{l=\lceil -\log_2 u_N \rceil}^{\infty} 2^{-\beta l} a = C 2^{-\beta\lceil -\log_2 u_N \rceil} a \leq C a u_N^\beta. \quad (3.262)$$

Returning to (3.256), we have proven that

$$\sup_{t\geq 0} \left\| W_1(\mu_t^N, \phi_t(\mu_0) \right\|_{L^1(\mathbb{P})} \leq Ca\left((N^{-\eta} + W_1(\mu_0^N, \mu_0)^\zeta) u_N^{2-p_2} + u_N^\beta\right). \quad (3.263)$$

If we now choose

$$u_N = (N^{-\eta} + W_1(\mu_0^N, \mu_0)^\zeta)^{1/(p_2+\beta-2)} \quad (3.264)$$

then both terms are of the same order, and we get

$$\begin{aligned} \sup_{t\geq 0} \left\| W_1(\mu_t^N, \phi_t(\mu_0)) \right\|_{L^1(\mathbb{P})} &\leq Ca(N^{-\eta} + W_1(\mu_0^N, \mu_0)^\zeta)^{\beta/(p_2+\beta-2)} \\ &\leq Ca\left(N^{-\eta\beta/(p_2+\beta-2)} + W_1(\mu_0^N, \mu_0)^{\zeta\beta/(p_2+\beta-2)}\right) \end{aligned} \quad (3.265)$$

which proves the desired result. The case for the local uniform estimate is similar, using Theorem 3.2 in place of Theorem 3.1. □

3.7 Proof of Theorem 3.4

The proof of Theorem 3.4 is based on the following heuristic argument. Fix N, and consider a Kac process (μ_t^N) on N particles. As $t \to \infty$, its law relaxes to the equilibrium distribution π_N, which is known to be the uniform distribution σ^N on $\mathcal{S}_N$. Since this measure assigns non-zero probability to regions R_N at macroscopic distance from the fixed point γ, given by

$$\gamma(dv) = \frac{e^{-\frac{d}{2}|v|^2}}{(2\pi d^{-1})^{d/2}} dv, \tag{3.266}$$

the process will almost surely hit R_N on an unbounded set of times. Meanwhile, the Boltzmann flow $\phi_t(\mu_0)$ will converge to γ. Therefore, at some large time, the particle system μ_t^N will have macroscopic distance from the Boltzmann flow $\phi_t(\mu_0^N)$.

It will be slightly more convenient here to work with labelled processes in order to make contact with the literature on the relaxation of the particle system. We recall that a labelled Kac process is the Markov process of velocities $\mathcal{V}_t^N = (V_t^1, ..., V_t^N)$ corresponding to the particle dynamics, in labelled Boltzmann sphere

$$\mathbb{S}^N = \left\{(v_1, ..., v_N) \in (\mathbb{R}^d)^N : \sum_{i=1}^N v_i = 0, \sum_{i=1}^N |v_i|^2 = N\right\}$$

and that we recover the unlabelled Kac process and unlabelled Boltzmann sphere via the map θ_N taking an N-tuple to its normalised empirical measure in $\mathcal{S}_N$.

Considered as a $((N-1)d-1)$-dimensional sphere in $(\mathbb{R}^d)^N$, $\mathbb{S}^N$ has a uniform (Hausdorff) distribution γ^N, which we push forward to a measure σ^N on $\mathcal{S}_N$:

$$\sigma^N(A) := \gamma^N \left\{(v_1, ...v_N) \in \mathbb{S}^d : \theta_N(v_1, ..., v_N) \in A\right\}. \tag{3.267}$$

We will use this definition to transfer the positivity of the measure γ^N forward to σ^N. We also note that, since γ^N is the equilbrium distribution for the (labelled) Kac process, it follows that σ^N is the equilibrium measure for the (unlabelled) Kac process.

As discussed in the literature review, the problem of relaxation to equilibrium for the Kac process is a subtle problem, and has been extensively studied. For our purposes, the following $L^2(d\sigma^N)$-convergence is sufficient:

Proposition 3.27. *Suppose that $(\mu_t^N)_{t\geq 0}$ is a hard-spheres Kac process, where the law of the initial data* $\text{Law}(\mu_0^N)$ *has a density $h_0^N \in L^2(\sigma^N)$ with respect to σ^N. Then at all positive times $t \geq 0$, the law* $\text{Law}(\mu_t^N)$ *has a density $h_t^N \in L^2(\sigma^N)$ with respect to σ^N, and for some universal constant $\lambda_0 > 0$, we have*

$$\left\|h_t^N - 1\right\|_{L^2(\sigma^N)} \leq e^{-\lambda_0 t} \left\|h_0^N - 1\right\|_{L^2(\sigma^N)}. \tag{3.268}$$

A version of this, for the labelled Kac process, appears as [142, Theorem 6.8 and corollary]; the result stated above follows by a pushforward argument. This is sufficient to prove the following weak ergodic theorem:

Lemma 3.28. *Let $(\mu_t^N)_{t\geq 0}$ be a hard-spheres Kac process on N particles, started from $\mu_0^N \sim \sigma^N$. Let $R_N \subset \mathcal{S}_N$ be such that $p = \sigma^N(R_N) > 0$. Then*

$$\frac{1}{t}\int_0^t \mathbb{1}(\mu_s^N \in R_N)ds \to p \tag{3.269}$$

in $L^2(\mathbb{P})$. In particular, almost surely, μ_t^N visits R_N on an unbounded set of times.

Proof. Observe that, since the process is in equilibrium, the law of μ_t^N is σ^N for all $t \geq 0$, and so

$$\mathbb{E}_{\sigma^N}\left[\frac{1}{t}\int_0^t \mathbb{1}(\mu_s^N \in R_N)ds\right] = \frac{1}{t}\int_0^t \mathbb{P}_{\sigma^N}(\mu_s^N \in R_N)ds = p \tag{3.270}$$

so our claim reduces to bounding the variance.

For times $t \geq 0$, write $A(t)$ as the event $A(t) = \{\mu_t^N \in R_N\}$; we will compute the covariance of the random variables $\mathbb{1}_{A(s_1)}$ and $\mathbb{1}_{A(s_2)}$, for $0 \leq s_1 \leq s_2$. Observe that

$$\mathbb{E}_{\sigma^N}\left[\mathbb{1}_{A(s_1)}(\mathbb{1}_{A(s_2)} - p)\right] = p\left(\mathbb{P}\left(A(s_2)|A(s_1)\right) - p\right). \tag{3.271}$$

Conditional on $A(s_1)$, the law of $\mu_{s_1}^N$ has a conditional density $h_{s_1}^N \propto \mathbb{1}_{R_N}$ with respect to σ^N. By Proposition 3.27, conditional on $A(s_1)$, the law of $\mu_{s_2}^N$ has a density $h_{s_2}^N$, and we can bound

$$\begin{aligned} |\mathbb{P}_{\sigma^N}(A(s_2)|A(s_1)) - p| &\leq \int_{\mathcal{S}_N} |h_{s_2}^N - 1|\mathbb{1}_{R_N}(\mu^N)\sigma^N(d\mu^N) \\ &\leq \left\|h_{s_2}^N - 1\right\|_{L^2(\sigma^N)} \leq 2p^{-1/2}e^{-\lambda_0(s_2-s_1)} \end{aligned} \tag{3.272}$$

since we note $\|h_{s_1}^N\|_{L^2(\sigma_N)} \leq p^{-1/2}$. Hence

$$\mathbb{E}_{\sigma^N}\left[(\mathbb{1}_{A(s_1)} - p)(\mathbb{1}_{A(s_2)} - p)\right] = p(\mathbb{P}(A(s_2)|A(s_1)) - p) \leq 2p^{1/2}e^{-\lambda_0(s_2-s_1)}. \tag{3.273}$$

We can now integrate to bound the variance:

$$\begin{aligned} \mathrm{Var}_{\sigma^N}\left(\frac{1}{t}\int_0^t \mathbb{1}(\mu_s^N \in R_N)ds\right) &= \frac{2}{t^2}\int_0^t ds_1 \int_{s_1}^t ds_2\ \mathbb{E}\left[(\mathbb{1}_{A(s_1)} - p)(\mathbb{1}_{A(s_2)} - p)\right] \\ &\leq \frac{4p^{1/2}}{t^2}\int_0^t ds_1 \int_{s_1}^\infty ds_2\ e^{-\lambda_0(s_2-s_1)} \\ &\leq \frac{4p^{1/2}}{\lambda_0 t} \to 0. \end{aligned} \tag{3.274}$$

For the second assertion, the $L^2(\mathbb{P})$ convergence implies $\mathbb{P}$-almost sure convergence along some subsequence $t_n \to \infty$, and on this event it holds that μ_t^N visits R_N on an unbounded set of times, as claimed. □

An immediate corollary is that the long-run deviation must be bounded *below* by the essential supremum of the deviation under the invariant measure:

Corollary 3.29. *Let $(\mu_t^N)_{t\geq 0}$ be a N- particle Kac process in equilibrium. Then, almost surely,*

$$\limsup_{t\to\infty} W_1(\mu_t^N,\gamma) \geq \|W_1(\cdot,\gamma)\|_{L^\infty(\sigma^N)} = \operatorname*{ess\,sup}_{\sigma^N(d\mu)} W_1(\mu,\gamma). \tag{3.275}$$

Proof. For ease of notation, write W^* as the essential supremum appearing on the right hand side. For any $\epsilon > 0$, let $R_{N,\epsilon} = \{\mu \in \mathcal{S}_N : W_1(\mu,\gamma) > W^* - \epsilon\}$; it is immediate that $\sigma^N(R_{N,\epsilon}) > 0$. By the remark in Lemma 3.28, almost surely, μ_t^N visits $R_{N,\epsilon}$ on an unbounded set of times, and so

$$\limsup_{t\to\infty} W_1(\mu_t^N,\gamma) \geq W^* - \epsilon. \tag{3.276}$$

The conclusion now follows on taking an intersection over some sequence $\epsilon_n \downarrow 0$. □

To prove Theorem 3.4, it now only remains to show a lower bound on the essential supremum.

Lemma 3.30. *Let f be given by*

$$f(v) = (1+|v|^2)\min\left(\frac{|v|}{\sqrt{N/2}},1\right). \tag{3.277}$$

Then $f \in \mathcal{A}$, and

$$\|\langle f,\mu-\gamma\rangle\|_{L^\infty(\sigma^N)} \geq 1 - \frac{C}{\sqrt{N}} \tag{3.278}$$

for some constant $C = C(d)$. In particular, this is a lower bound for the essential supremum W^*, and so for the long-run deviation.

Proof. It is easy to see that $f \in \mathcal{A}_1$. Moreover, the region

$$\widetilde{R}_N = \left\{(v_1,...v_N) \in \mathbb{S}^N : \langle f, \theta_N(v_1,...,v_N)\rangle > 1\right\} = \theta_N^{-1}\left\{\mu \in \mathcal{S}^N : \langle f,\mu\rangle > 1\right\} \tag{3.279}$$

is an open subset of $\mathbb{S}^N$, which contains $\sqrt{N/2}\,(e_1,-e_1,0,..,0)$ and is therefore nonempty. By positivity of the uniform measure γ^N on $\mathbb{S}^N$, it follows that $\gamma^N(\widetilde{R}_N) > 0$. The corresponding region in $\mathcal{S}_N$ is

$$R_N = \{\mu^N \in \mathcal{S}_N : \langle f,\mu^N\rangle > 1\}. \tag{3.280}$$

By definition (3.267) of σ^N, we have

$$\sigma^N(R_N) = \gamma^N(\widetilde{R}_N) > 0. \tag{3.281}$$

For all $\mu^N \in R_N$, we have

$$W_1(\mu^N,\gamma) \geq \langle f,\mu^N - \gamma\rangle \geq 1 - \sqrt{2/N}\langle(1+|v|^2)|v|,\gamma\rangle. \tag{3.282}$$

Since R_N has positive measure, taking $C = \sqrt{2}\langle(1+|v|^2)|v|,\gamma\rangle$, we can conclude that

$$W^* \geq 1 - \frac{C}{\sqrt{N}} \tag{3.283}$$

and we are done. □

Proof of Theorem 3.4. From the previous two lemmas, we know that for all $N \geq 2$, and for σ^N- almost all μ^N,

$$\mathbb{P}_{\mu^N}\left(\limsup_{t\to\infty} W_1(\mu^N_t, \gamma) \geq 1 - \frac{C}{\sqrt{N}}\right) = 1 \tag{3.284}$$

where $\mathbb{P}_{\mu^N}$ denotes the law of a Kac process started at μ^N.

Let $N \geq 2, p > 2$ and $a > 1$. The region $R_{\star,N}$ of the labelled sphere such that $\Lambda_p(\theta_N(\mathcal{V})) < a$ is an open set; to conclude that it has positive σ^N-measure, which is readily seen to be nonempty by constructing $(V^1, .., V^N) \in \mathbb{S}_N$ where each $|V^i| = 1$, and by the positivity of γ^N, we have $\gamma^N(R_{\star,N}) > 0$. The positivity transfers to the corresponding region of $\mathcal{S}_N$:

$$\sigma^N\left\{\mu^N \in \mathcal{S}_N : \Lambda_p(\mu^N) < a\right\} = \gamma^N(R_{N,\star}) > 0. \tag{3.285}$$

Hence, for any $N \geq 2$, we can choose an initial datum $\mu^N_0 = \mu^N$, with $\Lambda_p(\mu^N_0) < a$, such that (3.284) holds. Observing that

$$W_1(\phi_t(\mu^N_0), \gamma) \leq \|\phi_t(\mu^N_0) - \phi_t(\gamma)\|_{\mathrm{TV}+2} \to 0 \tag{3.286}$$

it follows that, $\mathbb{P}_{\mu^N}$- almost surely

$$\limsup_{t\to\infty} W_1(\mu^N_t, \gamma) = \limsup_{t\to\infty} W_1(\mu^N_t, \phi_t(\mu^N_0)) \geq 1 - \frac{C}{\sqrt{N}}. \tag{3.287}$$

□

Remark 3.31. *i). The proof of Lemma 3.28 leaves open the possibility that there is a non-empty 'exceptional set' of initial data μ^N where (3.284) does not hold. A stronger assertion would be positive Harris recurrence, as defined in [110], which allows a similar ergodic theorem for* any *initial data μ^N. This is not necessary for our purposes.*

ii). Such 'highly unlikley' events, where the Boltzmann equation fails by a macroscopic margin, are the purview of Large Deviation theory, which we explore in Chapter 6. Instead of the 'bad sets' above, we could have considered regions $R_{N,\varepsilon}$ given by

$$R_{N,e} = \{\mathcal{V} \in \mathbb{S}_N : W_1(\theta_N(\mathcal{V}), \nu) < \varepsilon\}$$

for any $\nu \in \mathcal{S}$ and $\varepsilon > 0$; using the same ideas as Proposition 2.4, these are nonempty for N sufficiently large. In this case, we get the asymptotic

$$p_N = \gamma^N(R_{N,\varepsilon}) \geq e^{-N(H(\nu|\gamma)+\varepsilon)}$$

for N large enough, depending on $\varepsilon > 0$, where $H(\cdot|\gamma)$ is the relative entropy by Sanov's Theorem (see Section 1.2.4). In particular, the same proof as above shows *that, starting from equilibrium, we reach these regions on time-scales at most of the order* $e^{3N(H(\nu|\gamma)+\delta)/2}$ *for any* $\delta > 0$*, provided that* $H(\nu|\gamma) < \infty$*.*

In Chapter 6, we will also see behaviour where macroscopic proportion of the energy concentrates in $o(N)$ *particles, with probability* $\sim e^{-cN}$ *under* $\gamma^{\otimes N}$*. Pushing forward by* θ_N *produces regions*

$$R_{N,\varepsilon} = \left\{\mathcal{V}^N \in \mathbb{S}_N : \mathcal{W}_{1,1}(\theta_N(\mathcal{V}^N), \gamma^-) < \varepsilon\right\}$$

for γ^- *a Maxwellian of average energy* $\langle |v|^2, \gamma^-\rangle < 1$ *and some* $\varepsilon > 0$*, with* $\gamma^N(R_{N,\varepsilon}) \sim e^{-cN}$ *for some* $c < \infty$*, which is the same kind of behaviour we saw in the proof above. It follows that this kind of behaviour also occurs on scales at most exponentially long in* N*.*

Appendix

3.A Calculus of Martingales

We also review some basic facts and inequalities for martingales associated to the Kac process. All of these facts are true for general Markov chains, see [49].

Let μ^N_t be a Kac process, and write m^N, $\overline{m}^N$ for the jump measure and compensator in Definition 3.2.2. Then, for any bounded and measurable $F^N : [0, t_{\text{fin}}] \times \mathcal{S}_N \to \mathbb{R}$, the process

$$\mathcal{M}^N_t = \int_{(0,t]\times\mathcal{S}_N} \left\{F^N_s(\mu^N) - F^N_s(\mu^N_{s-})\right\}(m^N - \overline{m}^N)(ds, d\mu^N), \quad 0 \le t \le t_{\text{fin}} \tag{3.288}$$

is a martingale for the natural filtration $(\mathfrak{F}^N_t)_{t\ge 0}$ of the process. We have the $L^2(\mathbb{P})$ control

$$\|\mathcal{M}^N_t\|_2^2 = \mathbb{E}\left\{\int_{(0,t]\times\mathcal{S}_N} \left\{F^N_s(\mu^N) - F^N_s(\mu^N_{s-})\right\}^2 \overline{m}^N(ds, d\mu^N)\right\}. \tag{3.289}$$

We will also use another special case of Itô's isometry for the measure $m^N - \overline{m}^N$ for a similar form of martingale. If F^N is bounded and measurable on $[0,T]\times\mathcal{S}_N$, then for $t \le t_{\text{fin}}$,

$$\left\|\int_0^t F^N_s(\mu^N_{s-})(m^N - \overline{m}^N)(ds, \mathcal{S}_N)\right\|_2^2 = \mathbb{E}\left\{\int_0^t F^N_s(\mu^N_s)^2\, \overline{m}^N(ds, \mathcal{S}_N)\right\}. \tag{3.290}$$

For the local uniform case, Theorem 3.2, it will be necessary to control martingales of the form (3.288) in general $L^q(\mathbb{P})$ spaces, rather than simply $L^2(\mathbb{P})$. Since $\mathcal{M}^N$ of this form are finite variation martingales, the quadratic variation is given by

$$\left[\mathcal{M}^N\right]_t = \int_{(0,t]\times\mathcal{S}_N} \left\{F^N_s(\mu^N) - F^N_s(\mu^N_{s-})\right\}^2 m^N(ds, d\mu^N), \quad 0 \le t \le t_{\text{fin}}. \tag{3.291}$$

Our analysis in $L^q(\mathbb{P})$ is based on Burkholder's inequality for càdlàg martingales, which we state here for the class of martingales constructed above:

Lemma 3.32. *Suppose that $(\mathcal{M}^N_t)_{t=0}^{t_{\text{fin}}}$ is the process given by (3.288), and let $q \ge 2$. Then there exists a constant $C = C(p) < \infty$ such that for all $t \le t_{\text{fin}}$, we have the $L^q(\mathbb{P})$ control*

$$\left\|\sup_{s\le t}\left|\mathcal{M}^N_s\right|\right\|_{L^q(\mathbb{P})}^q \le C(q)\mathbb{E}\left[\left(\int_0^t \left\{F^N_s(\mu^N) - F^N_s(\mu^N_{s-})\right\}^2 m^N(ds, d\mu^N)\right)^{q/2}\right]. \tag{3.292}$$

Chapter 4

Non-Cutoff Hard Potentials with a Moderate Angular Singularity

4.1 Main Results

In this chapter, we consider the case of the Boltzmann equation and Kac process in the case of a noncutoff hard potential kernel B of the form (NCHP), and will prove the assertions of Theorem 2. Throughout, we assume the notation of (NCHP) and work in the φ, z parametrisation (see Lemma 2.5); objects defined for kernels with Grad's angular cutoff will be denoted with a subscript or superscript $\cdot^K, \cdot_K$.

The central approach of this chapter, summarised in Theorems 4.1, 4.4, is a coupling argument, which allows us to couple either Boltzmann processes (stBE) in the spirit of Tanaka's stochastic interpretation of the Boltzmann equation [177], which we rewrite (stBE^a) in Section 4.2 using the parametrisation in terms of z, φ given in Section 2.4. The same coupling also allows us to obtain any noncutoff Kac process μ^N_t on N as the limit of cutoff process $\mu^{N,K}_t$ of cutoff processes as $K \to \infty$, with a rate uniform in N. The coupling of Boltzmann processes will imply a similar result for the cutoff and non-cutoff Boltzmann equations described below, and the coupling of Kac processes implies the well-posedness of the martingale problem for the generator $\mathcal{G}^N$ given by (1.31) for the unlabelled Kac process and the convergence of the Kac process.

In this chapter in particular, we will also work with labelled Kac processes $\mathcal{V}^N_t, \mathcal{V}^{N,K}_t$, as this is natural for the coupling: we pair the i^{th} particle in each of the two systems. Since all of our results are naturally phrased at the level of the empirical measures, we will move between labelled and unlabelled processes $\mathcal{V}^N_t, \mu^N_t$, which is justified by Proposition 4.7.

As remarked below (stBE), for Boltzmann processes and for all $f : \mathbb{R}^d \to \mathbb{R}$ Lipschitz, the process $f(V_t) - f(V_0) - \int_0^t \int_{\mathbb{R}^d} \mathcal{L}_{\mathrm{B}} f(v, v_*) \mu_s(dv_*) ds$ is a martingale, where $\mathcal{L}_{\mathrm{B}}$ is the non-local generator given by (1.19); we will make frequent use of the fact that this characterises solutions to (stBE), see [70]. To avoid repeatedly writing subscripts, as there is no possibility of confusion, we write $\mathcal{L} = \mathcal{L}_{\mathrm{B}}$ throughout this chapter.

We now give precise formulations of our main results, which correspond to Theorem 2. Our first result gives a coupling of solutions to Tanaka's SDE (stBE) or its equivalent form (stBE^a), which gives a quantitative stability result for solutions to the Boltzmann equation. We work throughout with the optimal transportation costs w_p defined in Section 2.1, which are equivalent to $\mathcal{W}_{p+2}$ or to weak convergence, plus convergence of moments of orders up to $p+2$, and which are defined as the optimal transport cost for the function $d_p : \mathbb{R}^d \times \mathbb{R}^d \to [0, \infty)$.

Theorem 4.1. *Let B be a kernel of the form (NCHP). There exists $p_0 = p_0(B, d)$ and, for $p > p_0$, $K_0 = K_0(B, p, d)$ such that, whenever $p > p_0$ and $K \geq K_0$, there exists $C = C(B, p, d)$ such that the following holds.*

Let $(\mu_t)_{t\geq 0}, (\nu_t)_{t\geq 0}$ be weak solutions to (BE, BE$_K$) respectively, with $K \in [K_0, \infty]$, starting at $\mu_0, \nu_0 \in \mathcal{S}^{p+2}$, and for some $a_1 \geq 1$, we have $\Lambda_{p+\gamma}(\mu_0, \nu_0) \leq a_1$. If $K < \infty$, let us assume further that, for some $a_2 \geq 1$, we have the moment estimate $\Lambda_l(\nu_0) \leq a_2$. Then there exists a stochastic process $(V_t, \tilde{V}_t)_{t\geq 0}$ such that $\pi_t = \mathrm{Law}(V_t, \tilde{V}_t)$ is a coupling $\pi_t \in \Pi(\mu_t, \nu_t)$ for all $t \geq 0$, and such that $V_t, \tilde{V}_t$ solve (stBE^a, stBE$^a{}_K$) respectively. Furthermore, for some constant $C = C(B, p, d)$, we have the estimates, for $K \in [K_0, \infty)$,

$$\mathbb{E}[d_p(V_t, \tilde{V}_t)] \leq e^{Ca_1(1+t)} \left(w_p(\mu_0, \nu_0) + a_2 t K^{1-1/\nu}\right) \tag{4.1}$$

or, if $K = \infty$, then

$$\mathbb{E}[d_p(V_t, \tilde{V}_t)] \leq e^{Ca_1(1+t)} w_p(\mu_0, \nu_0). \tag{4.2}$$

Since $(V_t, \tilde{V}_t)$ produces a coupling of μ_t, ν_t for all $t \geq 0$, it follows that $\mathbb{E}[d_p(V_t, \tilde{V}_t)] \geq w_p(\mu_t, \nu_t)$, and we obtain the following as a corollary.

Corollary 4.2 (Wasserstein Stability of the Boltzmann Flow)**.** *Let B be a kernel of the form (NCHP), and let $p > p_0$ $K_0 = K_0(B, p, d)$, $l = p + 2 + \gamma$ be as in Theorem 4.4. For any $\mu_0 \in \mathcal{S}^{p+2}$, there exists a unique energy-conserving solution $(\mu_t)_{t\geq 0}$ to the Boltzmann equation starting at μ_0, which we write as $\mu_t = \phi_t(\mu_0)$. Moreover, for some constant $C = C(B, p, d)$,*

i). Whenever $\mu_0, \nu_0 \in \mathcal{S}^{p+2}$ satisfy the moment bound $\Lambda_{p+\gamma}(\mu_0, \nu_0) \leq a$, we have the continuity estimate

$$w_p\left(\phi_t(\mu_0), \phi_t(\nu_0)\right) \leq e^{Ca(1+t)} w_p(\mu_0, \nu_0). \tag{4.3}$$

ii). Whenever $\mu_0 \in \mathcal{S}^l$, the solution $\phi_t(\mu)$ is the w_p-limit of the solutions $\phi^K_t(\mu_0)$ to the K-cutoff Boltzmann Equations (BE$_K$) starting at μ_0, as the cutoff parameter

$K \to \infty$. More precisely, if $\Lambda_{p+\gamma}(\mu_0) \le a_1, \Lambda_l(\mu_0) \le a_2$, for some $a_1, a_2 \ge 1$, then for $K \ge K_0$ we have

$$w_p(\phi_t^K(\mu_0), \phi_t(\mu_0)) \le e^{Ca_1(1+t)} a_2 t K^{1-1/\nu}. \tag{4.4}$$

As well as a direct deduction from Theorem 4.1, we will also offer a different proof in Section 4.9, which uses the coupling of Kac processes in Theorem 4.4 below as an intermediate step. We will discuss the two possible approaches below. We also note that this is a much stronger well-posedness estimate than exists in the literature; see the discussion in the literature review below. Using the machinery developed in the proof of Theorem 4.1, we also rederive the following result of Fournier [88], which shows that Tanaka's SDE (stBE, stBE^a) is equivalent to (BE). Although it is not our main result, it is satisfying to know that the two are (unconditionally) equivalent.

Theorem 4.3. *Let B be a kernel of the form (NCHP). Let $(V_t)_{t\ge0}$ be a solution to (stBE^a) or (stBE$^a{}_K$) with $\mu_t = \mathrm{Law}(V_t) \in \mathcal{S}$ for all $t \ge 0$. Then $(\mu_t)_{t\ge0}$ is a solution to (BE) or (BE$_K$) respectively. Conversely, if $(\mu_t)_{t\ge0} \subset \mathcal{S}$ is a solution to (BE, BE$_K$), then there exists a solution $(V_t)_{t\ge0}$ to (stBE^a, stBE$^a{}_K$) respectively with $\mathrm{Law}(V_t) = \mu_t$ for all $t \ge 0$.*

In the same spirit as Theorem 4.1, we will exhibit a coupling of the many-particle Kac process, uniformly in N.

Theorem 4.4 (Tanaka Coupling of Kac Processes)**.** *Let B be a kernel of the form (NCHP) and let $p \ge p_0, K \in [K_0, \infty], l = p+2+\gamma$ be as in Theorem 4.1. Fix $a_1, a_2, a_3 \ge 1$, $N \ge 2$, and let $\mu_0^N, \widetilde{\mu}_0^{N,K} \in \mathcal{S}_N$ be empirical measures satisfying*

$$\Lambda_{p+\gamma}\left(\mu_0^N, \widetilde{\mu}_0^{N,K}\right) \le a_1; \qquad \Lambda_l\left(\mu_0^N, \widetilde{\mu}_0^{N,K}\right) \le a_2; \qquad \Lambda_q\left(\mu_0^N, \widetilde{\mu}_0^{N,K}\right) \le a_3 \tag{4.5}$$

where $l = p+2+\gamma, q = 2p+4+2\gamma$. If $K < \infty$, there exists a coupling of a noncutoff Kac process μ_t^N starting at μ_0^N and a K-cutoff Kac process $\widetilde{\mu}_t^{N,K}$ starting at $\widetilde{\mu}_0^{N,K}$ such that, for all $t \ge 0$,

$$\mathbb{E}\left[w_p\left(\mu_t^N, \widetilde{\mu}_t^{N,K}\right)\right] \le e^{Ca_1(1+t)}\left(w_p\left(\mu_0^N, \widetilde{\mu}_0^{N,K}\right) + a_2 K^{1-1/\nu}\right) + Ca_3^2 t N^{-1/2} \tag{4.6}$$

and, for all $t_{\text{fin}} \ge 0$,

$$\mathbb{E}\left[\sup_{t\le t_{\text{fin}}} w_p\left(\mu_t^N, \widetilde{\mu}_t^{N,K}\right)\right] \le e^{Ca_1(1+t_{\text{fin}})}\left(w_p\left(\mu_0^N, \widetilde{\mu}_0^{N,K}\right) + a_2 t K^{1-1/\nu} \cdots + \frac{Ca_3^2(1+t_{\text{fin}})^2}{N^{1/2}}\right). \tag{4.7}$$

The same is true with a coupling of noncutoff Kac processes if $K = \infty$, omitting the terms proportional to $K^{1-1/\nu}$.

Our final result is to study the convergence of the noncutoff Kac process in the large number limit $N \to \infty$.

Theorem 4.5. *Let B be a kernel of the form (NCHP). For all N, the N-particle Kac process defined by the generator (1.31) has uniqueness in law. Moreover, if p, q are as in Theorem 4.4, $a \geq 1$ and $t_{\text{fin}} \geq 0$, then whenever $\mu_0 \in \mathcal{S}^q$ has a moment $\Lambda_q(\mu_0) \leq a$ and μ^N_t is a N-particle Kac process with initial data satisfying $\Lambda_q(\mu^N_0) \leq a$ almost surely then we have the estimate*

$$\mathbb{E}\left[\sup_{t\leq t_{\text{fin}}} w_p\left(\phi_t(\mu_0), \mu^N_t\right)\right] \leq e^{Ca(1+t_{\text{fin}})}\left((\log N)^{1-1/\nu} + \mathbb{E}\left[w_p(\mu^N_0, \mu_0)\right]\right). \tag{4.8}$$

Thanks to the general considerations in Chapter 2, this proves propagation of chaos on compact time intervals, with an explicitable rate.

4.1.1 Plan of the Chapter

Our programme will be as follows.

i). The remainder of this section is a discussion, comparing the results of this chapter to the existing results and techniques from the literature and to the other chapters.

ii). In Section 4.2, we re-write the definition (stBE) of Boltzmann processes in terms of the coefficients $a(v, v_*, z, \varphi)$ introduced in Section 2.4, which is natural for our coupling arguments. We similarly formulate (labelled) Kac processes as solutions to stochastic differential equation in $(\mathbb{R}^d)^N$ driven by Poisson random measures, and state results on the well-posedness of the resulting stochastic differential equation and its relationship to the unlabelled Kac process; the proofs are deferred to Appendix 4.A for ease of readability.

iii). Section 4.3 introduces the key 'Tanaka trick' (Lemma 4.8), which compensates for the fact that the coefficients a parametrising the jumps are not continuous. We present the key 'Tanaka-Povzner' calculation (Lemma 4.10), which will be used in both the coupling of Boltzmann processes and the Kac process. The proof is deferred until Section 4.10 for ease of readability.

iv). In Section 4.4, we will check some properties of Boltzmann processes given by the stochastic differential equations (stBE, stBE^a) for later use.

v). Section 4.5 is dedicated to the construction of the processes $(V_t, \tilde{V}_t)_{t\geq 0}$, which we characterise by a nonlinear jump SDE (4.50).

vi). Section 4.6 gives the proof of Theorem 4.1, applying the Tanaka-Povzner estimate to the coupling produced in Section 4.5.

vii). Section 4.7 gives the proof of Theorem 4.3, based on tools which have already been developed in the proof of Theorem 4.1.

viii). In Section 4.8, we apply the Tanaka coupling to the many-particle Kac process. We prove Theorem 4.4 and, by comparing with the convergence of cutoff Kac processes in Chapter 3, we deduce Theorem 4.5.

ix). In Section 4.9, we offer an alternative proof of Corollary 4.2 which does not use Theorem 4.1, based on the Tanaka coupling of the Kac processes.

x). Section 4.10 presents the main calculations on our Tanaka coupling, deferred from Section 4.3.

xi). Finally, Appendices 4.A, 4.B deal with some technical issues concerning the well-posedness for unlabelled Kac processes deferred from Section 4.2, and a variant on the construction of the coefficients, which we use in constructing the couplings $(V_t, \tilde{V}_t)$.

4.1.2 Literature Review

We will now briefly discuss related works and their relationship to our work.

1. Tanaka's Coupling. The key idea in this chapter and the next is a coupling pioneered by Tanaka [177] in the case of Maxwell Molecules (GMM) in terms of the stochastic differential equation (stBE). This was generalised by Fournier and Méléard [91, 79] to include the cases without cutoff, and for non-Maxwellian molecules and used to show uniqueness for the Boltzmann equation with Maxwell molecules [176].

Even without cutoff, this case is significantly easier, because the same arguments as in Section 4.3 give a 'compensated Lipschitz property' for the maps $(v, v_*) \to a(v, v_*, z, \varphi)$, uniformly in z, φ, see [79, Lemma 2.6]. In terms of the parametrisation in Section 2.4, this is because the deflection angle $\theta(v, v_*, z)$ is now a function only of z. The classical result of Tanaka [177] is equivalent to Theorem 4.1 above, now omitting the exponential factor and with $p = 0, w_0 = \mathcal{W}_2$. In general, as soon as $\gamma > 0$, this additional dependence introduces further error terms into the coupling estimates, which cannot be estimates using the Grönwall lemma; in our calculation, these are the terms $\mathcal{T}_2, \mathcal{T}_3$ in Section 4.10, and see also the proof of Proposition 4.18. In this case, we need $p > 2$ potentially large to cancel these terms with our 'Tanaka-Pozvner calculation'.

Let us mention two particular works to which our approach can be compared. The main calculations in Sections 4.3, 4.10 were inspired by Fournier and Mischler [92] on the Nanbu particle system, in which only one particle jumps at a time. In our notation, the cited paper produces estimates in w_0; the major novelty of this work is that, by working in w_p for p large enough, and exploiting symmetries of the Tanaka or Kac processes, we are able to obtain a desirable cancellation of 'bad' terms. Let us also remark that that the main result of Rousset [167] is very similar to Theorem 4.4 in obtaining a coupling of Kac processes with error uniform in N in the case of Maxwell molecules. In this case, one can find an additional negative term, even without the weighting we use here, which corresponds to refining the estimate we use on $\mathcal{T}_2$. For our cases, this is not helpful, because the additional negative term is already weaker than $-\mathbb{E}[w_0(\mu_t^N, \mu_t^{N,K})]$, see the remark above [167, Equation 0.11], which are themselves weaker than the terms in our expansion for $\gamma > 0$. Rousset uses this coupling to investigate relaxation to equilibrium for the Maxwell Molecules system in the limit $t \to \infty$, uniformly in N. In our case, we are unable to make the coupling uniform in time to investigate relaxation to equilibrium, but use a similar coupling to prove propagation of chaos.

2. Well-Posedness of the Boltzmann Equation. As mentioned in the introduction, the Cauchy theory for the case of noncutoff hard potentials (NCHP) is substantially less complete than in the case of Maxwell Molecules (GMM) or hard spheres and cutoff hard potentials (HS, CHP_K), see also Proposition 3.15. Fournier [81] examined the case where $|v - v_\star|^\gamma$ is replaced by a bounded function Φ, and results for the case of full hard potentials have been found by Desvillettes and Mouhot [56] and in the case of measure solutions by Fournier and Mouhot [93]. Let us note that the uniqueness and stability statement in Theorem 4.1 assumes only a finite number of moments, rather than a finite exponential moment $\langle e^{\epsilon|v|^\gamma}, \mu_0\rangle < \infty$ as does the result of [93], which is recalled in Proposition 4.18 below; correspondingly, our quantitative stability result is stronger, as we can use Grönwall's lemma rather than the Yudovich lemma [121]. The result of [56] requires the initial data μ_0 to have a density $f_0 \in W^{1,1}(\mathbb{R}^d, (1 + |v|^2)dv)$, and so requires fewer moments than our results but much more regularity.

3. Propagation of Chaos for the Kac Process. Thanks to the general considerations in Chapter 2, we can understand the conclusion of Theorem 4.5 as proving propagation of chaos for the case of non-cutoff hard potentials (NCHP), with an explicit rate. Although the rate is fairly slow, we believe that this is the first such result for these kernels; let us also mention some works on related models. Fournier and Guilin [88] consider a related particle system which approximates the Landau equation for hard potentials, and Fournier [89] deals with this model for soft potentials. The work [92] which we have already mentioned considers the asymmetric *Nanbu process* in which only one particle jumps at a time, and shows propagation of chaos for this system for Maxwell molecules and hard potentials; a recent work of Salem [169] extends this to the case of soft potentials with a moderate angular singularity.

The ideas of this chapter draw primarily from the literature on the Tanaka coupling, including Fournier and Mischler [92], and the proof of Theorem 4.5 combines the N-uniform approximation in Theorem 4.4 with the convergence for the Kac process in the cutoff cases (CHP_K) proven in Lemma 3.14, which comes from Norris [157].

We also remark that the rate obtained in Theorem 4.5 is equivalent to that of [55], and is likely very far from optimal. It may be possible to improve on this by using the *regularising effect of grazing collisions* [7, 8, 60] to improve the estimates in terms of the branching process representation or linearised Boltzmann equation described in Section 3.2, but we will not explore this here.

Let us remark that the original proof of Corollary 4.2, given in [112] and reproduced in Section 4.9 offers an entirely probabilistic 'bottom-up' proof of the stability of the Boltzmann equation by propagating the coupling of Kac process in Theorem 4.4 to the Boltzmann equation. Since the argument uses the same fundamental ingredients but combined in a different way, and introduces a different argument (the existence of a stable solution map implies uniqueness), it is included here for completeness. Let us remark that, while the uniqueness of the Boltzmann equation does not require coupling the many-particle systems and the coupling of the Boltzmann processes is, in a sense, more fundamental, it is much easier to set up a coupling for Kac processes than the equivalent coupling for the Boltzmann processes, and the symmetry arguments are more manifest.

Significant parts of the argument leading to Corollary 4.2 can be replaced by a deterministic (i.e., without using probability theory) argument, once one has established the existence of solutions to a certain coupled Boltzmann equation (4.54). In this case, the proof can be shortened significantly, completely bypassing the analysis of the stochastic Boltzmann processes in Section 4.4 and bypassing some of the steps in Lemmas 4.16 and the final proof; the argument for the Landau case in the joint work by the author [90] follows exactly this argument. In this case, we would only find a coupling $\pi_t \in \Pi(\mu_t, \nu_t)$ for each t, without finding or describing a stochastic process $(V_t, \tilde{V}_t)_{t\geq 0}$ whose marginals are π_t; in keeping with the overall stochastic theme of this thesis, it is interesting to keep take this more probabilistic approach.

4.2 Jump Stochastic Differential Equations Associated to Boltzmann Processes

We begin by formulating both Boltzmann processes and (labelled) Kac process as a stochastic differential equation in $\mathbb{R}^d, (\mathbb{R}^d)^N$ respectively driven by Poisson random measures, using the coefficients $a(v, v_*, z, \varphi)$ described in Section 2.4 for the case (NCHP). For the Boltzmann processes, we rewrite the stochastic differential equation (stBE) using the parametrisation of Section 2.4 as

$$V_t = V_0 + \int_{(0,t]\times\mathbb{R}^d\times\mathbb{S}^{d-2}\times(0,\infty)} a(V_{s-}, v_*, z, \varphi)\mathcal{N}(ds, dv_*, d\varphi, dz) \tag{stBE^a}$$

for a Poisson random measure $\mathcal{N}$ on $(0,\infty)\times\mathbb{R}^d\times\mathbb{S}^{d-2}\times(0,\infty)$ of intensity $2dt\mu_t(dv_\star)d\varphi dz$, and with nonlinear dependence through $\mu_t = \mathrm{Law}(V_t)$; using the representation Theorem of El Karoui [70] and Lemma 2.5, this can be seen to be equivalent to the original formulation (stBE). In the case with cutoff, we will write $a_K, \mathcal{L}_K$ for the corresponding objects, and (stBE$^a{}_K$) for the corresponding equation.

We also construct a labelled Kac process to be the solution to an SDE with Poisson noise. For unordered pairs $\{ij\} = \{ji\}$ of distinct indexes $i, j = 1, ..., N$, let $\mathcal{N}^{\{ij\}}$ be independent Poisson random measures on $(0,\infty)\times\mathbb{S}^{d-2}\times(0,\infty)$, with intensity $2N^{-1}dsd\varphi dz$, and consider solutions $\mathcal{V}^N_t = (V^1_t, ..., V^N_t)$ to the system of stochastic differential equations

$$V^i_t = V^i_0 + \sum_{j\neq i}\int_{(0,t]\times\mathbb{S}^{d-2}\times(0,\infty)} a(V^i_{s-}, V^j_{s-}, z, \varphi)\,\mathcal{N}^{\{ij\}}(ds, d\varphi, dz) \tag{LK}$$

where the index i runs over $1, ..., N$. The factor of 2 in the rate corresponds to working with unlabelled, rather than labelled, pairs of particles. Moreover, thanks to the antisymmetry of a in the first two arguments, and recalling that $\mathcal{N}^{\{ij\}} = \mathcal{N}^{\{ji\}}$, we see that a jump in the i^{th} particle $V^i_t \neq V^i_{t-}$ matches a jump in some j^{th} particle, $j \neq i$. Clasically [127], weak solutions to the stochastic differential equation (LK) are Markov processes with the generator

$$\begin{aligned}(\mathcal{G}^{\mathrm{L}}\widehat{F})(\mathcal{V}^N) \\ = \frac{2}{N}\sum_{\{ij\}}\int_0^\infty dz\int_{\mathbb{S}^{d-2}} d\varphi\,\left(\widehat{F}\left(\mathcal{V}^N + a(V^i, V^j, z, \varphi)(\mathbf{e}_i - \mathbf{e}_j)\right) - \widehat{F}(\mathcal{V}^N)\right)\end{aligned} \tag{4.9}$$

where $h\mathbf{e}_i = (0, \ldots 0, h, 0\ldots 0) \in (\mathbb{R}^d)^N$ has h in the i^{th} place. Using the same computations as in Section 2.4, the integral can be rewritten to be exactly (1.12) in the introduction, so that (LK) corresponds exactly the labelled Kac processes.

We will use the following result.

Proposition 4.6. *For all $\mathcal{V}^N_0 \in \mathbb{S}_N$, there exists a unique-in-law labelled Kac process $\mathcal{V}^N_t, t \geq 0$, that is, a weak solution to (LK), starting at $\mathcal{V}^N_0$.*

This proposition is largely standard, see for instance the proof of a very similar proposition [92, Proposition 1.2ii)]; for existence, one proves a tightness property for the cutoff equivalent (cLK) below, and identifies subsequential limits as solutions to (LK). While the coefficients a are not Lipschitz, one can use a coupling argument to show that every solution $(\mathcal{V}^N_t)_{t\ge 0}$ to the associated martingale problem is the limit of solutions $(\mathcal{V}^{N,K}_t)_{t\ge 0}$ to the cutoff martingale problems, with some N-dependent rate; since the law of $\mathcal{V}^{N,K}_t$ is unique, by finiteness of the rates, it follows that the same is true of $\mathcal{V}^N_t$. Since the coupling argument is similar to one we develop anyway (with N-uniform rates) for Theorem 4.4 in Section 4.8, we will defer the proof until then, and remark that no intermediate steps use existence so there is no circularity.

We justify moving between the labelled and unlabelled dynamics, with the following proposition.

Proposition 4.7. *i). Suppose $\mathcal{V}^N_t$ is a solution to to the stochastic differential equation (LK). Then the empirical measures $\mu^N_t = \theta_N(\mathcal{V}^N_t)$ are unlabelled Kac process.*

ii). Every Kac process arises in this way: if $(\widetilde{\mu}^N_t, t \ge 0)$ is a Kac process starting at μ^N_0, pick $\mathcal{V}^N_0 \in \theta_N^{-1}(\mu^N_0)$ uniformly at random. Then there exists a weak solution to the stochastic differential equation (LK), starting at $\mathcal{V}^N_0$, such that $(\mu^N_t, t \ge 0) = (\theta_N(\mathcal{V}^N_t), t \ge 0)$ has the same law as $(\widetilde{\mu}^N_t, t \ge 0)$.

iii). For any $\mu^N_0 \in \mathcal{S}_N$, then there exists a unique-in-law, N-particle (unlabelled) Kac process starting at μ^N_0.

For ease of readability, the proof is deferred to Appendix 4.A. The first item is elementary, and uses the $\mathrm{Sym}(N)$-symmetry of the labelled dynamics; the second item amounts to the careful application of a theorem due to Kurtz [126, 127], and we conclude item iii) using Proposition 4.6. Again, there is no circularity.

We also construct a cutoff version $\mathcal{V}^{N,K}_t = (V^{1,K}_t, ...V^{N,K}_t)$ of these processes as follows. In analogy to the definition above, set

$$a_K(v, v_\star, z, \varphi) = a(v, v_\star, z, \varphi)\,\mathbb{1}\left(z \le K|v - v_\star|^\gamma\right). \tag{4.10}$$

The K-cutoff version of (LK), corresponding to the kernel B_K with Grad's angular cutoff defined in (CHP_K) is now

$$V^{i,K}_t = V^{i,K}_0 + \sum_{j \ne i} \int_{(0,t]\times\mathbb{S}^{d-2}\times(0,\infty)} a_K(V^{i,K}_{s-}, V^{j,K}_{s-}, z, \varphi)\, \mathcal{N}^{\{ij\}}(ds, d\varphi, dz). \tag{cLK}$$

In the notation of Section 2.4, $\theta_0(K) = G(K) \to 0$ as $K \to \infty$. Let us remark that the statements equivalent to Propositions 4.6, 4.7, for the cutoff differential equation (cLK) and the corresponding cutoff Kac process $\mu^{N,K}_t$ are elementary, as in both cases the overall jump rates are uniformly bounded.

4.3 Accurate Tanaka Trick and Tanaka-Povzner Lemma

In this section, we will exhibit the key coupling, which will be applied to both Boltzmann processes and the Kac process, and a family of lemmas which will describe, in either case, how fast the coupled processes tend to move away from each other.

4.3.1 Accurate Tanaka's Trick

We begin with the following 'accurate Tanaka Lemma', which generalises that of [85]. Our result is slightly more general, in that we allow any $d \geq 3$, while the result cited applies for only $d = 3$.

Lemma 4.8. *[Accurate Tanaka's Trick] There exists a measurable function* $R : \mathbb{R}^d \times \mathbb{R}^d \to$ *Isom*$(\mathbb{S}^{d-2})$ *such that, for all* $X, Y \in \mathbb{R}^d$ *and* $\varphi \in \mathbb{S}^{d-2}$*, we have*

$$\Gamma(X, \varphi) \cdot \Gamma(Y, R(X, Y)\varphi) \geq X \cdot Y. \tag{4.11}$$

Proof. The cases where either $X = 0, Y = 0$ are vacuous; for the remainder of the proof, let us assume that both X, Y are nonzero. Let us write, throughout, S_X for the set

$$S_X = \{u \in \mathbb{R}^d : |u| = |X|, u \cdot X = 0\}. \tag{4.12}$$

By considering separately the cases where X, Y are and are not colinear, we observe that we may choose j^1_X, j^1_Y such that

$$\dim \text{Span}(X, Y, j^1_X, j^1_Y) = 2, \qquad j^1_X \in S_X, j^1_Y \in S_Y; \qquad j^1_X \cdot j^1_Y = X \cdot Y. \tag{4.13}$$

With some care, the map $(X, Y) \mapsto (j^1_X, j^1_Y)$ can further be constructed to be measurable. We now construct, in a measurable way, $u_2, ...u_{d-1}$ as an orthonormal basis for $\text{Span}(X, j^1_X)^\perp = \text{Span}(Y, j^1_Y)^\perp$, and set

$$j^2_X = |X|u_2, \; j^3_X = |X|u_3, \; ...., \; j^{d-1}_X = |X|u_{d-1}; \tag{4.14}$$

$$j^2_Y = |Y|u_2, \; j^3_Y = |Y|u_3, \; ...., \; j^{d-1}_Y = |Y|u_{d-1}. \tag{4.15}$$

Now, $\{j^1_X, ...j^{d-1}_X\}$ are orthonogal, and lie in S_X, so there is a unique isometry $P_X \in$ Isom$(\mathbb{S}^{d-2})$ such that

$$\Gamma(X, P_X e_k) = j^k_X, \qquad k = 1, ..., d-1 \tag{4.16}$$

and similarly for P_Y. We now observe, for all $\varphi \in \mathbb{S}^{d-2}$,

$$\begin{aligned} \Gamma(X, P_X\varphi) \cdot \Gamma(Y, P_Y\varphi) = \sum_{k=1}^{d-1} \phi_k^2 \, j_X \cdot j_Y &= \phi_1^2 \, (X \cdot Y) + (1 - \varphi_1^2)|X||Y| \\ &\geq \phi_1^2 \, (X \cdot Y) + (1 - \varphi_1^2)(X \cdot Y) \\ &= X \cdot Y \end{aligned} \tag{4.17}$$

which implies the result when we define $R(X, Y) = P_Y P_X^{-1}$. ∎

We will use a version of this in which the maps ι, Γ defined in Section 2.4 are given an additional parameter $\alpha \in (0,1)$, and the maps R now also depend on α accordingly.

Lemma 4.9. *There exists a map $\iota : \mathbb{R}^d \times (0,1) \to (\mathbb{R}^d)^{d-1}$ such that, for all $\alpha \in (0,1)$, $j,k \in \{1,...,d-1\}$, such that $\iota_\alpha(-X) = -\iota_\alpha(X)$ and $X \in \mathbb{R}^d$,*

$$\iota_{\alpha,j}(X) \cdot \iota_{\alpha,k}(X) = |X|^2 \cdot \mathbb{1}_{j=k}. \tag{4.18}$$

There also exists a map $R : \mathbb{R}^d \times \mathbb{R}^d \times (0,1) \to Isom(\mathbb{S}^{d-2})$ such that, for all α, X, Y, $R_\alpha(X,Y)$ satisfies the conclusion (4.11) of Lemma 4.8 for the maps Γ_α defined with ι_α in place of ι. Finally, ι_α, R_α can be chosen such that, for all $X, Y \in \mathbb{R}^d \setminus \{0\}$, the sets

$$\{\alpha : \iota_\alpha \text{ or } \Gamma_\alpha \text{ is discontinuous at } X\}; \tag{4.19}$$

$$\{\alpha : R_\alpha \text{ is discontinuous at } (X,Y)\} \tag{4.20}$$

have $d\alpha$-measure 0.

Since the specific construction is not important for the arguments of this chapter, it is deffered to Appendix 4.B. The averaging over α will be important in Section 4.5, in order to ensure that the generator of a nonlinear Markov process is continuous.

4.3.2 Tanaka-Povzner Lemmata

The key tool at the heart of our results is the following variant of some calculations in [92, Lemmas 3.1, 3.3]. The key point is the appearance of a large negative term, similar to that arising in the Povzner inequalities, which ensures the cancellation of 'bad' terms and leads to a Grönwall inequality.

Lemma 4.10. *For all $X, Y \in \mathbb{R}^d$, let $R(X,Y) \in Isom(\mathbb{S}^{d-2})$ be an isometry satisfying the conclusion of Lemma 4.8. Let us write, for $v, \widetilde{v}, v_\star, \widetilde{v}_\star \in \mathbb{R}^d$, $z \in (0,\infty), \varphi \in \mathbb{S}^{d-2}$ and $K < \infty$,*

$$a = a(v, v_\star, z, \varphi); \qquad v' = v + a; \qquad v'_K = v + a_K(v, v_\star, z, \varphi); \tag{4.21}$$

$$\widetilde{a}_K = a_K(\widetilde{v}, \widetilde{v}_\star, z, R(v - v_\star, \widetilde{v} - \widetilde{v}_\star)\varphi); \qquad \widetilde{v}'_K = \widetilde{v} + \widetilde{a}_K. \tag{4.22}$$

Define

$$\mathcal{E}_{p,K}(v, \widetilde{v}, v_\star, \widetilde{v}_\star) = \int_0^\infty dz \int_{\mathbb{S}^{d-2}} d\varphi \left(d_p(v', \widetilde{v}'_K) - d_p(v, \widetilde{v})\right) \tag{4.23}$$

and, for $p \geq 2$,

$$\lambda_p = \int_0^{\pi/2} \left(1 - \left(\frac{1+\cos\theta}{2}\right)^{p/2}\right) \beta(\theta) d\theta. \tag{4.24}$$

Then there exists $K_0(p)$, constants $c = c(G,d)$ and $C = C(G,d,p)$, such that, whenever $K \geq K_0(p)$, we have

$$\begin{aligned}\mathcal{E}_{p,K}(v,\widetilde{v},v_\star,\widetilde{v}_\star) \leq \left(c - \frac{\lambda_p}{2}\right) d_{p+\gamma}(v,\tilde{v}) + c d_{p+\gamma}(v_*,\tilde{v}_*) \\ + C\left(|v_\star|^{p+\gamma} + |\widetilde{v}_\star|^{p+\gamma}\right) d_p\left(v,\widetilde{v}\right) \\ + C\left(|v|^{p+\gamma} + |\widetilde{v}|^{p+\gamma}\right) d_p\left(v_\star,\widetilde{v}_\star\right) \\ + CK^{1-1/\nu}(1+|v|^l+|v_\star|^l+|\widetilde{v}|^l+|\widetilde{v}_\star|^l)\end{aligned} \tag{4.25}$$

where $l = p+2+\gamma$. In the case where a, a_K have the additional parameter α as in Lemma 4.9, the same holds for $\mathcal{E}_{p,K}(v,\tilde{v},v_*,\tilde{v}_*,\alpha)$ for the corresponding maps R_α, with constants independent of α.

Remark 4.11. *Let us motivate this lemma, which is not necessarily transparent. We obtain, in expanding $\mathcal{E}_{p,K}$, the noise term in the final line, and terms proportional to $|v-\widetilde{v}|^2, |v_\star - \widetilde{v}_\star|^2$ with all possible polynomial weightings of order $p+\gamma$. The difficult terms are those like $d_{p+\gamma}(v,\tilde{v})$, which prevent a Grönwall estimate (see also the sketch proof of Proposition 4.18 below). However, we ensure that the coefficients of such terms are independent of p, which allows us to cancel all such terms by the negative 'Povzner term' appearing in the first line by making p large.*

We will also use the following variant, which will be used to prove a local uniform estimate on our coupling.

Lemma 4.12. *In the notation of the previous lemma, define also*

$$\mathcal{Q}_K(v,\widetilde{v},v_\star,\widetilde{v}_\star) = \int_0^\infty dz \int_{\mathbb{S}^{d-2}} d\varphi \left(d_p(v',\widetilde{v}'_K) + d_p(v'_\star,\widetilde{v}'_{\star K}) - d_p(v,\widetilde{v}) - d_p(v_\star,\widetilde{v}_\star)\right)^2. \tag{4.26}$$

Then, for some C=C(G,d,p), we have

$$\mathcal{Q}_K(v,\widetilde{v},v_\star,\widetilde{v}_\star) \leq C(1+|v|^{2l}+|v_\star|^{2l}+|\widetilde{v}|^{2l}+|\widetilde{v}_\star|^{2l}). \tag{4.27}$$

where $l = p+2+\gamma$ is as in Lemma 4.10.

4.4 Some Properties of Boltzmann Processes

Before giving the proof of Theorem 4.1 based on Tanaka's stochastic interpretation (stBE^a) of the Boltzmann equation, we will first some properties which we will use later. Our first property is a tightness property, together with an identification of limits.

Lemma 4.13. *For $n \geq 1$, let $(V_t^n)_{t\geq 0}$ be a solution to (stBE^a) with law $\mathrm{Law}(X_t^n) = \mu_t^n \in \mathcal{S}$ the corresponding solution to (BE). Then the processes V^n are tight in the Skorokhod topology of $\mathbb{D}([0,\infty),\mathbb{R}^d)$. Moreover, if any V_t is any subsequential limit point such that $\mu_t = \mathrm{Law}(V_t) \in \mathcal{S}$ for all t, then it is again a solution to (stBE^a).*

We will use the following simple fact in the proof.

Lemma 4.14. *Let (E,d) be a metric space, $\mu^n \in \mathcal{P}(E)$ converging to $\mu \in \mathcal{P}(E)$ in the weak topology, and $f : E \times E \to \mathbb{R}$ a continuously supported function such that, for some $K_1 \subset E$ compact, $f(x,y) = 0$ whenever $y \notin K_1$. Then, for any $K \subset E$ compact, we have*

$$\sup_{x\in K}\left|\int_E f(x,y)\mu^n(dy) - \int_E f(x,y)\mu(dy)\right| \to 0. \tag{4.28}$$

Proof of Lemma 4.13. Using the criterion of Aldous [4], to prove tightness it is sufficient to show that, for all $t_{\text{fin}} \geq 0$,

$$\sup_n \mathbb{E}\left[\sup_{t\leq t_{\text{fin}}} |V_t^n|\right] < \infty; \tag{4.29}$$

$$\lim_{\delta\to 0}\sup_n \sup_{(T,T')\in\mathcal{S}(\delta,t_{\text{fin}})} |V_{T'}^n - V_T^n| \to 0 \tag{4.30}$$

where $\mathcal{S}(\delta, t_{\text{fin}})$ is the set of pairs of stopping times (T,T') with $0 \leq T \leq T' \leq T+\delta \leq T$.

Step 1. Local Uniform Moment Estimate For the local supremum $\sup_{t\leq t_{\text{fin}}} |V_t^n|$, we observe that, at jumps,

$$\begin{aligned}|V_t^n| - |V_{t-}^n| &\leq |V_{t-}^n - v|\sin G\left(\frac{z}{|V_{t-}^n - v|^\gamma}\right)\\ &\leq (|V_{t-}^n| + |v|)\sin G\left(\frac{z}{|V_{t-}^n - v|^\gamma}\right)\end{aligned} \tag{4.31}$$

where (t,v,φ,z) is the corresponding point of the Poisson random measure $\mathcal{N}$ driving V_t^n. Integrating,

$$\sup_{s\leq t}|V_s^n| \leq |V_0^n| + \int_{(0,t]\times\mathbb{R}^d\times\mathbb{S}^{d-2}\times(0,\infty)} (|V_{s-}^n| + |v|)\sin G\left(\frac{z}{|V_{s-}^n - v|^\gamma}\right)\mathcal{N}^n(ds,dv,d\varphi,dz). \tag{4.32}$$

Therefore

$$\sup_{s\le t}|V^n_s| - \int_{(0,t]\times\mathbb{R}^d\times(0,\infty)} (|V^n_s|+|v|)\sin G\left(\frac{z}{|V^n_s-v|^\gamma}\right) ds\mu^n_s(dv)dz \tag{4.33}$$

is a supermartingale. When we integrate over z, we obtain an extra factor $|V^n_s - v|^\gamma\int_0^\infty \sin G(z)dz$ and recall that $\int_0^\infty \sin G(z)dz<\infty$, so for some C,

$$\sup_{s\le t}|V^n_s| - C\int_{(0,t]\times\mathbb{R}^d}(1+|V^n_s|^{1+\gamma}+|v|^{1+\gamma})\mu^n_s(dv)ds \tag{4.34}$$

is also a supermartingale. Taking expectations and recalling that $1+\gamma\le 2$, the second term is controlled by the hypothesised second moment estimate $\mathbb{E}[|V^n_t|^2]=\Lambda_2(\mu_t)=1$ and we conclude that $\mathbb{E}[\sup_{s\le t}|V^n_s|]<\infty$ is bounded, uniformly in n and locally uniformly in t.

We further remark that a similar argument holds for the local suprema $\sup_{t\in(\delta,t_{\text{fin}}]}|V^n_t|^2$. In this case, we observe that $\mathbb{E}[|V^n_\delta|^2]=\Lambda_2(\mu^n_\delta)=1$, and for some C, $\sup_{\delta\le s\le t}|V^n_s|^2 - C\int_{(\delta,t]\times\mathbb{R}^d}(1+|V^n_s|^{2+\gamma}+|v|^{2+\gamma})\mu^n_s(dv)ds$ is a supermartingale. Using the moment bounds in Proposition 2.6i), there exists C, independently of n,δ,t_{fin}, such that $\Lambda_{2+\gamma}(\mu^n_s)\le C(1+s^{-1})$, and hence $\int_\delta^{t_{\text{fin}}}\Lambda_{2+\gamma}(\mu^n_s)ds\le C(t_{\text{fin}}+\log t_{\text{fin}}-\log\delta)$, and we find that

$$\mathbb{E}\left[\sup_{\delta\le t\le t_{\text{fin}}}|V^n_t|^2\right]\le 1+C(t_{\text{fin}}+\log t_{\text{fin}}-\log\delta). \tag{4.35}$$

Step 2. Equicontinuity Property We next check (4.30). Fixing t_{fin},δ,n and such T,T', we argue as in the previous step that

$$|V^n_{T'}-V^n_T|\le\int_{(0,t_{\text{fin}}]\times\mathbb{R}^d\times\mathbb{S}^{d-2}\times(0,\infty)}(|V^n_{s-}|+|v|)\sin G\left(\frac{z}{|V^n_{s-}-v|^\gamma}\right)1\!\!1_{T<s\le T'}\mathcal{N}^n(ds,dv,d\varphi,dz). \tag{4.36}$$

Now, we observe that the process

$$Z^n_t=\int_{(0,t]\times\mathbb{R}^d\times\mathbb{S}^{d-2}\times(0,\infty)}(|V^n_{s-}|+|v|)\sin G\left(\frac{z}{|V^n_{s-}-v|^\gamma}\right) \cdots\times(\mathcal{N}^n(ds,dv,d\varphi,dz)-ds\mu^n_s(dv)d\varphi dz) \tag{4.37}$$

is a martingale, and we write

$$|V^n_{T'}-V^n_T|\le Z^N_{T'}-Z^N_T+\int_{(0,t_{\text{fin}}]\times\mathbb{R}^d\times\mathbb{S}^{d-2}\times(0,\infty)}(|V^n_{s-}|+|v|)\sin G\left(\frac{z}{|V^n_{s-}-v|^\gamma}\right)1\!\!1_{T<s\le T+\delta} \cdots\times ds\mu^n_s(dv)d\varphi dzd\alpha. \tag{4.38}$$

Taking expectations, the first two terms cancel by the optional stopping theorem. Carrying out the integral over z and using again that $\int_0^\infty \sin G dz$ is finite, we find that

$$\begin{aligned}\mathbb{E}[|V^n_{T'} - V^n_T|] &\le C\mathbb{E}\left[\int_T^{T'}\int_{\mathbb{R}^d}(|V^n_s|^{1+\gamma} + |v|^{1+\gamma})\mu^n_s(dv)ds\right] \\ &\le C\mathbb{E}\left[\int_T^{T'}\int_{\mathbb{R}^d}(|V^n_s|^{2} + |v|^{2})\mu^n_s(dv)ds\right]\end{aligned} \tag{4.39}$$

where, in the final line, we recall that $1+\gamma \le 2$. For the integral in v, we recall that that $\Lambda_2(\mu^n_t) = 1$ to find a contribution $C\delta$, while the term in V^n_s is bounded by recalling that $T' - T \le \delta$, so that

$$\int_T^{T'} |V^n_s|^2 ds \le \int_0^\delta |V^n_s|^2 ds + \delta \sup_{\delta \le t \le t_{\text{fin}}} |V^n_s|^2. \tag{4.40}$$

Taking expectations, the first term gives $\mathbb{E}[|V^n_s|^2] = 1$ inside the integral and we control the second term using (4.35) to obtain

$$\mathbb{E}\left[\int_T^{T'} |V^n_s|^2 ds\right] \le \delta + C\delta(t_{\text{fin}} + \log t_{\text{fin}} - \log \delta). \tag{4.41}$$

Putting everything together, we have

$$\mathbb{E}[|V^n_{T'} - V^n_T|] \le C\delta(t_{\text{fin}} + 1 + \log t_{\text{fin}} - \log \delta). \tag{4.42}$$

The right-hand side is now independent of n and $(T, T') \in \mathcal{S}(t_{\text{fin}}, \delta)$, and converges to 0 as $\delta \to 0$, so the claim (4.30) is proven.

Step 3. Characterisation of Limits We finally check that any subsequential limit is also a solution to (stBE^a) for some choice of Poisson random measure $\mathcal{N}$. Let us suppose that $(V_t)_{t\ge 0}$ is any process extracted from $(V^n_t)_{t\ge 0}, n \ge 1$ as the limit in distribution, for the Skorokhod topology, along some subsequence. Using Skorokhod's representation theorem, we can replace V^n_t, V_t by processes with the same law with almost sure Skorokhod convergence; further, Step 2 shows that, for all t, $\mathbb{P}(V_t \ne V_{t-}) = 0$, so $V^n_t \to V_t$ almost surely for any fixed t. Thanks to the representation theorem of El Karoui and Lepeltier [70] it is sufficient to show that, for any $f \in C^1_b(\mathbb{R}^d)$, the process

$$M^f_t = f(V_t) - f(V_0) - \int_0^t \int_{\mathbb{R}^d} \mathcal{L}f(V_s, v)\mu_s(dv)ds \tag{4.43}$$

is a martingale. For each n, the corresponding processes $M^{n,f}_t$, defined with V^n_t, μ^n_t in place of V_t, μ_t, is a martingale.

Let us fix $f \in C^1_b(\mathbb{R}^d)$ and observe first some properties of $\mathcal{L}f$. It is straightforward to see that each $\mathcal{L}_K f$ is continuous on $\mathbb{R}^d\times\mathbb{R}^d$, and the straightforward estimate $|v'-v|\le |v-v_\star|\sin G(z/|v-v_*|^\gamma)$ implies the growth bound

$$|\mathcal{L}f(v,v_\star)|\le C(f)|v-v_\star|\int_{(0,\infty)}\sin G\left(\frac{z}{|v-v_*|^\gamma}\right)dz\le C(f)|v-v_\star|^{1+\gamma} \tag{4.44}$$

for some constant $C=C(f)$, depending only on the Lipschitz constant of f, and similarly for $\mathcal{L}_K$. The same argument also shows that

$$|(\mathcal{L}f)(v,v_\star)-(\mathcal{L}_K f)(v,v_\star)|\le C(f)\epsilon_K|v-v_\star|^{1+\gamma}; \tag{4.45}$$

$$\epsilon_K=\int_K^\infty \sin G(z)dz\to 0 \tag{4.46}$$

so that $\mathcal{L}_K f\to\mathcal{L}f$, uniformly on compact subsets of $\mathbb{R}^d\times\mathbb{R}^d$; it therefore follows that $\mathcal{L}f$ is continuous. Recalling that $1+\gamma\le 2$, it holds that $|\mathcal{L}f(v,v_*)|\le C(1+|v|^{1+\gamma})(1+|v_*|^2)$, and using the moment estimates Proposition 2.6i) on μ_t^n, there exists some C such that $\Lambda_{2+\gamma/2}(\mu^n_t)\le C(1+t^{-1/2})$ for all n, and by lower semicontinuity the same is true for μ_t. We now use a Chebychev estimate to see that

$$\begin{aligned}\int_{\mathbb{R}^d}|\mathcal{L}f(v,v_*)|\mathbb{1}_{|v_*|\ge R}\mu_t^n(dv_*)&\le C(1+|v|^{1+\gamma})R^{-\gamma/2}\Lambda_{2+\gamma/2}(\mu_t^n)\\ &\le C(1+|v|^{1+\gamma})R^{-\gamma/2}(1+t^{-1/2})\end{aligned} \tag{4.47}$$

and similarly for μ_t. We now choose a continuous, compactly supported function $\psi:\mathbb{R}^d\to[0,1]$, which is 1 when $|v_*|\le R$. Since $\mathcal{L}f(v,v_*)\psi(v_*)$ is continuous and compactly supported in v_*, it follows from Lemma 4.14 that, for any given t,

$$\int_{\mathbb{R}^d}\mathcal{L}f(v,v_*)\psi(v_*)\mu_t^n(dv_*)\to\int_{\mathbb{R}^d}\mathcal{L}f(v,v_*)\psi(v_*)\mu_t(dv_*)$$

uniformly on compact sets in v. In particular, on compact time intervals, we can choose a compact set containing the image of $V_t^n, V_t, t\le t_{\text{fin}}$ for all n to see that

$$\int_{\mathbb{R}^d}\mathcal{L}f(V_t^n,v_*)\psi(v_*)(\mu_t^n-\mu_t)(dv_*)\to 0.$$

Meanwhile, $\mathcal{L}f(v,v_*)\psi(v_*)$ is uniformly continuous on compact regions in v, so for all t such that $V_t^n\to V_t$, it follows that

$$\int_{\mathbb{R}^d}\mathcal{L}f(V_t^n,v_*)\psi(v_*)\mu_t(dv_*)\to\int_{\mathbb{R}^d}\mathcal{L}f(V_t,v_*)\psi(v_*)\mu_t(dv_*)$$

and we conclude that, for such t,

$$\int_{\mathbb{R}^d}\mathcal{L}f(V_t^n,v_*)\psi(v_*)\mu_t(dv_*)\to\int_{\mathbb{R}^d}\mathcal{L}f(V_t,v_*)\psi(v_*)\mu_t(dv_*).$$

Almost surely, this convergence holds for all t at which V_t is continuous, which excludes only a dt-measure 0 set. It then follows from bounded convergence that the same holds when we replace each side by its integral from 0 to t by bounded convergence. Recalling (4.47), we find that, for all $t \geq 0$, almost surely

$$\limsup_n \left|\int_0^t \int_{\mathbb{R}^d} \mathcal{L}f(V_s^n, v_*)\mu_s^n(dv_*)ds - \int_0^t \int_{\mathbb{R}^d} \mathcal{L}f(V_s, v_*)\mu_s(dv_*)ds\right| \\ \leq CR^{-\gamma/2} \sup_n \sup_{s\leq t}(1 + |V_s^n|^{1+\gamma} + |V_s|^{1+\gamma})(t + t^{1/2}). \tag{4.48}$$

We observe that Skorokhod convergence implies that the local supremum $\sup_n \sup_{s\leq t}|V_s^n|$ is almost surely finite, so we can take $R \to \infty$ in the right-hand side to conclude that, almost surely

$$\int_0^t \int_{\mathbb{R}^d} \mathcal{L}f(V_s^n, v_*)\mu_s^n(dv_*) \to \int_{\mathbb{R}^d} \mathcal{L}f(V_s, v_*)\mu_s(dv_*)$$

uniformly on compact time intervals. Similarly, since f is continuous, $f(V_t^n) - f(V_0^n) \to f(V_t) - f(V_0)$ almost surely in the Skorokhod topology, and almost surely for each fixed t. To upgrade from almost sure convergence to $L^1(\mathbb{P})$ convergence, observe that $f(V_t^n), f(V_0^n)$ are uniformly bounded, and so uniformly integrable random variables. For the integral term, we note that $\frac{2+(\gamma/2)}{1+\gamma} = \frac{4+\gamma}{2+2\gamma} > 1$ for any $\gamma \in (0, 1]$, and we estimate

$$\begin{aligned}
\mathbb{E}&\left[\left|\int_0^t \int_{\mathbb{R}^d} \mathcal{L}f(V_s^n, v)\mu_s^n(dv)ds\right|^{(4+\gamma)/(2+2\gamma)}\right] \\
&\leq t^{(2-\gamma)/(2+2\gamma)} \int_0^t \mathbb{E}\left[|\mathcal{L}f(V_s^n, v)|^{(4+\gamma)/(2+2\gamma)}\right] \mu_s^n(ds) \\
&\leq Ct^{(2-\gamma)/(2+2\gamma)} \int_0^t \int_{\mathbb{R}^d} (1 + \mathbb{E}[|V_s^n|^{2+(\gamma/2)}] + |v|^{2+(\gamma/2)})\mu_s^n(ds) \\
&\leq Ct^{(2-\gamma)/(2+2\gamma)} \int_0^t \Lambda_{2+(\gamma/2)}(\mu_s^n)ds \\
&\leq Ct^{(2-\gamma)/(2+2\gamma)} \int_0^t (1 + s^{-1/2})ds.
\end{aligned} \tag{4.49}$$

The final integral is bounded, uniformly in n, and it follows that $M_t^{n,f}$ are bounded, uniformly in $L^{(4+\gamma)/(2+2\gamma)}(\mathbb{P})$ and locally uniformly in time. For fixed t, f, $M_t^{n,f}$ are uniformy integrable, and so converge to M_t^f in $L^1(\mathbb{P})$. The limit M_t^f is therefore a martingale, and the step is complete. □

The other property we will use is the following uniqueness result.

Lemma 4.15. *Suppose that $(V_t)_{t\geq 0}, (\tilde{V}_t)_{t\geq 0}$ be two solutions to (stBE^a) with* $\text{Law}(V_t) = \mu_t = \text{Law}(\tilde{V}_t)$ *for all $t \geq 0$. Then* $\text{Law}((V_t)_{t\geq 0}) = \text{Law}((\tilde{V}_t)_{t\geq 0})$.

Sketch Proof. Let us remark that this result can be found in the literature; we sketch the argument of [81, Proposition 3.4]. By a disintegration argument, it is enough to know that, for all $v \in \mathbb{R}^d$, there is a unique-in-law process V_t such that $V_0 = v$ and, for all $f \in C_c^1(\mathbb{R}^d)$, $f(V_t) - f(v) - \int_0^t A_s f(V_s)ds$ is a martingale, where

$$A_s f(v) := \int_{\mathbb{R}^d} \mathcal{L}f(v, v_*)\mu_s(dv_*).$$

Let us call this martingale problem (MP). If one replaces the jumps a with the truncation $a(H_n(v), v_*, \varphi, z)$, $H_n(v) = (n\wedge|v|/|v|)v$, then any solution to the corresponding, truncated martingale problem (MP$_n$) can be approximated by solutions to the martingale problem with truncation and angular cutoff (MP$_{n,K}$), using the same Tanaka-style coupling as we develop in Section 4.5; see [81, Proposition 3.4, Step 6]. Uniqueness certainly holds for (MP$_{n,K}$), since the rates are finite, and so by taking limits it holds that uniqueness holds for (MP$_n$), see [83, Remark 9.7]. Finally, let T_n be the first time that $|V_t| \geq n$; since V is càdlàg, $T_n \to \infty$ almost surely. On the other hand, $V_t^{T_n}$ has the same law as a solution V_t^n to (MP$_n$), stopped at the respective time, and we conclude that the original martingale problem (MP) has uniqueness in law, as desired. □

4.5 Tanaka Coupling of Boltzmann Processes

In this section, we will set up a Tanaka coupling of Boltzmann processes $(V_t, \tilde{V}_t)_{t\geq 0}$ in such a way as to be able to apply Lemma 4.10. We write $C^1_p(\mathbb{R}^d \times \mathbb{R}^d)$ for continuously differentiable functions on $\mathbb{R}^d \times \mathbb{R}^d$ whose first derivative is of at most polynomial growth. The main step is the following.

Lemma 4.16. *Assume the notation of Lemma 4.9. Let $(\mu_t)_{t\geq 0}, (\nu_t)_{t\geq 0} \subset \mathcal{S}$ solutions of the noncutoff and cutoff Boltzmann equations (BE, BE$_K$) respectively, for some $K \in [1,\infty]$. Suppose that μ_0 satisfies an exponential initial moment condition $\int_{\mathbb{R}^d} e^{\varepsilon|v|^\gamma}\mu_0 < \infty$ for some $\varepsilon > 0$, and that all moments of ν_0 are finite (if $K < \infty$), or that $\int_{\mathbb{R}^d} e^{\varepsilon|v|^\gamma}\nu_0 < \infty$ for some $\varepsilon > 0$ if $K = \infty$.*

Fix $\pi_0 \in \Pi(\mu_0, \nu_0)$. Then there exists a stochastic process $(V_t, \tilde{V}_t)_{t\geq 0}$ such that, for all $t \geq 0$, $\pi_t = \mathrm{Law}(V_t, \tilde{V}_t) \in \Pi(\mu_t, \nu_t)$, and $(V_t, \tilde{V}_t)_{t\geq 0}$ solves the nonlinear jump SDE

$$\begin{cases} V_t = V_0 + \int_E a(V_{s-}, v_*, z, \varphi, \alpha)\, \mathbb{1}_{s\leq t}\, \mathcal{N}(ds, dv_*, d\tilde{v}_*, d\varphi, dz, d\alpha); \\ \tilde{V}_t = \tilde{V}_0 + \int_E a_K(\tilde{V}_{s-}, \tilde{v}_*, z, R_\alpha(V_{s-} - v_*, \tilde{V}_{s-} - \tilde{v}_*)\varphi, \alpha)\, \mathbb{1}_{s\leq t} \\ \qquad\qquad \cdots \times \mathcal{N}(ds, dv_*, d\tilde{v}_*, d\varphi, dz, d\alpha) \end{cases} \tag{4.50}$$

where $\mathcal{N}$ is a Poisson random measure on $E = (0,\infty) \times (\mathbb{R}^d)^2 \times \mathbb{S}^{d-2} \times (0,\infty) \times (0,1)$ of intensity $2dt\, \pi_t(dv_*, d\tilde{v}_*)\, d\varphi\, dz\, d\alpha$. In particular, for any $f \in C^1_p(\mathbb{R}^d \times \mathbb{R}^d)$ the process

$$M^f_t = f(V_t, \tilde{V}_t) - \int_0^t \int_{\mathbb{R}^d\times\mathbb{R}^d} \mathcal{A}_K f(V_s, \tilde{V}_s, v_*, \tilde{v}_*)\pi_s(dv_*, d\tilde{v}_*)ds \tag{4.51}$$

is a martingale, where we define

$$\mathcal{A}_K f(v, \tilde{v}, v_*, \tilde{v}_*) = 2\int_{\mathbb{S}^{d-2}} d\varphi \int_{(0,\infty)} dz \int_{(0,1)} d\alpha \big(f((v,\tilde{v}) + A) - f(v,\tilde{v})\big); \tag{4.52}$$

with the shorthand

$$A = \begin{pmatrix} a(v, v_*, \varphi, z, \alpha) \\ a_K(\tilde{v}, \tilde{v}_*, R_\alpha(v - v_*, \tilde{v} - \tilde{v}_*)\varphi, z, \alpha) \end{pmatrix} \tag{4.53}$$

Remark 4.17. *We make the following remarks.*

i) We will check, in 4.20 below, that V_t solves (stBE^a) and that $\tilde{V}_t$ solves (stBE$^a{}_K$). It therefore follows that this produces the coupling claimed in Theorem 4.1 in the case where μ_0, ν_0 have the necessary exponential moments.

ii) The randomisation over $\alpha \in (0,1)$ will be important at this stage, to ensure that the generator $\mathcal{A}_K f$ given by (4.52) is continuous for $f \in C^1_p(\mathbb{R}^d \times \mathbb{R}^d)$, in order to apply a weak compactness argument. This is why we had to introduce the additional dependence in the coefficients $a(\cdot,\alpha), R_\alpha$ in Lemma 4.9; if we attempted to repeat the same arguments with the 'non-randomised' a, R, we would find that $\mathcal{A}_K f$ could fail to be continuous on a set of a set of 0 Lebesgue measure in $\mathbb{R}^d \times \mathbb{R}^d$, which is however not obviously of 0 measure with respect to π_t.

iii) To rederive the stability results at the level of the solutions $(\mu_t)_{t\ge0}, (\nu_t)_{t\ge0}$, we could work at the level of the coupled Boltzmann equation for π_t, given by specifying, for $f \in C^1_p(\mathbb{R}^d \times \mathbb{R}^d)$,

$$\langle f, \pi_t\rangle = \langle f, \pi_0\rangle + \int_0^t \int_{(\mathbb{R}^d)^2\times(\mathbb{R}^d)^2} \mathcal{A}_K f(x,y)\pi_s(dx)\pi_s(dy). \tag{4.54}$$

It is interesting to give a stochastic statement of the theorem, where we prove the a coupling of the processes and not just of the marginals.

iv) The form of the coupling is essential in order for the estimates obtained from Lemma 4.10 to close; for instance, if we fixed a coupling $\pi_t \in \Pi(\mu_t, \nu_t)$ attaining the minimum of the optimal transport problem for $w_p(\mu_t, \nu_t)$ and treated (4.50) as a linear jump SDE, the estimates would not close: we would find additional terms

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} d_{p+\gamma}(v,\tilde{v})\pi_t(dv, d\tilde{v})$$

which we cannot cancel against the Povzner term $-\lambda_p \mathbb{E}[d_{p+\gamma}(V_t, \tilde{V}_t)]$. With this bilinear form, so that $\pi_t = \mathrm{Law}(V_t, \tilde{V}_t)$, the two terms can be absorbed provided that λ_p is large enough.

v) We will not need uniqueness of solutions to (4.54), or uniqueness in law for (4.50): existence is sufficient for our proof.

We will use in the proof the following weaker fact on well-posedness of (BE), from [93, Corollary 2.3iii)].

Proposition 4.18. *Suppose $\mu_0 \in \mathcal{S}$ satisfies, for some $\epsilon > 0$, $\langle e^{\epsilon|v|^\gamma}, \mu_0\rangle < \infty$. Then there exists at most one solution to either the noncutoff the Boltzmann Equation (BE) taking values in $\mathcal{S}$ and starting at at μ_0.*

For completeness, and since it is in a similar spirit to the argument we will present later, we outline the important steps of the proof, following [93, Theorem 2.2].

Sketch proof of Proposition 4.18. Let us fix two solutions $(\mu_t)_{t\geq 0}, (\nu_t)_{t\geq 0}$ to (BE), and consider, for $f:\mathbb{R}^d\to\mathbb{R}$ with Lipschitz constant at most 1, $h^f_t=\langle f, \mu_t-\nu_t\rangle$. By duality (2.22), we can choose a coupling $\pi_t\in\Pi(\mu_t,\nu_t)$ such that, for all $t\geq 0$,

$$u_t := \mathcal{W}_1(\mu_t,\nu_t)=\sup\left\{h^f_t:\ \text{for all } v,w, |h(v)-h(w)|\leq |v-w|\right\}=\int_{\mathbb{R}^d\times\mathbb{R}^d}|v-\tilde{v}|\pi_t(dv,d\tilde{v}). \tag{4.55}$$

It follows from (BE) that, using the same map R as in Lemma 4.8 but now working with the (θ,φ)-parametrisation, we have

$$\begin{aligned}\frac{d}{dt}h^f_t =&2\int_{(\mathbb{R}^d\times\mathbb{R}^d)^2}\pi_t(dv,d\tilde{v})\pi_t(dv_*,d\tilde{v}_*)\int_0^\pi\beta(\theta)d\theta\int_{\mathbb{S}^{d-2}}d\varphi\\ &\qquad\ldots\Big(|v-v_*|^\gamma|(f(v'(v,v_*,\theta,\varphi))-f(v))\\ &\qquad\qquad\cdots-|\tilde{v}-\tilde{v}_*|^\gamma|(f(v'(\tilde{v},\tilde{v}_*,\theta,R(v-v_*,\tilde{v}-\tilde{v}_*)\varphi))-f(\tilde{v}))\Big).\end{aligned} \tag{4.56}$$

By breaking up the integral and using the properties of the map R, one finds, for any $\delta>0, A<\infty$, [93, Equation 3.5]

$$\frac{d}{dt}h^f_t\leq H_t+\Gamma_{\delta,A}(t)+AS_\delta\left(u_t-h^f_t\right) \tag{4.57}$$

where

$$\begin{aligned}H_t=C\int_{(\mathbb{R}^d\times\mathbb{R}^d)^2}\pi_t(dv,d\tilde{v})\pi_t(dv_*,d\tilde{v}_*)\Big(&\min(|v-v_*|^\gamma,|\tilde{v}-\tilde{v}_*|^\gamma)|v-\tilde{v}|\\ &+(|v-v_*|^\gamma-|\tilde{v}-\tilde{v}_*|^\gamma)_+|v-v_*|\\ &+(|\tilde{v}-\tilde{v}_*|^\gamma-|v-v_*|^\gamma)_+|\tilde{v}-\tilde{v}_*|\Big)\end{aligned}$$

and where, in our notation, $\Gamma_{\delta,A}(s)\leq 2\int_0^\delta\theta\beta(\theta)d\theta+C_AS_\delta$, $S_\delta=\int_\delta^\pi\beta(\theta)d\theta$ and finally

$$C_A=\sup_{t\geq 0}\int_{|v|>A}|v|^2(\mu_t(dv)+\nu_t(dv))\to 0$$

using Proposition 2.15. From (4.57), one obtains first a bound for h^f_t in terms of d_t; taking the supremum, and recalling (4.55), we get

$$u_te^{AS_\delta t}\leq u_0e^{AS_\delta t}+e^{AS_\delta t}\int_0^t(H_s+\Gamma_{\delta,A}(s))ds.$$

Cancelling the exponential, one now takes first $A\to\infty$ and then $\delta\to 0$; uniformly in s, one has $\lim_{\delta\to 0}\lim_{A\to\infty}\Gamma_{\delta,A}(s)=0$, and we conclude

$$u_t\leq u_0+\int_0^tH_sds. \tag{4.58}$$

Bounding all the terms appearing in the integrand of H_s in the spirit of Lemma 4.10, one gets

$$H_t \le C \int_{(\mathbb{R}^d\times\mathbb{R}^d)^2} \pi_t(dv, d\tilde{v})\pi_t(dv_*, d\tilde{v}_*)(|v|^\gamma + |v_*|^\gamma + |\tilde{v}|^\gamma + |\tilde{v}_*|^\gamma)|v-\tilde{v}|.$$

The second and fourth terms integrate to produce u_t again, recalling again (4.55) while, for any $a>0$, we split the other terms into the region where both $|v|, |\tilde{v}| \le a$, or where one exceeds a:

$$\begin{aligned}\int_{\mathbb{R}^d\times\mathbb{R}^d} (|v|^\gamma + |\tilde{v}|^\gamma)|v-\tilde{v}| &\le 2a^\gamma u_t + \int_{\mathbb{R}^d\times\mathbb{R}^d} \pi_t(dv, d\tilde{v})(|v|^{1+\gamma} + |\tilde{v}|^{1+\gamma})\mathbb{1}(|v|>a \text{ or } |\tilde{v}|>a) \\ &\le 2a^\gamma u_t + L_\varepsilon e^{-\varepsilon a^\gamma/2}\int_{\mathbb{R}^d} e^{\varepsilon |v|^\gamma}(\mu_t+\nu_t)(dv)\end{aligned} \tag{4.59}$$

where $\varepsilon>0$ is as in the statement, and L_ε depends only on $\varepsilon>0$. The exponential moment appearing in the final line is finite at 0 by hypothesis, and is propagated to be finite at future times [93, Lemma 4.1], so can be absorbed into C; we now set a so that $a^\gamma = |2(\log u_t)/\varepsilon|$. We finally find, for some C depending on ε and on the exponential moments of $(\mu_t)_{t\ge0}, (\nu_t)_{t\ge0}$,

$$u_t \le u_0 + C\int_0^t u_s(1+|\log u_s|).$$

If we now start from $\nu_0=\mu_0$, then $u_0=0$. Since finally $\int_0^1 \frac{1}{x(1+|\log x|)}dx = \infty$, we apply the Yudovitch lemma [121] to conclude that $u_t = \mathcal{W}_1(\mu_t,\nu_t)=0$ for all $t\ge0$, which concludes the proof. □

Remark 4.19. *We make the following remarks.*

i). The point of the Tanaka-Povzner coupling and the optimal transport function d_p which we use is to cancel the terms like $|v|^\gamma|v-\tilde{v}|$, which prevented us from using Grönwall's lemma on (4.58) in the previous proof.

ii). The key difference between this and our proof is that this proof applies the Tanaka estimate to any two solutions $(\mu_t)_{t\ge0}, (\nu_t)_{t\ge0}$. In our proof, on the other hand, we construct a coupling π_t of two (potentially new) solutions $(\widetilde{\mu}_t)_{t\ge0}, (\widetilde{\nu}_t)_{t\ge0}$ to which we can apply our estimates; we will apply this proposition at the end of the proof to show that we recover the prescribed solutions $\widetilde{\mu}_t=\mu_t, \widetilde{\nu}_t=\nu_t$.

With this uniqueness in hand, we can now give the proof of Lemma 4.16.

Proof of Lemma 4.16. We follow the argument of [83, Lemma 9.4], which generalises well to our case. We will write, throughout, $x=(v,\tilde{v})$ for variables in $\mathbb{R}^d\times\mathbb{R}^d$ to shorten notation.

Step 1. Approximate Problem Let us define, for $n \geq 1$, the truncated jumps

$$a_{(n)}(v, v_*, z, \varphi, \alpha) := \left(-\frac{1}{2}\left(1 - \cos G\left(\frac{z}{n \wedge |v - v_\star|^\gamma}\right)\right)(v - v_*) \right.$$
$$\left. + \frac{1}{2}\sin G\left(\frac{z}{n \wedge |v - v_\star|^\gamma}\right)\Gamma_\alpha(v - v_*, \varphi)\right)\mathbb{1}_{z \leq n(n \wedge |v - v_*|^\gamma)}$$

and similarly $a_{(n,K)}$, where the final indicator is replaced by $\mathbb{1}_{z \leq (n \wedge K)(n \wedge |v - v_*|^\gamma)}$. We now define the two-level truncated jumps

$$A_n(x, x_\star, \varphi, z, \alpha) := \begin{pmatrix} a_{(n)}(v, v_*, z, \varphi, \alpha) \\ a_{(n,K)}(\tilde{v}, \tilde{v}_*, z, R_\alpha(v - v_*, \tilde{v} - \tilde{v}_*)\varphi, \alpha) \end{pmatrix}$$

and similarly $A(x, x_\star, \varphi, z, \alpha)$ with the truncation removed. If we now define $\mathcal{A}_{n,K}$ as in (4.52) with these truncations, then equation corresponding to (4.54) is

$$\begin{aligned} \langle f, \pi_t^n \rangle &= \langle f, \pi_0^n \rangle + \int_0^t \int_{(\mathbb{R}^d \times \mathbb{R}^d)^2} \mathcal{A}_{n,K} f(x, x_\star) \pi_s^n(dx) \pi_s^n(dx_\star) ds \\ &=: \langle f, \pi_0^n \rangle + \int_0^t \langle f, Q_{n,K}^{(2)}(\pi_s^n, \pi_s^n) \rangle ds \end{aligned} \tag{4.60}$$

where the last equality defines the two-level truncated collision operator $Q_{n,K}^{(2)}$. Since the rates are bounded, the map $\pi \mapsto Q_{n,K}^{(2)}(\pi, \pi)$ is continuous in the total variation norm of $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$, and in particular, this equation is readily seen to have a unique solution $\pi_t^n, t \geq 0$ starting at π_0.

To construct the stochastic processes $X_t^n = (V_t^n, \tilde{V}_t^n)$ associated to π_t^n, let $\mathcal{N}^n$ be a Poisson random measure on E of intensity $dt\pi_t^n(dx)d\varphi dz d\alpha$, and independently let X_0^n be a sample from π_0. We define

$$X_t^n := X_0^n + \int_E A_n(X_{s-}^n, x, \varphi, z, \alpha)\, \mathbb{1}_{s \leq t}\, \mathcal{N}^n(ds, dx, d\varphi, dz, d\alpha)$$

and note that this is well-defined: indeed, since A_n is supported on $\{z \leq n^2\}$, this is simply a finite recurrence relation. Moreover, for any bounded, Lipschitz $f : (\mathbb{R}^d)^2 \to \mathbb{R}$, it is readily seen that

$$M_t^{f,n} = f(X_t^n) - f(X_0^n) - \int_0^t \int_{(\mathbb{R}^d)^2} \mathcal{A}_{n,K} f(X_s^n, x) \pi_s^n(dx) ds \tag{4.61}$$

is a martingale. Writing $\widetilde{\pi}_t^n := \mathrm{Law}(X_t^n)$ and taking expectations, we have that $\widetilde{\pi}_0^n = \pi_0$ and solves the linear equivalent of (4.60)

$$\langle f, \widetilde{\pi}_t^n \rangle = \langle f, \pi_0 \rangle + \int_0^t \langle f, Q_{n,K}^{(2)}(\widetilde{\pi}_s^n, \pi_s^n) \rangle ds. \tag{4.62}$$

Again using the finiteness of the rate, the linear operator $Q_{n,K}^{(2)}(\cdot, \pi_s^n)$ on the right-hand side is a continuous linear map for the total variation norm, and so this equation has unique solutions. Since π_t^n solves this equation, we conclude that $\widetilde{\pi}_t^n = \pi_t^n$, so that X_t^n solves the nonlinear jump SDE equivalent to (4.50) with the truncated coefficients $a_{(n)}, a_{(n,K)}$.

2. Moment Estimates As a first step towards proving a tightness property, we will develop some moment estimates for X_t^n and its law π_t^n. Let us write $\mu_t^n = \text{Law}(V_t^n), \nu_t^n = \text{Law}(\tilde{V}_t^n)$ for the marginals of π_t^n; we claim that μ_t^n satisfies a Boltzmann-type equation. Indeed, for any $f : \mathbb{R}^d \to \mathbb{R}$ bounded, we apply (4.60) with $g(x) = f(v)$ and observe that

$$\mathcal{A}_{n,K}g(x, x_\star) = 2\int_{\mathbb{S}^{d-2}} d\varphi \int_{(0,\infty)} dz \int_{(0,1)} d\alpha (f(v + a_{(n)}(v, v_*, z, \varphi, \alpha)) - f(v)).$$

For any fixed $\alpha \in (0, 1)$, we repeat the calculations of Section 4.2 to rewrite the integrals over z and $\mathbb{S}^{d-2}$ as an integral over $\mathbb{S}^{d-1}$:

$$\begin{aligned}\mathcal{A}_{n,K}g(x, x_\star) &= 2\int_{(0,1)} d\alpha \int_{\mathbb{S}^{d-1}} (f(v') - f(v))1\!\!1_{\theta \geq \theta_0(n)}(n \wedge |v - v_*|^\gamma) b(\cos\theta) d\sigma \\ &= 2\int_{\mathbb{S}^{d-1}} (f(v') - f(v))1\!\!1_{\theta \geq \theta_0(n)}(n \wedge |v - v_*|^\gamma) b(\cos\theta) d\sigma.\end{aligned}$$

By symmetry, when we integrate, we find

$$\langle g, Q^{(2)}_{n,K}(\pi_t^n, \pi_t^n)\rangle = \langle f, Q_{(n)}(\mu_t^n)\rangle$$

where $Q_{(n)}$ is the usual Boltzmann collision operator in one variable for the kernel B_n with cutoff at n and the kinetic factor replaced by $n \wedge |v - v_*|^\gamma$. It follows that μ_t^n solves a Boltzmann-type equation and preserves energy, using the boundedness of the kinetic factor. Following the same calculations leading to (2.62) we see that, for any $p \geq 4$ and some $C = C(p)$ depending only on p and varying line to line,

$$\begin{aligned}\frac{d}{dt}\Lambda_p(\mu_t^n) &\leq C\int_{\mathbb{R}^d\times\mathbb{R}^d} (|v|^{p-1}|v_*| + |v_*|^{p-1}|v|)(n \wedge |v - v_*|^\gamma)\mu_t^n(dv)\mu_t^n(dv_*) \\ &\leq C\Lambda_p(\mu_t^n).\end{aligned} \tag{4.63}$$

Using the initial exponential moment condition, it follows that all moments of V_t^n are bounded, uniformly in n, K and locally uniformly in time. Identical arguments hold for $\tilde{V}_t^n$.

Step 3. Tightness We claim that the processes $X_t^n = (V_t^n, \tilde{V}_t^n)$ are tight in the Skorokhod topology of $\mathbb{D}([0, \infty), (\mathbb{R}^d)^2)$. Indeed, it is sufficient to check that each component $V_t^n, \tilde{V}_t^n$ are tight in the Skorokhod topology of $\mathbb{D}([0, \infty), \mathbb{R}^d)$. For these components, we can repeat the arguments of Lemma 4.13, replacing the moment creation property where necessary by the control over moments of all orders, to conclude that both coordinates $V_t^n, \tilde{V}_t^n$ are tight. Thanks to Prohorov's theorem, we may pass to a subsequence under which $(X_t^n)_{t\geq 0}$ converge in distribution for the Skorokhod topology to a limiting process $(X_t)_{t\geq 0}$, and using a bound analagous to (4.30) again, it follows that for any fixed $t \geq 0$, $\mathbb{P}(X_t \neq X_{t-}) = 0$, so X has no fixed discontinuities. Further applying Skorokhod's representation theorem, we can find $X^{n'}, X'$ with the same laws as X^n, X with almost sure convergence in the Skorokhod topology. We will work with these new processes and, by an abuse of notation, we omit the $'$ to ease notation.

Step 4: Identification of the Limit We now show that any limiting process $X_t = (V_t, \tilde{V}_t)$ constructed in the previous step is a solution to (4.50), for some choice of Poisson random measure $\mathcal{N}$ on E of the correct intensity. First of all, taking the limits of the conclusion of Step 2, it follows that $\mathbb{E}[\sup_{s\le t}|X_s|^p] < \infty$ for all $t \ge 0, p \ge 0$, and since X has no fixed discontinuities, it follows that $X^n_t \to X_t$ almost surely, for any fixed $t \ge 0$, and in paritcular $\pi^n_t = \mathrm{Law}(X^n_t) \to \pi_t = \mathrm{Law}(X_t)$ weakly.

In the rest of the step, we show that the process X solves the nonlinear stochastic differential equation (4.50) for some choice of Poisson random measure $\mathcal{N}$. Thanks to the representation theorem [70], it is sufficient to show that the processes M^f_t given by (4.51) are martingales for all $f \in C^1_p(\mathbb{R}^d \times \mathbb{R}^d)$; let us fix such f, and q such that $|\nabla f(x)| \le C_f(1+|x|^q)$. We start by an analysis of $\mathcal{A}_K f$, in the same way we did for $\mathcal{L}$ in (4.44) in the previous section. We observe that, for any $x = (v, \tilde{v}), x_\star = (v_*, \tilde{v}_*)$, letting $x'_n = x + A_n(x, x_\star, \varphi, z, \alpha) = (v'_n, \tilde{v}'_n)$, we have,

$$|v - v'_n| \le |v - v_\star| \sin G\left(\frac{z}{|v - v_*|^\gamma}\right);$$
$$|\tilde{v} - \tilde{v}'_n| \le |\tilde{v} - \tilde{v}_*| \sin G\left(\frac{z}{|\tilde{v} - \tilde{v}_*|^\gamma}\right).$$

It follows that $|x'_n| \le |x| + |x_\star|$, and we bound

$$\begin{aligned} |f(x'_n) - f(x)| &\le C(1+|x|^q + |x_\star|^q)|x - x'| \\ &\le C(1+|x|^q+|x_\star|^q)\left(|v - v_\star| \sin G\left(\frac{z}{|v-v_*|^\gamma}\right) + |\tilde{v} - \tilde{v}_\star| \sin G\left(\frac{z}{|\tilde{v}-\tilde{v}_*|^\gamma}\right)\right) \end{aligned} \tag{4.64}$$

again uniformly in n. Integrating with respect to α, z, φ and using again that $\int_0^\infty \sin G dz = \int_{\mathbb{S}^{d-1}} \sin\theta b(\cos\theta) d\sigma < \infty$ then produces

$$\begin{aligned} \mathcal{A}_{n,K} f(x, x_\star) &\le C(1+|x|^q+|x_\star|^q)(|v - v_\star|^{1+\gamma} + |\tilde{v} - \tilde{v}_*|^{1+\gamma}) \\ &\le C(1+|x|^q+|x_\star|^q)|x - x_\star|^{1+\gamma} \end{aligned} \tag{4.65}$$

for some C depending only on f, and the same argument applies for $\mathcal{A}_K f$. For convergence of $\mathcal{A}_{n,K}$ to $\mathcal{A}_K$, we fix $R < \infty$, and consider the compact region $|x|, |x_\star| \le R$. For all such $x, x_\star$ and $n \ge (2R)^\gamma$, we continue in the notation above, writing $x' = (v', \tilde{v}') = x + A(x, x_\star, \varphi, z, \alpha)$. We now observe that, for such n, $v'_n = v'$ whenever $z \le n|v - v_*|^\gamma$, and otherwise $|v' - v'_n| = |v' - v|$ which we bound as above. An identical argument holds for $\tilde{v}'_n, \tilde{v}$, checking the cases $z \le K|\tilde{v} - \tilde{v}_*|^\gamma, z > K|\tilde{v} - \tilde{v}_*|^\gamma$ separately, and together we find that

$$|x' - x'_n| \le |v - v_*| \sin G\left(\frac{z}{|v - v_*|^\gamma}\right) \mathbb{I}_{z \ge n|v - v_*|^\gamma} + |\tilde{v} - \tilde{v}_*| \sin G\left(\frac{z}{|\tilde{v} - \tilde{v}_*|^\gamma}\right) \mathbb{I}_{z \ge n|\tilde{v} - \tilde{v}_*|^\gamma}. \tag{4.66}$$

Arguing as in (4.64), we conclude that

$$\begin{aligned}|f(x')-f(x'_n)|\le CR^q\Bigg(&|v-v_*|\sin G\left(\frac{z}{|v-v_*|^\gamma}\right)\mathbb{1}_{z\ge n|v-v_*|^\gamma}\\&+|\tilde{v}-\tilde{v}_*|\sin G\left(\frac{z}{|\tilde{v}-\tilde{v}_*|^\gamma}\right)\mathbb{1}_{z\ge n|\tilde{v}-\tilde{v}_*|^\gamma}\Bigg).\end{aligned} \tag{4.67}$$

If we now integrate over z,α,φ, we find that, for all such $x, x_\star$ and $n\ge 2R$,

$$\begin{aligned}|\mathcal{A}_K f-\mathcal{A}_{n,K}f|(x,x_\star)&\le CR^q\left(|v-v_*|^{1+\gamma}+|\tilde{v}-\tilde{v}_*|^{1+\gamma}\right)\int_n^\infty \sin G dz\\&\le CR^{1+q+\gamma}\int_n^\infty \sin G dz\end{aligned} \tag{4.68}$$

for some $C=C(f)$, and the right-hand side converges to 0. We thus conclude that $\mathcal{A}_{n,K}f\to\mathcal{A}_K f$, uniformly on compact sets of $\mathbb{R}^d\times\mathbb{R}^d$.

We next check continuity. For any given n, let $(x^m,x^m_\star)=((v^m,\tilde{v}^m),(v^m_*,\tilde{v}^m_*))\to(x,x_\star)=((v,\tilde{v}),(v_*,\tilde{v}_*))$ be any convergent sequence in $\mathbb{R}^d\times\mathbb{R}^d$, and suppose that neither component of $x-x_\star=(v-v_*,\tilde{v}-\tilde{v}_*)$ is 0. In this case, we use Lemma 4.9 to see that, for $d\varphi\, dz\, d\alpha$-almost all (φ,z,α), it holds that $A_n(x^m,x^m_\star,\varphi,z,\alpha)\to A_n(x,x_\star,\varphi,z,\alpha)$, and hence $f(x^m+A_n(x^m,x^m_\star,\varphi,z,\alpha))\to f(x+A_n(x,x_\star,\varphi,z,\alpha))$. Further, since $|x^m+A_n(x^m,x^m_\star,\varphi,z,\alpha)|$ is bounded, uniformly in m,α,z,φ, so is $f(x^m+A_n(x^m,x^m_\star,\varphi,z,\alpha))$, and we can apply bounded convergence on the region $\{z\le n^2\}$ to conclude that

$$\begin{aligned}\mathcal{A}_{n,K}f(x^m,x^m_\star)&=2\int_{\mathbb{S}^{d-2}\times(0,\infty)\times(0,1)} d\varphi dz d\alpha(f(x^m+A_n(x^m,x^m_\star,\varphi,z,\alpha)-f(x^m))\\&\to 2\int_{\mathbb{S}^{d-2}\times(0,\infty)\times(0,1)} d\varphi dz d\alpha(f(x^m+A_n(x,x_\star,\varphi,z,\alpha)-f(x))\\&=:\mathcal{A}_{n,K}f(x,x_\star)\end{aligned} \tag{4.69}$$

so that $\mathcal{A}_{n,K}$ is continuous at such points. The same argument applies in the case where $\tilde{v}-\tilde{v}_*=0$: even though we may not have convergence of $R_\alpha(v^m-v^m_*,\tilde{v}^m-\tilde{v}^m_*)$, it still holds that

$$|a_{(n,K)}(\tilde{v}^m,\tilde{v}^m_*,R_\alpha(v^m-v^m_*,\tilde{v}^m-\tilde{v}^m_*)\varphi,z)|\le|\tilde{v}^m-\tilde{v}^m_*|\to 0$$

and so

$$a_{(n,K)}(\tilde{v}^m,\tilde{v}^m_*,R_\alpha(v^m-v^m_*,\tilde{v}^m-\tilde{v}^m_*)\varphi,z)\to 0=a_{(n,K)}(\tilde{v},\tilde{v}_*,R_\alpha(v-v_*,\tilde{v}-\tilde{v}_*)\varphi,z).$$

In the case where $v-v_*=0$, we observe that $(v^m,\tilde{v}^m+a_{(n,K)}(\tilde{v}^m,\tilde{v}^m_*,z,R_\alpha(v-v_*,\tilde{v}-\tilde{v}_*)\varphi,\alpha))$ differs from $x^m+A_n(x^m,x^m_\star,\varphi,z,\alpha)$ by at most $|v^m-v^m_*|\to 0$ in the first component, and it follows, using the Lipschitz property and dominated convergence in

$\{z \leq n^2\}$ that

$$\begin{aligned}\mathcal{A}_{n,K}f(x^m, x^m_\star) \\ - 2\int_{\mathbb{S}^{d-2}\times(0,\infty)\times(0,1)} d\varphi dz d\alpha (f(v^m, \tilde{v}^m + a_{(n,K)}(\tilde{v}^m, \tilde{v}^m_*, z, R_\alpha(v - v_*, \tilde{v} - \tilde{v}_*)\varphi, \alpha)) \\ - f(x^m)) \\ \to 0.\end{aligned} \tag{4.70}$$

Using the fact that each R_α preserves the uniform measure $d\varphi$, we rewrite the integral as

$$\begin{aligned}2\int_{\mathbb{S}^{d-2}\times(0,\infty)\times(0,1)} d\varphi dz d\alpha (f(v^m, \tilde{v}^m + a_{n,K}(\tilde{v}^m, \tilde{v}^m_*, z, R_\alpha(v - v_*, \tilde{v} - \tilde{v}_*)\varphi, \alpha)) - f(x^m)) \\ = 2\int_{\mathbb{S}^{d-2}\times(0,\infty)\times(0,1)} d\varphi dz d\alpha (f(v^m, \tilde{v}^m + a_{n,K}(\tilde{v}^m, \tilde{v}^m_*, z, \varphi, \alpha)) - f(x^m)).\end{aligned} \tag{4.71}$$

Arguing using the continuity of $a_{(n,K)}$ at $(\tilde{v}, \tilde{v}_*)$ for $d\varphi dz d\alpha$-lamost all φ, z, α as above, it follows that we have the convergence

$$\begin{aligned}\mathcal{A}_{n,K}f(x^m, x^m_\star) \to 2\int_{\mathbb{S}^{d-2}\times(0,\infty)\times(0,1)} d\varphi dz d\alpha (f(v, \tilde{v} + a_{n,K}(\tilde{v}, \tilde{v}_*, \varphi, z, \alpha)) - f(x)) \\ = 2\int_{\mathbb{S}^{d-2}\times(0,\infty)\times(0,1)} d\varphi dz d\alpha (f(v, \tilde{v} + a_{n,K}(\tilde{v}, \tilde{v}_*, R_\alpha(0, \tilde{v} - \tilde{v}_*)\varphi, z, \alpha)) - f(x)) \\ = \mathcal{A}_{n,K}f(x, x_\star).\end{aligned} \tag{4.72}$$

Since we have checked all possible cases, it follows that $\mathcal{A}_{n,k}f$ is continuous everywhere. Together with the local uniform convergence, it also follows that $\mathcal{A}_K f$ is also continuous.

We now check convergence of the martingales $M^{n,f}_t$ to the martingales identified by (4.51). By a truncation argument using (4.65) and using Lemma 4.14, it follows that $\int_{\mathbb{R}^d\times\mathbb{R}^d} \mathcal{A}_{n,K}f(x,y)\pi^n_t(dy) \to \int_{\mathbb{R}^d\times\mathbb{R}^d} \mathcal{A}_K f(x,y)\pi_t(dy)$, uniformly on compact sets in x. Further using that $\mathcal{A}_K f(x,y)$ is continuous and a further truncation argument, at all points of continuity t of the limit process X, we have that $X^n_t \to X_t$ and so $\int_{\mathbb{R}^d\times\mathbb{R}^d} \mathcal{A}_K f(X^n_t, y)\pi_t(dy) \to \int_{\mathbb{R}^d\times\mathbb{R}^d} \mathcal{A}_K f(X_t, y)\pi_t(dy)$. Combining everything, almost surely, for dt-almost all t,

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \mathcal{A}_{n,K}f(X^n_t, y)\pi^n_t(dy) \to \int_{\mathbb{R}^d\times\mathbb{R}^d} \mathcal{A}_K f(X^n_t, y)\pi_t(dy) \tag{4.73}$$

and so

$$\int_0^t \int_{\mathbb{R}^d\times\mathbb{R}^d} \mathcal{A}_{n,K}f(X^n_s, y)\pi^n_s(dy)ds \to \int_0^t \int_{\mathbb{R}^d\times\mathbb{R}^d} \mathcal{A}_K f(X^n_s, y)\pi_s(dy)ds \tag{4.74}$$

almost surely, uniforly on compact time intervals. Meanwhile, since f is continuous, it follows that $f(X^n_t) - f(X^n_0) \to f(X_t) - f(X_0)$ almost surely in the Skorokhod topology, and together we conclude that the martingales $M^{f,n}_t$ identified in Step 1 converge almost surely in the Skorokhod topology to

$$M^f_t = f(X_t) - f(X_0) - \int_0^t \int_{\mathbb{R}^d\times\mathbb{R}^d} \mathcal{A}_K f(X_s, y)\pi_s(dy)ds \tag{4.75}$$

as desired. Further, recalling again that for each $t \geq 0$, X is almost surely continuous at t, it follows that M^f is almost surely continuous at t, and the convergence of $M^{f,n}_t$ to M^f_t is almost sure. It is straightforward to check, using the moment estimates of all orders obtained in Step 2, that for fixed f, t, $M^{f,n}_t$ are bouded in $L^2(\mathbb{P})$, and hence converge in $L^1(\mathbb{P})$ to M^f_t. We finally conclude that $(M^f_t)_{t\geq 0}$ is a martingale, and the step is complete. The conclusion that π_t solves (4.54) follows by taking expectations.

Step 5: The Limit Process couples the given solutions We finally check that $\pi_t = \text{Law}(V_t, \tilde{V}_t)$ is a coupling of the prescribed solutions μ_t, ν_t. Let us write $\widetilde{\mu}_t, \widetilde{\nu}_t$ for the two marginals of π_t. For a fixed $f \in C^1_b(\mathbb{R}^d)$, we set $g(v, \tilde{v}) := f(v)$ and remark as in Step 2 that

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \mathcal{A}_K g(x, y)\pi_t(dx)\pi_t(dy) = \langle f, Q(\widetilde{\mu}_t)\rangle \tag{4.76}$$

for the usual (noncutoff) Boltzmann collision operator Q, and we obtain from (4.54) that $\langle f, \widetilde{\mu}_t\rangle = \langle f, \mu_0\rangle + \int_0^t \langle f, Q(\widetilde{\mu}_s)\rangle ds$ and $(\widetilde{\mu}_t)_{t\geq 0}$ is a solution to (BE) starting at μ_0. Further, each $\mu^n_t = \text{Law}(V^n_t)$ conserves energy, and we can take limits using the boundedness of higher moments to conclude that μ_t conserves energy and, in particular, takes values in $\mathcal{S}$. Since μ_0 is assumed to have an exponential moment $\int_{\mathbb{R}^d} e^{\varepsilon|v|^\gamma}\mu_0(dv) < \infty$, Proposition 4.18 applies, and we conclude that $\widetilde{\mu}_t = \mu_t$ as desired. Similarly, $\widetilde{\nu}_t$ solves (BE_K), using Corollary 3.13 if $K < \infty$ or arguing as above if $K = \infty$, and the conclusion that $\widetilde{\nu}_t = \nu_t$ is identical. □

We next check that this produces a coupling of Boltzmann processes. This is the content of the following lemma, which is adapted from a similar claim [92, Proposition 4.4].

Lemma 4.20. *Continue in the notation of Lemma 4.16, and let $(V_t, \tilde{V}_t)_{t\geq 0}$ be a solution to (4.50). Then $(V_t)_{t\geq 0}$ is a solution to (stBE^a), and $(\tilde{V}_t)_{t\geq 0}$ is a solution to (stBE$^a{}_K$).*

Proof. Let us argue for $(\tilde{V}_t)_{t\geq 0}$; the case for $(V_t)_{t\geq 0}$ is strictly simpler. For $(\tilde{V}_t)_{t\geq 0}$, let $\mathcal{N}$ be the Poisson random measure of intensity $2dt\pi_t(dx)d\varphi dz d\alpha$ driving the given solution $(V_t, \tilde{V}_t)$, and define random measures $\widehat{\mathcal{N}}$ on $(0,\infty)\times\mathbb{R}^d\times\mathbb{S}^{d-2}\times(0,\infty)$ by specifying, for bounded and compactly supported $f:(0,\infty)\times\mathbb{R}^d\times\mathbb{S}^{d-2}\times(0,\infty)\to\mathbb{R}$,

$$\int_{(0,\infty)\times\mathbb{R}^d\times\mathbb{S}^{d-2}\times(0,\infty)} f(t,v,\varphi,z)\widehat{\mathcal{N}}(dt,dv,d\varphi,dz)$$
$$= \int_{(0,\infty)\times(\mathbb{R}^d)^2\times\mathbb{S}^{d-2}\times(0,\infty)\times(0,1)} f(t,v,R_{\alpha,s-}(v,\tilde{v})\varphi,z)\mathcal{N}(dt,dv,d\tilde{v},d\varphi,dz,d\alpha) \tag{4.77}$$

where

$$R_{\alpha,t}(v,\tilde{v}) := R_\alpha(V_t-v,\tilde{V}_t-\tilde{v}).$$

It follows immediately that $\tilde{V}_t$ solves (stBE^a) for this measure, so we must now show that it is a Poisson random measure of the correct intensity. If H is a bounded and compactly supported previsible function on $(0,\infty)\times\mathbb{R}^d\times\mathbb{S}^{d-2}\times(0,\infty)$, then the function

$$\widehat{H}(s,v,\tilde{v},\varphi,z) := H(s,v,R_{\alpha,s-}(v,\tilde{v}),z)$$

is bounded and previsible on $(0,\infty)\times\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-2}\times(0,\infty)$. Recalling that the first marginal of π_t is μ_t and noting that $R_{\alpha,t}(v,\tilde{v})$ preserves the uniform measure $d\varphi$, we obtain

$$\int_{(0,t]\times\mathbb{R}^d\times\mathbb{S}^{d-2}\times(0,\infty)} H(s,v,\varphi,z)(\widehat{\mathcal{N}}(dt,dv,d\varphi,dz)-2ds\mu_s(dv)d\varphi dz)$$
$$= \int_{(0,t]\times\mathbb{R}^d\times\mathbb{S}^{d-2}\times(0,\infty)\times(0,1)} \widehat{H}(s,v,\tilde{v},\varphi,z)(\mathcal{N}(ds,dv,d\tilde{v},d\varphi,dz,d\alpha)-2ds\pi_s(dv,d\tilde{v})d\varphi dz d\alpha). \tag{4.78}$$

The latter process is a martingale by the stochastic calculus of Poisson processes, which implies that $\widehat{\mathcal{N}}$ is a Poisson random measure on $(0,\infty)\times\mathbb{R}^d\times\mathbb{S}^{d-2}\times(0,\infty)$ of rate $2dt\mu_t(dv)d\varphi dz$, as desired. □

4.6 Proof of Theorem 4.1

We now prove Theorem 4.1. We begin with the case where both μ_0,ν_0 have an exponential moment, and then carefully remove this hypothesis.

Lemma 4.21. *There exists $p_0=p_0(B,d)$ such that, whenever $p\geq p_0$ and $K\in[K_0(B,p,d),\infty]$, for K_0 found in Lemma 4.10, the following holds. For all weak solutions $(\mu_t)_{t\geq 0},(\nu_t)_{t\geq 0}\subset\mathcal{S}$ to (BE, BE$_K$) respectively and with exponential moments $\int_{\mathbb{R}^d}e^{\varepsilon|v|^\gamma}(\mu_0+\nu_0)(dv)<\infty$ for some $\varepsilon>0$, there exists $(V_t)_{t\geq 0},(\tilde{V}_t)_{t\geq 0}$ solving (stBE^a, stBE$^a{}_K$) respectively, such that $\pi_t=\mathrm{Law}(V_t,\tilde{V}_t)$ is a coupling of $\pi_t\in\Pi(\mu_t,\nu_t)$. Furthermore, this coupling achieves*

$$\mathbb{E}[d_p(V_t,\tilde{V}_t)]\leq e^{C(1+t)\Lambda_{p+\gamma}(\mu_0,\nu_0)}\left(w_p(\mu_0,\nu_0)+tK^{1-1/\nu}\Lambda_l(\mu_0,\nu_0)\right) \tag{4.79}$$

where $l=p+2+\gamma$, understanding the second term to be 0 in the noncutoff case $K=\infty$.

Proof. As remarked in Section 2.1, we can find $\pi_0 \in \Pi(\mu_0, \nu_0)$ attaining the minimum

$$w_p(\mu_0, \nu_0) = \int_{\mathbb{R}^d \times \mathbb{R}^d} d_p(v, \tilde{v}) \pi_0(dv, d\tilde{v}).$$

We can now apply Lemma 4.16 to construct the pair $(V_t, \tilde{V}_t)_{t \geq 0}$, and the statement that each component $(V_t)_{t\geq 0}, (\tilde{V}_t)_{t\geq 0}$ solves (stBE^a, stBE$^a{}_K$) is exactly Lemma 4.20, and it only remains to prove (4.79). We will write expressions as though $K < \infty$, understanding that negative powers of K are 0 if $K = \infty$.

We write

$$u_p(t) = \mathbb{E}[d_p(V_t, \tilde{V}_t)] = \int_{\mathbb{R}^d \times \mathbb{R}^d} d_p(v, \tilde{v}) \pi_t(dv, d\tilde{v}). \tag{4.80}$$

By (4.54), it follows that

$$u_p(t) = u_p(0) + \int_0^t \int_{(\mathbb{R}^d \times \mathbb{R}^d)^2} \mathcal{A}_K d_p(v, \tilde{v}, v_*, \tilde{v}_*) \pi_s(dv, d\tilde{v}) \pi_s(dv_*, d\tilde{v}_*) ds.$$

With the definitions in 4.10, we have

$$\mathcal{A}_K d_p(v, \tilde{v}, v_*, \tilde{v}_*) = \int_{(0,1)} \mathcal{E}_{p,K}(v, \tilde{v}, v_*, \tilde{v}_*, \alpha) d\alpha$$

and the integrand is bounded, uniformly in α, by (4.25). Integrating and grouping similar terms, we find, up to new choices of c, C,

$$\begin{aligned} u_p(t) \leq & w_p(\mu_0, \nu_0) + \int_0^t \int_{\mathbb{R}^d \times \mathbb{R}^d} \left(c - \frac{\lambda_p}{2}\right) d_{p+\gamma}(v, \tilde{v}) \pi_s(dv, d\tilde{v}) ds \\ & + C \int_{\mathbb{R}^d \times \mathbb{R}^d} \int_{\mathbb{R}^d \times \mathbb{R}^d} (|v_*|^{p+\gamma} + |\tilde{v}_*|^{p+\gamma}) d_p(v, \tilde{v}) \pi_s(dv, d\tilde{v}) \pi_s(dv_*, d\tilde{v}_*) ds \\ & + C K^{1-1/\nu} \int_0^t \int_{(\mathbb{R}^d \times \mathbb{R}^d)^2} (1 + |v|^l + |\tilde{v}|^l + |v_*|^l + |\tilde{v}_*|^l) \pi_s(dv, d\tilde{v}) \pi_s(dv_*, d\tilde{v}_*) ds \\ & =: w_p(\mu_0, \nu_0) + \int_0^t (I_1(s) + I_2(s) + I_3(s)) ds \end{aligned} \tag{4.81}$$

where

$$I_1(s) = \left(c - \frac{\lambda_p}{2}\right) \int_{\mathbb{R}^d \times \mathbb{R}^d} d_{p+\gamma}(v, \tilde{v}) \pi_s(dv, d\tilde{v}); \tag{4.82}$$

$$I_2(s) = C \int_{(\mathbb{R}^d \times \mathbb{R}^d)^2} (|v_*|^{p+\gamma} + |\tilde{v}_*|^{p+\gamma}) d_p(v, \tilde{v}) \pi_s(dv, d\tilde{v}) \pi_s(dv_*, d\tilde{v}_*); \tag{4.83}$$

$$I_3(s) = C K^{1-1/\nu} \int_{(\mathbb{R}^d \times \mathbb{R}^d)^2} (1 + |v|^l + |\tilde{v}|^l + |v_*|^l + |\tilde{v}_*|^l) \pi_s(dv, d\tilde{v}) \pi_s(dv_*, d\tilde{v}_*). \tag{4.84}$$

Let us now choose p. We recall that c does *not* depend on p, and return to the definition

$$\lambda_p := \int_0^{\pi/2} \left(1 - \left(\frac{1 + \cos\theta}{2}\right)^{p/2}\right) \beta(\theta) d\theta. \tag{4.85}$$

As $p\to\infty$, the term in parentheses converges up to 1 for any $\theta\neq 0$, and so λ_p converges to $\int_0^{\pi/2}\beta(\theta)d\theta=\infty$ by monotone convergence. In particular, there exists some p_0, depending only on B,d such that, for all $p>p_0$, $\lambda_p\geq c$, and for such p, $I_1(s)\leq 0$ for all s, so that $u_p(t)\leq u_p(0)+\int_0^t(I_2(s)+I_3(s))ds$. Meanwhile, recalling that $\pi_t\in\Pi(\mu_t,\nu_t)$, we control

$$I_2(s)\leq C\Lambda_{p+\gamma}(\mu_s,\nu_s)\int_{\mathbb{R}^d\times\mathbb{R}^d}d_p(v,\tilde{v})\pi_s(dv,d\tilde{v})=C\Lambda_{p+\gamma}(\mu_s,\nu_s)u_p(s) \tag{4.86}$$

and

$$I_3(s)=CK^{1-1/\nu}\Lambda_l(\mu_s,\nu_s). \tag{4.87}$$

We thus find that

$$u_p(t)\leq w_p(\mu_0,\nu_0)+C\int_0^t((1+\Lambda_{p+\gamma}(\mu_s,\nu_s))u_p(s)+K^{1-1/\nu}\Lambda_l(\mu_s,\nu_s))ds$$

and by the Grönwall lemma,

$$u_p(t)\leq \exp\left(C\int_0^t(1+\Lambda_{p+\gamma}(\mu_s,\nu_s))ds\right)\left(w_p(\mu_0,\nu_0)+CK^{1-1/\nu}\int_0^t(1+\Lambda_l(\mu_s,\nu_s))ds\right).$$

Finally, using the moment propagation properties for either (BE, BE_K) in Section 2.5, we can replace $\Lambda_{p+\gamma}(\mu_s,\nu_s)\leq C\Lambda_{p+\gamma}(\mu_0,\nu_0)$ and $\Lambda_l(\mu_s,\nu_s)\leq C\Lambda_l(\mu_0,\nu_0)$. Absorbing all constants C into the exponent, we finally find (4.79) as desired. □

It finally remains to relax the assumption that μ_0,ν_0 have exponential moments, which we used in the construction of the coupling. In order to do so, we will use the exponential moment creation property in Proposition 2.13 to apply the previous result at time s, and carefully take a limit along some sequence $s\downarrow 0$. We will use the following intermediate lemma.

Lemma 4.22. *Let $(\mu_t)_{t\geq 0}\subset\mathcal{S}$ be a weak solution to the Boltzmann equation (BE) or (BE_K), suppose that $\Lambda_{p+2}(\mu_0)<\infty$ for some $p>0$. Then $w_p(\mu_t,\mu_0)\to 0$.*

Proof. Let us fix $\varepsilon>0$. Using (BE) and the duality (2.25) and the estimate $\langle f,Q(\mu_s)\rangle\leq C(f)$, for all Lipschitz f and $s\geq 0$, it follows that $\mathcal{W}_1(\mu_t,\mu_0)\to 0$ as $t\downarrow 0$, so we can find a coupling $\rho_t\in\Pi(\mu_t,\mu_0)$ attaining the infimum $\mathcal{W}_1(\mu_t,\mu_0)=\int_{\mathbb{R}^d\times\mathbb{R}^d}|v-\tilde{v}|\rho_t(dv,d\tilde{v})$. Now, we apply Proposition 2.15 on the moments of order $p+2$ to find $R>0$ and $t_0>0$ such that, for all $t\in[0,t_0)$,

$$\int_{\mathbb{R}^d}(1+|v|^{p+2})\mu_t(dv)<\varepsilon. \tag{4.88}$$

Next, by considering cases separately where $|v|,|w|\leq R$ or where one or both exceed R, we observe that, for some absolute constant C,

$$d_p(v,w)\leq C(1+R^{p+1})|v-w|+C(1+|v|^{p+2})\mathbb{1}_{|v|\geq R}+C(1+|w|^{p+2})\mathbb{1}_{|w|\geq R}. \tag{4.89}$$

Integrating with respect to $\rho_t(dv,dw)$, we find that

$$\begin{aligned} w_p(\mu_t,\mu_0) &\le \int_{\mathbb{R}^d\times\mathbb{R}^d} d_p(v,w)\rho_t(dv,dw) \\ &\le C(1+R^{p+1})\mathcal{W}_1(\mu_t,\mu_0) + C\int_{\mathbb{R}^d\times\mathbb{R}^d}(1+|v|^{p+2})\mathbb{1}_{|v|>R}\rho_t(dv,dw) \\ &\qquad + C\int_{\mathbb{R}^d\times\mathbb{R}^d}(1+|w|^{p+2})\mathbb{1}_{|w|>R}\rho_t(dv,dw) \\ &\le C(1+R^{p+1})\mathcal{W}_1(\mu_t,\mu_0) + C\int_{\mathbb{R}^d}(1+|v|^{p+2})\mathbb{1}_{|v|>R}\mu_0(dv) \\ &\qquad + C\int_{\mathbb{R}^d\times\mathbb{R}^d}(1+|w|^{p+2})\mathbb{1}_{|w|>R}\mu_t(dw) \end{aligned} \tag{4.90}$$

using, in the last equality, that the marginals of ρ_t are μ_0,μ_t respectively. We conclude that, for $t\le t_0$,

$$w_p(\mu_0,\mu_t) \le C(1+R^{p+1})\mathcal{W}_1(\mu_t,\mu_0) + 2C\varepsilon$$

and hence $\limsup_{t\to 0} w_p(\mu_t,\mu_0)\le 2C\varepsilon$. Since $\varepsilon>0$ was arbitrary, we are done. □

We are now ready to give the proof of Theorem 4.1.

Proof of Theorem 4.1. Let us fix $p_0=p_0(B,d)$ as in Lemma 4.21 and choose $p>p_0, K\ge K_0(B,p,d)$. Let $(\mu_t)_{t\ge 0},(\nu_t)_{t\ge 0}$ be two solutions such that the initial data μ_0,ν_0 have $p+2$ moments. Let us first consider the case $K=\infty$ where we couple noncutoff solutions; the case where $K<\infty$ will be discussed at the end. As in the theorem, let us write $a_1\ge\Lambda_{p+\gamma}(\mu_0,\nu_0)$ for an upper bound on the initial $(p+\gamma)^{\text{th}}$ moments.

Fix $t>0$ and $0<s<1$. Thanks to the exponential moment creation Proposition 2.13, we have $\int_{\mathbb{R}^d} e^{\varepsilon|v|^\gamma}(\mu_s+\nu_s)(dv)<\infty$ for some $\varepsilon=\varepsilon(s)>0$, and we can apply Lemma 4.16 starting at μ_s,ν_s, to find $(V^s_t,\tilde V^s_t)$, where each coordinate solves (stBE^a) and $\pi^s_t=\text{Law}(V^s_t,\tilde V^s_t)\in\Pi(\mu_{s+t},\nu_{s+t})$, and by Lemma 4.21, we have

$$\mathbb{E}[d_p(V^s_t,\tilde V^s_t)] \le e^{C(1+t)\Lambda_{p+\gamma}(\mu_s,\nu_s)}w_p(\mu_s,\nu_s) \le e^{C(1+t)a_1}w_p(\mu_s,\nu_s) \tag{4.91}$$

by using the moment propagation property and modifying C if necessary. Now, Lemma 4.13 applies to each process $(V^s_t)_{t\ge0},(\tilde V^s_t)_{t\ge 0}$, so we can find a sequence $s_n\to 0$ such that $(V^{s_n}_t)_{t\ge 0}$ converge in distribution in the Skorokhod topology of $\mathbb{D}([0,\infty),\mathbb{R}^d)$ to a limit $(V_t)_{t\ge0}$, and similarly $(\tilde V^{s_n}_t)_{t\ge 0}\to(\tilde V_t)_{t\ge 0}$; using Lemma 4.13 again, both are solutions to (stBE^a). Further, since $(V_t,\tilde V_t)$ has no fixed discontinuities, for each t it holds that $(V^{s_n}_t,\tilde V^{s_n}_t)\to(V_t,\tilde V_t)$ in distribution, so the laws $\pi^{s_n}_t\to\pi_t=\text{Law}(V_t,\tilde V_t)$. We deduce that $\mu_{s_n+t}=\text{Law}(V^{s_n}_t)\to\text{Law}(V_t)$ in the weak topology, and since $u\mapsto\mu_u$ is continuous for the weak topology, $\text{Law}(V_t)=\mu_t$. Similarly, $\text{Law}(\tilde V_t)=\nu_t$, so $\pi_t=\text{Law}(V_t,\tilde V_t)\in\Pi(\mu_t,\nu_t)$.

We now take the limit of (4.91). Recalling the relaxed triangle inequality (2.19), we have for some $C = C(p)$

$$w_p(\mu_s, \nu_s) \le C(w_p(\mu_s, \mu_0) + w_p(\mu_0, \nu_0) + w_p(\nu_0, \nu_s)) \tag{4.92}$$

and by Lemma 4.22, the first and third terms converge to 0, and so

$$\limsup_{s\to 0} w_p(\mu_s, \nu_s) \le C w_p(\mu_0, \nu_0).$$

Meanwhile, for $t > 0$, we can use the boundedness of moments of all orders of $\mathrm{Law}(V_t^{s_n}) = \mu_{t+s_n}$ and $\mathrm{Law}(\tilde{V}_t^{s_n}) = \nu_{t+s_n}$ to see that $\mathbb{E}[d_p(V_t^{s_n}, \tilde{V}_t^{s_n})] \to \mathbb{E}[d_p(V_t, \tilde{V}_t)]$, and take the limit of (4.91) to conclude that

$$\mathbb{E}[d_p(V_t, \tilde{V}_t)] \le e^{Ca_1(1+t)} w_p(\mu_0, \nu_0)$$

which is exactly (4.2) as desired.

Let us now deal with the case $K < \infty$, so that ν_t solves a cutoff Boltzmann equation (BE_K); in this case, let $l = p + 2 + \gamma$ as in Lemma 4.21 and let $a_2 \ge 1$ be an upper bound for $\Lambda_l(\nu_0)$. We could apply the previous argument wholesale, using Proposition 2.6 to verify the integrability condition on ν_s, which would produce a final bound depending on the l^{th} moments of both μ_0, ν_0; instead, we will go via an intermediate solution $\widetilde{\nu}_t$, which has the same initial data as ν_t but no cutoff.

As remarked in the introduction, existence of solutions to (BE) for kernels (NCHP) is known, so we can let $(\widetilde{\nu}_t)_{t\ge 0}$ be a solution to the full, non-cutoff Boltzmann equation with the same initial data $\widetilde{\nu}_0 = \nu_0$. We then repeat the arguments above on $(\widetilde{\nu}_t, \nu_t)$, using Proposition 2.6 to verify the integrability condition on ν_s; in place of (4.91) we find

$$\mathbb{E}[d_p(U_t^s, \tilde{U}_t^s)] \le e^{Ca_1(1+t)}(w_p(\widetilde{\nu}_s, \nu_s) + a_2 t K^{1-1/\nu}) \tag{4.93}$$

where again we have replaced the moments of $\widetilde{\nu}_s, \nu_s$ by those of $\widetilde{\nu}_0 = \nu_0$, up to a new choice of C, and we extract a convergent subsequence $s_n \to 0$ exactly as before to find $(U_t, \tilde{U}_t)_{t\ge 0}$, whose components solve (stBE^a, $\mathrm{stBE}^a{}_K$) respectively, such that $\mathrm{Law}(U_t, \tilde{U}_t) \in \Pi(\widetilde{\nu}_t, \nu_t)$, and taking the limit of (4.93) gives

$$\mathbb{E}[d_p(U_t, \tilde{U}_t)] \le e^{C(1+t)a_1} t K^{1-1/\nu}.$$

Using the case without cutoff as we did previously, we find $(W_t, \tilde{W}_t)_{t\ge 0}$ for $(\mu_t)_{t\ge 0}, (\widetilde{\nu}_t)_{t\ge 0}$, and with the same conclusion

$$\mathbb{E}[d_p(W_t, \tilde{W}_t)] \le e^{C(1+t)a_1} w_p(\mu_0, \nu_0).$$

Now, by Lemma 4.15, since U_t and $\tilde{W}_t$ both solve (stBE^a) with $\mathrm{Law}(U_t) = \mathrm{Law}(\tilde{W}_t) = \widetilde{\nu}_t$, it follows that $\mathrm{Law}((U_t)_{t\ge 0}) = \mathrm{Law}((\tilde{W}_t)_{t\ge 0})$. Using the gluing lemma [191, Lemma 7.6], there exists a triple of stochastic processes $(V_t), (\widehat{V}_t), (\tilde{V}_t)$, such that the law of $(V_t, \widehat{V}_t)_{t\geq 0}$ is the same as that of $(W_t, \tilde{W}_t)_{t\geq 0}$ and the law of $(\widehat{V}_t, \tilde{V}_t)_{t\geq 0}$ is the same as that of $(U_t, \tilde{U}_t)$. The coupling $(V_t, \tilde{V}_t)_{t\geq 0}$ therefore has $\pi_t = \text{Law}(V_t, \tilde{V}_t) \in \Pi(\mu_t, \nu_t)$, and the components solve (stBE^a, stBE$^a{}_K$) respectively. Recalling again the relaxed triangle inequality (2.17) for d_p, we combine the previous two displays and absorb the additional factor into the exponent to find

$$\mathbb{E}[d_p(V_t, \tilde{V}_t)] \leq e^{Ca_1(1+t)}(w_p(\mu_0, \nu_0) + a_2 t K^{1-1/\nu})$$

as desired. □

4.7 Equivalence of the Boltzmann Equation and Bolzmann Processes

We now give the proof of Theorem 4.3, as we have already developed the tools to do so in the course of proving Theorem 4.1. As already remarked, it is straightforward to see that, if (V_t) is a solution to (stBE^a) with $\mu_t = \text{Law}(V_t) \in \mathcal{S}$ for all t, then μ_t satisfies (BE), for instance by taking expectations of the martingales $M_t^f = f(V_t) - f(V_0) - \int_0^t \mathcal{L}f(V_s, v_*)\mu_s(dv_*)ds$ and recalling that $\int_{\mathbb{R}^d\times\mathbb{R}^d} \mathcal{L}f(v, v_*)\mu_s(dv)\mu_s(dv_*) = \langle f, Q(\mu_s)\rangle$. We now prove the other implication.

Proof of Theorem 4.3. Let $(\mu_t)_{t\geq 0}$ be a weak solution to (BE), taking values in $\mathcal{S}$. For all $s > 0$, μ_s has an exponential moment $\int_{\mathbb{R}^d} e^{\varepsilon|v|^\gamma}\mu_s(dv) < \infty$, for some $\varepsilon > 0$, by Proposition 2.13, and so we can find a solution $(V_t^s)_{t\geq 0}$ to (stBE^a) with $\text{Law}(V_t^s) = \mu_{s+t}$, either repeating the arguments of Lemma 4.16 or by applying the cited result with $\nu_t \in \mathcal{S}$ an arbitrary solution to (BE). Thanks to Lemma 4.13, we can find $s_n \to 0$ and a process $(V_t)_{t\geq 0}$ which is the limit in distribution of $(V_t^{s_n})_{t\geq 0}$ in the Skrokhod topology of $\mathbb{D}([0,\infty), \mathbb{R}^d)$. By Lemma 4.13 again, $(V_t)_{t\geq 0}$ is a solution to (stBE^a), and we can take weak limits of $\text{Law}(V_t^{s_n}) = \mu_{t+s_n}$ to obtain $\text{Law}(V_t) = \mu_t$ as in the proof of Theorem 4.1 above. □

4.8 Tanaka Coupling of the Kac Process

We now apply the Tanaka coupling we have developed to the Kac process, which will prove the two main Theorems 4.4, 4.5.

4.8.1 Coupling of the Kac Process

We first give a Tanaka-coupling of the Kac processes, in the same spirit as Lemma 4.16 for Boltzmann processes. Let $\mathcal{V}^N_t=(V^1_t,...,V^N_t)$ be a noncutoff Kac process, and let $\mathcal{N}^{\{ij\}}$, $1\le i\neq j\le N$, be the Poisson random measures of intensity $2N^{-1}dtd\varphi dz$ driving $\mathcal{V}^N_t$, so that

$$V^i_t=V^i_0+\sum_{j\neq i}\int_{(0,t]\times\mathbb{S}^{d-2}\times(0,\infty)} a(V^i_{s-},V^j_{s-},z,\varphi)\,\mathcal{N}^{\{ij\}}(ds,d\varphi,dz), \qquad i=1,...,N. \tag{4.94}$$

Let us fix $\widetilde{\mathcal{V}}^{N,K}_0=(\widetilde{V}^{1,K}_0,....,\widetilde{V}^{N,K}_0)$ and define $\widetilde{\mathcal{V}}^{N,K}_t=(\widetilde{V}^{1,K}_t,...\widetilde{V}^{N,K}_t)$ by

$$\widetilde{V}^{i,K}_t=\widetilde{V}^{i,K}_0+\sum_{j\neq i}\int_{(0,t]\times\mathbb{S}^{d-2}\times(0,\infty)} a_K(\widetilde{V}^{i,K}_{s-},\widetilde{V}^{j,K}_{s-},z,R^{i,j}_{s-}\varphi)\,\mathcal{N}^{\{ij\}}(ds,d\varphi,dz); \tag{4.95}$$

$$R^{i,j}_t:=R(V^i_t-V^j_t,\widetilde{V}^{i,K}_t-\widetilde{V}^{j,K}_t) \tag{4.96}$$

where $R:\mathbb{R}^d\times\mathbb{R}^d\to \mathrm{Isom}(\mathbb{S}^{d-2})$ is the isometry on $\mathbb{S}^{d-2}$ constructed in Lemma 4.8. We remark first that the rates of $\mathcal{N}^{\{ij\}}$ are all finite on the support of a_K, so that the stochastic differential equation (4.95, 4.96) is really a recurrence relation; in particular, $\widetilde{\mathcal{V}}^{N,K}_t$ is uniquely defined by the above equations. Next, we claim that $\widetilde{\mathcal{V}}^{N,K}_t$ is a K-cutoff Kac process on N particles. It certainly holds that $\widetilde{\mathcal{V}}^{N,K}_t$ satisfies (cLK) for the measures $\widetilde{\mathcal{N}}^{\{ij\}}$ given by specifying, for bounded and compactly supported $f:(0,\infty)\times\mathbb{S}^{d-2}\times(0,\infty)\to\mathbb{R}$,

$$\begin{aligned}\int_{(0,\infty)\times\mathbb{S}^{d-2}\times(0,\infty)} & f(s,\varphi,z)\widetilde{\mathcal{N}}^{\{ij\}}(ds,d\varphi,dz)\\ &:=\int_{(0,\infty)\times\mathbb{S}^{d-2}\times(0,\infty)} f(s,R^{ij}_{s-}\varphi,z)\mathcal{N}^{\{ij\}}(ds,d\varphi,dz).\end{aligned} \tag{4.97}$$

Repeating the arguments of Lemma 4.20 and using the fact that $R^{i,j}_{s-}$ are previsible, we see that $\widetilde{\mathcal{N}}^{\{ij\}}$ are Poisson random measures on $(0,\infty)\times\mathbb{S}^{d-2}\times(0,\infty)$ of the correct intensity $2dtd\varphi dz$, so that $\widetilde{\mathcal{V}}^{N,K}_t$ is a cutoff Kac process as desired.

Our first result on the coupling is the following, which proves Theorem 4.4.

Lemma 4.23 (Convergence of the Tanaka Coupling)**.** *For the same $p_0=p_0(B,d)$ found in Lemma 4.21 and, for $p>p_0$, $K\in[K_0(B,p,d),\infty)$, for K_0 as in Lemma 4.10 such that, whenever $p>p_0$ and $K>K_0$, we have the following estimates.*

Let $\mathcal{V}_t^N$ be a noncutoff labelled Kac process and $\widetilde{\mathcal{V}}_0^{N,K} \in \mathbb{S}_N$. Let $\widetilde{\mathcal{V}}_t^{N,K}$ be the cutoff Kac process constructed in (4.95), and define

$$\overline{d}_p(t) := \frac{1}{N}\sum_{i=1}^N d_p\left(V_t^i, \widetilde{V}_t^{i,K}\right). \tag{4.98}$$

Suppose the initial data $\mathcal{V}_0^N, \widetilde{\mathcal{V}}_0^{N,K}$ are such that the associated empirical measures $\mu_0^N, \widetilde{\mu}_0^{N,K}$ satisfy moment bounds

$$\max\left(\Lambda_l(\mu_0^N), \Lambda_l(\widetilde{\mu}_0^{N,K})\right) \leq a_2; \tag{4.99}$$

$$\max\left(\Lambda_q(\mu_0^N), \Lambda_q(\widetilde{\mu}_0^{N,K})\right) \leq a_3; \tag{4.100}$$

with l as in Lemma 4.10 and $q = 2l$, and for some $a_2, a_3 > 1$. Fix $b > 1$, and let T_b^N be the stopping time (2.104) for the empirical measures μ_t^N of $\mathcal{V}_t^N$, with $p+\gamma$ in place of p, and similarly $T_b^{N,K}$ for $\mathcal{V}_t^{N,K}$. Then there exists $C = C(p, G, d)$ such that, for all $t \geq 0$,

$$\mathbb{E}\left[\overline{d}_p(t)\right] \leq e^{Cb(1+t)}\left(\overline{d}_p(0) + a_2 t K^{1-1/\nu}\right) + a_3 C \mathbb{P}(T_b^{N,K} \wedge T_b^N \leq t)^{1/2} \tag{4.101}$$

and, for all $t_{\text{fin}} \geq 0$,

$$\begin{aligned}\mathbb{E}\left[\sup_{t\leq t_{\text{fin}}} \overline{d}_p(t)\right] \leq e^{Cb(1+t_{\text{fin}})}&\left(\overline{d}_p(0) + a_2 t K^{1-1/\nu} + \frac{C a_3 t_{\text{fin}}^{1/2}}{N^{1/2}}\right) \\ &+ a_3 C(1+t_{\text{fin}})\mathbb{P}(T_b^{N,K} \wedge T_b^N \leq t)^{1/2}.\end{aligned} \tag{4.102}$$

This is the key result from which Theorems 4.4, 4.5 follow. Let us make the following remarks.

Remark 4.24. *i). We will show below that this essentially establishes Theorem 4.4. The form presented here, where we are free to choose b, is useful for dealing with the well-posedness issues deferred from Proposition 4.7.*

ii). In principle, one could perform a finer analysis for $\mathcal{Q}_K$ in Lemma 4.12, to replace the third term in (4.102) with an error in terms of $\overline{d}_p(0), K^{-\alpha}$, for some $\alpha > 0$. In this way, we would obtain an estimate for the uniform convergence on compacts in probability of $\widetilde{\mathcal{V}}^{N,K}$, as $K \to \infty$ with N fixed, and which is uniform in N. Since we are mostly interested in a limit where $N, K \to \infty$ simultaneously, we will not explore this.

Proof. Let $p \geq 0$ to be decided later, and consider the processes

$$M_t^i = d_p(V_t^i, V_t^{i,K}) - d_p(V_0^i, \widetilde{V}_0^{i,K}) - \frac{2}{N}\int_0^t \sum_{j=1}^N \mathcal{E}_{p,K}(V_s^i, V_s^{i,K}, V_s^j, V_s^{j,K})ds \tag{4.103}$$

for $1 \le i \le N$, and their average

$$\overline{M}_t = \frac{1}{N}\sum_{i=1}^N M^i_t = \overline{d}_p(t) - \int_0^t \overline{\mathcal{E}}_{p,K}(s)ds \tag{4.104}$$

where we define

$$\overline{\mathcal{E}}_{p,K}(t) := \frac{2}{N^2}\sum_{i,j=1}^N \mathcal{E}_{p,K}(V^i_t, \widetilde{V}^{i,K}_t, V^j_t, \widetilde{V}^{j,K}_t). \tag{4.105}$$

By classical results in the theory of Markov chains [49], each M^i_t is a martingale, and hence so is $\overline{M}$. By Lemma 4.10, provided K is large enough, depending on G, p, d, we have, for some $c = c(G,d), C = C(G,d,p)$,

$$\begin{aligned}\mathcal{E}_{p,K}(V^i_t, \widetilde{V}^{i,K}_t, V^j_t, \widetilde{V}^{j,K}_t) \le \left(c - \frac{\lambda_p}{2}\right) d_{p+\gamma}\left(V^i_t, \widetilde{V}^{i,K}_t\right) + c\, d_{p+\gamma}\left(V^j_t, \widetilde{V}^{j,K}_t\right) \\ + C\left(|V^j_t|^{p+\gamma} + |\widetilde{V}^{j,K}_t|^{p+\gamma}\right) d_p\left(V^i_t, \widetilde{V}^{i,K}_t\right) \\ + C\left(|V^i_t|^{p+\gamma} + |\widetilde{V}^{i,K}_t|^{p+\gamma}\right) d_p\left(V^j_t, \widetilde{V}^{j,K}_t\right) \\ + CK^{1-1/\nu}(1 + |V^i_t|^l + |V^j_t|^l + |\widetilde{V}^{j,K}_t|^l + |\widetilde{V}^{i,K}_t|^l).\end{aligned} \tag{4.106}$$

Let us now take the average over all i, j, which repeats in the context of the Kac process the same calculations in Lemma 4.21. The two terms on the first line can be absorbed together, as can the the terms on the second and third lines; for the same constants c, C as appearing in (4.81), we have

$$\begin{aligned}\overline{\mathcal{E}}_{p,K}(t) \le \frac{1}{N}\sum_{i=1}^N \left(c - \frac{\lambda_p}{2}\right) d_{p+\gamma}\left(V^i_t, \widetilde{V}^{i,K}_t\right) + C\left(\Lambda_{p+\gamma}(\mu^N_t) + \Lambda_{p+\gamma}(\widetilde{\mu}^{N,K}_t)\right)\overline{d}_p(t) \\ + CK^{1-1/\nu}\left(\Lambda_l(\mu^N_t) + \Lambda_l(\widetilde{\mu}^{N,K}_t)\right).\end{aligned} \tag{4.107}$$

In particular, for the same choice of $p_0 = p_0(B,d)$ as in Lemma 4.21, for all $p > p_0$, $\lambda_p \ge 2c$, and for such p, the first line of (4.107) is nonpositive, so the same argument as in Lemma 4.21 produces

$$\overline{\mathcal{E}}_{p,K}(t) \le C\left(\Lambda_{p+\gamma}(\mu^N_t) + \Lambda_{p+\gamma}(\widetilde{\mu}^{N,K}_t)\right)\overline{d}_p(t) + CK^{1-1/\nu}\left(\Lambda_l(\mu^N_t) + \Lambda_l(\widetilde{\mu}^{N,K}_t)\right) \tag{4.108}$$

whence

$$\begin{aligned}\overline{d}_p(t) \le \overline{d}_p(0) + C\int_0^t \left(\Lambda_{p+\gamma}(\mu^N_s) + \Lambda_{p+\gamma}(\widetilde{\mu}^{N,K}_s)\right)\overline{d}_p(s)ds \\ + CK^{1-1/\nu}\int_0^t \left(\Lambda_l(\mu^N_s) + \Lambda_l(\widetilde{\mu}^{N,K}_s)\right)ds + \overline{M}_t.\end{aligned} \tag{4.109}$$

Let us now write $T := T^N_b \wedge T^{N,K}_b$ for the stopping times $T^N_b, T^{N,K}_b$ defined in the statement, and consider the moment prefactor in (4.108, 4.109). We recall from Proposition 2.10iii) that, almost surely, for all $t \ge 0$,

$$\Lambda_{p+\gamma}(\mu^N_t) \le 2^{\frac{p+\gamma}{2}+1}\Lambda_{p+\gamma}(\mu^N_{t-}) \tag{4.110}$$

and similarly for $\widetilde{\mu}_t^{N,K}$. The moment factor is therefore at most $2b$ for all $s \leq T$, and so we obtain, for all $t \geq 0$,

$$\int_0^{t\wedge T} \left(\Lambda_{p+\gamma}(\mu_s^N) + \Lambda_{p+\gamma}(\widetilde{\mu}_s^{N,K})\right) \overline{d}_p(s)ds \leq 2b \int_0^t \overline{d}_p(s\wedge T)ds. \tag{4.111}$$

Stopping (4.109) at T, we therefore obtain, for all $t \geq 0$,

$$\begin{aligned}\overline{d}_p(t\wedge T) \leq \overline{d}_p(0) + Cb\int_0^t \overline{d}_p(s\wedge T)ds \\ + CK^{1-1/\nu}\int_0^t \left(\Lambda_l(\mu_s^N) + \Lambda_l(\widetilde{\mu}_s^{N,K})\right) ds + \overline{M}_{t\wedge T}.\end{aligned} \tag{4.112}$$

For the first item, we fix $t \geq 0$, and take expectations of (4.112). By optional stopping, $\mathbb{E}[\overline{M}_{t\wedge T}] = 0$, and we use the moment estimates in Propositions 2.10 to control the first term on the second line:

$$\mathbb{E}\left[\int_0^t (\Lambda_l(\mu_s^N) + \Lambda_l(\widetilde{\mu}_s^{N,K}))ds\right] \leq Cta_2. \tag{4.113}$$

We therefore use Grönwall's Lemma to obtain

$$\mathbb{E}\left[\overline{d}_p(t\wedge T)\right] \leq e^{Cbt}\left(\overline{d}_p(0) + Cta_2K^{1-1/\nu}\right). \tag{4.114}$$

Next, we observe that

$$\overline{d}_p(t) \leq \overline{d}_p(t\wedge T) + \overline{d}_p(t)1\!\!1_{T\leq t}. \tag{4.115}$$

We now estimate the second term. From the bound $d_p(v,w) \leq c(1+|v|^{p+2}+|w|^{p+2})$ we see that

$$\overline{d}_p(t) \leq c\left(\Lambda_{p+2}(\mu_t^N) + \Lambda_{p+2}(\widetilde{\mu}_t^{N,K})\right) \tag{4.116}$$

We use Hölder's inequality with indexes $\frac{q}{p+2}$ and $\frac{q}{p+2+\gamma} \leq 2$, to obtain

$$\begin{aligned}\mathbb{E}\left[\overline{d}_p(t)1\!\!1_{T\leq t}\right] &\leq c\,\mathbb{P}(T\leq t)^{(p+2+\gamma)/q}\,\mathbb{E}\left[\Lambda_q(\mu_t^N) + \Lambda_q(\widetilde{\mu}_t^{N,K})\right]^{(p+2)/q} \\ &\leq C\,\mathbb{P}(T\leq t)^{1/2}a_3\end{aligned} \tag{4.117}$$

thanks to the moment bounds in Proposition 2.10 and the choice of initial data. Combining with the previous term (4.114) now proves the first claim.

For the second item, we return to the martingale $\overline{M}_t$ constructed above. From [49, Lemma 8.7], the process

$$L_t = \overline{M}_t^2 - \frac{2}{N^3}\sum_{\{ij\}}\int_0^t \mathcal{Q}_K(V_s^i, \widetilde{V}_s^{i,K}, V_s^j, \widetilde{V}_s^{j,K})ds \tag{4.118}$$

is also a martingale, where the sum now runs over unordered pairs $\{ij\}$ of indexes. Thanks to the bound computed in Lemma 4.12, we find

$$\begin{aligned}\mathbb{E}\left[\overline{M}_{t_{\text{fin}}}^2\right] &\leq \frac{C}{N^3}\mathbb{E}\left[\sum_{\{ij\}}\int_0^{t_{\text{fin}}}(1+|V_s^i|^q + |V_s^j|^q + |\widetilde{V}_s^{i,K}|^q + |\widetilde{V}_s^{j,K}|^q)ds\right] \\ &\leq \frac{C}{N}\int_0^{t_{\text{fin}}}\mathbb{E}(\Lambda_q(\mu_s^N) + \Lambda_q(\widetilde{\mu}_s^{N,K}))ds.\end{aligned} \tag{4.119}$$

Using the moment propagation estimate in Proposition 2.10 and Doob's L^2 inequality, we conclude that

$$\mathbb{E}\left[\sup_{s\le t_{\mathrm{fin}}} |\overline{M}_t|\right] \le \frac{Ca_3 t_{\mathrm{fin}}^{1/2}}{N^{1/2}}. \tag{4.120}$$

With this estimate, we return to the argument above. Applying Grönwall to (4.112), we obtain a pathwise estimate

$$\sup_{t\le t_{\mathrm{fin}}} \overline{d}_p(t\wedge T) \le e^{Cbt}\left(\overline{d}_p(0) + K^{1-1/\nu}\int_0^{t_{\mathrm{fin}}} \left(\Lambda_l(\mu^N_s) + \Lambda_l(\widetilde{\mu}^{N,K}_s)\right) ds + \sup_{t\le t_{\mathrm{fin}}}|\overline{M}_t|\right). \tag{4.121}$$

Taking expectations, we conclude that

$$\mathbb{E}\left[\sup_{t\le t_{\mathrm{fin}}} \overline{d}_p(t\wedge T)\right] \le e^{Cbt}\left(\overline{d}_p(0) + K^{1-1/\nu}Ct_{\mathrm{fin}}a_2 + \frac{Ca_3 t_{\mathrm{fin}}^{1/2}}{N^{1/2}}\right). \tag{4.122}$$

Following the same argument as in (4.117) we also bound

$$\begin{aligned}\mathbb{E}\left[\left(\sup_{t\le t_{\mathrm{fin}}} \overline{d}_p(t)\right)\mathbb{1}[T\le t_{\mathrm{fin}}]\right] &\le C\mathbb{P}\left(T\le t_{\mathrm{fin}}\right)^{1/2}\mathbb{E}\left[\sup_{t\le t_{\mathrm{fin}}}\left(\Lambda_q(\mu^N_t) + \Lambda_q(\widetilde{\mu}^{N,K}_t)\right)\right] \\ &\le C\mathbb{P}(T\le t_{\mathrm{fin}})^{1/2} t_{\mathrm{fin}} a_3.\end{aligned} \tag{4.123}$$

Combining (4.122, 4.123), we obtain the desired result. □

We now prove Proposition 4.6, as the uniqueness-in-law uses the coupling we have built. We remark that we have not used existence and uniqueness at any point, so there is no circularity. We do not seek any estimates uniformly in N, and we can replace moment estimates with the trivial bound $|V^i_t| \le \sqrt{N}$. For ease of presentation, we will use the estimates we have already developed in this chapter, although those from the literature [92] would work equally well.

Proof of Proposition 4.6. Let us fix $\mathcal{V}^N_0$; for each K, let $\mathcal{V}^{N,K}_t$ be a solution to (cLK), starting at $\mathcal{V}^N_0$, with cutoff parameter K. Since the rates are finite, such processes can be constructed elementarily, and have uniqueness in law. We check tightness via Aldous' criterion; thanks to the energy constraint, each $\mathcal{V}^{N,K}_t$ takes values in $[-N^{1/2}, N^{1/2}]^{Nd}$, and for equicontinuity, we use the same argument as in Lemma 4.13. Similar arguments to those that show that $\mathcal{L}f$ is continuous show that, for all Lipschitz $F:\mathbb{S}_N\to\mathbb{R}$, $\mathcal{G}^L F$ is continuous, and is the uniform limit of the cutoff generators $\mathcal{G}^L_K F$, recalling that $\mathcal{G}^L$ is the generator of the labelled dynamics given by (4.9) and writing $\mathcal{G}^L_K$ for the cutoff analogue. We can therefore take a limit to see that any subsequential limit point of $\mathcal{V}^{N,K}_t$ as $K\to\infty$ is a solution to the martingale problem for $\mathcal{G}^L$, and hence is a weak solution to (LK) by the representation theorem again [70].

For uniqueness in law, let $\mathcal{V}^N_t$ be any solution to (LK) starting at $\mathcal{V}^N_0$. We now apply

Lemma 4.23; fix $p > p_0(G, d), K > K_0(G, p, d)$ as in the statement, and $0 \le t_1 < ... < t_m$, we take $b = N^{(p+\gamma)/2}$, so that $T_b^N = T_b^{N,K} = \infty$. The cited lemma now shows that $(\mathcal{V}^N_{t_i})_{i\le m}$ is the limit in probability, of $(\mathcal{V}^{N,K}_{t_i})_{i\le m}$, for cutoff labelled Kac processes $\mathcal{V}^{N,K}_t$ starting at $\mathcal{V}^N_0$, as $K \to \infty$. Since the law of each $\mathcal{V}^{N,K}_t$ is uniquely determined, the same is true of the m-tuple $(\mathcal{V}^N_{t_i})_{i\le m}$. Since t_i were arbitrary and finite marginal distributions characterise the laws of càdlàg processes, we conclude that the law of $\mathcal{V}^N_t$ is unique, as claimed. □

4.8.2 Proof of Theorem 4.4

We will now deduce Theorem 4.4. Roughly speaking, it is sufficient to take $b = Ca_1$ in the previous lemma, using the concentration of moments in Proposition 2.10iv).

Proof of Theorem 4.4. Let $p_0(G, d), K_0(B, p, d)$ be as above, and fix $p > p_0, K > K_0$. We deal first with the case $K < \infty$, and will consider $K = \infty$ at the end.

Let us fix μ^N_0, $\widetilde{\mu}^{N,K}_0, a_1, a_2, a_3$ as in the statement of the Theorem, and choose $\mathcal{V}^N_0 \in \theta_N^{-1}(\mu^N_0), \widetilde{\mathcal{V}}^{N,K'}_0 \in \theta_N^{-1}(\widetilde{\mu}^{N,K}_0)$ corresponding to $\mu^{N,K}_0, \widetilde{\mu}^{N,K'}_0$ which achieve the optimal coupling

$$w_p(\mu^{N,K}_0, \widetilde{\mu}^{N,K'}_0) = \frac{1}{N}\sum_{i=1}^N d_p(V^i_0, \widetilde{V}^{i,K'}_0). \tag{4.124}$$

Now, let $\mathcal{V}^N_t$ be a noncutoff labelled Kac process starting at $\mathcal{V}^N_0$, and write $\mu^N_t = \theta_N(\mathcal{V}^N_t)$ for the process of empirical measures. Let $\mathcal{V}^{N,K}_t$ be the Kac processes constructed by Lemma 4.23 for the initial data $\mathcal{V}^N_0$ respectively with cutoff parameter K, and let $\mu^{N,K}_t$ be the associated empirical measures. We observe that

$$w_p\left(\mu^N_t, \widetilde{\mu}^{N,K}_t\right) \le \frac{1}{N}\sum_{i=1}^N d_p\left(V^i_t, \widetilde{V}^{i,K}_t\right) = \overline{d}_p(t) \tag{4.125}$$

which we control by the previous lemma to obtain, for some C and all $t \ge 0, b > 1$

$$\begin{aligned}\mathbb{E}\left[w_p\left(\mu^N_t, \widetilde{\mu}^{N,K}_0\right)\right] \le e^{Cb(1+t)}\Big(w_p\left(\mu^N_0, \widetilde{\mu}^{N,K}_0\right) + a_2K^{1-1/\nu}\Big) \\ + Ca_3\mathbb{P}\left(T^N_b \wedge T^{N,K}_b \le t\right)^{1/2}.\end{aligned} \tag{4.126}$$

where $T^N_b, T^{N,K}_b$ are as above. Now, taking $b = Ca_1$ for some large $C = C(p)$, we use Proposition 2.10iv) to control the final term and obtain, for some C,

$$\begin{aligned}\mathbb{P}\left(T^N_{Ca_1} \wedge T^{N,K}_{Ca_1} \le t\right) &\le \mathbb{P}\left(T^N_{Ca_1} \le t\right) + \mathbb{P}\left(T^{N,K}_{Ca_1} \le t\right) \\ &\le Cta_3N^{-1}\end{aligned} \tag{4.127}$$

as desired. We obtain (4.7) from (4.102) for the same processes $\mu^N_t, \widetilde{\mu}^{N,K}_t$ in exactly the same way.

We finally deal with the case $K=\infty$. In this case, we build the couplings $\widetilde{\mathcal{V}}_t^{N,K}$ as above for all $K\geq K_0$; arguing as in Proposition 4.6 above, these processes are tight, and hence so are $(\mathcal{V}_t^N,\widetilde{\mathcal{V}}_t^{N,K})_{t\geq 0}$. If we pass to a subsequential limit $(\mathcal{V}_t^N,\widetilde{\mathcal{V}}_t^N)_{t\geq 0}$, the same proof as in Proposition 4.6 above shows that $\widetilde{\mathcal{V}}_t^N$ is a labelled Kac process, and the desired estimates hold by taking limits of the cutoff case. □

4.8.3 Proof of Theorem 4.5

We now prove the Theorem 4.5 concerning the convergence of the full, non-cutoff Kac process to the solution to the Boltzmann equation in the many-particle limit $N\to\infty$. We will interpolate between the coupling given by the coupling in Theorem 4.4 and the K-dependent convergence of the cutoff Kac process in Lemma 3.14.

Proof of Theorem 4.5. The uniqueness in law follows from Proposition 4.7, which is discussed in Appendix 4.A.

For the convergence estimate, let $\mu_t^N, t\geq 0$ be any unlabelled Kac process, and consider the case $\mu_0=\mu_0^N$. Fix t_{fin} and let $K\in[K_0,\infty)$ to be chosen later; for this K, let $\widetilde{\mu}_t^N,\widetilde{\mu}_t^{N,K}$ be the coupling of noncutoff and cutoff Kac processes, both starting at μ_0^N given Theorem 4.4. By uniqueness in law, it is sufficient to prove the estimate with $\widetilde{\mu}_t^N$ in place of μ_t^N. For some constants $C=C(p,q),\alpha=\alpha(p,q)$, we have the following estimates. By Theorem 4.4,

$$\mathbb{E}\left[\sup_{t\leq t_{\text{fin}}} w_p\left(\widetilde{\mu}_t^N,\widetilde{\mu}_t^{N,K}\right)\right]\leq e^{Ca(1+t_{\text{fin}})}(K^{1-1/\nu}+N^{-1/2}); \tag{4.128}$$

while we recall from lemma 3.14

$$\mathbb{E}\left[\sup_{t\leq t_{\text{fin}}} w_p\left(\widetilde{\mu}_t^{N,K},\phi_t^K(\mu_0^N)\right)\right]\leq \exp\left(CaK(1+t_{\text{fin}})\right)\ N^{-\alpha} \tag{4.129}$$

and by Corollary 4.2,

$$\sup_{t\leq t_{\text{fin}}}\ w_p\left(\phi_t^K(\mu_0^N),\phi_t(\mu_0^N)\right)\leq e^{Ca(1+t_{\text{fin}})}K^{1-1/\nu}. \tag{4.130}$$

Combining, and keeping the worst terms, we have the estimate

$$\mathbb{E}\left[\sup_{t\leq t_{\text{fin}}}\ w_p\left(\widetilde{\mu}_t^N,\phi_t(\mu_0^N)\right)\right]\leq e^{Ca(1+t_{\text{fin}})}K^{1-1/\nu}+e^{CaK(1+t_{\text{fin}})}\ N^{-\alpha}. \tag{4.131}$$

We now choose

$$K=\max\left(K_0,\frac{1}{2Ca(1+t_{\text{fin}})}\log(N^\alpha)\right) \tag{4.132}$$

to conclude that

$$\mathbb{E}\left[\sup_{t\leq t_{\text{fin}}}\ w_p\left(\widetilde{\mu}_t^N,\phi_t(\mu_0^N)\right)\right]\leq e^{Ca(1+t_{\text{fin}})}\left(\log N\right)^{1-1/\nu}. \tag{4.133}$$

Finally, by Corollary 4.2 again, we have

$$\mathbb{E}\left[\sup_{t\le t_{\text{fin}}} w_p\left(\phi_t(\mu_0^N),\phi_t(\mu_0)\right)\right] \le e^{Ca(1+t_{\text{fin}})}\,\mathbb{E}\left[w_p(\mu_0^N,\mu_0)\right] \tag{4.134}$$

and combining gives the claimed bound. □

4.9 Alternative Proof of Corollary 4.2

We will now give the alternative proof of Corollary 4.2, based on the Tanaka coupling of Kac processes presented in Lemma 4.23. The proof is broken down into a series of lemmas; in order to give an overview of the strategy, we will state all the intermediate steps before turning to the proofs. Our first step is to use Theorem 4.4 to compare two *cutoff* Kac processes, with different values of the cutoff and different initial data.

Lemma 4.25. *[Coupling of Cutoff Kac Processes] Let p, q, l, K_0, C be as in Lemma 4.23, and let $K' \ge K > K_0(G,p,d)$. Let $\mu_0^{N,K}, \widetilde{\mu}_0^{N,K'} \in \mathcal{S}_N$, with moments*

$$\Lambda_{p+\gamma}\left(\mu_0^N, \widetilde{\mu}_0^{N,K}\right) \le a_1; \qquad \Lambda_l\left(\mu_0^N, \widetilde{\mu}_0^{N,K}\right) \le a_2; \qquad \Lambda_q\left(\mu_0^N, \widetilde{\mu}_0^{N,K}\right) \le a_3. \tag{4.135}$$

Then there exists a coupling of cutoff Kac processes $\mu_t^{N,K}, \widetilde{\mu}_t^{N,K'}$ with cutoff parameters K, K' respectively, such that

$$\mathbb{E}\left[w_p(\mu_t^{N,K}, \widetilde{\mu}_t^{N,K'})\right] \le e^{Ca_1(1+t)}(w_p(\mu_0^{N,K}, \widetilde{\mu}_0^{N,K}) + a_2 t K^{1-1/\nu}) + a_3^2 C N^{-1/2} t. \tag{4.136}$$

We next transfer this coupling to solutions to the cutoff Boltzmann equation, potentially with different cutoff parameters and different initial data.

Lemma 4.26. *Let $p > p_0(G,d)$ and $l = p+2+\gamma$. Then there exist a constant $C = C(G,p,d)$ such that, whenever $K' \ge K > K_0(G,p,d), a_1, a_2 \ge 1$ and $\mu_0, \nu_0 \in \mathcal{S}$ satisfy moment bounds*

$$\Lambda_{p+\gamma}(\mu_0,\nu_0) \le a_1; \qquad \Lambda_l(\mu_0,\nu_0) \le a_2 \tag{4.137}$$

then the solution maps ϕ_t^K to the cutoff Boltzmann equation (BE$_K$) satisfy, for all $t \ge 0$,

$$w_p\left(\phi_t^K(\mu_0), \phi_t^{K'}(\nu_0)\right) \le e^{Ca_1(1+t)}\left(w_p(\mu_0,\nu_0) + a_2 t K^{1-1/\nu}\right). \tag{4.138}$$

As a next step, we show that the solutions $\phi_t^K(\mu_0)$ to the cutoff Boltzmann equations converge, as $K \to \infty$, to a solution of the noncutoff equation (BE).

Lemma 4.27. *Let p, l be as above, and let $\mu_0 \in \mathcal{S}$ satisfy moment assumptions*

$$\Lambda_{p+\gamma}(\mu_0) \le a_1, \qquad \Lambda_l(\mu_0) \le a_2 \tag{4.139}$$

for some $a_1, a_2 \ge 1$. Then, for some $(\phi_t(\mu))_{t\ge 0} \subset \mathcal{S}$ and some $C = C(G,p,d)$,

$$w_p(\phi_t^K(\mu), \phi_t(\mu)) \le e^{Ca_1(1+t)} t a_2 K^{1-1/\nu} \tag{4.140}$$

for all $K > K_0(G,p,d)$. Moreover, if $\nu_0 \in \mathcal{S}$ is another measure with the same moment estimates, we have the continuity

$$w_p(\phi_t(\mu_0), \phi_t(\nu_0)) \le e^{Ca_1(1+t)} w_p(\mu_0,\nu_0). \tag{4.141}$$

Finally, $(\phi_t(\mu_0) : t \ge 0)$ is a solution to the noncutoff Boltzmann equation (BE), and satisfies the moment estimates in Proposition 2.6.

We next extend the maps ϕ_t defined above to all of $\mathcal{S}^{p+2}$, and obtain the claimed continuity estimate in this context

Lemma 4.28. *Let p, l be as above. The solution maps $\phi_t : \mathcal{S}^l \to \mathcal{S}$ defined above can be extended to $\mathcal{S}^{p+2}$, such that, for all $\mu_0 \in \mathcal{S}^{p+2}$, $(\phi_t(\mu_0) : t \geq 0)$ is a solution to the Boltzmann Equation (BE), and so that (4.141) holds whenever $\mu_0, \nu_0 \in \mathcal{S}^{p+2}$ satisfy a moment estimate $\Lambda_{p+\gamma}(\mu_0, \nu_0) \leq a$, for some $a \geq 1$.*

To conclude Corollary 4.2, we must show that the solutions obtained in this way are the unique solutions to (BE) as soon as $\mu_0 \in \mathcal{S}^{p+2}$. Let us now show how these results imply the claimed result.

Proof of Theorem 4.1. In light of Lemma 4.28 above, it remains only to prove that the solutions constructed above are unique. Let us fix $\mu_0 \in \mathcal{S}^p$ and such that $\Lambda_{p+2}(\mu_0) \leq a$ for some $a \geq 1$. Let $(\mu_t)_{t\geq 0} \subset \mathcal{S}$ be any solution to (BE) starting at μ_0; we will now show that $\mu_t = \phi_t(\mu_0)$ for all $t \geq 0$.

Fix $s > 0, t \geq 0$. Thanks to the appearance of exponential moments in Proposition 2.13, there exists $\epsilon = \epsilon_s > 0$ such that $\langle e^{\epsilon|v|^\gamma}, \mu_s\rangle < \infty$, and by Proposition 4.18, there exists at most one energy-conserving solution starting at μ_s. Since both $(\phi_u(\mu_s))_{u\geq 0}$ and $(\mu_{u+s})_{u\geq 0}$ are such solutions, we conclude that $\phi_t(\mu_s) = \mu_{t+s}$ for all such t, s.

Let us now take the limit $s \downarrow 0$. As in Lemma 4.22, $w_p(\mu_s, \mu_0) \to 0$ as $s \downarrow 0$. Lemma 4.28 now shows that, up to a new choice of C,

$$w_p(\phi_t(\mu_s), \phi_t(\mu_0)) \leq e^{Ca(1+t)} w_p(\mu_s, \mu_0) \to 0. \tag{4.142}$$

Using the same argument again, $w_p(\mu_{t+s}, \mu_t) \to 0$, and we conclude that

$$w_p\left(\mu_t, \phi_t(\mu_0)\right) \leq \limsup_{s\downarrow 0} \left[C w_p\left(\phi_t(\mu_s), \phi_t(\mu_0)\right) + C w_p\left(\mu_{t+s}, \mu_t\right)\right] = 0 \tag{4.143}$$

and so we have the desired uniqueness. □

4.9.1 Proof of Lemmas

Sketch Proof of Lemma 4.25. The proof is very similar to the proof of Theorem 4.4, and we will sketch the main points. Let us construct $\mathcal{V}_0^N \in \theta_N^{-1}(\mu_0^N)$ and $\widetilde{\mathcal{V}}_0^{N,K'} \in \theta_N^{-1}(\widetilde{\mu}_0^{N,K'})$ as in the proof of Theorem 4.4, and following the previous proof, construct a *non*cutoff labelled process $\mathcal{V}_t^N$ starting at $\mathcal{V}_0^N$ and a K'-cutoff $\widetilde{\mathcal{V}}_t^{N,K'}$ starting at $\widetilde{\mathcal{V}}_0^{N,K'}$, with a control over $N^{-1}\sum_i d_p(V_t^i, \widetilde{V}_t^{i,K})$. We take $\mu_t^N, \widetilde{\mu}_t^{N,K'}$ to be the associated empirical measures, which are (unlabelled) Kac processes, and which inherit the control from $w_p(\mu_t^N, \widetilde{\mu}_t^{N,K'}) \leq N^{-1}\sum_i d_p(V_t^i, \widetilde{V}_t^{i,K})$.

We now repeat this argument to construct a K-cutoff process $\mathcal{V}_t^{N,K}$ starting at the same point $\mathcal{V}_0^N = \mathcal{V}_0^{N,K}$, and let $\mu_t^{N,K}$ be the associated empirical measures. The same argument as the proof of Theorem 4.4 establishes controls on

$$\mathbb{E}\left[w_p\left(\mu_t^N, \widetilde{\mu}_t^{N,K'}\right)\right]; \qquad \mathbb{E}\left[w_p\left(\mu_t^N, \mu_t^{N,K}\right)\right]. \tag{4.144}$$

Recalling the relaxed triangle inequality (2.17), we combine these to find the desired estimate. □

Proof of Lemma 4.26. Let us consider the case first where the initial data μ_0, ν_0 have a finite q^{th} moment $\Lambda_q(\mu_0, \nu_0) \leq a_3$ for some $a_3 \geq 1$. Applying Proposition 2.4, we can choose N-particle empirical measures $\mu_0^N \in \mathcal{S}_N$ such that $w_p(\mu_0^N, \mu_0) \to 0$ and such that the $l^{\text{th}}, q^{\text{th}}$ moments converge: $\Lambda_l(\mu_0^N) \to \Lambda_l(\mu_0), \Lambda_q(\mu_0^N) \to \Lambda_q(\mu_0)$; construct ν_0^N similarly for ν_0. Using the relaxed triangle inequality (2.17), it follows that, for some $C = C(p)$,

$$\limsup_{N\to\infty} w_p(\mu_0^N, \nu_0^N) \leq C\, w_p(\mu_0, \nu_0). \tag{4.145}$$

Let us now take $\mu_t^{N,K}$, $\nu_t^{N,K'}$ be the cutoff Kac processes constructed in Corollary 4.25 started at these initial data; fix $t \geq 0$, and consider

$$w_p\left(\phi_t^K(\mu_0), \phi_t^{K'}(\nu_0)\right) \leq C\,\mathbb{E}\Big[w_p\left(\phi_t^K(\mu_0), \mu_t^{N,K}\right) + w_p\left(\mu_t^{N,K}, \nu_t^{N,K'}\right) + w_p\left(\nu_t^{N,K'}, \phi_t^{K'}(\nu_0)\right)\Big]. \tag{4.146}$$

Using Corollary 4.25 to bound the middle term, we have

$$\begin{aligned} w_p\left(\phi_t^K(\mu_0), \phi_t^{K'}(\nu_0)\right) \leq C\,\mathbb{E}&\left[w_p\left(\phi_t^K(\mu_0), \mu_t^{N,K}\right) + w_p\left(\nu_t^{N,K'}, \phi_t^{K'}(\nu_0)\right)\right] \\ &+ e^{Ca_1(1+t)}\left(w_p\left(\mu_0^N, \nu_0^N\right) + a_2 K^{1-1/\nu}\right) \\ &+ a_3^2 C N^{-1/2}. \end{aligned} \tag{4.147}$$

Let us now take $N \to \infty$. Thanks to Lemma 3.14, both terms on the first line converge to 0, as does the final term. Using (4.145), we conclude that

$$w_p\left(\phi_t^K(\mu_0), \phi_t^{K'}(\nu_0)\right) \leq e^{Ca_1(1+t)}\left(w_p\left(\mu_0^N, \nu_0^N\right) + a_2 K^{1-1/\nu}\right) \tag{4.148}$$

as desired.

Let us now show how this extends to initial data μ_0, ν_0 with only $l = p + \gamma + 2$ moments as in the statement. In this case, we use Proposition 2.4 again, with l in place of q, to construct $\mu_0^N \in \mathcal{S}_N$ such that

$$w_p(\mu_0^N, \mu_0) \to 0, \qquad \Lambda_l(\mu_0^N) \to \Lambda_l(\mu_0) \tag{4.149}$$

and similarly ν_0^N for ν_0. Since μ_0^N, ν_0^N are compactly supported, the previous estimate applies so that

$$w_p\left(\phi_t^K(\mu_0^N), \phi_t^{K'}(\nu_0^N)\right) \leq e^{Ca_1(1+t)}\left(w_p\left(\mu_0^N, \nu_0^N\right) + a_2 K^{1-1/\nu}\right). \tag{4.150}$$

Using Lemma 3.13,

$$w_p\left(\phi_t^K(\mu_0^N), \phi_t^K(\mu_0)\right) \to 0; \qquad w_p\left(\phi_t^{K'}(\nu_0^N), \phi_t^{K'}(\nu_0)\right) \to 0. \tag{4.151}$$

The same argument as above therefore allows us to take $N \to \infty$ in (4.150), noting that no moments higher than l^{th} appear, to conclude that

$$w_p\left(\phi_t^K(\mu_0), \phi_t^{K'}(\nu_0)\right) \le Ce^{Ca_1(1+t)}\left(w_p\left(\mu_0, \nu_0\right) + a_2 K^{1-1/\nu}\right) \tag{4.152}$$

and we absorb the prefactor C into the exponent. □

Proof of Lemma 4.27. Let us fix $\mu_0 \in \mathcal{S}^l$ and consider the space

$$\mathcal{C} = C([0,\infty), (\{\mu \in \mathcal{P}_2 : \Lambda_2(\mu) \le 1\}, \mathcal{W}_1))$$

equipped with a metric inducing uniform convergence on compact time intervals. Since $(\{\mu \in \mathcal{P}_2 : \Lambda_2(\mu) \le 1\}, \mathcal{W}_1)$ is complete, so is $\mathcal{C}$. Recalling that $\mathcal{W}_1 \le w_p$, the previous observation shows that $(\phi^K(\mu_0), t \ge 0)_{K\ge 1}$ are Cauchy in $\mathcal{C}$, and hence converge to some process $(\phi_t(\mu_0), t \ge 0)$. Moreover, thanks to Proposition 2.6, the l^{th} moments of $\phi_t^K(\mu_0)$ are bounded, uniformly in K, and since $l > 2$, we conclude that $1 = \Lambda_2(\phi_t^K(\mu_0)) \to \Lambda_2(\phi_t(\mu_0))$, and similarly for $0 = \langle v, \phi_t^K(\mu_0)\rangle \to \langle v, \phi_t(\mu_0)\rangle$, to get that $(\phi_t(\mu_0))_{t\ge 0} \subset \mathcal{S}$.

Next, let us show that $\phi_t^K(\mu_0) \to \phi_t(\mu_0)$ in w_p. For $t = 0$, $\phi_0^K(\mu_0) = \mu_0$, and so there is nothing to prove. If $t > 0$ then, by point iii) of Proposition 2.6, there exists $\lambda_{p+3} = \lambda_{p+3}(t) < \infty$ such that, for all $K \ge 1$,

$$\Lambda_{p+3}\left(\phi_t^K(\mu_0)\right) \le \lambda_{p+3}(t). \tag{4.153}$$

By lower semicontinuity of $\mu \mapsto \Lambda_{p+3}(\mu)$ in $\mathcal{W}_1$, the same is true for the limit $\phi_t(\mu_0)$, and using the estimates in Section 2.1,

$$w_p\left(\phi_t^K(\mu_0), \phi_t(\mu_0)\right) \le \Lambda_{p+3}\left(\phi_t^K(\mu_0), \phi_t(\mu_0)\right)\ \mathcal{W}_1\left(\phi_t^K(\mu_0), \phi_t(\mu_0)\right)^\alpha \tag{4.154}$$

for some $\alpha > 0$. By construction, the second term on the right-hand side converges to 0, and the first term is bounded, so the left-hand side converges to 0 as desired. We now conclude the bound (4.140): if $K > K_0(G, p, d)$, then for all $K' \ge K$,

$$\begin{aligned} w_p\left(\phi_t^K(\mu_0), \phi_t(\mu_0)\right) &\le C\left(w_p\left(\phi_t^K(\mu_0), \phi_t^{K'}(\mu_0)\right) + w_p\left(\phi_t^{K'}(\mu_0), \phi_t(\mu_0)\right)\right) \\ &\le C\left(e^{Ca_1(1+t)} a_2 t K^{1-1/\nu} + w_p\left(\phi_t^{K'}(\mu_0), \phi_t(\mu_0)\right)\right). \end{aligned} \tag{4.155}$$

Taking $K' \to \infty$, the second term on the final line converges to 0, and the desired bound follows, absorbing the prefactor $C = C(p)$ into the exponent. The bound (4.141) is similar: if μ_0, ν_0 in $\mathcal{S}^l$ satisfy

$$\Lambda_{p+\gamma}(\mu_0, \nu_0) \le a_1; \qquad \Lambda_l(\mu_0, \nu_0) \le a_2 \tag{4.156}$$

then we bound, for any K,

$$\begin{aligned} w_p\left(\phi_t(\mu_0),\phi_t(\nu_0)\right) &\le C\big(w_p\left(\phi_t^K(\mu_0),\phi_t(\mu_0)\right)+w_p\left(\phi_t^K(\mu_0),\phi_t^K(\nu_0)\right) \\ &\qquad + w_p\left(\phi_t^K(\nu_0),\phi_t(\nu_0)\right)\big) \\ &\le Ce^{Ca_1(1+t)}\left(3a_2tK^{1-1/\nu}+w_p(\mu_0,\nu_0)\right) \end{aligned} \tag{4.157}$$

where, in the second line, we have used Lemma 4.26 to compare $\phi_t^K(\mu_0),\phi_t^K(\nu_0)$ and used the previous part to estimate the other two terms. Taking $K\to\infty$, we conclude the desired bound, again up to a new choice of C.

It remains to show that $\phi_t(\mu_0), t\ge 0$ solves the full, noncutoff Boltzmann equation (BE). To shorten notation, write $\mu_t^K := \phi_t^K(\mu_0), \mu_t = \phi_t(\mu_0)$. We fix a bounded, Lipschitz function $f:\mathbb{R}^d\to\mathbb{R}$ and recall, from Lemma 4.13, the growth bound

$$|\mathcal{L}f(v,v_\star)|\le C(f)|v-v_\star|\int_{\mathbb{S}^{d-1}}\sin\theta B(v-v_\star,\sigma)d\sigma\le C(f)|v-v_\star|^{1+\gamma}$$

for some constant $C=C(f)$, depending only on the Lipschitz constant of f, and similarly for B_K; recall also that $\mathcal{L}f$ is continuous, and that $|\mathcal{L}_Kf-\mathcal{L}f|\le\varepsilon_K|v-v_*|^{1+\gamma}$, for some $\varepsilon_K\downarrow 0$. We now claim that, for all $t\ge 0$,

$$\langle f, Q_K(\mu_t^K)\rangle\to\langle f,Q(\mu_0)\rangle. \tag{4.158}$$

Fix $t\ge 0$. For all $R\ge 0$, let $\psi_R:\mathbb{R}^d\times\mathbb{R}^d\to[0,1]$ be a smooth, compactly supported cutoff function, such that $\psi_R(v,v_\star)=1$ on the ball $\{|v|^2+|v_\star|^2\le R\}$. We estimate, uniformly in K,

$$\begin{aligned} &\int_{\mathbb{R}^d\times\mathbb{R}^d}|\mathcal{L}f(v,v_\star)(1-\psi_R(v,v_\star))|\mu_t^K(dv)\mu_t^K(dv_\star) \\ &\qquad\le C(f)\int_{\mathbb{R}^d\times\mathbb{R}^d}(1+|v|^2+|v_\star|^2)\mathbb{1}_{|v|^2+|v_\star|^2\ge R}\mu_t^K(dv)\mu_t^K(dv_\star) \\ &\qquad\le C(f)R^{-p}a_2\,\Lambda_l\left(\mu_0\right) \end{aligned} \tag{4.159}$$

where, in the final line, we used the moment hypothesis on μ_0, with $l>p+2$, and the moment propagation result in Proposition 2.6; the same argument holds for the limit μ_t. It is elementary to show that the Wasserstein convergence $\mathcal{W}_1(\mu_t^K,\mu_t)\to 0$ implies that, for all compactly supported, continuous $g:\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}$, we have

$$\int_{\mathbb{R}^d\times\mathbb{R}^d}g(v,v_\star)\left(\mu_t^K(dv)\mu_t^K(dv_\star)-\mu_t(dv)\mu_t(dv_\star)\right)\to 0 \tag{4.160}$$

and, in particular, this holds with $g=(\mathcal{L}f)(v,v_\star)\psi_R$. We now write

$$\begin{aligned} &\left|\left\langle f, Q(\mu_t)-Q(\mu_t^K)\right\rangle\right| \\ &\qquad\le\int_{\mathbb{R}^d\times\mathbb{R}^d}|(\mathcal{L}f)|(1-\psi_R)(v,v_\star)(\mu_t^K(dv)\mu_t^K(dv_\star)+\mu_t(dv)\mu_t(dv_\star)) \\ &\qquad+\left|\int_{\mathbb{R}^d\times\mathbb{R}^d}(\mathcal{L}f)\psi_R(v,v_\star)(\mu_t^K(dv)\mu_t^K(dv_\star)-\mu_t(dv)\mu_t(dv_\star))\right|. \end{aligned} \tag{4.161}$$

The second term converges to 0 by (4.160), so using (4.159) twice on the first term,

$$\limsup_{K\to\infty} \left|\langle f, Q(\mu_t) - Q(\mu_t^K)\rangle\right| \le CR^{-p}a_2\Lambda_l(\mu_0) \tag{4.162}$$

and, since R was arbitrary, we have shown that

$$\langle f, Q(\mu_t^K)\rangle \to \langle f, Q(\mu_t)\rangle\,. \tag{4.163}$$

Finally, integrating the bound on $|\mathcal{L}f - \mathcal{L}_K f|$, we find

$$\begin{aligned} \left|\langle f, Q(\mu_t^K) - Q_K(\mu_t^K)\rangle\right| &\le \int_{\mathbb{R}^d\times\mathbb{R}^d} |\mathcal{L}f - \mathcal{L}_K f|(v, v_\star)\mu_t^K(dv)\mu_t^K(dv_\star) \\ &\le C(f)\epsilon_K \int_{\mathbb{R}^d\times\mathbb{R}^d} (1+|v|^2+|v_\star|^2)\mu_t^K(dv)\mu_t^K(dv_\star) \\ &\le C(f)\epsilon_K \to 0 \end{aligned} \tag{4.164}$$

and, combining with (4.163), we see that $\langle f, Q(\mu_t^K)\rangle \to \langle f, Q(\mu_t)\rangle$ for all $t \ge 0$ as claimed.

We now conclude. For any $t \ge 0$ and any bounded, Lipschitz f, we have

$$\langle f, \mu_t^K\rangle = \langle f, \mu_0\rangle + \int_0^t \langle f, Q_K(\mu_s^K)\rangle ds. \tag{4.165}$$

The integrand $\langle f, Q_K(\mu_s^K)\rangle$ is bounded, uniformly in $s \le t$ and $K \ge 1$, and converges to $\langle f, Q(\mu_s)\rangle$ for all s, while the left-hand side converges to $\langle f, \mu_t\rangle$. We therefore take the limit $K \to \infty$ to conclude that, for all bounded, Lipschitz f and all $t \ge 0$

$$\langle f, \mu_t\rangle = \langle f, \mu_0\rangle + \int_0^t \langle f, Q(\mu_s)\rangle ds \tag{4.166}$$

as desired. □

Finally, Lemma 4.28 follows much the same pattern as above.

Proof of Lemma 4.28. To extend the maps ϕ_t, fix $\mu_0 \in \mathcal{S}^{p+2}$. Using Proposition 2.4 again, let $\mu_0^N \in \mathcal{S}_N$ be a sequence of discrete measures such that $w_p(\mu_0^N, \mu_0) \to 0$ and $\Lambda_{p+\gamma}(\mu_0^N) \to \Lambda_{p+\gamma}(\mu_0), \Lambda_{p+2}(\mu_0^N) \to \Lambda_{p+2}(\mu_0)$; in particular, $\Lambda_{p+\gamma}(\mu_0^N) \le 2a$ for all N large enough. The bound (4.141) obtained in the previous lemma applies to show that, for all such N and all $t \ge 0$,

$$\mathcal{W}_1(\phi_t(\mu_0^N), \phi_t(\mu_0^{N'})) \le w_p\left(\phi_t(\mu_0^N), \phi_t(\mu_0^{N'})\right) \le e^{Ca(1+t)} w_p(\mu_0^N, \mu_0^{N'}). \tag{4.167}$$

The same argument as in Lemma 4.22 shows that right-hand side converges to 0 as $N, N' \to \infty$, which implies that $\phi_t(\mu_0^N)$ converges, uniformly in $\mathcal{W}_1$ on compact time intervals, to some limit. If we now define $\phi_t(\mu_0)$ to be this limit, a similar calculation shows that the limit $\phi_t(\mu_0)$ takes values in $\mathcal{S}$ and is independent of the choice of limiting sequence, and the same argument as in Lemma 4.27 above shows that $(\phi_t(\mu_0), t \geq 0)$ is again a solution to the noncutoff Boltzmann equation (BE). Finally, if μ_0, ν_0 are two such measures, one applies the conclusion (4.141) of the previous lemma to approximating sequences μ_0^N, ν_0^N and passes to the limit $N \to \infty$ to obtain the same result for μ_0, ν_0, again up to a new constant C in the exponent. □

4.10 Main Calculations on the Tanaka Coupling

We now give the proof of Lemmas 4.10, 4.12, which we deferred earlier.

4.10.1 Some Estimates for G

In preparation for the proofs of Lemma 4.10, we will first record some basic estimates concerning the regularity and integrability of G.

Lemma 4.29. *i.) Let G be as above. Then, for some constants $0 < c_1 \le c_2 < \infty$, we have*

$$c_1(1+z)^{-1/\nu} \le G(z) \le c_2(1+z)^{-1/\nu}. \tag{4.168}$$

Moreover, G is continuously differentiable, and c_1, c_2 above can be chosen such that

$$c_1(1+z)^{-1-1/\nu} \le |G'(z)| \le c_2(1+z)^{-1-1/\nu}. \tag{4.169}$$

ii.) We have

$$\int_0^\infty z\left|\frac{d}{dz}\left(1-\cos G(z)\right)\right| dz < \infty. \tag{4.170}$$

iii.) There exists a constant $c < \infty$ such that, for all $x, y > 0$,

$$\int_0^\infty \left(G\left(\frac{z}{x}\right) - G\left(\frac{z}{y}\right)\right)^2 dz \le c\frac{|x-y|^2}{x+y}. \tag{4.171}$$

Proof. i). For the first claim, we use the definition of H and the asymptotics of b in (NCHP) to see that, for some constants $c_1, c_2 \in (0,\infty)$ and all $\theta \in (0, \pi/2)$,

$$c_1\int_\theta^{\pi/2} x^{-1-\nu}dx \le H(\theta) \le c_2\int_\theta^{\pi/2} x^{-1-\nu}dx \tag{4.172}$$

so that

$$\frac{c_1}{\nu}\left(\theta^{-\nu} - \left(\frac{\pi}{2}\right)^{-\nu}\right) \le H(\theta) \le \frac{c_2}{\nu}\left(\theta^{-\nu} - \left(\frac{\pi}{2}\right)^{-\nu}\right). \tag{4.173}$$

The first claim now follows, potentially for a new choice of c_1, c_2. The differentiability is an immediate consequence of the inverse function theorem. Indeed, we have

$$G'(z) = \frac{1}{H'(G(z))} = -\frac{1}{\beta(G(z))} \tag{4.174}$$

and so the second claim follows from the first, using (NCHP) again.

ii). We have

$$z\frac{d}{dz}((1-\cos G(z))) = z\left(\sin G(z)\right) G'(z) \tag{4.175}$$

and so

$$\left|z\frac{d}{dz}(1-\cos G(z))\right| \le zG(z)|G'(z)|. \tag{4.176}$$

Using the bounds from the previous part, it follows that the right-hand side is bounded by $c_2(1+z)^{-2/\nu}$ for some $c_2 < \infty$, which is integrable because $\nu \in (0,1)$.

iii). We follow [85, Lemma 1.1]. Recalling that G is decreasing, and integrating the bound on G' found in part i)., we see that, for all $0 \le z \le w$ and some $c < \infty$, we have

$$0 \le G(z) - G(w) \le c\left((1+z)^{-1/\nu} - (1+w)^{-1/\nu}\right). \tag{4.177}$$

We also recall that, for all $a > b > 0$, we have

$$a^{1/\nu} - b^{1/\nu} \le c\frac{a-b}{a^{1-1/\nu} + b^{1-1/\nu}}. \tag{4.178}$$

For any $z > 0$, $0 < y < x$, we apply this bound with $a = (1+z/x)^{-1}, b = (1+z/y)^{-1}$ to obtain

$$\begin{aligned} 0 \le G\left(\frac{z}{x}\right) - G\left(\frac{z}{y}\right) &\le c\left((1+z/x)^{-1/\nu} - (1+z/y)^{-1/\nu})\right) \\ &\le c\left|\frac{x}{x+z} - \frac{y}{y+z}\right|\left(1+\frac{z}{x}\right)^{1-1/\nu} \\ &\le c|x-y|(x+z)^{-1/\nu}x^{-1+1/\nu}. \end{aligned} \tag{4.179}$$

We square and integrate over z, to obtain for all $x > y > 0$,

$$\begin{aligned} \int_0^\infty \left(G\left(\frac{z}{x}\right) - G\left(\frac{z}{y}\right)\right)^2 dz &\le c|x-y|^2 x^{1-2/\nu} x^{-2+2/\nu} \\ &= c\frac{|x-y|^2}{x} \le c\frac{|x-y|^2}{x+y} \end{aligned} \tag{4.180}$$

where in the final equality we recall that $y < x$ so $\frac{1}{x} \le \frac{2}{x+y}$. This concludes the proof of both claimed bounds in the case $x > y > 0$; for $y > x$, we reverse the roles of $x \leftrightarrow y$.

□

4.10.2 Proof of Lemma 4.10

We now turn to the proof of Lemma 4.10, which was deferred earlier. In order to avoid unnecessarily unwieldy expressions, we introduce some notation. We work with fixed α, and omit this from all notation; no calculations will depend on this choice, so all bounds are independent of α. We define $x = |v - v_\star|, \widetilde{x} = |\widetilde{v} - \widetilde{v}_\star|$, and write L for the cuttoff $L = K\widetilde{x}^\gamma$. We will also write R for $R_\alpha(v - v_\star, \widetilde{v} - \widetilde{v}_\star)$, and suppress the dependence of $a, \widetilde{a}_K, \mathcal{E}_{p,K}$ on $v, \widetilde{v}, v_\star, \widetilde{v}_\star, \alpha$. Throughout, c will denote a constant which is allowed to depend only on B through G and on d, and C will denote a constant which is also allowed to depend on p; both are understood to vary from line to line as necessary. We will also write expressions as though $K < \infty$, understanding that negative powers of K or integrals $\int_L^\infty ...dz$ are 0 if $K = \infty$.

Our first lemma is the following, which gives us control over the 'Povzner' term, similar to the estimates in Section 2.5. Since this estimate produces the key negative term in Lemma 4.10 and is essential for subsequent calculations, it is presented as a separate lemma.

Lemma 4.30. *For all $v, v_\star, z$, we have the bound*

$$\begin{aligned} |v+a|^p \le |v|^p \left(\frac{1+\cos G(z/x^\gamma)}{2}\right)^{p/2} + |v_\star|^p \left(\frac{\sin G(z/x^\gamma)}{2}\right)^{p/2} \\ + C\left(|v|^{p-1}|v_\star| + |v||v_\star|^{p-1}\right)\sin G(z/x^\gamma) \\ =: f_p(|v|,|v_\star|,z,x). \end{aligned} \tag{4.181}$$

Proof. Let us start from

$$v+a = v\left(\frac{1+\cos G(z/x^\gamma)}{2}\right) + v_\star\left(\frac{1-\cos G(z/x^\gamma)}{2}\right) + \frac{\sin G(z/x^\gamma)}{2}\Gamma(v-v_\star,\varphi). \tag{4.182}$$

We now take the norm of both sides, recalling that $|\Gamma(v-v_\star,\varphi)| = |v-v_\star|$:

$$\begin{aligned} |v+a|^2 = \left(\frac{1+\cos G(z/x^\gamma)}{2}\right)^2 |v|^2 + \left(\frac{1-\cos G(z/x^\gamma)}{2}\right)^2 |v_\star|^2 \\ + \left(\frac{\sin G(z/x^\gamma)}{2}\right)^2 (|v|^2 + |v_\star|^2 + 2|v||v_\star|) \\ + \left(\frac{\sin G(z/x^\gamma)}{2}\right)\left(\frac{1+\cos G(z/x^\gamma)}{2}\right) v\cdot\Gamma(v-v_\star,\varphi) \\ + \left(\frac{\sin G(z/x^\gamma)}{2}\right)\left(\frac{1-\cos G(z/x^\gamma)}{2}\right) v_\star\cdot\Gamma(v-v_\star,\varphi) \\ + \left(\frac{1-\cos G(z/x^\gamma)}{2}\right)\left(\frac{1+\cos G(z/x^\gamma)}{2}\right) v\cdot v_\star. \end{aligned} \tag{4.183}$$

For the third and fourth lines, we use orthogonality to see that $v\cdot\Gamma(v-v_\star,\varphi) = v_\star\cdot\Gamma(v-v_\star,\varphi)$. It follows that

$$|v\cdot\Gamma(v-v_\star,\varphi)| \le \min(|v|,|v_\star|)(|v|+|v_\star|) \le 2|v||v_\star|. \tag{4.184}$$

Using the inequality $1-\cos G(z) \le \sin G(z)$, we now group similar terms to obtain

$$\begin{aligned} |v+a|^2 \le \left(\frac{1+\cos G(z/x^\gamma)}{2}\right)|v|^2 + \left(\frac{1-\cos G(z/x^\gamma)}{2}\right)|v_\star|^2 + C\sin G(z/x^\gamma)|v||v_\star| \\ := h_1 + h_2 + h_3. \end{aligned} \tag{4.185}$$

We now raise both sides to the $(p/2)^{\text{th}}$ power, recalling the inequality $(x+y)^{p/2} \le x^{p/2} + y^{p/2} + C(xy^{p/2-1} + x^{p/2-1}y)$, valid for all $x,y>0$. It is straightforward to see that the cross terms are dominated by the final term in (4.181):

$$h_1^{p/2-1}(h_2+h_3) + h_1(h_2+h_3)^{p/2-1} \le C(|v|^{p-1}|v_\star| + |v||v_\star|^{p-1})\sin G(z/x^\gamma); \tag{4.186}$$

$$h_2^{p/2-1}h_3 + h_2 h_3^{p/2-1} \le C(|v|^{p-1}|v_\star| + |v||v_\star|^{p-1})\sin G(z/x^\gamma); \tag{4.187}$$

$$h_3^{p/2} \le C(|v|^{p-1}|v_\star| + |v||v_\star|^{p-1})\sin G(z/x^\gamma). \tag{4.188}$$

Using the inequality twice, we thus obtain

$$\begin{aligned} |v+a|^p &\le h_1^{p/2} + (h_2+h_3)^{p/2} + C(|v|^{p-1}|v_\star| + |v||v_\star|^{p-1})\sin G(z/x^\gamma) \\ &\le h_1^{p/2} + h_2^{p/2} + h_3^{p/2} + C(|v|^{p-1}|v_\star| + |v||v_\star|^{p-1})\sin G(z/x^\gamma) \\ &\le h_1^{p/2} + h_2^{p/2} + C(|v|^{p-1}|v_\star| + |v||v_\star|^{p-1})\sin G(z/x^\gamma) \end{aligned} \tag{4.189}$$

which gives the desired bound on substituting the definitions of h_1, h_2. □

We now break up $\mathcal{E}_{p,K}$ as follows. We define

$$\mathcal{E}^1_{p,K} = \int_0^\infty dz \int_{\mathbb{S}^{d-2}} d\varphi \left(|v'|^p|v' - \widetilde{v}'_K|^2 - |v|^p|v-\widetilde{v}|^2\right); \tag{4.190}$$

$$\mathcal{E}^2_{p,K}(v,\widetilde{v},v_\star,\widetilde{v}_\star) = \int_0^\infty dz \int_{\mathbb{S}^{d-2}} d\varphi \left(|\widetilde{v}'_K|^p|v' - \widetilde{v}'_K|^2 - |\widetilde{v}|^p|v-\widetilde{v}|^2\right); \tag{4.191}$$

$$\mathcal{E}^3_{p,K}(v,\widetilde{v},v_\star,\widetilde{v}_\star) = \int_0^\infty dz \int_{\mathbb{S}^{d-2}} d\varphi \left(|v' - \widetilde{v}'_K|^2 - |v-\widetilde{v}|^2\right). \tag{4.192}$$

In this way, using the definition of d_p, it follows that $\mathcal{E}_{p,K} = \mathcal{E}^1_{p,K} + \mathcal{E}^2_{p,K} + \mathcal{E}^3_{p,K}$. It therefore suffices to prove the following estimates.

Lemma 4.31. *For some constants $K_0 = K_0(p), c = c(G,d)$ and $C = C(G,d,p)$, and $q = p+2+\gamma$, whenever $K \ge K_0(p)$, we have*

$$\begin{aligned} \mathcal{E}^1_{p,K}(v,\widetilde{v},v_\star,\widetilde{v}_\star) \le &\left(c + \left(c - \frac{\lambda_p}{2}\right)|v|^{p+\gamma} + c|\widetilde{v}|^{p+\gamma}\right)|v-\widetilde{v}|^2 \\ &+ \left(c|v_\star|^{p+\gamma} + c|\widetilde{v}_\star|^{p+\gamma}\right)|v_\star - \widetilde{v}_\star|^2 \\ &+ C\left(|v_\star|^{p+\gamma} + |\widetilde{v}_\star|^{p+\gamma}\right)(1+|v|^p+|\widetilde{v}|^p)|v-\widetilde{v}|^2 \\ &+ C\left(|v|^{p+\gamma} + |\widetilde{v}|^{p+\gamma}\right)(1+|v_\star|^p+|\widetilde{v}_\star|^p)|v_\star-\widetilde{v}_\star|^2 \\ &+ CK^{1-1/\nu}(1+|v|^l+|v_\star|^l+|\widetilde{v}|^l+|\widetilde{v}_\star|^l); \end{aligned} \tag{4.193}$$

$$\begin{aligned} \mathcal{E}^2_{p,K}(v,\widetilde{v},v_\star,\widetilde{v}_\star) \le &\left(c + \left(c - \frac{\lambda_p}{2}\right)|\widetilde{v}|^{p+\gamma} + c|v|^{p+\gamma}\right)|v-\widetilde{v}|^2 \\ &+ \left(c|v_\star|^{p+\gamma} + c|\widetilde{v}_\star|^{p+\gamma}\right)|v_\star - \widetilde{v}_\star|^2 \\ &+ C\left(|v_\star|^{p+\gamma} + |\widetilde{v}_\star|^{p+\gamma}\right)(1+|v|^p+|\widetilde{v}|^p)|v-\widetilde{v}|^2 \\ &+ C\left(|v|^{p+\gamma} + |\widetilde{v}|^{p+\gamma}\right)(1+|v_\star|^p+|\widetilde{v}_\star|^p)|v_\star-\widetilde{v}_\star|^2 \\ &+ CK^{1-1/\nu}(1+|v|^l+|v_\star|^l+|\widetilde{v}|^l+|\widetilde{v}_\star|^l) \end{aligned} \tag{4.194}$$

and

$$\begin{aligned}\mathcal{E}^3_{p,K}(v,\widetilde{v},v_\star,\widetilde{v}_\star) \le c(1+|v|^\gamma+|\widetilde{v}|^\gamma+|v_\star|^\gamma+|\widetilde{v}_\star|^\gamma)(|v-\widetilde{v}|^2+|v_\star-\widetilde{v}_\star|^2) \\ +CK^{1-1/\nu}(1+|v|^l+|\widetilde{v}|^l+|v_\star|^l+|\widetilde{v}_\star|^l).\end{aligned} \tag{4.195}$$

Proof of Lemmas 4.10, 4.31. Let us begin from the bound (4.181), and define also

$$f_p^\star(|v|,|v_\star|) = |v|^p + C(|v||v_\star|^{p-1}+|v|^{p-1}|v_\star|)+|v_\star|^p \tag{4.196}$$

which is an upper bound for f_p, uniformly in z,x. We therefore find

$$\mathcal{E}^1_{p,K} \le \int_0^\infty dz \int_{\mathbb{S}^{d-2}} d\varphi\, (f_p(|v|,|v_\star|,z,x)|v-\widetilde{v}+a-\widetilde{a}|^2 - |v|^p|v-\widetilde{v}|^2). \tag{4.197}$$

Let us also introduce

$$\widehat{a} = a(\widetilde{v},\widetilde{v}_\star,z,R(v-v_\star,\widetilde{v}-\widetilde{v}_\star)\varphi) \tag{4.198}$$

so that $\widetilde{a} = \widehat{a}\mathbb{1}(z\le L)$. We can therefore replace $\widetilde{a}$ by $\widehat{a}$, introducing a further error:

$$\begin{aligned}\mathcal{E}^1_{p,K} \le \int_0^\infty dz \int_{\mathbb{S}^{d-2}} d\varphi\, (f_p(|v|,|v_\star|,z,x)|v-\widetilde{v}+a-\widehat{a}|^2 - |v|^p|v-\widetilde{v}|^2) \\ + \int_L^\infty dz \int_{\mathbb{S}^{d-2}} d\varphi\, f_p(|v|,|v_\star|,z,x)\left(|v-\widetilde{v}+a|^2 - |v-\widetilde{v}+a-\widehat{a}|^2\right).\end{aligned} \tag{4.199}$$

Finally, we expand the squared norm $|v-\widetilde{v}+a-\widehat{a}|^2$ in the first line to obtain the decomposition

$$\mathcal{E}^1_{p,K} \le \mathcal{T}_1+\mathcal{T}_2+\mathcal{T}_3+\mathcal{T}_4 \tag{4.200}$$

where we define

$$\mathcal{T}_1 := \int_0^\infty dz \int_{\mathbb{S}^{d-2}} d\varphi\, \left(f_p(|v|,|v_\star|,z,x)-|v|^p\right)|v-\widetilde{v}|^2; \tag{4.201}$$

$$\mathcal{T}_2 := 2\int_0^\infty dz \int_{\mathbb{S}^{d-2}} d\varphi\, f_p(|v|,|v_\star|,z,x)(v-\widetilde{v})\cdot(a-\widehat{a}); \tag{4.202}$$

$$\mathcal{T}_3 := \int_0^\infty dz \int_{\mathbb{S}^{d-2}} d\varphi\, f_p^\star(|v|,|v_\star|)|a-\widehat{a}|^2; \tag{4.203}$$

$$\mathcal{T}_4 := \int_L^\infty dz \int_{\mathbb{S}^{d-2}} d\varphi\, f_p(|v|,|v_\star|,z,x)\left(|v+a-\widetilde{v}|^2 - |v+a-\widetilde{v}-\widehat{a}|^2\right). \tag{4.204}$$

We will now analyse this bound for $\mathcal{E}^1_{p,K}$ in detail, and an equivalent analysis of $\mathcal{E}^2_{p,K},\mathcal{E}^3_{p,K}$ will be discussed at the end of the proof. Let us now deal with these terms one by one.

1. Analysis of $\mathcal{T}_1$. Recalling the construction of G, the moment integral in $\mathcal{T}_1$ can be reparametrised in terms of θ:

$$\begin{aligned}\int_0^\infty (f_p(|v|,|v_\star|,z,x)-|v|^p)dz \\ &= -|v-v_\star|^\gamma|v|^p\int_0^{\pi/2}\beta(\theta)\left(1-\left(\frac{1+\cos\theta}{2}\right)^{p/2}\right)d\theta \\ &\quad + |v-v_\star|^\gamma(|v_\star|^p + C(|v||v_\star|^{p-1}+|v|^{p-1}|v_\star|))\int_0^{\pi/2}\beta(\theta)\sin(\theta)d\theta \\ &\le -\lambda_p|v-v_\star|^\gamma|v|^p + C|v-v_\star|^\gamma(|v_\star|^p+|v||v_\star|^{p-1}+|v|^{p-1}|v_\star|).\end{aligned} \tag{4.205}$$

On the negative term, we use the bound $|v|^\gamma-|v_\star|^\gamma\le|v-v_\star|^\gamma$ and Young's inequality to see that

$$\begin{aligned}-|v-v_\star|^\gamma|v|^p &\le -|v|^{p+\gamma}+|v_\star|^\gamma|v|^p \\ &\le -|v|^{p+\gamma}+\frac{1}{4}|v|^{p+\gamma}+4^{p/\gamma}|v_\star|^{p+\gamma}.\end{aligned} \tag{4.206}$$

For the positive term in (4.205), we use $|v-v_\star|^\gamma\le|v|^\gamma+|v_\star|^\gamma$ to obtain

$$\begin{aligned}|v-v_\star|^\gamma(|v_\star|^p+|v||v_\star|^{p-1}+|v|^{p-1}|v_\star|) &\le |v|^\gamma|v_\star|^p+|v|^{\gamma+1}|v_\star|^{p-1}+|v|^{p+\gamma-1}|v| \\ &\quad +|v_\star|^{\gamma+p}+|v||v_\star|^{p+\gamma-1}+|v|^{p-1}|v_\star|^{p+\gamma}+|v|^{p-1}|v_\star|^{1+\gamma}.\end{aligned} \tag{4.207}$$

We now use Young's inequality on each term to obtain

$$C|v-v_\star|^\gamma(|v_\star|^p+|v||v_\star|^{p-1}+|v|^{p-1}|v_\star|)\le\frac{\lambda_p}{4}|v|^{p+\gamma}+C|v_\star|^{p+\gamma} \tag{4.208}$$

Combining, we have shown that

$$\int_0^\infty (f_p(|v|,|v_\star|,z,x)-|v|^p)dz\le-\frac{\lambda_p}{2}|v|^{p+\gamma}+C|v_\star|^{p+\gamma} \tag{4.209}$$

and so

$$\mathcal{T}_1\le-\frac{\lambda_p}{2}|v|^{p+\gamma}|v-\widetilde{v}|^2+C|v_\star|^{p+\gamma}|v-\widetilde{v}|^2. \tag{4.210}$$

2. Analysis of $\mathcal{T}_2$. We first observe that

$$\int_{\mathbb{S}^{d-2}}d\varphi\,(a-\widehat{a})=-\frac{1}{2}(1-\cos G(z/x^\gamma))(v-v_\star)+\frac{1}{2}(1-\cos G(z/\widetilde{x}^\gamma))(\widetilde{v}-\widetilde{v}_\star). \tag{4.211}$$

It therefore follows that

$$\mathcal{T}_2=(v-\widetilde{v})\cdot\{\Phi(\widetilde{x},|v|,|v_\star|,x)(\widetilde{v}-\widetilde{v}_\star)-\Phi(x,|v|,|v_\star|,x)(v-v_\star)\} \tag{4.212}$$

where we define, for any $y,u,v,w>0$,

$$\begin{aligned}\Phi(y,u,v,w)&=\int_0^\infty dz\;f_p(u,v,z,w)(1-\cos G(z/y^\gamma))dz \\ &:=\Psi(y^\gamma,u,v,w).\end{aligned} \tag{4.213}$$

We differentiate the function thus defined to obtain

$$\begin{aligned}\frac{\partial}{\partial y}\Psi(y,u,v,w) &= \int_0^\infty f_p(u,v,z,w)\left(-\frac{z}{y}\right)\frac{d}{dz}\left(1-\cos G\left(\frac{z}{y}\right)\right)dz \\ &= \int_0^\infty f_p(u,v,yz,w)\left(z\frac{d}{dz}(1-\cos G(z))\right)dz\end{aligned} \tag{4.214}$$

where the final line follows by an integration by substitution $z \mapsto yz$. From the calculations in Lemma 4.29, we therefore conclude that

$$\left|\frac{\partial}{\partial y}\Psi(y,u,v,w)\right| \le c f_p^\star(u,v). \tag{4.215}$$

Now, using the bound $|x^\gamma - y^\gamma| \le 2|x-y|/(x^{1-\gamma}+y^{1-\gamma})$, we obtain

$$|\Phi(x,|v|,|v_\star|,x) - \Phi(\widetilde{x},|v|,|v_\star|,x)| \le \frac{c|x-\widetilde{x}|}{x^{1-\gamma}+\widetilde{x}^{1-\gamma}} f_p^\star(|v|,|v_\star|) \tag{4.216}$$

and, for all $y > 0$,

$$|\Phi(x,|v|,|v_\star|,y)| \le c y^\gamma f_p^\star(|v|,|v_\star|). \tag{4.217}$$

We therefore obtain the bound

$$\begin{aligned}|\mathcal{T}_2| \le |v-\widetilde{v}|\Big\{&|v-v_\star-\widetilde{v}+\widetilde{v}_\star||\Phi(x,|v|,|v_\star|,x)+\Phi(\widetilde{x},|v|,|v_\star|,x)| \\ &+ (|v-v_\star|+|\widetilde{v}-\widetilde{v}_\star|)\,|\Phi(x,|v|,|v_\star|,x)-\Phi(\widetilde{x},|v|,|v_\star|,x)|\Big\} \\ \le c\left(|v-\widetilde{v}|^2+|v_\star-\widetilde{v}_\star|^2\right)&(|v|^\gamma+|v_\star|^\gamma+|\widetilde{v}|^\gamma+|\widetilde{v}_\star|^\gamma) f_p^\star(|v|,|v_\star|).\end{aligned} \tag{4.218}$$

3. Analysis of $\mathcal{T}_3$. We now turn to the term $\mathcal{T}_3$, and begin by noting that

$$\begin{aligned}a\cdot\widehat{a} = &\frac{1}{4}(1-\cos G(z/x^\gamma))(1-\cos G(z/\widetilde{x}^\gamma))(v-v_\star)\cdot(\widetilde{v}-\widetilde{v}_\star) \\ &-\frac{1}{4}(1-\cos G(z/x^\gamma))\sin G(z/\widetilde{x}^\gamma)(v-v_\star)\cdot\Gamma(\widetilde{v}-\widetilde{v}_\star,R\varphi) \\ &-\frac{1}{4}(1-\cos G(z/\widetilde{x}^\gamma))\sin G(z/x^\gamma)\Gamma(v-v_\star,\varphi)\cdot(\widetilde{v}-\widetilde{v}_\star) \\ &+\frac{1}{4}\sin G(z/x^\gamma)\sin G(z/\widetilde{x}^\gamma)\Gamma(v-v_\star,\varphi)\cdot\Gamma(\widetilde{v}-\widetilde{v}_\star,R\varphi).\end{aligned} \tag{4.219}$$

We now integrate over $\varphi \in \mathbb{S}^{d-2}$. Since $\int_{\mathbb{S}^{d-2}}\Gamma(u,\varphi)d\varphi = 0$ and R preserves the uniform measure $d\varphi$, the middle two lines integrate to 0. We also recall, from the construction of $R = R(v-v_\star,\widetilde{v}-\widetilde{v}_\star)$ in Lemma 4.8, that $\Gamma(v-v_\star,\varphi)\cdot\Gamma(\widetilde{v}-\widetilde{v}_\star,R\varphi) \ge (v-v_\star)\cdot(\widetilde{v}-\widetilde{v}_\star)$, and so integrating (4.219) gives

$$\begin{aligned}\int_{\mathbb{S}^{d-2}} a\cdot\widehat{a}\,d\varphi \ge &\frac{1}{4}\Big[(1-\cos G(z/x^\gamma))(1-\cos G(z/\widetilde{x}^\gamma)) \\ &+\sin G(z/x^\gamma)\sin G(z/\widetilde{x}^\gamma)\Big](v-v_\star)\cdot(\widetilde{v}-\widetilde{v}_\star) \\ = &\frac{1}{4}\Big[(1-\cos G(z/x^\gamma))+(1-\cos G(z/\widetilde{x}^\gamma)) \\ &-\Big(1-\cos\left(G(z/x^\gamma)-G(z/\widetilde{x}^\gamma)\right)\Big)\Big](v-v_\star)\cdot(\widetilde{v}-\widetilde{v}_\star).\end{aligned} \tag{4.220}$$

Similar, elementary calculations show that

$$|a|^2 = \frac{1}{2}(1-\cos G(z/x^\gamma))|v-v_\star|^2; \qquad |\widehat{a}|^2 = \frac{1}{2}(1-\cos G(z/\widetilde{x}^\gamma))|\widetilde{v}-\widetilde{v}_\star|^2. \tag{4.221}$$

We now observe that

$$\int_0^\infty (1-\cos G(z/x^\gamma))dz = cx^\gamma \tag{4.222}$$

and so, from (4.220, 4.221), we obtain

$$\begin{aligned}\int_0^\infty dz \int_{\mathbb{S}^{d-2}} d\varphi\, |a-\widehat{a}|^2 \le \frac{c}{2}\left(x^{2+\gamma}+\widetilde{x}^{2+\gamma}-(v-v_\star)\cdot(\widetilde{v}-\widetilde{v}_\star)(x^\gamma+\widetilde{x}^\gamma)\right)\\ +\frac{x\widetilde{x}}{4}\int_0^\infty \left(G(z/x^\gamma)-G(z/\widetilde{x}^\gamma)\right)^2 dz.\end{aligned} \tag{4.223}$$

Recalling that $x^2=(v-v_\star)\cdot(v-v_\star)$ and $\widetilde{x}^2=(\widetilde{v}-\widetilde{v}_\star)\cdot(\widetilde{v}-\widetilde{v}_\star)$, the term in parentheses on the first line rearranges to

$$\begin{aligned} &x^{2+\gamma}+\widetilde{x}^{2+\gamma}-(v-v_\star)\cdot(\widetilde{v}-\widetilde{v}_\star)(x^\gamma+\widetilde{x}^\gamma)\\ &\quad=(v-v_\star)\cdot\left[(v-v_\star)-(\widetilde{v}-\widetilde{v}_\star)\right]x^\gamma\\ &\qquad+(\widetilde{v}-\widetilde{v}_\star)\cdot\left[(\widetilde{v}-\widetilde{v}_\star)-(v-v_\star)\right]\widetilde{x}^\gamma\\ &\quad=\left((v-\widetilde{v})-(v_\star-\widetilde{v}_\star)\right)\cdot\left[(v-v_\star)x^\gamma-(\widetilde{v}-\widetilde{v}_\star)\widetilde{x}^\gamma\right].\end{aligned} \tag{4.224}$$

We argue similarly to (4.218), now with Φ replaced by $x^\gamma, \widetilde{x}^\gamma$, and controlling $|x^\gamma-\widetilde{x}^\gamma|$ as in (4.216). We thus obtain

$$\begin{aligned}&x^{2+\gamma}+\widetilde{x}^{2+\gamma}-(v-v_\star)\cdot(\widetilde{v}-\widetilde{v}_\star)(x^\gamma+\widetilde{x}^\gamma)\\ &\qquad\le c\left(|v-\widetilde{v}|^2+|v_\star-\widetilde{v}_\star|^2\right)\left(|v|^\gamma+|v_\star|^\gamma+|\widetilde{v}|^\gamma+|\widetilde{v}_\star|^\gamma\right).\end{aligned} \tag{4.225}$$

Let us now consider the final line of (4.223). By Lemma 4.29, we have the bound

$$\int_0^\infty \left(G(z/x^\gamma)-G(z/\widetilde{x}^\gamma)\right)^2 dz \le c\frac{|x^\gamma-\widetilde{x}^\gamma|^2}{x^\gamma+\widetilde{x}^\gamma}. \tag{4.226}$$

We therefore obtain

$$\begin{aligned}x\widetilde{x}\int_0^\infty \left(G(z/x^\gamma)-G(z/\widetilde{x}^\gamma)\right)^2 dz &\le c\frac{\min(x,\widetilde{x})}{\max(x,\widetilde{x})^{1-\gamma}}|x-\widetilde{x}|^2\\ &\le c(|v|^\gamma+|v_\star|^\gamma+|\widetilde{v}|^\gamma+|\widetilde{v}_\star|^\gamma)(|v-\widetilde{v}|^2+|v_\star-\widetilde{v}_\star|^2).\end{aligned} \tag{4.227}$$

Combining (4.223, 4.225, 4.227), we have shown that

$$\mathcal{T}_3 \le c(|v|^\gamma+|v_\star|^\gamma+|\widetilde{v}|^\gamma+|\widetilde{v}_\star|^\gamma)(|v-\widetilde{v}|^2+|v_\star-\widetilde{v}_\star|^2)f_p^\star(|v|,|v_\star|). \tag{4.228}$$

4. Analysis of $\mathcal{T}_4$. The final error term is the term $\mathcal{T}_4$, which corresponds to collisions in the noncutoff system with no corresponding event in the cutoff system. As a result, we anticipate that $\mathcal{T}_4$ will not be bounded in terms of $v-\widetilde{v}, v_\star-\widetilde{v}_\star$, but will be small in the limit $K\to\infty$. Let us recall that the integration limit L is defined as $L:=K\widetilde{x}^\gamma$. By expanding out the norms, we bound the integrand, for $z\ge L$,

$$\left|f_p(|v|,|v_\star|,z,x)(|v+a-\widetilde{v}|^2-|v+a-\widetilde{v}-\widehat{a}|^2)\right|\le cf_p^\star(|v|,|v_\star|)|\widehat{a}|(|v|+|\widetilde{v}|+|v_\star|+|\widetilde{v}_\star|). \tag{4.229}$$

As above, we have

$$|\widehat{a}|=\sqrt{\frac{1}{2}\left(1-\cos G\left(\frac{z}{\widetilde{x}^\gamma}\right)\right)}|\widetilde{v}-\widetilde{v}_\star|\le\frac{1}{2}G\left(\frac{z}{\widetilde{x}^\gamma}\right)|\widetilde{v}-\widetilde{v}_\star|. \tag{4.230}$$

We therefore obtain the bound

$$\mathcal{T}_4\le cf_p^\star(|v|,|v_\star|)(|v|^2+|v_\star|^2+|\widetilde{v}|^2+|\widetilde{v}_\star|^2)\int_L^\infty G\left(\frac{z}{\widetilde{x}^\gamma}\right)dz. \tag{4.231}$$

Recalling the definition of $L=K\widetilde{x}^\gamma$, the integral evalues to

$$\int_L^\infty G\left(\frac{z}{\widetilde{x}^\gamma}\right)dz=\widetilde{x}^\gamma\int_K^\infty G(z)dz\le c\widetilde{x}^\gamma K^{1-1/\nu}. \tag{4.232}$$

We therefore find

$$\mathcal{T}_4\le CK^{1-1/\nu}(|v|^{p+2+\gamma}+|\widetilde{v}|^{p+2+\gamma}+|v_\star|^{p+2+\gamma}+|\widetilde{v}_\star|^{p+2+\gamma}) \tag{4.233}$$

Recalling that $l:=p+2+\gamma$, this is exactly the error claimed.

5. Converting into the form desired. Combining (4.210, 4.218, 4.228, 4.233), we see that

$$\begin{aligned}\mathcal{E}^1_{p,K}\le&\left((c-\frac{\lambda_p}{2})|v|^{p+\gamma}+C|v_\star|^{p+\gamma}\right)|v-\widetilde{v}|^2\\&+c(|v|^\gamma+|\widetilde{v}|^\gamma+|v_\star|^\gamma+|\widetilde{v}_\star|^\gamma)f_p^\star(|v|,|v_\star|)(|v-\widetilde{v}|^2+|v_\star-\widetilde{v}_\star|^2)\\&+C(|v|^l+|\widetilde{v}|^l+|v_\star|^l+|\widetilde{v}_\star|^l)K^{1-1/\nu}.\end{aligned} \tag{4.234}$$

The first and last lines are already in the form desired in the statement of the lemma. Let us now examine the middle term. Using Young on the cross-terms in $f_p^\star$, we see that

$$f_p^\star(|v|,|v_\star|)\le 2|v|^p+C|v_\star|^p \tag{4.235}$$

and so

$$\begin{aligned}&(|v|^\gamma+|v_\star|^\gamma+|\widetilde{v}|^\gamma+|\widetilde{v}_\star|^\gamma)f_p^\star(|v|,|v_\star|)\\&\quad\le c(|v|^{p+\gamma}+|v|^p|v_\star|^\gamma+|v|^p|\widetilde{v}|^\gamma+|v|^p|\widetilde{v}_\star|^\gamma)\\&\qquad+C(|v_\star|^{p+\gamma}+|v_\star|^p|v|^\gamma+|v_\star|^p|\widetilde{v}|^\gamma+|v_\star|^p|\widetilde{v}_\star|^\gamma)\end{aligned} \tag{4.236}$$

We now use Young's inequality on all terms appearing in this expression; for the second term, we use Peter-Paul to find

$$C(|v_\star|^{p+\gamma}+|v_\star|^p|v|^\gamma+|v_\star|^p|\widetilde{v}|^\gamma+|v_\star|^p|\widetilde{v}_\star|^\gamma)\le c|v|^{p+\gamma}+c|v_\star|^{p+\gamma}+C(|v_\star|^{p+\gamma}+|\widetilde{v}_\star|^{p+\gamma}). \tag{4.237}$$

Therefore,

$$\begin{aligned}(|v|^\gamma+|v_\star|^\gamma+|\widetilde{v}|^\gamma+|\widetilde{v}_\star|^\gamma)f_p^\star(|v|,|v_\star|)\\ \le c|v|^{p+\gamma}+c|\widetilde{v}|^{p+\gamma}+C(1+|v_\star|^{p+\gamma}+|\widetilde{v}_\star|^{p+\gamma})(1+|v|^p+|\widetilde{v}|^p).\end{aligned} \tag{4.238}$$

We use this inequality for the term multiplying $|v-\widetilde{v}|^2$ in the second line of (4.234), and reverse the roles of $v\leftrightarrow v_\star, \widetilde{v}\leftrightarrow\widetilde{v}_\star$ for the term involving $|v_\star-\widetilde{v}_\star|^2$. Together, we see that

$$\begin{aligned}&(|v|^\gamma+|\widetilde{v}|^\gamma+|v_\star|^\gamma+|\widetilde{v}_\star|^\gamma)f_p^\star(|v|,|v_\star|)(|v-\widetilde{v}|^2+|v_\star-\widetilde{v}_\star|^2)\\ &\qquad\le c(|v|^{p+\gamma}+|\widetilde{v}|^{p+\gamma})|v-\widetilde{v}|^2+c(|v_\star|^{p+\gamma}+|\widetilde{v}_\star|^{p+\gamma})|v_\star-\widetilde{v}_\star|^2\\ &\qquad+C(1+|v_\star|^{p+\gamma}+|\widetilde{v}_\star|^{p+\gamma})(1+|v|^p+|\widetilde{v}|^p)|v-\widetilde{v}|^2\\ &\qquad+C(1+|v|^{p+\gamma}+|\widetilde{v}|^{p+\gamma})(1+|v_\star|^p+|\widetilde{v}_\star|^p)|v_\star-\widetilde{v}_\star|^2\end{aligned} \tag{4.239}$$

which gives the bound desired for $\mathcal{E}^1_{p,K}$.

6. Estimate on $\mathcal{E}^2_{p,K}$. We now turn to the analysis of $\mathcal{E}^2_{p,K}$, which follows a similar pattern to $\mathcal{E}^1_{p,K}$ above. In this case, we use the bound

$$|\widetilde{v}+\widetilde{a}_K|^p\le f_{p,L}(|\widetilde{v}|,|\widetilde{v}_\star|,z,\widetilde{x})=\begin{cases}f_p(|\widetilde{v}|,|\widetilde{v}_\star|,z,\widetilde{x}), & z\le L;\\ |\widetilde{v}|^p, & z>L\end{cases} \tag{4.240}$$

which has the same upper bound $f_p^\star$. We therefore obtain a decomposition equivalent to (4.200):

$$\mathcal{E}^2_{p,K}\le\widetilde{\mathcal{T}}_1+\widetilde{\mathcal{T}}_2+\widetilde{\mathcal{T}}_3+\widetilde{\mathcal{T}}_4 \tag{4.241}$$

where

$$\widetilde{\mathcal{T}}_1:=\int_0^\infty dz\,(f_{p,L}(|\widetilde{v}|,|\widetilde{v}_\star|,z,x)-|\widetilde{v}|^p)|v-\widetilde{v}|^2; \tag{4.242}$$

$$\widetilde{\mathcal{T}}_2:=2\int_0^L dz\int_{\mathbb{S}^{d-2}}d\varphi\, f_{p,L}(|\widetilde{v}|,|\widetilde{v}_\star|,z,x)(v-\widetilde{v})\cdot(a-\widehat{a}); \tag{4.243}$$

$$\widetilde{\mathcal{T}}_3:=\int_0^\infty dz\int_{\mathbb{S}^{d-2}}d\varphi\, f_p^\star(|\widetilde{v}|,|\widetilde{v}_\star|)|a-\widehat{a}|^2; \tag{4.244}$$

$$\widetilde{\mathcal{T}}_4:=\int_L^\infty dz\int_{\mathbb{S}^{d-2}}d\varphi\, f_{p,L}(|\widetilde{v}|,|\widetilde{v}_\star|,z,\widetilde{x})\left|2(v-\widetilde{v})\cdot\widehat{a}+|\widehat{a}|^2\right| \tag{4.245}$$

The analyses of $\widetilde{\mathcal{T}}_3,\widetilde{\mathcal{T}}_4$ are identical to the arguments above, and we will now discuss the necessary modifications for $\widetilde{\mathcal{T}}_1,\widetilde{\mathcal{T}}_2$.

6a. Analysis of $\widetilde{\mathcal{T}}_1$. Let us begin with $\widetilde{\mathcal{T}}_1$. The same reparametrisation gives

$$\int_0^\infty (f_{p,L}(|\widetilde{v}|, |\widetilde{v}_\star|z, \widetilde{x}) - |\widetilde{v}|^p)dz \le -|\widetilde{v} - \widetilde{v}_\star|^\gamma |\widetilde{v}|^p \int_{\theta_0(K)}^{\pi/2} \left(1 - \left(\frac{1+\cos\theta}{2}\right)^{p/2}\right)\beta(\theta)d\theta \\ + C|\widetilde{v} - \widetilde{v}_\star|^\gamma (|\widetilde{v}_\star|^p + |\widetilde{v}|^{p-1}|\widetilde{v}_\star| + |\widetilde{v}||\widetilde{v}_\star|^{p-1}). \tag{4.246}$$

We therefore obtain

$$\int_0^\infty (f_{p,L}(|\widetilde{v}|, |\widetilde{v}_\star|z, \widetilde{x}) - |\widetilde{v}|^p)dz \\ \le -|\widetilde{v} - \widetilde{v}_\star|^\gamma |\widetilde{v}|^p \lambda_{p,K} + |\widetilde{v} - \widetilde{v}_\star|^\gamma (|\widetilde{v}_\star|^p + C(|\widetilde{v}|^{p-1}|\widetilde{v}_\star| + |\widetilde{v}||\widetilde{v}_\star|^{p-1})) \\ \le -\lambda_{p,K}|\widetilde{v}|^{p+\gamma} + \lambda_p |\widetilde{v}_\star|^\gamma |\widetilde{v}|^p + C|\widetilde{v} - \widetilde{v}_\star|^\gamma (|\widetilde{v}_\star|^p + |\widetilde{v}|^{p-1}|\widetilde{v}_\star| + |\widetilde{v}||\widetilde{v}_\star|^{p-1}) \tag{4.247}$$

where

$$\lambda_{p,K} := \int_{\theta_0(K)}^{\pi/2} \left(1 - \left(\frac{1+\cos\theta}{2}\right)^{p/2}\right)\beta(\theta)d\theta \le \lambda_p. \tag{4.248}$$

We now use Peter-Paul on the positive terms, independently of K, to obtain

$$\lambda_p |\widetilde{v}_\star|^\gamma |\widetilde{v}|^p + C|\widetilde{v} - \widetilde{v}_\star|^\gamma (|\widetilde{v}_\star|^p + |\widetilde{v}|^{p-1}|\widetilde{v}_\star| + |\widetilde{v}||\widetilde{v}_\star|^{p-1}) \le \frac{\lambda_p}{3}|\widetilde{v}|^{p+\gamma} + C|\widetilde{v}_\star|^{p+\gamma}. \tag{4.249}$$

By monotone convergence, $\lambda_{p,K} \to \lambda_p$ as $K \to \infty$ with p fixed; in particular, for some $K_0 = K_0(G,p,d)$ and all $K \ge K_0(G,p,d)$, $\lambda_{p,K} \ge \frac{5}{6}\lambda_p$. For such K, we have shown that

$$\widetilde{\mathcal{T}}_1 \le -\frac{\lambda_p}{2}|\widetilde{v}|^{p+\gamma}|v - \widetilde{v}|^2 + C|\widetilde{v}_\star|^{p+\gamma}|v - \widetilde{v}|^2. \tag{4.250}$$

6b. Analysis of $\widetilde{\mathcal{T}}_2$. Following the same manipulations as (4.212), we obtain

$$\widetilde{\mathcal{T}}_2 = (v - \widetilde{v}) \cdot \{(\Psi_{0L} + \Psi_{L\infty})(\widetilde{x}^\gamma, |\widetilde{v}|, |\widetilde{v}_\star|, \widetilde{x})(\widetilde{v} - \widetilde{v}_\star) - (\Psi_{0L} + \Psi_{L\infty})(\widetilde{x}^\gamma, |\widetilde{v}|, |\widetilde{v}_\star|, \widetilde{x})(v - v_\star)\} \tag{4.251}$$

where we define

$$\Psi_{0L}(y,u,v,w) = \int_0^L f_p(u,v,z,w)(1 - \cos G(z/y))dz \tag{4.252}$$

and

$$\Psi_{L\infty}(y,u,v,w) = \int_L^\infty v^p(1 - \cos G(z/y))dz. \tag{4.253}$$

One then repeats the differentiation (4.214) for each part separately, to obtain a bound

$$\left|\frac{\partial}{\partial y}\Psi_{0L}(y,u,v,w)\right| + \left|\frac{\partial}{\partial y}\Psi_{L\infty}(y,u,v,w)\right| \le cf_p^\star(u,v) \tag{4.254}$$

and the rest of the argument follows as for $\mathcal{T}_2$.

7. Bound on $\mathcal{E}^3_{p,K}$. Finally, let us mention $\mathcal{E}^3_{p,K}$. This term is strictly easier than the two above: there is no term analagous to $\mathcal{T}_1$, and one can omit the moment prefactors in the remaining terms. Alternatively, one may note that $\mathcal{E}^3_{p,K}$ is exactly that analysed in [92, Lemma 3.1], and the claimed bound is exactly the content of [92, Lemma 5.1]. □

4.10.3 Proof of Lemma 4.12

We now turn to the proof of the quadratic bound Lemma 4.12, where we replace the integrand of $\mathcal{E}_{p,K}$ with its square. In this case, the integrand is nonnegative, and there is no hope of exploiting cancellations in the way we did above. On the other hand, the statement we seek to prove is much weaker; we ask only for local boundedness of $\mathcal{Q}_K$, rather than being small in a suitable sense when $|v-\widetilde{v}|, |v_\star - \widetilde{v}_\star|$ are small. It will be sufficient to prove the following slightly simpler lemma, which breaks up $\mathcal{Q}_K$ in a similar way to the decomposition $\mathcal{E}_{p,K} = \mathcal{E}^1_{p,K} + \mathcal{E}^2_{p,K} + \mathcal{E}^3_{p,K}$ above.

Lemma 4.32. *Define*

$$\mathcal{Q}^1_K = \int_0^\infty dz \int_{\mathbb{S}^{d-2}} d\varphi \, \Big(d_p(v', \widetilde{v}'_K) - d_p(v, \widetilde{v})\Big)^2; \tag{4.255}$$

$$\mathcal{Q}^2_K = \int_0^\infty dz \int_{\mathbb{S}^{d-2}} d\varphi \, \Big(d_p(v'_\star, \widetilde{v}'_{\star K}) - d_p(v_\star, \widetilde{v}_\star)\Big)^2. \tag{4.256}$$

Then the estimate (4.27) holds with either $\mathcal{Q}^1_K$ or $\mathcal{Q}^2_K$ in place of $\mathcal{Q}_K$.

Once we have established these estimates, the second point of Lemma 4.12 follows from the easy comparison $\mathcal{Q}_K \le 2\mathcal{Q}^1_K + 2\mathcal{Q}^2_K$.

Proof of Lemmas 4.12. We use the same notation as above, and start from a decomposition similar to (4.200):

$$\begin{aligned} d_p(v', \widetilde{v}'_K)^2 - d_p(v,\widetilde{v}) = (|v'|^p + |\widetilde{v}'_K|^p - |v|^p - |\widetilde{v}|^p)|v-\widetilde{v}|^2 \\ + (1+|v'|^p + |\widetilde{v}'_K|^p)(2(a-\widehat{a})\cdot(v-\widetilde{v}) + |a-\widehat{a}|^2) \\ + (1+|v'|^p + |\widetilde{v}'_K|^p)(2\widehat{a}\cdot(v+a-\widetilde{v}) + |\widehat{a}|^2)\mathbb{1}(z \ge L). \end{aligned} \tag{4.257}$$

We now square each term, and use the crude bounds $|a| \le |v| + |v_\star|, |\widehat{a}| \le |\widetilde{v}| + |\widetilde{v}_\star|$ to see that

$$\begin{aligned} (d_p(v', \widetilde{v}'_K)^2 - d_p(v,\widetilde{v}))^2 \le c(|v'|^p + |\widetilde{v}'_K|^p - |v|^p - |\widetilde{v}|^p)^2|v-\widetilde{v}|^4 \\ + c(1+|v'|^p + |\widetilde{v}'_K|^p)^2(|v|^2 + |\widetilde{v}|^2 + |v_\star|^2 + |\widetilde{v}|^2)|a-\widehat{a}|^2 \\ + c(1+|v'|^p + |\widetilde{v}'_K|^p)^2(|v|^2 + |\widetilde{v}|^2 + |v_\star|^2 + |\widetilde{v}|^2)|\widehat{a}|^2\mathbb{1}(z\ge L)). \end{aligned} \tag{4.258}$$

We can now replace every instance of $|v'|^p \le C(|v|^p + |v_\star|^p)$, and similarly for $\widetilde{v}'_K$, and drop the factor $\mathbb{1}(z \ge L)$ in the final term. In this way, we obtain

$$\mathcal{Q}^1_K \le C\,(\mathcal{T}_5 + \mathcal{T}_6 + \mathcal{T}_7)\,; \tag{4.259}$$

where the three terms are

$$\mathcal{T}_5 := \int_0^\infty dz \int_{\mathbb{S}^{d-2}} d\varphi \, ||v'|^p - |v|^p + |\widetilde{v}'_K|^p - |\widetilde{v}|^p| \cdots \times (1 + |v|^{p+4} + |v_\star|^{p+4} + |\widetilde{v}|^{p+4} + |\widetilde{v}_\star|^{p+4}); \tag{4.260}$$

$$\mathcal{T}_6 := \int_0^\infty dz \int_{\mathbb{S}^{d-2}} d\varphi (1 + |v|^{2p+2} + |v_\star|^{2p+2} + |\widetilde{v}^{p+2}| + |\widetilde{v}_\star|^{2p+2})|a - \widehat{a}|^2; \tag{4.261}$$

$$\mathcal{T}_7 := \int_0^\infty dz \int_{\mathbb{S}^{d-2}} d\varphi (1 + |v|^{2p+2} + |v_\star|^{2p+2} + |\widetilde{v}|^{p+2} + |\widetilde{v}_\star|^{2p+2})|\widehat{a}|^2. \tag{4.262}$$

Let us now analyse these integrals one by one. The analysis of $\mathcal{T}_5$ is similar to that of $\mathcal{T}_1$, although with an absolute value, and the integrals appearing in $\mathcal{T}_6, \mathcal{T}_7$ can be reduced to the calculations for $\mathcal{T}_3, \mathcal{T}_4$ in the previous proof.

1. Analysis of $\mathcal{T}_5$. We start from the observation that, for all $v, w \in \mathbb{R}^d$, we have

$$||v|^p - |w|^p| \le C(1 + |v|^{p-1} + |w|^{p-1})|v - w|. \tag{4.263}$$

It follows that

$$\begin{aligned} ||v'|^p - |v|^p| &\le C(1 + |v|^{p-1} + |v + a|^{p-1})|a| \\ &\le C(1 + |v|^{p-1}| + |v_\star|^{p-1})(|v| + |v_\star|) G\left(\frac{z}{x^\gamma}\right). \end{aligned} \tag{4.264}$$

Integrating, we find that

$$\begin{aligned} \int_0^\infty dz \int_{\mathbb{S}^{d-2}} d\varphi \, ||v'|^p - |v|^p| &\le C(1 + |v|^p + |v_\star|^p) \int_0^\infty G(z/x^\gamma) dz \\ &\le C(1 + |v|^{p+\gamma} + |v_\star|^{p+\gamma}). \end{aligned} \tag{4.265}$$

A similar argument applies for $||\widetilde{v}'_K|^p - |\widetilde{v}|^p|$. Including the moment prefactors, we obtain

$$\mathcal{T}_5 \le C(1 + |v|^{2p+4+\gamma} + |v_\star|^{2p+4+\gamma} + |\widetilde{v}|^{2p+4+\gamma} + |\widetilde{v}_\star|^{2p+4+\gamma}). \tag{4.266}$$

2. Analysis of $\mathcal{T}_6$. For $\mathcal{T}_6$, we note that the moment prefactor is constant over the integral, and that we already analysed $\int_0^\infty dz \int_{\mathbb{S}^{d-2}} |a - \widehat{a}|^2$ when analysing $\mathcal{T}_3$ in the previous proof. Absorbing the terms $|v - \widetilde{v}|^2$ and $|v_\star - \widetilde{v}_\star|^2$, the same calculations as above therefore give

$$\mathcal{Q}_2 \le C(1 + |v|^{2p+4+\gamma} + |v_\star|^{2p+4+\gamma} + |\widetilde{v}|^{2p+4+\gamma} + |\widetilde{v}_\star|^{2p+4+\gamma}). \tag{4.267}$$

3. Analysis of $\mathcal{T}_7$. As above, the moment prefactor is independent of the integration variables z, φ, and the problem reduces to estimating $\int_L^\infty \int_{\mathbb{S}^{d-2}} |\widehat{a}|^2$, which is analagous to $\mathcal{T}_4$. We recall that

$$|\widehat{a}|^2 = \frac{1}{2}|\widetilde{v} - \widetilde{v}_\star|^2 \left(1 - \cos G\left(\frac{z}{\widetilde{x}}\right)\right) \le \frac{1}{4}|v - \widetilde{v}|^2 G\left(\frac{z}{\widetilde{x}^\gamma}\right)^2. \tag{4.268}$$

Therefore,

$$\int_0^\infty dz \int_{\mathbb{S}^{d-2}} d\varphi |\widehat{a}|^2 \le C|v - \widetilde{v}|^2 |v - \widetilde{v}|^\gamma \int_0^\infty G(z)^2 dz. \tag{4.269}$$

The final integral is finite, thanks to the estimates established in Subsection 4.10.1, so we conclude

$$\mathcal{T}_7 \le C(1 + |v|^{2p+4+\gamma} + |v_\star|^{2p+4+\gamma} + |\widetilde{v}|^{2p+4+\gamma} + |\widetilde{v}_\star|^{2p+4+\gamma}). \tag{4.270}$$

Combining (4.266, 4.267, 4.270) gives the claimed result. □

Appendix

4.A Proof of Proposition 4.7

We now address the proof of Proposition 4.7, which was deferred earlier. The first item is elementary, and relies on a consistency between the unlabelled and labelled generators $\mathcal{G}, \mathcal{G}^{\mathrm{L}}$; for the second item, we carefully state a result of Kurtz [126, 127] and show how it applies in our case.

Let us recall some notation which will be needed. We will frequently move between objects defined on the labelled Kac sphere

$$\mathbb{S}_N = \left\{ \mathcal{V}^N = (V^1, ...V^N) \in (\mathbb{R}^d)^N, \ \sum_{i=1}^N V^i = 0, \ \sum_{i=1}^N |V^i|^2 = N \right\} \tag{4.271}$$

and the unlabelled state space $\mathcal{S}_N$; we recall that $\theta_N : \mathbb{S}_N \to \mathcal{S}_N$ is the map

$$\mathcal{V}^N = (V^1, ..., V^N) \mapsto \frac{1}{N} \sum_{i=1}^N \delta_{V^i}. \tag{4.272}$$

For clarity, we will indicate functions on $\mathbb{S}_N$ with a $\widehat{\cdot}$ to distinguish them from those on $\mathcal{S}_N$. We will equip $\mathcal{S}^N$ with the distance

$$|\mathcal{V}^N - \mathcal{W}^N| := \sum_{i=1}^N |V^i - W^i| \tag{4.273}$$

where the right-hand side is the Euclidean norm on $\mathbb{R}^d$. We will write $W^{1,\infty}(\mathbb{S}_N)$ for the Sobolev space of functions $\widehat{F} : \mathbb{S}_N \to \mathbb{R}$ which are Lipschitz with respect to this distance, equipped with the norm

$$\|\widehat{F}\|_{W^{1,\infty}(\mathbb{S}_N)} := \max \left(\sup_{\mathcal{V}^N} |\widehat{F}(\mathcal{V}^N)|, \sup_{\mathcal{V}^N \neq \mathcal{W}^N} \frac{|\widehat{F}(\mathcal{V}^N) - \widehat{F}(\mathcal{V}^N)|}{|\mathcal{V}^N - \mathcal{W}^N|} \right) \tag{4.274}$$

and define $W^{1,\infty}(\mathcal{S}_N)$ similarly, equipping $\mathcal{S}_N$ with the Wasserstein$_1$ distance $\mathcal{W}_1$. It is elementary to show that these spaces are separable. Let us also recall, for convenience, the generators of the labelled and unlabelled dynamics, given respectively by

$$(\mathcal{G}^N F)(\mu^N) = N \int_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{S}^{d-1}} (F(\mu^{N,v,v_\star,\sigma}) - F(\mu^N)) B(v - v_\star, \sigma) \mu^N(dv) \mu^N(dv_\star) d\sigma; \tag{4.275}$$

$$(\mathcal{G}^{\mathrm{L}}\widehat{F})(\mathcal{V}^N) = \frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N\int_{\mathbb{S}^{d-1}}\left(\widehat{F}(\mathcal{V}^N_{i,j,\sigma}) - \widehat{F}(\mathcal{V}^N)\right)d\sigma \tag{4.276}$$

for Lipschitz functions $F \in W^{1,\infty}(\mathcal{S}_N), \widehat{F} \in W^{1,\infty}(\mathbb{S}_N)$ respectively. With this notation fixed, we turn to the proof of the two propositions.

Proof of Proposition 4.7. For item i)., observe the following consistency between the unlabelled generator (4.275) and labelled generator (4.276), which follows from the $\mathrm{Sym}(N)$ symmetry of the labelled dynamics: if $F \in W^{1,\infty}(\mathcal{S}_N)$, then $\widehat{F} := F \circ \theta_N \in W^{1,\infty}(\mathbb{S}_N)$, and

$$\mathcal{G}^L\left(F\circ\theta_N\right) = (\mathcal{G}F)\circ\theta_N. \tag{4.277}$$

Now, let $\mathcal{V}^N_t$ be a labelled Kac process, for some filtration $(\mathfrak{F}_t)_{t\ge0}$; it follows that $\mathcal{V}^N_t$ solves the martingale problem for (4.276) for the same filtration. Now, let $\mu^N_t = \theta_N(\mathcal{V}^N_t)$ be the associated empirical measures, and fix $F \in W^{1,\infty}(\mathcal{S}_N)$. For $\widehat{F} = F\circ\theta_N$ as above, the consistency (4.277) gives

$$F(\mu^N_t) - F(\mu^N_0) - \int_0^t (\mathcal{G}F)(\mu^N_s)ds = \widehat{F}(\mathcal{V}^N_t) - \widehat{F}(\mathcal{V}^N_0) - \int_0^t(\mathcal{G}^{\mathrm{L}}\widehat{F})(\mathcal{V}^N_s)ds. \tag{4.278}$$

The right-hand side is a martingale by assumption, and hence μ^N_t solves the martingale problem for (4.275) in the filtration $(\mathfrak{F}_t)_{t\ge0}$, as desired; in particular, μ^N_t is a Markov process with generator (4.275).

For item ii), we will use the following result, which generalises the implication needed, due to Kurtz [126, 127]. Let us first fix some terminology. For a topological space E, let us write $\overline{C}(E)$ for the space of bounded, continuous functions on E, $B(E)$ for the space of bounded, Borel-measurable functions on E, and $\mathcal{P}(E)$ for the space of Borel probability measures. Given another such space E_0, a transition function α from E_0 to E is a mapping from $E_0 \to \mathcal{P}(E)$ such that, for all Borel sets $A \subset E$, the map $y \mapsto \alpha(y, A)$ is a Borel function on E_0; for such α and $f \in B(E)$, define $\alpha f \in B(E_0)$ by

$$(\alpha f)(y) := \int_E f(z)\alpha(y, dz). \tag{4.279}$$

We will write $M_E[0,\infty), D_E[0,\infty)$ for the measurable, respectively càdàg functions from $[0,\infty)$ to E.

Let us say that a linear operator $\mathcal{A} \subset B(E)\times B(E)$ is seperable if there exists a countable subset $\{f_\beta, \beta \ge 1\} \subset \mathcal{D}(\mathcal{A})$ such that, for all $(f,g) \in \mathcal{A}$, there exists a subsequence $\beta_i \to \infty$ such that $(f_i, \mathcal{A}f_i)$ are bounded uniformly in i, and converge pointwise to (f,g). We say that a linear operator $\mathcal{A}$ is a pregenerator if it is dissipative, and there exists a sequence of functions $q_n : E \to \mathcal{P}(E), r_n : E \to [0,\infty)$ such that, for all $f \in \mathcal{D}(\mathcal{A})$, we have the pointwise convergence

$$r_n(x)\int_E (f(y) - f(x))q_n(x, dy) \to (\mathcal{A}f)(x) \qquad \text{for all } x \in E. \tag{4.280}$$

With these definitions, we can state the following result, which appears as part of [127, Theorem 1.4]

Proposition 4.33. *Let $(E,r),(E_0,r_0)$ be complete, separable metric spaces. Let $\mathcal{A}\subset \overline{C}(E)\times\overline{C}(E)$ be a linear operator which is seperable and a pre-generator, and whose domain $\mathcal{D}(\mathcal{A})$ separates points in E. Suppose that $\theta: E\to E_0$ is Borel measureable, and α is a transition function from E_0 to E satisfying the compatibility condition $\alpha(y,\theta^{-1}(y))=1$ for all $y\in E_0$. Let $\mathcal{A}^\theta$ be the linear operator*

$$\mathcal{A}^\theta=\{(\alpha f,\alpha(\mathcal{A}f)): f\in\mathcal{D}(\mathcal{A})\}\subset B(E_0)\times B(E_0). \tag{4.281}$$

Let $\mathcal{L}_0\in\mathcal{P}(E_0)$, and let $\widetilde{\mathcal{L}}_0=\alpha_\#\mathcal{L}_0\in\mathcal{P}(E)$ be given by

$$\widetilde{\mathcal{L}}_0(A)=\int_{E_0}\alpha(y,A)\mathcal{L}_0(dy). \tag{4.282}$$

If $\widetilde{\mu}=(\widetilde{\mu}_t)_{t\ge0}$ is a solution of the martingale problem for $(\mathcal{A}^\theta,\mathcal{L}_0)$, then there exists a solution $\mathcal{V}$ of the martingale problem for $(\mathcal{A},\widetilde{\mathcal{L}}_0)$ such that $\widetilde{\mu}$ has the same law on $M_{E_0}[0,\infty)$ as $\mu=\theta\circ\mathcal{V}$. Further, if $\widetilde{\mu}$, and hence μ, has a modification with sample paths in $D_{E_0}[0,\infty)$, then the modified $\widetilde{\mu},\mu$ have the same law on $D_{E_0}[0,\infty)$.

Let us now show how this applies in our case. We will take E,E_0 to be the labelled and unlabelled Kac spheres $E=\mathbb{S}_N, E_0=\mathcal{S}_N$ respectively, equipped with the metrics as above. We take $\mathcal{A}$ to be the labelled generator $\mathcal{G}^L$ given by (4.9), defined on $F\in W^{1,\infty}(\mathbb{S}_N)$, and let $\theta=\theta_N$ be given by (4.272). We define α as the average over the preimage

$$\alpha(\mu^N)=\frac{1}{\#\theta_N^{-1}(\mu^N)}\sum_{\mathcal{V}^N\in\theta_N^{-1}(\mu^N)}\delta_{\mathcal{V}^N}. \tag{4.283}$$

We remark that, if $\mu^N\in\mathcal{S}_N$ and $\mathcal{V}^N\in\theta_N^{-1}(\mu^N)$, then $\alpha(\mu^N)$ can be rewritten

$$\alpha(\mu^N)=\frac{1}{N!}\sum_{\pi\in\mathrm{Sym}(N)}\delta_{\mathcal{V}^{N,\pi}} \tag{4.284}$$

where $\mathcal{V}^{N,\pi}$ denotes the action of $\pi\in\mathrm{Sym}(N)$ permuting the N components $V^1,..,V^N\in\mathbb{R}^d$ of $\mathcal{V}^N$. It is elementary, if somewhat tedious, to check that with these choices, the linear operator $\mathcal{A}^\theta$ is exactly the unlabelled generator $\mathcal{G}$, defined on $W^{1,\infty}(\mathcal{S}_N)$; the inclusion $\mathcal{G}\subset\mathcal{A}^\theta$ is exactly the statement (4.277), and for the other inclusion $\mathcal{A}^\theta\subset\mathcal{G}$, we use (4.284) to check that, for $\widehat{F}:\mathbb{S}_N\to\mathbb{R}$ Lipschitz, $\alpha\widehat{F}:\mathcal{S}_N\to\mathbb{R}$ is Lipschitz, and straightforward calculations show that $\mathcal{G}(\alpha\widehat{F})=\alpha(\mathcal{G}^L\widehat{F})$ as desired.

To see that $\mathcal{A}=\mathcal{G}^L$ is separable, we note that $W^{1,\infty}(\mathbb{S}_N)$ is separable, and $\mathcal{G}^{\mathrm{L}}: W^{1,\infty}(\mathbb{S}_N)\to L^\infty(\mathbb{S}_N)$ is a bounded linear map. Its graph is therefore separable in the stronger topology induced by $\mathcal{G}^{\mathrm{L}}\subset W^{1,\infty}(\mathbb{S}_N)\times L^\infty(\mathbb{S}_N)$, and so is separable in the topology of bounded pointwise convergence in the definition above.

To see that $\mathcal{G}^{\mathrm{L}}$ is a pregenerator, let us define $\mathcal{G}^{\mathrm{L}}_K$ to be the cutoff equivalent, replacing B by the cutoff kernel B_K given by (CHP_K). It is straightforward to write $\mathcal{G}^{\mathrm{L}}_K$ in the form desired, and $\mathcal{G}^{\mathrm{L}}_K \to \mathcal{G}_K$ in the space of bounded linear maps $\mathcal{B}(W^{1,\infty}(\mathbb{S}_N), L^\infty(\mathbb{S}_N))$. Elementarily, each $\mathcal{G}^{\mathrm{L}}_K$ is the generator of a cutoff, labelled Kac process, and so generates a semigroup of contraction mappings; by the Lumer-Phillips Theorem, they are therefore dissipative; we can then take a limit to conclude that $\mathcal{G}^{\mathrm{L}}$ is dissipative, and so is a a pregenerator.

We can now apply the conclusion of Proposition 4.33 above. Let us fix $\mu^N_0 \in \mathcal{S}_N$, and let $(\widetilde{\mu}^N_t)_{t\ge 0}$ be a solution to the martingale problem for the unlabelled generator (1.31) starting at μ^N_0. The law $\widetilde{\mathcal{L}}_0$ given by Proposition 4.33 exactly corresponds to picking $\mathcal{V}^N_0 \in \theta_N^{-1}(\mu^N_0)$ uniformly at random, as in the statement of the proposition, and by the result quoted above, there exists a solution to the martingale problem for (4.9), starting at $\mathcal{V}^N_0$ such that $\widetilde{\mu}^N_t$ has the same law as $\theta_N(\mathcal{V}^N_t)$. $\mathcal{V}^N_t$ is therefore a weak solution to the stochastic differential equation (LK), and so we have proven the claim of item ii). □

4.B Proof of Lemma 4.9

We now sketch the proof of Lemma 4.9, which were deferred from earlier. We will demonstrate a single construction which provides both ι_α, R_α with the desired properties.

Sketch Proof. Let us view $(0,1)$ as an auxiliary probability space, equipped with the Borel σ-algebra $\mathcal{B}(0,1)$ and Lebesgue measure $d\alpha$. Let us write λ for the Haar measure on the group $\mathrm{Isom}(\mathbb{S}^{d-2})$, and construct random variables $Z_1(\alpha), ...Z_d(\alpha)$, independently distributed uniformly over the sphere $\mathbb{S}^{d-1}$, $\zeta(\alpha)$ distributed uniformly on $(0,\pi/3)$ and $S_n(\alpha)$ samples from λ, all independently of each other.

With these variables given, we construct the maps as follows. If $\{Z_1(\alpha), Z_2(\alpha), ..., Z_d(\alpha)\}$ is a basis of $\mathbb{R}^d$, then let $\chi_i(\alpha) : \mathbb{R}^d \to \mathbb{R}$ be the dual basis, so any X can be written as $X = \sum_{i=1}^d Z_i(\alpha)\chi_i(\alpha, X)$. Now, if $\chi_1(\alpha, X) > 0$, we can find an orthogonal basis by applying Gram-Schmidt to $\{X, Z_2(\alpha), ..., Z_d(\alpha)\}$, and produce $\iota_\alpha(X)$ by scaling so that elements have norm $|X|$. If $\chi_1(\alpha, X) < 0$, we define $\iota_\alpha(-X)$ by the previous construction and define $\iota_\alpha(X) := -\iota_\alpha(-X)$. We construct ι_α on the remaining hyperplane $\{X : \chi_1(\alpha, X) = 0\}$ by setting $\iota_\alpha(0) = 0$ and otherwise, picking the first coordinate j with $\chi_j(\alpha, X) \neq 0$; if $\chi_j(\alpha, X) > 0$, apply Gram-Schmidt to $\{X, Z_2(\alpha), ..., Z_{j-1}(\alpha), Z_{j+1}(\alpha), ...Z_d(\alpha)\}$, and if $\chi_j(\alpha, X) < 0$, define $\iota_\alpha(X) := -\iota_\alpha(-X)$, which is defined by the previous case. Finally, if $\{Z_i(\alpha)\}$ do not form a basis, repeat the same construction with the canonical basis $\{e_1, ..., e_d\}$ of $\mathbb{R}^d$.

Let us check the claimed continuity property. Any ι_α with the desired properties is immediately continuous at 0, thanks to the normalisation $|\iota_{\alpha,j}(X)| = |X|$. For any $X \neq 0$, it holds $d\alpha$-almost surely that $\{Z_i(\alpha)\}$ form a basis and that $X \notin \text{Span}(Z_2(\alpha), ...Z_{d-1}(\alpha))$, so $\chi_1(\alpha, X) \neq 0$. The Gram-Schmidt construction on

$$\{\pm X, Z_2(\alpha), ..., Z_d(\alpha)\}$$

is continuous in X by explicit construction away from $\{\chi_1(\alpha, X) = 0\}$, so we conclude that

$$\begin{aligned}\{\alpha :\ \iota_\alpha \text{discontinuous at } X\} \subset &\left\{\alpha : \{Z_i(\alpha)\} \text{ not a basis}\right\} \\ &\cup \left\{\alpha : X \in \text{Span}(Z_2(\alpha), ..., Z_d(\alpha))\right\}\end{aligned} \tag{4.285}$$

which has $d\alpha$-measure 0. The same immediately applies to $\Gamma_\alpha, a(\cdot, \alpha)$.

For $R_\alpha(X, Y)$, we construct R separately depending on whether $X \cdot Y$ is greater than or less than $-|X||Y| \cos\zeta(\alpha)$. If $X \cdot Y = |X||Y|$, so X, Y are colinear, then the unique possible choice of R is to be the identity on $\mathbb{S}^{d-2}$. Otherwise, if $X \cdot Y \geq -|X||Y| \cos\zeta$, we obtain $j^1_X(\alpha)$ by applying Gram-Schmidt orthogonalisation to $\{X, Y\}$, and similarly for Y. We now find $u_2(\alpha), ...u_{d-1}(\alpha)$ by applying Gram-Schmidt to $\{X, j^1_X, Z_2(\alpha), ..., Z_d(\alpha)\}$ whenever this is a spanning set for $\mathbb{R}^d$, and arbitrarily otherwise. We now let $R_\alpha(X, Y)$ be the isometry constructed in the proof Lemma 4.8, and observe that, for any given X, Y which are not colinear,

$$\begin{aligned}\{\alpha : R_\alpha \text{ discontinuous at } (X, Y)\} \cap \{\alpha : X \cdot Y \geq -|X||Y| \cos\zeta(\alpha)\} \\ \subset \{\alpha : \{X, Y, Z_3(\alpha), ..., Z_d(\alpha)\} \text{ not spanning}\}\end{aligned} \tag{4.286}$$

which has $d\alpha$-measure 0. In the case where X, Y are colinear and have the same sense, if α is such that ι is continuous at X, Y and $(X^n, Y^n) \to (X, Y)$ and if ι is continuous at both X, Y, then for any subsequence we can find a further subsequence with $R_\alpha(X^n, Y^n) \to r$, for some $r \in \text{Isom}(\mathbb{S}^{d-2})$ by compactness, then we can take limits to check that this r also satisfies the requirement(4.11), and so is the identity $r = \text{Id}_{\mathbb{S}^{d-2}} = R_\alpha(X, Y)$, and we conclude that R_α is continuous at colinear (X, Y) with the same sense.

We finally deal with the case $X \cdot Y < -|X||Y| \cos\zeta(\alpha)$ For any X, Y, α, let us write $\mathcal{R}_\alpha(X, Y)$ for the set of all isometries $R \in \text{Isom}(\mathbb{S}^{d-2})$ satisfying the conclusion (4.11) for the maps Γ_α; thanks to Lemma 4.8, this is always a nonempty set. Further, if $X \cdot Y < 0$, then we can examine the proof of Lemma 4.8 to see that $\mathcal{R}_\alpha(X, Y)$ contains an open neighbourhood of the isometry constructed in the cited proof, so that $\mathcal{R}_\alpha(X, Y)$ has nontrivial interior and hence $\lambda(\mathcal{R}_\alpha(X, Y)) > 0$, while the boundary $\partial\mathcal{R}_\alpha(X, Y)$ is a union of codimension 1 submanifolds of $\text{Isom}(\mathbb{S}^{d-2})$, and in particular has λ-measure 0. Now, in the case where $X \cdot Y < -|X||Y| \cos\zeta(\alpha) < 0$, we set $N(\alpha, X, Y) = \min(n : S_n(\alpha) \in \mathcal{R}_\alpha(X, Y)\} \in \mathbb{N} \cup \{\infty\}$, and $R_\alpha(X, Y) = S_{N(\alpha,X,Y)}(\alpha)$ whenever $N(\alpha, X, Y)$ is finite, and construct $R_\alpha(X, Y)$ with the desired properties arbitrarily otherwise. In this way, for any fixed X, Y, let α be such that $X \cdot Y < -|X||Y| \cos\zeta(\alpha)$ and such that ι_α is continuous at X, Y. The set of such α where $N(\alpha, X, Y)$ is infinite or $S_{N(\alpha,X,Y)} \in \partial\mathcal{R}_\alpha(X, Y)$ has

$d\alpha$-measure 0, and for all other α, some thought shows R_α is locally constant at X, Y and in particular continuous. Summing up, we have shown that

$$\begin{aligned}\{\alpha : R_\alpha \text{ discontinuous at } (X,Y)\} \cap \{\alpha : X \cdot Y < -|X||Y|\cos\zeta(\alpha)\} \\ \subset \{\alpha : (\alpha, X, Y) = \infty \text{ or } S_{N(\alpha,X,Y)} \in \partial\mathcal{R}_\alpha(X,Y)\} \\ \cup \{\alpha : \iota_\alpha \text{ discontinuous at } X \text{ or } Y\}\end{aligned} \tag{4.287}$$

which has $d\alpha$-measure 0. Finally, for given X, Y, the edge case $\{\alpha : X \cdot Y = -|X||Y|\cos\zeta(\alpha)\}$ has $d\alpha$-measure 0, because ζ has a density, and we are done. □

Chapter 5

The Hard Potential Landau Equation

5.1 Introduction & Main Results

In this chapter, we study the spatially homogeneous Landau equation (LE) in dimension $d = 3$, with hard potentials ($0 < \gamma \leq 1$), proving all the assertions of Theorem 3. This chapter is based on the work [90], jointly with Prof. Nicolas Fournier.

In this context, we continue with the ideas of Chapter 4. The central result is a new stability and well-posedness result (Theorem 3ii-iii).), analagous to Theorem 4.1. In case of the Landau equation, we can significantly refine the estimates in Sections 4.3,4.10 to show that *any* $p > 2$ suffices for the well-posedness of the Landau equation, whereas previous results have required either exponential moments [88] or additional regularity [58], and where we needed a (large, but finite) p in Chapter 4. As a result of the new uniqueness theorem, we extend previous regularity results, to show that they apply to *all* weak solutions to (LE), aside from the degenerate case of point masses, whereas the corresponding result of Desvillettes and Villani [58] only shows that such solutions exist, and later results require strong *a priori* regularity conditions. Finally, we give an existence result for $\mu \in \mathcal{P}_2(\mathbb{R}^3)$, which is the most general possible for the definition of weak solutions.

5.1.1 Notation

Let us briefly introduce some notation which is specific to this chapter. First, we introduce some notation regarding the regularity theory of solutions. For $k, s \geq 0$, we define the weighted Sobolev norm

$$\|f\|^2_{H^k_s(\mathbb{R}^3)} = \sum_{|\alpha|\leq k} \int_{\mathbb{R}^3} |\partial_\alpha f(v)|^2 (1+|v|^2)^{s/2} dv$$

and the weighted Sobolev space $H^k_s(\mathbb{R}^3)$ for those f where this is finite. By an abuse of notation, we say that $\mu \in \mathcal{P}(\mathbb{R}^3)$ belongs to $H^k_s(\mathbb{R}^3)$ if μ admits a density f with respect to the Lebesgue measure with $f \in H^k_s(\mathbb{R}^3)$, and in this case we write $\|\mu\|_{H^k_s(\mathbb{R}^3)} = \|f\|_{H^k_s(\mathbb{R}^3)}$. Similarly, we say that $\mu \in \mathcal{P}(\mathbb{R}^3)$ is analytic if it admits an analytic density.

Let us also mention some other points of notation which we will use in this chapter. Since we are (exclusively) interested in the Landau equation, we write $\mathcal{L} = \mathcal{L}_{\mathrm{L}}$ and omit the subscript. Similarly, since we only work with $d = 3$, we write $\mathcal{S} = \mathcal{S}(\mathbb{R}^3)$ everywhere.

5.1.2 Main Results

We now give precise formulations of our results, corresponding to Theorem 3. Throughout this chapter, we will work with the transport costs $w_{p,\varepsilon}$ for the functions $d_{p,\varepsilon}$ given in Section 2.1. Our first result is the following stability result, which corresponds to Theorem 4.1.

Theorem 5.1. *Fix $\gamma \in (0,1]$ and $p > 2$ and two weak solutions $(\mu_t)_{t\ge 0}, (\nu_t)_{t\ge 0} \subset \mathcal{S}$ to (LE) starting from μ_0 and ν_0, both belonging to $\mathcal{S}^p$. Then there exists a stochastic process $(V_t, \tilde{V}_t)_{t\ge 0}$, such that $\pi_t = \mathrm{Law}(V_t, \tilde{V}_t) \in \Pi(\mu_t, \nu_t)$, and that $V_t, \tilde{V}_t$ solve (stLE) for μ_t, ν_t respectively, for different white noises. Furthermore, for some constant C, depending only on p and γ, the coupling satisfies for all $t \ge 0$,*

$$\mathbb{E}[d_{p,1}(V_t, \tilde{V}_t)] \le w_p(\mu_0, \nu_0) \exp\Big(C(1+t)\Lambda_p(\mu_0, \nu_0)^2\Big). \tag{5.1}$$

Since $\pi_t = \mathrm{Law}(V_t, \tilde{V}_t)$ is a coupling of μ_t, ν_t, the right-hand side is an upper bound for $w_{p,1}(\mu_t, \nu_t)$.

As a byproduct of our analysis, we also obtain the following, which extends the equivalence between (LE) to (stLE) to include all solutions.

Theorem 5.2. *If $(V_t)_{t\ge 0}$ is any solution to (stLE) with $\mu_t = \mathrm{Law}(V_t) \in L^1_{loc}([0,\infty), \mathcal{S}^{2+\gamma})$, then $(\mu_t)_{t\ge 0}$ is a weak solution to (LE). Conversely, if $(\mu_t)_{t\ge 0} \subset \mathcal{S}$ is a weak solution to (LE), then there exists a solution $(V_t)_{t\ge 0}$ to (stLE) with $\mathrm{Law}(V_t) = \mu_t$ for all t.*

Concerning existence, the following will prove Theorem 3iv).

Theorem 5.3. *Let $\gamma \in (0,1]$ and $\mu_0 \in \mathcal{S}$. Then there exists a stochastic process $(V_t)_{t\ge 0}$ solving (stLE) with $\mathrm{Law}(V_0) = \mu_0$ and such that the laws $\mu_t = \mathrm{Law}(V_t)$ satisfy $\mu \in L^1_{loc}([0,\infty), \mathcal{S}^{2+\gamma})$. In particular, $(\mu_t)_{t\ge 0} \subset \mathcal{S}$ is a weak solution to (LE) starting at μ_0.*

Taken together, Theorems 5.1, 5.2, 5.3 show that all weak solutions arise via the SDE (stLE), that solutions to (stLE) exist as soon as $\mu_0 \in \mathcal{S}$, and provide a quantitative stability estimate as soon as the initial data belong to $\mathcal{S}^{2+} = \cup_{p>2}\mathcal{S}^p$.

Finally, we consider the regularity of solutions. The following regularity proves the assertion of Theorem 3iii), and arises as a result of our uniqueness theorem, combined with previous regularity results from the literature.

Theorem 5.4. *Fix $\gamma \in (0,1]$. Let $(\mu_t)_{t\geq 0} \subset \mathcal{S}$ be any weak solution to (LE). Then we have*

$$\text{for all } k, s \geq 0 \text{ and all } t_0 > 0, \quad \sup_{t\geq t_0} \|\mu_t\|_{H^k_s(\mathbb{R}^3)} < \infty. \tag{5.2}$$

and for all $t > 0$, μ_t is analytic and has a finite entropy.

Let us remark that we do not need to exclude the case of point masses as in the corresponding theorem in [90], since we already excluded such cases by normalising to $\mathcal{S}$. In any case, if $\mu_0 = \delta_{v_0}$ is a point mass, then conservation of energy and momentum ensures that $\mu_t = \delta_{v_0}$ for all $t > 0$, and there is no hope of regularity.

5.1.3 Strategy

We first work towards the stability and uniqueness result Theorem 5.1, which follows a pattern similar to the second proof given in the Boltzmann case in Chapter 4. The principle remains the same, of using negative 'Povzner terms' to counteract the higher-order terms which prevent a Grönwall estimate (see also the sketch proof of Proposition 4.18). In this case, we will find explicit, rather than explicit*able* constants; indeed, choose $p = 2 + \epsilon$ will suffice to cancel the difficult terms. To allow for initial data with only $\mu_0 \in \mathcal{S}^{2+}$, we will use the transport costs $w_{p,1}$ introduced in Section 2.1, based on the cost $(1 + |v|^p + |\tilde{v}|^p)|v - \tilde{v}|^2/(1 + |v - \tilde{v}|^2)$, rather than $w_p = w_{p,0}$ as we did in the previous chapter, since $w_{p,1}$ is defined on $\mathcal{P}_p(\mathbb{R}^3)$ rather than $\mathcal{P}_{p+2}(\mathbb{R}^3)$.

5.1.4 Plan of the Chapter

The chapter is structured as follows.

i). In Section 5.2, we will present some preliminary calculations on the coefficients a, b of (1.8) which are used throughout the chapter.

ii). Section 5.3 is a self-contained result regarding a tightness property for solutions to (stLE), analagous to Section 4.4, which will be used in the same way to construct solutions via extraction of a subsequence converging in distribution.

iii). Section 5.4 introduces the Tanaka-style coupling, equivalent to Section 4.3, and presents the key estimate without proof.

iv). Section 5.5 gives the proof of the stability and uniqueness result Theorem 5.1. We first prove the result in Lemma 5.10 under Gaussian moment conditions, which are then carefully relaxed.

v). Section 5.6 gives the proof of Theorem 5.2, using tools which have already been developed in constructing the coupling in Theorem 5.1 and in Section 5.3, following the same argument as in the Boltzmann case in 4.7.

vi). Section 5.7 consists of a self-contained proof of our existence result Theorem 5.3, building only on an existence result of Desvilettes and Villani, and using the de La Vallée Poussin theorem and the compactness property in Section 5.3.

vii). In Section 5.8, we prove Theorem 5.4 about smoothness. We show a very mild regularity result (Lemma 5.15): solutions do not remain concentrated on lines. This allows us to apply results from the literature on the *existence* of regular solutions, exploiting the uniqueness from Theorem 5.1.

viii). Finally, Section 5.9 contains the proof of the estimate Lemma 5.9.

5.1.5 Literature Review & Discussion

Let us discuss our results in the context of the literature on the Landau equation and the other work in this thesis.

1. Well-Posedness of the Landau Equation We emphasise that the key result of this chapter is the stability and uniqueness result Theorem 5.1, which continues the study of stability of the hard potential Landau equation by Arsenv and Buryak [13], Desvillettes and Villani [58], and by Fournier and Guillin [88]. As in Chapter 3 and the works [142, 143], stability estimates can be used to prove propagation of chaos for the many-particle system, and the estimate we establish here could be a first step in this direction. In this context, it is particularly advantageous that our result requires neither regularity nor exponential moments, as these are not readily applicable to the empirical measures of the particle system. However, let us remark that the result we obtain here on its own is not (yet) sufficient to use arguments analagous to Chapter 3 or the abstract result of [143]; these would need a 'second order' stability result, comparable to Proposition 3.15, whereas our current stability result is only first order.

As in the Boltzmann case, the Tanaka-Povzner argument we use here leads to Theorem 5.1 being stronger and more general than those found in the literature. The uniqueness result of Desvillettes and Villani [58] requires that the initial data μ_0 has a density f_0 satisfying

$$\int_{\mathbb{R}^3} (1+|v|^2)^{p/2} f_0^2(v) dv < \infty \quad \text{for some } p > 15 + 5\gamma, \tag{5.3}$$

while the result of the [88] recalled in Proposition 5.8 below allows measure solutions, but requires a finite exponential moment. Our result therefore allows much less localisation than either of the results above, while also not requiring any regularity on the initial data μ_0, ν_0.

Let us also remark that the condition $\mu_0 \in \mathcal{S}^{2+\varepsilon}, \varepsilon > 0$ in Theorem 5.1 appears to be almost optimal for $\gamma > 0$. As in the Boltzmann case, the Maxwell molecule case $\gamma = 0$ of the Landau equation is particularly tractable, and results of Villani [187] show that existence and uniqueness hold, assuming only that $\mu_0 \in \mathcal{P}_2(\mathbb{R}^3)$. Moreover, in Theorem 3.6, we prove a similar result for the Boltzmann equation for hard spheres (HS) or cutoff hard potentials (CHP_K), again requiring $p > 2$ moments. With only $p = 2$ moments, the cutoff Boltzmann equation is known to have unique energy-conserving solutions [144], but no quantitative stability estimate, and the technique fails as cutoff is removed. This limitation is therefore consistent with the state-of-the-art for the (cutoff) Boltzmann equation.

Moreover, since the Landau equation is known to converge exponentially to equilibrium [41], we could interpolate between the short-time result here and the long-time convergence to equilibrium to obtain a uniform-in-time Hölder stability estimate as we did in Theorem 3.6. However, the estimate obtained is fairly weak, as the resulting Hölder exponent depends on the moments of the solutions, and so is much weaker than Theorem 3.6.

2. Coupling Arguments for the Landau Equation As discussed in the introduction, coupling arguments to prove stability go back as far as Tanaka [177] for the Boltzmann equation in the case of Maxwell molecules, and have subsequently been applied to both hard potentials (as in the previous chapter) and soft potentials for the Boltzmann equation. The idea was imported to the Landau equation by Funaki [95] and has previously been applied by Fournier [88] in the context of stability and propagation of chaos, as well as Guérin [105, 106]. It also seems possible that the technique we develop here could be applied to prove propagation of chaos via the coupling method of [88], but we will not explore this here.

3. Existence Regarding existence, Theorem 5.3 extends the existence result of [58], recalled in Proposition 5.12 below, as we only assume $\mu_0 \in \mathcal{S}$, which is clearly necessary for the definition of weak solutions, instead of $\mu_0 \in \mathcal{P}_{2+}(\mathbb{R}^3)$. Theorem 5.2 similarly extends, without moment hypotheses, a similar result of Fournier and Guillin [88] for a stochastic differential equation driven by a Brownian motion, which requires 4 moments on the initial data; as in the Boltzmann case it is satisfying to know that solving (LE) is (unconditionally) equivalent to solving (stLE).

4. Regularity Regarding regularity, Theorem 5.4 shows the smoothness of *any* weak solution, assuming only that $\mu_0 \in \mathcal{S}$, instead of showing the existence of one smooth solution when $\mu_0 \in \mathcal{P}_{2+}(\mathbb{R}^3)$ as in [58] and [45, 46]. Another possible approach to regularity results is the use of Malliavin calculus, see Guérin [105], although this proof is already long in the case of Maxwell molecules, while the proof of Theorem 5.4 is comparatively short, based on Theorem 5.1 and the known results on the existence of regular solutions. Let us finally mention that in the case $\gamma = 0$, a stronger 'ulta-analytic' regularity is known, see Morimoto, Pravda-Starov and Xu [145]; in this case, one has the advantage that the coefficients of (LE) are already analytic (polynomial) functions.

Let us remark that we do not include a similar result in the case of the non-cutoff hard potentials (NCHP). In this case, we do not obtain a regularity result as an application of the uniqueness, as we do here. In the case $d = 3$, Alexandre et al. [8] showed that, if μ_0 admits a density, then the density f_t of μ_t has $\sqrt{f_t} \in H^{\nu/2}_{\text{loc}}(\mathbb{R}^3)$, which Chen and He [47] improved to $(1+|v|^2)\sqrt{f_t} \in H^{\nu/2}(\mathbb{R}^3)$. Fournier [83] showed that the entropy immediately becomes finite, which implies the existence of a density f_t for μ_t for $t > 0$, and that the density belongs to a certain Besov space. For the case of *regularised hard potentials*, where the kinetic factor is replaced by something like $(1 + |v|^2)^{\gamma/2}$, Desvillettes and Wennberg [60] proved that, provided $\mu_0 \in \mathcal{S}$ has finite entropy, then there exists a solution $(\mu_t)_{t\geq 0}$ admitting a density f_t in the Schwarz space for all $t > 0$. If we modified the ideas of Chapter 4 and of Fournier [83] to cover this case, we could use the same argument as in the Landau case to prove that this applies to *all* solutions for this choice of kernel.

5. Other Landau Equations As mentioned in the introduction, the Landau equation makes sense for the full range of the parameters $\gamma \in [-3, 1]$, whereas the results of this chapter apply exclusively to $\gamma \in (0, 1]$. Indeed, for $\gamma > 0$ the coefficients fail to be Lipschitz at infinity, and the technique we use is exactly designed to compensate for this failure. On the other hand, when $\gamma < 0$, the coefficients fail to be Lipschitz close to the diagonal $v = v_\star$, and our current technique produces no compensation. Let us mention the state of the art for the other cases of the Landau equation: the cases $\gamma \in (-3, 0)$ have been studied by Fournier, Guèrin and Hauray [85, 89]. As in the introduction, the case $\gamma = -3$ is the most physically relevant case and corresponds to particles interacting by repulsive Coulomb interactions, where the Boltzmann collision operator no longer makes sense ($\nu = -2$). This case has been studied by Villani [186], Desvillettes [54] and Fournier [82]. We also study only the spatially homogeneous case; let us mention works by by Guo [107], He and Yang [109], Golse, Imbert, Mouhot and Vasseur [99] and Mouhot [148] on the Cauchy problem for the full, spatially inhomogeneous, Landau equation. Finally, the Landau-Fermi-Dirac equation has been recently studied by Alonso, Bagland and Lods [10].

6. Dimension constraint We also remark on the constraint $d = 3$, which is specific to this chapter. While the same principle would hold in general dimensions $d \geq 3$, the estimates on the parameters in Section 5.2 and hence the key calculations in Section 5.9 hold only for the particular choice $d = 3$. In general, the same technique would lead to a requirement that $p > p_0$ for some $p_0(d)$, as in Chapter 4, and so may not coincide with the almost sharp results we obtain here.

5.2 Some Preliminary Calculations

We introduce a few notation and handle some computations of constant use. We denote by $|\cdot|$ the Euclidean norm on $\mathbb{R}^3$ and for A and B two 3×3 matrices, we put $\|A\|^2 = \mathrm{Tr}(AA^*)$ and $\langle\langle A, B\rangle\rangle = \mathrm{Tr}(AB^*)$.

For $x \in \mathbb{R}^3$, we introduce

$$\sigma(x) = [a(x)]^{1/2} = |x|^{1+\gamma/2}\Pi_{x^\perp}.$$

For $x, \tilde{x} \in \mathbb{R}^3$, it holds that

$$||\sigma(x)||^2 = 2|x|^{\gamma+2} \quad \text{and} \quad \langle\langle \sigma(x), \sigma(\tilde{x})\rangle\rangle = |x|^{1+\gamma/2}|\tilde{x}|^{1+\gamma/2}\Big(1+\frac{(x\cdot\tilde{x})^2}{|x|^2|\tilde{x}|^2}\Big) \geq 2|x|^{\gamma/2}|\tilde{x}|^{\gamma/2}(x\cdot\tilde{x}). \tag{5.4}$$

Indeed, it suffices to justify the second assertion, and a simple computation shows that $\Pi_{x^\perp}\Pi_{\tilde{x}^\perp} = \mathbf{I}_3 - |x|^{-2}xx^* - |\tilde{x}|^{-2}\tilde{x}\tilde{x}^* + |x|^{-2}|\tilde{x}|^{-2}(x\cdot\tilde{x})x\tilde{x}^*$, from which we conclude that

$$\langle\langle \sigma(x), \sigma(\tilde{x})\rangle\rangle = |x|^{1+\gamma/2}|\tilde{x}|^{1+\gamma/2}\mathrm{Tr}\,(\Pi_{x^\perp}\Pi_{\tilde{x}^\perp}) = |x|^{1+\gamma/2}|\tilde{x}|^{1+\gamma/2}[1 + |x|^{-2}|\tilde{x}|^{-2}(x\cdot\tilde{x})^2],$$

which is greater than $2|x|^{\gamma/2}|\tilde{x}|^{\gamma/2}(x\cdot\tilde{x})$ because $1 + a^2 \geq 2a$.

We note a useful inequality we will frequently use. For $a, b \geq 0$ and $\alpha \in (0, 1)$, we have that

$$|a^\alpha - b^\alpha| \leq (a \vee b)^{\alpha-1}|a - b|. \tag{5.5}$$

To see this, assume without loss of generality that $a \geq b$, and the claimed inequality follows from $a^\alpha - b^\alpha = a^\alpha[1 - (b/a)^\alpha] \leq a^\alpha(1 - b/a) = a^{\alpha-1}(a - b)$. The case $b \geq a$ follows by symmetry.

For $x, \tilde{x} \in \mathbb{R}^3$, recalling that $b(x) = -2|x|^\gamma x$, we have

$$|b(x) - b(\tilde{x})| \leq 2|x|^\gamma|x - \tilde{x}| + 2|\tilde{x}|||x|^\gamma - |\tilde{x}|^\gamma| \leq 2(|x|^\gamma + |\tilde{x}|^\gamma)|x - \tilde{x}|, \tag{5.6}$$

because $|\tilde{x}|||x|^\gamma - |\tilde{x}|^\gamma| \leq |\tilde{x}|(|x| \vee |\tilde{x}|)^{\gamma-1}|x - \tilde{x}| \leq |\tilde{x}|^\gamma|x - \tilde{x}|$ by (5.5). We also have, thanks to (5.4),

$$||\sigma(x) - \sigma(\tilde{x})||^2 \leq 2|x|^{\gamma+2} + 2|\tilde{x}|^{\gamma+2} - 4|x|^{\gamma/2}|\tilde{x}|^{\gamma/2}(x\cdot\tilde{x}) = 2||x|^{\gamma/2}x - |\tilde{x}|^{\gamma/2}\tilde{x}|^2. \tag{5.7}$$

Proceeding as for (5.6), we deduce that

$$||\sigma(x) - \sigma(\tilde{x})||^2 \leq 2(|x|^{\gamma/2}|x - \tilde{x}| + |\tilde{x}|||x|^{\gamma/2} - |\tilde{x}|^{\gamma/2}|)^2 \leq 2(|x|^{\gamma/2} + |\tilde{x}|^{\gamma/2}|)^2|x - \tilde{x}|^2. \tag{5.8}$$

Finally, for $v, v_* \in \mathbb{R}^3$, $\sigma(v - v_*)v = \sigma(v - v_*)v_*$, because $\Pi_{(v-v_*)^\perp}(v - v_*) = 0$, and so

$$|\sigma(v-v_*)v| \leq C||\sigma(v-v_*)||(|v|\wedge|v_*|) \leq C|v-v_*|^{1+\gamma/2}(|v|\wedge|v_*|) \leq C|v-v_*|^{\gamma/2}|v||v_*|, \tag{5.9}$$

because $|v - v_*|(|v| \wedge |v_*|) \leq (|v| + |v_*|)(|v| \wedge |v_*|) \leq 2|v||v_*|$.

5.3 A Tightness Property for Solutions of the Nonlinear SDE

We first prove the following tightness property for solutions to (stLE), which will be in frequent use later. We remark that this step is not necessary when arguing at the level of weak solutions $(\mu_t)_{t\geq 0}$ to (LE), but will be necessary in proving the 'stochastic' form of the results Theorems 5.1, 5.2, 5.3; in preparation for applying the result in the course of Theorems 5.2, 5.3, we are careful that the *only* moment requirements are some uniform integrability property for $\mu^n \in L^\infty_{\text{loc}}([0,\infty),\mathcal{P}_2(\mathbb{R}^3))\cap L^1_{\text{loc}}([0,\infty),\mathcal{P}_{2+\gamma})$.

Lemma 5.5. *Let $(\mu^n_t)_{t\geq 0}$ be a family of weak solution to (LE) with the uniform integrability conditions*

$$\sup_{t_{\text{fin}}\geq 0}\lim_{R\to\infty}\sup_n\sup_{t\leq t_{\text{fin}}}\int_{\mathbb{R}^3}(1+|v|^2)\mathbb{1}_{|v|\geq R}\mu^n_t(dv)=0; \tag{5.10}$$

$$\sup_{t\geq 0}\lim_{R\to\infty}\sup_n\int_0^t\int_{\mathbb{R}^3}(1+|v|^{2+\gamma})\mathbb{1}_{|v|\geq z}\mu^n_s(dv)ds=0. \tag{5.11}$$

For each n, let $(V^n_t)_{t\geq 0}$ be a solution to (stLE) with $\text{Law}(V^n_t)=\mu^n_t$ *for all $t\geq 0$. Then the processes $(V^n_t)_{t\geq 0}$ are tight in the local uniform topology of $C([0,\infty),\mathbb{R}^3)$, and any subsequential limit point $(V_t)_{t\geq 0}$ is a solution to (stLE).*

As in the Boltzmann case, we will use Lemma 4.14 in the proof. Compared to the Boltzmann case Lemma 4.13, this lemma is much more borderline and difficult, because we have the growth $|\mathcal{L}f|\leq C(1+|v|^{2+\gamma}+|v_*|^{2+\gamma})$, and the exponent $2+\gamma$ is the borderline case for local integrability estimates $\Lambda_p(\mu_t)\in L^1_{\text{loc}}([0,\infty))$ from Proposition 2.12, unless we make additional moment assumptions. By contrast, in the Boltzmann case, we had the exponent $1+\gamma\leq 2$, which we dominated using *a priori* esimtates on $\Lambda_{2+(\gamma/2)}(\mu_t)\in L^1_{\text{loc}}([0,\infty))$.

Proof of Lemma 5.5. First, (5.10) gives immediately that the second moments $\Lambda_2(\mu^n_t)$ are bounded, and (5.11) implies that $\int_0^t\Lambda_{2+\gamma}(\mu^n_u)du$ is bounded in n, locally uniformly in t.

Step 1: Tightness We first prove tightness. Let us fix $\epsilon>0$ and a time horizon t_{fin}. In what follows, A,C will be be allowed to vary from line to line, depending only on $\varepsilon, t_{\text{fin}}$ and the uniform integrablity estimates, but not on n.

Returning to (stLE), we can write V^n_t as

$$V^n_t=V_0+M^n_t+\int_0^t\int_{\mathbb{R}^3}b(V^n_u-v_\star)\mu^n_u(dv_\star)du \tag{5.12}$$

where M is a $\mathbb{R}^3$-valued martingale with quadratic covariation

$$[M^n]_t = \int_0^t \int_{\mathbb{R}^3} \sigma(V_u^n - v_\star)\sigma^\star(V_u^n - v_\star)\mu_u^n(dv_\star)du. \tag{5.13}$$

Recalling (5.4) and that $\mathrm{Law}(V_u^n) = \mu_u^n$, we integrate to find that, for some absolute constant C,

$$\begin{aligned}\mathbb{E}\left(\mathrm{Tr}[M^n]_t\right) &= \int_0^t \int_{\mathbb{R}^3} \mathbb{E}[\langle\!\langle \sigma(V_u^n - v_\star), \sigma(V_u^n - v_\star)\rangle\!\rangle]\mu_u^n(dv_\star)du \\ &= 2\int_0^t \int_{\mathbb{R}^3} \mathbb{E}[|V_u^n - v_\star|^{2+\gamma}]\mu_u^n(dv_\star)du \\ &\le C\int_0^t \int_{\mathbb{R}^3} (\mathbb{E}[|V_u^n|^{2+\gamma} + |v_\star|^{2+\gamma})\mu_u^n(dv_\star)du \le 2C\int_0^t \Lambda_{2+\gamma}(\mu_u^n)du.\end{aligned} \tag{5.14}$$

By the uniform integrability hypothesis, the last expression is bounded, uniformly in $n \ge 1$ and $t \le t_{\mathrm{fin}}$. In particular, M_t^n are $L^2(\mathbb{P})$-bounded, uniformly in n, and there exists A, such that, for all n

$$\mathbb{P}\left(\sup_{0\le t\le t_{\mathrm{fin}}} |M_t^n| > A\right) \le \frac{\varepsilon}{3}. \tag{5.15}$$

For the drift term, we similarly estimate

$$\begin{aligned}\mathbb{E}\left[\int_0^t \int_{\mathbb{R}^3} |b(V_u^n - v_\star)|\mu_u^n(dv_\star)\right] &\le C\int_0^t \int_{\mathbb{R}^3} (\mathbb{E}[|V_u^n|^{1+\gamma}] + |v_\star|^{1+\gamma})du \\ &\le 2C\int_0^t \Lambda_2(\mu_u^n)du = 2Ct\Lambda_2(\mu_0^n).\end{aligned} \tag{5.16}$$

The final expression is bounded, uniformly in n, and so, possibly making A larger, we can arrange that

$$\mathbb{P}\left(\sup_{t\le t_{\mathrm{fin}}} \left|\int_0^t \int_{\mathbb{R}^3} b(V_u^n - v_\star)\mu_u^n(du)\right| > A\right) < \frac{\varepsilon}{3} \tag{5.17}$$

again uniformly in n. Finally, using the boundedness of the second moments $\mathbb{E}[|V_0^n|^2] = 1$, we can also choose A so that $\mathbb{P}(|V_0^n| > A) < \varepsilon/3$, and combining everything, for a new A and uniformly in n,

$$\mathbb{P}\left(\sup_{0\le t\le t_{\mathrm{fin}}} |V_t^n| > A\right) < \epsilon. \tag{5.18}$$

Let us now check equicontinuity with high probability on this event. For the drift term, observe that, on this event, for any $0 \le t \le t' \le t_{\mathrm{fin}}$, we have the pathwise and almost sure bound

$$\begin{aligned}\left|\int_t^{t'} \int_{\mathbb{R}^3} b(V_u^n - v_\star)\mu_u^n(dv_\star)\right| &\le C\int_t^{t'} (|V_u^n| + |v_\star|)\mu_u^n(dv_\star) \\ &\le C(t'-t)(A^{1+\gamma} + \Lambda_2(\mu_0^n)) \\ &\le C(t'-t)A^{1+\gamma}\end{aligned} \tag{5.19}$$

for a new choice of A in the final line. For the martingale term, we use the Damins-Dubbins-Schwarz theorem to write the ith component $M^{n,i}$ of M^n as

$$M^{n,i}_t = B^{n,i}_{[M^{s,i}]_t} \tag{5.20}$$

for three (not necessarily independent) Brownian motions $B^{n,i}$. On the event in (5.18), we have the bound

$$[M^{n,i}]_{t_{\text{fin}}} \le C\left(t_{\text{fin}}A^{2+\gamma} + \int_0^{t_{\text{fin}}} \Lambda_{2+\gamma}(\mu^n_u)du\right) \le z \tag{5.21}$$

for some $z<\infty$, independent of n, and thanks to the usual pathwise regularity properties of Brownian motion, there exists α, depending only on z,ε such that

$$\mathbb{P}\left(|B^{n,i}_u - B^{n,i}_v| \le \alpha|u-v|^{1/4}/3 \text{ for all } u,v\le z,\ i=1,2,3\right) > 1-\varepsilon. \tag{5.22}$$

Returning to the analysis leading to (5.14) and still working on the event in (5.18), we have the almost sure, nonrandom upper bound, for all $0\le t\le t'\le t_{\text{fin}}$,

$$\begin{aligned}[M^{n,i}]_{t'} - [M^{n,i}]_t &\le 2\int_t^{t'}\int_{\mathbb{R}^3} |V^n_u - v_\star|^{2+\gamma}\mu^n_u(dv_\star)du \\ &\le C\left(A^{2+\gamma}(t'-t) + \int_t^{t'}\Lambda_{2+\gamma}(\mu^n_u)du\right).\end{aligned} \tag{5.23}$$

Let us now define $\vartheta:[0,\infty)\to[0,\infty)$ to be the modulus of continuity

$$\vartheta(x) := \sup\left\{\int_t^{t'}\Lambda_{2+\gamma}(\mu^n_u)du : 0\le t\le t'\le t_{\text{fin}},\ t'-t\le x,\ n\ge 1\right\}.$$

Clearly $\vartheta(0)=0$. Moreover, given $\epsilon>0$, we use (5.11) to find R such that, uniformly in n, $\int_0^{t_{\text{fin}}}\int_{\mathbb{R}^3}|v|^{2+\gamma}1\!\!1_{|v|\ge R}\mu^n_u(dv)du < \varepsilon/2$, and choose $x<\varepsilon/2(R^{2+\gamma})$. For any $0\le t\le t'\le t_{\text{fin}}, |t-t'|\le x$ and any n, we have

$$\begin{aligned}\int_t^{t'}\Lambda_{2+\gamma}(\mu^n_u)du &\le \int_0^{t_{\text{fin}}}\int_{\mathbb{R}^3}|v|^{2+\gamma}1\!\!1_{|v|\ge R}\mu^n_u(du)du + \int_t^{t'}\int_{\mathbb{R}^3}|v|^{2+\gamma}1\!\!1_{|v|\le R}\mu^n_u(dv)du \\ &< \frac{\varepsilon}{2} + R^{2+\gamma}(t'-t) < \epsilon\end{aligned} \tag{5.24}$$

and we see that ϑ is continuous at 0. Returning to (5.23), still working on the event in (5.18), we have

$$|[M^{n,i}]_t - [M^{n,i}]_t| \le C\left(A^{2+\gamma}(t'-t) + \vartheta(t'-t)\right). \tag{5.25}$$

Consequently, combining (5.18) and (5.22), we conclude that

$$\begin{aligned}\mathbb{P}\Big(&|V^n_t|\le A \text{ and } |M^n_t - M^n_{t'}| \le \alpha\left(CA^{2+\gamma}(t'-t)+\vartheta(t'-t)\right)^{1/4} \\ &\qquad\qquad \text{for all } 0\le t\le t'\le t_{\text{fin}}\Big) > 1-2\epsilon.\end{aligned} \tag{5.26}$$

Finally, combining with (5.19) and defining the set

$$\mathcal{K} = \Big\{ v \in C([0,t_{\text{fin}}],\mathbb{R}^3) : |v_t| \le A, |v_t - v_{t'}| \le \alpha(CA^{2+\gamma}(t'-t) + \vartheta(t'-t))^{1/4} \\ \cdots + CA^{1+\gamma}(t'-t) \text{ for all } 0 \le t \le t' \le t_{\text{fin}} \Big\} \tag{5.27}$$

we have proven that, uniformly in n,

$$\mathbb{P}\left((V_t^n)_{0\le t\le t_{\text{fin}}} \notin \mathcal{K}\right) < 2\epsilon. \tag{5.28}$$

The sets $\mathcal{K}$ are compact in the uniform topology of $C([0,t_{\text{fin}}],\mathbb{R}^3)$ by the classical Arzelà-Ascoli Theorem, so it follows that $(V_t^n)_{0\le t\le t_{\text{fin}}}$ are tight for the uniform topology on any compact time interval, and therefore the processes $(V_t^n)_{t\ge0}$ are tight in the local uniform topology.

Step 2: Characterisation of the Limits Let us now suppose that $(V_t)_{t\ge0}$ is any process extracted from $(V_t^n)_{t\ge0}$ as the limit in distribution, for the local uniform topology, along some subsequence; to ease notation, we will not relabel the subsequence. Using Skorokhod's representation theorem, we can replace $(V_t^n)_{t\ge0}$ by processes with the same law so that the convergence is almost sure; we will use the same notation for the new processes. Writing $\mu_t = \text{Law}(V_t)$, since $\text{Law}(V_t^n) = \mu_t^n$ and $V_t^n \to V_t$ almost surely, we see that $\mu_t^n \to \mu_t$ in the weak topology.

To see that $(V_t)_{t\ge0}$ solves the SDE (stLE), it is sufficient to show that $V_t - \int_0^t \int_{\mathbb{R}^3} b(V_u - v_\star)\mu_u(dv_\star)$ is a martingale with the correct covariation, equivalent to the expression (5.13) we wrote for M^n above, in order to apply general results on nonlinear processes driven by white noise [71]. Let us fix a time horizon t_{fin}, and a sequence of continuous functions $v^n \in C([0,t_{\text{fin}}],\mathbb{R}^3)$, converging uniformly to $v \in C([0,t_{\text{fin}}] \times \mathbb{R}^3)$. Thanks to uniform convergence, we can find a compact set K containing the images of v^n, v, for all n; let us fix $\epsilon > 0$. We will now show convergence of the integrated coefficients in this (deterministic) setting.

For the drift term, we start by observing that $|b(v-v_\star)| \le C(|v|^{1+\gamma} + |v_\star|^{1+\gamma}) \le C(1+|v|^2)(1+|v_\star|^2)$ and hence, by the uniform integrability hypothesis (5.10), there exists R such that

$$\sup_{t\le t_{\text{fin}}} \sup_n \sup_{v\in K} \int_{\mathbb{R}^3} |b(v-v_\star)| \mathbb{1}_{|v_\star|\ge R} \mu_t^n(dv_\star) \\ \le \sup_{v\in K}(1+|v|^2) \sup_n \sup_{t\le t_{\text{fin}}} \int_{\mathbb{R}^3} (1+|v_\star|^2)\mathbb{1}_{|v|_\star \ge R}\mu_t^n(dv_\star) < \varepsilon. \tag{5.29}$$

Now, we choose a continuous functions $\varphi : \mathbb{R}^3 \to [0,1]$ such that $\varphi(v) = 1$ if $|v| \le R$ and $\varphi(v) = 0$ if $|v| \ge R+1$. Using (5.29), it follows that

$$\sup_n \sup_{t\in[0,t_{\text{fin}}]} \sup_{v\in K} \left| \int_{\mathbb{R}^3} b(v-v_\star)(1-\varphi)(v_\star)\mu_t^n(dv_\star) \right| < \varepsilon \tag{5.30}$$

and the same holds for μ_t by taking a limit of (5.29) and using lower semicontinuity. Let us now fix $t \in [0, t_{\text{fin}}]$. Applying Lemma 4.14, there exists $n_0 > 0$ such that, for all $n \geq n_0$ and all $v \in K$, we have

$$\left| \int_{\mathbb{R}^3} b(v - v_\star)\varphi(v_\star)\mu_t^n(dv_\star) - \int_{\mathbb{R}^3} b(v - v_\star)\varphi(v_\star)\mu_t(dv_\star) \right| < \epsilon. \tag{5.31}$$

Moreover, it is easy to see that the map $(v, v_\star) \mapsto b(v - v_\star)\varphi(v_\star)$ is uniformly continuous on $K \times \mathbb{R}^3$, and so there exists $\delta > 0$ such that if $v, w \in K$ with $|v - w| < \delta$,

$$\left| \int_{\mathbb{R}^3} b(w - v_\star)\varphi(v_\star)\mu_t(dv_\star) - \int_{\mathbb{R}^3} b(v - v_\star)\varphi(v_\star)\mu_t(dv_\star) \right| < \epsilon. \tag{5.32}$$

In particular, as soon as $n \geq n_0$ is large enough that $|v_t^n - v_t| < \delta$, we combine the three previous displays to obtain

$$\left| \int_{\mathbb{R}^3} b(v_t^n - v_\star)\mu_t^n(dv_\star) - \int_{\mathbb{R}^3} b(v_t - v_\star)\mu_t(dv_\star) \right| < 4\epsilon. \tag{5.33}$$

In particular, we have shown convergence

$$\int_{\mathbb{R}^3} b(v_t^n - v_\star)\mu_t^n(dv_\star) \to \int_{\mathbb{R}^3} b(v_t - v_\star)\mu_t(dv_\star) \tag{5.34}$$

for any fixed t. To obtain convergence when we integrate in time, we use the same argument as (5.29) to obtain the n-uniform bound, uniformly in $t \leq t_{\text{fin}}$

$$\left| \int_{\mathbb{R}^3} b(v_t^n - v_\star)\mu_t^n(dv_\star) \right| \leq \sup_{v \in K}(1 + |v|^2) \sup_m \sup_{u \leq t_{\text{fin}}} \int_{\mathbb{R}^3} (1 + |v_\star|^2)\mu_u^m(dv_\star) < \infty. \tag{5.35}$$

Therefore, by bounded convergence, it follows that

$$\int_0^t \int_{\mathbb{R}^3} b(v_t^n - v_\star)\mu_t^n(dv_\star) \to \int_0^t \int_{\mathbb{R}^3} b(v_t - v_\star)\mu_t(dv_\star) \tag{5.36}$$

uniformly on compact time intervals. We next prove a similar property for the noise term. For v^n, v as above, we observe that $\|\sigma^\star\sigma(v - v_\star)\| \leq 2(1 + |v|^{2+\gamma})(1 + |v_\star|^{2+\gamma})$ and using (5.11), we find R such that, uniformly in n,

$$\begin{aligned} &\sup_{v \in K} \int_0^{t_{\text{fin}}} \int_{\mathbb{R}^3} \|\sigma\sigma^\star(v - v_\star)\| \mathbb{1}_{|v|_\star \geq R}\mu_u^n(dv_\star)du \\ &\qquad \leq 2 \sup_{v \in K}(1 + |v|^{2+\gamma}) \sup_n \int_0^{t_{\text{fin}}} \int_{\mathbb{R}^3} (1 + |v_\star|^{2+\gamma})\mathbb{1}_{|v_\star| \geq R}\mu_u^n(dv_\star)du < \varepsilon. \end{aligned} \tag{5.37}$$

We now choose φ as for the previous case with the new choice of R, from which it follows that

$$\sup_n \sup_{t \in [0, t_{\text{fin}}]} \int_0^t \int_{\mathbb{R}^3} \|\sigma\sigma^\star(v_u^n - v_\star)\|(1 - \varphi)(v_\star)\mu_u^n(dv_\star)du < \epsilon \tag{5.38}$$

and similarly for v_u, μ_u. We again fix $t \leq t_{\text{fin}}$; applying Lemma 4.14 to each entry of the matrix, we have

$$\sup_{v \in K} \left\| \int_{\mathbb{R}^3} \sigma\sigma^\star(v - v_\star)\varphi(v_\star)\mu_t^n(dv_\star) - \int_{\mathbb{R}^3} \sigma\sigma^\star(v - v_\star)\varphi(v_\star)\mu_t(dv_\star) \right\| \to 0. \tag{5.39}$$

Now, using bounded convergence, it follows that

$$\left\|\int_0^t \int_{\mathbb{R}^3} \sigma\sigma^\star(v^n_u - v_\star)\varphi(v_\star)\mu^n_u(dv_\star)du - \int_0^t \int_{\mathbb{R}^3} \sigma\sigma^\star(v^n_u - v_\star)\varphi(v_\star)\mu_u(dv_\star)du\right\| \to 0 \quad (5.40)$$

as $n \to \infty$, uniformly on $[0, t_{\text{fin}}]$. Using uniform continuity of $(v, v_\star) \to \sigma\sigma^\star(v - v_\star)\varphi(v_\star)$ on $(v, v_\star) \in K \times \mathbb{R}^3$ as we did for b, it also follows that

$$\sup_{t \le t_{\text{fin}}} \left\|\int_{\mathbb{R}^3} \sigma\sigma^\star(v^n_t - v_\star)\varphi(v_\star)\mu_t(dv_\star) - \int_{\mathbb{R}^3} \sigma\sigma^\star(v_t - v_\star)\varphi(v_\star)\mu_t(dv_\star)\right\| \to 0. \quad (5.41)$$

Combining, it follows that

$$\left\|\int_0^t \int_{\mathbb{R}^3} \sigma\sigma^\star(v^n_u - v_\star)\varphi(v_\star)\mu^n_u(dv_\star) - \int_0^t \int_{\mathbb{R}^3} \sigma\sigma^\star(v_u - v_\star)\varphi(v_\star)\mu_u(dv_\star)\right\| \to 0 \quad (5.42)$$

uniformly in $t \le t_{\text{fin}}$. Returning to (5.38), we have proven that

$$\limsup_n \sup_{t \le t_{\text{fin}}} \left\|\int_0^t \int_{\mathbb{R}^3} \sigma\sigma^\star(v^n_t - v_\star)\mu^n_u(dv_\star)du - \int_0^t \int_{\mathbb{R}^3} \sigma\sigma^\star(v_u - v_\star)\mu_u(dv_\star)du\right\| \le 2\epsilon \quad (5.43)$$

and hence conclude that

$$\int_0^t \sigma\sigma^\star(v^n_u - v_\star)\mu^n_u(dv_\star)du \to \int_0^t \sigma\sigma^\star(v_u - v_\star)\mu_u(dv_\star)du \quad (5.44)$$

uniformly on compact time intervals. With these convergences in hand, we return to the stochastic processes. The previous convergences apply pathwise with $v^n_t = V^n_t, v_t = V_t$, so that almost surely

$$\int_0^t b(V^n_u - v_\star)\mu^n_u(dv_\star)du \to \int_0^t b(V_u - v_\star)\mu_u(dv_\star)du; \quad (5.45)$$

$$\int_0^t \sigma\sigma^\star(V^n_u - v_\star)\mu^n_u(dv_\star)du \to \int_0^t \sigma\sigma^\star(V_u - v_\star)\mu_u(dv_\star)du \quad (5.46)$$

uniformly on compact time intervals. Therefore, the martingales $M^n_t = V^n_t - V^n_0 - \int_0^t \int_{\mathbb{R}^3} b(V^n_u - v_\star)\mu^n_u(dv_\star)du$ converge, uniformly on compact time intervals, to a process M_t. As proven in Step 1, $[M^n]$ are $L^2(\mathbb{P})$-bounded on the compact time interval $[0, t_{\text{fin}}]$, and hence so is M; it therefore follows that M is a true, $L^2(\mathbb{P})$-bounded martingale. Similarly, the (true) martingales $Z^n_t = M^n_t \otimes M^n_t - \int_0^t \int_{\mathbb{R}^3} \sigma\sigma^\star(V^n_u - v_\star)\mu^n_u(dv_\star)du$ converge almost surely, uniformly on compact time intervals, to

$$Z_t = M_t \otimes M_t - \int_0^t \int_{\mathbb{R}^3} \sigma\sigma^\star(V_u - v_\star)\mu_u(dv_\star)du \quad (5.47)$$

which is therefore a local martingale. We have therefore identified the quadratic variation $[M]_t = \int_0^t \int_{\mathbb{R}^3} \sigma\sigma^\star(V_u - v_\star)\mu_u(dv_\star)du$, and the step is complete. □

5.4 Tanaka-style Coupling of Landau Processes

We now set up the coupling of solutions which will lead to Theorem 5.1, corresponding to the Tanaka coupling we used in the Boltzmann case in Chapter 4. For $E=\mathbb{R}^3$ or $\mathbb{R}^3\times\mathbb{R}^3$, we denote by $C^2_p(E)$ the set of C^2 functions on E of which the derivatives of order 0 to 2 have at most polynomial growth.

Lemma 5.6. *Fix $\gamma\in(0,1]$, and consider two weak solutions $(\mu_t)_{t\ge0},(\nu_t)_{t\ge0}\subset\mathcal{S}$ to (LE) such that $\int_{\mathbb{R}^3}e^{a|v|^2}(\mu_0+\nu_0)(dv)<\infty$ for some $a>0$, and fix $\pi_0\in\Pi(\mu_0,\nu_0)$. Then there exists a stochastic process $(V_t,\tilde V_t)_{t\ge0}$, taking values in $(\mathbb{R}^3)^2$, such that $\pi_t=\mathrm{Law}(V_t,\tilde V_t)\in\Pi(\mu_t,\nu_t)$, and*

$$\begin{cases} V_t = V_0+\int_0^t\int_{\mathbb{R}^3\times\mathbb{R}^3} b(V_s-v_*)\pi_s(dv_*,d\tilde v_*)ds+\int_0^t\int_{\mathbb{R}^3\times\mathbb{R}^3}\sigma(V_s-v_*)N(dv_*,d\tilde v_*,ds);\\ \tilde V_t = \tilde V_0+\int_0^t\int_{\mathbb{R}^3\times\mathbb{R}^3} b(\tilde V_s-\tilde v_*)\pi_s(dv_*,d\tilde v_*)ds+\int_0^t\int_{\mathbb{R}^3\times\mathbb{R}^3}\sigma(\tilde V_s-\tilde v_*)N(dv_*,d\tilde v_*,ds)\end{cases} \tag{5.48}$$

where $N=(N^1,N^2,N^3)$ is a 3D-white noise on $\mathbb{R}^3\times\mathbb{R}^3\times[0,\infty)$ with covariance measure $\pi_s(dv_*,d\tilde v_*)ds$. In particular, each coordinate $(V_t)_{t\ge0},(\tilde V_t)_{t\ge0}$ solves the nonlinear stochastic differential equation (stLE), and for all $f\in C^2_p(\mathbb{R}^3\times\mathbb{R}^3)$, the process

$$M^f_t=f(V_t,\tilde V_t)-f(V_0,\tilde V_0)-\int_0^t\int_{\mathbb{R}^3\times\mathbb{R}^3}\mathcal{A}f(V_s,\tilde V_s,v_*,\tilde v_*)\pi_s(dv,dv_\star)ds \tag{5.49}$$

is a martingale, where

$$\begin{aligned}\mathcal{A}f(v,\tilde v,v_*,\tilde v_*)=&\sum_{k=1}^3[b_k(v-v_*)\partial_{v_k}f(v,\tilde v)+b_k(\tilde v-\tilde v_*)\partial_{\tilde v_k}f(v,\tilde v)]\\ &+\frac12\sum_{k,\ell=1}^3[a_{k\ell}(v-v_*)\partial^2_{v_kv_\ell}f(v,\tilde v)+a_{k\ell}(\tilde v-\tilde v_*)\partial^2_{\tilde v_k\tilde v_\ell}f(v,\tilde v)]\\ &+\sum_{j,k,\ell=1}^3\sigma_{kj}(v-v_*)\sigma_{\ell j}(\tilde v-\tilde v_*)\partial^2_{v_k\tilde v_\ell}f(v,\tilde v).\end{aligned} \tag{5.50}$$

Remark 5.7. *Let us make the following remark.*

i). This will produce the coupling desired for prove Theorem 5.1 in the cases where the initial data have Gaussian initial moments.

ii). Since $\pi_s\in\Pi(\mu_s,\nu_s)$, we have $\int_{\mathbb{R}^3\times\mathbb{R}^3}b(V_s-v_*)\pi_s(dv_*,d\tilde v_*)ds=\int_{\mathbb{R}^3}b(V_s-v_*)\mu_s(dv_*)ds$. Similarly, $\int_0^t\int_{\mathbb{R}^3\times\mathbb{R}^3}\sigma(V_s-v_*)N(dv_*,d\tilde v_*,ds)=\int_0^t\int_{\mathbb{R}^3}\sigma(V_s-v_*)W(dv_*,ds)$, for some 3D-white noise on $\mathbb{R}^3\times[0,\infty)$ of covariance measure $\mu_s(dv_*)ds$. Hence in law, *the first SDE (for $(V_t)_{t\ge0}$) does not depend on $(\nu_t)_{t\ge0}$.*

iii). To prove the stability result purely at the level of the laws $(\mu_t)_{t\ge0}$, we could instead work at the level of the PDE and prove the existence of a solution $\pi_t\in\Pi(\mu_t,\nu_t)$ to *the coupled Landau equation*

$$\langle f, \pi_t \rangle = \langle f, \pi_0 \rangle + \int_0^t \int_{(\mathbb{R}^3)^2 \times (\mathbb{R}^{3^2})} \mathcal{A}f(x,y)\pi_s(dx)\pi_s(dy)ds. \tag{5.51}$$

This is exactly the argument presented in [90]. We argue here at the stochastic level to show how the coupling can be achieved in a dynamic way by coupling solutions to (stLE)*.*

iv). As in Chapter 4, this form of the coupling is crucial, rather than taking π_t to be (say) a $w_{p,1}$-optimal coupling of μ_t, ν_t. As in the Boltzmann case, we will use the fact that $\pi_t = \mathrm{Law}(V_t, \tilde{V}_t)$ for a symmetry argument, which allows us to ensure that the relevant estimates close.

v). We do not claim the uniqueness of solutions to (5.51) *or uniqueness in law for $(V_t, \tilde{V}_t)$; existence is sufficient for our needs.*

Along the way, we will use the following proposition, which plays the same rôle here that Proposition 4.18 does in the Boltzmann case.

Proposition 5.8 ([88], Theorem 2)**.** *Fix $\gamma \in (0,1]$ and let $\mu_0 \in \mathcal{S}$ be such that*

$$\int_{\mathbb{R}^3} e^{|v|^\alpha} \mu_0(dv) < \infty \quad \textit{for some } \alpha > \gamma. \tag{5.52}$$

Then there exists a unique weak solution $(\mu_t)_{t\geq 0}$ to (LE) *starting at μ_0.*

As in the Boltzmann case in Section 4.5, we will outline the important points of the proof of our auxiliary uniqueness result, following [88, Theorem 2]. Since this proof builds in turn on a similar stochastic representation to that of Proposition 5.6, it is deferred until after that proof. we remark now that the proof will only use the *existence* of a Landau process, which can be proven by following Steps 1-5 of the argument below, while we need this proposition to complete step 6. In particular, there is no circularity.

Proof of Proposition 5.6. We first prove the result for the case where the coefficients are Lipschitz, sketching the argument of Guérin [105]. Since the coefficients b, σ appearing in (5.48) are not Lipschitz, we will then use a compactness argument to produce solutions from solutions of a truncated equation.

Step 1: Construction of a Truncated Equation We fix $k \geq 1$ and define the truncated *two level* coefficients $B_k : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3 \times \mathbb{R}^3$ and $\Sigma_k : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathcal{M}_{6\times 3}(\mathbb{R})$ by

$$B_k \begin{pmatrix} x \\ \tilde{x} \end{pmatrix} = \begin{pmatrix} b_k(x) \\ b_k(\tilde{x}) \end{pmatrix}; \qquad \Sigma_k \begin{pmatrix} x \\ \tilde{x} \end{pmatrix} = \begin{pmatrix} \sigma_k(x) \\ \sigma_k(\tilde{x}) \end{pmatrix},$$

where $b_k(x) = -2(|x| \wedge k)^\gamma x$ and $\sigma_k(x) = (|x| \wedge k)^{\gamma/2}|x|\Pi_{x^\perp}$, and define similarly

$$B\begin{pmatrix} x \\ \tilde{x} \end{pmatrix} = \begin{pmatrix} b(x) \\ b(\tilde{x}) \end{pmatrix}; \qquad \Sigma_k\begin{pmatrix} x \\ \tilde{x} \end{pmatrix} = \begin{pmatrix} \sigma(x) \\ \sigma(\tilde{x}) \end{pmatrix},$$

Exactly the same arguments as lead to (5.6) and (5.8) show that B_k and Σ_k are globally Lipschitz continuous on $\mathbb{R}^3 \times \mathbb{R}^3$.

With this notation, we can compactly rewrite (5.48) as an equation for $X_t = (V_t, \tilde{V}_t)$:

$$X_t = X_0 + \int_0^t \int_{\mathbb{R}^3\times\mathbb{R}^3} B(X_s - x)\pi_s(dx)ds + \int_0^t \int_{\mathbb{R}^3\times\mathbb{R}^3} \Sigma(X_s - x)N(dx, ds)$$

and approximating equations with Lipschitz coefficients

$$X^k_t = X^k_0 + \int_0^t \int_{\mathbb{R}^3\times\mathbb{R}^3} B_k(X^k_s - x)\pi_s(dx)ds + \int_0^t \int_{\mathbb{R}^3\times\mathbb{R}^3} B_k(X^k - x)N(dx, ds). \quad (5.53)$$

Step 2: Existence for the Truncated Equation We now prove the existence of solutions to (5.53), following the arguments of Guérin, which adapt the usual argument for stochastic differential equations to the white noise case. We consider the auxiliary probability space $((0,1), \mathcal{B}(0,1), d\alpha)$, and we will write a subscript $_\alpha$ to denote objects defined on this probability space. We define the spaces $\mathcal{C}^2$ of continuous, adapted processes X on the underlying probability space $(\Omega, \mathfrak{F}, (\mathfrak{F}_t)_{t\ge 0}, \mathbb{P})$ to $\mathbb{R}^3 \times \mathbb{R}^3$ such that, for all $t \ge 0$, $\mathbb{E}[\sup_{s\le t}|X_s|^2] < \infty$, and similarly $\mathcal{C}^2_\alpha$ for continuous processes on the auxiliary space, which we equip with the norm

$$\|X\|_{2,t} := \mathbb{E}\left[\sup_{s\le t}|X_s|^2\right]^{1/2}.$$

Now, let $W = (W^1, W^2, W^3)$ be a white noise on $[0,\infty) \times (0,1)$ with covariance measure $dsd\alpha$, and $X_0 \sim \pi_0$ be independent of W, and consider the map Φ_k on $\mathcal{C}^2_\alpha \times \mathcal{C}^2$, given by

$$\Phi_k(Y,Z)_t := X_0 + \int_0^t \int_{(0,1)} B_k(Z_s - Y_s(\alpha))d\alpha ds + \int_0^t \int_{(0,1)} \Sigma_k(Z_s - Y_s(\alpha))W(ds, d\alpha). \quad (5.54)$$

It is immediate that $\Phi_k(Y,Z)$ defines a continuous process, and Itô's isometry shows that $\Phi_k(Y,Z) \in \mathcal{C}^2$. Since the coefficients are Lipschitz, the standard arguments for stochastic differential equations lead to

$$\|\Phi_k(Y,Z) - \Phi_k(Y',Z')\|^2_{2,t} \le C_k(1+t)\left(\int_0^t \|Z - Z'\|^2_{2,s}ds + \int_0^t \|Y - Y'\|^2_{2,s,\alpha}ds\right) \quad (5.55)$$

for some constant C_k.

We now use a nonlinear version of the usual Picard iteration argument. For $n = 0$, we set $X^{k,0}_t := X_0 \in \mathcal{C}^2$. For the iterative step, given $X^{k,0}, ..., X^{k,n} \in \mathcal{C}^2$ and $Y^{k,0}, \dots, Y^{k,n-1} \in \mathcal{C}^2_\alpha$

such that $\mathrm{Law}(X^{k,0},\dots,X^{k,n-1}) = \mathrm{Law}_\alpha(Y^{k,0},\dots,Y^{k,n-1})$, we can use the disintegration theorem and Skorokhod's representation theorem to construct $Y^{k,n}\in\mathcal{C}^2_\alpha$ such that $\mathrm{Law}(X^{k,0},...,X^{k,n}) = \mathrm{Law}_\alpha(Y^{k,0},\dots,Y^{k,n})$, and now define $X^{k,n+1} := \Phi_k(Y^{k,n},X^{k,n})$. In particular, we have $\mathrm{Law}(X^{k,n-1},X^{k,n}) = \mathrm{Law}_\alpha(Y^{k,n-1},Y^{k,n})$, and so (5.55) gives

$$\begin{aligned}\|X^{k,n+1}-X^{k,n}\|^2_{2,t} &\le C_k(1+t)\left(\int_0^t \|X^{k,n}-X^{k,n-1}\|^2_{2,s}+\|Y^{k,n}-Y^{k,n-1}\|^2_{2,s,\alpha}ds\right)\\ &\le 2C_k(1+t)\left(\int_0^t\|X^{k,n}-X^{k,n-1}\|^2_{2,s}ds\right).\end{aligned}\tag{5.56}$$

Iterating this bound produces

$$\begin{aligned}\|X^{k,n+1}-X^{k,n}\|^2_{2,t} &\le (2C_k(1+t))^n\int_0^t dt_1....\int_0^{t_{n-1}}dt_n\|X^{k,1}-X^{k,0}\|^2_{2,t_n}\\ &\le \frac{(2C_kt(1+t))^n}{n!}\|X^{k,1}-X^{k,0}\|^2_{2,t}.\end{aligned}\tag{5.57}$$

Thanks to the $n!$ in the denominator, the right-hand side is summable, so completeness of $(\mathcal{C}^2,\|\cdot\|_{2,t})$ implies that there exists $X^k\in\mathcal{C}^2$ such that $\|X^{k,n}-X^k\|_{2,t}\to 0$ for all t, and similarly $Y^k\in\mathcal{C}^2_\alpha$. Since $\mathrm{Law}(X^{k,n}) = \mathrm{Law}_\alpha(Y^{k,n})$ for all n, it follows that $\mathrm{Law}(X^k)=\mathrm{Law}_\alpha(Y^k)$. Using (5.55), we also see that $X^{k,n+1}=\Phi_k(Y^{k,n},X^{k,n})$ converges in $\|\cdot\|_{2,t}$ to $\Phi_k(Y^k,X^k)$, for all t, so we finally conclude that X^k satisfies

$$X^k_t = X^k_0+\int_0^t\int_{(0,1)}B_k(X^k_s-Y^k_s(\alpha))\,d\alpha ds+\int_0^t\int_{(0,1)}\Sigma_k(X^k_s-Y^k_s(\alpha))W(ds,d\alpha)$$

for a copy Y^k of X^k.

Step 3: Characterisation of the laws We next identify an appropriate generator, corresponding to (5.50), for the approximating processes X^k_t, and hence characterise the laws $\pi^k_t = \mathrm{Law}(X^k_t) = \mathrm{Law}_\alpha(Y^k_t)$. For any $f\in C^2_p(\mathbb{R}^3\times\mathbb{R}^3)$, we apply Itô's formula to find that

$$\begin{aligned}f(X^k_t) = f(X^k_0)+M^{f,k}_t &+\int_0^t\int_{(0,1)}\nabla f(X^k_s)\cdot B_k(X^k_s-Y^k_s(\alpha))d\alpha ds\\ &+\int_0^t\int_{(0,1)}\frac12\sum_{i,j=1}^6\partial_{ij}f(X^k_s)[\Sigma_k(X^k_s-Y^k_s(\alpha))\Sigma^*_k(X^k_s-Y^k_s(\alpha))]_{ij}]d\alpha ds\end{aligned}\tag{5.58}$$

for some martingale $M^{k,f}$. Next, we observe that the integrated terms are exactly

$$\int_0^t\int_{(0,1)}\mathcal{A}_kf(X^k_s,Y^k_s(\alpha))d\alpha ds$$

where $\mathcal{A}_kf$ is defined as $\mathcal{A}f$ in (5.50), replacing b, σ and $a=\sigma\sigma^*$ by b_k, σ_k and $a_k=\sigma_k\sigma^*_k$ respectively. Taking expectations and using that $\pi^k_t=\mathrm{Law}(X^k_t)=\mathrm{Law}_\alpha(Y^k_t)$, we find

$$\begin{aligned}\int_{\mathbb{R}^3\times\mathbb{R}^3}f(v,\tilde v)\pi^k_t(dv,d\tilde v) &= \int_{\mathbb{R}^3\times\mathbb{R}^3}f(v,\tilde v)\pi_0(dv,d\tilde v)\\ &\quad+\int_0^t\int_{\mathbb{R}^3\times\mathbb{R}^3}\int_{\mathbb{R}^3\times\mathbb{R}^3}\mathcal{A}_kf(v,\tilde v,v_*,\tilde v_*)\pi^k_s(dv,d\tilde v)\pi^k_s(dv_*,d\tilde v_*)ds,\end{aligned}$$

For functions of the form $f(v,\tilde{v}) = f_1(v) + f_2(\tilde{v})$, we have $\mathcal{A}_k f(v,\tilde{v},v_*,\tilde{v}_*) = \mathcal{L}_k f_1(v,v_*) + \mathcal{L}_k f_2(\tilde{v},\tilde{v}_*)$, where $\mathcal{L}_k f_i$ is defined as $\mathcal{L} f_i = \mathcal{L}_{\mathrm{L}} f_i$, replacing b and a by b_k and a_k. It is then straightforward to check that the approximate equation (5.59) conserves energy and propagates moments, uniformly in k, using arguments similar to those of [58, Theorem 3] or Step 1 of the proof of Proposition 2.13ii). In particular, under our initial Gaussian moment assumption, all moments of π_t^k are bounded, uniformly $k \geq 1$, locally uniformly in $t \geq 0$, and $\mathbb{E}[|V_t^k|^2] = \mathbb{E}[|V_0^k|^2]$ for all $t \geq 0$, and similarly for $\tilde{V}_t$.

Step 4: A Compactness Argument We now argue that the laws of X^k are tight in $C([0,\infty),\mathbb{R}^3 \times \mathbb{R}^3)$ using similar arguments to Lemma 5.5, which will justify passing to a subsequence. Using the propagation of moments remarked in Step 3, it is straightforward to show that $\mathbb{E}[\sup_{s\leq t}|X_s^k|^2]$ is bounded, uniformly in $k \geq 1$ and locally uniformly in t. Next, using standard calculations, for any $0 \leq s < t \leq t_{\mathrm{fin}}$, we have

$$\begin{aligned} |X_t^k - X_s^k|^2 \leq 2(t-s)\int_s^t \int_{(0,1)} |B_k(X_u^k) - Y_u^k(\alpha)|^2 d\alpha ds \\ + 2\left(\int_s^t \int_{(0,1)} \Sigma_k(X_u^k) - Y_u^k(\alpha))W(ds,d\alpha)\right). \end{aligned} \tag{5.59}$$

Taking expectations and using Itô's isometry produces

$$\begin{aligned} \mathbb{E}[|X_t^k - X_s^k|^2] \leq 2(t-s)\int_s^t \int_{(0,1)} \mathbb{E}[|B_k(X_u^k) - Y_u^k(\alpha))|^2] d\alpha ds \\ + 2\mathbb{E}\left[\int_s^t \int_{(0,1)} \|\Sigma_k(X_u^k) - Y_u^k(\alpha))\|^2 ds d\alpha\right] \end{aligned} \tag{5.60}$$

Using the same estimates as in Section 5.2, we estimate

$$|B_k(X_u^k - Y_u^k(\alpha))|^2 \leq C(|X_u^k|^{2+2\gamma} + |Y_u^k(\alpha)|^{2+2\gamma});$$

$$\|\Sigma_k(X_u^k) - Y_u^k(\alpha))\|^2 \leq C(|X_u^k|^{2+\gamma} + |Y_u^k(\alpha)|^{2+\gamma})$$

for some constant C independent of k. We finally conclude that, allowing C to vary line to line, independently of k,

$$\mathbb{E}[|X_t^k - X_s^k|^2] \leq C\int_s^t \int_{\mathbb{R}^d\times\mathbb{R}^d} |x|^{2+2\gamma}\pi_u^k(dx) \leq C(t-s). \tag{5.61}$$

Therefore, by Kolmogorov's continuity criterion, there exist $G_{k,t_{\mathrm{fin}}} \in L^2$, with $\mathbb{E}[G_{k,t_{\mathrm{fin}}}^2]$ bounded uniformly in k, such that $|X_t^k - X_s^k| \leq G_{k,t_{\mathrm{fin}}}|t-s|^{1/4}$ for all $s,t \leq t_{\mathrm{fin}}$. It now follows, using the classical Arzelà-Ascoli theorem, that the laws of $(X_t^k)_{0\leq t\leq t_{\mathrm{fin}}}$ are tight on $C([0,t_{\mathrm{fin}}],\mathbb{R}^3 \times \mathbb{R}^3)$, and hence tight in the local uniform topology of $C([0,\infty),\mathbb{R}^3 \times \mathbb{R}^3)$. Therefore, using Prohorov's theorem, we can find a (not relabelled) subsequence under which X^k converges in distribution to a random variable X in the local uniform topology of $C([0,\infty),\mathbb{R}^3 \times \mathbb{R}^3)$. Finally, using Skorokhod's representation theorem, we can find random variables $X^{k,\prime}, X'$ with the same laws as X^k, X, with almost sure convergence in the local uniform topology; by an abuse of notation, we omit the $'$ to ease notation. In this way, we gain almost sure convergence, but X^k are no longer driven by the same white noise W.

Step 5: Identification of the Limit We next analyse the limiting process $X_t = (V_t, \tilde{V}_t)$ constructed above. First, for any $p \geq 0$, we have $\mathbb{E}[|X^k_t|^p]$ bounded, uniformly in k and locally uniformly in $t \geq 0$, and taking the limit implies that $\mathbb{E}[|X_t|^p] < \infty$, uniformly on compact time intervals. Next, for any given $f \in C^2_p(\mathbb{R}^3 \times \mathbb{R}^3)$, we observe that, for some $C = C_f$ and some p,

$$|\mathcal{A}_k f(x,y)| \leq C_f(1 + |x|^p + |y|^p) \tag{5.62}$$

and that $\mathcal{A}_k f \to \mathcal{A} f$, uniformly in compact subsets of $\mathbb{R}^3 \times \mathbb{R}^3$. By a truncation argument and using Lemma 4.14, it therefore follows that $\int_{\mathbb{R}^3\times\mathbb{R}^3} \mathcal{A}_k f(x,y)\pi^k_t(dy) \to \int_{\mathbb{R}^3\times\mathbb{R}^3} \mathcal{A} f(x,y)\pi_t(dy)$, uniformly on compact regions in $\mathbb{R}^3$ and on compact time intervals; see Lemma 5.5 for a similar argument with fewer moment estimates. It follows that

$$\int_{\mathbb{R}^3\times\mathbb{R}^3} \mathcal{A}_k f(X^k, y)\pi^k(dy) \to \int_{\mathbb{R}^3\times\mathbb{R}^3} \mathcal{A} f(X, y)\pi(dy) \tag{5.63}$$

uniformly on compact time intervals, almost surely. Moreover, for a new constant C_f, we have

$$\left|\int_{\mathbb{R}^3\times\mathbb{R}^3} \mathcal{A}_k f(x,y)\pi^k(dy)\right| \leq C_f(1 + |x|^p) \tag{5.64}$$

uniformly in k, and similarly for $\int_{\mathbb{R}^3\times\mathbb{R}^3} \mathcal{A} f(x,y)\pi(dy)$. We can therefore use a further truncation argument to see that (5.63) also holds for convergence in $L^1(\mathbb{P})$, uniformly on compact time intervals. We then take the limit of the martingales

$$M^{f,k}_t = f(X^k_t) - f(X^k_0) - \int_0^t \int_{\mathbb{R}^3\times\mathbb{R}^3} \mathcal{A}_k f(X^k_s, y)\pi^k_s(dy)ds \tag{5.65}$$

found in step 3 to see that

$$M^f_t = f(X_t) - f(X_0) - \int_0^t \int_{\mathbb{R}^3\times\mathbb{R}^3} \mathcal{A} f(X_s, y)\pi_s(dy)ds \tag{5.66}$$

is a martingale, and correspondingly (5.51) holds. As in Lemma 5.5, the conclusion that the desired form (5.48) holds for some white noise N of the correct covariance is exactly the result [71, Proposition 4.1], which shows that the martingale $M_t = X_t - \int_0^t \int_{\mathbb{R}^3\times\mathbb{R}^3} B(X_s - y)\pi_s(dy)ds$ can be disintegrating into the correct white noise integral.

Step 6: The constructed process couples the given solutions Finally, we address the claim that π_t is a coupling $\pi_t \in \Pi(\mu_t, \nu_t)$. Let us write $\tilde{\mu}_t, \tilde{\nu}_t$ for the two marginals of π_t.

For any $f_1 \in C_b^2(\mathbb{R}^3)$, we set $f(v,\tilde{v}) = f_1(v)$ and observe that $\mathcal{A}f(v,\tilde{v},v_*,\tilde{v}_*) = \mathcal{L}f_1(v,v_*)$, so that (5.51) tells us that

$$\int_{\mathbb{R}^3} f_1(v)\tilde{\mu}_t(dv) = \int_{\mathbb{R}^3} f_1(v)\mu_0(dv) + \int_0^t \int_{\mathbb{R}^3} \mathcal{L}f_1(v,v_*)\tilde{\mu}_s(dv_*)\tilde{\mu}_s(dv)ds.$$

Moreover, using the bounds on higher moments, we can take limits of the conservation of energy $\mathbb{E}[|V_t^k|^2] = \mathbb{E}[|V_0^k|^2]$ to obtain $\Lambda_2(\tilde{\mu}_t) = \Lambda_2(\tilde{\mu}_0)$ for all $t \geq 0$. Therefore, $(\tilde{\mu}_t)_{t\geq 0}$ is a weak solution to (LE) which starts at μ_0. Since μ_0 is assumed to have a Gaussian moment, the uniqueness result Proposition 5.8 applies and so $(\tilde{\mu}_t)_{t\geq 0} = (\mu_t)_{t\geq 0}$ as desired. The argument that $(\tilde{\nu}_t)_{t\geq 0} = (\nu_t)_{t\geq 0}$ is identical. □

We now give sketch the important points of the proof of Proposition 5.8, which we deferred earlier. We discuss the main points of [88, Theorem 2].

Sketch Proof of Proposition 5.8. Let us fix $\mu_0 \in \mathcal{S}$ with $\langle e^{|v|^\alpha}, \mu_0\rangle < \infty$, for some $\alpha > \gamma$. First, we follow Steps 1-5 of the previous proof (with some arbitrary ν_0, or following the arguments verbatim for processes in $\mathbb{R}^3$) to find *a* Landau process $(V_t)_{t\geq 0}$ solving (stLE) and with $\mathrm{Law}(V_0) = \mu_0$. Setting $\mu_t = \mathrm{Law}(V_t)$, we get that $\mu \in L^1_{\mathrm{loc}}([0,\infty), \mathcal{S}^{2+\gamma})$ and is a weak solution to (LE). We claim that this solution is unique; let $(\nu_t)_{t\geq 0}$ be any other solution with $\nu_0 = \mu_0$.

Step 1: Itô representation of (stLE) Since we will not use the full strength of the white-noise representation of (stLE), it is convenient to rewrite it as a SDE driven by a Brownian motion. We first define averaged coefficients; for $v \in \mathbb{R}^3$ and $\mu \in \mathcal{P}_{2+\gamma}(\mathbb{R}^3)$, define

$$b(v,\mu) = \int_{\mathbb{R}^3} b(v-v_*)\mu(dv_*), \qquad a(v,\mu) = \int_{\mathbb{R}^3} a(v-v_*)\mu(dv_*)$$

and let $\sigma(v,\mu)$ be a square root of $a(v,\mu)$. Now, using the martingale representation theorem, it follows that V_t can be written as

$$V_t = V_0 + \int_0^t b(V_s,\mu_s)ds + \int_0^t \sigma(V_s,\mu_s)d\beta_s \tag{5.67}$$

for a 3-dimensional Brownian motion β_t.

Step 2: Tanaka Process for ν_t We now build a similar representation of ν_t. We set $W_0 = V_0$, recalling that $\nu_0 = \mu_0$. For positive definite, symmetric 3×3 matrices A, B, set $U(A,B)$ to the the orthogonal matrix $U(A,B) = B^{-1/2}A^{-1/2}(A^{1/2}BA^{1/2})^{1/2}$, and for nonnegative definite matrices and $\varepsilon > 0$, set $U_\varepsilon(A,B) := U(A+\varepsilon I, B+\varepsilon I)$; it is straightforward to see that the map $(A,B) \to U_\varepsilon(A,B)$ is locally Lipschitz in the space of 3×3, nonnegative-definite matrices. We now consider, for $\varepsilon > 0$,

$$W_t^\varepsilon = W_t^\varepsilon + \int_0^t b(W_s^\varepsilon,\nu_s)ds + \int_0^t \sigma(W_s^\varepsilon,\nu_s)U_\varepsilon(a(V_s,\mu_s), a(W_s^\varepsilon,\nu_s))d\beta_s.$$

Since the coefficients are locally Lipschitz, this SDE has pathwise unique, maximal solutions, and using the orthogonality of U_ε, the process $\widehat{\beta}^\varepsilon_t = \int_0^t U_\varepsilon(a(V_s,\mu_s), a(W^\varepsilon_s,\nu_s))d\beta_s$ is again a 3D-Brownian motion, so this is equivalent to (5.67) with ν_s in place of μ_s. In particular, the solution is globally defined, and by proving uniqueness for the *linear* evolution equation for $\mathrm{Law}(W^\varepsilon_t)$ (see [88, Proposition 10]), one finds that $\mathrm{Law}(W^\varepsilon_t) = \nu_t$.

Step 3: Tanaka Estimate Next, we derive an estimate for $u^\varepsilon_t = \mathbb{E}[|V_t - W^\varepsilon_t|^2]$. For $\pi \in \mathcal{P}_{2+\gamma}(\mathbb{R}^3 \times \mathbb{R}^3)$ and $\mu, \nu \in \mathcal{P}_{2+\gamma}(\mathbb{R}^3)$, we set $\mathcal{E}_\varepsilon(\pi,\mu,\nu)$ to be

$$\mathcal{E}_\varepsilon(\pi,\mu,\nu) = \int_{\mathbb{R}^3\times\mathbb{R}^3}\Big(\|\sigma(v,\mu) - \sigma(w,\nu)U_\varepsilon(\sigma(v,\mu),\sigma(w,\nu))\|^2 \\ + 2(v-w)\cdot(b(v,\mu)-b(w,\nu))\Big)\pi(dv,dw).$$

If we now set $\pi^\varepsilon_t = \mathrm{Law}(V_t, W^\varepsilon_t)$, it follows from the Itô calculus of (5.67) that $\frac{d}{dt}u^\varepsilon_t = \mathcal{E}_\varepsilon(\pi_t,\mu_t,\nu_t)$, see also the derivation of Lemma 5.9 below, where we repeat this argument with our weighted distances. For this case, one finds [88, Proposition 12] that, for some $\kappa > 0$ and some C, and any $M > 0$,

$$\mathcal{E}_\varepsilon(\pi_t,\mu_t,\nu_t) \le C\sqrt{\varepsilon}(1+\Lambda_{2+\gamma}(\mu_t,\nu_t)) + Mu^\varepsilon_t + C\left(1+\Lambda_{2+\gamma}(\nu_t) + \langle e^{|v|^\alpha},\mu_t\rangle\right)e^{-\kappa M^{\alpha/\gamma}}.$$

In the same spirit as Proposition 2.13i), the exponential moments on μ_t are propagated to later times, and by Proposition 2.12, all polynomial moments are finite, bounded in terms of the moments of μ_0. Recalling that $u^\varepsilon_0 = 0$, since $\nu_0 = \mu_0$, we absorb all moments into a constant C and find by Grönwall's Lemma that

$$u^\varepsilon_t \le Ce^{Mt}\left(\sqrt{\varepsilon} + te^{-\kappa M^{\alpha/\gamma}}\right).$$

Since π_t is a coupling of μ_t, ν_t, the right-hand side is an upper bound for $\mathcal{W}_2^2(\mu_t,\nu_t)$. Taking $\varepsilon \to 0$ with M fixed, the first term converges to 0, and then taking $M \to \infty$, the second term converges to 0, since $\kappa > 0$ and $\alpha > \gamma$. We conclude that $\mathcal{W}_2(\mu_t,\nu_t) = 0$ for all $t \ge 0$, and the uniqueness is complete. ☐

We finally state the following central inequality, which will play the same rôle for the Landau case as Lemma 4.10 does for the Boltzmann case. In terms of the proof above, we refine the estimates on $\mathcal{E}$, now working in terms of the white-noise representation, and using our weighted distance, to avoid the compensation using exponential moments.

Lemma 5.9. *Define, for the coupled generator $\mathcal{A}$ defined at (5.50) in Proposition 5.6,*

$$\mathcal{E}_{p,\varepsilon}(v,v_*,\tilde{v},\tilde{v}_*) = \mathcal{A}f_{p,\varepsilon}(v,v_*,\tilde{v},\tilde{v}_*) \qquad f_{p,\varepsilon}(v,\tilde{v}) = d_{p,\varepsilon}(v,\tilde{v}). \tag{5.68}$$

Then there is a constant C, depending only on $p \geq 2$ and $\gamma \in (0,1]$, such that for all $\varepsilon \in (0,1]$, all $v, v_*, \tilde{v}, \tilde{v}_* \in \mathbb{R}^3$,

$$\begin{aligned}\mathcal{E}_{p,\varepsilon}(v, v_*, \tilde{v}, \tilde{v}_*) \leq & [2-p] d_{p+\gamma,\varepsilon}(v,\tilde{v}) \\ & + C\sqrt{\varepsilon}(1+|v_*|^p + |\tilde{v}_*|^p) d_{p+\gamma,\varepsilon}(v,\tilde{v}) \\ & + C\sqrt{\varepsilon}(1+|v|^p + |\tilde{v}|^p) d_{p+\gamma,\varepsilon}(v_*,\tilde{v}_*) \\ & + \frac{C}{\sqrt{\varepsilon}}(1+|v_*|^{p+\gamma} + |\tilde{v}_*|^{p+\gamma}) d_{p,\varepsilon}(v,\tilde{v}) \\ & + \frac{C}{\sqrt{\varepsilon}}(1+|v|^{p+\gamma} + |\tilde{v}|^{p+\gamma}) d_{p,\varepsilon}(v_*,\tilde{v}_*).\end{aligned}$$

We recognise the same fundamental structure as in Lemma 4.10, with the appearance of terms higher order terms $d_{p+\gamma,\varepsilon}(v,\tilde{v}), d_{p+\gamma,\varepsilon}(v_*,\tilde{v}_*)$ which prevent a Grönwall estimate, a negative 'Povzner term', and lower-order cross-terms. Correspondingly, the strategy is similar: we use the negative Povzner term to cancel the terms in the second and third lines, using a symmetry argument, and the estimate using only the remaining term closes to lead to a Grönwall argument. In this case, thanks to the exact Itô calculus, we can find explicit, rather than the explicit*able* constant c in Lemma 4.10. We will also have to choose $\varepsilon > 0$ small, depending on p and the p^{th} moments of μ_t, ν_t, in order to ensure the cancellation still holds. Relative to Chapter 4, this additional complication arises because we work with $d_{p,\varepsilon}$ rather than $d_p = d_{p,0}$, which is necessary for the argument to hold with only $2+\epsilon$ moments.

Let us mention that a rather direct computation, with $\varepsilon = 0$, i.e. with the cost $d_{p,0}(v,\tilde{v}) = (1+|v|^p + |\tilde{v}|^p)|v-\tilde{v}|^2$, relying on the simple estimates (5.6), (5.8) and (5.9), shows that

$$\begin{aligned}\mathcal{E}_{p,0}(v, v_*, \tilde{v}, \tilde{v}_*) \leq & [32-p] d_{p+\gamma,0}(v,\tilde{v}) + C(1+|v_*|^{p+\gamma} + |\tilde{v}_*|^{p+\gamma}) d_{p,0}(v,\tilde{v}) \\ & + C(1+|v|^{p+\gamma} + |\tilde{v}|^{p+\gamma}) d_{p,0}(v_*,\tilde{v}_*).\end{aligned}$$

The same arguments as in Chapter 4 then give a stability result in $(\mathcal{P}_{45}, w_{32,0})$. Our estimate is rather finer; we have to be very careful and to use many cancelations to replace the Povzner term $[32-p]$ by $[2-p]$. Moreover, we have to deal with $d_{p,\varepsilon}$ with $\varepsilon > 0$ instead of $d_{p,0}$, because $w_{p,0}$ requires moments of order $p+2$ to be well-defined. All this is crucial to obtain a stability result in $\mathcal{P}_p(\mathbb{R}^3)$, for any $p > 2$. Since the proof is rather lengthy, it is deferred to Section 5.9 for the ease of readability.

5.5 Proof of Theorem 5.1

We now give the proof of our stability estimate. As with Lemma 4.21 in Section 4.6, we first deal with the case when the initial data have a finite exponential (in this case: Gaussian) moment, and then carefully relax this assumption.

Lemma 5.10. *Fix $\gamma \in (0,1]$ and let $(\mu_t)_{t\geq 0}, (\nu_t)_{t\geq 0} \subset \mathcal{S}$ be weak solutions to (LE) with initial moments $\int_{\mathbb{R}^3} e^{a|v|^2}(\mu_0+\nu_0)(dv) < \infty$ for some $a > 0$. Then there exists $(V_t)_{t\geq 0}, (\tilde{V}_t)_{t\geq 0}$ solving (stLE), such that $\pi_t = \mathrm{Law}(V_t, \tilde{V}_t) \in \Pi(\mu_t, \nu_t)$ for any $t \geq 0$, and such that (5.1) holds.*

Proof. We fix $p > 2$, consider $\varepsilon \in (0,1]$ to be chosen later and introduce $\pi_0 \in \Pi(\mu_0, \nu_0)$ such that

$$w_{p,\varepsilon}(\mu_0,\nu_0) = \int_{\mathbb{R}^3\times\mathbb{R}^3} d_{p,\varepsilon}(v,\tilde{v})\pi_0(dv,d\tilde{v}).$$

Since we have a Gaussian moment condition, we can apply Lemma 5.6 to construct $(V_t, \tilde{V}_t)_{t\geq 0}$ using and remark that these individually solve (stLE), and $\pi_t = \mathrm{Law}(V_t,\tilde{V}_t) \in \Pi(\mu_t,\nu_t)$ by the cited Lemma. We now define

$$u_\varepsilon(t) = \mathbb{E}[d_{p,\varepsilon}(V_t,\tilde{V}_t)] = \int_{\mathbb{R}^3\times\mathbb{R}^3} d_{p,\varepsilon}(v,\tilde{v})\pi_t(dv,d\tilde{v}). \tag{5.69}$$

By Lemma 5.6, and since $u_\varepsilon(0) = w_{p,\varepsilon}(\mu_0,\nu_0)$, it holds that for all $t \geq 0$,

$$u_\varepsilon(t) = w_{p,\varepsilon}(\mu_0,\nu_0) + \int_0^t \int_{\mathbb{R}^3\times\mathbb{R}^3}\int_{\mathbb{R}^3\times\mathbb{R}^3} \mathcal{E}_{p,\varepsilon}(v,v_*,\tilde{v},\tilde{v}_*)\pi_s(dv_*,d\tilde{v}_*)\pi_s(dv,d\tilde{v})ds.$$

Using next Lemma 5.9 and a symmetry argument, we find that

$$u_\varepsilon(t) \leq w_{p,\varepsilon}(\mu_0,\nu_0) + \int_0^t (I_{1,\varepsilon}(s) + I_{2,\varepsilon}(s) + I_{3,\varepsilon}(s))ds,$$

where, for some constant $C > 0$ depending only on p and γ,

$$\begin{aligned}
I_{1,\varepsilon}(s) =& [2-p]\int_{\mathbb{R}^3\times\mathbb{R}^3} d_{p+\gamma,\varepsilon}(v,\tilde{v})\pi_s(dv,d\tilde{v}),\\
I_{2,\varepsilon}(s) =& C\sqrt{\varepsilon}\int_{\mathbb{R}^3\times\mathbb{R}^3}\int_{\mathbb{R}^3\times\mathbb{R}^3}(1+|v_*|^p+|\tilde{v}_*|^p)d_{p+\gamma,\varepsilon}(v,\tilde{v})\pi_s(dv_*,d\tilde{v}_*)\pi_s(dv,d\tilde{v}),\\
I_{3,\varepsilon}(s) =& \frac{C}{\sqrt{\varepsilon}}\int_{\mathbb{R}^3\times\mathbb{R}^3}\int_{\mathbb{R}^3\times\mathbb{R}^3}(1+|v_*|^{p+\gamma}+|\tilde{v}_*|^{p+\gamma})d_{p,\varepsilon}(v,\tilde{v})\pi_s(dv_*,d\tilde{v}_*)\pi_s(dv,d\tilde{v}).
\end{aligned}$$

Using that $\pi_s \in \Pi(\mu_s,\nu_s)$, we use the moment estimates in Proposition 2.12 to see that, up to a new choice of C,

$$\begin{aligned}
I_{2,\varepsilon}(s) \leq& C\sqrt{\varepsilon}(1+\Lambda_p(\mu_s,\nu_s))\int_{\mathbb{R}^3\times\mathbb{R}^3} d_{p+\gamma,\varepsilon}(v,\tilde{v})\pi_s(dv,d\tilde{v})\\
&\leq C\sqrt{\varepsilon}(1+\Lambda_p(\mu_0+\nu_0))\int_{\mathbb{R}^3\times\mathbb{R}^3} d_{p+\gamma,\varepsilon}(v,\tilde{v})\pi_s(dv,d\tilde{v});
\end{aligned}$$

and

$$I_{3,\varepsilon}(s) \le \frac{C}{\sqrt{\varepsilon}}(1+\Lambda_{p+\gamma}(\mu_s,\nu_s))u_\varepsilon(s).$$

We now fix $t>0$ and work on $[0,t]$. Choosing, for the value of C in the previous two displays,

$$\varepsilon = \Big[\frac{p-2}{p-2+C(1+\Lambda_p(\mu_0,\nu_0))}\Big]^2$$

we have $\varepsilon\in(0,1]$, and we can absorb $I_{2,\varepsilon}$ into the Povzner term so that $I_{1,\varepsilon}(s)+I_{2,\varepsilon}(s)\le 0$ for all $s\in[0,t]$. We now have the integral inequality

$$u_\varepsilon(r)\le w_{p,\varepsilon}(\mu_0,\nu_0)+\frac{C}{\sqrt{\varepsilon}}\int_0^r(1+\Lambda_{p+\gamma}(\mu_s,\nu_s))u_\varepsilon(s)ds$$

for all $r\in[0,t]$, and we apply the Grönwall lemma and Proposition 2.12ii) to obtain, again changing C,

$$\begin{aligned}u_\varepsilon(t)&\le w_{p,\varepsilon}(\mu_0,\nu_0)\exp\Big(\frac{C}{\sqrt{\varepsilon}}\int_0^t(1+\Lambda_{p+\gamma}(\mu_s,\nu_s))ds\Big)\\&\le w_{p,\varepsilon}(\mu_0,\nu_0)\exp\Big(\frac{C}{\sqrt{\varepsilon}}(1+t)\Lambda_p(\mu_0)\Big).\end{aligned}$$

We now convert everything back to the distances $d_{p,1},w_{p,1}$. Recalling that $d_{p,1}\le d_{p,\varepsilon}\le \varepsilon^{-1}d_{p,1}$, we deduce that

$$u_1(t)\le u_\varepsilon(t)\quad\text{and}\quad w_{p,\varepsilon}(\mu_0,\nu_0)\le\frac{1}{\varepsilon}w_{p,1}(\mu_0,\nu_0).$$

We thus end with

$$\mathbb{E}[d_{p,1}(V_t,\tilde{V}_t)]\le\frac{1}{\varepsilon}w_{p,1}(\mu_0,\nu_0)\exp\Big(\frac{C}{\sqrt{\varepsilon}}\Lambda_p(\mu_0)(1+t)\Big).$$

Recalling our choice for ε and allowing the value of C, still depending only on p and γ, to change from line to line, we find that

$$\begin{aligned}&\mathbb{E}[d_{p,1}(V_t,\tilde{V}_t)]\\&\quad\le C(1+\Lambda_p(\mu_0,\nu_0))^2w_{p,1}(\mu_0,\nu_0)\exp\Big(C(1+\Lambda_p(\mu_0,\nu_0))\Lambda_p(\mu_0,\nu_0)(1+t)\Big)\\&\quad\le w_p(\mu_0,\nu_0)\exp\Big(C\Lambda_p(\mu_0,\nu_0)^2(1+t)\Big)\end{aligned}$$

where, in the final line, we changed C again to move the moment prefactor into the exponent. This was our goal, and the lemma is complete. □

In order to relax the initial Gaussian moment condition, we will use the following convergence, which replaces Lemma 4.22.

Lemma 5.11. *Fix $\gamma\in(0,1]$ and $p>2$. Let $(\mu_t)_{t\ge0}$ be a weak solution to (LE), with initial moment $\Lambda_p(\mu_0)<\infty$. Then $w_{p,1}(\mu_t,\mu_0)\to0$ as $t\to0$.*

Proof. First, thanks to the density of $C_b^2(\mathbb{R}^3)$ in $C_b(\mathbb{R}^3)$, we deduce from (1.8) that $\mu_t \to \mu_0$ weakly. It classically follows that $\lim_{t\to 0} \mathcal{W}_{1,1}(\mu_t, \mu_0) = 0$, where $\mathcal{W}_{1,1}$ is the (Monge-Kantorovich-)Wasserstein distance defined in Section 2.1. Moreover, as remarked in Section 2.1, there exists $\rho_t \in \Pi(\mu_0, \mu_t)$ attaining $\mathcal{W}_{1,1}(\mu_t, \mu_0) = \int_{\mathbb{R}^3\times\mathbb{R}^3}(1 \wedge |v-w|)\rho_t(dv, dw)$. Now, fix $\epsilon > 0$; by Lemma 2.15, there exists $R < \infty$ and $t_0 > 0$ such that

$$\int_{\mathbb{R}^3} (1 + |v|^p)1\!\!1_{\{|v|>R\}}\mu_t(dv) < \epsilon \quad \text{for all } t \in [0, t_0].$$

Since now

$$d_{p,1}(v, w) \le (1 + |v|^p + |w|^p)(|v - w| \wedge 1) \le (1 + |v|^p)(|v - w| \wedge 1) + (1 + |w|^p)(|v - w| \wedge 1)$$

we have

$$\begin{aligned} w_{p,1}(\mu_t, \mu_0) \le& \int_{\mathbb{R}^3\times\mathbb{R}^3} d_{p,1}(v, w)\rho_t(dv, dw) \\ \le&(1 + R^p)\mathcal{W}_{1,1}(\mu_t, \mu_0) + \int_{\mathbb{R}^3\times\mathbb{R}^3} (1 + |v|^p)1\!\!1_{\{|v|>R\}}\rho_t(dv, dw) \\ &+ (1 + R^p)\mathcal{W}_{1,1}(\mu_t, \mu_0) + \int_{\mathbb{R}^3\times\mathbb{R}^3} (1 + |w|^p)1\!\!1_{\{|w|>R\}}\rho_t(dv, dw) \\ =&2(1 + R^p)\mathcal{W}_{1,1}(\mu_t, \mu_0) + \int_{\mathbb{R}^3} (1 + |v|^p)1\!\!1_{\{|v|>R\}}\mu_t(dv) \\ &+ \int_{\mathbb{R}^3} (1 + |w|^p)1\!\!1_{\{|w|>R\}}\mu_0(dw), \end{aligned}$$

where in the last line we recall $\rho_t \in \Pi(\mu_t, \mu_0)$. We conclude that for all $t \in [0, t_0]$,

$$w_{p,1}(\mu_t, \mu_0) \le 2(1 + R^p)\mathcal{W}_{1,1}(\mu_t, \mu_0) + 2\varepsilon,$$

so that $\limsup_{t\to 0} w_{p,1}(\mu_t, \mu_0) \le 2\varepsilon$ and we are done, as $\epsilon > 0$ was arbitrary. □

The strategy of proving the full statement of Theorem 5.1 is now as follows. We fix $p > 2$ and two solutions $(\mu_t)_{t\ge 0}, (\nu_t)_{t\ge 0} \subset \mathcal{S}$ to (LE) with finite p^{th} moments. By the Gaussian moment creation Proposition 2.13, we can apply Lemma 5.10 by considering instead $(\mu_{s+t})_{t\ge 0}, (\nu_{s+t})_{t\ge 0}$, which produces a coupling $(V_t^s, \tilde{V}_t^s)_{t\ge 0}$. We then construct the coupling $(V_t, \tilde{V}_t)_{t\ge 0}$ as a weak limit point, using Lemma 5.5.

We are now ready to remove the additional assumptions and prove the full stability and coupling statement.

Proof of Theorem 5.1. We fix $\gamma \in (0, 1]$, $p > 2$ and we consider two weak solutions $(\mu_t)_{t\ge 0}$ and $(\nu_t)_{t\ge 0}$ to (LE) such that $\Lambda_p(\mu_0, \nu_0) < \infty$.

Step 1: Approximation Procedure Fix $t > 0$ and let $0 < s \le 1$; thanks to Proposition 2.13, we have $\int_{\mathbb{R}^3} e^{a|v|^2}(\mu_s + \nu_s)(dv) < \infty$ for some $a > 0$. Lemma 5.10 therefore applies to $(\mu_u)_{u\ge s}, (\nu_u)_{u\ge s}$, so that there exists a coupling $(V^s_t, \tilde{V}^s_t)_{t\ge 0}$, where each coordinate solves (stLE), with $\pi^s_t = \text{Law}(V^s_t, \tilde{V}^s_t) \in \Pi(\mu_{t+s}, \nu_{t+s})$, and such that

$$\begin{aligned}\mathbb{E}[d_{p,1}(V^s_t, \tilde{V}^s_t)] \le & w_{p,1}(\mu_s, \nu_s) \exp\left(C\Lambda_p(\mu_s, \nu_s)^2(1+t)\right) \\ \le & w_{p,1}(\mu_s, \nu_s) \exp\left(C\Lambda_p(\mu_0, \nu_0)^2(1+t)\right).\end{aligned} \tag{5.70}$$

Step 2: Extraction of a Subsequence We now show that the compactness result Lemma 5.5 applies to each coordinate $(V^{1/n}_t)_{t\ge 0}, (\tilde{V}^{1/n}_t)_{t\ge 0}$, with $\mu^n_t = \mu_{t+1/n}, \nu^n_t = \nu_{t+1/n}$ respectively. For either coordinate, the first uniform integrability result (5.10) follows immediately from Proposition 2.15. For (5.11), for any t,

$$\sup_n \int_0^t \int_{\mathbb{R}^3} (1+|v|^{2+\gamma})\mathbb{1}_{|v|\ge R}\mu^n_t(dv) \le \int_0^{t+1} \int_{\mathbb{R}^3} (1+|v|^{2+\gamma})\mathbb{1}_{|v|\ge R}\mu^n_t(dv) \to 0$$

as $R \to \infty$, using the fact that $\int_0^{t+1} \Lambda_{2+\gamma}(\mu_u)du < \infty$ and the dominated convergence theorem; the case for ν^n_t is identical. It therefore follows that $(V^{s_n}_t, \tilde{V}^{s_n}_t)_{t\ge 0}$ is also tight, so we can find a sequence $s_n \to 0$ under which $(V^{s_n}_t, \tilde{V}_t)_{t\ge 0}$ converges in distribution to a limiting process $(V_t, \tilde{V}_t)_{t\ge 0}$. It follows that both $(V^{s_n}_t)_{t\ge 0} \to (V_t)_{t\ge 0}, (\tilde{V}^{s_n}_t)_{t\ge 0} \to (\tilde{V}_t)_{t\ge 0}$ in distribution, so we apply the second part of Lemma 5.5 to conclude that $(V_t)_{t\ge 0}, (\tilde{V}_t)_{t\ge 0}$ solve (stLE) for some choices of white noise as desired.

Step 3: Conclusion We now show that the limit point $(V_t, \tilde{V}_t)_{t\ge 0}$ found above satisfies all the desired properties. Let us fix $t \ge 0$; for all n, $\text{Law}(V^{s_n}_t) = \mu_{t+s_n}$ and, as already remarked in Lemma 5.11, we have $\mu_{t+s_n} \to \mu_t$ in the weak topology, so we take the limit to conclude that $\text{Law}(V_t) = \mu_t$. The same argument applies to $\tilde{V}_t$, and so we see that $\pi_t = \text{Law}(V_t, \tilde{V}_t) \in \Pi(\mu_t, \nu_t)$ as desired. For the bound (5.1), we recall the relaxed triangle inequality (2.19), we have, for some constant C depending only on p,

$$w_{p,1}(\mu_s, \nu_s) \le C[w_{p,1}(\mu_s, \mu_0) + w_{p,1}(\mu_0, \nu_0) + w_{p,1}(\nu_0, \nu_s)]$$

and as $s \to 0$, the first and third terms converge to 0 by Lemma 5.11, so

$$\limsup_{s\to 0} w_{p,1}(\mu_s, \nu_s) \le C w_{p,1}(\mu_0, \nu_0).$$

An identical argument for the couplings $\pi^{s_n}_t = \text{Law}(V^{s_n}_t, \tilde{V}^{s_n}_t)$ shows that $\mathbb{E}[d_{p,1}(V^{s_n}_t, \tilde{V}^{s_n}_t)] \to \mathbb{E}[d_{p,1}(V_t, \tilde{V}_t)]$, and we can take the limit in (5.70) to obtain the desired result. □

5.6 Equivalence of The Landau Equation and Landau Processes

We now prove Theorem 5.2. As remarked in the introduction, it is already clear, by applying Itô's formula as in Lemma 5.6, that if $(V_t)_{t\geq 0}$ is a solution of (stLE) such that the laws μ_t belong to $L^1_{\text{loc}}([0,\infty), \mathcal{S}^{2+\gamma})$, then $(\mu_t)_{t\geq 0}$ is a weak solution to (LE). Let us now apply the results we have already obtained to prove the other implication, following exactly the same argument as Theorem 4.3 in the Boltzmann case.

Proof of Theorem 5.2. Let $(\mu_t)_{t\geq 0} \subset \mathcal{S}$ be any weak solution to (LE). For all $s > 0$, μ_s has Gaussian moment, and in particular, we can apply Lemma 5.6 to construct a solution $(V^s_t)_{t\geq 0}$ to (stLE) with $\text{Law}(V^s_t) = \mu_{s+t}$ for all $t \geq 0$ - either by taking ν_t to be an arbitrary solution to (LE), or by repeating the arguments of Lemma 5.6 verbatim. We apply Lemma 5.5 exactly as in Step 2 of the proof of Theorem 5.1 to find $s_n \to 0$ and $(V_t)_{t\geq 0}$ solving (stLE), which is the limit in distribution of $(V^{s_n}_t)_{t\geq 0}$. The conclusion that $\mu_t = \text{Law}(V_t)$ is exactly as in Step 3 of the proof of Theorem 5.1, and we are done. □

5.7 Proof of Theorem 5.3

Our starting point for existence is the following result, due to Desvillettes and Villani, which proves existence with a p^{th} moment, for any $p > 2$.

Proposition 5.12. *Suppose that $\mu_0 \in \mathcal{P}_p(\mathbb{R}^3)$ for some $p > 2$. Then a weak solution starting at μ_0 exists.*

The cited result also assumes that μ_0 has a density, but this is only used to show that the solution constructed also has a density: the construction of a solution remains valid without this assumption. The novelty of Theorem 5.3 lies in relaxing the additional moment requirement, which we replace by the de La Vallée Poussin theorem to achieve control sharper than Λ_2 without additional assumptions on the initial data.

Proof of Theorem 5.3. Let us start from $\mu_0 \in \mathcal{S}$. By the de La Vallée Poussin theorem, there exists a C^2-function $h : [0,\infty) \to [0,\infty)$ such that $h'' \geq 0$, $h'(\infty) = \infty$ and

$$\int_{\mathbb{R}^3} h(|v|^2)\mu_0(dv) < \infty. \tag{5.71}$$

We can also impose that $h'' \leq 1$ and that $h'(0) = 1$, and argue similarly to Step 3 of Proposition 2.8.

Step 1: Approximation Procedure We consider $n_0 \geq 1$ such that for all $n \geq n_0$, $\alpha_n = \int_{\mathbb{R}^3} 1\!\!1_{\{|v|\leq n\}}\mu_0(dv) \geq 1/2$ and set, for $n \geq n_0$,

$$\mu_0^n(dv) = \alpha_n^{-1} 1\!\!1_{\{|v|\leq n\}}\mu_0(dv) \in \mathcal{P}(\mathbb{R}^3).$$

Since μ_0^n is compactly supported, it has all moments finite and there exists a weak solution $(\mu_t^n)_{t\geq 0}$ to (LE) starting at μ_0^n by Proposition 5.12. Of course, μ_0^n converges weakly to μ_0 as $n \to \infty$, and $\Lambda_2(\mu_0^n) \to 1$. Moreover, we can use Theorem 5.2 to find $(V_t^n)_{t\geq 0}$ solving (stLE) with $\text{Law}(V_t^n) = \mu_t^n$ for all $t \geq 0$.

Step 2: A Uniform Integrability Property We now show a uniform integrability property for the solutions μ_t^n found above, using the hypothesised function h. By Proposition 2.12, all polynomial moments of μ_t^n are bounded, uniformly in $t \geq 0$ (but not necessarily in n). We can therefore apply the weak formulation (1.8) of (LE) to the function $f(v) = h(|v|^2)$: arguing as in (2.126),

$$\partial_k f(v) = 2v_k h'(|v|^2); \qquad \partial^2_{k\ell} f(|v|^2) = 2h'(|v|^2) 1\!\!1_{\{k=\ell\}} + 4v_k v_l h''(|v|^2)$$

and so, setting $x = v - v_*$ as usual,

$$\mathcal{L}f(v, v_*) = h'(|v|^2)[2v \cdot b(x) + \text{Tr}a(x)] + 2|\sigma(x)v|^2 h''(|v|^2).$$

Recalling (5.9) and that $0 \leq h'' \leq 1$, we bound the second order term by

$$2|\sigma(x)v|^2 h''(|v|^2) \leq C|x|^\gamma |v|^2 |v_*|^2 \leq C(|v|^{2+\gamma}|v_*|^2 + |v|^2|v_*|^{2+\gamma}).$$

Meanwhile, since $b(x) = -2|x|^\gamma x$ and $\text{Tr}a(x) = 2|x|^{\gamma+2}$, the first term is

$$\begin{aligned} h'(|v|^2)[2v \cdot b(x) + \text{Tr}a(x)] &= 2h'(|v|^2)[-|x|^\gamma|v|^2 + |x|^\gamma|v_*|^2] \\ &\leq -2h'(|v|^2)|v|^{2+\gamma} + 2h'(|v|^2)|v_*|^\gamma|v|^2 + 2h'(|v|^2)|v|^\gamma|v_*|^2 + 2h'(|v|^2)|v_*|^{2+\gamma} \\ &\leq -h'(|v|^2)|v|^{2+\gamma} + C(1+|v|^2)|v_*|^{\gamma+2}. \end{aligned}$$

We used that $|x|^\gamma \geq |v|^\gamma - |v_*|^\gamma$, that $|x|^\gamma \leq |v|^\gamma - |v_*|^\gamma$ and, for the last inequality, that there is $C > 0$ such that $|v_*|^\gamma|v|^2 + |v|^\gamma|v_*|^2 \leq \frac{1}{2}|v|^{2+\gamma} + C|v_*|^{2+\gamma}$ and that $h'(r) \leq 1 + r$. All in all,

$$\mathcal{L}f(v, v_*) \leq -h'(|v|^2)|v|^{2+\gamma} + C(1+|v|^2)|v_*|^{\gamma+2} + C(1+|v_*|^2)|v|^{\gamma+2}.$$

We thus find, by (1.8), recalling that $\Lambda_2(\mu_t^n) = \Lambda_2(\mu_0^n)$, that

$$\begin{aligned} \int_{\mathbb{R}^3} h(|v|^2)\mu_t^n(dv) + \int_0^t \int_{\mathbb{R}^3} h'(|v|^2)|v|^{2+\gamma}\mu_s^n(dv)ds \\ \leq \int_{\mathbb{R}^3} h(|v|^2)\mu_0^n(dv) + 2C(1+\Lambda_2(\mu_0^n)) \int_0^t \int_{\mathbb{R}^3} |v|^{2+\gamma}\mu_s^n(dv)ds \\ \leq 2\int_{\mathbb{R}^3} h(|v|^2)\mu_0(dv) + 6C \int_0^t \int_{\mathbb{R}^3} |v|^{2+\gamma}\mu_s^n(dv)ds, \end{aligned}$$

since $\mu_0^n \le 2\mu_0$ and $\Lambda_2(\mu_0) = 1$. But since $h'(\infty) = \infty$, there is a constant κ such that $6C|v|^{2+\gamma} \le \frac{1}{2}h'(|v|^2)|v|^{2+\gamma} + \kappa$ for all $v \in \mathbb{R}^3$. We finally get

$$\int_{\mathbb{R}^3} h(|v|^2)\mu_t^n(dv) + \frac{1}{2}\int_0^t \int_{\mathbb{R}^3} h'(|v|^2)|v|^{2+\gamma}\mu_s^n(dv)ds \le 2\int_{\mathbb{R}^3} h(|v|^2)\mu_0(dv) + \kappa t$$

and we conclude that for all $t_{\text{fin}} < \infty$, we have the uniform integrability

$$K_{t_{\text{fin}}} := \sup_n \left\{ \sup_{t\in[0,t_{\text{fin}}]} \int_{\mathbb{R}^3} h(|v|^2)\mu_t^n(dv) + \int_0^{t_{\text{fin}}} \int_{\mathbb{R}^3} |v|^{2+\gamma} h'(|v|^2)\mu_t^n(dv)dt \right\} < \infty. \tag{5.72}$$

Step 3: Construction of a Limit We now construct $(V_t)_{t\ge 0}$ solving (stLE) starting at μ_0 by using Lemma 5.5 to extract a subsequence converging in law in $n \ge n_0$. Let us fix t_{fin} and let $K_{t_{\text{fin}}}$ be the (finite) supremum appearing in (5.72). For (5.10), we have

$$\begin{aligned} \sup_{t\le t_{\text{fin}}} \sup_n \int_{\mathbb{R}^3} (1+|v|^2)1\!\!1_{|v|\ge R}\mu_t^n(dv) &\le \frac{1+R^2}{h(R^2)} \sup_{t\le t_{\text{fin}}} \sup_n \int_{\mathbb{R}^3} h(|v|^2)\mu_t^n(dv) \\ &\le \frac{1+R^2}{h(R^2)} K_{t_{\text{fin}}} \end{aligned} \tag{5.73}$$

where we note that $h(x)/x$ is increasing, by convexity of h. Since $h'(\infty) = \infty$, it follows that $h(x)/x \to \infty$ as $x \to \infty$, so the right-hand side converges to 0 as $R \to \infty$, and (5.10) is proven. For the second item, we argue similarly: we have

$$\begin{aligned} \sup_n \int_0^{t_{\text{fin}}} \int_{\mathbb{R}^3} (1+|v|^{2+\gamma})1\!\!1_{|v|\ge R}\mu_t^n(dv)dt &\le \frac{1+R^{2+\gamma}}{R^{2+\gamma}h'(R^2)} \sup_n \int_0^{t_{\text{fin}}} \int_{\mathbb{R}^3} |v|^{2+\gamma}h'(|v|^2)\mu_t^n(dv)dt \\ &\le \frac{1+R^{2+\gamma}}{R^{2+\gamma}h'(R^2)} K_{t_{\text{fin}}} \to 0 \end{aligned} \tag{5.74}$$

as $R \to \infty$. We now apply Lemma 5.5 to find $n_k \ge n_0$, $n_k \to \infty$ along which $(V_t^{n_k})_{t\ge 0}$ converge in distribution to a limit $(V_t)_{t\ge 0}$, which solves (stLE). We observe that $\text{Law}(V_0^{n_k}) = \mu_0^{n_k}$ converges to $\text{Law}(V_0)$, so we conclude that $\text{Law}(V_0) = \mu_0$ as desired. Taking the weak limits $\mu_t^{n_k} \to \mu_t$ in (5.72) and using lower semicontinuity,

$$\sup_{t\in[0,t_{\text{fin}}]} \int_{\mathbb{R}^3} h(|v|^2)\mu_t(dv) + \int_0^t \int_{\mathbb{R}^3} |v|^{2+\gamma}h'(|v|^2)\mu_t(dv)dt \le K_{t_{\text{fin}}} < \infty \tag{5.75}$$

for all $t_{\text{fin}} \ge 0$, which implies that $\mu \in L^\infty_{\text{loc}}([0,\infty), \mathcal{P}_2(\mathbb{R}^3)) \cap L^1_{\text{loc}}([0,\infty), \mathcal{P}_{2+\gamma}(\mathbb{R}^3))$. It finally remains to check that μ_t conserves energy; let us fix $t \ge 0$. Recalling that $\Lambda_2(\mu_t^n) = \Lambda_2(\mu_0^n)$ for all n and that $\Lambda_2(\mu_0^n) \to \Lambda_2(\mu_0) = 1$, we immediately have from lower semicontinuity

$$\begin{aligned} \int_{\mathbb{R}^3} |v|^2\mu_t(dv) \le \liminf_k \int_{\mathbb{R}^3} |v|^2\mu_t^{n_k}(dv) &= \liminf_k \Lambda_2(\mu_t^{n_k}) \\ &= \liminf_k \Lambda_2(\mu_0^{n_k}) = 1. \end{aligned} \tag{5.76}$$

For any t and any ε, we use (5.73) to find $R < \infty$ such that, uniformly in n,

$$\int_{\mathbb{R}^3} |v|^2 \mathbb{1}_{|v|\geq R} \mu_t^n(dv) < \varepsilon \tag{5.77}$$

and similarly for μ_t. Letting now $\chi_R : \mathbb{R}^3 \to [0,1]$ be a continuous, compactly supported function with $\chi_R(v) = 1$ when $|v| \leq R$, it follows that $\int_{\mathbb{R}^3} |v|^2 \chi_R(v) \mu_t^{n_k}(dv) \to \int_{\mathbb{R}^3} |v|^2 \chi_R(v) \mu_t(dv)$, while for each k,

$$\begin{aligned} \int_{\mathbb{R}^3} |v|^2 \chi_R(v) \mu_t^{n_k}(dv) &\geq \int_{\mathbb{R}^3} |v|^2 \mu_t^{n_k}(dv) - \int_{\mathbb{R}^3} |v|^2 \mathbb{1}_{|v|\geq R} \mu_t^{n_k}(dv) \\ &= \Lambda_2(\mu_t^{n_k}) - \int_{\mathbb{R}^3} |v|^2 \mathbb{1}_{|v|\geq R} \mu_t^{n_k}(dv) \\ &\geq 1 - \varepsilon \end{aligned} \tag{5.78}$$

so that taking a limit produces $\int_{\mathbb{R}^3} |v|^2 \mu_t(dv) \geq 1 - \varepsilon$. Since $\varepsilon > 0$ was arbitrary, we conclude that $\Lambda_2(\mu_t) = 1 = \Lambda_2(\mu_0)$ for all $t \geq 0$, so $\mu \in L^1_{\text{loc}}([0,\infty), \mathcal{S}^{2+\gamma})$ and the proof is complete. □

5.8 Proof of Theorem 5.4

We now prove our regularity result Theorem 5.4. As a starting point, we recall some results from the literature, on which we will build.

Proposition 5.13 (Theorem 6, [58]). *Suppose that $\mu_0 \in \mathcal{P}_p(\mathbb{R}^3)$ for some $p > 2$, and that the support of μ_0 is not a line:*

$$\text{for all } v_0, v_1 \in \mathbb{R}^3, \qquad \mu_0(\{v_0 + \lambda v_1 : \lambda \in \mathbb{R}\}) < 1. \tag{5.79}$$

Then there exists a weak solution $(\mu_t)_{t\geq 0}$ starting at μ_0 such that for all $t > 0$, μ_t has finite entropy $H(\mu_t) < \infty$ and

$$\text{for all } k, s \geq 0 \text{ and all } t_0 > 0, \quad \sup_{t\geq t_0} \|\mu_t\|_{H^k_s(\mathbb{R}^3)} < \infty. \tag{5.80}$$

When (5.79) holds, we say that μ_0 is not concentrated on a line. Let us remark that the cited theorem as it is stated assumes that μ_0 has a density, but that this can be removed in favour of (5.79); see the remark below [58, Lemma 9]. The same conclusion was later extended by Chen, Li and Xu [46] as follows.

Proposition 5.14 ([46], Theorem 1.1). *Fix $\gamma \in (0,1]$. Let $(\mu_t)_{t\geq 0} \subset \mathcal{S}$ be a weak solution to (LE) such that the estimate (5.80) holds. Then μ_t is analytic for all $t > 0$.*

We begin with the following very mild regularity principle, which guarantees that the hypotheses of Proposition 5.13 apply at some small time, provided that μ_0 has 4 moments. We then 'bootstrap' to the claimed result, using Propositions 2.12 and 5.14 and our uniqueness result.

Lemma 5.15. *Let $\gamma \in (0,1]$ and $\mu_0 \in \mathcal{S}^4$, and let $(\mu_t)_{t\geq 0}$ be the weak solution to (LE) starting at μ_0. Then, for any $t_0 > 0$, there exists $t_1 \in [0, t_0)$ such that μ_{t_1} is not concentrated on a line.*

Proof. If μ_0 is already not concentrated on a line, there is nothing to prove, since we can choose $t_1 = 0$ and the conclusion is immediate. We thus assume that μ_0 concentrates on a line and, by translational and rotational invariance, that μ_0 concentrates on the z-axis $L_0 = \{(0,0,z) : z \in \mathbb{R}\}$. Further, the assumption that $\mu_0 \in \mathcal{S}$ implies that it is not a point mass, so we can find two disjoint compact intervals $K_1, K_2 \subset L_0$ such that $\mu_0(K_1) > 0$ and $\mu_0(K_2) > 0$. Thanks to Theorem 5.2, we can find a solution $(V_t)_{t\geq 0}$ to (stLE) such that $\mathrm{Law}(V_t) = \mu_t$ for all $t \geq 0$.

Step 1: Itô representation of (stLE) We return to the representation of (stLE) and of V_t in terms of a Brownian motion from Proposition 5.8. The averaged coefficients are given by, for $v \in \mathbb{R}^3$ and $\mu \in \mathcal{P}_{2+\gamma}(\mathbb{R}^3)$, define

$$b(v,\mu) = \int_{\mathbb{R}^3} b(v - v_*)\mu(dv_*), \qquad a(v,\mu) = \int_{\mathbb{R}^3} a(v - v_*)\mu(dv_*)$$

and $\sigma(v,\mu)$ is a square root of $a(v,\mu)$. We recall that V_t can be written as

$$V_t = V_0 + \int_0^t b(V_s, \mu_s)ds + \int_0^t \sigma(V_s, \mu_s)d\beta_s \tag{5.81}$$

for a 3-dimensional Brownian motion β_t. Moreover, using the disintegration theorem, we find probability measures $\mathbb{P}_{v_0}, v_0 \in \mathbb{R}^3$, under which V_t solves (5.81) with the deterministic initial condition $V_0 = v_0$, and write $\mathbb{E}_{v_0}$ for the corresponding expectations. We then have

$$\mu_t(A) = \mathbb{P}(V_t \in A) = \int_{\mathbb{R}^3} \mathbb{P}_{v_0}(V_t \in A)\mu_0(dv)$$

for any $A \in \mathcal{B}(\mathbb{R}^3)$ and any $t \geq 0$.

Step 2: Small-Time Scaling Limit We now claim that if $F : \mathbb{R}^3 \to \mathbb{R}$ is bounded and continuous and $Z \sim \mathcal{N}(0, I_3)$, then

$$\lim_{\varepsilon \to 0} \sup_{v_0 \in K_1} \left| \mathbb{E}_{v_0}\left[F\left(\frac{V_\varepsilon - v_0}{\sqrt{\varepsilon}}\right)\right] - \mathbb{E}\left[F\left(\sigma(v_0, \mu_0)Z\right)\right]\right| = 0. \tag{5.82}$$

Let U be an open ball for the Euclidean norm containing K_1, and for $v \in \mathbb{R}^3$, let $\pi(v)$ be the minimiser of $|v - \tilde{v}|$ over $\tilde{v} \in \overline{U}$, which is unique by strict convexity; moreover, the map $v \mapsto \pi(v)$ is readily seen to be Lipschitz with constant 1. Recalling the growth bounds

$$|b(v - v_*)| \leq C|v - v_*|^{1+\gamma}, \qquad \|a(v - v_*)\| \leq C|v - v_*|^{2+\gamma},$$

that $\sup_{t\ge 0}\Lambda_4(\mu_t)<\infty$ by Proposition 2.12i), one checks that $|b(v,\mu_s)|+||\sigma(v,\mu_s)||\le C(1+|v|^{1+\gamma})$ and, since $\mu_t\to\mu_0$ weakly as $t\to 0$, that $a(v,\mu_t)\to a(v,\mu_0)$, and thus $\sigma(v,\mu_t)\to\sigma(v,\mu_0)$, uniformly over $v\in\overline{U}$, as $t\to 0$. We now define

$$b_t(v)=b(\pi(v),\mu_t);\qquad \sigma_t(v)=\sigma(\pi(v),\mu_t)$$

so that $b_t(v)$ and $\sigma_t(v)$ are bounded, globally Lipschitz in v, agree with $b(v,\mu_t),\sigma(v,\mu_t)$ for $v\in\overline{U}$ and $\sigma_t(v)$ converges uniformly on $\mathbb{R}^3$ as $t\downarrow 0$. Now, let $\tilde{V}_t$ be the solution to the stochastic differential equation (5.81), driven by the same Brownian motion β with these coefficients in place of $b(v,\mu_t)$ and $\sigma(v,\mu_t)$, started at $\tilde{V}_0=V_0$; this is licit as the coefficients are globally Lipschitz. Let T be the stopping time when $\tilde{V}_t$ first leaves U; by uniqueness, we have $V_t=\tilde{V}_t$ for all $t\in[0,T]$. Using now that b_t and σ_t are bounded, that $\sigma_t\to\sigma_0$ uniformly and that $\tilde{V}_t\to v_0$ as $t\to 0$, we see that

$$\begin{aligned}&\limsup_{\varepsilon\to 0}\sup_{v_0\in K_1}\mathbb{E}_{v_0}\Big[\Big|\frac{\tilde{V}_\varepsilon-v_0}{\sqrt{\varepsilon}}-\sigma_0(v_0)\frac{\beta_\varepsilon}{\sqrt{\varepsilon}}\Big|^2\Big] \\ \le&\limsup_{\varepsilon\to 0}\sup_{v_0\in K_1}\frac{1}{\varepsilon}\mathbb{E}_{v_0}\Big[2\Big(\int_0^\varepsilon b_s(\tilde{V}_s)ds\Big)^2+2\Big(\int_0^\varepsilon(\sigma_s(\tilde{V}_s)-\sigma_0(v_0))d\beta_s\Big)^2\Big]=0.\end{aligned}\tag{5.83}$$

Recalling that $\sigma_0(v_0)=\sigma(v_0,\mu_0)$ when $v_0\in K_1$ and that $\frac{\beta_\varepsilon}{\sqrt{\varepsilon}}\sim\mathcal{N}(0,I_3)$, we conclude that

$$\begin{aligned}&\sup_{v_0\in K_1}\Big|\mathbb{E}_{v_0}\Big[F\Big(\frac{V_\varepsilon-v_0}{\sqrt{\varepsilon}}\Big)\Big]-\mathbb{E}_{v_0}[F(\sigma(v_0,\mu_0)Z)]\Big| \\ \le&\sup_{v_0\in K_1}\Big|\mathbb{E}_{v_0}\Big[F\Big(\frac{\tilde{V}_\varepsilon-v_0}{\sqrt{\varepsilon}}\Big)\Big]-\mathbb{E}_{v_0}\Big[F\Big(\sigma_0(v_0)\frac{\beta_\varepsilon}{\sqrt{\varepsilon}}\Big)\Big]\Big|+2\|F\|_\infty\sup_{v_0\in K_1}\mathbb{P}_{v_0}(T<\varepsilon)\to 0\end{aligned}$$

where the final convergence follows (5.83) and the fact that $\sup_{v_0\in K_1}\mathbb{P}_{v_0}(T<\varepsilon)\to 0$ because $d(K_1,U^c)=\inf\{|v-\tilde{v}|:v\in K_1,\tilde{v}\notin U\}>0$ and because b_t and σ_t are bounded. The proof of the claim is complete.

Step 3. We now construct three test functions F_i to which apply Step 2: let $B_i\subset\mathbb{R}^2, i=1,2,3$ be disjoint open balls in the plane such that no line (in the plane) meets all three, and let $\chi_i:\mathbb{R}^2\to[0,1]$ be nonzero, smooth bump functions, supported on each B_i. Now, we define $\rho:\mathbb{R}^3\to\mathbb{R}^2$ the projection $\rho(v_1,v_2,v_3)=(v_1,v_2)$. We then introduce the bounded smooth functions $F_i:\mathbb{R}^3\to[0,1]$ defined by $F_i(v)=\chi_i(\rho(v))$. Observe that $F_i(v)\le 1\!\!1_{\{\rho(v)\in B_i\}}$.

Since μ_0 concentrates on the z-axis L_0, denoting by $e_3=(0,0,1)$, we have, for all $v_0\in L_0$,

$$a(v_0,\mu_0)=\int_{\mathbb{R}^3}|v_0-v|^{\gamma+2}\Pi_{(v-v_0)^\perp}\mu_0(dv)=h(v_0)\Pi_{e_3^\perp}=h(v_0)\begin{pmatrix}1&0&0\\0&1&0\\0&0&0\end{pmatrix},$$

where $h(v_0)=\int_{\mathbb{R}^3}|v_0-v|^{\gamma+2}\mu_0(dv)$. One easily checks that h is bounded from above and from below on K_1, since $\sup_{v_0\in K_1}h(v_0)\le C(1+\Lambda_{2+\gamma}(\mu_0))$ and $\inf_{v_0\in K_1}h(v_0)\ge$

$\alpha^{\gamma+2}\mu_0(K_2)$, where $\alpha > 0$ is the distance between K_1 and K_2. Since $\sigma(v_0,\mu_0) = [a(v_0,\mu_0)]^{1/2}$ and since $\rho(Z) \sim \mathcal{N}(0, I_2)$, we deduce that for some $\delta > 0$ and all $i = 1, 2, 3$,

$$\inf_{v_0\in K_1} \mathbb{E}_{v_0}[F_i(\sigma(v_0,\mu_0)Z)] = \inf_{v_0\in K_1} \mathbb{E}[\chi_i(h^{1/2}(v_0)\rho(Z))] \geq 2\delta > 0.$$

Thanks to (5.82), we can find $\varepsilon_0 > 0$ such that for all $\varepsilon \in (0, \varepsilon_0)$, all $i = 1, 2, 3$,

$$\inf_{v_0\in K_1} \mathbb{E}_{v_0}\Big[F_i\Big(\frac{V_\varepsilon - v_0}{\sqrt{\varepsilon}}\Big)\Big] \geq \delta \quad \text{whence} \quad \inf_{v_0\in K_1} \mathbb{P}_{v_0}\Big(\rho\Big(\frac{V_\varepsilon - v_0}{\sqrt{\varepsilon}}\Big) \in B_i\Big) \geq \delta.$$

Step 4. Now, we fix $t_0 > 0$ as in the statement, and consider $t_1 \in (0, \varepsilon_0 \wedge t_0)$. For a given line $L = \{x_0 + \lambda u_0 : \lambda \in \mathbb{R}\} \subset \mathbb{R}^3$ and for $v_0 \in K_1$, we denote by $L_{t_1,v_0} = \rho((L - v_0)/\sqrt{t_1})$, which is a line (or a point) in $\mathbb{R}^2$. There is $i \in \{1, 2, 3\}$, possibly depending on t_1, such that $L_{t_1,v_0} \cap B_i = \emptyset$, so that

$$\begin{aligned}\mathbb{P}_{v_0}(V_{t_1} \in L) =& \mathbb{P}_{v_0}\Big(\frac{V_{t_1} - v_0}{\sqrt{t_1}} \in \frac{L - v_0}{\sqrt{t_1}}\Big) \\ \leq& \mathbb{P}_{v_0}\Big(\rho\Big(\frac{V_{t_1} - v_0}{\sqrt{t_1}}\Big) \in L_{t_1,v_0}\Big) \\ \leq& 1 - \mathbb{P}_{v_0}\Big(\rho\Big(\frac{V_{t_1} - v_0}{\sqrt{t_1}}\Big) \in B_i\Big) \\ \leq& 1 - \delta\end{aligned}$$

by Step 3. In other words, for all $v_0 \in K_1$, $\mathbb{P}_{v_0}(V_{t_1} \in \mathbb{R}^3 \setminus L) \geq \delta$, whence

$$\mu_{t_1}(\mathbb{R}^3 \setminus L) = \int_{\mathbb{R}^3} \mathbb{P}_{v_0}(V_{t_1} \notin L)\mu_0(dv_0) \geq \delta\mu_0(K_1) > 0.$$

The proof is complete. □

We now prove our claimed result.

Proof of Theorem 5.4. Let $\mu_0 \in \mathcal{S}$, and let $(\mu_t)_{t\geq 0}$ be any weak solution to (LE) starting at μ_0. Fix $t_0 > 0$. By Proposition 2.12i), picking $t_1 \in (0, t_0)$ arbitrarily, we have $\Lambda_4(\mu_{t_1}) < \infty$, so $\mu_{t_1} \in \mathcal{S}^4$. We can therefore apply Lemma 5.15 to find $t_2 \in [t_1, t_0)$ such that μ_{t_2} is not concentrated on a line, and we also have $\Lambda_4(\mu_{t_2}) < \infty$, still by Proposition 2.12i), because $t_2 > 0$.

Now, by Proposition 5.13 there exists *a* solution $(\nu_t)_{t\geq 0}$ to (LE) starting at $\nu_0 = \mu_{t_2}$ such that, for all $s, k \geq 0$ and $\delta > 0$,

$$\sup_{t\geq\delta} \|\nu_t\|_{H^k_s(\mathbb{R}^3)} < \infty$$

and such that $H(\nu_t) < \infty$ for all $t > 0$; by Proposition 5.14, $\tilde{\nu}_t$ is further analytic for all $t > 0$.

Since the fourth moment $\Lambda_4(\mu_{t_2}) < \infty$, we can apply the uniqueness in Theorem 5.1 to see that there is a unique weak solution to (LE) starting at $\nu_0 = \mu_{t_2}$, so we must have $\nu_t = \mu_{t_2+t}$ for all $t \geq 0$. In particular, $\mu_{t_0} = \nu_{t_0-t_2}$ is analytic and has finite entropy and choosing $\delta = t_0 - t_2$, for all $s, k \geq 0$, $\sup_{t\geq t_0} \|\mu_t\|_{H^k_s(\mathbb{R}^3)} < \infty$. □

5.9 Proof of the central inequality

We now turn to the proof of Lemma 5.9. It is convenient to first break up $\mathcal{E}_{p,\varepsilon}$, which we accomplish with the following Itô lemma, carefully applying the coupling operator to our cost function.

Lemma 5.16. *Adopt the notation of Lemma 5.9 and fix $p \geq 2$ and $\varepsilon \in [0,1]$, and let $d_{p,\epsilon}$ be the transport cost defined in Section 2.1. For $v, v_*, \tilde{v}, \tilde{v}_* \in \mathbb{R}^3$,*

$$\begin{aligned}\mathcal{E}_{p,\varepsilon}(v,\tilde{v},v_*,\tilde{v}_*) \leq & \mathcal{T}_1(v,v_*,\tilde{v},\tilde{v}_*) + \mathcal{T}_2(v,v_*,\tilde{v},\tilde{v}_*) + \mathcal{T}_2(\tilde{v},\tilde{v}_*,v,v_*) \\ & + \mathcal{T}_3(v,v_*,\tilde{v},\tilde{v}_*) + \mathcal{T}_3(\tilde{v},\tilde{v}_*,v,v_*),\end{aligned}$$

where, setting $x = v - v_*$ and $\tilde{x} = \tilde{v} - \tilde{v}_*$,

$$\begin{aligned}\mathcal{T}_1(v,v_*,\tilde{v},\tilde{v}_*) =& (1+|v|^p+|\tilde{v}|^p)\varphi_\varepsilon'(|v-\tilde{v}|^2)\Big[2(v-\tilde{v})\cdot(b(x)-b(\tilde{x})) + ||\sigma(x)-\sigma(\tilde{x})||^2\Big], \\ \mathcal{T}_2(v,v_*,\tilde{v},\tilde{v}_*) =& \varphi_\varepsilon(|v-\tilde{v}|^2)\Big[p|v|^{p-2}v\cdot b(x) + \frac{p}{2}|v|^{p-2}||\sigma(x)||^2 + \frac{p(p-2)}{2}|v|^{p-4}|\sigma(x)v|^2\Big], \\ \mathcal{T}_3(v,v_*,\tilde{v},\tilde{v}_*) =& 2p|v|^{p-2}\varphi_\varepsilon'(|v-\tilde{v}|^2)[\sigma(x)v]\cdot[(\sigma(x)-\sigma(\tilde{x}))(v-\tilde{v})].\end{aligned}$$

Proof. Let us fix $p \geq 2, \varepsilon > 0$, and recall that $d_{p,\varepsilon}$ is given by

$$d_{p,\varepsilon}(v,\tilde{v}) = (1+|v|^p+|\tilde{v}|^p)\varphi_\varepsilon(|v-\tilde{v}|^2) \tag{5.84}$$

where the function $\varphi_\varepsilon(r) = r/(1+\varepsilon r)$ satisfies

$$0 \leq \varphi_e' \leq 1; \qquad \varphi_e'' \leq 0. \tag{5.85}$$

Let $f(v,\tilde{v}) = d_{p,\varepsilon}(v,\tilde{v}) = (1+|v|^p+|\tilde{v}|^p)\varphi_\varepsilon(|v-\tilde{v}|^2)$. We have

$$\partial_{v_k} f(v,\tilde{v}) = p|v|^{p-2}v_k\varphi_\varepsilon(|v-\tilde{v}|^2) + 2(v_k-\tilde{v}_k)(1+|v|^p+|\tilde{v}|^p)\varphi_\varepsilon'(|v-\tilde{v}|^2)$$

and a symmetric expression for $\partial_{\tilde{v}_k} f(v,\tilde{v})$. Differentiating again, we find

$$\begin{aligned}\partial^2_{v_k v_\ell} f(v,\tilde{v}) =& p|v|^{p-2}\mathbb{1}_{\{k=\ell\}}\varphi_\varepsilon(|v-\tilde{v}|^2) + p(p-2)|v|^{p-4}v_k v_\ell\varphi_\varepsilon(|v-\tilde{v}|^2) \\ & + 2p|v|^{p-2}v_k(v_\ell-\tilde{v}_\ell)\varphi_\varepsilon'(|v-\tilde{v}|^2) + 2\mathbb{1}_{\{k=\ell\}}(1+|v|^p+|\tilde{v}|^p)\varphi_\varepsilon'(|v-\tilde{v}|^2) \\ & + 4(v_k-\tilde{v}_k)(v_\ell-\tilde{v}_\ell)(1+|v|^p+|\tilde{v}|^p)\varphi_\varepsilon''(|v-\tilde{v}|^2) \\ & + 2p|v|^{p-2}(v_k-\tilde{v}_k)v_\ell\varphi_\varepsilon'(|v-\tilde{v}|^2)\end{aligned}$$

and a symmetric expression for $\partial^2_{\tilde{v}_k\tilde{v}_\ell} f(v,\tilde{v})$. Concerning the cross terms,

$$\begin{aligned}\partial^2_{v_k\tilde{v}_\ell} f(v,\tilde{v}) =& 2p|v|^{p-2}v_k(\tilde{v}_\ell-v_\ell)\varphi_\varepsilon'(|v-\tilde{v}|^2) + 2p|\tilde{v}|^{p-2}(v_k-\tilde{v}_k)\tilde{v}_\ell\varphi_\varepsilon'(|v-\tilde{v}|^2) \\ & - 4(v_k-\tilde{v}_k)(v_\ell-\tilde{v}_\ell)(1+|v|^p+|\tilde{v}|^p)\varphi_\varepsilon''(|v-\tilde{v}|^2) \\ & - 2\mathbb{1}_{\{k=\ell\}}(1+|v|^p+|\tilde{v}|^p)\varphi_\varepsilon'(|v-\tilde{v}|^2).\end{aligned}$$

Let us now examine the sums in the definition of $\mathcal{A}f$ one by one. First,

$$\begin{aligned}
&\sum_{k=1}^{3}[b_k(v-v_*)\partial_{v_k}f(v,\tilde{v})+b_k(\tilde{v}-\tilde{v}_*)\partial_{\tilde{v}_k}f(v,\tilde{v})] \\
=&p|v|^{p-2}v\cdot b(v-v_*)\varphi_\varepsilon(|v-\tilde{v}|^2) && (=A_1)\\
&+p|\tilde{v}|^{p-2}\tilde{v}\cdot b(\tilde{v}-\tilde{v}_*)\varphi_\varepsilon(|v-\tilde{v}|^2) && (=A_2)\\
&+2(1+|v|^p+|\tilde{v}|^p)(v-\tilde{v})\cdot(b(v-v_*)-b(\tilde{v}-\tilde{v}_*))\varphi'_\varepsilon(|v-\tilde{v}|^2). && (=A_3)
\end{aligned}$$

Next, using that for $x,y,z\in\mathbb{R}^3$, Tr $a(x)=||\sigma(x)||^2$ and $\sum_{k,\ell=1}^3 a_{k\ell}(x)y_kz_\ell=[\sigma(x)y]\cdot[\sigma(x)z]$,

$$\begin{aligned}
\frac{1}{2}\sum_{k,\ell=1}^{3}a_{k\ell}(v-v_*)\partial^2_{v_kv_\ell}f(v,\tilde{v}) =& \frac{p}{2}|v|^{p-2}\|\sigma(v-v_*)\|^2\varphi_\varepsilon(|v-\tilde{v}|^2) && (=B_1)\\
&+\frac{p(p-2)}{2}|v|^{p-4}|\sigma(v-v_*)v|^2\varphi_\varepsilon(|v-\tilde{v}|^2) && (=B_2)\\
&+2p|v|^{p-2}[\sigma(v-v_*)v]\cdot[\sigma(v-v_*)(v-\tilde{v})]\varphi'_\varepsilon(|v-\tilde{v}|^2) && (=B_3)\\
&+(1+|v|^p+|\tilde{v}|^p)\|\sigma(v-v_*)\|^2\varphi'_\varepsilon(|v-\tilde{v}|^2) && (=B_4)\\
&+2(1+|v|^p+|\tilde{v}|^p)|\sigma(v-v_*)(v-\tilde{v})|^2\varphi''_\varepsilon(|v-\tilde{v}|^2). && (=B_5)
\end{aligned}$$

Similarly,

$$\begin{aligned}
\frac{1}{2}\sum_{k,\ell=1}^{3}a_{k\ell}(\tilde{v}-\tilde{v}_*)\partial^2_{\tilde{v}_k\tilde{v}_\ell}f(v,\tilde{v}) =& \frac{p}{2}|\tilde{v}|^{p-2}\|\sigma(\tilde{v}-\tilde{v}_*)\|^2\varphi_\varepsilon(|v-\tilde{v}|^2) && (=C_1)\\
&+\frac{p(p-2)}{2}|\tilde{v}|^{p-4}|\sigma(\tilde{v}-\tilde{v}_*)\tilde{v}|^2\varphi_\varepsilon(|v-\tilde{v}|^2) && (=C_2)\\
&+2p|\tilde{v}|^{p-2}[\sigma(\tilde{v}-\tilde{v}_*)\tilde{v}]\cdot[\sigma(\tilde{v}-\tilde{v}_*)(\tilde{v}-v)]\varphi'_\varepsilon(|v-\tilde{v}|^2) && (=C_3)\\
&+(1+|v|^p+|\tilde{v}|^p)\|\sigma(\tilde{v}-\tilde{v}_*)\|^2\varphi'_\varepsilon(|v-\tilde{v}|^2) && (=C_4)\\
&+2(1+|v|^p+|\tilde{v}|^p)|\sigma(\tilde{v}-\tilde{v}_*)(\tilde{v}-v)|^2\varphi''_\varepsilon(|v-\tilde{v}|^2). && (=C_5)
\end{aligned}$$

Finally, we look at the cross-terms:

$$\begin{aligned}
&\sum_{j,k,\ell=1}^{3}\sigma_{kj}(v-v_*)\sigma_{\ell j}(\tilde{v}-\tilde{v}_*)\partial^2_{v_k\tilde{v}_\ell}f(v,\tilde{v})\\
=&-2p|v|^{p-2}[\sigma(v-v_*)v]\cdot[\sigma(\tilde{v}-\tilde{v}_*)(v-\tilde{v})]\varphi'_\varepsilon(|v-\tilde{v}|^2) && (=D_1)\\
&+2p|\tilde{v}|^{p-2}[\sigma(v-v_*)(v-\tilde{v})]\cdot[\sigma(\tilde{v}-\tilde{v}_*)\tilde{v}]\varphi'_\varepsilon(|v-\tilde{v}|^2) && (=D_2)\\
&-4(1+|v|^p+|\tilde{v}|^p)[\sigma(v-v_*)(v-\tilde{v})]\cdot[\sigma(\tilde{v}-\tilde{v}_*)(v-\tilde{v})]\varphi''_\varepsilon(|v-\tilde{v}|^2) && (=D_3)\\
&-2(1+|v|^p+|\tilde{v}|^p)\langle\!\langle\sigma(v-v_*),\sigma(\tilde{v}-\tilde{v}_*)\rangle\!\rangle\varphi'_\varepsilon(|v-\tilde{v}|^2). && (=D_4)
\end{aligned}$$

Recalling the notation $x = v - v_*$ and $\tilde{x} = \tilde{v} - \tilde{v}_*$, we find that

$$\begin{aligned}
A_3 + B_4 + C_4 + D_4 =& \mathcal{T}_1(v, v_*, \tilde{v}, \tilde{v}_*),\\
A_1 + B_1 + B_2 =& \mathcal{T}_2(v, v_*, \tilde{v}, \tilde{v}_*),\\
A_2 + C_1 + C_2 =& \mathcal{T}_2(\tilde{v}, \tilde{v}_*, v, v_*),\\
B_3 + D_1 =& \mathcal{T}_3(v, v_*, \tilde{v}, \tilde{v}_*),\\
C_3 + D_2 =& \mathcal{T}_3(\tilde{v}, \tilde{v}_*, v, v_*),
\end{aligned}$$

and finally that

$$B_5 + C_5 + D_3 = 2(1 + |v|^p + |\tilde{v}|^p)|(\sigma(x) - \sigma(\tilde{x}))(v - \tilde{v})|^2 \varphi''_\varepsilon(|v - \tilde{v}|^2) \le 0$$

since φ''_ε is nonpositive, see (5.85). □

We now analyse the terms one by one.

Proof of Lemma 5.9. We introduce the shortened notation $x = v - v_*$, $\tilde{x} = \tilde{v} - \tilde{v}_*$ and start from the bound in the previous lemma

$$\mathcal{E}_{p,\varepsilon} \le \mathcal{T}_1 + \mathcal{T}_2 + \widetilde{\mathcal{T}}_2 + \mathcal{T}_3 + \widetilde{\mathcal{T}}_3, \tag{5.86}$$

where $\mathcal{T}_1 = \mathcal{T}_1(v, v_*, \tilde{v}, \tilde{v}_*)$, $\mathcal{T}_2 = \mathcal{T}_2(v, v_*, \tilde{v}, \tilde{v}_*)$, $\widetilde{\mathcal{T}}_2 = \mathcal{T}_2(\tilde{v}, \tilde{v}_*, v, v_*)$, etc. In the whole proof, C is allowed to change from line to line and to depend (only) on p and γ.

Step 1. We begin with $\mathcal{T}_1$, which is the most difficult term. We start from

$$\mathcal{T}_1 = (1 + |v|^p + |\tilde{v}|^p)\varphi'_\varepsilon(|v - \tilde{v}|^2)[g_1 + g_2 + g_3],$$

where

$$\begin{aligned}
g_1 =& [(v - \tilde{v}) - (v_* - \tilde{v}_*)] \cdot (b(x) - b(\tilde{x})) + ||\sigma(x) - \sigma(\tilde{x})||^2,\\
g_2 =& (v - \tilde{v}) \cdot (b(x) - b(\tilde{x})),\\
g_3 =& (v_* - \tilde{v}_*) \cdot (b(x) - b(\tilde{x})).
\end{aligned}$$

Step 1.1. Recalling that $b(x) = -2|x|^\gamma x$ and using (5.7), we find

$$\begin{aligned}
g_1 \le& 2(x - \tilde{x}) \cdot [-|x|^\gamma x + |\tilde{x}|^\gamma \tilde{x}] + 2|x|^{\gamma+2} + 2|\tilde{x}|^{\gamma+2} - 4|x|^{\gamma/2}|\tilde{x}|^{\gamma/2}(x \cdot \tilde{x})\\
=& 2(|x|^\gamma + |\tilde{x}|^\gamma)(x \cdot \tilde{x}) - 4|x|^{\gamma/2}|\tilde{x}|^{\gamma/2}(x \cdot \tilde{x})\\
=& 2(x \cdot \tilde{x})(|x|^{\gamma/2} - |\tilde{x}|^{\gamma/2})^2.
\end{aligned}$$

Using now (5.5) with $\alpha = \gamma/2$,

$$g_1 \le 2|x||\tilde{x}|(|x| \vee |\tilde{x}|)^{\gamma-2}(|x| - |\tilde{x}|)^2 = 2(|x| \wedge |\tilde{x}|)(|x| \vee |\tilde{x}|)^{\gamma-1}(|x| - |\tilde{x}|)^2 \le 2(|x| \wedge |\tilde{x}|)^\gamma |x - \tilde{x}|^2.$$

Since $|x-\tilde{x}| = |(v-\tilde{v})-(v_*-\tilde{v}_*)|$, we end with

$$g_1 \le 2(|x|\wedge|\tilde{x}|)^\gamma|v-\tilde{v}|^2 + 2(|x|\wedge|\tilde{x}|)^\gamma(2|v-\tilde{v}||v_*-\tilde{v}_*| + |v_*-\tilde{v}_*|^2).$$

Step 1.2. We next study g_2, assuming without loss of generality that $|x| \ge |\tilde{x}|$. We write, using (5.5) with $\alpha=\gamma$,

$$\begin{aligned} g_2 =&2(v-\tilde{v})\cdot[-|x|^\gamma(x-\tilde{x}) + (|\tilde{x}|^\gamma - |x|^\gamma)\tilde{x}] \\ \le& -2|x|^\gamma(v-\tilde{v})\cdot(x-\tilde{x}) + 2|v-\tilde{v}||\tilde{x}|(|x|\vee|\tilde{x}|)^{\gamma-1}||x|-|\tilde{x}|| \\ \le& -2|x|^\gamma(v-\tilde{v})\cdot(x-\tilde{x}) + 2|v-\tilde{v}||\tilde{x}|^\gamma|x-\tilde{x}|. \end{aligned}$$

Since now $x=v-v_*$ and $\tilde{x}=\tilde{v}-\tilde{v}_*$, we see that

$$\begin{aligned} g_2 \le& -2|x|^\gamma|v-\tilde{v}|^2 + 2|x|^\gamma|v-\tilde{v}||v_*-\tilde{v}_*| + 2|\tilde{x}|^\gamma[|v-\tilde{v}|^2 + |v-\tilde{v}||v_*-\tilde{v}_*|] \\ \le& 2(|x|^\gamma + |\tilde{x}|^\gamma)|v-\tilde{v}||v_*-\tilde{v}_*| \end{aligned}$$

since $|x|\ge|\tilde{x}|$ by assumption. By symmetry, the same bound holds when $|x|\le|\tilde{x}|$.

Step 1.3. Using now (5.6), we see that

$$g_3 \le 2|v_*-\tilde{v}_*|[|x|^\gamma+|\tilde{x}|^\gamma]|x-\tilde{x}| \le 2(|x|^\gamma+|\tilde{x}|^\gamma)[|v-\tilde{v}||v_*-\tilde{v}_*| + |v_*-\tilde{v}_*|^2].$$

Step 1.4. Gathering Steps 1.1, 1.2, 1.3, we have checked that

$$\mathcal{T}_1 \le (1+|v|^p+|\tilde{v}|^p)\varphi_\varepsilon'(|v-\tilde{v}|^2)\Big[2(|x|\wedge|\tilde{x}|)^\gamma|v-\tilde{v}|^2 + C(|x|^\gamma+|\tilde{x}|^\gamma)(|v-\tilde{v}||v_*-\tilde{v}_*| + |v_*-\tilde{v}_*|^2)\Big].$$

Recalling that $r\varphi_\varepsilon'(r)\le\varphi_\varepsilon(r)$ by (2.11) and that $|x|^\gamma \le |v|^\gamma+|v_*|^\gamma$ and $|\tilde{x}|^\gamma\le|\tilde{v}|^\gamma+|\tilde{v}_*|^\gamma$, we may write $\mathcal{T}_1\le\mathcal{T}_{1,1}+\mathcal{T}_{1,2}$, where

$$\begin{aligned} \mathcal{T}_{1,1} =&2(1+|v|^p+|\tilde{v}|^p)[(|v|^\gamma+|v_*|^\gamma)\wedge(|\tilde{v}|^\gamma+|\tilde{v}_*|^\gamma)]\varphi_\varepsilon(|v-\tilde{v}|^2), \\ \mathcal{T}_{1,2} =&C(1+|v|^p+|\tilde{v}|^p)(|v|^\gamma+|v_*|^\gamma+|\tilde{v}|^\gamma+|\tilde{v}_*|^\gamma)\varphi_\varepsilon'(|v-\tilde{v}|^2)(|v-\tilde{v}||v_*-\tilde{v}_*| + |v_*-\tilde{v}_*|^2). \end{aligned}$$

First,

$$\begin{aligned} \mathcal{T}_{1,1} \le& 2(|v|^\gamma+|v_*|^\gamma)\varphi_\varepsilon(|v-\tilde{v}|^2) + 2|v|^p(|v|^\gamma+|v_*|^\gamma)\varphi_\varepsilon(|v-\tilde{v}|^2) + 2|\tilde{v}|^p(|\tilde{v}|^\gamma+|\tilde{v}_*|^\gamma)\varphi_\varepsilon(|v-\tilde{v}|^2) \\ =&2(|v|^{p+\gamma}+|\tilde{v}|^{p+\gamma})\varphi_\varepsilon(|v-\tilde{v}|^2) + 2(|v|^\gamma+|v_*|^\gamma+|v|^p|v_*|^\gamma+|\tilde{v}|^p|\tilde{v}_*|^\gamma)\varphi_\varepsilon(|v-\tilde{v}|^2) \\ \le& 2(|v|^{p+\gamma}+|\tilde{v}|^{p+\gamma})\varphi_\varepsilon(|v-\tilde{v}|^2) + C(1+|v_*|^\gamma+|\tilde{v}_*|^\gamma)(1+|v|^p+|\tilde{v}|^p)\varphi_\varepsilon(|v-\tilde{v}|^2) \\ =&2d_{p+\gamma,\varepsilon}(v,\tilde{v}) + C(1+|v_*|^{p+\gamma}+|\tilde{v}_*|^{p+\gamma})d_{p,\varepsilon}(v,\tilde{v}). \end{aligned}$$

We next use that $ab\le\varepsilon^{1/2}a^2+\varepsilon^{-1/2}b^2$ to write

$$\begin{aligned} \mathcal{T}_{1,2} \le& C\sqrt{\varepsilon}(1+|v|^p+|\tilde{v}|^p)(|v|^\gamma+|v_*|^\gamma+|\tilde{v}|^\gamma+|\tilde{v}_*|^\gamma)\varphi_\varepsilon'(|v-\tilde{v}|^2)|v-\tilde{v}|^2 \\ &+\frac{C}{\sqrt{\varepsilon}}(1+|v|^p+|\tilde{v}|^p)(|v|^\gamma+|v_*|^\gamma+|\tilde{v}|^\gamma+|\tilde{v}_*|^\gamma)\varphi_\varepsilon'(|v-\tilde{v}|^2)|v_*-\tilde{v}_*|^2 \\ \le& C\sqrt{\varepsilon}(1+|v|^p+|\tilde{v}|^p)(|v|^\gamma+|v_*|^\gamma+|\tilde{v}|^\gamma+|\tilde{v}_*|^\gamma)\varphi_\varepsilon(|v-\tilde{v}|^2) \\ &+\frac{C}{\sqrt{\varepsilon}}(1+|v|^p+|\tilde{v}|^p)(|v|^\gamma+|v_*|^\gamma+|\tilde{v}|^\gamma+|\tilde{v}_*|^\gamma)|v_*-\tilde{v}_*|^2, \end{aligned}$$

because $r\varphi'_\varepsilon(r) \le \varphi_\varepsilon(r)$ and $\varphi'_\varepsilon(r) \le 1$ by (2.11). We carry on with

$$\begin{aligned}
\mathcal{T}_{1,2} \le & C\sqrt{\varepsilon}(1+|v|^{p+\gamma}+|\tilde v|^{p+\gamma})\varphi_\varepsilon(|v-\tilde v|^2) + C\sqrt{\varepsilon}(1+|v|^p+|\tilde v|^p)(|v_*|^\gamma+|\tilde v_*|^\gamma)\varphi_\varepsilon(|v-\tilde v|^2)\\
& + \frac{C}{\sqrt{\varepsilon}}(1+|v|^p+|\tilde v|^p)(|v|^\gamma+|v_*|^\gamma+|\tilde v|^\gamma+|\tilde v_*|^\gamma)(1+\varepsilon|v_*-\tilde v_*|^2)\varphi_\varepsilon(|v_*-\tilde v_*|^2)\\
\le & C\sqrt{\varepsilon}d_{p+\gamma,\varepsilon}(v,\tilde v) + C\sqrt{\varepsilon}(1+|v_*|^\gamma+|\tilde v_*|^\gamma)d_{p,\varepsilon}(v,\tilde v)\\
& + \frac{C}{\sqrt{\varepsilon}}(1+|v|^p+|\tilde v|^p)(|v|^\gamma+|v_*|^\gamma+|\tilde v|^\gamma+|\tilde v_*|^\gamma)\varphi_\varepsilon(|v_*-\tilde v_*|^2)\\
& + C\sqrt{\varepsilon}(1+|v|^p+|\tilde v|^p)(|v|^\gamma+|v_*|^\gamma+|\tilde v|^\gamma+|\tilde v_*|^\gamma)(|v_*|^2+|\tilde v_*|^2)\varphi_\varepsilon(|v_*-\tilde v_*|^2)\\
\le & C\sqrt{\varepsilon}d_{p+\gamma,\varepsilon}(v,\tilde v) + C\sqrt{\varepsilon}(1+|v_*|^{p+\gamma}+|\tilde v_*|^{p+\gamma})d_{p,\varepsilon}(v,\tilde v)\\
& + \frac{C}{\sqrt{\varepsilon}}(1+|v|^{p+\gamma}+|\tilde v|^{p+\gamma})(1+|v_*|^\gamma+|\tilde v_*|^\gamma)\varphi_\varepsilon(|v_*-\tilde v_*|^2)\\
& + C\sqrt{\varepsilon}(1+|v|^{p+\gamma}+|\tilde v|^{p+\gamma})(|v_*|^2+|\tilde v_*|^2)\varphi_\varepsilon(|v_*-\tilde v_*|^2)\\
& + C\sqrt{\varepsilon}(1+|v|^p+|\tilde v|^p)(1+|v_*|^{2+\gamma}+|\tilde v_*|^{2+\gamma})\varphi_\varepsilon(|v_*-\tilde v_*|^2).
\end{aligned}$$

Since $p \ge 2$, since $\gamma \in (0,1)$ and since $\varepsilon \in (0,1]$, we end with

$$\begin{aligned}
\mathcal{T}_{1,2} \le & C\sqrt{\varepsilon}d_{p+\gamma,\varepsilon}(v,\tilde v) + C(1+|v_*|^{p+\gamma}+|\tilde v_*|^{p+\gamma})d_{p,\varepsilon}(v,\tilde v)\\
& + \frac{C}{\sqrt{\varepsilon}}(1+|v|^{p+\gamma}+|\tilde v|^{p+\gamma})d_{p,\varepsilon}(v_*,\tilde v_*)\\
& + C\sqrt{\varepsilon}(1+|v|^p+|\tilde v|^p)d_{p+\gamma,\varepsilon}(v_*,\tilde v_*).
\end{aligned}$$

Summing the bounds on $\mathcal{T}_{1,1}$ and $\mathcal{T}_{1,2}$, we have proven that

$$\begin{aligned}
\mathcal{T}_1 \le & 2d_{p+\gamma,\varepsilon}(v,\tilde v)\\
& + C\sqrt{\varepsilon}(1+|v_*|^p+|\tilde v_*|^p)d_{p+\gamma,\varepsilon}(v,\tilde v)\\
& + C\sqrt{\varepsilon}(1+|v|^p+|\tilde v|^p)d_{p+\gamma,\varepsilon}(v_*,\tilde v_*)\\
& + \frac{C}{\sqrt{\varepsilon}}(1+|v_*|^{p+\gamma}+|\tilde v_*|^{p+\gamma})d_{p,\varepsilon}(v,\tilde v)\\
& + \frac{C}{\sqrt{\varepsilon}}(1+|v|^{p+\gamma}+|\tilde v|^{p+\gamma})d_{p,\varepsilon}(v_*,\tilde v_*).
\end{aligned} \tag{5.87}$$

Step 2. We next turn to $\mathcal{T}_2$. By (2.126)-(2.127) and by definition of $\mathcal{T}_2$, we see that

$$\begin{aligned}
\mathcal{T}_2 \le & \varphi_\varepsilon(|v-\tilde v|^2)\Big[-p|v|^{p+\gamma}+p|v|^p|v_*|^\gamma + Cp^2(|v|^{p-2+\gamma}|v_*|^2+|v|^{p-2}|v_*|^{2+\gamma})\Big].\\
\le & \varphi_\varepsilon(|v-\tilde v|^2)\Big[-p|v|^{p+\gamma}+C(1+|v_*|^{2+\gamma})(1+|v|^p)\Big].
\end{aligned}$$

It follows that

$$\mathcal{T}_2 \le -p|v|^{p+\gamma}\varphi_\varepsilon(|v-\tilde v|^2) + C(1+|v_*|^{p+\gamma})d_{p,\varepsilon}(v,\tilde v), \tag{5.88}$$

and, still allowing C to change from line to line and to depend on p, that

$$\begin{aligned}\mathcal{T}_2+\tilde{\mathcal{T}}_2 \le& -p(|v|^{p+\gamma}+|\tilde{v}|^{p+\gamma})\varphi_\varepsilon(|v-\tilde{v}|^2)+C(1+|v_*|^{p+\gamma}+|\tilde{v}_*|^{p+\gamma})d_{p,\varepsilon}(v,\tilde{v}) \\ =& -p(d_{p+\gamma}(v,\tilde{v})-\varphi_\varepsilon(|v-\tilde{v}|^2))+C(1+|v_*|^{p+\gamma}+|\tilde{v}_*|^{p+\gamma})d_{p,\varepsilon}(v,\tilde{v}) \\ \le& -pd_{p+\gamma}(v,\tilde{v})+C(1+|v_*|^{p+\gamma}+|\tilde{v}_*|^{p+\gamma})d_{p,\varepsilon}(v,\tilde{v}),\end{aligned} \tag{5.89}$$

where the equality uses the definition (2.10) of $d_{p+\gamma,\varepsilon}$, and in the final line we absorb $\varphi(|v-\tilde{v}|^2)\le d_{p,\varepsilon}(v,\tilde{v})$ into the second term.

Step 3. We finally deal with $\mathcal{T}_3,\tilde{\mathcal{T}}_3$; by symmetry, it suffices to treat the case of $\mathcal{T}_3$. Recalling that $|\sigma(x)v|\le C|x|^{\gamma/2}|v||v_*|$ by (5.9), and that $||\sigma(x)-\sigma(\tilde{x})||\le C(|x|^{\gamma/2}+|\tilde{x}|^{\gamma/2})|x-\tilde{x}|$ by (5.8), we directly find

$$\begin{aligned}\mathcal{T}_3 \le& C|v|^{p-1}|v_*||x|^{\gamma/2}(|x|^{\gamma/2}+|\tilde{x}|^{\gamma/2})|x-\tilde{x}||v-\tilde{v}|\varphi'_\varepsilon(|v-\tilde{v}|^2) \\ \le& C|v|^{p-1}|v_*|(|v|^\gamma+|\tilde{v}|^\gamma+|v_*|^\gamma+|\tilde{v}_*|^\gamma)(|v-\tilde{v}|^2+|v-\tilde{v}||v_*-\tilde{v}_*|)\varphi'_\varepsilon(|v-\tilde{v}|^2) \\ =& \mathcal{T}_{3,1}+\mathcal{T}_{3,2},\end{aligned}$$

where

$$\begin{aligned}\mathcal{T}_{3,1} =& C|v|^{p-1}|v_*|(|v|^\gamma+|\tilde{v}|^\gamma+|v_*|^\gamma+|\tilde{v}_*|^\gamma)|v-\tilde{v}|^2\varphi'_\varepsilon(|v-\tilde{v}|^2), \\ \mathcal{T}_{3,2} =& C|v|^{p-1}|v_*|(|v|^\gamma+|\tilde{v}|^\gamma+|v_*|^\gamma+|\tilde{v}_*|^\gamma)|v-\tilde{v}||v_*-\tilde{v}_*|\varphi'_\varepsilon(|v-\tilde{v}|^2).\end{aligned}$$

Since $r\varphi'_\varepsilon(r)\le\varphi_\varepsilon(r)$ by (2.11), we have

$$\begin{aligned}\mathcal{T}_{3,1} \le& C(1+|v_*|^{1+\gamma}+|\tilde{v}_*|^{1+\gamma})(1+|v|^{p-1+\gamma}+|\tilde{v}|^{p-1+\gamma})\varphi_\varepsilon(|v-\tilde{v}|^2) \\ \le& C(1+|v_*|^{p+\gamma}+|\tilde{v}_*|^{p+\gamma})d_{p,\varepsilon}(v,\tilde{v}).\end{aligned}$$

Next, we use that, with $a=|v-\tilde{v}|$ and $a_*=|v_*-\tilde{v}_*|$, since $a[\varphi'_\varepsilon(a^2)]\le\sqrt{a^2\varphi'_\varepsilon(a^2)}\le\sqrt{\varphi_\varepsilon(a^2)}$ by (2.11),

$$aa_*\varphi'_\varepsilon(a^2)\le\sqrt{\varphi_\varepsilon(a^2)}\sqrt{\varphi_\varepsilon(a_*^2)(1+\varepsilon a_*^2)}\le[\varphi_\varepsilon(a^2)+\varphi_\varepsilon(a_*^2)](1+\sqrt{\varepsilon}a_*)$$

to write $\mathcal{T}_{3,2}\le\mathcal{T}_{3,2,1}+\mathcal{T}_{3,2,2}+\mathcal{T}_{3,2,3}+\mathcal{T}_{3,2,4}$, where

$$\begin{aligned}\mathcal{T}_{3,2,1} =& C|v|^{p-1}|v_*|(|v|^\gamma+|\tilde{v}|^\gamma+|v_*|^\gamma+|\tilde{v}_*|^\gamma)\varphi_\varepsilon(|v-\tilde{v}|^2), \\ \mathcal{T}_{3,2,2} =& C|v|^{p-1}|v_*|(|v|^\gamma+|\tilde{v}|^\gamma+|v_*|^\gamma+|\tilde{v}_*|^\gamma)\varphi_\varepsilon(|v_*-\tilde{v}_*|^2), \\ \mathcal{T}_{3,2,3} =& C\sqrt{\varepsilon}|v|^{p-1}|v_*|(|v|^\gamma+|\tilde{v}|^\gamma+|v_*|^\gamma+|\tilde{v}_*|^\gamma)|v_*-\tilde{v}_*|\varphi_\varepsilon(|v-\tilde{v}|^2), \\ \mathcal{T}_{3,2,4} =& C\sqrt{\varepsilon}|v|^{p-1}|v_*|(|v|^\gamma+|\tilde{v}|^\gamma+|v_*|^\gamma+|\tilde{v}_*|^\gamma)|v_*-\tilde{v}_*|\varphi_\varepsilon(|v_*-\tilde{v}_*|^2).\end{aligned}$$

We have

$$\begin{aligned}\mathcal{T}_{3,2,1} \le& C(1+|v_*|^{1+\gamma}+|\tilde{v}_*|^{1+\gamma})(1+|v|^{p-1+\gamma}+|\tilde{v}|^{p-1+\gamma})\varphi_\varepsilon(|v-\tilde{v}|^2) \\ \le& C(1+|v_*|^{p+\gamma}+|\tilde{v}_*|^{p+\gamma})d_{p,\varepsilon}(v,\tilde{v}),\end{aligned}$$

as well as

$$
\begin{aligned}
\mathcal{T}_{3,2,2} \leq & C(1+|v_*|^{1+\gamma}+|\tilde{v}_*|^{1+\gamma})(1+|v|^{p-1+\gamma}+|\tilde{v}|^{p-1+\gamma})\varphi_\varepsilon(|v_*-\tilde{v}_*|^2) \\
\leq & C(1+|v|^{p+\gamma}+|\tilde{v}|^{p+\gamma})d_{p,\varepsilon}(v_*,\tilde{v}_*),
\end{aligned}
$$

and, dropping $\sqrt{\varepsilon}$ and using that $|v_*-\tilde{v}_*| \leq |v_*|+|\tilde{v}_*|$,

$$
\begin{aligned}
\mathcal{T}_{3,2,3} \leq & C(1+|v_*|^{2+\gamma}+|\tilde{v}_*|^{2+\gamma})(1+|v|^{p-1+\gamma}+|\tilde{v}|^{p-1+\gamma})\varphi_\varepsilon(|v-\tilde{v}|^2) \\
\leq & C(1+|v_*|^{p+\gamma}+|\tilde{v}_*|^{p+\gamma})d_{p,\varepsilon}(v,\tilde{v}).
\end{aligned}
$$

Finally, using again the bound $|v_*-\tilde{v}_*| \leq |v_*|+|\tilde{v}_*|$,

$$
\begin{aligned}
\mathcal{T}_{3,2,4} \leq & C\sqrt{\varepsilon}(1+|v|^{p-1+\gamma}+|\tilde{v}|^{p-1+\gamma})(1+|v_*|^{2+\gamma}+|\tilde{v}_*|^{2+\gamma})\varphi_\varepsilon(|v_*-\tilde{v}_*|^2) \\
\leq & C\sqrt{\varepsilon}(1+|v|^p+|\tilde{v}|^p)d_{p+\gamma,\varepsilon}(v_*,\tilde{v}_*).
\end{aligned}
$$

Summing the bounds on $\mathcal{T}_{3,1}$, $\mathcal{T}_{3,2,1}$, $\mathcal{T}_{3,2,2}$, $\mathcal{T}_{3,2,3}$ and $\mathcal{T}_{3,2,4}$, we conclude that

$$
\begin{aligned}
\mathcal{T}_3+\tilde{\mathcal{T}}_3 \leq & C(1+|v_*|^{p+\gamma}+|\tilde{v}_*|^{p+\gamma})d_{p,\varepsilon}(v,\tilde{v}) \\
& + C(1+|v|^{p+\gamma}+|\tilde{v}|^{p+\gamma})d_{p,\varepsilon}(v_*,\tilde{v}_*) \\
& + C\sqrt{\varepsilon}(1+|v_*|^p+|\tilde{v}_*|^p)d_{p+\gamma,\varepsilon}(v,\tilde{v}) \\
& + C\sqrt{\varepsilon}(1+|v|^p+|\tilde{v}|^p)d_{p+\gamma,\varepsilon}(v_*,\tilde{v}_*).
\end{aligned}
\tag{5.90}
$$

Gathering (5.86), (5.87), (5.89) and (5.90) completes the proof since $\varepsilon \in (0,1]$. □

Chapter 6

Large Deviations of the Kac Process

6.1 Introduction & Main Results

This chapter is dedicated to the investigation of the large deviations of the (unlabelled) Kac system. Throughout, we will work with a labelled Kac process (μ^N_t) arising from a labelled process $\mathcal{V}^N_t$. As in the introduction we fix, forever, a time horizon $t_{\text{fin}} \in (0, \infty)$, and write $\bullet$ for processes indexed by $t \in [0, t_{\text{fin}}]$, so that $\mu^N_\bullet = (\mu^N_t)_{0 \le t \le t_{\text{fin}}}$.

For this chapter, it will be convenient to parametrise collisions by the 'ω-representation' from Section 2.4, which we here write as

$$v'(v, v_\star, \sigma) = v - ((v - v_\star) \cdot \sigma)\sigma; \qquad v'_\star(v, v_\star, \sigma) = v_\star + ((v - v_\star) \cdot \sigma)\sigma \tag{6.1}$$

which has the advantage that the map $\mathcal{T}_\sigma : (v, v_*) \to (v', v'_*)$ is a self-inverse linear isometry of $(\mathbb{R}^d)^2$. With this parametrisation, the regularised hard spheres (rHS) and cutoff Maxwell Molecules (GMM) kernels in the introduction are[1]

$$B(v, \sigma) = \begin{cases} 1 + |v| & \text{(rHS)} \\ 1 & \text{(MM)}. \end{cases} \tag{6.2}$$

As usual, we write $\Psi(|v|)$ for the kinetic factor, given by $1 + |v|, 1$ in the two cases respectively. Also uniquely for this chapter, we will not normalise to $\mu^N_t \in \mathcal{S}$, and will take initial velocities sampled independently from a reference measure $\mu^\star_0 \in \mathcal{S}$, which satisfies some further conditions (Hypothesis 6.1 below). For example, while the normalisation procedure in Section 2.3 is innocuous at the level of the law of large numbers, it can make a significant difference at the level of large deviations, and by insisting on independence we have access to Sanov's Theorem [51]. We will comment further on possible hypotheses in the discussion section.

[1]Properly, the kernels with the usual 'σ-representation' are the pushforward of these kernels under the map taking this representation to the usual one. This abuse of terminology will not cause problems in the sequel.

6.1.1 Framework of Large Deviations

We recall that we carefully interpret the informal estimate (1.33) by asking that, for Kac processes $\mu^N_\bullet$ and associated empirical fluxes w^N,

$$\limsup_N \frac{1}{N}\log \mathbb{P}\left((\mu^N_\bullet, w^N)\in\mathcal{A}\right) \le -\inf\left\{\mathcal{I}(\mu_\bullet, w) : (\mu_\bullet, w)\in\mathcal{A}\right\}; \tag{6.3}$$

$$\liminf_N \frac{1}{N}\log \mathbb{P}\left((\mu^N_\bullet, w^N)\in\mathcal{U}\right) \ge -\inf\left\{\mathcal{I}(\mu_\bullet, w) : (\mu_\bullet, w)\in\mathcal{U}\right\} \tag{6.4}$$

for any closed set $\mathcal{A}$ and open set $\mathcal{U}$; in order to state our main results, we must carefully introduce the functional spaces and topology we use, and specify our (candidate) rate function $\mathcal{I}$. Formally, we consider the space $\mathcal{P}_2$ of probability measures on $\mathbb{R}^d$ with finite second moment, equipped with the Monge-Kantorovich-Wasserstein distance $\mathcal{W}_{1,1}$ defined in Section 2.1. We write $\mathcal{P}_2^N$ for the subspace consisting of empirical measures on N points, and $\mathcal{D}$ for the Skorokhod space

$$\mathcal{D} := \left\{\mu_\bullet \in \mathbb{D}([0,t_{\text{fin}}],(\mathcal{P}_2,\mathcal{W}_{1,1})) : \sup_{t\le t_{\text{fin}}}\langle |v|^2,\mu_t\rangle < \infty,\ \mu_{t_{\text{fin}}} = \mu_{t_{\text{fin}}-}\right\} \tag{6.5}$$

which we equip with a metric inducing the Skorokhod J_1-topology; see Appendix 6.A. For the empirical fluxes, we recall, and write throughout, the notation E for the parameter space of collisions $E=(0,t_{\text{fin}}]\times\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}$ and $E_t=(0,t]\times\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}\subset E$. We write $\mathcal{M}(E)$ for the space of finite Borel measures on E satisfying[2] $w(\{t_{\text{fin}}\}\times\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1})=0$, which we equip with the Wasserstein$_1$ metric ρ_1 described in Section 2.1, now with $S=E$ with the metric induced from $E\subset\mathbb{R}^{3d+1}$, which is given by

$$\rho_1(w,w') = \sup\left\{\langle g, w-w'\rangle : \sup_E |g|\le 1,\ \sup_{p,q\in E, p\neq q}\frac{|g(p)-g(q)|}{|p-q|}\le 1\right\} \tag{6.6}$$

where $|\cdot|$ is the Euclidean norm on $E\subset\mathbb{R}^{3d+1}$. We also recall that (μ^N_t, w^N_t) is a Markov process in $\mathcal{P}_2^N\times\mathcal{M}(E)$ with time-dependent generator given on bounded functions by

$$\begin{aligned}\mathcal{G}^N_t F(\mu^N, w^N) = N\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}} &(F(\mu^{N,v,v_\star,\sigma}, w^{N,t,v,v_\star,\sigma}) - F(\mu^N,w^N)) \\ &\cdots\times B(v-v_\star,\sigma)\mu^N(dv)\mu^N(dv_\star)d\sigma\end{aligned} \tag{6.7}$$

where the changes in the measure and the flux are

$$\mu^{N,v,v_\star,\sigma} := \mu^N + \frac{1}{N}\Delta(v,v_\star,\sigma); \qquad w^{N,t,v,v_\star,\sigma} := w^N + \frac{1}{N}\delta_{(t,v,v_\star,\sigma)}. \tag{6.8}$$

where we have introduced the notation

$$\Delta(v,v_*,\sigma) = \delta_{v'}+\delta_{v'_*}-\delta_v-\delta_{v_*}.$$

We will also frequently use the notation

$$\Delta f(v,v_*,\sigma) := \langle f, \Delta(v,v_*,\sigma)\rangle = f(v')+f(v'_*)-f(v)-f(v_*).$$

We will make the following further hypotheses on the reference measure $\mu_0^\star$:

[2] This condition on the empirical flux, and the left-continuity of $\mu^N_\bullet$ at t_{fin}, can be imposed on the sample paths with modification on a set of probability 0.

Hypothesis 6.1. *Let us write $\mathcal{E}_z(\mu)$ for the Gaussian moment functional $\mathcal{E}_z(\mu) := \langle e^{z|v|^2}, \mu\rangle \in [0,\infty]$. Then we assume that $\mu_0^\star \in \mathcal{S}$, and that the following hold.*

i). Gaussian upper bound: *there exists $z_1 > 0$ such that $\mathcal{E}_{z_1}(\mu_0^\star) < \infty$.*

ii). Gaussian Lower Bound: there exists $z_2 < \infty$ such that $\mathcal{E}_z(\mu_0^\star) < \infty$ for $z < z_2$, and $\mathcal{E}_z(\mu_0^\star) \to \infty$ as $z \uparrow z_2$.

iii). Continuous Density: *$\mu_0^\star$ has a continuous density $f_0^\star$ with respect to the Lebesgue measure, and for some $z_3 \in (0,\infty)$ and $c > 0$,*

$$f_0^\star \geq c e^{-z_3|v|^2}. \tag{6.9}$$

Under the initial conditions described above, Sanov's Theorem [51] applies to show that the initial data satisfy a large deviation function in $(\mathcal{P}_2, W)$ with rate function given by the relative entropy

$$H(\mu_0|\mu_0^\star) := \begin{cases} \int_{\mathbb{R}^d} \frac{d\mu_0}{d\mu_0^\star} \log\left(\frac{d\mu_0}{d\mu_0^\star}\right) \mu_0^\star(dv) & \text{if } \mu_0 \ll \mu_0^\star; \\ \infty & \text{else.} \end{cases} \tag{6.10}$$

6.1.2 A Proposed Rate Function

Let us review a candidate rate function identified by Léonard [131] for exactly this problem. For $\mu_\bullet \in \mathcal{D}$ we define $\overline{w}_\mu \in \mathcal{M}(E)$ by

$$\overline{w}_\mu(dt, dv, dv_\star, d\sigma) = B(v - v_*, \sigma) dt \mu_t(dv)\mu_t(dv_\star) d\sigma. \tag{6.11}$$

We say that $(\mu_\bullet, w) \in \mathcal{D} \times \mathcal{M}(E)$ is a *measure-flux pair* if $w \ll \overline{w}_\mu$ and if they solve the *continuity equation*: for all $0 \leq t \leq t_{\text{fin}}$,

$$\mu_t = \mu_0 + \int_{E_t} \Delta(v, v_\star, \sigma)\, w(ds, dv, dv_\star, d\sigma) = \mu_0 + \int_E \Delta(v, v_\star, \sigma)\, 1\!\!1_{\{s\leq t\}}\, w(ds, dv, dv_\star, d\sigma). \tag{CE}$$

We will use, throughout, the notation K for the density $\frac{dw}{d\overline{w}_\mu}$, which we call a tilting function. With this notation, if $(\mu_\bullet, w)$ is a measure-flux pair, then $\mu_\bullet$ solves a controlled Boltzmann equation

$$\mu_t = \mu_0 + \int_{E_t} \Delta(v, v_\star, \sigma) K(s, v, v_\star, \sigma) B(v - v_*, \sigma) ds \mu_s(dv)\mu_s(dv_\star) d\sigma \tag{CBE$_K$}$$

in the sense of Bochner integrals. Equivalently, given $\mu_\bullet$ solving (CBE$_K$) for some $K \in L^1(\overline{w}_\mu)$, one can define $w = K\overline{w}_\mu$ and $(\mu_\bullet, w)$ is a measure-flux pair. We define the dynamic cost of a trajectory $(\mu_\bullet, w) \in \mathcal{D} \times \mathcal{M}(E)$ to be

$$\mathcal{J}(\mu_\bullet, w) := \begin{cases} \int_E \tau\left(\frac{dw}{d\overline{w}_\mu}\right) \overline{w}_\mu(ds, dv, dv_\star, d\sigma) & \text{if } (\mu_\bullet, w) \text{ is a measure-flux pair;} \\ \infty & \text{else} \end{cases} \tag{6.12}$$

where $\tau : [0, \infty] \to [0, \infty]$ is the function $\tau(k) = k \log k - k + 1$, and define the full rate function to be

$$\mathcal{I}(\mu_\bullet, w) := H(\mu_0|\mu_0^\star) + \mathcal{J}(\mu_\bullet, w). \tag{6.13}$$

An analagous upper bound, which can be recovered from this rate function using the contraction principle on $(\mu_\bullet, w) \to \mu_\bullet$ is obtained by Léonard [131] in a different topology, and the same rate function has been found in other contexts for kinetic large deviations; see the comments in the literature review.

6.1.3 Main Results

We are now in a position to carefully state our results, corresponding to Theorem 4. Our first result collects some useful facts on the proposed rate function $\mathcal{I}$ and on the exponential tightness.

Proposition 6.1 (Exponential Tightness and Semicontinuity). *Fix a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$. For $N \geq 2$, let $\mu_\bullet^N$ be either regularised hard sphere or Maxwell Molecule Kac processes with initial velocities drawn independently from a measure $\mu_0^\star$ satisfying Hypothesis 6.1i). Then the following hold.*

i). The random variables $(\mu_\bullet^N, w) \in \mathcal{D} \times \mathcal{M}(E)$ are exponentially tight: for any $M > 0$, there exists a compact set $\mathcal{K} \subset \mathcal{D} \times \mathcal{M}(E)$ such that

$$\limsup_N \frac{1}{N} \log \mathbb{P}\left((\mu_\bullet^N, w^N) \notin \mathcal{K}\right) \leq -M. \tag{6.14}$$

ii). The function $\mathcal{I}$ is lower semicontinuous on $\mathcal{D} \times \mathcal{M}(E)$: the lower sub-level sets

$$\{(\mu_\bullet, w) \in \mathcal{D} \times \mathcal{M}(E) : \mathcal{I}(\mu_\bullet, w) \leq a\} \subset \mathcal{D} \times \mathcal{M}(E) \tag{6.15}$$

are closed.

We emphasise that we do not claim that $\mathcal{I}$ is 'good', in that the sub-level sets are compact; indeed, Theorem 6.5 suggests that this is false.

Our main positive result is as follows. We rederive, in our context, the upper bound with our rate function, which reproduces the result of Léonard [131], and prove a lower bound with the same rate function on a restricted set. In this way, the proposed rate function captures at least some of the correct large deviation behaviour of the Kac process.

Theorem 6.2. *Let B be either the regularised hard spheres or Maxwell molecules kernel, and for $N \geq 2$ let $(\mu_\bullet^N, w^N)$ be a Kac process and its flux, with particles drawn initially from $\mu_0^\star$ satisfying Hypothesis 6.1, and let $\mathcal{I}$ be the rate function given above. Then* *i). For all* $\mathcal{A} \subset \mathcal{D} \times \mathcal{M}(E)$ *closed, we have*

$$\limsup_N \frac{1}{N} \log \mathbb{P}\left((\mu^N_\bullet, w^N) \in \mathcal{A}\right) \leq -\inf\left\{\mathcal{I}(\mu_\bullet, w) : (\mu_\bullet, w) \in \mathcal{A}\right\}. \tag{6.16}$$

ii). For all $\mathcal{U} \subset \mathcal{D} \times \mathcal{M}(E)$ *open, we have*

$$\liminf_N \frac{1}{N} \log \mathbb{P}\left((\mu^N_\bullet, w^N) \in \mathcal{U}\right) \geq -\inf\left\{\mathcal{I}(\mu_\bullet, w) : (\mu_\bullet, w) \in \mathcal{U} \cap \mathcal{R}\right\} \tag{6.17}$$

where $\mathcal{R} = \{(\mu_\bullet, w) \in \mathcal{D} \times \mathcal{M}(E) : \langle 1 + |v|^2 + |v_\star|^2, w\rangle < \infty\}$.

Next, let us show some applications of the positive large deviations result at the level of nonrandom $(\mu_\bullet, w) \in \mathcal{D} \times \mathcal{M}(E)$. We first show how the entropy plays the role of a *quasipotential* for the Kac dynamics and Boltzmann equation. Let us refer the reader to [31, Section 3.3] for a general discussion of such results.

Corollary 6.3 (Entropy as a Quasipotential)**.** *Let* B *be either the regularised hard spheres or Maxwell molecules kernel, and fix* $\mu \in \mathcal{P}_2$*. Then*

$$H(\mu|\gamma) \geq \inf\left\{H(\nu_0|\gamma) + \int_E \tau(K) d\overline{w}_\nu : \nu \in \mathcal{D}, \nu \text{ solves } (\mathrm{CBE}_K), \nu_{t_{\mathrm{fin}}} = \mu\right\} \tag{6.18}$$

and

$$H(\mu|\gamma) \leq \inf\left\{H(\nu_0|\gamma) + \int_E \tau(K) d\overline{w}_\nu : \nu \in \mathcal{D}, \nu \text{ solves } (\mathrm{CBE}_K), \nu_T = \mu, \right. \\ \left. \text{and } \int_E (|v|^2 + |v_\star|^2) K d\overline{w}_\nu < \infty\right\}. \tag{6.19}$$

In this sense, we view $\tau(K)$ as the entropic cost of moving to a higher-entropy state by controlled Boltzmann dynamics (CBE_K). The second item also implies a weak form of the H-Theorem (H), as it shows that $H(\mu_t|\gamma)$ is strictly decreasing along solutions $\mu_\bullet$ to (BE) such that $(\mu_\bullet, \overline{w}_\mu) \in \mathcal{R}$, but the upper bound would be false without this second moment condition due to Theorem 6.5 below.

Our second application is a short result concerning the *time-reversal* of the large deviations theory we construct. We define the time reversal operation on $(\mu_\bullet, w) \in \mathcal{D} \times \mathcal{M}(E)$ by writing $\mathbb{T}(\mu_\bullet, w) = (\mathbb{T}\mu_\bullet, \mathbb{T}w)$, where the time reversal of the path is

$$\mathbb{T}\mu^N_\bullet = \left(\mu^N_{(t_{\mathrm{fin}} - t)-}\right)_{0 \leq t \leq t_{\mathrm{fin}}}; \tag{6.20}$$

and the time reversal of the flux w is given by specifying, for bounded measurable $g : E \to \mathbb{R}$,

$$\int_E g(t, v, v_*, \sigma)(\mathbb{T}w)(dt, dv, dv_*, d\sigma) := \int_E g(t_{\mathrm{fin}} - t, v', v'_*, \sigma) w(dt, dv, dv_*, d\sigma). \tag{6.21}$$

It is well-known that the Kac process is reversible in equilibrium, so that if $\mu^N_\bullet$ is formed by sampling independently from γ, then the law of $\mathbb{T}\mu^N_\bullet$ is the same as that of $\mu^N_\bullet$. Our result on the time-reversibility of large deviations is as follows.

Proposition 6.4 (Time-Reversibility of Large Deviations). *Let $\mu^N_\bullet$ be a Kac process for either (rHS, GMM) formed by sampling initial velocities independently from γ and let w^N be its associated empirical flux.*

i). The law of the time-reversal $\mathbb{T}(\mu^N_\bullet, w^N)$ is the same as that of $(\mu^N_\bullet, w^N)$.

ii). For any $(\mu_\bullet, w) \in \mathcal{R}$, we have $\mathcal{I}(\mathbb{T}\mu_\bullet, \mathbb{T}w) = \mathcal{I}(\mu_\bullet, w)$, for the same set $\mathcal{R}$ as in Theorem 6.5.

iii). Let $(\mu_\bullet, w) \in \mathcal{D} \times \mathcal{M}(E)$ be a measure-flux pair, with tilting function K, such that μ_t, μ_{t-} all admit strictly positive densities $f_t, f_{t-} > 0$ with respect to the Lebesgue measure for all $t \in [0, t_{\text{fin}}]$. Then the time-reversal $(\mathbb{T}\mu_\bullet, \mathbb{T}w)$ is a measure-flux pair with the new tilting function

$$\mathbb{T}K(t, v, v_*, \sigma) = K(t_{\text{fin}} - t, v', v'_*, \sigma)\frac{f_{t_{\text{fin}}-t}(v')f_{t_{\text{fin}}-t}(v'_*)}{f_{t_{\text{fin}}-t}(v)f_{t_{\text{fin}}-t}(v_*)}.$$

iv). Let $(\mu_\bullet, \overline{w}_\mu) \in \mathcal{R}$ be an energy-conserving solution to (BE) and its associated flux, and suppose that μ_t admits a strictly positive density f_t for all $t \in [0, t_{\text{fin}}]$. Then the dynamic cost of the time-reversal is

$$\mathcal{J}(\mathbb{T}\mu_\bullet, \mathbb{T}\overline{w}_\mu) = \int_0^{t_{\text{fin}}} D(f_t)dt = \int_E \log\left(\frac{f_t(v)f_t(v_*)}{f_t(v')f_t(v'_*)}\right)\overline{w}_\mu(dt, dv, dv_*, d\sigma)$$

where $D(f_t)$ is the entropy dissipation, defined for nonnegative $f \in L^1(dv)$ by

$$D(f) = \int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}} \log\left(\frac{f(v)f(v_*)}{f(v')f(v'_*)}\right) B(v - v_*, \sigma)f(v)f(v_*)dvdv_*d\sigma.$$

We therefore conclude Boltzmann's H-Theorem: for such solutions,

$$H(\mu_{t_{\text{fin}}}|\gamma) + \int_0^{t_{\text{fin}}} D(f_s)ds = H(\mu_0|\gamma).$$

As discussed in the literature review, the proposed rate function is a very natural candidate for describing the large deviations behaviour, and one might expect to be able to find a 'true' lower bound (6.4). In this context, the restricted lower bound presented here is somewhat dissatisfying, as it leaves open the question of which open sets $\mathcal{U}$ are such that $\inf_{\mathcal{U}} \mathcal{I} = \inf_{\mathcal{U}\cap\mathcal{R}} \mathcal{I}$, or the possibility that a better upper bound may be possible. The restriction to a set $\mathcal{R}$ of 'good' paths, as here, is necessary for the paths in question to be approximated by paths which can be recovered by a Girsanov transform; see Lemma 6.16. Key to this argument is that these paths should be uniquely specified by the initial data and tilting K, so that the path is the unique possible hydrodynamic limit of 'tilted' dynamics along any subsequence. However, at the level of the Boltzmann equation (BE), this uniqueness is known *not* to hold: solutions with increasing energy have been constructed by Lu and Wennberg [133]. Since the energy $\langle |v|^2, \mu^N_t\rangle$ is almost surely conserved by the paths of the stochastic Kac process, one might hope that such solutions are somehow spurious and can be excluded from the large deviation analysis by redefining $\mathcal{I}$ to be ∞ on such paths. However, we prove the following theorem, which shows that such solutions can be reached with finite exponential cost, and so cannot be excluded from the large deviation analysis, but the occurrence of such paths is not correctly predicted by the proposed rate function. Equivalently, this example can be understood as producing explicitable open sets $\mathcal{U}$ such that the infima of the rate function over $\mathcal{U}$ and $\mathcal{U} \cap \mathcal{R}$ do not coincide.

Theorem 6.5. *Assume the notation of Proposition 6.1.*

i). Suppose B is the regularised hard spheres kernel, and the reference measure $\mu_0^\star$ satisfies Hypothesis 6.1i-ii). Let $\Theta : [0, t_{\mathrm{fin}}] \to (0, \infty)$ be increasing and left-continuous, with $\Theta(0) = 1$ and such that, for some closed set $P \subset [0, t_{\mathrm{fin}}], 0 \in P, t_{\mathrm{fin}} \notin P$ with null interior, Θ is locally constant on $[0, t_{\mathrm{fin}}] \setminus P$. For some constant $\alpha = \alpha(\Theta(t_{\mathrm{fin}}))$, define

$$A(t) := \alpha \left(\inf_{s \in P : s < t} (t - s) \right)^{-2} \in (0, \infty] \tag{6.22}$$

and consider the set $\mathcal{A}_\Theta$ given by

$$\mathcal{A}_\Theta := \Big\{ (\mu_\bullet, w) \in \mathcal{D} \times \mathcal{M}(E) : \mu_\bullet \text{ solves (BE)},\ w = \overline{w}_\mu,\ \mu_0 = \mu_0^\star, \\ \text{and for all } t \geq 0,\ \langle |v|^2, \mu_t \rangle = \Theta(t) \text{ and } \langle |v|^4, \mu_t \rangle \leq A(t) \Big\}. \tag{6.23}$$

Then $\mathcal{A}_\Theta$ is compact, nonempty, and $\mathcal{I}(\mu_\bullet, w) = 0$ on $\mathcal{A}_\Theta$. We have, for the same z_2 as in Hypothesis 6.1,

$$\inf_{\mathcal{U} \supset \mathcal{A}_\Theta} \liminf_N \frac{1}{N} \log \mathbb{P}\left((\mu_\bullet^N, w^N) \in \mathcal{U} \right) \geq -\Theta(t_{\mathrm{fin}}) z_2 \tag{6.24}$$

where the infimum runs over all open sets $\mathcal{U} \subset \mathcal{D} \times \mathcal{M}(E)$ containing $\mathcal{A}_\Theta$, and there exists an open set $\mathcal{V} \supset \mathcal{A}_\Theta$ such that

$$\liminf_N \frac{1}{N} \log \mathbb{P}\left((\mu_\bullet^N, w^N) \in \mathcal{V} \right) < 0. \tag{6.25}$$

ii). Suppose instead that B is the cutoff Maxwell Molecules kernel. For all $\delta > 0$ and Θ as above, and with A as above with α depending on δ as well as $\Theta(t_{\mathrm{fin}})$, define the set

$$\mathcal{A}_{\Theta,\delta} := \Big\{ (\mu_\bullet, w) \in \mathcal{D} \times \mathcal{M}(E) : (\mu_\bullet, w) \text{ is a measure-flux pair with} \\ K(t, v, v_\star, \sigma) = 1 + \delta |v - v_\star|, \mu_0 = \mu_0^\star, \text{ and for all } t > 0, \\ \langle |v|^2, \mu_t \rangle = \Theta(t_{\mathrm{fin}}) \text{ and } \langle |v|^4, \mu_t \rangle \leq A(t) \Big\}. \tag{6.26}$$

The sets $\mathcal{A}_{\Theta,\delta}$ are compact and nonempty, and $\mathcal{I}(\mu_\bullet, w) \le 4\delta^2\Theta(t_{\text{fin}})t_{\text{fin}}$ for all $(\mu_\bullet, w) \in \mathcal{A}_{\Theta,\delta}$. We have

$$\inf_{\mathcal{U}\supset\mathcal{A}_{\Theta,\delta}} \liminf_N \frac{1}{N}\log \mathbb{P}\left((\mu^N_\bullet, w^N)\in\mathcal{U}\right) \ge -\Theta(t_{\text{fin}})(z_2 + C\delta) \tag{6.27}$$

where, as above, the infimum runs over all open sets $\mathcal{U}$ containing $\mathcal{A}_{\Theta,\delta}$. However, there exist open sets $\mathcal{V}_\delta \supset \mathcal{A}_{\Theta,\delta}$ such that, for any Θ,

$$\limsup_{\delta\downarrow 0}\, \liminf_N \frac{1}{N}\mathbb{P}\left((\mu^N_\bullet, w^N)\in\mathcal{V}_\delta\right) < 0. \tag{6.28}$$

In both cases, the first point shows that such behaviour cannot be excluded by superexponential estimates, and so is a form of behaviour which must be taken into account in the large deviation theory; the second point shows that the rate function on such paths is not that predicted above. In the first case, we will construct changes of measure $\mathbb{Q}^N \ll \mathbb{P}$, with an exponential cost associated to changing the initial data and sub-exponential cost associated to modifying the dynamics, so that $o(N)$ particles containing $\mathcal{O}(1)$ energy are temporarily 'frozen' and, under these new measures, the Kac processes concentrate on the set $\mathcal{A}_\Theta$ given. The argument for the Maxwell molecule case is similar, with an additional exponential cost $\mathcal{O}(e^{N\delta})$ necessary to modify the dynamics. In these cases, the behaviour of $o(N)$ particles has a macroscopic effect on the evolution of the whole process, meaning that the large deviation behaviour is not purely captured by the empirical measure and control K. Possible generalisations of this phenomenon will be discussed in Section 6.1.5.2 below

Let us now examine some consequences of the counterexamples presented in Theorem 6.5. One might hope that it is possible to prove a true large deviation principle under well-chosen initial conditions where one puts in 'by hand' that there is no such concentration initially. The following easy corollary suggests that, even under such well-chosen conditions, the same concentration of energy can occur as a result of the binary collisions.

Corollary 6.6. *Let us take $\mu_0^\star = \gamma$, and fix a decreasing, right-continuous function Θ, $\Theta(t_{\text{fin}}) = 1$, which is locally constant aside from at a closed set $P \subset [0, t_{\text{fin}}]$ with empty interior, $t_{\text{fin}} \in P, 0 \notin P$. For either Maxwell molecules or hard spheres, there exists an explicitable function A such*

$$\mathcal{B} = \left\{(\mu_\bullet, w)\in\mathcal{D}\times\mathcal{M}(E) : \langle |v|^2, \mu_t\rangle = \Theta(t) \text{ for all } t\in[0,t_{\text{fin}}] \text{ and } \langle |v|^4, \mu_t\rangle \le A(t)\right\} \tag{6.29}$$

satisfies

$$\inf_{\mathcal{U}\supset\mathcal{B}} \liminf_N \frac{1}{N}\log\mathbb{P}\left((\mu^N_\bullet, w^N)\in\mathcal{U}\right) > -\infty. \tag{6.30}$$

where, as above, the outer infimum runs over open $\mathcal{U}\subset\mathcal{D}$ containing $\mathcal{B}$.

As a result, it is not possible to find a superexponential estimate to prevent the accumulation of energy in $o(N)$ particles at future times.

Since the stochastic processes $\mu^N_\bullet$ are exponentially tight in $\mathcal{D}$, it follows that one can extract subsequences satisfying a *true* large deviation principle in the sense of (6.3, 6.4). As a consequence of Theorem 6.5, no such subsequence can avoid the bad paths we have constructed.

Corollary 6.7. *Let $\mu_0^\star$ be a reference measure satisfying Hypothesis 6.1i-ii), and let $\mu^N_\bullet, w^N$ be N-particle Kac processes, either for the regularised hard spheres or Maxwell molecules case. Suppose that $L \subset \mathbb{N}$ is an infinite subsequence such that $(\mu^N_\bullet, w^N)_{N\in L}$ satisfy a large deviation principle in $\mathcal{D} \times \mathcal{M}(E)$ with some rate function $\widetilde{\mathcal{I}}$. Then there exists $(\mu_\bullet, w)$ in $\mathcal{D} \times \mathcal{M}(E)$ such that $\widetilde{\mathcal{I}}(\mu_\bullet, w) < \infty$ but such that $t \mapsto \langle |v|^2, \mu_t\rangle$ is not constant.*

6.1.4 Plan of the Chapter

This chapter is structured as follows.

i). In the remainder of this section, we will review some recent works on large deviations and related problems, and make some remarks on the hypothesis and functional framework.

ii). In Section 6.2, we derive an upper bound Theorem 6.2i); in doing so, we will prove Proposition 6.1, using a variational representation of the rate function $\mathcal{I}$ for the lower semi-continuity.

iii). Section 6.3 presents, without detailed proof, a change-of-measure formula for the Kac process which we will use in the lower bound and in Theorem 6.5.

iv). Section 6.4 proves the restricted lower bound Theorem 6.2ii), based on an approximation argument for paths belonging to $\mathcal{R}$ and a standard 'tilting' argument.

v). Section 6.5 gives the proof of Corollary 6.3 and Proposition 6.4 based on Theorem 6.5.

vi). The proof of Theorem 6.5 is given in Section 6.6, based on the moment properties of the Kac process in Section 2.5 and a careful analysis of Cramér bounds.

vii). Section 6.7 gives the proof of the negative applications Corollary 6.7, and Corollary 6.6, which is based on the same time-reversal principle as Proposition 6.4.

viii). Finally, Appendix 6.A is a self-contained appendix on the Skorokhod topology and Appendix 6.B contains a justification of the change-of-measure formula.

6.1.5 Literature Review & Discussion

1. Large Deviations for Jump Particle Systems The theory of large deviations for Markov processes in the small-noise limit goes back to Freidlin and Wentzell [94]. The seminal work of Feng & Kurz [77] developed tools based on a comparison principle for Hamilton-Jacobi equation in infinite dimensions, which are general but hard to verify. The analysis is somewhat different in the case where the dynamics are driven by diffusive rather than jump noise, see the discussion in Léonard [131]. In this context let us mention the recent works [15, 32, 33, 152].

Within collisional kinetic theory, previous works have reported upper bounds of a similar form. The work of Léonard [131] already cited considers the same case of the energy-preserving Kac model, and produces a rate function exactly given by $\mathcal{I}(\mu_\bullet) = \inf_w \mathcal{I}(\mu_\bullet, w)$, albeit for a different topology. Rezakhanlou [166] considers a collisional model for a spatially inhomogeneous gas, where the positions take values in the unit circle $\mathbb{R}/\mathbb{Z}$ and the velocities take values only in a finite set, and finds an upper bound and a restricted lower bound, where the infimum runs only over a subset $\mathcal{R} \cap \mathcal{U}$ as in Theorem 6.2 rather than over the full open set $\mathcal{U}$ as in (6.4). The rate function is analagous to the variational form (c.f. Lemma 6.11 or [131, Theorem 7.1]). Bodineau et al. [23] consider the full spatially inhomogeneous Boltzmann–Grad limit with random initial data and deterministic dynamics for local interactions; the rate function is again given in a variational form, and the lower bound is again restricted to sufficiently good paths. Most recently, a very similar large deviation result for a range of 'Kac–like' processes has been found by Basile et al. [16]. The lower bound is again of the restricted form as in Theorem 6.2, and is proven only in cases where the collisions preserve momentum but not energy.

Outside of kinetic theory, analagous rate functions have been found for large deviations of jump processes, for instance [62]. A number of works [160, 159, 165] have considered the case of 'reaction networks', which formally includes the Kac/Boltzmann dynamics considered here by viewing $\mathbb{R}^d$ as a continuum of particle species; these works are the origin of considering the pair $(\mu^N_\bullet, w^N)$ which significantly eases the analysis. Other works [66, 124] in the context of reaction networks or mean-field dynamics exploit a control representation of the dynamics, leading to an equation similar to (CBE$_K$) with random controls K and the same cost function τ, and using weak convergence method due to Dupuis [65]. In this weak convergence method, it is essential that the control uniquely determines possible limiting paths (see a similar argument in the proof of Theorem 6.5ii) in Section 6.4), whereas this type of uniqueness is known not to be hold in the Boltzmann case, even in the most advantageous possible case of Maxwell molecules. In the work [124], the key to removing the restriction on regular paths is an approximation argument so that paths are perturbed to lie in the *interior* of the space of probability measures on the space of the finite space of species S, which is clearly impossible in the infinite-dimensional setting here.

Let us mention that this is a very natural form for the rate function jump processes. One recognises τ as the dynamic cost of controlling a Poisson random measure: for a Poisson process of unit intensity, a straightforward argument using Stirling's formula shows that $\mathbb{P}(N^{-1}Z_N \approx z) \asymp \exp(-N\tau(z))$ in the same sense as (6.3, 6.4). Similarly, if one fixes a finite space S and a probability measure μ and forms X_N as a Poisson random measure of intensity $N\mu$, then one has the equivalent

$$\mathbb{P}\left(N^{-1}X_N \approx \nu\right) \asymp \exp\left(-N\sum_S \tau\left(\frac{\nu(x)}{\mu(x)}\right)\mu(x)\right) = \exp\left(-N\int_S \tau\left(\frac{d\nu}{d\mu}\right)\mu(dx)\right). \tag{6.31}$$

The proposed rate function above would correspond to the intuition that, given $\mu_t^N \approx \mu_t$, the instantaneous distribution of jumps is approximately Poisson, with intensity $\approx B(v-v_*, \sigma)dt\mu_t(dv)\mu_t(dv_\star)d\sigma$.

As remarked above, several other works [62, 166, 23, 16] have encountered the same problem that the lower bound can only be proven over a class of good paths. Both the works [62, 166] conjecture that a 'true' lower bound should hold in the respective frameworks. In the works cited above, such a conjecture has only been proven in the cases of reaction networks with a finite set of species [124, 160, 159, 165] or mean-field dynamics with finite state space [66], which are very far from the Kac/Bolzmann dynamics we consider. To the best of our knowledge, the current work represents the first time that such a hypothesis has been falsified. It is also interesting to note that this is *different* from what one finds in the Freidlin-Wentzell theory of stochastic scalar conservation laws, see [136]. In this case, the limiting scalar conservation law admits many weak solutions, and a large deviation principle holds with 0 rate on all solutions to the limiting equation. For this case, the cost of approximating a solution which is not the unique entropic solution is subexponential, whereas in the Kac process we consider here, there is a nonzero exponential cost required to reach the non-energy-conserving solutions to (BE).

2. Remarks on the Hypotheses & Functional Framework We make the following remarks on the Hypothesis 6.1 and on the functional framework. Firstly, the hypotheses allow the very natural choice of taking $\mu_0^\star$ to be the equilibrium distribution

$$\mu_0^\star(dv) = \gamma(dv) = \frac{1}{(2\pi d^{-1})^{d/2}} e^{-d|v|^2/2} dv \tag{6.32}$$

but Hypothesis 6.1ii). disallows measures of the form $\mu_0^\star(dv) \propto (1+|v|^2)^{-m}\gamma(dv), m > \frac{d}{2}$. In general, the condition that μ_0^N be given by drawing particles independently from a reference measure $\mu_0^\star$ will not propagate in time. However, this is natural in order to ensure that μ_0^N satisfies a large deviation principle; elementary counterexamples can be found to show that the more usual conditions, that the initial data be chaotic or entropically chaotic [108], do not imply a large deviation principle for μ_0^N. Moreover, in the most important case $\mu_0^\star = \gamma$, the independence *is* propagated, as $\gamma^{\otimes N}$ is an equilibrium distribution for the N-particle dynamics.

Regarding the functional framework, while $(\mathcal{P}_2, \mathcal{W}_{1,1})$ is not complete, the choice of metric $\mathcal{W}_{1,1}$ and Skorokhod space $\mathcal{D}$ are natural to guarantee that $(\mu_\bullet^N, w^N)$ are exponentially tight. One could alternatively equip $\mathcal{P}_2$ with the Wasserstein$_2$ metric $\mathcal{W}_2$ or the weighted metrics W_γ defined in Section 2.1. In this case, the map $\mu \mapsto \langle |v|^2, \mu\rangle$ continuous, and one can take a limit of the pathwise energy conservation $\langle |v|^2, \mu_t^N\rangle = \langle |v|^2, \mu_0^N\rangle$ to conclude that all possible large deviation paths still conserve energy. However, carefully following the arguments leading to our counterexamples in Section 6.6 proves that the initial measures μ_0^N then fail to be exponentially tight, as does the whole process $(\mu_\bullet^N, w^N)$. Since large deviations techniques rely heavily on such tightness to prove the existence of subsequential limits under the change of measures, we have been unable to determine $\mathcal{I}$ correctly determines the large deviations in this framework. In light of this, we interpret Theorem 6.5 as showing the existence of a different kind of large deviations behaviour, where macroscopic energy concentrates in $o(N)$ particles, which is not captured by convergence in $(\mathcal{P}_2, \mathcal{W}_2)$.

In future works, it may be interesting to consider the large deviations in the functional framework similar to that of Léonard [131]. Recalling the notation $\mathcal{A}_0$ for the continuous functions of quadratic growth on $\mathbb{R}^d$, we write $\mathfrak{P}_2$ for the space of linear maps $\mathfrak{m} : \mathcal{A}_0 \to \mathbb{R}$ satisfying $\mathfrak{m}(1) = 1$, $\mathfrak{m}(f) \ge 0$ whenever $f \ge 0$, and such that there exists $\mu = j[\mathfrak{m}] \in \mathcal{P}_2$ with $\langle f, \mu\rangle = \mathfrak{m}(f)$ for all bounded $f \in C_b(\mathbb{R}^d)$. We then equip $\mathfrak{P}_2$ with the product topology from the inclusion $\mathfrak{P}_2 \subset \mathbb{R}^{\mathcal{A}_0}$, and we can view $\mathcal{P}_2 \subset \mathfrak{P}_2$ via the identification $\iota : \mathcal{P}_2 \to \mathfrak{P}_2$, $\iota(\mu)(\varphi) := \langle \varphi, \mu\rangle$, so that the Kac process can be understood as taking values in $\mathfrak{P}_2$. Moreover, thanks to the classical Tychonoff theorem, the sets

$$\mathfrak{K} = \{\mathfrak{m} \in \mathfrak{P}_2 : \text{ for all } f \in \mathcal{A}_0, |\mathfrak{m}(f)| \le a\} \tag{6.33}$$

are compact for all $a \in [0, \infty)$. In this framework, one has both exponential tightness, and continuity of the map $\mathfrak{m} \mapsto \mathfrak{m}(|v|^2)$. On the other hand, we warn the reader that elements of $\mathfrak{P}_2$ are typically *not* measures, since $j[\mathfrak{m}] = \mu$ does not imply that $\mathfrak{m} = \iota(\mu)$. Indeed, following the construction of the initial data in Section 6.6 leading to Theorem 6.5 produces limits with $\mathfrak{m}(|v|^2) = \Theta(t_{\text{fin}}) > \langle |v|^2, j[\mathfrak{m}]\rangle = 1$.

3. Generality of the Phenomenon In keeping with the rest of the thesis, we consider *only* the Kac collisional process. However, we remark that the key ingredients of the counterexample Theorem 6.5 may generalise to other large deviation systems. Although we will not explore the general case in more detail, the key points we require generalse to an interacting particle system on a state space S as follows:

1. **Conserved Quantity:** There exists $\varphi : S \to [0, \infty)$ such that, almost surely, $\langle \varphi, \mu_t^N \rangle$ is constant along sample paths;

2. **Criticality:** The initial distributions are such that the limit

$$F(z) := \lim_N N^{-1} \log \mathbb{E}[e^{Nz\langle \varphi, \mu_0^N \rangle}]$$

exists in $[0, \infty]$ for all z. Moreover, the function $F(z)$ is finite on a neighbourhood I of the origin, but diverges to infinity as $z \uparrow \sup I < \infty$.

3. **Delocalisation Mechanism:** For some ψ with $\psi \geq \varphi, \sup \psi/\varphi = \infty$, one **either** has

   (a) Uniformly in N, for all $t > 0$ and all starting points μ_0^N, $\mathbb{E}\langle \psi, \mu_t^N \rangle < \infty$ can be controlled only in terms of t and $\langle \varphi, \mu_0^N \rangle$, uniformly in compact subsets of $t \in (0, \infty)$; **or**

   (b) For some μ_0^N and for all $\delta > 0$, one can find changes of measure $\mathbb{Q}^N \ll \mathbb{P}$ by modifying only the dynamics, such that $\sup_N \mathbb{E}_{\mathbb{Q}^N} \langle \psi, \mu_t^N \rangle$ can be controlled in terms of $t, \langle \varphi, \mu_0^N \rangle, \delta$, uniformly in N, uniformly in compact subsets of $t \in (0, \infty)$, and the perturbation is small in the sense that $\mathbb{Q}^N(\frac{d\mathbb{Q}^N}{d\mathbb{P}} > e^{N\delta a}) < \frac{1}{2}$ for N large enough, for some a depending only on $\langle \varphi, \mu_0^N \rangle$.

In our case, the conserved quantity φ is the energy $\varphi(v) = |v|^2$, and this would apply to any system with stochastic, energy-preserving dynamics. The second point is natural for Gibbs distributions in statistical mechanics with no interaction potential, where the density with respect to some Lebesgue measure is given by $\propto e^{-H} = e^{-N\langle \varphi, \mu_0^N \rangle}$. The third point says that, potentially under a small perturbation of the dynamics, the system rapidly distributes φ among all particles. Let us remark that, thanks to item 2, no exponential moments for $\langle \psi, \mu_0^N \rangle$ can be hoped for, so that bounds on $\mathbb{E}_{\mathbb{Q}^N} \langle \psi, \mu_t^N \rangle$ will not hold under typical changes of measure $\mathbb{Q}^N \ll \mathbb{P}$: we only ask that one such change of measure can be found. In our case, this rôle will be played by the moment creation property and Povzner estimates with $\psi = |v|^p, p > 2$, see Proposition 2.10; case a) corresponds to regularised hard spheres, and case b) to Maxwell molecules.

In either case, since both φ is necessarily unbounded, the dynamics cannot only be captured by a weakly continuous function of the empirical measure. Item 2 allows cases where a macroscopic pertubation of $\langle \varphi, \mu_0^N \rangle$ is achieved with only a small pertubation of μ_0^N in the weak topology, and playing the pathwise conservation (item 1) against the delocalisation mechanism (item 3) instantaneously spreads this pertubation to the whole empirical measure. This leads to a law of large numbers for paths $\mu_\bullet$ along which $\langle \varphi, \mu_t \rangle$ is not conserved, and is a given, nonconstant function $\Theta(t)$ which is constant aside from a jump discontinuity at 0; more general Θ could be found with further assumptions on the dynamics. The dynamic cost required for such paths is either 0, or $e^{\mathcal{O}(N\delta)}$, by following exactly the arguments in Section 6.6; the conclusion that the large deviation occurrence of such limit paths is not correctly predicted by the naïve rate function then follows by exploiting the conflict between the non-conservative limit paths and conservative finite-N paths (item 1).

4. Relationship to Other Problems We mention some other aspects of the Boltzmann/Kac dynamics which are related to the current work.

4a. Boltzmann's H-Theorem and Relaxation to Equilibrium As already mentioned above, large deviations give a probabilistic meaning to Boltzmann's Entropy functional $H(\cdot|\gamma)$; the H-Theorem, which guarantees that entropy increases along solutions to (BE) or its spatially inhomogeneous version, goes back to the foundations of kinetic theory; quantitative versions of this increase, and hence qualifying the convergence to equilibrium, have been a major topic in the analysis of the Boltzmann Equation (among many others, [43, 181, 21, 188, 190, 57]). Let us also mention the work of Mischler and Mouhot [142], which gives a probabilistic proof of the H-theorem via entropic chaos of the Kac process; however, as remarked above, entropic chaos does not lead to the large deviations considered here. Proposition 6.4 gives (another) proof of the well-known H-Theorem, based on large deviations; it is also satisfying that we can give a large-deviations meaning to the entropy dissipation functional $D(f_t)$ as the dynamic cost of reversing the Boltzmann path, see also paragraph 5 below.

4b. Gradient Descent Following the seminal work of Jordan, Kinderlehrer and Otto [120], it has been shown that many equations of mean-field type can be understood as the gradient flow of the entropy for a metric adapted to the particular problem, so that the dynamics not only increase entropy, but do so in the most efficient way possible. Further, it is known that such gradient flow properties can be derived from large deviation principles [3, 64, 2, 140, 74]. Since such a gradient descent formulation of the Boltzmann equation is already known [73, 31, 16], we will not explore this here.

4c. Energy Non-Conserving Solutions to the Boltzmann Equation The proof of Theorem 6.5 follows the construction of energy non-conserving solutions to the Boltzmann equation by Lu and Wennberg [133], using the moment production properties of the Kac process (in Proposition 2.10) in place of those of the Boltzmann equation, now keeping track of the exponential change of measure necessary to obtain such paths as large deviations. The conclusion of Theorem 6.5 is exactly that such paths cannot be avoided when attempting to classify all events of exponentially small probability $\mathbb{P}(A_N) \geq e^{-cN}$, and that the occurrence of such paths is strictly more rare than predicted by the candidate rate function $\mathcal{I}$.

Concerning the negative result Theorem 6.5, identical arguments would hold with the true hard spheres kernel (HS), the cutoff hard potentials (CHP_K), or with some modification for noncutoff Maxwell molecules. The difficulty with the cases (HS, CHP_K) in this chapter lies exclusively in the proof of the lower bound of Theorem 6.2ii), where we need an approximation argument; in these cases, the kinetic factor Ψ vanishes near 0, and $\log \Psi(|v - v_*|)$ fails to be continuous near the diagonal $v = v_*$. It is possible that our arguments in Section 6.4 could be modified to deal with these cases, but this would require refinement of an already difficult proof. In the case of *non*cutoff hard potentials (NCHP), the framework of the empirical flux no longer makes sense, since w is almost surely an infinite measure, and one must be more careful with the formulation of the continuity equation, but the same arguments leading to a proof of Theorem 6.5, now phrased only in terms of the empirical measures, would remain valid.

5. Reversibility Let us refer to the recent work of Bouchet [31] which discusses the classical paradox of reversibility with an analysis based on large deviations, with a rate function analagous to (6.13) above. The same paper also studies the properties the entropy $H(\cdot|\gamma)$ as the quasipotential driving the Boltzmann dynamics for the inhomogeneous case. Bouchet remarked that the relevant rate function is reversible in time, and proposed that the irreversibility of the Boltzmann equation arises from considering *only* the "most probable" evolution or the law of large numbers rather than all possible evolutions. Let us remark that something similar happens in our case; Proposition 6.4 shows that, when one reverses a Boltzmann path, some of the overall cost is moved from the cost $H(\cdot|\gamma)$ of the initial data onto the dynamical cost, which is coherent with Corollary 6.3. At the probabilistic level, starting from equilibrium, one obtains solutions to the Boltzmann equation with $\mu_0 \neq \gamma$ by making a change of measure depending only on μ_0^N, whereas if one applies the same change of measure at some future time $t > 0$, the same overall change of measure is split between an initial cost and a dynamic cost, and one finds limiting paths which instead satisfy a controlled Boltzmann equation (CBE_K) up to time t.

6. Comparison to Other Parts of the Thesis We remark that this chapter has a somewhat different nature from those preceeding. In Chapters 3-4 in particular, we had the freedom to play with as many polynomial moments as were necessary, and even in Chapter 5 we were able to use control of some higher moment and the moment creation property. By contrast, at the large deviation level one would need estimates of the form $\mathbb{P}(\langle f, \mu_t^N \rangle > a) \leq e^{-\varepsilon N}$, for some $f \geq 0$. In the case where the processes is in equilibrium $\mu_0^\star = \gamma$, the classical Cramér theorem shows us that such estimates hold if, and only if, $\int_{\mathbb{R}^d} e^{\delta f(v)} \gamma(dv) < \infty$ for some $\delta > 0$, which is false as soon as f has super-quadratic growth at infinity. We also cannot use the de la Vallée-Poussin Theorem for *a priori* estimates as we did in Theorem 5.3, for the same reason. This argument would show that, for any limit path $(\mu_\bullet, w)$, that there is a smooth, convex h, with $h'(\infty) = \infty$, such that $\langle h(|v|^2), \mu_0\rangle < \infty$. On the other hand, the arguments of Section 6.6 show that there exists events of probabilities at least e^{-cN}, for some finite c, conditional on which, $\langle h(|v|^2), \mu_0^N\rangle \to \infty$ almost surely, simultaneously for *all* such h. Thus, if we take a 'bottom up' approach to large deviations and attempt to classify all sequences of events A_N of negative exponential probability $\mathbb{P}(A_N) \geq e^{-cN}$, then we are really constrained to what can be done with 2 moments.

Let us remark that a similar phenomenon takes place at the level of the limiting paths, since the wide range of possible controls K in the controlled Boltzmann equation (CBE$_K$) destroys all hope repeating the kinds of calculations in Section 2.5. Indeed, for any p, we could set

$$K(t, v, v_*, \sigma) = \mathbb{1}\{|v'|^p + |v'_*|^p > |v|^p + |v_*|^p\}$$

which guarantees that the p^{th} moment always increases, unlike the behaviour without the control K in Chapter 2.

We also remark that, when considering the convergence of the Kac process under a change of measure, as an intermediate step (see Lemma 6.24 and the proof of Theorem 6.2ii) at the start of Section 6.4), we are restricted to qualitative techniques, using tightness and an identification of the limit paths, similar to the methods of Sznitman [173]. As remarked above, we do not have access to moment estimates, even for 'nice' controls, which prevents us adapting some of the more recent quantitative techniques (Mischler and Mouhot [142], Norris [157], or the arguments of Chapters 3-4). Indeed, we even develop, in the course of Theorem 6.5, a law of large numbers for solutions to (BE) which do not conserve energy, and for which all moments higher than second become unbounded. In any case, quantifying the rate of convergence in these steps would not change the main results, which are concerned with the asymptotics of the negative N-exponential.

Finally, let us remark that we could seek some qualitative bounds of the form

$$\mathbb{P}(\sup_{t \leq t_{\text{fin}}} \mathcal{W}_{1,1}(\mu_t^N, \mu_t) > \varepsilon) \leq e^{-\delta N}$$

for some explicit or explicitable $\delta = \delta(\varepsilon)$ and $N \geq N_0(\varepsilon)$, which is the approach taken in the work by Bolley, Guillin and Villani [26]. In the case (GMM), this could be achieved using some of the techniques of Section 3.2 and improving the control of the martingales to an exponential control, as described the review paper [49]. However, in the much more realisting and interesting case (rHS, HS), this is hindered by the same lack of control of moments higher than second of the particle system μ_t^N with exponentially high probability as described above.

6.2 Exponential Tightness & Upper Bound

In this section, we will prove Propositon 6.1 and the upper bound Theorem 6.2i). We first verify exponential tightness in Subsection 6.2.1. In Subsection 6.2.2 we introduce a variational form for the rate function, similar to that of Léonard [131, Theorem 3.1, Theorem 7.1] and prove equivalence of the two formulations; this leads to a simple proof of lower semicontinuity in Proposition 6.1ii) and the fact that (CE) defines a closed set for the topology of $\mathcal{D}\times\mathcal{M}(E)$. Finally, we use the variational formulation to prove the upper bound in Section 6.2.1, based on standard martingale techniques and a covering argument.

6.2.1 Exponential Tightness

We first prove that $(\mu^N_\bullet, w^N)$ are exponentially tight in $\mathcal{D}\times\mathcal{M}(E)$, which proves the first assertion Proposition 6.1i.

Lemma 6.8 (Verification of Conditions for Exponential Tightness)**.** *For any $M>0$, the following hold.*

a). For $\lambda>0$, set $K_\lambda=\{\mu\in\mathcal{P}_2 : \langle |v|^2,\mu\rangle\le\lambda\}$ and $\mathcal{D}_\lambda := \{\mu_\bullet\in\mathcal{D} : \mu_t\in K_\lambda \text{ for all } t\}$. There exists $\lambda\in(0,\infty)$ such that

$$\limsup_N \frac{1}{N}\log\mathbb{P}\left(\mu^N_\bullet\not\in\mathcal{D}_\lambda\right)\le -M. \tag{6.34}$$

b). For all $\delta>0$, define $q^N(\delta)=\sup(\mathcal{W}_{1,1}(\mu^N_s,\mu^N_t) : |s-t|<\delta)$. For all $\epsilon>0$, there exists $\delta>0$ such that

$$\limsup_N \frac{1}{N}\log\mathbb{P}\left(q^N(\delta)>\epsilon\right)\le -M. \tag{6.35}$$

c). There exists $C>0$ such that

$$\limsup_N \frac{1}{N}\log\mathbb{P}\left(w^N(E)>C\right)\le -M. \tag{6.36}$$

Let us remark that the first item proves that, for each fixed $t\in[0,t_{\text{fin}}]$, μ^N_t are exponentially tight because K_λ are compact for the metric $\mathcal{W}_{1,1}$. Together, the first two conditions verify the well-known criteria for exponential tightness in the Skorokhod space $\mathcal{D}$ due to Feng and Kurz [77, Theorem 4.1]. In the third item, the sets $\{w\in\mathcal{M}(E) : w(E)\le C\}$ are compact for the metric ρ_1, which induces the weak topology, and hence the third item shows that w^N are exponentially tight in $\mathcal{M}(E)$. Together, these prove that the pair $(\mu^N_\bullet, w^N)$ together are exponentially tight in $\mathcal{D}\times\mathcal{M}(E)$.

Proof of Lemma 6.8. Fix M throughout. We start with the first point, and begin by noting that, thanks to Hypothesis 6.1i) and a Chebychev bound, for $z_1 > 0$ sufficiently small and all $\lambda > 0$,

$$\mathbb{P}\left(\langle |v|^2, \mu_0^N\rangle > \lambda\right) \le e^{-N\lambda z_1}\mathbb{E}\left[e^{Nz_1\langle |v|^2, \mu_0^N\rangle}\right] = \exp\left(-N(\lambda z_1 - \log \mathcal{E}_{z_1}(\mu_0^\star))\right)$$

where, in the right-hand side, we recall that μ_0^N is given by sampling particles independently from $\mu_0^\star$. Taking $\lambda_M = z_1^{-1}(M + \log \mathcal{E}_{z_1}(\mu_0^\star))$, we conclude that

$$\mathbb{P}\left(\mu_0^N \notin K_{\lambda_M}\right) = \mathbb{P}\left(\langle |v|^2, \mu_0^N\rangle > \lambda_M\right) \le e^{-MN}.$$

To extend this to the whole process we note that the kinetic energy $\langle |v|^2, \mu_t^N\rangle$ is constant in time, so that $\mu_t^N \in \mathcal{D}_{\lambda_M}$ if, and only if, $\mu_0^N \in K_{\lambda_M}$. Therefore

$$\mathbb{P}\left(\mu_\bullet^N \notin \mathcal{D}_{\lambda_M}\right) = \mathbb{P}\left(\mu_0^N \notin K_{\lambda_M}\right) \le e^{-MN}$$

and the first item follows. For the second item, we observe that the instantaneous rate of the Kac process, in either the regularised hard spheres (rHS) or Maxwell molecules (GMM) case, is bounded by

$$\begin{aligned} N\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}} B(v-v_*,\sigma)\mu_t^N(dv)\mu_t^N(dv_\star)d\sigma &\le N\int_{\mathbb{R}^d\times\mathbb{R}^d}(3+|v|^2+|v_\star|^2)\mu_t^N(dv)\mu_t^N(dv_\star) \\ &\le 3N(1+\langle |v|^2, \mu_0^N\rangle) \end{aligned} \tag{6.37}$$

where we note that, for either kernel, $B(v-v_\star,\sigma) \le 1+|v|+|v_\star| \le 3+|v|^2+|v_\star|^2$, and in the final inequality, we recall again that the second moment $\langle |v|^2, \mu_t^N\rangle$ is independent of time. It therefore follows that we can construct a time-homogenous Poisson process $\widetilde{w}_t^N$, of constant, random rate $3N(1+\langle |v|^2, \mu_0^N\rangle)$, such that $\widetilde{w}_t^N$ has jumps on a superset of the times when μ_t^N jumps, see also the equivalent construction above (3.231). This leads to the bound, for any $s \le t$,

$$(w_t^N - w_s^N)([0, t_{\text{fin}}] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{S}^{d-1}) \le \frac{1}{N}(\widetilde{w}_t^N - \widetilde{w}_s^N).$$

For each δ, we now pick a partition $0 = t_0 < t_1 ... < t_m$ of size $\lceil t_{\text{fin}}/\delta\rceil$ by taking constant steps of size δ. We now observe that $\mathcal{W}_{1,1}(\mu_t^N, \mu_{t-}^N) \le 4/N$ at times when μ_t^N jumps, and for any $|s-t| \le \delta$, the interval $[s,t]$ is contained in at most two adjacent intervals $[t_{i-1}, t_{i+1}]$. Together, we conclude that

$$q^N(\delta) \le 8 \max_i \frac{1}{N}(\widetilde{w}_{t_i}^N - \widetilde{w}_{t_{i-1}}^N). \tag{6.38}$$

For λ_M as above and for any $z > 0$, we bound

$$\begin{aligned} \mathbb{P}\left(\frac{1}{N}(\widetilde{w}_{t_i}^N - \widetilde{w}_{t_{i-1}}^N) > \frac{\epsilon}{8} \,\middle|\, \mu_0^N \in K_{\lambda_M}\right) &\le e^{-zN\epsilon/8}\mathbb{E}\left[e^{z(\widetilde{w}_{t_i}^N - \widetilde{w}_{t_{i-1}}^N)} \,\middle|\, \mu_0^N \in K_{\lambda_M}\right] \\ &\le \exp\left(-N\left(\frac{z\epsilon}{8} - 3(1+\lambda_M)\delta(e^z-1)\right)\right) \end{aligned} \tag{6.39}$$

where, in the second line, we use the bound that the rate of $\widetilde{w}^N$ is at most $3N(1+\lambda_M)$ if $\mu_0^N \in K_{\lambda_M}$. We now choose $z = 8(M+1)/\epsilon$, and $\delta > 0$ small enough, depending on z, λ_M, so that $3(1+\lambda_M)\delta(e^z-1) < 1$. For this choice of δ, the final expression in (6.39) is e^{-NM}, for each interval. Finally, we take a union bound:

$$\{q^N(\delta) > \epsilon\} \subset \{\mu_0^N \notin K_{\lambda_M}\} \cup \bigcup_{i \le \lceil t_{\text{fin}}/\delta \rceil} \left\{\widetilde{w}^N_{t_i} - \widetilde{w}^N_{t_{i-1}} > \frac{N\epsilon}{8},\ \mu_0^N \in K_{\lambda_M}\right\}.$$

By the choices of λ_M and δ,

$$\mathbb{P}\left(q^N(\delta) > \epsilon\right) \le (1 + \lceil t_{\text{fin}}/\delta \rceil)e^{-NM}$$

and the second item now follows. The final item follows in exactly the same way: following (6.39), for all $C > 0$ we bound

$$\begin{aligned} \mathbb{P}\left(w^N(E) > C \middle| \mu_0^N \in K_{\lambda_M}\right) &\le \mathbb{P}\left(\widetilde{w}^N_{t_{\text{fin}}} > CN \middle| \mu_0^N \in K_{\lambda_M}\right) \\ &\le e^{-CN}\mathbb{E}\left[e^{\widetilde{w}^N_{t_{\text{fin}}}} \middle| \mu_0^N \in K_{\lambda_M}\right] \\ &\le \exp\Big(-N\big(C - 3(1+\lambda_M)t_{\text{fin}}(e-1)\big)\Big) \end{aligned} \tag{6.40}$$

and choosing $C = M + 3(1+\lambda_M)t_{\text{fin}}(e-1)$ makes the final probability at most e^{-MN}. Using a union bound,

$$\mathbb{P}\left(w^N(E) > C\right) \le \mathbb{P}\left(w^N(E) > C \middle| \mu_0^N \in K_{\lambda_M}\right) + \mathbb{P}\left(\mu_0^N \notin K_{\lambda_M}\right) \le 2e^{-MN} \tag{6.41}$$

from which (6.36) follows. □

Let us also record, for later use, the following corollary.

Corollary 6.9. *Let $\mathbb{Q}^N \ll \mathbb{P}$ be changes of measure such that*

$$\lim_{a\to\infty} \liminf_N \mathbb{Q}^N\left(\frac{d\mathbb{Q}^N}{d\mathbb{P}} \le e^{Na}\right) = 1. \tag{6.42}$$

Then the laws of $(\mu^N_\bullet, w^N)$ under $\mathbb{Q}^N$ are tight: for all $\epsilon > 0$ there exists a compact set $\mathcal{K} \subset \mathcal{D} \times \mathcal{M}(E)$ such that

$$\sup_N \mathbb{Q}^N\left((\mu^N_\bullet, w^N) \notin \mathcal{K}\right) < \epsilon. \tag{6.43}$$

Proof. This follows from Proposition 6.1i) and the hypothesis (6.42) by purely general considerations. Let us fix $\epsilon > 0$; thanks to (6.42) we can choose a such that, for all but finitely many N,

$$\mathbb{Q}^N\left(\frac{d\mathbb{Q}^N}{d\mathbb{P}} > e^{Na}\right) < \frac{\epsilon}{2} \tag{6.44}$$

and, changing a if necessary, we can arrange that (6.44) holds for all N. We now choose $M = a - \log(\epsilon/2)$; by Proposition 6.1i), there exists a compact set $\mathcal{K}$ such that, $\limsup_N N^{-1} \log \mathbb{P}((\mu^N_\bullet, w^N) \notin \mathcal{K}) < -M$ which implies that, for all N sufficiently large,

$$\mathbb{P}\left((\mu^N_\bullet, w^N) \notin \mathcal{K}\right) \le e^{-MN}. \tag{6.45}$$

Since the space $\mathcal{D} \times \mathcal{M}(E)$ is a separable metric space, each $(\mu^N_\bullet, w^N)$ is tight so $\mathcal{K}$ can be replaced with a larger compact set such that (6.45) again holds for all N. Together, (6.44, 6.45) imply that

$$\begin{aligned} \mathbb{Q}^N\left((\mu^N_\bullet, w^N) \notin \mathcal{K}\right) &\le \mathbb{Q}^N\left(\frac{d\mathbb{Q}^N}{d\mathbb{P}} \le e^{Na}, (\mu^N_\bullet, w^N) \notin \mathcal{K}\right) + \mathbb{Q}^N\left(\frac{d\mathbb{Q}^N}{d\mathbb{P}} > e^{Na}\right) \\ &< e^{Na}\mathbb{P}\left((\mu^N_\bullet, w^N) \notin \mathcal{K}\right) + \frac{\epsilon}{2} \le e^{Na}e^{-N(a+\log(\epsilon/2))} + \frac{\epsilon}{2} \le \epsilon \end{aligned} \tag{6.46}$$

and we are done. □

6.2.2 Variational Formulation of the Rate Function

In preparation for the upper bound, we will now present a variational formulation of the rate function. This will also allow us to prove the lower semicontinuity in Proposition 6.1. We are aided in this equivalence by the inclusion of the flux in the large deviation principle: the choice of K, if it exists, is unique, which allows us to significantly simplify the proof of Léonard [131].

We begin with the following construction. Let us write $C^{1,0}_{0,b}([0, t_{\text{fin}}] \times \mathbb{R}^d)$ for those functions $f : [0, t_{\text{fin}}] \times \mathbb{R}^d \to \mathbb{R}$ which are continuous and bounded, with a continuous and bounded time derivative $\partial_t f_t$, such that both $f_t, \partial_t f_t$ are Lipschitz in the $\mathbb{R}^d$-variable, and such that $f_0 = f_{t_{\text{fin}}} = 0$. For $\varphi \in C_b(\mathbb{R}^d), f \in C^{1,0}_{0,b}([0, t_{\text{fin}}] \times \mathbb{R}^d)$ and $g \in C_c(E)$ and $t \in [0, t_{\text{fin}}]$, we define

$$\Xi(\mu_\bullet, w, \varphi, f, g)_t = \Xi_0(\mu_\bullet, \varphi) + \Xi_{1,t}(\mu_\bullet, w, f) + \Xi_{2,t}(\mu_\bullet, w, g) \tag{6.47}$$

where

$$\Xi_0(\mu_\bullet, \varphi) = \langle \varphi, \mu_0 \rangle - \log\langle e^\varphi, \mu^\star_0 \rangle; \tag{6.48}$$

$$\Xi_{1,t}(\mu_\bullet, w, f) := \langle f_t, \mu_t \rangle - \int_0^t \langle \partial_s f_s, \mu_s \rangle ds - \int_{E_t} \Delta f(s, v, v_\star, \sigma) w(ds, dv, dv_\star, d\sigma) \tag{6.49}$$

and

$$\begin{aligned} \Xi_{2,t}(\mu_\bullet, w, g) : = &\int_{E_t} (g(s, v, v_\star, \sigma) w(ds, dv, dv_\star, d\sigma) \\ &- \int_{E_t} (e^g - 1)(s, v, v_\star, \sigma) \overline{w}_\mu(ds, dv, dv_\star, d\sigma)). \end{aligned} \tag{6.50}$$

We write $\Xi(\mu_\bullet, w, \varphi, f, g)$ for the terminal value $\Xi(\mu_\bullet, w, \varphi, f, g) = \Xi(\mu_\bullet, w, \varphi, f, g)_{t_{\text{fin}}}$. Let us note that these processes make sense at the level of the particle system $\mu^N_\bullet, w^N$. The function f here entering into $\Xi_{1,t}(\mu_\bullet, w, f)$ will play the rôle of a Lagrange multiplier to enforce the constraint of the continuity equation. This is made precise by the following result.

Lemma 6.10. *Fix* $(\mu_\bullet, w) \in \mathcal{D} \times \mathcal{M}(E)$. *Then*

$$\sup\left\{\Xi_{1,t_{\text{fin}}}(\mu_\bullet, w, f) : f \in C^{1,0}_{0,b}([0, t_{\text{fin}}] \times \mathbb{R}^d)\right\} = \begin{cases} 0 & \text{if } (\mu_\bullet, w) \text{ solves (CE);} \\ \infty & \text{else.} \end{cases} \tag{6.51}$$

Proof. For the case where $(\mu_\bullet, w)$ solves (CE), we will show that, for all $f \in C^{1,0}_{0,b}([0, t_{\text{fin}}] \times \mathbb{R}^d)$ and all $t \in [0, t_{\text{fin}}]$, we have the time-dependent equivalent of (CE):

$$\langle f, \mu_t\rangle = \int_0^t \langle \partial_s f_s, \mu_s\rangle ds + \int_{E_t} \Delta f(s, v, v_\star, \sigma) w(ds, dv, dv_\star, d\sigma). \tag{6.52}$$

This will immediately imply that $\Xi_{1,t}(\mu_\bullet, w, f) = 0$ for all f and all t, which implies the claim. The proof of this formulation is slightly complicated by the lack of regularity, since we only assume *a priori* that $\mu_\bullet$ is càdlàg rather than continuous; we will instead use the facts about càdlàg paths from Proposition 6.28. Since (6.52) is linear in f, we can assume that $f_t, \partial_t f_t$ have Lipschitz norm bounded by 1: $\|f_t\|_{0,1}, \|\partial_t f_t\|_{0,1} \le 1$, in the notation of (2.6). Fix $t \in [0, t_{\text{fin}}]$, $\epsilon > 0$, $\delta > 0$.

Let us write $P \subset [0, t_{\text{fin}}]$ for those $s \in [0, t_{\text{fin}}]$ with $\mathcal{W}_{1,1}(\mu_{s-}, \mu_s) \ge \epsilon$; thanks to Proposition 6.28a), P is finite, and we write $m = |P| < \infty$ for its cardinality. Possibly making $\delta > 0$ smaller, Proposition 6.28b) guarantees that δ can be chosen so that any interval $[u, v)$ of length δ either contains a point of P, or for all $s \in [u, v)$, $\mathcal{W}_{1,1}(\mu_s, \mu_u) < \epsilon$. Now, for such δ, we decompose $(0, t]$ into intervals $(t_i, t_{i+1}]$ of length at most δ, and add

$$\langle f_{t_{i+1}}, \mu_{t_{i+1}} - \mu_{t_i}\rangle = \int_{E_{t_{i+1}} \setminus E_{t_i}} \Delta f(t_{i+1}, v, v_\star, \sigma) w(ds, dv, dv_\star, d\sigma); \tag{6.53}$$

$$\langle f_{t_{i+1}} - f_{t_i}, \mu_{t_i}\rangle = \int_{t_i}^{t_{i+1}} \langle \partial_s f_s, \mu_{t_i}\rangle ds \tag{6.54}$$

to obtain

$$\begin{aligned} \langle f_{t_{i+1}}, \mu_{t_{i+1}}\rangle - \langle f_{t_i}, \mu_{t_i}\rangle = &\int_{t_i}^{t_{i+1}} \langle \partial_s f_s, \mu_{t_i}\rangle ds \\ &+ \int_{E_{t_{i+1}} \setminus E_{t_i}} \Delta f(t_{i+1}, v, v_\star, \sigma) w(ds, dv, dv_\star, d\sigma). \end{aligned} \tag{6.55}$$

We approximate the two terms by

$$\begin{aligned} \left|\int_{t_i}^{t_{i+1}} \langle \partial_s f_s, \mu_{t_i}\rangle ds - \int_{t_i}^{t_{i+1}} \langle \partial_s f_s, \mu_s\rangle ds\right| &\le \int_{t_i}^{t_{i+1}} \mathcal{W}_{1,1}(\mu_s, \mu_{t_i}) ds \\ &\le 2\epsilon(t_{i+1} - t_i) + \delta \cdot \mathbb{I}_{P \cap [t_i, t_{i+1}) \neq \emptyset} \end{aligned} \tag{6.56}$$

since, by the choice of δ and the noramlisation of f, either $\langle \partial_s f_s, \mu_s, \mu_{t_i}\rangle \le 2\mathcal{W}_{1,1}(\mu_s, \mu_{t_i}) \le \epsilon$ for all $s \in [t_i, t_{i+1}]$, or there is a point of P in $[t_i, t_{i+1})$, in which case we use the trivial bound $\mathcal{W}_{1,1}(\mu_s, \mu_{t_i}) \le 1$ and recall that the interval is of length at most δ. For the second term

$$\begin{aligned}\left|\int_{E_{t_{i+1}}\setminus E_{t_i}} (\Delta f(t_{i+1}, v, v_\star, \sigma) - \Delta f(s, v, v_\star, \sigma))\, w(ds, dv, dv_\star, d\sigma)\right| \\ \le \left|\int_{E_{t_{i+1}}\setminus E_{t_i}} 4\|f_{t_{i+1}} - f_s\|_\infty w(ds, dv, dv_\star, d\sigma)\right| \\ \le 4\delta w\left(E_{t_{i+1}} \setminus E_{t_i}\right)\end{aligned} \tag{6.57}$$

where in the final line we recall that we have scaled so that $\|\partial_t f_t\|_\infty \le 1$. Adding, we conclude that

$$\begin{aligned}\Bigg|\langle f_{t_{i+1}}, \mu_{t_{i+1}}\rangle - \langle f_{t_i}, \mu_{t_i}\rangle - \int_{t_i}^{t_{i+1}} \langle \partial_s f_s, \mu_s\rangle ds \\ - \int_{E_{t_{i+1}}\setminus E_{t_i}} \Delta f(s, v, v_\star, \sigma) w(ds, dv, dv_\star, d\sigma)\Bigg| \\ \le 2\epsilon(t_{i+1} - t_i) + \delta \cdot 1\!\!1_{P\cap[t_i, t_{i+1})\neq\emptyset} + 4\delta w\left(E_{t_{i+1}} \setminus E_{t_i}\right).\end{aligned} \tag{6.58}$$

Summing over all such intervals $(t_i, t_{i+1}]$ covering $(0, t]$, and recalling that $f_0 \equiv 0$, we obtain

$$\begin{aligned}\left|\langle f_t, \mu_t\rangle - \int_0^t \langle \partial_s f_s, \mu_s\rangle ds - \int_{E_t} \Delta f(s, v, v_\star, \sigma) w(ds, dv, dv_\star, d\sigma)\right| \\ \le 2\epsilon t + m\delta + 4\delta w(E)\end{aligned} \tag{6.59}$$

and the right-hand side can be made arbitrarily small by taking $\epsilon, \delta \to 0$, recalling that w is a finite measure by hypothesis, so the claim (6.52) is proven, and we conclude that $\sup_f \Xi_{1,t_{\text{fin}}}(\mu_\bullet, w, f) = 0$ as claimed. Otherwise, if (CE) fails, there exists some $t_0 \in (0, t_{\text{fin}}]$ and some $g \in C_c(\mathbb{R}^d)$ such that

$$\langle g, \mu_{t_0}\rangle - \langle g, \mu_0\rangle - \int_{E_{t_0}} \Delta g(v, v_\star, \sigma) w(ds, dv, dv_\star, d\sigma) > 1. \tag{6.60}$$

Further, if $t_0 = t_{\text{fin}}$, then the same is true at some $t_0' < t_{\text{fin}}$, since μ is continuous at t_{fin}, and $w(t_{\text{fin}} \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{S}^{d-1}) = 0$ by hypothesis, so we assume that $t_0 \in (0, t_{\text{fin}})$. We now fix a smooth, increasing function $\chi : \mathbb{R} \to \mathbb{R}$ such that $\chi = 0$ on $(-\infty, 0]$ and $\chi = 1$ on $[1, \infty)$, and for $0 < \delta < \min(t_0, t_{\text{fin}} - t_0)$, we construct $f^\delta \in C^{1,0}_{0,b}([0, t_{\text{fin}}] \times \mathbb{R}^d)$ by defining

$$f^\delta_t(v) := \begin{cases} \chi(t/\delta) g(v) & \text{if } t \in [0, \delta]; \\ g(v) & \text{if } \delta < t \le t_0; \\ \chi(1 - (t - t_0)/\delta) g(v) & \text{if } t_0 < t < t_0 + \delta \\ 0 & \text{else.} \end{cases} \tag{6.61}$$

Thanks to right-continuity of μ_t, we observe that $\int_0^{t_{\text{fin}}} \langle \partial_t f^\delta_t, \mu_t \rangle dt \to \langle g, \mu_0 \rangle - \langle g, \mu_{t_0} \rangle$, and using dominated convergence,

$$\int_E \Delta f^\delta(s, v, v_\star, \sigma) w(ds, dv, dv_\star, d\sigma) \to \int_{E_{t_0}} \Delta g(v, v_\star, \sigma) w(ds, dv, dv_\star, d\sigma). \tag{6.62}$$

Therefore, $\Xi_{1,t_{\text{fin}}}(\mu_\bullet, w, f^\delta)$ converges to

$$\Xi_{1,t_{\text{fin}}}(\mu_\bullet, w, f^\delta) \to \langle g, \mu_{t_0} \rangle - \langle g, \mu_0 \rangle - \int_{E_{t_0}} \Delta g(v, v_\star, \sigma) w(ds, dv, dv_\star, d\sigma) > 1 \tag{6.63}$$

and in particular, we can choose $\delta > 0$ small enough that $\Xi_{1,t_{\text{fin}}}(\mu_\bullet, w, f^\delta) > 1$. By linearity, for all $\lambda > 0$, $\Xi_{1,t_{\text{fin}}}(\mu_\bullet, w, \lambda f^\delta) > \lambda$, and so the supremum is infinite, as claimed. □

We now use this equality to show that the functions Ξ above give a variational formulation of the rate function $\mathcal{I}$ given in the introduction.

Lemma 6.11. *For $(\mu_\bullet, w) \in \mathcal{D} \times \mathcal{M}(E)$, we have*

$$\mathcal{I}(\mu_\bullet, w) = \sup \left\{ \Xi(\mu_\bullet, w, \varphi, f, g) : \varphi \in C_b(\mathbb{R}^d), f \in C^{1,0}_{0,b}([0, t_{\text{fin}}] \times \mathbb{R}^d), g \in C_c(E) \right\}. \tag{6.64}$$

Proof. Let us write $\widetilde{\mathcal{I}}$ for the right-hand side. Since Ξ_0 depends only on φ, Ξ_1 only on f and Ξ_2 only on g, the supremum decomposes as

$$\widetilde{\mathcal{I}}(\mu_\bullet, w) = \sup_\varphi \Xi_0(\mu_\bullet, \varphi) + \sup_f \Xi_{1,t_{\text{fin}}}(\mu_\bullet, w, f) + \sup_g \Xi_{2,t_{\text{fin}}}(\mu_\bullet, w, g) \tag{6.65}$$

where the suprema run over the same sets as above. Optimising over φ produces the well-known variational formulation $\sup_\varphi \Xi_0(\mu_0, \varphi) = H(\mu_0|\mu_0^\star)$ of the relative entropy. This identity can be found in [123, Appendix 1], or derived using essentially the same argument as for $\Xi_{2,t_{\text{fin}}}$ below. Thanks to Lemma 6.10, the supremum over f is infinite unless the continuity equation (CE) holds, in which case this term vanishes.

We now deal with the third term. If $w \not\ll \overline{w}_\mu$, there is a compact set $E' \subset E$ with $w(E') > 0$ but $\overline{w}_\mu(E') = 0$, and since E is a metric space and w is a Borel measure, we can find open $U_n \downarrow E'$ and closed $A_n \supset E', A_n \subset U_n$ with $\overline{w}_\mu(U_n) \downarrow 0$. We now choose $g_n \in C_c(E)$ so that $0 \leq g_n \leq 1$, $g_n = 1$ on A_n, and $= 0$ except on U_n, and bound for $\lambda > 0$,

$$\int_E \lambda g_n w(ds, dv, dv_\star, d\sigma) \geq \lambda w(E'); \tag{6.66}$$

$$\int_E (e^{\lambda g_n} - 1) \overline{w}_\mu(ds, dv, dv_\star, d\sigma) \leq (e^\lambda - 1) \overline{w}_\mu(U_n) \tag{6.67}$$

so that

$$\Xi_{2,t_{\text{fin}}}(\mu_\bullet, w, \lambda g_n) \geq \lambda w(E') - (e^\lambda - 1) \overline{w}_\mu(U_n). \tag{6.68}$$

By taking $\lambda = \lambda_n \to \infty$ slowly enough, the right-hand side can be made arbitrarily large as $n \to \infty$, so in this case $\sup_g \Xi_{2,t_{\text{fin}}}(\mu_\bullet, w, g) = \infty$. On the other hand, if $w \ll \overline{w}_\mu$, let us write K for the tilting function $\frac{dw}{d\overline{w}_\mu}$, so that

$$\Xi_{2,t_{\text{fin}}}(\mu_\bullet, w, g) = \int_E (Kg - e^g + 1)(s, v, v_\star, \sigma)\overline{w}_\mu(ds, dv, dv_\star, d\sigma). \tag{6.69}$$

Observing that, for all $x \in \mathbb{R}, y \geq 0$, it holds that $xy \leq (e^x - 1) + (y \ln y - y + 1) = (e^x - 1) + \tau(y)$, the first term $\int Kg$ can be bounded by

$$\begin{aligned} &\int_E Kg(s, v, v_\star, \sigma)\overline{w}_\mu(ds, dv, dv_\star, d\sigma) \\ &\qquad \leq \int_E \tau(K)\overline{w}_\mu(ds, dv, dv_\star, d\sigma) + \int_E (e^g - 1)\overline{w}_\mu(ds, dv, dv_\star, d\sigma) \end{aligned} \tag{6.70}$$

which leads to the bound, uniformly in $g \in C_c(E)$,

$$\Xi_{2,t_{\text{fin}}}(\mu_\bullet, w, g) \leq \int_E \tau(K(s, v, v_\star, \sigma))\overline{w}_\mu(ds, dv, dv_\star, d\sigma) \tag{6.71}$$

whether or not the right-hand side is finite. On the other hand, let us fix M. By Lusin's theorem, we can construct continuous, bounded $g_n \in C_c(E)$ with $g_n \to \ln K \wedge M$ for $\overline{w}_\mu$-almost all $(t, v, v_\star, \sigma)$, so that

$$Kg_n - e^{g_n} + 1 \leq MK + 1 \tag{6.72}$$

and

$$Kg_n - e^{g_n} + 1 \to K(\ln K \wedge M) - (K \wedge e^M) + 1 \tag{6.73}$$

for $\overline{w}_\mu$-almost all $(s, v, v_\star, \sigma)$. Since $K \in L^1(\overline{w}_\mu)$, we can use dominated convergence to obtain

$$\Xi_{2,t_{\text{fin}}}(\mu_\bullet, w, g_n) \to \int_E (K(\ln K \wedge M) - (K \wedge e^M) + 1)\overline{w}_\mu(ds, dv, dv_\star, d\sigma) \tag{6.74}$$

and the supremum is at least the right-hand side. The integrand is increasing in M, and converges to $\tau(K)$ pointwise, so the whole integral converges to $\int_E \tau(K) d\overline{w}_\mu$. We conclude that

$$\sup \{\Xi_{2,t_{\text{fin}}}(\mu_\bullet, w, g) : g \in C_c(E)\} \geq \int_E \tau(K(s, v, v_\star, \sigma))\overline{w}_\mu(ds, dv, dv_\star, d\sigma) \tag{6.75}$$

and (6.71) shows that this is an equality. Putting everything together, we have shown that

$$\begin{aligned} &\sup_{f,g} \{\Xi_{1,t_{\text{fin}}}(\mu_\bullet, w, f) + \Xi_{2,t_{\text{fin}}}(\mu_\bullet, w, g)\} \\ &\qquad = \begin{cases} \int_E \tau(K)\overline{w}_\mu(ds, dv, dv_\star, d\sigma) & \text{if } (\mu_\bullet, w) \text{is a measure flux pair;} \\ \infty & \text{else} \end{cases} \end{aligned} \tag{6.76}$$

and the right-hand side is exactly the definition of $\mathcal{J}(\mu_\bullet, w)$. Returning to (6.65), we have proven that

$$\widetilde{\mathcal{I}}(\mu_\bullet, w) = H(\mu_0|\mu_0^\star) + \mathcal{J}(\mu_\bullet, w) = \mathcal{I}(\mu_\bullet, w) \tag{6.77}$$

as desired. □

Thanks to this variational form, we readily obtain the lower semi-continuity claimed in Proposition 6.1. We first record, as a separate lemma, a result which will be helpful later.

Lemma 6.12. *For fixed $f \in L^\infty([0, t_{\text{fin}}], C_b(\mathbb{R}^d))$ and $g \in C_c(E)$, the maps*

$$\mu_\bullet \mapsto \int_0^{t_{\text{fin}}} \langle f_t, \mu_t\rangle dt; \qquad \mu_\bullet \mapsto \int_E g\overline{w}_\mu(ds, dv, dv_\star, d\sigma) \tag{6.78}$$

are continuous in the topology of $\mathcal{D}$.

Proof. Noting that the topology of $\mathcal{D}$ is induced by a metric, it is sufficient to prove sequential continuity: let us fix $\mu_\bullet^{(n)} \to \mu_\bullet$. By Proposition 6.29, it follows that $\mathcal{W}_{1,1}(\mu_t^{(n)}, \mu_t) \to 0$ for dt-almost all t, and for all such t, we also have the weak convergence $\mu_t^{(n)} \otimes \mu_t^{(n)} \to \mu_t \otimes \mu_t$. Since g has compact support in E, for any fixed σ and for such t, the map $(v, v_\star) \mapsto g(t, v, v_\star, \sigma)B(v - v_*, \sigma)$ is bounded and continuous, and so we have the convergences

$$\langle f_t, \mu_t^{(n)}\rangle \to \langle f_t, \mu_t\rangle; \tag{6.79}$$

$$\begin{aligned}\int_{\mathbb{R}^d\times\mathbb{R}^d} g(t, v, v_\star, \sigma)B(v - v_*, \sigma)\mu_t^{(n)}(dv)\mu_t^{(n)}(dv_\star) \\ \to \int_{\mathbb{R}^d\times\mathbb{R}^d} g(t, v, v_\star, \sigma)B(v - v_*, \sigma)\mu_t(dv)\mu_t(dv_\star).\end{aligned} \tag{6.80}$$

Since these hold for all σ and dt-almost all t, we can integrate and use bounded convergence to find that

$$\int_0^{t_{\text{fin}}} \langle f_t, \mu_t^{(n)}\rangle dt \to \int_0^{t_{\text{fin}}} \langle f_t, \mu_t\rangle dt; \tag{6.81}$$

$$\int_E g\overline{w}_{\mu^{(n)}}(ds, dv, dv_\star, d\sigma) \to \int_E g\overline{w}_\mu(ds, dv, dv_\star, d\sigma) \tag{6.82}$$

and we are done. □

Lemma 6.13. *For fixed $\varphi \in C_b(\mathbb{R}^d)$, $f \in C_{0,b}^{1,0}([0, t_{\text{fin}}] \times \mathbb{R}^d)$ and $g \in C_c(\mathbb{R}^d)$, the maps*

$$(\mu_\bullet, w) \to \Xi_0(\mu_\bullet, \varphi); \qquad (\mu_\bullet, w) \to \Xi_{1,t_{\text{fin}}}(\mu_\bullet, \varphi); \qquad (\mu_\bullet, w) \to \Xi_{2,t_{\text{fin}}}(\mu_\bullet, \varphi); \tag{6.83}$$

$$(\mu_\bullet, w) \to \Xi(\mu_\bullet, w, \varphi, f, g) \tag{6.84}$$

are continuous for the topology of $\mathcal{D} \times \mathcal{M}(E)$. In particular, the sub-level sets $\{\mathcal{I} \le a\} \subset \mathcal{D} \times \mathcal{M}(E)$ are closed for all $a \in [0, \infty)$, as is the set of pairs $(\mu_\bullet, w)$ for which (CE) holds, and $\{\mu \in \mathcal{P}_2 : H(\mu|\mu_0^\star) \le a\}$ in the topology of $(\mathcal{P}_2, W)$.

Proof. With the choices of topologies on $\mathcal{D}, \mathcal{M}(E)$, the maps

$$w \mapsto \int_E gw(ds, dv, dv_\star, d\sigma); \qquad w \mapsto \int_E \Delta f(s, v, v_\star, \sigma) w(ds, dv, dv_\star, d\sigma)$$

are immediately continuous, and thanks to Proposition 6.29a), so are $\mu_\bullet \mapsto \langle \varphi, \mu_0 \rangle; \mu_\bullet \mapsto \langle f_{t_{\text{fin}}}, \mu_{t_{\text{fin}}} \rangle$. Combining with Lemma 6.12, with f_t replaced by $\partial_t f_t$, each expression appearing in the definitions of $\Xi_{i,t_{\text{fin}}}$ is continuous, and we conclude the claimed continuity of the stated maps. For the second point, we use Lemma 6.11 to write the sublevel sets, for any $a \in [0, \infty]$, as

$$\{\mathcal{I} \le a\} = \bigcap_{\varphi \in C_b(\mathbb{R}^d), f \in C^{1,0}_{0,b}([0,t_{\text{fin}}] \times \mathbb{R}^d), g \in C_c(E)} \{(\mu_\bullet, w) : \Xi(\mu_\bullet, w, \varphi, f, g)_{t_{\text{fin}}} \le a\}. \tag{6.85}$$

Each set in the intersection is closed, and hence so is the left-hand side, which proves lower semi-continuity; the assertion for $H(\cdot|\mu_0^\star)$ is identical, recalling again that $H(\mu|\mu_0^\star) = \sup_\varphi \Xi_0(\varphi, \mu_0)$. The remaining assertion is similar: using Lemma 6.10,

$$\{(\mu_\bullet, w) : (\text{CE}) \text{ holds}\} = \bigcap_{f \in C^{1,0}_{0,b}([0,t_{\text{fin}}] \times \mathbb{R}^d)} \{(\mu_\bullet, w) : \Xi_{1,t_{\text{fin}}}(\mu_\bullet, w, f) = 0\} \tag{6.86}$$

which is an intersection of closed sets, and hence closed. □

6.2.3 Upper Bound

Using the variational formulation above, we now prove the upper bound in Theorem 6.2. We begin with a local version of the result.

Lemma 6.14. *Fix $(\mu_\bullet, w) \in \mathcal{D} \times \mathcal{M}(E)$ with finite rate $\mathcal{I}(\mu_\bullet, w) < \infty$, and fix $\epsilon > 0$. Then there exists an open set $\mathcal{U} \ni (\mu_\bullet, w)$ such that*

$$\limsup_N \frac{1}{N} \log \mathbb{P}\left((\mu^N_\bullet, w^N) \in \mathcal{U}\right) \le -\mathcal{I}(\mu_\bullet, w) + \epsilon. \tag{6.87}$$

If instead $\mathcal{I}(\mu_\bullet, w) = \infty$ and $M < \infty$ then there exists an open set $\mathcal{U} \ni (\mu_\bullet, w)$ such that

$$\limsup_N \frac{1}{N} \log \mathbb{P}\left((\mu^N_\bullet, w^N) \in \mathcal{U}\right) \le -M. \tag{6.88}$$

Proof. Let us consider the first case; the second is essentially identical. Thanks to Lemma 6.11, we can choose $\varphi \in C_b(\mathbb{R}^d)$, $f \in C^{1,0}_{0,b}([0, t_{\text{fin}}] \times \mathbb{R}^d)$ and $g \in C_c(E)$ such that

$$\mathcal{I}(\mu_\bullet, w) < \Xi(\mu_\bullet, w, \varphi, f, g) + \frac{\epsilon}{2} \tag{6.89}$$

and, thanks to Lemma 6.13, we can find open $\mathcal{U} \ni (\mu_\bullet, w)$ such that, for all $(\mu'_\bullet, w') \in \mathcal{U}$, we have

$$\Xi(\mu'_\bullet, w' \varphi, f, g) > \Xi(\mu_\bullet, w, \varphi, f, g) - \frac{\epsilon}{2} > \mathcal{I}(\mu_\bullet, w) - \epsilon. \tag{6.90}$$

We consider the processes

$$Z^N_t := \exp\left(N\Xi(\mu^N_\bullet, w, \varphi, f, g)_t\right). \tag{6.91}$$

We first observe that, since $\mu^N_\bullet, w^N$ satisfy the continuity equation (CE), Lemma 6.10 shows that, for all $t \geq 0$,

$$\Xi_{1,t}(\mu^N_\bullet, w^N, \varphi, f, g) = 0. \tag{6.92}$$

Next, we show that Z^N is a martingale, following arguments of [49]. We observe that at points $(t, v, v_\star, \sigma)$ of w^N, Z^N_t jumps by

$$Z^N_t - Z^N_{t-} = Z^N_{t-}\left(e^{g(t,v,v_\star,\sigma)} - 1\right) \tag{6.93}$$

while between jumps, Z^N_t is differentiable and

$$\partial_t Z^N_t = -N\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}} Z^N_t\left(e^{g(t,v,v_\star,\sigma)} - 1\right)B(v-v_*,\sigma)\mu^N_t(dv)\mu^N_t(dv_\star)d\sigma. \tag{6.94}$$

Together, Z^N_t admits the representation

$$Z^N_t = Z^N_0 + N\int_{E_t} Z^N_{s-}\left(e^{g(s,v,v_\star,\sigma)} - 1\right)(w^N - \overline{w}_{\mu^N})(ds, dv, dv_\star, d\sigma). \tag{6.95}$$

Recalling the generator (1.31), Z^N_t is a local martingale, and since it is clearly positive, a supermartingale, and at time 0,

$$\mathbb{E}\left[e^{N\Xi_0(\mu_\bullet,\varphi)}\right] = \frac{\mathbb{E}\left[e^{N\langle\varphi,\mu^N_0\rangle}\right]}{\langle e^\varphi, \mu^\star_0\rangle^N} = 1 \tag{6.96}$$

where we recall that μ^N_0 is formed by independent samples from $\mu^\star_0$. We now take the expectation of

$$\begin{aligned}\mathbb{1}\left((\mu^N_\bullet, w^N) \in \mathcal{U}\right) &\leq Z^N_{t_{\text{fin}}}\exp\left(-N\inf\{\Xi(\mu'_\bullet, w', \varphi, f, g) : (\mu'_\bullet, w') \in \mathcal{U}\}\right)\\ &\leq Z^N_{t_{\text{fin}}}\exp\left(-N(\mathcal{I}(\mu_\bullet, w) - \epsilon)\right)\end{aligned} \tag{6.97}$$

to obtain

$$\begin{aligned}\mathbb{P}\left((\mu^N_\bullet, w^N) \in \mathcal{U}\right) &\leq \mathbb{E}[Z^N_{t_{\text{fin}}}]\exp\left(-N(\mathcal{I}(\mu_\bullet, w) - \epsilon)\right)\\ &\leq \exp\left(-N(\mathcal{I}(\mu_\bullet, w) - \epsilon)\right)\end{aligned} \tag{6.98}$$

to produce the desired result. The case where $\mathcal{I}(\mu_\bullet, w) = \infty$ is essentially identical. □

We now give the proof of the global upper bound.

Proof of Theorem 6.5i). Let $\mathcal{A}$ be any closed subset of $\mathcal{D}\times\mathcal{M}(E)$ and fix $\epsilon \in (0,1]$. Let us assume that $\mathcal{A}$ is nonempty, and that $\inf_{\mathcal{A}}\mathcal{I} < \infty$. Choosing $M = \inf_{\mathcal{A}}\mathcal{I} + 1$, by Proposition 6.1i) there exists a compact set $\mathcal{K} \subset \mathcal{D}\times\mathcal{M}(E)$ such that

$$\limsup_N \frac{1}{N}\mathbb{P}\left((\mu^N_\bullet, w^N) \notin \mathcal{K}\right) \leq -M. \tag{6.99}$$

Now, $\mathcal{A}\cap\mathcal{K}$ is compact, since $\mathcal{A}$ was assumed to be closed. For all $(\mu_\bullet, w)\in\mathcal{A}$, we now use Lemma 6.14 to construct $\mathcal{U}(\mu_\bullet,w)\ni(\mu_\bullet,w)$: if $\mathcal{I}(\mu_\bullet,w)<\infty$ then choose $\mathcal{U}(\mu_\bullet,w)$ such that

$$\limsup_N \frac{1}{N}\log\mathbb{P}\left((\mu^N_\bullet,w^N)\in\mathcal{U}(\mu_\bullet,w)\right)\le -\mathcal{I}(\mu_\bullet,w)+\epsilon \tag{6.100}$$

or if $\mathcal{I}(\mu_\bullet,w)=\infty$, then choose $\mathcal{U}(\mu_\bullet,w)$ such that

$$\limsup_N \frac{1}{N}\log\mathbb{P}\left((\mu^N_\bullet,w^N)\in\mathcal{U}(\mu_\bullet,w)\right)\le -M. \tag{6.101}$$

The sets $\{\mathcal{U}(\mu_\bullet,w):(\mu_\bullet,w)\in\mathcal{A}\cap\mathcal{K}\}$ are an open cover of $\mathcal{A}\cap\mathcal{K}$, so by compactness we can find $n<\infty$ and $(\mu^{(i)}_\bullet,w^{(i)})\in\mathcal{A}\cap\mathcal{K}$ such that $\mathcal{A}\cap\mathcal{K}$ is covered by $\mathcal{U}_i=\mathcal{U}(\mu^{(i)}_\bullet,w^{(i)}), i\le n$ and conclude that for each N,

$$\mathbb{P}\left((\mu^N_\bullet,w^N)\in\mathcal{A}\right)\le\mathbb{P}\left((\mu^N_\bullet,w^N)\notin\mathcal{K}\right)+\sum_{i=1}^n\mathbb{P}\left((\mu^N_\bullet,w^N)\in\mathcal{U}_i\right). \tag{6.102}$$

The limiting exponential behaviour for the upper bound is then given by the maximum of the exponentials in each term:

$$\begin{aligned}\limsup_N &\frac{1}{N}\log\mathbb{P}\left((\mu^N_\bullet,w^N)\in\mathcal{A}\right)\\ &\le\max\left(\limsup\frac{1}{N}\log\mathbb{P}\left((\mu^N_\bullet,w^N)\in\mathcal{V}\right):\mathcal{V}=\mathcal{K}^c,\mathcal{U}_1,...,\mathcal{U}_n\right).\end{aligned} \tag{6.103}$$

The terms appearing in the right hand side are all *either* bounded by $-M\le-\inf_{\mathcal{A}}\mathcal{I}$, for the cases where $\mathcal{V}=\mathcal{K}^c$ or $\mathcal{V}=\mathcal{U}_i$, for a path $(\mu^{(i)}_\bullet,w^{(i)})$ with $\mathcal{I}(\mu^{(i)}_\bullet,w^{(i)})=\infty$, *or* at most $-\mathcal{I}(\mu^{(i)}_\bullet,w^{(i)})+\epsilon\le-\inf_{\mathcal{A}}\mathcal{I}+\epsilon$. All together, we conclude that

$$\limsup_N\frac{1}{N}\log\mathbb{P}\left(\mu^N_\bullet,w^N)\notin\mathcal{A}\right)\le-\inf\{\mathcal{I}(\mu_\bullet,w):(\mu_\bullet,w)\in\mathcal{A}\}+\epsilon \tag{6.104}$$

and taking $\epsilon\to0$ concludes the proof in the case where the infimum is finite. The case where the infimum is infinite is essentially identical: we now keep M as a free parameter, choose a compact set $\mathcal{K}$ such that $\limsup_N N^{-1}\log\mathbb{P}((\mu^N_\bullet,w^N)\notin\mathcal{K})\le-M$, and cover $\mathcal{A}\cap\mathcal{K}$ with open sets $\mathcal{U}(\mu_\bullet,w)\ni(\mu_\bullet,w)$ satisfying (6.101). The same covering argument then gives

$$\limsup_N\frac{1}{N}\log\mathbb{P}\left((\mu^N_\bullet,w^N)\notin\mathcal{A}\right)\le-M \tag{6.105}$$

and the conclusion follows by taking $M\to\infty$. □

6.3 Change of Measure for the Kac Process

In order to prove Theorem 6.5 and its consequences, we will use the following change of measure necessary to perturb the initial data and dynamics. The changes of measure we will use are as follows.

Proposition 6.15 (Kac process under change of measure). *Let μ_t^N be a Kac process with collision kernel B, and velocities initially sampled independently from $\mu_0^\star$, which is a Markov process on a filtered probability space $(\Omega, \mathfrak{F}, (\mathfrak{F}_t)_{t\geq 0}, \mathbb{P})$, and let w_t^N be the associated empirical flux. Let $\varphi : \mathbb{R}^d \to \mathbb{R}$ be such that $\int e^{\varphi(v)}\mu_0^\star(dv) = 1$, $A_0 \in \mathfrak{F}_0$ such that $c_N = \mathbb{E}[\mathbb{1}_A e^{N\langle \varphi, \mu_0^N\rangle}] > 0$, and let $K : \mathcal{P}_2^N \times E \to [0,\infty)$ be measurable and such that $K/(1+|v|+|v_\star|)$ is uniformly bounded. Define a new measure $\mathbb{Q}$ by*

$$\begin{aligned}\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp\Bigg(& N\langle \varphi, \mu_0^N\rangle + N\langle \log K(\mu_0^N, \cdot), w^N\rangle \\ & - N\int_0^{t_{\text{fin}}}\int_E (K-1)(\mu_0^N, t, v, v_\star, \sigma)\overline{w}_{\mu^N}(dt, dv, dv_\star, d\sigma)\Bigg) c_N^{-1}\mathbb{1}_A\end{aligned} \tag{6.106}$$

where we understand the right-hand side to be 0 if $supp(w^N) \cap \{K = 0\} \neq \emptyset$. Then $\mathbb{Q}$ is a probability measure, under which μ_0^N is given as the empirical measure of N independent draws from $e^{\varphi(v)}\mu_0^\star(dv)$ conditioned on $A_0 \in \mathfrak{F}_0$, and under which $(\mu_0^N, \mu_t^N, w_t^N)$ is a time-inhomogeneous Markov process, with time-dependent generator, for bounded $F : \mathcal{P}_2^N \times \mathcal{P}_2^N \times \mathcal{M}(E) \to \mathbb{R}$,

$$\begin{aligned}\mathcal{G}_t F(\nu, \mu^N, w^N) = N\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}} & (F(\nu, \mu^{N,v,v_\star,\sigma}, w^{N,t,v,v_\star,\sigma}) - F(\nu, \mu^N, w^N)) \\ & \cdots \times K(\nu, t, v, v_\star, \sigma)B(v-v_*, \sigma)\mu^N(dv)\mu^N(dv_\star)d\sigma.\end{aligned} \tag{6.107}$$

This is a version of the Girsanov theorem for jump processes; see, for example, [123, Appendix 1, Theorem 7.3], which is tailor-made for our purposes. The hypotheses on the growth of K are probably not the most general possible, but are sufficient for the applications in this paper in Sections 6.4, 6.6. Since this does not appear to be standard, a proof is given in Appendix 6.B.

6.4 Restricted Lower Bound

We now give a proof of the lower bound with the additional integrability hypothesis. The restricted lower bound is based on the following approximation lemma.

Lemma 6.16 (Approximation by Regular Paths)**.** *Let $(\mu_\bullet, w)$ be a measure-flux pair such that*

$$\mathcal{I}(\mu_\bullet, w) < \infty; \qquad \langle 1 + |v|^2 + |v_\star|^2, w\rangle < \infty.$$

Then there exists a sequence $(\mu^{(n)}_\bullet, w^{(n)})$ of measure-flux pairs whose tilting functions $K^{(n)}$ are continuous, such that and $K^{(n)}(t, v, v_\star, \sigma)B(v - v_\star, \sigma)$ is bounded and bounded away from 0, such that $\mu^{(n)}_0$ admits a bounded density with respect to $\mu^\star_0$, and such that

$$\sup_{t\le t_{\text{fin}}} \mathcal{W}_{1,1}(\mu^{(n)}_t, \mu_t) + \rho_1(w^{(n)}, w) \to 0; \qquad \limsup_n \mathcal{I}(\mu^{(n)}_\bullet, w) \le \mathcal{I}(\mu_\bullet, w). \tag{6.108}$$

Moreover, each $(\mu^{(n)}_\bullet, w^{(n)})$ is the unique measure-flux pair starting from $\mu^{(n)}_0$ and with tilting function $K^{(n)}$.

Throughout, we write the indexes $^{(n)}$ in the superscripts in brackets, to distinguish them from similar notation for the Kac process $\mu^N_\bullet, w^N$.The proof of this lemma is rather technical, and so is deferred until Subsection 6.4.1. Once this lemma is in hand, the restricted lower bound Theorem 6.2ii) follows straightforwardly from standard 'tilting' arguments, using the change-of-measure given in Proposition 6.15 via the following law of large numbers.

Lemma 6.17. *Let $(\mu_\bullet, w)$ be a measure-flux pair whose tilting function K is continuous and such that $KB(v - v_*, \sigma)$ is bounded and bounded away from 0, such that μ_0 has a bounded density with respect to $\mu^\star_0$, and such that the pair $(\mu_\bullet, w)$ is the unique measure-flux pair with this tilting function and this initial measure. Let $\mathbb{Q}^N$ be the measures given by Proposition 6.15 with $\varphi = \log \frac{d\mu_0}{d\mu^\star_0}$ and $K : E \to (0, \infty)$ the tilting function for $(\mu_\bullet, w)$. Then for all open sets $\mathcal{U} \ni (\mu_\bullet, w)$ and $\epsilon > 0$, we have*

$$\mathbb{Q}^N\left((\mu^N_\bullet, w^N) \in \mathcal{U}, \left|\frac{1}{N}\log\frac{d\mathbb{Q}^N}{d\mathbb{P}} - \mathcal{I}(\mu_\bullet, w)\right| < \epsilon\right) \to 1. \tag{6.109}$$

Proof. We start by applying Proposition 6.15. Since K is a function only $E \to \mathbb{R}$, (μ^N_t, w^N_t) is a Markov process with generator given by (6.107) applied to functions $F : \mathcal{P}^N_2 \times \mathcal{M}(E) \to \mathbb{R}$. For the initial data, μ^N_0 is given, under $\mathbb{Q}^N$, by sampling N particles independently with common law $e^{\log d\mu_0/d\mu^\star_0}\mu^\star_0 = \mu_0$.

Step 1: Functional Law of Large Numbers We begin by show that, under $\mathbb{Q}^N$, the pairs $(\mu^N_\bullet, w^N)$ converge in probability to $(\mu_\bullet, w)$. Since K is bounded and φ is bounded above, Corollary 6.9 applies and the laws $\mathbb{Q}^N \circ (\mu^N_\bullet, w^N)^{-1}$ are tight on $\mathcal{D} \times \mathcal{M}(E)$, so every subsequence has a further subsequence converging weakly on $\mathcal{D}\times\mathcal{M}(E)$. We will now prove that the only possible subsequential limit is $\delta_{(\mu_\bullet,w)}$, which implies that the whole sequence $\mathbb{Q}^N\circ(\mu^N_\bullet,w^N)^{-1}$ converges weakly to this limit, and hence

$$\mathbb{Q}^N\left((\mu^N_\bullet,w^N)\in\mathcal{U}\right)\to 1. \tag{6.110}$$

Let $L\subset\mathbb{N}$ be any subsequence along which $\mathbb{Q}^N\circ(\mu^N_\bullet,w^N)^{-1}$ converges weakly. thanks to Skorokhod's representation theorem, we can realise all $(\mu^N_\bullet,w^N), N\in L$ with these laws on a common probability space, with probability measure $\mathbb{Q}$, converging $\mathbb{Q}$-almost surely to a limit $(\widetilde{\mu}_\bullet,\widetilde{w})$. For each N, $(\mu^N_\bullet,w^N)$ almost surely lies in the set of pairs satisfies the continuity equation, which is closed by Lemma 6.13, and hence $(\widetilde{\mu}_\bullet,\widetilde{w})$ almost surely satisfies (CE). We now show that the limit is almost surely a measure-flux pair with tilting K: for all $g\in C_c(E)$, the processes

$$M^{N,g}_t=\langle g,w^N_t\rangle-\int_{E_t}g(s,v,v_\star,\sigma)K(s,v,v_\star,\sigma)B(v-v_*,\sigma)ds\mu^N_s(dv)\mu^N_s(dv_\star)d\sigma \tag{6.111}$$

is a càdlàg martingale, with previsible, increasing quadratic variation

$$\begin{aligned}[M^{N,g}]_t&=\frac{1}{N}\int_{E_t}g(s,v,v_\star,\sigma)^2K(s,v,v_\star,\sigma)B(v-v_*,\sigma)ds\mu^N_s(dv)\mu^N_s(dv_\star)d\sigma\\&\le\|g\|^2_\infty\sup_E(B(v-v_*,\sigma)K)t_{\text{fin}}/N,\end{aligned} \tag{6.112}$$

see, for instance, [49, 157]. In particular, since $B(v-v_*,\sigma)K$ is bounded by construction, the constant in the final expression is finite. Therefore, for all such g,

$$\mathbb{E}_\mathbb{Q}\left|\langle g,w^N\rangle-\int_E gK(s,v,v_\star,\sigma)B(v-v_*,\sigma)ds\mu^N_s(dv)\mu^N_s(dv_\star)d\sigma\right|\le\frac{C_g}{\sqrt{N}} \tag{6.113}$$

for some constant C_g. Taking $N\to\infty$ through S, the first term in the expectation converges almost surely to $\langle g,\widetilde{w}\rangle$, and the second term converges to $\int gKd\overline{w}_{\widetilde{\mu}}$ by Lemma 6.12 applied to gK. We now take $N\to\infty$ through L to obtain

$$\mathbb{E}_\mathbb{Q}\left|\langle g,\widetilde{w}\rangle-\int_E g(s,v,v_\star,\sigma)K(s,v,v_\star,\sigma)\overline{w}_{\widetilde{\mu}}(ds,dv,dv_\star,d\sigma)\right|=0 \tag{6.114}$$

and so the integrand is 0, $\mathbb{Q}$-almost surely. Taking a union bound over a countable dense set in $C_c(E)$, we conclude that $\widetilde{w}=K\overline{w}_{\widetilde{\mu}}$ almost surely, and the limit is a measure-flux pair with the prescribed rate function K. Since the initial velocities are drawn independently from μ_0 under $\mathbb{Q}$, we have the elementary convergence $\mu^N_0\to\widetilde{\mu}_0$ in $\mathbb{Q}$-probability, which implies that $\mu_0=\widetilde{\mu}_0$. By hypothesis, these properties uniquely characterise the desired limit $(\mu_\bullet,w)$, so $\mathbb{Q}((\widetilde{\mu}_\bullet,\widetilde{w})=(\mu_\bullet,w))=1$ and the step is complete.

Step 2: Law of Large Numbers for the Dynamic Cost. We will now show that the (random) exponential cost induced by the change of measure (6.106) converges under $\mathbb{Q}^N$: for all $\epsilon>0$,

$$\mathbb{Q}^N\left(\left|\frac{1}{N}\log\frac{d\mathbb{Q}^N}{d\mathbb{P}}-\mathcal{I}(\mu_\bullet,w)\right|>\epsilon\right)\to 0. \tag{6.115}$$

We begin by using the definitions of $\mathcal{I}$ and (6.106) to rewrite the difference as

$$\frac{1}{N}\log\frac{d\mathbb{Q}^N}{d\mathbb{P}}-\mathcal{I}(\mu_\bullet,w)=\langle\varphi,\mu_0^N\rangle-H(\mu_0|\mu_0^\star)+\langle\log K,w^N-w\rangle \\ -\int_E (K-1)(t,v,v_\star,\sigma)B(v-v_*,\sigma)\left(\mu_t^N(dv)\mu_t^N(dv_\star)-\mu_t(dv)\mu_t(dv_\star)\right)d\sigma \tag{6.116}$$

and examine the terms one by one. Fix, throughout, $\epsilon,\epsilon'>0$.

Step 2a: Cost of the Initial Data For the cost of the initial data, $\langle\varphi,\mu_0^N\rangle$ is the empirical mean of $\log\frac{d\mu_0}{d\mu_0^\star}$, sampled at N independent draws from μ_0. The mean of each draw is exactly $\int_{\mathbb{R}^d}\log\frac{d\mu_0}{d\mu_0^\star}(v)\mu_0(dv)=:H(\mu_0|\mu_0^\star)$, so by the weak law of large numbers, for all N large enough

$$\mathbb{Q}^N\left(|\langle\varphi,\mu_0^N\rangle-H(\mu_0|\mu_0^\star)|>\epsilon/4\right)<\epsilon'/3. \tag{6.117}$$

Step 2b: Integral against Empirical Flux Let us now examine the second term. By the choice of K, $\log K$ is continuous, bounded above, and bounded below by $\log(c/B(v-v_*,\sigma))$ for some constant $c>0$. We can further bound this below by

$$\log\frac{1}{B(v-v_*,\sigma)}\geq\log\frac{1}{(1+|v|)(1+|v_\star)}\geq -c(|v|+|v_\star|) \tag{6.118}$$

for a new constant c: in particular, $|\log K|\leq C(1+|v|+|v_\star|)$ is continuous, and of at most linear growth. Recalling that $B(v-v_*,\sigma)K$ is bounded, we also estimate, uniformly in N,

$$\mathbb{E}_{\mathbb{Q}^N}\left[\langle 1+|v|^2+|v_\star|^2,w^N\rangle\right] \\ =\mathbb{E}_{\mathbb{Q}^N}\left[\int_E(1+|v|^2+|v_\star|^2)B(v-v_*,\sigma)K(t,v,v_\star,\sigma)\mu_t^N(dv)\mu_t^N(dv_\star)d\sigma\right] \\ \leq C\mathbb{E}_{\mathbb{Q}^N}\langle 1+|v|^2,\mu_0^N\rangle=C\langle 1+|v|^2,\mu_0\rangle. \tag{6.119}$$

Elementary Chebychev estimates produce $R<\infty$ such that, uniformly in N,

$$\mathbb{Q}^N\left(\langle(1+|v|+|v_\star|)1\!\!1_{|v|>R\text{ or }|v_\star|>R},w^N\rangle>\epsilon/12C\right)<\epsilon'/6; \tag{6.120}$$

and similarly, using the boundedness of $B(v-v_*,\sigma)K$ and finiteness of the second moments, the second moment $\langle 1+|v|^2+|v_\star|^2,w\rangle<\infty$ is also finite, and so we can additionally choose R so that

$$\langle(1+|v|+|v_\star|)1\!\!1_{|v|>R\text{ or }|v_\star|>R},w\rangle<\frac{\epsilon}{12C} \tag{6.121}$$

and construct a continuous, compactly supported function $g:E\to\mathbb{R}$ such that $|g-\log K|\leq C(1+|v|+|v_\star|)$ and which agrees with $\log K$ when both $|v|,|v_\star|\leq R$. We therefore find from (6.120) that

$$\mathbb{Q}^N\left(\langle|g-\log K|,w^N\rangle>\epsilon/12\right)<\epsilon'/6; \qquad \langle|g-\log K|,w\rangle<\frac{\epsilon}{12}. \tag{6.122}$$

Thanks to the convergence in distribution, for N large enough,

$$\mathbb{Q}^N\left(|\langle g, w^N - w\rangle| > \epsilon/12\right) < \epsilon'/6 \tag{6.123}$$

and we find from (6.122,6.123) that

$$\mathbb{Q}^N\left(\left|\langle \log K, w^N - w\rangle\right| > \epsilon/4\right) < \epsilon'/3. \tag{6.124}$$

Step 2c: Integral against Compensator We finally deal with the third term in (6.116). Since $B(v-v_*,\sigma)K$ is bounded and K is continuous, it follows that $B(v-v_*,\sigma)(K-1)$ is of at most linear growth, so there exists C such that $|B(v-v_*,\sigma)(K-1)| \le C(1+|v|+|v_\star|)$, and as in the previous step, we can choose R such that, uniformly in N,

$$\mathbb{Q}^N\left(\sup_{t\le t_{\text{fin}}}\int_{\mathbb{R}^d\times\mathbb{R}^d}(1+|v|+|v_\star|)\mathbb{1}_{|v|>R \text{ or } |v_\star|>R}\mu^N_t(dv)\mu^N_t(dv_\star) > \epsilon/12Ct_{\text{fin}}\right) < \epsilon'/6 \tag{6.125}$$

for the new meaning of C, and similarly such that

$$\sup_{t\le t_{\text{fin}}}\int_{\mathbb{R}^d\times\mathbb{R}^d}(1+|v|+|v_\star|)\mathbb{1}_{|v|>R \text{ or } |v_\star|>R}\mu_t(dv)\mu_t(dv_\star) \le \epsilon/12Ct_{\text{fin}}. \tag{6.126}$$

We again truncate, with a proxy $h : E \to \mathbb{R}$ which is continuous, compactly supported, agrees with $(K-1)B(v-v_*,\sigma)$ if both $|v|, |v_\star| \le R$, and such that $|(K-1)B(v-v_*,\sigma)-h| \le C(1+|v|+|v_\star|)$ for the same constant C. Using Lemma 6.12 again,

$$\mathbb{Q}^N\left(\left|\int_E h(t,v,v_\star,\sigma)dt(\mu^N_t(dv)\mu^N_t(dv_\star) - \mu_t(dv)\mu_t(dv_\star))d\sigma\right| > \epsilon/12\right) < \epsilon'/6 \tag{6.127}$$

for N large enough, while (6.125) implies that

$$\mathbb{Q}^N\left(\left|\int_E (h - B(v-v_*,\sigma)(K-1))dt\mu^N_t(dv)\mu^N_t(dv_\star)d\sigma\right| > \frac{\epsilon}{12}\right) < \epsilon'/6 \tag{6.128}$$

and (6.126) implies that

$$\left|\int_E (h - B(v-v_*,\sigma)(K-1))dt\mu_t(dv)\mu_t(dv_\star)\right| < \frac{\epsilon}{12}. \tag{6.129}$$

Gathering (6.127, 6.128, 6.129), we conclude that, for N large enough,

$$\mathbb{Q}^N\left(\left|\int_E (K-1)B(v-v_*,\sigma)dt(\mu^N_t(dv)\mu^N_t(dv_\star) - \mu_t(dv)\mu_t(dv_\star))d\sigma\right| > \epsilon/4\right) < \epsilon'/3. \tag{6.130}$$

Returning to (6.116), we combine (6.117, 6.124, 6.130) to obtain, for all $\epsilon, \epsilon' > 0$, and all N large enough, depending on ϵ, ϵ',

$$\mathbb{Q}^N\left(\left|\frac{1}{N}\log\frac{d\mathbb{Q}^N}{d\mathbb{P}} - \mathcal{I}(\mu_\bullet, w)\right| > \epsilon\right) < \epsilon' \tag{6.131}$$

and we have proven the desired convergence (6.115). Together with the previous step, the proof is complete. □

We can now prove the restricted lower bound.

Proof of Theorem 6.2ii). Let us fix a Skorokhod-open set $\mathcal{U}$, a path $(\mu_\bullet, w) \in \mathcal{U} \cap \mathcal{R}$, and $\epsilon > 0$. Let us assume that $\mathcal{I}(\mu_\bullet, w) < \infty$. Thanks to Lemma 6.16, there exists a measure-flux pair $(\mu'_\bullet, w') \in \mathcal{U}$ with overall cost $\mathcal{I}(\mu'_\bullet, w') < \mathcal{I}(\mu_\bullet, w) + \epsilon$, satisfying the conclusions of Lemma 6.16. For the changes of measure $\mathbb{Q}^N \ll \mathbb{P}$ as in Lemma 6.17, we then have, for all N large enough,

$$\mathbb{Q}^N\left((\mu^N_\bullet, w^N) \in \mathcal{U}, \frac{d\mathbb{Q}^N}{d\mathbb{P}} \le \exp\left(N(\mathcal{I}(\mu'_\bullet, w') + \epsilon)\right)\right) \ge \frac{1}{2} \tag{6.132}$$

which implies that

$$\begin{aligned}
&\mathbb{P}\left((\mu^N_\bullet, w^N) \in \mathcal{U}\right) \\
&\qquad \ge \mathbb{E}_{\mathbb{Q}^N}\left[\left(\frac{d\mathbb{Q}^N}{d\mathbb{P}}\right)^{-1} 1\!\!1\left(\frac{d\mathbb{Q}^N}{d\mathbb{P}} \le \exp\left(N(\mathcal{I}(\mu'_\bullet, w') + \epsilon)\right), (\mu^N_\bullet, w^N) \in \mathcal{U}\right)\right] \\
&\qquad \ge \frac{1}{2}\exp\left(-N(\mathcal{I}(\mu'_\bullet, w') + \epsilon)\right) \ge \frac{1}{2}\exp\left(-N(\mathcal{I}(\mu_\bullet, w) + 2\epsilon)\right).
\end{aligned} \tag{6.133}$$

Taking the logarithm and the limit $N \to \infty$, and then $\epsilon \to 0$,

$$\liminf_N \frac{1}{N}\log \mathbb{P}\left((\mu^N_\bullet, w^N) \in \mathcal{U}\right) \ge -\mathcal{I}(\mu_\bullet, w). \tag{6.134}$$

Of course, (6.134) trivially holds if $\mathcal{I}(\mu_\bullet, w) = \infty$, and so applies to any $(\mu_\bullet, w) \in \mathcal{U} \cap \mathcal{R}$, and the result is proven. □

6.4.1 Proof of Approximation Lemma

We will now present the proof of the approximation lemma as a number of intermediate steps. We will present the statements here, to give an overview of the proof of the overall approximation lemma, and the proofs in Subsection 6.4.2. We begin with the following definition.

Definition 6.4.1. *Let $(\mu_\bullet, w)$ be a measure-flux pair, and $\lambda > 0$. Let g_λ be the Gaussian in $\mathbb{R}^d$*

$$g_\lambda(x) := \frac{1}{(2\pi\lambda)^{d/2}}\exp\left(-|x|^2/2\lambda\right). \tag{6.135}$$

We define the convolutions $(g_\lambda \star \mu_\bullet), g_\lambda \star w$ by

$$(g_\lambda \star \mu_\bullet)_t(dv) = (g_\lambda \star \mu_t)(dv) = \int_{\mathbb{R}^d} g_\lambda(v-u)\mu_t(du)dv; \tag{6.136}$$

$$(g_\lambda \star w)(dt, dv, dv_\star, d\sigma) = \int_{\mathbb{R}^d\times\mathbb{R}^d} g_\lambda(v-u)g_\lambda(v_\star - u_\star)dvdv_* w(dt, du, du_\star, d\sigma). \tag{6.137}$$

The measures $g_\lambda \star \mu_t$ are absolutely continuous with respect to the Lebesgue measure; we will alternatively use the notation $g_\lambda \star \mu_t$ for their density on $\mathbb{R}^d$.

Remark 6.18. *Let us note that the choice of Gaussian mollification is essential here, as it is the unique mollifier which is invariant under changing between the pre- and post-collisional velocities; see (6.152).*

Lemma 6.19 (Approximation by Convolution)**.** *Suppose $(\mu_\bullet, w)$ is a measure-flux pair with a bounded tilting function K, such that*

$$\langle |v|^2+|v_\star|^2, w\rangle<\infty; \qquad \mathcal{J}(\mu_\bullet, w)<\infty; \qquad \sup_{t\le t_{\text{fin}}}\langle |v|^2, \mu_t\rangle = \langle |v|^2, \mu_0\rangle<\infty. \tag{6.138}$$

Then, for all $\lambda>0$, $(\mu_\bullet\star g_\lambda, w\star g_\lambda)$ is a measure-flux pair. Furthermore, there exists a continuous function $\vartheta:[0,1]\to[0,\infty)$, which is continuous at 0 and $\vartheta(0)=0$, and a constant C, which only depends on upper bounds for the quantities in (6.138) *and not the boundedness of K, such that for all $\lambda\in(0,1]$,*

$$\mathcal{J}(g_\lambda\star\mu_\bullet, g_\lambda\star w)\le \mathcal{J}(\mu_\bullet, w)+C\vartheta(\lambda). \tag{6.139}$$

Finally, the tilting function K^λ satisfies

$$\sup_{t,v,v_\star,\sigma} K^\lambda(t,v,v_\star,\sigma)B(v-v_*,\sigma)\le \sup_{t,v,v_\star,\sigma} K(t,v,v_\star,\sigma)B(v-v_*,\sigma) \tag{6.140}$$

We now apply this to produce some approximation results.

Lemma 6.20. *Let $\mu_\bullet, w$ be as in Lemma 6.16. Then there exist measure-flux pairs $\mu^{(n)}_\bullet, w^{(n)}$ such that*

$$\mathcal{J}(\mu^{(n)}_\bullet, w^{(n)})\to\mathcal{J}(\mu_\bullet, w) \tag{6.141}$$

$$\sup_{t\le t_{\text{fin}}}\|(1+|v|^2)(\mu^{(n)}_t-\mu_t)\|_{\text{TV}}+\|(1+|v|^2+|v_\star|^2)(w^{(n)}-w)\|_{\text{TV}}\to 0 \tag{6.142}$$

and, for each n, the tilting function $K^{(n)}$ is such that $K^{(n)}(t,v,v_\star,\sigma)B(v-v_*,\sigma)$ is bounded and $K^{(n)}$ is continuous in $v, v_\star$. Furthermore, the starting points $\mu^{(n)}_0$ can be taken to be of the form

$$\mu^{(n)}_0(dv)=c_n^{-1}\left(\mu_0(dv)+\nu^{(n)}(dv)\right) \tag{6.143}$$

for a suitable normalising constant $c_n\to 1$ and measures $\nu^{(n)}$ with $\langle 1+|v|^2, \nu^{(n)}\rangle\to 0$.

Lemma 6.21. *Let $\mu_\bullet, w$ be as in Lemma 6.16. Then there exist measure-flux pairs $\mu^{(n)}_\bullet, w^{(n)}$ such that*

$$\mathcal{I}(\mu^{(n)}_\bullet, w^{(n)})\to\mathcal{I}(\mu_\bullet, w); \qquad \sup_{t\le t_{\text{fin}}}\mathcal{W}_{1,1}\left(\mu^{(n)}_t, \mu_t\right)+\rho_1(w^{(n)}, w)\to 0 \tag{6.144}$$

and, for each n, the tilting function $K^{(n)}$ is such that $K^{(n)}(t,v,v_\star,\sigma)B(v-v_*,\sigma)$ is bounded and $K^{(n)}$ is continuous in $v, v_\star$.

Lemma 6.22. *Let $\mu_\bullet, w$ be a measure-flux pair with finite rate $\mathcal{I}(\mu_\bullet, w) < \infty$, such that $K(t,v,v_\star,\sigma)B(v-v_*,\sigma)$ is bounded and K is continuous in $v, v_\star$. Then there exist measure-flux pairs $\mu^{(n)}_\bullet, w^{(n)}$ with*

$$\mathcal{I}(\mu^{(n)}_\bullet, w) \to \mathcal{I}(\mu_\bullet, w); \qquad \sup_{t\le t_{\mathrm{fin}}} \|\mu^{(n)}_t - \mu_t\|_{\mathrm{TV}} + \|w^{(n)} - w\|_{\mathrm{TV}} \to 0 \tag{6.145}$$

and additionally, for each n,

$$\sup_{(t,v,v_\star,\sigma)} K^{(n)}(t,v,v_\star,\sigma)B(v-v_*,\sigma) \le \sup_{(t,v,v_\star,\sigma)} K(t,v,v_\star,\sigma)B(v-v_*,\sigma) + 1; \tag{6.146}$$

$$\inf_{(t,v,v_\star,\sigma)} K^{(n)}(t,v,v_\star,\sigma)B(v-v_*,\sigma) > 0; \tag{6.147}$$

$$\sup \frac{d\mu^{(n)}_0}{d\mu^\star_0} < \infty \tag{6.148}$$

and such that $K^{(n)}$ are continuous functions on E. Moreover, the approximations are uniquely characterised among measure-flux pairs by the initial value $\mu^{(n)}_0$ and the tilting function $K^{(n)}$.

Equipped with these lemmas, the stated result Lemma 6.16 follows by a standard diagonal argument.

Proof of Lemma 6.16. Let us fix $\mu_\bullet, w$ as given, and construct a sequence of approximating measure-flux pairs $\mu^{(n)}_\bullet, w^{(n)}$ as follows. By Lemma 6.21, there exists a pair $\mu^{(n,1)}_\bullet, w^{(n,1)}$ whose tilting function $K^{(n,1)}$ is continuous in $v, v_\star$ and $K^{(n,1)}B(v-v_*,\sigma)$ is bounded, and such that

$$\sup_{t\le t_{\mathrm{fin}}} \mathcal{W}_{1,1}\left(\mu^{(n,1)}_t, \mu_t\right) + \rho_1(w^{(n,1)}, w) + \left|\mathcal{I}(\mu^{(n,1)}_\bullet, w^{(n,1)}) - \mathcal{I}(\mu_\bullet, w)\right| < \frac{1}{n}. \tag{6.149}$$

Thanks to Lemma 6.22, we can approximate $\mu^{(n,1)}_\bullet, w^{(n,1)}$ by a further pair $\mu^{(n,2)}_\bullet, w^{(n,2)}$, whose tilting function $K^{(n,2)}$ is continuous and so that $K^{(n,2)}B(v-v_*,\sigma)$ is still bounded and bounded away from 0, and where $\mu^{(n,2)}_0$ has a bounded density with respect to $\mu^\star_0$, which is uniquely characterised among measure-flux pairs by the initial data $\mu^{(n,2)}_0$ and tilting function $K^{(n,2)}$, with further error

$$\sup_{t\le t_{\mathrm{fin}}} \|\mu^{(n,2)}_t - \mu^{(n,1)}_t\|_{\mathrm{TV}} + \|w^{(n,2)} - w^{(n,1)}\|_{\mathrm{TV}} + \left|\mathcal{I}(\mu^{(n,2)}_\bullet, w^{(n,2)}) - \mathcal{I}(\mu^{(n,1)}_\bullet, w^{(n,1)})\right| < \frac{1}{n}. \tag{6.150}$$

Combining (6.149, 6.150), we recall that the total variation distance on measures on $\mathbb{R}^d$, respectively E, dominates the Wasserstein$_1$ distance $\mathcal{W}_{1,1}$, respectively ρ_1, so the sequence $\mu^{(n)}_\bullet = \mu^{(n,2)}_\bullet, w^{(n)} = w^{(n,2)}$ has the desired properties. □

6.4.2 Proof of Lemmas

We start with the convolution lemma, which is the most difficult step.

Proof of Lemma 6.19. We divide the proof into several steps. Throughout, C will denote a constant, which may vary from line to line, but is allowed to depend *only* on the quantities specified in (6.138).

Step 1: $(g_\lambda \star \mu_\bullet, g_\lambda \star w)$ solves the continuity equation. This property is fairly well-known, see Erbar [73] or Basile [16], and we include a proof for completeness. If we fix $f \in C_b(\mathbb{R}^d)$, let us denote $g_\lambda \star f$ the convolution $(g_\lambda \star f)(v) := \int_{\mathbb{R}^d} g_\lambda(v-w)f(w)dw$, and observe that $\langle f, g_\lambda \star \mu_t\rangle = \langle g_\lambda \star f, \mu_t\rangle$ for all $t \in [0, t_{\text{fin}}]$. Now, using the continuity equation for $(\mu_\bullet, w)$ with the test function $g_\lambda \star f$, for all $t \in [0, t_{\text{fin}}]$,

$$\begin{aligned}\langle f, g_\lambda \star \mu_t\rangle - \langle f, g_\lambda \star \mu_0\rangle &= \langle g_\lambda \star f, \mu_t\rangle - \langle g_\lambda \star f, \mu_0\rangle \\ &= \int_{E_t} \Delta(g_\lambda \star f)(v, v_\star, \sigma) w(ds, dv, dv_\star, d\sigma).\end{aligned} \tag{6.151}$$

Let us now fix $v, v_\star, \sigma$, and observe that the map $\mathcal{T}_\sigma : (v, v_\star) \to (v', v'_\star)$ is a linear isometry of Euclidean distance on $(\mathbb{R}^d)^2$; for variables $(u', u'_\star) \in (\mathbb{R}^d)^2$, let us write $(u, u_\star)$ for the preimage under $\mathcal{T}_\sigma$. We therefore have

$$\begin{aligned}&(g_\lambda \star f)(v') + (g_\lambda \star f)(v'_\star) \\ &= \frac{1}{(2\pi\lambda)^d}\int_{\mathbb{R}^d\times\mathbb{R}^d} (f(u') + f(u'_\star)) \exp\left(-\frac{|u'-v'|^2 + |u'_\star - v'_\star|^2}{2\lambda}\right) du'du'_\star \\ &= \frac{1}{(2\pi\lambda)^d}\int_{\mathbb{R}^d\times\mathbb{R}^d} (f(u') + f(u'_\star)) \exp\left(-\frac{|u-v|^2 + |u_\star - v_\star|^2}{2\lambda}\right) du'du'_\star \\ &= \frac{1}{(2\pi\lambda)^d}\int_{\mathbb{R}^d\times\mathbb{R}^d} (f(u') + f(u'_\star)) \exp\left(-\frac{|u-v|^2 + |u_\star - v_\star|^2}{2\lambda}\right) dudu_\star \\ &= \int_{\mathbb{R}^d\times\mathbb{R}^d} (f(u') + f(u'_\star)) g_\lambda(u-v) g_\lambda(u_\star - v_\star) dudu_\star\end{aligned} \tag{6.152}$$

where the penultimate line makes the change of variables $(u', u'_\star) = \mathcal{T}_\sigma(u, u_\star)$, with unit determinant. We now substitute the resulting identity

$$\begin{aligned}\Delta(g_\lambda \star f)(v, v_\star, \sigma) &= \int_{\mathbb{R}^d\times\mathbb{R}^d} (f(u') + f(u'_\star) - f(u) - f(u_\star)) g_\lambda(u-v) g_\lambda(u_\star - v_\star) dudu_\star \\ &= \int_{\mathbb{R}^d\times\mathbb{R}^d} (\Delta f)(u, u_\star, \sigma) g_\lambda(u-v) g_\lambda(u_\star - v_\star) dudu_\star\end{aligned} \tag{6.153}$$

into (6.151) to obtain

$$\begin{aligned}\langle f, g_\lambda \star \mu_t\rangle - \langle f, g_\lambda \star \mu_0\rangle \\ &= \int_{E_t}\int_{\mathbb{R}^d\times\mathbb{R}^d} (\Delta f)(u,u_\star,\sigma)g_\lambda(u-v)g_\lambda(u_\star - v_\star)w(ds,dv,dv_\star,d\sigma)dudu_\star \\ &=: \int_{E_t} (\Delta f)(u,u_\star,\sigma)(g_\lambda\star w)(ds,du,du_\star,d\sigma)\end{aligned} \tag{6.154}$$

and, since f is arbitrary, we conclude that $(g_\lambda\star\mu_\bullet, g_\lambda\star w)$ satisfies the continuity equation (CE) as desired.

Step 2: Identification of the Tilting Function. To show that $(\mu_\bullet, w)$ is a measure-flux pair, and in preparation for estimating the dynamic cost, we will now explicitly find a tilting function. Let us write K for the tilting function for the pair $(\mu_\bullet, w)$, so that $w = K\overline{w}_\mu$. For any Borel subset $A\subset E$, we observe that

$$\begin{aligned}(g_\lambda\star w)(A) &= \int_E\int_{\mathbb{R}^d\times\mathbb{R}^d} \mathbb{I}_A(t,u,u_\star,\sigma)g_\lambda(u-v)g_\lambda(u_\star-v_\star)K(t,v,v_\star,\sigma) \\ &\qquad \cdots\times\overline{w}_\mu(dt,dv,dv_\star,d\sigma)dudu_\star \\ &= \int_E \mathbb{I}_A(t,u,u_\star,\sigma)\left(\int_{\mathbb{R}^d\times\mathbb{R}^d} g_\lambda(u-v)g_\lambda(u_\star-v_\star)K(t,v,v_\star,\sigma)B(v-v_*,\sigma)\mu_t(dv)\mu_t(dv_\star)\right) \\ &\qquad \cdots\times dudu_\star dtd\sigma.\end{aligned} \tag{6.155}$$

Now, let us define K^λ by

$$K^\lambda(t,u,u_\star,\sigma) := \frac{\int_{\mathbb{R}^d\times\mathbb{R}^d} g_\lambda(u-v)g_\lambda(u_\star-v_\star)K(t,v,v_\star,\sigma)B(v-v_*,\sigma)\mu_t(dv)\mu_t(dv_\star)}{B(u-u_\star)(g_\lambda\star\mu_t)(u)(g_\lambda\star\mu_t)(u_\star)} \tag{6.156}$$

and observe that this is a well-defined function, since K was assumed to be bounded and B is bounded away from 0, and where in the denominator $(g_\lambda\star\mu_t)(u)>0$ denotes the density of the measure $g_\lambda\star\mu_t$ with respect to the Lebesgue measure. Returning to (6.155), this definition yields

$$\begin{aligned}(g_\lambda\star w)(A) &= \int_E \mathbb{I}_A(t,u,u_\star,\sigma)K^\lambda(t,u,u_\star,\sigma)B(u-u_\star)(g_\lambda\star\mu_t)(u)du \\ &\qquad \cdots\times(g_\lambda\star\mu_t)(u_\star)du_\star dtd\sigma \\ &= \int_E \mathbb{I}_A(t,u,u_\star,\sigma)K^\lambda(t,u,u_\star,\sigma)(g_\lambda\star\mu_t)(du)(g_\lambda\star\mu_t)(du_\star)dtd\sigma \\ &= \int_E \mathbb{I}_A(t,u,u_\star,\sigma)K^\lambda(t,u,u_\star,\sigma)\overline{w}_{g_\lambda\star\mu}(dt,du,du_\star,d\sigma).\end{aligned} \tag{6.157}$$

We conclude that K^λ is a tilting function for $(g_\lambda\star\mu_\bullet, g_\lambda\star w)$, and so this is a measure-flux pair as claimed, and the bound (6.140) is immediate. For future convenience, we will now define

$$\overline{r}^\lambda_t(u,u_\star) := \int_{\mathbb{R}^d\times\mathbb{R}^d} g_\lambda(u-v)g_\lambda(u_\star-v_\star)\Psi(|v-v_*|)\mu_t(dv)\mu_t(dv_\star) \tag{6.158}$$

recalling that $\Psi(|v|)$ is the kinetic factor in the kernel, and introduce the proxy to K^λ given by

$$\begin{aligned}\overline{K}^\lambda(t,u,u_\star,\sigma) :&= \frac{\int_{\mathbb{R}^d\times\mathbb{R}^d} g_\lambda(u-v)g_\lambda(u_\star-v_\star)K(t,v,v_\star,\sigma)B(v-v_*,\sigma)\mu_t(dv)\mu_t(dv_\star)}{\overline{r}^\lambda_t(u,u_\star)} \\ &= \int_{\mathbb{R}^d\times\mathbb{R}^d} K(t,v,v_\star,\sigma)\nu^\lambda_{(t,u,u_\star)}(dv,dv_\star)\end{aligned} \tag{6.159}$$

which we have written in terms of the integral against the probability measures

$$\nu^\lambda_{(t,u,u_\star)}(dv,dv_\star) = \frac{\Psi(|v-v_*|)g_\lambda(v-u)\mu_t(dv)g_\lambda(v_\star-u_\star)\mu_t(dv_\star)}{\overline{r}^\lambda_t(u,u_\star)} \tag{6.160}$$

recalling that the kernels B of interest take the form $B(v,\sigma)=\Psi(|v|)$. For any $u,u_\star,\sigma$, the quotient is given by the function

$$\begin{aligned}\frac{K^\lambda}{\overline{K}^\lambda}(t,u,u_\star,\sigma) = \psi_\lambda(t,u,u_\star) &:= \frac{\overline{r}^\lambda_t(u,u_\star)}{\Psi(|u-u_\star|)(g_\lambda\star\mu_t)(u)(g_\lambda\star\mu_t)(u_\star)} \\ &= \frac{1}{\Psi(|u-u_\star|)}\int_{\mathbb{R}^d\times\mathbb{R}^d}\Psi(|v-v_*|)\frac{g_\lambda(v-u)\mu_t(dv)}{(g_\lambda\star\mu_t)(u)}\frac{g_\lambda(v_\star-u_\star)\mu_t(dv_\star)}{(g_\lambda\star\mu_t)(u_\star)}.\end{aligned} \tag{6.161}$$

which depends only on μ_t, and not on w.

Step 3: Decomposition of the Rate Function We now break the rate function up into several parts which can be more easily manipulated. We start from

$$\mathcal{J}(g_\lambda\star\mu_\bullet, g_\lambda\star w) = \int_E (\tau-1)(K^\lambda(t,v,v_\star,\sigma))\overline{w}_{g_\lambda\star\mu}(dt,dv,dv_\star,d\sigma) + \overline{w}_{g_\lambda\star\mu}(E) \tag{6.162}$$

where we recall that $(\tau-1)(x) = x\log x - x$ is a convex function on $[0,\infty)$. Next, we observe that

$$(\tau-1)(K^\lambda) = K^\lambda\log\psi_\lambda + \psi_\lambda(\tau-1)(\overline{K}^\lambda). \tag{6.163}$$

For the first term, we start with the observation that

$$\begin{aligned}&K^\lambda(t,u,u_\star,\sigma)\overline{w}_{g_\lambda\star\mu}(dt,du,du_\star,d\sigma) \\ &\quad= \int_{\mathbb{R}^d\times\mathbb{R}^d} K(t,v,v_\star,\sigma)B(v-v_*,\sigma)\mu_t(dv)\mu_t(dv_\star)g_\lambda(u-v)g_\lambda(u_\star-v_\star)dtd\sigma dudu_\star\end{aligned} \tag{6.164}$$

from which it follows that we can rewrite the integral of the first term as an error term

$$\begin{aligned}\mathcal{T}_\lambda(\mu_\bullet, w) &:= \int_E K^\lambda(t,u,u_\star,\sigma)\log\psi_\lambda(t,u,u_\star)\overline{w}_{g_\lambda\star\mu}(dt,du,du_\star,d\sigma) \\ &= \int_{E\times\mathbb{R}^d\times\mathbb{R}^d} K(t,v,v_\star,\sigma)B(v-v_*,\sigma)g_\lambda(u-v)g_\lambda(u_\star-v_\star)\log\psi_\lambda(t,u,u_\star)dt \\ &\qquad\cdots\times\mu_t(dv)\mu_t(dv_\star)d\sigma du du_\star.\end{aligned} \tag{6.165}$$

To integrate the second term, we note that

$$\psi_\lambda(t,u,u_\star)\overline{w}_{g_\lambda\star\mu}(dt,du,du_\star,d\sigma) = \overline{r}^\lambda_t(u,u_\star)dt du du_\star d\sigma. \tag{6.166}$$

Since $\nu^\lambda_{(t,u,u_\star)}$ are probability measures, we can apply Jensen to (6.159) with the convex function $\tau-1$ to find

$$\begin{aligned}(\tau-1)(\overline{K}^\lambda(t,u,u_\star,\sigma)) &\le \int_{\mathbb{R}^d\times\mathbb{R}^d}(\tau-1)(K(t,v,v_\star,\sigma))\nu^\lambda_{(t,u,u_\star)}(dv,dv_\star) \\ &= \frac{1}{\overline{r}^\lambda_t(u,u_\star)}\int_{\mathbb{R}^d\times\mathbb{R}^d}(\tau-1)(K(t,v,v_\star,\sigma))B(v-v_*,\sigma)g_\lambda(u-v)g_\lambda(u_\star-v_\star)\mu_t(dv)\mu_t(dv_\star).\end{aligned} \tag{6.167}$$

Gathering (6.166, 6.167), we obtain

$$\begin{aligned}&\int_E \psi_\lambda(t,u,u_\star)(\tau-1)(\overline{K}^\lambda)(t,u,u_\star,\sigma)\overline{w}_{g_\lambda\star\mu}(dt,du,du_\star,d\sigma) \\ &\quad\le \int_{E\times\mathbb{R}^d\times\mathbb{R}^d}(\tau-1)(K(t,v,v_\star,\sigma))B(v-v_*,\sigma)g_\lambda(u-v)g_\lambda(u_\star-v_\star)dt\mu_t(dv)\mu_t(dv_\star)d\sigma \\ &\quad= \int_E(\tau-1)(K(t,v,v_\star,\sigma))B(v-v_*,\sigma)dt\mu_t(dv)\mu_t(dv_\star)d\sigma \\ &\quad= \int_E(\tau-1)(K(t,v,v_\star,\sigma))\overline{w}_\mu(dt,dv,dv_\star,d\sigma).\end{aligned} \tag{6.168}$$

Returning to (6.162) and using the analagous equation for $\mu_\bullet$, we finally obtain the decomposition

$$\mathcal{J}(g_\lambda\star\mu_\bullet, g_\lambda\star w) \le \mathcal{J}(\mu_\bullet, w) + (\overline{w}_{g_\lambda\star\mu}-\overline{w}_\mu)(E) + \mathcal{T}_\lambda(\mu_\bullet, w). \tag{6.169}$$

It is very straightforward to show that the second term converges to 0 as $\lambda\to 0$, with a rate depending only on $\sup_t\langle|v|^2,\mu_t\rangle$.

Step 4: Analysis of ψ_λ. We now turn to the error term $\mathcal{T}_\lambda$ identified in (6.165), which depends on the continuity of $\log\Psi$. We remark first that this term cannot be avoided purely on general considerations; consider, for example, the kernel $\Psi(|v|) = 1\!\!1_{|v|\ge 1}$ in which case $\mathcal{J}(g_\lambda\star\mu_\bullet, g_\lambda\star w)$ can become infinite due to contributions in the region $\{|v-v_\star|<1\}$.

We start with an upper bound for ψ_λ. We define first the measures

$$\xi^\lambda_{(t,u)}(dv) := \frac{g_\lambda(v-u)\mu_t(dv)}{(g_\lambda\star\mu_t)(u)}. \tag{6.170}$$

Setting $R:=\sqrt{2\sup_s\langle |v|^2,\mu_s\rangle}$, a Chebychev bound shows that $\mu_t(|v|\le R)\ge \frac{1}{2}$ for all t, which leads to the lower bound

$$(g_\lambda\star\mu_t)(u)\ge \frac{1}{2}\exp\left(-(|u|+R)^2)/2\lambda\right)/(2\pi\lambda)^{d/2}. \tag{6.171}$$

We now estimate

$$\begin{aligned}\left\langle |v|^2 1\!\!1_{|v-u|>R+|u|},\xi^\lambda_{(t,u)}\right\rangle &= \frac{\int_{|v-u|>R+|u|} e^{-|v-u|^2/2\lambda}|v|^2\mu_t(dv)/(2\pi\lambda)^{d/2}}{(g_\lambda\star\mu_t)(u)} \\ &\le \frac{e^{-(|u|+R)^2/2\lambda}/(2\pi\lambda)^{d/2}}{(g_\lambda\star\mu_t)(u)}\int_{|v-u|>R+|u|}|v|^2\mu_t(dv) \\ &\le 2\sup_s\langle |v|^2,\mu_s\rangle = R^2.\end{aligned} \tag{6.172}$$

Together with a trivial bound for the remaining region, we conclude that

$$\left\langle |v|^2,\xi^\lambda_{(t,u)}\right\rangle \le 3R^2+4|u|^2\le C(1+|u|^2) \tag{6.173}$$

since $R\ge 1$ depends only on the quantities in (6.138). Using the lower bound $\Psi(|u-u_\star|)\ge 1$ and the upper bound $\Psi(|v-v_*|)\le 1+|v|+|v_\star|\le C(1+|v|^2+|v_\star|^2)$, we now return to (6.161) to obtain

$$\psi_\lambda(t,u,u_\star)\le C(1+|u|^2+|u_\star|^2). \tag{6.174}$$

This bound will be useful in general, for $(u,u_\star)$ where ψ_λ cannot be shown to be close to 1. We complement this with a bound which will show that, for most points $(u,u_\star)$ in the support of $(g_\lambda\star\mu_t)^{\otimes 2}$, ψ_λ is not too much bigger than 1. We will exploit, repeatedly, the observation that, for the kernels of interest,

$$\Psi(|v-v_*|)\le (1+|v-u|+|v_\star-u_\star|)\Psi(|u-u_\star|). \tag{6.175}$$

Fix $t\ge 0$, and suppose that $v^0,v^0_\star$ are such that there exists a sets $U\ni v^0, U_\star\ni v^0_\star$ of diameter $c\sqrt{\lambda}$ and $\mu_t(U),\mu_t(U_\star)\ge \epsilon\lambda^{d/2}$, and $(u,u_\star)$ is such that $|(u,u_\star)-(v^0,v^0_\star)|<x\sqrt{\lambda}$, for some constant c and parameters $\epsilon>0, x<\infty$ to be chosen later. In this case, we bound the denominator below by observing that

$$(g_\lambda\star\mu_t)(u)\ge \frac{1}{(2\pi\lambda)^{d/2}}e^{-(x\sqrt{\lambda}+c\sqrt{\lambda})^2/2\lambda}\mu_t(U)\ge \frac{\epsilon}{(2\pi)^{d/2}}\exp\left(-x^2-c^2\right) \tag{6.176}$$

and similarly for $u_\star$. In this setting, we return to (6.161) and split the integral defining ψ_λ into the regions depending on whether $|(v,v_\star)-(u,u_\star)|<\lambda^{1/3}+3x\sqrt{\lambda}$ or not. In the first case,

$$\Psi(|v-v_*|)\le (1+2\lambda^{1/3}+6x\sqrt{\lambda})\Psi(|u-u_\star|) \tag{6.177}$$

and so

$$\begin{aligned}\int_{|(v,v_\star)-(u,u_\star)|<\lambda^{1/3}+3x\sqrt{\lambda}} \frac{\Psi(|v-v_*|)}{\Psi(|u-u_\star|)}\,\frac{g_\lambda(v-u)\mu_t(dv)}{(g_\lambda\star\mu_t)(u)}\,\frac{g_\lambda(v_\star-u_\star)\mu_t(dv_\star)}{(g_\lambda\star\mu_t)(u_\star)} \\ \le 1+2\lambda^{1/3}+6x\sqrt{\lambda}.\end{aligned} \tag{6.178}$$

On the other hand, using (6.176) and the trivial bounds $\Psi(|v-v_*|)\le 1+|v|+|v_\star|$, $\Psi(|u-u_\star|)\ge 1$, we bound the term from the second region by

$$\begin{aligned}&\int_{|(v,v_\star)-(u,u_\star)|\ge\lambda^{1/3}+3x\sqrt{\lambda}} \frac{\Psi(|v-v_*|)}{\Psi(|u-u_\star|)}\,\frac{g_\lambda(v-u)\mu_t(dv)}{(g_\lambda\star\mu_t)(u)}\,\frac{g_\lambda(v_\star-u_\star)\mu_t(dv_\star)}{(g_\lambda\star\mu_t)(u_\star)} \\ &\qquad\le C\int_{\mathbb{R}^d\times\mathbb{R}^d}(1+|v|+|v_\star|)\lambda^{-d}\exp\left(-\frac{1}{2\lambda}(9x^2\lambda+\lambda^{2/3})\right)\epsilon^{-2}\exp(2x^2+2c^2) \\ &\qquad\qquad\cdots\times\mu_t(dv)\mu_t(dv_\star) \\ &\qquad\le C\exp\left(2c^2-x^2\right)\epsilon^{-2}.\end{aligned} \tag{6.179}$$

where, in the final line, we recall that $\lambda^{-d}\exp(-\lambda^{-1/6}/2)$ is uniformly bounded on $(0,\infty)$, and that the remaining integral is controlled in terms of the second moments of μ_t, so can be absorbed into C. Gathering (6.177, 6.179), we conclude that for $(u,u_\star), x, \epsilon$ as above,

$$\begin{aligned}\psi_\lambda(t,u,u_\star)&\le 1+2\lambda^{1/3}+6x\sqrt{\lambda}+C\exp(2c^2-x^2)\epsilon^{-2} \\ &\le (1+2\lambda^{1/3}+6x\sqrt{\lambda})\left(1+C\exp(2c^2-x^2)\epsilon^{-2}\right).\end{aligned} \tag{6.180}$$

Step 5: Analysis of $\mathcal{T}_\lambda$. Equipped with this preliminary analysis of ψ_λ in the previous step, we bound the final term $\mathcal{T}_\lambda$ appearing in (6.169). Together with the observation under (6.169), this suffices to prove (6.139) and finish the proof of the lemma.

We break up the integration space $E\times\mathbb{R}^d\times\mathbb{R}^d$ in the definition of $\mathcal{T}_\lambda$ as follows. For $M\ge 2, R\in[3,\infty)\cap\sqrt{\lambda}\mathbb{N}, x\in(0,\infty),\epsilon\in(0,\infty)$ to be chosen later, we form a partition $\mathfrak{P}$ of $(-R,R]^d$ into $(2R/\sqrt{\lambda})^d$ translates of $(0,\sqrt{\lambda}]^d$, and for $v\in(-R,R]^d$, write $\mathcal{B}(v)$ for the unique $\mathcal{B}\in\mathfrak{P}$ containing v. We now consider the partition of $E\times\mathbb{R}^d\times\mathbb{R}^d$ given by

$$A_1 := \Big\{v,v_\star\in(-R,R]^d, K(t,v,v_\star,\sigma)\le M, \mu_t(\mathcal{B}(v))\ge\epsilon\lambda^{d/2}, \mu_t(\mathcal{B}(v_\star))\ge\epsilon\lambda^{d/2}, \\ |(u,u_\star)-(v,v_\star)|<x\sqrt{\lambda}\Big\};$$

$$A_2 := \Big\{v, v_\star \in (-R,R]^d, K(t,v,v_\star,\sigma) \le M, \mu_t(\mathcal{B}(v)) \ge \epsilon\lambda^{d/2}, \mu_t(\mathcal{B}(v_\star)) \ge \epsilon\lambda^{d/2},$$

$$|(u,u_\star)-(v,v_\star)| \ge x\sqrt{\lambda}\Big\};$$

$$A_3 := \left\{v, v_\star \in (-R,R]^d, K(t,v,v_\star,\sigma) \le M, \mu_t(\mathcal{B}(v)) < \epsilon\lambda^{d/2} \text{ or } \mu_t(\mathcal{B}(v_\star)) < \epsilon\lambda^{d/2}\right\};$$

$$A_4 := \left\{v, v_\star \in (-R,R]^d, K(t,v,v_\star,\sigma) > M\right\};$$

$$A_5 := \left\{(v,v_\star) \notin (-R,R]^{2d}\right\}. \tag{6.181}$$

We analyse the contributions from these regions one-by-one. Roughly, A_1 is the 'good' region, containing most of the contributions from the integrating measure, where $\log\psi_\lambda$ is small by (6.180), and the remaining terms are small, depending on the parameters M, R, x, ϵ; at the end, we will optimise, so that $M, R, x \to \infty$ and $\epsilon \to 0$ as functions of $\lambda \to 0$.

Step 5a: Contribution from A_1 For the region A_1, we observe that the hypotheses leading to (6.180) hold, with $U = \mathcal{B}(v), U_\star = \mathcal{B}(v_\star)$ and $c = \sqrt{d}$ is an absolute constant. Further, $K < M$ and $B(v - v_*, \sigma) \le 1 + 2R \le CR$, so we integrate (6.180) to find

$$\int_{A_1} K(t,v,v_\star,\sigma)B(v-v_*,\sigma)g_\lambda(u-v)g_\lambda(u_\star-v_\star)\log\psi_\lambda(t,u,u_\star)dt\mu_t(dv)\mu_t(dv_\star)d\sigma du du_\star$$
$$\le CMR\left(\log(1+2\lambda^{1/3}+6x\sqrt{\lambda}) + \log(1+\epsilon^{-2}e^{-x^2})\right). \tag{6.182}$$

Step 5b: Contribution from A_2. For A_2, we use the general upper bound (6.174), which is valid without restriction on $u, u_\star$. For fixed $v, v_\star$ we use Hölder's inequality to see that

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \log\psi_\lambda(t,u,u_\star)\mathbb{I}_{|(u,u_\star)-(v,v_\star)|\ge x\sqrt{\lambda}}g_\lambda(u-v)g_\lambda(u_\star-v_\star)du du_\star$$
$$\le C\int_{\mathbb{R}^d\times\mathbb{R}^d} \log(1+|u|^2+|u_\star|^2)\mathbb{I}_{|(u,u_\star)-(v,v_\star)|\ge x\sqrt{\lambda}}g_\lambda(u-v)g_\lambda(u_\star-v_\star)du du_\star$$
$$\le C\left(\int_{\mathbb{R}^d\times\mathbb{R}^d}(1+|u|^2+|u_\star|^2)g_\lambda(u-v)g_\lambda(u_\star-v_\star)du du_\star\right)^{1/2}$$
$$\cdots\times\left(\int_{\mathbb{R}^d\times\mathbb{R}^d}\mathbb{I}[|(u',u'_\star)| > x\sqrt{\lambda}]g_\lambda(u')g_\lambda(u'_\star)du'du'_\star\right)^{1/2}$$
$$\le C(1+|v|+|v_\star|)\exp\left(-x^2/16d\right) \tag{6.183}$$

where for the first factor in the final line, we integrated $\int |u|^2 g_\lambda(u-v)du = |u|^2 + \lambda^2 \le |u|^2+1$, and for the second factor we used standard tail estimates for the normal distribution, absorbing constants into the prefactor C. Bounding $B(v-v_*,\sigma), K$ as above, and integrating over $t, v, v_\star, \sigma$, we find

$$\int_{A_2} K(t,v,v_\star,\sigma)B(v-v_*,\sigma)g_\lambda(u-v)g_\lambda(u_\star-v_\star)\log\psi_\lambda(t,u,u_\star)dt\mu_t(dv)\mu_t(dv_\star)d\sigma du du_\star \\ \le CMR\exp\left(-x^2/16d\right). \tag{6.184}$$

Step 5c: Contribution from A_3. Similarly to the previous step, we start from a bound on the integrals over $u, u_\star$, with t and $v, v_\star$ fixed. We split the integral over $u, u_\star$ into $\{|(u,u_\star)-(v,v_\star)| \le (1+|v|+|v_\star|)\}$ and $\{|(u,u_\star)-(v,v_\star)| > (1+|v|+|v_\star|)\}$. In the first region, thanks to (6.175),

$$\log\psi_\lambda(t,u,u_\star) \le C\log(1+|u|^2+|u_\star|^2) \le C\log(1+|v|^2+|v_\star|^2) \tag{6.185}$$

while the contribution from the second region is controlled by using Hölder's inequality in the same way as (6.183) to obtain

$$\int_{|(u,u_\star)-(v,v_\star)|>(1+|v|+|v_\star|)} \log\psi_\lambda(t,u,u_\star)g_\lambda(u-v)g_\lambda(u_\star-v_\star)du du_\star \\ \le C(1+|v|+|v_\star|)\exp(-(1+|v|+|v_\star|)^2/16d). \tag{6.186}$$

This term can be absorbed into the contribution from (6.185), and we conclude that, for all t and $v, v_\star \in \mathbb{R}^d$,

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \log\psi_\lambda(t,u,u_\star)g_\lambda(u-v)g_\lambda(u_\star-v_\star)du du_\star \le C\log(1+|v|^2+|v_\star|^2). \tag{6.187}$$

In particular, when $v, v_\star \in (-R,R]^d$, the right-hand side can be replaced by $C\log R$, and $B(v-v_*,\sigma) \le CR$. We now integrate over A_3 to find

$$\int_{A_3} K(t,v,v_\star,\sigma)B(v-v_*,\sigma)g_\lambda(u-v)g_\lambda(u_\star-v_\star)\log\psi_\lambda(t,u,u_\star)dt\mu_t(dv)\mu_t(dv_\star)d\sigma du du_\star \\ \le CM(R\log R)\int_0^{t_{\mathrm{fin}}} \mu_t^{\otimes 2}\left((v,v_\star)\in(-R,R]^{2d} : \mu_t(\mathcal{B}(v)) < \epsilon\lambda^{d/2} \text{ or } \mu_t(\mathcal{B}(v_\star)) < \epsilon\lambda^{d/2}\right)dt \\ \le CM(R\log R)\int_0^{t_{\mathrm{fin}}} \mu_t\left(v\in(-R,R]^d : \mu_t(\mathcal{B}(v)) < \epsilon\lambda^{d/2}\right)dt \tag{6.188}$$

where the last line follows using a union bound, absorbing the factor of 2 into C. The integrand is now

$$\mu_t\left(v\in(-R,R]^d : \mu_t(\mathcal{B}(v)) < \epsilon\lambda^{d/2}\right) = \sum_{\mathcal{B}\in\mathfrak{P}} \mu_t(\mathcal{B})\mathbb{1}(\mu_t(\mathcal{B}) < \epsilon\lambda^{d/2}) \\ \le \epsilon\lambda^{d/2}(\#\mathfrak{P}) = \epsilon(2R)^d \tag{6.189}$$

recalling that $\#\mathfrak{P} = (2R/\sqrt{\lambda})^d$. Substituting this bound back into (6.188) we conclude that

$$\int_{A_3} K(t,v,v_\star,\sigma)B(v-v_*,\sigma)g_\lambda(u-v)g_\lambda(u_\star-v_\star)\log\psi_\lambda(t,u,u_\star)dt\mu_t(dv)\mu_t(dv_\star)d\sigma dudu_\star \\ \leq CM(R^{d+1}\log R)\epsilon. \tag{6.190}$$

Step 5d: Contribution from A_4. In A_4, we use the same bound (6.187) on $\int \log\psi_\lambda g_\lambda(u-v)g_\lambda(u_\star-v_\star)dudu_\star$, and observe that, on A_4, $K(t,v,v_\star,\sigma) \leq \frac{M}{\tau(M)}\tau(K)$. Integrating, it follows that

$$\begin{aligned}\int_{A_4} K(t,v,v_\star,\sigma)&B(v-v_*,\sigma)g_\lambda(u-v)g_\lambda(u_\star-v_\star)\log\psi_\lambda(t,u,u_\star)dt\mu_t(du)\mu_t(dv_\star)d\sigma dudu_\star \\ &\leq C(\log R)\left(\frac{M}{\tau(M)}\right)\int_E \tau(K(t,v,v_\star,\sigma))\overline{w}_\mu(dt,dv,dv_\star,d\sigma) \\ &= C(\log R)\left(\frac{M}{\tau(M)}\right)\mathcal{J}(\mu_\bullet,w) = C(\log R)\left(\frac{M}{\tau(M)}\right)\end{aligned} \tag{6.191}$$

since C is allowed to depend on an upper bound for $\mathcal{J}(\mu_\bullet,w)$.

Step 5e: Contribution from A_5. We finally turn to the contribution from A_5. Thanks to (6.187), for any $(v,v_\star)\notin(-R,R]^{2d}$, we have

$$\begin{aligned}\int_{\mathbb{R}^d\times\mathbb{R}^d} \log\psi_\lambda(t,u,u_\star)g_\lambda(u-v)g_\lambda(u_\star-v_\star)dudu_\star &\leq C\log(1+|v|^2+|v_\star|^2) \\ &\leq C\frac{\log(R^2)}{R^2}(1+|v|^2+|v_\star|^2)\end{aligned} \tag{6.192}$$

as $(\log x)/x$ is decreasing on $[R,\infty)\subset[e,\infty)$, and $1+|v|^2+|v_\star|^2 \geq R^2$. Integrating over $t,v,v_\star,\sigma$, we obtain

$$\begin{aligned}\int_{A_5} K(t,v,v_\star,\sigma)&B(v-v_*,\sigma)g_\lambda(u-v)g_\lambda(u-v_\star)\log\psi_\lambda(t,u,u_\star)dt\mu_t(dv)\mu_t(dv_\star)d\sigma dudu_\star \\ &\leq C\left(\frac{\log R^2}{R^2}\right)\int_E(|v|^2+|v_\star|^2)K(t,v,v_\star,\sigma)\overline{w}_\mu(dt,dv,dv_\star,d\sigma) \\ &= C\left(\frac{\log R}{R^2}\right)\langle|v|^2+|v_\star|^2,w\rangle = C\left(\frac{\log R}{R^2}\right)\end{aligned} \tag{6.193}$$

recalling again that the second moment of w is one of the quantities (6.138) on which C is allowed to depend.

Step 5f: Conclusion Gathering (6.182, 6.184, 6.190, 6.191, 6.193), we conclude that, for any M, R, x, ϵ as above,

$$\mathcal{T}_\lambda(\mu_\bullet, w) \le C\Bigg(MR\log(1+2\lambda^{1/3}+6x\sqrt{\lambda}) + MR\log\left(1+e^{-x^2}\epsilon^{-2}\right) \\ \cdots + MRe^{-x^2/16d} + M(R^{d+1}\log R)\epsilon + (\log R)\left(\frac{M}{\tau(M)}\right) + \frac{\log R}{R^2}\Bigg). \tag{6.194}$$

We now define $\vartheta_1(\lambda)$ to be the infimum of the term in parantheses over the possible choices of M, R, ϵ, x described at the start of Step 5 for $\lambda > 0$, and $\vartheta_1(0) = 0$. Although this expression is somewhat complicated to optimise directly, it is straightforward to see that $\vartheta_1(\lambda) \to 0$ as $\lambda \downarrow 0$: given a target $\eta > 0$, we can choose R such that the last term is at $< \eta/6$, independently of M, ϵ, x, λ; with R thus fixed, we choose M such that the second-last term is $< \eta/6$ for all ϵ, x, λ, and so on. Returning to (6.169), one obtains an additional error, corresponding to the term $\overline{w}_{g_\lambda \star \mu}(E) - \overline{w}_\mu(E)$, which can easily be controlled, giving another term $C\vartheta_2(\lambda)$. Adding the two, the lemma is proven, with a new function ϑ. □

Proof of Lemma 6.20. We now prove Lemma 6.20 based on the following truncation argument.

Step 1: Definition For $n \ge 1$, let B_n be the set $\{v \in \mathbb{R}^d : |v| \le n\}$ and

$$E_{(n)} = \{(t, v, v_\star, \sigma) : B(v - v_*, \sigma)K(t, v, v_\star, \sigma) \le n, \text{ and } v, v_\star, v', v'_\star \in B_n\} \tag{6.195}$$

and define

$$\nu^{(n)} = \int_{E^c_{(n)}} (\delta_{v'} 1\!\!1_{v' \in B_n} + \delta_{v'_\star} 1\!\!1_{v'_\star \in B_n})\, w(dt, dv, dv_\star, d\sigma); \tag{6.196}$$

$$c_n = (\mu_0 + \nu^{(n)})(\mathbb{R}^d); \tag{6.197}$$

and let the flux be

$$\begin{aligned} w^{(n)}(dt, dv, dv_\star, d\sigma) &= c_n^{-1}\, K(t, v, v_\star, \sigma)\, 1\!\!1_{E_{(n)}}\, \overline{w}_\mu(dt, dv, dv_\star, d\sigma) \\ &= c_n^{-1} 1\!\!1_{E_{(n)}} w(dt, dv, dv_\star, d\sigma) \end{aligned} \tag{6.198}$$

Let $\mu^{(n)}_\bullet$ be given by

$$\mu^{(n)}_t = \frac{\mu_0 + \nu^{(n)}}{c_n} + \int_{E_t} \Delta(v, v_\star, \sigma)\;\, w^{(n)}(ds, dv, dv_\star, d\sigma). \tag{6.199}$$

This definition gives a signed measure with $\mu_t^{(n)} 1\!\!1_{B_n^c} = c_n^{-1}\mu_0 1\!\!1_{B_n^c} \geq 0$, and we further observe that for any Borel $A \subset B_n$ and $t \leq t_{\mathrm{fin}}$,

$$\begin{aligned}\mu_t^{(n)}(A) &\geq c_n^{-1}\mu_0(A) \\ &\quad + c_n^{-1}\int_{E_t}\left((\Delta 1\!\!1_A)(v,v_\star,\sigma)1\!\!1_{E_{(n)}} + (1\!\!1_A(v') + 1\!\!1_A(v'_\star))1\!\!1_{E^c_{(n)}}\right) w(ds,dv,dv_\star,d\sigma) \\ &\geq c_n^{-1}\left(\mu_0 + \int_{E_t}\Delta 1\!\!1_A(v,v_\star,\sigma)\, w(ds,dv,dv_\star,d\sigma)\right) = c_n^{-1}\mu_t(A).\end{aligned} \tag{6.200}$$

It follows that $\mu_t^{(n)}$ is a positive measure for all $t \leq t_{\mathrm{fin}}$, and thanks to the normalisation by c_n, it follows that $\mu_t^{(n)}$ is a probability measure. Moreover, it also follows that $\mu_t 1\!\!1_{B_n}$ is absolutely continuous with respect to $\mu_t^{(n)}$, and that

$$\frac{d(\mu_t 1\!\!1_{B_n})}{d\mu_t^{(n)}} \leq c_n \qquad \mu_t^{(n)}\text{-almost everywhere} \tag{6.201}$$

and the form (6.143) of $\mu_0^{(n)}$ is immediate by construction. Further, the conntinuity equation (CE) follows immediately by construction.

Step 2: Convergence of the Truncated Measure-Flux Firstly, we show that $\mu_\bullet^{(n)}$ approximates $\mu_\bullet$ uniformly in the weighted total variation norm $\|\xi\|_{\mathrm{TV}+2} = \langle 1+|v|^2, |\xi|\rangle$. At time 0,

$$\|\mu_0^{(n)} - \mu_0\|_{\mathrm{TV}+2} \leq c_n^{-1}\langle 1+|v|^2, \nu^{(n)}\rangle + \frac{|1-c_n|}{c_n}\langle 1+|v|^2, \mu_0\rangle. \tag{6.202}$$

In the first term,

$$\langle 1+|v|^2, \nu^{(n)}\rangle = \int_E ((1+|v'|^2)1\!\!1_{v'\in B_n} + (1+|v'_\star|^2)1\!\!1_{v_\star \in B_n})1\!\!1_{E^c_{(n)}} w(dt,dv,dv_\star,d\sigma) \to 0 \tag{6.203}$$

by applying dominated convergence: the integrand is at most $2(1+|v|^2+|v_\star|^2)$ by energy conservation, and by hypothesis, $\langle 2+|v|^2+|v_\star|^2, w\rangle < \infty$. It follows already from these estimates that $c_n \to 1$, and the second term on the right-hand side of (6.202) converges to 0. Similarly, we estimate

$$\begin{aligned}\left\|(\mu_t^{(n)} - \mu_0^{(n)}) - (\mu_t - \mu_0)\right\|_{\mathrm{TV}+2} \\ &\leq 4c_n^{-1}\int_{E^c_{(n)}}(1+|v|^2+|v_\star|^2)1\!\!1_{E_n^c} w(ds,dv,dv_\star,d\sigma) \\ &\quad + 4|1-c_n^{-1}|\int_E (1+|v|^2+|v_\star|^2) w(ds,dv,dv_\star,d\sigma) \\ &\to 0\end{aligned} \tag{6.204}$$

and we conclude that $\sup_{t\leq t_{\mathrm{fin}}} \|\mu_t^{(n)} - \mu_t\|_{\mathrm{TV}+2} \to 0$. A similar argument shows that

$$\|w^{(n)} - w\|_{\mathrm{TV}+2} \to 0 \tag{6.205}$$

as desired. Let us remark that this would still be true, working in TV rather than TV + 2, even without the additional integrability assumption, but we will crucially use the finiteness of $\langle 1+|v|^2, \nu^{(n)}\rangle$ in controlling the cost of the initial data after a convolution later.

Step 3: Tilting Function for the Truncated Pair We now construct the tilting function $K^{(n)}$, which completes the proof that $\mu^{(n)}_\bullet, w^{(n)}$ is a measure-flux pair. By construction, we have

$$\mu_t^{(n)} - \mu_0^{(n)} = \int_{E_t} \Delta(v, v_\star, \sigma) w^{(n)}(ds, dv, dv_\star, d\sigma) \tag{6.206}$$

and

$$\begin{aligned} w^{(n)}(dt, dv, dv_\star, d\sigma) &= c_n^{-1} K(t, v, v_\star, \sigma) 1\!\!1_{E_{(n)}}(t, v, v_\star, \sigma) B(v - v_\star, d\sigma)\mu_t(dv)\mu_t(dv_\star)dt \\ &= c_n^{-1} K(t, v, v_\star, \sigma) 1\!\!1_{E_{(n)}}(t, v, v_\star, \sigma) B(v - v_\star, d\sigma)(\mu_t 1\!\!1_{B_n})(dv)(\mu_t 1\!\!1_{B_n})(dv_\star)dt \end{aligned} \tag{6.207}$$

where, in the last line, we observe that $1\!\!1_{E_{(n)}} = 1\!\!1_{E_{(n)}} 1\!\!1_{B_n}(v) 1\!\!1_{B_n}(v_\star)$. Recalling the absolute continuity (6.201), we have

$$w^{(n)}(dt, dv, dv_\star, d\sigma) = K^{(n)}(t, v, v_\star, \sigma)\overline{w}_{\mu^{(n)}}(dt, dv, dv_\star, d\sigma) \tag{6.208}$$

where $K^{(n)}$ is given by

$$K^{(n)}(t, v, v_\star, \sigma) = c_n^{-1} K(t, v, v_\star, \sigma) 1\!\!1_{E_{(n)}}(t, v, v_\star, \sigma) \left(\frac{d(\mu_t 1\!\!1_{B_n})}{d\mu_t^{(n)}}\right)(v) \left(\frac{d(\mu_t 1\!\!1_{B_n})}{d\mu_t^{(n)}}\right)(v_\star). \tag{6.209}$$

From (6.201) and the definition of $E_{(n)}$,

$$B(v - v_*, \sigma) K^{(n)}(t, v, v_\star, \sigma) \le n c_n \tag{6.210}$$

is bounded, as claimed.

Step 4: Convergence of the Dynamic Cost It remains to show that $\mathcal{J}(\mu^{(n)}, w^{(n)}) \to \mathcal{J}(\mu, w)$. From the total variation convergence proven above, it follows that $(\mu^{(n)}_\bullet, w^{(n)}) \to (\mu_\bullet, w)$ in the topology of $\mathcal{D} \times \mathcal{M}(E)$. This implies that $\liminf_n \mathcal{J}(\mu^{(n)}, w^{(n)}) \ge \mathcal{J}(\mu, w)$ by lower semicontinuity (Lemma 6.13), and so it suffices to prove an upper bound. We start by observing that, by construction

$$\begin{aligned} (\tau - 1)(K^{(n)})\overline{w}_{\mu^{(n)}}(dt, dv, dv_\star, d\sigma) &= (\log K^{(n)} - 1) w^{(n)}(dt, dv, dv_\star, d\sigma) \\ &= c_n^{-1} 1\!\!1_{E_{(n)}} (\log K^{(n)} - 1) w(dt, dv, dv_\star, d\sigma) \end{aligned} \tag{6.211}$$

and that, on $E_{(n)}$, $K^{(n)} \le c_n K$, so

$$\begin{aligned} (\tau - 1)(K^{(n)})\overline{w}_{\mu^{(n)}}(dt, dv, dv_\star, d\sigma) &\le c_n^{-1}(\log K - 1 + \log c_n) w(dt, dv, dv_\star, d\sigma) \\ &= c_n^{-1}((\tau - 1)(K) + \log c_n)\overline{w}_\mu(dt, dv, dv_\star, d\sigma). \end{aligned} \tag{6.212}$$

Integrating, and recalling the definition of $\mathcal{J}$, we see that

$$\begin{aligned}\mathcal{J}(\mu^{(n)}_\bullet, w^{(n)}) - \overline{w}_{\mu^{(n)}}(E) &= \int_E (\tau-1)(K^{(n)})\overline{w}_{\mu^{(n)}}(dt,dv,dv_\star,d\sigma) \\ &\le c_n^{-1}(\mathcal{J}(\mu_\bullet,w)-\overline{w}_\mu(E)) + (c_n^{-1}\log c_n)\overline{w}_\mu(E).\end{aligned} \tag{6.213}$$

Using the weighted total variation convergence, it is straightforward to see that $\overline{w}_{\mu^{(n)}}(E) \to \overline{w}_\mu(E)$. Since $c_n \to 1$, we conclude that $\limsup_n \mathcal{J}(\mu^{(n)}_\bullet, w^{(n)}) \le \mathcal{J}(\mu,w)$ and we are done. □

Combining the previous two results, we prove Lemma 6.21. The main difficulty with the construction above is that the presence of $\nu^{(n)}$ may make the cost of the initial data large: *a priori* $\nu^{(n)}$ could be singular, which would give $H(\mu^{(n)}_0|\mu^\star_0)=\infty$. To avoid this, we will convolve with the mollifiers g_λ, at a scale $\lambda=\lambda_n$ to be chosen. For this reason, it is import to have the uniform convergence of the cost function in Lemma 6.19.

Proof of Lemma 6.21. Let $\mu_\bullet, w$ be as given, and let $\mu^{(n,0)}, w^{(n,0)}$ be the approximations produced by Lemma 6.20. We observe first that, thanks to the strong convergence (6.141, 6.142),

$$\sup_n \langle 1+|v|^2+|v_\star|^2, w^{(n,0)}\rangle < \infty; \qquad \sup_n \mathcal{J}(\mu^{(n,0)}_\bullet, w^{(n,0)}) < \infty; \tag{6.214}$$

$$\sup_n \sup_{t\le t_{\text{fin}}} \langle |v|^2, \mu^{(n,0)}_t\rangle < \infty. \tag{6.215}$$

For any $\lambda>0$, let $\mu^{(n,\lambda)}_\bullet, w^{(n,\lambda)}$ be the convolutions

$$\mu^{(n,\lambda)}_\bullet := g_\lambda \star \mu^{(n,0)}_\bullet; \qquad w^{(n,\lambda)} := g_\lambda \star w^{(n,0)}. \tag{6.216}$$

Thanks to Lemma 6.19 and (6.214, 6.215), there exists some C, uniform in n, such that

$$\mathcal{J}(\mu^{(n,\lambda)}, w^{(n,\lambda)}) \le \mathcal{J}(\mu^{(n,0)}, w^{(n,0)}) + C\vartheta(\lambda). \tag{6.217}$$

We consider now the cost due to the initial data. Firstly, we write

$$\mu^{(n,0)}_0 = (1-p_n)(g_\lambda \star \mu_0) + p_n(g_\lambda \star \xi^{(n)}) \tag{6.218}$$

with $p_n = \nu^{(n)}(\mathbb{R}^d)/c_n \to 0$, and $\xi^{(n)} := \nu^{(n)}/\nu^{(n)}(\mathbb{R}^d)$. Using the convexity of $H(\cdot|\mu^\star_0)$, we immediately have

$$H\left(\mu^{(n,\lambda)}|\mu^\star_0\right) \le (1-p_n)H(g_\lambda\star\mu_0|\mu^\star_0) + p_n H(g_\lambda \star \xi^{(n)}|\mu^\star_0). \tag{6.219}$$

We investigate these terms one at a time.

Step 1: Entropy of $H(g_\lambda \star \mu_0|\mu_0^\star)$. We first show that

$$\limsup_{\lambda\to 0} H(g_\lambda \star \mu_0|\mu_0^\star) \le H(\mu_0|\mu_0^\star) \tag{6.220}$$

where we recall that, since $\mathcal{I}(\mu_\bullet, w) < \infty$ by hypothesis, the right-hand side is a finite limit. Since μ_0 is absolutely continuous with respect to $\mu_0^\star$, it is absolutely continuous with respect to the Lebesgue measure; let us write f_0 for its density, and recall the notation $f_0^\star$ for the density of $\mu_0^\star$. We can then write $H(\mu_0|\mu_0^\star) = \int f_0 \log(f_0/f_0^\star) < \infty$, and, recalling that $f_0^\star \ge ce^{-z_3|v|^2}$ by Hypothesis 6.1iii), $\log f_0 \le \log(f_0/f_0^\star) - \log c + z_3|v|^2$. Since μ_0 has a finite second moment, we see that $\int f_0 \log f_0 dv < \infty$. Further, bounding

$$-\log f_0^\star \le \log c + z_3|v|^2$$

and

$$(\log f_0^\star)1\!\!1_{f_0^\star \ge 1} f_0 \le f_0^\star \exp(1\!\!1_{f_0^\star\ge 1}) + f_0 \log f_0$$

we conclude that $\int |\log f_0^\star| f_0 < \infty$. We now write, as a difference of finite integrals,

$$H(\mu_0|\mu_0^\star) = \int_{\mathbb{R}^d} f_0 \log f_0 dv + \int_{\mathbb{R}^d} (-\log f_0^\star) f_0(v) dv = \int_{\mathbb{R}^d} f_0 \log f_0 dv + \int_{\mathbb{R}^d} (-\log f_0^\star)\mu_0(dv) \tag{6.221}$$

and similarly

$$\begin{aligned} H(g_\lambda \star \mu_0|\mu_0^\star) &= \int_{\mathbb{R}^d} (g_\lambda \star f_0)\log(g_\lambda \star f_0) dv + \int_{\mathbb{R}^d} (g_\lambda \star f_0)(-\log f_0^\star) dv \\ &= \int_{\mathbb{R}^d} (g_\lambda \star f_0)\log(g_\lambda \star f_0) dv + \int_{\mathbb{R}^d} (-\log f_0^\star)(g_\lambda \star \mu_0)(dv). \end{aligned} \tag{6.222}$$

Let us fix $\epsilon > 0$. For the first term, we recall that the function $x \log x$ is convex on $[0,\infty)$, which implies that, for all $\lambda > 0$,

$$\int_{\mathbb{R}^d} (g_\lambda \star f_0)\log(g_\lambda \star f_0) dv \le \int_{\mathbb{R}^d} f_0 \log f_0 dv. \tag{6.223}$$

For the second term, we recall that $-\log f_0^\star$ is continuous, and $-\log f_0^\star \le -\log c + z_3|v|^2$ for some $c > 0$ and $z_3 < \infty$. Using the fact $\langle |v|^2, g_\lambda \star \mu_0\rangle = \langle |v|^2, \mu_t\rangle + d\lambda \to \langle |v|^2, \mu_0\rangle < \infty$ and $g_\lambda \star \mu_0 \to \mu_0$ weakly, we use the same argument as in Lemma 2.15 to check the uniform integrability

$$\limsup_M \limsup_{\lambda \to 0} \langle |v|^2 1\!\!1_{|v|\ge M}, g_\lambda \star \mu_0\rangle = 0 \tag{6.224}$$

whence there exists $M < \infty$ and $\lambda_0 > 0$ such that, for all $\lambda < \lambda_0$,

$$\int_{|v|>M} (-\log f_0^\star)_+ (g_\lambda \star \mu_0)(dv) \le \int_{|v|>M} \left|\log c + z_3|v|^2\right| (g_\lambda \star \mu_0)(dv) < \frac{\epsilon}{4} \tag{6.225}$$

where $_+$ denotes the positive part. Using the weak convergence $g_\lambda \star \mu_0 \to \mu_0$, there exists $\lambda_1 < \lambda_0$ such that, for all $\lambda < \lambda_1$,

$$\left| \int_{|v|\le M} (-\log f_0^\star)(g_\lambda \star \mu_0)(dv) - \int_{|v|\le M} (-\log f_0^\star)\mu_0(dv)\right| < \frac{\epsilon}{3} \tag{6.226}$$

since the indicator $1\!\!1_{|v|\le M}$ is discontinuous on a μ_0-measure set, by absolute continuity. Finally, observe that the map

$$\mu \mapsto \langle 1\!\!1_{|v|>M, f_0^\star\ge 1}(\log f_0^\star), \mu\rangle \tag{6.227}$$

is lower semicontinuous for the weak convergence, since the integrand is nonnegative, and is finite for $\mu = \mu_0^\star$ as noted above. Therefore, we can find $\lambda_2 < \lambda_1$ such that, for all $\lambda < \lambda_2$,

$$\int_{|v|>M, f_0^\star\ge 1} (\log f_0^\star)(g_\lambda \star \mu_0)(dv) > \int_{|v|>M, f_0^\star\ge 1} (\log f_0^\star)\mu_0(dv) - \frac{\epsilon}{3}. \tag{6.228}$$

For such λ, we split the second integral in (6.229) into regions $\{|v|\le M\}$, $\{|v|>M, f_0^\star \ge 1\}$ and $\{|v|>M, f_0^\star < 1\}$ to obtain

$$\begin{aligned} H(g_\lambda\star\mu_0|\mu_0^\star) \le & \int_{\mathbb{R}^d} f_0 \log f_0 dv + \int_{|v|\le M} (-\log f_0^\star)(g_\lambda\star\mu_0)(dv) \\ & - \int_{|v|>M, f_0^\star\ge 1} (\log f_0^\star)(g_\lambda\star\mu_0)(dv) + \int_{|v|>M, f_0^\star<1} |\log f_0^\star|(g_\lambda\star\mu_0)(dv) \\ \le & \int_{\mathbb{R}^d} f_0\log f_0(dv) + \left(\int_{|v|<M}(-\log f_0^\star)\mu_0(dv) + \frac{\epsilon}{4}\right) \\ & - \left(\int_{|v|>M, f_0^\star\ge 1}(\log f_0^\star)(g_\lambda\star\mu_0)(dv) - \frac{\epsilon}{4}\right) + \frac{\epsilon}{4} \\ = & \int_{\mathbb{R}^d} f_0 \log f_0 dv - \int_{|v|\le M \text{ or } f_0^\star \ge 1} (\log f_0^\star)\mu_0(dv) + \epsilon \\ \le & H(\mu_0|\mu_0^\star) + \epsilon \end{aligned} \tag{6.229}$$

and we have proven (6.220).

Step 2: Entropy of Remainder Term. We next turn to the convolution $g_\lambda \star \xi^{(n)}$. On the one hand, the density of $g_\lambda\star\xi^{(n)}$ is at most $g_\lambda(0)$; on the other hand, taking $R^2 = 2\langle |v|^2, \xi^{(n)}\rangle$, it follows by Chebychev that

$$\xi^{(n)}(|v|\le R) \ge 1 - R^{-2}\langle |v|^2, \xi^{(n)}\rangle = \frac{1}{2}. \tag{6.230}$$

For any fixed u, if $|v|\le R$ then $g_\lambda(u-v) \ge g_\lambda(0)\exp(-(|u|^2+R^2)/\lambda)$, and integrating over this region gives

$$(g_\lambda\star\xi^{(n)})(v) \ge \frac{1}{2}\exp\left(-(|u|^2 + 2\langle |v|^2, \xi^{(n)}\rangle)/\lambda\right) g_\lambda(0). \tag{6.231}$$

Together with Hypothesis 6.1iii), there exists a constant a_λ such that

$$\left|\log \frac{d(g_\lambda\star\xi^{(n)})}{d\mu_0^\star}(u)\right| \le c_\lambda\left(1 + |u|^2 + \langle |v|^2, \xi_0^{(n)}\rangle\right). \tag{6.232}$$

We now integrate, and recall that the second moment of $\langle |v|^2, g_\lambda \star \xi^{(n)}\rangle = d\lambda + \langle |v|^2, \xi^{(n)}\rangle$, to find

$$H(g_\lambda \star \xi^{(n)}|\mu_0^\star) = \int_{\mathbb{R}^d} \log \frac{d(g_\lambda \star \xi^{(n)})}{d\mu_0^\star}(v)(g_\lambda \star \xi^{(n)})(dv) \le a_\lambda \left\langle 1+|v|^2, \xi^{(n)}\right\rangle \tag{6.233}$$

potentially for a new choice of a_λ.

Step 3: Control of Overall Cost We now combine (6.217, 6.219, 6.220, 6.233) to see that, for some constant C and a_λ,

$$\begin{aligned}\mathcal{I}(\mu_\bullet^{(n,\lambda)}, w^{(n,\lambda)}) \le \mathcal{I}(\mu_\bullet, w) &+ \left(\mathcal{J}(\mu_\bullet^{(n,0)}, w^{(n,0)}) - \mathcal{J}(\mu_\bullet, w)\right) + C\vartheta(\lambda) \\ &+ \left(H(g_\lambda \star \mu_0|\mu_0^\star) - H(\mu_0|\mu_0^\star)\right) + a_\lambda\langle 1+|v|^2, p_n\xi^{(n)}\rangle.\end{aligned} \tag{6.234}$$

By the definitions of $p_n, \xi^{(n)}$, it follows that $p_n\xi^{(n)} = \nu^{(n)}/c_n$; by Lemma 6.20, $\langle 1 + |v|^2, \nu^{(n)}\rangle \to 0$, $c_n \to 1$, so for fixed $\lambda > 0$, the last term converges to 0 as $n \to \infty$. We can therefore choose a sequence $\lambda_n \to 0$ which decays slowly enough that $a_{\lambda_n}\langle 1+|v|^2, \nu^{(n)}\rangle \to 0$. We now define $\mu_\bullet^{(n)} := \mu_\bullet^{(n,\lambda_n)}, w^{(n)} := w^{(n,\lambda_n)}$. Every term except the first on the right-hand side of (6.234) converges to 0, and in particular $\limsup_n \mathcal{I}(\mu_\bullet^{(n)}, w^{(n)}) \le \mathcal{I}(\mu_\bullet, w)$.

Step 4: Conclusion. We now check that the diagonal sequence extracted has all the desired properties. First, thanks to (6.140), the convolution with g_{λ_n} preserves the boundedness, so

$$\sup_{(t,v,v_\star,\sigma)} K^{(n)}(t,v,v_\star,\sigma)B(v-v_*,\sigma) \le \sup_{(t,v,v_\star,\sigma)} K^{(n,0)}(t,v,v_\star,\sigma)B(v-v_*,\sigma) < \infty. \tag{6.235}$$

To see convergence of the overall sequence, note that $\mathcal{W}_{1,1}(\mu, g_\lambda \star \mu) \le C\sqrt{\lambda}$ for all measures μ, and since $\mathcal{W}_{1,1}$ is dominated by the total variation distance,

$$\begin{aligned}\sup_{t\le t_{\rm fin}} \mathcal{W}_{1,1}(\mu_t^{(n)}, \mu_t) &\le \sup_{t\le t_{\rm fin}} \mathcal{W}_{1,1}(g_{\lambda_n} \star \mu_t^{(n,0)}, \mu_t^{(n,0)}) + \sup_{t\le t_{\rm fin}} \left\|\mu_t^{(n,0)} - \mu_t\right\|_{\rm TV} \\ &\le C\sqrt{\lambda_n} + \sup_{t\le t_{\rm fin}} \left\|(1+|v|^2)(\mu_t^{(n,0)} - \mu_t)\right\|_{\rm TV} \to 0.\end{aligned} \tag{6.236}$$

Similarly, $\rho_1(g_{\lambda_n} \star w^{(n,0)}, w^{(n,0)}) \le C\sqrt{\lambda_n}$, so that $\rho_1(w^{(n)}, w) \to 0$. □

Finally, we prove Lemma 6.22, which allows us to impose an asymptotic lower bound on K, so we control how fast $|\log K|$ grows as $v, v_\star \to \infty$.

Proof of Lemma 6.22. Let us consider the space $\mathcal{MS}_2(\mathbb{R}^d)$ of signed measures with finite second moment $\langle 1 + |v|^2, |\xi|\rangle < \infty$, equipped with the complete distance given by the weighted total variation norm $\|\xi\|_{\mathrm{TV}+2} := \|(1+|v|^2)\xi\|_{\mathrm{TV}}$. We start from a measure-flux pair $(\mu_\bullet, w)$ as in the statement, so that the tilting function K is continuous in $v, v_\star$, and $B(v-v_*,\sigma)K$ is bounded; since B is bounded away from 0, this implies the same for K.

Step 1: Construction of K We begin with a family of mollifiers. Let us fix a smooth function $\eta : \mathbb{R} \to [0,\infty)$, supported on $[-1,1]$ and such that $\int \eta ds = 1$, and for $t \in [0, t_{\text{fin}}], \lambda > 0$, define

$$\eta_\lambda(s,t) = \frac{\eta((s-t)/\lambda)}{\int_0^{t_{\text{fin}}} \eta((u-t)/\lambda)du} \tag{6.237}$$

so that η_λ is continuous in both arguments, $\eta_\lambda(\cdot,t)$ is supported on $[0,t_{\text{fin}}]\cap[t-\lambda,t+\lambda]$, and $\int_0^{t_{\text{fin}}} \eta_\lambda(s,t)ds = 1$. For the spherical directions, let $h_\lambda(\cdot,\cdot)$ be the heat kernel on $\mathbb{S}^{d-1}$, so that $h_\lambda(\cdot,\sigma)$ is a smooth function on $\mathbb{S}^{d-1}$ which integrates to 1, and $h_\lambda(\sigma',\sigma)d\sigma' \to \delta_\sigma(d\sigma')$ weakly as $\lambda \to 0$. With these fixed, we define $K^{(n)}$ by

$$K^{(n)}(t,v,v_\star,\sigma) := \int_{[0,t_{\text{fin}}]\times\mathbb{S}^{d-1}} K(s,v,v_\star,\sigma')\eta_{1/n}(s,t)h_{1/n}(\sigma',\sigma)dsd\sigma' + \frac{1}{nB(v-v_*,\sigma)}. \tag{6.238}$$

From the construction, the continuity of K in $v, v_\star$ implies that each $K^{(n)}$ is continuous on E. $K^{(n)}$ also inherit the upper bound: there exists M such that

$$\sup_n \sup_{t,v,v_\star,\sigma} B(v-v_*,\sigma)K^{(n)}(t,v,v_\star,\sigma) \le M; \qquad \sup_{t,v,v_\star,\sigma} B(v-v_*,\sigma)K(t,v,v_\star,\sigma) \le M \tag{6.239}$$

and by construction $\inf_E B(v-v_*,\sigma)K^{(n)} \ge \frac{1}{n} > 0$. Finally, for all $(v,v_\star)$ fixed, $dtd\sigma$ almost everywhere, $K^{(n)}(t,v,v_\star,\sigma) \to K(t,v,v_\star,\sigma)$.

Step 2: Construction of $\mu^{(n)}_\bullet$ by Picard-Lindelöf We now construct processes $\mu^{(n)}_\bullet$, which at this stage may be signed measures, via the machinery of the Picard-Lindelöf theorem. For any $t \in [0,t_{\text{fin}}]$, $\xi \in \mathcal{MS}_2$, define the signed measures

$$\Phi(t,\xi) := \int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}} \Delta(v,v_\star,\sigma)B(v-v_*,\sigma)K(t,v,v_\star,\sigma)\xi(dv)\xi(dv_\star)d\sigma; \tag{6.240}$$

$$\Phi_n(t,\xi) := \int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}} \Delta(v,v_\star,\sigma)B(v-v_*,\sigma)K^{(n)}(t,v,v_\star,\sigma)\xi(dv)\xi(dv_\star)d\sigma. \tag{6.241}$$

Using the uniform boundedness of $B(v-v_*,\sigma)K, B(v-v_*,\sigma)K^{(n)}$, it is easy to see that

$$\|(\Phi_n(t,\xi)-\Phi_n(t,\xi')\|_{\text{TV+2}} \le C\|(\xi-\xi')\|_{\text{TV+2}}(\|\xi\|_{\text{TV+2}}+\|\xi'\|_{\text{TV+2}}) \tag{6.242}$$

for some constant C, uniformly in n, and similarly for Φ. It then follows from the Picard-Lindelöf theorem that, for any $\xi^{(n)}_0, \xi_0$, there exist unique local solutions to the integral equations

$$\xi^{(n)}_t = \xi^{(n)}_0 + \int_0^t \Phi_n(s,\xi^{(n)}_s)ds; \qquad \xi_t = \xi_0 + \int_0^t \Phi(s,\xi_s)ds. \tag{6.243}$$

Further, observing that $\|\Phi_n(t,\xi)\|_{\text{TV}} \le C\|\xi\|_{\text{TV}}$, it follows that $\|\xi^{(n)}_t\|_{\text{TV}}$ grows at most exponentially in time; using the similar estimate that $\|\Phi_n(t,\xi)\|_{\text{TV+2}} \le C\|\xi\|_{\text{TV+2}}\|\xi\|_{\text{TV}}$ by taking $\xi' = 0$ above, the same holds for $\|\xi^{(n)}_t\|_{\text{TV+2}}$, so the solutions are globally defined.

Let us now consider these equations to construct our approximations. For the initial data, we recall that the finiteness of the entropy $H(\mu_0|\mu_0^\star) \le \mathcal{I}(\mu_\bullet, w) < \infty$ implies that μ_0 has a density with respect to $\mu_0^\star$, and we set

$$\mu_0^{(n)}(dv) = c_n \left(\frac{d\mu_0}{d\mu_0^\star} \wedge n \right) \mu_0^\star(dv) \tag{6.244}$$

for a normalising constant $c_n \to 1$ which makes $\mu_0^{(n)}$ a probability measure. We now take $\mu_\bullet^{(n)}$ to be the unique solution $\xi_t^{(n)}$ produced to $\partial_t \xi_t^{(n)} = \Phi_n(t, \xi_t^{(n)})$ for this choice of initial data. It follows by definition of K that the process μ_t given satisfies $\partial_t \mu_t = \Phi(t, \mu_t)$, which must then by the unique solution.

Step 3: Positivity of $\mu_\bullet^{(n)}$. To see that this gives positive measures, we use an integrating factor introduced by Norris [155] in the context of a similar construction for the Smoluchowski equation. We define

$$\theta_t^{(n)}(v) := \exp \left(\int_0^t \int_{\mathbb{R}^d \times \mathbb{S}^{d-1}} (K^{(n)}(s, v, v_\star, \sigma) + K^{(n)}(s, v_\star, v, \sigma)) B(v - v_*, \sigma) \mu_s^{(n)}(dv_\star) d\sigma ds \right) \tag{6.245}$$

and

$$\Phi_n^+(t, \xi) = \int_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{S}^{d-1}} (\theta_t^{(n)}(v') \delta_{v'} + \theta_t^{(n)}(v'_\star) \delta_{v'_\star}) B(v - v_*, \sigma) K^{(n)}(t, v, v_\star, \sigma) \xi(dv) \xi(dv_\star) d\sigma. \tag{6.246}$$

Thanks to the boundedness, $\theta_t^{(n)}$ is bounded and bounded away from 0, uniformly on compact time intervals, and $\Phi_n^+(\xi) \ge 0$ whenever $\xi \ge 0$. These integrating factors are such that $\partial_t(\theta_t^{(n)} \mu_t^{(n)}) = \Phi_n^+(t, \theta_t^{(n)} \mu_t^{(n)})$, while applying the same arguments as above in the smaller space $(\mathcal{P}_2, \|\cdot\|_{\mathrm{TV}+2})$ shows that the unique solution to $\partial_t \nu_t = \Phi_n^+(t, \nu_t)$ remains positive if $\nu_0 \in \mathcal{P}_2$ is a positive measure. It follows that $\theta_t^{(n)} \mu_t^{(n)} \ge 0$ are positive measures, and hence so are $\mu_t^{(n)}$; recalling again the boundedness, energy conservation implies that $\langle |v|^2, \mu_t^{(n)} \rangle = \langle |v|^2, \mu_0^{(n)} \rangle < \infty$ is constant for each n, and we conclude that $\mu_\bullet^{(n)} \in \mathcal{D}$. We define the corresponding flux $w^{(n)}$ by

$$w^{(n)}(dt, dv, dv_\star, d\sigma) := K^{(n)}(t, v, v_\star, \sigma) \overline{w}_{\mu^{(n)}}(dt, dv, dv_\star, d\sigma) \tag{6.247}$$

so that $\mu_\bullet^{(n)}, w^{(n)}$ is a measure-flux pair. Moreover, if $(\mu'_\bullet, w')$ is any measure-flux pair with $\mu'_0 = \mu_0^{(n)}$ and with tilting function $K^{(n)}$, then $\partial_t \mu'_t = \Phi_n(t, \mu'_t)$, which implies that $\mu'_t = \mu_t^{(n)}$ by the uniqueness in step 2, and $w' = K\overline{w}_{\mu'} = K\overline{w}_{\mu^{(n)}} = w$, so each approximating pair $(\mu_\bullet^{(n)}, w^{(n)})$ is uniquely characterised by the initial value and tilting function, as claimed.

Step 4: Convergence of the Approximations. Let us now show that the measure-flux pairs constructed above converge as $n \to \infty$. We start from

$$\begin{aligned}\|\Phi_n(t,\mu_t^{(n)})-\Phi(t,\mu_t)\|_{\mathrm{TV}+2} &\le \|\Phi_n(t,\mu_t^{(n)})-\Phi_n(t,\mu_t)\|_{\mathrm{TV}+2}+\|\Phi_n(t,\mu_t)-\Phi(t,\mu_t)\|_{\mathrm{TV}+2} \\ &\le C\|\mu_t^{(n)}-\mu_t\|_{\mathrm{TV}+2}(\|\mu_t^{(n)}\|_{\mathrm{TV}+2}+\|\mu_t\|_{\mathrm{TV}+2})+\|\Phi_n(t,\mu_t)-\Phi(t,\mu_t)\|_{\mathrm{TV}+2}\end{aligned} \tag{6.248}$$

using (6.242). In the first term, we observe that $\|\mu_t^{(n)}\|_{\mathrm{TV}+2}=\langle 1+|v|^2,\mu_0^{(n)}\rangle$ is bounded uniformly in n,t, thanks to energy conservation and the construction of $\mu_0^{(n)}$, and we absorb this constant factor into C. We can now use Grönwall's Lemma to obtain

$$\sup_{t\le t_{\mathrm{fin}}}\|\mu_t^{(n)}-\mu_t\|_{\mathrm{TV}+2}\le e^{Ct_{\mathrm{fin}}}\left(\|\mu_0^{(n)}-\mu_0\|_{\mathrm{TV}+2}+\int_0^{t_{\mathrm{fin}}}\|\Phi_n(t,\mu_t)-\Phi(t,\mu_t)\|_{\mathrm{TV}+2}dt\right). \tag{6.249}$$

The first term is readily seen to converge to 0 using the construction (6.244) of $\mu_0^{(n)}$, recalling that μ_0 has finite second moment. For the second term, we return to the definition of Φ,Φ_n to see that

$$\begin{aligned}&\|\Phi_n(t,\mu_t)-\Phi(t,\mu_t)\|_{\mathrm{TV}+2}\\ &\le 2\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}}(1+|v|^2+|v_\star|^2)(K^{(n)}-K)(t,v,v_\star,\sigma)B(v-v_*,\sigma)\mu_t(dv)\mu_t(dv_\star)d\sigma\end{aligned} \tag{6.250}$$

and integrating over $t\in[0,t_{\mathrm{fin}}]$ produces

$$\begin{aligned}&\int_0^{t_{\mathrm{fin}}}\|\Phi_n(t,\mu_t)-\Phi(t,\mu_t)\|_{\mathrm{TV}+2}dt\\ &\qquad\le 2\int_E(1+|v|^2+|v_\star|^2)(K^{(n)}-K)B(v-v_*,\sigma)dt\mu_t(dv)\mu_t(dv_\star)d\sigma.\end{aligned} \tag{6.251}$$

We now apply dominated convergence to see that the right-hand side converges to 0, since $B(v-v_*,\sigma)(K^{(n)}-K)$ is bounded by (6.239), and converges to 0 for $dt\mu_t(dv)\mu_t(dv_\star)d\sigma$ almost all $(t,v,v_\star,\sigma)$, while μ_t has constant, finite second moment. Returning to (6.249), we conclude that $\|\mu_t^{(n)}-\mu_t\|_{\mathrm{TV}+2}\to 0$, which is stronger than the required convergence. For the flux, we estimate

$$\begin{aligned}\|w^{(n)}-w\|_{\mathrm{TV}}\le&\int_E|K^{(n)}-K|B(v-v_*,\sigma)dt\mu_t^{(n)}(dv)\mu_t^{(n)}(dv_\star)d\sigma\\ &+\int_E KB(v-v_*,\sigma)dt|\mu_t(dv)\mu_t(dv_\star)-\mu_t^{(n)}(dv)\mu_t^{(n)}(dv_\star)|d\sigma.\end{aligned} \tag{6.252}$$

The first term converges to 0 as above, and recalling (6.239), the second term is bounded by

$$\int_E KB(v-v_*,\sigma)dt|\mu_t(dv)\mu_t(dv_\star)-\mu_t^{(n)}(dv)\mu_t^{(n)}(dv_\star)|d\sigma\le 2C\int_0^{t_{\mathrm{fin}}}\|\mu_t^{(n)}-\mu_t\|_{\mathrm{TV}}dt\to 0 \tag{6.253}$$

and we have proven that $\|w^{(n)}-w\|_{\mathrm{TV}}\to 0$ as desired.

Step 5: Convergence of the Cost Function We finally check the convergence of the rate function $\mathcal{I}$ along our subsequence. First, for the cost of the initial data, the construction of $\mu_0^{(n)}$ gives

$$H(\mu_0^{(n)}|\mu_0^\star) \le \log c_n + H(\mu_0|\mu_0^\star)$$

and $c_n \to 1$ by construction, so we have $\limsup_n H(\mu_0^{(n)}|\mu_0^\star) \le H(\mu_0|\mu_0^\star)$. It therefore suffices, by the usual lower semicontinuity, to prove the same thing for the dynamic cost $\mathcal{J}$. We start by writing

$$\begin{aligned} \mathcal{J}(\mu_\bullet^{(n)}, w^{(n)}) &= \int_E \tau(K^{(n)})\overline{w}_{\mu^{(n)}}(dt, dv, dv_\star, d\sigma) \\ &\le \mathcal{J}(\mu_\bullet, w) + \int_E \tau(K^{(n)})(\overline{w}_{\mu^{(n)}} - \overline{w}_\mu)(dt, dv, dv_\star, d\sigma) \\ &\quad + \int_E (\tau(K^{(n)}) - \tau(K))\overline{w}_\mu(dt, dv, dv_\star, d\sigma). \end{aligned} \tag{6.254}$$

In the second term, $B(v - v_*, \sigma)K^{(n)} \le M$ everywhere, and since $B \ge 1$, this implies the same bound for $K^{(n)}$ and hence the bound $\tau(K^{(n)}) \le 1 + \tau(M)$, uniformly in n. The second term is now at most $(1 + \tau(M))\|w^{(n)} - w\|_{\mathrm{TV}} \to 0$. Similarly, $\tau(K^{(n)}) \to \tau(K)$ converges $dt\mu_t(dv)\mu_t(dv_\star)d\sigma$ almost everywhere, and hence $\overline{w}_\mu$ almost everywhere, with the same uniform bound as above. Since $\overline{w}_\mu(E) < \infty$, we can apply dominated convergence to see that the third term $\to 0$, and we are done. □

6.5 Applications of the Positive Result 6.2

We now prove Corollary 6.3 and Proposition 6.4, which are applications of Theorem 6.2.

6.5.1 Entropy as a Quasipotential

We begin with the proof of Corollary 6.3.

Proof of Corollary 6.3. We apply a contraction principle (see [77, 147]) argument to Theorem 6.2. Since we do not have a 'true' large deviation principle, and must further compensate for the failure of the rate function $\mathcal{I}$ to be good, the arguments do not follow from any statement of the contraction principle we have found in the literature, and we present the arguments in detail.

Let us fix $\mu \in \mathcal{P}_2$ and $\epsilon > 0$ and consider

$$\mathcal{U}_\epsilon := \{(\mu_\bullet, w) \in \mathcal{D} \times \mathcal{M}(E) : \mathcal{W}_{1,1}(\mu_{t_{\text{fin}}}, \mu) < \epsilon\} \tag{6.255}$$

so that the closure is

$$\overline{\mathcal{U}}_\epsilon := \{(\mu_\bullet, w) \in \mathcal{D} \times \mathcal{M}(E) : \mathcal{W}_{1,1}(\mu_{t_{\text{fin}}}, \mu) \le \epsilon\} \tag{6.256}$$

Let us take $\mu_0^\star = \gamma$ and, for each N, sample initial velocities independently from γ. In this case, the N-particle system is in equilibrium, so that the distribution of $\mu^N_{t_{\text{fin}}}$ is that of a N-particle independent sample from γ. In particular, Sanov's theorem [51] applies, so that $\mu^N_{t_{\text{fin}}}$ satisfies a large deviation principle with rate function $H(\cdot|\gamma)$. We first prove the first item (6.18): by Sanov's Theorem

$$\begin{aligned} \liminf_N \frac{1}{N} \log \mathbb{P}\left((\mu^N_\bullet, w^N) \in \mathcal{U}_\epsilon\right) &= \liminf_N \frac{1}{N} \log \mathbb{P}\left(\mathcal{W}_{1,1}(\mu^N_{t_{\text{fin}}}, \mu) < \epsilon\right) \\ &\ge -\inf\{H(\nu|\gamma) : \mathcal{W}_{1,1}(\nu, \mu) < \epsilon\} \\ &\ge -H(\mu|\gamma) \end{aligned} \tag{6.257}$$

while, immediately

$$\liminf_N \frac{1}{N} \log \mathbb{P}\left((\mu^N_\bullet, w^N) \in \mathcal{U}_\epsilon\right) \le \limsup_N \frac{1}{N} \log \mathbb{P}\left((\mu^N_\bullet, w^N) \in \overline{\mathcal{U}}_\epsilon\right). \tag{6.258}$$

If $H(\mu|\gamma) = \infty$ there is, of course, nothing to prove; otherwise, we choose $M > H(\mu|\gamma)$ and using Proposition 6.1i), pick a compact set $\mathcal{K} \subset \mathcal{D} \times \mathcal{M}(E)$ such that

$$\limsup_N N^{-1} \mathbb{P}((\mu^N_\bullet, w^N) \notin \mathcal{K}) \le -M.$$

From (6.258) and applying Theorem 6.2i),

$$\begin{aligned}\limsup_N \frac{1}{N}\log\mathbb{P}\left((\mu^N_\bullet, w^N)\in\overline{\mathcal{U}}_\epsilon\right) &\le \max\left(\limsup_N \frac{1}{N}\log\mathbb{P}\left((\mu^N_\bullet, w^N)\in\overline{\mathcal{U}}_\epsilon\cap\mathcal{K}\right), \limsup_N \frac{1}{N}\log\mathbb{P}\left((\mu^N_\bullet, w^N)\notin\mathcal{K}\right)\right) \\ &\le \max\left(-\inf\left\{\mathcal{I}(\nu_\bullet, w) : (\nu_\bullet, w)\in\overline{\mathcal{U}}_\epsilon\cap\mathcal{K})\right\}, -M\right).\end{aligned} \tag{6.259}$$

Comparing against (6.257), we must have that

$$H(\mu|\gamma)\ge\min\left(\inf\{\mathcal{I}(\nu_\bullet, w) : (\nu_\bullet, w)\in\overline{\mathcal{U}}_\epsilon\cap\mathcal{K}\}, M\right) \tag{6.260}$$

and since $M > H(\mu|\gamma)$ by construction,

$$H(\mu|\gamma)\ge\inf\{\mathcal{I}(\nu_\bullet, w) : (\nu_\bullet, w)\in\overline{\mathcal{U}}_\epsilon\cap\mathcal{K}\} \tag{6.261}$$

We claim that the right-hand side converges as $\epsilon\downarrow 0$:

$$\inf\left\{\mathcal{I}(\nu_\bullet, w) : (\nu_\bullet, w)\in\overline{\mathcal{U}}_\epsilon\cap\mathcal{K})\right\}\to\inf\{\mathcal{I}(\nu_\bullet, w) : \nu_{t_{\text{fin}}}=\mu, (\nu_\bullet, w)\in\mathcal{K}\}. \tag{6.262}$$

It is immediate that the left-hand side is increasing as $\epsilon\downarrow 0$ and that the right-hand side is an upper bound; it is therefore sufficient to prove convergence on a subsequence. For each n, pick $(\nu^{(n)}_\bullet, w^{(n)})\in\overline{\mathcal{U}}_{1/n}\cap\mathcal{K}$ with error at most $1/n$ from the infimum. Since $\mathcal{K}$ is compact, we can pass to a subsequence with $(\nu^{(n)}_\bullet, w^{(n)})\to(\nu_\bullet, w)$; the limit has $\nu_{t_{\text{fin}}}=\mu$ and $(\nu_\bullet, w)\in\mathcal{K}$. By lower-semicontinuity from Proposition 6.1ii), we have

$$\mathcal{I}(\nu_\bullet, w)\le\liminf_n\mathcal{I}(\nu^{(n)}_\bullet, w^{(n)})\le\liminf_n\left(\inf\left\{\mathcal{I}(\nu_\bullet, w) : (\nu_\bullet, w)\in\overline{\mathcal{U}}_{1/n}\cap\mathcal{K})\right\}+\frac{1}{n}\right) \tag{6.263}$$

so that

$$\inf\{\mathcal{I}(\nu_\bullet, w) : \nu_{t_{\text{fin}}}=\mu, (\nu_\bullet, w)\in\mathcal{K}\}\le\liminf_n\left(\inf\left\{\mathcal{I}(\nu_\bullet, w) : (\nu_\bullet, w)\in\overline{\mathcal{U}}_{1/n}\cap\mathcal{K})\right\}\right) \tag{6.264}$$

which proves the claim (6.262). Returning to (6.261), we take $\epsilon\to 0$ to find

$$\begin{aligned}H(\mu|\gamma) &\ge\inf\left\{\mathcal{I}(\mu_\bullet, w) : \nu_{t_{\text{fin}}}=\mu, (\nu_\bullet, w)\in\mathcal{K}\right\} \\ &\ge\inf\left\{\mathcal{I}(\mu_\bullet, w) : \nu_{t_{\text{fin}}}=\mu\right\}\end{aligned} \tag{6.265}$$

and observe that the right-hand side is exactly the claimed bound in (6.18). For the second item (6.19), we apply the lower bound of Sanov:

$$\begin{aligned}\limsup_N \frac{1}{N}\log\mathbb{P}\left((\mu^N_\bullet, w^N)\in\overline{\mathcal{U}}_\epsilon\right) &= \limsup_N \frac{1}{N}\log\mathbb{P}\left(\mathcal{W}_{1,1}(\mu^N_{t_{\text{fin}}}, \mu)\le\epsilon\right) \\ &\le -\inf\{H(\nu|\gamma) : \mathcal{W}_{1,1}(\nu, \mu)\le\epsilon\}.\end{aligned} \tag{6.266}$$

On the other hand,

$$\begin{aligned}\limsup_N \frac{1}{N}\log \mathbb{P}\left((\mu^N_\bullet, w^N)\in \overline{\mathcal{U}}_\epsilon\right) &\geq \liminf_N \frac{1}{N}\log \mathbb{P}\left((\mu^N_\bullet, w^N)\in \mathcal{U}_\epsilon\right) \\ &\geq -\inf\left\{\mathcal{I}(\nu_\bullet, w):(\nu_\bullet, w)\in \mathcal{U}_\epsilon\cap\mathcal{R}\right\} \\ &\geq -\inf\left\{\mathcal{I}(\nu_\bullet, w): \nu_{t_{\mathrm{fin}}}=\mu, (\nu_\bullet, w)\in\mathcal{R}\right\}.\end{aligned} \tag{6.267}$$

We conclude that

$$\inf\{H(\nu|\gamma):\mathcal{W}_{1,1}(\nu,\mu)\leq\epsilon\}\leq \inf\left\{\mathcal{I}(\nu_\bullet, w): \nu_{t_{\mathrm{fin}}}=\mu, (\nu_\bullet, w)\in\mathcal{R}\right\}. \tag{6.268}$$

As $\epsilon\to 0$, the left-hand side converges to $H(\mu|\gamma)$ by the lower semi-continuity of entropy (cf. Lemma 6.13), and the right-hand side is exactly the right-hand side of (6.19), so we are done. □

6.5.2 Time Reversal

We now give the proof of Proposition 6.4, based on the time-reversibility of the Kac process. Throughout, we work with $\mu_0^\star=\gamma$, and recall the definitions of the time reversals from (6.20, 6.21), given by

$$\mathbb{T}\mu_\bullet := \left(\mu_{(t_{\mathrm{fin}}-t)-}\right)_{0\leq t\leq t_{\mathrm{fin}}}$$

and by specifying, for all bounded, measurable $g: E\to\mathbb{R}$,

$$\langle g, \mathbb{T}w\rangle = \int_E g(t_{\mathrm{fin}}-t, v', v'_*, \sigma)w(dt,dv,dv_*,d\sigma)$$

where v', v'_* are understood as functions of (v, v_*, σ) through the representation (6.1).

Proof of Proposition 6.4. We deal with the assertions one by one.

i). For the first assertion, we start from the well-known fact that the law of $\mathbb{T}\mu^N_\bullet$ on $\mathcal{D}$ is the same as that of $\mu_\bullet$; for instance, this follows from the fact that the (labelled) generator $\mathcal{G}^{\mathrm{L}}$ (1.12) is self-adjoint in the space $L^2((\mathbb{R}^d)^N, \gamma^{\otimes N})$, which implies reversibility in equilibrium. To deduce the same for the pair $(\mathbb{T}\mu^N_\bullet, \mathbb{T}w^N)$ we observe that, conditional on $\mu^N_\bullet$, w^N places a point of mass N^{-1}, selecting one of the four possible parameter choices for the incoming velocities and deflection angle (v, v_*, σ) uniformly at random. In the time reversal, $\mathbb{T}w^N$ places the same mass at the corresponding *outgoing* velocities, which are the incoming velocities for the corresponding change in the time-reversal $\mathbb{T}\mu^N_\bullet$, and again selecting one possibility at random. This is exactly the same as the law of w^N conditional on $\mu^N_\bullet$, so we are done.

ii). For the second point, we apply Theorem 6.2 together with the previous point. Let us note first that $\mathbb{T}$ is self-inverse, preserves the topology of $\mathcal{D}\times\mathcal{M}(E)$, and takes $\mathcal{R}$ to itself. Let us assume that $(\mu_\bullet, w)\in\mathcal{R}$, and that $\mathcal{I}(\mu_\bullet, w)<\infty$. We apply lower-semicontinuity, and recall that the topology is induced by a metric, to find an open neighbourhood $\mathcal{U}$ such that $\mathcal{I}\ge \mathcal{I}(\mu_\bullet,w)-\varepsilon$ on the closure $\overline{\mathcal{U}}$. Now

$$\frac{1}{N}\log\mathbb{P}\left((\mu^N_\bullet, w^N)\in\mathcal{U}\right)\le \frac{1}{N}\log\mathbb{P}\left((\mu^N_\bullet, w^N)\in\overline{\mathcal{U}}\right) \tag{6.269}$$

while using reversibility

$$\begin{aligned}\frac{1}{N}\log\mathbb{P}\left((\mu^N_\bullet, w)\in\mathcal{U}\right)&=\frac{1}{N}\log\mathbb{P}\left((\mathbb{T}\mu^N_\bullet, \mathbb{T}w^N)\in\mathcal{U}\right)\\ &=\frac{1}{N}\log\mathbb{P}\left((\mu^N_\bullet, w^N)\in\mathbb{T}^{-1}\mathcal{U}\right).\end{aligned} \tag{6.270}$$

The preimage $\mathbb{T}^{-1}\mathcal{U}$ is an open neighbourhood of $\mathbb{T}^{-1}(\mu_\bullet,w)=(\mathbb{T}\mu_\bullet,\mathbb{T}w)$, and so the limit inferior of the left-hand side of (6.269) is at least $-\mathcal{I}(\mathbb{T}\mu_\bullet,\mathbb{T}w)$, using Theorem 6.2ii). On the other hand, the limit superior of the right-hand side is at most $-\mathcal{I}(\mu_\bullet,w)+\varepsilon$, by the choice of $\mathcal{U}$, and we conclude that

$$\mathcal{I}(\mathbb{T}\mu_\bullet,\mathcal{T}w)\ge\mathcal{I}(\mu_\bullet,w)-\varepsilon$$

and since $\varepsilon>0$ was arbitrary,

$$\mathcal{I}(\mathbb{T}\mu_\bullet,\mathcal{T}w)\ge\mathcal{I}(\mu_\bullet,w). \tag{6.271}$$

A similar argument holds if $\mathcal{I}(\mu_\bullet,w)=\infty$, now letting $M>0$ be arbitrarily large and choosing $\mathcal{U}$ so that $\mathcal{I}\ge M$ on $\overline{\mathcal{U}}$, so that (6.271) holds for any $(\mu_\bullet,w)\in\mathcal{R}$. Finally, $\mathbb{T}$ takes $\mathcal{R}$ into itself and is self-inverse, so we conclude the reverse inequality by applying the same thing to $(\mathbb{T}\mu_\bullet,\mathbb{T}w)\in\mathcal{R}$.

iii). The third item now identifies the tilting function K when we reverse a measure-flux pair such that μ_t admits a strictly positive density f_t. First, it is straightforward to check that if $(\mu_\bullet,w)$ satisfies the continuity equation, so does $(\mathbb{T}\mu_\bullet,\mathbb{T}w)$, for instance, using the variational formulation in Lemma 6.10. We let K be the tilting function for $(\mu_\bullet,w)$, write $\mathbb{T}K$ for the putative tilting function for the reversed pair given in the statement. Fixing $g:E\to\mathbb{R}$ a bounded and measureable function, we write

$$\langle g,\mathbb{T}w\rangle=\int_E g(t_{\text{fin}}-t,v',v'_*,\sigma)K(t,v,v_*,\sigma)B(v-v_*,\sigma)f_t(v)f_t(v_*)dtdvdv_*d\sigma \tag{6.272}$$

using the definitions of $\mathbb{T}w, K$ and f_t. We next make the change of variables $(t,v,v_*,\sigma)\to(t_{\text{fin}}-t,v',v'_*,\sigma)$, which is self-inverse and has unit Jacobian (as in Lemma 6.19). Therefore, the previous expression can be rewritten

$$\begin{aligned}\langle g,\mathbb{T}w\rangle&=\int_E g(t,v,v_*,\sigma)K(t_{\text{fin}}-t,v\,v_*\,\sigma)B(v'-v'_*,\sigma)f_{t_{\text{fin}}-t}(v')f_{t_{\text{fin}}-t}(v'_*)dtdvdv_*d\sigma\\ &=\int_E g(t,v,v_*,\sigma)K(t_{\text{fin}}-t,v',v'_*,\sigma)B(v-v_*,\sigma)f_{t_{\text{fin}}-t}(v')f_{t_{\text{fin}}-t}(v'_*)dtdvdv_*d\sigma.\end{aligned} \tag{6.273}$$

Since $f_t > 0$ everywhere, we insert

$$f_{t_{\text{fin}}-t}(v')f_{t_{\text{fin}}-t}(v'_*) = \left(\frac{f_{t_{\text{fin}}-t}(v')f_{t_{\text{fin}}-t}(v'_*)}{f_{(t_{\text{fin}}-t)-}(v)f_{(t_{\text{fin}}-t)-}(v_*)}\right) f_{(t_{\text{fin}}-t)-}(v)f_{(t_{\text{fin}}-t)-}(v_*) \quad (6.274)$$

and recall the definition

$$\mathbb{T}K := \left(\frac{f_{t_{\text{fin}}-t}(v')f_{t_{\text{fin}}-t}(v'_*)}{f_{(t_{\text{fin}}-t)-}(v)f_{(t_{\text{fin}}-t)-}(v_*)}\right) K(t_{\text{fin}} - t, v', v'_*, \sigma) \quad (6.275)$$

to write

$$\langle g, \mathbb{T}w\rangle = \int_E g(t, v, v_*, \sigma)(\mathbb{T}K)(t, v, v_*, \sigma)B(v-v_*, \sigma)f_{(t_{\text{fin}}-t)-}(v)dv\ f_{(t_{\text{fin}}-t)-}(v_*)dv_*dtd\sigma. \quad (6.276)$$

Since $f_{(t_{\text{fin}}-t)-}$ is the density of $(\mathbb{T}\mu_\bullet)_t$, the last factors are exactly the definition of $\overline{w}_{\mathbb{T}\mu}$, and since g is arbitrary, we conclude that

$$\mathbb{T}w = (\mathbb{T}K)\overline{w}_{\mathbb{T}\mu} \quad (6.277)$$

for the claimed function $\mathbb{T}K$, and the claim is proven.

iv). For the third point additionally assume that $(\mu_\bullet, \overline{w}_\mu) \in \mathcal{R}$ is an energy-conserving solution to (BE), together with its associated flux. It is immediate that $\mu_t = \mu_{t-}$ for all t and $K = 1$, so we can omit the $-$ in the time index to find the tilting function is

$$\mathbb{T}K(t, v, v', \sigma) = \frac{f_{t_{\text{fin}}-t}(v')f_{t_{\text{fin}}-t}(v'_*)}{f_{t_{\text{fin}}-t}(v)f_{t_{\text{fin}}-t}(v_*)}. \quad (6.278)$$

Let us now compute the dynamic cost. We consider first

$$\begin{aligned}\int_E (\mathbb{T}K)\log(\mathbb{T}K)\overline{w}_{\mathbb{T}\mu}(dt, dv, dv_*, d\sigma) \\ = \int_E f_t(v')f_t(v'_*)\log\left(\frac{f_t(v')f_t(v'_*)}{f_t(v)f_t(v_*)}\right) B(v - v_*, \sigma)dtdvdv_*d\sigma\end{aligned} \quad (6.279)$$

by making the change of variables $t \to t_{\text{fin}}-t$ and cancelling the factors of $f_t(v)f_t(v_*)$. Using again the change of variables between $(v, v_*) \to (v', v'_*)$ for each fixed σ and that $B(v' - v'_*, \sigma) = B(v - v_*, \sigma)$ this is exactly

$$\int_E f_t(v)f_t(v_*)\log\left(\frac{f_t(v)f_t(v_*)}{f_t(v')f_t(v'_*)}\right) B(v - v_*, \sigma)dtdvdv_*d\sigma = \int_0^{t_{\text{fin}}} D(f_t)dt. \quad (6.280)$$

For the other terms, the same argument shows that

$$\begin{aligned}\int_E (\mathbb{T}K)\overline{w}_{\mathbb{T}\mu}(dt, dv, dv_*, d\sigma) &= \int_E f_t(v')f_t(v'_*)B(v - v_*, \sigma)dtdvdv_*d\sigma \\ &= \int_E f_t(v)f_t(v_*)B(v - v_*, \sigma)dtdvdv_*d\sigma \\ &= \int_E \overline{w}_{\mathbb{T}\mu}(dt, dv, dv_*, d\sigma)\end{aligned} \quad (6.281)$$

where we used again, in the third line, the change of variables $(v, v_*) \to (v', v'_*)$ and that $B(v' - v'_*, \sigma) = B(v - v_*, \sigma)$. The contributions from the second and third terms in $\tau(x) = x \log x - x + 1$ therefore cancel, and we are left with

$$\begin{aligned} \mathcal{J}(\mathbb{T}\mu_\bullet, \mathbb{T}\overline{w}_\mu) &= \int_E f_t(v) f_t(v_*) \log \left(\frac{f_t(v) f_t(v_*)}{f_t(v') f_t(v'_*)} \right) B(v - v_*, \sigma) dt dv dv_* d\sigma \\ &= \int_0^{t_{\text{fin}}} D(f_t) dt \end{aligned} \tag{6.282}$$

as claimed. Finally, using item ii) and the hypothesis that $(\mu_\bullet, \overline{w}_\mu) \in \mathcal{R}$, the previous item gives $\mathcal{I}(\mu_\bullet, \overline{w}_\mu) = \mathcal{I}(\mathbb{T}\mu_\bullet, \mathbb{T}\overline{w}_\mu)$, which we now express

$$H(\mu_0|\gamma) = H(\mu_{t_{\text{fin}}}|\gamma) + \int_0^{t_{\text{fin}}} D(f_t) dt \tag{6.283}$$

as desired.

□

Let us mention that we will use this same time-reversal principle to derive Corollary 6.6 from Theorem 6.5 in Section 6.7.

6.6 Proof of Theorem 6.5

We now turn to the proof of the main counterexample Theorem 6.5. Let us fix, throughout, Θ and P as in the theorem. We first present the proofs in detail in the case of the regularised hard sphere potential $B = 1 + |v|$: we will first carefully construct a change of measures $\mathbb{Q}^N$ using the general form in Proposition 6.15. We then prove a law of large numbers for the modified measures in Lemmas 6.24, showing that any subsequential limit in distribution under the new measures almost surely lands in $\mathcal{A}_\Theta$; the proof is further broken down into Lemmas 6.25 - 6.27, and we finally show how this implies the stated conclusion. We will discuss at the end the necessary modifications for the Maxwell Molecule case $B = 1$.

6.6.1 Construction of a change of measure $\mathbb{Q}^N$.

Throughout, let us fix $(\Omega, \mathfrak{F}, (\mathfrak{F}_t)_{t\geq 0}, \mathbb{P})$ on which are defined regularised hard sphere Kac processes $\mu^N_\bullet$ and their empirical fluxes w^N. We now use the Girsanov formula recalled in Proposition 6.15, and construct the tilting φ of the initial data and dynamic modification of K of the dynamics separately.

Step 1. Construction of Initial Data. Let us consider the random variables $X_M, M \geq 0$, which describe the initial localisation of the energy in the initial data:

$$X_M = \left\langle |v|^2 1\!\!1_{|v|\geq M}, \mu_0^N \right\rangle. \tag{6.284}$$

Since the particles are sampled independently from $\mu_0^\star$, we can write X_M as the mean of N independent variables, which each have the distribution $Y_M = |V|^2 1\!\!1_{|V|\geq M}; V \sim \mu_0^\star$. We write ψ_M for the cumulant generating function for Y_M, and $\psi^\star_M$ for the associated Legendre transform:

$$\psi_M(\lambda) = \log \mathbb{E}\left[e^{\lambda Y_M}\right]; \qquad \psi^\star_M(a) = \sup\left\{a\lambda - \psi_M(\lambda)\right\}.$$

By Hypothesis 6.1ii), it follows that $\psi_M(\lambda) = \infty$ for all $\lambda \geq z_2$ and all M, which implies that $\psi^\star_M(a) \leq az_2$, uniformly in M.

For $M > 0$ and $\lambda \in [0, z_2)$ to be chosen later, we will take $\varphi_{M,\lambda}$ to be the function

$$\varphi_{M,\lambda}(v) = \lambda|v|^2 1\!\!1_{|v|\geq M} - \psi_M(\lambda) \tag{6.285}$$

so that, under any change of measure of the form (6.106) for this choice of φ, each initial velocity V_0^i is distributed independently with law

$$\mu^\star_{0,\lambda,M}(dv) = \exp\left(\lambda|v|^2 1\!\!1_{|v|\geq M} - \psi_M(\lambda)\right)\mu_0^\star(dv). \tag{6.286}$$

Step 2: Choice of λ We now choose λ as a function of M. For fixed M, it is standard to check that

$$\mathcal{E}_M(\lambda,\mu_0^\star)=\int_{\mathbb{R}^d}|v|^2\exp\left(\lambda|v|^2\mathbb{1}_{|v|\geq M}-\psi_M(\lambda)\right)\mu_0^\star(dv)=\langle|v|^2,\mu^\star_{0,\lambda,M}\rangle$$

is continuous on $[0,z_2)$, $\mathcal{E}_M(0,\mu_0^\star)=1$, and observe that $\mathcal{E}_M(\lambda,\mu_0^\star)$ diverges to infinity as $\lambda\uparrow z_2$ thanks to Hypothesis 6.1ii). In particular, we can choose $\lambda=\lambda(M)\in(0,z_2)$ such that

$$\mathcal{E}_M(\lambda_M,\mu_0^\star)=\langle|v|^2,\mu^\star_{0,\lambda_M,M}\rangle=\Theta(t_{\text{fin}})\in(1,\infty).$$

With this choice of λ, we write $\varphi_M=\varphi_{M,\lambda_M}$ and $\mu^\star_{0,M}$ for $\mu^\star_{0,\lambda_M,M}$.

Step 3: Choice of K. We next choose the dynamic tilting function $K=K^{M,r,N}$, depending on the same parameter M and an additional parameter r, to be chosen later. Given $r\in\mathbb{N}$, let $0=t_0^{(r)}\leq t_1^{(r)}\leq ...\leq t_r^{(r)}=t_{\text{fin}}$ be the partition given by

$$t_i^{(r)}=\inf\left\{t\in[0,t_{\text{fin}}]:\Theta(t)\geq\left(1-\frac{i}{r}\right)\Theta(0)+\frac{i}{r}\Theta(t_{\text{fin}})\right\}\in P. \tag{6.287}$$

By Hypothesis 6.1iii), $\mu^\star_{0,M}$ has a density, and in particular the function $r\mapsto\langle|v|^2\mathbb{1}_{|v|\leq r},\mu^\star_{0,M}\rangle$ is continuous. We can therefore choose $M_0\leq M_1\leq M_2\leq M_{r-1}\leq M_r=\infty$ such that, for all $i=0,1,..r-1$,

$$\left\langle|v|^2\mathbb{1}_{|v|\leq M_i},\mu^\star_{0,M}\right\rangle=\Theta(t_i+) \tag{6.288}$$

and observe that $M_0>M$. We now construct a tilting function $K=K^{M,r,N}$ by setting, for $t_{i-1}^{(r)}\leq t<t_i^{(r)}$,

$$K^{M,r,N}(\mu_0^N,t,v,v_\star,\sigma)=\begin{cases}0 & \text{if either } v,v_\star\in\text{Supp}(\mu_0^N)\cap\{|v|\geq M_{i-1}\}=S_t;\\ N\mathbb{1}_{N_t\geq 1}/N_t & \text{else}\end{cases} \tag{6.289}$$

where $N_t=N_t(M,r)$ is the number of particles not in the special set, which is constant on $[t_{i-1}^{(r)},t_i^{(r)})$:

$$N_t=N-N\langle\mathbb{1}_{|v|\geq M_{i-1}},\mu_0^N\rangle=N\langle\mathbb{1}_{|v|<M_i},\mu_0^N\rangle=N-\#S_t. \tag{6.290}$$

Throughout, we will suppress the dependence of $K^{M,r,N}$ on the initial data μ_0^N. In this way, particles with initial velocity $|v|\in[M_{i-1},M_i)$ are 'frozen' until time $t_i^{(r)}$. Moreover, since the special set S_t is finite, almost surely, no particles ever enter S_t, and so under the new measures, all particles whose initial velocity is $\leq M_{i-1}$ interact as a Kac process on N_t particles on $[t_{i-1}^{(r)},t_i^{(r)})$. Let us also remark that $K^{M,r,N}$ satisfies the hypotheses of Proposition 6.15, since N_t depends only on μ_0^N, with the uniform bound $K^{M,r,N}\leq N$.

With this choice of K and $\varphi=\varphi_M$ as in steps 1-2, we now take $\mathbb{Q}^N_{M,r}$ to be the change of measure given by Proposition 6.15.

Step 4: Choice of M, r. By the law of large numbers, as $N \to \infty$ with M fixed,

$$\mathbb{Q}^N_{M,r}\left(\mathcal{W}_{1,1}(\mu^N_0, \mu^\star_{0,M}) > \delta\right) \to 0 \tag{6.291}$$

for any $\delta > 0$ and, with M, r fixed, for all $\delta > 0$,

$$\mathbb{Q}^N_{M,r}\left(\text{For all } i = 0, ..., r, \left|\langle |v|^2 \mathbb{1}[|v| \leq M_i], \mu^N_0\rangle - \Theta(t^{(r)}_i+)\right| < \delta\right) \to 1; \tag{6.292}$$

$$\mathbb{Q}^N_{M,r}\left(\left|\frac{N_0(M,r)}{N} - \langle \mathbb{1}_{|v|\leq M_0}, \mu^\star_{0,M}\rangle\right| < \delta\right) \to 1. \tag{6.293}$$

Now, we compare the equations

$$e^{\psi_M(\lambda_M)} = \int_{\mathbb{R}^d} e^{\lambda_M |v|^2 \mathbb{1}_{v|\geq M}} \mu^\star_0(dv); \tag{6.294}$$

$$\Theta(t_{\text{fin}}) e^{\psi_M(\lambda_M)} = \int_{\mathbb{R}^d} e^{\lambda_M |v|^2 \mathbb{1}_{|v|\geq M}} |v|^2 \mu^\star_0(dv) \tag{6.295}$$

to obtain

$$\begin{aligned} e^{\psi_M(\lambda_M)} &= \int_{|v|<M} \mu^\star_0(dv) + \int_{|v|\geq M} e^{\lambda_M |v|^2} \mu^\star_0(dv) \\ &\leq 1 + \frac{1}{M^2} \int_{\mathbb{R}^d} |v|^2 e^{\lambda_M |v|^2 \mathbb{1}[|v|\geq M]} \mu^\star_0(dv) \end{aligned} \tag{6.296}$$

which implies that

$$e^{\psi_M(\lambda_M)} \leq 1 + \frac{\Theta(t_{\text{fin}})}{M^2} e^{\psi_M(\lambda_M)} \tag{6.297}$$

and hence $\psi_M(\lambda_M) \to 0$ as $M \to \infty$. This implies the convergence of $\mu^\star_{0,M}$ to $\mu^\star_0$: for any f with $|f| \leq 1$, we estimate

$$\begin{aligned} \left|\langle f, \mu^\star_0 - \mu^\star_{0,M}\rangle\right| &\leq \int_{|v|<M} |f(v)| \left|1 - e^{-\psi_M(\lambda_M)}\right| \mu^\star_0(dv) + \frac{\langle |v|^2 |f|, \mu^\star_0 + \mu^\star_{0,M}\rangle}{M^2} \\ &\leq \left|1 - e^{-\psi_M(\lambda_M)}\right| + \frac{\Theta(t_{\text{fin}}) + 1}{M^2} \to 0 \end{aligned}$$

and, since f was arbitrary, the right-hand side is a bound for $\|\mu^\star_0 - \mu^\star_{0,M}\|_{\text{TV}} \geq \mathcal{W}_{1,1}(\mu^\star_0, \mu^\star_{0,M})$. Similarly, we observe that

$$\langle \mathbb{1}[|v| \leq M_0], \mu^\star_{0,M}\rangle \geq 1 - \frac{\langle |v|^2, \mu^\star_{0,M}\rangle}{M_0^2} \geq 1 - \frac{\Theta(t_{\text{fin}})}{M^2} \tag{6.298}$$

using that $M_0 \geq M$. Combining everything, and using a diagonal argument, we can construct a sequence $M_N \to \infty, r_N \to \infty$ slowly enough that, for all $\delta > 0$,

$$\mathbb{Q}^N_{M_N,r_N}\left(\max_{i\leq r_N} \left|\langle |v|^2 \mathbb{1}[|v| \leq M_{N,i}], \mu^N_0\rangle - \Theta(t^{(r)}_i+)\right| < \delta\right) \to 1 \tag{6.299}$$

$$\mathbb{Q}^N_{M_N,r_N}\left(\inf_{t\in[0,t_{\text{fin}}]} \frac{N_t(M_N, r_N)}{N} < 1 - \delta\right) \to 0 \tag{6.300}$$

$$\mathbb{Q}^N_{M_N,r_N}(\mathcal{W}_{1,1}(\mu^N_0, \mu^\star_{0,M_N}) > \delta) \to 0 \tag{6.301}$$

where $M_{N,i}$ are the values constructed in Step 3 associated to M_N. We now set $\widetilde{\mathbb{Q}}^N := \mathbb{Q}^N_{M_N,r_N}$ and define $\mathbb{Q}^N$ by conditioning

$$\mathbb{Q}^N(A) := c_N^{-1}\widetilde{\mathbb{Q}}^N\left(A\cap\left\{\langle|v|^2,\mu_0^N\rangle \le 2\Theta(t_{\mathrm{fin}}), \frac{N_0}{N}\ge\frac{1}{2}\right\}\right)$$

where

$$c_N = \widetilde{\mathbb{Q}}^N\left(\left\{\langle|v|^2,\mu_0^N\rangle \le 2\Theta(t_{\mathrm{fin}}), \frac{N_0}{N}\ge\frac{1}{2}\right\}\right)\to 1$$

are the appropriate normalising constants. We write throughout K^N for K^{N,M_N,r_N}, and we remark that, since $\mathbb{Q}^N$ is the conditioning of $\widetilde{\mathbb{Q}}^N$ to events of high $\widetilde{\mathbb{Q}}^N$-probability, the same convergences (6.299-6.301) hold with $\mathbb{Q}^N$ in place of $\widetilde{\mathbb{Q}}^N = \mathbb{Q}^N_{M_N,r_N}$. By Proposition 6.15, under the measures $\mathbb{Q}^N$, the particles are initially sampled independently from $\mu^\star_{0,M_N}$, conditional on $\frac{N_0}{N}\ge\frac{1}{2}$ and $\langle|v|^2,\mu_0^N\rangle\le 2\Theta(t_{\mathrm{fin}})$, and the dynamics are then governed by the inhomogeneous generator (6.107). We begin with the following preparatory lemma.

Lemma 6.23 (Estimate on the Radon-Nidoykm Derivative). *For the changes of measure $\mathbb{Q}^N \ll \mathbb{P}$ constructed above, and for all $\epsilon>0$,*

$$\mathbb{Q}^N\left(\frac{1}{N}\log\frac{d\mathbb{Q}^N}{d\mathbb{P}} > z_2\Theta(t_{\mathrm{fin}})+\epsilon\right)\to 0. \tag{6.302}$$

Proof. By definition, the change of measure is

$$\begin{aligned}\frac{1}{N}\log\frac{d\mathbb{Q}^N}{d\mathbb{P}} \le \langle\varphi_{M_N},\mu_0^N\rangle + \langle\log K^N, w^N\rangle - \int_E (K^N-1)(t,v,v_\star,\sigma)\overline{w}_{\mu^N}(dt,dv,dv_\star,d\sigma)\\ -\log\widetilde{\mathbb{Q}}^N\left(\langle|v|^2,\mu_0^N\rangle\le 2\Theta(t_{\mathrm{fin}}), \frac{N_0}{N}\ge\frac{1}{2}\right).\end{aligned} \tag{6.303}$$

The final term immediately converges to 0, thanks to (6.299). For the first term, recall that $\phi_{M_N} = \phi_{M_N,\lambda_{M_N}} \le \lambda_{M_N}|v|^2$, and that $\lambda_{M_N}\le z_2$ is bounded, uniformly in N, and (6.299) gives

$$\mathbb{Q}^N\left(\langle\varphi_{M_N},\mu_0^N\rangle > z_2\Theta(t_{\mathrm{fin}})+\epsilon/3\right)\le\mathbb{Q}^N\left(\langle|v|^2,\mu_0^N\rangle>\Theta(t_{\mathrm{fin}})+\epsilon/3z_2\right)\to 0. \tag{6.304}$$

In the second term, observe that $\log K^{N,M_N,r_N}\le\log N/N_0(M_N,r_N)$ on the support of w^N, $\mathbb{Q}^N$-almost surely, since by definition of w^N, no points in the support of w^N have either $v,v_\star$ belonging to the special set S_t. Thanks the conditioning in the definition of $\mathbb{Q}^N$, we have, $\mathbb{Q}^N$-almost surely,

$$\sup K^N\le 2, \qquad \langle|v|^2,\mu_0^N\rangle\le 2\Theta(t_{\mathrm{fin}})$$

and the same arguments as in Section 6.2 show that we can bound $Nw^N(E_t)$ by a $\mathbb{Q}^N$-Poisson process of rate NC, for some constant C. All together, there exists a new constant C, depending only on $\Theta(t_{\mathrm{fin}})$, on such that

$$\mathbb{Q}^N\left(w^N(E)>C\right)\to 0. \tag{6.305}$$

Using (6.301) again,

$$\mathbb{Q}^N\left(\log N/N_0(M_N,r_N)>\epsilon/3C\right)\to 0 \tag{6.306}$$

and, together with (6.305),

$$\mathbb{Q}^N\left(\langle \log K^N, w^N\rangle > \epsilon/3\right)\to 0. \tag{6.307}$$

For the final term, we will find an upper bound for $\int_E |K^{N,M_N,r_N}-1|d\overline{w}_{\mu^N}$. We split the integral into cases where neither $v, v_\star\in S_t$ and its complement. In the first case $(v,v_\star,\sigma)\in S_t^c\times S_t^c\times\mathbb{S}^{d-1}$, we have

$$|1-K^N(t,v,v_\star,\sigma)|\le \frac{N}{N_t(M_N,r_N)}-1.$$

On the other hand, observing that $S_t\subset\{|v|\ge M_N\}$, the contributions from $v\notin S_t$ and $v_\star\notin S_t$ can be controlled by straightforward Markov inequalities: for some constant C,

$$\int_{\mathbb{R}^d\times\mathbb{R}^d}(1+|v|+|v_\star|)(\mathbb{1}_{v\in S_t}+\mathbb{1}_{v_\star\in S_t})\mu^N_t(dv)\mu^N_t(dv_\star)\le CM_N^{-1}\langle 1+|v|^2,\mu_0^N\rangle^2. \tag{6.308}$$

Together we obtain

$$\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}}\left|1-K^N(t,v,v_\star,\sigma)\right|B(v-v_*,\sigma)\mu_t^N(dv)\mu_t^N(dv_\star)d\sigma$$
$$\le C\left(\left(\frac{N}{N_t(M_N,r_N)}-1\right)+\frac{1}{M_N}\right)\langle 1+|v|^2,\mu_t^N\rangle^2. \tag{6.309}$$

We recall that $M_N\to\infty$, use (6.300) to bound the first factor, and return to the definition of $\mathbb{Q}^N$ to bound the moment factor by $(1+\Theta(t_{\text{fin}}))^2$, $\mathbb{Q}^N$-almost surely. It follows that that

$$\mathbb{Q}^N\left(\int_E|1-K^N(t,v,v_\star,\sigma)|\overline{w}_{\mu^N}(dt,dv,dv_\star,d\sigma)>\epsilon/3\right)\to 0. \tag{6.310}$$

Gathering (6.304, 6.307, 6.310) and returning to (6.303), the lemma is proven. □

6.6.2 Law of Large Numbers

We next prove the following law of large numbers under the new measures for the sets $\mathcal{A}_\Theta$ given in the theorem.

Lemma 6.24. *Let $\mathbb{Q}^N$ be the probability measures constructed above, and suppose $L\subset\mathbb{N}$ is an infinite subsequence such that, the laws $\mathbb{Q}^N\circ(\mu^N_\bullet,w^N)^{-1}$ converges weakly on $\mathcal{D}\times\mathcal{M}(E)$, and let $(\mu_\bullet,w)$ be a random variable, defined with respect to a new probability space $(\Omega,\mathfrak{F},\mathbb{Q})$ and whose distribution is the limit. Then, for $\mathcal{A}_\Theta$ as in Theorem 6.5 for some $\alpha>0$ to be chosen,*

$$\mathbb{Q}\left((\mu_\bullet,w)\in\mathcal{A}_\Theta\right)=1. \tag{6.311}$$

It will be convenient, throughout, to realise all $(\mu_\bullet^N, w^N), N \in L$ and $(\mu_\bullet, w)$ on a common probability space with probability measure $\mathbb{Q}$, such that the law of $(\mu_\bullet^N, w^N)$ under $\mathbb{Q}$ is the same as under $\mathbb{Q}^N$, and such that $\mu_\bullet^N \to \mu_\bullet$ and $w^N \to w$ almost surely, and such that (6.299 - 6.301) hold with almost sure convergence as $N \to \infty$ through L:

$$\mathbb{Q}\left(\max_{i \le r_N} \left| \langle |v|^2 1\!\!1[|v| \le M_{N,i}], \mu_0^N \rangle - \Theta(t_i^{(r)}+) \right| \to 0 \right) = 1; \tag{6.312}$$

$$\mathbb{Q}\left(\inf_{t \in [0,t_{\text{fin}}]} \frac{N_t(M_N, r_N)}{N} \to 1\right) = 1 \tag{6.313}$$

$$\mathbb{Q}(\mathcal{W}_{1,1}(\mu_0^N, \mu_{0,M_N}^\star) \to 0) = 1. \tag{6.314}$$

We will write $\mathbb{E}_\mathbb{Q}$ for the expectation under this probability measure. For clarity, we will subdivide the argument into three smaller steps. We also observe that each $\mu_\bullet^N$ has jumps of size at most $4/N$, and so the limit $\mu_\bullet$ is $\mathbb{Q}$-almost surely continuous and the Skorokhod convergence can be upgraded to uniform convergence by Proposition 6.29c),

$$\mathbb{Q}\left(\sup_{t \le t_{\text{fin}}} \mathcal{W}_{1,1}(\mu_t^N, \mu_t) \to 0 \text{ as } N \to \infty \text{ through } L\right) = 1. \tag{6.315}$$

We also note immediately from (6.301) that

$$\mathcal{W}_{1,1}(\mu_0^N, \mu_0^\star) \le \mathcal{W}_{1,1}(\mu_0^N, \mu_{0,M_N}^\star) + \mathcal{W}_{1,1}(\mu_{0,M_N}^\star, \mu_0^\star) \to 0$$

$\mathbb{Q}$-almost surely, so that $\mu_0 = \mu_0^\star$ almost surely. We now prove the remaining properties defining $\mathcal{A}_\Theta$ one by one. First, we prove that the limit process $(\mu_\bullet, w)$ is almost surely a measure-flux pair.

Lemma 6.25 (Limiting Path as a Measure-Flux Pair)**.** *Continue in the notation following Lemma 6.24. Then*

$$\mathbb{Q}\left((\mu_\bullet, w) \text{ is a measure-flux pair, } w = \overline{w}_\mu\right) = 1. \tag{6.316}$$

In particular

$$\mathbb{Q}\left(\mu_\bullet \text{ is a solution to (BE) with } \mu_0 = \mu_0^\star\right) = 1. \tag{6.317}$$

Proof. This lemma is similar to Step 1 in the proof of the lower bound in Section 6.4. As in the cited proof, the continuity equation (CE) holds for the finite paths $(\mu_\bullet^N, w^N)$ $\mathbb{Q}$-almost surely, and since the set of pairs $(\mu_\bullet, w)$ satisfying the continuity equation is closed by Lemma 6.13, it follows that $(\mu_\bullet, w)$ solves (CE) almost surely.

We next show that $w = \overline{w}_\mu$, almost surely. Let us fix $g : E \to \mathbb{R}$ continuous and compactly supported, and start by observing that the process

$$M_t^{N,g} = \int_{E_t} g(s, v, v_\star, \sigma)(w^N - K^N \overline{w}_{\mu^N})(ds, dv, dv_\star, d\sigma) \tag{6.318}$$

is a $\mathbb{Q}$-martingale, with previsible quadratic variation at most

$$\begin{aligned}[M^{N,g}]_t &= N^{-1}\int_{E_t} g^2 K^N(s,v,v_\star,\sigma)B(v-v_*,\sigma)\mu_s^N(dv)\mu_s^N(dv_\star)dsd\sigma \\ &\le N_0(M_N,r_N)^{-1}t_{\text{fin}}\|g\|_\infty^2(3+2\langle|v|^2,\mu_0^N\rangle) \\ &\le N_0(M_N,r_N)^{-1}t_{\text{fin}}\|g\|_\infty^2(3+4\Theta(t_{\text{fin}})).\end{aligned} \tag{6.319}$$

We now observe that $M_0^{N,g}=0$ and $[M^{N,g}]_{t_{\text{fin}}}\to 0$ $\mathbb{Q}$-almost surely, which implies by standard martingale estimates that

$$\sup_{t\le t_{\text{fin}}} M_t^{N,g}\to 0 \qquad \text{in } \mathbb{Q}\text{-probability.} \tag{6.320}$$

We now investigate the difference between these martingales and the equivalent processes with $K\equiv 1$:

$$\begin{aligned}M_{t_{\text{fin}}}^{N,g} - \int_E g(s,v,v_\star,\sigma)(w^N-\overline{w}_{\mu^N})(ds,dv,dv_\star,d\sigma) \\ = \int_E g(s,v,v_\star,\sigma)(1-K^N)\overline{w}_{\mu^N}(ds,dv,dv_\star,d\sigma)\end{aligned} \tag{6.321}$$

We already controlled the integral of $1-K^N$ at (6.309) in the previous proof; integrating over time,

$$\begin{aligned}&\left|\int_E g(w^N-\overline{w}_{\mu^N})(ds,dv,dv_\star,d\sigma) - M_t^{N,g}\right| \\ &\qquad\le C_g t_{\text{fin}}\left(\left(\frac{N}{N_0(M_N,r_N)}-1\right)+\frac{1}{M_N}\right)\langle 1+|v|^2,\mu_0^N\rangle^2 \\ &\qquad\le C_g t_{\text{fin}}\left(\left(\frac{N}{N_0(M_N,r_N)}-1\right)+\frac{1}{M_N}\right)(1+2\Theta(t_{\text{fin}}))^2\end{aligned} \tag{6.322}$$

and, by the choice of r_N, M_N, the right-hand side converges to 0, almost surely. Finally, using Lemma 6.12, $\mathbb{Q}$-almost surely,

$$\left|\int_E g(t,v,v_\star,\sigma)(\overline{w}_{\mu^N}-\overline{w}_\mu)(ds,dv,dv_\star,d\sigma)\right|\to 0. \tag{6.323}$$

We now gather (6.320, 6.322, 6.323) to conclude that, $\mathbb{Q}$-almost surely, $\langle g, w-\overline{w}_\mu\rangle=0$. This extends to all $g\in C_c(E)$ simultaneously by taking a union bound over a countable dense subset of $C_c(E)$ to conclude that $\mathbb{Q}(w=\overline{w}_\mu)=1$, and the lemma is proven. □

Next, we prove that the second moment $\langle|v|^2,\mu_t\rangle$ coincides everywhere with the function $\Theta(t)$ given.

Lemma 6.26 (Second Moment of Limiting Path)**.** *We continue in the notation following Lemma 6.24. Then*

$$\mathbb{Q}\left(\langle|v|^2,\mu_t\rangle=\Theta(t) \text{ for all } t\le t_{\text{fin}}\right)=1. \tag{6.324}$$

Proof. We start by decomposing

$$\mu_t^N = \xi_t^N + \frac{N_t}{N}\nu_t^N \tag{6.325}$$

where ξ_t^N is the empirical measure of frozen particles $\xi_t^N = 1_{S_t}\mu_0^N$, which is constant on each time interval $[t_{i-1}^{(r_N)}, t_i^{(r_N)})$, and on each such time interval, ν_t^N is a Kac process on N_t particles. Thanks to the conditioning in the construction of $\mathbb{Q}^N$, we have the $\mathbb{Q}$-almost sure bound

$$\langle |v|^2, \nu^N_{t_{i-1}^{(r)}}\rangle \le \frac{N}{N_{t_{i-1}^{(r)}}}\langle |v|^2 1\!\!1_{|v|\le M_{N,i-1}}, \mu_0^N\rangle \le 4\Theta(t_{\text{fin}}). \tag{6.326}$$

We now fix an interval $I \subset [0, t_{\text{fin}}] \setminus P$ and $\ell > 0$ such that

$$\inf_{t\in I, s\in P, s<t}(t-s) \ge \ell > 0. \tag{6.327}$$

For each N, the points $t_i^{(r_N)}$ all belong to P, and so do not lie in I; we may therefore apply the moment creation property in Proposition 2.10i) to obtain, for all N large enough,

$$\mathbb{E}_{\mathbb{Q}}\left[\sup_{t\in I}\langle |v|^4, \nu_t^N\rangle\right] \le C\ell^{-2}\,\mathbb{E}_{\mathbb{Q}}\left[\langle |v|^2, \nu^N_{t_i^{(r_N)}}\rangle\right] = C\,\ell^{-2} \tag{6.328}$$

for some C depending only on $\Theta(t_{\text{fin}})$, uniformly in N. We next observe that $\|\xi_t^N\|_{\text{TV}} \le 1 - N_0/N \to 0$, uniformly in time $\mathbb{Q}$-almost surely, and so it follows from (6.315) that

$$\mathbb{Q}\left(\sup_{t\le t_{\text{fin}}} \mathcal{W}_{1,1}(\nu_t^N, \mu_t) \to 0 \text{ as } N\to\infty \text{ through } L\right) = 1. \tag{6.329}$$

Using Fatou's lemma and the lower semicontinuity of moments, we may now send $N \to \infty$ through L in (6.328) to obtain, for I as before, the same estimate on $\sup_I \langle |v|^4, \mu_t\rangle$, and together

$$\mathbb{E}_{\mathbb{Q}}\left[\sup_{t\in I}\left(\langle |v|^4, \nu_t^N\rangle + \langle |v|^4, \mu_t\rangle\right)\right] \le C\ell^{-2} \tag{6.330}$$

so we can find a large R, depending on ℓ, such that, for all N,

$$\mathbb{Q}\left(\sup_{t\in I}\langle |v|^2 1\!\!1_{|v|\ge R}, \nu_t^N + \mu_t\rangle \ge \epsilon/3\right) < \epsilon'/3. \tag{6.331}$$

Now, let f_R be a continuous, compactly supported function with $0 \le f_R \le |v|^2$ and $f_R = |v|^2$ when $|v| \le R$. By the uniform convergence (6.315), for $N \in L$ large enough,

$$\mathbb{Q}\left(\sup_{t\in I}|\langle f_R, \nu_t^N - \mu_t\rangle| \ge \epsilon/4\right) < \epsilon'/3 \tag{6.332}$$

and thanks to (6.331),

$$\mathbb{Q}\left(\sup_{t\in I}\langle |v|^2 - f_R, \nu_t^N + \mu_t\rangle \ge \epsilon/3\right) < \epsilon'/3 \tag{6.333}$$

and together, for $N \in L$ large enough,

$$\mathbb{Q}\left(\sup_{t\in I}\left|\langle |v|^2, \nu^N_t\rangle - \langle |v|^2, \mu_t\rangle\right| > 2\epsilon/3\right) \le 2\epsilon'/3. \tag{6.334}$$

For each N, the interval I lies in some $[t^{(r)}_{i-1}, t^{(r)}_i)$, $r = r_N, i = i_N$, as the endpoints of all such intervals always belong to P, and in particular, ν^N_t is a conservative Kac process on this interval, with

$$\langle |v|^2, \nu^N_{t^{(r)}_{i-1}}\rangle = \frac{N}{N_{t^{(r)}_{i-1}}}\langle |v|^2 \mathbb{1}_{|v|\le M_{N,i(N)-1}}, \mu^N_0\rangle \tag{6.335}$$

Thanks to (6.300), the first factor converges to 1, $\mathbb{Q}$-almost surely, and using (6.301), for all $\epsilon > 0$, we obtain

$$\mathbb{Q}\left(\sup_{t\in I}\left|\langle |v|^2, \nu^N_t\rangle - \Theta(t^{(r_N)}_{i(N)-1}+)\right| > \epsilon\right) \to 0. \tag{6.336}$$

Since I is an interval disjoint from P, Θ is constant on I, and by the construction of the points $t^{(r)}_i$, we have the nonrandom bound

$$\sup_{t\in I}\left|\Theta(t) - \Theta(t^{(r_N)}_{i(N)-1}+)\right| \le \frac{1}{r_N} \to 0 \tag{6.337}$$

and we conclude that, for $N \in L$ large enough,

$$\mathbb{Q}\left(\sup_{t\in I}\left|\langle |v|^2, \nu^N_t\rangle - \Theta(t)\right| \ge \epsilon/3\right) < \epsilon'/3. \tag{6.338}$$

Combining (6.334, 6.338), we have shown that, for all $\epsilon, \epsilon' > 0$,

$$\mathbb{Q}\left(\sup_{t\in I}\left|\langle |v|^2, \mu_t\rangle - \Theta(t)\right| > \epsilon\right) \le \epsilon' \tag{6.339}$$

so that $\langle |v|^2, \mu_t\rangle = \Theta(t)$ for all $t \in I$, $\mathbb{Q}$-almost surely. We can now cover $[0, t_{\text{fin}}] \setminus P$ by a countable collection of intervals of this form, so that this conclusion holds for all $t \notin P$ almost surely.

We now show that, on a single almost sure event, this also holds for $t \in P$. As remarked in Lemma 6.25, there is a $\mathbb{Q}$-almost sure event on which $\mu_\bullet$ is a solution to (BE) with $\mu_0 = \mu^\star_0$, and in particular $\mu_\bullet$ is continuous and the energy $\langle |v|^2, \mu_t\rangle$ is nondecreasing by Proposition 2.14. On this event, the equality $\langle |v|^2, \mu_0\rangle = \langle |v|^2, \mu^\star_0\rangle = 1 = \Theta(0)$ certainly holds at time $t = 0$, and on the intersection of this event and the event where the second moment equality holds for $t \notin P$, then any $t \in P \setminus \{0\}$ can be approached from below by $s \in [0, t_{\text{fin}}] \setminus P$. By left-continuity of Θ,

$$\Theta(t) = \limsup_{s\uparrow t, s\notin P} \Theta(s) = \limsup_{s\uparrow t, s\notin P} \langle |v|^2, \mu_s\rangle \le \langle |v|^2, \mu_t\rangle. \tag{6.340}$$

For the other inequality, on the same almost sure event as above, fix $t \in P$ and $\epsilon > 0$. By monotone convergence, we can find a continuous, compactly supported function $0 \le f \le |v|^2$ such that $\langle |v|^2, \mu_t\rangle < \langle f, \mu_t\rangle + \epsilon$. Using continuity in $\mathcal{W}_{1,1}$,

$$\langle |v|^2, \mu_t\rangle < \langle f, \mu_t\rangle + \epsilon = \limsup_{s\uparrow t, s\notin P} \langle f, \mu_s\rangle + \epsilon \le \limsup_{s\uparrow t} \Theta(s) + \epsilon = \Theta(t) + \epsilon \tag{6.341}$$

and, since $\epsilon > 0$ was arbitrary, we have equality at t. We emphasise again that the almost sure event used here does not depend on $t \in P$, and so the equality holds for all t simultaneously with $\mathbb{Q}$-probability 1, as desired. □

Finally we check the fourth moment conditions.

Lemma 6.27 (Fourth Moment of Limiting Path)**.** *Continue in the notation above. For $A(t)$ as in (6.22), for some $\alpha > 0$ to be chosen, we have*

$$\mathbb{Q}\left(\langle |v|^4, \mu_t\rangle \le A(t) \text{ for all } t \in [0, t_{\text{fin}}]\right) = 1. \tag{6.342}$$

Proof. Since $A = \infty$ on P, there is nothing to prove for such times. Let us fix $I = [u, v] \subset (0, t_{\text{fin}}]$ disjoint from P, and let $u' = \max(s : s \in P, s < u)$, which always exists, belongs to P and is strictly less than u, because P is closed and $0 \in P, 0 < u$. For any $J_a = [a, v] \supset I$ with $u' < a \le u$, we apply Lemma 6.25 to see that,$\mathbb{Q}$-almost surely, $(\mu_t)_{t\in J_a}$ is a solution to (BE), with constant energy given by $\langle |v|^2, \mu_t\rangle = \Theta(t) = \Theta(a) \le \Theta(t_{\text{fin}})$, because J_a is disjoint from P. Proposition 2.6 now applies pathwise, and for some absolute constant C,

$$\mathbb{Q}\left(\langle |v|^4, \mu_t\rangle \le C\Theta(t_{\text{fin}})(t-a)^{-2} \text{ for all } t \in I\right) = 1. \tag{6.343}$$

We now take $a \downarrow u'$ to obtain

$$\mathbb{Q}\left(\langle |v|^4, \mu_t\rangle \le C\Theta(t_{\text{fin}})(t-u')^{-2} \text{ for all } t \in I\right) = 1. \tag{6.344}$$

Choosing $\alpha = C\Theta(t_{\text{fin}})$, the bound is exactly $A(t)$, because $t-u' = \min(t-s : s \in P, s < t)$ for all $t \in I$. We now cover $[0, t_{\text{fin}}] \setminus P$ with countably many such I, and the claim is proven. □

Together, Lemmas 6.25, 6.26, 6.27 prove the conclusions of Lemma 6.24

6.6.3 Proof of Theorem

We now give the proof in the case of the regularised hard spheres kernel.

Proof of Theorem 6.5a. We first check that $\mathcal{A}_\Theta \subset \mathcal{D} \times \mathcal{M}(E)$ are compact. This follows almost exactly the same argument as Lemma 6.24 above: fix $(\mu^{(n)}_\bullet, w^{(n)}) \in \mathcal{A}_\Theta$. Since the spaces $\{\mu \in \mathcal{P}_2 : \langle |v|^2, \mu\rangle \le \Theta(t_{\text{fin}})\}$ are compact for W, and using the Boltzmann equation (BE) and the second moment bound to check equicontinuity, we can pass to a subsequence converging to a limit $(\mu_\bullet, w)$. First, since continuous functions are closed for Skorokhod convergence, $\mu_\bullet$ must also be continuous, so one can upgrade to uniform convergence $\sup_t \mathcal{W}_{1,1}(\mu^{(n)}_t, \mu_t) \to 0$ by Proposition 6.29c). Immediately, $\mu_0 = \mu^\star_0$, and the lower semicontinuity of moments gives $\langle |v|^4, \mu_t\rangle \le A(t), \sup_t\langle |v|^2, \mu_t\rangle \le \Theta(t_{\text{fin}})$. Using the same argument as (6.323) and the second moment bound, $\overline{w}_{\mu^{(n)}} = w^{(n)} \to \overline{w}_\mu$ so that

$w = \overline{w}_\mu$, and the same argument as before allows us to take the limit of the continuity equation to conclude that $\mu_\bullet$ solves (BE). Finally, repeating the arguments of Lemma 6.26, $\langle |v|^2, \mu_t\rangle$ can be found as the limit of $\langle |v|^2, \mu^{(n)}_t\rangle = \Theta(t)$ away from P to obtain $\langle |v|^2, \mu_t\rangle, t \notin P$. Since $\mu_\bullet$ solve (BE), $\langle |v|^2, \mu_t\rangle$ is nondecreasing by Proposition 2.14, and we may take left-limits to extend the equality to $t \in P$.

For the rate function, we return to the definition (6.13): all $\mu_\bullet \in \mathcal{A}_\Theta$ start at $\mu_0 = \mu^\star_0$, we have $H(\mu_0|\mu^\star_0) = 0$, and the unique choice $K = 1$ gives $\tau(K) = 0$, so $\mathcal{J}(\mu_\bullet, w) = 0$ and $\mathcal{I}(\mu_\bullet, w) = 0$ as desired.

We now prove (6.24,6.25). For the first item, let $\mathcal{U} \supset \mathcal{A}_\Theta$ be any open set, and $L' \subset \mathbb{N}$ a subsequence such that

$$\lim_{N\to\infty, N\in L'} \frac{1}{N} \log \mathbb{P}\left((\mu^N_\bullet, w^N) \in \mathcal{U}\right) = \liminf_{N\in\mathbb{N}} \frac{1}{N} \log \mathbb{P}\left((\mu^N_\bullet, w^N) \in \mathcal{U}\right). \tag{6.345}$$

For the changes of measure $\mathbb{Q}^N$ constructed above, we recall Lemma 6.23 to see that Corollary 6.9 applies, so that the laws $\mathbb{Q}^N \circ (\mu^N_\bullet, w^N)^{-1}$ are tight. We can therefore use Prohorov's Theorem pass to a further subsequence $L \subset L'$ such that the laws $\mathbb{Q}^N \circ (\mu^N_\bullet, w^N)^{-1}$ converge to the law of a new random variable $(\mu_\bullet, w)$ under a new probability measure $\mathbb{Q}$. This is exactly the setting of Lemma 6.24, from which $\mathbb{Q}((\mu_\bullet, w) \in \mathcal{A}_\Theta) = 1$, which certainly implies that $\mathcal{A}_\Theta$ is nonempty. We then have

$$\liminf_{N\in L} \mathbb{Q}^N\left((\mu^N_\bullet, w^N) \in \mathcal{U}\right) \geq \mathbb{Q}((\mu_\bullet, w) \in \mathcal{U}) \geq \mathbb{Q}((\mu_\bullet, w) \in \mathcal{A}_\Theta) = 1$$

since $\mathcal{U} \supset \mathcal{A}_\Theta$. Fixing $\epsilon > 0$ and recalling Lemma 6.23 again, we see that, for $N \in L$ large enough,

$$\mathbb{Q}^N\left((\mu^N_\bullet, w^N) \in \mathcal{U}, \frac{d\mathbb{Q}^N}{d\mathbb{P}} \leq e^{N(\Theta(t_{\text{fin}})z_2+\epsilon)}\right) > \frac{1}{2}.$$

It follows that, for $N \in L$ large enough,

$$\begin{aligned}\mathbb{P}\left((\mu^N_\bullet, w^N) \in \mathcal{U}\right) &= \mathbb{E}_{\mathbb{Q}^N}\left[\left(\frac{d\mathbb{Q}^N}{d\mathbb{P}}\right)^{-1} \mathbb{1}(\mu^N_\bullet, w^N) \in \mathcal{U}]\right] \\ &\geq e^{-N(\Theta(t_{\text{fin}})z_2+\epsilon)}\mathbb{Q}^N\left(\mu^N_\bullet \in \mathcal{U}, \frac{d\mathbb{Q}^N}{d\mathbb{P}} \leq e^{N(z_2\Theta(t_{\text{fin}})+\epsilon)}\right) \\ &\geq \frac{1}{2}e^{-N(\Theta(t_{\text{fin}})z_2+\epsilon)}.\end{aligned} \tag{6.346}$$

Taking the logarithm and the limit $N \to \infty$ through $L \subset L'$ and then the limit $\epsilon \downarrow 0$, we conclude that

$$\lim_{N\to\infty, N\in L'} \frac{1}{N} \log \mathbb{P}\left((\mu^N_\bullet, w^N) \in \mathcal{U}\right) \geq -\Theta(t_{\text{fin}})z_2$$

and by the choice (6.345), we have proven the same bound for the limit inferior over the full sequence $N \in \mathbb{N}$. The lower bound is independent of $\mathcal{U} \supset \mathcal{A}_\Theta$, and so we have proven the claim (6.24).

For the second item (6.25), we observe that Θ is locally constant at t_{fin}, so we can find an interval $I \ni t_{\text{fin}}$, with $\inf(t-s : t \in I, s \in P) > 0$ and such that $\Theta(t) = \Theta(t_{\text{fin}}) > 1$ for all $t \in I$. Thanks to the fourth moment bound in the construction of $\mathcal{A}_\Theta$, we can choose $R < \infty$ and a continuous, compactly supported function $0 \le f_R(v) \le |v|^2$ such that, for all $\mu_\bullet \in \mathcal{A}_\Theta$,

$$\inf_{t\in I}\langle f_R, \mu_t\rangle > \frac{1+\Theta(t_{\text{fin}})}{2}. \tag{6.347}$$

Now, writing $|I|$ for the Lebesgue measure of I, we choose $\mathcal{V}$ to be the set

$$\mathcal{V} = \left\{(\mu_\bullet, w) \in \mathcal{D}\times\mathcal{M}(E) : \int_I \langle f_R, \mu_t\rangle dt > \frac{1+\Theta(t_{\text{fin}})}{2}|I|\right\}. \tag{6.348}$$

$\mathcal{V}$ is open in $\mathcal{D}\times\mathcal{M}(E)$ by Lemma 6.12, and $\mathcal{A}_\Theta \subset \mathcal{V}$ by construction. However, for all N, we have the bound $\langle f_R, \mu^N_t\rangle \le \langle |v|^2, \mu^N_0\rangle$ for all t, because the kinetic energy is constant in time, so

$$\mathbb{P}\left((\mu^N_\bullet, w^N) \in \mathcal{V}\right) \le \mathbb{P}\left(\langle |v|^2, \mu^N_0\rangle > \frac{1+\Theta(t_{\text{fin}})}{2}\right). \tag{6.349}$$

We now apply Cranmér's theorem. Recalling the notation $\psi_0, \psi_0^\star$ defined in Subsection 6.6.1, we recall that $\psi_0^\star(a) > 0$ for all $a \neq \langle |v|^2, \mu_0^\star\rangle = 1$, and

$$\begin{aligned}\liminf_N \left[\frac{1}{N}\log\mathbb{P}\left((\mu^N_\bullet, w^N)\in\mathcal{V}\right)\right] &\le \liminf_N\left[\frac{1}{N}\log\mathbb{P}\left(\langle |v|^2, \mu^N_0\rangle > \frac{1+\Theta(t_{\text{fin}})}{2}\right)\right] \\ &= -\psi_0^\star\left(\frac{1+\Theta(t_{\text{fin}})}{2}\right) < 0.\end{aligned} \tag{6.350}$$

□

6.6.4 Maxwell Molecules Case

We now give the proof in the case of Maxwell molecules. In this case, since the kernel B is bounded, the moment creation property no longer holds; we also change measure so that, under $\mathbb{Q}^N$, the Kac process has a kernel $\widetilde{B} = \widetilde{B}_\delta$ with linear growth. The previous argument then applies, albeit with an additional (small) exponential cost. Since the argument is almost identical, we will discuss only the essential modifications relative to the regularised hard spheres case. As before, let us fix $(\Omega, \mathfrak{F}, (\mathfrak{F}_t)_{t\ge 0}, \mathbb{P})$ on which are defined Maxwell molecule Kac processes $\mu^N_\bullet$ and their empirical fluxes w^N.

Proof of Theorem 6.5b). Fix $\Theta, E, \delta > 0$ as in the statement. We construct the modification of the initial data via φ_M exactly as for the case of hard spheres above. With the same notation on $M_i, M_{N,i}, t_i^{(r)}$ and the special set of 'frozen' particles S_t, we now choose K to be given by

$$K^{N,M,r}(t, v, v_\star, \sigma) = \begin{cases} 0 & \text{if either } v, v_\star \in S_t; \\ N(1+\delta|v-v_\star|)\mathbb{1}_{N_t\ge 1}/N_t & \text{else} \end{cases} \tag{6.351}$$

where, again, we supress the argument μ_0^N. In this way, the non-frozen particles interact as a Kac process with kernel $(1+\delta|v|)$ on N_t particles on each time interval $[t_{i-1}^{(r)}, t_i^{(r)})$. We choose $M_N, r_N \to \infty$ in exactly the same way as before, and write $\widetilde{\mathbb{Q}}^N$ for the resulting changes of measure via Proposition 6.15:

$$\frac{d\widetilde{\mathbb{Q}}^N}{d\mathbb{P}} = \exp\Bigg(N\langle \varphi_{M_N}, \mu_0^N\rangle + \langle \log K^N, w^N\rangle \\ - N\int_0^{t_{\text{fin}}}\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}} (K^N-1)\overline{w}_{\mu^N}(dt,dv,dv_\star,d\sigma)\Bigg). \tag{6.352}$$

Again, we construct $\mathbb{Q}^N \ll \widetilde{\mathbb{Q}}^N$ by conditioning to the event of high probability where $\langle |v|^2, \mu_0^N\rangle \le 2\Theta(t_{\text{fin}})$ and $N_0/N \ge \frac{1}{2}$. We will write $K = (1+\delta|v-v_\star|)$ for the limiting tilting function. The strategy is now similar to the previous case. The law of large numbers follows in the same way for the new definition of $\mathcal{A}_\Theta$ without essential modification, allowing α to depend on δ and arguing in the same was as leading to (6.309) to obtain

$$\int_E |K^N - K|(t,v,v_\star,\sigma)\overline{w}_{\mu^N}(dt,dv,dv_\star,d\sigma) \\ \le Ct_{\text{fin}}\left(\left(\frac{N}{N_0(M_N,r_N)}-1\right)+\frac{1}{M_N}\right)\langle 1+|v|^2,\mu_0^N\rangle^2. \tag{6.353}$$

This is, in fact, the same estimate as before, up to the inclusion of δ; the linear factor $(1+\delta|v-v_\star|)$ now included in K, K^N replaces the equivalent one previously in the measure $\overline{w}_{\mu^N} = B(v-v_*,\sigma)\mu_t^N(dv)\mu_t^N(dv_\star)d\sigma dt$ so that the previous calculations are unchanged. With these modifications, the proof of the law of large numbers works exactly as before.

We again estimate the change of measure $\frac{1}{N}\log\frac{d\mathbb{Q}^N}{d\mathbb{P}}$. In this case, we will only find an estimate which asymptotically holds with sufficiently large probability, rather than with probabilities converging to 1 as we did before; this will not affect the final result. We recall that

$$\frac{1}{N}\log\frac{d\mathbb{Q}^N}{d\mathbb{P}} = \langle\varphi_{M_N},\mu_0^N\rangle + \langle \log K^N, w^N\rangle - \int_E (K^N-1)(t,v,v_\star,\sigma)\overline{w}_{\mu^N}(dt,dv,dv_\star,d\sigma) \\ - \log\widetilde{\mathbb{Q}}^N\left(\langle |v|^2,\mu_0^N\rangle \le 2\Theta(t_{\text{fin}}), \frac{N_0}{N}\ge\frac{1}{2}\right). \tag{6.354}$$

As before, the final term converges to 0. Let us fix $\epsilon > 0$. The term from the change of initial data is exactly as in the hard spheres case:

$$\mathbb{Q}^N\left(\langle\varphi_{M_N},\mu_0^N\rangle > z_2\Theta(t_{\text{fin}}) + \frac{\epsilon}{3}\right) \to 0. \tag{6.355}$$

In the second term, we now use the upper bound

$$\log K^N \le \log N/N_0(M_N,r_N) + \log(1+\delta|v-v_\star|) \le \log N/N_0(M_N,r_N) + \delta(|v|+|v_\star|). \tag{6.356}$$

As in the hard spheres case, the first term contributes at most $(\log N/N_0)w^N(E) \leq \epsilon/2$ with high $\mathbb{Q}^N$-probability, and in the second term, observe that

$$\delta\langle |v|+|v_\star|, w_t^N\rangle - \delta\int_{E_t}\frac{N}{N_0}(1+\delta|v-v_\star|)(|v|+|v_\star|)ds\mu_s^N(dv)\mu_s^N(dv_\star)d\sigma \tag{6.357}$$

is a $\mathbb{Q}^N$-supermartingale, so there exists a constant C such that

$$\begin{aligned}\mathbb{E}_{\mathbb{Q}^N}\left[\delta\langle |v|+|v_\star|, w^N\rangle\right] &\leq \delta\mathbb{E}_{\mathbb{Q}^N}\left[\int_E 2(|v|+|v_\star|+2\delta|v|^2+2\delta|v_\star|^2)ds\mu_s^N(dv)\mu_s^N(dv_\star)d\sigma\right] \\ &\leq \delta C\mathbb{E}_{\mathbb{Q}^N}\left[\int_0^{t_{\text{fin}}}\langle |v|^2, \mu_s^N\rangle ds\right] = \delta C t_{\text{fin}}\Theta(t_{\text{fin}}).\end{aligned} \tag{6.358}$$

Therefore, up to a new choice of C, for all N,

$$\mathbb{Q}^N\left(\delta\langle |v|+|v_\star|, w^N\rangle > \delta C t_{\text{fin}}\Theta(t_{\text{fin}})\right) \leq \frac{2}{9} \tag{6.359}$$

and including the term $\log N/N_0$, we conclude that

$$\mathbb{Q}^N\left(\langle \log K^N, w^N\rangle > \delta C t_{\text{fin}}\Theta(t_{\text{fin}})+\epsilon/3\right) < \frac{1}{3}. \tag{6.360}$$

For the final term of the first line of (6.354), we observe that $|K^N-1| \leq |K^N-K| + |K-1| \leq |K^N-K| + \delta(|v|+|v_\star|)$, and arguing from (6.353), for all N sufficiently large,

$$\mathbb{Q}^N\left(\int_E |K^N-K|\overline{w}_{\mu^N}(dt,dv,dv_\star,d\sigma) > \epsilon/3\right) \to 0 \tag{6.361}$$

while in the second term, we have the pathwise inequality

$$\int_E \delta(|v|+|v_\star|)\overline{w}_{\mu^N}(dt,dv,dv_\star,d\sigma) \leq 2\delta\int_0^{t_{\text{fin}}}\langle |v|^2, \mu_t^N\rangle dt \leq 4\delta t_{\text{fin}}\Theta(t_{\text{fin}}) \tag{6.362}$$

which implies

$$\mathbb{Q}^N\left(\int_E |K^N-1|\overline{w}_{\mu^N}(dt,dv,dv_\star,d\sigma) > 4\delta t_{\text{fin}}\Theta(t_{\text{fin}})+\epsilon/3\right) \to 0. \tag{6.363}$$

Gathering (6.355, 6.360, 6.363), we conclude that, for some absolute constant C and all N large enough,

$$\mathbb{Q}^N\left(\frac{1}{N}\log\frac{d\mathbb{Q}^N}{d\mathbb{P}} > \Theta(t_{\text{fin}})(z_2+C\delta)+\epsilon\right) < \frac{1}{3}. \tag{6.364}$$

Exactly the same argument also implies that

$$\lim_{a\to\infty}\limsup_N \mathbb{Q}^N\left(\frac{d\mathbb{Q}^N}{d\mathbb{P}} > e^{Na}\right) = 0 \tag{6.365}$$

so that Corollary 6.9 applies.

The conclusions of the theorem now follow in the same pattern as the hard spheres case. For the dynamic cost of any $(\mu_\bullet, w) \in \mathcal{A}_{\Theta,\delta}$, one bounds $\tau(k) \le (k-1)^2$ to obtain

$$\begin{aligned}\mathcal{J}(\mu_\bullet, w) = \int_E \tau(1+\delta|v-v_\star|)dt\mu_t(dv)\mu_t(dv_\star)d\sigma &\le 2\delta^2 \int_E (|v|^2+|v_\star|^2)dt\mu_t(dv)\mu_t(dv_\star)d\sigma \\ &= 4\delta^2 \int_0^{t_{\text{fin}}} \Theta(t)dt \le 4\delta^2 t_{\text{fin}}\Theta(t_{\text{fin}})\end{aligned} \tag{6.366}$$

and recalling that $\mu_0 = \mu_0^\star$ for all such $\mu_\bullet$, we conclude that the same bound holds for $\mathcal{I}(\mu_\bullet, w)$. If we now fix an open set $\mathcal{U} \supset \mathcal{A}_{\Theta,\delta}$, we let L' be an infinite subsequence along which $N^{-1}\log \mathbb{P}((\mu^N_\bullet, w^N) \in \mathcal{U})$ converges to its lim inf; using Corollary 6.9 to prove tightness, we can pass to a further subsequence L such that the laws $\mathbb{Q}^N \circ (\mu^N_\bullet, w^N)^{-1}$ converge weakly for the changes of measure above. By the law of large numbers, for $N \in L$ large enough,

$$\mathbb{Q}^N\left((\mu^N_\bullet, w^N) \in \mathcal{U}\right) \ge \frac{1}{2} \tag{6.367}$$

and, for $\epsilon > 0$ fixed, combining with (6.364), for all sufficiently large $N \in L$,

$$\mathbb{Q}^N\left((\mu^N_\bullet, w^N) \in \mathcal{U}, \frac{1}{N}\log\frac{d\mathbb{Q}^N}{d\mathbb{P}} \le \Theta(t_{\text{fin}})(z_2 + C\delta) + \epsilon\right) \ge \frac{1}{6}. \tag{6.368}$$

For such $N \in L$, we invert in the usual way to find

$$\mathbb{P}\left((\mu^N_\bullet, w^N) \in \mathcal{U}\right) \ge \frac{1}{4}\exp\left(-N(\Theta(t_{\text{fin}})(z_2 + C\delta) + \epsilon)\right). \tag{6.369}$$

Since $L \subset L'$ attains the limit inferior, we take the logarithm and send $N \to \infty$ through L to obtain

$$\liminf_N \frac{1}{N}\log \mathbb{P}\left((\mu^N_\bullet, w^N) \in \mathcal{U}\right) \ge -\Theta(t_{\text{fin}})(z_2 + C\delta) - \epsilon \tag{6.370}$$

and taking $\epsilon \to 0$ proves the claim. The final item, regarding open $\mathcal{V}_\delta$, follows in the same way as in the hard spheres case: we fix an open interval $I \ni t_{\text{fin}}$ on which Θ is constant, and bounded away from E, and write $|I|$ for its Lebesgue measure. Recalling that the fourth moment condition on $\mathcal{A}_{\Theta,\delta}$ depends on δ, we can choose $R = R_\delta$ and a continuous, compactly supported $0 \le f_\delta \le |v|^2$, which coincides on $|v|^2$ when $|v| \le R_\delta$ such that, for all $(\mu_\bullet, w) \in \mathcal{A}_{\Theta,\delta}$,

$$\inf_{t\in I}\langle f_\delta, \mu_t\rangle > \frac{1+\Theta(t_{\text{fin}})}{2} \tag{6.371}$$

and, following the previous case, take

$$\mathcal{V}_\delta := \left\{(\mu_\bullet, w) \in \mathcal{D}\times\mathcal{M}(E) : \int_I \langle f_\delta, \mu_t\rangle dt > \left(\frac{1+\Theta(t_{\text{fin}})}{2}\right)|I|\right\}. \tag{6.372}$$

Using Lemma 6.12 as before, these are open and contain $\mathcal{A}_{\Theta,\delta}$ by construction, and uniformly in δ,

$$\frac{1}{N}\log \mathbb{P}\left((\mu^N_\bullet, w^N) \in \mathcal{V}_\delta\right) \le \frac{1}{N}\log \mathbb{P}\left(\langle |v|^2, \mu^N_0\rangle > \frac{1+\Theta(t_{\text{fin}})}{2}\right) \tag{6.373}$$

so by Cramér,

$$\begin{aligned}\liminf_N \frac{1}{N} \log \mathbb{P}\left((\mu^N_\bullet, w^N) \in \mathcal{V}_\delta\right) &\le \liminf_N \frac{1}{N} \log \mathbb{P}\left(\langle |v|^2, \mu^N_0\rangle > \frac{1+\Theta(t_{\mathrm{fin}})}{2}\right) \\ &\le -\psi^\star_0\left(\frac{1+\Theta(t_{\mathrm{fin}})}{2}\right) < 0.\end{aligned} \tag{6.374}$$

The final bound is uniform in $\delta > 0$, and the theorem is complete. □

6.7 Applications of the Negative Result Theorem 6.5

We now give the two corollaries 6.7, 6.6, which are applications of the negative result Theorem 6.5.

Proof of Corollary 6.7. Throughout, fix Θ arbitrarily as in the statement of Theorem 6.5, and, in the case of Maxwell molecules, pick $\delta > 0$ arbitrarily, and let $\mathcal{A} = \mathcal{A}_\Theta, \mathcal{A}_{\Theta,\delta}$ be the resulting 'bad' set from Theorem 6.5 in either case. For a contradiction, let $(\mu^N_\bullet)_{N\in L}$ be an infinite subsequence which satisfies a large deviation principle with a rate function $\widetilde{\mathcal{I}}$ such that $\widetilde{\mathcal{I}}(\mu_\bullet, w) = \infty$ if $\mu_\bullet$ does not conserve energy. Since no paths in $\mathcal{A}$ conserve energy, we know that $\mathcal{I}(\mu_\bullet, w) = \infty$ for all $(\mu_\bullet, w) \in \mathcal{A}$ by construction. Due to exponential tightness in Proposition 6.1, the rate function must be good; that is, the sublevel sets $\{(\mu_\bullet, w) \in \mathcal{D} \times \mathcal{M}(E) : \widetilde{\mathcal{I}}(\mu_\bullet, w) \le a\}$ are compact in $\mathcal{D} \times \mathcal{M}(E)$ for any $a \in [0, \infty)$. Now, for any a, $\{\widetilde{\mathcal{I}} \le a\}$ is disjoint from $\mathcal{A}$, and since $\mathcal{D} \times \mathcal{M}(E)$ is a normal topological space, there exists an open set $\mathcal{U}_a \supset \mathcal{A}$ whose closure $\overline{\mathcal{U}}_a$ is disjoint from $\{\widetilde{\mathcal{I}} \le a\}$. By hypothesis,

$$\begin{aligned}\limsup_{N\in S} \left[\frac{1}{N} \log \mathbb{P}\left((\mu^N_\bullet, w^N) \in \overline{\mathcal{U}}_a\right)\right] &\le -\inf\left\{\widetilde{\mathcal{I}}(\mu_\bullet, w) : \mu_\bullet \in \overline{\mathcal{U}}_a\right\} \\ &\le -a.\end{aligned} \tag{6.375}$$

This is inconsistent with the conclusions (6.24,6.27) of Theorem 6.5 for a large enough, and we have the desired contradiction. □

We conclude with the proof of Corollary 6.6, which shows that the same behaviour, in which energy concentrates in a few particles, can also arise as a result of the binary collisions, even if the initial data are well-controlled. We will use a time-reversal argument, setting $\mu^\star_0 = \gamma$, and recalling the definitions of $\mathbb{T}\mu_\bullet, \mathbb{T}w$ from (6.20,6.21) respectively. We work with either regularised hard spheres or Maxwell molecules.

Proof of Corollary 6.6. Let Θ be the function given, and set $\Theta_\mathbb{T}$ to be the time reversed function $\Theta_\mathbb{T}(t) := \Theta(t_{\text{fin}} - t)$. By hypothesis, $\Theta_\mathbb{T}$ satisfies the conditions required in Theorem 6.5; in the case of Maxwell Molecules, choose $\delta > 0$ arbitrarily, and in either case set $A_\mathbb{T}$ to be the fourth moment bound given by Theorem 6.5 and $\mathcal{A}_{\mathbb{T},\Theta}$ the resulting bad set constructed by Theorem 6.5. We now set $A(t) := A_\mathbb{T}(t_{\text{fin}} - t)$ and set $\widehat{\mathcal{A}}_\mathbb{T}$ to be the projection

$$\widehat{\mathcal{A}}_\mathbb{T} = \{\mu_\bullet : (\mu_\bullet, w) \in \mathcal{A}_{\mathbb{T},\Theta}\}. \tag{6.376}$$

Since $\mathcal{A}_{\mathbb{T},\Theta}$ are compact and $\mathbb{T}$ preserves the Skorokhod topology of $\mathcal{D}$, it follows that $\widehat{\mathcal{A}}_\mathbb{T}$ are also compact, as are $\widehat{\mathcal{A}} := \{\mathbb{T}\mu_\bullet : \mu_\bullet \in \widehat{\mathcal{A}}_\mathbb{T}\}$ and by construction, the set desired can be written as $\mathcal{B} = \widehat{\mathcal{A}} \times \mathcal{M}(E)$.

Let us now fix $\mathcal{U} \supset \mathcal{B}$ open and $M > 0$ to be chosen later. Thanks to Lemma 6.8, we can choose a compact $\mathcal{K} \subset \mathcal{M}(E)$ such that $\mathbb{P}(w^N \notin \mathcal{K}) \leq e^{-MN}$ for all N, and since $\widehat{\mathcal{A}} \times \mathcal{K}$ is compact, we can choose $\mathcal{U}_1, \mathcal{U}_2$, open in $\mathcal{D}, \mathcal{M}(E)$ respectively, such that $\widehat{\mathcal{A}} \times \mathcal{K} \subset \mathcal{U}_1 \times \mathcal{U}_2 \subset \mathcal{U}$. Now, $\mathbb{T}\mathcal{U}_1$ is open and contains $\widehat{\mathcal{A}}_{\mathbb{T}}$ and using reversibility,

$$\mathbb{P}(\mu^N_\bullet \in \mathbb{T}\mathcal{U}_1) = \mathbb{P}(\mathbb{T}\mu^N_\bullet \in \mathcal{U}_1) = \mathbb{P}(\mu^N_\bullet \in \mathcal{U}_1). \tag{6.377}$$

Using Theorem 6.5 in either of the two cases on the open set $\mathbb{T}\mathcal{U}_1 \times \mathcal{M}(E) \supset \mathcal{A}_{\mathbb{T},\Theta}$, for some finite C, independent of M, it holds that

$$\liminf_N \frac{1}{N} \log \mathbb{P}\left(\mu^N_\bullet \in \mathbb{T}\mathcal{U}_1\right) = \liminf_N \frac{1}{N} \log \mathbb{P}\left((\mu^N_\bullet, w^N) \in \mathbb{T}\mathcal{U}_1 \times \mathcal{M}(E)\right) \geq -C \tag{6.378}$$

and thanks to (6.377), the same holds with $\mathcal{U}_1$ in place of $\mathbb{T}\mathcal{U}_1$. We now observe that

$$\begin{aligned} -C &\leq \liminf_N \frac{1}{N} \log \mathbb{P}\left((\mu^N_\bullet, w^N) \in \mathcal{U}_1 \times \mathcal{M}(E)\right) \\ &\leq \max\left(\liminf_N \frac{1}{N} \log \mathbb{P}\left((\mu^N_\bullet, w^N) \in \mathcal{U}_1 \times \mathcal{U}_2\right), \liminf_N \frac{1}{N} \log \mathbb{P}\left(w^N \notin \mathcal{U}_2\right)\right) \\ &\leq \max\left(\liminf_N \frac{1}{N} \log \mathbb{P}\left((\mu^N_\bullet, w^N) \in \mathcal{U}\right), -M\right) \end{aligned} \tag{6.379}$$

where, in the final line, we use the choice of $\mathcal{K}$ and recall that $\mathcal{U}_2 \supset \mathcal{K}$. If we now choose $M > C$, we must have that

$$\liminf_N \frac{1}{N} \log \mathbb{P}\left((\mu^N_\bullet, w^N) \in \mathcal{U}\right) \geq -C > -\infty \tag{6.380}$$

as claimed. □

Appendix

6.A Some Properties of Skorohod Paths

We will now recall some facts about right-continuous, left-limited (càdlàg) paths, and the resulting Skorohod topology. For a fixed metric space (X, d) and t_{fin}, we write $\mathbb{D}([0, t_{\text{fin}}], (X, d))$ for the set of all such functions $x_\bullet : [0, t_{\text{fin}}] \to X$, which we equip with the metric

$$\rho(x_\bullet, y_\bullet) = \inf\left\{\max\left(\sup_{t\le t_{\text{fin}}} d(x(t), y(\iota(t))), \sup_{t\le t_{\text{fin}}} |t-\iota(t)|\right) : \iota \in \Lambda\right\} \tag{6.381}$$

where the infimum runs over the set Λ of increasing, continuous bijections $\iota : [0, t_{\text{fin}}] \to [0, t_{\text{fin}}]$. We say that x has a jump of size at least $\epsilon > 0$ at t if $d(x(t), x(t-)) \ge \epsilon$.

Our first result is a replacement for uniform continuity in the context of such paths.

Proposition 6.28. *Let $x_\bullet \in \mathbb{D}([0, t_{\text{fin}}], (X, d))$ and fix $\epsilon > 0$. Then*

a). There exists at most finitely many $t \in [0, t_{\text{fin}}]$ such that $d(x(t), x(t-)) > \epsilon$.

b). There exists $\delta > 0$ such that, for all $t \in [0, t_{\text{fin}}]$, either *there exists $s \in [t, t+\delta) \cap [0, t_{\text{fin}}]$ with a jump discontinuity of size at least $d(x(s-), x(s)) \ge \epsilon$* or*, for all $s \in [t, t+\delta) \cap [0, t_{\text{fin}}]$, we have $d(x(t), x(s)) < \epsilon$.*

Proof. For the first item, suppose that we can find a countable sequence of distinct $t_n \in (0, t_{\text{fin}}]$ such that $d(x(t_n-), x(t_n)) > \epsilon$, and up to passing to an infinite subsequence, we can also arrange that t_n converges monotonically, either increasingly or decreasingly, to a limit $t \in [0, t_{\text{fin}}]$. We consider the two cases separately:

1. If $t_n \uparrow t$, we can pick $s_n \in [t_n - n^{-1}, t_n]$ such that $d(x(s_n), x(t_n)) > \epsilon$, which contradicts the fact that both $x(s_n), x(t_n) \to x(t-)$ by the left-limitedness.

2. If $t_n \downarrow t$, we can pick $s_n \in (t, t_n)$, still so that $d(x(s_n), x(t_n)) > \epsilon$, and obtain the same contradiction by the convergence $x(t_n), x(s_n) \to x(t)$ by right-continuity.

In either case, we have a contradiction, so the claim is proven.

We now prove item b). Suppose, for a contradiction, that the conclusion is false, so that we can construct t_n, s_n, with $t_n < s_n < t_n + n^{-1}$, such that there is no jump of size $\geq \epsilon$ in $[t_n, s_n)$, but such that $d(x(t_n), x(s_n)) \geq \epsilon$. As before, by passing to a infinite subsequence, we can arrange that either $t_n \downarrow t$ or $t_n \uparrow t$. We again deal with the cases separately.

1. If $t_n \uparrow t$, we split further into cases, depending on whether $s_n > t$ infinitely often or not.
   (a) If $s_n > t$ infinitely often, we can pass to a further subsequence so that $t_n \uparrow t, s_n \downarrow t$, so that $d(x(t_n), x(s_n)) \to d(x(t-), x(t))$. Since $d(x(t_n), x(s_n)) \geq \epsilon$ for all n by construction, we conclude that there is a jump discontinuity of size $\geq \epsilon$ at t, which contradicts the hypothesis that $[t_n, s_n)$ contains no such jumps.
   (b) Otherwise, $s_n \leq t$ eventually, so by passing to a subsequence, $t_n, s_n \uparrow t$ and $x(t_n), x(s_n) \to x(t-)$, which contradicts having $d(x(t_n), x(s_n)) \geq \varepsilon$.
2. If $t_n \downarrow t$, then $s_n \downarrow t$ and $x(t_n), x(s_n) \to x(t)$, contradicting that $d(x(t_n), x(s_n)) \geq \epsilon$.

Since all possible cases lead to a contradiction, the claim is proven. □

We next classify some continuity properties for the Skorokhod convergence. These results are standard and included for completeness.

Proposition 6.29. *a). The maps $x_\bullet \mapsto x(0)$, $x_\bullet \mapsto x(t_{\text{fin}})$ are continuous with respect to the metric ρ.*

b). If $x^n_\bullet \in \mathbb{D}([0, t_{\text{fin}}], (X, d))$ converge to $x_\bullet$ with respect to ρ, then for all but countably many $t \in [0, t_{\text{fin}}]$, $d(x^n(t), x(t)) \to 0$.

c). If, in b), the limit path $x_\bullet$ is continuous, then we additionally have the uniform convergence $\sup_t d(x^n(t), x(t)) \to 0$.

Proof. For the first item, observe that $\iota(0) = 0, \iota(t_{\text{fin}}) = t_{\text{fin}}$ for all $\iota \in \Lambda$, which implies that $d(x(0), y(0)) \leq \rho(x_\bullet, y_\bullet)$ for all $x_\bullet, y_\bullet$, and similarly at t_{fin}. For the second item, from the previous proposition, $x_\bullet$ is continuous at all but countably many $t \in [0, t_{\text{fin}}]$. For points of continuity t of $x_\bullet$, fix $\epsilon > 0$: there exists $\delta > 0$ such that, for all s with $|s - t| < \delta$, $|x(s) - x(t)| < \epsilon/2$. For all n sufficiently large, we have $\rho(x^n_\bullet, x_\bullet) < \min(\delta, \epsilon/2)$ and so we can pick $\iota \in \Lambda$ such that $\sup |t - \iota(t)| < \delta$ and $\sup_t |x^n(\iota(t)) - x(t)| < \epsilon/2$. We now conclude: we have $|t - \iota^{-1}(t)| < \delta$, and so

$$|x^n(t) - x(t)| \leq |x^n(t) - x(\iota^{-1}(t))| + |x(\iota^{-1}(t)) - x(t)| < \epsilon/2 + \epsilon/2 = \epsilon \tag{6.382}$$

and we are done. The final item also follows, noting that as $x_\bullet$ is continuous, it is uniformly continuous, which implies that δ, and hence n, can be chosen independently of $t \in [0, t_{\text{fin}}]$. □

6.B A Singular Girsanov Theorem for Jump Processes

We now justify the changes of measure in Proposition 6.15. We start from a filtered probability space $(\Omega, \mathfrak{F}, (\mathfrak{F}_t)_{t\geq 0}, \mathbb{P})$, on which is defined a Kac process $\mu^N_\bullet$ its empirical flux w^N_t. We have a deterministic tilting $\varphi : \mathbb{R}^d \to \mathbb{R}$ of the initial data, such that $\int e^{\varphi(v)}\mu_0^\star(dv) = 1$, and $A_0 \in \mathfrak{F}_0$ such that $\alpha_N = \mathbb{E}[e^{N\langle \varphi, \mu^N_0\rangle}\mathbb{I}_{A_0}] > 0$. By an abuse of notation, we understand A_0 as a subset of $\mathcal{P}^2_N$. Regaring the dynamics, we have a measurable $K : \mathcal{P}^N_2 \times E \to [0,\infty)$, enjoying a bound $K/(1+|v|+|v_\star|) \leq C$, for some absolute constant C. The modification of the dynamics with therefore be random, depending on the initial value μ^N_0: our new measures are given by

$$\begin{aligned}\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp\Bigg(& N\langle \varphi, \mu^N_0\rangle + N\langle \log K(\mu^N_0, \cdot), w^N_{t_{\text{fin}}}\rangle \\ & - N\int_E (K-1)(\mu^N_0, t, v, v_\star, \sigma)\overline{w}_{\mu^N}(dt, dv, dv_\star, d\sigma)\Bigg)\mathbb{I}_{A_0}/\alpha_N\end{aligned} \tag{6.383}$$

where, if w^N has any point with $K=0$, then the integral $\langle \log K(\mu^N_0, \cdot), w^N\rangle = -\infty$ and the density is understood to be 0.

We start from a disintegration of $\mathbb{P}$: let L_0 be the law of μ^N_0 on $\mathcal{P}^N_2$, and for any $\nu \in \mathcal{P}^N_2$, let $\mathbb{P}_\nu$ be the law of the Kac process started from ν, so that, for any $A \in \mathfrak{F}$,

$$\mathbb{P}(A) = \int_{\mathcal{P}^N_2} \mathbb{P}_\nu(A) L_0(d\nu). \tag{6.384}$$

We now write, again for any $A \in \mathfrak{F}$,

$$\mathbb{Q}(A) = \int_{\mathcal{P}^N_2} \mathbb{E}_\nu\left[Z^{\delta,\nu}_{t_{\text{fin}}}\mathbb{I}_A\right]\left(\frac{e^{N\langle\varphi,\nu\rangle}\mathbb{I}_{A_0}(\nu)}{\alpha_N}L_0\right)(d\nu) = \int_{\mathcal{P}^N_2}\mathbb{Q}_\nu(A)\widetilde{L}_0(d\nu) \tag{6.385}$$

where we define the modified law $\widetilde{L}_0$ by

$$\widetilde{L}_0(d\nu) = \frac{e^{N\langle\varphi,\nu\rangle}\mathbb{I}_{A_0}(\nu)}{\alpha_N}L_0(d\nu) \tag{6.386}$$

and modify the conditional law $\mathbb{P}_\nu$ by

$$Z^\nu_t = \exp\left(N\langle \log K(\nu,\cdot), w^N_t\rangle - N\int_E \mathbb{I}_{s\leq t}(K^\delta - 1)(\nu, s, v, v_\star, \sigma)\overline{w}_{\mu^N}(ds, dv, dv_\star, d\sigma)\right) \tag{6.387}$$

where we again set $Z^\nu_t = 0$ if there are any point with $K=0$, and $\mathbb{Q}_\nu = Z^\nu_{t_{\text{fin}}}\mathbb{P}_\nu$.

For the initial law, L_0 is the pushforward of $(\mu^\star_0)^{\otimes N}$ by the map $\theta_N : (v_1, .., v_N) \to N^{-1}\sum \delta_{v_i}$, and note that $\exp(N\langle\varphi,\nu\rangle) = \exp(\sum\varphi(v_i))$. It therefore follows that $\widetilde{L}_0$ is the pushforward of the measure

$$\exp(\sum\varphi(v_i))\prod_i \mu^\star_0(dv_i) = \prod_i e^{\varphi(v_i)}\mu^\star_0(dv_i)\mathbb{I}_{A_0}[\theta_N(v_1, ... v_N)]/\alpha_N$$

by θ_N. Since φ was chosen so that $\int e^{\varphi} d\mu_0^\star = 1$, each factor is a probability measure, and so $\widetilde{L}_0$ is the probability measure on $\mathcal{P}_2^N$ for the empirical measure of sampling N particles independently from the probability measure $e^{\varphi}\mu_0^\star$ and conditioning that the empirical measure lies in A_0, as claimed.

We now consider the modification of each $\mathbb{P}_\nu$. We observe that the conservation of energy guarantees that there exists $M = M_\nu$ such that, $\mathbb{P}_\nu$-almost surely, μ_t^N is supported on $[-M, M]^d$ for all t, and w^N is supported on $(t, v, v_\star, \sigma) \in E$ with $|v|, |v_\star| \leq M$. In particular, thanks to the hypothesised bound, one finds the upper bound

$$\sup_{t \leq t_{\text{fin}}} Z_t^\nu \leq \exp\left(CNM_\nu(1 + w^N(E))\right) \tag{6.388}$$

and the right-hand side has all moments finite, since $Nw_t^N(E)$ can be dominated by a Poisson process of rate $3(1 + \langle |v|^2, \nu\rangle)$, as in Section 6.2.1. We now observe that, at collisions, Z_t^ν changes by

$$Z_t^\nu - Z_{t-}^{\delta,\nu} = Z_{t-}^\nu \left(e^{\log K(\nu,t,v,v_\star,\sigma)} - 1\right) = Z_{t-}^\nu \left(K(\nu, t, v, v_\star, \sigma) - 1\right) \tag{6.389}$$

which is valid whether or not $K(\nu, t, v, v_\star, \sigma)$ is strictly positive, while in between collisions, Z_t^ν is differentiable, with

$$\frac{d}{dt} Z_t^\nu = -N \int_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{S}^{d-1}} (K - 1)(\nu, t, v, v_\star, \sigma) B(v - v_*, \sigma) \mu_t^N(dv) \mu_t^N(dv_\star). \tag{6.390}$$

Together, we obtain

$$Z_t^\nu = Z_0^\nu + \int_E \mathbb{I}_{s \leq t} N Z_{s-}^\nu \left(K(\nu, s, v, v_\star, \sigma) - 1\right)(w^N - \overline{w}_{\mu^N})(ds, dv, dv_\star, d\sigma) \tag{6.391}$$

which is a $\mathbb{P}_\nu$-local martingale, and hence a true martingale using the upper bound (6.388), with constant mean $Z_0^\nu = 1$. It follows that each $\mathbb{Q}_\nu$ is a probability measure, and hence so is $\mathbb{Q}$.

Let us now describe the dynamics under each $\mathbb{Q}_\nu$. Let us fix a bounded, measurable function $F : \mathcal{P}_N^2 \times \mathcal{P}_N^2 \times \mathcal{M}(E) \to \mathbb{R}$, and let A_t be given by

$$\begin{aligned} A_t =& N \int_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{S}^{d-1}} \left(F(\nu, \mu^{N,v,v_\star,\sigma}, w^{N,t,v,v_\star,\sigma}) - F(\nu, \mu^N, w^N)\right) \\ & \qquad \cdots \times K(\nu, t, v, v_\star, \sigma) B(v - v_*, \sigma) \mu^N(dv) \mu^N(dv_\star) d\sigma. \end{aligned} \tag{6.392}$$

We now consider $Y_t^\nu := Z_t^\nu(F(\nu, \mu_t^N, w_t^N) - \int_0^t A_s ds)$. The changes at jumps are given by

$$\begin{aligned} Y_t^\nu - Y_{t-}^\nu &= Z_{t-}^\nu \Bigg(K(\nu, t, v, v_\star, \sigma) - F(\nu, \mu_{t-}^N, w_{t-}^N) - F(\nu, \mu_{t-}^N, w_{t-}^N) \\ & \qquad \cdots - (K(\nu, t, v, v_*, \sigma) - 1) \int_0^t A_s ds \Bigg) \\ &= Y_{t-}^\nu (K(\nu, t, v, v_*, \sigma) - 1) + K(\nu, t, v, v_*, \sigma) Z_{t-}^\nu (F(\nu, \mu_t^N, w_t^N) - F(\nu, \mu_{t-}^N, w_{t-}^N)) \end{aligned} \tag{6.393}$$

while the drift between jumps is

$$\frac{d}{dt}Y^\nu_t = Z^\nu_t\Bigg(- N\left(F(\nu,\mu^N_t,w^N_t) - \int_0^t A_s ds\right)\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}}(K(\nu,t,v,v_\star,\sigma)-1) \\ \cdots\times B(v-v_*,\sigma)\mu^N_t(dv)\mu^N_t(dv_\star)d\sigma\Bigg) - A_tZ^\nu_t \\ = -NY^\nu_t\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}}(K(\nu,t,v,v_*,\sigma)-1)B(v-v_*,\sigma)\mu^N_t(dv)\mu^N_t(dv_*)d\sigma \\ - NZ^\nu_t\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}}(F(\nu,\mu^N_{t-}+\frac{1}{N}\Delta(v,v_*,\sigma),w^N_{t-}+\frac{1}{N}\delta_{(t,v,v_*\sigma)}) - F(\nu,\mu^N_{t-},w^N_{t-})) \\ \cdots\times K(\nu,t,v,v_*,\sigma)B(v-v_*,\sigma)\mu^N_{t-}(dv)\mu^N_{t-}(dv_*)d\sigma \\ = -N\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}} H(\nu,t,v,v_*,\sigma)B(v-v_*,\sigma)\mu^N_t(dv)\mu^N_t(dv_*)d\sigma \tag{6.394}$$

where the final line defines H. Combining the previous two displays, we write Y^ν_t as the stochastic integral

$$Y^\nu_t - Y^\nu_0 = N\int_{E_t} H(\nu,s,v,v_*,\sigma)(w^N-\overline{w}_{\mu^N})(ds,dv,dv_\star,d\sigma) \tag{6.395}$$

H is then locally bounded and previsible, so we conclude from (6.395) that Y^ν which is again a $\mathbb{P}_\nu$-martingale, using almost sure bound on the supports of $\overline{w}_\mu, w^N$ under $\mathbb{P}_\nu$ as commented above. It follows that $F(\nu,\mu^N_t,w^N_t)-\int_0^t A_s ds$ is a $\mathbb{Q}_\nu$-martingale, and we conclude that (μ^N_0,μ^N_t,w^N_t) is a $(\Omega,\mathfrak{F},(\mathfrak{F}_t)_{t\geq 0},\mathbb{Q}_\nu)$-Markov process with time-dependent generator

$$\mathcal{G}_tF(\nu',\mu^N,w^N) = N\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}}(F(\nu',\mu^{N,v,v_\star,\sigma},w^{N,t,v,v_\star,\sigma}) - F(\nu',\mu^N,w^N)) \\ \cdots\times K(\nu',t,v,v_\star,\sigma)B(v-v_*,\sigma)\mu^N(dv)\mu^N(dv_\star)d\sigma. \tag{6.396}$$

Using the same boundedness arguments as before, this generator characterises a unique semigroup $P^K_{s,t}$ of transition kernels on $\mathcal{P}^N_2\times\mathcal{P}^N_2\times\mathcal{M}(E)$, so that for any $0=t_0<t_1....<t_n$ and Borel sets $A_i\subset\mathcal{P}^N_2\times\mathcal{P}^N_2\times\mathcal{M}(E)$, we have

$$\mathbb{Q}_\nu\left((\mu^N_0,\mu^N_{t_i},w^N_{t_i})\in A_i, i=0,...,n\right) \\ = \int_{A_0\times...\times A_n}\delta_{(\nu,\nu,0)}(dx_0)P^K_{t_0,t_1}(x_0,dx_1)....P^K_{t_{n-1},t_n}(x_{n-1},dx_n). \tag{6.397}$$

Returning to (6.385), we conclude

$$\mathbb{Q}\left((\mu^N_0,\mu^N_{t_i},w^N_{t_i})\in A_i, i=0,...,n\right) \\ = \int_{A_0\times...\times A_n}\delta_{(\nu,\nu,0)}(dx_0)P^K_{t_0,t_1}(x_0,dx_1)....P^K_{t_{n-1},t_n}(x_{n-1},dx_n)\widetilde{L}_0(d\nu) \tag{6.398}$$

which is exactly the statement that, under $\mathbb{Q}$, $(\mu_0^N, \mu_t^N, w_t^N)$ is the Markov process with generator (6.396), and initial data $(\mu_0^N, \mu_0^N, 0)$, with μ_0^N sampled from $\widetilde{L}_0$ as above.

Chapter 7

Bilinear Coagulation Equations

7.1 Introduction & Main Results

This chapter is dedicated to the proof of Theorem 5, concerning the behaviour of the coagulative interaction clusters for the Kac process. As already mentioned in the introduction, this analysis does not rely at all on the way we obtained the coagulative particle system from the underlying Boltzmann dynamics; in fact, with the different structure of the equations, the coagulative equations we consider in this chapter have very different properties from (BE), and so this chapter will have a very different nature from the others. Since the proofs will not rely closely on the underlying Boltzmann dynamics, we will build a general framework of *bilinear coagulation systems* which promotes the key properties of the equations (Sm, Fl) in the introduction to axioms. We will state and prove Theorems 7.2 - 7.3 concerning the corresponding particle systems and limiting equations for this more general setting, and we will show how these apply to the Boltzmann case in Section 7.2

7.1.1 Definitions

We begin by precisely introducing our notion of a bilinear coagulation space. Our analysis rests on the *bilinear* form of the total rate of merger $\overline{K}(x,y)$ between two particles, which allows us to connect the Smoluchowski equation to random graphs in Section 7.5. The following definition makes this precise.

Definition 7.1.1. *A* Bilinear Coagulation System *is a 6-tuple* $(S, R, \pi, K, J, (J^N)_{N\geq 1})$ *consisting of a complete metric space* S*, a continuous involution* R *on* S*, a finite collection of continuous maps* $\pi = (\pi_i)_{0\leq i\leq n+m}$*,* $n \geq 1, m \geq 0$ *from* S *to* $\mathbb{R}$*, a nonnegative ('coagulation') symmetric kernel* $K : S \times S \to \mathcal{M}(S)$ *and a family of ('evolution') kernels* $J^N, J : S \to \mathcal{M}(S)$*, such that the following hold.*

i). For all $0 \le i \le n+m$ and all $x, y \in S$,

$$\pi_i = \pi_i(x) + \pi_i(y) \qquad K(x,y,\cdot)\text{- almost everywhere} \tag{7.1}$$

and for all $x \in S, N \ge 1$,

$$\pi_i = \pi_i(x) \qquad J(x,\cdot)\text{- almost everywhere and } J^N(x,\cdot)\text{- almost everywhere.} \tag{7.2}$$

ii). For $1 \le i \le n$, the map $\pi_i : S \to \mathbb{R}$ takes only nonnegative values, and π_0 takes values in the positive integers $\mathbb{N}$.

iii). The involution R satisfies

$$\pi_i \circ R = \begin{cases} \pi_i & 0 \le i \le n; \\ -\pi_i, & n+1 \le i \le n+m \end{cases} \tag{7.3}$$

and, for all $x, y \in S$ and $N \ge 1$,

$$K(Rx, Ry, \cdot) = R_\# K(x,y,\cdot); \quad J(Rx,\cdot) = R_\# J(x,\cdot); \quad J^N(Rx,\cdot) = R_\# J^N(x,\cdot) \tag{7.4}$$

where we write $_\#$ for the pushforward of a measure.

iv). There exists a constant C such that, for all $x \in S$,

$$\sum_{i=n+1}^{m} \pi_i(x)^2 \le C\varphi(x)^2 \tag{7.5}$$

where $\varphi(x) = \sum_{i=0}^n \pi_i(x)$. Moreover, any sublevel set $S_\xi = \{x \in S : \varphi(x) \le \xi\}$ is a countable union of compact sets, for any $\xi \in [0,\infty)$.

v). For all $x, y \in S$, the total rate $\overline{K}(x,y) = K(x,y,S)$ may be expressed as

$$\overline{K}(x,y) = \sum_{1 \le i,j \le n+m} a_{ij}\pi_i(x)\pi_j(y) \tag{7.6}$$

for a fixed $(n+m)\times(n+m)$ symmetric real matrix $A = (a_{ij})_{1 \le i,j \le n+m}$. Moreover, the matrix A is of the block-diagonal form

$$A = \begin{pmatrix} A^+ & 0 \\ 0 & A^{\mathrm{par}} \end{pmatrix} \tag{7.7}$$

where A^+, A^{par} are $n \times n$ and $m \times m$ square matrices respectively, and all entries of A^+ are nonnegative. Finally, for all $1 \le i \le n+m$, there exists $1 \le j \le n+m$ such that $a_{ij} > 0$, so that no row or column of A vanishes.

vi). Regarding the evolution kernels J^N, J, we ask that the total rate $\overline{J}(x) = J(x,S), \overline{J}^N(x)$ satisfy $\sup_x \frac{\overline{J}(x)}{\varphi(x)} < \infty$ and $\sup_N \sup_x \frac{\overline{J}^N(x)}{\varphi(x)} < \infty$. Further, for any $\xi > 0$, we have the convergence $\sup_{x \in S_\xi} \|J^N(x,\cdot) - J(x,\cdot)\|_{\mathrm{TV}} \to 0$, where the convergence is uniform over the sublevel sets S_ξ in item iv).

vii). For $f \in C_b(S)$, the maps

$$Kf : S \times S \to \mathbb{R}, \qquad (x, y) \mapsto \int_S f(z)K(x, y, dz); \tag{7.8}$$

$$Jf : S \to \mathbb{R}, \qquad x \mapsto \int_S f(z)J(x, dz) \tag{7.9}$$

are continuous.

Remark 7.1. *We think of $\pi_0(x)$ as counting the number of particles at time 0 which have been absorbed into x. As a result, we will ask in (A5.) below that our initial measure λ_0 is supported on $\{\pi_0\} = 1$, and π_0 artificially introduces monodisperse initial conditions.*

If we are given a space S equipped only with $\pi_1, ..., \pi_{n+m}$, we can replace S by $\mathbb{N} \times S$, and setting $\pi_0(a, x) = a, \pi_i(a, x) = \pi_i(x), i = 1, ..., n + m, (a, x) \in \mathbb{N} \times S$. In this way, and since π_0 does not enter the total rate $\overline{K}(x, y)$, the artificial requirements on π_0 above do not restrict the physics of the coagulation system.

Stochastic Particle Systems. With the setting defined above, we can introduce the interacting particle systems under consideration.

We study a system of coagulating particles $(x_j^N(t) : j \leq l^N(t))$, and the associated empirical measure

$$\lambda_t^N = \frac{1}{N} \sum_{j=1}^{l^N(t)} \delta_{x_j^N(t)} \tag{7.10}$$

with the following dynamical rules.

i). The rate at which unordered pairs of particles $\{x, y\}$ in S merge to form a new particle in $A \subset S$ is $2K(x, y, A)/N$.

ii). A particle of type x evolves can a particle of type $y \in A \subset S$ with a total rate $J^N(x, A)$.

This is a generalisation of a Marcus–Lushnikov coagulation process [134] on S, which we will refer to as the *stochastic coagulant*. Note that a $1/N$ scaling of the pair interaction rate is used, which ensures that each molecule has a total evolution rate of order 1. Dividing jump rates by N is equivalent to accelerating time by the same factor and this alternative formulation means that the jump rates in the definition of the "stochastic coalescent" in [5] as well as of the "stochastic K-coagulant" in [156] omit the $1/N$ from the rates and rescale time when taking the $N \to \infty$ limit.

Limiting kinetic equations. Let us first define a convenient class of test functions for a weak formulation of the kinetic equations. We define

$$\mathcal{A} := \{f \in C_b(S) : \sup\{\varphi(x) : x \in \mathrm{supp}(f)\} < \infty\}$$

so that any $f \in \mathcal{A}$ is supported on some S_ξ, $\xi < \infty$. We now consider various forms of the limiting Smoluchowski equation. Define a drift operator L, by specifying for all $f \in \mathcal{A}$,

$$\begin{aligned}\langle f, L(\lambda)\rangle = \int_{S^3} \{f(z) - f(x) - f(y)\} K(x, y, dz)\lambda(dx)\lambda(dy) \\ + \int_{S^2} \{f(y) - f(x)\} J(x, dy).\end{aligned} \tag{7.11}$$

Let us first remark that all the terms here make sense as soon as $f \in \mathcal{A}$, thanks to the form of the total rates of K, J in Definition 7.1.1. Compared to the dynamical rules for the stochastic coagulant above, the absence of the factor of 2 in the coagulation term here compensates for the fact that every unordered pair of particles x, y in the stochastic coagulant appears twice in the integral. The weak form of the Smolochowski equation for a process of measures $(\lambda_t)_{t<T}$ on S is

$$\forall f \in \mathcal{A},\ t < T, \qquad \langle f, \lambda_t\rangle = \langle f, \lambda_0\rangle + \int_0^t \langle f, L(\lambda_s)\rangle ds. \tag{Sm}$$

As discussed in the introduction, equation (Sm) captures the effects of coagulations between finite clusters; in order to include the possibility of a non-inert macroscopic component, or *gel*, we must move to a Flory-style equation. We modify the drift operator by setting, for $f \in \mathcal{A}$,

$$\langle f, L_{\mathrm{g}}(\lambda_t)\rangle = \langle f, L(\lambda_t)\rangle - \int_S f(x)\overline{K}(x, y)\lambda_t(dx)(\lambda_0 - \lambda_t)(dy). \tag{7.12}$$

In our general context, our Flory equation is now to ask

$$\forall f \in \mathcal{A},\ t < T, \qquad \langle f, \lambda_t\rangle = \langle f, \lambda_0\rangle + \int_0^t \langle f, L_{\mathrm{g}}(\lambda_s)\rangle ds. \tag{Fl}$$

As in the introduction, the additional term comes into play only after λ_t ceases to conserve the quantities $\langle \pi_i, \lambda_t\rangle, 1 \le i \le n + m$, and the extra term represents the interaction with the gel. This generalises the Smoluchowski coagulation equations [171] in a way analogous to Flory [199], and we use the term 'K-coagulant' for a solution to (Fl), following [156].

Precise conditions on measurability and integrability required to interpret these equations concretely are given in Appendix 7.A.

We write

$$g_t = (M_t, E_t, P_t) = \langle \pi, \lambda_0 - \lambda_t\rangle = (\langle \pi_i, \lambda_0 - \lambda_t\rangle)_{i=0}^{n+m} \tag{7.13}$$

for the gel data, where M_t, E_t, P_t are the 0^{th}, $1^{\text{st}} - n^{\text{th}}$, and $(n+1)^{\text{th}} - (n+m)^{\text{th}}$ coordinates, respectively. Following remarks in [156], one may show that if λ_t is a solution to (Fl), then the maps $t \mapsto \langle \pi_i, \lambda_t \rangle, i \leq n$ are non-increasing, which guarantees that $M_t, E_t \geq 0$. We write S^Π for the state space of gel data, given by

$$S^\Pi = \mathbb{N} \times \mathbb{R}^n \times \mathbb{R}^m \tag{7.14}$$

and use the same notation $\pi_i, 0 \leq i \leq n+m$ for the projections onto the factors. Similarly, we define $R^\Pi : S^\Pi \to S^\Pi$, by

$$R^\Pi(m, e, p) = (m, e, -p). \tag{7.15}$$

This definition is consistent in the sense that $R^\Pi(\pi x) = \pi(Rx)$ for all $x \in S$, by point iii) of Definition 7.1.1. When $x \in S$ and $g \in S^\Pi$, we use $\overline{K}(x, g)$ for the rate of absorption, given by (7.6) with the new meanings of $\pi_i(g)$. We will also write φ for the linear combination $\varphi = \sum_{i \leq n} \pi_i$, defined on both S and S^Π.

Definition 7.1.2 (Conservative Solutions)**.** *Let S be a bilinear coagulation system. We say that a solution $(\lambda_t)_{t<T}$ to either* (Sm) *or* (Fl) *is* conservative *if all the functions $t \mapsto \langle \pi_i, \lambda_t \rangle, 0 \leq i \leq n+m$ are constant on* $[0, T)$.

Thus, any solution to (Sm) or (Fl) is conservative up to some time $0 \leq t_{\mathrm{g}} \leq \infty$, and non-conservative thereafter.

We will usually impose symmetry requirements (A1.) on the initial data which guarantee that $\langle \pi_i, \lambda_t \rangle = 0$ for all t, for all $i = n+1, .., n+m$. As noted above, the functions $t \mapsto \langle \pi_i, \lambda_t \rangle, i \leq n$ are non-increasing, whenever $(\lambda_t)_{t<T}$ is a local solution to either equation. Therefore, under hypothesis (A1.), a solution $(\lambda_t)_{t<T}$ to either equation is conservative if, and only if, the map $t \mapsto \langle \varphi, \lambda_t \rangle$ is constant on $[0, T)$.

We will work in the space $\mathcal{M}_{\leq 1}(S)$ of (Borel) measures on S with total mass at most 1, which we equip with the metric ρ_1 defined in Section 2.1 for this choice of (S, d).

7.1.2 Statement of Results

We will make the following hypotheses on the initial data λ_0.

Hypothesis 7.1. *We will ask that the initial data λ_0 is a sub-probability measure on a bilinear coagulation space S, satisfying the following hypotheses.*

(A1.) The measure λ_0 is even under the transformation R: $R_\# \lambda_0 = \lambda_0$.

(A2.) For all $i \leq n$, we have $\langle \pi_i^3, \lambda_0 \rangle < \infty$.

(A3.) The set $\{\pi_i : 1 \leq i \leq n\}$ is linearly independent in the space $L^2(\lambda_0)$. In particular, none of the functions $\pi_i : 1 \leq i \leq n$ are 0 λ_0-almost everywhere.

(A4.) *The kernel K is λ_0-irreducible: if $A \subset S$ is such that, for all $x \in A$ and $y \in A^c$, $\overline{K}(x, y) = 0$, then either $\lambda_0(A) = 0$ or $\lambda_0(A^c) = 0$. Moreover, λ_0 is not a point mass.*

(A5.) *The initial data λ_0 is supported on $\{x \in S : \pi_0(x) = 1\}$.*

Let us remark that if (A3.) is false, we can repeat all of the arguments in terms of a smaller subset of $\{\pi_i : i \le n\}$ and recover the integrals of the removed components through linear dependence. If (A4.) is false, the space S can be decomposed into smaller, irreducible spaces, and particles in different components do not interact; we can therefore recover solutions of the limiting equations through linear combinations of the limiting equation on each irreducible component.

We summarise our results on the analysis of the Flory equation (Fl) as follows.

Theorem 7.2. *Let S be a π_0-bilinear coagulation system, and let λ_0 be a sub-probability measure on S satisfying Hypothesis 7.1. Then the equation (Fl) has a unique solution $(\lambda_t)_{t\ge 0}$ starting at λ_0; we write $g_t = (M_t, E_t, P_t)$ for the gel data defined in (7.13). This solution has the following properties.*

1. Phase Transition. *Let t_g be the first time at which the solution λ_t fails to be conservative, that is:*

$$t_g := \inf\{t \ge 0 : \langle \pi_i, \lambda_t\rangle \ne \langle \pi_i, \lambda_0\rangle \text{ for some } 0 \le i \le n + m\} = \inf\{t \ge 0 : \langle \varphi, \lambda_t\rangle < \langle \varphi, \lambda_0\rangle\}. \tag{7.16}$$

Then $t_g \in (0, \infty)$, and can be given explicitly in terms of the moments of λ_0 as

$$t_g = \sigma_1(\mathcal{Z}(\lambda_0))^{-1} \tag{7.17}$$

where $\sigma_1(\cdot)$ denotes the largest eigenvalue of a matrix, and $\mathcal{Z}(\lambda_0)$ is the $n \times n$ matrix of moments

$$\mathcal{Z}(\lambda_0)_{ij} = 2\langle (A\pi)_i \pi_j, \lambda_0\rangle, \qquad 1 \le i, j \le n. \tag{7.18}$$

2. Behaviour of the Second Moment. *Consider the second moments*

$$\mathcal{Q}(t) = (\langle \pi_i \pi_j, \lambda_t\rangle)_{i,j=0}^n; \qquad \mathcal{E}(t) = \langle \varphi^2, \lambda_t\rangle. \tag{7.19}$$

Then

i). $\mathcal{Q}(t)$ is finite and continuous, and so locally bounded, on $[0, \infty) \setminus \{t_g\}$.

ii). On $[0, t_g)$, each moment $\mathcal{Q}_{ij}$ is monotonically increasing, as is $\mathcal{E}$.

iii). At the gelation time, $\mathcal{E}(t_g) = \infty$, and $\mathcal{E}(t) \to \infty$ as $t \to t_g$.

3. Representation of Gel Data. *For each $t \geq 0$, there exists a unique maximal n-tuple $c_t = (c_t^i)_{i=1}^n \geq 0$ such that, for all $x \in S$,*

$$\sum_{i=1}^n c_t^i \pi_i(x) = 2t \int_S \left(1 - \exp\left(-\sum_{i=1}^n c_t^i \pi_i(y)\right)\right) \overline{K}(x,y)\lambda_0(dy). \tag{7.20}$$

c_t undergoes a phase transition at time t_g: if $t \leq t_g$, then $c_t = 0$, and if $t > t_g$ then at least one component of c_t is strictly positive. Moreover, the map $t \mapsto c_t$ is continuous.

The gel data are given in terms of c_t by

$$g_t^i = \int_S \pi_i(x) \left(1 - \exp\left(-\sum_{j=1}^n c_t^j \pi_j(x)\right)\right) \lambda_0(dx), \qquad 1 \leq i \leq n+m. \tag{7.21}$$

Therefore, if $t > t_g$ then $M_t > 0$, and $E_t > 0$ componentwise, while $P_t = 0$ for all $t \geq 0$. Moreover, the map $t \mapsto g_t$ is continuous, and $g_{t_g} = 0$.

4. Gel Dynamics. *The map $t \mapsto g_t$ is differentiable on $t \in (t_g, \infty)$, and*

$$\frac{d}{dt} g_t^i = \sum_{j,k=1}^n \langle \pi_i \pi_j, \lambda_t \rangle a_{jk} g_t^k, \qquad 1 \leq i \leq n. \tag{7.22}$$

5. Order of the Phase Transition, and the Size-Biasing Effect. *The map $t \mapsto c_t$ is right-differentiable at t_g, and as a consequence, the phase transition of g_t is first order. That is, the right-derivatives of the gel data $g_t^i, i = 0, 1, ..., n$ exist and are strictly positive at t_g. Moreover, there exist $\theta_i \geq 0, i = 1, .., n$, such that $\sum_i \theta_i = 1$ and such that*

$$\sum_{i=1}^n \theta_i (g'_{t_g+})_i \geq \left(\frac{\sum_{i=1}^n \langle \theta_i \pi_i, \lambda_0 \rangle}{\langle \pi_0, \lambda_0 \rangle} \right) (g'_{t_g+})_0. \tag{7.23}$$

We call this a size-biasing *effect: the average of the linear combination $\sum_i \theta_i \pi_i$ over particles in the early gel is at least the average over all particles. Let us define also the total interaction rate, which will quantify the inhomogeneity of the initial data λ_0:*

$$s(x) = \int_S \overline{K}(x,y)\lambda_0(dy). \tag{7.24}$$

If s is not constant λ_0-almost everywhere, then θ_i can be chosen so that the inequality in (7.23) is strict.

We also prove the following theorem, which is a law of large numbers result for the coagulating particle system $(x_j^N(t) : j \leq l^N(t))$. Firstly, following ideas of [156], we show that the empirical measure λ_t^N converges to the limiting solution $(\lambda_t)_{t\geq 0}$ in the weak topology, uniformly in time. The second part of the result is that the *stochastic gel*

$g_t^N = N^{-1}\pi(x_1^N(t))$ itself satisfies a law of large numbers, converging to the true gel g_t as $N \to \infty$, where we order[1] the particles so that x_1^N is the largest particle by π_0.

We make the following hypotheses for the law of large numbers. These are naturally satisfied when, for example, the initial particles $(x_i^N(0) : 1 \le i \le l^N(0))$ are sampled as a Poisson random measure with intensity $N\lambda_0$. However, it is useful for some intermediate results to give these results in the more general form used here.

Hypothesis 7.2. *Let λ_0^N be the initial data the stochastic coagulant, and let λ_0 be the initial data of the limiting Flory equation.*

(B1.) As $N \to \infty$, the initial measures $\lambda_0^N = \frac{1}{N}\sum_{i \le l^N(0)} \delta_{x_i}$ converge in probability to λ_0 in ρ_1, that is:

$$\rho_1(\lambda_0^N, \lambda_0) \to 0 \qquad \text{in probability.} \tag{7.25}$$

Moreover, λ_0^N is supported on the set $\{\pi_0 = 1\}$.

(B2.) We also have the convergence

$$\langle \pi_i, \lambda_0^N \rangle \to \langle \pi_i, \lambda_0 \rangle \qquad \text{in probability} \tag{7.26}$$

for all $0 \le i \le n+m$, and the uniform integrability

$$\sup_{N \ge 1} \mathbb{E}\langle \varphi^2, \lambda_0^N \rangle < \infty; \qquad \sup_{N \ge 1} \mathbb{E}\left[\langle \varphi^2, \lambda_0^N \rangle 1\!\!1\left(\langle \varphi^2, \lambda_0^N \rangle \ge M\right)\right] \to 0 \text{ as } M \to \infty. \tag{7.27}$$

Theorem 7.3. *Let λ_0 be a sub-probability measure on S satisfying Hypothesis 7.1, and let $(\lambda_t)_{t \ge 0}, (g_t)_{t \ge 0}$ be the associated solution to (Fl) and corresponding gel. For $N \ge 1$, let λ_t^N be the stochastic coagulant with initial data satisfying Hypothesis 7.2, and write $(x_j^N(t) : j \le l^N(t))$ for the particles of the stochastic system, sorted in decreasing order of π_0. Let $g_t^N = N^{-1}(\pi_i(x_1^N(t)))_{i=0}^n$ be the data of the largest particle in the stochastic system, normalised by N^{-1}. Then we have the convergence*

$$\sup_{t \ge 0} \left(\rho_1\left(\lambda_t^N, \lambda_t\right) + \left|g_t^N - g_t\right|\right) \to 0 \tag{7.28}$$

in probability. In particular, we have the following phase transition:

i). If $t \le t_{\mathrm{g}}$, then the largest particle has gel data of the order $o_{\mathrm{p}}(N)$;

ii). If $t > t_{\mathrm{g}}$, the largest particle has gel data of the order $\Theta_{\mathrm{p}}(N)$.

Moreover, if ξ_N is any sequence with $\xi_N \to \infty$ and $\frac{\xi_N}{N} \to 0$, then we may define $\widetilde{g}_t^N$ by summing the data of all particles $x_j^N(t)$ with $\pi_0(x_j^N(t)) \ge \xi_N$, and normalising by N. Then the same result holds when we replace g_t^N by $\widetilde{g}_t^N$ in (7.28).

Here, and throughout, we use the notation $o_{\mathrm{p}}(\cdot), \mathcal{O}_{\mathrm{p}}(\cdot), \Theta_{\mathrm{p}}(\cdot)$ for the probabilistic equivalents of $o(\cdot), \mathcal{O}(\cdot), \Theta(\cdot)$, and say that an event[2] holds *with high probability* if relevant probabilities converge to 1 as $N \to \infty$. Precise definitions can be found in [117].

[1] With an arbitrary rule for tie-breaks.

[2] Or, more formally, a sequence of events indexed by N.

7.1.3 Plan of the Chapter.

Our programme will be as follows.

i). In the remainder of this section, we will discuss other works on coagulating particle systems in the literature, and how they relate to our results.

ii). Before beginning on the analysis of the more general framework of bilinear coagulation systems, we discuss in Section 7.2 how the general framework can be applied to the specific case of the Kac interaction clusters. This will show how Theorems 7.2 - 7.3 imply Theorem 5 and link this chapter back to the ideas discussed in the introduction.

iii). In Section 7.3, we will prove that the limiting equation (Fl) has unique, globally defined solutions, based on a truncation argument from [155, 156].

iv). In Section 7.4, we prove an initial result, Lemma 7.7, on the convergence of the stochastic coagulant, using the ideas of [156, Theorem 4.1]. This will later be used to prove later points of Theorem 7.2 based on probabilistic arguments for the empirical measures λ^N_t, and the random graphs G^N_t introduced in Section 7.5.

v). In Section 7.5, we show how the stochastic coagulant can be coupled to a family of inhomogenous random graphs defined in [28]. Key results for these graphs are recalled in Appendix 7.B. The critical time t_c for these graphs may be found exactly, leading to the explicit expression in Theorem 7.2.

vi). A weakness of the preceding sections is that, a priori, the critical time t_c for the graph processes may differ from the gelation time t_g; in Section 7.6, we show that this cannot happen. This is based on a preliminary version of Theorem 7.3, which shows convergence of (λ^N_t, g^N_t) at a single fixed time $t \geq 0$.

vii). Section 7.7 is dedicated to a proof of item 2 of Theorem 7.2, concerning the second moments $\mathcal{Q}_{ij}(t) = \langle \pi_i\pi_j, \lambda_t\rangle, \mathcal{E}(t) = \langle \varphi^2, \lambda_t\rangle$. The statements about the subcritical and critical cases $t < t_\mathrm{g}, t = t_\mathrm{g}$ follow general ideas in [155, 156], while the statement about the supercritical case $t > t_\mathrm{g}$ uses additional ideas from the theory of random graphs.

viii). Section 7.8 uses the ideas of previous sections to prove items 3 and 4 of Theorem 7.2, concerning the gel data g_t beyond the critical point.

ix). Section 7.9 uses the analysis of the gel to extend Lemma 7.7 to show that convergence is uniform in time.

x). Section 7.10 proves item 5 of Theorem 7.2, concerning the behaviour near the critical point. This completes the proof of this theorem.

xi). To finish the proof of Theorem 7.3, we revisit the ideas of Section 7.6 to prove convergence of the stochastic gel $g_t^N, \widetilde{g}_t^N$, uniformly in time. This is the focus of Section 7.11, and builds further on ideas of previous sections.

7.1.4 Literature Review

Let us discuss some literature connected to the coagulation equations we discuss here and which is specific to this chapter. Smoluchowski originally considered the case where particles have only a mass in $\mathbb{N}$, and where the coagulation $x, y \mapsto x+y$ has a general rate $\overline{K}(x,y)$. In this case, identifying measures $\lambda \in \mathcal{M}_{\leq 1}(\mathbb{N})$ with a summable sequence, the equation analagous to (Sm) reads

$$\frac{d}{dt}\lambda_t(x) = \sum_{y<x} \overline{K}(y, x-y)\lambda_t(y)\lambda_t(x-y)dy - 2\lambda_t(x)\sum_{y=1}^{\infty}\overline{K}(x,y)\lambda_t(y)dy, \tag{7.29}$$

For an extensive review we refer to [5]. The case $\overline{K}(x,y) := xy$ is known as the *multiplicative* coagulation kernel and in this case with $\lambda_0 = \delta_1$, the solution of (7.29) exhibits gelation at $t_{\mathrm{g}} = \frac{1}{2}$.

The existence and value of the gelation time has been studied for a range of $\overline{K}$. For particles with integer masses and $\epsilon(xy)^\alpha \leq \overline{K}(x,y) \leq Mxy,\ M \in \mathbb{R}_+, \alpha \in (\frac{1}{2}, 1)$ Jeon [119] proved the existence of a gelation phase transition and provided an upper bound on the gelation time.

Müller [151] first introduced the study of coagulation equations on a continuous state space. Norris [155, 156] introduced a more general form, analagous to (Sm) on a general space S, allowing particles with internal structure and where, for any pair of particles, there are multiple possible coagulation products, in terms of a general conserved function φ, which is the spirit of our current investigation. This work provided showed that, if the merger rate goes like $\overline{K}(x,y) \leq \varphi(x) + \varphi(y)$, then $t_{\mathrm{g}} = \infty$, so no gelation occurs, and that, if the kernel is bounded above and below by multiples of $\varphi(x)\varphi(y)$, that gelation occurs at a finite, positive time, coinciding with the blow-up of the second moment. Under the second assumption, [156, Theorem 2.2] finds upper and lower bounds for the gelation time; however, these bounds do not coincide in general. Although these ideas do not immediately apply to our bilinear case, we will use this as an intermediate step in the derivation of Theorem 7.2.2. Normand [153] obtained explicit results concerning the blowup of a second moment for a sexed model which gives a lower bound on the gelation time, and in a later work [154] finds explicit expressions for the gelation time for a selection of models with arms. Consequently, this is one of the first (families of) models for which the gelation time can be found semi-explicitly; moreover, several aspects of our analysis extend what was previously known about the Smoluchowski equation, using the connection to random graphs [28].

The study of gelation as the formation of a very large connected structure by joining basic building blocks goes back at least to Flory [78] whose motivation was hydrocarbon polymerisation in the manufacture of plastics. Flory understood polymerisation as the formation of a random graph, rather than in terms of coagulation, and was aware of a sharp phase transition at the emergence of a giant connected structure, which he termed 'gel'. A rigorous proof of the random graph phase transition was provided by Erdős and Rényi [75] for the multiplicative coagulant. In terms of the random graphs we will use throughout this chapter, the case of the multiplicative coagulant corresponds exactly to sparse Erdős-Réyni graphs. The existence of a phase transition corresponding to the formation of a giant particle, which corresponds to the phase transition in Theorem 7.3, was first discussed by Lushnikov [134], who uses this to explain the explosion of the second moment, corresponding to item 2 of Theorem 7.2, in the particular case of the multiplicative kernel. The first connection between random graph and particle approaches appears in [34], where the phase transition is proved for the particle coagulation process and an interpretation as a new proof for a phase transition in the Erdős-Rényi random graph is noted; this is also discussed in the survey article [5]. We extend this connection, and show that the bilinear form of the merger rate allows us to couple the stochastic coagulant process to *inhomogeneous* random graphs as considered by [28].

As in the introduction, the original motivation of this chapter was to study a concept of interaction clusters introduced by Gabrielov et al. [96] in the context of the spatially inhomogeneous Boltzmann equation (spBE). The distribution of the sizes of the interaction clusters is formally derived in [161] in terms of the solution of the Boltzmann equation. Reducing to the case of cutoff Maxwell molecules for the spatially homogeneous Boltzmann equation, the phase transition observed in [96] can be identified precisely and the cluster size distributions observed to match those arising from the Smoluchowski coagulation equation with product kernel [134, 5, 161]. This current chapter is in the spirit of [162], where the clusters were studied for the Kac process as a softer problem, and which considers kernels including the hard sphere case to investigate the dependence of the gelation phenomena on the kinetic factor. It was formally shown in this work that, in a large particle number limit, the distribution of the cluster sizes converges to a version of the Smoluchowski coagulation equation with a time-dependent product kernel. This work also conjectured that gelation happens before the mean free time, which we verify for the kernels ($\mathrm{Q}_{a,b}$) here.

Let us finally remark that the systems we study are pure coagulation systems, rather than coagulation-fragmentation systems (see, for example [1, 17, 139]); once particles are merged, they remain merged forever. This simplification leads to very different behaviour; for example, the long-time behaviour of both the particle system and the limiting equation is degenerate, as all the mass is eventually absorbed into a single particle. This is why we can obtain uniform-in-time convergence in Theorem 7.3 without contradicting the sorts of arguments we saw in Section 3.7 for the Kac process.

7.2 Application to the Boltzmann Equation and Kac Process

Before turning to the analysis of coagulation systems in general, let us show how Theorems 7.2 and 7.3 apply in the case in the introduction to the coagulative structure of interaction clusters.

Let us first construct a space S which is rich enough to record all the information of collision histories in a cluster. We define the space of k-particle clusters $\mathbb{CL}_k$ to be the space of connected multigraphs[3] x on $\{1, \ldots, k\}$, where each vertex i is equipped with an ordered $c(x)_i = c_i = \deg_x(i) + 1$-tuple of vectors $v_i^1, \ldots v_i^{c_i} \in \mathbb{R}^d$, and the edges $E(x)$ have a partial ordering $\prec$ such that any edges with a common vertex are comparable. The edges will represent collision events between particles, with $e \prec e'$ meaning that the collision event represented by e takes place before that represented by e'; the tuples $v_i^1, \ldots, v_i^{c_x(i)}$ represent the velocities taken by particle i, initially and then after successive collisions, and we can extract the most recent velocities by keeping only the last entry of each tuple: $\mathbf{v}(x) = (v_i^{c_x(i)})_{i=1}^k$. All $x \in \mathbb{CL}_k$ are understood up to isomorphism, that is, up to a relabelling of $\{1, \ldots, k\}$ preserving the edges, the total ordering and the map v. This admits a complete metric, given by

$$d_k(x, y) = \min\left\{1, \inf_\tau \sum_{i=1}^k \sum_{j \le c_x(i)} |v_i^j - w_{\tau(i)}^j|\right\}$$

where the infimum runs over the (possibly empty) set of relabelings $\tau \in \mathrm{Sym}(1, .., k)$ which map the edges of x, with their partial order, onto those of y, and where we write v_i^j, w_i^j for the tuples associated to x, y. The overall state space for the coagulation is then the disjoint union

$$S = \bigsqcup_{k \ge 1} \mathbb{CL}_k$$

which we give a metric by setting $d(x, y) = d_k(x, y)$ if both $x, y \in \mathbb{CL}_k$ for some k, and $d(x, y) = 1$ otherwise. Let us also say that x is a *tree* if the underlying graph is a tree, that is, contains no cycles.

2. Coagulation & Evolution Kernels We next specify the transition maps K, J, J^N representing coagulation events and internal evolution respectively. For K, for each $x \in \mathbb{CL}_p, y \in \mathbb{CL}_q$ with most recent velocities $\mathbf{v}(x) = (v_i), \mathbf{v}(y) = (w_i)$ respectively and $i \le p, j \le q$, let us form a family of new graphs as follows. We form a new graph on $\{1, \ldots, p+q\}$ by adding p to the labels attached to y, keeping the edge structure and partial ordering the same. We add a new edge between i and $j + p$, and the new partial

[3] i.e. allowing multiple edges between the same vertexes.

ordering is the minimal extension of the previous partial orderings such that the new edge on $(i, j+p)$ lies above all previous edges in x, y with which it shares a vertex. Finally, the tuples of velocities at $i, j+p$ are changed by appending $v_i^{1+c_x(i)} = v'(v_i, w_j, \sigma)$ and $w_{j+p}^{1+c_y(j+p)} = v'_*(v_i, w_j, \sigma)$, for $\sigma \in \mathbb{S}^{d-1}$ and the usual post-collisional velocities v', v'_*. Calling this new graph[4] $M(x, y, i, j, \sigma)$, the coagulation kernel K is formally given by specifying, for $f \in C_c(S)$,

$$\int_S f(z)K(x, y, dz) := \sum_{i\le p, j\le q} \int_{\mathbb{S}^{d-1}} f\left(M(x, y, i, j, \sigma)\right) B(v_i - w_j, \sigma)d\sigma. \tag{7.30}$$

The other kind of behaviour is internal evolution, which in the Boltzmann model corresponds to a collision between particles already belonging to the same cluster. Continuing in the notation for x above, we define the family of new graphs for internal structure changes as follows. For $i, j \le p$, we add a new edge between vertexes i, j, extending the partial ordering so that this lies above all other edges with which it shares a vertex. The tuples of velocities of i, j are changed by adding $v_i^{1+c_x(i)} = v'(v_i, v_j, \sigma)$ and $v_j^{1+c_x(j)} = v'_*(v_i, v_j, \sigma)$ respectively; we call the new graph $U(x, i, j, \sigma)$. The corresponding kernel at the level of the Kac process is given by specifying

$$\int_S f(y)J^N(x, dy) = N^{-1} \sum_{i,j\le p} \int_{\mathbb{S}^{d-1}} f\left(U(x, i, j, \sigma)\right) B(v_i - v_j, \sigma)d\sigma$$

which forces the choice $J = 0$ in the limit.

3. Involution and Conserved Quantities We next specify the involution R and the conserved quantities π_i. For R, we simply negate all the velocities in a graph x, keeping the graph structure and partial ordering fixed. We set $n = 2$ and $m = d$ the dimension of the underlying space, and define for $x \in \mathbb{CL}_p$,

$$\pi_0(x) = \pi_1(x) = p; \qquad \pi_2(x) = \sum_{j=1}^p |v_j|^2; \qquad \pi_{i+2}(x) = \sum_{j=1}^p (v_j)_i$$

where $v_j = v_j^{c_x(j)}$ are the most recent velocities associated to x, and where $_i$ refers to the i^{th} component of the vector in $\mathbb{R}^d$.

4. Verification of requirements Let us now check that these are compatible with Definition 7.1.1. The requirement i) that π_i add at coagulation and are preserved under internal evolution follows immediately from the fact that binary collisions $(v, v_*) \to (v', v'_*)$ preserve energy and momentum; item ii) is true by the definitions, and iii) follows by the rotational invariance of the kernel. iv). is, in this context, simply the Cauchy-Schwarz inequality, and items vi-vii) are easy to check.

[4]Formally, an equivalence class of such graphs.

The only item which requires some thought, and for which we must restrict to the kernels ($Q_{a,b}$), is item v), where we require that the total interaction rate has a certain specific (exact) form. As in the introduction, we expand the total rate of the coagulation kernel K by setting $f = 1$ in (7.30) to find, in the same notation as above,

$$\begin{aligned}\overline{K}(x,y) &= \sum_{i\le \pi_0(x), j\le \pi_0(y)} \int_{\mathbb{S}^{d-1}} B(v_i - w_j, \sigma)d\sigma \\ &= \sum_{i\le \pi_0(x), j\le \pi_0(y)} (a + b|v_i|^2 - 2bv_i\cdot w_j + b|w_j|^2) \\ &= a\pi_1(x)\pi_1(y) + b(\pi_1(x)\pi_2(y) + \pi_2(x)\pi_1(y)) - 2b\sum_{i=3}^{d} \pi_i(x)\pi_i(y)\end{aligned} \tag{7.31}$$

which is of the form desired, with the block-diagonal matrix

$$A = \begin{pmatrix} a & b & 0 & 0 & \dots & 0 \\ b & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & -2b & 0 & \dots & 0 \\ 0 & 0 & 0 & -2b & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -2d \end{pmatrix}. \tag{7.32}$$

With this framework, let us now see how the statement of Theorem 5 in the introduction corresponds to Theorems 7.2, 7.3 which we will prove in this chapter. The conditions that $\langle |v|^6, \mu_0\rangle < \infty$ and the symmetry of μ_0 imply that the Hypotheses 7.1 holds, whereas Hypothesis 7.2 follows from the convergence $\mu_0^N \to \mu_0$, $\langle |v|^4, \mu_0^N\rangle \to \langle |v|^4, \mu_0\rangle$ hold, when we obtain λ_0^N, λ_0 as the pushforward of μ_0^N, μ_0 under the map which takes v to the graph x_v^1 on $\{1\}$, with no edges and decorated with the assignment $v_1 = v$.

With these in hand, almost all of Theorem 5 follows immediately. To compute the gelation time explicitly, we recall that $\mu_0 \in \mathcal{S}$ and that λ_0 is its pushforward, as above, to get

$$\langle \pi_1\pi_1, \lambda_0\rangle = 1; \quad \langle \pi_1\pi_2, \lambda_0\rangle = \langle |v|^2, \mu_0\rangle = 1; \quad \langle \pi_2\pi_2, \lambda_0\rangle = \langle |v|^4, \mu_0\rangle = \Lambda_4(\mu_0).$$

The requirement that $\langle \pi_1^3, \lambda_0\rangle$ is trivial, and $\langle \pi_2^3, \lambda_0\rangle = \langle |v|^6, \mu_0\rangle$ is finite by assumption. The matrix of moments $\mathcal{Z}(\lambda_0)_{ij} = 2\langle (A\pi)_i\pi_j, \lambda_0\rangle, i,j \le n$ is therefore

$$\mathcal{Z}(\lambda_0) = \begin{pmatrix} 2a+2b & 2a+2b\Lambda_4(\mu_0) \\ 2b & 2b \end{pmatrix} \tag{7.33}$$

The eigenvalues are the two roots of a quadratic equation which simplifies to

$$x^2 - 2x(a+2b) + 4b^2(1-\Lambda_4(\mu_0)) = 0$$

of which the larger is is the positive root, producing

$$t_{\mathrm{g}}^{-1} = a + 2b + \sqrt{(a+2b)^2 + 4b^2(\Lambda_4(\mu_0) - 1)}$$

which is the value claimed in Theorem 5. For comparison, the reciprocal of the mean free time is the mean collision rate:

$$t_{\text{mf}}^{-1} = 2\int_{\mathbb{R}^d\times\mathbb{R}^d}(a+b|v-w|^2)\mu_0(dv)\mu_0(dw) = 2(a+2b)$$

which is strictly less than t_{g}^{-1} unless $b^2(\Lambda_4(\mu_0)-1)=0$, which gives exactly the two special cases in Theorem 5: either $b=0$ or $\Lambda_4(\mu_0)=1$, and in the latter case the normalisation $\Lambda_2(\mu_0)=1$ implies that $\mu_0(|v|=1)=1$. We also remark that the special case of point 5 of Theorem 7.2 occurs only when the map

$$v\mapsto \int_{\mathbb{R}^d}(a+b|v-w|^2)\mu_0(dw) = a+b+b|v|^2$$

is constant $\mu_0(dv)$-almost everywhere, which is exactly the same two special cases: either $b=0$ or $|v|$ is constant μ_0-almost everywhere.

We must also check that the solution $(\lambda_t)_{t\geq 0}$ to the corresponding Flory equation (Fl) concentrates on tree-clusters. For $k\geq 1$, let us choose a continuous, nonincreasing function g_k on $[0,\infty)$ which is supported on $[0,k+1]$ and equal to 1 on $[0,k]$. We now define the functions

$$f_k(x) = \begin{cases} g_k(\pi_2(x)) & \text{if } x\in\mathbb{CL}_p \text{ is not a tree, } p\leq k; \\ 0 & \text{else} \end{cases}$$

so that each f_k belongs to the class $\mathcal{A}$. We next observe that, for any x,y,i,j,σ, the graph with one edge added $M(x,y,i,j,\sigma)$ is a tree if and only if both x,y are trees, since we only add one edge between the respective copies of x,y. It follows that, if $f_k(M(x,y,i,j,\sigma))>0$, then at least one of x,y has $\pi_1\leq k, \pi_2\leq k$ and fails to be a tree; without loss of generality, let us suppose that this is x. Since g_k is nonincreasing, we get

$$f_k(M(x,y,i,j,\sigma)) = g_k(\pi_2(x)+\pi_2(y)) \leq g_k(\pi_2(x)) = f_k(x).$$

Overall it follows that, in any case,

$$f_k(M(x,y,i,j,\sigma)) - f_k(x) - f_k(y) \leq 0$$

for any x,y,i,j,σ, and recalling that $J=0$ for this application, $\langle f_k, L(\lambda)\rangle\leq 0$ for any λ. In particular, along our solution λ_t we have $\langle f_k, L_{\text{g}}(\lambda_t)\rangle\leq 0$, so $\langle f_k,\lambda_t\rangle$ is nonincreasing. This starts at 0 because all singletons are trees, and since $f_k\geq 0$ we conclude that $\langle f_k,\lambda_t\rangle = 0$ for all t. Now sending $k\uparrow\infty$ for any fixed t, the functions f_k converge upwards to the indicator on clusters which are not trees, so

$$\lambda_t(x : x \text{ is not a tree}) = \lim_k \langle f_k,\lambda_t\rangle = 0$$

and we conclude that λ_t is supported on trees as claimed.

Let us finally check the claim regarding recovering a solution to the Boltzmann equation;

let λ_0 be obtained by pushing μ_0 forward as above, and let λ_t be the corresponding solution to (Fl); in Theorem 5, we claimed that the measures μ_t given by forgetting the cluster structure (and keeping only the most recent velocities) produced an energy-conserving solution to (BE) on $[0, t_{\mathrm{g}}]$, but were not probability measures for $t > t_{\mathrm{g}}$. We deal with the claims $t \le t_{\mathrm{g}}$, $t > t_{\mathrm{g}}$ separately.

Step 1. $t \le t_{\mathrm{g}}$ Let us fix $f \in C_b(\mathbb{R}^d)$ with $\|f\|_\infty \le 1$, and fix for the moment $t < t_{\mathrm{g}}$. Let us set $g = \mathcal{F}^\star f$ be the function on S which evaluates f at all the most recent velocities:

$$g(x) = (\mathcal{F}^\star f)(x) := \sum_{i=1}^{\pi_0(x)} f(v_i); \qquad (v_1, ..., v_{\pi_0(x)}) = \mathbf{v}(x).$$

Let us now fix $x \in \mathbb{CL}_p, y \in \mathbb{CL}_q$ with most recent velocities $(v_i), (w_j)$ as above. We observe, as in the introduction, that

$$\begin{aligned} &g(M(x,y,i,j,\sigma)) - g(x) - g(y) \\ &\qquad = f(v'(v_i,w_j,\sigma)) - f(v'_*(v_i,w_j,\sigma)) - f(v_i) - f(w_j) \end{aligned}$$

so that

$$\begin{aligned} &\int_S (g(z) - g(x) - g(y))K(x,y,dz) \\ &\quad = \sum_{i\le p, j\le q} \int_{\mathbb{S}^{d-1}} (f(v'(v_i,w_j,\sigma)) - f(v'_*(v_i,w_j,\sigma)) - f(v_i) - f(w_j))B(v_i - w_j,\sigma)d\sigma. \end{aligned} \tag{7.34}$$

We now use a truncation argument. For $R \ge 1$, let ψ_R be a continuous function which is equal to 1 on $[0, R]$ and supported only on $[0, R+1]$. We now set $g_R(x) := g(x)\psi_R(\pi_1(x) + \pi_2(x))$, so that we may use g_R as test functions in (Sm, Fl). For $t < t_{\mathrm{g}}$, only $L(\lambda_t)$ counts, and we have

$$\langle g_R, \lambda_t\rangle = \langle g_R, \lambda_0\rangle + \int_0^t \langle g_R, L(\lambda_s)\rangle ds, \qquad t < t_{\mathrm{g}}.$$

We now take $R \to \infty$; since $|g_R| \le \pi_1$ is integrated by all λ_t, we get $\langle g_R, \lambda_t\rangle \to \langle g, \lambda_t\rangle$. For the dynamic term, we write

$$\int_0^t \langle g_R, L(\lambda_s)\rangle ds = \int_0^t \int_{S\times S\times S} (g_R(z) - g_R(x) - g_R(y))\lambda_s(dx)\lambda_s(dy)K(x,y,dz).$$

For each x, y, z and $s \le t$, the integrand converges to $g(z) - g(x) - g(y)$; to use dominated convergence we write, uniformly in R, $|g_R(z) - g_R(x) - g_R(y)| \le 2(\pi_1(x) + \pi_1(y))$. We then dominate $\int_S K(x,y,dz) \le C(\pi_1(x) + \pi_2(x))(\pi_1(y) + \pi_2(y))$, for some C, so that for all x, y, uniformly in R,

$$\begin{aligned} &\int_S |g_R(z) - g_R(x) - g_R(y)|K(x,y,dz) \\ &\qquad \le C(\pi_1(x) + \pi_2(x))^2(\pi_1(y) + \pi_2(y)) + C(\pi_1(x) + \pi_2(x))(\pi_1(y) + \pi_2(y))^2. \end{aligned}$$

The right-hand side is integrable on $[0,t]$, because $t < t_g$, using Theorem 7.2.2.i) to see that $\langle(\pi_1+\pi_2)^2, \lambda_s\rangle$ is bounded on $s \in [0,t]$, and $\langle \pi_1+\pi_2, \lambda_s\rangle$ is bounded globally. We can therefore use dominated convergence to see that

$$\int_0^t \langle g_R, L(\lambda_s)\rangle ds \to \int_0^t \int_{S\times S\times S} (g(z)-g(x)-g(y))K(x,y,dz)\lambda_s(dx)\lambda_s(dy)ds$$

for all $t < t_g$, and we conclude that

$$\langle g, \lambda_t\rangle = \langle g, \lambda_0\rangle + \int_0^t \int_{S\times S\times S} (g(z)-g(y)-g(x))K(x,y,dz)\lambda_s(dx)\lambda_s(dy)ds. \tag{7.35}$$

Finally, we express everything in terms of the original function f and $\mu_t = \mathcal{F}\lambda_t$. We have $\langle g, \lambda_t\rangle = \langle f, \mu_t\rangle$ by definition, and integrating (7.34) shows that

$$\int_{S\times S\times S} (g(z)-g(x)-g(y))K(x,y,dz)\lambda_s(dx)\lambda_s(dy) = \langle f, Q(\mu_s)\rangle.$$

We therefore conclude from (7.35) that, for any $t < t_g$,

$$\langle f, \mu_t\rangle = \langle f, \mu_0\rangle + \int_0^t \langle f, Q(\mu_s)\rangle ds \tag{7.36}$$

so that $(\mu_t)_{t<t_g}$ solves (BE) as desired. We also have that $\langle 1, \mu_0 - \mu_t\rangle = \langle \pi_1, \lambda_0 - \lambda_t\rangle = 0$, so we have a probability measure, and $\langle |v|^2, \mu_0 - \mu_t\rangle = \langle \pi_2, \lambda_0 - \lambda_t\rangle = 0$, so we conserve energy, for $t \le t_g$.

To extend (7.36) to include the critical point, we observe that $\langle \pi_1, \lambda_{t_g}\rangle = 1$, so for all $\varepsilon > 0$, there exists $R < \infty$ such that $\langle \pi_1 \mathbb{1}\{\pi_1 > R\}, \lambda_{t_g}\rangle < \varepsilon$; it is straightforward to see that this integral can only increase on $[0, t_g]$, so the same holds for $t < t_g$. For the same truncated functions g_R as earlier, we have that $\langle g_R, \lambda_t\rangle \to \langle g_R, \lambda_{t_g}\rangle$ as $t \uparrow t_g$, using (Fl), and $\langle |g - g_R|, \lambda_t\rangle \le \langle \pi_1 \mathbb{1}\{\pi_1 > R\}, \lambda_t\rangle < \varepsilon$ for $t \le t_g$, so we conclude that $\limsup_{t\uparrow t_g} |\langle g, \lambda_t - \lambda_{t_g}\rangle| \le 2\varepsilon$. Since $\varepsilon > 0$ was arbitrary, we conclude that $\langle f, \mu_t\rangle = \langle g, \lambda_t\rangle$ converges to $\langle f, \mu_{t_g}\rangle = \langle g, \lambda_{t_g}\rangle$ as $t \uparrow t_g$. It is an easy calculation to see that $\langle f, Q(\mu_t)\rangle$ is bounded, we can take a limit $t \uparrow t_g$ of (7.36) above to see that the same equality holds at $t = t_g$.

Step 2. $t > t_g$. In the case $t > t_g$, we set $f = 1$ in the notation above, which produces $g = \mathcal{F}^\star f = \pi_1$. We therefore have, as remarked above,

$$\langle 1, \mu_t\rangle = \langle \pi_1, \lambda_t\rangle = 1 - M_t$$

which is valid for all $t \ge 0$. As soon as $t > t_g$, Theorem 7.2.3 shows that $M_t > 0$, and μ_t is a strictly sub-probability measure.

Let us conclude with a remark on the choice of S. We chose here a fairly complicated space, which is rich enough to encode not only the most recent velocities but also the entire collisional history. Most of the applications would be the same if we could extract only the most recent velocities or even only the overall particle number, energy and momentum of each cluster, to work on (much) simpler spaces.

7.3 Well-Posedness of the Flory Equation

This chapter is dedicated to a first analysis of the Smoluchowski equations (Sm, Fl), following Norris [155, 156]. Our goal in this section is to prove the following lemma on the well-posedness of (Fl).

Lemma 7.4. *For any measure $\lambda_0 \in \mathcal{M}_{\leq 1}(S)$ satisfying (A1.), the equation with gel (Fl) has a unique global solution $(\lambda_t)_{t\geq 0}$ starting at λ_0. Moreover, $P_t = 0$ for all $t \geq 0$.*

Corollary 7.5. *Suppose $(\lambda'_t)_{t<T}$ is a conservative local solution to the equation without gel, (Sm), starting at λ_0. Then $\lambda_t = \lambda'_t$ for all $t < T$, and $T < t_{\mathrm{g}}$. Hence, (Sm) has a unique maximal conservative solution, given by $(\lambda_t)_{t<t_{\mathrm{g}}}$.*

Our proof of Lemma 7.4 is an adaptation of the arguments in [155, Section 2] and [156, Section 2] and is based on a truncation argument. Recalling that $\varphi = \sum_{i=0}^n \pi_i$, we see that $\overline{K}(x,y) \leq \Delta\varphi(x)\varphi(y)$ for some $\Delta = \Delta(A)$. For all $\xi > 0$, we define the truncated particle space

$$S_\xi = \{x \in S : \varphi(x) \leq \xi\}. \tag{7.37}$$

We consider the following 'truncation at level ξ': in the empirical measure, we track only those particles inside S_ξ, and consider all other particles to belong to a 'truncated gel'. Although the particles in the truncated gel affect the dynamics in S_ξ, these contributions depend only on the total data g^ξ of the truncated gel, due to the bilinear form of the kernel. This leads to an ordinary differential equation with Lipschitz coefficients in an infinite dimensional space.

We formalise this intuition as follows. For a measure λ^ξ supported on S_ξ and $g^\xi \in S_{\mathrm{g}}$, we define a signed measure $L^\xi_{\mathrm{g}}(\lambda^\xi, g^\xi)$ on S_ξ by specifying, for all $f \in C_b(S)$,

$$\begin{aligned}&\langle f, L^\xi_{\mathrm{g}}(\lambda^\xi, g^\xi)\rangle \\ &\quad= \int_{S_\xi\times S_\xi\times S}[f(z)\mathbb{1}[\varphi(z)\leq\xi] - f(x) - f(y)]K(x,y,dz)\lambda^\xi(dx)\lambda^\xi(dy) \\ &\qquad\qquad + \int_{S_\xi}(f(y)-f(x))J(x,dy)\lambda^\xi(dx) - \int_{S_\xi} f(x)\overline{K}(x,g^\xi)\lambda^\xi(dx).\end{aligned} \tag{7.38}$$

This corresponds to the dynamics of particles inside S_ξ. The rate of change of the truncated gel data is given by

$$\begin{aligned}R^\xi_{\mathrm{g}}(\lambda^\xi, g^\xi) = &\int_{S_\xi^2}\pi(x+y)\mathbb{1}[\varphi(x)+\varphi(y)>\xi]\overline{K}(x,y)\lambda^\xi(dx)\lambda^\xi(dy) \\ &+ \int_{S_\xi}\pi(x)\overline{K}(x,g^\xi)\lambda^\xi(dx).\end{aligned} \tag{7.39}$$

We now seek measures λ^ξ_t supported on S_ξ and gel data $g^\xi_t = (M^\xi_t, E^\xi_t, P^\xi_t) \in S_{\mathrm{g}}$ such that, for all bounded measurable f on S_ξ,

$$\langle f, \lambda^\xi_t\rangle = \langle f, \lambda^\xi_0\rangle + \int_0^t \langle f, L^\xi_{\mathrm{g}}(\lambda^\xi_s, g^\xi_s)\rangle\, ds; \tag{Fl$|^1_\xi$}$$

$$g_t^\xi = g_0^\xi + \int_0^t R_{\rm g}^\xi(\lambda_s^\xi, g_s^\xi)ds. \tag{Fl$|_\xi^2$}$$

We will use the following existence and uniqueness result for the restricted dynamics (Fl$|_\xi^1$, Fl$|_\xi^2$).

Lemma 7.6. *[Existence and Uniqueness of Restricted Dynamics] Suppose λ_0^ξ is a finite measure on S_ξ which satisfies (A1.), and $g_0^\xi \in S_{\rm g}$ satisfies $\pi_i(g_0^\xi) = 0$ for all $i > n$. Then there exists a unique map (λ_t^ξ, g_t^ξ) on $[0, \infty)$, which solves the restricted dynamics (Fl$|_\xi^1$, Fl$|_\xi^2$). Moreover, for all $t \geq 0$, λ_t^ξ is a positive, finite measure on S_ξ, $P_t^\xi = 0$ for all times $t \geq 0$, and $g_t^\xi \in S_{\rm g}$.*

Sketch Proof of Lemma 7.6. This may be proved by a straightforward modification of the arguments in [155, Proposition 2.2]. We define Picard iterates $((\lambda_t^{(\xi,n)}, g_t^{(\xi,n)}) : n \geq 0, t \geq 0)$ by

$$(\lambda_t^{(\xi,0)}, g_t^{(\xi,0)}) = (\lambda_0^\xi, g_0^\xi); \tag{7.40}$$

$$\left(\lambda_t^{(\xi,n+1)}, g_t^{(\xi,n+1)}\right) = (\lambda_0^\xi, g_0^\xi) + \int_0^t (L_{\rm g}^\xi, R_{\rm g}^\xi)\left(\lambda_s^{(n,\xi)}, g_s^{(n,\xi)}\right) ds. \tag{7.41}$$

One then uses bilinear continuity arguments in total variation norm $\|\cdot\|_{\rm TV}$ to show that, given a bound $\langle \varphi, \lambda_0^\xi\rangle + \varphi(g_0^\xi) \leq C$, there is a positive time $T = T(\xi, C) > 0$ such that the Picard iterates $(\lambda_t^{(\xi,n)})_{t\leq T}$ converge uniformly in total variation on $[0, T]$, and that the limit λ_t^ξ solves (Fl$|_\xi^1$, Fl$|_\xi^2$), possibly allowing λ_t^ξ to be a signed measure. This argument also implies that the solution is unique on this interval. Now, we note that the quantity $\langle \varphi, \lambda_t^\xi\rangle + \varphi(g_t^\xi)$ is constant in time, and therefore this construction can be repeated on $[T, 2T]$, $[2T, 3T]$, etc, which proves global existence and uniqueness. Finally, an integrating factor is introduced (see also Step 3 of Lemma 6.22) to argue that λ_t is a positive measure. In our case, it is also straightforward to see that the gel data $M_t^\xi, E_t^\xi \geq 0$, and that $P_t^\xi = 0$, thanks to the symmetry (A1.). □

Proof of Lemma 7.4. We first show existence. For all $\xi < \infty$, we let (λ_t^ξ, g_t^ξ) be the solution to the dynamics (Fl$|_\xi^1$, Fl$|_\xi^2$) restricted to S_ξ, with initial data

$$\lambda_0^\xi(dx) = 1\!\!1[x \in S_\xi]\, \lambda_0(dx); \qquad g_0^\xi = \int_{x \notin S_\xi} \pi(x)\lambda_0(dx). \tag{7.42}$$

Observe that, if $\xi < \xi'$, then $\widetilde{\lambda}_t^\xi, \widetilde{g}_t^\xi$ given by

$$\widetilde{\lambda}_t^\xi(dx) = 1\!\!1_{x\in S_\xi}\, \lambda_t^{\xi'}(dx); \qquad \widetilde{g}_t^\xi = g_t^{\xi'} + \int_{x\in S_{\xi'}\setminus S_\xi} \pi(x)\lambda_t^{\xi'}(dx) \tag{7.43}$$

solve the dymanics (Fl$|_\xi^1$,Fl$|_\xi^2$) with the same initial data λ_0^ξ, g_0^ξ. From uniqueness in Lemma 7.6, it follows that $\widetilde{\lambda}_t^\xi = \lambda_t^\xi; \widetilde{g}_t^\xi = g_t^\xi$. This shows that the measures λ_t^ξ are increasing in

ξ, while the gel data M_t^ξ, E_t^ξ are decreasing, and P_t^ξ is identically 0, by symmetry (A1.). Therefore, the limits

$$\lambda_t = \lim_{\xi\uparrow\infty} \lambda_t^\xi; \qquad M_t = \lim_{\xi\to\infty} M_t^\xi; \qquad E_t = \lim_{\xi\to\infty} E_t^\xi \tag{7.44}$$

exist in the sense of monotone limits; one can then check that λ_t and $g_t = (M_t, E_t, 0)$ satisfy the full equation (Fl), with initial values λ_0 and $g_0 = 0$.

To see uniqueness, let λ_t be the solution constructed above and write $g_t = (M_t, E_t, P_t)$ for the data of the gel. Let $\widetilde{\lambda}_t$ be any solution to (Fl) starting at λ_0, and let $\widetilde{g}_t = (\tilde{M}_t, \tilde{E}_t, \tilde{P}_t)$ be the associated data of the gel. For all $\xi < \infty$, it is simple to verify that

$$\widetilde{\lambda}_t^\xi(dx) = 1\!\!1_{x\in S_\xi}\, \widetilde{\lambda}_t(dx); \qquad \widetilde{g}_t^\xi = \widetilde{g}_t + \int_{S_\xi^c} \pi(x)\widetilde{\lambda}_t(dx) \tag{7.45}$$

is a solution to the dynamics ($\mathrm{Fl}|^1_\xi$, $\mathrm{Fl}|^2_\xi$) on S_ξ. By uniqueness in Lemma 7.6, it follows that $\widetilde{\lambda}_t^\xi = \lambda_t^\xi$, and taking monotone limits, we see that $\widetilde{\lambda}_t = \lim_{\xi\to\infty} \widetilde{\lambda}_t^\xi = \lim_{\xi\to\infty} \lambda_t^\xi = \lambda_t$. The argument for $\widetilde{g}$ is identical. □

7.4 Convergence of the Stochastic Coagulant

We now turn to a preliminary version of Theorem 7.3. In this section, we will outline the proof of the convergence of the stochastic coagulant λ_t^N to a solution λ_t of (Fl), locally uniformly in time. Most of the arguments are well-known for the Smoluchowski equation [155, 156], and for brevity, we will sketch the proof with an indication of the nontrivial technical details. Throughout, we fix λ_0 satisfying Hypothesis 7.1, and λ_t^N with initial data λ_0^N satisfying Hypothesis 7.2. Our result is as follows.

Lemma 7.7. *Suppose λ_0 satisfies Hypothesis 7.1, and let $(\lambda_t)_{t\ge0}$ be the solution to (Fl) starting at λ_0. Let $(\lambda_t^N)_{t\ge0}$ be stochastic coalescents with initial data λ_0^N satisfying Hypothesis 7.2. Then we have the local uniform convergence*

$$\forall t_{\mathrm{f}} \ge 0 \qquad \sup_{t\le t_{\mathrm{f}}} \rho_1(\lambda_t^N, \lambda_t) \to 0 \text{ in probability} \tag{7.46}$$

where recall that ρ_1 is a complete metric inducing the weak topology.

Remark 7.8. *We will later upgrade the* local *uniform convergence to* full *uniform convergence in Lemma 7.37. We also remark that this does not immediately imply the convergence of the gel terms in Theorem 7.3, as the test functions involved neither belong to $\mathcal{A}$, nor are even bounded. This will be dealt with in Sections 7.6, 7.11, where the proofs build on this result.*

Proof. Since we seek only qualitative convergence, we use the well-known method of proving tightness and identifying possible limit paths.

Step 1. Tightness Firstly, the jump rates can bounded, uniformly in time, in terms of the initial second moment $\langle \varphi^2, \lambda^N_0\rangle$, with a total jump rate of the form $cN(1+\langle \varphi^2, \lambda^N_0\rangle)$. Thanks to (B2.), the kinetic factors these are stochastically bounded: $\langle \varphi^2, \lambda^N_0\rangle \in \mathcal{O}_\mathrm{p}(1)$ as $N\to\infty$, while the jumps are all of the order N^{-1} in the metric d. As a result, it follows that for all $t_\mathrm{f}\ge 0$, the processes $(\lambda^N_t)_{0\le t\le t_\mathrm{f}}$ are tight in the Skorohod topology of $\mathbb{D}([0,t_\mathrm{f}],(\mathcal{M}_{\le 1}(S),\rho_1))$, see for instance [116, Theorem 4.6] or [4].

Step 2. Identification of the Limit Let us now prove that the only possible limit path is the desired path $(\lambda_t)_{t\le t_\mathrm{fin}}$ given by (Fl). Let us fix a subsequence $L\subset\mathbb{N}$, along which $(\lambda^N_t)_{t\le t_\mathrm{fin}}$ converges in distribution in the Skorokhod topology to a limit $(\overline{\lambda}_t)_{t\le t_\mathrm{fin}}$. We observe that, for almost all $\xi>0$, the following hold:

i). Almost surely, for almost all $t\le t_\mathrm{f}$,

$$\overline{\lambda}_t\left(\{x\colon \varphi(x)=\xi\}\right)+\overline{\lambda}_t\otimes\overline{\lambda}_t\left(\{(x,y)\colon \varphi(x+y)=\xi\}\right)=0; \tag{C1.}$$

ii). This also holds for $t=0$. That is, almost surely,

$$\overline{\lambda}_0\left(\{x\colon \varphi(x)=\xi\}\right)+\overline{\lambda}_0\otimes\overline{\lambda}_0\left(\{(x,y)\colon \varphi(x+y)=\xi\}\right)=0. \tag{C2.}$$

For such ξ, we consider the pair

$$\lambda^{N,\xi}_t=\lambda^N_t 1\!\!1_{S_\xi};\qquad g^{N,\xi}_t=\langle \pi,\lambda^N_t-\lambda^{N,\xi}_t\rangle \tag{7.47}$$

and take the limit $N\to\infty$. We first observe that $\overline{\lambda}_0=\lambda_0$ using (B1.), and that $g^{N,\xi}_t\to \langle \pi,\lambda_0-\overline{\lambda}^\xi_t\rangle=\overline{g}^\xi_t$ using (C1.-C2.). For any $f\in\mathcal{A}$, we observe that the process

$$\begin{aligned} M^{N,f}_t=\langle f,\lambda^{N,\xi}_t-\lambda^{N,\xi}_0\rangle &-\int_0^t\int_{S_\xi\times S_\xi\times S}(f(z)1\!\!1_{z\in S_\xi}-f(x)-f(y))K(x,y,dz)\lambda^N_s(dx)\lambda^N_s(dy)ds\\ &-\int_{S_\xi} f(x)\overline{K}(x,g^{N,\xi}_s)\lambda^N_s(dx)ds\\ &-\int_0^t\int_{S_\xi\times S_\xi}(f(y)-f(x))J^N(x,dy)\lambda^N_s(dx)ds \end{aligned} \tag{7.48}$$

is a martingale, of quadratic variation on the order $\mathbb{E}[M^{N,f}]_{t_\mathrm{fin}}\le CN^{-1}$, and so $M^{N,f}_t$ converge in probability in the Skorokhod topology to the 0 process. On the other hand, on the subsequence L, we take the limit $N\to\infty$. This is licit, since we can use (C1., C2.) to show that the discontinuity of the indicators in z,x,y does not change the convergence, since the sets where the integrand is discontinuous receive 0 mass in the limit. We therefore have the convergence in distribution

$$M^{N,f}_t\to\langle f,\overline{\lambda}^\xi_t-\overline{\lambda}^\xi_0\rangle-\int_0^t\langle f,L^\xi_\mathrm{g}(\overline{\lambda}^\xi_t,\overline{g}^\xi_t)\rangle ds \tag{7.49}$$

and it follows that the right-hand side is 0 almost surely. Taking an intersection over a countable set of functions in $C_b(S_\xi)$ which are dense for the topology of uniform converge on compact sets, and since (Fl$|^1_\xi$) depends only on the behaviour on the set S_ξ, we conclude that $\overline{\lambda}_t 1\!\!1_{S_\xi}, \overline{g}^\xi_t = \langle \pi, \lambda_0 - \overline{\lambda}^\xi_0 \rangle$ satisfy the truncated equation (Fl$|^1_\xi$) with $\overline{\lambda}^\xi_0 = \lambda^\xi_0$ and $\overline{g}^\xi_0 = 0$, and a similar argument holds for (Fl$|^2_\xi$). Thanks to the construction of solutions to the global equation (Fl) in Lemma 7.4, we know that for all such ξ, $\overline{\lambda}_t 1\!\!1_{S_\xi}$ coincides with $\lambda_t 1\!\!1_{S_\xi}$ almost surely. Finally, we take a limit of such $\xi \uparrow \infty$, to conclude the equality $\overline{\lambda}_t = \lambda_t, t \le t_\mathrm{f}$. We have now identified the only possible limit process as the (nonrandom) solution $(\lambda_t)_{t\ge 0}$ to (Fl) starting at λ_0, so we must have $\lambda^N_t \to \lambda_t$ in the Skorokhod topology in probability.

Step 3. Uniformity of Convergence Since the limit process $(\lambda_t)_{0\le t\le t_\mathrm{f}}$ is continuous in the weak topology $(\mathcal{M}_{\le 1}(S), \rho_1)$, it follows that we may upgrade from Skorohod to uniform convergence:

$$\sup_{0\le t\le t_\mathrm{f}} \rho_1\left(\lambda^N_t, \lambda_t\right) \to 0 \quad \text{in probability} \tag{7.50}$$

as claimed. □

7.5 Coupling of the Stochastic Coagulant to Random Graphs

In this section, we will show that the stochastic coagulant defined in the introduction may be coupled to a *dynamic* version of the random graphs $\mathcal{G}^{\mathcal{V}}(N, tk/N)$ considered in [28]. This allows us to apply some results of that paper, which we summarise in Appendix 7.B, to analyse the stochastic coagulant process and the limit equation.

Definition 7.5.1. *[Dynamic Inhomogenous Random Graphs] Fix a measure λ_0 satisfying Hypothesis 7.1. Let $\mathbf{x}_N = (x_i, i = 1, 2, ..., l^N)$ be a collection random points in S of potentially random length l^N, and sample $\tau_e \sim$ Exponential(1), independently of each other, for $e = (ij), 1 \le i, j \le l^N$, and independently of $\mathbf{x}_N$. We define the kernel*

$$k(v, w) = 2\overline{K}(x, y) \tag{7.51}$$

where the right-hand side is the total mass of the interaction kernel $\overline{K}(x, y) = K(x, y, S)$. We form the random graphs $(G^N_t)_{t\ge 0}$ on $\{1, 2, ..., l^N\}$ by including the edge $e = (ij)$ if

$$t \ge \frac{N\tau_e}{k(x_i, x_j)}. \tag{7.52}$$

We write $G^N_t \sim \mathcal{G}(\mathbf{x}_N, \frac{tK}{N})$ for the distribution of G^N_t, for a single fixed $t \ge 0$. We say that G^N_0 satisfy Hypothesis 7.2 for λ_0 if the same is true of the empirical measures $\lambda^N_0 = N^{-1}\sum_{i\le l^N} \delta_{x_i}$. We emphasise that the x_i do not *change during the dynamics.*

This has the following immediate consequences. Firstly, the conditions in Hypothesis 7.2 guarantee that $\mathcal{V} = (S, \lambda_0, (\mathbf{x}_N)_{N\geq 1})$, is a generalised vertex space in the sense of [28], which is recalled in Definition 7.B.1, and k is an irreducible kernel as described in Definition 7.B.2, thanks to (A4.). Using both parts of (B2.), one can also show that the kernel k is *graphical* in the sense of Definition 7.B.4.

For all times t, G^N_t is an instance of the inhomogeneous random graph from Definition 7.B.3. Moreover, the process $(G^N_t)_{t\geq 0}$ is increasing, and is a Markov process, by the memoryless property of the exponential variables τ_e. We write T for the convolution operator

$$(Tf)(x) = \int_S f(y)k(x,y)\lambda_0(dy) \tag{7.53}$$

and $\|T\|$ for the associated operator norm in $L^2(\lambda_0)$.

We write also $t_c = \|T\|^{-1}$. The following is the basic statement of a phase transition for G^N_t, which follows from Theorem 7.45.

Lemma 7.9. *Let λ_0 satisfy Hypothesis 7.1, and let G^N_t be the random graphs constructed above, such that G^N_0 satisfy Hypothesis 7.2. Write $C_1(G^N_t)$ for the size of the largest component of G^N_t. Then we have the following phase transition:*

i). If $t \leq t_c$, then $N^{-1}C_1(G^N_t) \to 0$ in probability.

ii). If $t > t_c$, then there exists $c = c(t)$ such that, with high probability, $C_1(G^N_t) \geq cN$.

We write $\mathcal{C}_1(G), ...\mathcal{C}_j(G), ...$ for the connected components, which we also call *clusters*, of G, in decreasing order of size, allowing $\mathcal{C}_j = \emptyset$ if G has fewer than j components and $C_j(G)$ for the number of vertices in $\mathcal{C}_j(G)$. For a cluster $\mathcal{C}$ of the graph G^N_t, we will write

$$M(\mathcal{C}) = \sum_{i\in\mathcal{C}} \pi_0(x_i), \quad E(\mathcal{C}) = \left(\sum_{i\in\mathcal{C}} \pi_j(x_i)\right)^n_{j=1}, \quad P(\mathcal{C}) = \left(\sum_{i\in\mathcal{C}} \pi_j(x_i)\right)^{n+m}_{j=n+1} \tag{7.54}$$

for the unnormalised data, and

$$\pi(\mathcal{C}) = \sum_{i\in\mathcal{C}} \pi(x_i) = (M(\mathcal{C}), E(\mathcal{C}), P(\mathcal{C})), \quad \varphi(\mathcal{C}) = \sum_{i\in\mathcal{C}}\sum_{j=0}^n \pi_j(x_i). \tag{7.55}$$

We write $\delta(\mathcal{C})$ for the point mass $\delta(\mathcal{C}) = \delta_{\pi(\mathcal{C})}$, and $\pi_\star(G^N_t)$ for the normalised empirical measure

$$\pi_\star(G^N_t) = \frac{1}{N}\sum_{\text{Clusters}} \delta(\mathcal{C}) \tag{7.56}$$

where the sum is over all clusters $\mathcal{C}$ of G^N_t. This is connected to the stochastic coagulants as follows:

Lemma 7.10 (Coupling of Random Graphs and Stochastic Coagulants)**.** *Fix points* $\mathbf{x}_N = (x_1, ..., x_{l^N(0)})$ *in* S*, and let* $(G_t^N)_{t\geq 0}$ *be the random graph process described in Definition 7.5.1 for this choice of vertex data. Consider also a stochastic coagulant* $(\lambda_t^N)_{t\geq 0}$ *started from* $\lambda_0^N = \frac{1}{N}\sum_{i\leq l^N(0)} \delta_{x_i}$*. Then the processes* $\pi_\star(G_t^N)$ *and* $\pi_\#\lambda_t^N$ *are equal in law.*

Remark 7.11. *This is the key result which makes much of our analysis possible. Many of the remaining points of Theorem 7.2 above concern only the moments* $\langle \pi_i, \lambda_t\rangle$, $\langle \varphi^2, \lambda_t\rangle$ *which depend on* λ_t *only through the pushforward* $\pi_\#\lambda_t$*. By applying Lemma 7.7 in the space* S^Π*, we can use the pushforwards* $\pi_\#\lambda_t^N$ *as stochastic proxies to* $\pi_\#\lambda_t$*, and thanks to Lemma 7.10, the measures* $\pi_\#\lambda_t^N$ *can be realised as* $\pi_\star(G_t^N)$ *for a random graph process* G_t^N*. In this way, we can apply results from the theory of random graphs [28] to deduce results about solutions* (λ_t) *to the Smoluchowski equation (Fl).*

Further, the new kernel k *here represents the rate of merger of an* unordered *pair* $\{x, y\}$*, since there is only one edge: "*x *merges with* y*" is the same as "*y *merges with* x*". The (graph) kernel* k *therefore has to be the same merger rate as for the stochastic particle system, forcing the inclusion of the factor* 2.

Sketch of proof of Lemma 7.10. Let us fix $\mathbf{x}_N$. Firstly, both processes are Markov: for $\pi_\#\lambda_t^N$, the follows because the total rate (7.6) depends only on $\pi(x), \pi(y)$, and similarly for $\pi_\star(G_t^N)$. One may also verify that the two processes undergo the same transitions at the same rates, again thanks to (7.6), and that the total rate is bounded in terms of $\mathbf{x}_N$. The boundedness of the total rate implies the uniqueness in law for the corresponding Markov generator, which concludes the proof. □

Combining this with the approximation result Lemma 7.7 for the stochastic coagulant, we may connect the random graph process to the limit equation as follows.

Lemma 7.12 (Convergence of the Random Graphs)**.** *Let* λ_0 *be a measure on* S *satisfying Hypothesis 7.1, and let* $(G_t^N)_{t\geq 0}$ *be the random graph processes constructed above with initial data* $\mathbf{x}_N = (x_1, ...x_{l^N})$ *which satisfies Hypothesis 7.2. Let* $(\lambda_t)_{t\geq 0}$ *be the solution to the Smoluchowski Equation (Fl) starting at* λ_0*; then we have the local uniform convergence*

$$\sup_{t\leq t_\mathrm{f}} \rho_{1,\Pi}(\pi_\star(G_t^N), \pi_\#\lambda_t) \to 0 \tag{7.57}$$

in probability, for all $t_\mathrm{f} < \infty$*, where* $\rho_{1,\Pi}$ *is defined as for* ρ_1 *on the projected state space* $S^\Pi \subset \mathbb{R}^{n+m+1}$.

We can also compute the critical time associated to G_t^N explicitly:

Lemma 7.13 (Computation of critical time)**.** *Let* λ_0 *be a measure satisfying Hypothesis 7.1, and let* G_t^N *be random graphs satisfying Hypothesis 7.2. Then the convolution operator* T *constructed above is a bounded linear map on* $L^2(\lambda_0)$ *and the inverse of the critical time* *for the graph phase transition, t_c^{-1}, is the largest eigenvalue of the $n \times n$ matrix $\mathcal{Z}(\lambda_0)$ given by $\mathcal{Z}(\lambda_0)_{ij} = \langle (A\pi)_i \pi_j, \lambda_0 \rangle$. In particular, $t_c \in (0, \infty)$.*

Remark 7.14. *This is exactly the form claimed for t_g in Theorem 7.2. However, we have not yet established that $t_c = t_g$; this is the content of Lemma 7.16.*

Proof of Lemma 7.13. Firstly, by (A2.), it is easy to see that $k \in L^2(S \times S, \lambda_0 \times \lambda_0)$, and so, by Lemma 7.47, $\|T\| = t_c^{-1}$ is the largest eigenvalue of T; its eigenspace is one-dimensional and consists of functions that are single signed, λ_0- almost everywhere. Since $0 < \|T\| < \infty$ we have $0 < t_c < \infty$.

In order to reduce from the operator T to the matrix $\mathcal{Z}(\lambda_0)$ we construct a basis $\{e_i\}_{i \geq 1}$ of $L^2(\lambda_0)$ such that

$$e_i(x) = \pi_i(x), \quad i = 1, 2, ..n + m \tag{7.58}$$

and, for $i > n + m$, e_i is orthogonal to $E = \text{Span}(e_1, ..., e_{n+m})$. Note that π_0 plays no special rôle in the basis, because it does not appear in the rate $\overline{K}$. We also write $E_+ = \text{Span}(e_1, ...e_n)$ and $E_{\text{Sym}} = \text{Span}(e_{n+1}, ..., e_{n+m})$. By expanding the total rate $\overline{K}(x, y)$, we see that, for all $f \in L^2(\lambda_0)$,

$$(Tf)(x) = 2 \sum_{i,j=1}^{n+m} a_{ij} \langle f, \pi_i \rangle_{L^2(\lambda_0)} \pi_j(x) \tag{7.59}$$

where $\langle \cdot, \cdot \rangle_{L^2(\lambda_0)}$ denotes the $L^2(\lambda_0)$ inner product, and the factor of 2 comes from the definition (7.51). Therefore, T maps into the subspace E, and is 0 on its orthogonal complement. We further note that the subspaces E_+, E_{sym} are orthogonal, and are invariant under T. Therefore, the eigenspace E^λ corresponding to $\lambda = t_c^{-1}$ is a direct sum $E_+^\lambda \oplus E_{\text{sym}}^\lambda$ of eigenspaces contained within E_+, E_{sym}.

Since E^λ is one-dimensional, one summand must be trivial, and so either $E^\lambda = E_+^\lambda \subseteq E_+$, or $E^\lambda \subseteq E_{\text{sym}}$. To exclude the second possibility, we note that any $f \in E_{\text{sym}}$ satisfies $f(Rx) = -f(x)$ for all x by Definition 7.1.1, while eigenfunctions of T are single-signed λ_0-almost everywhere. It therefore follows that $E^\lambda \subseteq E_+$ and that t_c^{-1} is the largest eigenvalue of $T|_{E_+}$.

The result is now immediate since (7.59) shows that $\mathcal{Z}(\lambda_0)$ is the matrix representation of $T|_{E_+}$ respect to the basis introduced above. $\square$

We also define κ_t as the survival function from Lemma 7.44, given by the maximal solution to

$$\kappa_t(x) = 1 - \exp\left(-t(T\kappa_t)(x)\right). \tag{7.60}$$

We note, for future use, the following properties where k is the kernel given above.

Lemma 7.15. *The survival function $\kappa_t(v) = \rho(tk, x)$ takes the form*

$$\kappa_t(x) = 1 - \exp\left(-\sum_{i=1}^{n} c_t^i \pi_i(x)\right) \tag{7.61}$$

for some $c_t^i \geq 0$. Moreover, the functions $t \mapsto c_t^i$ are continuous.

This proves the first two assertions of item 4 of Theorem 7.2.

Proof. Using the symmetry $k(Rx, Ry) = k(x, y)$ and Hypothesis (A1.), it is simple to verify that $\tilde{\rho}(x) := \kappa_t(Rx)$ also satisfies the fixed point equation (7.60). By maximality of κ_t, we must have $\kappa_t(Rx) \leq \kappa_t(x)$ for all $x \in S$, which implies that κ_t is even under R.

Using the identification of the range of T as in Lemma 7.13, we see that there exist $c_t^i : 1 \leq i \leq n+m$ such that

$$t(T\kappa_t)(x) = \sum_{i=1}^{n+m} c_t^i \pi_i(x) \tag{7.62}$$

and expanding k as in (7.59), the coefficients are given explicitly by

$$c_t^i = 2t \sum_{j=1}^{n} a_{ij} \langle \pi_j \kappa_t, \lambda_0 \rangle. \tag{7.63}$$

Since κ_t is even, we have $c_t^i = 0$ for $i > n$, and since $\kappa_t \geq 0$, $c_t^i \geq 0$ for $i = 1, ..., n$. Using (7.60) again, we obtain the claimed representation

$$\kappa_t(x) = 1 - \exp\left(-\sum_{i=1}^{n} c_t^i \pi_i(x)\right). \tag{7.64}$$

The continuity follows by applying dominated convergence to (7.63), and using the continuity of κ_t established in Theorem 7.49. ☐

7.6 Equality of the Critical Times

In this section, we will prove that the critical time t_c for the graph process, introduced in Section 7.5, coincides with the gelation time for the limiting equation, defined in Section 7.3 as the time at which mass and energy begin to escape to infinity.

Lemma 7.16. *Let λ_0 be a measure on S satisfying Hypothesis 7.1. Let $(\lambda_t)_{t\geq 0}$ be the solution to (Fl) starting at λ_0, with associated data M_t, E_t of the gel; recall that t_g is defined by*

$$t_g := \inf\{t \geq 0 : \langle \varphi, \lambda_t \rangle < \langle \varphi, \lambda_0 \rangle\} = \inf\{t \geq 0 : g_t \neq 0\}. \tag{7.65}$$

Let (G_t^N) be the random graph processes constructed above, and suppose that Hypothesis 7.2 holds for G_0^N, λ_0. Then the critical time t_c for the graph transition process coincides with the gelation time t_g.

The following is a straightforward corollary.

Corollary 7.17. *Let λ_0 satisfy Hypothesis 7.1, and let $(\lambda_t)_{t\geq 0}$ be the solution to (Fl) starting at λ_0, with gelation at t_{g}. Then t_{g} is given explicitly by (7.17).*

Proof of Corollary 7.17. Let us form $\mathbf{x}_N$ by sampling points as a Poisson random measure with intensity $N\lambda_0$. It is immediate that the resulting data $\mathbf{x}_N$ satisfies Hypothesis 7.2 for the measure λ_0, and the critical time t_{c} of the associated graphs G^N_t is given by the claimed expression (7.17). From the previous lemma, it now follows that the gelation time $t_{\mathrm{g}} = t_{\mathrm{c}}$, which proves the claimed result. □

The proof of Lemma 7.16 is based on the following weak version of the convergence of the gel in Theorem 7.3, which will be revisited in Section 7.11 to establish uniform convergence.

Lemma 7.18. *Let $(\lambda_t)_{t\geq 0}, M_t, E_t$ be as in Lemma 7.16 and G^N_t be as in the proof of Corollary 7.17. Fix $t > 0$, and write g^N_t for the scaled mass and energy of the largest particle in G^N_t, as in Section 7.5:*

$$g^N_t = \frac{1}{N}\,\pi(\mathcal{C}_1(G^N_t)) = \left(\frac{1}{N}\sum_{i\in\mathcal{C}_1(G^N_t)}\pi_j(x_i)\right)_{j=0}^{n+m} = (M^N_t, E^N_t, P^N_t). \tag{7.66}$$

Then $M^N_t \to M_t$ and $E^N_t \to E_t$ in probability.

We first show that Lemma 7.18 implies Lemma 7.16; the remainder of this section is dedicated to the proof of Lemma 7.18.

Proof of Lemma 7.16. Let us assume, for the moment, that Lemma 7.18 holds. Throughout, let $(x_i)_{i=1}^{l^N}$ be the vertex data of the random graph process, which we recall are independent of time.

Firstly, suppose for a contradiction that $t_{\mathrm{g}} < t_{\mathrm{c}}$. Then $\varphi(g_{t_{\mathrm{c}}}) > 0$, but we bound

$$\varphi(g^N_{t_{\mathrm{c}}}) \leq \left(\frac{1}{N}C_1(G^N_{t_{\mathrm{c}}})\right)^{\frac{1}{2}}\left(\frac{1}{N}\sum_{i=1}^{l^N}\varphi(x_i)^2\right)^{\frac{1}{2}}. \tag{7.67}$$

The first term converges to 0 in probability, by definition of the phase transition in Theorem 7.45, and the second term is bounded in $L^2(\mathbb{P})$ by hypothesis (B2.). This implies that $\varphi(g^N_{t_{\mathrm{c}}}) \to 0$ in probability, which contradicts Lemma 7.18; we must therefore have that $t_{\mathrm{g}} \geq t_{\mathrm{c}}$.

Conversely, if $t < t_{\mathrm{g}}$, then $M_t = 0$ by definition. Now, the convergence

$$\frac{1}{N}C_1(G^N_t) = M^N_t \to 0 \tag{7.68}$$

in probability implies that the largest cluster is of the order $o_{\mathrm{p}}(N)$, which is only possible if $t \le t_{\mathrm{c}}$ by Lemma 7.9. Since $t < t_{\mathrm{g}}$ was arbitrary, we must have $t_{\mathrm{g}} \le t_{\mathrm{c}}$, and together with the previous argument, we have shown that $t_{\mathrm{g}} = t_{\mathrm{c}}$ as claimed. □

The proof of Lemma 7.18 is based on the following argument. We know, from Theorem 7.48, that any 'mesoscopic' clusters contain negligable mass; thanks to the integrability assumption (A2.), the same is true for the energy. Therefore, almost all mass and energy either belongs to the 'microscopic' scale, whose convergence is quantified by Lemma 7.7, or the giant component, whose convergence is the subject of interest here. Therefore, with a suitable approximation argument, the claimed convergence will follow from the quoted results.

We begin with a preparatory lemma; throughout, we will assume the notation of Lemma 7.18. For the proof of of Lemma 7.16, and later Theorem 7.3, we will wish to study the convergence of integrals $\langle \varphi f, \lambda^N_t\rangle$, for functions f which do not belong to $\mathcal{A}$. However, the convergence result Lemma 7.7 only gives us information about weak convergence, which requires bounded and continuous test functions. Our second preparatory lemma allows us to approximate the integrals $\langle \varphi f, \lambda^N_t\rangle$ for functions f whose support is bounded in the π_0-direction.

Lemma 7.19 (A step towards uniform integrability). *Let λ_0, $(\lambda^N_t)_{t\ge 0}$ be as in the previous lemma. Then, for every $r > 0$,*

$$\beta(r,\eta) := \sup_{N\ge 1} \mathbb{E}\left[\sup_{t\ge 0}\left\langle \varphi 1\!\!1[\varphi(x) > \eta, \pi_0(x) \le r], \lambda^N_t\right\rangle\right] \to 0 \qquad \textit{as } \eta\to\infty. \tag{7.69}$$

Proof. We note that $\langle \varphi 1\!\!1[\varphi(x) > \eta, \pi_0(x) \le r], \lambda^N_t\rangle$ depends on λ^N_t only through the pushforward $\pi_\# \lambda^N_t$, since the integrand only depends on the values of π at the different particles. From Lemma 7.10, we can find random graphs G^N_t, such that $\mathbf{x}_N$ is an enumeration of the atoms of λ^N_0 and $\pi_\star(G^N_t) = \pi_\#\lambda^N_t$ for all times t. With this coupling, we express the integral as follows:

$$\begin{aligned}\langle \varphi 1\!\!1[\varphi(x) > \eta, \pi_0(x) \le r], \lambda^N_t\rangle &= \frac{1}{N}\sum_{\text{Clusters } \mathcal{C}\subset G^N_t} \varphi(\mathcal{C}) 1\!\!1[\varphi(\mathcal{C}) > \eta, \pi_0(\mathcal{C}) \le r] \\ &= \frac{1}{N}\sum_{j=1}^{l^N(t)} \sum_{i\in\mathcal{C}_j(G^N_t)} \varphi(x_i) 1\!\!1\left[\varphi(\mathcal{C}_j(G^N_t)) > \eta, \pi_0(\mathcal{C}_j(G^N_t)) \le r\right].\end{aligned} \tag{7.70}$$

Using Cauchy-Schwarz, we bound

$$\begin{aligned}&\sup_{t\geq 0}\left\langle \varphi \mathbb{1}[\varphi(x)>\eta, \pi_0(x)\leq r], \lambda^N_t\right\rangle \\ &\quad\leq \left(\frac{1}{N}\sum_{j=1}^{l^N(0)}\varphi(x_j)^2\right)^{\frac{1}{2}}\left(\sup_{t\geq 0}\frac{1}{N}\sum_{j=1}^{l^N(t)}\sum_{i\in\mathcal{C}_j(G^N_t)}\mathbb{1}\left[\varphi(\mathcal{C}_j(G^N_t))>\eta, C_j(G^N_t)\leq r\right]\right)^{\frac{1}{2}} \\ &\quad= \left(\frac{1}{N}\sum_{i=1}^{l^N(0)}\varphi(x_i)^2\right)^{\frac{1}{2}}\left(\sup_{t\geq 0}\left\langle \pi_0\mathbb{1}[\varphi(x)>\eta, \pi_0(x)\leq r], \lambda^N_t\right\rangle\right)^{\frac{1}{2}}.\end{aligned} \tag{7.71}$$

As remarked in Definition 7.5.1, the data x_i associated with the graph nodes are constant in time, so the first factor is independent of $t\geq 0$, and is bounded in $L^2(\mathbb{P})$ by the second assertion of (B2.). Therefore, it is sufficient to prove the claim with φ replaced by π_0.

Now we note that with probability one

$$\sup_{t\geq 0}\left\langle \pi_0\mathbb{1}[\varphi(x)>\eta, \pi_0(x)\leq r], \lambda^N_t\right\rangle \leq r\sup_{t\geq 0}\left\langle \mathbb{1}[\varphi(x)>\eta], \lambda^N_t\right\rangle \leq \frac{r}{\eta}\sup_{t\geq 0}\left\langle \varphi, \lambda^N_t\right\rangle = \frac{r}{\eta}\left\langle \varphi, \lambda^N_0\right\rangle$$

and the result follows from (B2.). □

Using the preparatory lemma developed above, we now prove Lemma 7.18.

Proof of Lemma 7.18. Throughout, we let $(\lambda^N_t)_{t\geq 0}$ be a stochastic coagulant coupled to a random graph process $(G^N_t)_{t\geq 0}$, as described in Section 7.5 with vertex data $\mathbf{x}_N=(x_i)_{i=1}^{l^N(0)}$; thanks to this construction, M^N_t is exactly the size of the largest cluster in G^N_t, and E^N_t are the sums

$$E^N_t = \left(N^{-1}\sum_{j\in\mathcal{C}_1(G^N_t)}\pi_i(x_j)\right)_{i=1}^{n}. \tag{7.72}$$

The case $t=0$ is trivial, and can be omitted. We deal first with the 0^{th} coordinate M^N_t; the cases for the $1^{\text{st}}, ..., n^{\text{th}}$ coordinates E^N_t are entirely analagous.

Fix $t>0$, and let ξ_N be a sequence, to be constructed later, such that

$$\xi_N\to\infty; \qquad \frac{\xi_N}{N}\to 0. \tag{7.73}$$

We now construct 'bump functions' as follows. Let $\eta_r\to\infty$ be a sequence growing sufficiently fast that, in the notation of Lemma 7.19, $\beta(r,\eta_r)\to 0$, and let

$$S_{(r)} := \{x\in S: \pi_0(x)<r, \varphi(x)\leq\eta_r\}. \tag{7.74}$$

Let $\widetilde{h}_r$ be the indicator $\widetilde{h}_r=\mathbb{1}[\pi_0(x)<r]$, and construct a continuous function $\widetilde{f}_r$ such that

$$0\leq\widetilde{f}_r\leq 1; \qquad \widetilde{f}_r=1 \text{ on } S_{(r)}; \qquad \widetilde{f}_r(x)=0 \text{ if } \pi_0(x)\geq r. \tag{7.75}$$

The final condition is compatible with continuity because $\pi_0 : S \to \mathbb{N}$ is continuous and integer valued. We define $f_N = \widetilde{f}_{\xi_N}$ and $h_N = \widetilde{h}_{\xi_N}$. We now decompose the difference $M^N_t - M_t$:

$$\begin{aligned} M^N_t - M_t = &\underbrace{(\langle \pi_0, \lambda_t\rangle - \langle \pi_0 f_N, \lambda_t\rangle)}_{:=\mathcal{T}^1_N} + \underbrace{\langle \pi_0 f_N, \lambda_t - \lambda^N_t\rangle}_{:=\mathcal{T}^2_N} \\ &+ \underbrace{\langle \pi_0(f_N - h_N), \lambda^N_t\rangle}_{:=\mathcal{T}^3_N} + \underbrace{\langle \pi_0 h_N, \lambda^N_t\rangle - (\langle \pi_0, \lambda^N_0\rangle - M^N_t)}_{:=\mathcal{T}^4_N} \\ &+ \underbrace{\langle \pi_0, \lambda^N_0 - \lambda_0\rangle}_{:=\mathcal{T}^5_N}. \end{aligned} \tag{7.76}$$

where we recall that $M_t = \langle \pi_0, \lambda_0 - \lambda_t\rangle$. We now estimate the errors $\mathcal{T}^i_N$, $i = 1, 3, 4, 5$; the remaining term $\mathcal{T}^2_N$ will be dealt with separately, and requires careful construction of the sequence ξ_N.

Step 1. Estimate on $\mathcal{T}^1_N$. Let z_N be the lower bound $z_N = 1\!\!1_{S_{(\xi_N)}}$, so that $z_N \leq f_N \leq 1$. As $N \to \infty$, $\pi_0 z_N \uparrow \pi_0$, and so by monotone convergence, $\langle \pi_0 z_N, \lambda_t\rangle \uparrow \langle \pi_0, \lambda_t\rangle$. This implies that the (nonrandom) error $\mathcal{T}^1_N \to 0$.

Step 2. Estimate on $\mathcal{T}^3_N$. From the definitions of f_N, h_N, we observe that

$$|\mathcal{T}^3_N(t)| = \langle \pi_0(h_N - f_N), \lambda^N_t\rangle \leq \langle \pi_0 1\!\!1[\pi_0(x) < \xi_N, \varphi(x) > \eta_{\xi_N}], \lambda^N_t\rangle. \tag{7.77}$$

Therefore, in the notation of Lemma 7.19, $\mathbb{E}\left[\sup_{t\geq 0} |\mathcal{T}^3_N(t)|\right] \leq \beta(\xi_N, \eta_{\xi_N})$. By construction of η_r, and since $\xi_N \to \infty$, it follows that $\mathbb{E}[\sup_{t\geq 0} |\mathcal{T}^3_N(t)|] \to 0$, which implies convergence to 0 in probability.

Step 3. Estimate on $\mathcal{T}^4_N$. Recalling that $h_N(x) = 1\!\!1[\pi_0(x) < \xi_N]$ and using the coupling to random graphs, we have the equality

$$\langle \pi_0 h_N, \lambda^N_t\rangle = \langle \pi_0, \lambda^N_0\rangle - M^N_t 1\!\!1\left[M^N_t \geq \frac{\xi_N}{N}\right] - \frac{1}{N} \sum_{j\geq 2: C_j(G^N_t)\geq \xi_N} \; \sum_{i\in C_j(G^N_t)} \pi_0(x_i) \tag{7.78}$$

which gives the equality

$$\mathcal{T}^4_N = -M^N_t 1\!\!1\left(M^N_t < \frac{\xi_N}{N}\right) - \frac{1}{N} \sum_{j\geq 2: C_j(G^N_t)\geq \xi_N} \pi_0(\mathcal{C}_j(G^N_t)). \tag{7.79}$$

Using Cauchy-Schwarz, we bound

$$\left|\mathcal{T}^4_N(t)\right| \leq \left(\frac{1}{N} \sum_{j\geq 2: C_j(G^N_t)\geq \xi_N} C_j(G^N_t)\right)^{\frac{1}{2}} \left(\frac{1}{N} \sum_{i=1}^{l^N(0)} \varphi(x_i)^2\right)^{\frac{1}{2}} + \frac{\xi_N}{N}. \tag{7.80}$$

The first term converges to 0 in probability by Theorem 7.48 and (B2.), and the second converges to 0 since $\xi_N \ll N$. Together, these imply that $\mathcal{T}^4_N(t) \to 0$ in probability.

Step 4. Estimate on $\mathcal{T}^5_N$. Using the first part of (B2.), we have the convergence in distribution

$$\langle \pi_0, \lambda_0^N \rangle \to \langle \pi_0, \lambda_0 \rangle \tag{7.81}$$

which implies that $\mathcal{T}^5_N \to 0$ in probability as desired.

Step 5. Construction of ξ_N, and convergence of $\mathcal{T}^2_N$. It remains to show how a sequence ξ_N can be constructed such that $\mathcal{T}^2_N \to 0$ in probability and such that (7.73) holds. Recalling the definition of $\tilde{f}_r$ above, let $A^1_{r,N}$ be the events $A^1_{r,N} = \{|\langle \varphi \widetilde{f}_r, \lambda^N_t - \lambda_t\rangle| < \frac{1}{r}\}$; as $N \to \infty$ with r fixed, both $\mathbb{P}(A^1_{r,N}) \to 1$, by Lemma 7.7. We now define N_r inductively for $r \geq 1$ by setting $N_1 = 1$, and letting N_{r+1} be the minimal $N > \max(N_r, (r+1)^2)$ such that, for all $N' \geq N$, $\mathbb{P}(A^1_{r+1,N'}) > \frac{r}{r+1}$. Now, we set $\xi_N = r$ for $N \in [N_r, N_{r+1})$. It follows that $\xi_N \to \infty$ and $\xi_N \leq \sqrt{N} \ll N$, and

$$\mathbb{P}\left(C_1(G^N_t)) \geq \xi_N\right) \geq 1 - \frac{1}{\xi_N} \to 1. \tag{7.82}$$

Therefore, ξ_N satisfies the requirements (7.73) above. Moreover,

$$\mathbb{P}\left(|\mathcal{T}^2_N| < \frac{1}{\xi_N}\right) \geq \mathbb{P}\left(A^1_{\xi_N,N}\right) > 1 - \frac{1}{\xi_N} \to 1 \tag{7.83}$$

and so, with this choice of ξ_N, $\mathcal{T}^2_N \to 0$ in probability. Since we have now dealt with every term appearing in the decomposition (7.76), it follows that $M^N_t \to M_t$ in probability, as claimed.

The arguments for the $1^{\text{st}} - n^{\text{th}}$ components E^N_t are identical to those above, using the same bound (7.80) on $\mathcal{T}^4_N$. ☐

We also note, for future use, an important corollary of this argument.

Corollary 7.20. *At the instant of gelation, the gel is negligible:* $g_{t_{\mathrm{g}}} = 0$.

Proof. For the $0^{\text{th}} - n^{\text{th}}$ components, this follows from the critical case of Theorem 7.45 exactly as in (7.67). The remaining m components $g^i_t, i > n$ are identically 0 by the symmetry (A1.), as in Lemma 7.4. ☐

7.7 Behaviour of the Second Moments

In this section, we consider part 2 of Theorem 7.2, concerning the behaviour of the second moments $\mathcal{Q}(t)_{ij} = \langle \pi_i \pi_j, \lambda_t \rangle, 0 \leq i,j \leq n$ and $\mathcal{E}(t) = \langle \varphi^2, \lambda_t \rangle$. Following [134, 156], one might expect that the gelation time t_{g} corresponds to a divergence of $\mathcal{E}(t)$ as $t \uparrow t_{\mathrm{g}}$; by an approximation argument, we will show that this is indeed the case. We also introduce a *duality argument*, corresponding to Theorem 7.50, which allows us to prove that $\mathcal{E}$ is finite on (t_{g}, ∞). The final assertion follows from the fact that $g_{t_{\mathrm{g}}} = 0$, which is the content of Corollary 7.20.

7.7.1 Subcritical Regime

We first deal with the subcritical regime $[0, t_{\mathrm{g}})$, to show that the second moments $\mathcal{Q}_{ij}(t), \mathcal{E}(t)$ are finite and increasing on this interval, and that t_{g} is exactly the first time at which $\mathcal{E}$ diverges.

Lemma 7.21. *Suppose λ_0 satisfies Hypothesis 7.1, and let $(\lambda_t)_{t\geq 0}$ be the corresponding solution to (Fl). The second moments $\mathcal{Q}(t)_{ij} = \langle \pi_i\pi_j, \lambda_t\rangle, 0 \leq i, j \leq n$, $\mathcal{E}(t) = \langle \varphi^2, \lambda_t\rangle$ are finite, continuous and increasing on $[0, t_{\mathrm{g}})$, and $\mathcal{E}(t) = \langle \varphi^2, \lambda_t\rangle$ increases to infinity as $t \uparrow t_{\mathrm{g}}$, where t_{g} is the associated gelation time.*

The ideas of this argument follow [156], where there is a similar result for *approximately multiplicative* kernels, for which the total rate $\overline{K}(x, y)$ is bounded above *and below* by nonzero multiples of $\widetilde{\varphi}(x)\widetilde{\varphi}(y)$, where $\widetilde{\varphi}$ is a mass function playing the same rôle as our φ. Unfortunately, this cannot be applied directly, for two reasons.

i). Firstly, the total rate in (7.6) contains the terms $a_{ij}\pi_i(x)\pi_j(y), n \leq i, j \leq n + m$ of indefinite sign.

ii). Secondly, the remaining combination of $\pi_i, 1 \leq i \leq n$ is not *a priori* of approximately multiplicative form: particles where some π_i are small, and others large, will in general prevent such a bound from holding.

Our strategy will be as follows.

1. Firstly, we will show that if $(\lambda_t)_{t\geq 0}$ solves (Fl), then the pushforward measures $(\pi_\#\lambda_t)_{t<t_{\mathrm{g}}(\lambda_0)}$ solve a modified equation (mΠFl) on the simpler space $S^\Pi = \mathbb{N} \times [0, \infty)^n \times \mathbb{R}^m$, with a reduced kernel $K^{\Pi,\mathrm{m}}$. This allows us to eliminate the terms of indefinite sign mentioned above. This new equation has unique solutions, and so $\nu_t = \pi_\#\lambda_t$ is the unique solution starting at $\nu_0 = \pi_\#\lambda_0$; in particular, the second moments $\langle \varphi^2, \nu_t\rangle$, $\langle \varphi^2, \lambda_t\rangle$ coincide, and gelation takes place at the same time $t_{\mathrm{g}}(\lambda_0) = t_{\mathrm{g}}(\nu_0)$. Therefore, we can prove the desired result working solely at the level of (mΠFl).

2. Thanks to results of Norris [156, Theorem 2.1], if $(\nu_t)_{t\geq 0}$ is a solution to (mΠFl) with $\langle \varphi^2, \nu_0\rangle < \infty$, then there exists $t_{\mathrm{e}} = t_{\mathrm{e}}(\nu_0) > 0$ such that $\langle \varphi^2, \nu_t\rangle$ is locally integrable on $[0, t_{\mathrm{e}})$ and such that $\langle \varphi^2, \nu_t\rangle \uparrow \infty$ as $t \uparrow t_{\mathrm{e}}$.

3. We introduce a truncated state space S^Π_ϵ, which excludes particles where any $\pi_i/\pi_0, 1 \leq i \leq n$ is either very large or very small, and construct new initial data ν_0^ϵ which are supported in this space. In this context, the kernel $K^{\Pi,\mathrm{m}}$ is approximately multiplicative, and so [156, Theorem 2.2] guarantees that the solutions $(\nu_t^\epsilon)_{t\geq 0}$ undergo gelation at exactly the blow-up time $t_{\mathrm{e}}(\nu_0^\epsilon)$.

4. We argue, from the characterisation of the gelation time in Section 7.5, that our construction gives an approximation of the gelation times: $t_\mathrm{g}(\nu_0^\epsilon) \to t_\mathrm{g}(\nu_0)$. We will argue, based on a system of ordinary differential equations for the moments $\langle \pi_i\pi_j, \nu_t\rangle = \langle \pi_i\pi_j, \lambda_t\rangle$, that the blowup time is also continuous: $t_\mathrm{e}(\nu_0^\epsilon) \to t_\mathrm{e}(\nu_0)$. Together with the previous points, this proves the claimed result.

We begin by introducing the modified equation.

Lemma 7.22. *Let K^Π be the kernel on $S^\Pi = \mathbb{N} \times [0,\infty)^n \times \mathbb{R}^m$ given by $K^\Pi(p,q,dr) = \sum_{i,j=1}^{n+m} a_{ij}p_iq_j\delta_{p+q}(dr)$, and let $K^{\Pi,\mathrm{m}}$ be the symmetrised kernel*

$$K^{\Pi,\mathrm{m}}(p,q,dr) = \frac{1}{4}K^\Pi(R^\Pi p, q, dr) + \frac{1}{2}K^\Pi(p,q,dr) + \frac{1}{4}K^\Pi(p, R^\Pi q, dr) \tag{7.84}$$

where we recall that $R^\Pi : S^\Pi \to S^\Pi$ is given by (7.15). Consider the corresponding equation incorporating gel, for measures on S^Π, which we write as

$$\nu_t = \nu_0 + \int_0^t L^\mathrm{m}_\mathrm{g}(\nu_s)ds. \tag{mΠFl}$$

Let λ_0 be a measure on S satisfying Hypothesis 7.1, and let $(\lambda_t)_{t\ge 0}$ be the corresponding solution to (Fl). Then the pushforward measures $\nu_t = \pi_\#\lambda_t$ are the unique solution to (mΠFl) starting at $\nu_0 = \pi_\#\lambda_0$.

Remark 7.23. *Under the new kernel $K^{\Pi,\mathrm{m}}$, the quantities π_i are still conserved for $0 \le i \le n$, but not for $n+1 \le i \le n+m$. However, since we seek to analyse $\langle \varphi^2, \lambda_t\rangle, \varphi = \sum_{0\le i\le n}\pi_i$, we will not need any conservation properties of π_i for $i > n$ in this section.*

Sketch Proof of Lemma 7.22. Much of the proof consists of algebraic manipulations, using the definitions and hypotheses in Definition 7.1.1. In the interest of brevity, such manipulations will omitted.

Let us first consider the reflected measures $R_\#\lambda_t = \lambda_t \circ R^{-1}$ on S. By (A1.), $R_\#\lambda_0 = \lambda_0$, and using part iii) of Definition 7.1.1, one can show that for all $t \ge 0$, all finite measures λ on S and all bounded, measurable functions f on S, $\langle f\circ R, L(\lambda)\rangle = \langle f, L(R_\#\lambda)\rangle$. From this, and performing a similar manipulation for the gel term, it follows that $(R_\#\lambda_t)_{t\ge0}$ also solves the equation (Fl) with the same initial data which implies, by uniqueness, we must have $\lambda_t = R_\#\lambda_t$ for all $t \ge 0$. Using this, one can now similarly prove that, for all t and f as above,

$$\begin{aligned}\langle f, L(\lambda_t)\rangle &= \int_{S^3}(f(z)-f(x)-f(y))K(Rx,y,dz)\lambda_t(dx)\lambda_t(dy)\\ &= \int_{S^3}(f(z)-f(x)-f(y))K(x,Ry,dz)\lambda_t(dx)\lambda_t(dy).\end{aligned} \tag{7.85}$$

Taking a linear combination, and again performing a similar manipulation for the gel term, it follows that λ_t solves the equation analagous to (Fl) for the symmetrised kernel

$$K^{\mathrm{Sym}}(x,y,dz) = \frac{1}{4}K(Rx,y,dz) + \frac{1}{2}K(x,y,dz) + \frac{1}{4}K(x,Ry,dz). \tag{7.86}$$

Since the coagulation rate $K^{\mathrm{Sym}}(x, y, S)$ only depends on $\pi(x), \pi(y)$, one can verify that the pushforward measures $\pi_{\#}\lambda_t$ on S^{Π} solve the projected equation (mΠFl) as claimed. □

We now turn to the second point, which concerns the moment behaviour of the solutions to (mΠFl). The following result follows from ideas of [156], which we will briefly sketch.

Lemma 7.24. *Let ν_0 be a measure on S^{Π} satisfying Hypothesis 7.1, and let $(\nu_t)_{t\geq 0}$ be the corresponding solution to (mΠFl). Then there exists $t_{\mathrm{e}} = t_{\mathrm{e}}(\nu_0) > 0$ such that $t \mapsto \langle \varphi^2, \nu_t\rangle$ is finite and increasing on $[0, t_{\mathrm{e}})$, and $\langle \varphi^2, \nu_t\rangle \uparrow \infty$ as $t \uparrow t_{\mathrm{e}}$. Moreover, $(\nu_t)_{t<t_{\mathrm{e}}}$ is conservative, and so $t_{\mathrm{e}}(\nu_0) \leq t_{\mathrm{g}}(\nu_0)$.*

The subscript $_{\mathrm{e}}$ here denotes 'explosion': t_{e} is the first time the second moment diverges to ∞.

Sketch Proof of Lemma 7.24. This argument applies different results from [156] to our case. We say that a local solution $(\nu_t)_{t<T}$ to (mΠFl) is *strong* if the map $t \mapsto \langle \varphi^2, \nu_t\rangle$ is integrable on compact subsets of $[0, T)$. Applying the results of [156, Theorem 2.1], there exists a unique maximal strong solution $(\nu'_t)_{t<t_{\mathrm{e}}(\nu_0)}$ to (mΠFl), which is conservative and that $t_{\mathrm{e}}(\nu_0) \geq C\langle \varphi^2, \nu_0\rangle^{-1}$ for some constant C depending on A.

We next apply Corollary 7.5 to see that this solution must be an initial segment of $(\nu_t)_{t<t_{\mathrm{g}}(\nu_0)}$: that is, $t_{\mathrm{e}}(\nu_0) \leq t_{\mathrm{g}}(\nu_0)$, and $\nu'_t = \nu_t$ for all $t \leq t_{\mathrm{e}}(\nu_0)$. Therefore, the results of [156] will apply to our process $(\nu_t)_{t<t_{\mathrm{e}}(\nu_0)}$.

Since $(\nu_t)_{t<t_{\mathrm{e}}(\nu_0)}$ is conservative, we follow the ideas of [156, Proposition 2.7], to obtain the integral relations, for all $t < t_{\mathrm{e}}$ and $0 \leq i, j \leq n$,

$$\langle \pi_i\pi_j, \nu_t\rangle = \langle \pi_i\pi_j, \nu_0\rangle + 2\int_0^t \sum_{k,l=1}^n \langle \pi_i\pi_k, \nu_s\rangle a_{kl}\langle \pi_l\pi_j, \nu_s\rangle ds. \tag{7.87}$$

These immediately imply that $\langle \varphi^2, \nu_t\rangle$ is bounded on compact subsets of $[0, t_{\mathrm{e}})$, and in particular does not diverge before t_{e}. Moreover, since all terms on the right-hand side are nonnegative, these relations imply that all moments $\langle \pi_i\pi_j, \nu_t\rangle$ and $\langle \varphi^2, \nu_t\rangle$ are increasing on $[0, t_{\mathrm{e}})$.

Finally, we show that $\langle \varphi^2, \nu_t\rangle$ diverges near $t_{\mathrm{e}}(\nu_0)$. This follows from the time-of-existence estimate quoted above: for $t < t_{\mathrm{e}}$, the unique maximal strong solution starting at ν_t is precisely $(\nu_{s+t})_{s<t_{\mathrm{e}}-t}$, and so for some $C = C(A) < \infty$,

$$t_{\mathrm{e}} - t \geq C\langle \varphi^2, \nu_t\rangle^{-1}. \tag{7.88}$$

This rearranges to show that $\langle \varphi^2, \nu_t\rangle \geq C(t_{\mathrm{e}}-t)^{-1}$ which diverges as $t \uparrow t_{\mathrm{e}}$, as claimed. □

In order to obtain the full connection of the explosion and gelation times, we modify the setting to exclude the problematic particles in the remark above, which prevent $K^{\Pi,\mathrm{m}}$ from being approximately multiplicative. Let

$$S^{\Pi}_{\epsilon} = \{p \in S^{\Pi} : \epsilon\pi_0(p) \leq \pi_i(p) \leq (\epsilon^{-1} + \epsilon)\pi_0(p) \text{ for all } 1 \leq i \leq n\}. \tag{7.89}$$

Note that this state space is preserved under the kernel $K^{\Pi,\mathrm{m}}$. Moreover, on the reduced state space S^Π_ϵ, the modified kernel $K^{\Pi,\mathrm{m}}$ is *approximately multiplicative* [156] in the sense that, for some $\delta_\epsilon > 0$ and $\Delta_\epsilon < \infty$, we have

$$\delta_\epsilon\, \varphi(p)\varphi(q) \le \overline{K^{\Pi,\mathrm{m}}}(p,q) \le \Delta_\epsilon\, \varphi(p)\varphi(q) \tag{7.90}$$

for all $p, q \in S^\Pi_\epsilon$.

We now construct approximations ν_0^ϵ to ν_0 which are supported on S^Π_ϵ. Let us fix λ_0 satisfying Hypothesis 7.1 and $\nu_0 = \pi_\#\lambda_0$; for any $\epsilon > 0$, let ν_0^ϵ be given by specifying, for all bounded measurable functions h on S^Π,

$$\int_{S^\Pi} h(p)\nu_0^\epsilon(dp) \\ = \int_{S^\Pi} h(p_0, p_1+\epsilon, ...p_n+\epsilon, p_{n+1}, ...p_{n+m})\mathbb{1}[p_i \le \epsilon^{-1} \text{ for all } 1 \le i \le n]\nu_0(dp). \tag{7.91}$$

In this way, we shift ν_0 slightly away from the axes, while also truncating when any π_i becomes large. It follows, from existence and uniqueness, that the solution $(\nu_t^\epsilon)_{t\ge0}$ to (mΠFl) starting at ν_0^ϵ is supported on S^Π_ϵ for all $t \ge 0$. We can now apply [156, Theorem 2.2] to obtain the connection between gelation and explosion for these solutions:

Lemma 7.25. *Let $(\nu_t^\epsilon)_{t\ge0}$ be the solution to (mΠFl) starting at the measure ν_0^ϵ constructed above. Let $t_\mathrm{e}(\nu_0^\epsilon)$ be the explosion time of the second moment, as above, and $t_\mathrm{g}(\nu_0^\epsilon)$ the first time that ν_t^ϵ fails to be conservative. Then $t_\mathrm{e}(\nu_0^\epsilon) = t_\mathrm{g}(\nu_0^\epsilon)$.*

This then connects the gelation phenomenon to the blowup of the second moment, as desired, but only for the special case of the truncated and shifted initial distribution. We now seek to remove this restriction to obtain the result for the original measures λ_0, ν_0. To do this, we will show that $t_\mathrm{g}(\nu_0^\epsilon) \to t_\mathrm{g}(\nu_0)$ and $t_\mathrm{e}(\nu_0^\epsilon) \to t_\mathrm{e}(\nu_0)$ as we take $\epsilon \downarrow 0$.

Lemma 7.26 (Convergence of Gelation Times)**.** *Let ν_0, ν_0^ϵ be the measures constructed above, and $t_\mathrm{g}(\nu_0), t_\mathrm{g}(\nu_0^\epsilon)$ the corresponding gelation times. Then, as $\epsilon \downarrow 0$, $t_\mathrm{g}(\nu_0^\epsilon) \to t_\mathrm{g}(\nu_0)$.*

Proof. First, we recall that $\pi_1, ...\pi_n$ are linearly independent in $L^2(\lambda_0)$, and hence in $L^2(\nu_0)$, by hypothesis. Using the convergence $\langle \pi_i\pi_j, \nu_0^\epsilon\rangle \to \langle \pi_i\pi_j, \nu_0\rangle$, it follows that for $\epsilon > 0$ small enough, and any z with $\sum_i |z_i| = 1$, we have $\langle (\sum_i z_i\pi_i)^2, \nu_0^\epsilon\rangle > 0$. This, in turn, guarantees that $\pi_1, ...\pi_n$ are linearly independent in $L^2(\nu_0^\epsilon)$, for all $\epsilon > 0$ small enough.

We can now apply the explicit characterisation of t_g obtained in Lemma 7.13 for the measures ν_0^ϵ:

$$t_\mathrm{g}(\nu_0^\epsilon) = \sigma_1(\mathcal{Z}(\nu_0^\epsilon))^{-1} \tag{7.92}$$

where $\mathcal{Z}(\nu_0^\epsilon)$ is the matrix $\mathcal{Z}(\nu_0^\epsilon)_{ij} = 2\sum_{k=1}^{n+m}\langle \pi_i\pi_k, \nu_0^\epsilon\rangle a_{kj}, 1 \le i, j \le n$ and $\sigma_1(\cdot)$ here denotes the largest eigenvalue of a matrix. Moreover, as $\epsilon \downarrow 0$, the coefficients of the matrices $\mathcal{Z}(\nu_0^\epsilon)$ converge to the analagous matrix $\mathcal{Z}(\nu_0)$ for the measure ν_0.

It is well-known, following for instance from [197], that as the coefficients of a matrix vary continuously, so to do the associated eigenvalues, meaning that

$$\sigma_1(\mathcal{Z}(\nu_0^\epsilon)) \to \sigma_1(\mathcal{Z}(\nu_0)) \tag{7.93}$$

as $\epsilon \downarrow 0$. Combining this with the characterisation of t_g above, it follows that

$$\begin{aligned} t_\mathrm{g}(\nu_0^\epsilon) = \sigma_1(\mathcal{Z}(\nu_0^\epsilon))^{-1} &\to \sigma_1(\mathcal{Z}(\nu_0))^{-1} \\ &= t_\mathrm{g}(\nu_0) \end{aligned} \tag{7.94}$$

as desired. □

Finally, we show the same result for the explosion times. Thanks to Lemma 7.24 and (7.87), the matrix of second moments $q_{ij}(t) = \langle \pi_i\pi_j, \nu_t\rangle, 1 \le i, j \le n$ satisfies a closed system of differential equations, with locally Lipschitz coefficients, on $[0, t_\mathrm{e})$. We will now show that t_e is exactly the time of existence of a solution started at q_0.

Lemma 7.27. *Consider the ordinary differential equations*

$$\dot{q}_t = b(q_t); \qquad b(q) = 2qA^+q, \qquad q \in \mathrm{Mat}_n(\mathbb{R}); \tag{Q1}$$

$$\dot{z}_t = w(q_t)z_t, \qquad w : \mathrm{Mat}_n(\mathbb{R}) \to \mathrm{Mat}_{n+1}(\mathbb{R}) \textit{ linear}; \qquad z \in \mathbb{R}^{n+1} \tag{Q2}$$

where we recall in (Q1) that A^+ is the upper $n\times n$ block of the matrix A in (7.7). Then, for all $(z_0, q_0) \in \mathbb{R}^{n+1} \times \mathrm{Mat}_n(\mathbb{R})$, there exists a unique maximal solution $\chi(t, z_0, q_0), \psi(t, q_0)$ starting at (z_0, q_0), defined until the time $\zeta(q_0)$ where (Q1) blows up.

Then, for any measure ν_0 on S^Π, the time of existence is exactly the explosion time:

$$t_\mathrm{e}(\nu_0) = \zeta(q_0), \qquad (q_0)_{ij} = \langle \pi_i\pi_j, \nu_0\rangle, 1 \le i, j \le n. \tag{7.95}$$

Proof. Firstly, it is straightforward to verify that q_t does not depend on the initial data z_0, since (Q1) only depends on q; in particular, the blowup time ζ is a function only of q_0. It is also straightforward to verify that (Q2) cannot blow up before $\zeta(q_0)$, since on compact subsets $[0, t] \subset [0, \zeta(q_0))$, the coefficients of (Q2) are Lipschitz, uniformly over this time interval. As a result, the time of existence for the pair (Q1, Q2) is exactly the time of existence $\zeta(q_0)$, as claimed.

To link the explosion times t_e and the time of existence $\zeta(q_0)$, the equations (7.87) show that the matrix $q_{ij}(t) = \langle \pi_i\pi_j, \nu_t\rangle, 1 \le i, j \le n$ and the vector $z_t = \langle \pi_0\pi_i, \nu_t\rangle_{0\le i\le n}$ solve the system (Q1, Q2) on $0 \le t < t_\mathrm{e}(\nu_0)$, which implies that $t_\mathrm{e}(\nu_0) \le \zeta(q_0)$. For the converse, for $t < t_\mathrm{e}$, we have the equality

$$\chi(z_0, q_0)_0 + \sum_{i=1}^n \psi(q_0, t)_{ii} = \langle \pi_0^2, \nu_t\rangle + \sum_{i=1}^n \langle \pi_i^2, \nu_t\rangle \tag{7.96}$$

where the initial data are

$$q_0 = (\langle \pi_i\pi_j, \nu_0\rangle)_{i,j=1}^n \qquad z_0 = (\langle \pi_i\pi_0, \nu_0\rangle)_{0\le i\le n}. \tag{7.97}$$

The left hand side is bounded on compact subsets of $[0, \zeta(q_0))$ and the right-hand side dominates $\langle \varphi^2, \nu_t\rangle$ up to a constant C, which leads to a contradiction if we assume that $t_\mathrm{e} < \zeta(q_0)$, since $\langle \varphi^2, \nu_t\rangle \uparrow \infty$ as $t \uparrow t_\mathrm{e}(\nu_0)$. We therefore have $\zeta(q_0) \le t_\mathrm{e}(\nu_0)$ which proves the equality desired. □

We will now analyse the pair of equations presented above. This will prove the desired continuity of t_e, and some points which will be helpful for later reference.

Lemma 7.28. *Consider the differential equations (Q1, Q2) in the previous lemma, and the sets*

$$E = \mathrm{Mat}_n([0,\infty)); \qquad E_\delta = \{q \in E : \forall i, q_{ii} > \delta\}; \qquad E^\circ = \cup_{\delta>0} E_\delta; \tag{7.98}$$

$$E_\mathrm{cs} = \{q \in E : \textit{ for all } i,j \le n \textit{ and } t < \zeta(q),\ \psi(q,t)_{ij}^2 \le \psi(q,t)_{ii}\psi(q,t)_{jj}\}. \tag{7.99}$$

Then, if $q_0 \in E_\delta$, $(\psi(q_0,t))_{t<\zeta(q_0)} \subset E_\delta$, and similarly if $q_0 \in E_\mathrm{cs}$, then $(\psi(q_0,t))_{t<\zeta(q_0)} \subset E_\mathrm{cs}$. We have the following properties:

i). Let J_ϵ be the set

$$J_\epsilon = \{q \in E :\ \zeta(q) \ge \epsilon\}. \tag{7.100}$$

Then for all $\epsilon, \delta > 0$, the set $J_\epsilon \cap E_\delta \cap E_\mathrm{cs}$ is bounded.

ii). Suppose $q_0^\epsilon \in E_\mathrm{cs}, \epsilon > 0$ and $q_0^\epsilon \to q_0 \in E_\mathrm{cs} \cap E^\circ$ as $\varepsilon \to 0$. Then $\zeta(q_0^\epsilon) \to \zeta(q_0)$.

iii). Suppose $I \subset \mathbb{R}_+$ is an open interval, and the map $(z_0, q_0) : I \to \mathbb{R}^{n+1} \times (E_\mathrm{cs} \cap E^\circ)$ is continuous, and such that $t < \zeta(q_0(t))$ for all $t \ge 0$. Then the maps $t \mapsto \psi(q_0(t), t)$ and $t \mapsto \chi(z_0(t), q_0(t), t)$ are continuous on I.

Proof. i). Let us first fix $q \in E$. First of all write $a_\star = \min\{a_{ij} : a_{ij} > 0\}$ and let i, j be such that $a_{ij} > 0$. We now estimate

$$\frac{d}{dt}\psi(t,q)_{ij} \ge 2a_{ij}\psi(t,q)_{ij}^2 \ge 2a_\star q_{ij}^2. \tag{7.101}$$

This differential inequality may be integrated to obtain

$$\psi(t,q)_{ij} \ge \frac{q_{ij}}{1 - 2ta_\star q_{ij}}. \tag{7.102}$$

In particular, this gives the upper bound $\zeta(q) \le (2a_\star q_{ij})^{-1}$, which implies the claimed boundedness of J_ϵ in the $(i,j)^\mathrm{th}$ coordinate whenever $a_{ij} > 0$.

We will now extend this boundedness to all n^2 coordinates when we restrict to $q \in E_\delta \cap J_\epsilon \cap E_{\mathrm{cs}}$. Let M be the maximum diagonal entry of q:

$$M = \max_{1\le i\le n} q_{ii} \tag{7.103}$$

and fix i where this maximum is attained; by hypothesis on A, there exists $j \le n$ such that $a_{ij} \ge a_\star > 0$. It is straightforward to see that the derivative $\frac{d}{dt}\psi(t,q)_{ij}$ is increasing along the solution, which implies the estimate

$$\psi\left(\frac{\epsilon}{2}, q\right)_{ij} \ge \frac{\epsilon}{2} b(q)_{ij} = \epsilon \sum_{k,l\le n} q_{ik}a_{kl}q_{lj} \ge \epsilon q_{ii}a_{ij}q_{jj} \ge \epsilon\delta a_\star M. \tag{7.104}$$

By hypothesis, $\zeta(q) \ge \epsilon$, so $\zeta(\psi(\frac{\epsilon}{2}, q)) \ge \frac{\epsilon}{2}$. Applying the bound on ζ above, we find that

$$\frac{\epsilon}{2} \le \frac{1}{2a_\star^2 \epsilon\delta M}. \tag{7.105}$$

Finally, since we chose $q \in E_{\mathrm{cs}}$, we have the uniform bound

$$\max_{ij} q_{ij} \le M \le (a_\star^2\epsilon^2\delta)^{-1}. \tag{7.106}$$

ii). The lower semicontinuity of explosion times is standard, and follows from the continuous dependence on the initial data. Therefore, it is sufficient to prove that $\limsup_{\epsilon\to 0}\zeta(q^\epsilon) \le \zeta(q)$.

Suppose, for a contradiction, that for some $\eta > 0$, we have $\limsup_{\epsilon\to 0}\zeta(q^\epsilon) > \zeta(q) + \eta$; by passing to a subsequence, we may assume that $\zeta(q^\epsilon) > \tau + \eta$ for all ϵ, where we write $\tau = \zeta(q)$. Moreover, since $q_0^\epsilon \in E_{\mathrm{cs}}$ and $q^\epsilon \to q \in E^\circ$, we may assume that $q^\epsilon, q \in E_\delta \cap E_{\mathrm{cs}}$ for all ϵ, for some $\delta > 0$, which implies that $\psi(q^\epsilon, t) \in E_\delta \cap E_{\mathrm{cs}}$ for all $t < \zeta(q^\epsilon)$ and all $\epsilon > 0$.

Now, if $t \le \tau$, we have $\zeta(\psi(t, q^\epsilon)) = \zeta(q^\epsilon) - t \ge \eta$, which implies the containment

$$\{\psi(t, q^\epsilon) : t \le \tau, \epsilon > 0\} \subset E_\delta \cap J_\eta \cap E_{\mathrm{cs}} \tag{7.107}$$

which we know, from item i)., to be bounded: for some $C < \infty$,

$$\{\psi(t, q^\epsilon) : t \le \tau, \epsilon > 0\} \subset \mathrm{Mat}_n([0, C]). \tag{7.108}$$

By the lemma of leaving compact sets, there exists $s < \tau$ such that, for all $t \in (s, \tau)$, $\psi_t(q) \notin \mathrm{Mat}_n([0, C])$. However, if we pick $t \in (s, \tau)$, we have $\psi_t(q^\epsilon) \to \psi_t(q)$, by the continuity of the dependence in the initial conditions, which is a contradiction. Therefore, $\limsup_{\epsilon\to 0}\zeta(q^\epsilon) \le \zeta(q)$, which proves the claimed convergence.

iii). Let us first establish the claim for ψ. Firstly, we note that by ii)., the map $t \mapsto \zeta(q_0(t))$ is continuous on I. Therefore, fixing $t \in I$, we may choose choose $\epsilon, \delta > 0$

such that, if $|t-s|\le\delta$, then $s\in I$ and $s<\min\left(\zeta\left(q_0(s)\right),\zeta\left(q_0(t)\right)\right)-\epsilon$. Now, we observe that, for $s\in[t-\delta,t+\delta]$,

$$|\psi(t,q_0(t))-\psi(s,q_0(s))|\le|\psi(t,q_0(t))-\psi(t,q_0(s))|+|\psi(t,q_0(s))-\psi(s,q_0(s))|. \tag{7.109}$$

As $s\to t$, the first term converges to 0 by continuity of the solution $\psi(t,q)$ in the initial data q_0; it is therefore sufficient to control the second term. By the choice of δ, for all $s\in[t-\delta,t+\delta]$, we have

$$\zeta(\psi(s,q_0(s)))=\zeta(q_0(s))-s>\epsilon \tag{7.110}$$

so that $\psi(s,q_0(s))\in J_\epsilon$. Moreover, by compactness of $[t-\delta,t+\delta]$, there exists some $\eta>0$ such that $q_0(s)\in E_\eta$ for all $s\in[t-\delta,t+\delta]$, and since $q_0(s)\in E_{\mathrm{cs}}$ and these sets are preserved under the flow, we have $\psi(u,q_0(s))\in E_\eta\cap E_{\mathrm{cs}}$ for all $0\le u<\zeta(q_0(s))$. However, we showed in point i). above that that the intersection of these three regions is compact and so there exists a constant $M=M(\epsilon)$: for all $s\in[t-\delta,t+\delta]$, and for all $u\le t+\delta$,

$$u<\zeta(q_0(s));\qquad |b(\psi(u,q_0(s))|\le M. \tag{7.111}$$

This implies the bound, for all $s\in[t-\delta,t+\delta]$,

$$|\psi(t,q_0(s))-\psi(s,q_0(s))|\le M|t-s| \tag{7.112}$$

which implies the claimed continuity.

The case for $\chi(z_0(t),q_0(t),t)$ is similar. Let us fix $t\in I$; following the same argument leading to (7.111), there exists $\delta>0, M<\infty$ such that, if $s\in[t-\delta,t+\delta]$ then $s\in I$ and for all $u\le s$, $\psi(u,q_0(u))\in\mathrm{Mat}_n([0,M])$. The equation (Q2) can now be integrated directly to obtain, for $s\in[t-\delta,t+\delta]$,

$$\chi(s,z_0(s),q_0(s))=\exp\left(\int_0^s w\left(\psi(u,q_0(u))\right)du\right)z_0(s). \tag{7.113}$$

In particular, it follows that $\chi(s,z_0(s),q_0(s))$ is bounded as s varies in $[t-\delta,t+\delta]$. With this, the argument for ψ can be modified to prove the same result

□

We can finally combine the previous lemmas to prove Lemma 7.21.

Proof of Lemma 7.21. Let us fix λ_0 satisfying Hypothesis 7.1, and let ν_0 be its push-forward $\nu_0=\pi_\#\lambda_0$; let $(\lambda_t)_{t\ge0}$ and $(\nu_t)_{t\ge0}$ be the solutions to (Fl, mΠFl) with these starting points, respectively. By Lemma 7.22, ν_t is given by $\nu_t=\pi_\#\lambda_t$ and in particular, $\mathcal{E}(t)=\langle\varphi^2,\lambda_t\rangle=\langle\varphi^2,\nu_t\rangle$, $\mathcal{Q}_{ij}(t)=\langle\pi_i\pi_j,\lambda_t\rangle=\langle\pi_i\pi_j,\nu_t\rangle$ and $t_{\mathrm{g}}(\nu_0)=t_{\mathrm{g}}(\lambda_0)$.

From Lemma 7.24, we know that there exists $t_{\mathrm{e}}=t_{\mathrm{e}}(\nu_0)>0$ such that $\mathcal{E}(t)=\langle\varphi^2,\nu_t\rangle$

is finite, continuous and increasing on $[0, t_e)$, and diverges to infinity as $t \uparrow t_e$. Moreover, thanks to the differential equations (7.87), all components of $\mathcal{Q}(t)$ are continuous and increasing on $[0, t_e)$.

Consider next the shifted initial data ν_0^ϵ given by (7.91); thanks to Lemma 7.25, we know that $t_g(\nu_0^\epsilon) = t_e(\nu_0^\epsilon)$. By Lemma 7.26, $t_g(\nu_0^\epsilon) \to t_g(\nu_0)$. For the explosion times, we know from Lemma 7.27 that $t_e(\nu_0^\epsilon) = \zeta(q_0^\epsilon)$ and $t_e(\nu_0) = \zeta(q_0)$, where $q_0^\epsilon, q_0 \in E$ are the matrixes

$$(q_0)_{ij} = \langle \pi_i \pi_j, \nu_0 \rangle; \qquad (q_0^\epsilon)_{ij} = \langle \pi_i \pi_j, \nu_0^\epsilon \rangle. \tag{7.114}$$

By dominated convergence, $q_0^\epsilon \to q_0$; by hypothesis (A3.), each $(q_0)_{ii} = \langle \pi_i^2, \nu_0 \rangle = \langle \pi_i^2, \lambda_0 \rangle > 0$, so $q_0 \in E_\delta$ for some $\delta > 0$. Finally, for all $t < \zeta(q_0) = t_e(\nu_0)$, $\psi(t, q_0)_{ij} = \langle \pi_i \pi_j, \nu_t \rangle$ which certainly satisfies the desired Cauchy-Schwarz inequality $\psi(t, q_0)_{ij}^2 \leq \psi(t, q_0)_{ii} \psi(t, q_0)_{jj}$, so $q_0 \in E_{cs}$. A similar argument shows that $q_0^\epsilon \in E_{cs}$ for all $\epsilon > 0$, so Lemma 7.28 shows that $t_e(\nu_0^\epsilon) = \zeta(q_0^\epsilon) \to \zeta(q_0) = t_e(\nu_0)$. Comparing these two limits, $t_g(\nu_0) = t_e(\nu_0)$, concluding the proof. □

7.7.2 The Critical Point

Using the concepts introduced above, we next consider the behaviour at and near the critical time t_g.

Lemma 7.29. *In the notation of Lemma 7.21, we have*

$$\mathcal{E}(t_g) = \infty = \lim_{t \to t_g} \mathcal{E}(t). \tag{7.115}$$

Proof. We first show that $\mathcal{E}(t_g) = \infty$. Suppose, for a contradiction, that $\mathcal{E}(t_g) < \infty$. Then, applying [156, Proposition 2.7] as in Lemma 7.24, we see that, for some positive $\delta > 0$, there exists a strong solution $(\nu_t)_{t<\delta}$ to (Sm), starting at λ_{t_g}. This solution is conservative, so is an initial segment of the solution $(\nu_t)_{t \geq 0}$ to (Fl) starting at λ_{t_g}. By Corollary 7.20, $\langle \varphi, \lambda_{t_g} \rangle = \langle \varphi, \lambda_0 \rangle$, which implies that $(\lambda_{t_g+t})_{t \geq 0}$ solves (Fl) starting at λ_{t_g}. By uniqueness in Lemma 7.4, we conclude

$$\nu_t = \lambda_{t_g+t} \qquad \text{for all } t \geq 0. \tag{7.116}$$

On the other hand. by definition of t_g,

$$\langle \varphi, \lambda_{t_g+t} \rangle < \langle \varphi, \lambda_0 \rangle = \langle \varphi, \lambda_{t_g} \rangle \text{ for all } t > 0. \tag{7.117}$$

This contradicts the fact that $(\nu_t)_{t<\delta}$ is strong, which therefore shows that $\mathcal{E}(t) = \infty$.

The second point follows, because $t \mapsto \lambda_t$ is continuous, and $\lambda \mapsto \langle \varphi^2, \lambda \rangle$ is lower semi-continuous, when $\mathcal{M}_{\leq 1}(S)$ is equipped with the weak topology. □

7.7.3 The Supercritical Regime

We finally turn to the supercritical case; our result is as follows.

Lemma 7.30. *In the notation of Lemma 7.21, the map $t \mapsto \mathcal{E}(t)$ is finite and continuous, and therefore locally bounded, on (t_{g}, ∞).*

The proof is based on a *duality argument* following Theorem 7.50, which connects the measures in the supercritical regime to an auxiliary process in the subcritical case. Let $(G_t^N)_{t\geq 0}$ be the random graph processes described in Section 7.5 with points $\mathbf{x}_N$ sampled as a Poisson random measure of intensity $N\lambda_0$; it is straightforward to see that Hypothesis 7.2 holds. Fix $t > t_{\mathrm{g}}$, and let $\widetilde{G}_t^N$ be the graph G_t^N with the giant component deleted.

Let $\kappa_t(x)$ be the survival function defined in Lemmas 7.15, 7.44, and let $\widehat{\lambda}_0^t(dx) = (1 - \kappa_t(x))\lambda_0(dx)$. By Lemma 7.4, there exists a unique solution $(\widehat{\lambda}_s^t)_{s\geq 0}$ to the equation (Fl) starting at $\widehat{\lambda}_0^t$; write $\widehat{t}_{\mathrm{g}}(t)$ for its gelation time.

Let $\mathbf{y}_N = (y_i : i \leq \widehat{l}^N)$ be an enumeration of the vertexes x_i not belonging to the giant component in G_t^N. By Theorem 7.50, we can construct a random graph $\widehat{G}_t^N$ on $\{1, ..., \widehat{l}^N\}$ with distribution $\mathcal{G}(\mathbf{y}_N, tk/N)$, such that, with high probability, $\widehat{G}_t^N$ is the graph $\widetilde{G}_t^N$ formed by deleting the largest component of G_t^N.

In order to appeal to Lemmas 7.12, 7.16, we will now verify that the desired Hypothesiss 7.1, 7.2 hold for the vertex space $\widehat{\mathcal{V}}$.

Lemma 7.31. *Fix $t > 0$, and let $\lambda_0, G_t^N, \widehat{\lambda}_0^t$ and $\widehat{\mathcal{V}}$ be as described above. Then Hypothesis 7.2 hold for $\mathbf{y}_N$ and $\widehat{\lambda}_0^t$.*

Proof. To ease notation, we write $\widehat{\lambda}_0$ for $\widehat{\lambda}_0^t$, λ_0^N for the initial empirical measure of the unmodified process corresponding to $\mathbf{x}_N$, and $\widehat{\lambda}_0^N$ for the reduced empirical measure corresponding to $\mathbf{y}_N$:

$$\widehat{\lambda}_0^N = \frac{1}{N}\sum_{i=1}^{\widehat{l}^N} \delta_{y_i}. \tag{7.118}$$

It is straightforward to see that $\widehat{\lambda}_0^t$ inherits the properties in Hypothesis 7.1 from λ_0, and so it is sufficient to establish Hypothesis 7.2.

For (B1.), we note that part of the content of Theorem 7.50 is the weak convergence

$$\widehat{\lambda}_0^N = \frac{1}{N}\sum_{i=1}^{\widehat{l}^N} \delta_{y_i} \to \widehat{\lambda} \qquad \text{weakly, in probability} \tag{7.119}$$

as desired. Moreover, by construction, $\mathrm{Supp}(\widehat{\lambda}_0^N) \subset \mathrm{Supp}(\lambda_0^N)$, so it follows from (B1.) that $\widehat{\lambda}_0^N$ is supported on $\{\pi_0 = 1\}$ as required.

We will now show that (B2.) follows from the previous point, together with the moment estimates for the original initial measure λ_0^N.

Fix $\xi < \infty$, and let $\chi \in C_b(S)$ be such that $1\!\!1_{S_\xi} \le \chi \le 1\!\!1_{S_{\xi+1}}$. We observe that

$$\begin{aligned} \left|\langle \pi, \widehat{\lambda}_0^N\rangle - \langle \pi, \widehat{\lambda}_0\rangle\right| &\le \left|\langle \pi\chi, \widehat{\lambda}_0^N - \widehat{\lambda}_0\rangle\right| + \langle |\pi|\, 1\!\!1_{S_\xi^c}, \widehat{\lambda}_0^N\rangle + \langle |\pi|\, 1\!\!1_{S_\xi^c}, \widehat{\lambda}_0\rangle \\ &\le \left|\langle \pi\chi, \widehat{\lambda}_0^N - \widehat{\lambda}_0\rangle\right| + \frac{C}{\xi}\langle \varphi^2, \lambda_0^N\rangle + \frac{C}{\xi}\langle \varphi^2, \lambda_0\rangle \end{aligned} \tag{7.120}$$

for some constant C, thanks to the bound in part iv) of the definition (7.1.1). We now fix $\epsilon, \delta > 0$. Thanks to (A2., B2.), $\langle \varphi^2, \lambda_0^N\rangle$ is bounded in L^1 and $\langle \varphi^2, \lambda_0\rangle < \infty$, and so we may choose $\xi < \infty$ such that the second and third terms are at most $\epsilon/3$ with probability exceeding $1 - \delta/2$, for all N. For this choice of ξ, the first term vanishes as $N \to \infty$ by weak convergence in probability, and so is at most $\frac{\epsilon}{3}$ with probability exceeding $1 - \delta/2$ for all N large enough. Therefore, for all such N, we have

$$\mathbb{P}\left(|\langle \pi, \widehat{\lambda}_0^N - \widehat{\lambda}_0\rangle| > \epsilon\right) \le \delta \tag{7.121}$$

which proves the desired convergence in probability.

For the second assertion of (B2.), we note that $\langle \varphi^2, \widehat{\lambda}_0^N\rangle \le \langle \varphi^2, \lambda_0^N\rangle$ by the construction of $\mathbf{y}_N$, and $\langle \varphi^2, \lambda_0^N\rangle$ is uniformly integrable by the hypothesis (B2.). □

We now use this preparatory result to prove Lemma 7.30.

Proof of Lemma 7.30. Let $G_t^N, \widetilde{G}_t^N, \widehat{G}_t^N$ be as above. Recalling that we consider equality of graphs to include equality of the vertex data, it follows from Theorem 7.50 that

$$\mathbb{P}(\pi_\star(\widehat{G}_t^N) = \pi_\star(\widetilde{G}_t^N)) \to 1. \tag{7.122}$$

From Lemmas 7.12, 7.31, we obtain the following convergences in probability:

$$\pi_\star(G_t^N) \to \pi_\#\lambda_t; \qquad \pi_\star(\widehat{G}_t^N) \to \pi_\#\widehat{\lambda}_t^t \tag{7.123}$$

in the weak topology, in probability. Moreover, the difference

$$\pi_\star(G_t^N) - \pi_\star(\widetilde{G}_t^N) = \frac{1}{N}\delta(\mathcal{C}_1(G_t^N)) \tag{7.124}$$

converges to 0 in the weak topology in probability, since the mass of the difference converges to 0. It follows that

$$\pi_\star(\widetilde{G}_t^N) \to \pi_\#\lambda_t \tag{7.125}$$

in the weak topology, in probability, and by uniqueness of limits, we have $\pi_\#\widehat{\lambda}_t^t = \pi_\#\lambda_t$. In particular, it follows that

$$\langle \varphi^2, \lambda_t\rangle = \langle \varphi^2, \pi_\#\lambda_t\rangle = \langle \varphi^2, \pi_\#\widehat{\lambda}_t^t\rangle = \langle \varphi^2, \widehat{\lambda}_t^t\rangle. \tag{7.126}$$

Using assumption (A2.), we can see that $tk \in L^2(S \times S, \lambda_0 \times \lambda_0)$, and so it follows from Theorem 7.50 that the graphs $\widehat{G}_t^N$ are subcritical. By Lemma 7.16, it follows that that $t < \widehat{t}_g(t)$, and so by Lemma 7.21, we have

$$\langle \varphi^2, \lambda_t\rangle = \langle \varphi^2, \widehat{\lambda}_t^t\rangle < \infty. \tag{7.127}$$

Using Theorem 7.49 and dominated convergence, the map

$$\begin{aligned} t \mapsto q_0^t &= \left(\left\langle \pi_i\pi_j, \widehat{\lambda}_0^t \right\rangle\right)_{i,j=1}^n = (\langle (1-\kappa_t)\pi_i\pi_j, \lambda_0\rangle)_{i,j=1}^n ; \\ t \mapsto z_0^t &= \left(\left\langle \pi_i\pi_0, \widehat{\lambda}_0^t \right\rangle\right)_{i=0}^n = (\langle (1-\kappa_t)\pi_i\pi_0, \lambda_0\rangle)_{i=0}^n \end{aligned} \tag{7.128}$$

are continuous, and q_0^t takes values in E°. Therefore, by the general ODE considerations in Lemma 7.28 point iii)., it follows that the maps

$$t \mapsto q^t(t) = \psi(t, q_0^t) = \left(\langle \pi_i\pi_j, \widehat{\lambda}_t^t\rangle\right)_{i,j=1}^n ; \qquad t \mapsto z_t^t = \chi(t, z_0^t, q_0^t) = (\langle \pi_i\pi_0, \widehat{\lambda}_t^t\rangle)_{i=0}^n \tag{7.129}$$

are finite and continuous on (t_{g}, ∞). Since $\pi_\# \widehat{\lambda}_t^t = \pi_\# \lambda_t$, item iii) of Lemma 7.28 shows that the maps $t \mapsto \mathcal{Q}(t)_{ij}, 0 \le i, j \le n$ are finite and continuous on (t_{g}, ∞), which implies that they are bounded on compact subsets. □

Remark 7.32. *The same argument also shows that $t \mapsto \widehat{t}_{\mathrm{g}}(t)$ is continuous. This fact will be used later in the proof of Lemma 7.42.*

7.8 Representation and Dynamics of the Gel

7.8.1 Representation Formula

The duality construction used in the proof of Lemma 7.30 gives us a natural way to relate the gel data g_t to the survival function κ_t. This is the content of the following lemma.

Lemma 7.33. *Let λ_0 be an initial data satisfying Hypothesis 7.1, and let $g_t = (M_t, E_t, 0)$ be the gel data for the corresponding solution to (Fl). Let $\kappa_t(\cdot)$ be the corresponding survival function defined in Section 7.5 and Appendix 7.B. Then we have the equality*

$$g_t = \langle \kappa_t \pi, \lambda_0 \rangle. \tag{7.130}$$

In particular, $t \mapsto g_t$ is continuous and if $t > t_{\mathrm{g}}$ then $M_t > 0$, and $E_t > 0$ componentwise.

Together with the identification of κ_t in Lemma 7.15, this proves part 3 of Theorem 7.2.

Proof. We deal with the supercritical and subcritical/critical cases, $t > t_{\mathrm{g}}, t \le t_{\mathrm{g}}$ separately.

Step 1. Supercritical Case $t > t_{\mathrm{g}}$. Let $(\widehat{\lambda}_s^t)_{s \ge 0}$ and $\widehat{t}_{\mathrm{g}}(t)$ be as in the proof of Theorem 7.30. Then, since $(\widehat{\lambda}_s^t)_{s\ge 0}$ is conservative on $[0, \widehat{t}_{\mathrm{g}})$, and $t < \widehat{t}_{\mathrm{g}}(t)$, we have, for all $0 \le i \le n+m$,

$$\langle \pi_i, \widehat{\lambda}_t^t \rangle = \langle \pi_i, \widehat{\lambda}_0^t \rangle = \int_S \pi_i(x)(1-\kappa_t(x))\lambda_0(dx). \tag{7.131}$$

As shown in Lemma 7.30, $\pi_\# \lambda_t = \pi_\# \widehat{\lambda}^t_t$, so we have

$$\begin{aligned} g^i_t :&= \langle \pi_i, \lambda_0\rangle - \langle \pi_i, \lambda_t\rangle = \langle \pi_i, \lambda_0\rangle - \langle \pi_i, \widehat{\lambda}^t_t\rangle \\ &= \langle \pi_i \kappa_t, \lambda_0\rangle \end{aligned} \tag{7.132}$$

as claimed.

Step 2. Subcritical and Critical Cases $t \le t_{\mathrm{g}}$. For $t < t_{\mathrm{g}}$, the result is immediate: we have g_t by definition of t_{g}, and $\kappa_t = 0$ by Theorem 7.44. The critical case is identical, recalling from Corollary 7.20 that $g_{t_{\mathrm{g}}} = 0$.

Continuity follows from Theorem 7.49 by using dominated convergence. For the final claim, if $t > t_{\mathrm{g}}$ then $\kappa_t(x) > 0$ λ_0 - almost everywhere, by Lemma 7.44. By hypothesis (A3.), for all $i = 1, ..., n$, $\pi_i > 0$ on a set of positive λ_0 measure. Together, these imply that $\langle \kappa_t \pi_i, \lambda_0\rangle > 0$, as claimed. □

7.8.2 Gel Dynamics Beyond the Critical Time

We now obtain point 4 of Theorem 7.2 as a consequence of the previous results. We have already proven the continuity of g_t on the whole time interval $[0, \infty)$ and the finiteness of the second moments $q_t = (\langle \pi_i \pi_j, \lambda_t\rangle)^n_{i,j=1}$ in the supercritical regime. Therefore, it is sufficient to prove the following result.

Lemma 7.34. *In the notation of Theorem 7.2, let g_t be the data of the gel associated to $(\lambda_t)_{t\ge 0}$. Then, for $t \ge t_{\mathrm{g}}$, we have*

$$g^i_t = \int_{t_{\mathrm{g}}}^t \sum_{j,k=1}^n \langle \pi_i \pi_j, \lambda_t\rangle a_{jk} g^k_s ds. \tag{7.133}$$

Thanks to the continuity of the second moments above t_{g}, this has the differential form, holding in the classical sense,

$$\frac{d}{dt} g^i_t = \sum_{j,k=1}^n \langle \pi_i \pi_j, \lambda_t\rangle a_{jk} g^k_t. \tag{7.134}$$

Remark 7.35. *In proving Lemma 7.34, we will split the growth of the gel into two terms $\mathcal{T}_1 + \mathcal{T}_2$, where $\mathcal{T}_1$ represents the absorption of particles into the gel, and $\mathcal{T}_2$ represents the coagulation of smaller particles. We will show that $\mathcal{T}_2 = 0$, giving the claimed result; this may be expected following the relationship between gelation and blowup of the second moment $\mathcal{E}(t)$ in Lemma 7.21, and the finiteness of $\mathcal{E}$ in the supercritical regime.*

Proof. We return to the truncated dynamics $(\mathrm{Fl}|^1_\xi, \mathrm{Fl}|^2_\xi)$ used in the proof of Lemma 7.4. We recall that, starting at

$$\lambda^\xi_0 = 1\!\!1_{S_\xi} \lambda_0; \qquad g^\xi_0 = \int_{x \notin S_\xi} x \lambda_0(dx) \tag{7.135}$$

the solution (λ^ξ_t, g^ξ_t) to $(\mathrm{Fl}|^1_\xi, \mathrm{Fl}|^2_\xi)$ exists and is unique, and we have

$$\lambda^\xi_t = \lambda_t 1\!\!1_{S_\xi}; \qquad (M^\xi_t, E^\xi_t) \downarrow (M_t, E_t) \text{ as } \xi \uparrow \infty. \tag{7.136}$$

where $(\lambda_t)_{t\geq 0}$ is the solution to (Fl) starting at λ_0, and (M_t, E_t) are the nonzero components of the associated gel data.

Fix s, t such that $t_\mathrm{g} < s < t$. Rewriting $(\mathrm{Fl}|^2_\xi)$ and using that $P^\xi_t = 0$, we have that

$$\begin{aligned} g^{\xi,i}_t - g^{\xi,i}_s = &\int_s^t \sum_{j,k=1}^n \langle \pi_i\pi_j, \lambda^\xi_u\rangle a_{jk} g^{\xi,k}_u du \\ &+ \frac{1}{2}\int_s^t \int_{S^2_\xi} \pi_i(x+y) 1\!\!1[\varphi(x+y) > \xi]\overline{K}(x,y)\lambda_u(dx)\lambda_u(dy)du. \end{aligned} \tag{7.137}$$

Let us write $\mathcal{T}_1(\xi), \mathcal{T}_2(\xi)$ for the two terms appearing in (7.137) for ease of notation.

We first show that $\mathcal{T}_1(\xi)$ converges to the expression analagous to the claimed limit in (7.133). By the monotonicity $\lambda^\xi_u \leq \lambda_u$, and local boundedness in Lemma 7.30, each $\langle \pi_i\pi_j, \lambda^\xi_u\rangle$ is bounded, uniformly in $\xi < \infty$ and $u \in [s,t]$. It is also straightforward to see that the truncated gel data are bounded by $g^{\xi,i}_u \leq \langle \pi_i, \lambda_0\rangle$, so the integrand appearing in $\mathcal{T}_1(\xi)$ is bounded. Using (7.136) and bounded convergence, we take the limit $\xi \to \infty$ to obtain

$$\mathcal{T}_1(\xi) \to \int_s^t \sum_{j,k=1}^n \langle \pi_i\pi_j, \lambda_u\rangle a_{jk} g^k_u du. \tag{7.138}$$

We now deal with the second term $\mathcal{T}_2(\xi)$, which we claim converges to 0. Expanding the total rate $\overline{K}$, we have

$$\mathcal{T}_2(\xi) = \int_s^t \sum_{j,k=1}^n \int_{S^2} \pi_i(x)\pi_j(x)\pi_k(y) 1\!\!1\left[\varphi(x+y) > \xi\right] \lambda_u(dx)\lambda_u(dy). \tag{7.139}$$

The integrand converges to 0 pointwise as $\xi \to \infty$, and is dominated by $\pi_i(x)\pi_j(x)\pi_k(y)$. By Lemma 7.30,

$$\sup_{u\in[s,t]} \int_{S^2} \pi_i(x)\pi_j(x)\pi_k(y)\lambda_u(dx)\lambda_u(dy)du \leq \sup_{u\in[s,t]} \langle \varphi^2, \lambda_u\rangle\langle \pi_k, \lambda_0\rangle < \infty. \tag{7.140}$$

Therefore, by dominated convergence, $\mathcal{T}_2(\xi) \to 0$ as $\xi \to \infty$, as claimed. Combining this with the analysis of the first term, we have shown that

$$g^i_t - g^i_s = \int_s^t \sum_{j,k=1}^n \langle \pi_i\pi_j, \lambda_u\rangle a_{jk} g^k_u du. \tag{7.141}$$

Taking $s \downarrow t_\mathrm{g}$, and using the continuity $g_s \downarrow 0$ established in Lemma 7.33, we obtain the claimed result. $\square$

7.9 Uniform Convergence of the Stochastic Coagulant

We now show how previous results, describing the dynamics of g_t, imply convergence to their maximum values $\langle \pi_0, \lambda_0\rangle$ as $t \to \infty$. Using this, we will be able to upgrade the previous result, Lemma 7.7, on the convergence of the stochastic coagulant to *uniform* convergence.

Lemma 7.36. *Let λ_0 be an initial measure satisfying Hypothesis 7.1, and let g_t be the gel data for the associated solution $(\lambda_t)_{t\geq 0})$ to (Fl). As $t \uparrow \infty$, we have*

$$g_t^i \to g_\infty^i = \langle \pi_i, \lambda_0\rangle \tag{7.142}$$

for $i = 0, ..., n$.

Proof. Let us fix $1 \leq i \leq n$, and write g_∞^i for the claimed limit $\langle \pi_i, \lambda_0\rangle$; it is immediate that $g_t^i \leq g_\infty^i$ for all $t \geq 0$. Choose $t_0 > t_{\rm g}$ and $1 \leq j \leq n$ such that $a_{ij} > 0$. Thanks to Lemma 7.33, $\epsilon = a_{ij} g_{t_0}^j > 0$, and note also that g_t^j is increasing, so that this bound holds uniformly in $t \geq t_0$. Applying Lemma 7.34 and taking $t \to \infty$, we obtain the integral inequality

$$\begin{aligned} \lim_{t\to\infty} \left(g_t^i - g_{t_0}^i\right) &\geq \int_{t_0}^\infty \langle \pi_i^2, \lambda_s\rangle a_{ij} g_s^j ds \geq \epsilon \int_{t_0}^\infty \langle \pi_i, \lambda_s\rangle^2 ds \\ &\geq \epsilon \int_{t_0}^\infty \left(g_\infty^i - g_s^i\right)^2 ds \end{aligned} \tag{7.143}$$

where the limit on the left hand side exists since g_t^i is increasing. Recalling that g_t^i is bounded, the integral appearing on the right-hand side must converge, and since the integrand is decreasing in s, this is only possible if $(g_\infty^i - g_s^i)^2 \to 0$ as $s \to \infty$, as desired.

We must deal separately with π_0, since π_0 does not appear in the dynamics explicitly and the argument above does not apply. For this case, we note that the monotonicity $\kappa_s \leq \kappa_t$ whenever $s \leq t$ implies that κ_t converges pointwise to a limit $\kappa_\infty \leq 1$. Using Lemma 7.33 and dominated convergence, we have, for all $i = 1, ..., n$

$$\langle \pi_i \kappa_\infty, \lambda_0\rangle = \lim_{t\to\infty} \langle \pi_i \kappa_t, \lambda_0\rangle = \lim_{t\to\infty} g_t^i = \langle \pi_i, \lambda_0\rangle. \tag{7.144}$$

This implies the containment

$$\{\kappa_\infty < 1\} \subset \{\pi_i = 0\} \cup \mathcal{N}_i \tag{7.145}$$

for a λ_0-null set $\mathcal{N}_i$, for each $i = 1, ...n$. Taking an intersection, and since $\lambda_0(\pi_i = 0 \text{ for all } i = 1, ..n) = 0$ by irreducibility (A4.), we see that $\kappa_\infty = 1$, λ_0-almost everywhere. By Lemma 7.33 and dominated convergence again,

$$M_t = g_t^0 = \langle \pi_0 \kappa_t, \lambda_0\rangle \to \langle \pi_0, \lambda_0\rangle \tag{7.146}$$

which is the claimed limit. ∎

Lemma 7.37. *Fix a measure λ_0 satisfying Hypothesis 7.1, and let $(\lambda_t)_{t\geq 0}$ be the associated solution to (Fl). Let λ^N_t be the stochastic coagulants, with initial data λ^N_0 satisfying Hypothesis 7.2. Then we have the* uniform *convergence*

$$\sup_{t\geq 0} \rho_1(\lambda^N_t, \lambda_t) \to 0 \tag{7.147}$$

in probability.

Proof. Fix $\epsilon > 0$. By Lemma 7.36, we can find $t_+ \in (t_{\mathrm{g}}, \infty)$ such that $M_{t_+} > \langle \pi_0, \lambda_0\rangle - \frac{\epsilon}{3}$. Let A^1_N be the event

$$A^1_N = \left\{ M^N_{t_+} > \langle \pi_0, \lambda_0\rangle - \frac{\epsilon}{2}; \quad \langle \pi_0, \lambda^N_0\rangle \leq \langle \pi_0, \lambda_0\rangle + \frac{\epsilon}{2}\right\}. \tag{7.148}$$

By Lemma 7.18 and condition (B2.), it follows that $\mathbb{P}(A^1_N) \to 1$. On this event, for any bounded, Lipschitz $f : S \to \mathbb{R}$ we have

$$\begin{aligned} \sup_{t\geq 0} \langle f, \lambda^N_t - \lambda_t\rangle &\leq \sup_{0\leq t\leq t_+} \langle f, \lambda^N_t - \lambda_t\rangle + \sup_{t> t_+} \langle f, \lambda^N_t - \lambda_t\rangle \\ &\leq \sup_{0\leq t\leq t_+} \langle f, \lambda^N_t - \lambda_t\rangle + \left(\langle \pi_0, \lambda^N_0\rangle - M^N_{t_+}\right) + \left(\langle \pi_0, \lambda_0\rangle - M_{t_+}\right) \\ &\leq \sup_{0\leq t\leq t_+} \langle f, \lambda^N_t - \lambda_t\rangle + \epsilon. \end{aligned} \tag{7.149}$$

Now, taking the supremum over all $f : S \to \mathbb{R}$ which are 1-bounded and 1-Lipschitz, we obtain on the same event

$$\sup_{t\geq 0} \rho_1(\lambda^N_t, \lambda_t) \leq \sup_{t\leq t_{\mathrm{fin}}} \rho_1(\lambda^N_t, \lambda_t) + \varepsilon \tag{7.150}$$

and the first term converges to 0 in probability by Lemma 7.7. We conclude that, with probability converging to 1, $\sup_{t\geq 0} \rho_1(\lambda^N_t, \lambda_t) < 2\varepsilon$, and since $\varepsilon > 0$ was arbitrary, we are done. □

7.10 Behaviour Near the Critical Point

We now prove item 5 of Theorem 7.2, concerning the phase transition: we will show that the gel data $g_t = (g^i_t)$ have nonnegative right-derivatives at the gelation time t_{g}. We start from the nonlinear fixed point equation (7.20), which we rewrite as

$$c_t = tF(c_t); \qquad F(c)_i = 2\int_S \left(1 - \exp\left(-\sum_{k=1}^n c_k \pi_k(x)\right)\right) \sum_{j=1}^n a_{ij}\pi_j(x)\lambda_0(dx). \tag{7.151}$$

The following proof is a modification of the arguments in [28, Theorem 3.17], which itself generalises an analagous, well-known result for the phase transition of Erdős-Réyni graphs.

Lemma 7.38. *Suppose that λ_0 satisfies Hypothesis 7.1, and let c_t be as in Lemma 7.15. Then c_t is right-differentiable at t_g, and the right-derivative $c'_{t_g^+} > 0$ is componentwise positive.*

Proof. Let us equip $\mathbb{R}^n$ with the inner product

$$(c, c')_{\lambda_0} = \sum_{i,j=1}^{n} c_i c'_j \langle \pi_i \pi_j, \lambda_0 \rangle \tag{7.152}$$

which is the pullback of the $L^2(\lambda_0)$ inner product under $c \mapsto \sum_i c_i \pi_i$, and write $|\cdot|_{\lambda_0}$ for the associated norm on $\mathbb{R}^n$. Differentiating under the integral sign twice, and using (A2.), we write

$$F(c) = \mathcal{Z}c - \Sigma(c) + R(c) \tag{7.153}$$

where $\mathcal{Z} = \mathcal{Z}(\lambda_0)$ is the $n \times n$ matrix found in Lemma 7.13, $\Sigma(\cdot)$ is a quadratic term, and R is a remainder:

$$\mathcal{Z}_{ij} = 2 \sum_{k=1}^{n} a_{ik} \langle \pi_k \pi_j, \lambda_0 \rangle; \tag{7.154}$$

$$\Sigma(c)_i = \sum_{j,k,l=1}^{n} a_{ij} \langle \pi_j \pi_k \pi_l, \lambda_0 \rangle c_k c_l \tag{7.155}$$

$$|R(c)|_{\lambda_0} = o\left(|c|^2_{\lambda_0}\right) \text{ as } |c| \to 0. \tag{7.156}$$

The signs here are chosen to guarantee that, if $c > 0$, then $\mathcal{Z}c, \Sigma(c) > 0$, and $\mathcal{Z}$ is self-adjoint with respect to $(\cdot,\cdot)_{\lambda_0}$. We also recall from Lemma 7.13 that the largest eigenvalue of $\mathcal{Z}$ is precisely t_g^{-1}, and the corresponding eigenspace is 1-dimensional. Let ψ be an associated eigenvector, scaled so that $|\psi|_{\lambda_0} = 1$. We note that $\sum_i \psi_i \pi_i$ is an eigenfunction of T, and in particular, the sign of ψ can be chosen so that $\sum_i \psi_i \pi_i > 0$ is strictly positive λ_0-almost everywhere; using (7.59) it follows that $\psi_i > 0$ for all $i = 1, ..., n$. From Lemma 7.15, Theorem 7.45 and Theorem 7.49, we know that $c_{t_g} = 0$, that $c_{t_g+\epsilon} \in [0,\infty)^n \setminus 0$ for all $\epsilon > 0$, and that $t \mapsto c_t$ is continuous at t_g.

Let us write $\psi^\perp$ for the orthogonal compliment of $\mathrm{Span}(\psi)$ with respect to $(\cdot,\cdot)_{\lambda_0}$. Since $\mathrm{Span}(\psi)$ is exactly the eigenspace $\mathrm{Ker}(\mathcal{Z} - t_g^{-1}1)$ of $\mathcal{Z}$ corresponding the largest eigenvalue t_g^{-1}, it follows from the self-adjointness of $\mathcal{Z}$ that $\mathcal{Z}$ maps $\psi^\perp$ into itself. Moreover, since ψ spans the eigenspace of $\mathcal{Z}$ for t_g^{-1}, it follows that, for $t > t_g$ small enough, $(t\mathcal{Z} - 1)|_{\psi^\perp}$ is invertible, and that the operator norm $\|(t\mathcal{Z} - 1)|_{\psi^\perp}^{-1}\|_{\lambda_0 \to \lambda_0}$ with respect to $|\cdot|_{\lambda_0}$ on $\psi^\perp \subset \mathbb{R}^n$ is bounded as $t \downarrow t_g$.

Let $P : \mathbb{R}^n \to \mathbb{R}^n$ be the orthogonal projection onto $\psi^\perp$ with respect to $(\cdot,\cdot)_{\lambda_0}$, and write $c_t^* = Pc_t$ so that we have the orthogonal decomposition

$$c_t = \alpha_t \psi + c_t^* \tag{7.157}$$

for some $\alpha_t \in \mathbb{R}$ and $c_t^* \in \psi^\perp$. Noting that $\mathcal{Z}P = P\mathcal{Z}$, it follows from (7.151, 7.157) that

$$c_t^\star = P(tF(c_t)) = t\mathcal{Z}c_t^\star + tP\left(-\Sigma(c_t) + R(c_t)\right). \tag{7.158}$$

The function $-\Sigma(c) + R(c)$ is of quadratic growth as $|c|_{\lambda_0} \to 0$, and using the invertibility of $(t\mathcal{Z} - I)|_{\psi^\perp}$ described above, it follows that there exists $\beta > 0$ such that $|c_t^*|_{\lambda_0} \leq \beta |c_t|^2_{\lambda_0}$ whenever $|c_t|_{\lambda_0} \leq 1$. In turn, it follows that $|c_t|_{\lambda_0} \sim \alpha_t$ as $t \downarrow t_{\mathrm{g}}$. Now, using (7.151) and the self-adjointness of $\mathcal{Z}$, we obtain

$$\begin{aligned} \alpha_t &= (\psi, c_t)_{\lambda_0} = (t_{\mathrm{g}}\mathcal{Z}\psi, c_t)_{\lambda_0} = t_{\mathrm{g}}(\psi, \mathcal{Z}c_t)_{\lambda_0} \\ &= \frac{t_{\mathrm{g}}}{t}(\psi, c_t)_{\lambda_0} - t_{\mathrm{g}}\left(\psi, -\Sigma(c_t) + R(c_t)\right)_{\lambda_0} \\ &= \frac{t_{\mathrm{g}}}{t}\alpha_t - t_{\mathrm{g}}\left(\psi, -\Sigma(c_t) + R(c_t)\right)_{\lambda_0}. \end{aligned} \tag{7.159}$$

We now expand to second order in α_t; for clarity, we will number the error terms $\mathcal{T}_t^i$. Since $|c_t|_{\lambda_0} \sim \alpha_t$, it follows that that $|c_t^*|_{\lambda_0} = \mathcal{O}(\alpha_t^2)$ and that $R(c_t) = o(\alpha_t^2)$. Expanding $\Sigma(c_t)$ using (7.157),

$$-\Sigma(c_t) + R(c_t) = -\alpha_t^2\Sigma(\psi) + \mathcal{T}_t^1; \qquad |\mathcal{T}_t^1|_{\lambda_0} = o(\alpha_t^2). \tag{7.160}$$

It therefore follows that

$$\alpha_t = t_{\mathrm{g}}\left(\frac{\alpha_t}{t} + \alpha_t^2(\psi, \Sigma(\psi))_{\lambda_0}\right) + \mathcal{T}_t^2; \qquad \mathcal{T}_t^2 = o(\alpha_t^2). \tag{7.161}$$

For $t > t_{\mathrm{g}}$ small enough, $\alpha_t > 0$, and we may rearrange to find

$$t - t_{\mathrm{g}} = t\, t_{\mathrm{g}}\, \alpha_t(\psi, \Sigma(\psi))_{\lambda_0} + \mathcal{T}_t^3; \qquad \mathcal{T}_t^3 = o(\alpha_t) \tag{7.162}$$

and in particular $\alpha_t \sim t - t_{\mathrm{g}}$ as $t \downarrow t_{\mathrm{g}}$, since

$$(\psi, \Sigma(\psi))_{\lambda_0} = \sum_{i,j,k,l=1}^n a_{ij}\psi_i\psi_k\psi_l\langle \pi_j\pi_k\pi_l, \lambda_0\rangle > 0. \tag{7.163}$$

Finally, we obtain

$$\frac{\alpha_t}{t - t_{\mathrm{g}}} \to \frac{1}{t_{\mathrm{g}}^2(\psi, \Sigma(\psi))_{\lambda_0}} \qquad \text{as } t \downarrow t_{\mathrm{g}}. \tag{7.164}$$

The calculations above show that $|c_t - \alpha_t\psi| = \mathcal{O}((t - t_{\mathrm{g}})^2)$, and the claimed right-differentiability now follows. Finally, since $\psi_i > 0$ is strictly positive componentwise and $\alpha'_{t_{\mathrm{g}}+} > 0$, it follows that $c'_{t_{\mathrm{g}}+} > 0$ componentwise. □

We now show how this implies item 5 of Theorem 7.2. From Lemmas 7.15, 7.33, we have, for all $i = 0, 1, ..., n$

$$g_t^i = \int_S \left(1 - \exp\left(-\sum_{j=1}^n c_t^j\pi_j(x)\right)\right)\pi_i(x)\lambda_0(dx) \tag{7.165}$$

Differentiating under the integral sign using hypothesis (A2.), we obtain

$$g_t^i = \sum_{j=1}^n c_t^j \langle \pi_i \pi_j, \lambda_0 \rangle + o(c_t). \tag{7.166}$$

In the notation of the previous lemma, we see that for $t > t_{\mathrm{g}}$,

$$\begin{aligned} g_t^i &= (t - t_{\mathrm{g}}) \sum_{j=1}^n (c'_{t_{\mathrm{g}}+})_j \langle \pi_i \pi_j, \lambda_0 \rangle + o(t - t_{\mathrm{g}}) \\ &= \alpha'_{t_{\mathrm{g}}+}(t - t_{\mathrm{g}}) \left\langle \sum_{j=1}^n \psi_j \pi_j \pi_i, \lambda_0 \right\rangle + o(t - t_{\mathrm{g}}). \end{aligned} \tag{7.167}$$

which proves the desired right-differentiability. For the positivity, since all components of $c'_{t_{\mathrm{g}}+}$ are strictly positive, we have the lower bound for $i = 1, ..., n$

$$(g'_{t_{\mathrm{g}}+})_i \geq (c'_{t_{\mathrm{g}}+})_i \langle \pi_i^2, \lambda_0 \rangle > 0. \tag{7.168}$$

A similar argument holds for the 0^{th} component.

Finally, we address the size-biasing effect. We wish to choose a convex combination $\theta_i : i = 1, ..., n$ such that

$$\frac{\sum_{i=1}^n \theta_i (g'_{t_{\mathrm{g}}+})_i}{(g'_{t_{\mathrm{g}}+})_0} \geq \frac{\sum_{i=1}^n \theta_i \langle \pi_i, \lambda_0 \rangle}{\langle \pi_0, \lambda_0 \rangle}. \tag{7.169}$$

Thanks to the calculation above, this is equivalent to proving that

$$\frac{\sum_{i,j=1}^n \theta_i \psi_j \langle \pi_i \pi_j, \lambda_0 \rangle}{\sum_{k=1}^n \psi_k \langle \pi_k, \lambda_0 \rangle} \geq \frac{\sum_{i=1}^n \theta_i \langle \pi_i, \lambda_0 \rangle}{\langle \pi_0, \lambda_0 \rangle}. \tag{7.170}$$

If we choose $\theta_i = \psi_i / \sum_j \psi_j$, then these follow from the Cauchy-Schwarz inequality

$$\begin{aligned} \left\langle \sum_i \psi_i \pi_i, \lambda_0 \right\rangle^2 &\leq \left\langle (\sum_i \psi_i \pi_i)^2, \lambda_0 \right\rangle \langle 1, \lambda_0 \rangle \\ &= \left\langle (\sum_i \psi_i \pi_i)^2, \lambda_0 \right\rangle \langle \pi_0, \lambda_0 \rangle. \end{aligned} \tag{7.171}$$

We recall that the linear combination $f = \sum_i \psi_i \pi_i$ is an eigenfunction of T, and so can only be constant λ_0-almost everywhere if $s(x) = (T1)(x)$ is constant. In particular, if s is not constant λ_0-almost everywhere, the inequality (7.171) is strict, and hence so is (7.169), as desired.

7.11 Convergence of the Gel

We now prove the remaining part of Theorem 7.3, concerning the *uniform* convergence of the stochastic gel, drawing on other results we have proven. We recall that g_t^N are the data of the largest particle in the stochastic coagulant λ^N_t; to conclude the proof of Theorem 7.3, we must extend Lemma 7.18, to show uniform convergence in time, in probability.

Throughout this section, let λ_0 be an initial measure satisfying Hypothesis 7.1, and λ^N_t be stochastic coagulants satisfying Hypothesis 7.2 for this choice of λ_0. We will also let G^N_t be random graphs coupled to λ^N_t as described in Section 7.5, so that g^N_t is the data of the largest component in G^N_t.

This subsection is structured as follows. We recall that, in the proof of Lemma 7.18, we used the result on mesoscopic clusters from [28]: if $\xi_N \to \infty$ and $\frac{\xi_N}{N} \to 0$, then for all $t \geq 0$,

$$\frac{1}{N}\sum_{j\geq 2: C_j(G^N_t)\geq \xi_N} C_j(G^N_t) \to 0 \tag{7.172}$$

in probability. We will first state a lemma which extends this convergence to be uniform in time. Equipped with this lemma, and previous results, we will show how the proof of Lemma 7.18 can be modified to establish uniform convergence, and prove the analagous result when we sum over all clusters exceeding a deterministic size $\xi_N \ll N$. Finally, we return to prove the preliminary lemma.

The key lemma which we will require is the following, which generalises the result of Bollobás et al. recalled in Lemma 7.48.

Lemma 7.39. *Let G^N_t be as above, and let ξ_N be any sequence such that $\xi_N \to \infty$, $\frac{\xi_N}{N} \to 0$. Then we have the uniform estimate*

$$\sup_{t\geq 0}\left[\frac{1}{N}\sum_{j\geq 2: C_j(G^N_t)\geq \xi_N} C_j(G^N_t)\right] \to 0 \qquad \textit{in probability.} \tag{7.173}$$

The proof of this lemma will be deferred until Subsection 7.11.2

7.11.1 Proof of Theorem 7.3

It remains to prove that the convergence of the stochastic approximations g^N_t, $\widetilde{g}^N_t$ to the gel, given by the gel data of the largest cluster, and of all clusters exceeding a certain scale ξ_N respectively. This is the content of the following two lemmas.

Lemma 7.40. *In the notation above, we have the uniform convergence*

$$\sup_{t\geq 0}\left|g^N_t - g_t\right| \to 0 \qquad \textit{in probability.} \tag{7.174}$$

Lemma 7.41. *Fix a sequence ξ_N such that $\xi_N \to \infty$ and $\frac{\xi_N}{N} \to 0$, and let $\widetilde{g}^N_t$ be given by*

$$\widetilde{g}^N_t = \frac{1}{N}\sum_{j\geq 1: C_j(G^N_t)\geq \xi_N} \pi(\mathcal{C}_j(G^N_t)) = \left(\langle \pi_i 1\!\!1[\pi_0 \geq \xi_N], \lambda^N_t\rangle\right)_{i=0}^{n+m}. \tag{7.175}$$

Then

$$\sup_{t\geq 0} \left|\widetilde{g}_t^N - g_t^N\right| \to 0 \qquad \textit{in probability.} \tag{7.176}$$

We now prove these two lemmas, looking primarily at the 0^{th} coordinate. The other coordinates follow, with minor modifications which will be discussed later.

Proof of Lemma 7.40. This is an extension of the proof of Lemma 7.18, from where much of the notation is taken. We deal first with the 0^{th} coordinate $M_t^N - M_t$. Let η_r be a fast-growing sequence such that $\beta(r, \eta_r) \to 0$ in the notation of Lemma 7.19, and let $S_{(r)}, \widetilde{f}_r, \widetilde{h}_r$ be as in Lemma 7.18. Let also ξ_N be a sequence, to be constructed later, such that

$$\xi_N \to \infty; \qquad \frac{\xi_N}{N} \to 0 \tag{7.177}$$

and write $f_N = \widetilde{f}_{\xi_N}, h_N = \widetilde{h}_{\xi_N}$. We recall also the decomposition (7.76)

$$M_t^N - M_t = \sum_{i=1}^{5} \mathcal{T}_N^i(t) \tag{7.178}$$

where the definitions of the error terms are given in (7.76). The bounds obtained on $\mathcal{T}_N^3(t), \mathcal{T}_N^5(t)$ in the proof of Lemma 7.18 are already uniform in time; we will now show how the previous proof can be modified to estimate the other terms uniformly in time.

Step 1. Estimate on $T_N^1(t)$ $\mathcal{T}_N^1(t)$ is the nonrandom error $\langle \pi_0, \lambda_t\rangle - \langle \pi_0 f_N, \lambda_t\rangle$. The estimate in Lemma 7.18 shows that $\langle \pi_0 f_N, \lambda_t\rangle \uparrow \langle \pi_0, \lambda_t\rangle$ for each fixed $t \geq 0$. The maps $t \mapsto \langle \pi_0 f_N, \lambda_t\rangle$, $t \mapsto \langle \pi_0, \lambda_t\rangle$ are both continuous on $[0, \infty)$, by the definition of the Flory dynamics (Fl) and Lemma 7.33 respectively. Let us extend both of these maps to $[0, \infty]$ by defining both to be 0 at $t = \infty$; the extensions are continuous, by Lemma 7.36. Therefore, by Dini's theorem, it follows that $\langle \pi_0 f_N, \lambda_t\rangle \to \langle \pi_0, \lambda_t\rangle$ uniformly, which implies that $\sup_{t\geq 0} |T_N^1(t)| \to 0$ as desired.

Step 2. Estimate on $\mathcal{T}_N^4$. As in (7.79), we have the equality, for all $t \geq 0$

$$\mathcal{T}_N^4(t) = -M_t^N \mathbb{1}\left(M_t^N \leq \frac{\xi_N}{N}\right) - \frac{1}{N} \sum_{j\geq 2: C_j(G_t^N)\geq \xi_N} \pi_0(\mathcal{C}_j(G_t^N)) \tag{7.179}$$

where we have used the coupling of the random graphs $(G_t^N)_{t\geq 0}$ to the stochastic coagulant. Therefore, we have the uniform bound

$$\sup_{t\geq 0} \left|\mathcal{T}_N^4(t)\right| = \langle \varphi^2, \lambda_0^N\rangle^{\frac{1}{2}} \left(\sup_{t\geq 0} \frac{1}{N} \sum_{j\geq 2: C_j(G_t^N)\geq \xi_N} C_j(G_t^N)\right)^{\frac{1}{2}} + \frac{\xi_N}{N} \tag{7.180}$$

which converges to 0, by Lemma 7.39, (B2.), and because $\xi_N \ll N$.

Step 3. Construction of ξ_N, and convergence of $\mathcal{T}_N^2$. To conclude the proof of the supercritical case, it remains to show how a sequence ξ_N can be constructed such that $\mathcal{T}_N^2 \to 0$ uniformly, in probability. Recalling the definitions of $\tilde{f}_r$ above, let $A_{r,N}$ be the events

$$A_{r,N} = \left\{ \sup_{t\geq 0} |\langle \pi_0 \widetilde{f}_r, \lambda_t^N - \lambda_t \rangle| < \frac{1}{r} \right\}. \tag{7.181}$$

Then, as $N \to \infty$ with r fixed, $\mathbb{P}(A^1_{r,N}) \to 1$ by Lemma 7.37. We now define N_r inductively for $r \geq 1$ inductively, as in Lemma 7.18, by setting $N_1 = 1$ and letting N_{r+1} be the minimal $N > N_r$ such that, for all $N' \geq N$,

$$N \geq (r+1)^2; \qquad \mathbb{P}(A_{r+1,N'}) > \frac{r}{r+1}. \tag{7.182}$$

Now, we set $\xi_N = r$ for $N \in [N_r, N_{r+1})$. It follows that ξ_N satisfies the requirements above, and

$$\mathbb{P}\left(\sup_{t\geq 0} |\mathcal{T}_N^2| < \frac{1}{\xi_N} \right) \geq \mathbb{P}\left(A^1_{\xi_N,N} \right) > 1 - \frac{1}{\xi_N} \to 1 \tag{7.183}$$

Therefore, with this choice of ξ_N, $\mathcal{T}_N^2 \to 0$ uniformly in probability on $t \geq 0$.

This concludes the proof for the 0^{th} coordinate M_t^N; the 1^{st} - n^{th} coordinates are identical. For the remaining m coordinates, we replace f_N by $\frac{1}{2}(f_N(x) + f_N(Rx))$, which makes $\mathcal{T}_N^1$ identically 0 by symmetry, and use the bound $\pi_i(x)^2 \leq c\varphi(x)^2$ in estimating $\mathcal{T}_N^4$. □

Proof of Lemma 7.41. We now turn to the case where, instead of considering the largest cluster, we sum over the (possibly empty) set of clusters of size at least ξ_N, for a deterministic sequence ξ_N. In this way, we have

$$\widetilde{g}_t^N = \langle \pi \mathbb{1}[\pi_0 \geq \xi_N], \lambda_t^N \rangle. \tag{7.184}$$

Let us write $h_N(x) = \mathbb{1}[\pi_0(x) < \xi_N]$, so that $\widetilde{g}_t^N = \langle \pi, \lambda_0^N \rangle - \langle \pi h_N, \lambda_t^N \rangle$. With this notation,

$$g_t^N - \widetilde{g}_t^N = \langle \pi h_N, \lambda_t^N \rangle - \left(\langle \pi, \lambda_0^N \rangle - g_t^N \right) \tag{7.185}$$

is exactly the term $\mathcal{T}_N^4$ estimated in the proofs of Lemma 7.18, 7.40, for the new choice of ξ_N. The estimate (7.180) therefore applies to bound $\sup_{t\geq 0} \left| g_t^N - \widetilde{g}_t^N \right|$, and the hypotheses on ξ_N are sufficient to guarantee that the right-hand side converges to 0 in probability. □

7.11.2 Proof of Lemma 7.39

We now turn to the proof of Lemma 7.39; our strategy is as follows. First, we prove uniform convergence on compact subsets $I \subset (t_g, \infty)$ in Lemma 7.42. We will then show how this may be extended to the whole interval $[0, \infty)$, by arguing separately for an initial interval $[0, t_-]$ and for large times $[t_+, \infty)$.

Lemma 7.42. *Let G^N_t and ξ_N be as above. Fix a compact subset $I \subset (t_{\rm g}, \infty)$. Then we have the convergence*

$$\sup_{t\in I}\left[\frac{1}{N}\sum_{j\geq 2: C_j(G^N_t)\geq \xi_N} C_j(G^N_t)\right] \to 0 \qquad \textit{in probability.} \tag{7.186}$$

Proof of Lemma 7.42. It is sufficient to show that for every $t > t_{\rm g}$ the claim holds for some I of the form $I = (t_-, t_+) \subset (t_{\rm g}, \infty)$ containing t. As in Theorem 7.30, let $\widehat{\lambda}^t_0$ be the measure on S given by $\widehat{\lambda}^t_0(dx) = (1 - \kappa_t(x))\lambda_0(dx)$. We also write $\widehat{t}_{\rm g}(t)$ for the gelation time of the solution $(\widehat{\lambda}^t_s)_{s\geq 0}$ to (Fl) starting at $\widehat{\lambda}^t_0$. We showed in the proof of Theorem 7.30 that, for all $t > t_{\rm g}$, $\widehat{t}_{\rm g}(t) > t$, and the map $t \mapsto \widehat{t}_{\rm g}(t)$ is continuous. Therefore, for any $t > t_{\rm g}$, we can choose $t_\pm$ such that

$$t_{\rm g} < t_- < t < t_+ < \widehat{t}_{\rm g}(t_-). \tag{7.187}$$

We form $\widetilde{G}^N_{t_-}$ from G^N_{t-} by deleting all vertexes of the giant component of $C_1(G^N_{t_-})$. We now form a new graph, $\widetilde{G}^N_{t_-,t_+}$ by including all edges between vertexes of $\widetilde{G}^N_{t_-}$ which are present in the graph $G^N_{t_+}$.

From Theorem 7.50 and Lemma 7.31, we can construct a sequence $\mathbf{y}_N, N \geq 1$ satisfying Hypothesis 7.2 for $\widehat{\lambda}^{t_-}_0$ and random graphs $\widehat{G}^N_{t_-} \sim \mathcal{G}(\mathbf{y}_N, t_-K/N)$, such that

$$\mathbb{P}\left(\widehat{G}^N_{t_-} = \widetilde{G}^N_{t_-}\right) \to 1. \tag{7.188}$$

We now form $\widehat{G}^N_{t_-,t_+}$ from $\widehat{G}^N_{t_-}$ by adding those edges present in $G^N_{t_+}$. By the Markov property of the graph process $(G^N_s)_{t\geq 0}$, these edges are independent of the construction of $\widehat{G}^N_{t_-}$, and so $\widehat{G}^N_{t_-,t_+} \sim \mathcal{G}(\mathbf{y}_N, t_+K/N)$.

Since Hypothesis 7.2 applies to $\mathbf{y}_N$ and $\widehat{\lambda}^{t_-}_0$, Lemma 7.16 shows that the critical time for $\mathcal{G}(\mathbf{y}_N, tK/N)$ is exactly the gelation time of $(\widehat{\lambda}^{t_-}_s)_{s\geq 0}$, which we have written as $\widehat{t}_{\rm g}(t_-)$. By the choices of $t_\pm$, $t_+ < \widehat{t}_{\rm g}(t_-)$, and in particular, $\widehat{G}^N_{t_-,t_+}$ is still subcritical. By construction,

$$\mathbb{P}\left(\widehat{G}^N_{t_-,t_+} = \widetilde{G}^N_{t_-,t_+}\right) \to 1. \tag{7.189}$$

For $s \in [t_-, t_+]$, let $\mathcal{C}'_1(G^N_s)$ be the connected component of G^N_s which contains $\mathcal{C}_1(G^N_{t_-})$, and let $C'_1(G^N_s)$ be its size. By definition, $C'_1(G^N_s) \leq C_1(G^N_s)$ and so

$$\sum_{j\geq 2} C_j(G^N_s)1\!\!1\left[C_j(G^N_s) \geq \xi_N\right] \quad\leq\quad \sum_{j\geq 1} C_j(G^N_s)1\!\!1\left[C_j(G^N_s) \geq \xi_N, \mathcal{C}_j(G^N_s) \neq \mathcal{C}'_1(G^N_s)\right]. \tag{7.190}$$

Moreover, the right-hand side is increasing as s runs over $[t_-, t_+]$, since it can be rewritten as

$$.... = \sum_{i=1}^{l^N} 1\!\!1\left[\exists j : i \in \mathcal{C}_j(G^N_s),\ C_j(G^N_s) \geq \xi_N,\ i \notin \mathcal{C}_1(G^N_{t_-})\right] \tag{7.191}$$

and each summand can only increase in s as the clusters grow. Evaluating at the endpoint t_+, the construction of $\widetilde{G}^N_{t_-,t_+}$ gives

$$\sum_{j\geq 1} C_j(G^N_{t_+})\mathbb{1}\left[C_j(G^N_{t_+} \geq \xi_N, \mathcal{C}_j(G^N_{t_+}) \neq \mathcal{C}_1'(G^N_{t_+})\right] = \sum_{j\geq 1} C_j(\widetilde{G}^N_{t_-,t_+})\mathbb{1}\left[C_j(\widetilde{G}^N_{t_-,t_+}) \geq \xi_N\right]. \tag{7.192}$$

Combining (7.189, 7.190, 7.192) we see that, with high probability,

$$\sup_{s\in[t_-,t_+]}\left[\frac{1}{N}\sum_{j\geq 2: C_j(G^N_s)\geq \xi_N} C_j(G^N_s)\right] \leq \frac{1}{N}C_1(\widehat{G}^N_{t_-,t_+}) + \frac{1}{N}\sum_{j\geq 2: C_j(\widehat{G}^N_{t_-,t_+})\geq\xi_N} C_j(\widehat{G}^N_{t_-,t_+}). \tag{7.193}$$

The first term of the right-hand side converges to 0 in probability because $\widehat{G}^N_{t_-,t_+}$ is subcritical, and the second term converges to 0 in probability by Theorem 7.48. □

Proof of Lemma 7.39. For λ_0 as in the hypothesis, let M_t be the mass of the gel associated to the solution $(\lambda_t)_{t\geq 0}$ to (Fl). Fix $\epsilon > 0$; without loss of generality, assume that $\epsilon < 1$. By continuity from Lemma 7.33 and Lemma 7.36, we can choose $t_\pm \in (t_{\mathrm{g}}, \infty)$ such that

$$M_{t_-} < \frac{\epsilon}{3}; \qquad M_{t_+} > \lambda_0(S) - \frac{\epsilon}{3}. \tag{7.194}$$

Consider now the events

$$A^1_N = \left\{\frac{1}{N}C_1(G^N_{t_-}) < \frac{2\epsilon}{3}; \quad \frac{1}{N}C_1(G^N_{t_+}) > \lambda_0(S) - \frac{\epsilon}{2}; \quad \langle \pi_0, \lambda^N_0\rangle < \lambda_0(S) + \frac{\epsilon}{2}\right\}; \tag{7.195}$$

$$A^2_N = \left\{\frac{1}{N}\sum_{j\geq 2: C_j(G^N_{t_-})\geq\xi_N} C_j(G^N_{t_-}) < \frac{\epsilon}{3}\right\}. \tag{7.196}$$

Thanks to the coupling described in Section 7.5, Lemma 7.18 implies that $\mathbb{P}(A^1_N) \to 1$, and $\mathbb{P}(A^2_N) \to 1$ from Theorem 7.48. On the event $A^1_N \cap A^2_N$, we bound as follows.

i). For the initial interval $[0, t_-]$, an argument similar to that of Lemma 7.42 shows that, on this event,

$$\sup_{t\in[0,t_-]}\frac{1}{N}\sum_{\substack{j\geq 2:\\ C_j(G^N_t)\geq\xi_N}} C_j(G^N_t) \leq \frac{1}{N}\sum_{\substack{j\geq 1:\\ C_j(G^N_{t_-})\geq\xi_N}} C_j(G^N_{t_-}) = \frac{1}{N}C_1(G^N_{t_-}) + \frac{1}{N}\sum_{\substack{j\geq 2:\\ C_j(G^N_{t_-})\geq\xi_N}} C_j(G^N_{t_-}) < \epsilon. \tag{7.197}$$

ii). For late times $t \in [t_+, \infty)$, the largest cluster $\mathcal{C}_1(G^N_t)$ is at least the size of the cluster containing $\mathcal{C}_1(G^N_{t_+})$. Therefore,

$$\inf_{t\geq t_+}\frac{1}{N}C_1(G^N_t) \geq \frac{1}{N}C_1(G^N_{t_+}) > \lambda_0(S) - \frac{\epsilon}{2} \tag{7.198}$$

and so

$$\sup_{t\geq t_+}\left[\frac{1}{N}\sum_{j\geq 2:C_j(G_t^N)\geq \xi_N} C_j(G_t^N)\right] \leq \sup_{t\geq t_+}\left[\frac{1}{N}\sum_{j\geq 2} C_j(G_t^N)\right] \leq \langle \pi_0, \lambda_0^N\rangle - \frac{1}{N}C_1(G_{t_+}^N) < \epsilon. \tag{7.199}$$

Now, consider the events

$$A_N^3 = \left\{\sup_{t\in[t_-,t_+]}\left[\frac{1}{N}\sum_{j\geq 2:C_j(G_t^N)\geq \xi_N} C_j(G_t^N)\right] < \epsilon\right\}; \tag{7.200}$$

$$A_N = A_N^1 \cap A_N^2 \cap A_N^3. \tag{7.201}$$

By Lemma 7.42, $\mathbb{P}(A_N^3) \to 1$, and so $\mathbb{P}(A_N) \to 1$. On the event A_N, we have

$$\sup_{t\geq 0}\left[\frac{1}{N}\sum_{j\geq 2:C_j(G_t^N)\geq \xi_N} C_j(G_t^N)\right] < \epsilon \tag{7.202}$$

which proves the claimed convergence in probability. □

Appendix

7.A Weak Formulation of Smoluchowski and Flory Equations

Throughout, we work with the weak formulation of the Smoluchowski and Flory equations described in the introduction. In order to make sense of every term for a putative solution $(\lambda_t)_{t<T}$, we ask for the following conditions to hold.

i). For all Borel sets $A \subset S$, the map $t \mapsto \lambda_t(A)$ is measurable;

ii). For all bounded, measurable functions $f : S \to \mathbb{R}_+$ belonging to $\mathcal{A}$, $\langle f, \lambda_0 \rangle < \infty$;

iii). For all compact subsets $S' \subset S$ and all $t < T$,

$$\int_0^t ds \int_{S' \times S} \overline{K}(x, y)\lambda_s(dx)\lambda_s(dy) < \infty; \tag{7.203}$$

If these hold, then we say can make sense of the following weak form of the Smoluchowski equation (Sm).

iv). For all $f \in \mathcal{A}$ and $t < T$,

$$\langle f, \lambda_t \rangle = \langle f, \lambda_0 \rangle + \int_0^t \langle f, L(\lambda_s) \rangle ds. \tag{7.204}$$

7.B Introduction to Inhomogenous Random Graphs

As discussed in the introduction, the connection between gelation and random graphs is well-understood, and the multiplicative kernel corresponds to the well-known Erdős-Réyni random graphs [78, 75, 5]. However, for our purposes, not all particles are equal: particles with large values of $\pi_i(x)$ will undergo more collisions and exhibit quantitatively different behaviour, and so we will need a more sophisticated model of random graphs to accommodate this inhomogeneity. In this section, we will review the theory of *inhomogenous random graphs* developed in [28], which will play the same rôle for our model that the Erdős-Réyni model does for the multiplicative kernel. We now summarise the key definitions and results from [28] which we use in our work.

Definition 7.B.1. *A* generalised vertex space *is a triple* $\mathcal{V} = (\mathcal{S}, m, (\mathbf{x}_N)_{N\geq 1})$*, consisting of*

- *A separable metric space* $\mathcal{S}$*, equipped with its Borel* σ*-algebra;*
- *A measure* m *on* $\mathcal{S}$*, with* $m(\mathcal{S}) \in (0, \infty)$*;*
- *A family of random variables* $\mathbf{x}_N = (x_1^{(N)}, ..., x_{l^N}^{(N)})$ *taking values in* $\mathcal{S}$*, and of potentially random length* l^N*, such that the empirical measures*

$$m_N = \frac{1}{N}\sum_{k=1}^{l^N} \delta_{x_k^{(N)}} \tag{7.205}$$

converge to m *in the weak topology* $\mathcal{F}(C_b(\mathcal{S}))$*, in probability.*

In the special case where $m(\mathcal{S}) = 1$ *and* $l^N = N$*, we say that* $(\mathcal{S}, m, (\mathbf{x}_N)_{N\geq 1})$ *is a* vertex space.

Definition 7.B.2. *A* kernel *is a symmetric, measurable map* $k : \mathcal{S} \times \mathcal{S} \to [0, \infty)$*. We say that* k *is* irreducible *if, whenever* $A \subset S$ *is such that* $k(x, y) = 0$ *for all* $x \in A$ *and* $y \in A^{\mathrm{c}}$*, then either* $m(A) = 0$ *or* $m(A^{\mathrm{c}}) = 0$*.*

Definition 7.B.3 (Inhomogenous random graphs)**.** *Given a kernel* k *and a generalised vertex space* $\mathcal{V}$*, we let* G^N *be a random graph on* $\{1, 2, .., l^N\}$ *given as follows. Conditional on the values of* $\mathbf{x}_N$*, the edge* $e = (ij)$ *is included with probability*

$$p_{ij} = 1 - \exp\left(-\frac{k(x_i^{(N)}, x_j^{(N)})}{N}\right) \tag{7.206}$$

and such that the presence of different edges is (conditionally) independent. We write $G^N \sim \mathcal{G}^{\mathcal{V}}(N, k/N)$*. We also consider the* vertex data $\mathbf{x}_N = (x_i^{(N)})_{i=1}^{l^N}$ *to be part of the data of* G_t^N*, so that an equality of random graphs* $G = G'$ *includes the equality of the vertex data.*

Remark 7.43. *This differs slightly from the main definition in [28], but is rather one of the alternatives considered in [28, Remark 2.4]*

To treat a general class of kernels k, additional regularity is required, to prevent pathologies. This is the content of the following defintion:

Definition 7.B.4 (Graphical Kernel)**.** *We say that a kernel* k *on a vertex space* $\mathcal{V} = (\mathcal{S}, m, (\mathbf{v}_N)_{N\geq 1})$ *is* graphical *if the following hold.*

i). k *is almost everywhere continuous on* $\mathcal{S} \times \mathcal{S}$;

ii). $k \in L^1(\mathcal{S} \times \mathcal{S}, m \otimes m)$*;*

iii). If $G^N \sim \mathcal{G}^{\mathcal{V}}(N, k/N)$, then

$$\frac{1}{N}\mathbb{E}\left[e\left(G^N\right)\right] \to \frac{1}{2}\int_{\mathcal{S}\times\mathcal{S}} k(v,w)m(dv)m(dw) \tag{7.207}$$

where $e(\cdot)$ denotes the number of edges of the graph.

Definition 7.B.5. *Given a graph G, we write $\mathcal{C}_j(G) : j = 1, 2...$ for the connected components of G, in decreasing order of their sizes $\#\mathcal{C}_j(G) = C_j(G)$. If there are fewer than j connected components, then $\mathcal{C}_j(G) = \emptyset$ and $C_j(G) = 0$.*

The phase transition is given in terms of the convolution operator

$$(Tf)(v) = \int_S k(v,w)f(w)m(dw) \tag{7.208}$$

for functions f such that the right-hand side is defined (i.e., finite or $+\infty$) for m-almost all v; for instance, if $f \geq 0$ then Tf is well-defined, possibly taking the value ∞. We define

$$\|T\| = \sup\{\|Tf\|_{L^2(m)} : \|f\|_{L^2(m)} \leq 1, f \geq 0\}. \tag{7.209}$$

If T defines a bounded linear map from $L^2(m)$ to itself, then $\|T\|$ is precisely its operator norm in this setting; otherwise, $\|T\| = \infty$. It is straightforward to show that if $k \in L^2(S \times S, m \otimes m)$ then $T : L^2(m) \to L^2(m)$ is a Hilbert-Schmidt operator, and that $\|T\|_{\mathrm{HS}} = \|k\|_{L^2(m)} < \infty$. In this case, $\|T\|$ is certainly finite, and is the operator norm of $T : L^2(m) \to L^2(m)$. The example of interest to us will fall into this case.

The analysis of the random graphs uses a branching process, similar to that used in the standard analysis of Erdős-Rényi graphs. Many quantities of the graph can be expressed in terms of the 'survival probability' $\kappa(k, v)$ when the data v of the first vertex in the branching process is given. To avoid the unnecessary complication of making this into a precise definition, we use the following characterisation, which is equivalent by [28, Theorem 6.2].

Theorem 7.44. *Let k be an irreducible kernel on a generalised vertex space $\mathcal{V}$, such that $k \in L^1(\mathcal{S} \times \mathcal{S}, m \times m)$, and such that, for all x,*

$$\int_S k(x,y)m(dy) < \infty. \tag{7.210}$$

Consider the nonlinear fixed-point equation

$$\forall x \in S, \qquad \kappa(x) = 1 - e^{-(T\kappa)(x)} \tag{7.211}$$

where T is the convolution operator (7.53). Then (7.211) has a maximal solution $\kappa_k(x) = \kappa(k; x)$; that is, for any other solution $\tilde{\kappa}$,

$$\forall x \in S, \qquad \tilde{\kappa}(x) \leq \kappa(k, x). \tag{7.212}$$

It therefore follows that $0 \leq \kappa_k(x) \leq 1$ for all x. The maximal solution is necessarily unique, and so this uniquely defines κ_k. Moreover, we have the following dichotomy:

i). If $\|T\| \leq 1$, then $\kappa(k,x)=0$ for all x;

ii). If $\|T\|>1$, then $\kappa(k,x)>0$ for all m-almost all x.

This can be stated dynamically as follows. Consider the survival function 'at time t', given by $\kappa(tk,x)$, which we will write throughout as $\kappa_t(x)$. Then

- *If $t \leq \|T\|^{-1}$, then $\kappa_t(x)=0$ for all x;*
- *If $t > \|T\|^{-1}$, then $\kappa_t(x)>0$ for all x.*

We can now state the main results on the phase transition, given by [28, Theorem 3.1 and Corollary 3.2].

Theorem 7.45 (Phase Transition)**.** *Let k be a graphical and irreducible kernel for a vertex space $\mathcal{V}$, with $0<\|T\|<\infty$. Let $G^N \sim \mathcal{G}^{\mathcal{V}}(N,k/N)$ be random graphs on a common probability space. Then we have the convergence*

$$\frac{1}{N}C_1(G_t^N) \to \int_{\mathcal{S}} \kappa(tk,v)m(dv) \qquad \text{in probability.} \tag{7.213}$$

Therefore, if $(G_t^N)_{t\geq 0}$ is a dynamic family of random graphs $G_t^N \sim \mathcal{G}^{\mathcal{V}}(N,tk)$, then we have the following dichotomy:

i). If $t \leq t_{\mathrm{c}} = \|T\|^{-1}$, then there is no giant component, in particular

$$\frac{C_1(G_t^N)}{N} \to 0 \tag{7.214}$$

in probability.

ii). If $t > t_{\mathrm{c}} = \|T\|^{-1}$, then there is a giant component: there exists $c=c(t)>0$ such that

$$\mathbb{P}(C_1(G_t^N) > cN) \to 1. \tag{7.215}$$

Remark 7.46. *Following [28], based on this dichotomy, we say that*

i). G^N is subcritical *if $\|T\|<1$;*

ii). G^N is critical *if $\|T\|=1$;*

iii). G^N is supercritical *if $\|T\|>1$.*

The next result characterises t_{g} in terms of the point spectrum $\sigma_{\mathrm{p}}(T)$ as an operator on $L^2(m)$, and appears as [28, Lemma 5.15]

Theorem 7.47 (Spectrum of T)**.** *Let $\mathcal{V}$ be a generalised vertex space and k be a graphical, irreducible kernel on $\mathcal{V}$ such that $k \in L^2(S \times S, m \times m)$. Then the operator T defined in (7.53) has an eigenvalue $t_c^{-1} = \|T\|$ in $L^2(m)$, and the corresponding eigenspace is 1-dimensional. Moreover, there exists an eigenfunction f such that $f > 0$ m-almost everywhere.*

The third result we will recall is [28, Theorem 3.6], which considers clusters of a scale $\xi_N \ll N$, excluding the largest cluster. We term these *mesoscopic* clusters.

Theorem 7.48. *Let $G^N \sim \mathcal{G}^{\mathcal{V}}(N, k/N)$, for a (generalised) vertex space $\mathcal{V}$ and an irreducible graphical kernel k. Let ξ_N be a sequence with*

$$\xi_N \to \infty; \qquad \frac{\xi_N}{N} \to 0. \tag{7.216}$$

Then

$$\frac{1}{N} \sum_{j \geq 2: C_j(G^N) \geq \xi_N} C_j(G^N) \to 0 \tag{7.217}$$

in probability.

We will also make use of the following monotonicity and continuity properties, from [28, Theorem 6.4].

Theorem 7.49. *Let k be a kernel on a vertex space $\mathcal{V}$, and let $\kappa_t(\cdot) = \kappa(tk, \cdot)$ be the survival function defined above. Then the map $t \mapsto \kappa_t(\cdot)$ is monotonically increasing, in the sense that for all $0 \leq s \leq t$ and for all x, $\kappa_s(x) \leq \kappa_t(x)$. We also have the following continuity property. Let $t_n \to t$ be a monotone sequence, either increasing or decreasing. Then*

$$\kappa_{t_n}(x) \to \kappa_t(x) \qquad \text{for } m\text{- almost all } x, \text{ and} \tag{7.218}$$

$$\int_S \kappa_{t_n}(x) m(dx) \to \int_S \kappa_t(x) m(dx). \tag{7.219}$$

The final result which we will need is a 'duality' result, connecting the supercritical and subcritical behaviours. This is given by [28, Theorem 12.1].

Theorem 7.50. *Let k be an irreducible graphical kernel on a generalised vertex space $\mathcal{V}$, such that $\|T\| > 1$. Let $G^N \sim \mathcal{G}^{\mathcal{V}}(N, k/N)$, and form $\widetilde{G}^N$ by deleting all vertexes in the largest component $\mathcal{C}_1(G^N)$. Then, defined on the same underlying probability space, there is a generalised vertex space $\widehat{\mathcal{V}} = (\mathcal{S}, \widehat{m}, (\mathbf{y}_N)_{N\geq 1})$ with*

$$\widehat{m}(dx) = (1 - \rho(k; x)) m(dx) \tag{7.220}$$

and such that $\mathbf{y}_N$ is an enumeration of those x_i not belonging to the component $\mathcal{C}_1(G^N)$, and a random graph $\widehat{G}^N \sim \mathcal{G}^{\widehat{\mathcal{V}}}(N, k/N)$ such that

$$\mathbb{P}(\widetilde{G}^N = \widehat{G}^N) \to 1. \tag{7.221}$$

Furthermore, if $k \in L^2(\mathcal{S} \times \mathcal{S}, m \otimes m)$, then $\widehat{G}^N$ is subcritical.

We emphasise here that we have defined the equality $\widetilde{G}^N = \widehat{G}^N$ to include equality of the values x_i associated to each vertex; this follows from the construction in [28], since the values $\mathbf{y}_N$ associated to $\widehat{G}^N$ are exactly those x_i not belonging to the giant component. This generalises the standard 'duality result' of Bollobás [27] for Erdős-Rényi graphs.

Bibliography

[1] A. S. Ackleh, R. Lyons, and N. Saintier. A structured coagulation-fragmentation equation in the space of radon measures. In *2021 Joint Mathematics Meetings (JMM)*. AMS, 2021.

[2] S. Adams, N. Dirr, M. Peletier, and J. Zimmer. Large deviations and gradient flows. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 371(2005):20120341, 17, 2013.

[3] S. Adams, N. Dirr, M. A. Peletier, and J. Zimmer. From a large-deviations principle to the Wasserstein gradient flow: a new micro-macro passage. *Comm. Math. Phys.*, 307(3):791–815, 2011.

[4] D. Aldous. Stopping times and tightness. II. *Ann. Probab.*, 17(2):586–595, 1989.

[5] D. Aldous. Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists. *Bernoulli*, 5(1):3–48, 1999.

[6] D. J. Aldous. Book Review: Markov processes: Characterization and convergence. *Bull. Amer. Math. Soc. (N.S.)*, 16(2):315–318, 1987.

[7] R. Alexandre. Remarks on 3D Boltzmann linear equation without cutoff. *Transport Theory Statist. Phys.*, 28(5):433–473, 1999.

[8] R. Alexandre, L. Desvillettes, C. Villani, and B. Wennberg. Entropy dissipation and long-range interactions. *Arch. Ration. Mech. Anal.*, 152(4):327–355, 2000.

[9] R. Alexandre and C. Villani. On the Landau approximation in plasma physics. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 21(1):61–95, 2004.

[10] R. Alonso, V. Bagland, and B. Lods. Long time dynamics for the Landau-Fermi-Dirac equation with hard potentials. *J. Differential Equations*, 270:596–663, 2021.

[11] R. Alonso, I. M. Gamba, and M. Tasković. Exponentially-tailed regularity and time asymptotic for the homogeneous boltzmann equation. *arXiv preprint arXiv:1711.06596*, 2017.

[12] L. Arkeryd. On the Boltzmann equation. I. Existence. *Arch. Rational Mech. Anal.*, 45:1–16, 1972.

[13] A. A. Arsenev and O. E. Buryak. On a connection between the solution of the boltzmann equation and the solution of the Landau-Fokker-Planck equation. *Mat. Sb.*, 181(4):435–446, 1990.

[14] A. A. Arsenev and N. V. Peskov. Existence of a generalized solution to the landau equation. *Zhurnal Vychislitelnoi Matematiki i Matematicheskoi Fiziki*, 17:1063–1068, 1977.

[15] S. Banerjee, A. Budhiraja, and M. Perlmutter. A new approach to large deviations for the Ginzburg-Landau model. *Electron. J. Probab.*, 25:Paper No. 26, 51, 2020.

[16] G. Basile, D. Benedetto, L. Bertini, and C. Orrieri. Large deviations for Kac-like walks. *J. Stat. Phys.*, 184(1):Paper No. 10, 27, 2021.

[17] P. J. Blatz and A. V. Tobolsky. Note on the kinetics of systems manifesting simultaneous polymerization-depolymerization phenomena. *The journal of physical chemistry*, 49(2):77–80, 1945.

[18] A. V. Bobylëv. Exact solutions of the nonlinear Boltzmann equation and the theory of relaxation of a Maxwell gas. *Teoret. Mat. Fiz.*, 60(2):280–310, 1984.

[19] A. V. Bobylëv. The theory of the nonlinear spatially uniform Boltzmann equation for Maxwell molecules. In *Mathematical physics reviews, Vol. 7*, volume 7 of *Soviet Sci. Rev. Sect. C: Math. Phys. Rev.*, pages 111–233. Harwood Academic Publ., Chur, 1988.

[20] A. V. Bobylëv. Moment inequalities for the Boltzmann equation and applications to spatially homogeneous problems. *J. Statist. Phys.*, 88(5-6):1183–1214, 1997.

[21] A. V. Bobylëv and C. Cercignani. On the rate of entropy production for the Boltzmann equation. *J. Statist. Phys.*, 94(3-4):603–618, 1999.

[22] T. Bodineau, I. Gallagher, L. Saint-Raymond, and S. Simonella. One-sided convergence in the Boltzmann-Grad limit. *Ann. Fac. Sci. Toulouse Math. (6)*, 27(5):985–1022, 2018.

[23] T. Bodineau, I. Gallagher, L. Saint-Raymond, and S. Simonella. Fluctuation theory in the Boltzmann-Grad limit. *J. Stat. Phys.*, 180(1-6):873–895, 2020.

[24] T. Bodineau, I. Gallagher, L. Saint-Raymond, and S. Simonella. Statistical dynamics of a hard sphere gas: fluctuating boltzmann equation and large deviations. *arXiv preprint arXiv:2008.10403*, 2020.

[25] N. N. Bogoliubov. Problems of a dynamical theory in statistical physics. In *Studies in Statistical Mechanics, Vol. I*, pages 1–118. North-Holland, Amsterdam; Interscience, New York, 1962.

[26] F. Bolley, A. Guillin, and C. Villani. Quantitative concentration inequalities for empirical measures on non-compact spaces. *Probab. Theory Related Fields*, 137(3-4):541–593, 2007.

[27] B. Bollobás. The evolution of random graphs. *Trans. Amer. Math. Soc.*, 286(1):257–274, 1984.

[28] B. Bollobás, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Structures Algorithms*, 31(1):3–122, 2007.

[29] L. Boltzmann. *Lectures on gas theory.* University of California Press, Berkeley-Los Angeles, Calif., 1964. Translated by Stephen G. Brush.

[30] Ludwig Boltzmann. Weitere studien über das wärmegleichgewicht unter gasmolekülen. In *Kinetische Theorie II*, pages 115–225. Springer, 1970.

[31] F. Bouchet. Is the Boltzmann equation reversible? A large deviation perspective on the irreversibility paradox. *J. Stat. Phys.*, 181(2):515–550, 2020.

[32] A. Budhiraja, Y. Chen, and L. Xu. Large deviations of the entropy production rate for a class of Gaussian processes. *J. Math. Phys.*, 62(5):Paper No. 052702, 25, 2021.

[33] A. Budhiraja and M. Conroy. Empirical measure and small noise asymptotics under large deviation scaling for interacting diffusions. *Journal of Theoretical Probability*, pages 1–55, 2021.

[34] E. Buffet and J. V. Pulé. Polymers and random graphs. *J. Statist. Phys.*, 64(1-2):87–110, 1991.

[35] Russel E. Caflisch. The Boltzmann equation with a soft potential. I. Linear, spatially-homogeneous. *Comm. Math. Phys.*, 74(1):71–95, 1980.

[36] E. Carlen, M. C. Carvalho, and M. Loss. Many-body aspects of approach to equilibrium. In *Journées "Équations aux Dérivées Partielles" (La Chapelle sur Erdre, 2000)*, pages Exp. No. XI, 12. Univ. Nantes, Nantes, 2000.

[37] E. A. Carlen, M. C. Carvalho, and A. Einav. Entropy production inequalities for the Kac walk. *Kinet. Relat. Models*, 11(2):219–238, 2018.

[38] E. A. Carlen, M. C. Carvalho, J. Le Roux, M. Loss, and C. Villani. Entropy and chaos in the Kac model. *Kinet. Relat. Models*, 3(1):85–122, 2010.

[39] E. A. Carlen, M. C. Carvalho, and M. Loss. Determination of the spectral gap for Kac's master equation and related stochastic evolution. *Acta Math.*, 191(1):1–54, 2003.

[40] E. A. Carlen, J. S. Geronimo, and M. Loss. Determination of the spectral gap in the Kac model for physical momentum and energy-conserving collisions. *SIAM J. Math. Anal.*, 40(1):327–364, 2008.

[41] K. Carrapatoso. Exponential convergence to equilibrium for the homogeneous Landau equation with hard potentials. *Bull. Sci. Math.*, 139(7):777–805, 2015.

[42] J. A. Carrillo, M. G. Delgadino, L. Desvillettes, and J. Wu. The landau equation as a gradient flow. *arXiv preprint arXiv:2007.08591*, 2020.

[43] C. Cercignani. H-theorem and trend to equilibrium in the kinetic theory of gases. *Arch. Mech. (Arch. Mech. Stos.)*, 34(3):231–241 (1983), 1982.

[44] C. Cercignani. *The Boltzmann equation and its applications*, volume 67 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1988.

[45] H. Chen, W. Li, and C. Xu. Gevrey regularity for solution of the spatially homogeneous Landau equation. *Acta Math. Sci. Ser. B (Engl. Ed.)*, 29(3):673–686, 2009.

[46] H. Chen, W. Li, and C. Xu. Analytic smoothness effect of solutions for spatially homogeneous Landau equation. *J. Differential Equations*, 248(1):77–94, 2010.

[47] Y. Chen and L. He. Smoothing estimates for Boltzmann equation with full-range interactions: spatially homogeneous case. *Arch. Ration. Mech. Anal.*, 201(2):501–548, 2011.

[48] R. Cortez and J. Fontbona. Quantitative uniform propagation of chaos for Maxwell molecules. *Comm. Math. Phys.*, 357(3):913–941, 2018.

[49] R. W. R. Darling and J. R. Norris. Differential equation approximations for Markov chains. *Probab. Surv.*, 5:37–79, 2008.

[50] P. Degond and B. Lucquin-Desreux. The Fokker-Planck asymptotics of the Boltzmann collision operator in the Coulomb case. *Math. Models Methods Appl. Sci.*, 2(2):167–182, 1992.

[51] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*, volume 38 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 1998.

[52] L. Desvillettes. On asymptotics of the Boltzmann equation when the collisions become grazing. *Transport Theory Statist. Phys.*, 21(3):259–276, 1992.

[53] L. Desvillettes. Some applications of the method of moments for the homogeneous Boltzmann and Kac equations. *Arch. Rational Mech. Anal.*, 123(4):387–404, 1993.

[54] L. Desvillettes. Entropy dissipation estimates for the Landau equation in the Coulomb case and applications. *J. Funct. Anal.*, 269(5):1359–1403, 2015.

[55] L. Desvillettes, C. Graham, and S. Méléard. Probabilistic interpretation and numerical approximation of a Kac equation without cutoff. *Stochastic Process. Appl.*, 84(1):115–135, 1999.

[56] L. Desvillettes and C. Mouhot. Stability and uniqueness for the spatially homogeneous Boltzmann equation with long-range interactions. *Arch. Ration. Mech. Anal.*, 193(2):227–253, 2009.

[57] L. Desvillettes, C. Mouhot, and C. Villani. Celebrating Cercignani's conjecture for the Boltzmann equation. *Kinet. Relat. Models*, 4(1):277–294, 2011.

[58] L. Desvillettes and C. Villani. On the spatially homogeneous Landau equation for hard potentials. I. Existence, uniqueness and smoothness. *Comm. Partial Differential Equations*, 25(1-2):179–259, 2000.

[59] L. Desvillettes and C. Villani. On the spatially homogeneous Landau equation for hard potentials. II. H-theorem and applications. *Comm. Partial Differential Equations*, 25(1-2):261–298, 2000.

[60] L. Desvillettes and B. Wennberg. Smoothness of the solution of the spatially homogeneous Boltzmann equation without cutoff. *Comm. Partial Differential Equations*, 29(1-2):133–155, 2004.

[61] G. Di Blasio. Differentiability of spatially homogeneous solutions of the Boltzmann equation in the non Maxwellian case. *Comm. Math. Phys.*, 38:331–340, 1974.

[62] B. Djehiche and A. Schied. Large deviations for hierarchical systems of interacting jump processes. *J. Theoret. Probab.*, 11(1):1–24, 1998.

[63] R. L. Dobrušin. Vlasov equations. *Funktsional. Anal. i Prilozhen.*, 13(2):48–58, 96, 1979.

[64] M. H. Duong, V. Laschos, and M. Renger. Wasserstein gradient flows from large deviations of many-particle limits. *ESAIM Control Optim. Calc. Var.*, 19(4):1166–1188, 2013.

[65] P. Dupuis and R. S. Ellis. *A weak convergence approach to the theory of large deviations.* Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, 1997. A Wiley-Interscience Publication.

[66] P. Dupuis, K. Ramanan, and W. Wu. Large deviation principle for finite-state mean field interacting particle systems. *arXiv preprint arXiv:1601.06219*, 2016.

[67] A. Einav. An improved upper bound on the entropy production for the kac master equation. *Kinetic and Related Models*, 2010.

[68] A. Einav. On Villani's conjecture concerning entropy production for the Kac master equation. *Kinet. Relat. Models*, 4(2):479–497, 2011.

[69] A. Einav. A counter example to Cercignani's conjecture for the d dimensional Kac model. *J. Stat. Phys.*, 148(6):1076–1103, 2012.

[70] N. El Karoui and J. Lepeltier. Représentation des processus ponctuels multivariés à l'aide d'un processus de Poisson. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 39(2):111–133, 1977.

[71] N. El Karoui and S. Méléard. Martingale measures and stochastic calculus. *Probab. Theory Related Fields*, 84(1):83–101, 1990.

[72] T. Elmroth. Global boundedness of moments of solutions of the Boltzmann equation for forces of infinite range. *Arch. Rational Mech. Anal.*, 82(1):1–12, 1983.

[73] M. Erbar. A gradient flow approach to the boltzmann equation. *arXiv preprint arXiv:1603.00540*, 2016.

[74] M. Erbar, J. Maas, and D. R. M. Renger. From large deviations to Wasserstein gradient flows in multiple dimensions. *Electron. Commun. Probab.*, 20:no. 89, 12, 2015.

[75] P. Erdős and A. Rényi. On the evolution of random graphs. *Bull. Inst. Internat. Statist.*, 38:343–347, 1961.

[76] M. Escobedo and S. Mischler. Scalings for a ballistic aggregation equation. *J. Stat. Phys.*, 141(3):422–458, 2010.

[77] J. Feng and T. G. Kurtz. *Large deviations for stochastic processes*, volume 131 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2006.

[78] P. J. Flory. Molecular size distribution in three dimensional polymers. i. gelation1. *Journal of the American Chemical Society*, 63(11):3083–3090, 1941.

[79] F. Fournier and S. Méléard. A stochastic particle numerical method for 3D Boltzmann equations without cutoff. *Math. Comp.*, 71(238):583–604, 2002.

[80] N. Fournier. Existence and regularity study for two-dimensional Kac equation without cutoff by a probabilistic approach. *Ann. Appl. Probab.*, 10(2):434–462, 2000.

[81] N. Fournier. Uniqueness for a class of spatially homogeneous Boltzmann equations without angular cutoff. *J. Stat. Phys.*, 125(4):927–946, 2006.

[82] N. Fournier. Uniqueness of bounded solutions for the homogeneous Landau equation with a Coulomb potential. *Comm. Math. Phys.*, 299(3):765–782, 2010.

[83] N. Fournier. Finiteness of entropy for the homogeneous Boltzmann equation with measure initial condition. *Ann. Appl. Probab.*, 25(2):860–897, 2015.

[84] N. Fournier. On Exponential Moments of the Homogeneous Boltzmann Equation for Hard Potentials Without Cutoff. *Comm. Math. Phys.*, 387(2):973–994, 2021.

[85] N. Fournier and H. Guérin. On the uniqueness for the spatially homogeneous Boltzmann equation with a strong angular singularity. *J. Stat. Phys.*, 131(4):749–781, 2008.

[86] N. Fournier and H. Guérin. Well-posedness of the spatially homogeneous Landau equation for soft potentials. *J. Funct. Anal.*, 256(8):2542–2560, 2009.

[87] N. Fournier and A. Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields*, 162(3-4):707–738, 2015.

[88] N. Fournier and A. Guillin. From a Kac-like particle system to the Landau equation for hard potentials and Maxwell molecules. *Ann. Sci. Éc. Norm. Supér. (4)*, 50(1):157–199, 2017.

[89] N. Fournier and M. Hauray. Propagation of chaos for the Landau equation with moderately soft potentials. *Ann. Probab.*, 44(6):3581–3660, 2016.

[90] N. Fournier and D. Heydecker. Stability, well-posedness and regularity of the homogeneous landau equation for hard potentials. In *Annales de l'Institut Henri Poincaré C, Analyse non linéaire*. Elsevier, 2021.

[91] N. Fournier and S. Méléard. A Markov process associated with a Boltzmann equation without cutoff and for non-Maxwell molecules. *J. Statist. Phys.*, 104(1-2):359–385, 2001.

[92] N. Fournier and S. Mischler. Rate of convergence of the Nanbu particle system for hard potentials and Maxwell molecules. *Ann. Probab.*, 44(1):589–627, 2016.

[93] N. Fournier and C. Mouhot. On the well-posedness of the spatially homogeneous Boltzmann equation with a moderate angular singularity. *Comm. Math. Phys.*, 289(3):803–824, 2009.

[94] M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, second edition, 1998. Translated from the 1979 Russian original by Joseph Szücs.

[95] T. Funaki. The diffusion approximation of the spatially homogeneous Boltzmann equation. *Duke Math. J.*, 52(1):1–23, 1985.

[96] A. Gabrielov, V. Keilis-Borok, Y. Sinai, and I. Zaliapin. Statistical properties of the cluster dynamics of the systems of statistical mechanics. In *Boltzmann's legacy*, ESI Lect. Math. Phys., pages 203–215. Eur. Math. Soc., Zürich, 2008.

[97] I. Gallagher, L. Saint-Raymond, and B. Texier. From newton to boltzmann: the case of short-range potentials. *Preprint*, 2012.

[98] I. Gallagher, L. Saint-Raymond, and B. Texier. *From Newton to Boltzmann: hard spheres and short-range potentials.* Zurich Lectures in Advanced Mathematics. European Mathematical Society (EMS), Zürich, 2013.

[99] F. Golse, C. Imbert, C. Mouhot, and A. F. Vasseur. Harnack inequality for kinetic Fokker-Planck equations with rough coefficients and application to the Landau equation. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, 19(1):253–295, 2019.

[100] F. Golse and F. Poupaud. Un résultat de compacité pour l'équation de Boltzmann avec potentiel mou. Application au problème de demi-espace. *C. R. Acad. Sci. Paris Sér. I Math.*, 303(12):583–586, 1986.

[101] T. Goudon. On Boltzmann equations and Fokker-Planck asymptotics: influence of grazing collisions. *J. Statist. Phys.*, 89(3-4):751–776, 1997.

[102] H. Grad. Asymptotic theory of the Boltzmann equation. *Phys. Fluids*, 6:147–181, 1963.

[103] S. E. Graversen and G. Peskir. Maximal inequalities for the Ornstein-Uhlenbeck process. *Proc. Amer. Math. Soc.*, 128(10):3035–3041, 2000.

[104] F. A. Grünbaum. Propagation of chaos for the Boltzmann equation. *Arch. Rational Mech. Anal.*, 42:323–345, 1971.

[105] H. Guérin. Existence and regularity of a weak function-solution for some Landau equations with a stochastic approach. *Stochastic Process. Appl.*, 101(2):303–325, 2002.

[106] H. Guérin. Solving Landau equation for some soft potentials through a probabilistic approach. *Ann. Appl. Probab.*, 13(2):515–539, 2003.

[107] Y. Guo. The Landau equation in a periodic box. *Comm. Math. Phys.*, 231(3):391–434, 2002.

[108] M. Hauray and S. Mischler. On Kac's chaos and related problems. *J. Funct. Anal.*, 266(10):6055–6157, 2014.

[109] L. He and X. Yang. Well-posedness and asymptotics of grazing collisions limit of Boltzmann equation with Coulomb interaction. *SIAM J. Math. Anal.*, 46(6):4104–4165, 2014.

[110] O. Hernández-Lerma and J. B. Lasserre. Further criteria for positive Harris recurrence of Markov chains. *Proc. Amer. Math. Soc.*, 129(5):1521–1524, 2001.

[111] D. Heydecker. Pathwise convergence of the hard spheres Kac process. *Ann. Appl. Probab.*, 29(5):3062–3127, 2019.

[112] D. Heydecker. Kac's process with hard potentials and a moderate angular singularity. *arXiv preprint arXiv:2008.12943*, 2020.

[113] D. Heydecker. Large deviations of kac's conservative particle system and energy non-conserving solutions to the boltzmann equation: A counterexample to the predicted rate function. *arXiv preprint arXiv:2103.14550*, 2021.

[114] D. Heydecker and R. I. A. Patterson. Bilinear coagulation equations. *arXiv preprint arXiv:1902.07686*, 2019.

[115] E. Ikenberry and C. Truesdell. On the pressures and the flux of energy in a gas according to Maxwell's kinetic theory. I. *J. Rational Mech. Anal.*, 5:1–54, 1956.

[116] A. Jakubowski. On the Skorokhod topology. *Ann. Inst. H. Poincaré Probab. Statist.*, 22(3):263–285, 1986.

[117] S. Janson, T. Ł uczak, and A. Rucinski. *Random graphs.* Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

[118] E. Janvresse. Spectral gap for Kac's model of Boltzmann equation. *Ann. Probab.*, 29(1):288–304, 2001.

[119] I. Jeon. Existence of gelling solutions for coagulation-fragmentation equations. *Comm. Math. Phys.*, 194(3):541–567, 1998.

[120] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.

[121] V. I. Judovič. Non-stationary flows of an ideal incompressible fluid. *Ž. Vyčisl. Mat i Mat. Fiz.*, 3:1032–1066, 1963.

[122] M. Kac. Foundations of kinetic theory. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. III*, pages 171–197. University of California Press, Berkeley and Los Angeles, Calif., 1956.

[123] C. Kipnis and C. Landim. *Scaling limits of interacting particle systems*, volume 320 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1999.

[124] R. C. Kraaij. Flux large deviations of weakly interacting jump processes via well-posedness of an associated Hamilton-Jacobi equation. *Bernoulli*, 27(3):1496–1528, 2021.

[125] T. G. Kurtz. *Approximation of population processes*, volume 36 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa., 1981.

[126] T. G. Kurtz. Martingale problems for conditional distributions of Markov processes. *Electron. J. Probab.*, 3:no. 9, 29, 1998.

[127] T. G. Kurtz. Equivalence of stochastic equations and martingale problems. In *Stochastic analysis 2010*, pages 113–130. Springer, Heidelberg, 2011.

[128] H. J. Kushner and G. G. Yin. *Stochastic approximation and recursive algorithms and applications*, volume 35 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 2003. Stochastic Modelling and Applied Probability.

[129] O. E. Lanford, III. Time evolution of large classical systems. In *Dynamical systems, theory and applications (Rencontres, Battelle Res. Inst., Seattle, Wash., 1974)*, pages 1–111. Lecture Notes in Phys., Vol. 38. Springer, Berlin, Heidelberg, 1975.

[130] O. E. Lanford, III. On a derivation of the Boltzmann equation. In *International Conference on Dynamical Systems in Mathematical Physics (Rennes, 1975)*, pages 117–137. Astérisque, No. 40. SociétGe mathématique de France, 1976.

[131] C. Léonard. On large deviations for particle systems associated with spatially homogeneous Boltzmann type equations. *Probab. Theory Related Fields*, 101(1):1–44, 1995.

[132] X. Lu and C. Mouhot. On measure solutions of the Boltzmann equation, part I: moment production and stability estimates. *J. Differential Equations*, 252(4):3305–3363, 2012.

[133] X. Lu and B. Wennberg. Solutions with increasing energy for the spatially homogeneous Boltzmann equation. *Nonlinear Anal. Real World Appl.*, 3(2):243–258, 2002.

[134] A. A. Lushnikov. Gelation in coagulating systems. *Phys. D*, 222(1-2):37–53, 2006.

[135] F. Malrieu. Convergence to equilibrium for granular media equations and their Euler schemes. *Ann. Appl. Probab.*, 13(2):540–560, 2003.

[136] M. Mariani. Large deviations principles for stochastic scalar conservation laws. *Probab. Theory Related Fields*, 147(3-4):607–648, 2010.

[137] D. K. Maslen. The eigenvalues of Kac's master equation. *Math. Z.*, 243(2):291–331, 2003.

[138] H. P. McKean, Jr. An exponential formula for solving Boltmann's equation for a Maxwellian gas. *J. Combinatorial Theory*, 2:358–382, 1967.

[139] ZA87880 Melzak. A scalar transport equation. *Transactions of the American Mathematical Society*, 85(2):547–560, 1957.

[140] A. Mielke, M. A. Peletier, and D. R. M. Renger. On the relation between gradient flows and the large-deviation principle, with applications to Markov chains and diffusion. *Potential Anal.*, 41(4):1293–1327, 2014.

[141] S. Mischler. Sur le programme de Kac concernant les limites de champ moyen. In *Seminaire: Equations aux Dérivées Partielles. 2009–2010*, Sémin. Équ. Dériv. Partielles, pages Exp. No. XXXIII, 19. École Polytech., Palaiseau, 2012.

[142] S. Mischler and C. Mouhot. Kac's program in kinetic theory. *Invent. Math.*, 193(1):1–147, 2013.

[143] S. Mischler, C. Mouhot, and B. Wennberg. A new approach to quantitative propagation of chaos for drift, diffusion and jump processes. *Probab. Theory Related Fields*, 161(1-2):1–59, 2015.

[144] S. Mischler and B. Wennberg. On the spatially homogeneous Boltzmann equation. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 16(4):467–501, 1999.

[145] Y. Morimoto, K. Pravda-Starov, and C. Xu. A remark on the ultra-analytic smoothing properties of the spatially homogeneous Landau equation. *Kinet. Relat. Models*, 6(4):715–727, 2013.

[146] Y. Morimoto, S. Ukai, C. Xu, and T. Yang. Regularity of solutions to the spatially homogeneous Boltzmann equation without angular cutoff. *Discrete Contin. Dyn. Syst.*, 24(1):187–212, 2009.

[147] Peter Mörters. Introduction to large deviations. *October 19th*, 2010.

[148] C. Mouhot. Explicit coercivity estimates for the linearized Boltzmann and Landau operators. *Comm. Partial Differential Equations*, 31(7-9):1321–1348, 2006.

[149] C. Mouhot. Rate of convergence to equilibrium for the spatially homogeneous Boltzmann equation with hard potentials. *Comm. Math. Phys.*, 261(3):629–672, 2006.

[150] C. Mouhot and C. Villani. Regularity theory for the spatially homogeneous Boltzmann equation with cut-off. *Arch. Ration. Mech. Anal.*, 173(2):169–212, 2004.

[151] H. Müller. Zur allgemeinen theorie ser raschen koagulation. *Kolloidchemische Beihefte*, 27(6-12):223–250, 1928.

[152] N. N. Nguyen and G. Yin. Large deviation principles for Langevin equations in random environment and applications. *J. Math. Phys.*, 62(8):Paper No. 083301, 26, 2021.

[153] R. Normand. A model for coagulation with mating. *J. Stat. Phys.*, 137(2):343–371, 2009.

[154] R. Normand and L. Zambotti. Uniqueness of post-gelation solutions of a class of coagulation equations. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 28(2):189–215, 2011.

[155] J. R. Norris. Smoluchowski's coagulation equation: uniqueness, nonuniqueness and a hydrodynamic limit for the stochastic coalescent. *Ann. Appl. Probab.*, 9(1):78–109, 1999.

[156] J. R. Norris. Cluster coagulation. *Comm. Math. Phys.*, 209(2):407–435, 2000.

[157] J. R. Norris. A consistency estimate for Kac's model of elastic collisions in a dilute gas. *Ann. Appl. Probab.*, 26(2):1029–1081, 2016.

[158] J. R. Norris. Fluid limits in long time - working notes. *Private Communication*, 2016.

[159] R. I. A. Patterson and D. R. M. Renger. Dynamical large deviations of countable reaction networks under a weak reversibility condition. *Preprint*, 2016.

[160] R. I. A. Patterson and D. R. M. Renger. Large deviations of jump process fluxes. *Math. Phys. Anal. Geom.*, 22(3):Paper No. 21, 32, 2019.

[161] R. I. A. Patterson, S. Simonella, and W. Wagner. Kinetic theory of cluster dynamics. *Phys. D*, 335:26–32, 2016.

[162] R. I. A. Patterson, S. Simonella, and W. Wagner. A kinetic equation for the distribution of interaction clusters in rarefied gases. *J. Stat. Phys.*, 169(1):126–167, 2017.

[163] A. J. Povzner. On the Boltzmann equation in the kinetic theory of gases. *Mat. Sb. (N.S.)*, 58 (100):65–86, 1962.

[164] M. Pulvirenti and S. Simonella. The Boltzmann-Grad limit of a hard sphere system: analysis of the correlation error. *Invent. Math.*, 207(3):1135–1237, 2017.

[165] D. R. M. Renger. Flux large deviations of independent and reacting particle systems, with implications for macroscopic fluctuation theory. *J. Stat. Phys.*, 172(5):1291–1326, 2018.

[166] F. Rezakhanlou. Large deviations from a kinetic limit. *Ann. Probab.*, 26(3):1259–1340, 1998.

[167] M. Rousset. A n-uniform quantitative tanaka's theorem for the conservative kac's n-particle system with maxwell molecules. *arXiv preprint arXiv:1407.1965*, 2014.

[168] D. Ruelle. *Statistical mechanics: Rigorous results.* W. A. Benjamin, Inc., New York-Amsterdam, 1969.

[169] S. Salem. Propagation of chaos for the boltzmann equation with moderately soft potentials. *arXiv preprint arXiv:1910.01883*, 2019.

[170] A. Shwartz and A. Weiss. *Large deviations for performance analysis.* Stochastic Modeling Series. Chapman & Hall, London, 1995. Queues, communications, and computing, With an appendix by Robert J. Vanderbei.

[171] M. von Smoluchowski. Drei vortrage uber diffusion, brownsche bewegung und koagulation von kolloidteilchen. *Zeitschrift fur Physik*, 17:557–585, 1916.

[172] A. Sznitman. Équations de type de Boltzmann, spatialement homogènes. *Z. Wahrsch. Verw. Gebiete*, 66(4):559–592, 1984.

[173] A. Sznitman. Topics in propagation of chaos. In *École d'Été de Probabilités de Saint-Flour XIX—1989*, volume 1464 of *Lecture Notes in Math.*, pages 165–251. Springer, Berlin, 1991.

[174] M. Talagrand. Matching random samples in many dimensions. *Ann. Appl. Probab.*, 2(4):846–856, 1992.

[175] M. Talagrand. The transportation cost from the uniform measure to the empirical measure in dimension ≥ 3. *Ann. Probab.*, 22(2):919–959, 1994.

[176] H. Tanaka. On the uniqueness of Markov process associated with the Boltzmann equation of Maxwellian molecules. In *Proceedings of the International Symposium on Stochastic Differential Equations (Res. Inst. Math. Sci., Kyoto Univ., Kyoto, 1976)*, pages 409–425. Wiley, New York-Chichester-Brisbane, 1978.

[177] H. Tanaka. Probabilistic treatment of the Boltzmann equation of Maxwellian molecules. *Z. Wahrsch. Verw. Gebiete*, 46(1):67–105, 1978/79.

[178] H. Tanaka. Some probabilistic problems in the spatially homogeneous Boltzmann equation. In *Theory and application of random fields (Bangalore, 1982)*, volume 49 of *Lect. Notes Control Inf. Sci.*, pages 258–267. Springer, Berlin, 1983.

[179] M. Tasković, R. J. Alonso, I. M. Gamba, and N. Pavlović. On Mittag-Leffler moments for the Boltzmann equation for hard potentials without cutoff. *SIAM J. Math. Anal.*, 50(1):834–869, 2018.

[180] G. Toscani and C. Villani. Probability metrics and uniqueness of the solution to the Boltzmann equation for a Maxwell gas. *J. Statist. Phys.*, 94(3-4):619–637, 1999.

[181] G. Toscani and C. Villani. Sharp entropy dissipation bounds and explicit rate of trend to equilibrium for the spatially homogeneous Boltzmann equation. *Comm. Math. Phys.*, 203(3):667–706, 1999.

[182] G. Toscani and C. Villani. On the trend to equilibrium for some dissipative systems with slowly increasing a priori bounds. *J. Statist. Phys.*, 98(5-6):1279–1309, 2000.

[183] I. Tristani. Exponential convergence to equilibrium for the homogeneous Boltzmann equation for hard potentials without cut-off. *J. Stat. Phys.*, 157(3):474–496, 2014.

[184] C. Truesdell. On the pressures and the flux of energy in a gas according to Maxwell's kinetic theory. II. *J. Rational Mech. Anal.*, 5:55–128, 1956.

[185] C. Villani. *Contribution À l'Étude mathématique des Équations de Boltzmann et de Landau en théorie cinétique des gaz et des plasmas.* PhD thesis, Université Paris-Dauphine à Paris, 1998. Thèse de doctorat dirigée par Lions, Pierre-Louis Mathématiques Paris 9 1998.

[186] C. Villani. On a new class of weak solutions to the spatially homogeneous Boltzmann and Landau equations. *Arch. Rational Mech. Anal.*, 143(3):273–307, 1998.

[187] C. Villani. On the spatially homogeneous Landau equation for Maxwellian molecules. *Math. Models Methods Appl. Sci.*, 8(6):957–983, 1998.

[188] C. Villani. On the trend to equilibrium for solutions of the boltzmann equation: quantitative versions of boltzmannś h-theorem. *Unpublished Review Paper*, 1999.

[189] C. Villani. A review of mathematical topics in collisional kinetic theory. In *Handbook of mathematical fluid dynamics, Vol. I*, pages 71–305. North-Holland, Amsterdam, 2002.

[190] C. Villani. Cercignani's conjecture is sometimes true and always almost true. *Comm. Math. Phys.*, 234(3):455–490, 2003.

[191] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[192] C. Villani. H-theorem and beyond: Boltzmann's entropy in today's mathematics. In *Boltzmann's legacy*, ESI Lect. Math. Phys., pages 129–143. Eur. Math. Soc., Zürich, 2008.

[193] J. B. Walsh. An introduction to stochastic partial differential equations. In *École d'été de probabilités de Saint-Flour, XIV—1984*, volume 1180 of *Lecture Notes in Math.*, pages 265–439. Springer, Berlin, 1986.

[194] B. Wennberg. Entropy dissipation and moment production for the Boltzmann equation. *J. Statist. Phys.*, 86(5-6):1053–1066, 1997.

[195] E. Wild. On Boltzmann's equation in the kinetic theory of gases. *Proc. Cambridge Philos. Soc.*, 47:602–609, 1951.

[196] L. Xu. Uniqueness and propagation of chaos for the Boltzmann equation with moderately soft potentials. *Ann. Appl. Probab.*, 28(2):1136–1189, 2018.

[197] M. Zedek. Continuity and location of zeros of linear combinations of polynomials. *Proc. Amer. Math. Soc.*, 16:78–84, 1965.

[198] E. Zermelo. Über einen satz der dynamik und die mechanische wärmetheorie. *Annalen der Physik*, 293(3):485–494, 1896.

[199] R. M. Ziff. Kinetics of polymerization. *J. Statist. Phys.*, 23(2):241–263, 1980.