Ann. Inst. H. Poincaré C
Anal. Non Linéaire 43 (2026), 273–343
DOI 10.4171/AIHPC/157

The mean-field limit of sparse networks of integrate-and-fire neurons

Pierre-Emmanuel Jabin and Datong Zhou

Abstract. We study the mean-field limit of a model of biological neuron networks based on the so-called stochastic integrate-and-fire (IF) dynamics. However, we do not assume any structure on the graph of interactions but consider instead any connection weights between neurons that obey a generic mean-field scaling. To address this, we propose a novel notion of observables that naturally extends the marginals laws in studying classical many-particle systems. We prove the stability of the network in terms of observables, by applying a novel commutator estimate in weak norms to a tree-indexed extension of the BBGKY hierarchy. While we require non-vanishing diffusion, this approach notably addresses the challenges of sparse interacting graphs/matrices and singular interactions from Poisson jumps, and requires no additional regularity on the initial distribution.

1. Introduction

This article derives a continuous limit for the large-scale behavior of networks of neurons following a type of dynamics known as integrate-and-fire (IF). It is a natural example of *multi-agent systems*, where each agent (neuron) could influence others and be influenced in return. However, because each neuron has a priori different connections to other neurons, it is also an important example of non-exchangeable systems.

We focus on IF systems for a large number of agents or neurons, typically 86×10^9 in a human brain for example. This makes it quite challenging to study the original system, either numerically or analytically. Instead one can try to approach the large-scale behavior of such multi-agent systems through the concept of the *mean-field limit*. In classical exchangeable systems, the mean-field limit consists in replacing the exact influence exerted on one particle by its expectation or mean. It is hence connected to the famous notion of propagation of chaos which allows the use of a law of large numbers to rigorously justify this approximation. However, in non-exchangeable systems, the derivation of the mean-field limit also requires a way to capture the limit of the non-identical interactions between particles or agents.

Mathematics Subject Classification 2020: 35Q70 (primary); 05C90, 35Q49, 35Q83, 35Q92, 35R02, 35R06, 92B20 (secondary).
Keywords: mean-field limit, graph limit, integrate-and-fire neurons, neural networks, non-exchangeable systems, sparse networks, Poisson jumps, BBGKY hierarchy, negative Sobolev norms.

This article introduces a novel strategy based on a new concept of observables that are well-chosen linear combinations of empirical laws of agents or neurons. This family solves a tree-indexed hierarchy, approximately for a finite number of neurons and exactly at the limit; a key feature of this hierarchy is that the connection weights between neurons do not appear explicitly anymore. As a consequence, the mean-field limit can be derived directly by passing to the limit in the hierarchy, bypassing a priori structural assumptions on the connection weights. In particular, our result is entirely compatible with *sparse* connection weights, as supported by experimental findings in neuroscience. However, the IF-type dynamics involve jump processes in time, which inevitably introduce discontinuities. Therefore, at the technical level, a major contribution of this article is the development of well-adapted weak norms that provide quantitative stability estimates.

1.1. An IF neuron network with non-identical sparse connections

We focus in this article on a type of stochastic integrate-and-fire model. In this model, neurons interact through "spikes" that represent short electrical pulses in the membrane potential, typically lasting 1–2 ms. A broad range of IF models adopt the following theoretical simplification that dates back to the earliest mathematical model of a neuron; see [63] as well as [51, 66].

Spikes occur at distinct points in time, initiating what is typically referred to as a "fire" event. For a network of IF neurons, at the exact time when the ith neuron fires,

$$\text{for all } j \text{ connected to } i,\ X^j \text{ jumps by } w_{j,i},$$

where X^j describes the membrane potential of the jth neuron and $w_{j,i}$ represents the *synaptic connection* from i to j. The case of no synaptic connection is represented by $w_{j,i} = 0$.

There exists a large variety of models with various rules to determine when a neuron is firing and what is the evolution of the membrane potential between spikes. In the seminal work [63], the firing of neuron i is predicted at the time X^i reaches a certain hard threshold value X_F. According to IF dynamics, at such a time point each X^j jumps by $w_{j,i}$ and X^i is reset to zero. However, in the present paper we consider instead a notion of soft threshold where the firing of each neuron follows an independent Poisson process with a rate that depends on the membrane potential.

When there is no firing, the "pre-spike" dynamics of membrane potential is usually given by a simple ODE or SDE, which we may write in our case as

$$\mathrm{d}X^i(t) = \mu(X^i(t))\,\mathrm{d}t + \sigma(X^i(t))\,\mathrm{d}\mathbf{B}_t^i.$$

As mentioned earlier, there exists a large variety of IF models in terms of the equations for pre-spike dynamics and criteria for firing. From the point of view of the mathematical analysis developed in this paper, both the stochasticity in the SDE equation on $X^i(t)$ and the soft threshold are needed.

The non-linearity of pre-spike dynamics has been observed in modern experimental studies such as [3], and stochasticity was noted in [38,39,57]. Although biophysical models such as Hodgkin–Huxley [52] and FitzHugh–Nagumo [35, 68] are available for more accurately representing the shape of each spike, IF-type dynamics are frequently preferred for their perceived precision when investigating multiple-neuron networks. Nevertheless, the present is still a compromise between mathematical succinctness and biological plausibility. Some extended mathematical models that aim to capture more complex neuronal phenomena have also been studied, for example, in [16, 70, 72]. For a more extensive discussion of IF models in the context of neuroscience, we refer to [14, 40, 41] and the references therein. For a more thorough exploration of the biological considerations, we direct interested readers to references [41, 77].

To complete the definition at endpoints, it is conventional to define $X^i(t)$ at a firing time as the value *after* the jump or reset, making each $X^i(t)$ right continuous with left limit (càdlàg functions). This allows a precise mathematical definition of the dynamics to be given. Let $(X^i_t)_{i=1}^N$ be the $\mathbb{R}$-valued càdlàg processes representing the membrane potential changes of the N neurons and let $w_N := (w_{i,j;N})_{i,j=1}^N$ be the interaction matrix describing the synaptic connection between these neurons. The IF-type dynamics of neurons are characterized by the following SDE in integral form holding for all $i \in \{1,\dots,N\}$:

$$\begin{aligned} X_t^{i;N} &= X_0^{i;N} + \int_0^t \mu(X_{s^-}^{i;N})\,\mathrm{d}s + \int_0^t \sigma(X_{s^-}^{i;N})\,\mathrm{d}\mathbf{B}_s^i \\ &\quad + \sum_{j\neq i} w_{i,j;N} \int_0^t \int_0^\infty \mathbb{1}\{z \le \nu(X_{s^-}^{j;N})\}\mathbf{N}^j(\mathrm{d}z,\mathrm{d}s) \\ &\quad - \int_0^t \int_0^\infty X_{s^-}^{i;N}\mathbb{1}\{z \le \nu(X_{s^-}^{i;N})\}\mathbf{N}^i(\mathrm{d}z,\mathrm{d}s), \end{aligned} \tag{1.1}$$

where

$\{\mathbf{N}^i\}_{i=1}^N$ are homogeneous spatial Poisson processes w.r.t. Lebesgue measure,
$\{\mathbf{B}^i\}_{i=1}^N$ are standard Wiener processes, and the $2N$ processes are independent.

For the target neuron i, the term $\mu(X_{s^-}^{i;N})\,\mathrm{d}s$ summarizes its pre-spike dynamics and $\sigma(X_{s^-}^{i;N})\,\mathrm{d}\mathbf{B}_s^i$ adds a Brownian noise. It experiences a jump of $w_{i,j;N}$ when another neuron j fires and is reset to zero when it fires itself. Neuron i firing occurs with a likelihood depending on the membrane potential, which we denote by $\nu(X_{s^-}^{i;N})$ and we introduce the Poisson processes $\mathbf{N}^i(\mathrm{d}z,\mathrm{d}s)$.

For the simplified case that the connections between neurons are all identical, i.e. $w_{i,j;N} = 1/N$, for all $\{i,j\} \in \{1,\dots,N\}$, the mean-field limit of (1.1) or its variations can be expressed as a PDE about the (time-varying) density function $f(t,x)$, where $x \in \mathbb{R}$ represents the membrane potential. We mention [75] which employs a PDE-based approach, and [27,29,34] which each offer a distinct probabilistic perspective. Though significantly different from (1.1), Hawkes processes give another type of popular model for biological neuron networks and their mean-field limit has also been studied, as in [19,30]. We also cite [4, 12] for the study of large biophysical models with Hodgkin–Huxley and FitzHugh–Nagumo equations for the neurons, together with [71] which derives an IF model from biophysical models in a mean-field setting. Even in the case of identical connections, we emphasize that some neuron models may contain singularities that lead to important mathematical challenges when deriving the mean-field limit.

While assuming identical connections is a significant simplification, the derived mean-field limits have nonetheless provided useful insights into our understanding of large biological neuron networks. For some limiting models, the mean-field equations can for example exhibit blow-up in finite time, which may represent some large-scale synchronization within the network; see for instance [15,17,18] from a PDE perspective, and [28] from a probability point of view. The issue of convergence to equilibrium in the mean-field limit is also an important question, for which we refer for example to [37] and [33]. Other studies, such as [24–26], have explored the spectral conditions sufficient for the existence of periodic solutions near the invariant measure through a Hopf bifurcation.

Systems with non-identical connections remain less understood, despite their relevance to applications in neuroscience, as noted for instance in [74]. This is also supported by recent progress in experimental biology that makes detailed connection graphs for large neuron networks available [53]. Mathematically, having non-identical connections fundamentally alters the dynamics of coupled ODEs or SDEs like (1.1), rendering them *non-exchangeable* and making many established tools for exchangeable systems lose their applicability.

Despite these challenges, there exists a wide range of results that are able to handle systems with certain types of non-identical connections, provided some structural assumptions are made. A first example assumes that connections follow the algebraic constraint $\sum_j w_{i,j;N} \equiv 1$ and that the initial data $(X_0^i)_{i=1}^N$ are i.i.d.; the same mean-field equation as in the exchangeable case is then obtained, see for instance [54]. Another well-known case is found when the connections smoothly depend on the physical location of each neuron: A typical assumption is that $w_{i,j;N} = W(y_{i;N}, y_{j;N})$, where $y_{i;N} \in \mathbb{R}^d$ denotes the spatial location of the ith neuron and $W(\cdot,\cdot)$ is a smooth function. This case leads to some version of the well-known neural field equations; see [1, 8, 47–49, 79]. Within this type of assumption on connections, the mean-field limit has also been investigated in [20] for a model based on the Hawkes processes. Another well-known setting consists in taking random connections, typically corresponding to some classical random graph. This can of course be an attractive assumption when the connections remain mostly unknown. The mean-field limit has been rigorously derived with several types of random connections including the Erdős–Rényi type, as shown in [46]. We also mention [23, 67, 69] which obtain mean-field limits of other multi-agent systems, still with random connections.

It is also enlightening to draw a comparison with the wider spectrum of results on general non-exchangeable systems and not specifically IF models. Many approaches rely on graphon theory, such as [56], which derives the mean-field limit for the Kuramoto model (originally introduced in [59]), while subsequent explorations of the dynamics were conducted in [21, 22]. The use of graphons and cut distance was later adopted in the study of multi-agent systems with stochasticity; for example, see [5, 7, 65]. Graphons are natural tools for describing the graph limit of connections $w_{i,j;N}$ without requiring a priori knowledge of additional regularity. This approach was initially developed in the seminal works [11, 64] for $L^\infty([0,1]^2)$ kernels, which assume a dense scaling for the weight matrices, typically with $\max_{i,j} |w_{i,j;N}| \sim O(1/N)$. More recently, this framework was extended in [9, 10] to $L^p([0,1]^2)$ kernels, allowing it to handle limits for connection weights that are potentially sparser. There are also results on adaptations of standard graphons that address even sparser networks, where the limit kernel may no longer be a Lebesgue function. We mention [61] based on some concept of weak convergence on graphs, or [42–44, 58] which are based on extensions of graphons such as graph-op. While those results still require a priori knowledge of some additional convergence of $w_{i,j;N}$, the case of sparse connections without a priori regularity has been recently studied in [55].

Let us remark that, when we discuss sparse connection weights in the present article, we mean that the support of the corresponding kernels can be concentrated, i.e. there can be a significant number of zero entries in the discrete matrix. In much of the literature, the definition of sparsity is more stringent. For example, it may refer to graphs with bounded degree (or the number of non-zero entries in every column and every row is capped). As will soon become apparent, the discussions in this article do not apply to such "extremely sparse" connection weights. This is because the averaging effect in the binary interactions becomes insufficient, making it impossible to justify mean-field interactions. This regime should be described using different approaches and cannot be adequately addressed in this article. For a discussion of such cases, we refer instead to [61, 69] and the references therein.

In the present article, we keep the same general assumptions on connections as [55] namely, the following:

- The $w_{i,j;N}$ may be completely different for every pair of neurons.
- The $w_{i,j;N}$ can be positive or negative with corresponding excitation or inhibition between neurons, and are not symmetric.
- The number of neurons is assumed to be very large, $N \gg 1$. We recall in particular that the human brain for example contains approximately 8.6×10^{10} neurons.
- The $w_{i,j;N}$ satisfy the following scaling:

$$\max\left(\max_i \sum_j |w_{i,j;N}|, \max_j \sum_i |w_{i,j;N}|\right) \sim O(1), \quad \max_{i,j} |w_{i,j;N}| \ll 1.$$

  This scaling allows each neuron i to be connected to a large population of neurons j, while keeping the network sparsely connected. This again seems to fit with the average of 7000 synaptic connections per neuron in the human brain.

However, we introduce several new key ideas with respect to [55] which we emphasize below:

- We introduce a novel notion of observables at the level of the many-particle system. In contrast, the observables in [55] could only be defined after propagation of chaos is obtained for the system of coupled PDEs. This leads to significant differences in the present article: First of all, we do not require propagation of chaos in the classical sense. Some notion of independence is still implicitly assumed in our result but it is much weaker and would for example allow for some strong correlations between a small subset of neurons.
- We develop the first theory of commutator estimates for hierarchies of PDEs. This is both of interest in itself and a necessity given the new observables mentioned above: contrary to [55], the new observables do not solve the limiting hierarchy and in fact the jumps introduce several singular terms. This means that we cannot derive the limit just by propagating strong L^p-norms in a straightforward manner. Commutator estimates are of course the key step in the theory of renormalized solutions to obtain uniqueness of weak solutions to advection equations. The extension to infinite systems of coupled PDEs, like the hierarchies in the present paper, is however challenging: the commutator estimate for one equation involves the next observables in the hierarchy so that each estimate needs to be compatible with the estimates at the next level.

A main contribution of the present article is to derive a mean-field limit while only asking that the connection weights obey the extended mean-field scaling above, without any further assumptions. A motivation for this scaling is to be able to consider general sparse graphs, though it is not the only one. We note that the mean-field limit has been established for certain sparse graph structures. For instance, classical graphon theory can already be applied to specific classes of sparse random graphs (though typically with some renormalization). But while other results can handle sparse graphs with a similar scaling to the one here, they also typically require additional assumptions on the behavior of the connection weights, for example [44]. Furthermore, heterogeneity may play a significant role so that sparsity is not necessarily the most accurate characterization of the challenge. For example, one may consider for connection weights the sum of a dense graph and some sparse graph. The resulting w_{ij} are not sparse but will fail to satisfy the classical graphon scaling, or many other extensions.

As a last remark, to convey the essence more straightforwardly, we decided to state our results for deterministic weights, although it is indeed possible to introduce randomness to the weights. This is because, from the perspective of probability, our results are neither "annealed results" nor "quenched results" but rather "semi-quenched" (we can treat the randomness of the graph as fixed or "frozen", but should consider the law of the agents' solutions rather than every individual realization). Therefore, attempting to state the strongest possible results in the random context would feel cumbersome. Additionally, the motivation for our results is not to better handle connection weights sampled from some sufficiently smooth graphon, but rather to include more general connection weights. Conceptually, this should allow us to consider data from experimental neuroscience without presuming that the connection weights exhibit regularity within any specific phase space.

Our results nevertheless hold for certain sparse Erdős–Rényi graphs. For instance, consider edge weights that are independent Bernoulli random variables, each taking the value $1/(pN)$ with probability p and 0 with probability $1 - p$. Here, p decreases as N goes to infinity. This generates sparse weighted graphs that cannot be well described by standard graphon theory, which requires L^p regularity for $p > 1$ (see [10], where graphon theory is extended only to $p > 1$; there is a natural barrier at $p = 1$, as seen in numerous analysis problems). However, this requires p to decay slowly with respect to N. If p decays too quickly, there is a non-zero probability that the maximum of row summation (or the column summation) blows up, causing even our kernel to fail.

1.2. The marginal laws and BBGKY hierarchy for exchangeable systems

A classical way to address this mean-field limit of large SDE systems like (1.1) is to shift our focus from tracking trajectories to examining the joint law of various subsets of neurons.

For clarity, let us first mention some of the notation that we are using. We denote by $\mathcal{M}(\mathbb{R}^k)$ the space of signed Borel measures with bounded *total variation norm* on $\mathbb{R}^k$. Here, $\mathcal{M}_+(\mathbb{R}^k)$ stands for the subset of non-negative measures, and $\mathcal{P}(\mathbb{R}^k)$ stands for the subset of probability measures. When choosing a topology on $\mathcal{M}(\mathbb{R}^k)$, we will mostly use the classical notion of *weak-** convergence. Note that we will also have bounds on some exponential moments, so that together with those estimates, weak-* convergence will typically imply tight convergence.

We now introduce the classical concept of marginals for exchangeable systems, where we emphasize the following steps to highlight the difference from non-exchangeable systems:

- For any distinct indices $i_1, \dots, i_k \in \{1, \dots, N\}$, denote the marginal law of the agents $X_t^{i_1;N}, \dots, X_t^{i_k;N}$ by
$$f_N^{i_1,\dots,i_k}(t,\cdot) := \mathrm{Law}(X_t^{i_1;N}, \dots, X_t^{i_k;N}) \in \mathcal{P}(\mathbb{R}^k).$$
- Formally define $f_N^{i_1,\dots,i_k} \equiv 0$ if there are duplicated indices among $i_1, \dots, i_k$.
- For the full joint law, adopt the simplified notation that
$$f_N(t,\cdot) := f_N^{1,\dots,N}(t,\cdot) = \mathrm{Law}(X_t^{1;N}, \dots, X_t^{N;N}) \in \mathcal{P}(\mathbb{R}^N).$$

In the context of exchangeable systems (identical connections, $w_{i,j;N} = w(N)$), it is straightforward that, if $(X_t^{1;N}, \dots, X_t^{N;N})$ is a solution of the system, then any permutation $(X_t^{i_1;N}, \dots, X_t^{i_N;N})$ solves the same system. This implies that the full joint law equation is symmetric, so it suffices to consider that marginals of the same order are identical, namely,
$$f_N^{i_1,\dots,i_k}(t,\cdot) = f_N^{j_1,\dots,j_k}(t,\cdot) \in \mathcal{P}(\mathbb{R}^k),$$
if the indices $1 \le i_1, \dots, i_k \le$ are distinct and $1 \le j_1, \dots, j_k \le N$ are also distinct.

Given this property, it is natural to define the unique k-marginal

$$f_{N,k}(t,\cdot) := f_N^{1,\dots,k}(t,\cdot) \in \mathcal{P}(\mathbb{R}^k).$$

The marginals are solutions to the famous BBGKY hierarchy of equations, in which the equation for each $f_{N,k}$ depends on itself and the next marginal $f_{N,k+1}$ recursively.

One of the key concepts to obtain the mean-field limit is the notion of (Kac's) chaos, which can be defined in various equivalent ways. One possible definition, which is the one we use in this article, involves the marginals: We have chaos iff the k-marginals of random variables $(X^{1;N},\dots,X^{N;N})$ converge weak-* to the tensorization of a certain one-particle distribution $f \in \mathcal{P}(\mathbb{R})$ as $N \to \infty$, namely,

$$f_{N,k} \overset{*}{\rightharpoonup} f^{\otimes k} \in \mathcal{P}(\mathbb{R}^k), \quad f^{\otimes k}(z_1,\dots,z_k) := \prod_{m=1}^{k} f(z_m), \quad \text{for all fixed } k \in \mathbb{N}.$$

At least for smooth enough dynamics, it is possible to show that chaos on the initial data implies chaos at every later time, which is the famous propagation of chaos. Among the various strategies for proving propagation of chaos and for obtaining the Vlasov equation as a mean-field limit, we highlight the following one given its similarities with the approach we will follow:

- Pass to the limit in the BBGKY hierarchy to the Vlasov hierarchy as $N \to \infty$, which yields $f_{N,k}(t,\cdot) \overset{*}{\rightharpoonup} f_{\infty,k}(t,\cdot) \in \mathcal{P}(\mathbb{R}^k)$, where $f_{\infty,k}(t,\cdot)$ represents a solution to the Vlasov hierarchy with initial data in tensorized form, namely $f_{\infty,k}(0,\cdot) = f_0^{\otimes k}$.
- Notice that if the one-particle distribution $f(t,\cdot)$ solves the Vlasov equation with initial data f_0, then the k-marginals in tensorized form $f^{\otimes k}(t,\cdot)$ are a solution to the Vlasov hierarchy with the same initial data $f_{\infty,k}(0,\cdot) = f_0^{\otimes k}$.
- Prove the uniqueness of the solution of the Vlasov hierarchy, which allows one to conclude that at all time $t \geq 0$, $f_{\infty,k}(t,\cdot) = f^{\otimes k}(t,\cdot)$.

A variation of this argument involves directly obtaining stability estimates between the BBGKY hierarchy and the Vlasov hierarchy, yet quantifies the deviation of the N-particle SDE system to the Vlasov equation on the level of marginal laws. In general, deriving the mean-field limit can be challenging, especially when the interaction between particles is singular or when there is no diffusion in the dynamics. Not surprisingly, the above approach usually requires smoothness on the dynamics: from analytic in [76] to Lipschitz in [45]. However, recent results such as [60] and [13] have shown how to take advantage of non-vanishing diffusion to handle interactions through a kernel merely in respectively only L^∞ (more precisely some exponential Orlicz space) and only L^p for $p > 1$.

We hope to implement a similar strategy for non-exchangeable systems, such as our (1.1). However, given a solution $(X_t^{1;N},\dots,X_t^{N;N})$ of (1.1), a permutation $(X_t^{i_1;N},\dots,X_t^{i_N;N})$ is not in general also a solution since $w_{1,2;N}$ is in general not equal to $w_{i_1,i_2;N}$ for example. Consequently, the concept of k-marginals does not actually exist and, instead, we have to consider the more complicated situation where for a fixed k, each marginal law $f_N^{i_1,\dots,i_k}$ might differ.

It is, however, not even the most significant obstacle. The more intricate issue lies in the fact that any direct generalization of the BBGKY hierarchy would depend explicitly on the coefficients $w_{i,j;N}$. Hence, passing to the limit in the hierarchy would require passing to the limit in some appropriate sense in the coefficients $w_{i,j;N}$. If the $w_{i,j;N}$ are of order $O(1/N)$, one can potentially apply the graphon theory [64] to achieve this, as has been done for the Kuramoto model in [56]. Unfortunately, we are considering potentially sparse networks without any a priori smoothness and we have no idea how to generalize graphon theory in that case.

1.3. The novel notion of observables for non-exchangeable systems

A main contribution of the paper is to introduce a novel concept of *observables* in Definition 1.2, which not only incorporates into the marginal laws $f_N^{i_1,\dots,i_k}$ but also takes into account the effect of connectivity $w_N = (w_{i,j})_{i,j=1}^N$ in (1.1).

Those observables satisfy an approximate hierarchy that extends in some sense the BBGKY hierarchy but which does not involve any explicit dependence on the connection weights. This new hierarchy hence offers a promising framework for obtaining the mean-field limit, as it will be enough to pass to the limit in a countable family of observables and equations.

Its structure however remains more complex. The main idea behind the definition of the new observables, is to track all possible interactions between any finite number of neurons. In the exchangeable case, it does not matter in which order these interactions take place, so that our observables would reduce to the marginals and only depend on the total number of neurons under consideration. But in the non-exchangeable case such as here, it is necessary to keep track of which neuron is interacting with which. To achieve this, we use tree graphs to index our observables, and establish a natural correspondence between adding a leaf on a node of the tree and interacting with a particular agent among the k selected ones.

Definition 1.1. Define $\mathcal{T}$ as a set of directed labeled graphs (trees) constructed recursively in the following manner:

- Denoting by $|T|$ the total number of vertices in T, index the vertices in T from $1, \dots, |T|$.
- The graph of a single node (indexed by 1) belongs to $\mathcal{T}$.
- All other elements of $\mathcal{T}$ are constructed recursively: For any $T \in \mathcal{T}$ and any $1 \le m \le |T|$, the graph $T + m$ belongs to $\mathcal{T}$, where $T + m$ is obtained by adding a leaf to vertex #m namely by adding a node indexed by $|T| + 1$ and adding $(m, |T| + 1)$ as an edge to T.

The family $\mathcal{T}$ corresponds to all trees up to isomorphisms, but it is equipped with a natural orientation. The root of the tree is always labeled 1, and $(l, m) \in \mathcal{E}(T)$ if there exists an edge connecting l and m and if l is closer to the root than m. This family enables us to define our observables.

Definition 1.2. Consider any connectivity matrix $w_N = (w_{i,j})_{i,j=1}^N$ and a collection of random processes $(X_t^{1;N}, \ldots, X_t^{N;N})$. We define the observable $\tau_N(T, w_N, f_N)(t, \cdot) \in \mathcal{M}(\mathbb{R}^{|T|})$, $T \in \mathcal{T}$ as the weighted sum of marginals

$$\tau_N(T, w_N, f_N)(t, \mathrm{d}z) \\ := \frac{1}{N} \sum_{i_1,\ldots,i_{|T|}=1}^{N} w_{N,T}(i_1, \ldots, i_{|T|}) f_N^{i_1,\ldots,i_{|T|}}(t, \mathrm{d}z_1, \ldots, \mathrm{d}z_{|T|}) \tag{1.2}$$

where the weight of each marginal is given by

$$w_{N,T}(i_1, \ldots, i_{|T|}) := \prod_{(l,m)\in\mathcal{E}(T)} w_{i_l,i_m;N} \in \mathbb{R}.$$

We also define the absolute observable $|\tau_N|(T, w_N, f_N)(t, \cdot) \in \mathcal{M}_+(\mathbb{R}^{|T|})$, $T \in \mathcal{T}$, as

$$|\tau_N|(T, w_N, f_N)(t, \mathrm{d}z) := \frac{1}{N} \sum_{i_1,\ldots,i_{|T|}=1}^{N} |w_{N,T}(i_1, \ldots, i_{|T|})| f_N^{i_1,\ldots,i_{|T|}}(t, \mathrm{d}z_1, \ldots, \mathrm{d}z_{|T|}).$$

As we can see, if $T_1, T_2 \in \mathcal{T}$ are isomorphic as tree graphs, the corresponding observables are also identical up to permutation. In this sense, we can say our observables are indexed by trees. It will be apparent later that the weights are chosen in a natural way so that, in the evolution of observable T, the observable $T + m$ accounts for the interaction with the mth agent among the $|T|$ selected ones.

There does not appear to be an immediate interpretation for most observables, with the obvious exception of the first one. If we take $T = T_1$, the first trivial tree with only one vertex and no edge, then the observable is the 1-particle distribution which is just the average of all marginals of order 1,

$$\tau_N(T_1, w_N, f_N)(t, \mathrm{d}z_1) = \frac{1}{N} \sum_{i=1}^{N} f_N^i(t, \mathrm{d}z_1).$$

Hence obtaining the limit of the observables directly provides the limit of the 1-particle distribution.

We also emphasize that, in contrast to the marginals, our observables are not probability measures. They are neither necessarily normalized to a total mass of 1, nor guaranteed to be non-negative. But the scaling of w_N still ensures the total variation of any observable is at most $O(1)$.

Lemma 1.3. *For any $T \in \mathcal{T}$, we have that*

$$\| |\tau_N|(T, w_N, f_N)(t, \cdot)\|_{\mathcal{M}(\mathbb{R}^{|T|})} \le \left(\max_i \sum_{j=1}^{N} |w_{i,j;N}| \right)^{|T|-1}.$$

Proof. Recall that any marginal law has total mass 1 by definition, thus,

$$\| |\tau_N|(T, w_N, f_N)(t,\cdot)\|_{\mathcal{M}(\mathbb{R}^{|T|})} \leq \frac{1}{N} \sum_{i_1,\ldots,i_{|T|}=1}^{N} |w_{N,T}(i_1,\ldots,i_{|T|})|.$$

If $|T| = 1$, the right-hand side equals 1 trivially, concluding the proof.

When $|T| \geq 2$, we can assume $T = T' + m$ and argue recursively:

$$\begin{aligned}\frac{1}{N} \sum_{i_1,\ldots,i_{|T|}=1}^{N} |w_{N,T}(i_1,\ldots,i_{|T|})| &= \frac{1}{N} \sum_{i_1,\ldots,i_{|T|}=1}^{N} |w_{N,T'}(i_1,\ldots,i_{|T|-1})| |w_{i_m,i_{|T|};N}| \\ &\leq \Big(\frac{1}{N} \sum_{i_1,\ldots,i_{|T|-1}=1}^{N} |w_{N,T'}(i_1,\ldots,i_{|T|-1})|\Big) \max_i \sum_{j=1}^{N} |w_{i,j;N}|.\end{aligned}$$

■

Remark 1.4. While in Definition 1.2 the laws and observables are only assumed to be measures, and hence are denoted by $f(\mathrm{d}z)$, we may adopt the abuse of notation $f(z)$ in later discussions. Many of the forthcoming equations, such as the Vlasov equation (1.3), are indeed classically written on densities.

Given the non-exchangeability of the system (1.1), the limiting behavior as $N \to \infty$ cannot be approximated by just a function $f(t,x)$, with $x \in \mathbb{R}$. Following the idea in [56] and [55], we introduce the so-called extended density $f(t,\xi,x)$ instead, where the additional variable $\xi \in [0,1]$ accounts for the non-exchangeable indices $i \in \{1,\ldots,N\}$ in the mean-field limit. The non-identical interactions in the limit are described by a kernel that we denote by $w(\xi,\zeta)$, $(\xi,\zeta) \in [0,1]^2$, and the Vlasov equation corresponding to (1.1) is given by

$$\begin{aligned}\partial_t f(t,\xi,x) + \partial_x(\mu_f^*(t,\xi,x) f(t,\xi,x)) - \frac{\sigma^2}{2}\partial_{xx} f(t,\xi,x) \\ + \nu(x) f(t,\xi,x) - \delta_0(x) J_f(t,\xi) = 0,\end{aligned} \tag{1.3}$$

where the mean firing rate and the mean-field drift are defined as

$$\begin{aligned}J_f(t,\xi) &:= \int_{\mathbb{R}} \nu(x) f(t,\xi,x)\,\mathrm{d}x, \\ \mu_f^*(t,\xi,x) &:= \mu(x) + \int_0^1 w(\xi,\zeta) J_f(t,\zeta)\,\mathrm{d}\zeta.\end{aligned} \tag{1.4}$$

In our context, $w(\xi,\zeta)$ should be the limit object of the sparsely connected $w_N := (w_{i,j;N})_{i,j=1}^N$ that we described in Section 1.1. As a consequence, we are forced to consider singular kernels $w(\xi,\zeta)$ and the only property we can inherit from w_N is the $O(1)$ scaling of

$$\max\Big(\max_i \sum_j |w_{i,j;N}|, \max_j \sum_i |w_{i,j;N}|\Big) = \max(\|w_N\|_{\ell^\infty\to\ell^\infty}, \|w_N\|_{\ell^1\to\ell^1}).$$

To extend this norm for $N \times N$ connectivity matrices to the kernel on $(\xi, \zeta) \in [0,1]^2$, we define the Banach space $L^\infty_\xi([0,1], \mathcal{M}_\zeta[0,1])$ as the topological dual of the (strong) Bochner space $L^1_\xi([0,1], C_\zeta[0,1])$. Then, since $\mathcal{M}_{\xi,\zeta}([0,1]^2)$ is the topological dual of $C_{\xi,\zeta}([0,1]^2)$ and the canonical embedding

$$C_{\xi,\zeta}([0,1]^2) \to L^1_\xi([0,1], C_\zeta[0,1])$$

is continuous with dense image, one can consider

$$L^\infty_\xi([0,1], \mathcal{M}_\zeta[0,1]) \subset \mathcal{M}_{\xi,\zeta}([0,1]^2).$$

This leads to the main Banach space $\mathcal{W}$ for the kernels

$$\begin{aligned} \mathcal{W} := \big\{ w \in \mathcal{M}([0,1]^2) : w(\xi, \mathrm{d}\zeta) \in L^\infty_\xi([0,1],\ \mathcal{M}_\zeta[0,1]), \\ w(\mathrm{d}\xi, \zeta) \in L^\infty_\zeta([0,1],\ \mathcal{M}_\xi[0,1]) \big\}, \\ \|w\|_{\mathcal{W}} := \max\big\{ \|w\|_{L^\infty_\xi \mathcal{M}_\zeta}, \|w\|_{L^\infty_\zeta \mathcal{M}_\xi} \big\}. \end{aligned} \tag{1.5}$$

The proper definition of the kernel space $\mathcal{W}$ allows us to correctly define the conjectured limiting observables from the extended density. Here, we illustrate an instructive example of elements in $\mathcal{W}$ that do not belong to any $L^p([0,1]^2)$, following [55]. We also present how this limit can appear from finite connection weight matrices that satisfy our assumptions.

Example 1.5. Let $\Phi\colon [0,1] \to [0,1]$ be any measure-preserving, bijective map. Then $w(\xi, \zeta) = \delta(\Phi(\xi) - \zeta)$ belongs to $\mathcal{W}$. This kernel can be interpreted in three equivalent ways:

- First, as an element in $L^\infty_\xi \mathcal{M}_\zeta$, i.e. a parametrized family $\xi \in [0,1] \mapsto \delta_{\Phi(\xi)}(\zeta) \in \mathcal{M}([0,1])$ of measures on $\zeta \in [0,1]$. Specifically,

$$\int_{[0,1]} \phi(\zeta) w(\xi, \mathrm{d}\zeta) = \phi(\Phi(\xi)), \quad \text{for all } \phi \in C([0,1]).$$

- Second, as an element in $L^\infty_\zeta \mathcal{M}_\xi$, i.e. a parametrized family $\zeta \in [0,1] \mapsto \delta_{\Phi^{-1}(\zeta)}(\xi) \in \mathcal{M}([0,1])$ of measures on $\xi \in [0,1]$. Here, we have used the change of variables $\zeta = \Phi(\xi)$ (since Φ is one-to-one and measure preserving). In this interpretation,

$$\int_{[0,1]} \phi(\xi) w(\mathrm{d}\xi, \zeta) = \phi(\Phi^{-1}(\zeta)), \quad \text{for all } \phi \in C([0,1]).$$

- Finally, as the measure in $\mathcal{M}([0,1]^2)$. Namely,

$$\int_{[0,1]^2} \varphi(\xi, \zeta) w(\mathrm{d}\xi, \mathrm{d}\zeta) = \int_{[0,1]} \varphi(\xi, \Phi(\xi))\, \mathrm{d}\xi, \quad \text{for all } \varphi \in C([0,1]^2).$$

These definitions provide a straightforward framework where one can switch seamlessly between interpretations as needed, since for all $\varphi \in C([0,1]^2)$, we have

$$\int_{[0,1]^2} \varphi(\xi,\zeta) w(d\xi, d\zeta) = \int_{[0,1]} \varphi(\xi, \Phi(\xi))\, d\xi = \int_{[0,1]^2} \varphi(\xi,\zeta) w(d\xi, d\zeta).$$

Consider $w_N = (w_{i,j;N})_{i,j=1}^N$ which is an ad hoc discretization of $w(\xi,\zeta) = \delta(\Phi(\xi) - \zeta)$. Specifically, define

$$\begin{aligned} w_{i,j;N} &= \frac{N}{k^2} \int_{[\frac{i-k}{N}, \frac{i}{N}) \times [\frac{j-k}{N}, \frac{j}{N}) \cap [0,1]^2} \varphi(\xi,\zeta) w(d\xi, d\zeta) \\ &= \frac{N}{k^2} \Big| \Big[\frac{i-k}{N}, \frac{i}{N}\Big) \cap \Phi\Big(\Big[\frac{j-k}{N}, \frac{j}{N}\Big) \cap [0,1]\Big) \Big|. \end{aligned}$$

Let $N, k \to \infty$ with $N/k \to \infty$. It is straightforward to verify that, when w_N is viewed as an element of $\mathcal{W}$, it converges in the weak-* topology to $w \in \mathcal{W}$. Moreover, $\max_{1 \le i,j \le N} |w_{i,j;N}| \to 0$.

In the following sections of this paper, when performing a priori estimates, we use $[0, t_*]$ to denote the time dimension's endpoints. As will become apparent later, this choice is solely due to our estimates being bounded on a bounded interval, and it remains valid for any later time. That is, we can select any $t_* > 0$. Additionally, we note that for the a priori estimate of $f(t,\xi,x)$, we use the usual strong Bochner spaces $L^\infty([0,t_*] \times [0,1]; \mathcal{M}(\mathbb{R}))$, which differ from the definition of $\mathcal{W}$ based on duality. We employ $\mathcal{M}_+(\mathbb{R})$ to indicate that the density remains non-negative, which can be straightforwardly derived from our equations.

Definition 1.6. Consider a connectivity kernel $w(\xi,\zeta) \in \mathcal{W}$, $(\xi,\zeta) \in [0,1]^2$ and some extended density $f \in L^\infty([0,t_*] \times [0,1]; \mathcal{M}_+(\mathbb{R}))$. Define the observables $\tau_\infty(T, w, f)(t,\cdot) \in \mathcal{M}(\mathbb{R}^{|T|})$, $T \in \mathcal{T}$, as

$$\tau_\infty(T, w, f)(t,z) := \int_{[0,1]^{|T|}} w_T(\xi_1, \dots, \xi_{|T|}) \prod_{m=1}^{|T|} f(t, \xi_m, z_m)\, d\xi_1, \dots, d\xi_{|T|}, \tag{1.6}$$

where

$$w_T(\xi_1, \dots, \xi_{|T|}) := \prod_{(l,m) \in \mathcal{E}(T)} w(\xi_l, \xi_m).$$

It is easy to check the validity of integrals in (1.4) and (1.6) if the kernel $w(\xi,\zeta)$ is smooth or when $w \in L^\infty$. At the present, it may not be clear yet why the integrations with respect to $\xi \in [0,1]$ involved in (1.4) and (1.6) make sense when we only have $w \in \mathcal{W}$. We prove in Section 4 that it is possible to extend the bounds in Lemma 1.3 through a density argument. We note that a definition akin to τ_∞ along with a similar argument on integrability has been addressed in [55].

1.4. Main result

Our main result states that the large-scale dynamics of (1.1) described in terms of observables $\tau_N(T, w_N, f_N)$ can indeed be approximated by the mean-field limit, provided the initial observables $\tau_N(t=0)$ are approximated by the initial $\tau_\infty(t=0)$.

Theorem 1.7. *Assume that $\mu, \nu \in W^{1,\infty}$ and constant $\sigma > 0$. For a sequence of $N \to \infty$, let $(X_t^{i;N})_{i=1}^N$ be solutions of the non-exchangeable SDE system (1.1) with connectivity matrices $w_N := (w_{i,j;N})_{i,j=1}^N$. In addition, let $f \in L^\infty([0,t_*] \times [0,1]; \mathcal{M}_+(\mathbb{R}))$ be a solution of the Vlasov equation (1.3)–(1.4) with connectivity kernel $w \in \mathcal{W}$. Assume that the following holds:*

- *The connectivity matrices are uniformly bounded: For some $C_{\mathcal{W}} > 0$,*

$$\sup_N \max\Big(\max_i \sum_j |w_{i,j;N}|, \max_j \sum_i |w_{i,j;N}|\Big) \le C_{\mathcal{W}}. \tag{1.7}$$

- *The interaction of each pair of agents vanishes:*

$$\max_{1\le i,j\le N} |w_{i,j;N}| \to 0 \quad \text{as } N \to \infty. \tag{1.8}$$

- *The hierarchy of observables and the extended density are initially bounded by an exponential scale: There exist some $a > 0$, $M_a > 0$, such that*

$$\begin{aligned} \sup_N \int_{\mathbb{R}^{|T|}} \exp\Big(a \sum_{m=1}^{|T|} |z_m|\Big) |\tau_N|(T, w_N, f_N)(0,z)\,\mathrm{d}z \le M_a^{|T|}, \quad \forall T \in \mathcal{T}, \\ \operatorname*{ess\,sup}_{\xi\in[0,1]} \int_{\mathbb{R}} e\exp(a|x|) f(0,\xi,x)\,\mathrm{d}x \le M_a. \end{aligned} \tag{1.9}$$

- *The hierarchy of observables initially converges in weak-* topology:*

$$\begin{aligned} &\tau_N(T, w_N, f_N)(0,\cdot) \\ &\quad \overset{*}{\rightharpoonup} \tau_\infty(T, w, f)(0,\cdot) \in \mathcal{M}(\mathbb{R}^{|T|}) \quad \text{as } N \to \infty, \forall T \in \mathcal{T}. \end{aligned} \tag{1.10}$$

Then the hierarchy of observables converges at any time, in weak- topology:*

$$\begin{aligned} &\tau_N(T, w_N, f_N)(t,\cdot) \\ &\quad \overset{*}{\rightharpoonup} \tau_\infty(T, w, f)(t,\cdot) \in \mathcal{M}(\mathbb{R}^{|T|}) \quad \text{as } N \to \infty, \forall t \in [0,t_*], T \in \mathcal{T}. \end{aligned} \tag{1.11}$$

While we state Theorem 1.7 in terms of the observables τ_N from non-exchangeable systems converging to the limiting observables τ_∞ in weak-* topology, our approach is inherently quantitative. We state, in the next section, a precise and quantitative version of Theorem 1.7, namely Theorem 2.6.

We recall that the first observable immediately corresponds to the 1-particle distribution so that Theorem 1.7 provides the limit of this 1-particle distribution. It would in fact be possible to derive the limit of other well-known statistical objects, the 2-particle distribution and correlations for example. To do that, we would build another family of new observables starting from the 2-particle distribution in addition to the 1-particle distribution. This would also require stronger assumptions with the initial convergence on both families instead of only (1.10). However, we did not want to further add to our approach or our statements and confine ourselves to the limit of the 1-particle distribution.

The only non-straightforward assumption in Theorem 1.7 is (1.10) about whether the $\tau_\infty(T)(0,\cdot)$, for all $T \in \mathcal{T}$, come from a pair of extended density $f(0,\xi,x)$ and $w \in \mathcal{W}$ as defined in Definition 1.6. It would be possible to formulate a version of Theorem 1.7 without this assumption. The sequence of initial data $\tau_N(T, w_N, f_N)(0,\cdot)$ is obviously precompact as $N \to \infty$, so that we could extract a converging subsequence. The proof of Theorem 1.7 would then imply that the limiting τ_∞ are exact solutions to a limiting, tree-indexed hierarchy. However, without (1.10), we cannot identify the limiting τ_∞ as being obtained through some solution $f(t,\xi,x)$ to the limiting Vlasov equation.

It is fortunately straightforward to show that (1.10) directly follows when the initial $X_0^{i,N} = X^{i,N}(t=0)$ are independent. When the initial data $(X_0^{1;N},\dots,X_0^{N;N})$ are independent random variables with $f_{N,0}^i = \mathrm{Law}(X_0^{i;N})$ for all $1 \le i \le N$, the marginal laws are of the form $f_{N,0}^{i_1,\dots,i_k} = \prod_{m=1}^k f_{N,0}^{i_m}$ for $1 \le i_1,\dots,i_k \le N$ which are distinct. We can then define a graphon-like kernel and the extended density as

$$\begin{aligned} \widetilde{w}_N(\xi,\zeta) &= \sum_{i,j=1}^N N w_{i,j;N} \mathbb{1}_{[\frac{i-1}{N},\frac{i}{N})}(\xi)\mathbb{1}_{[\frac{j-1}{N},\frac{j}{N})}(\zeta), \\ \tilde{f}_N(x,\xi) &= \sum_{i=1}^N f_N^i(x)\mathbb{1}_{[\frac{i-1}{N},\frac{i}{N})}(\xi). \end{aligned} \tag{1.12}$$

It becomes straightforward to show that the initial observables $\tau_N(T, w_N, f_N, t=0)$ are approximated by $\tau_\infty(T,\widetilde{w}_N,\tilde{f}_N, t=0)$ up to an error of $O(\max_{1\le i,j\le N}|w_{i,j;N}|)$. We in particular state the following proposition, whose proof is postponed to Section 4.

Proposition 1.8. *For a sequence of $N \to \infty$, consider $(X^{1;N},\dots,X^{N;N})$ as independent random variables and $w_N = (w_{i,j})_{i,j=1}^N \in \mathbb{R}^{N\times N}$. Denote the marginal laws as $f_N^i =$ $\mathrm{Law}(X^{i;N})$ for each N and $1 \le i \le N$. Further, let $\widetilde{w}_N$, $\tilde{f}_N$ be the kernel and extended density as defined in (1.12). Assume that the following hold:*

- *The connectivity matrices are uniformly bounded: For some $C_{\mathcal{W}} > 0$,*

$$\sup_N \max\left(\max_i \sum_j |w_{i,j;N}|, \max_j \sum_i |w_{i,j;N}|\right) \le C_{\mathcal{W}}.$$

- *The interaction of each pair of agents vanishes:*

$$\overline{w}_N := \max_{1\le i,j\le N} |w_{i,j;N}| \to 0, \quad \text{as } N \to \infty.$$

- *The laws are bounded by an exponential scale: There exist some $a > 0$, $M_a > 0$, such that*

$$\sup_N \max_{1\le i\le N} \int_{\mathbb{R}} \exp(a|z|) f_N^i(z)\,\mathrm{d}z \le M_a.$$

Then the difference between the observables $\tau_N(T, w_N, f_N)$ and their approximations $\tau_\infty(T, \widetilde{w}_N, \tilde{f}_N)$, as formulated by (1.6) and (1.12), is quantified by

$$\begin{aligned}\int_{\mathbb{R}^{|T|}} &\exp\Big(a\sum_{m=1}^{|T|}|z_m|\Big)|\tau_\infty(T,\widetilde{w}_N,\tilde{f}_N)(z)-\tau_N(T,w_N,f_N)(z)|\,\mathrm{d}z\\ &\le \max_{1\le i,j\le N}|w_{i,j;N}|\max\Big(\max_i\sum_j|w_{i,j;N}|,\max_j\sum_i|w_{i,j;N}|\Big)^{|T|-2}|T|^2M_a^{|T|}. \quad (1.13)\end{aligned}$$

Moreover, by extracting a subsequence (which we still index by N for simplicity), there exists a pair of kernel $w \in \mathcal{W}$ and extended density $f \in L^\infty([0,1];\mathcal{M}_+(\mathbb{R}))$, such that the hierarchy of approximate observables $\tau_\infty(T, \widetilde{w}_N, \tilde{f}_N)$ converges weak- to the limit hierarchy $\tau_\infty(T, w, f)$:*

$$\tau_\infty(T,\widetilde{w}_N,\tilde{f}_N) \overset{*}{\rightharpoonup} \tau_\infty(T,w,f) \in \mathcal{M}(\mathbb{R}^{|T|}) \quad \text{as } N\to\infty,\ \forall T\in\mathcal{T}. \quad (1.14)$$

In addition, such an extended density f satisfies the bound

$$\operatorname*{ess\,sup}_{\xi\in[0,1]} \int_{\mathbb{R}} \exp(a|x|) f(\xi,x)\,\mathrm{d}x \le M_a.$$

When combined with Proposition 1.8, Theorem 1.7 yields the mean-field limit for independent initial $X_0^{i,N}$ with only some appropriate moments bounds and no other structural assumptions on the $w_{i,j,N}$. For the integrate-and-fire model discussed here, as well as for many other models with sufficiently smooth dynamics, the asymptotic independence of agents $X^{i,N}$ is a property that can propagate to any later time $t > 0$. The proof follows an argument similar to that in [55]. A recent work [62], which also extended the BBGKY hierarchy to non-exchangeable systems, provided sharp quantitative estimates for the propagation of independence in terms of relative entropy.

However, we do emphasize that the proof of our main result, Theorem 1.7, does not rely on the property of asymptotic independence. This is not a necessary condition but merely a convenient situation where the existence of a limit for the initial data is guaranteed. In fact, for non-exchangeable systems, the convergence of observables can in general be much less demanding than independence. It is a very different situation from exchangeable systems where chaos (or approximate independence) is essentially equivalent to the asymptotic tensorization of the marginal. Counterexamples, where observables converge but asymptotic independence of agents does not strictly hold, are easy to construct.

Example 1.9. For convenience, consider the number of agents is $2N$ (i.e. it is even). Let $I_1 = \{1,\dots,N\}$ and $I_2 = \{N+1,\dots,2N\}$. We then take

$$w_{i,j;N} = \begin{cases} 0 & \text{if } (i,j) \in (I_1 \times I_2) \cup (I_2 \times I_1), \\ \dfrac{1}{N} & \text{if } (i,j) \in (I_1 \times I_1) \cup (I_2 \times I_2). \end{cases}$$

In such a case, there are no interactions between neurons in I_1 and neurons in I_2. We can then easily satisfy assumption (1.10) by having the $X_0^{i,2N}$ i.i.d. for $i = 1,\dots,N$, but $X_0^{i+N,2N} = X_0^{i,2N}$ for $i = 1,\dots,N$.

In the exchangeable case, the strong correlation between agents, such as $X_0^{i+N,2N} = X_0^{i,2N}$, does not pose a essential difficulty. One can introduce a random index permutation π and define $Y_0^{i,2N} = X_0^{\pi(i),2N}$, thereby diluting the strong correlation across indices and reducing the system back to the asymptotically i.i.d. (chaotic) case. In fact, any data converging to a deterministic mean-field limit can always be reduced to asymptotic independence. However, in the non-exchangeable case, such a random index permutation does not guarantee that $Y_0^{i,2N}$ remains a solution of the original system. (For the specific example, a carefully chosen permutation could guarantee this, but the loophole is a trade-off we make for the simplicity of the argument.) As a result, we cannot reduce the system to an asymptotically independent setting. This example can obviously be generalized to any arbitrary fixed number of subsets and it is possible to construct even more intricate examples. But this already shows that the optimal assumptions on the initial $X_0^{i,N}$ have to depend intrinsically on the structure of the connections in non-exchangeable cases. In that regard, we conjecture that assumption (1.10) is both necessary and sufficient to have the convergence of the 1-particle distribution.

Theorem 1.7 is the first rigorous result to obtain the mean-field limit for networks of neurons interacting through integrate-and-fire models. The approach through an extended hierarchy solved by observables has very few comparisons in the literature, having only been used previously in [55]. In comparison with the previous [55] however, we put forward several new key ideas with notably the following:

- We introduce the observables directly at the level of the marginals. Instead, the notion of observables in [55] was only valid for almost independent variables, which required first the propagation of independence. There are hence several advantages to our new definition, first as per the discussion above about independence, but also by providing a much more immediate notion of the statistical distribution in the system.
- We develop a new approach for the quantitative estimates on the hierarchy, based on weak norms. This is again in contrast to [55] which used strong L^2-norms. This is a critical point, because the jumps in integrate-and-fire models lead to discontinuities so that we cannot have convergence in the hierarchy for our system for any strong norm. On the other hand, the use of weak norms forces a different method in the analysis, as propagating weak norms necessarily creates intricate commutator estimates. An

important technical contribution of the present paper is to introduce the "right" weak norms and a novel approach to handle those commutators.

There are however many remaining open questions. First of all, the statistical approach followed here does not seem to allow us to obtain the limit of any individual trajectory. This is again in contrast with classical exchangeable systems where obtaining the limit of the 1-particle distribution allows one to have the limit of typical (in some sense) trajectories. Another important question is whether it is possible to connect the additional variable ξ to some properties of individual neurons, which could lead to classifying neurons in terms of their role in the dynamics. We mention as a final example of an open problem, the issue of including learning in the models. In the setting of (1.1), learning can be simply incorporated in the model by considering time-dependent synaptic weights $w_{i,j;N}(t)$ together with some equation prescribing the evolution of those weights. This was recognized as being a critical mechanism as early as the famous Hebb rule in [50]. But it is unclear how to model this kind of learning appropriately while keeping sparse connections and a mean-field scaling, or whether the present approach would remain valid for such models. The mean-field limit has been derived [73,78] for neuron networks incorporating learning mechanisms, and also in [2] for an opinion dynamics model. But those results impose the strong algebraic constraint that $w_{i_1,j;N} = w_{i_2,j;N}$, for all $i_1, i_2 \neq j$.

The rest of the paper is structured as follows. In Section 2 we present our approach of directly obtaining stability estimates, starting from the extended BBGKY hierarchy from the non-exchangeable system (1.1), the corresponding Vlasov hierarchy, and their a priori estimates. The main stability result, as a quantitative version of (1.11), is stated as Theorem 2.6.

The subsequent sections are about rigorously proving the results in Section 2. We discuss in Section 3 the properties of the weak norms denoted by $H_\eta^{-1\otimes k}$ that we use throughout the quantitative estimates. In Section 4 we revisit the limiting observables $\tau_\infty(T, w, f)$, $T \in \mathcal{T}$, to show that they are well defined. Finally, with the preliminaries done in Sections 3 and 4, Section 5 is devoted to the proofs of the main results of Section 2, including Theorem 2.6.

2. Quantitative stability estimates

2.1. A tensorized negative Sobolev norm

This subsection is dedicated to the introduction of the $H_\eta^{-1\otimes k}$-norm along with its basic properties. While it is straightforward, the specific choice of this norm plays a key role in our later estimates as it leads to good commutator estimates. Introducing the mollification kernel

$$K(x) := \frac{1}{\pi} \int_0^\infty \exp(-|x| \cosh(\xi))\, d\xi,$$

we may define the $H_\eta^{-1\otimes k}$-norm as follows.

Definition 2.1. For any function F defined on $\mathbb{R}$, denote its tensorization to $\mathbb{R}^k$ by

$$F^{\otimes k}(z_1, \dots, z_k) := \prod_{m=1}^{k} F(z_m), \quad \forall (z_1, \dots, z_k) \in \mathbb{R}^k.$$

We then define

$$\|g\|_{H^{-1 \otimes k}} := \|K^{\otimes k} \star g\|_{L^2(\mathbb{R}^k)},$$

and for any weight function η on $\mathbb{R}$,

$$\|g\|_{H_\eta^{-1 \otimes k}} := \|K^{\otimes k} \star (g\eta^{\otimes k})\|_{L^2(\mathbb{R}^k)}.$$

The introduction of the weight η is motivated by the need for some control on the decay of the solutions at infinity since we work on the whole of $\mathbb{R}$. We simply choose some $\alpha > 0$ and define

$$\eta(x) = \eta_\alpha(x) := C_\alpha \exp\bigl(\sqrt{1 + \alpha^2 x^2}\bigr), \quad C_\alpha = \int_{\mathbb{R}} \exp\bigl(-\sqrt{1 + \alpha^2 x^2}\bigr)\, \mathrm{d}x.$$

Our definition of $H_\eta^{-1 \otimes k}$ leads to a topology that is equivalent to the classical weak-* topology of $\mathcal{M}(\mathbb{R}^k)$.

Lemma 2.2. *Consider any $a > 0$, $C > 0$, $0 < \alpha < a$ (which determines $\eta = \eta_\alpha$) and any sequence*

$$\{g_n\}_{n=1}^{\infty} \subset \bigl\{g \in \mathcal{M}(\mathbb{R}^k) : \textstyle\int_{\mathbb{R}^k} \exp(a \sum_{m=1}^{k} |z_m|)|g|(z)\, \mathrm{d}z \le C\bigr\}.$$

Then the following are equivalent:

- $g_n \overset{*}{\rightharpoonup} g_\infty$ *under the weak-* topology of* $\mathcal{M}(\mathbb{R}^k)$.
- $\|g_n - g_\infty\|_{H_\eta^{-1 \otimes k}} \to 0$.

The proof of Lemma 2.2 is postponed to Section 3, where we also conduct a deeper examination of the relationship between the $H_\eta^{-1 \otimes k}$-norm and classical negative Sobolev norms. The use of weak distances such as Wasserstein distances is classical in the derivation of the mean-field limit, in particular when looking at the notion of empirical measures.

However, our observables are bounded functions at any $t > 0$, for which we can even prove bounds, and a main motivation for the use of weak norms stems from the singularity introduced by the Poisson jump processes. The usefulness of negative Sobolev norms in that context has been highlighted in works such as [75]. We also mention [36] which considers a somewhat relaxed IF model with connections depending on the spatial structure of neurons. However, instead of studying the 1-particle distribution, we use tensorized $H_\eta^{-1 \otimes k}$-norms to investigate the joint law $f_N^{i_1,\dots,i_k}$ and the observables, which seems to be a novel approach in this context.

2.2. From the original SDE system to the extended BBGKY hierarchy

We show in this subsection that the observables, as defined in Definition 1.2, satisfy an extended BBGKY hierarchy.

We first recall the Liouville or forward Kolmogorov equation that is satisfied by the full joint law f_N of solutions to the SDE (1.1),

$$\partial_t f_N(t,x) + \sum_{i=1}^{N} \Big[\partial_{x_i}(\mu(x_i) f_N(t,x)) - \frac{\sigma^2}{2} \partial_{x_i}^2 f_N(t,x) + \nu(x_i) f_N(t,x) - \delta_0(x_i) \Big(\int_{\mathbb{R}} \nu(y_i) f_N(t, y - (w_N)_{\cdot,i}^{\top})\, \mathrm{d}y_i \Big) \Big|_{\substack{\forall j \neq i \\ y_j = x_j}} \Big] = 0, \tag{2.1}$$

$$(w_N)_{\cdot,i}^{\top} = (w_{1,i;N}, \dots, w_{N,i;N}) \in \mathbb{R}^N, \quad \forall 1 \le i \le N,$$

where δ_0 is the Dirac delta function at the origin. The "spike vector" $(w_N)_{\cdot,i}^{\top}$ corresponds to the ith column of connectivity matrix w_N that account for the jumps when the ith neuron fires.

From the Kolmogorov equation, we may derive equations on each observable.

Proposition 2.3. *Assume that $\mu, \nu \in W^{1,\infty}$ and constant $\sigma > 0$. Let $w_N := (w_{i,j;N})_{i,j=1}^N$ be the connectivity matrix and $(X_0^{1;N}, \dots, X_0^{N;N})$ be the initial data with $g_N = \mathrm{Law}(X_0^{1;N}, \dots, X_0^{N;N})$.*

Then there exists a unique solution $(X_t^{1;N}, \dots, X_t^{N;N})$ solving SDE (1.1) for all $t \ge 0$, whose law

$$f_N(t,\cdot) = \mathrm{Law}(X_t^{1;N}, \dots, X_t^{N;N})$$

is the unique distributional solution of the Liouville equation (2.1) with initial data g_N. In addition, the observables

$$\tau_N(T) = \tau_N(T, w_N, f_N), \quad \forall T \in \mathcal{T}$$

solve the extended version of the BBGKY hierarchy with remainder terms: For all $T \in \mathcal{T}$,

$$\begin{aligned} &\partial_t \tau_N(T)(t,z) \\ &= \sum_{m=1}^{|T|} \Big\{ \Big[-\partial_{z_m}(\mu(z_m)\tau_N(T)(t,z)) + \frac{\sigma^2}{2}\partial_{z_m}^2 \tau_N(T)(t,z) - \nu(z_m)\tau_N(T)(t,z) \\ &\quad + \delta_0(z_m) \Big(\int_{\mathbb{R}} \nu(u_m)(\tau_N(T)(t,u) + \mathcal{R}_{N,T,m}(t,u))\, \mathrm{d}u_m \Big) \Big|_{\substack{\forall n \neq m \\ u_n = z_n}} \Big] \\ &\quad - \partial_{z_m} \Big[\int_{\mathbb{R}} \nu(z_{|T|+1})(\tau_N(T+m)(t,z) + \widetilde{\mathcal{R}}_{N,T+m,|T|+1}(t,z))\, \mathrm{d}z_{|T|+1} \Big] \Big\}, \end{aligned} \tag{2.2}$$

where the remainder terms are given by

$$\begin{aligned}\mathcal{R}_{N,T,m}(t,z) &:= \frac{1}{N}\sum_{i_1,\ldots,i_{|T|}=1}^{N} w_{N,T}(i_1,\ldots,i_{|T|}) \\ &\qquad\times \big(f_N^{i_1,\ldots,i_{|T|}}(t,z-w_{N;i_m}^{i_1,\ldots,i_{|T|}}) - f_N^{i_1,\ldots,i_{|T|}}(t,z)\big), \\ \widetilde{\mathcal{R}}_{N,T,m}(t,z) &:= \int_0^1 \frac{1}{N}\sum_{i_1,\ldots,i_{|T|}=1}^{N} w_{N,T}(i_1,\ldots,i_{|T|}) \\ &\qquad\times \big(f_N^{i_1,\ldots,i_{|T|}}(t,z-rw_{N;i_m}^{i_1,\ldots,i_{|T|}}) - f_N^{i_1,\ldots,i_{|T|}}(t,z)\big)\,\mathrm{d}r,\end{aligned} \tag{2.3}$$

and the $w_{N;j}^{i_1,\ldots,i_k}$ *are defined as the restriction of the "spike vector"* $(w_N)_{\cdot,j}^{\top}$ *to the marginal space, namely*

$$w_{N;j}^{i_1,\ldots,i_k} := (w_{i_n,j;N})_{n=1}^{k} = (w_{i_1,j;N},\ldots,w_{i_k,j;N}) \in \mathbb{R}^k.$$

The proof of the proposition will be done in Section 5.1. Unlike the standard BBGKY hierarchy which usually gives a closed equation involving $f_{N,k}$ and the next marginal $f_{N,k+1}$, the hierarchy of equations derived here is only approximate as the remainder terms do not only depend on our observables. Thus, an essential part of our approach is to prove that as the strength of pairwise interaction $\max_{1\le i,j\le N}|w_{i,j;N}|$ goes to 0 (which is assumption (1.8) in Theorem 1.7), those remainder terms $\mathcal{R}$ and $\widetilde{\mathcal{R}}$ vanish in the $H_\eta^{-1\otimes k}$ sense. As we mentioned earlier, it is a main motivation for choosing $H_\eta^{-1\otimes k}$ as its specific form. This result is precisely formulated in Proposition 3.6 in the next section.

We also note that the presence of the remainder terms $\mathcal{R}$ and $\widetilde{\mathcal{R}}$ is not only a consequence of the Poisson jump process. Consider the more classical first-order dynamics

$$\begin{aligned}X_t^{i;N} &= X_0^{i;N} + \int_0^t \mu(X_s^{i;N})\,\mathrm{d}s + \int_0^t \sigma(X_s^{i;N})\,\mathrm{d}\mathbf{B}_s^i \\ &\quad + \sum_{j\neq i} w_{i,j;N}\int_0^t \nu(X_s^{i;N},X_s^{j;N})\,\mathrm{d}s.\end{aligned}$$

Depending on the specific form of $\nu(\cdot,\cdot)$, the term $\widetilde{\mathcal{R}}_{N,T+m,|T|+1}$ may vanish, but the term $\mathcal{R}_{N,T,m}$ is always present. More than the specific form of the dynamics, the remainders reflect the more essential difficulty that interaction between the first k neurons $i_1,\ldots,i_k$ cannot be fully described by the observables as defined in Definition 1.2.

This is also one of the crucial distinctions that separates the method in this article from [55]. The observables in [55] are similar to the limiting observables τ_∞ in this article, but are constructed from the solutions of the McKean–Vlasov SDE where the interaction felt by one agent $X^{i;N}$ is determined not by the exact $X^{j;N}$, but the $\mathrm{Law}(X^{j;N})$. This leads to a simplified hierarchy without remainders. On the other hand, in this article all the observables are constructed directly from the solution of (1.1), hence the extended, approximate BBGKY hierarchy (2.2) reflects the dynamics of the original non-exchangeable system.

We conclude the subsection with a priori estimates of the absolute observables $|\tau_N|$ whose proof is also postponed to Section 5.1.

Proposition 2.4. *Let $N \geq 1$, $t_* > 0$, and $\alpha > 0$ (which determines $\eta = \eta_\alpha$). Assume that the connectivity matrix $w_N := (w_{i,j;N})_{i,j=1}^N$ and joint law $f_N \in L^\infty([0,t_*]; \mathcal{M}_+(\mathbb{R}^N))$ solves the Kolmogorov equation (2.1) in the sense of distributions. For any $T \in \mathcal{T}$, assume that at $t = 0$,*

$$\| |\tau_N|(T)(0,\cdot)\eta^{\otimes|T|}\|_{\mathcal{M}(\mathbb{R}^{|T|})} \leq C_\eta(T) < \infty.$$

Then there exists $A_\eta > 0$ only depending on α, $\|\mu\|_{W^{1,\infty}}$, $\|\nu\|_{W^{1,\infty}}$, σ, and

$$\max\left(\max_i \sum_j |w_{i,j;N}|, \max_j \sum_i |w_{i,j;N}|\right),$$

such that

$$\| |\tau_N|(T)(t,\cdot)\eta^{\otimes|T|}\|_{\mathcal{M}(\mathbb{R}^{|T|})} \leq C_\eta(T)(\exp(A_\eta t_*))^{|T|}, \quad \forall t \in [0,t_*],$$

and

$$\| |\tau_N|(T)(t,\cdot)\|_{H_\eta^{-1\otimes|T|}} \leq C_\eta(T)(\|K\|_{L^2(\mathbb{R})}\exp(A_\eta t_*))^{|T|}, \quad \forall t \in [0,t_*]. \tag{2.4}$$

Let us emphasize again that this proposition is about $|\tau_N|$ the absolute observables, which are non-negative measures obtained by linear combinations of laws $f_N^{i_1,\ldots,i_k}$, $1 \leq i_1,\ldots,i_k \leq N$. We do not expect a straightforward extension to the τ_N as the potential cancellations of positive and negative terms in the dynamics makes the problem much less tractable.

2.3. From the limiting Vlasov equation to the limiting hierarchy

The following proposition states that the limiting observables τ_∞ defined from the limiting Vlasov equation (1.3)–(1.4) satisfy the limiting hierarchy (2.5), which is similar to the BBGKY hierarchy (2.2) in Proposition 2.3 but without the remainder terms $\mathcal{R}$ and $\widetilde{\mathcal{R}}$. In that sense the limiting hierarchy provides closed recursive relations of the family $\tau_\infty(T)$, for all $T \in \mathcal{T}$. In particular, the quantitative estimates proved later would imply the uniqueness of solutions to the hierarchy for a given choice of initial data.

Proposition 2.5. *Assume that $\mu, \nu \in W^{1,\infty}$ and constant $\sigma > 0$. Then for any $t_* > 0$, $\alpha > 0$ (which determines $\eta = \eta_\alpha$), any connectivity kernel $w \in \mathcal{W}$, and any initial extended density $g \in L^\infty([0,1]; H_\eta^{-1} \cap \mathcal{M}_+(\mathbb{R}))$, there exists a unique*

$$f \in L^\infty([0,t_*] \times [0,1]; H_\eta^{-1} \cap \mathcal{M}_+(\mathbb{R}))$$

solving the Vlasov equation (1.3)–(1.4) in the sense of distributions. Furthermore, the observables $\tau_\infty(T) = \tau_\infty(T, w, f)$, for all $T \in \mathcal{T}$ are bounded by

$$\|\tau_\infty(T,w,f)(t,\cdot)\|_{H_\eta^{-1\otimes|T|}} \leq \|w\|_{\mathcal{W}}^{|T|-1}\|f\|_{L^\infty_{t,\xi}(H_\eta^{-1})_x}^{|T|}, \quad \forall t \in [0,t_*],\ T \in \mathcal{T},$$

and solve the following non-exchangeable extended version of the Vlasov hierarchy: For all $T \in \mathcal{T}$,

$$\begin{aligned}
&\partial_t \tau_\infty(T)(t,z) \\
&\quad = \sum_{m=1}^{|T|} \bigg\{ \Big[-\partial_{z_m}(\mu(z_m)\tau_\infty(T)(t,z)) + \frac{\sigma^2}{2}\partial_{z_m}^2 \tau_\infty(T)(t,z) \\
&\qquad\qquad - \nu(z_m)\tau_\infty(T)(t,z) + \delta_0(z_m)\Big(\int_{\mathbb{R}} \nu(u_m)\tau_\infty(T)(t,u)\,\mathrm{d}u_m\Big)\Big|_{\substack{\forall n\neq m\\ u_n = z_n}} \Big] \\
&\qquad - \partial_{z_m}\Big[\int_{\mathbb{R}} \nu(z_{|T|+1})\tau_\infty(T+m)(t,z)\,\mathrm{d}z_{|T|+1}\Big]\bigg\}. \qquad (2.5)
\end{aligned}$$

The proof of the proposition is again done in Section 5.1.

2.4. Quantitative stability estimates between the hierarchies

We are now ready to state the main quantitative result in this paper, which compares the observables $\tau_N(T, w_N, f_N)$ satisfying the approximate hierarchy (2.2)–(2.3) to $\tau_\infty(T)$ satisfying the limiting hierarchy (2.5). The proof of the theorem and the exact derivation of constants C_1, C_2 in the estimate are performed in Section 5.2.

Theorem 2.6. *Assume that $\mu, \nu \in W^{1,\infty}$, constant $\sigma > 0$ and $N \geq 1$. Let $w_N := (w_{i,j;N})_{i,j=1}^N \in \mathbb{R}^{N\times N}$ be a connectivity matrix and $f_N^{i_1,\dots,i_k}$, for all $\{i_1,\dots,i_k\} \subset \{1,\dots,N\}$ be marginal laws, from which the hierarchy of observables $\tau_N(T, w_N, f_N)$ and the absolute observables $|\tau_N|(T, w_N, f_N)$ are defined and satisfy (2.2)–(2.3) in the distributional sense. Denote the strength of pairwise interaction as*

$$\bar{w}_N := \max_{1\leq i,j\leq N} |w_{i,j;N}|.$$

In addition, let $\tau_\infty(T) \in L^\infty([0,t_*]; \mathcal{M}(\mathbb{R}^{|T|}))$, for all $T \in \mathcal{T}$ satisfy (2.5) in the distributional sense.

For some choice of $\lambda > 0$ and $\alpha > 0$ (which determines $\eta = \eta_\alpha$), assume that there exists $n \in \mathbb{N}$ such that

$$\bar{\varepsilon} := C_1[\exp((2+2\alpha)n\bar{w}_N) - 1] + (1/4)^n < 1,$$

where C_1 is a constant depending only on $\|\mu\|_{W^{1,\infty}}$, $\|\nu\|_{W^{1,\infty}}$, σ and the scaling factor $\lambda > 0$. Then the following estimate holds: For any tree $T_* \in \mathcal{T}$,

$$\begin{aligned}
&\sup_{t\leq t_*} (\lambda/8)^{|T_*|} \|\tau_N(T_*, w_N, f_N)(t,\cdot) - \tau_\infty(T_*)(t,\cdot)\|_{H_\eta^{-1\otimes|T_*|}}^2 \qquad (2.6)\\
&\leq C_2 C_{\lambda;\eta}^2 \Big\{ \max\Big(\bar{\varepsilon}, \max_{\substack{|T|\leq\\ \max(n,|T_*|)}} \frac{(\lambda/8)^{|T|}}{C_{\lambda;\eta}^2} \|\tau_N(T, w_N, f_N)(0,\cdot) - \tau_\infty(T)(0,\cdot)\|_{H_\eta^{-1\otimes|T|}}^2\Big)\Big\}^{\frac{1}{C_2}},
\end{aligned}$$

where C_2 depends only on t_*, $\|\mu\|_{W^{1,\infty}}$, $\|\nu\|_{W^{1,\infty}}$, σ, and $\lambda > 0$, and where $C_{\lambda;\eta}$ depends on the following a priori estimate:

$$\sup_{t \le t_*} \max_{\substack{|T| \le \\ \max(n,|T_*|)}} \lambda^{\frac{|T|}{2}} \left(\| |\tau_N|(T, w_N, f_N)(t,\cdot)\|_{H_\eta^{-1 \otimes |T|}} + \|\tau_\infty(T)(t,\cdot)\|_{H_\eta^{-1\otimes|T|}} \right) \le C_{\lambda;\eta}. \tag{2.7}$$

Remark 2.7. The values of λ, α, and n must be chosen carefully for this result to be useful. The scaling factors λ and α need to be selected so that the various norms in the theorem are finite, to fit with the existing a priori estimates. Also, we need to have n such that $\bar{\varepsilon}$ is small enough, which would typically lead to taking $n \sim \frac{|\log \bar{w}_N|}{\bar{w}_N}$. However, n also enters into the definition of $C_{\lambda:\eta}$ in an implicit way as a larger value of n forces one to take the max over more trees T. Hence the actual optimal value of n is not so easy to determine unless (2.7) is a priori given where the maximum is replaced by the supremum over all trees $T \in \mathcal{T}$.

Stability and uniqueness estimates on the kind of generalized hierarchy that we are dealing with here are notoriously difficult, with only limited results available. As we mentioned before there are obvious similarities between our approach and the hierarchy derived in [55] or the strong estimates on the classical BBGKY hierarchy in [13] (leading for example to the mean-field limit to the Vlasov–Fokker–Planck–Poisson equation). We also mention results around the wave kinetic equation in [31, 32].

A major difference in Theorem 2.6 is that the observables τ_N do not solve an exact hierarchy and the remainder terms only vanish in some weak norms. As we briefly explained earlier, this forces the use of the $H_\eta^{-1\otimes|T|}$ norm both to control the remainders and to have appropriate commutator estimates, which is the main technical innovation in the paper.

We also emphasize that the general method used to derive stability estimates relies on recursive inequalities, which often leads to a blow-up in finite time. This does not occur here because we can derive a priori estimates, namely (2.7) from Proposition 2.4 and Proposition 2.5, that are strong enough with respect to the weak norms that we are using.

2.5. Proving Theorem 1.7 from our quantitative estimates

We conclude this subsection by explaining how Theorem 1.7 follows from all the estimates presented here.

Proof of Theorem 1.7. The first step is to make sure that we can apply Theorem 2.6 from the assumptions (1.7)–(1.10) in Theorem 1.7. More precisely, we tend to show that (2.7) in Theorem 2.6 holds for some well-chosen $\lambda > 0$ and $C_{\lambda;\eta} > 0$, and the maximum over $|T| \le \max(n, |T_*|)$ can actually be replaced by the supremum over all trees $T \in \mathcal{T}$.

Recall that for any $k \geq 1$,

$$\eta_a^{\otimes k}(z_1, \ldots, z_k) = C_a^k \exp\Bigg(\sum_{m=1}^{k} \sqrt{1 + a^2 z_m^2}\Bigg),$$

hence

$$\exp\Bigg(\sum_{m=1}^{k} a|z_m|\Bigg) \leq \eta_a^{\otimes k}(z_1, \ldots, z_k) \leq (C_a \exp(1))^k \exp\Bigg(\sum_{m=1}^{k} a|z_m|\Bigg).$$

Thus, from assumption (1.9) in Theorem 1.7, the following two inequalities for the initial data can immediately be derived:

$$\sup_N \||\tau_N|(T)(0,\cdot)\eta_a^{\otimes|T|}\|_{\mathcal{M}(\mathbb{R}^{|T|})} \leq (M_a C_a \exp(1))^{|T|}, \quad \forall T \in \mathcal{T},$$
$$\|f(0,\cdot,\cdot)\|_{L_\xi^\infty(H_{\eta a}^{-1})_x} \leq \|K\|_{L^2(\mathbb{R})} M_a C_a \exp(1).$$

Now, applying Propositions 2.4 and 2.5 to the two initial bounds, we obtain the exponential moment bound

$$\sup_N \int_{\mathbb{R}^{|T|}} \exp\Bigg(a \sum_{m=1}^{|T|} |z_m|\Bigg) |\tau_N|(T)(t, \mathrm{d}z)$$
$$\leq \||\tau_N|(T)(t,\cdot)\eta_a^{\otimes|T|}\|_{\mathcal{M}(\mathbb{R}^{|T|})} \leq \big(M_a C_a \exp(1)\exp(A_\eta t_*)\big)^{|T|}, \quad \forall t \in [0, t_*], T \in \mathcal{T},$$

and a priori energy bounds

$$\||\tau_N|(T)(t,\cdot)\|_{H_{\eta a}^{-1 \otimes |T|}} \leq (\|K\|_{L^2(\mathbb{R})} M_a C_a \exp(1)\exp(A_\eta t_*))^{|T|}, \quad \forall t \in [0, t_*],\ T \in \mathcal{T},$$
$$\|\tau_\infty(T, w, f)(t,\cdot)\|_{H_{\eta a}^{-1 \otimes |T|}} \leq \|w\|_{\mathcal{W}}^{|T|-1} \|f\|_{L_{t,\xi}^\infty(H_{\eta a}^{-1})_x}^{|T|}, \quad \forall t \in [0, t_*],\ T \in \mathcal{T},$$

where the coefficient A_η inside the exponent now only depends on a, $\|\mu\|_{W^{1,\infty}}$, $\|\nu\|_{W^{1,\infty}}$, σ, and

$$\max\Bigg(\max_i \sum_j |w_{i,j;N}|,\ \max_j \sum_i |w_{i,j;N}|\Bigg).$$

This guarantees (2.7) where the maximum over $|T| \leq \max(n, |T_*|)$ is replaced by the supremum over all trees $T \in \mathcal{T}$, with λ, $C_{\lambda;\eta}$ chosen as

$$\lambda = \min\big((\|K\|_{L^2(\mathbb{R})} M_a C_a \exp(1)\exp(A_\eta t_*))^{-2},$$
$$(\max(\|w\|_{\mathcal{W}}, 1)\|f\|_{L_{t,\xi}^\infty(H_\eta^{-1})_x})^{-2}\big), \quad C_{\lambda;\eta} = 1.$$

Hence the assumptions of Theorem 2.6 are satisfied and we apply it to the following point:

- Using (1.7) and (1.9), we choose the coefficients $\alpha \in (0, a)$, $\lambda > 0$, $C_{\lambda;\eta} > 0$ in (2.7) independent of N, and the supremum in (2.7) is taken over all possible $T \in \mathcal{T}$. This implies in particular a uniform bound on exponential moments with coefficient $a > 0$ so that Lemma 2.2 applies.

- Fix $T_* \in \mathcal{T}$. For any $\varepsilon > 0$, choose sufficiently large n, such that

$$(\lambda/8)^{-\frac{|T_*|}{2}} \sqrt{C_2} C_{\lambda;\eta} [2(1/4)^n]^{1/2C_2} \leq \varepsilon.$$

- By (1.8), we choose a sufficiently large N_1, such that for all $N \geq N_1$ the corresponding

$$\bar{w}_N := \max_{1 \leq i,j \leq N} |w_{i,j;N}|$$

is sufficiently small such that

$$C_1[\exp((2+2\alpha)n\bar{w}_N) - 1] \leq (1/4)^n.$$

- Notice that there are only finitely many $T \in \mathcal{T}$ satisfying $|T| \leq \max(n, |T_*|)$. By (1.10) on the weak-* convergence of initial data, and by Lemma 2.2, choose a sufficiently large $N_2 \geq 1$ such that for all $N \geq N_2$,

$$\max_{\substack{|T| \leq \\ \max(n,|T_*|)}} (\lambda/8)^{|T|} \|\tau_N(T, w_N, f_N)(0,\cdot) - \tau_\infty(T)(0,\cdot)\|^2_{H_\eta^{-1 \otimes |T|}} / (4C^2_{\lambda;\eta}) \leq 2(1/4)^n.$$

- In summary, for any $T_* \in \mathcal{T}$ and any $\varepsilon > 0$, by taking $N \geq \max(N_1, N_2)$ according to our previous discussion and applying Theorem 2.6, we obtain that

$$\sup_{t \in [0,t_*]} \|\tau_N(T_*, w_N, f_N)(t,\cdot) - \tau_\infty(T_*)(t,\cdot)\|_{H_\eta^{-1 \otimes |T_*|}} \leq \varepsilon.$$

- Invoking Lemma 2.2 again, we finally deduce that

$$\lim_{N \to \infty} \tau_N(T, w_N, f_N)(t,\cdot) = \tau_\infty(T)(t,\cdot), \quad \forall T \in \mathcal{T}$$

in the weak-* topology of $L^\infty([0,t_*], \mathcal{M})$. ∎

3. The weak norm and the exponential moments

3.1. Basic properties

We first revisit our definition of the kernel K, and introduce another kernel, denoted as Λ, as follows:

$$K(x) := \frac{1}{\pi} \int_0^\infty \exp(-|x|\cosh(\xi))\,\mathrm{d}\xi, \quad \Lambda(x) := \frac{1}{2}\exp(-|x|), \quad \forall x \in \mathbb{R}.$$

For $x > 0$, the kernel K is, in fact, the zeroth order modified Bessel function of the second type. From the known properties of Bessel functions, K is a non-negative, radially decreasing L^2 function, and satisfies

$$K \star K = \Lambda, \quad \hat{K}(\xi) = \int_{\mathbb{R}} K(x)\exp(-2\pi i x\xi)\,\mathrm{d}x = \frac{1}{\sqrt{1+4\pi^2\xi^2}}.$$

It is easy to extend the identity $K \star K = \Lambda$ to the tensorized kernels $K^{\otimes k} \star K^{\otimes k} = \Lambda^{\otimes k}$, which yields the following equivalent formalism of $H^{-1\otimes k}$ by Fourier analysis:

$$\begin{aligned}\|f\|_{H^{-1\otimes k}}^2 &= \int_{z\in\mathbb{R}^k} [K^{\otimes k} \star f(z)]^2 \,\mathrm{d}z = \int_{z\in\mathbb{R}^k} f(z)[\Lambda^{\otimes k} \star f(z)]\,\mathrm{d}z \\ &= \int_{\xi\in\mathbb{R}^k} \Big(\prod_{m=1}^{k} \frac{1}{1+4\pi^2\xi_m^2}\Big) \hat{f}(\xi)\hat{f}(\xi)\,\mathrm{d}\xi.\end{aligned}$$

In one dimension, it is straightforward that our notion of the $H^{-1\otimes k}$-norm for $k = 1$ is equivalent to the negative Sobolev norm of $H^{-1}(\mathbb{R})$, i.e.

$$\|f\|_{H^{-1\otimes 1}} = \|f\|_{H^{-1}(\mathbb{R})},$$

provided we define $H^s(\mathbb{R})$ as

$$\|g\|_{H^s(\mathbb{R})}^2 := \int_{\mathbb{R}} (1+4\pi^2\xi^2)^s |\hat{g}(\xi)|^2 \,\mathrm{d}\xi,$$

for any $s \in \mathbb{R}$.

This also gives us the duality formula

$$\|f\|_{H^{-s}(\mathbb{R})} = \sup_{\|g\|_{H^s(\mathbb{R})}\le 1} \Big| \int_{\mathbb{R}} f(x)g(x)\,\mathrm{d}x \Big|,$$

and the inequality from the Leibniz rule for $s = 1$,

$$\begin{aligned}\|\nu f\|_{H^{-1}(\mathbb{R})} = \sup_{\|g\|_{H^1}\le 1} \Big| \int_{\mathbb{R}} g(x)\nu(x)f(x)\,\mathrm{d}x \Big| &\le \sup_{\|g\|_{H^1}\le 1} \|g\nu\|_{H^1} \|f\|_{H^{-1}} \\ &\le 2\|\nu\|_{W^{1,\infty}} \|f\|_{H^{-1}(\mathbb{R})}.\end{aligned}$$

3.2. Tensorization properties

In higher dimensions, our notion of the $H^{-1\otimes k}$-norm is the tensorization of the $H^{-1}(\mathbb{R})$-norm to $\mathbb{R}^k$.

Lemma 3.1. *For any weight function $\eta: \mathbb{R} \to \mathbb{R}_+$, one has*

$$\|f^{\otimes k}\|_{H^{-1\otimes k}} = (\|f\|_{H^{-1}(\mathbb{R})})^k, \quad \|f^{\otimes k}\|_{H_\eta^{-1\otimes k}} = (\|f\|_{H_\eta^{-1}(\mathbb{R})})^k.$$

Proof. One has that

$$\begin{aligned}\|f^{\otimes k}\|_{H_\eta^{-1\otimes k}}^2 &= \int_{z\in\mathbb{R}^k} [K^{\otimes k} \star (f^{\otimes k}\eta^{\otimes k})(z)]^2 \,\mathrm{d}z = \prod_{m=1}^{k} \int_{z_m\in\mathbb{R}} [K \star (f\eta)(z_m)]^2 \,\mathrm{d}z_m \\ &= (\|f\|_{H_\eta^{-1}(\mathbb{R})})^{2k}.\end{aligned}$$

The unweighted case of $H^{-1\otimes k}$ is naturally included by choosing $\eta \equiv 1$. ∎

It is important to emphasize however that the tensorized $H^{-1\otimes k}$-norm is weaker than the standard $H^{-1}(\mathbb{R}^k)$-norm since in the Fourier domain

$$\prod_{m=1}^{k} \frac{1}{1+4\pi^2\xi_m^2} \ll \frac{1}{1+4\pi^2\sum_{m=1}^{k}\xi_m^2}.$$

This shows that the energy distributed along the diagonals of the Fourier domain have a much smaller contribution to the tensorized $H^{-1\otimes k}$-norm than to the $H^{-1}(\mathbb{R}^k)$-norm.

Similarly, while it is possible to include $\mathcal{M}(\mathbb{R}^k)$ into the standard $H^{-s}(\mathbb{R}^k)$, the order $s>0$ in such Sobolev inequalities depends on the dimension k, namely $s>k/2$. On the other hand, the following lemma holds for our notion of $H^{-1\otimes k}$-norm.

Lemma 3.2. *Consider $g \in \mathcal{M}(\mathbb{R}^k)$ and any weight function $\eta \in L^1(\mathbb{R},\mathbb{R}_+)$ such that $\eta^{\otimes k}$ is integrable against g. Then*

$$\|g\|_{H^{-1\otimes k}} := \|K^{\otimes k} \star g\|_{L^2(\mathbb{R}^k)} \le \|K\|_{L^2(\mathbb{R})}^k \|g\|_{\mathcal{M}(\mathbb{R}^k)},$$
$$\|g\|_{H_\eta^{-1\otimes k}} := \|K^{\otimes k} \star (g\eta^{\otimes k})\|_{L^2(\mathbb{R}^k)} \le \|K\|_{L^2(\mathbb{R})}^k \|g\eta^{\otimes k}\|_{\mathcal{M}(\mathbb{R}^k)}.$$

Proof. The proof is a simple application of the convolutional inequality. ■

Hence $\mathcal{M}(\mathbb{R}^k)$ is naturally included in $H^{-1\otimes k}$, and can also be included in $H_\eta^{-1\otimes k}$, provided that the measure has the right moment bound.

The next lemma extends the inequality from the Leibniz rule to any dimension.

Lemma 3.3. *Consider ν_m of the form*

$$\nu_m = 1 \otimes \cdots \otimes \nu \otimes \cdots \otimes 1,$$

where $\nu \in W^{1,\infty}(\mathbb{R})$ appears in the mth coordinate, i.e. $\nu_m(z) = \nu(z_m)$. Then for any $f \in \mathcal{M}(\mathbb{R}^k) \cap H^{-1\otimes k}$, the following inequality holds:

$$\|\nu_m f\|_{H^{-1\otimes k}} \le 2\|\nu\|_{W^{1,\infty}(\mathbb{R})} \|f\|_{H^{-1\otimes k}},$$

while for $f \in \mathcal{M}(\mathbb{R}^k) \cap H_\eta^{-1\otimes k}$, we have the corresponding

$$\|\nu_m f\|_{H_\eta^{-1\otimes k}} \le 2\|\nu\|_{W^{1,\infty}(\mathbb{R})} \|f\|_{H_\eta^{-1\otimes k}}.$$

Proof. Let us first discuss the unweighted inequality and without loss of generality consider ν_k that is non-constant in the kth dimension. Let us introduce the Fourier transform on the first $k-1$ dimensions,

$$\mathcal{F}^{\otimes k-1} \otimes I : \mathbb{R}^{k-1} \times \mathbb{R} \to \mathbb{R}^{k-1} \times \mathbb{R}.$$

It is easy to verify that

$$\begin{aligned}&(\mathcal{F}^{\otimes k-1} \otimes I)(K^{\otimes k} \star (\nu_k f))(\xi_1,\dots,\xi_{k-1},z_k)\\ &\quad= \Big(\prod_{m=1}^{k-1} \frac{1}{\sqrt{1+4\pi^2\xi_m^2}}\Big)(K \star_k (\nu_k \mathcal{F}^{\otimes k-1} f))(\xi_1,\dots,\xi_{k-1},z_k).\end{aligned}$$

By the Plancherel identity,

$$
\begin{aligned}
&\|\nu_k f\|^2_{H^{-1\otimes k}} \\
&= \int \left| \left(\prod_{m=1}^{k-1} \frac{1}{\sqrt{1+4\pi^2\xi_m^2}} \right) (K \star_k (\nu_k \mathcal{F}^{\otimes k-1} f))(\xi_1, \dots, \xi_{k-1}, z_k) \right|^2 \mathrm{d}\xi_1, \dots, \mathrm{d}\xi_{k-1} \mathrm{d}z_k \\
&= \int \left(\prod_{m=1}^{k-1} \frac{1}{1+4\pi^2\xi_m^2} \right) \|(\nu_k \mathcal{F}^{\otimes k-1} f)(\xi_1, \dots, \xi_{k-1}, \cdot)\|^2_{H^{-1}(\mathbb{R})} \mathrm{d}\xi_1, \dots, \mathrm{d}\xi_{k-1}.
\end{aligned}
$$

Since $\nu \in W^{1,\infty}(\mathbb{R})$,

$$
\begin{aligned}
&\|(\nu_k \mathcal{F}^{\otimes k-1} f)(\xi_1, \dots, \xi_{k-1}, \cdot)\|_{H^{-1}(\mathbb{R})} \\
&\qquad \le 2\|\nu\|_{W^{1,\infty}(\mathbb{R})} \|\mathcal{F}^{\otimes k-1} f(\xi_1, \dots, \xi_{k-1}, \cdot)\|_{H^{-1}(\mathbb{R})}.
\end{aligned}
$$

Hence

$$
\begin{aligned}
&\|\nu_k f\|^2_{H^{-1\otimes k}} \\
&\le 4\|\nu\|^2_{W^{1,\infty}(\mathbb{R})} \int \left(\prod_{m=1}^{k-1} \frac{1}{1+4\pi^2\xi_m^2} \right) \|\mathcal{F}^{\otimes k-1} f(\xi_1, \dots, \xi_{k-1}, \cdot)\|^2_{H^{-1}(\mathbb{R})} \mathrm{d}\xi_1, \dots, \mathrm{d}\xi_{k-1} \\
&= 4\|\nu\|^2_{W^{1,\infty}(\mathbb{R})} \|f\|^2_{H^{-1\otimes k}},
\end{aligned}
$$

which completes the proof of the unweighted inequality. Finally, for the weighted inequality, we can apply the unweighted inequality to obtain

$$
\begin{aligned}
\|\nu_m f\|_{H_\eta^{-1\otimes k}} = \|\nu_m f \eta^{\otimes k}\|_{H^{-1\otimes k}} &\le 2\|\nu\|_{W^{1,\infty}} \|f\eta^{\otimes k}\|_{H^{-1\otimes k}} \\
&= 2\|\nu\|_{W^{1,\infty}} \|f\|_{H_\eta^{-1\otimes k}}.
\end{aligned}
$$

■

3.3. The weak-* topology on measures

Now we proceed to the proof of Lemma 2.2, restated here.

Lemma 3.4. *Consider any $a > 0$, $C_a > 0$, $0 < \alpha < a$ (which determines $\eta = \eta_\alpha$) and any sequence*

$$
\{g_n\}_{n=1}^\infty \subset \left\{ g \in \mathcal{M}(\mathbb{R}^k) : \int_{\mathbb{R}^k} \exp\left(a \sum_{m=1}^k |z_m|\right) |g|(\mathrm{d}z) \le C_a \right\}. \tag{3.1}
$$

Then the following are equivalent:

- $g_n \overset{*}{\rightharpoonup} g_\infty$ *under the weak-* topology of* $\mathcal{M}(\mathbb{R}^k)$.
- $\|g_n - g_\infty\|_{H_\eta^{-1\otimes k}} \to 0$.

Proof of Lemma 2.2. A sequence $\{g_n\}_{n=1}^\infty$ satisfying (3.1) is uniformly tight and bounded in total variation norm. By Prokhorov's theorem, $\{g_n\}_{n=1}^\infty$ is sequentially precompact in the weak-* topology. Assuming now that $\|g_n - g_\infty\|_{H_\eta^{-1\otimes k}} \to 0$, the definition of $H_\eta^{-1\otimes k}$ directly implies that $(g_n - g_\infty)\eta^{\otimes k}$ converges to 0 in the sense of distributions. Since $\eta = \eta_\alpha$ is smooth, bounded from below and from above on any compact subset, it further yields that g_n converges to g_∞, still in the sense of distributions. Hence we immediately have that $g_n \overset{*}{\rightharpoonup} g_\infty$ under the weak-* topology of $\mathcal{M}(\mathbb{R}^k)$.

Assume now only that $g_n \overset{*}{\rightharpoonup} g_\infty$ under the weak-* topology of $\mathcal{M}(\mathbb{R}^k)$. First recall that

$$\eta^{\otimes k}(z_1, \dots, z_k) = C_\alpha^k \exp\left(\sum_{m=1}^k \sqrt{1+\alpha^2 z_m^2}\right) \le (C_\alpha \exp(1))^k \exp\left(\sum_{m=1}^k \alpha |z_m|\right).$$

The kernel $\Lambda^{\otimes k}$ is Lipschitz. Hence the convolution $\Lambda^{\otimes k} \star (g_n \eta^{\otimes k})$ is also Lipschitz, by

$$\|\Lambda^{\otimes k} \star (g_n \eta^{\otimes k})\|_{W^{1,\infty}} \le \|\Lambda^{\otimes k}\|_{W^{1,\infty}} \|g_n \eta^{\otimes k}\|_{\mathcal{M}}.$$

By the exponential moment bound (3.1), we have

$$\begin{aligned} \|g_n \eta^{\otimes k}\|_{\mathcal{M}} &= (C_\alpha \exp(1))^k \int_{\mathbb{R}^k} \exp\left(-(a-\alpha)\sum_{m=1}^k |z_m|\right) \exp\left(a \sum_{m=1}^k |z_m|\right) |g_n| \,(\mathrm{d}z) \\ &\le (C_\alpha \exp(1))^k C_a. \end{aligned}$$

This implies that $g_n \eta^{\otimes k}$ is precompact and hence converges to $g_\infty \eta^{\otimes k}$, so that

$$\Lambda^{\otimes k} \star (g_n \eta^{\otimes k}) \to \phi = \Lambda^{\otimes k} \star (g_\infty \eta^{\otimes k}) \in C(\mathbb{R}^k)$$

uniformly on all compact subsets of $\mathbb{R}^k$. Let $\rho \in C_c(\mathbb{R})$ such that $0 \le \rho \le 1$, $\rho([-1,1]) \equiv 1$, $\operatorname{supp} \rho \subset [-2,2]$, and denote $\rho_R(x) = \rho(x/R)$. Then

$$\begin{aligned} \|g_n\|_{H_\eta^{-1\otimes k}}^2 &= \int_{z \in \mathbb{R}^k} (g_n \eta^{\otimes k})(z) [\Lambda^{\otimes k} \star (g_n \eta^{\otimes k})(z)] \,\mathrm{d}z \\ &\le \int_{z \in \mathbb{R}^k} (g_n \eta^{\otimes k})(z) (\phi \rho_R^{\otimes k})(z) \,\mathrm{d}z \\ &\quad + \int_{z \in \mathbb{R}^k} (g_n \eta^{\otimes k})(z) \big((\Lambda^{\otimes k} \star (g_n \eta^{\otimes k}) - \phi) \rho_R^{\otimes k}\big)(z) \,\mathrm{d}z \\ &\quad + \int_{z \in \mathbb{R}^k} (g_n \eta^{\otimes k})(z) \big((\Lambda^{\otimes k} \star (g_n \eta^{\otimes k}))(1 - \rho_R^{\otimes k})\big)(z) \,\mathrm{d}z \\ &=: L_1 + L_2 + L_3. \end{aligned}$$

We note that $\phi \rho_R^{\otimes k}$ is continuous and compactly supported so that, for a fixed R, L_1 converges to 0 from the weak-* convergence of g_n. Also, L_2 directly converges to 0 for a fixed R from the uniform convergence of $\Lambda^{\otimes k} \star (g_n \eta^{\otimes k})$ to ϕ on compact sets.

Finally, for any $\varepsilon > 0$, choose sufficiently large $R > 0$ such that

$$[C_\alpha \exp(1 - (a-\alpha)R)]^k \le \frac{\varepsilon/6}{\|\Lambda^{\otimes k}\|_{L^\infty} (C_\alpha \exp(1))^k C_a^2}.$$

Then

$$L_3 \leq \|\Lambda^{\otimes k}\|_{L^\infty} (C_\alpha \exp(1))^k C_a \int_{z\in\mathbb{R}^k} |g_n \eta^{\otimes k}|(z)(1-\rho_R^{\otimes k})(z)\,\mathrm{d}z$$
$$\leq \|\Lambda^{\otimes k}\|_{L^\infty} (C_\alpha \exp(1))^k C_a [C_\alpha \exp(1-(a-\alpha)R)]^k C_a \leq \varepsilon/6.$$

This shows that L_3 converges to 0 as $R \to \infty$ uniformly in n, which concludes. ■

3.4. Bounding the remainder terms

As a first example of an application of our weak norms, we can derive a quantified weak convergence of the remainder terms $\mathcal{R}$ and $\widetilde{\mathcal{R}}$ in (2.3). However, L^p-norms are too sensitive to the pointwise density of the distribution, which makes it difficult to quantify vanishing translations. The following lemma shows how such translations are smoothed when mollified by $\Lambda^{\otimes k}$, making the behavior of $\mathcal{R}$ and $\widetilde{\mathcal{R}}$ milder in the $H^{-1\otimes k}$ sense and laying the ground for our future commutator estimates.

Lemma 3.5. *For any non-negative measure $f \in \mathcal{M}_+(\mathbb{R}^k)$ and vector $w \in \mathbb{R}^k$, the following pointwise estimate holds:*

$$|(\Lambda^{\otimes k} \star f)(z-w) - (\Lambda^{\otimes k} \star f)(z)| \leq [\exp(\|w\|_{\ell^1}) - 1](\Lambda^{\otimes k} \star f)(z), \quad \forall z \in \mathbb{R}^k.$$

Proof. It is straightforward that

$$|(\Lambda^{\otimes k} \star f)(z-w) - (\Lambda^{\otimes k} \star f)(z)| \leq \int_{\mathbb{R}^k} |\Lambda^{\otimes k}(z-w-y) - \Lambda^{\otimes k}(z-y)| f(y)\,\mathrm{d}y.$$

From the formula,

$$\Lambda^{\otimes k}(z) = \frac{1}{2}\exp\Big(-\sum_{m=1}^{k} |z_m|\Big),$$

we have that

$$|\Lambda^{\otimes k}(z-w-y) - \Lambda^{\otimes k}(z-y)| \leq [\exp(\|w\|_{\ell^1}) - 1]\Lambda^{\otimes k}(z-y).$$

We conclude the lemma by multiplying both sides by $f(y)$ and integrating by y. ■

The following proposition summarizes the estimates of $\mathcal{R}$ and $\widetilde{\mathcal{R}}$ terms.

Proposition 3.6. *Consider any $\alpha > 0$ (which determines $\eta = \eta_\alpha$), any connectivity matrix $w_N \in \mathbb{R}^{N\times N}$, and any joint law $f_N \in \mathcal{M}_+(\mathbb{R}^N)$. Let $\mathcal{R}_{N,T,m}$ and $\widetilde{\mathcal{R}}_{N,T,m}$ be the remainder terms as in* (2.3) *and let $|\tau_N|(T) = |\tau_N|(T, w_N, f_N)$ as in Definition* 1.2 *(where the variable t will be neglected). Then the following estimate holds:*

$$\max(\|\mathcal{R}_{N,T,m}\|^2_{H_\eta^{-1\otimes|T|}}, \|\widetilde{\mathcal{R}}_{N,T,m}\|^2_{H_\eta^{-1\otimes|T|}})$$
$$\leq [\exp((2+2\alpha)c(w_N, |T|)) - 1]\||\tau_N|(T)\|^2_{H_\eta^{-1\otimes|T|}},$$

where

$$c(w_N, |T|) := \min\Big(|T|(\max_{i,j}|w_{i,j;N}|), \max\Big(\max_j \sum_i |w_{i,j;N}|, \max_i \sum_j |w_{i,j;N}|\Big)\Big).$$

Notice that the right-hand side of the inequality is the "absolute" observables $|\tau_N|$ instead of τ_N as non-negativity plays a role in the proof. The constant $\alpha > 0$ takes the effect of weight $\eta = \eta_\alpha$ into account.

Proof of Proposition 3.6. Once we obtain the bound for $\mathcal{R}_{N,T,m}$, we can derive the same bound for $\widetilde{\mathcal{R}}_{N,T,m}$ by the Minkowski inequality. Hence, let us only consider $\mathcal{R}_{N,T,m}$. For simplicity, we also omit the t variable in the proof.

By definition,

$$\begin{aligned}\|\mathcal{R}_{N,T,m}\|^2_{H_\eta^{-1\otimes|T|}} &= \int_{\mathbb{R}^{|T|}} [(\mathcal{R}_{N,T,m}\eta^{\otimes n})(z)][\Lambda^{\otimes n} \star (\mathcal{R}_{N,T,m}\eta^{\otimes n})(z)]\,\mathrm{d}z \\ &\le \int_{\mathbb{R}^{|T|}} |(\mathcal{R}_{N,T,m}\eta^{\otimes n})(z)|\,|\Lambda^{\otimes n} \star (\mathcal{R}_{N,T,m}\eta^{\otimes n})(z)|\,\mathrm{d}z.\end{aligned}$$

We recall the notation

$$w_{N;j}^{i_1,\dots,i_{|T|}} = (w_{i_l,j;N})_{l=1}^{|T|},$$

so that

$$\begin{aligned}\|w_{N;j}^{i_1,\dots,i_{|T|}}\|_{\ell^1} &\le \min\Big(|T|(\max_{i,j}|w_{i,j;N}|), \max\Big(\max_j \sum_i |w_{i,j;N}|, \max_i \sum_j |w_{i,j;N}|\Big)\Big) \\ &= c(w, |T|).\end{aligned}$$

By Lemma 3.5, since the marginals are non-negative,

$$\begin{aligned}&|\Lambda^{\otimes n} \star (\mathcal{R}_{N,T,m}\eta^{\otimes n})(z)| \\ &\quad\le \frac{1}{N}\sum_{i_1,\dots,i_{|T|}=1}^N |w_{N,T}(i_1,\dots,i_{|T|})| \\ &\qquad\qquad \times \big|\Lambda^{\otimes n} \star \big((f_N^{i_1,\dots,i_{|T|}}(\cdot - w_{N;i_m}^{i_1,\dots,i_{|T|}}) - f_N^{i_1,\dots,i_{|T|}}(\cdot))\eta^{\otimes n}\big)(z)\big| \\ &\quad\le \frac{1}{N}\sum_{i_1,\dots,i_{|T|}=1}^N |w_{N,T}(i_1,\dots,i_{|T|})|[\exp((1+\alpha)c(w,|T|)) - 1] \\ &\qquad\qquad \times [\Lambda^{\otimes n} \star (f_N^{i_1,\dots,i_{|T|}}\eta^{\otimes n})(z)] \\ &\quad= [\exp((1+\alpha)c(w,|T|)) - 1][\Lambda^{\otimes n} \star (|\tau_N|(T)\eta^{\otimes n})(z)].\end{aligned}$$

Then

$$\begin{aligned}\|\mathcal{R}_{N,T,m}\|^2_{H_\eta^{-1\otimes|T|}} &\le [\exp((1+\alpha)c(w,|T|)) - 1] \\ &\qquad \times \int_{\mathbb{R}^{|T|}} |(\mathcal{R}_{N,T,m}\eta^{\otimes n})(z)|[\Lambda^{\otimes n} \star (|\tau_N|(T)\eta^{\otimes n})(z)]\,\mathrm{d}z\end{aligned}$$

$$\leq [\exp((1+\alpha)c(w,|T|))-1] \times \int_{\mathbb{R}^{|T|}} [\Lambda^{\otimes n} \star |\mathcal{R}_{N,T,m}\eta^{\otimes n}|(z)][(|\tau_N|(T)\eta^{\otimes n})(z)]\,\mathrm{d}z.$$

We hence need also to bound $\Lambda^{\otimes n} \star |\mathcal{R}_{N,T,m}\eta^{\otimes n}|$ with the absolute value inside but

$$\begin{aligned}
&(\Lambda^{\otimes n} \star |\mathcal{R}_{N,T,m}\eta^{\otimes n}|)(z)\\
&= \frac{1}{N}\sum_{i_1,\dots,i_{|T|}=1}^{N} |w_{N,T}(i_1,\dots,i_{|T|})|\\
&\qquad \times \big[\Lambda^{\otimes n} \star \big((f_N^{i_1,\dots,i_{|T|}}(\cdot - w_{N;i_m}^{i_1,\dots,i_{|T|}}) + f_N^{i_1,\dots,i_{|T|}}(\cdot))\eta^{\otimes n}\big)(z)\big]\\
&= 2\Lambda^{\otimes n} \star (|\tau_N|(T)\eta^{\otimes n})(z)\\
&\quad + \frac{1}{N}\sum_{i_1,\dots,i_{|T|}=1}^{N} |w_{N,T}(i_1,\dots,i_{|T|})|\\
&\qquad \times \big[\Lambda^{\otimes n} \star \big((f_N^{i_1,\dots,i_{|T|}}(\cdot - w_{N;i_m}^{i_1,\dots,i_{|T|}}) - f_N^{i_1,\dots,i_{|T|}}(\cdot))\eta^{\otimes n}\big)(z)\big].
\end{aligned}$$

Hence again by Lemma 3.5,

$$\begin{aligned}
&(\Lambda^{\otimes n} \star |\mathcal{R}_{N,T,m}\eta^{\otimes n}|)(z)\\
&\quad \leq 2\,\Lambda^{\otimes n} \star (|\tau_N|(T)\eta^{\otimes n})(z)\\
&\qquad + \frac{1}{N}\sum_{i_1,\dots,i_{|T|}=1}^{N} |w_{N,T}(i_1,\dots,i_{|T|})|[\exp((1+\alpha)c(w,|T|))-1]\\
&\qquad\qquad \times [\Lambda^{\otimes n} \star (f_N^{i_1,\dots,i_{|T|}}\eta^{\otimes n})(z)]\\
&\quad = [\exp((1+\alpha)c(w,|T|))+1][\Lambda^{\otimes n} \star (|\tau_N|(T)\eta^{\otimes n})(z)].
\end{aligned}$$

In conclusion,

$$\begin{aligned}
\|\mathcal{R}_{N,T,m}\|^2_{H_\eta^{-1\otimes|T|}} &\leq [\exp((1+\alpha)c(w,|T|))^2-1]\\
&\quad \times \int_{\mathbb{R}^{|T|}} [\Lambda^{\otimes n} \star (|\tau_N|(T)\eta^{\otimes n})(z)][(|\tau_N|(T)\eta^{\otimes n})(z)]\,\mathrm{d}z\\
&= [\exp((2+2\alpha)c(w,|T|))-1]\||\tau_N|(T)\|^2_{H_\eta^{-1\otimes|T|}}. \qquad \blacksquare
\end{aligned}$$

3.5. Bounding the firing rate through exponential moments

We present here another set of technical results which show how to handle the weight function in our subsequent commutator estimates.

Lemma 3.7. *Consider the weight function $\eta = \eta_\alpha$ and any signed measure $f \in \mathcal{M}(\mathbb{R})$. The following estimate holds:*

$$\left|\int_{\mathbb{R}} K \star (\nu f)\,\mathrm{d}x\right| = \left|\int_{\mathbb{R}} \nu f\,\mathrm{d}x\right| \leq C(\alpha)\|\nu\|_{W^{1,\infty}}\|f\|_{H_\eta^{-1}},$$

where $C(\alpha)$ only depends on $\alpha > 0$.

Proof. Only the inequality in the statement is not trivial. Choose now a non-negative, smooth function φ with compact support $\operatorname{supp}\varphi \subset [-1,1]$, such that $\varphi_i = \varphi(\cdot - i)$, $i \in \mathbb{N}$ form a partition of unity of $\mathbb{R}$ in the usual sense that

$$\sum_{i=-\infty}^{\infty} \varphi(x-i) \equiv 1, \quad \forall x \in \mathbb{R}.$$

It is easy to verify that

$$\int_{\mathbb{R}} \varphi \, \mathrm{d}x = 1.$$

Then

$$\begin{aligned}\left|\int_{\mathbb{R}} \nu f \, \mathrm{d}x\right| = \left|\int_{\mathbb{R}} (\nu/\eta) f \eta \, \mathrm{d}x\right| &\leq \sum_{i=-\infty}^{\infty} \left|\int_{\mathbb{R}} (\nu/\eta) f \eta \varphi_i \, \mathrm{d}x\right| \\ &= \sum_{i=-\infty}^{\infty} \left|\int_{\mathbb{R}} \varphi \star ((\nu/\eta) f \eta \varphi_i) \, \mathrm{d}x\right| \\ &\leq C \sum_{i=-\infty}^{\infty} \left(\int_{\mathbb{R}} |\varphi \star ((\nu/\eta) f \eta \varphi_i)|^2 \, \mathrm{d}x\right)^{\frac{1}{2}},\end{aligned}$$

where in the last line we use that each integrand is supported in $[-2+i, 2+i]$.

From the smoothness of φ, its Fourier transform can be bounded by

$$\hat{\varphi}(\xi) \leq \frac{C}{\sqrt{1+4\pi^2\xi^2}} = C\,\widehat{K}(\xi).$$

Hence we further have from Lemma 3.1,

$$\begin{aligned}\left|\int_{\mathbb{R}} \nu f \, \mathrm{d}x\right| &\leq C \sum_{i=-\infty}^{\infty} \left(\int_{\mathbb{R}} |K \star ((\nu/\eta) f \eta \varphi_i)|^2 \, \mathrm{d}x\right)^{\frac{1}{2}}, \\ &\leq C \sum_{i=-\infty}^{\infty} \|(\nu/\eta)\varphi_i\|_{W^{1,\infty}} \left(\int_{\mathbb{R}} |K \star (f\eta)|^2 \, \mathrm{d}x\right)^{\frac{1}{2}} \\ &\leq C \left(\sum_{i=-\infty}^{\infty} \|\varphi_i/\eta\|_{W^{1,\infty}}\right) \|\nu\|_{W^{1,\infty}} \|f\eta\|_{H^{-1}},\end{aligned}$$

where the constant C is some universal constant which may change line by line.

Since each φ_i is a translation of φ and has support in $[-1+i, 1+i]$, it is easy to check the uniform bound

$$\sum_{i=-\infty}^{\infty} \|\varphi_i/\eta\|_{W^{1,\infty}} \leq C(1+\alpha) \sum_{i=-\infty}^{\infty} \exp(-\alpha|i|) < \infty,$$

where the constant only depends on the particular choice of φ, which concludes the proof. ∎

This lemma also admits the following tensorization.

Lemma 3.8. *For $f \in \mathcal{M}(\mathbb{R}^k) \cap H_\eta^{-1\otimes k}$,*

$$\int_{\mathbb{R}^{k-1}} \left(\int_{\mathbb{R}} K^{\otimes k} \star ((\nu_m/\eta_m) f \eta^{\otimes k})(t,z)\, \mathrm{d}z_m \right)^2 \prod_{n \neq m} \mathrm{d}z_n$$
$$\leq C(\alpha)^2 \|\nu\|_{W^{1,\infty}}^2 \|f\|_{H_\eta^{-1\otimes k}}^2,$$

where we recall the notation $\nu_m = \nu(z_m)$ and $\eta_m = \eta(z_m)$.

Proof. Without any loss of generality, we may assume $m = k$ and define

$$\begin{aligned} g(z) &= [K^{\otimes(k-1)} \star_{1,\dots,(k-1)} (f \eta^{\otimes(k-1)})](z) \\ &= \int_{\mathbb{R}^{k-1}} \prod_{n=1}^{k-1} K(u_n - z_n) f(u_1, \dots, u_{k-1}, z_k) \prod_{n=1}^{k-1} \eta(u_n)\, \mathrm{d}u_n. \end{aligned}$$

Then, from the previous lemma,

$$\begin{aligned} &\int_{\mathbb{R}^{k-1}} \left(\int_{\mathbb{R}} K^{\otimes k} \star ((\nu_k/\eta_k) f \eta^{\otimes k})(t,z)\, \mathrm{d}z_k \right)^2 \prod_{n=1}^{k-1} \mathrm{d}z_n \\ &\quad = \int_{\mathbb{R}^{k-1}} \left(\int_{\mathbb{R}} \nu(z_k) g(t,z)\, \mathrm{d}z_k \right)^2 \prod_{n=1}^{k-1} \mathrm{d}z_n \\ &\quad \leq \int_{\mathbb{R}^{k-1}} C(\alpha)^2 \|\nu\|_{W^{1,\infty}}^2 \int_{\mathbb{R}} ([K \star_k (g\eta_k)](t, z_1, \dots, z_k))^2\, \mathrm{d}z_k \prod_{n=1}^{k-1} \mathrm{d}z_n \\ &\quad = C(\alpha)^2 \|\nu\|_{W^{1,\infty}}^2 \int_{\mathbb{R}^k} ([K^{\otimes k} \star (f \eta^{\otimes k})](t, z_1, \dots, z_k))^2 \prod_{n=1}^{k} \mathrm{d}z_n, \end{aligned}$$

which concludes. ∎

4. The limiting observables from the Vlasov equation

This section is centered on the limiting observables $\tau_\infty(T, w, f)$, $T \in \mathcal{T}$. We first show that Definition 1.6 is still correct when the kernel and extended density are merely $w \in \mathcal{W}$ and $f \in L^\infty([0, t_*] \times [0, 1]; \mathcal{M}_+(\mathbb{R}))$. We also prove Proposition 1.8, which shows the compactness can be attained not only at the level of weak-* topology of each limiting observable τ_∞, $T \in \mathcal{T}$, but also directly at the level of w and f.

Contrary to the rest of the paper, this section owes much to the technical framework developed in [55], which extends to our setting.

4.1. Revisiting the definition of limiting observables

The main kernel space $\mathcal{W}$ in this article was defined in (1.5), and is restated here:

$$\begin{aligned}\mathcal{W} := \big\{w \in \mathcal{M}([0,1]^2) : w(\xi, \mathrm{d}\zeta) \in L^\infty_\xi([0,1],\ \mathcal{M}_\zeta[0,1]),\\ w(\mathrm{d}\xi, \zeta) \in L^\infty_\zeta([0,1],\ \mathcal{M}_\xi[0,1])\big\},\\ \|w\|_{\mathcal{W}} := \max\big\{\|w\|_{L^\infty_\xi \mathcal{M}_\zeta}, \|w\|_{L^\infty_\zeta \mathcal{M}_\xi}\big\}.\end{aligned}$$

The motivation for introducing this Banach space in its current form lies in its capability to act as an $L^p \to L^p$ mapping.

Lemma 4.1. *Consider the following bounded linear operator:*

$$\begin{aligned}\mathcal{W} \times C([0,1]; B) &\to L^\infty([0,1]; B),\\ (w, \phi) &\mapsto \int_{[0,1]} \phi(\zeta) w(\cdot, \mathrm{d}\zeta),\end{aligned}$$

where B stands for any Banach space such as $L^p(\mathbb{R})$. Then this operator can be uniquely extended to $\mathcal{W} \times L^\infty([0,1]; B) \to L^\infty([0,1]; B)$ with

$$\Big\| \int_{[0,1]} \phi(\zeta) w(\cdot, \mathrm{d}\zeta) \Big\|_{L^p([0,1];B)} \le \|w\|_{\mathcal{W}} \|\phi\|_{L^p([0,1];B)}.$$

Proof. The cases for $p = 1$ and $p = \infty$ can be checked through a careful but straightforward density argument, for which we refer to [55, Lemma 3.8]. Extending the result to $1 < p < \infty$ is an application of a textbook result of interpolation between Banach spaces, which can be found in [6] for instance. ∎

The integrals appearing in Definition 1.6 can then be made rigorous by sequentially consider the integrations as operations $L^p \to L^p$. To assist such an argument, we again follow [55] and introduce the following countable algebra, which, as we will see later, contains all the necessary information to reproduce the limiting observables $\tau_\infty(T, w, f)$, $T \in \mathcal{T}$.

Definition 4.2 (A countable algebra). We denote by $\mathcal{T}$ the countable algebra of transforms over spaces of arbitrarily large dimensions which is built as follows: For each transform $F \in \mathcal{T}$ there exists $k \in \mathbb{N}$ (called the rank of F) so that F maps each couple (w, f) into a signed measure $F(w, f) \in L^\infty([0,1]; \mathcal{M}(\mathbb{R}^k))$. The full algebra $\mathcal{T}$ is obtained in a recursive way according to the following three rules:

(i) (Seed). The elementary 1-rank transform $F_0 \colon (w, f) \mapsto f$ belongs to the algebra $\mathcal{T}$.

(ii) (Graft). Let $F_1 \in \mathcal{T}$ and $F_2 \in \mathcal{T}$ be k_1-rank and k_2-rank transforms respectively. Then the following $(k_1 + k_2)$-rank transform $(F_1 \otimes F_2)$ also belongs to $\mathcal{T}$:

$$\begin{aligned}&(F_1 \otimes F_2)(w, f):\\ &(\xi, z_1, \ldots, z_{k_1+k_2}) \mapsto F_1(w, f)(\xi, z_1, \ldots, z_{k_1}) F_2(w, f)(\xi, z_{k_1+1}, \ldots, z_{k_1+k_2}).\end{aligned}$$

(iii) (Grow). Let $F \in \mathcal{T}$ be a k-rank transform. Then the following k-rank transform F^* also belongs to $\mathcal{T}$:

$$F^*(w, f):\\ (\xi, z_1, \ldots, z_k) \mapsto \int_{[0,1]} F(w, f)(\zeta, z_1, \ldots, z_k) w(\xi, \mathrm{d}\zeta).$$

The following lemma shows that the transforms of the countable algebra $\mathcal{T}$ are well defined on $\mathcal{W}$.

Lemma 4.3. *Consider any kernel $w \in \mathcal{W}$ and extended density $f \in L^\infty([0, 1]; H_\eta^{-1} \cap \mathcal{M}_+(\mathbb{R}))$. Then for each $F \in \mathcal{T}$, the signed measure $F(w, f)$ is well defined and belongs to $L^\infty([0, 1]; H_\eta^{-1\otimes k} \cap \mathcal{M}_+(\mathbb{R}^k))$ for some $k \in \mathbb{N}$. Moreover, as $n \to \infty$,*

$$F(w^{(n)}, f^{(n)}) \to F(w, f) \quad \text{in } L^2([0, 1]; H_\eta^{-1\otimes k})$$

for any fixed $F \in \mathcal{T}$, any sequence $\{f^{(n)}\}_{n=1}^\infty$ uniformly bounded in $L^\infty([0, 1]; H_\eta^{-1} \cap \mathcal{M}_+(\mathbb{R}))$, and any sequence $\{w^{(n)}\}_{n=1}^\infty$ uniformly bounded in $\mathcal{W}$, satisfying

$$\begin{aligned} f^{(n)} &\to f && \text{in } L^\infty([0, 1]; H_\eta^{-1}(\mathbb{R})),\\ w^{(n)}(\xi, \zeta) &\to w(\xi, \zeta) && \text{in } L_\xi^2 H_\zeta^{-1} \cap L_\zeta^2 H_\xi^{-1}. \end{aligned}$$

We note that since $\zeta \in [0, 1]$, we have that $L_\xi^2 H_\zeta^{-1} \subset L_\xi^\infty \mathcal{M}_\zeta$ with compact embedding.

Proof of Lemma lem:tree_algebra. We use an induction argument based on the recursive rules in Definition 4.2.

(i) The seed element $F_0(w, f) = f$ is well defined and belongs to $L^\infty([0, 1]; H_\eta^{-1} \cap \mathcal{M}_+(\mathbb{R}))$.

(ii) Consider two elements $F_1(w, f)$ and $F_2(w, f)$ that are well defined and satisfy

$$F_i(w, f) \in L^\infty([0, 1]; H_\eta^{-1\otimes k_i} \cap \mathcal{M}(\mathbb{R}^{k_i})), \quad i = 1, 2.$$

Because both norms are stable under tensorization, for the combined element we have

$$\begin{aligned} &\|(F_1 \otimes F_2)(w, f)\|_{L^\infty([0,1];B_1\otimes B_2)}\\ &\qquad \le \|F_1(w, f)\|_{L^\infty([0,1];B_1)} \|F_2(w, f)\|_{L^\infty([0,1];B_2)}, \end{aligned}$$

where we may choose either $B_1 = H_\eta^{-1\otimes k_1}$, $B_2 = H_\eta^{-1\otimes k_2}$, $B_1 \otimes B_2 = H_\eta^{-1\otimes(k_1+k_2)}$, or $B_1 = \mathcal{M}(\mathbb{R}^{k_1})$, $B_2 = \mathcal{M}(\mathbb{R}^{k_2})$, $B_1 \otimes B_2 = \mathcal{M}(\mathbb{R}^{k_1+k_2})$. Hence,

$$(F_1 \otimes F_2)(w, f) \in L^\infty([0, 1]; H_\eta^{-1\otimes(k_1+k_2)} \cap \mathcal{M}(\mathbb{R}^{k_1+k_2})).$$

(iii) Consider an element $F(w, f)$ that is well defined and satisfies

$$F(w, f) \in L^\infty([0, 1]; H_\eta^{-1\otimes k} \cap \mathcal{M}(\mathbb{R}^k)).$$

Applying Lemma 4.1 with $p = \infty$ with either $B = H_\eta^{-1\otimes k}$ or $B = \mathcal{M}(\mathbb{R}^k)$, for the grow element we have

$$\|F^*(w, f)\|_{L^\infty([0,1];B)} \le \|w\|_{\mathcal{W}} \|F(w, f)\|_{L^\infty([0,1];B)}.$$

Hence,

$$F^*(w, f) \in L^\infty([0, 1]; H_\eta^{-1\otimes k} \cap \mathcal{M}_+(\mathbb{R}^k)).$$

Since $\mathcal{T}$ is generated by the three rules in Definition 4.2, the above argument shows that any $F(w, f)$, $F \in \mathcal{T}$ is well defined.

We can use a similar argument to prove the convergence $F(w^{(n)}, f^{(n)}) \to F(w, f)$ for any fixed $F \in \mathcal{T}$.

(i) For the seed sequence, $F_0(w^{(n)}, f^{(n)}) = f^{(n)} \to f$ in $L^\infty([0, 1]; H_\eta^{-1}(\mathbb{R}))$, hence the convergence also holds in $L^2([0, 1]; H_\eta^{-1}(\mathbb{R}))$.

(ii) Consider the two sequences $F_1(w^{(n)}, f^{(n)})$ and $F_2(w^{(n)}, f^{(n)})$ satisfying

$$F_i(w^{(n)}, f^{(n)}) \to F_i(w, f) \quad \text{in } L^2([0, 1]; H_\eta^{-1\otimes k_i}),\ i = 1, 2.$$

Then by introducing the intermediary element $F_1(w^{(n)}, f^{(n)}) \otimes_z F_2(w, f)$ and by applying the triangular inequality, we have

$$\begin{aligned}
&\|(F_1 \otimes F_2)(w^{(n)}, f^{(n)}) - (F_1 \otimes F_2)(w, f)\|_{L^2([0,1];H_\eta^{-1\otimes(k_1+k_2)})} \\
&\quad \le \|F_1(w^{(n)}, f^{(n)})\|_{L^\infty([0,1];H_\eta^{-1\otimes k_1})} \|F_2(w^{(n)}, f^{(n)}) - F_2(w, f)\|_{L^2([0,1];H_\eta^{-1\otimes k_2})} \\
&\qquad + \|F_1(w^{(n)}, f^{(n)}) - F_1(w, f)\|_{L^2([0,1];H_\eta^{-1\otimes k_1})} \|F_2(w, f)\|_{L^\infty([0,1];H_\eta^{-1\otimes k_2})}.
\end{aligned}$$

As $n \to \infty$, we immediately have

$$(F_1 \otimes F_2)(w^{(n)}, f^{(n)}) \to (F_1 \otimes F_2)(w, f) \quad \text{in } L^2([0, 1]; H_\eta^{-1\otimes(k_1+k_2)}).$$

(iii) (Grow). Consider a sequence $F(w^{(n)}, f^{(n)})$ satisfying

$$F(w^{(n)}, f^{(n)}) \to F(w, f) \quad \text{in } L^2([0, 1]; H_\eta^{-1\otimes k}).$$

The difference between the grow sequences is given by

$$\begin{aligned}
&\|F^*(w^{(n)}, f^{(n)}) - F^*(w, f)\|_{L^2([0,1];H_\eta^{-1\otimes k})} \\
&\quad = \left\| \int_{[0,1]} F(w^{(n)}, f^{(n)})(\zeta, \cdot) w^{(n)}(\xi, d\zeta) - \int_{[0,1]} F(w, f)(\zeta, \cdot) w(\xi, d\zeta) \right\|_{L^2([0,1];H_\eta^{-1\otimes k})}.
\end{aligned}$$

Introduce any $\phi_\varepsilon \in H^1([0, 1]; H_\eta^{-1\otimes k})$ approximating $F(w, f)$ in $L^2([0, 1]; H_\eta^{-1\otimes k})$ and the intermediary elements

$$\int_{[0,1]} \phi_\varepsilon(\zeta, \cdot) w^{(n)}(\xi, d\zeta), \quad \int_{[0,1]} \phi_\varepsilon(\zeta, \cdot) w(\xi, d\zeta).$$

Then, applying the triangular inequality and Lemma 4.1 with $p = 2$, $B = H_\eta^{-1\otimes k}$, we have

$$\begin{aligned}\|F^*(w^{(n)}, f^{(n)}) - F^*(w, f)\|_{L^2([0,1];H_\eta^{-1\otimes k})} \\ \le \|w^{(n)}\|_{\mathcal{W}} \|F(w^{(n)}, f^{(n)}) - \phi_\varepsilon\|_{L^2([0,1];H_\eta^{-1\otimes k})} \\ + \|w\|_{\mathcal{W}} \|F(w, f) - \phi_\varepsilon\|_{L^2([0,1];H_\eta^{-1\otimes k})} \\ + \|w^{(n)} - w\|_{L^2H^{-1}} \|\phi_\varepsilon\|_{H^1([0,1];H_\eta^{-1\otimes k})}.\end{aligned}$$

Letting $n \to \infty$ and $\varepsilon \to 0$, we conclude that

$$F^*(w^{(n)}, f^{(n)}) \to F^*(w, f) \quad \text{in } L^2([0, 1]; H_\eta^{-1\otimes k}).$$

■

The following lemma shows that it is possible to recover the limiting observables $\tau_\infty(T, w, f)$, $T \in \mathcal{T}$ from $F(w, f)$, $F \in \mathfrak{T}$.

Lemma 4.4. *For any tree $T \in \mathcal{T}$, there exists a transform $F \in \mathfrak{T}$ such that*

$$\begin{aligned}&F(w, f)(\zeta, z_1, \dots, z_{|T|}) \\ &\quad = \left(\int_{[0,1]^{|T|-1}} w_T(\xi_1, \dots, \xi_{|T|}) \prod_{m=1}^{|T|} f(\xi_m, z_m)\, \mathrm{d}\xi_2 \dots \mathrm{d}\xi_{|T|}\right)\Bigg|_{\xi_1=\zeta},\end{aligned}$$

where the variable ξ_1 corresponding to the root of T is not integrated. As a consequence,

$$\tau_\infty(T, w, f) = \int_{[0,1]} F(w, f)(\zeta, z_1, \dots, z_{|T|})\, \mathrm{d}\zeta.$$

Proof of Lemma 4.4. For the tree $T_1 \in \mathcal{T}$ with only one node, the corresponding transform in $\mathfrak{T}$ is the seed element F_0. It is easy to verify that

$$F_0(w, f)(\zeta, z_1) = f(\zeta, z_1), \quad \tau_\infty(T_0, w, f)(z_1) = \int_{[0,1]} f(\zeta, z_1)\, \mathrm{d}\zeta.$$

For any tree $T \in \mathcal{T}$ with more than one node, let $i_1, \dots, i_k \in \{2, \dots, |T|\}$ be all the nodes that are directly connected to the root 1, and let $T_1, \dots, T_k$ be the subtrees of T taking $i_1, \dots, i_k$ as their roots. Suppose by induction that we have found corresponding transforms $F_1, \dots, F_k$ for $T_1, \dots, T_k$. Then

$$\begin{aligned}&\int_{[0,1]^{|T|-1}} w_T(\xi_1, \dots, \xi_{|T|}) \prod_{m=1}^{|T|} f(\xi_m, z_m)\, \mathrm{d}\xi_2, \dots, \mathrm{d}\xi_{|T|} \\ &= f(\xi_1, z_1) \prod_{l=1}^{k} \left(\int_{[0,1]^{|T_l|}} w(\xi_1, \xi_{i_l}) \prod_{(j,j')\in\mathcal{E}(T_l)} w(\xi_j, \xi_{j'}) \prod_{m\in T_l} f(\xi_m, z_m)\, \mathrm{d}\xi_m\right) \\ &= f(\xi_1, z_1) \prod_{l=1}^{k} \int_{[0,1]} F_l(w, f)(\xi_{i_l}, z_{i_l}, \dots) w(\xi_1, \mathrm{d}\xi_{i_l}) = F(w, f)(\xi_1, z_1, \dots, z_{|T|}).\end{aligned}$$

Up to an index permutation (so that if $i \in T_l$, $i' \in T_{l'}$, $l < l'$, then $i < i'$), it can be reformulated into the more straightforward form

$$F(w, f) = \Big[F_0 \otimes \bigotimes_{l=1}^{k} (F_l)^* \Big](w, f),$$

showing that F is obtained by making each F_l grow (by rule (iii)) with depth 1, then grafting (by rule (ii)) all of them together with another seed element F_0 (by rule (i)). ∎

4.2. Compactness of the limiting observables

We now turn to the proof of Proposition 1.8.

Proof of Proposition 1.8. To prove (1.13), let us define $S^m(N) := \{1, \dots, N\}^m$ and

$$S^m_{\text{diag}}(N) := \big\{ (i_1, \dots, i_m) \in \{1, \dots, N\}^m : \exists j \neq k \text{ s.t. } i_j = i_k \big\}.$$

Recall that

$$\widetilde{w}_N(\xi, \zeta) = \sum_{i,j=1}^{N} N w_{i,j;N} \mathbb{1}_{[\frac{i-1}{N}, \frac{i}{N})}(\xi) \mathbb{1}_{[\frac{j-1}{N}, \frac{j}{N})}(\zeta),$$

$$\tilde{f}_N(x, \xi) = \sum_{i=1}^{N} f_N^i(x) \mathbb{1}_{[\frac{i-1}{N}, \frac{i}{N})}(\xi).$$

Since we have independence, it is straightforward that

$$\begin{aligned} \tau_\infty(T, \widetilde{w}_N, \tilde{f}_N)(\mathrm{d}z) &= \int_{[0,1]^{|T|}} \prod_{(l,l') \in \mathcal{E}(T)} \widetilde{w}_N(\xi_l, \xi_{l'}) \prod_{m=1}^{|T|} \tilde{f}_N(t, \xi_m, \mathrm{d}z_m)\, \mathrm{d}\xi_1, \dots, \mathrm{d}\xi_{|T|} \\ &= \frac{1}{N} \sum_{\substack{(i_1, \dots, i_{|T|}) \in \\ S^{|T|}(N)}} \prod_{(l,l') \in \mathcal{E}(T)} w_{i_l, i_{l'};N} \prod_{m=1}^{|T|} f_N^{i_m}(\mathrm{d}z_m). \end{aligned}$$

On the other hand, again from independence,

$$\tau_N(T, w_N, f_N)(\mathrm{d}z) = \frac{1}{N} \sum_{\substack{(i_1, \dots, i_{|T|}) \in \\ S^{|T|}(N) \setminus S^{|T|}_{\text{diag}}(N)}} \prod_{(l,l') \in \mathcal{E}(T)} w_{i_l, i_{l'};N} \prod_{m=1}^{|T|} f_N^{i_m}(\mathrm{d}z_m),$$

where the terms involving repeated indexes are excluded from the summation, contrary to the case of τ_∞.

Therefore the difference is controlled by

$$\tau_\infty(T,\widetilde{w}_N,\tilde{f}_N)(\mathrm{d}z)-\tau_N(T,w_N,f_N)(\mathrm{d}z) = \frac{1}{N}\sum_{\substack{(i_1,\dots,i_{|T|})\in\\ S^{|T|}_{\mathrm{diag}}(N)}}\prod_{(l,l')\in\mathcal{E}(T)} w_{i_l,i_{l'};N}\prod_{m=1}^{|T|} f_N^{i_m}(\mathrm{d}z_m),$$

whose (weighted) total variation norm is bounded by

$$\begin{aligned}&\int_{\mathbb{R}^{|T|}}\exp\Big(a\sum_{m=1}^{|T|}|z_m|\Big)|\tau_\infty(T,\widetilde{w}_N,\tilde{f}_N)(\mathrm{d}z)-\tau_N(T,w_N,f_N)(\mathrm{d}z)|\\ &\quad=\int_{\mathbb{R}^{|T|}}\exp\Big(a\sum_{m=1}^{|T|}|z_m|\Big)\Big|\frac{1}{N}\sum_{\substack{(i_1,\dots,i_{|T|})\in\\ S^{|T|}_{\mathrm{diag}}(N)}}\prod_{(l,l')\in\mathcal{E}(T)} w_{i_l,i_{l'};N}\prod_{m=1}^{|T|} f_N^{i_m}(\mathrm{d}z_m)\Big|\\ &\quad\le\frac{1}{N}\sum_{\substack{(i_1,\dots,i_{|T|})\in\\ S^{|T|}_{\mathrm{diag}}(N)}}\prod_{(l,l')\in\mathcal{E}(T)} |w_{i_l,i_{l'};N}|\int_{\mathbb{R}^{|T|}}\prod_{m=1}^{|T|}\exp(a|z_m|)f_N^{i_m}(\mathrm{d}z_m).\end{aligned}$$

When $|T|=1$, this term is zero as $S^{|T|}_{\mathrm{diag}}(N)=\varnothing$, while for $|T|\ge 2$ we use the following lemma.

Lemma 4.5. *The following bound holds:*

$$\begin{aligned}&\frac{1}{N}\sum_{\substack{(i_1,\dots,i_{|T|})\in\\ S^{|T|}_{\mathrm{diag}}(N)}}\prod_{(l,l')\in\mathcal{E}(T)} |w_{i_l,i_{l'};N}|\\ &\quad\le\max_{1\le i,j\le N}|w_{i,j;N}|\max\Big(\max_i\sum_j|w_{i,j;N}|,\max_j\sum_i|w_{i,j;N}|\Big)^{|T|-2}|T|^2.\end{aligned}$$

Once we prove Lemma 4.5, we immediately obtain (1.13),

$$\begin{aligned}&\int_{\mathbb{R}^{|T|}}\exp\Big(a\sum_{m=1}^{|T|}|z_m|\Big)|\tau_\infty(T,\widetilde{w}_N,\tilde{f}_N)(\mathrm{d}z)-\tau_N(T,w_N,f_N)(\mathrm{d}z)|\\ &\quad\le\max_{1\le i,j\le N}|w_{i,j;N}|\max\Big(\max_i\sum_j|w_{i,j;N}|,\max_j\sum_i|w_{i,j;N}|\Big)^{|T|-2}|T|^2M_a^{|T|}.\end{aligned}$$

Proof of Lemma 4.5. Let us consider

$$\sum_{\substack{(i_1,\dots,i_{|T|})\in\\ S^{|T|}(N)}}\mathbb{1}_{\{i_m=i_{m'}=i\}}\prod_{(l,l')\in\mathcal{E}(T)}|w_{i_l,i_{l'};N}|$$

for any $1 \le m, m' \le |T|$ and $1 \le i \le N$. We introduce the path P which is the set of indices n on the unique path connecting m and m'. We can immediately remove from the sum the indices not in P as before,

$$\begin{aligned}
&\sum_{\substack{(i_1,\dots,i_{|T|})\in\\ S^{|T|}(N)}} \mathbb{1}_{\{i_m=i_{m'}=i\}} \prod_{(l,l')\in\mathcal{E}(T)} |w_{i_l,i_{l'};N}| \\
&\quad\le \max\Big(\max_i \sum_j |w_{i,j;N}|, \max_j \sum_i |w_{i,j;N}|\Big)^{|T|-|P|} \\
&\qquad\times \sum_{\substack{(i_{n_1},\dots,i_{n_{|P|}})\in\\ S^{|P|}(N)}} \mathbb{1}_{\{i_m=i_{m'}=i\}} \prod_{(l,l')\in\mathcal{E}(P)} |w_{i_l,i_{l'};N}|,
\end{aligned}$$

where we denote $P = \{n_1, \dots, n_{|P|}\}$ with $n_1 = m$ and $n_{|P|} = m'$.

The path P connecting m and m' naturally goes up in the tree first (to reach the parent vertex that is shared by m and m') and then down. Denote by k the number of indices for which the path goes up (with possibly $k = 1$ if m is a parent of m') and write

$$\begin{aligned}
&\sum_{\substack{(i_{n_1},\dots,i_{n_{|P|}})\in\\ S^{|P|}(N)}} \mathbb{1}_{\{i_m=i_{m'}=i\}} \prod_{(l,l')\in\mathcal{E}(P)} |w_{i_l,i_{l'};N}| \\
&\quad= \sum_{1\le j_1,\dots,j_{|P|}\le N} \mathbb{1}_{\{j_1=j_{|P|}=i\}} \prod_{n=1}^{k-1} |w_{j_{n+1},j_n;N}| \prod_{n=k}^{|P|-1} |w_{j_n,j_{n+1};N}| \\
&\quad= \max_{1\le i,j\le N} |w_{i,j;N}| \max\Big(\max_i \sum_j |w_{i,j;N}|, \max_j \sum_i |w_{i,j;N}|\Big)^{|P|-2}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\sum_{\substack{(i_1,\dots,i_{|T|})\in\\ S^m(N)}} \mathbb{1}\{i_m = i_{m'} = i\} \prod_{(l,l')\in\mathcal{E}(T)} |w_{i_l,i_{l'};N}| \\
&\quad\le \max_{1\le i,j\le N} |w_{i,j;N}| \max\Big(\max_i \sum_j |w_{i,j;N}|, \max_j \sum_i |w_{i,j;N}|\Big)^{|T|-2}.
\end{aligned}$$

As a consequence,

$$\begin{aligned}
&\frac{1}{N} \sum_{\substack{(i_1,\dots,i_{|T|})\in\\ S^m_{\mathrm{diag}}(N)}} \prod_{(l,l')\in\mathcal{E}(T)} |w_{i_l,i_{l'};N}| \\
&\quad\le \frac{1}{N} \sum_{i=1}^N \sum_{1\le m,m'\le|T|} \sum_{\substack{(i_1,\dots,i_{|T|})\in\\ S^m(N)}} \mathbb{1}_{\{i_m=i_{m'}=i\}} \prod_{(l,l')\in\mathcal{E}(T)} |w_{i_l,i_{l'};N}|
\end{aligned}$$

$$\leq \max_{1\leq i,j\leq N} |w_{i,j;N}| \max\Big(\max_i \sum_j |w_{i,j;N}|, \max_j \sum_i |w_{i,j;N}|\Big)^{|T|-2} |T|^2,$$

which concludes the proof. ■

It remains to prove (1.14), for which we first invoke [55, Corollary 4.9].

Lemma 4.6 ([55, Corollary 4.9]). *Consider any sequence g_n in $L^\infty([0,1])$. Then there exists $\Phi\colon [0,1] \to [0,1]$, a.e. injective, measure preserving, such that the following estimate is verified:*

$$\int_{[0,1]} |(g_n \circ \Phi)(\xi) - (g_n \circ \Phi)(\xi + h)| \, d\xi \leq 2^n \|g_n\|_{L^\infty} 2^{-C\sqrt{\log \frac{1}{|h|}}}$$

for any $n \in \mathbb{N}$, $0 < |h| < 1$, and some universal constant C.

This lemma tells us that, at the cost of a measure-preserving rearrangement, a minimum regularity of L^∞ functions on $[0,1]$ can be obtained.

In order to apply Lemma 4.6, we need to first check the stability of the algebra $F(w, f)$, $F \in \mathfrak{T}$ under measure-preserving rearrangements.

Lemma 4.7. *Consider any $w \in \mathcal{W}$ and $f \in L^\infty([0,1]; \mathcal{M}_+(\mathbb{R}))$ and any a.e. injective, measure-preserving $\Phi\colon [0,1] \to [0,1]$. Define the push-forward kernel and measure*

$$w_\#(\xi, d\zeta) := \Phi_\#^{-1} w(\Phi(\xi), \cdot)(d\zeta), \qquad f_\#(\xi, dz) := f(\Phi(\xi), dz),$$

where Φ^{-1} is any a.e. defined left inverse of Φ. Then the algebra $F(w, f)$, $F \in \mathfrak{T}$ is stable under Φ in the sense that

$$F(w_\#, f_\#)(\xi, dz_1, \dots, dz_k) = F(w, f)(\Phi(\xi), dz_1, \dots, dz_k)$$

for any transform $F \in \mathfrak{T}$ and for a.e. $\xi \in [0,1]$. Moreover, $\tau_\infty(T, w_\#, f_\#) = \tau_\infty(T, w, f)$ for any $T \in \mathcal{T}$.

Proof. The proof is again done by an induction argument based on the recursive rules defining $F(w, f)$, $F \in \mathfrak{T}$.

(i) For the seed element $F_0(w, f)$ the property is obvious.

(ii) Consider two elements $F_1(w, f)$ and $F_2(w, f)$ stable under Φ. Then the grafted element satisfies

$$\begin{aligned}
&(F_1 \otimes F_2)(w_\#, f_\#)(\xi, dz_1, \dots, dz_{k_1+k_2}) \\
&\quad = F_1(w_\#, f_\#)(\xi, dz_1, \dots, dz_{k_1}) F_2(w_\#, f_\#)(\xi, dz_{k_1+1}, \dots, dz_{k_1+k_2}) \\
&\quad = F_1(w, f)(\Phi(\xi), dz_1, \dots, dz_{k_1}) F_2(w, f)(\Phi(\xi), dz_{k_1+1}, \dots, dz_{k_1+k_2}) \\
&\quad = F(w, f)(\Phi(\xi), dz_1, \dots, dz_{k_1+k_2}),
\end{aligned}$$

which is the stated stability under Φ.

(iii) Consider an element $F(w, f)$ stable under Φ. Then the grow element satisfies the stability property

$$\begin{aligned} F^*(w_\#, f_\#)(\xi, \mathrm{d}z_1, \ldots, \mathrm{d}z_k) &= \int_{\zeta \in [0,1]} F(w_\#, f_\#)(\zeta, \mathrm{d}z_1, \ldots, \mathrm{d}z_k) w_\#(\xi, \mathrm{d}\zeta) \\ &= \int_{\zeta \in [0,1]} F(w, f)(\Phi(\zeta), \mathrm{d}z_1, \ldots, \mathrm{d}z_k) w_\#(\xi, \mathrm{d}\zeta) \\ &= \int_{\zeta \in [0,1]} F(w, f)(\zeta, \mathrm{d}z_1, \ldots, \mathrm{d}z_k) w(\Phi(\xi), \mathrm{d}\zeta) \\ &= F^*(w, f)(\Phi(\xi), \mathrm{d}z_1, \ldots, \mathrm{d}z_k). \end{aligned}$$

Finally, for any $T \in \mathcal{T}$, take $F \in \mathfrak{T}$ as claimed in Lemma 4.4. Then

$$\begin{aligned} \tau_\infty(T, w_\#, f_\#)(\mathrm{d}z_1, \ldots, \mathrm{d}z_{|T|}) &= \int_{\xi \in [0,1]} F(w_\#, f_\#)(\xi, \mathrm{d}z_1, \ldots, \mathrm{d}z_{|T|})\, \mathrm{d}\xi \\ &= \int_{\xi \in [0,1]} F(w, f)(\xi, \mathrm{d}z_1, \ldots, \mathrm{d}z_{|T|})\, \mathrm{d}\xi \\ &= \tau_\infty(T, w, f)(\mathrm{d}z_1, \ldots, \mathrm{d}z_{|T|}), \end{aligned}$$

which finishes the proof. ∎

The next step is to derive the compactness of the algebra $F(w, f)$, $F \in \mathfrak{T}$ and to identify the limit, which we summarize here.

Lemma 4.8. *Under the assumptions of Proposition 1.8, there exist measure-preserving maps $\Phi_N : [0,1] \to [0,1]$ for the sequence of $N \to \infty$ and $w \in \mathcal{W}$, $f \in L^\infty([0,1]; \mathcal{M}_+(\mathbb{R}))$, such that convergence in the following strong-weak-* sense holds: For all $F \in \mathfrak{T}$ and all $\varphi \in C_c(\mathbb{R}^k)$, where k is the rank of F,*

$$\begin{aligned} \lim_{N \to \infty} \int_{z \in \mathbb{R}^k} & \varphi(z_1, \ldots, z_k) F(\widetilde{w}_N, \tilde{f}_N)(\Phi_N(\xi), \mathrm{d}z_1, \ldots, \mathrm{d}z_k) \\ &= \int_{z \in \mathbb{R}^k} \varphi(z_1, \ldots, z_k) F(w, f)(\xi, \mathrm{d}z_1, \ldots, \mathrm{d}z_k) \end{aligned} \tag{4.1}$$

in any $L^p_\xi([0,1])$, $1 \le p < \infty$.

Proof. Since the algebra is countable, we may index the elements as $\mathfrak{T} = \{F_m : m \in \mathbb{N}\}$. For each $m \in \mathbb{N}$, let k_m be the rank of F_m and let $\{\varphi_{m,l}\}_{l \in \mathbb{N}}$ be any countable dense set of $C_c(\mathbb{R}^{k_m})$. Define the functions

$$g^N_{m,l}(\xi) := \int_{z \in \mathbb{R}^k} \varphi_{m,l}(z_1, \ldots, z_{k_m}) F_m(\widetilde{w}_N, \tilde{f}_N)(\xi, \mathrm{d}z_1, \ldots, \mathrm{d}z_{k_m}), \quad \forall m, l, N.$$

It is straightforward that $\sup_N \|g_{m,l}\|_{L^\infty([0,1])} < \infty$ from the bounds on $F_m(\widetilde{w}_N, \tilde{f}_N)$ in the space $L^\infty([0,1]; \mathcal{M}(\mathbb{R}^{k_m}))$ that follow from Lemma 1.3 and the identification provided by Lemma 4.4.

Thus, by Lemma 4.6, there exists $\Phi_N\colon[0,1]\to[0,1]$ for the sequence $N\to\infty$, so that the rearrangements

$$\begin{aligned}\tilde{g}^N_{m,l}(\xi) &= (g^N_{m,l}\circ\Phi_N)(\xi)\\ &= \int_{z\in\mathbb{R}^k}\varphi_{m,l}(z_1,\dots,z_{k_m})F_m(\widetilde{w}_N,\tilde{f}_N)(\Phi_N(\xi),\mathrm{d}z_1,\dots,\mathrm{d}z_{k_m})\end{aligned}$$

fulfill the estimates

$$\int_{[0,1]}|\tilde{g}^N_{m,l}(\xi)-\tilde{g}^N_{m,l}(\xi+h)|\,\mathrm{d}\xi\le C_{m,l}2^{-C\sqrt{\log\frac{1}{|h|}}},\quad \forall 0<|h|<1$$

for some universal constant $C>0$ and $C_{m,l}>0$ depending on the two indexes only.

By the Fréchet–Kolmogorov theorem and using a diagonal extraction there exists some subsequence of N (which we still denote by N for simplicity) and for all $m,l\in\mathbb{N}$, there exists $\tilde{g}_{m,l}\in L^\infty([0,1])$ such that as $N\to\infty$,

$$\tilde{g}^N_{m,l}\to\tilde{g}_{m,l}\quad\text{in any } L^p([0,1]),\ 1\le p<\infty.$$

Let us define, for any N in the subsequence, any $F\in\mathcal{T}$ and $\varphi\in C_c(\mathbb{R}^k)$, where k is the rank of F,

$$\begin{aligned}\tilde{g}^N_{F,\varphi} &:= \int_{z\in\mathbb{R}^k}\varphi(z_1,\dots,z_k)F(\widetilde{w}_N,\tilde{f}_N)(\Phi_N(\xi),\mathrm{d}z_1,\dots,\mathrm{d}z_k)\\ &= \int_{z\in\mathbb{R}^k}\varphi(z_1,\dots,z_k)F(\widetilde{w}_{N;\#},\tilde{f}_{N;\#})(\xi,\mathrm{d}z_1,\dots,\mathrm{d}z_k),\end{aligned}$$

where we again apply the following notation for the rearrangement:

$$\widetilde{w}_{N;\#}(\xi,\mathrm{d}\zeta):=\Phi_\#^{-1}\widetilde{w}_N(\Phi(\xi),\cdot)(\mathrm{d}\zeta),\qquad \tilde{f}_{N;\#}(\xi,\mathrm{d}z):=\tilde{f}_N(\Phi(\xi),\mathrm{d}z).$$

By a density argument of $C_c(\mathbb{R}^k)$, we conclude that for any $F\in\mathcal{T}$ and $\varphi\in C_c(\mathbb{R}^k)$, there exists $\tilde{g}_{F,\varphi}\in L^\infty([0,1])$ such that as $N\to\infty$,

$$\tilde{g}^N_{F,\varphi}\to\tilde{g}_{F,\varphi}\quad\text{in any } L^p([0,1]),\ 1\le p<\infty.$$

It remains to identify $w\in\mathcal{W}$ and $f\in L^\infty([0,1];\mathcal{M}_+(\mathbb{R}))$ for the limit. Recall that we have defined the kernel space $\mathcal{W}$ as

$$\begin{aligned}\mathcal{W}:=\big\{w\in\mathcal{M}([0,1]^2): w(\xi,\mathrm{d}\zeta)\in L^\infty_\xi([0,1],\mathcal{M}_\zeta[0,1]),\\ w(\mathrm{d}\xi,\zeta)\in L^\infty_\zeta([0,1],\mathcal{M}_\xi[0,1])\big\},\end{aligned}$$

where $L^\infty_\xi([0,1],\mathcal{M}_\zeta[0,1])$ denotes the topological dual of $L^1_\xi([0,1],C_\zeta[0,1])$.

Hence there exists a subsequence (which we still index by N) and $w\in\mathcal{W}$, $f\in L^\infty([0,1];\mathcal{M}_+(\mathbb{R}))$, such that

$$\widetilde{w}_{N;\#}\overset{*}{\rightharpoonup}w,\quad \tilde{f}_{N;\#}\overset{*}{\rightharpoonup}f.$$

By passing to the limit we can immediately obtain the exponential moment bound

$$\operatorname*{ess\,sup}_{\xi\in[0,1]} \int_{\mathbb{R}} \exp(a|x|) f(\xi, \mathrm{d}x) \le M_a.$$

Let us define, for any $F \in \mathcal{T}$ and $\varphi \in C_c(\mathbb{R}^k)$,

$$g_{F,\varphi}(\xi) := \int_{z\in\mathbb{R}^k} \varphi(z_1, \dots, z_k) F(w, f)(\xi, \mathrm{d}z_1, \dots, \mathrm{d}z_k).$$

It is straightforward that $g_{F,\varphi} \in L^\infty([0,1])$ and (4.1) can be simply restated as

$$\tilde{g}_{F,\varphi} = g_{F,\varphi}. \tag{4.2}$$

We apply another induction argument based on the recursive rules.

(i) For the seed element $F_0(w, f) = f$, it is straightforward that for any $\psi \in C([0,1])$, $\phi \in C_c(\mathbb{R})$,

$$\begin{aligned}\int_{[0,1]} \psi(\xi) g_{F_0,\varphi}(\xi)\,\mathrm{d}\xi &= \int_{[0,1]} \psi(\xi) \int_{z\in\mathbb{R}} \varphi(z) f(\xi, \mathrm{d}z)\,\mathrm{d}\xi \\ &\overset{*}{=} \lim_{N\to\infty} \int_{[0,1]} \psi(\xi) \int_{z\in\mathbb{R}} \varphi(z) \tilde{f}_{N;\#}(\xi, \mathrm{d}z)\,\mathrm{d}\xi \\ &= \int_{[0,1]} \psi(\xi) \tilde{g}_{F_0,\varphi}(\xi)\,\mathrm{d}\xi,\end{aligned}$$

where the equality $\overset{*}{=}$ is due to the weak-* convergence $\tilde{f}_{N;\#} \overset{*}{\rightharpoonup} f$. Hence, the identity (4.2) holds for F_0.

(ii) Consider two elements $F_1, F_2 \in \mathcal{T}$ satisfying (4.2). Then for any $\phi_1 \in C_c(\mathbb{R}^{k_1})$, $\phi_2 \in C_c(\mathbb{R}^{k_2})$,

$$\begin{aligned}&g_{(F_1\otimes F_2),(\varphi_1\otimes\varphi_2)}(\xi) \\ &\quad= \int_{z\in\mathbb{R}^{k_1+k_2}} (\varphi_1 \otimes \varphi_2)(z_1, \dots, z_{k_1+k_2})(F_1 \otimes F_2)(w, f)(\xi, \mathrm{d}z_1, \dots, \mathrm{d}z_{k_1+k_2}) \\ &\quad= g_{F_1,\varphi_1}(\xi) g_{F_2,\varphi_2}(\xi),\end{aligned}$$

hence $g_{(F_1\otimes F_2),(\varphi_1\otimes\varphi_2)} = g_{F_1,\varphi_1} g_{F_2,\varphi_2}$.

By a similar argument, $\tilde{g}^N_{(F_1\otimes F_2),(\varphi_1\otimes\varphi_2)} = \tilde{g}^N_{F_1,\varphi_1} \tilde{g}^N_{F_2,\varphi_2}$ for all N. Passing to the limit (in any L^p, $1 \le p < \infty$) as $N \to \infty$ we obtain $\tilde{g}_{(F_1\otimes F_2),(\varphi_1\otimes\varphi_2)} = \tilde{g}_{F_1,\varphi_1} \tilde{g}_{F_2,\varphi_2}$. Therefore, one can conclude

$$g_{(F_1\otimes F_2),(\varphi_1\otimes\varphi_2)} = \tilde{g}_{(F_1\otimes F_2),(\varphi_1\otimes\varphi_2)},$$

which is (4.2) for $F = (F_1 \otimes F_2)$ when $\varphi \in C_c(\mathbb{R}^{k_1+k_2})$ is in the tensorized form $\varphi = \varphi_1 \otimes \varphi_2$.

Finally, any $\varphi \in C_c(\mathbb{R}^{k_1+k_2})$ can be approximated by a sum of tensorized functions so that we derive (4.2) for $F = (F_1 \otimes F_2)$ with any arbitrary $\varphi \in C_c(\mathbb{R}^{k_1+k_2})$.

(iii) Consider an element $F \in \mathfrak{T}$ satisfying (4.2). Then for any $\psi \in C([0,1])$, $\phi \in C_c(\mathbb{R}^k)$,

$$\begin{aligned}
&\int_{[0,1]} \psi(\xi) g_{F^*,\varphi}(\xi)\,\mathrm{d}\xi \\
&\quad = \int_{[0,1]} \psi(\xi) \int_{z\in\mathbb{R}^k} \varphi(z_1,\dots,z_k) F^*(w,f)(\xi,\mathrm{d}z_1,\dots,\mathrm{d}z_k)\,\mathrm{d}\xi \\
&\quad = \int_{\xi\in[0,1]} \psi(\xi) \int_{z\in\mathbb{R}^k} \varphi(z_1,\dots,z_k) \int_{\zeta\in[0,1]} F(w,f)(\zeta,\mathrm{d}z_1,\dots,\mathrm{d}z_k) w(\mathrm{d}\xi,\zeta)\,\mathrm{d}\zeta \\
&\quad = \int_{\xi\in[0,1]} \psi(\xi) g_{F,\varphi}(\zeta) w(\mathrm{d}\xi,\zeta)\,\mathrm{d}\zeta.
\end{aligned}$$

By a similar argument, for all N,

$$\int_{[0,1]} \psi(\xi) \tilde{g}^N_{F^*,\varphi}(\xi)\,\mathrm{d}\xi = \int_{\xi\in[0,1]} \psi(\xi) \tilde{g}^N_{F,\varphi}(\zeta) \widetilde{w}_{N;\#}(\mathrm{d}\xi,\zeta)\,\mathrm{d}\zeta.$$

Next, by the convergence

$$\begin{aligned}
\psi(\xi)\tilde{g}^N_{F,\varphi}(\zeta) &\to \psi(\xi) g_{F,\varphi}(\zeta) && \text{in } L^1_\zeta([0,1], C_\xi[0,1]), \\
\widetilde{w}_{N;\#}(\mathrm{d}\xi,\zeta)\,\mathrm{d}\zeta &\overset{*}{\rightharpoonup} w(\mathrm{d}\xi,\zeta)\,\mathrm{d}\zeta && \text{in } L^\infty_\zeta([0,1], \mathcal{M}_\xi[0,1]),
\end{aligned}$$

we obtain that

$$\lim_{N\to\infty} \int_{\xi\in[0,1]} \psi(\xi) \tilde{g}^N_{F,\varphi}(\zeta) \widetilde{w}_{N;\#}(\mathrm{d}\xi,\zeta)\,\mathrm{d}\zeta = \int_{\xi\in[0,1]} \psi(\xi) g_{F,\varphi}(\zeta) w(\mathrm{d}\xi,\zeta)\,\mathrm{d}\zeta.$$

Hence $g_{F^*,\varphi} = \tilde{g}_{F^*,\varphi}$, which is (4.2) for F^*. ∎

We may now conclude the proof of Proposition 1.8. For any $T \in \mathcal{T}$, there exists $F \in \mathfrak{T}$ such that

$$\begin{aligned}
\tau_\infty(T, \widetilde{w}_{N;\#}, \tilde{f}_{N;\#}) &= \int_{[0,1]} F(\widetilde{w}_N, \tilde{f}_N)(\Phi_N(\xi), \mathrm{d}z_1,\dots,\mathrm{d}z_{|T|})\,\mathrm{d}\xi, \\
\tau_\infty(T, w, f) &= \int_{[0,1]} F(w,f)(\xi, \mathrm{d}z_1,\dots,\mathrm{d}z_{|T|})\,\mathrm{d}\xi.
\end{aligned}$$

For any $\varphi \in C_c(\mathbb{R}^{|T|})$, by Lemma 4.8,

$$\begin{aligned}
&\lim_{N\to\infty} \int_{z\in\mathbb{R}^{|T|}} \tau_\infty(T, \widetilde{w}_{N;\#}, \tilde{f}_{N;\#})(\mathrm{d}z_1,\dots,\mathrm{d}z_{|T|}) \\
&\quad = \lim_{N\to\infty} \int_{[0,1]} \int_{z\in\mathbb{R}^{|T|}} \varphi(z_1,\dots,z_{|T|}) F(\widetilde{w}_N, \tilde{f}_N)(\Phi_N(\xi), \mathrm{d}z_1,\dots,\mathrm{d}z_{|T|})\,\mathrm{d}\xi \\
&\quad = \int_{[0,1]} \int_{z\in\mathbb{R}^{|T|}} \varphi(z_1,\dots,z_{|T|}) F(w,f)(\xi, \mathrm{d}z_1,\dots,\mathrm{d}z_{|T|})\,\mathrm{d}\xi \\
&\quad = \int_{z\in\mathbb{R}^{|T|}} \tau_\infty(T, w, f)(\mathrm{d}z_1,\dots,\mathrm{d}z_{|T|}).
\end{aligned}$$

Since $\varphi \in C_c(\mathbb{R}^{|T|})$ is arbitrary we conclude (1.14), restated here:

$$\tau_\infty(T, \widetilde{w}_N, \tilde{f}_N) \overset{*}{\rightharpoonup} \tau_\infty(T, w, f) \in \mathcal{M}(\mathbb{R}^{|T|}), \quad \forall T \in \mathcal{T}.$$ ∎

5. Proofs of the quantitative results

5.1. The hierarchy of equations

This subsection provides the main proofs of Propositions 2.3, 2.4, and 2.5, which derive the hierarchy of equations from the Liouville equation (2.1) and the Vlasov equation (1.3)–(1.4).

We begin with the proof of Proposition 2.3, showing that the observables corresponding to the laws of $(X_0^{1;N},\dots,X_0^{N;N})$ solving (1.1) satisfy the extended BBGKY hierarchy (2.2)–(2.3).

Proof of Proposition 2.3. Since the coefficients are bounded Lipschitz, the well-posedness of the SDE system (1.1) and the Liouville-type equation (2.1) are classical results. For simplicity of the presentation, we avoid using weak formulations but only present a formal calculation.

Consider any distinct indexes $i_1,\dots,i_k \in \{1,\dots,N\}$. It is easy to verify the following identity deriving the marginal laws from the full joint law:

$$\begin{aligned} f_N^{i_1,\dots,i_k}(t,z_1,\dots,z_k) &:= \mathrm{Law}(X_t^{i_1;N},\dots,X_t^{i_k;N}) \\ &= \left(\int_{\mathbb{R}^{N-k}} f_N(t,x_1,\dots,x_N) \prod_{i\neq i_1,\dots,i_k} \mathrm{d}x_i\right)\Bigg|_{\forall l=1,\dots,k,\ x_{i_l}=z_l}. \end{aligned}$$

By integrating the Liouville equation (2.1) along spatial directions $i \notin \{i_1,\dots,i_k\}$ and calculate the summation $i \in \{i_1,\dots,i_k\}$ and $i \notin \{i_1,\dots,i_k\}$ separately, we obtain equations for the marginals:

$$\begin{aligned} &\partial_t f_N^{i_1,\dots,i_k}(t,z_1,\dots,z_k) \qquad (5.1)\\ &= \sum_{m=1}^k \Bigg\{\Bigg[-\partial_{z_m}(\mu(z_m) f_N^{i_1,\dots,i_k}(t,z)) + \frac{\sigma^2}{2}\partial_{z_m}^2 f_N^{i_1,\dots,i_k}(t,z) - \nu(z_m) f_N^{i_1,\dots,i_k}(t,z) \\ &\qquad + \delta_0(z_m)\left(\int_{\mathbb{R}} \nu(u_m) f_N^{i_1,\dots,i_k}(t,u-w_{N;i_m}^{i_1,\dots,i_k})\,\mathrm{d}u_m\right)\Bigg|_{\substack{\forall n\neq m\\ u_n=z_n}}\Bigg]\Bigg\} \\ &\quad + \sum_{i\neq i_1,\dots,i_k} \int_{\mathbb{R}} \nu(z_{k+1})\left(f_N^{i_1,\dots,i_k,i}(t,z-w_{N;i}^{i_1,\dots,i_k,i}) - f_N^{i_1,\dots,i_k,i}(t,z)\right)\mathrm{d}z_{k+1}. \end{aligned}$$

We can reformulate the last line as

$$\begin{aligned} &\sum_{i\neq i_1,\dots,i_k} \int_{\mathbb{R}} \nu(z_{k+1})(f_N^{i_1,\dots,i_k,i}(t,z-w_{N;i}^{i_1,\dots,i_k,i}) - f_N^{i_1,\dots,i_k,i}(t,z))\,\mathrm{d}z_{k+1} \\ &= \sum_{i\neq i_1,\dots,i_k} \int_{\mathbb{R}} \nu(z_{k+1})\left(\int_0^1 \sum_{m=1}^k -w_{i_m,i}\partial_{z_m} f_N^{i_1,\dots,i_k,i}(t,z-rw_{N;i}^{i_1,\dots,i_k,i})\,\mathrm{d}r\right)\mathrm{d}z_{k+1} \\ &= \sum_{m=1}^k -\partial_{z_m}\left[\sum_{i\neq i_1,\dots,i_k} w_{i_m,i;N}\int_{\mathbb{R}} \nu(z_{k+1})\left(\int_0^1 f_N^{i_1,\dots,i_k,i}(t,z-rw_{N;i}^{i_1,\dots,i_k,i})\,\mathrm{d}r\right)\mathrm{d}z_{k+1}\right], \end{aligned}$$

changing it into an additional advection term $\partial_{z_m}[\dots]$ to the equation.

Introduce the simple identity

$$f_N^{i_1,\dots,i_k}(u - w_{N;i_m}^{i_1,\dots,i_k}) = f_N^{i_1,\dots,i_k}(u) - \{f_N^{i_1,\dots,i_k}(u) - f_N^{i_1,\dots,i_k}(u - w_{N;i_m}^{i_1,\dots,i_k})\},$$

and proceed to do the same for $f_N^{i_1,\dots,i_k,i}(z - r w_{N;i}^{i_1,\dots,i_k,i})$, so that the marginal equations (5.1) now read

$$\begin{aligned}
&\partial_t f_N^{i_1,\dots,i_k}(z_1,\dots,z_k) \\
&= \sum_{m=1}^{k} \Bigg\{ \Bigg[-\partial_{z_m}(\mu(z_m) f_N^{i_1,\dots,i_k}(z)) + \frac{\sigma^2}{2}\partial_{z_m}^2 f_N^{i_1,\dots,i_k}(z) - \nu(z_m) f_N^{i_1,\dots,i_k}(z) \\
&\qquad + \delta_0(z_m)\Bigg(\int_{\mathbb{R}} \nu(u_m)\big(f_N^{i_1,\dots,i_k}(u) - \{f_N^{i_1,\dots,i_k}(u) \\
&\qquad\qquad - f_N^{i_1,\dots,i_k}(u - w_{N;i_m}^{i_1,\dots,i_k})\}\big)\,\mathrm{d}u_m \Bigg)\Bigg|_{\substack{\forall n\neq m\\ u_n=z_n}} \Bigg] \\
&\quad - \partial_{z_m}\Bigg[\sum_{i\neq i_1,\dots,i_k} w_{i_m,i;N} \\
&\qquad \times \int_{\mathbb{R}} \nu(z_{k+1}) \Bigg(\int_0^1 f_N^{i_1,\dots,i_k,i}(z) - \{f_N^{i_1,\dots,i_k,i}(z) \\
&\qquad\qquad - f_N^{i_1,\dots,i_k,i}(z - r w_{N;i}^{i_1,\dots,i_k,i})\}\,\mathrm{d}r \Bigg)\,\mathrm{d}z_{k+1} \Bigg]\Bigg\},
\end{aligned} \tag{5.2}$$

where we omit the variable t for simplicity.

By taking the time derivative of the definition of observables (1.2), restated here,

$$\tau_N(T, w_N, f_N)(t,z) := \frac{1}{N}\sum_{i_1,\dots,i_{|T|}=1}^{N} w_{N,T}(i_1,\dots,i_{|T|}) f_N^{i_1,\dots,i_{|T|}}(t, z_1,\dots,z_{|T|}),$$

and substituting the right-hand side $\partial_t f_N^{i_1,\dots,i_{|T|}}$ of the marginal equation (5.2) with $k = |T|$, we obtain

$$\begin{aligned}
&\partial_t\Bigg(\frac{1}{N}\sum_{i_1,\dots,i_{|T|}=1}^{N} w_{N,T}(i_1,\dots,i_{|T|}) f_N^{i_1,\dots,i_{|T|}}(z_1,\dots,z_{|T|}) \Bigg) \\
&\quad = \frac{1}{N}\sum_{i_1,\dots,i_{|T|}=1}^{N} w_{N,T}(i_1,\dots,i_{|T|}) \\
&\qquad \times \sum_{m=1}^{|T|} \Bigg\{ \Bigg[-\partial_{z_m}(\mu(z_m) f_N^{i_1,\dots,i_{|T|}}(z)) + \frac{\sigma^2}{2}\partial_{z_m}^2 f_N^{i_1,\dots,i_{|T|}}(z) - \nu(z_m) f_N^{i_1,\dots,i_{|T|}}(z) \\
&\qquad\quad + \delta_0(z_m)\Bigg(\int_{\mathbb{R}} \nu(u_m)\big(f_N^{i_1,\dots,i_{|T|}}(u) - \{f_N^{i_1,\dots,i_{|T|}}(u) \\
&\qquad\qquad - f_N^{i_1,\dots,i_{|T|}}(u - w_{N;i_m}^{i_1,\dots,i_{|T|}})\}\big)\mathrm{d}u_m \Bigg)\Bigg|_{\substack{\forall n\neq m\\ u_n=z_n}} \Bigg]
\end{aligned}$$

$$- \partial_{z_m} \Big[\sum_{i \neq i_1,\dots,i_k} w_{i_m,i;N} \int_{\mathbb{R}} \nu(z_{|T|+1}) \times \Big(\int_0^1 f_N^{i_1,\dots,i_{|T|},i}(z) \{ f_N^{i_1,\dots,i_{|T|},i}(z) - f_N^{i_1,\dots,i_{|T|},i}(z - r w_{N;i}^{i_1,\dots,i_{|T|},i}) \} \mathrm{d}r \Big) \, \mathrm{d}z_{|T|+1} \Big] \Big\}.$$

Noticing the identity $w_{N,T+j}(i_1,\dots,i_{|T|+1}) = w_{N,T}(i_1,\dots,i_{|T|}) w_{i_j,i_{|T|+1}}$, we see that all the marginals, except the two terms of form $\{f_N^{\cdots}(\cdot) - f_N^{\cdots}(\cdot - w)\}$, are expressed in the right way so they can be rewritten as observables, obtaining (2.2) as the approximate hierarchy and (2.3) as the explicit form of the remainders. ∎

We now turn to the proof of Proposition 2.4. It is worth noting that the main Gronwall estimate could also be written in the probabilistic language of Itô calculus. However, we prefer to keep an approach and notation similar to the rest of the proofs presented.

Proof of Proposition 2.4. To simplify the argument, we only present a formal calculation where the tensorized weight $\eta^{\otimes|T|}$ is directly used as the test function, while, strictly speaking, the valid test functions for distributional solutions should have compact support. Given that the remaining coefficients are bounded Lipschitz and all terms in the subsequent calculation are non-negative, passing the limit to justify the use of unbounded weight on the dual side poses no problems.

The weighted total variation $\| |\tau_N|(T)\eta^{\otimes|T|}\|_{\mathcal{M}(\mathbb{R}^{|T|})}$ can be decomposed as

$$\begin{aligned}
&\| |\tau|(T)(t,\cdot)\eta^{\otimes|T|}\|_{\mathcal{M}(\mathbb{R}^{|T|})} \\
&= \int_{\mathbb{R}^{|T|}} \frac{1}{N} \sum_{i_1,\dots,i_{|T|}=1}^{N} |w_{N,T}(i_1,\dots,i_{|T|})| f_N^{i_1,\dots,i_{|T|}}(t,z)\eta^{\otimes|T|}(z)\,\mathrm{d}z \\
&= \frac{1}{N} \sum_{i_1,\dots,i_{|T|}=1}^{N} |w_{N,T}(i_1,\dots,i_{|T|})| \int_{\mathbb{R}^{|T|}} f_N^{i_1,\dots,i_{|T|}}(t,z)\eta^{\otimes|T|}(z)\,\mathrm{d}z.
\end{aligned}$$

For any distinct indexes $i_1,\dots,i_k$, we have

$$\int_{\mathbb{R}^{|T|}} f_N^{i_1,\dots,i_{|T|}}(t,z)\eta^{\otimes|T|}(z)\,\mathrm{d}z = \int_{\mathbb{R}^N} f_N(t,x) \prod_{l=1}^{|T|} \eta(x_{i_l})\,\mathrm{d}x.$$

The forthcoming estimate is not exclusive to our specific choice $\eta = \eta_\alpha$, but for any weight function adhering to the form

$$\eta(x) = \exp(h(x)), \quad \forall x \in \mathbb{R}$$

such that $\|h'\|_{L^\infty}, \|h''\|_{L^\infty}$ are bounded and $h(0) \le h(x)$. Our choice of $\eta = \eta_\alpha$ is clearly included by choosing $h(x) = \sqrt{1+\alpha^2 x^2}$, resulting in $\|h'\|_{L^\infty} \le \alpha$ and $\|h''\|_{L^\infty} \le \alpha^2$. The following inequalities are immediate results by the chain rule and the fundamental theorem of calculus.

Lemma 5.1. *For any weight function of form $\eta(x) = \exp(h(x))$ such that $\|h'\|_{L^\infty}$, $\|h''\|_{L^\infty}$ are bounded and $h(0) \le h(x)$, one has that*

$$|\eta'/\eta|(x) \le \|h'\|_{L^\infty}, \quad |\eta''/\eta|(x) \le \|h''\|_{L^\infty} + \|h'\|_{L^\infty}^2,$$

and

$$\eta(x+y) - \eta(x) \le \|h'\|_{L^\infty}|y| \exp(\|h'\|_{L^\infty}|y|)\eta(x).$$

The last inequality can be extended to the tensorized case $\eta^{\otimes k}(x) = \prod_{l=1}^{k} \eta(x_{i_l})$ as

$$\eta^{\otimes k}(x+y) - \eta^{\otimes k}(x) \le \|h'\|_{L^\infty}\|y\|_{\ell^1} \exp(\|h'\|_{L^\infty}\|y\|_{\ell^1})\eta^{\otimes k}(x).$$

We are now ready to prove Proposition 2.4 under the more general assumption that $\eta(x) = \exp(h(x))$. Since f_N solves (2.1) in the distributional sense, it is easy to verify that

$$\begin{aligned}
&\int_{\mathbb{R}^N} f_N(t,x) \prod_{l=1}^{|T|} \eta(x_{i_l})\,\mathrm{d}x \\
&= \int_{\mathbb{R}^N} f_N(0,x) \prod_{l=1}^{|T|} \eta(x_{i_l})\,\mathrm{d}x \\
&\quad + \int_0^t \int_{\mathbb{R}^N} f_N(s,x) \Bigg[\sum_{m=1}^{|T|} \Big(\mu(x_{i_m})(\eta'/\eta)(x_{i_m}) + \frac{1}{2}\sigma^2 (\eta''/\eta)(x_{i_m}) \Big) \prod_{l=1}^{|T|} \eta(x_{i_l}) \\
&\qquad + \sum_{j=i_1,\ldots,i_{|T|}} \nu(x_j) \Big(\frac{\eta(0)}{\eta(x_j)} \prod_{l=1}^{|T|} \eta(x_{i_l} + w_{i_l,j;N}) - \prod_{l=1}^{|T|} \eta(x_{i_l}) \Big) \\
&\qquad + \sum_{j\neq i_1,\ldots,i_{|T|}} \nu(x_j) \Big(\prod_{l=1}^{|T|} \eta(x_{i_l} + w_{i_l,j;N}) - \prod_{l=1}^{|T|} \eta(x_{i_l}) \Big) \Bigg] \mathrm{d}x\,\mathrm{d}s.
\end{aligned}$$

By Lemma 5.1, we have

$$\begin{aligned}
&\int_{\mathbb{R}^N} f_N(t,x) \prod_{l=1}^{|T|} \eta(x_{i_l})\,\mathrm{d}x \\
&\le \int_{\mathbb{R}^N} f_N(0,x) \prod_{l=1}^{|T|} \eta(x_{i_l})\,\mathrm{d}x \\
&\quad + \int_0^t \int_{\mathbb{R}^N} f_N(s,x) \Bigg[\sum_{m=1}^{|T|} \Big(\|\mu\|_{L^\infty}\|h'\|_{L^\infty} + \frac{1}{2}\sigma^2 (\|h''\|_{L^\infty} + \|h'\|_{L^\infty}^2) \Big) \prod_{l=1}^{|T|} \eta(x_{i_l}) \\
&\quad + \sum_{j=1}^{N} \|\nu\|_{L^\infty}\|h'\|_{L^\infty} \sum_{m=1}^{|T|} |w_{i_m,j;N}| \exp\Big(\|h'\|_{L^\infty} \max_j \sum_i |w_{i,j;N}| \Big) \prod_{l=1}^{|T|} \eta(x_{i_l}) \Bigg] \mathrm{d}x\,\mathrm{d}s
\end{aligned}$$

$$
\begin{aligned}
&= \int_{\mathbb{R}^N} f_N(0,x) \prod_{l=1}^{|T|} \eta(x_{i_l})\,\mathrm{d}x + \Bigg[\sum_{m=1}^{|T|} \Big(\|\mu\|_{L^\infty} \|h'\|_{L^\infty} + \frac{1}{2}\sigma^2 (\|h''\|_{L^\infty} + \|h'\|_{L^\infty}^2) \Big) \\
&\quad + \sum_{j=1}^{N} \sum_{m=1}^{|T|} |w_{i_m,j;N}| \|\nu\|_{L^\infty} \|h'\|_{L^\infty} \exp\Big(\|h'\|_{L^\infty} \max_j \sum_i |w_{i,j;N}| \Big) \Bigg] \\
&\quad \times \int_0^t \int_{\mathbb{R}^N} f_N(s,x) \prod_{l=1}^{|T|} \eta(x_{i_l})\,\mathrm{d}x\,\mathrm{d}s,
\end{aligned}
$$

where the summations of $j = i_1, \dots, i_{|T|}$ and $j \neq i_1, \dots, i_{|T|}$ are combined by the simple fact that $h(0) \leq h(x_j)$, hence $\eta(0)/\eta(x_j) \leq 1$.

Furthermore, we have that

$$
\sum_{j=1}^{N} \sum_{m=1}^{|T|} |w_{i_m,j;N}| \leq |T| \max_i \sum_j |w_{i,j;N}|.
$$

Hence by choosing

$$
\begin{aligned}
C_{\mathbb{W}} &= \max\Big(\max_i \sum_j |w_{i,j;N}|, \max_j \sum_i |w_{i,j;N}| \Big), \\
A_\eta &= \Big(\|\mu\|_{L^\infty} \|h'\|_{L^\infty} + \frac{1}{2}\sigma^2 (\|h''\|_{L^\infty} + \|h'\|_{L^\infty}^2) \\
&\quad + \|\nu\|_{L^\infty} \|h'\|_{L^\infty} C_{\mathbb{W}} \exp(\|h'\|_{L^\infty} C_{\mathbb{W}}) \Big),
\end{aligned}
$$

we conclude that

$$
\begin{aligned}
&\int_{\mathbb{R}^N} f_N(t,x) \prod_{l=1}^{|T|} \eta(x_{i_l})\,\mathrm{d}x \\
&\quad \leq \int_{\mathbb{R}^N} f_N(0,x) \prod_{l=1}^{|T|} \eta(x_{i_l})\,\mathrm{d}x + \int_0^t |T| A_\eta \int_{\mathbb{R}^N} f_N(s,x) \prod_{l=1}^{|T|} \eta(x_{i_l})\,\mathrm{d}x\,\mathrm{d}s.
\end{aligned}
$$

By the Gronwall lemma, this implies that

$$
\int_{\mathbb{R}^N} f_N(t,x) \prod_{l=1}^{|T|} \eta(x_{i_l})\,\mathrm{d}x \leq \exp(|T| A_\eta t) \int_{\mathbb{R}^N} f_N(0,x) \prod_{l=1}^{|T|} \eta(x_{i_l})\,\mathrm{d}x.
$$

Taking the summation over $i_1, \dots, i_{|T|}$, we have

$$
\begin{aligned}
\| |\tau|(T) \eta^{\otimes |T|}(t,\cdot) \|_{\mathcal{M}(\mathbb{R}^{|T|})} &\leq \exp(|T| A_\eta t) \| |\tau|(T) \eta^{\otimes |T|}(0,\cdot) \|_{\mathcal{M}(\mathbb{R}^{|T|})} \\
&\leq C_\eta (M_\eta \exp(A_\eta t_*))^{|T|}
\end{aligned}
$$

Finally, by applying Lemma 3.2 to the left-hand side, we immediately obtain (2.4), restated here:

$$\||\tau_N|(T)(t,\cdot)\|_{H_\eta^{-1\otimes|T|}} \le C_\eta(T)(\|K\|_{L^2(\mathbb{R})}\exp(A_\eta t_*))^{|T|}$$

for all $T \in \mathcal{T}, t \in [0,t_*]$. ∎

Finally, we give the proof of Proposition 2.5.

Proof of Proposition 2.5. We show the well-posedness of the Vlasov equation (1.3)–(1.4) by a classical fixed point argument. Let us first define the mapping $f \mapsto \mathcal{L}f$ as the solution of

$$\partial_t \mathcal{L}f(t,\xi,x) + \partial_x(\mu_f^*(t,\xi,x)\mathcal{L}f(t,\xi,x)) - \frac{\sigma^2}{2}\partial_{xx}(\mathcal{L}f(t,\xi,x)) + \nu(x)\mathcal{L}f(t,\xi,x) - \delta_0(x)J_f(t,\xi) = 0.$$

If f is given, then J_f and μ_f^* are determined, making the above identity a linear equation with respect to $\mathcal{L}f$. We are going to see that if $f \in L^\infty([0,t_*]\times[0,1]; H_\eta^{-1}\cap\mathcal{M}_+(\mathbb{R}))$, then $\mathcal{L}f$ belongs to the same space.

By multiplying the equation by the weight function η and applying the Leibniz formula, we obtain

$$\begin{aligned}
&\partial_t \mathcal{L}f(t,\xi,x)\eta(x)\\
&\quad= -\partial_x(\mu_f^*(t,\xi,x)\mathcal{L}f(t,\xi,x)\eta(x)) + \frac{\sigma^2}{2}\partial_{xx}(\mathcal{L}f(t,\xi,x)\eta(x))\\
&\qquad - \nu(x)\mathcal{L}f(t,\xi,x)\eta(x)\\
&\qquad + \delta_0(x)\eta(0)J_f(t,\xi) + \mu_f^*(t,\xi,x)(\eta'/\eta)(x)\mathcal{L}f(t,\xi,x)\eta(x)\\
&\qquad + \frac{\sigma^2}{2}\big[-\partial_x(2(\eta'/\eta)(x)\mathcal{L}f(t,\xi,x)\eta(x)) + (\eta''/\eta)(x)\mathcal{L}f(t,\xi,x)\eta(x)\big].
\end{aligned}$$

We start the a priori estimate of the linear mapping $\mathcal{L}$ by the total mass. It is straightforward to verify that

$$\begin{aligned}
\|\mathcal{L}f(t,\xi,\cdot)\|_{\mathcal{M}(\mathbb{R})} &\le \|f(0,\xi,\cdot)\|_{\mathcal{M}(\mathbb{R})} + \int_0^t J_f(s,\xi)\,\mathrm{d}s\\
&\le \|f(0,\xi,\cdot)\|_{\mathcal{M}(\mathbb{R})} + \int_0^t \|\nu\|_{L^\infty}\|f(s,\xi,\cdot)\|_{\mathcal{M}(\mathbb{R})}\,\mathrm{d}s. \qquad (5.3)
\end{aligned}$$

Note that by choosing $t_1 = 1/(2\|\nu\|_{L^\infty})$, we have

$$\begin{aligned}
&\sup_{t\in[0,t_1]}\|f(t,\xi,\cdot)\|_{\mathcal{M}(\mathbb{R})} \le 2\|f(0,\xi,\cdot)\|_{\mathcal{M}(\mathbb{R})}\\
&\qquad \Rightarrow \sup_{t\in[0,t_1]}\|\mathcal{L}f(t,\xi,\cdot)\|_{\mathcal{M}(\mathbb{R})} \le 2\|f(0,\xi,\cdot)\|_{\mathcal{M}(\mathbb{R})}.
\end{aligned}$$

Next, consider the η-weighted total moment,

$$
\begin{aligned}
&\|\mathcal{L} f(t,\xi,\cdot)\eta\|_{\mathcal{M}(\mathbb{R})} \\
&\quad\le \|f(0,\xi,\cdot)\eta\|_{\mathcal{M}(\mathbb{R})} + \int_0^t \Big\{\eta(0) J_f(s,\xi) \\
&\qquad\qquad + \Big[\|\mu_f^*(s,\xi,\cdot)\|_{L^\infty}\|\eta'/\eta\|_{L^\infty} + \frac{\sigma^2}{2}\|\eta''/\eta\|_{L^\infty}\Big] \\
&\qquad\qquad \times \|\mathcal{L} f(s,\xi,\cdot)\eta\|_{\mathcal{M}(\mathbb{R})}\Big\}\,\mathrm{d}s \\
&\quad\le \|f(0,\xi,\cdot)\eta\|_{\mathcal{M}(\mathbb{R})} + \int_0^t \Big\{\eta(0)\|\nu\|_{L^\infty}\|f(s,\xi,\cdot)\|_{\mathcal{M}(\mathbb{R})} \\
&\qquad\qquad + \Big[(\|\mu\|_{L^\infty} + \|w\|_{\mathscr{W}}\|\nu\|_{L^\infty}\|f(s,\cdot,\cdot)\|_{L^\infty_\xi \mathcal{M}_x})\|\eta'/\eta\|_{L^\infty} \\
&\qquad\qquad\quad + \frac{\sigma^2}{2}\|\eta''/\eta\|_{L^\infty}\Big]\|\mathcal{L} f(s,\xi,\cdot)\eta\|_{\mathcal{M}(\mathbb{R})}\Big\}\,\mathrm{d}s.
\end{aligned}
$$

By taking the supremum over $\xi \in [0,1]$, we have, for $t \in [0,t_1]$,

$$
\begin{aligned}
&\|\mathcal{L} f(t,\cdot,\cdot)\eta\|_{L^\infty_\xi \mathcal{M}_x} \\
&\quad\le \|f(0,\cdot,\cdot)\eta\|_{L^\infty_\xi \mathcal{M}_x} + \int_0^t \eta(0)\|\nu\|_{L^\infty}\|f\|_{L^\infty_{t,\xi}\mathcal{M}_x} \\
&\qquad + \Big[(\|\mu\|_{L^\infty} + \|w\|_{\mathscr{W}}\|\nu\|_{L^\infty}\|f\|_{L^\infty_{t,\xi}\mathcal{M}_x})\|\eta'/\eta\|_{L^\infty} + \frac{\sigma^2}{2}\|\eta''/\eta\|_{L^\infty}\Big] \\
&\qquad \times \|\mathcal{L} f(s,\cdot,\cdot)\eta\|_{L^\infty_\xi \mathcal{M}_x}\,\mathrm{d}s \\
&\quad\le \Big(\|f(0,\cdot,\cdot)\eta\|_{L^\infty_\xi \mathcal{M}_x} + \int_0^t \eta(0)\|\nu\|_{L^\infty}\|f\|_{L^\infty_{t,\xi}\mathcal{M}_x}\,\mathrm{d}s\Big) \\
&\qquad \times \exp\Big(\Big[(\|\mu\|_{L^\infty} + \|w\|_{\mathscr{W}}\|\nu\|_{L^\infty}\|f\|_{L^\infty_{t,\xi}\mathcal{M}_x})\|\eta'/\eta\|_{L^\infty} + \frac{\sigma^2}{2}\|\eta''/\eta\|_{L^\infty}\Big]t\Big), \qquad (5.4)
\end{aligned}
$$

where the L^∞_t should be understood as the supremum over $t \in [0,t_1]$.

We construct the invariance set and show $\mathcal{L}$-contractivity on the set by the following procedure: For any $R > R_0 := \|f(0,\cdot,\cdot)\eta\|_{L^\infty_\xi \mathcal{M}_x}$, and any $t_* > 0$, denote

$$
E_{R;t} := \big\{f \in \mathcal{M}_+ : \sup_{s\in[0,t]}\|f(s,\cdot,\cdot)\eta\|_{L^\infty_\xi \mathcal{M}_x} < R\big\}.
$$

By taking a sufficiently small t_2, for example

$$
\begin{aligned}
t_2 \le \min\Bigg(&\frac{1}{2\|\nu\|_{L^\infty}}, \frac{R-R_0}{2\eta(0)\|\nu\|_{L^\infty}R}, \\
&\frac{\log\frac{2R}{R+R_0}}{(\|\mu\|_{L^\infty} + \|w\|_{\mathscr{W}}\|\nu\|_{L^\infty}R)\|\eta'/\eta\|_{L^\infty} + \frac{\sigma^2}{2}\|\eta''/\eta\|_{L^\infty}}\Bigg),
\end{aligned}
$$

we can make $E_{R;t_2}$ an invariance set, i.e. $\mathcal{L}(E_{R;t_2}) \subset E_{R;t_2}$.

To show that $f \mapsto \mathcal{L} f$ is contracting in the H_η^{-1}-sense, we consider the following energy estimate: Along each fiber $\xi \in [0,1]$,

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}\int_{\mathbb{R}}[\Lambda \star ((\mathcal{L}f-\mathcal{L}g)\eta)](\mathcal{L}f-\mathcal{L}g)\eta\,\mathrm{d}x\right)\\
&\quad= \int_{\mathbb{R}}[\Lambda \star ((\mathcal{L}f-\mathcal{L}g)\eta)]\partial_t(\mathcal{L}f-\mathcal{L}g)\eta\,\mathrm{d}x\\
&\quad= \int_{\mathbb{R}} -\frac{\sigma^2}{2}[\Lambda \star \partial_x((\mathcal{L}f-\mathcal{L}g)\eta)][\partial_x((\mathcal{L}f-\mathcal{L}g)\eta)]\\
&\qquad + [\Lambda \star \partial_x((\mathcal{L}f-\mathcal{L}g)\eta)]\\
&\qquad \times\left[\mu_f^*(\mathcal{L}f-\mathcal{L}g)\eta + (\mu_f^*-\mu_g^*)(\mathcal{L}g)\eta + \sigma^2(\eta'/\eta)(\mathcal{L}f-\mathcal{L}g)\eta\right]\\
&\qquad + [\Lambda \star ((\mathcal{L}f-\mathcal{L}g)\eta)]\Big[-\nu(\mathcal{L}f-\mathcal{L}g)\eta + \delta_0\eta(0)(J_f-J_g)\\
&\qquad\qquad + \mu_f^*(\eta'/\eta)(\mathcal{L}f-\mathcal{L}g)\eta + (\mu_f^*-\mu_g^*)(\eta'/\eta)(\mathcal{L}g)\eta\\
&\qquad\qquad + \frac{\sigma^2}{2}(\eta''/\eta)(\mathcal{L}f-\mathcal{L}g)\eta\Big]\,\mathrm{d}x.
\end{aligned}
$$

Applying the Cauchy–Schwarz inequality, we obtain

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\left(\int_{\mathbb{R}}[\Lambda \star ((\mathcal{L}f-\mathcal{L}g)\eta)](\mathcal{L}f-\mathcal{L}g)\eta\,\mathrm{d}x\right)\\
&\quad\le \frac{4}{\sigma^2}\|\mu_f^*(\mathcal{L}f-\mathcal{L}g)\|_{H_\eta^{-1}}^2 + \frac{4}{\sigma^2}\|(\mu_f^*-\mu_g^*)(\mathcal{L}g)\|_{H_\eta^{-1}}^2 + 4\|(\eta'/\eta)(\mathcal{L}f-\mathcal{L}g)\|_{H_\eta^{-1}}^2\\
&\qquad + \left(4+\frac{\sigma^2}{2}\right)\|(\mathcal{L}f-\mathcal{L}g)\|_{H_\eta^{-1}}^2 + \|\nu(\mathcal{L}f-\mathcal{L}g)\|_{H_\eta^{-1}}^2 + \|\delta_0(J_f-J_g)\|_{H_\eta^{-1}}^2\\
&\qquad + \|\mu_f^*(\eta'/\eta)(\mathcal{L}f-\mathcal{L}g)\|_{H_\eta^{-1}}^2 + \|(\mu_f^*-\mu_g^*)(\eta'/\eta)(\mathcal{L}g)\|_{H_\eta^{-1}}^2\\
&\qquad + \frac{\sigma^2}{2}\|(\eta''/\eta)(\mathcal{L}f-\mathcal{L}g)\|_{H_\eta^{-1}}^2.
\end{aligned}
$$

Applying Lemma 3.3, we further have

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\|(\mathcal{L}f-\mathcal{L}g)\|_{H_\eta^{-1}}^2\\
&\quad\le \Big(\frac{16}{\sigma^2}\|\mu_f^*\|_{W^{1,\infty}}^2 + 16\|\eta'/\eta\|_{W^{1,\infty}}^2 + \left(4+\frac{\sigma^2}{2}\right)\\
&\qquad + 4\|\nu\|_{W^{1,\infty}}^2 + 4\|\mu_f^*\|_{W^{1,\infty}}^2\|\eta'/\eta\|_{W^{1,\infty}}^2 + 2\sigma^2\|\eta''/\eta\|_{W^{1,\infty}}^2\Big)\|(\mathcal{L}f-\mathcal{L}g)\|_{H_\eta^{-1}}^2\\
&\qquad + \left(\frac{4}{\sigma^2}\|(\mathcal{L}g)\|_{H_\eta^{-1}}^2 + 4\|\eta'/\eta\|_{W^{1,\infty}}^2\|(\mathcal{L}g)\|_{H_\eta^{-1}}^2\right)|\mu_f^*-\mu_g^*|^2 + \|\delta_0\|_{H_\eta^{-1}}^2|J_f-J_g|^2.
\end{aligned}
$$

Now let us consider the integration over $\xi \in [0,1]$. Firstly, using that $w \in \mathcal{W}$ combined with classical interpolation,

$$\int_{[0,1]} |\mu_f^*(t,\xi,x) - \mu_g^*(t,\xi,x)|^2 \,\mathrm{d}\xi = \int_{[0,1]} \left(\int_{[0,1]} w(\xi,\zeta)(J_f(t,\zeta) - J_g(t,\zeta))\,\mathrm{d}\zeta \right)^2 \mathrm{d}\xi$$
$$\leq \|w\|_{\mathcal{W}}^2 \|J_f(t,\cdot) - J_g(t,\cdot)\|_{L^2_\xi}^2.$$

Secondly, by Lemma 3.7,

$$|J_f(t,\xi) - J_g(t,\xi)| \leq \left| \int_{\mathbb{R}} \nu(x)(f(t,\xi,x) - g(t,\xi,x))\,\mathrm{d}x \right|$$
$$\leq C(\alpha)\|\nu\|_{W^{1,\infty}} \|f(t,\cdot,\xi) - g(t,\cdot,\xi)\|_{H_\eta^{-1}}.$$

Hence, we have

$$\|J_f(t,\cdot) - J_g(t,\cdot)\|_{L^2_\xi}^2 = \int_{[0,1]} |J_f(t,\xi) - J_g(t,\xi)|^2 \,\mathrm{d}\xi$$
$$\leq C(\alpha)^2 \|\nu\|_{W^{1,\infty}}^2 \int_{[0,1]} \|f(t,\xi,\cdot) - g(t,\xi,\cdot)\|_{H_\eta^{-1}}^2 \,\mathrm{d}\xi$$
$$= C(\alpha)^2 \|\nu\|_{W^{1,\infty}}^2 \|f - g\|_{L^2_\xi (H_\eta^{-1})_x}^2.$$

Therefore, by integrating over $\xi \in [0,1]$,

$$\|(\mathcal{L}f - \mathcal{L}g)(t,\cdot,\cdot)\|_{L^2_\xi (H_\eta^{-1})_x}^2$$
$$\leq \int_0^t M_0 \|(\mathcal{L}f - \mathcal{L}g)(s,\cdot,\cdot)\|_{L^2_\xi (H_\eta^{-1})_x}^2 + M_1 \|(f-g)(s,\cdot,\cdot)\|_{L^2_\xi (H_\eta^{-1})_x}^2 \,\mathrm{d}s$$
$$\leq \exp(M_0 t) \int_0^t M_1 \|(f-g)(s,\cdot,\cdot)\|_{L^2_\xi (H_\eta^{-1})_x}^2 \,\mathrm{d}s, \tag{5.5}$$

where M_0, M_1 are required to satisfy

$$M_0 \geq \sup_{t\in[0,t_2]} \left(\frac{16}{\sigma^2} \|\mu_f^*\|_{L^\infty_\xi W_x^{1,\infty}}^2 + 16\|\eta'/\eta\|_{W^{1,\infty}}^2 + \left(4 + \frac{\sigma^2}{2}\right) \right.$$
$$\left. + 4\|\nu\|_{W^{1,\infty}}^2 + 4\|\mu_f^*\|_{L^\infty_\xi W_x^{1,\infty}}^2 \|\eta'/\eta\|_{W^{1,\infty}}^2 + 2\sigma^2 \|\eta''/\eta\|_{W^{1,\infty}}^2 \right),$$

$$M_1 \geq \sup_{t\in[0,t_2]} \left[\left(\frac{4}{\sigma^2} \|(\mathcal{L}g)\|_{L^\infty_\xi (H_\eta^{-1})_x}^2 + 4\|\eta'/\eta\|_{W^{1,\infty}}^2 \|(\mathcal{L}g)\|_{L^\infty_\xi (H_\eta^{-1})_x}^2 \right) \|w\|_{\mathcal{W}}^2 \right.$$
$$\left. + \|\delta_0\|_{H_\eta^{-1}}^2 \right] C(\alpha)^2 \|\nu\|_{W^{1,\infty}}^2.$$

In addition, by $w \in \mathcal{W}$ and Lemma 3.7, we can derive

$$\|\mu_f^*\|_{L^\infty_\xi W_x^{1,\infty}} \leq \|\mu\|_{W_x^{1,\infty}} + \sup_{\xi\in[0,1]} \left| \int_0^1 w(\xi,\zeta) J_f(t,\zeta)\,\mathrm{d}\zeta \right|$$
$$\leq \|\mu\|_{W_x^{1,\infty}} + \|w\|_{\mathcal{W}} \|J_f(t,\cdot)\|_{L^\infty}$$
$$\leq \|\mu\|_{W_x^{1,\infty}} + \|w\|_{\mathcal{W}} C(\alpha) \|\nu\|_{W^{1,\infty}} \|f\|_{L^\infty_\xi (H_\eta^{-1})_x}.$$

When $f, g, \mathcal{L}f, \mathcal{L}g \in E_{R;t_2}$, by Lemma 3.2, we have

$$\|f\|_{L^\infty_\xi(H^{-1}_\eta)_x} \le \frac{R}{2}, \quad \|(\mathcal{L}g)\|_{L^\infty_\xi(H^{-1}_\eta)_x} \le \frac{R}{2},$$

for $t \in [0, t_2]$.

Hence M_0, M_1 in (5.5) can be chosen such that they only depend on R and the regularity of the various fixed coefficients in the system. By choosing sufficiently small $t_* > 0$, for example

$$t_* \le \max\Big(t_2, \frac{1}{3M_1}, \frac{\log 2}{M_0}\Big),$$

by (5.5) we conclude that $\mathcal{L}$ is contracting on the set $\mathcal{L}(E_{R;t_*})$ for the $L^2_\xi(H^{-1}_\eta)_x$-norm. Repeating the argument allows us to extend the weak solution to any finite time interval as usual, since the a priori estimates (5.3) and (5.4) do not blow up in finite time.

We now turn to the derivation of the limiting hierarchy. Taking the derivative of $\tau_\infty(T) = \tau_\infty(T, w, f)$ in Definition 1.6, we first obtain

$$\begin{aligned}
&\partial_t \tau_\infty(T, w, f)(t, z)\\
&= \sum_{m=1}^{|T|} \Big[-\partial_{z_m}(\mu(z_m)\tau_\infty(T)(t,z)) + \frac{\sigma^2}{2}\partial^2_{z_m}\tau_\infty(T)(t,z)\\
&\qquad - \nu(z_m)\tau_\infty(T)(t,z) + \delta_0(z_m)\Big(\int_{\mathbb{R}} \nu(u_m)\tau_\infty(T)(t,u)\Big)\Big|_{\substack{\forall n\neq m\\ u_n = z_n}}\\
&\qquad - \partial_{z_m}\Big(\int_{[0,1]^{|T|}} w_T(\xi_1,\dots,\xi_{|T|}) f^{\otimes|T|}(t,\xi_1,z_1,\dots,\xi_{|T|},z_{|T|})\\
&\qquad\qquad \times \Big(\int_0^1 w(\xi_m,\xi_{|T|+1}) \int_{\mathbb{R}} \nu(z_{|T|+1}) f(t,\xi_{|T|+1},z_{|T|+1})\, \mathrm{d}z_{|T|+1}\mathrm{d}\xi_{|T|+1}\Big)\\
&\qquad\qquad \times \mathrm{d}\xi_1,\dots,\xi_{|T|}\Big)\Big].
\end{aligned}$$

The last term can be rewritten by using the observables with one more leaf, resulting in the limiting hierarchy (2.5), restated here:

$$\begin{aligned}
\partial_t \tau_\infty(T)(t,z) = \sum_{m=1}^{|T|} \Big\{ \Big[&-\partial_{z_m}(\mu(z_m)\tau_\infty(T)(t,z)) + \frac{\sigma^2}{2}\partial^2_{z_m}\tau_\infty(T)(t,z)\\
&- \nu(z_m)\tau_\infty(T)(t,z)\\
&+ \delta_0(z_m)\Big(\int_{\mathbb{R}} \nu(u_m)\tau_\infty(T)(t,u)\,\mathrm{d}u_m\Big)\Big|_{\substack{\forall n\neq m\\ u_n = z_n}}\Big]\\
&- \partial_{z_m}\Big[\int_{\mathbb{R}} \nu(z_{|T|+1})\tau_\infty(T+m)(t,z)\,\mathrm{d}z_{|T|+1}\Big]\Big\}. \qquad \blacksquare
\end{aligned}$$

5.2. Quantitative stability

This subsection focuses on the proof of the main quantitative estimate of the article. The technical Lemma 5.2 about recursive differential inequalities is given separately in the next subsection.

Proof of Theorem 2.6. For simplicity, let us recall the notation

$$\nu_m = 1 \otimes \cdots \otimes \nu \otimes \cdots \otimes,$$

where ν appears in the mth coordinate, i.e. $\nu_m(z) = \nu(z_m)$. The same convention applies to μ and η.

Define the difference $\Delta_N(T)(t,z) := \tau_N(T)(t,z) - \tau_\infty(T)(t,z)$. By subtracting (2.5) from (2.2), one has

$$\begin{aligned}
&\partial_t \Delta_N(T)(t,z) \\
&= \sum_{m=1}^{|T|} \bigg\{ \bigg[-\partial_{z_m}(\mu(z_m)\Delta_N(T)(t,z)) + \frac{\sigma^2}{2}\partial^2_{z_m}\Delta_N(T)(t,z) - \nu(z_m)\Delta_N(T)(t,z) \\
&\qquad + \delta_0(z_m)\bigg(\int_{\mathbb{R}} \nu(u_m)(\Delta_N(T)(t,u) + \mathcal{R}_{N,T,m}(t,u))\,\mathrm{d}u_m\bigg)\bigg|_{\substack{\forall n\neq m\\ u_n=z_n}} \bigg] \\
&\qquad - \partial_{z_m}\bigg[\int_{\mathbb{R}} \nu(z_{|T|+1})(\Delta_N(T+m)(t,z) + \widetilde{\mathcal{R}}_{N,T+m,|T|+1}(t,z))\,\mathrm{d}z_{|T|+1}\bigg]\bigg\},
\end{aligned}$$

for all $T \in \mathcal{T}$. We highlight that, for any fixed $N < \infty$, the above equalities and the later inequalities involving $\Delta_N(T)$ can be understood as recursive relations that hold on all $T \in \mathcal{T}$. At first glance, one may think that the approximate hierarchy (2.2) is only defined for observables $\tau_N(T)$ with $|T| \le N$. Nevertheless, by our formal definition that $f_N^{i_1,\dots,i_k} \equiv 0$ if there are duplicated indices among $i_1,\dots,i_k$, it is easy to verify that for any tree T such that $|T| > N$,

$$\tau_N(T, w_N, f_N)(t,z) := \frac{1}{N}\sum_{i_1,\dots,i_{|T|}=1}^{N} w_{N,T}(i_1,\dots,i_{|T|}) f_N^{i_1,\dots,i_{|T|}}(t,z_1,\dots,z_{|T|}) \equiv 0,$$

as in each marginal there must be duplicated indices. By a similar discussion, we see that $\mathcal{R}_{N,T,m} \equiv 0$ and $\widetilde{\mathcal{R}}_{N,T+m,|T|+1} \equiv 0$ when $|T| > N$. With these formal definitions, it is then straightforward to show that the approximate hierarchy (2.2) holds for all $T \in \mathcal{T}$.

By multiplying by the weight function $\eta^{\otimes|T|}$ and integrating, we obtain

$$\begin{aligned}
&(\partial_t \Delta_N(T)(t,z))\eta^{\otimes|T|}(z) \\
&= \sum_{m=1}^{|T|} \bigg\{ -\partial_{z_m}(\mu_m \Delta_N(T)\eta^{\otimes|T|})(t,z) + \frac{\sigma^2}{2}\partial^2_{z_m}(\Delta_N(T)\eta^{\otimes|T|})(t,z) \\
&\qquad - (\nu_m\Delta_N(T)\eta^{\otimes|T|})(t,z) + (\mu_m(\eta_m'/\eta_m)\Delta_N(T)\eta^{\otimes|T|})(t,z) \\
&\qquad + \delta_0(z_m)\eta(z_m)\bigg(\int_{\mathbb{R}} ((\nu_m/\eta_m)(\Delta_N(T) + \mathcal{R}_{N,T,m})\eta^{\otimes|T|})(t,u)\,\mathrm{d}u_m\bigg)\bigg|_{\substack{\forall n\neq m\\ u_n=z_n}}
\end{aligned}$$

$$- \partial_{z_m}\Big[\int_{\mathbb{R}} \big((\nu_{|T|+1}/\eta_{|T|+1})(\Delta_N(T+m) + \widetilde{\mathcal{R}}_{N,T+m,|T|+1})\eta^{\otimes|T|+1}\big)(t,z)\,\mathrm{d}z_{|T|+1}\Big] + \frac{\sigma^2}{2}[\partial_{z_m}(-2(\eta'_m/\eta_m)\Delta_N(T)\eta^{\otimes|T|}) + (\eta''_m/\eta_m)\Delta_N(T)\eta^{\otimes|T|}](t,z)\Big\}.$$

Substituting $(\partial_t\Delta_N(T))\eta^{\otimes|T|}$ into the right-hand side of

$$\frac{\mathrm{d}}{\mathrm{d}t}\bigg(\frac{1}{2}\int_{\mathbb{R}^{|T|}}\big(K^{\otimes|T|}\star(\Delta_N(T)\eta^{\otimes|T|})(t,z)\big)^2\,\mathrm{d}z\bigg) = \int_{\mathbb{R}^{|T|}}\big(K^{\otimes|T|}\star(\Delta_N(T)\eta^{\otimes|T|})(t,z)\big)\big(K^{\otimes|T|}\star(\partial_t\Delta_N(T)\eta^{\otimes|T|})(t,z)\big)\,\mathrm{d}z,$$

yields the extensive expression

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\bigg(\frac{1}{2}\int_{\mathbb{R}^{|T|}}\big(K^{\otimes|T|}\star(\Delta_N(T)\eta^{\otimes|T|})(t,z)\big)^2\,\mathrm{d}z\bigg)\\
&= \int_{\mathbb{R}^{|T|}}\sum_{m=1}^{|T|}\bigg\{-\frac{\sigma^2}{2}[\partial_{z_m}K^{\otimes|T|}\star(\Delta_N(T)\eta^{\otimes|T|})(t,z)]^2\\
&\quad + [K^{\otimes|T|}\star(\Delta_N(T)\eta^{\otimes|T|})(t,z)]\\
&\quad \times\bigg[-K^{\otimes|T|}\star(\nu_m\Delta_N(T)\eta^{\otimes|T|})(t,z)\\
&\qquad + K(z_m)\eta(0)\bigg(\int_{\mathbb{R}}K^{\otimes|T|}\star((\nu_m/\eta_m)\Delta_N(T)\eta^{\otimes|T|})(t,u)\,\mathrm{d}u_m\bigg)\bigg|_{\substack{\forall n\neq m\\ u_n=z_n}}\\
&\qquad + K(z_m)\eta(0)\bigg(\int_{\mathbb{R}}K^{\otimes|T|}\star((\nu_m/\eta_m)\mathcal{R}_{N,T,m}\eta^{\otimes|T|})(t,u)\,\mathrm{d}u_m\bigg)\bigg|_{\substack{\forall n\neq m\\ u_n=z_n}}\\
&\qquad + K^{\otimes|T|}\star(\mu_m(\eta'_m/\eta_m)\Delta_N(T)\eta^{\otimes|T|})(t,z)\\
&\qquad + \frac{\sigma^2}{2}K^{\otimes|T|}\star((\eta''_m/\eta_m)\Delta_N(T)\eta^{\otimes|T|})(t,z)\bigg]\\
&\quad + [\partial_{z_m}K^{\otimes|T|}\star(\Delta_N(T)\eta^{\otimes|T|})(t,z)]\\
&\quad \times\bigg[K^{\otimes|T|}\star(\mu_m\Delta_N(T)\eta^{\otimes|T|})(t,z)\\
&\qquad + \int_{\mathbb{R}}K^{\otimes|T|+1}\star((\nu_{|T|+1}/\eta_{|T|+1})\Delta_N(T+m)\eta^{\otimes|T|+1})(t,z)\,\mathrm{d}z_{|T|+1}\\
&\qquad + \int_{\mathbb{R}}K^{\otimes|T|+1}\star((\nu_{|T|+1}/\eta_{|T|+1})\widetilde{\mathcal{R}}_{N,T+m,|T|+1}\eta^{\otimes|T|+1})(t,z)\,\mathrm{d}z_{|T|+1}\\
&\qquad + \frac{\sigma^2}{2}K^{\otimes|T|}\star(2(\eta'_m/\eta_m)\Delta_N(T)\eta^{\otimes|T|})(t,z)\bigg]\bigg\}\,\mathrm{d}z.
\end{aligned}$$

We then apply the Cauchy–Schwarz inequality to obtain

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\Big(\frac{1}{2}\|\Delta_N(T)\|^2_{H_\eta^{-1\otimes|T|}}\Big)\\
&\quad\le \sum_{m=1}^{|T|}\bigg\{\Big(2+\frac{\sigma^2}{4}\Big)\|\Delta_N(T)\|^2_{H_\eta^{-1\otimes|T|}} + \frac{1}{2}\|\nu_m\Delta_N(T)\|^2_{H_\eta^{-1\otimes|T|}}\\
&\qquad + \frac{1}{2}\|\mu_m(\eta'_m/\eta_m)\Delta_N(T)\|^2_{H_\eta^{-1\otimes|T|}} + \frac{\sigma^2}{4}\|(\eta''_m/\eta_m)\Delta_N(T)\|^2_{H_\eta^{-1\otimes|T|}}\\
&\qquad + \frac{1}{2}\|K\|^2_{L^2}\eta(0)^2\int_{\mathbb{R}^{|T|-1}}\Big(\int_{\mathbb{R}}K^{\otimes|T|}\star((\nu_m/\eta_m)\Delta_N(T)\eta^{\otimes|T|})(t,z)\,\mathrm{d}z_m\Big)^2\prod_{n\neq m}\mathrm{d}z_n\\
&\qquad + \frac{1}{2}\|K\|^2_{L^2}\eta(0)^2\int_{\mathbb{R}^{|T|-1}}\Big(\int_{\mathbb{R}}K^{\otimes|T|}\star((\nu_m/\eta_m)\mathcal{R}_{N,T,m}\eta^{\otimes|T|})(t,z)\,\mathrm{d}z_m\Big)^2\prod_{n\neq m}\mathrm{d}z_n\\
&\qquad + \frac{2}{\sigma^2}\|\mu_m\Delta_N(T)\|^2_{H_\eta^{-1\otimes|T|}} + \frac{\sigma^2}{2}\|2(\eta'_m/\eta_m)\Delta_N(T)\|^2_{H_\eta^{-1\otimes|T|}}\\
&\qquad + \frac{2}{\sigma^2}\int_{\mathbb{R}^{|T|}}\Big(\int_{\mathbb{R}}K^{\otimes|T|+1}\star((\nu_{|T|+1}/\eta_{|T|+1})\Delta_N(T+m)\eta^{\otimes|T|+1})(t,z)\\
&\qquad\qquad\qquad \times\mathrm{d}z_{|T|+1}\Big)^2\prod_{n=1}^{|T|}\mathrm{d}z_n\\
&\qquad + \frac{2}{\sigma^2}\int_{\mathbb{R}^{|T|}}\Big(\int_{\mathbb{R}}K^{\otimes|T|+1}\star((\nu_{|T|+1}/\eta_{|T|+1})\widetilde{\mathcal{R}}_{N,T+m,|T|+1}\eta^{\otimes|T|+1})(t,z)\\
&\qquad\qquad\qquad \times\mathrm{d}z_{|T|+1}\Big)^2\prod_{n=1}^{|T|}\mathrm{d}z_n\bigg\}.
\end{aligned}\tag{5.6}$$

This is where the proper choice of weak distance becomes critical, as we need to bound the various terms on the right-hand side by the norm $\|\Delta_N(T)\|^2_{H_\eta^{-1\otimes|T|}}$. The commutator estimate in Lemma 3.3 can directly bound all the terms with an explicit $H_\eta^{-1\otimes|T|}$-norm as the coefficients μ,ν are $W^{1,\infty}$ and η is smooth. For example

$$\|\nu_m\Delta_N(T)\|^2_{H_\eta^{-1\otimes|T|}} \le 4\|\nu\|^2_{W^{1,\infty}(\mathbb{R})}\,\|\Delta_N(T)\|^2_{H_\eta^{-1\otimes|T|}}.$$

This leads to the simplified expression for some constant $\widetilde{C}_0$,

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\Big(\frac{1}{2}\|\Delta_N(T)\|^2_{H_\eta^{-1\otimes|T|}}\Big)\\
&\quad\le \sum_{m=1}^{|T|}\bigg\{\widetilde{C}_0\|\Delta_N(T)\|^2_{H_\eta^{-1\otimes|T|}}\\
&\qquad + \frac{1}{2}\|K\|^2_{L^2}\eta(0)^2\int_{\mathbb{R}^{|T|-1}}\Big(\int_{\mathbb{R}}K^{\otimes|T|}\star((\nu_m/\eta_m)\Delta_N(T)\eta^{\otimes|T|})(t,z)\,\mathrm{d}z_m\Big)^2\prod_{n\neq m}\mathrm{d}z_n
\end{aligned}$$

$$+\frac{1}{2}\|K\|_{L^2}^2\eta(0)^2\int_{\mathbb{R}^{|T|-1}}\bigg(\int_{\mathbb{R}}K^{\otimes|T|}\star((\nu_m/\eta_m)\mathcal{R}_{N,T,m}\eta^{\otimes|T|})(t,z)\,\mathrm{d}z_m\bigg)^2\prod_{n\neq m}\mathrm{d}z_n$$
$$+\frac{2}{\sigma^2}\int_{\mathbb{R}^{|T|}}\bigg(\int_{\mathbb{R}}K^{\otimes|T|+1}\star((\nu_{|T|+1}/\eta_{|T|+1})\Delta_N(T+m)\eta^{\otimes|T|+1})(t,z)\times\mathrm{d}z_{|T|+1}\bigg)^2\prod_{n=1}^{|T|}\mathrm{d}z_n$$
$$+\frac{2}{\sigma^2}\int_{\mathbb{R}^{|T|}}\bigg(\int_{\mathbb{R}}K^{\otimes|T|+1}\star((\nu_{|T|+1}/\eta_{|T|+1})\widetilde{\mathcal{R}}_{N,T+m,|T|+1}\eta^{\otimes|T|+1})(t,z)\times\mathrm{d}z_{|T|+1}\bigg)^2\prod_{n=1}^{|T|}\mathrm{d}z_n\bigg\}. \tag{5.7}$$

The remaining integrals terms in (5.6) can be bounded by first applying Lemma 3.8 followed by Proposition 3.6. For example, consider the first remainder term and write, by Lemma 3.8,

$$\int_{\mathbb{R}^{|T|-1}}\bigg(\int_{\mathbb{R}}K^{\otimes|T|}\star((\nu_m/\eta_m)\mathcal{R}_{N,T,m}\eta^{\otimes|T|})(t,z)\,\mathrm{d}z_m\bigg)^2\prod_{n\neq m}\mathrm{d}z_n \le C(\alpha)^2\|\nu\|_{W^{1,\infty}}^2\|\mathcal{R}_{N,T,m}\|^2_{H_\eta^{-1\otimes|T|}}.$$

Next, apply Proposition 3.6 to the right-hand side to conclude that

$$\int_{\mathbb{R}^{|T|-1}}\bigg(\int_{\mathbb{R}}K^{\otimes|T|}\star((\nu_m/\eta_m)\mathcal{R}_{N,T,m}\eta^{\otimes|T|})(t,z)\,\mathrm{d}z_m\bigg)^2\prod_{n\neq m}\mathrm{d}z_n \le C(\alpha)^2\|\nu\|_{W^{1,\infty}}^2[\exp((2+2\alpha)c(w,|T|))-1]\||\tau_N|(T)\|^2_{H_\eta^{-1\otimes|T|}}.$$

The method applies for the other integrals terms in (5.7), which yields

$$\int_{\mathbb{R}^{|T|-1}}\bigg(\int_{\mathbb{R}}K^{\otimes|T|}\star((\nu_m/\eta_m)\Delta_N(T)\eta^{\otimes|T|})(t,z)\,\mathrm{d}z_m\bigg)^2\prod_{n\neq m}\mathrm{d}z_n \le C(\alpha)^2\|\nu\|_{W^{1,\infty}}^2\|\Delta_N(T)\|^2_{H_\eta^{-1\otimes|T|}},$$

$$\int_{\mathbb{R}^{|T|}}\bigg(\int_{\mathbb{R}}K^{\otimes|T|+1}\star((\nu_{|T|+1}/\eta_{|T|+1})\Delta_N(T+m)\eta^{\otimes|T|+1})(t,z)\,\mathrm{d}z_{|T|+1}\bigg)^2\prod_{n=1}^{|T|}\mathrm{d}z_n \le C(\alpha)^2\|\nu\|_{W^{1,\infty}}^2\|\Delta_N(T+m)\|^2_{H_\eta^{-1\otimes(|T|+1)}},$$

together with

$$\int_{\mathbb{R}^{|T|}}\bigg(\int_{\mathbb{R}}K^{\otimes|T|+1}\star((\nu_{|T|+1}/\eta_{|T|+1})\widetilde{\mathcal{R}}_{N,T+m,|T|+1}\eta^{\otimes|T|+1})(t,z)\,\mathrm{d}z_{|T|+1}\bigg)^2\prod_{n=1}^{|T|}\mathrm{d}z_n \le C(\alpha)^2\|\nu\|_{W^{1,\infty}}^2[\exp((2+2\alpha)c(w,|T|))-1]\||\tau_N|(T)\|^2_{H_\eta^{-1\otimes|T|}}.$$

Inserting those bounds into the energy estimate (5.7), we obtain a recursive differential inequality: For all $T \in \mathcal{T}$,

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}\|\Delta_N(T)(t,\cdot)\|^2_{H_\eta^{-1\otimes|T|}} \\ &\quad\leq \sum_{m=1}^{|T|}\Big\{\widetilde{C}_0\|\Delta_N(T)(t,\cdot)\|^2_{H_\eta^{-1\otimes|T|}} + \widetilde{C}_1\|\Delta_N(T+m)(t,\cdot)\|^2_{H_\eta^{-1\otimes(|T|+1)}} \\ &\qquad\quad + \varepsilon_0(T)\||\tau_N|(T)(t,\cdot)\|^2_{H_\eta^{-1\otimes|T|}} + \varepsilon_1(T)\||\tau_N|(T+m)(t,\cdot)\|^2_{H_\eta^{-1\otimes(|T|+1)}}\Big\}, \end{aligned} \tag{5.8}$$

where we can even provide explicit expressions for the constants:

$$\begin{aligned} \widetilde{C}_0 &= 4 + \frac{\sigma^2}{2} + 4\Big(\|\nu\|^2_{W^{1,\infty}} + \|\mu(\eta'/\eta)\|^2_{W^{1,\infty}} + \frac{\sigma^2}{2}\|\eta''/\eta\|^2_{W^{1,\infty}} \\ &\qquad\qquad + \frac{4}{\sigma^2}\|\mu\|^2_{W^{1,\infty}} + 2\sigma^2\|(\eta'/\eta)\|^2_{W^{1,\infty}}\Big) \\ &\quad + \|K\|^2_{L^2}\eta(0)^2C(\alpha)^2\|\nu\|^2_{W^{1,\infty}}, \\ \widetilde{C}_1 &= \frac{4C(\alpha)^2}{\sigma^2}\|\nu\|^2_{W^{1,\infty}}, \\ \varepsilon_0(T) &= \|K\|^2_{L^2}\eta(0)^2C(\alpha)^2\|\nu\|^2_{W^{1,\infty}}[\exp((2+2\alpha)c(w,|T|)) - 1], \\ \varepsilon_1(T) &= \frac{4C(\alpha)^2}{\sigma^2}\|\nu\|^2_{W^{1,\infty}}[\exp((2+2\alpha)c(w,|T|)) - 1]. \end{aligned}$$

We can now restrict the recursion relations by truncating them at any given depth $n \geq 1$, meaning that we only consider the inequalities (5.8) for all $T \in \mathcal{T}$ such that $|T| \leq n-1$. In such a case, since

$$c(w,|T|) \leq |T|(\max_{i,j}|w_{i,j;N}|) \leq n\bar{w}_N,$$

the coefficients ε_0, ε_1 can take the vanishing expression

$$\begin{aligned} \varepsilon_0(n) &= \|K\|^2_{L^2}\eta(0)^2C(\alpha)^2\|\nu\|^2_{W^{1,\infty}}[\exp((2+2\alpha)n\bar{w}_N) - 1], \\ \varepsilon_1(n) &= \frac{4C(\alpha)^2}{\sigma^2}\|\nu\|^2_{W^{1,\infty}}[\exp((2+2\alpha)n\bar{w}_N) - 1]. \end{aligned}$$

For a fixed depth $n \geq 1$, $\varepsilon_0(n)$ and $\varepsilon_1(n)$ now vanish as $\bar{w}_N \to 0$.

Let us now rescale the energy inequality through some $\lambda^{|T|}$ factor: For all $T \in \mathcal{T}$ such that $|T| \leq n-1$,

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}\lambda^{|T|}\|\Delta_N(T)(t,\cdot)\|^2_{H_\eta^{-1\otimes|T|}} \\ &\quad\leq \sum_{m=1}^{|T|}\Big\{\widetilde{C}_0\lambda^{|T|}\|\Delta_N(T)(t,\cdot)\|^2_{H_\eta^{-1\otimes|T|}} + (\widetilde{C}_1/\lambda)\lambda^{|T|+1}\|\Delta_N(T+m)(t,\cdot)\|^2_{H_\eta^{-1\otimes(|T|+1)}} \end{aligned}$$

$$+ \varepsilon_0(n)\lambda^{|T|} \||\tau_N|(T)(t,\cdot)\|^2_{H_\eta^{-1\otimes|T|}}$$
$$+ (\varepsilon_1(n)/\lambda)\lambda^{|T|+1} \||\tau_N|(T+m)(t,\cdot)\|^2_{H_\eta^{-1\otimes(|T|+1)}} \Big\}. \tag{5.9}$$

We also recall the a priori bound (2.7) for τ_N, τ_∞ assumed in Theorem 2.6:

$$\sup_{t\le t_*} \max_{|T|\le \max(n,|T_*|)} \lambda^{\frac{|T|}{2}} \left(\||\tau_N|(T, w_N, f_N)(t,\cdot)\|_{H_\eta^{-1\otimes|T|}} + \|\tau_\infty(T)(t,\cdot)\|_{H_\eta^{-1\otimes|T|}}\right) \le C_{\lambda;\eta},$$

where $T_* \in \mathcal{T}$ is the tree index in the final estimate (2.6). By the triangular inequality, this implies the following uniform bound of Δ_N:

$$\sup_{t\le t_*} \max_{|T|\le \max(n,|T_*|)} \lambda^{|T|} \|\Delta_N(T)(t,\cdot)\|^2_{H_\eta^{-1\otimes|T|}} \le C^2_{\lambda;\eta}. \tag{5.10}$$

Denote

$$\begin{aligned} M_k(t) &= \max_{|T|\le k} \lambda^{|T|} \|\Delta_N(T)(t,\cdot)\|^2_{H_\eta^{-1\otimes|T|}},\\ C &= \widetilde{C}_0 + \widetilde{C}_1/\lambda,\\ \varepsilon &= [\varepsilon_0(n) + \varepsilon_1(n)/\lambda] C^2_{\lambda;\eta},\\ L &= C^2_{\lambda;\eta},\\ n &= n,\ \ n' = |T_*|, \end{aligned}$$

so that (5.9) and (5.10) can be summarized as follows:

$$\frac{\mathrm{d}}{\mathrm{d}t} M_k(t) \le k(CM_{k+1}(t) + \varepsilon), \quad \forall 1 \le k \le n-1, \tag{5.11a}$$
$$M_k(t) \le L, \quad \forall 1 \le k \le \max(n, n'),\ t \in [0, t_*]. \tag{5.11b}$$

We now invoke the following result.

Lemma 5.2. *Consider a sequence of non-negative functions* $(M_k(t))_{k=1}^\infty$ *on* $t \in [0, t_*]$ *that satisfies the inequalities* (5.11a)–(5.11b) *with* $[\varepsilon/CL + (2\theta)^n] \le 1$. *Then*

$$\max_{1\le k\le \max(n,n')} [\theta^k M_k(t)]$$
$$\le L(Ct/\theta + 2) \max\Big([\varepsilon/CL + (2\theta)^n],\ \max_{1\le k\le n-1} [\theta^k M_k(0)]/L\Big)^{\frac{1}{p^{(Ct/\theta+1)}}}, \tag{5.12}$$

holds for any $1 < p < \infty$, $0 < \theta < 2^{-p'}$ *where* $1/p + 1/p' = 1$, *and any* $t \in [0, t_*]$.

Assume for the time being that Lemma 5.2 holds and apply it to (5.9) and (5.10). Choose $p = 2$, $\theta = 1/8$, and substitute ε, C, L with their explicit expressions to find that

$$\varepsilon/CL = \frac{\varepsilon_0(n) + \varepsilon_1(n)/\lambda}{\widetilde{C}_0 + \widetilde{C}_1/\lambda} = C_1[\exp((2 + 2\alpha)\overline{w}n) - 1],$$

where C_1 depends only on λ, the $W^{1,\infty}$-regularity of coefficients μ, ν and constant $\sigma > 0$ in (2.1), but neither on $\bar{w}_N$ nor on n. Choosing $C_0 = C/\theta$, and as $\bar{w}_N \to 0$ as $N \to \infty$, we deduce that for N large enough,

$$\bar{\varepsilon} = \varepsilon/CL + (2\theta)^n = C_1[\exp((2+2\alpha)n\bar{w}_N) - 1] + (1/4)^n \leq 1.$$

The conclusion of Lemma 5.2 hence holds, showing that

$$\begin{aligned} &\max_{|T|\leq \max(n,|T_*|)} (\lambda/8)^{|T|} \|\tau_N(T, w_N, f_N)(t,\cdot) - \tau_\infty(T)(t,\cdot)\|^2_{H_\eta^{-1\otimes|T|}} \\ &\leq C^2_{\lambda;\eta}(C_0 t + 2) \\ &\quad \times \max\Big(\bar{\varepsilon}, \max_{|T|\leq n-1} (\lambda/8)^{|T|} \|\tau_N(T, w_N, f_N)(0,\cdot) - \tau_\infty(T)(0,\cdot)\|^2_{H_\eta^{-1\otimes|T|}} / C^2_{\lambda;\eta}\Big)^{\frac{1}{2^{(C_0 t+1)}}}. \end{aligned}$$

This can be further simplified to (2.6) by relaxing the maximum on the left-hand side as $T = T_*$, taking the maximum on the right-hand side over $|T| \leq \max(n, |T_*|)$, and choosing C_2 in (2.6) as $C_2 = \max(C_0 t + 2, 2^{(C_0 t+1)})$. ■

5.3. Proof of Lemma 5.2

Proof of Lemma 5.2. Let us restate here the recursive differential inequality (5.11a),

$$\frac{\mathrm{d}}{\mathrm{d}t} M_k(t) \leq k(CM_{k+1}(t) + \varepsilon), \quad \forall 1 \leq k \leq n-1,$$

which directly yields

$$\frac{\mathrm{d}}{\mathrm{d}t}(M_k(t) + (\varepsilon/C)) \leq kC(M_{k+1}(t) + (\varepsilon/C)), \quad \forall 1 \leq k \leq n-1.$$

For any $1 \leq k \leq n-1$ and $t \in [0, t_*]$, by inductively integrating the inequalities in time, we obtain

$$\begin{aligned} (M_k(t) + (\varepsilon/C)) \leq C^{n-k} &\int_s^t \binom{n-1}{k-1} \frac{(t-r)^{n-k-1}}{n-k-1} (M_n(r) + (\varepsilon/C))\,\mathrm{d}r \\ &+ \sum_{l=k}^{n-1} C^{l-k} \binom{l-1}{k-1} (t-s)^{l-k} (M_l(s) + (\varepsilon/C)). \end{aligned}$$

We estimate the increase on M_k within time steps of size

$$t - s = \theta/C.$$

First, we bound the constant terms,

$$\begin{aligned} &C^{n-k} \int_s^t \binom{n-1}{k-1} \frac{(t-r)^{n-k-1}}{n-k-1} (\varepsilon/C)\,\mathrm{d}r + \sum_{l=k}^{n-1} C^{l-k} \binom{l-1}{k-1} (t-s)^{l-k} (\varepsilon/C) \\ &= (\varepsilon/C)\Big\{C^{n-k} \binom{n-1}{k-1} (t-s)^{n-k-1} + \sum_{l=k}^{n-1} C^{l-k} \binom{l-1}{k-1} (t-s)^{l-k}\Big\} \end{aligned}$$

$$= (\varepsilon/C)\sum_{l=k}^{n} C^{l-k}\binom{l-1}{k-1}(t-s)^{l-k}$$
$$\le \theta^{-k}(\varepsilon/C)\sum_{l=k}^{n}\binom{l-1}{k-1}\theta^{l},$$

where the last inequality uses our choice of time step $(t-s)\le\theta/C$.

On the other hand, for $\theta\le 1/2$,

$$\sum_{l=k}^{\infty}\binom{l-1}{k-1}\theta^{l} = \frac{1}{(\theta^{-1}-1)^{k}} \le 1.$$

Hence,

$$C^{n-k}\int_s^t \binom{n-1}{k-1}\frac{(t-r)^{n-k-1}}{n-k-1}(\varepsilon/C)\,\mathrm{d}r + \sum_{l=k}^{n-1} C^{l-k}\binom{l-1}{k-1}(t-s)^{l-k}(\varepsilon/C) \le \theta^{-k}\frac{\varepsilon}{C}.$$

We now turn to the terms involving $M_l(s)$ and $M_n(r)$ (with $s\le r\le t$). For $M_n(r)$ we have no choice but to take

$$M_n(r)\le L.$$

But for $M_l(s)$, $k\le l\le n-1$, we have

$$M_l(s)\le \min\Big(L, \max_{1\le m\le n-1}[\theta^m M_m(s)]\theta^{-l}\Big),$$

together with any geometric average between the two terms. Choose $\frac{1}{p}+\frac{1}{p'}=1$ so that

$$M_l(s)\le L^{\frac{1}{p'}}\Big(\max_{1\le m\le n-1}[\theta^m M_m(s)]\theta^{-l}\Big)^{\frac{1}{p}}$$
$$= L^{\frac{1}{p'}}\max_{1\le m\le n-1}[\theta^m M_m(s)]^{\frac{1}{p}}(\theta^{\frac{1}{p}})^{-l}.$$

Then we may write

$$C^{n-k}\int_s^t \binom{n-1}{k-1}\frac{(t-r)^{n-k-1}}{n-k-1}M_n(r)\,\mathrm{d}r + \sum_{l=k}^{n-1} C^{l-k}\binom{l-1}{k-1}(t-s)^{l-k}M_l(s),$$
$$\le C^{n-k}\binom{n-1}{k-1}(t-s)^{n-k}L$$
$$+\sum_{l=k}^{n-1} C^{l-k}\binom{l-1}{k-1}(t-s)^{l-k}L^{\frac{1}{p'}}\max_{1\le m\le n-1}[\theta^m M_m(s)]^{\frac{1}{p}}(\theta^{\frac{1}{p}})^{-l}$$
$$\le \theta^{-k}\Big\{L\binom{n-1}{k-1}\theta^{n} + L^{\frac{1}{p'}}\max_{1\le m\le n-1}[\theta^m M_m(s)]^{\frac{1}{p}}\sum_{l=k}^{n-1}\binom{l-1}{k-1}(\theta^{\frac{1}{p'}})^{-l}\Big\},$$

where we again use our choice of time step $(t-s)\le\theta/C$ in the last inequality.

Observe that

$$\binom{n-1}{k-1}\theta^n \le 2^{n-1}\theta^n, \quad \sum_{l=k}^{n-1}\binom{l-1}{k-1}(\theta^{\frac{1}{p'}})^{-l} \le \frac{1}{(\theta^{-\frac{1}{p'}}-1)^k} \le 1$$

when choosing $\theta^{\frac{1}{p'}} \le 1/2$, so that

$$C^{n-k}\int_s^t \binom{n-1}{k-1}\frac{(t-r)^{n-k-1}}{n-k-1}M_n(r)\,\mathrm{d}r + \sum_{l=k}^{n-1} C^{l-k}\binom{l-1}{k-1}(t-s)^{l-k}M_l(s)$$
$$\le \theta^{-k}\Big(L(2\theta)^n + L^{\frac{1}{p'}}\max_{1\le m\le n-1}[\theta^m M_m(s)]^{\frac{1}{p}}\Big).$$

Combining these bounds, provided that $\theta^{\frac{1}{p'}} \le 1/2$, we have, for all $1 \le k \le n-1$, that

$$M_k(t) \le \theta^{-k}\Big\{(\varepsilon/C) + L(2\theta)^n + L^{\frac{1}{p'}}\max_{1\le m\le n-1}[\theta^m M_m(s)]^{\frac{1}{p}}\Big\}.$$

On the other hand, for $n \le k \le \max(n, n')$, we simply have $M_k(t) \le L$. As $\theta^{-k+n} \ge 1$,

$$M_k(t) \le L \le \theta^{-k}\{L(2\theta)^n\},$$

and we can combine the two cases to obtain

$$\max_{1\le k\le \max(n,n')}[\theta^k M_k(t)] \le (\varepsilon/C) + L(2\theta)^n + L^{\frac{1}{p'}}\max_{1\le k\le n-1}[\theta^k M_k(s)]^{\frac{1}{p}}.$$

If $t \le \theta/C$ we are done, but otherwise we need to sum the various bounds. Letting $t_j = j\,\theta/C$ and recursively applying the preceding inequality at each step by taking $t = t_j$ and $s = t_{j-1}$, we have that

$$\max_{1\le k\le \max(n,n')}[\theta^k M_k(t_j)]$$
$$\le (\varepsilon/C) + L(2\theta)^n + L^{\frac{1}{p'}}\max_{1\le k\le n-1}[\theta^k M_k(t_{j-1})]^{\frac{1}{p}}$$
$$\le (\varepsilon/C) + L(2\theta)^n + L^{\frac{1}{p'}}\Big\{(\varepsilon/C) + L(2\theta)^n + L^{\frac{1}{p'}}\max_{1\le k\le n-1}[\theta^k M_k(t_{j-2})]^{\frac{1}{p}}\Big\}^{\frac{1}{p}}$$
$$\le (\varepsilon/C) + L(2\theta)^n + L^{\frac{1}{p'}}\{(\varepsilon/C) + L(2\theta)^n\}^{\frac{1}{p}} + L^{1-\frac{1}{p^2}}\max_{1\le k\le n-1}[\theta^k M_k(t_{j-2})]^{\frac{1}{p^2}}$$
$$\dots$$
$$\le \sum_{i=0}^{j-1} L^{1-\frac{1}{p^i}}\{(\varepsilon/C) + L(2\theta)^n\}^{\frac{1}{p^i}} + L^{1-\frac{1}{p^j}}\max_{1\le k\le n-1}[\theta^k M_k(0)]^{\frac{1}{p^j}},$$

where we use that $(a+b)^{1/p} \le a^{1/p} + b^{1/p}$ by concavity.

For any $t \geq 0$, we hence have with $j(t) = \lfloor \frac{Ct}{\theta} \rfloor + 1$,

$$\max_{1 \leq k \leq \max(n,n')} [\theta^k M_k(t)] \leq \sum_{i=0}^{j(t)-1} L^{1-\frac{1}{p^i}} \{\varepsilon/C + L(2\theta)^n\}^{\frac{1}{p^i}} + L^{1-\frac{1}{p^{j(t)}}} \max_{1 \leq k \leq n-1} [\theta^k M_k(0)]^{\frac{1}{p^{j(t)}}}. \tag{5.13}$$

Finally, by the assumption that $[\varepsilon/CL + (2\theta)^n] \leq 1$,

$$\forall i \leq j, \quad L^{1-\frac{1}{p^i}} \{\varepsilon/C + L(2\theta)^n\}^{\frac{1}{p^i}} = L\{\varepsilon/CL + (2\theta)^n\}^{\frac{1}{p^i}} \leq L\{\varepsilon/CL + (2\theta)^n\}^{\frac{1}{p^j}}.$$

Hence we can replace every i and every $j(t)$ in (5.13) by $(Ct/\theta + 1)$, which gives the looser bound (5.12), restated here:

$$\max_{1 \leq k \leq \max(n,n')} [\theta^k M_k(t)] \leq L(Ct/\theta + 2) \max\Big([\varepsilon/CL + (2\theta)^n], \sup_{1 \leq k \leq n-1} [\theta^k M_k(0)]/L\Big)^{\frac{1}{p^{(Ct/\theta+1)}}}. \qquad \blacksquare$$

Funding. P.-E. Jabin and D. Zhou were partially supported by NSF DMS grants 2205694 and 2219297. D. Zhou was supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 101034255.

References

[1] S.-i. Amari, Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol. Cybernet.* **27** (1977), no. 2, 77–87 Zbl 0367.92005 MR 0681526

[2] N. Ayi and N. Pouradier Duteil, Mean-field and graph limits for collective dynamics models with time-varying weights. *J. Differential Equations* **299** (2021), 65–110 Zbl 1472.34084 MR 4292107

[3] L. Badel, S. Lefort, T. K. Berger, C. C. H. Petersen, W. Gerstner, and M. J. E. Richardson, Extracting non-linear integrate-and-fire models from experimental data using dynamic $I-V$ curves. *Biol. Cybernet.* **99** (2008), no. 4-5, 361–370 Zbl 1161.92005 MR 2461097

[4] J. Baladron, D. Fasoli, O. Faugeras, and J. Touboul, Mean-field description and propagation of chaos in networks of Hodgkin–Huxley and FitzHugh–Nagumo neurons. *J. Math. Neurosci.* **2** (2012), article no. 10 Zbl 1322.60205 MR 2974499

[5] E. Bayraktar, S. Chakraborty, and R. Wu, Graphon mean field systems. *Ann. Appl. Probab.* **33** (2023), no. 5, 3587–3619 Zbl 1533.70013 MR 4663492

[6] J. Bergh and J. Löfström, *Interpolation spaces. An introduction*. Grundlehren Math. Wiss. 223, Springer, Berlin-New York, 1976 Zbl 0344.46071 MR 0482275

[7] G. Bet, F. Coppini, and F. R. Nardi, Weakly interacting oscillators on dense random graphs. *J. Appl. Probab.* **61** (2024), no. 1, 255–278 Zbl 1551.60127 MR 4706689

[8] R. L. Beurle, Properties of a mass of cells capable of regenerating pulses. *Philos. Trans. Roy. Soc. B* **240** (1956), 55–94

[9] C. Borgs, J. T. Chayes, H. Cohn, and Y. Zhao, An L^p theory of sparse graph convergence II: LD convergence, quotients and right convergence. *Ann. Probab.* **46** (2018), no. 1, 337–396 Zbl 1386.05099 MR 3758733

[10] C. Borgs, J. T. Chayes, H. Cohn, and Y. Zhao, An L^p theory of sparse graph convergence I: Limits, sparse random graph models, and power law distributions. *Trans. Amer. Math. Soc.* **372** (2019), no. 5, 3019–3062 Zbl 1417.05186 MR 3988601

[11] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi, Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.* **219** (2008), no. 6, 1801–1851 Zbl 1213.05161 MR 2455626

[12] M. Bossy, O. Faugeras, and D. Talay, Clarification and complement to "Mean-field description and propagation of chaos in networks of Hodgkin–Huxley and FitzHugh–Nagumo neurons". *J. Math. Neurosci.* **5** (2015), article no. 19 Zbl 1361.92014 MR 3392551

[13] D. Bresch, P.-E. Jabin, and J. Soler, A new approach to the mean-field limit of Vlasov–Fokker–Planck equations. *Anal. PDE* **18** (2025), no. 4, 1037–1064 Zbl 08023891 MR 4885878

[14] A. N. Burkitt, A review of the integrate-and-fire neuron model. I. Homogeneous synaptic input. *Biol. Cybernet.* **95** (2006), no. 1, 1–19 Zbl 1161.92315 MR 2237180

[15] M. J. Cáceres, J. A. Carrillo, and B. Perthame, Analysis of nonlinear noisy integrate & fire neuron models: Blow-up and steady states. *J. Math. Neurosci.* **1** (2011), article no. 7 Zbl 1259.35198 MR 2853216

[16] M. J. Cáceres and B. Perthame, Beyond blow-up in excitatory integrate and fire neuronal networks: Refractory period and spontaneous activity. *J. Theoret. Biol.* **350** (2014), 81–89 Zbl 1412.92001 MR 3190511

[17] J. A. Carrillo, M. D. M. González, M. P. Gualdani, and M. E. Schonbek, Classical solutions for a nonlinear Fokker–Planck equation arising in computational neuroscience. *Comm. Partial Differential Equations* **38** (2013), no. 3, 385–409 Zbl 1282.35382 MR 3019444

[18] J. A. Carrillo, B. Perthame, D. Salort, and D. Smets, Qualitative properties of solutions for the noisy integrate and fire model in computational neuroscience. *Nonlinearity* **28** (2015), no. 9, 3365–3388 Zbl 1336.35332 MR 3403402

[19] J. Chevallier, Mean-field limit of generalized Hawkes processes. *Stochastic Process. Appl.* **127** (2017), no. 12, 3870–3912 Zbl 1374.60090 MR 3718099

[20] J. Chevallier, A. Duarte, E. Löcherbach, and G. Ost, Mean field limits for nonlinear spatially extended Hawkes processes with exponential memory kernels. *Stochastic Process. Appl.* **129** (2019), no. 1, 1–27 Zbl 1404.60069 MR 3906989

[21] H. Chiba and G. S. Medvedev, The mean field analysis of the Kuramoto model on graphs I. The mean field equation and transition point formulas. *Discrete Contin. Dyn. Syst.* **39** (2019), no. 1, 131–155 Zbl 1406.34064 MR 3918168

[22] H. Chiba and G. S. Medvedev, The mean field analysis of the Kuramoto model on graphs II. Asymptotic stability of the incoherent state, center manifold reduction, and bifurcations. *Discrete Contin. Dyn. Syst.* **39** (2019), no. 7, 3897–3921 Zbl 1420.34062 MR 3960490

[23] F. Coppini, H. Dietert, and G. Giacomin, A law of large numbers and large deviations for interacting diffusions on Erdős–Rényi graphs. *Stoch. Dyn.* **20** (2020), no. 2, article no. 2050010 Zbl 1434.60279 MR 4080158

[24] Q. Cormier, A mean-field model of integrate-and-fire neurons: Non-linear stability of the stationary solutions. *Math. Neurosci. Appl.* **4** (2024), article no. 1 MR 4833513

[25] Q. Cormier, E. Tanré, and R. Veltz, Long time behavior of a mean-field model of interacting neurons. *Stochastic Process. Appl.* **130** (2020), no. 5, 2553–2595 Zbl 1437.35708 MR 4080722

[26] Q. Cormier, E. Tanré, and R. Veltz, Hopf bifurcation in a mean-field model of spiking neurons. *Electron. J. Probab.* **26** (2021), article no. 121 Zbl 1481.60201 MR 4316639

[27] A. De Masi, A. Galves, E. Löcherbach, and E. Presutti, Hydrodynamic limit for interacting neurons. *J. Stat. Phys.* **158** (2015), no. 4, 866–902 Zbl 1315.35222 MR 3311484

[28] F. Delarue, J. Inglis, S. Rubenthaler, and E. Tanré, Global solvability of a networked integrate-and-fire model of McKean–Vlasov type. *Ann. Appl. Probab.* **25** (2015), no. 4, 2096–2133 Zbl 1322.60085 MR 3349003

[29] F. Delarue, J. Inglis, S. Rubenthaler, and E. Tanré, Particle systems with a singular mean-field self-excitation. Application to neuronal networks. *Stochastic Process. Appl.* **125** (2015), no. 6, 2451–2492 Zbl 1328.60134 MR 3322871

[30] S. Delattre, N. Fournier, and M. Hoffmann, Hawkes processes on large networks. *Ann. Appl. Probab.* **26** (2016), no. 1, 216–261 Zbl 1334.60082 MR 3449317

[31] Y. Deng and Z. Hani, On the derivation of the wave kinetic equation for NLS. *Forum Math. Pi* **9** (2021), article no. e6 Zbl 1479.35771 MR 4291361

[32] Y. Deng and Z. Hani, Full derivation of the wave kinetic equation. *Invent. Math.* **233** (2023), no. 2, 543–724 Zbl 1530.35276 MR 4607721

[33] A. Drogoul and R. Veltz, Exponential stability of the stationary distribution of a mean field of spiking neural network. *J. Differential Equations* **270** (2021), 809–842 Zbl 1451.35025 MR 4152223

[34] X. Erny, E. Löcherbach, and D. Loukianova, Conditional propagation of chaos for mean field systems of interacting neurons. *Electron. J. Probab.* **26** (2021), article no. 20 Zbl 1480.60081 MR 4235471

[35] R. FitzHugh, Impulses and physiological states in theoretical models of nerve membrane. *Biophys. J.* **1** (1961), no. 6, 445–466

[36] F. Flandoli, E. Priola, and G. Zanco, A mean-field model with discontinuous coefficients for neurons with spatial interaction. *Discrete Contin. Dyn. Syst.* **39** (2019), no. 6, 3037–3067 Zbl 1415.60061 MR 3959420

[37] N. Fournier and E. Löcherbach, On a toy model of interacting neurons. *Ann. Inst. Henri Poincaré Probab. Stat.* **52** (2016), no. 4, 1844–1876 MR 3573298

[38] C. D. Geisler and J. M. Goldberg, A stochastic model of the repetitive activity of neurons. *Biophys. J.* **6** (1966), no. 1, 53–69

[39] G. L. Gerstein and B. Mandelbrot, Random walk models for the spike activity of a single neuron. *Biophys. J.* **4** (1964), no. 1, 41–68

[40] W. Gerstner and W. M. Kistler, *Spiking neuron models*. Cambridge University Press, Cambridge, 2002 Zbl 1100.92501 MR 1923120

[41] W. Gerstner, W. M. Kistler, R. Naud, and L. Paninski, *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press, 2014

[42] M. A. Gkogkas and C. Kuehn, Graphop mean-field limits for Kuramoto-type models. *SIAM J. Appl. Dyn. Syst.* **21** (2022), no. 1, 248–283 Zbl 1483.35270 MR 4366119

[43] M. A. Gkogkas, C. Kuehn, and C. Xu, Continuum limits for adaptive network dynamics. *Commun. Math. Sci.* **21** (2023), no. 1, 83–106 Zbl 1539.35269 MR 4530018

[44] M. A. Gkogkas, C. Kuehn, and C. Xu, Mean field limits of co-evolutionary signed heterogeneous networks. *European J. Appl. Math.* (2025) DOI 10.1017/S0956792524000858

[45] F. Golse, C. Mouhot, and V. Ricci, Empirical measures and Vlasov hierarchies. *Kinet. Relat. Models* **6** (2013), no. 4, 919–943 Zbl 1302.82063 MR 3177635

[46] P. Grazieschi, M. Leocata, C. Mascart, J. Chevallier, F. Delarue, and E. Tanré, Network of interacting neurons with random synaptic weights. In *CEMRACS 2017—Numerical methods for stochastic models: Control, uncertainty quantification, mean-field*, pp. 445–475, ESAIM Proc. Surveys 65, EDP Sci., Les Ulis, 2019 Zbl 1418.92006 MR 3968552

[47] J. S. Griffith, A field theory of neural nets. I: Derivation of field equations. *Bull. Math. Biophys.* **25** (1963), 111–120 Zbl 0129.34804 MR 0156712

[48] J. S. Griffith, On the stability of brain-like structures. *Biophys. J.* **3** (1963), 299–308 Zbl 0175.31703 MR 0157789

[49] J. S. Griffith, A field theory of neural nets. II. Properties of the field equations. *Bull. Math. Biophys.* **27** (1965), 187–195 Zbl 0139.14103 MR 0195597

[50] D. Hebb, *The organization of behavior*. Wiley New York, 1949

[51] A. V. Hill, Excitation and accommodation in nerve. *Proc. Roy. Soc. London B* **119** (1936), no. 814, 305–355

[52] A. L. Hodgkin and A. F. Huxley, A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* **117** (1952), no. 4, 500–544

[53] B. K. Hulse, H. Haberkern, R. Franconville, D. B. Turner-Evans, S. Y. Takemura, T. Wolff, M. Noorman, M. Dreher, C. Dan, R. Parekh, A. Hermundstad, G. M. Rubin, and V. Jayaraman, A connectome of the Drosophila central complex reveals network motifs suitable for flexible navigation and context-dependent action selection. *eLife* **10** (2021), article no. e66039

[54] J. Inglis and D. Talay, Mean-field limit of a stochastic particle system smoothly interacting through threshold hitting-times and applications to neural networks with dendritic component. *SIAM J. Math. Anal.* **47** (2015), no. 5, 3884–3916 Zbl 1325.60158 MR 3411723

[55] P.-E. Jabin, D. Poyato, and J. Soler, Mean-field limit of non-exchangeable systems. *Comm. Pure Appl. Math.* **78** (2025), no. 4, 651–741 Zbl 07988477 MR 4863199

[56] D. Kaliuzhnyi-Verbovetskyi and G. S. Medvedev, The mean field equation for the Kuramoto model on graph sequences with non-Lipschitz limit. *SIAM J. Math. Anal.* **50** (2018), no. 3, 2441–2465 Zbl 1406.35096 MR 3799057

[57] B. W. Knight, The relationship between the firing rate of a single neuron and the level of activity in a population of neurons: Experimental evidence for resonant enhancement in the population response. *J. Gen. Physiol.* **59** (1972), no. 6, 767–778

[58] C. Kuehn and C. Xu, Vlasov equations on digraph measures. *J. Differential Equations* **339** (2022), 261–349 Zbl 1498.05251 MR 4477050

[59] Y. Kuramoto, *International symposium on mathematical problems in theoretical physics*. Lecture Notes in Physics 30, 1975

[60] D. Lacker, Hierarchies, entropy, and quantitative propagation of chaos for mean field diffusions. *Probab. Math. Phys.* **4** (2023), no. 2, 377–432 Zbl 1515.82109 MR 4595391

[61] D. Lacker, K. Ramanan, and R. Wu, Local weak convergence for sparse networks of interacting processes. *Ann. Appl. Probab.* **33** (2023), no. 2, 843–888 Zbl 1525.60114 MR 4564415

[62] D. Lacker, L. C. Yeung, and F. Zhou, Quantitative propagation of chaos for non-exchangeable diffusions via first-passage percolation. 2024, arXiv:2409.08882v1

[63] L. Lapicque, Recherches quantitatives sur l'excitation électrique des nerfs traitée comme une polarisation. *J. Physiol. Pathol. Générale* **9** (1907), 567–578

[64] L. Lovász and B. Szegedy, Limits of dense graph sequences. *J. Combin. Theory Ser. B* **96** (2006), no. 6, 933–957 Zbl 1113.05092 MR 2274085

[65] E. Luçon, Quenched asymptotics for interacting diffusions on inhomogeneous random graphs. *Stochastic Process. Appl.* **130** (2020), no. 11, 6783–6842 Zbl 1454.60046 MR 4158803

[66] W. S. McCulloch and W. Pitts, A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **5** (1943), 115–133 Zbl 0063.03860 MR 0010388
[67] G. S. Medvedev, The continuum limit of the Kuramoto model on sparse random graphs. *Commun. Math. Sci.* **17** (2019), no. 4, 883–898 Zbl 1432.34045 MR 4030504
[68] J. Nagumo, S. Arimoto, and S. Yoshizawa, An active pulse transmission line simulating nerve axon. *Proceedings of the IRE* **50** (1962), no. 10, 2061–2070
[69] R. I. Oliveira, G. H. Reis, and L. M. Stolerman, Interacting diffusions on sparse graphs: Hydrodynamics from local weak limits. *Electron. J. Probab.* **25** (2020), article no. 110 Zbl 1459.60207 MR 4150522
[70] K. Pakdaman, B. Perthame, and D. Salort, Dynamics of a structured neuron population. *Nonlinearity* **23** (2010), no. 1, 55–75 Zbl 1267.35239 MR 2576373
[71] K. Pakdaman, M. Thieullen, and G. Wainrib, Fluid limit theorems for stochastic hybrid systems with application to neuron models. *Adv. in Appl. Probab.* **42** (2010), no. 3, 761–794 Zbl 1232.60019 MR 2779558
[72] B. Perthame and D. Salort, On a voltage-conductance kinetic system for integrate & fire neural networks. *Kinet. Relat. Models* **6** (2013), no. 4, 841–864 Zbl 1284.35103 MR 3177631
[73] B. Perthame, D. Salort, and G. Wainrib, Distributed synaptic weights in a LIF neural network and learning rules. *Phys. D* **353/354** (2017), 20–30 Zbl 1378.92006 MR 3669355
[74] J. Pham, K. Pakdaman, J. Champagnat, and J.-F. Vibert, Activity in sparsely connected excitatory neural networks: Effect of connectivity. *Neural Netw.* **11** (1998), no. 3, 415–434
[75] M. G. Riedler, M. Thieullen, and G. Wainrib, Limit theorems for infinite-dimensional piecewise deterministic Markov processes. Applications to stochastic excitable membrane models. *Electron. J. Probab.* **17** (2012), no. 55, 1–48 Zbl 1255.60126 MR 2955047
[76] H. Spohn, *Large scale dynamics of interacting particles*. Springer, 1991 Zbl 0742.76002
[77] O. Sporns, *Networks of the brain*. MIT Press, Cambridge, MA, 2010
[78] N. Torres and D. Salort, Dynamics of neural networks with elapsed time model and learning processes. *Acta Appl. Math.* **170** (2020), 1065–1099 Zbl 1464.35380 MR 4163271
[79] H. R. Wilson and J. D. Cowan, Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.* **12** (1972), no. 1, 1–24

Received 6 March 2024; revised 21 January 2025; accepted 9 May 2025.

Pierre-Emmanuel Jabin
Department of Mathematics, The Pennsylvania State University, 109 McAllister, University Park, 16802, USA; pejabin@psu.edu

Datong Zhou
Laboratoire Jacques-Louis Lions & Laboratoire de Probabilités, Statistique et Modélisation, Sorbonne Université, 4 Pl. Jussieu, 75005 Paris, France; datong.zhou@sorbonne-universite.fr