SIMULATION OF MCKEAN VLASOV SDES WITH SUPER LINEAR GROWTH

GONÇALO DOS REIS†, STEFAN ENGELHARDT‡, AND GREIG SMITH§

Abstract. We present two fully probabilistic numerical schemes, one explicit and one implicit, for the simulation of McKean-Vlasov Stochastic Differential Equations (MV-SDEs) with drifts of super-linear growth and random initial condition.

We provide a pathwise propagation of chaos result and show strong convergence for both schemes on the consequent particle system. The explicit scheme attains the standard \( \frac{1}{2} \) rate in stepsize. From a technical point of view, we successfully use stopping times to prove the convergence of the implicit method although we avoid them altogether for the explicit one. The combination of particle interactions and random initial condition makes the proofs technically more involved.

Numerical tests recover the theoretical convergence rates and illustrate a computational complexity advantage of the explicit over the implicit scheme. Comparative analysis is carried out on a stylized non Lipschitz MV-SDE and the neuron network model proposed in [1]. We provide numerical tests illustrating particle corruption effect where one single particle diverging can “corrupt” the whole system. Moreover, the more particles in the system the more likely this divergence is to occur.

Key words. McKean-Vlasov Stochastic Differential Equation, Interacting Particle System, Monte Carlo Simulation, Taming, Implicit and Explicit Schemes, Stochastic Neuron Networks

AMS subject classifications. 65C05 (Monte Carlo methods), 65C30 (Stochastic differential and integral equations), 65C35 (Stochastic particle methods)

1. Introduction. The aim of this paper is to develop a numerical scheme for simulating a McKean-Vlasov Stochastic Differential Equations (MV-SDEs) with drifts of super-linear growth and Lipschitz diffusion coefficients (with linear growth). MV-SDEs differ from standard SDEs by means of the presence of the law of the solution process in the coefficients.

\[
dX_t = b(t, X_t, \mu^X_t) dt + \sigma(t, X_t, \mu^X_t) dW_t, \quad X_0 \in L^m_0(\mathbb{R}^d),
\]

where \( \mu^X_t \) denotes the law of the process \( X_t \) at time \( t \). Similar to standard SDEs, MV-SDEs have been shown to have a unique strong solution in the super-linear growth setting in spatial parameter setting, see [9]. Of course, many mean-field models exhibit non globally Lipschitz growth, for example, mean-field models for neuronal activity (e.g. stochastic mean-field FitzHugh-Nagumo models or the network of Hodgkin-Huxley neurons) [1], [2], [3] appearing in biology or physics [12], [11]. We refer to the review in [1] for further motivation of the problem.

In general closed form solutions for such equations are rare, hence to fully utilize
MV-SDEs as a modeling tool, one needs a reliable way in which to simulate them. It is well known for standard SDEs that the explicit Euler scheme runs into difficulties in the super-linear growth setting, see [16], even though the SDE is known to have a unique strong solution. The original solution to this problem was to consider an implicit (or backwards) Euler scheme developed in [15]. Although implicit schemes allowed one to tackle more general SDEs they are slower especially in higher dimensions. The reason for this boils down to the fact that one is required to solve a fixed point equation at every time-step which can be computationally expensive. To solve this problem an explicit scheme was then developed in [17], a so-called Tamed Euler scheme. Since then several authors have built on this result and developed algorithms to deal with coefficients that grow super-linearly, see [8], [27], [13] for example.

There has been some work on improved Monte Carlo methods for MV-SDEs with super-linear drift, see e.g. [10].

An extra complication MV-SDEs offer over standard SDEs is the requirement to approximate $\mu$ at each time step. Although there are other techniques (see [14]) the most common is a so-called interacting particle system,

$$dX_t^{i,N} = b(t, X_t^{i,N}, \mu_t^{X,N})dt + \sigma(t, X_t^{i,N}, \mu_t^{X,N})dW_t^i,$$

where $\mu_t^{X,N}(dx) := \frac{1}{N} \sum_{j=1}^N \delta_{X_j^{i,N}}(dx)$ and $\delta_{X_j^{i,N}}$ is the Dirac measure at point $X_j^{i,N}$, and the independent Brownian motions $W_t^i, i = 1, \ldots, N$. Under Lipschitz type conditions this particle system is known to converge pathwise to the true solution of the MV-SDE. However, this convergence (with corresponding rate) in super-linear growth setting has thus far not been considered in full generality.

Closer to our work, we highlight: [5] develop an explicit Euler scheme to deal with a specific MV-SDE type equation; convergence is given but under Lipschitz conditions and constant diffusion coefficient. [21] studies an implicit Euler scheme in order to approximate a specific equation and requires constant diffusion coefficient, symmetry and uniform convexity of the interaction potential.

Our contribution. Firstly, we show that the above particle scheme converges in the super-linear growth case without coercivity/dissipativity (propagation of chaos). This result is crucial in showing convergence of the numerical scheme to the particle system rather than to the original MV-SDE, with corresponding rate.

The second contribution is the development and strong convergence of the explicit scheme to the MV-SDE, inspired by the explicit scheme originally developed in [17], [27]. We also obtain the classical 1/2 rate of convergence in the stepsize. Combining this with the propagation of chaos result gives an overall convergence rate for the explicit scheme.

The final contribution is to show strong convergence of an implicit scheme. This turns out to be a challenging problem since results involving implicit schemes rely on stopping time arguments. This causes several issues when generalizing results to the MV-SDE setting and we have had to make stronger assumptions on the coefficients in this setting in order for the arguments to continue to hold. On the other hand, we allow for random initial conditions and time dependent coefficients that to the best of our knowledge have not been fully treated in the standard SDE setting. We discuss these issues in Remarks 3.4 and 5.10. We only focus on strong convergence of this scheme and not the rate, mainly because the explicit scheme is in general superior (as our numerical testing shows) and such proof would lead to lengthy statements below without substantially enhancing the scope of our work.

From a technical point of view, we highlight the successful use of stopping time
arguments in combination with McKean-Vlasov equations and associated particle systems to show the convergence of the implicit scheme.

The paper is structured in the following way. In Section 2 we introduce the notation and our tamed particle scheme. In Section 3, we state our main result, namely, propagation of chaos and convergence results for the two schemes. Following that, in Section 4 we provide several numerical examples and highlight the particle corruption phenomena. This analysis implies one cannot hope to build a reliable scheme based on a standard Euler scheme. We further show the increased computational complexity associated with a MV-SDE makes the implicit scheme a less viable option than the explicit (tamed) scheme. Finally, the proofs are given in Section 5 and Appendix.

2. Preliminaries. Throughout the paper we work on a filtered probability space \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})\) satisfying the usual conditions, where \(\mathcal{F}_t\) is the augmented filtration of a standard multidimensional Brownian motion \(W\). We will work with \(\mathbb{R}^d\), the \(d\)-dimensional Euclidean space of real numbers, and for \(a = (a_1, \ldots, a_d) \in \mathbb{R}^d\) and \(b = (b_1, \ldots, b_d) \in \mathbb{R}^d\) we denote by \(|a|^2 = \sum_{i=1}^{d} a_i^2\) the usual Euclidean distance on \(\mathbb{R}^d\) and by \((a, b) = \sum_{i=1}^{d} a_ib_i\) the usual scalar product. For matrices \(V \in \mathbb{R}^{k \times l}\) we define \(|V| = \sup_{u \in \mathbb{R}^k, |u| \leq 1} |Vu|\).

We consider some finite terminal time \(T < \infty\) and let \(b = (b_1, \ldots, b_d) \in \mathbb{R}^d\) satisfy
\[
|b|^2 = \sum_{i=1}^{d} b_i^2 < \infty,
\]
for all \(t \in [0, T]\) and let \(\mu\) be a standard Euler scheme. We further show the increased computational complexity associated with a MV-SDE makes the implicit scheme a less viable option than the explicit (tamed) scheme. Finally, the proofs are given in Section 5 and Appendix.

2. Preliminaries. Throughout the paper we work on a filtered probability space \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})\) satisfying the usual conditions, where \(\mathcal{F}_t\) is the augmented filtration of a standard multidimensional Brownian motion \(W\). We will work with \(\mathbb{R}^d\), the \(d\)-dimensional Euclidean space of real numbers, and for \(a = (a_1, \ldots, a_d) \in \mathbb{R}^d\) and \(b = (b_1, \ldots, b_d) \in \mathbb{R}^d\) we denote by \(|a|^2 = \sum_{i=1}^{d} a_i^2\) the usual Euclidean distance on \(\mathbb{R}^d\) and by \((a, b) = \sum_{i=1}^{d} a_ib_i\) the usual scalar product. For matrices \(V \in \mathbb{R}^{k \times l}\) we define \(|V| = \sup_{u \in \mathbb{R}^k, |u| \leq 1} |Vu|\).

Given a measurable space \((\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))\), we denote by \(\mathcal{P}(\mathbb{R}^d)\) the set of probability measures on this space, and write \(\mu \in \mathcal{P}_2(\mathbb{R}^d)\) if \(\mu \in \mathcal{P}(\mathbb{R}^d)\) and for some \(x \in \mathbb{R}^d\), \(\int_{\mathbb{R}^d} |x - y|^2 \mu(dy) < \infty\). We then have the following metric on the space \(\mathcal{P}_2(\mathbb{R}^d)\) (Wasserstein metric) for \(\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)\) (see [9]),
\[
W^{(2)}(\mu, \nu) = \inf_{\pi} \left\{ \left( \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \pi(dx, dy) \right)^{1/2} : \pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) \text{ with marginals } \mu \text{ and } \nu \right\}.
\]

2.1. McKean-Vlasov stochastic differential equations. Let \(W\) be an \(l\)-dimensional Brownian motion and take the progressively measurable maps \(b : [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d\) and \(\sigma : [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^{d \times l}\). MV-SDEs are typically written in the form,
\[
\begin{align*}
\mathrm{d}X_t &= b(t, X_t, \mu_t^X) \mathrm{d}t + \sigma(t, X_t, \mu_t^X) \mathrm{d}W_t, \\
X_0 &= \mathcal{L}_0^p(\mathbb{R}^d),
\end{align*}
\]
where \(\mu_t^X\) denotes the law of the process \(X\) at time \(t\), i.e. \(\mu_t^X = \mathbb{P} \circ X_t^{-1}\). We make the following hypothesis on the coefficients throughout.

**Hypothesis 2.1.** Assume that \(\sigma\) is Lipschitz in the sense that there exists \(L > 0\) such that for all \(t \in [0, T]\) and all \(x, x' \in \mathbb{R}^d\) and \(\forall \mu, \mu' \in \mathcal{P}_2(\mathbb{R}^d)\) we have that
\[
|\sigma(t, x, \mu) - \sigma(t, x', \mu')| \leq L(|x - x'| + W^{(2)}(\mu, \mu')),
\]
and let \(b\) satisfy

1. One-sided Lipschitz in \(x\) and Lipschitz in law: there exist \(L_b, L > 0\) such that for all \(t \in [0, T]\), all \(x, x' \in \mathbb{R}^d\) and all \(\mu, \mu' \in \mathcal{P}_2(\mathbb{R}^d)\) we have that
\[
\begin{align*}
&\langle x - x', b(t, x, \mu) - b(t, x', \mu) \rangle \leq L_b |x - x'|^2 \\
&\quad \text{and } |b(t, x, \mu) - b(t, x, \mu')| \leq LW^{(2)}(\mu, \mu').
\end{align*}
\]
2. Locally Lipschitz with polynomial growth in $x$: there exists $q \in \mathbb{N}$ with $q > 1$ such that for all $t \in [0, T]$, $\forall \mu \in \mathcal{P}_2(\mathbb{R}^d)$ and all $x, \ x' \in \mathbb{R}^d$

$$|b(t, x, \mu) - b(t, x', \mu)| \leq L(1 + |x|^q + |x'|^q)|x - x'|.$$

**Hypothesis 2.2.** Assume that $b$ and $\sigma$ are 1/2-Hölder continuous in time.

Using the one-sided Lipschitz drift, a particularized version of [9, Theorem 3.3] provides a result for existence and uniqueness. Hypothesis 2.2 is not needed here.

**Theorem 2.3 ([9]).** Suppose that $b$ and $\sigma$ satisfy Hypothesis 2.1 and 2.2. Further, assume for some $m \geq 2$, $X_0 \in L_0^m(\mathbb{R}^d)$. Then there exists a unique solution for $X \in \mathbb{S}^m([0, T])$ to the MV-SDE (2.1). For some positive constant $C$ we have

$$\mathbb{E}\left[ \sup_{t \in [0, T]} |X_t|^m \right] \leq C \left( \mathbb{E}[|X_0|^m] + 1 \right) e^{CT}.$$

If the law $\mu^X$ is known beforehand, then the MV-SDE reduces to a “standard” SDE with added time-dependency. Typically this is not the case and usually the MV-SDE is approximated by a particle system.

**The interacting particle system approximation.** We approximate (2.1) (driven by the Brownian motion $W$), using an $N$-dimensional system of interacting particles. Let $i = 1, \ldots, N$ and consider $N$ particles $X_{i,N}$ satisfying the SDE with $X_{0,i,N} = X_{0,i}$ (since the initial condition is random, but independent of other particles)

$$dX_{t,i,N} = b\left(t, X_{t,i,N}, \mu_{t,i,N}\right)dt + \sigma\left(t, X_{t,i,N}, \mu_{t,i,N}\right)dW_i,$$

where $\mu_{t,i,N}(dx) := \frac{1}{N} \sum_{j=1}^N \delta_{X_{t,j,N}}(dx)$ and $\delta_{X_{t,j,N}}$ is the Dirac measure at point $X_{t,j,N}$, and the independent Brownian motions $W_i, i = 1, \ldots, N$ (also independent of the BM $W$ appearing in (2.1); with a slight abuse of notation to avoid re-defining the probability space’s Filtration).

**Propagation of chaos.** In order to show that the particle approximation is of use, one shows a pathwise propagation of chaos result. Although different types exist we are interested in strong error hence require a pathwise convergence result where we consider the system of non interacting particles

$$dX_t^i = b(t, X_t^i, \mu_t^X)dt + \sigma(t, X_t^i, \mu_t^X)dW_t^i, \quad X_0^i = X_0^i, \quad t \in [0, T],$$

which are of course just MV-SDEs and since the $X^i$s are independent, then $\mu_t^{X^i} = \mu_t^X$ for all $i$. Under global Lipschitz conditions, one can then prove the following convergence result (see [6, Theorem 1.10] for example)

$$\lim_{N \to \infty} \sup_{1 \leq i \leq N} \mathbb{E}\left[ \sup_{0 \leq t \leq T} |X_{t,N}^i - X_t^i|^2 \right] = 0.$$

All SDEs appearing below have initial condition $X_0^i$ and we work on the interval $[0, T]$.

**Standard Euler scheme particle system.** In general one cannot simulate (2.2) directly and therefore turns to a numerical scheme such as Euler. We partition the time interval $[0, T]$ into $M$ steps of size $h := T/M$, we then define $t_k := kh$ and recursively define the particle system for $k \in \{0, \ldots, M - 1\}$ as,

$$\tilde{X}_{t_{k+1}}^{i,N,M} = \tilde{X}_{t_k}^{i,N,M} + b\left(t_k, \tilde{X}_{t_k}^{i,N,M}, \tilde{\mu}_{t_k}^{X,N}\right)h + \sigma\left(t_k, \tilde{X}_{t_k}^{i,N,M}, \tilde{\mu}_{t_k}^{X,N}\right)\Delta W_{t_k}^i.$$
where \( \bar{\mu}^N_{t_k}(dx) := \frac{1}{N} \sum_{j=1}^{N} \delta_{X_{j,t_k}^N}(dx) \), \( \Delta W_{t_k}^i := W_{t_{k+1}}^i - W_{t_k}^i \) and \( \bar{X}_{0,i}^{1,N,M} := X_{0}^i \).

Under Lipschitz regularity it is well known that this scheme converges, see [4] or [19] (here a weak rate of convergence is shown under an additional regularity assumption).

**Euler particle system for the super-linear case: Explicit and Implicit.**

However, as discussed in works such as [16], [17], [27] one does not have convergence of the Euler scheme when we move away from the global Lipschitz setting. The goal of this paper is to therefore construct a suitable numerical schemes which converges. Inspired by the above works we consider a so-called *tamed* Euler scheme. With the notation above consider the following scheme

\[
\bar{X}_{i,t_{k+1}}^{1,N,M} = \bar{X}_{i,t_k}^{1,N,M} + \frac{b\left(t_k, \bar{X}_{i,t_k}^{1,N,M}, \bar{\mu}_{i,t_k}^N\right)}{1 + M^{-\alpha} b\left(t_k, \bar{X}_{i,t_k}^{1,N,M}, \bar{\mu}_{i,t_k}^N\right)} h + \sigma\left(t_k, \bar{X}_{i,t_k}^{1,N,M}, \bar{\mu}_{i,t_k}^N\right) \Delta W_{t_k}^i,
\]

where \( \bar{\mu}_{i,t_k}^N(dx) = \frac{1}{N} \sum_{j=1}^{N} \delta_{\bar{X}_{i,t_k}^N}(dx) \) and \( \alpha \in (0, 1/2] \) with \( \bar{X}_{0,i}^{1,N,M} = X_0^i \).

Of course, explicit schemes are not the only method one can deploy to solve this problem, we also consider the following implicit scheme

\[
\bar{X}_{i,t_{k+1}}^{1,N,M} = \bar{X}_{i,t_k}^{1,N,M} + b\left(t_k, \bar{X}_{i,t_k}^{1,N,M}, \bar{\mu}_{i,t_k}^N\right) h + \sigma\left(t_k, \bar{X}_{i,t_k}^{1,N,M}, \bar{\mu}_{i,t_k}^N\right) \Delta W_{t_k}^i,
\]

where \( \bar{\mu}_{i,t_k}^N(dx) := \frac{1}{N} \sum_{j=1}^{N} \delta_{\bar{X}_{i,t_k}^N}(dx) \) and \( \bar{X}_{0,i}^{1,N,M} = X_0^i \).

### 3. Main Results

We state our main results and hypothesis here, the proofs are postponed to Section 5. Recall that we want to associate a particle system to the MV-SDE and show its convergence, so-called *propagation of chaos*. We have the following result that holds under weaker assumptions than those in Theorem 3.3.

**Proposition 3.1 (Propagation of chaos).** *Let the hypothesis in Theorem 2.3 hold for \( m > 4 \). Then we have the following convergence result.*

\[
\sup_{1 \leq i \leq N} \mathbb{E}\left[ \sup_{0 \leq t \leq T} |X_t^i - X_{t}^{i,N,M}|^2 \right] \leq C \begin{cases} 
N^{-1/2} & \text{if } d < 4, \\
N^{-1/2} \log(N) & \text{if } d = 4, \\
N^{-2/d} & \text{if } d > 4.
\end{cases}
\]

Therefore, to show convergence between our numerical scheme and the MV-SDE, we only need to show that the “true” particle scheme and numerical version of the particle scheme converge.

**Explicit scheme.** We first introduce the continuous time version of the explicit scheme. Denote by \( \eta(t) := \sup\{s \in (0, h, 2h, \ldots, Mh) : s \leq t\} \) for all \( t \in [0, T] \), \( b_M(t, x, \nu) := \frac{b(t, x, \nu)}{1 + b(t, x, \nu)} \) with \( \alpha \in (0, 1/2] \) for all \( t \in [0, T], x \in \mathbb{R}^d, \nu \in \mathcal{P}_2(\mathbb{R}^d) \)

\[
X_{t}^{1,N,M} = X_0^i + \int_0^t b_M \left( \eta(s), X_{\eta(s)}^{1,N,M}, \mu_{\eta(s)}^{X,N,M} \right) ds + \int_0^t \sigma \left( \eta(s), X_{\eta(s)}^{1,N,M}, \mu_{\eta(s)}^{X,N,M} \right) dW_s, \quad \mu_t^{X,N,M}(dx) = \frac{1}{N} \sum_{j=1}^{N} \delta_{X_{\eta(t)}^j,N,M}(dx).
\]

Note that \( |b_M(t, x, \nu)| \leq \min(M^\alpha, |b(t, x, \nu)|) \) and that \( \bar{X}_{t_k}^{1,N,M} = X_{t_k}^{1,N,M} \) for all \( k \in \{0, 1, \ldots, M\} \) and hence \( X_{t}^{1,N,M} \) is a continuous version of \( \bar{X}_{t}^{1,N,M} \) from (2.4). We then obtain the following convergence result.
Proposition 3.2. Let the hypothesis in Theorem 3.3 hold. Then it holds that
\[
\sup_{1 \leq i \leq N} \mathbb{E} \left[ \sup_{0 \leq t \leq T} |X_{t}^{i,N} - X_{t}^{i,N,M}|^{2} \right] \leq C h.
\]

This then leads to our main explicit scheme convergence result.

Theorem 3.3 (Strong Convergence of Explicit). Let Hypothesis 2.1 and 2.2 hold, further let \(X_0 \in L^m(\mathbb{R}^d)\) for \(m \geq 4(1 + q)\) (note \(q > 1\)). Let \(X^i\) be the solution to (2.3), and \(X_{t}^{i,N,M}\) be (3.1). Then we obtain the following convergence result
\[
\sup_{1 \leq i \leq N} \mathbb{E} \left[ \sup_{0 \leq t \leq T} |X_t^i - X_t^{i,N,M}|^2 \right] \leq C \begin{cases} 
N^{-1/2} + h & \text{if } d < 4, \\
N^{-1/2} \log(N) + h & \text{if } d = 4, \\
N^{-2/d} + h & \text{if } d > 4.
\end{cases}
\]

Proof of Theorem 3.3. Theorem 3.3 is a consequence of Propositions 3.1 and 3.2.

Remark 3.4 (Issues using stopping times). The technique of using the stopping time \(\tau_R := \inf\{t \geq 0 : |X_{t}^{i,N,M}| \geq R \}\) to control the particles is suboptimal and several problems appear by introducing them. Namely, one can only consider stopping times that stop one particle since otherwise the convergence speed would decrease with a higher number of particles. However, applying a stopping time to a single particle does not allow us to fully bound the coefficients and moreover destroys the result of all particles being identically distributed.

The stopping times arguments used for the implicit scheme below require stronger assumptions in order to make the theory hold.

Implicit scheme. We have shown convergence of the explicit scheme for non Lipschitz coefficients, although this is indeed not the only method, there is another popular method known as implicit or backward Euler scheme. That being said, the implicit scheme has some well documented disadvantages, namely it is expensive compared to its explicit counterpart, we discuss this issue further in Section 4. One can consult [22] for example on the implicit scheme (and extensions) for standard SDEs.

Standard implicit scheme convergence results rely on the so called monotone growth condition, we therefore proceed with the following hypothesis.

Hypothesis 3.5. (H1). There exists a constant \(C\) such that, for all \(\mu \in \mathcal{P}_2(\mathbb{R}^d),\)
\[
|b(0,0,\mu)| + |\sigma(0,0,\mu)| \leq C.
\]

(H2). \(\sigma\) is only a function of time and space (does not have a measure dependence).

Although the main convergence theorem requires both H1 and H2, we only use H2 at the end of the proof of convergence. We present our auxiliary results requiring only H1 as we believe them to be of general independent interest.

Remark 3.6 (Monotone Growth). The combination of Hypothesis 2.1, 2.2 and H1, imply the monotone growth condition. Namely, there exist constants \(\alpha\) and \(\beta\) such \(\forall t \in [0,T], \mu \in \mathcal{P}_2(\mathbb{R}^d)\) with \(l\) being the dimension of the BM,
\[
\langle x, b(t,x,\mu) \rangle + \frac{1}{2} \sum_{a=1}^{l} |\sigma_a(t,x,\mu)|^2 \leq \alpha + \beta |x|^2.
\]

We now state the strong convergence of the implicit scheme (2.5) to (2.2).
Proposition 3.7. Let Hypothesis 2.1, 2.2 and 3.5 hold. Fix a timestep $h^* < 1 / \max(L_b, 2\beta)$ and assume $X_0 \in L^q(\mathbb{R}^d)$. Then, for any $T = Mh$ and $s \in [1, 2)$

$$\sup_{1 \leq i \leq N} \lim_{h \to 0} \mathbb{E}\left[|X_{T_i}^N - \tilde{X}_{T_i}^{N,M}|^s\right] = 0.$$ 

Theorem 3.8 (Strong Convergence of Implicit Scheme). Let the Hypothesis in Proposition 3.7 hold. Then, for any $T = Mh$ and $s \in [1, 2)$ one has

$$\lim_{N \to \infty} \sup_{1 \leq i \leq N} \lim_{h \to 0} \mathbb{E}\left[|X_{T_i}^N - \tilde{X}_{T_i}^{N,M}|^s\right] = 0.$$ 

Proof. The proof of this result follows by combing Proposition 3.1 and 3.7 and noting that the assertion in Proposition 3.7 is independent of $N$. 

4. Numerical testing and Examples. We illustrate immediately our results with numerical examples. We highlight the issues of using the standard Euler scheme in this setting and also compare the computational time and complexity of the explicit and implicit scheme. We juxtapose our findings to those in [1].

4.1. Particle Corruption. It is well known that the Euler scheme fails (diverges) when one moves outside the realm of linear growing coefficients, see [16]. We claim that this divergence is worse in the setting of MV-SDEs and associated particle system due to an effect we refer to as particle corruption.

The basic idea is that one particle becomes influential on all other particles, thus we are no longer in the setting of “weakly interacting”. This is of course not a problem for standard SDE simulation. We show two aspects of particle corruption in a simple example, firstly it exists i.e. one particle can cause the whole system to crash. Secondly and perhaps more profoundly, the more particles one has the more likely this is. This is of course a devastating issue when simulating a MV-SDE since accurately approximating the measure depends on having a large number of interacting particles.

To show this example we take a classical non-globally Lipschitz SDE, the stochastic Ginzburg Landau equation (see [28]) and add a simple mean field term to it,

$$dX_t = \left(\frac{\sigma^2}{2} X_t - X_t^3 + c\mathbb{E}[X_t]\right)dt + \sigma X_t dW_t, \quad X_0 = x.$$ 

This MV-SDE clearly satisfies the hypothesis to have a unique strong solution in $\mathbb{S}^p$ for all $p > 1$, hence in theory one could calculate $\varphi(t) := \mathbb{E}[X_t]$ and have a standard SDE with one-sided Lipschitz drift. The analysis carried out in [16] then implies that the Euler scheme diverges here.

Showing particle corruption exists. For our example we simulate $N = 5000$ particles with a time step $h = 0.05$, $T = 2$ and $X_0 = 1$, we also take $\sigma = 3/2$ and $c = 1/2$. We rerun this example until we observed a blow up and plotted the particle paths in Figure 1.

Figure 1 show the first part of the divergence, namely all particles are reasonably well behaved until one starts to oscillate rapidly. We have stopped plotting before the time boundary since this particle diverges shortly after this. We refer to this particle as the corrupt particle and it is fairly straightforward to see it will diverge. However, due to the interaction this single particle influences all the remaining particles and the whole system diverges shortly after.

Remark 4.1 (Why is particle corruption so pronounced?). The reason this effect is so dramatic is a simple consequence of the mean-field interaction. Typically, one
observes divergence of the Euler scheme via a handful of Monte Carlo simulations that return extremely large (or infinite) values. When one then looks to calculate the expected value of the SDEs at the terminal time for example, these few events completely dominate the other results. This is summed up in a statement of \([16]\), where an exponentially small probability event has a double exponential impact.

The difference in the MV-SDE (weakly interacting particle) case is that the expectation appears inside the simulation, hence a divergence of a single particle influences multiple particles simultaneously during the simulation and not just at the final time.

Convergence of Euler and propagation of chaos is impossible. The above shows that one particle diverging can cause the whole system to diverge, one may argue that using more particles would reduce the dependency between them and hence influence the system less. In fact as we shall see the opposite is true, the more particles the more likely a divergence is. To test this we use the same example as above but use \(N = [1000, 5000, 10000, 20000]\) particles and rerun each case 1000 times and record the total number of times we observe a divergence over the ensemble.

| Number of particles | 1000 | 5000 | 10000 | 20000 |
|---------------------|-----|------|--------|--------|
| Number of blow ups   | 3   | 32   | 43     | 108    |

Table 1: Number of divergences recorded at each particle level out of 1000 simulations.

The results in Table 1 show conclusively that the more particles the more likely a divergence is to occur. This is a real problem in this setting since in order to minimize the propagation of chaos error one should take \(N\) as large as possible, but in doing so makes the Euler scheme approximation (likelier to) diverge.

**Remark 4.2** (Euler cannot work). We have shown that naively applying the standard Euler scheme in the MV-SDE setting with non globally Lipschitz coefficient has issues. However, for standard SDEs there are some simple fixes one can apply and still obtain convergence e.g. removing paths that leave some ball as considered in [23].
Methods like this cannot work here since, we either take the ball “small” and therefore our approximation to the law is poor. Or we take a large ball, but then as the particles head towards the boundary they can “drag” other particles with them which again makes the system unstable.

The dependence on the measure (other particles) implies that the more crude approximation techniques cannot yield the strong convergence results we obtain with the more sophisticated techniques presented in this paper. In [1] the authors have a non-globally Lipschitz MV-SDE and simulate using standard Euler scheme. Since no divergence was observed in their simulations they conjectured that the Euler scheme works in their setting, however, they used a “small” diffusion coefficient \((\sigma \in [0, 0.5])\) and small particle number (in the order of hundreds), which makes divergence unlikely to be observed (but not impossible) and yields poorer approximation results. Again, our methods provide certainty in terms of convergence (and convergence rate).

4.2. Timing of Implicit vs Explicit: Size of cloud and spatial dimension.

It is well documented that implicit schemes are slower than explicit ones, mainly because one must solve a fixed point equation at each step. This operation is not “cheap” and moreover scales \(d^2\) in dimension, see [17]. Of course this analysis is carried out for standard SDEs, what we wish to consider is how the particle system affects the timing of both methods.

We consider the same example as previous (but take \(T = 1\)), we then consider a set of dimensions from 1 to 200 and number of particles from 100 to 20000. Plotting the time taken for both methods is given in Figure 2.

![Figure 2: Showing how the time (in seconds) of the explicit scheme (left; timescale \(\approx 60\) seconds) and implicit scheme (right; timescale \(\approx 10^4\) seconds) changes with particles and dimension.](image)

Firstly, we observe that the explicit scheme is two to three orders of magnitude faster than the implicit scheme. At the highest dimensional and particle number this difference is very apparent with the tamed scheme taking approximately 1 minute and the implicit 10 hours. Another note to make is the scaling of each method, both methods scale similarly with particle number, but the tamed scheme scales linearly with dimension, this is superior to the \(d^2\) scaling of the implicit scheme.

Even for the case \(d = 1\), \(N = 20000\) the tamed scheme takes approximately 7 seconds while the implicit scheme takes approximately 23 minutes. For many practical applications \(N = 20000\) is not enough for an acceptable level of accuracy, with this in mind and the dimension scaling, this makes the implicit scheme a very expensive method in this setting.
4.3. Explicit Vs Implicit Convergence: the Neuron Network Model.

We compare the convergence of the explicit and the implicit scheme. To this end we use the system in [1] where the authors develop a non globally Lipschitz MV-SDE to model neuron activity. In our notation their system with \( b : [0, T] \times \mathbb{R}^3 \times \mathcal{P}_2(\mathbb{R}^3) \rightarrow \mathbb{R}^3 \), \( \sigma : [0, T] \times \mathbb{R}^3 \times \mathcal{P}_2(\mathbb{R}^3) \rightarrow \mathbb{R}^{3 \times 3} \) reads for \( x = (x_1, x_2, x_3), z = (z_1, z_2, z_3) \in \mathbb{R}^3 \) as

\[
\begin{align*}
\sigma(t, x, \mu) &:= \left( \begin{array}{c}
\sigma_{ext} & 0 & -\int_{R^3} \sigma_f(x_1 - V_{rev}) z_3 d\mu(z) \\
0 & 0 & 0 \\
0 & \sigma_{32}(x) & 0 
\end{array} \right) \\
\sigma_{32}(x) &:= \mathbb{1}_{\{x_3 \in (0, 1)\}} \frac{a_r T_{max}(1 - x_3)}{1 + \exp(-\lambda(x_1 - V_T))} + a_d x_3 \Gamma \exp(-\Lambda/((1 - 2x_3 - 1)^2)),
\end{align*}
\]

\( T = 2 \) is chosen as the final time and

\[
X_0 \sim \mathcal{N} \left( \left( \begin{array}{c} V_0 \\
w_0 \\
y_0 \end{array} \right), \left( \begin{array}{ccc} \sigma_{V_0} & 0 & 0 \\
0 & \sigma_{w_0} & 0 \\
0 & 0 & \sigma_{y_0} \end{array} \right) \right),
\]

where the parameters have the values

\[
\begin{align*}
V_0 &= 0 & \sigma_{V_0} &= 0.4 & a &= 0.7 & b &= 0.8 & c &= 0.08 & I &= 0.5 & \sigma_{ext} &= 0.5 \\
w_0 &= 0.5 & \sigma_{w_0} &= 0.4 & V_{rev} &= 1 & a_r &= 1 & a_d &= 1 & T_{max} &= 1 & \lambda &= 0.2 \\
y_0 &= 0.3 & \sigma_{y_0} &= 0.05 & J &= 1 & \sigma_f &= 0.2 & V_T &= 2 & \Gamma &= 0.1 & \Lambda &= 0.5.
\end{align*}
\]

As the true solution is unknown to compare the convergence rates, we use as proxy the output of the explicit scheme with \( 2^{23} \) steps. Since the explicit scheme has convergence rate \( \sqrt{h} \) we know that \( 2^{16} \) steps and below yields one order of magnitude larger errors. The simulation for 1000 particles and average root mean square error of each particle is given in Figure 3.

One can observe that although initially the implicit scheme has a better rate of convergence, it levels off to yield the expected 1/2 rate. Making the explicit scheme the more computationally efficient. Of course our “true” was calculated from the explicit scheme, hence we additionally carried out a similar test with a “true” from the implicit, and the results were almost identical.

Remark 4.3 (Small Diffusion Setting). Above, we have taken \( \sigma_{ext} = 0.5 \), this goes against the example in [1] where \( \sigma_{ext} = 0 \). As it turns out, in the case \( \sigma_{ext} = 0 \), the implicit scheme has a convergence rate close to 1 (up to an error of around \( 10^{-4} \)), while the explicit scheme maintains the standard 1/2 rate. It is our belief that this is due to the fact that when \( \sigma_{ext} = 0 \) the diffusion coefficient makes little difference, hence both scheme revert close to their deterministic convergence rate. The explicit scheme of course still rate of order 1/2, while the implicit is order 1. It may therefore be that in the setting of small diffusion terms the implicit can yield superior results, of course though this is a special case and is not true in general.
Figure 3: Root mean square error of the explicit and implicit. The number of steps of the explicit scheme are $M \in \{2^2, 2^3, \ldots, 2^{16}\}$ and of the implicit scheme are $M \in \{2^2, 2^3, \ldots, 2^{11}\}$. We used 1000 particles and the true is calculated from the explicit with $2^{23}$ steps. Both schemes converge with rate $1/2$.

Figure 4: Approximate density of the first and second component of the MV-SDE at time $T = 1.2$. We used 10000 particles, $2^{20}$ steps and a bandwidth of 0.15 in the kernel smoothing.

**Obtaining the Density.** In some applications as well as the value of the MV-SDE at the terminal time, one may also be interested in the density (law). In [1, Section 4] the authors compare density estimation using both the Fokker-Plank equation and the histogram from the particle system. The approach using PDEs becomes computationally expensive here if one consider multiple populations of MV-SDE and hence the authors take a simple case (see [1, Section 4.3]). There are of course other drawbacks such as dimension scaling which often make stochastic techniques more favorable in this setting. Moreover, using the PDE one will only obtain the density, if one is further interested in calculating a “payoff” i.e. $\mathbb{E}[G(X_T)]$ for some function $G$. 
Then we would require an additional integral approximation or Metropolis Hastings style sampling scheme to calculate this expectation. While [1] apply a basic histogram approach when using MV-SDEs, this does not yield particularly nice results, namely, the resultant density is not a smooth surface. There are however, many statistical techniques one can use to improve this, see [18, Chapter 18.4] for further results and discussion. Taking the example in [1] (with $\sigma_{\text{ext}} = 0$) and applying MATLAB’s \texttt{kdensity} function we obtain Figure 4.

One can observe the similarity between our result using SDEs and the one obtained in [1, pg 31] using the (expensive) PDE approach.

**Conclusions and future work.** We have shown how one can apply the techniques from SDEs to the MV-SDE setting and some of its pitfalls and challenges that arise. The numerical testing carried out shows that the explicit scheme yields superior results (over the implicit scheme) in general.

Although we have been able to obtain convergence for the implicit scheme it is under stronger assumptions than the explicit scheme (the implicit scheme works very well in Section 4.3). The reason for these assumptions is that the implicit scheme is more challenging to bound than the explicit. The standard approach around this problem is to use stopping time arguments, however, as described in Remark 3.4 stopping times are harder to handle in the MV-SDE framework. Caution is needed to account for the extra technicalities that arise.

It is our belief that Hypothesis 3.5 although sufficient, is not necessary to guarantee the implicit scheme converges. As research is carried out into stopping times and MV-SDEs, future theoretical developments in this direction may allow this hypothesis to be weakened. We also leave open a proof for the convergence rate of the implicit scheme. Showing such a convergence rate in our framework is clearly possible but adds little in scope given the gains of the explicit over the implicit scheme. We leave the question open until a time a more resourceful implicit scheme can be designed.

Another interesting area which we have not discussed is sign preservation and the impact it has on the law. For example a MV-SDE may be known to be positive, however, if the numerical scheme takes the solution into the negative region how does the law dependence influence the remaining particles? One can consider the special case of $L_b < 0$ in Hypothesis 2.1, even though the MV-SDE could have a nonnegative solution, the numerical scheme may not preserve this feature.

5. **Proof of Main Results.** We shall use $C$ to denote a constant that can changes from line to line, but only depend on known quantities, $T$, $d$, the one-sided Lipschitz coefficients etc.

5.1. **Propagation of Chaos.** Let us show the propagation of chaos result.

**Proposition 3.1.** Let us fix $1 \leq i \leq N$, we then approach the proof in the usual way for dealing with one-sided Lipschitz coefficients, namely we apply Itô’s formula to the difference (note $X_0^i$ cancel),

$$
\begin{align*}
|X_t^i - X_t^{i,N}|^2 &= \int_0^t 2\langle X_s^i - X_s^{i,N}, b(s, X_s^i, \mu_s) - b(s, X_s^{i,N}, \bar{\mu}_s^N) \rangle ds \\
&\quad + \int_0^t 2\langle X_s^i - X_s^{i,N}, (\sigma(s, X_s^i, \mu_s) - \sigma(s, X_s^{i,N}, \bar{\mu}_s^N)) \rangle dW_s^i \\
&\quad + \sum_{a=1}^d \int_0^t |\sigma_a(s, X_s^i, \mu_s) - \sigma_a(s, X_s^{i,N}, \bar{\mu}_s^N)|^2 ds,
\end{align*}
$$

(5.1)
where \( \sigma_a \) is the \( a \)th column of matrix \( \sigma \), hence \( \sigma_a \) is a \( d \)-dimensional vector. Considering the first integral in (5.1),

\[
\begin{align*}
\langle X_s^i - X_s^{i,N}, b(s, X_s^i, \mu_s) - b(s, X_s^{i,N}, \bar{\mu}_s^N) \rangle \\
= \langle X_s^i - X_s^{i,N}, b(s, X_s^i, \mu_s) - b(s, X_s^{i,N}, \mu_s) \rangle \\
+ \langle X_s^i - X_s^{i,N}, b(s, X_s^{i,N}, \mu_s) - b(s, X_s^{i,N}, \bar{\mu}_s^N) \rangle.
\end{align*}
\]

Applying the one-sided Lipschitz property in space and \( W^{(2)} \) in measure along with Cauchy-Schwarz we obtain,

\[
\begin{align*}
\langle X_s^i - X_s^{i,N}, b(s, X_s^i, \mu_s) - b(s, X_s^{i,N}, \bar{\mu}_s^N) \rangle \\
\leq C |X_s^i - X_s^{i,N}|^2 + C |X_s^i - X_s^{i,N}| W^{(2)}(\mu_s, \bar{\mu}_s^N).
\end{align*}
\]

As is done in [6], we introduce the empirical measure constructed from the true solution i.e. \( \mu_s^N := \frac{1}{N} \sum_{j=1}^N \delta_{X_j^i} \). Since \( W^{(2)} \) is a metric (see [29, Chapter 6]), we have

\[
W^{(2)}(\mu_s, \bar{\mu}_s^N) \leq W^{(2)}(\mu_s, \mu_s^N) + W^{(2)}(\mu_s^N, \bar{\mu}_s^N).
\]

Since \( \mu_s^N, \bar{\mu}_s^N \) are empirical measures a standard result for Wasserstein metric is

\[
W^{(2)}(\mu_s^N, \bar{\mu}_s^N) \leq \left( \frac{1}{N} \sum_{j=1}^N |X_j^i - X_j^{j,N}|^2 \right)^{1/2}.
\]

We leave the other \( W^{(2)} \) term for the moment and consider the diffusion coefficient in the time integral. Since \( \sigma \) is globally Lipschitz and \( W^{(2)} \) for each \( a \) (by definition \( \sigma_a = \sigma e_a \), with \( e_a \) the basis vector, global Lipschitz follows from our norm).

\[
\begin{align*}
&|\sigma_a(s, X_s^i, \mu_s) - \sigma_a(s, X_s^{i,N}, \bar{\mu}_s^N)|^2 \\
\leq & C (|\sigma_a(s, X_s^i, \mu_s) - \sigma_a(s, X_s^{i,N}, \mu_s)|^2 + |\sigma_a(s, X_s^{i,N}, \mu_s) - \sigma_a(s, X_s^{i,N}, \bar{\mu}_s^N)|^2) \\
\leq & C (|X_s^i - X_s^{i,N}|^2 + W^{(2)}(\mu_s, \bar{\mu}_s^N)^2) \\
\leq & C (|X_s^i - X_s^{i,N}|^2 + \frac{1}{N} \sum_{j=1}^N |X_j^i - X_j^{j,N}|^2 + W^{(2)}(\mu_s, \bar{\mu}_s^N)^2).
\end{align*}
\]

One can note this is independent of \( a \). The final term to bound is the stochastic integral term, to do this though we take supremum and expectation to (5.1)

\[
\begin{align*}
E \left[ \sup_{t \leq [0,T]} |X_t^i - X_t^{i,N}|^2 \right] \\
\leq CE \left[ \sup_{t \leq [0,T]} \int_0^t |X_s^i - X_s^{i,N}|^2 + |X_s^i - X_s^{i,N}| W^{(2)}(\mu_s, \bar{\mu}_s^N) ds \right] \\
+ E \left[ \sup_{t \leq [0,T]} \int_0^t 2 \langle X_s^i - X_s^{i,N}, (\sigma(s, X_s^i, \mu_s) - \sigma(s, X_s^{i,N}, \bar{\mu}_s^N)) \rangle dW_s^i \right] \\
+ C E \left[ \sup_{t \leq [0,T]} \int_0^t |X_s^i - X_s^{i,N}|^2 + \frac{1}{N} \sum_{j=1}^N |X_j^i - X_j^{j,N}|^2 + W^{(2)}(\mu_s, \bar{\mu}_s^N)^2 ds \right].
\end{align*}
\]
For the stochastic integral,
\[
E \left[ \sup_{t \in [0,T]} \int_0^t 2(X_s^i - X_s^{i,N}, (\sigma(s, X_s^i, \mu_s) - \sigma(s, X_s^{i,N}, \mu_s^N))dW_s^i) \right] \\
\leq E \left[ \sup_{t \in [0,T]} \int_0^t 2(X_s^i - X_s^{i,N}, (\sigma(s, X_s^i, \mu_s) - \sigma(s, X_s^{i,N}, \mu_s^N))dW_s^i) \right] \\
\leq CE \left[ \left( \sum_{a=1}^l |\sigma_a(s, X_s^i, \mu_s) - \sigma_a(s, X_s^{i,N}, \mu_s^N)|^2 \right)(X_s^i - X_s^{i,N})^2 ds \right]^{1/2} \\
\leq E \left[ \left( \sup_{t \in [0,T]} |X_t^i - X_t^{i,N}|^2 C \int_0^T \sum_{a=1}^l |\sigma_a(s, X_s^i, \mu_s) - \sigma_a(s, X_s^{i,N}, \mu_s^N)|^2 ds \right) \right]^{1/2},
\]
where we have applied Burkholder–Davis-Gundy to remove the stochastic integral. Using Young’s inequality \(ab \leq a^2/2 + b^2/2\) we can bound this term by,
\[
E \left[ \frac{1}{2} \sup_{t \in [0,T]} |X_t^i - X_t^{i,N}|^2 + C \int_0^T \sum_{a=1}^l |\sigma_a(s, X_s^i, \mu_s) - \sigma_a(s, X_s^{i,N}, \mu_s^N)|^2 ds \right].
\]
Substituting into (5.2) yields,
\[
E \left[ \sup_{t \in [0,T]} |X_t^i - X_t^{i,N}|^2 \right] \\
\leq CE \left[ \sup_{t \in [0,T]} \int_0^t |X_s^i - X_s^{i,N}|^2 + |X_s^i - X_s^{i,N}|W^{(2)}(\mu_s, \mu_s^N)ds \right] \\
+ E \left[ \frac{1}{2} \sup_{t \in [0,T]} |X_t^i - X_t^{i,N}|^2 + C \int_0^T \sum_{a=1}^l |\sigma_a(s, X_s^i, \mu_s) - \sigma_a(s, X_s^{i,N}, \mu_s^N)|^2 ds \right] \\
+ CE \left[ \sup_{t \in [0,T]} \int_0^t |X_s^i - X_s^{i,N}|^2 + \frac{1}{N} \sum_{j=1}^N |X_s^j - X_s^{j,N}|^2 + W^{(2)}(\mu_s, \mu_s^N)^2 ds \right].
\]
Taking the \(\frac{1}{2} \sup_{t \in [0,T]} |X_t^i - X_t^{i,N}|^2\) to the other side, noting that the supremum value over the integrals is \(t = T\) and using the bound for the difference in \(\sigma\) we obtain,
\[
E \left[ \sup_{t \in [0,T]} |X_t^i - X_t^{i,N}|^2 \right] \\
\leq CE \left[ \int_0^T |X_s^i - X_s^{i,N}|^2 + |X_s^i - X_s^{i,N}|W^{(2)}(\mu_s, \mu_s^N)ds \right] \\
+ CE \left[ \int_0^T |X_s^i - X_s^{i,N}|^2 + \frac{1}{N} \sum_{j=1}^N |X_s^j - X_s^{j,N}|^2 + W^{(2)}(\mu_s, \mu_s^N)^2 ds \right].
\]
To deal with the summation term, observe that since all \(j\) are identically distributed,
\[
E \left[ \frac{1}{N} \sum_{j=1}^N |X_s^j - X_s^{j,N}|^2 \right] = E[|X_s^i - X_s^{i,N}|^2].
\]
Therefore, applying Young’s inequality to $|X_s^i - X_s^{i,N}|W_t^{(2)}(\mu_s, \bar{\mu}_s^N)$ and taking supremum over $i$,

$$
\sup_{1 \leq i \leq N} \mathbb{E} \left[ \sup_{t \in [0,T]} \left| X_t^i - X_t^{i,N} \right|^2 \right] \leq C \int_0^T \sup_{1 \leq i \leq N} \mathbb{E} \left[ \left| X_t^i - X_t^{i,N} \right|^2 \right] + \mathbb{E} \left[ W_t^{(2)}(\mu_s, \bar{\mu}_s^N)^2 \right] ds
$$

$$
\leq C \int_0^T \mathbb{E} \left[ W_t^{(2)}(\mu_s, \bar{\mu}_s^N)^2 \right] ds,
$$

where the final step follows from Grönwall’s inequality. At this point, one could conclude a pathwise propagation of chaos result, see [6, Lemma 1.9], however, here we are interested in the rate of convergence. This is well understood for $W_t^{(2)}$. We use the improved version [7, Theorem 5.8] of the classical convergence result [26, Chapter 10.2]. Provided $X_t^i \in L^p(\mathbb{R}^d)$ for any $p > 4$, then for any $s$,

$$
\mathbb{E} \left[ W_t^{(2)}(\mu_s, \bar{\mu}_s^N)^2 \right] \leq C \begin{cases} 
N^{-1/2} & \text{if } d < 4, \\
N^{-1/2} \log(N) & \text{if } d = 4, \\
N^{-2/d} & \text{if } d > 4.
\end{cases}
$$

Using the result in Theorem 2.3 with our hypothesis then completes the proof. \( \Box \)

**5.2. Proof of Explicit Convergence.** We detail the results to prove Proposition 3.2. To keep expressions as compact as possible for $s \in [0,T]$ we introduce,

$$
\Delta X_s^{i,N,M} := X_s^{i,N} - X_s^{i,N,M}.
$$

Further we use throughout the following result,

$$
\mathbb{E} \left[ \frac{1}{N} \sum_{j=1}^N \left| \Delta X_s^{j,N,M} \right|^2 \right] = \mathbb{E} \left[ \left| \Delta X_s^{i,N,M} \right|^2 \right] = \sup_{1 \leq j \leq N} \mathbb{E} \left[ \left| \Delta X_s^{j,N,M} \right|^2 \right],
$$

which holds because for every $i$ they are identically distributed, without mentioning.

**Lemma 5.1.** Suppose Hypothesis 2.1 and 2.2 are fulfilled and $X_0 \in L^2(\mathbb{R}^d)$, then there exists a constant $C$ which is independent of $N$ such that

$$
\sup_{M \geq 1} \sup_{1 \leq i \leq N} \sup_{0 \leq t \leq T} \mathbb{E} \left[ \left| X_t^{i,N,M} \right|^2 \right] < C.
$$

**Proof.** Applying Itô’s formula and restructuring the terms gives

$$
\left| X_t^{i,N,M} \right|^2 = \left| X_0^i \right|^2 + \int_0^t 2 \langle X_t^{i,N,M}, b_M(\kappa(s), X_t^{i,N,M}, \mu_{\kappa(s)}) \rangle ds
$$

$$
+ \sum_{a=1}^l \sigma_a(\kappa(s), X_t^{i,N,M}, \mu_{\kappa(s)})^2 ds
$$

$$
+ \int_0^t 2 \langle X_t^{i,N,M}, \sigma(\kappa(s), X_t^{i,N,M}, \mu_{\kappa(s)}) \rangle dW_s^i
$$

$$
+ \int_0^t 2 \langle X_t^{i,N,M} - X_t^{i,N,M}, \sigma(\kappa(s), X_t^{i,N,M}, \mu_{\kappa(s)}) \rangle ds.
$$
Observe that
\[
\left| \mathbb{E} \left[ \int_0^t (X^{i,N,M}_s - X^{i,N,M}_t) \, dB_M \left( \kappa(s), X^{i,N,M}_t, \mu^{X,N,M}_t \right) \right] \right|
\]
\[
\leq \mathbb{E} \left[ \int_0^t \left| \int_{\kappa(s)}^s b_M \left( \kappa(r), X^{i,N,M}_{\kappa(r)}, \mu^{X,N,M}_{\kappa(r)} \right) \, dr \right| \right. \\
\quad + \left. \int_{\kappa(s)}^s \sigma \left( \kappa(r), X^{i,N,M}_{\kappa(r)}, \mu^{X,N,M}_{\kappa(r)} \right) \, dW^i_r, b_M \left( \kappa(s), X^{i,N,M}_{\kappa(s)}, \mu^{X,N,M}_{\kappa(s)} \right) \right] \right| ds \\
\leq \sum_{k=0}^{M-1} \int_{t_k}^{t_{k+1}} \mathbb{E} \left[ \left| b_M \left( \kappa(s), X^{i,N,M}_{\kappa(s)}, \mu^{X,N,M}_{\kappa(s)} \right) \right| \right. \\
\quad + \left. \int_{t_k}^s \sigma \left( \kappa(r), X^{i,N,M}_{\kappa(r)}, \mu^{X,N,M}_{\kappa(r)} \right) \, dW^i_r \right] \right| ds \\
\leq \mathbb{E} \left[ \int_0^t \left| b_M \left( \kappa(s), X^{i,N,M}_{\kappa(s)}, \mu^{X,N,M}_{\kappa(s)} \right) \right| \right. \\
\quad + \left. \int_{t_k}^s \sigma \left( \kappa(r), X^{i,N,M}_{\kappa(r)}, \mu^{X,N,M}_{\kappa(r)} \right) \, dW^i_r \right] \right| ds \\
\leq t M^{2(\alpha-1)} \\
\leq t.
\]

Putting this together and using Hypothesis 2.1 and 2.2 we obtain
\[
\mathbb{E} \left[ |X^{i,N,M}_t|^2 \right]
\]
\[
\leq \mathbb{E} \left[ |X^{i,N}_0|^2 \right] + C \left( 1 + \mathbb{E} \left[ \int_0^t |X^{i,N,M}_s|^2 + \frac{1}{N} \sum_{j=1}^N |X^{j,N,M}_s|^2 \, ds \right] \right)
\]
\[
\leq \mathbb{E} \left[ |X^{i,N}_0|^2 \right] + C \left( 1 + \mathbb{E} \left[ \sup_{0 \leq u \leq s} \mathbb{E} \left[ |X^{i,N,M}_u|^2 \right] + \frac{1}{N} \sum_{j=1}^N |X^{j,N,M}_u|^2 \, ds \right] \right),
\]

which furthermore yields
\[
\sup_{1 \leq i \leq N} \sup_{0 \leq u \leq t} \mathbb{E} \left[ |X^{i,N,M}_t|^2 \right]
\]
\[
\leq C \left( 1 + \mathbb{E} \left[ |X^{i}_0|^2 \right] + \int_0^t \sup_{1 \leq i \leq N} \sup_{0 \leq u \leq s} \mathbb{E} \left[ |X^{i,N,M}_u|^2 \right] \, ds \right) < \infty,
\]

and hence by Grönwall's lemma
\[
\sup_{1 \leq i \leq N} \sup_{0 \leq u \leq t} \mathbb{E} \left[ |X^{i,N,M}_t|^2 \right] < C,
\]

where C is a constant which is independent of N and M.

\[
\text{Lemma 5.2. If Hypothesis 2.1 and 2.2 are fulfilled and } X_0 \in L^2(\mathbb{R}^d), \text{ then for all } p \in (0, 2) \text{ we have}
\]
\[
\left| X^{i,N,M}_t - X^{i,N,M}_{\kappa(t)} \right|^p \leq CM^{-p/2},
\]

(5.3)
and

\[
\sup_{1 \leq i \leq N} \sup_{0 \leq t \leq T} \mathbb{E} \left[ \left| X_{t}^{i,N,M} - X_{\kappa(t)}^{i,N,M} \right|^p \right] \leq C,
\]

where \( C \) is a positive constant independent of \( N \) and \( M \). Furthermore, if \( p > 2 \)

\[
\sup_{M \geq 1} \sup_{1 \leq i \leq N} \mathbb{E} \left[ \sup_{0 \leq t \leq T} \left| X_{t}^{i,N,M} \right|^p \right] < \infty,
\]

then the estimates (5.3) and (5.4) hold for those \( p \) as well.

Proof of Lemma 5.2. Using Hölder’s inequality we obtain for any \( p \geq 2 \)

\[
\left| \int_{\kappa(t)}^t b_M \left( \kappa(s), X_{\kappa(s)}^{i,N,M}, \mu_{\kappa(s)}^{X,N,M} \right) \, ds \right|^p \leq \left( \int_{\kappa(t)}^t \left| 1 \right|^{\frac{p}{p-1}} \, ds \right)^{p-1} \left( \int_{\kappa(t)}^t \left| b_M \left( \kappa(s), X_{\kappa(s)}^{i,N,M}, \mu_{\kappa(s)}^{X,N,M} \right) \right|^p \, ds \right)^{\frac{1}{p}} \leq \left( \frac{T}{M} \right)^{p-1} \frac{T}{M} M^{p\alpha} \leq T^p M^{-p/2},
\]

since \( |b_M| \leq M^{\alpha} \) and \( \alpha \leq 1/2 \). It is easy to see that in the case of \( p \in (0, 2] \)

\[
\mathbb{E} \left[ \left| X_{t}^{i,N,M} - X_{\kappa(t)}^{i,N,M} \right|^p \right] \leq \mathbb{E} \left[ \left| \int_{\kappa(t)}^t b_M \left( \kappa(s), X_{\kappa(s)}^{i,N,M}, \mu_{\kappa(s)}^{X,N,M} \right) \, ds + \int_{\kappa(t)}^t \sigma(\kappa(s), X_{\kappa(s)}^{i,N,M}, \mu_{\kappa(s)}^{X,N,M}) \, dW_s \right|^2 \right]^{\frac{p}{2}} \leq 2^{p/2} \mathbb{E} \left[ \left| \int_{\kappa(t)}^t b_M \left( \kappa(s), X_{\kappa(s)}^{i,N,M}, \mu_{\kappa(s)}^{X,N,M} \right) \, ds \right|^2 \right]^{\frac{p}{2}} + \left| \int_{\kappa(t)}^t \sigma(\kappa(s), X_{\kappa(s)}^{i,N,M}, \mu_{\kappa(s)}^{X,N,M}) \, dW_s \right|^2 \right]^{\frac{p}{2}},
\]

and due to Itô’s isometry and Lemma 5.1 for \( C \) independent of \( M \) and \( i \)

\[
\mathbb{E} \left[ \left| \int_{\kappa(t)}^t \sigma \left( \kappa(s), X_{\kappa(s)}^{i,N,M}, \mu_{\kappa(s)}^{X,N,M} \right) \, dW_s \right|^2 \right] \leq \mathbb{E} \left[ \int_{\kappa(t)}^t K \left( 1 + \left| X_{\kappa(s)}^{i,N,M} \right|^2 + \frac{1}{N} \sum_{j=1}^N \left| X_{\kappa(s)}^{j,N,M} \right|^2 \right) \, ds \right] \leq \sup_{1 \leq i \leq N} \sup_{s \in [\kappa(t), T]} \mathbb{E} \left[ \frac{T}{M} K \left( 1 + \left| X_{s}^{i,N,M} \right|^2 + \left| X_{s}^{i,N,M} \right|^2 \right) \right] \leq CM^{-1},
\]

which gives us combined with (5.5) that

\[
\sup_{0 \leq t \leq T} \mathbb{E} \left[ \left| X_{t}^{i,N,M} - X_{\kappa(t)}^{i,N,M} \right|^p \right] \leq CM^{-p/2},
\]
for all $p \in (0, 2]$. If additionally $\sup_{M \geq 1} \sup_{1 \leq i \leq N} \mathbb{E}\left[ \sup_{0 \leq t \leq T} |X_{t}^{i,N,M}|^{p} \right] < \infty$ for some $p > 2$, then

$$
\mathbb{E}\left[ |X_{t}^{i,N,M} - X_{\kappa(t)}^{i,N,M}|^{p} \right] \\
\leq C \mathbb{E}\left[ \int_{\kappa(t)}^{t} b_{M}(\kappa(s), X_{\kappa(s)}^{i,N,M}, \mu_{\kappa(s)}^{X,N,M}) ds \right]^{p} + \mathbb{E}\left[ \int_{\kappa(t)}^{t} \sigma(\kappa(s), X_{\kappa(s)}^{i,N,M}, \mu_{\kappa(s)}^{X,N,M}) dW_{s}^{i} \right]^{p} \\
\leq C \mathbb{E}\left[ T^{p} M^{p/2} + \int_{\kappa(t)}^{t} \sigma(\kappa(s), X_{\kappa(s)}^{i,N,M}, \mu_{\kappa(s)}^{X,N,M})^{2} ds \right]^{p/2},
$$

by the estimate (5.5) and the Burkholder-Davis-Gundy inequality. Since furthermore,

$$
\mathbb{E}\left[ \int_{\kappa(t)}^{t} \sigma(\kappa(s), X_{\kappa(s)}^{i,N,M}, \mu_{\kappa(s)}^{X,N,M})^{2} ds \right]^{p/2} \\
\leq \mathbb{E}\left[ \left( \frac{T}{M} \right)^{p/2} \sup_{s \in [\kappa(t), t]} \left( 1 + |X_{s}^{i,N,M}|^{p} + \left( \frac{1}{N} \sum_{j=1}^{N} |X_{s}^{j,N,M}|^{2} \right)^{p/2} \right) \right] \\
\leq \left( \frac{T}{M} \right)^{p/2} K \left( 1 + \mathbb{E}\left[ \sup_{0 \leq t \leq T} |X_{t}^{i,N,M}|^{p} \right] + \sup_{1 \leq i \leq N} \mathbb{E}\left[ \sup_{0 \leq t \leq T} |X_{t}^{i,N,M}|^{p} \right] \right) \\
\leq CM^{-p/2},
$$

we get the desired result here as well.

Finally,

$$
\mathbb{E}\left[ \left| X_{t}^{i,N,M} - X_{\kappa(t)}^{i,N,M} \right|^{p} \right] \\
\leq \mathbb{E}\left[ \left| X_{t}^{i,N,M} - X_{\kappa(t)}^{i,N,M} \right|^{p} \right] M^{p_{0}} \leq C,
$$

holds for any $t \in [0, T]$ and $1 \leq i \leq N$ by the former result.

**Lemma 5.3.** Suppose that Hypothesis 2.1 and 2.2 are fulfilled, then for every $p \geq 2$ with $X_{0} \in L^{p}(\mathbb{R}^{d})$ there exists a constant $C$ such that

$$
\sup_{M \geq 1} \sup_{1 \leq i \leq N} \mathbb{E}\left[ \sup_{0 \leq t \leq T} \left| X_{t}^{i,N,M} \right|^{p} \right] < C.
$$

**Proof.** Define $\hat{p} := \sup \{ p \geq 2 : \sup_{M \geq 1} \sup_{1 \leq i \leq N} \mathbb{E}\left[ \sup_{0 \leq t \leq T} \left| X_{t}^{i,N,M} \right|^{p} \right] < \infty \}$, where we use the convention $\sup\{\emptyset\} = -\infty$. Also we assume that $\hat{p} \geq 2$ since otherwise there is nothing to prove. Note that we can already apply Lemma 5.2 for $p \leq 2$.

We use an inductive argument and start with $p = 2$. In every step we set $q = 2p \wedge \hat{p}$.
By Itô’s formula we have
\[
\mathbb{E} \left[ \sup_{0 \leq s \leq t} |X^{i,N,M}_s|^q \right] 
\leq C \left( 1 + \mathbb{E} \left[ |X^{i,N,M}_0|^q \right] + \int_0^t \mathbb{E} \left[ |X^{i,N,M}_{\kappa(s)}|^q \right] ds 
+ \int_0^t \mathbb{E} \left[ \left| X^{i,N,M}_s - X^{i,N,M}_{\kappa(s)} \right|^{q/2} \sigma \left( \kappa(s), X^{i,N,M}_{\kappa(s)}, \mu_{\kappa(s)} \right) \right] ds 
\right)
\]
and the application of the Burkholder-Davis-Gundy inequality and Lemma 5.2 for \( p = q/2 \) yields
\[
\mathbb{E} \left[ \sup_{0 \leq s \leq t} |X^{i,N,M}_s|^q \right] \leq C \left( 1 + \mathbb{E} \left[ |X^{i,N,M}_0|^q \right] + \int_0^t \mathbb{E} \left[ \sup_{0 \leq u \leq s} |X^{i,N,M}_u|^q \right] ds 
\right)
\]
where \( C \) denotes in each case a constant that is independent of \( M \). With Young’s inequality in the form \( ab \leq \frac{1}{2} a^2 + \frac{1}{2} b^2 \), Hölder’s inequality and the estimate for \( \sigma \) we obtain
\[
\mathbb{E} \left[ \sup_{0 \leq s \leq t} |X^{i,N,M}_s|^q \right] 
\leq C \left( 1 + \mathbb{E} \left[ |X^{i,N,M}_0|^q \right] + \int_0^t \mathbb{E} \left[ \sup_{0 \leq u \leq s} |X^{i,N,M}_u|^q \right] ds + \frac{1}{2C} \mathbb{E} \left[ \sup_{0 \leq s \leq t} |X^{i,N,M}_s|^q \right] 
\right)
\]
and thus the application of Grönwall’s lemma yields that
\[
(5.6) \quad \mathbb{E} \left[ \sup_{1 \leq i \leq N} \sup_{0 \leq s \leq T} |X^{i,N,M}_s|^q \right] < C,
\]
holds for some positive constant \( C \) which is independent of \( N \) and \( M \).
Since (5.6) is proven for \( q \) we can set \( p = q \) and use this result in the next step of the iteration. Since the new \( q \) is at most twice as much as \( p \), Lemma 5.2 can again be applied for \( q/2 \). This iteration gets repeated until \( q = \hat{p} \). \( \square \)

Now we can complete the proof of Proposition 3.2.

**Proof of Proposition 3.2.** Using Itô’s formula we observe,

\[
\left| \Delta X_{i,N,M}^s \right|^2
= \int_0^t 2\langle \Delta X_{i,N,M}^s, b(s, X_{i,N,M}^s, \mu_{s,N,M}) - b_M \left( \kappa(s), X_{i,N,M}^s, \mu_{s,N,M} \right) \rangle \, ds
+ \sum_{a=1}^t \int_0^t \left| \sigma_a (s, X_{i,N,M}^s, \mu_{s,N,M}) - \sigma_a \left( \kappa(s), X_{i,N,M}^s, \mu_{s,N,M} \right) \right|^2 \, ds
+ \int_0^t 2\langle \Delta X_{i,N,M}^s, \left( \sigma (s, X_{i,N,M}^s, \mu_{s,N,M}) - \sigma \left( \kappa(s), X_{i,N,M}^s, \mu_{s,N,M} \right) \right) \rangle \, dW_s^s.
\]

Furthermore observe that

\[
\langle X_{i,N} - X_{i,N,M}^s, b(s, X_{i,N,M}^s, \mu_{s,N,M}) - b_M \left( \kappa(s), X_{i,N,M}^s, \mu_{s,N,M} \right) \rangle
= \langle \Delta X_{i,N,M}^s, b(s, X_{i,N,M}^s, \mu_{s,N,M}) - b_M \left( \kappa(s), X_{i,N,M}^s, \mu_{s,N,M} \right) \rangle
+ \langle \Delta X_{i,N,M}^s, b(s, X_{i,N,M}^s, \mu_{s,N,M}) - b_M \left( \kappa(s), X_{i,N,M}^s, \mu_{s,N,M} \right) \rangle
+ \langle \Delta X_{i,N,M}^s, b(s, X_{i,N,M}^s, \mu_{s,N,M}) - b_M \left( \kappa(s), X_{i,N,M}^s, \mu_{s,N,M} \right) \rangle
+ \langle \Delta X_{i,N,M}^s, b(s, X_{i,N,M}^s, \mu_{s,N,M}) - b_M \left( \kappa(s), X_{i,N,M}^s, \mu_{s,N,M} \right) \rangle
+ \langle \Delta X_{i,N,M}^s, b(s, X_{i,N,M}^s, \mu_{s,N,M}) - b_M \left( \kappa(s), X_{i,N,M}^s, \mu_{s,N,M} \right) \rangle
+ \langle \Delta X_{i,N,M}^s, b(s, X_{i,N,M}^s, \mu_{s,N,M}) - b_M \left( \kappa(s), X_{i,N,M}^s, \mu_{s,N,M} \right) \rangle,
\]

where we estimate every term on the right hand side in the following. Due to Hypothesis 2.1 we have

\[
\langle \Delta X_{i,N,M}^s, b(s, X_{i,N,M}^s, \mu_{s,N,M}) - b_M \left( \kappa(s), X_{i,N,M}^s, \mu_{s,N,M} \right) \rangle \leq L_b \left| \Delta X_{i,N,M}^s \right|^2,
\]

and

\[
\langle \Delta X_{i,N,M}^s, b(s, X_{i,N,M}^s, \mu_{s,N,M}) - b_M \left( \kappa(s), X_{i,N,M}^s, \mu_{s,N,M} \right) \rangle
\leq \left| \Delta X_{i,N,M}^s \right| \left| W^{(2)} \left( \mu_{s,N,M} \right) \right|
\leq \left| \Delta X_{i,N,M}^s \right| \frac{1}{\sqrt{N}} \left( \sum_{j=1}^N \left| \Delta X_{j,N,M}^s \right|^2 \right)^{1/2}
\leq \frac{1}{2} \left| \Delta X_{i,N,M}^s \right|^2 + \frac{1}{2N} \sum_{j=1}^N \left| \Delta X_{j,N,M}^s \right|^2,
\]

and

\[
\langle \Delta X_{i,N,M}^s, b(s, X_{i,N,M}^s, \mu_{s,N,M}) - b_M \left( \kappa(s), X_{i,N,M}^s, \mu_{s,N,M} \right) \rangle
\leq C \left| \Delta X_{i,N,M}^s \right| \left| s - \kappa(s) \right|^{1/2} \leq \frac{1}{2} \left| \Delta X_{i,N,M}^s \right|^2 + CM^{-1}.
\]
Further,

\[
\langle \Delta X^i,N,M \rangle, b \left( \kappa(s), X^i,N,M, \mu X,N,M \right) = b \left( \kappa(s), X^i,N,M, \mu X,N,M \right)
\]

\[
\leq \frac{1}{2} |\Delta X^i,N,M|^2 + \frac{1}{2} |b \left( \kappa(s), X^i,N,M, \mu X,N,M \right) - b \left( \kappa(s), X^i,N,M, \mu X,N,M \right)|^2,
\]

where we can furthermore estimate by using the polynomial growth of \( b \) with rate \( q \), Hölder’s inequality, Lemma 5.3 and Lemma 5.2

\[
\mathbb{E} \left[ \sup_{u \in [0, t]} \int_0^u \left| b \left( \kappa(s), X^i,N,M, \mu X,N,M \right) - b \left( \kappa(s), X^i,N,M, \mu X,N,M \right) \right|^2 ds \right]
\]

\[
\leq \int_0^t \mathbb{E} \left[ L \left( 1 + \left| X^i,N,M \right|^q + \left| X^i,N,M \right|^q \right) \left| X^i,N,M - X^i,N,M \right|^2 \right] ds
\]

\[
\leq \int_0^t \sqrt{CM^{-2}} ds \leq CM^{-1}
\]

since

\[
\sup_{M \geq 1} \sup_{1 \leq i \leq N} \mathbb{E} \left[ \sup_{0 \leq t \leq T} \left| X^i,N,M \right|^{4q} \right] \leq 1 + \sup_{M \geq 1} \sup_{1 \leq i \leq N} \mathbb{E} \left[ \sup_{0 \leq t \leq T} \left| X^i,N,M \right|^{4(1+q)} \right] < \infty.
\]

Again Hypothesis 2.1 yields

\[
\langle \Delta X^i,N,M \rangle, b \left( \kappa(s), X^i,N,M, \mu X,N,M \right) - b \left( \kappa(s), X^i,N,M, \mu X,N,M \right)
\]

\[
\leq |\Delta X^i,N,M| \frac{1}{\sqrt{N}} \left( \sum_{j=1}^N \left| X^j,N,M - X^j,N,M \right|^2 \right)^{1/2}
\]

\[
\leq \frac{1}{2} |\Delta X^i,N,M|^2 + \frac{1}{2} N \sum_{j=1}^N \left| X^j,N,M - X^j,N,M \right|^2,
\]

and the definition of \( b_M \) that

\[
\langle \Delta X^i,N,M \rangle, b \left( \kappa(s), X^i,N,M, \mu X,N,M \right) - b_M \left( \kappa(s), X^i,N,M, \mu X,N,M \right)
\]

\[
\leq \frac{1}{2} |\Delta X^i,N,M|^2 + \frac{1}{2} \left| b \left( \kappa(s), X^i,N,M, \mu X,N,M \right) - b_M \left( \kappa(s), X^i,N,M, \mu X,N,M \right) \right|^2
\]

\[
\leq \frac{1}{2} |\Delta X^i,N,M|^2 + \frac{1}{2} M^{-2\alpha} \left| b \left( \kappa(s), X^i,N,M, \mu X,N,M \right) \right|^4
\]

\[
\leq \frac{1}{2} |\Delta X^i,N,M|^2 + CM^{-2\alpha} \left( 1 + \left| X^i,N,M \right|^{4(1+q)} + \left( \frac{1}{N} \sum_{j=1}^N \left| X^j,N,M \right|^2 \right)^2 \right),
\]

where \( q \) is again the polynomial growth rate of \( b \). Also the Burkholder-Davis-Gundy
inequality yields
\[
E \left[ \sup_{u \in [0,t]} \int_0^u 2 (\Delta X_{i,N}^{i,N,M} - \sigma(s, X_{i,N}^{i,N,M}, \mu_{\kappa(s)}, X_{\kappa(s)}, M) \right) \right. dW^i_s] \\
\leq E \left[ (C \int_0^1 \left( \sum_{a=1}^1 [\sigma_a(s, X_{i,N}^{i,N,M}, \mu_{\kappa(s)} - \sigma_a(s, X_{\kappa(s)}, M)]^2 |\Delta X_{i,N}^{i,N,M}|^2 ds \right)^{\frac{1}{2}} \right. \\
\leq E \left[ \frac{1}{2} \sup_{u \in [0,t]} |\Delta X_{i,N}^{i,N,M}|^2 \right. \\
+ C \int_0^u \sum_{a=1}^1 |\sigma_a(s, X_{i,N}^{i,N,M}, \mu_{\kappa(s)} - \sigma_a(s, X_{\kappa(s)}, M)|^2 |\Delta X_{i,N}^{i,N,M}|^2 ds].
\]
and
\[
|\sigma_a(s, X_{i,N}^{i,N,M}, \mu_{\kappa(s)} - \sigma_a(s, X_{\kappa(s)}, M)|^2 \\
\leq C |s - \kappa(s)|^{2\beta_s} + C |X_{i,N}^{i,N,M}|^2 + CW(2) \left( \mu_{s, X_{\kappa(s)}, M} \mu_{X_{\kappa(s)}, M} \right)^2 \\
\leq CM^{-1} + C |X_{i,N}^{i,N,M} - X_{\kappa(s)}^{i,N,M}|^2 \\
+ C \sum_{j=1}^N \sum_{s} \left( |\Delta X_{i,N}^{i,N,M}|^2 + |\Delta X_{i,N}^{i,N,M}|^2 \right).
\]
By putting this together we obtain
\[
E \left[ \sup_{0 \leq u \leq t} |\Delta X_{i,N}^{i,N,M}|^2 \right] \\
\leq CE \left[ \int_0^t |\Delta X_{i,N}^{i,N,M}|^2 \right. \\
+ \frac{1}{2} \sum_{j=1}^N |X_{i,N}^{i,N,M} - X_{\kappa(s)}^{i,N,M}|^2 + M^{-1} + \frac{1}{N} \sum_{j=1}^N |\Delta X_{i,N}^{i,N,M}|^2 \\
+ |X_{i,N}^{i,N,M} - X_{\kappa(s)}^{i,N,M}|^2 + b(s, X_{i,N}^{i,N,M}, \mu_{s, X_{\kappa(s)}, M}) - b(s, X_{\kappa(s)}, M, X_{\kappa(s)}, M)^2 \\
+ M^{-2\alpha} \left( 1 + |X_{\kappa(s)}^{i,N,M}|^{4(1+q)} \right) + M^{-2\alpha} \left( \frac{1}{N} \sum_{j=1}^N |X_{\kappa(s)}^{i,N,M}|^2 \right)^2 \right. \\
+ \int_0^t \left( \Delta X_{i,N}^{i,N,M}, \sigma(s, X_{i,N}^{i,N,M}, \mu_{s, X_{\kappa(s)}, M}) - \sigma(s, X_{\kappa(s)}, M, X_{\kappa(s)}, M) \right) \right) ds \\
\leq C \left( \int_0^t E \left[ \sup_{0 \leq u \leq t} |\Delta X_{i,N}^{i,N,M}|^2 \right] ds + M^{-2\alpha} + M^{-1} \right),
\]
by Lemma 5.3 and since the \(X_{i,N}^{i,N,M}\) are identically distributed for all \(i \in \{1, \ldots, N\}\). This estimate holds for every \(i\) hence we can insert \(\sup_{1 \leq i \leq N} \) on both sides giving
\[
\sup_{1 \leq i \leq N} E \left[ \sup_{0 \leq u \leq t} |\Delta X_{i,N}^{i,N,M}|^2 \right] \\
\leq C \left( \int_0^t \sup_{1 \leq i \leq N} E \left[ \sup_{0 \leq u \leq t} |\Delta X_{i,N}^{i,N,M}|^2 \right] ds + M^{-2\alpha} + M^{-1} \right) < \infty,
\]
and finally by Grönwall’s lemma (using that \( \alpha = 1/2 \)),
\[
\sup_{1 \leq i \leq N} \mathbb{E} \left[ \sup_{0 \leq u \leq t} \left| X_{u}^{i,N} - X_{u}^{i,N,M} \right|^2 \right] \leq CM^{-1}.
\]

5.3. Proof of Implicit Convergence. The main goal here is to prove Proposition 3.7. We loosely follow [22], however, due to the extra dependencies on time and measure and further allowing for random initial conditions we require more refined arguments. We take \( N \) as some fixed positive integer. Before considering the implicit scheme, let us show a result on the particle system (2.2).

**Proposition 5.4.** Let Hypothesis 2.1, 2.2 and H1 (in Hypothesis 3.5) hold, further, let \( X_0 \in L^2(\mathbb{R}^d) \). Then the following bounds hold,
\[
\sup_{1 \leq i \leq N} \mathbb{E} \left[ \left| X_{T \land \tau_m}^{i,N} \right|^2 \right] \leq \left( \mathbb{E} \left[ \left| X_0 \right|^2 \right] + 2\alpha T \right) e^{2\beta T},
\]
and for \( \tau_m^i = \inf \{ t \geq 0 : \left| X_{t}^{i,N} \right| > m \} \)
\[
\sup_{1 \leq i \leq N} \mathbb{P}(\tau_m^i \leq T) \leq \frac{1}{m^2} \left( \mathbb{E} \left[ \left| X_0 \right|^2 \right] + 2\alpha T \right) e^{2\beta T}.
\]

**Proof.** Firstly, let us consider the stopped process \( X_{T \land \tau_m}^{i,N} \). Applying Itô to the square of this process and taking expectations yields
\[
\mathbb{E} \left[ \left| X_{T \land \tau_m}^{i,N} \right|^2 \right] = \mathbb{E} \left[ \left| X_0 \right|^2 \right] + \mathbb{E} \left[ \int_0^{T \land \tau_m^i} \left( \sum_{a=1}^l |\sigma_a(s, X_{s}^{i,N}, \mu_{s}^{X,N})|^2 ds \right) \right] \leq \mathbb{E} \left[ \left| X_0 \right|^2 \right] + 2\alpha T + \int_0^T 2\beta \mathbb{E} \left[ \left| X_{s \land \tau_m}^{i,N} \right| \right] ds \leq \left( \mathbb{E} \left[ \left| X_0 \right|^2 \right] + 2\alpha T \right) e^{2\beta T},
\]
where we have used the growth and stopping condition to remove the martingale term, then the monotone growth, uniform boundedness of \( b \) in the measure component \( b \) and Grönwall’s inequality to obtain the result.

Noting that the following lower bound also holds,
\[
\mathbb{E} \left[ \left| X_{T \land \tau_m}^{i,N} \right|^2 \right] \geq m^2 \mathbb{P}(\tau_m^i \leq T),
\]
hence we obtain,
\[
\mathbb{P}(\tau_m^i \leq T) \leq \frac{1}{m^2} \left( \mathbb{E} \left[ \left| X_0 \right|^2 \right] + 2\alpha T \right) e^{2\beta T}.
\]

Further, since \( \lim_{m \to \infty} \left| X_{T \land \tau_m}^{i,N} \right| = \left| X_T^{i,N} \right| \), we obtain by Fatou’s lemma,
\[
\mathbb{E} \left[ \left| X_T^{i,N} \right|^2 \right] \leq \liminf_{m \to \infty} \mathbb{E} \left[ \left| X_{T \land \tau_m}^{i,N} \right|^2 \right] \leq \left( \mathbb{E} \left[ \left| X_0 \right|^2 \right] + 2\alpha T \right) e^{2\beta T}.
\]

The result then follows by noting that \( \mathbb{E} \left[ \left| X_0 \right|^2 \right] = \mathbb{E} \left[ \left| X_0 \right|^2 \right] \) and hence the bounds are independent of \( i \), so we obtain the result for the supremum over \( i \).
Let us now return to the implicit scheme. At each time step \( t_i \) and for each particle \( i \) one needs to solve a fixed point equation,

\[
\tilde{X}_{t_{k+1}}^{i,N,M} - b(t_k, \tilde{X}_{t_{k+1}}^{i,N,M}, \tilde{\mu}_{t_k})h = \tilde{X}_{t_k}^{i,N,M} + \sigma(t_k, \tilde{X}_{t_k}^{i,N,M}, \tilde{\mu}_{t_k})\Delta W_{t_k}^i,
\]

this leads us to consider a function \( F \)

\[
F(t, x, \mu) := x - b(t, x, \mu)h.
\]

For the implicit scheme to have a solution the function \( F \) must have a unique inverse. The following lemma is crucial in proving convergence of the implicit scheme.

**Lemma 5.5.** Let Hypothesis 2.1, 2.2 and H1 (in Hypothesis 3.5) hold and fix \( h^* < 1/\max(L_b, 2\beta) \). Further, let \( 0 < h \leq h^* \) and take any \( t \in [0, T] \) and \( \mu \in \mathcal{P}_2(\mathbb{R}^d) \) fixed, then for all \( y \in \mathbb{R}^d \), there exists a unique \( x \) such that \( F(t, x, \mu) = y \). Hence the fixed point problem in (2.5) is well defined.

Moreover, for all \( t \in [0, T] \) and \( \mu \in \mathcal{P}_2(\mathbb{R}^d) \) the following bounds hold,

\[
|x|^2 \leq (1 - 2h\beta)^{-1}(|F(t, x, \mu)|^2 + 2h\alpha),
\]

and for any \( i \geq 1 \) the following recursive bound holds,

\[
|F(t_k, \tilde{X}_{t_k}^{i,N,M}, \tilde{\mu}_{t_k})|^2 \leq |F(t_{k-1}, \tilde{X}_{t_k}^{i,N,M}, \tilde{\mu}_{t_{k-1}})|^2 + \left( \sum_{a=1}^l |\sigma_a(t_k, \tilde{X}_{t_k}^{i,N,M}, \tilde{\mu}_{t_k})||\Delta W_{t_k}^i\right)_a \right)^2
\]

\[+ 2h\alpha + 2h\beta|\tilde{X}_{t_k}^{i,N,M}|^2 + 2\langle \tilde{X}_{t_k}^{i,N,M}, \sigma(t_k, \tilde{X}_{t_k}^{i,N,M}, \tilde{\mu}_{t_k})\Delta W_{t_k}^i \rangle,
\]

where \((\Delta W_{t_k}^i)_a\) is the \( a \)th entry of the vector.

**Proof.** Let us first prove there exists a unique solution to (5.7), in the sense that for all \( t \in [0, T] \) and \( \mu \in \mathcal{P}_2(\mathbb{R}^d) \) fixed, then there exists a unique \( x \in \mathbb{R}^d \) such that \( F(t, x, \mu) = y \) for a given \( y \in \mathbb{R}^d \), provided \( 0 < h < h^* \). This is a classical problem considered in [30, p.557] or see [20, p.2596], which require \( F \) to be continuous, monotone and coercive (in \( x \)). Clearly, since \( b \) is continuous, one has \( F \) is continuous.

For monotonicity in \( F \),

\[
\langle x - x', F(t, x, \mu) - F(t, x', \mu) \rangle = |x - x'|^2 - \langle x - x', b(t, x, \mu)h - b(t, x', \mu)h \rangle \\
\geq |x - x'|^2(1 - L_b h),
\]

which is clearly \( > 0 \) provided \( h < 1/L_b \). Coercivity follows similarly by the monotone growth condition in \( b \),

\[
\langle x, F(t, x, \mu) \rangle \geq |x|^2 - h(\alpha + \beta |x|^2),
\]

therefore,

\[
\lim_{|x| \to \infty} \frac{\langle x, F(t, x, \mu) \rangle}{|x|} = \infty, \quad \text{for } h < 1/\beta.
\]

Hence \( F(t, x, \mu) = y \) has a unique solution for \( F \) defined in (5.7) and therefore the numerical scheme (2.5) is well defined.
To show $x$ is bounded by $F(\cdot, x, \cdot)$, again fix some $t \in [0, T]$ and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, then,

$$|F(t, x, \mu)|^2 = |x|^2 - 2\langle x, b(t, x, \mu) \rangle h + |b(t, x, \mu)|^2 h^2 \geq |x|^2 - 2\langle x, b(t, x, \mu) \rangle h \geq (1 - 2h\beta)|x|^2 - 2h\alpha.$$

Since $h < 1/(2\beta)$, we obtain,

$$|x|^2 \leq (1 - 2h\beta)^{-1}(|F(t, x, \mu)|^2 + 2h\alpha).$$

This result is also useful since it holds for all $t \in [0, T]$ and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$. For the recursive bound it is useful to note,

$$F(t, \tilde{X}_{t_{k+1}}^{i,N,M}, \tilde{\mu}_{t_k}^{X,N,M}) = \tilde{X}_{t_{k+1}}^{i,N,M} - b(t, \tilde{X}_{t_{k+1}}^{i,N,M}, \tilde{\mu}_{t_k}^{X,N,M})h + \sigma(t, \tilde{X}_{t_{k+1}}^{i,N,M}, \tilde{\mu}_{t_k}^{X,N,M})\Delta W_{t_k}^i,$$

(5.9)

$$= F(t_{k-1}, \tilde{X}_{t_k}^{i,N,M}, \tilde{\mu}_{t_{k-1}}^{X,N,M}) + b(t_{k-1}, \tilde{X}_{t_k}^{i,N,M}, \tilde{\mu}_{t_{k-1}}^{X,N,M})h + \sigma(t, \tilde{X}_{t_k}^{i,N,M}, \tilde{\mu}_{t_{k-1}}^{X,N,M})\Delta W_{t_k}^i.$$

Of course, this recursion is only valid for $i \geq 1$, due to the appearance of $t_{k-1}$. Using this relation observe the following,

$$|F(t, \tilde{X}_{t_{k+1}}^{i,N,M}, \tilde{\mu}_{t_k}^{X,N,M})|^2$$

$$= |F(t_{k-1}, \tilde{X}_{t_k}^{i,N,M}, \tilde{\mu}_{t_{k-1}}^{X,N,M})|^2 + |b(t_{k-1}, \tilde{X}_{t_k}^{i,N,M}, \tilde{\mu}_{t_{k-1}}^{X,N,M})|^2 h^2 + 2\sigma(t, \tilde{X}_{t_k}^{i,N,M}, \tilde{\mu}_{t_{k-1}}^{X,N,M})\Delta W_{t_k}^i \Delta W_{t_k}^i$$

$$\leq 2\langle \tilde{X}_{t_k}^{i,N,M}, \sigma(t, \tilde{X}_{t_k}^{i,N,M}, \tilde{\mu}_{t_k}^{X,N,M})\Delta W_{t_k}^i \rangle.$$
Let us now show the first moment bound result, as is standard with implicit schemes we firstly do this under a stopping time, hence define,

\[(5.10) \quad \lambda_m^i = \inf\{k : |\hat{X}_{tk}^{i, N, M}| > m\}.\]

One should note that this stopping time does not actually bound \(\bar{X}\) at that point, the best one can do is bound the previous point i.e. for \(\lambda_m^i > 0\), we have \(|\hat{X}_{\lambda_m^i-1}^{i, N, M}| \leq m\).

**Lemma 5.6.** Let Hypothesis 2.1, 2.2 and H1 (in Hypothesis 3.5) hold and fix \(h^* < 1/\max(L_b, 2\beta)\). Then for any \(p \geq 2\) such that \(\mathbb{E}[|X_0|^p] = C(p) < \infty\), we also have,

\[
\sup_{1 \leq i \leq N} \mathbb{E}[|\hat{X}_{tk}^{i, N, M}|^p \mathbb{1}_{(k \leq \lambda_m^i)}] \leq C(p, m) \quad \forall k \leq M \text{ and } 0 < h \leq h^*.
\]

Using standard notation, \(C(a)\) denotes a constant that can depend on variable \(a\).

**Proof.** As it turns out the function \(F\) in (5.7) gives us a useful bound, from (5.9) we obtain,

\[
|F(t_k, \hat{X}_{tk+1}^{i, N, M}, \hat{\mu}_{tk}^{X, N, M})|^p \leq 2^{p-1}(\|X_{tk}^{i, N, M}|^p + |\sigma(t_k, \hat{X}_{tk}^{i, N, M}, \hat{\mu}_{tk}^{X, N, M})\Delta W_{tk}^i|^p).
\]

Hence, multiplying with the indicator and taking expected values yields,

\[
\mathbb{E}[|F(t_k, \hat{X}_{tk+1}^{i, N, M}, \hat{\mu}_{tk}^{X, N, M})|^p \mathbb{1}_{(i+1 \leq \lambda_m^i)}] \leq C(p) \left(m^p + \mathbb{E}[|\sigma(t_k, \hat{X}_{tk}^{i, N, M}, \hat{\mu}_{tk}^{X, N, M})\Delta W_{tk}^i|^p \mathbb{1}_{(i+1 \leq \lambda_m^i)}]\right).
\]

Then using,

\[
\mathbb{E}[|\sigma(t_k, \hat{X}_{tk}^{i, N, M}, \hat{\mu}_{tk}^{X, N, M})\Delta W_{tk}^i|^p \mathbb{1}_{(i+1 \leq \lambda_m^i)}] \leq \sum_{a=1}^{l} \mathbb{E}[|\sigma_a(t_k, \hat{X}_{tk}^{i, N, M}, \hat{\mu}_{tk}^{X, N, M})|^2 \mathbb{1}_{(i+1 \leq \lambda_m^i)}] + \mathbb{E}[|\Delta W_{tk}^i|_a|2^p] .
\]

Using the bounds on each coefficient of \(\sigma\), it is straightforward to observe,

\[
|\sigma_a(t_k, \hat{X}_{tk}^{i, N, M}, \hat{\mu}_{tk}^{X, N, M})|^2 \leq C(p)(1 + |\hat{X}_{tk}^{i, N, M}|^2) .
\]

Using this bound we obtain,

\[
\mathbb{E}[|F(t_k, \hat{X}_{tk+1}^{i, N, M}, \hat{\mu}_{tk}^{X, N, M})|^p \mathbb{1}_{(i+1 \leq \lambda_m^i)}] \leq C(p, m) .
\]

Rewriting the quantity we wish to bound as

\[
\mathbb{E}[|\hat{X}_{tk}^{i, N, M}|^p \mathbb{1}_{(k \leq \lambda_m^i)}] = \mathbb{E}[|\hat{X}_{tk+1}^{i, N, M}|^p \mathbb{1}_{(i+1 \leq \lambda_m^i)}] + \mathbb{E}[|\hat{X}_{tk+1}^{i, N, M}|^p \mathbb{1}_{(i=0, \lambda_m^i=0)}] \leq C(p, m) ,
\]

where the inequality follows from Lemma 5.5, our bound on \(F\), and the hypothesis that \(X_0 \in L^p(\mathbb{R}^d)\). Again, the corresponding bound is independent of the choice of \(i\), hence the result holds for the supremum over \(i\).

Although the previous bound is useful, the presence of the stopping time is inconvenient, we therefore remove it and show the second moment is bounded.
Proposition 5.7. Let Hypothesis 2.1, 2.2 and H1 (in Hypothesis 3.5) hold and fix $h^* < 1/\max(L_0, 2\beta)$. Further assume that $X_0 \in L^4(\mathbb{R}^d)$. Then,

$$\sup_{1 \leq k \leq M} \sup_{0 \leq t \leq N} \sup_{h \leq h^*} \mathbb{E}[|\tilde{X}_{t_k}^{i,N,M}|^2] \leq C.$$ 

Proof. Firstly let us take a nonnegative integer $K$, such that $Kh \leq T$. Now let us consider (5.8), one can note that this bound still holds where the $F$ terms are multiplied by $\mathbb{1}_{\{\lambda_m^i > 0\}}$ (since both sides are nonnegative and the indicator is bounded above by one). Summing both sides from $k = 1$ to $K \wedge \lambda_m^i$, noting that $F$ terms cancel, we obtain,

$$|F(t_{K \wedge \lambda_m^i}, \tilde{X}_{t_{K \wedge \lambda_m^i}+1}^{i,N,M}, \tilde{\mu}_{t_{K \wedge \lambda_m^i}}^{X,N,M})|^2 \mathbb{1}_{\{\lambda_m^i > 0\}}$$

$$\leq |F(t_0, \tilde{X}_{t_1}^{i,N,M}, \tilde{\mu}_{t_0}^{X,N,M})|^2 \mathbb{1}_{\{\lambda_m^i > 0\}} + \sum_{k=1}^{K \wedge \lambda_m^i} \left( \sum_{a=1}^{l} |\sigma_a(t_k, \tilde{X}_{t_k}^{i,N,M}, \tilde{\mu}_{t_k}^{X,N,M})||\Delta W_{t_k}^j\rangle \langle a| \mathbb{1}_{\{\lambda_m^i > 0\}} \right)^2$$

$$+ \sum_{k=1}^{K \wedge \lambda_m^i} 2\langle \tilde{X}_{t_k}^{i,N,M}, \sigma(t_k, \tilde{X}_{t_k}^{i,N,M}, \tilde{\mu}_{t_k}^{X,N,M})\Delta W_{t_k}^j \rangle \mathbb{1}_{\{\lambda_m^i > 0\}},$$

where we use the convention $\sum_{k=1}^0 a = 0$. Although the stopping time is useful it is not ideal that it appears on the sum, however, for nonnegative terms it is straightforward to take the stopping time into the coefficients and for the stochastic term we can rewrite as,

$$\sum_{k=1}^{K \wedge \lambda_m^i} 2\langle \tilde{X}_{t_k}^{i,N,M}, \sigma(t_k, \tilde{X}_{t_k}^{i,N,M}, \tilde{\mu}_{t_k}^{X,N,M})\Delta W_{t_k}^j \rangle$$

$$= \sum_{k=1}^{K} 2\langle \tilde{X}_{t_k}^{i,N,M}, \sigma(t_k, \tilde{X}_{t_k}^{i,N,M}, \tilde{\mu}_{t_k}^{X,N,M})\Delta W_{t_k}^j \rangle \mathbb{1}_{\{k \leq \lambda_m^i\}}.$$ 

Taking expectations of this and noting by Lemma 5.6 that $\tilde{X}_{t_k}^{i,N,M} \mathbb{1}_{\{k \leq \lambda_m^i\}} \in L^4_{t_k}(\mathbb{R}^d)$, hence, this term is a martingale. We therefore obtain the following bound,

$$\mathbb{E}[|F(t_{K \wedge \lambda_m^i}, \tilde{X}_{t_{K \wedge \lambda_m^i}+1}^{i,N,M}, \tilde{\mu}_{t_{K \wedge \lambda_m^i}}^{X,N,M})|^2 \mathbb{1}_{\{\lambda_m^i > 0\}}]$$

$$\leq \mathbb{E}[|F(t_0, \tilde{X}_{t_1}^{i,N,M}, \tilde{\mu}_{t_0}^{X,N,M})|^2] + 2\alpha T + \sum_{k=1}^{K} 2h \beta \mathbb{E}[|\tilde{X}_{t_k}^{i,N,M}|^2 \mathbb{1}_{\{\lambda_m^i > 0\}}]$$

$$+ \sum_{k=1}^{K} \mathbb{E}\left[\left( \sum_{a=1}^{l} |\sigma_a(t_k, \tilde{X}_{t_k}^{i,N,M}, \tilde{\mu}_{t_k}^{X,N,M})||\Delta W_{t_k}^j\rangle \langle a| \right)^2 \mathbb{1}_{\{\lambda_m^i > 0\}} \right].$$

The idea is to apply the discrete version of Grönwall’s inequality to this (see for example [24, pg 436] or [22, Lemma 3.4]), which requires our bound to be in terms of
F. Using arguments similar to previously,

\[
\mathbb{E}\left[ \sum_{a=1}^{l} |\sigma_a(t_{k\wedge\lambda_m}, \tilde{X}_{t_{k\wedge\lambda_m}}^{i,N,M}, \tilde{\mu}_{t_{k\wedge\lambda_m}}^{X,N,M})||((\Delta W_{t_{k\wedge\lambda_m}}^{i})_a)|^2 \mathbb{1}_{\{\lambda_m > 0\}} \right] \\
\leq C \sum_{a=1}^{l} \mathbb{E}\left[ |\sigma_a(t_{k\wedge\lambda_m}, \tilde{X}_{t_{k\wedge\lambda_m}}^{i,N,M}, \tilde{\mu}_{t_{k\wedge\lambda_m}}^{X,N,M})||((\Delta W_{t_{k\wedge\lambda_m}}^{i})_a)|^2 \mathbb{1}_{\{\lambda_m > 0\}} \right] \\
\leq C \sum_{a=1}^{l} h(1 + \mathbb{E}[|\tilde{X}_{t_{k\wedge\lambda_m}}^{i,N,M}|^2 \mathbb{1}_{\{\lambda_m > 0\}}]),
\]

where we have used independence of \( \sigma(\cdot) \mathbb{1}_{\{\lambda_m > 0\}} \) and \( \Delta W \) along with the growth bounds on \( \sigma \) to obtain the final inequality. Combing this with our previous bounds and appealing again to Lemma 5.5 (to bound \( \tilde{X} \) by \( F \)) we obtain,

\[
\mathbb{E}[|F(t_{k\wedge\lambda_m}, \tilde{X}_{t_{k\wedge\lambda_m}}^{i,N,M}, \tilde{\mu}_{t_{k\wedge\lambda_m}}^{X,N,M})|^2 \mathbb{1}_{\{\lambda_m > 0\}}] \\
\leq \mathbb{E}[|F(t_0, \tilde{X}_{t_0}^{i,N,M}, \tilde{\mu}_{t_0}^{X,N,M})|^2] + C + \sum_{k=1}^{K} C h \mathbb{E}[|\tilde{X}_{t_{k\wedge\lambda_m}}^{i,N,M}|^2 \mathbb{1}_{\{\lambda_m > 0\}}] \\
\leq \mathbb{E}[|F(t_0, \tilde{X}_{t_0}^{i,N,M}, \tilde{\mu}_{t_0}^{X,N,M})|^2] + C(1 + \frac{h}{1 - 2h\beta}) \\
+ \sum_{k=1}^{K} \frac{C h}{1 - 2h\beta} \mathbb{E}[|F(t_{(k\wedge\lambda_m)-1}, \tilde{X}_{t_{(k\wedge\lambda_m)-1}}^{i,N,M}, \tilde{\mu}_{t_{(k\wedge\lambda_m)-1}}^{X,N,M})|^2 \mathbb{1}_{\{\lambda_m > 0\}}].
\]

Applying a discrete version of Grönwall inequality and noting \( \sum_{k=1}^{K} 1 \leq T/h \) yields

\[
\mathbb{E}[|F(t_{k\wedge\lambda_m}, \tilde{X}_{t_{k\wedge\lambda_m}}^{i,N,M}, \tilde{\mu}_{t_{k\wedge\lambda_m}}^{X,N,M})|^2 \mathbb{1}_{\{\lambda_m > 0\}}] \\
\leq \left( \mathbb{E}[|F(t_0, \tilde{X}_{t_0}^{i,N,M}, \tilde{\mu}_{t_0}^{X,N,M})|^2] + C(1 + \frac{h}{1 - 2h\beta}) \right) \exp\left( \frac{C}{1 - 2h\beta} \right).
\]

Recalling (5.9), we can apply the same arguments as previous to obtain the bound

\[
\mathbb{E}[|F(t_0, \tilde{X}_{t_0}^{i,N,M}, \tilde{\mu}_{t_0}^{X,N,M})|^2] \leq C(1 + (1 + h) \mathbb{E}[|\tilde{X}_{t_0}^{i,N,M}|^2])
\]

Noting that our bound for \( F \) is now independent of \( m \), we can use Fatou’s lemma to take the limit and obtain (for \( K \geq 1 \),

\[
\mathbb{E}[|F(t_k, \tilde{X}_{t_{k+1}}^{i,N,M}, \tilde{\mu}_{t_k}^{X,N,M})|^2] \\
\leq C(1 + (1 + h) \mathbb{E}[|\tilde{X}_{t_0}^{i,N,M}|^2] + \frac{h}{1 - 2h\beta}) \exp\left( \frac{C}{1 - 2h\beta} \right).
\]

Again by Lemma 5.5, the LHS bounds \( \tilde{X}_{t_{k+1}}^{i,N,M} \) (with some constant) hence we obtain a bound for \( \tilde{X}_{t_k}^{i,N,M} \) for \( k \geq 2 \). Clearly \( \tilde{X}_{t_0}^{i,N,M} \) has second moment (by assumption), therefore we need to obtain a bound for \( \tilde{X}_{t_1}^{i,N,M} \). This is not difficult to obtain by again using that we can bound \( \tilde{X} \) as follows,

\[
\mathbb{E}[|\tilde{X}_{t_1}^{i,N,M}|^2] \leq (1 - 2h\beta)^{-1} \left( 2h\alpha + \mathbb{E}[|F(t_0, \tilde{X}_{t_0}^{i,N,M}, \tilde{\mu}_{t_0}^{X,N,M})|^2] \right),
\]
then we can apply the same bound on $F$ as above.

In order to complete the proof, we need to also show this bound exists for all $i$ and $0 < h < h^\ast$. One can see immediately that all bounds decrease as $h$ decreases, hence the supremum value is to set $h = h^\ast$, which is also finite since $h^\ast < 1/(2\beta)$. The supremum over $i$ follows from the fact that all bounds are independent of $i$. 

Now that we have established a bound on the second moment, we look to show convergence of this scheme to the true particle system solution. As always with discrete schemes it is beneficial to introduce their continuous counterpart. As it turns out doing it naively for implicit schemes leads to measurability problems, hence one introduces the so-called forward backward scheme

$$
\hat{X}_{i,k+1} = \hat{X}_{i,k} + b \left( t_{k-1 \lor 0}, \hat{X}_{i,k}, \tilde{\mu}_{t_{k-1 \lor 0}} \right) h + \sigma \left( t_{k}, \hat{X}_{i,k}, \tilde{\mu}_{t_{k}} \right) \Delta W_{i,k},
$$

where $\hat{X}_{0} = X_{i,0}$ and $\lor$ denotes the maximum. The scheme’s continuous time version is

$$
\hat{X}_{t} = X_{0} + \int_{0}^{t} b \left( \kappa(s) - h \right) \lor 0, \tilde{\mu}_{\kappa(s) - h} \lor 0 \right) ds
$$

(5.11)

$$
+ \int_{0}^{t} \sigma \left( \kappa(s), \tilde{\mu}_{\kappa(s)} \lor \tilde{\sigma}_{\kappa(s)} \right) dW_{s}.
$$

The first result we wish to present is that the discrete and continuous versions stay close to one another, up to the stopping time (5.10).

**Lemma 5.8.** Let Hypothesis $2.1$, $2.2$ and $H1$ (in Hypothesis $3.5$) hold and fix $h^\ast < 1 / \max(L_b, 2\beta)$. Further assume $X_{0} \in L^q(R^d)$. Then for $1 \leq p \leq 4$ the following holds for $0 < h < h^\ast$:

$$
\sup_{1 \leq i \leq N} \sup_{0 \leq k \leq M} \mathbb{E} \left[ |\hat{X}_{i,k} - \tilde{X}_{i,k} |^p 1_{k \leq \lambda_{m}} \right] \leq C(m, p) h^p.
$$

Moreover, we also have the following relation between $\hat{X}$ and $F$ for all $1 \leq k \leq M$,

$$
|\hat{X}_{i,k} - \tilde{X}_{i,k} |^2 \geq \frac{1}{2} |F(t_{k-1} \lor 0, \tilde{X}_{i,k} - \hat{X}_{i,k}, \tilde{\mu}_{t_{k-1}} \lor 0, \tilde{\mu}_{t_{k}} \lor 0) |^2 - |b(t_{0}, \hat{X}_{i,k}, \tilde{\mu}_{t_{0}} \lor 0, \hat{\mu}_{t_{0}}(t_{0}) \lor 0) h |^2.
$$

**Proof.** To show the first part we start by noting the following useful relation between (2.5) and (5.11), namely for $1 \leq k \leq M$,

$$
\hat{X}_{i,k} - \tilde{X}_{i,k} = \left( b(t_{0}, \hat{X}_{i,k}, \hat{\mu}_{t_{0}} \lor 0, \hat{\mu}_{t_{0}}) - b(t_{k-1} \lor 0, \hat{X}_{i,k} - \tilde{X}_{i,k}(t_{k-1} \lor 0, \hat{\mu}_{t_{k-1}} \lor 0) \right) h.
$$

Noting that one can bound,

$$
\left| b(t_{0}, \hat{X}_{i,k}, \hat{\mu}_{t_{0}} \lor 0, \hat{\mu}_{t_{0}}) - b(t_{k-1} \lor 0, \hat{X}_{i,k} - \tilde{X}_{i,k}(t_{k-1} \lor 0, \hat{\mu}_{t_{k-1}} \lor 0) \right| \leq C(1 + |t_{k}|^{1/2} + |\hat{X}_{i,k} - \tilde{X}_{i,k} |^{q+1} + |\hat{X}_{i,k} - \tilde{X}_{i,k} |^{q+1}),
$$

where we have used the growth bounds on the coefficient $b$, in particular Hypothesis $H1$. Hence,

$$
\mathbb{E} \left[ |\hat{X}_{i,k} - \tilde{X}_{i,k} |^p 1_{k \leq \lambda_{m}} \right] \leq C(p) h^p \left( 1 + |t_{k}|^{p/2} + \mathbb{E} \left[ |\hat{X}_{i,k} - \tilde{X}_{i,k} |^{q+1} 1_{k \leq \lambda_{m}} \right] \right),
$$

$$
\mathbb{E} \left[ |\hat{X}_{i,k} - \tilde{X}_{i,k} |^p 1_{k \leq \lambda_{m}} \right] \leq C(p) h^p \left( 1 + |t_{k}|^{p/2} + \mathbb{E} \left[ |\hat{X}_{i,k} - \tilde{X}_{i,k} |^{q+1} 1_{k \leq \lambda_{m}} \right] \right).$$
One observes that the terms on the RHS are bounded by $C(p,m)$ for $p \leq 4$ since $X_0 \in L^{4(q+1)}(\mathbb{R}^d)$ and Lemma 5.6. This completes the first part of the proof.

For the second part, recall from the relation between (2.5) and (5.11), one has,

$$\tilde{X}^{i,N,M}_{t_k} = b(t_0, \tilde{X}^{i,N,M}_{t_0}, \tilde{\mu}^{X,N,M}_{t_0})h + \tilde{X}^{i,N,M}_{t_k} - b(t_{k-1}, \tilde{X}^{i,N,M}_{t_{k-1}}, \tilde{\mu}^{X,N,M}_{t_{k-1}})h$$

$$= b(t_0, \tilde{X}^{i,N,M}_{t_0}, \tilde{\mu}^{X,N,M}_{t_0})h + F(t_{k-1}, \tilde{X}^{i,N,M}_{t_{k-1}}, \tilde{\mu}^{X,N,M}_{t_{k-1}}).$$

The result follows from squaring both sides and applying Young’s inequality.

The next result we wish to present is that both schemes do not blow up in finite time, for this we define a new stopping time,

$$\eta^i_m := \inf \{ t \geq 0 : |\tilde{X}^{i,N,M}_t| \geq m, \text{ or } |\tilde{X}^{i,N,M}_{\kappa(t)}| > m \}.$$ 

**Lemma 5.9.** Let Hypothesis 2.1, 2.2 and H1 (in Hypothesis 3.5) hold, fix $h^* < 1/\max(L_b, 2\beta)$ and assume $X_0 \in L^{4(q+1)}(\mathbb{R}^d)$. Then, for any $\epsilon > 0$, there exists a $m^*$ such that, for any $m \geq m^*$ we can find a $h_0^*(m)$ (note the dependence on $m$) so that,

$$\sup_{1 \leq i \leq N} \mathbb{P}(\eta^i_m < T) \leq \epsilon, \text{ for any } 0 < h \leq h_0^*(m).$$

**Proof.** Note due to the initial condition being random we must be careful with how we set $m$, we shall come back to this later. Let us start by applying Itô to the stopped version of (5.11),

$$|\tilde{X}^{i,N,M}_T|^{2} = |X_0|^2 + \int_0^T \left\{ \tilde{X}^{i,N,M}_t, b\left( (\kappa(s) - h) \vee 0, \tilde{X}^{i,N,M}_{\kappa(s)}, \tilde{\mu}^{X,N,M}_{\kappa(s)} \right) \right\} ds + \sum_{a=1}^l |\sigma_a(\kappa(s), \tilde{X}^{i,N,M}_{\kappa(s)}, \tilde{\mu}^{X,N,M}_{\kappa(s)})|^2 + \int_0^T \sum_{a=1}^l |\sigma_a(\kappa(s), \tilde{X}^{i,N,M}_{\kappa(s)}, \tilde{\mu}^{X,N,M}_{\kappa(s)})|^2 ds$$

$$+ \int_0^T \sum_{a=1}^l |\sigma_a(\kappa(s), \tilde{X}^{i,N,M}_{\kappa(s)}, \tilde{\mu}^{X,N,M}_{\kappa(s)})|^2 dW_s.$$

We now look to bound the various integrands, firstly one can observe

$$\langle \tilde{X}^{i,N,M}_t, b\left( (\kappa(s) - h) \vee 0, \tilde{X}^{i,N,M}_{\kappa(s)}, \tilde{\mu}^{X,N,M}_{\kappa(s)} \right) \rangle + \sum_{a=1}^l |\sigma_a(\kappa(s), \tilde{X}^{i,N,M}_{\kappa(s)}, \tilde{\mu}^{X,N,M}_{\kappa(s)})|^2$$

$$= \langle \tilde{X}^{i,N,M}_t - \tilde{X}^{i,N,M}_{\kappa(s)}, b((\kappa(s) - h) \vee 0, \tilde{X}^{i,N,M}_{\kappa(s)}, \tilde{\mu}^{X,N,M}_{\kappa(s)} - h) \vee 0) \rangle$$

$$+ \langle \tilde{X}^{i,N,M}_t, b((\kappa(s) - h) \vee 0, \tilde{X}^{i,N,M}_{\kappa(s)}, \tilde{\mu}^{X,N,M}_{\kappa(s)} - h) \vee 0) \rangle + \sum_{a=1}^l |\sigma_a(\kappa(s), \tilde{X}^{i,N,M}_{\kappa(s)}, \tilde{\mu}^{X,N,M}_{\kappa(s)})|^2$$

$$\leq C|\tilde{X}^{i,N,M}_t - \tilde{X}^{i,N,M}_{\kappa(s)}|\left( 1 + |\tilde{X}^{i,N,M}_{\kappa(s)}|^{q+1} + \alpha + \beta |\tilde{X}^{i,N,M}_{\kappa(s)}|^2 \right),$$

where we used Cauchy-Schwarz, polynomial growth bound and monotone growth to obtain the final inequality.

Taking expectations and noting that due to the stopping time the stochastic integral is a martingale i.e. has second moment, we obtain,

$$\mathbb{E}[|\tilde{X}^{i,N,M}_T|^2]$$

$$\leq \mathbb{E}[|X_0|^2] + \mathbb{E}\left[ \int_0^T C|\tilde{X}^{i,N,M}_s - \tilde{X}^{i,N,M}_{\kappa(s)}|\left( 1 + |\tilde{X}^{i,N,M}_{\kappa(s)}|^{q+1} + \alpha + \beta |\tilde{X}^{i,N,M}_{\kappa(s)}|^2 \right) ds \right].$$
To proceed we note the following, \( |\tilde{X}_{\kappa(s)}^{i,N,M}|^2 \leq 2(|\hat{X}_{\kappa(s)}^{i,N,M} - \hat{X}_{\kappa(s)}^{i,N,M}|^2 + |\hat{X}_{s}^{i,N,M}|^2) \) and also

\[
\int_0^{T \land \eta_m'} |\hat{X}_{s}^{i,N,M} - \hat{X}_{\kappa(s)}^{i,N,M}|^2 \, ds \leq C(m) \int_0^{T \land \eta_m'} |\hat{X}_{s}^{i,N,M} - \hat{X}_{\kappa(s)}^{i,N,M}| \, ds,
\]

where we used the fact that the stopping time ensures \( \hat{X} \) and \( \hat{X} \) are \( \leq m \) for \( s < \eta_m' \) and \( s = \eta_m' \) has measure zero. The same reasoning also implies,

\[
\int_0^{T \land \eta_m'} C|\hat{X}_{s}^{i,N,M} - \hat{X}_{\kappa(s)}^{i,N,M}|[(1 + |\hat{X}_{\kappa(s)}^{i,N,M}|^{q+1}) \, ds
\leq C(m) \int_0^{T \land \eta_m'} |\hat{X}_{s}^{i,N,M} - \hat{X}_{\kappa(s)}^{i,N,M}| \, ds.
\]

Hence the following result holds,

\[
E[|\hat{X}_{T \land \eta_m'}^{i,N,M}|^2] \leq E[|X_0|] + CE\left[ \int_0^{T \land \eta_m'} C(m)|\hat{X}_{s}^{i,N,M} - \hat{X}_{\kappa(s)}^{i,N,M}| + 1 + \beta|\hat{X}_{s}^{i,N,M}|^2 \, ds \right].
\]

The next step is of course to take the expectation inside the integral, let us start by noting the difference term can be bounded as,

\[
E\left[ \int_0^{T \land \eta_m'} |\hat{X}_{s}^{i,N,M} - \hat{X}_{\kappa(s)}^{i,N,M}| \, ds \right]
\leq E\left[ \int_0^{T \land \eta_m'} |\hat{X}_{s}^{i,N,M} - \hat{X}_{\kappa(s)}^{i,N,M}| \, ds + \int_0^{T \land \eta_m'} |\hat{X}_{s}^{i,N,M} - \hat{X}_{\kappa(s)}^{i,N,M}| \, ds \right]
\leq E\left[ \int_0^{T \land \eta_m'} \left| b \left( (\kappa(s) - h) \lor 0, \hat{X}_{\kappa(s)}^{i,N,M}, \hat{\mu}_{\kappa(s)}^{i,N,M} \right) \right| \, ds \right]
+ E\left[ \int_0^{T \land \eta_m'} \left| \sigma \left( \kappa(s), \hat{X}_{\kappa(s)}^{i,N,M}, \hat{\mu}_{\kappa(s)}^{i,N,M} \right) \left( W_s^i - W_{\kappa(s)}^i \right) \right| \, ds \right] + C(m)h^{1/2},
\]

where we have used Lemma 5.8 for the final inequality. For the other terms, one can note due to the growth assumptions on \( b \), that,

\[
E\left[ \int_0^{T \land \eta_m'} \left| b \left( (\kappa(s) - h) \lor 0, \hat{X}_{\kappa(s)}^{i,N,M}, \hat{\mu}_{\kappa(s)}^{i,N,M} \right) \right| \, ds \right] \leq C(m) h.
\]

The term involving \( \sigma \) is more complex, however, we can bound as follows,

\[
E\left[ \int_0^{T \land \eta_m'} \left| \sigma \left( \kappa(s), \hat{X}_{\kappa(s)}^{i,N,M}, \hat{\mu}_{\kappa(s)}^{i,N,M} \right) \left( W_s^i - W_{\kappa(s)}^i \right) \right| \, ds \right]
\leq C \int_0^T \sum_{a=1}^l E\left[ \left| \sigma_a \left( \kappa(s), \hat{X}_{\kappa(s)}^{i,N,M}, \hat{\mu}_{\kappa(s)}^{i,N,M} \right) \left( W_s^i - W_{\kappa(s)}^i \right) \right| \right] \, ds
\leq C \int_0^T \sum_{a=1}^l h^{1/2} (1 + E[|\hat{X}_{T \land \eta_m'}^{i,N,M}|^2]) \, ds \leq C(m) h^{1/2}.
\]
Further, since $|\hat{X}_{\eta_{m}}^{i,N,M}| \geq 0$, we obtain,

$$
E \left[ \int_{0}^{T \wedge \eta_{m}} |\hat{X}_{s}^{i,N,M}|^2 ds \right] \leq \int_{0}^{T} E \left[ |\hat{X}_{s}^{i,N,M}|^2 \right] ds.
$$

Hence,

$$
E[|\hat{X}_{T \wedge \eta_{m}}^{i,N,M}|^2] \leq E[|X_{0}|^2] + C(m)h^{1/2} + C \int_{0}^{T} 1 + \beta E[|\hat{X}_{s \wedge \eta_{m}}^{i,N,M}|^2] ds
$$

$$
\leq (E[|X_{0}|^2] + C(m)h^{1/2}) \exp(C/\beta T),
$$

(5.13)

where the final inequality follows from Grönwall.

In order to obtain an upper bound on the probability of the stopping time occurring we look to obtain a lower bound for (5.11) at the stopping time. For the moment let us take $X_{t}^{i} < m$, hence $\eta_{m} > 0$, there are now two possible ways the stopping time can be reached, if $\hat{X}$ hits the boundary first then we have $|\hat{X}_{\eta_{m}}^{i,N,M}| = m$ and if $\hat{X}$ hits the boundary first we have $|\hat{X}_{\eta_{m}}^{i,N,M}| > m$.

In the case that $\hat{X}$ hits the boundary first, the lower bound is obvious, namely $|\hat{X}_{\eta_{m}}^{i,N,M}| = m$. For the second case it is less obvious. Recalling (5.12) and Lemma 5.5 we obtain the following lower bound for,

$$
|\hat{X}_{t_{k}}^{i,N,M}|^2 \geq \frac{1}{2}((1 - 2h\beta)|\hat{X}_{t_{k}}^{i,N,M}|^2 - 2h\alpha) - |b(t_{0}, \hat{X}_{t_{k}}^{i,N,M}, \hat{\mu}_{t_{0}}^{X,N,M})h|^2,
$$

where again we are taking $k \geq 1$ here, but this is not a problem since we are assuming for the moment $X_{0}^{i} < m$. Observing that this lower bound holds independent of which process triggers the stopping condition we can say w.l.o.g. that,

$$
m^2 \geq |\hat{X}_{\eta_{m}}^{i,N,M}|^2 \mathbb{1}_{\{|X_{0}| < m\}}
$$

$$
\geq \frac{1}{2}((1 - 2h\beta)m^2 - 2h\alpha) \mathbb{1}_{\{|X_{0}| < m\}} - |b(t_{0}, \hat{X}_{t_{0}}^{i,N,M}, \hat{\mu}_{t_{0}}^{X,N,M})h|^2 \mathbb{1}_{\{|X_{0}| < m\}}.
$$

Therefore,

$$
|\hat{X}_{\eta_{m}}^{i,N,M}|^2 \mathbb{1}_{\{|X_{0}| < m\}} \geq (C_{1}m^2 - C_{2}h) \mathbb{1}_{\{|X_{0}| < m\}} - C(m)h^2 \mathbb{1}_{\{|X_{0}| < m\}},
$$

where $|b(t_{0}, \hat{X}_{t_{0}}^{i,N,M}, \hat{\mu}_{t_{0}}^{X,N,M})| \mathbb{1}_{\{|X_{0}| < m\}} \leq C(m) \mathbb{1}_{\{|X_{0}| < m\}}$ via the growth condition on $b$. Let us now combine these results to obtain an upper bound for the probability of the stopping time, notice that,

$$
E[|\hat{X}_{T \wedge \eta_{m}}^{i,N,M}|^2] \geq E[|X_{0}|^2 \mathbb{1}_{\{|X_{0}| \geq m\}}] + E[|\hat{X}_{\eta_{m}}^{i,N,M}|^2 \mathbb{1}_{\{0 < \eta_{m} < T\}}]
$$

$$
\geq \mathbb{P}(\eta_{m} = 0)
$$

$$
+ ((C_{1}m^2 - C_{2}h) - C(m)h^2) \mathbb{P}(\{|X_{0}| < m\} \cap \{0 < \eta_{m} < T\}).
$$

Leaving the second term for the moment, one observes that for any $\epsilon > 0$,

$$
\mathbb{P}(\eta_{m} = 0) \leq m\mathbb{P}(\{|X_{0}| \geq m\}) \leq E[|X_{0}| \mathbb{1}_{\{|X_{0}| \geq m\}}] \leq \frac{\epsilon}{3},
$$

for $m$ sufficiently large, call this point $m^{*}$, since $X_{0}^{i}$ is uniformly integrable. It is also useful to note that $\mathbb{P}(\{|X_{0}| < m\} \cap \{0 < \eta_{m} < T\}) = \mathbb{P}(\{0 < \eta_{m} < T\})$. It is clear
from our previous analysis that for \( m \) large enough and (5.13) the probability can be bounded by,

\[
P(0 < \eta_m < T) \leq \frac{\mathbb{E}[[\hat{X}_{T \wedge \eta_m}^{i,N,M}]^2]}{(C_1 m^2 - C_2 h - C(m)h^2)} \leq \frac{(\mathbb{E}[|X_0^i|^2] + C + C(m)h^{1/2}) \exp(C \beta T)}{C_1 m^2 - C_2 h - C(m)h^2}.
\]

Now the goal is to bound this by \( 2\epsilon/3 \), we already have taken \( m \) sufficiently large to obtain the last inequality, now consider for any given \( m \), \( h^*_{01}(m) \):

\[
C_2 h^*_{01}(m) + C(m)h^*_{01}(m)^2 \leq 1.
\]

It is clear for \( 0 < h < h^*_{01}(m) \) the same bound holds. Then for the same \( \epsilon \) as before choose \( m \) large enough such that,

\[
(\mathbb{E}[|X_0^i|^2] + C + C(m)h^1/2) \exp(C \beta T) \leq \epsilon/3.
\]

Redefine \( m^* \) as the corresponding maximum of this \( m \) and \( m^* \). Now for any \( m \geq m^* \), define \( h^*_{02}(m) \) such that,

\[
\frac{C(m)(h^*_{02})^{1/2} \exp(C \beta T)}{C_1 m^2 - 1} \leq \epsilon/3.
\]

Again for \( 0 < h < h^*_{02}(m) \) the above inequality holds. Hence for any \( m \geq m^* \) and any \( 0 < h < \min(h^*_{01}(m), h^*_{02}(m)) \), we have, \( P(\eta_m < T) \leq P(\eta_m = 0) + P(0 < \eta_m < T) \leq \epsilon \).

We now look towards showing our strong convergence result, firstly by showing convergence between (5.11) and (2.2) and then (2.5) and (2.2). From this point onwards we require H2 (in Hypothesis 3.5).

**Remark 5.10 (On the diffusion coefficient \( \sigma \) being independent of the measure).**

The reason we cannot allow \( \sigma \) to have measure dependence is because our stopping time arguments do not work. Namely in order for two diffusion coefficients to be similar we require all \( N \) particles to be close to one another, not just the \( i \)th particle.

As it turns out though, this is not a problem for the drift term, so we make no change to the measure dependence there.

Recalling the stopping time in Proposition 5.4, we now define \( \theta^i_m := \tau^i_m \wedge \eta^i_m \) and have the following convergence result.

**Lemma 5.11.** Let Hypothesis 2.1, 2.2, the full Hypothesis 3.5 hold, fix \( h^* < 1/\max(L_b, 2\beta) \) and assume \( X_0 \in L^{1(q+1)}(\mathbb{R}^d) \). Then

\[
\sup_{1 \leq i \leq N} \sup_{0 \leq t \leq T} \mathbb{E}[|X_{t \wedge \theta^i_m}^{i,N,M} - X_{t \wedge \theta^i_m}^{i,N}|^2] \leq C(m)h.
\]

**Proof.** For ease of presentation we denote by \( \pi(s) := (\kappa(s) - h) \vee 0 \). As is standard
we start by applying Itô to the difference to obtain,

$$|X_{t \wedge \theta_m}^{i,N} - \hat{X}_{t \wedge \theta_m}^{i,N,M}|^2$$

$$= \int_0^{t \wedge \theta_m} 2\langle X_s^{i,N} - \hat{X}_s^{i,N,M}, b(s, X_s^{i,N}, \mu_s^{X,N}) - b(\overline{\eta}(s), \hat{X}_s^{i,N,M}, \mu_{(\overline{\eta}(s)-h)}^{X,N,M}) \rangle$$

$$+ \sum_{a=1}^l |\sigma(s, X_s^{i,N}) - \sigma(\kappa(s), \hat{X}_s^{i,N,M})|^2 ds$$

$$+ \int_0^{t \wedge \theta_m} 2\langle X_s^{i,N} - \hat{X}_s^{i,N,M}, (\sigma(s, X_s^{i,N}) - \sigma(\kappa(s), \hat{X}_s^{i,N,M}))dW_s^j \rangle.$$ 

By writing out the drift term we have that,

$$\langle X_s^{i,N} - \hat{X}_s^{i,N,M}, b(s, X_s^{i,N}, \mu_s^{X,N}) - b(\overline{\eta}(s), \hat{X}_s^{i,N,M}, \mu_{(\overline{\eta}(s)-h)}^{X,N,M}) \rangle$$

$$= \langle X_s^{i,N} - \hat{X}_s^{i,N,M}, b(s, X_s^{i,N}, \mu_s^{X,N}) - b(s, \hat{X}_s^{i,N,M}, \mu_s^{X,N}) \rangle$$

$$+ \langle X_s^{i,N} - \hat{X}_s^{i,N,M}, b(s, \hat{X}_s^{i,N,M}, \mu_s^{X,N}) - b(\overline{\eta}(s), \hat{X}_s^{i,N,M}, \mu_{(\overline{\eta}(s)-h)}^{X,N,M}) \rangle$$

$$+ \langle X_s^{i,N} - \hat{X}_s^{i,N,M}, b(\overline{\eta}(s), \hat{X}_s^{i,N,M}, \mu_{(\overline{\eta}(s)-h)}^{X,N,M}) - b(\overline{\eta}(s), \hat{X}_s^{i,N,M}, \mu_s^{X,N}) \rangle$$

$$+ \langle X_s^{i,N} - \hat{X}_s^{i,N,M}, b(\overline{\eta}(s), \hat{X}_s^{i,N,M}, \mu_s^{X,N}) - b(\overline{\eta}(s), \hat{X}_s^{i,N,M}, \mu_{(\overline{\eta}(s)-h)}^{X,N,M}) \rangle$$

$$\leq C \left( |X_s^{i,N} - \hat{X}_s^{i,N,M}|^2 + |X_s^{i,N} - \hat{X}_s^{i,N,M}| + h \right.$$ 

$$+ (1 + |\hat{X}_s^{i,N,M}|^2 + |\hat{X}_{\kappa(s)}^{i,N,M}|^2) |\hat{X}_s^{i,N,M} - \hat{X}_{\kappa(s)}^{i,N,M}|$$

$$\left. + (1 + |\hat{X}_{\kappa(s)}^{i,N,M}|^2 + |\hat{X}_{\kappa(s)}^{i,N,M}|^2) |\hat{X}_{\kappa(s)}^{i,N,M} - \hat{X}_{\kappa(s)}^{i,N,M}| \right),$$

where we have used the growth bounds on $b$ (in particular bounded in measure) along with several applications of Cauchy-Schwarz and Young’s inequality. Similar arguments yield the following bound for the diffusion,

$$|\sigma(s, X_s^{i,N}) - \sigma(\kappa(s), \hat{X}_s^{i,N,M})|$$

$$\leq C(h^{1/2} + |X_s^{i,N} - \hat{X}_s^{i,N,M}| + |\hat{X}_s^{i,N,M} - \hat{X}_{\kappa(s)}^{i,N,M}| + |\hat{X}_{\kappa(s)}^{i,N,M} - \hat{X}_{\kappa(s)}^{i,N,M}|).$$

Ultimately we need to take supremum and expected values, hence we wish to bound

$$E \left[ \sup_{0 \leq \tau \leq t \wedge \theta_m} \int_0^\tau 2\langle X_s^{i,N} - \hat{X}_s^{i,N,M}, (\sigma(s, X_s^{i,N}) - \sigma(\kappa(s), \hat{X}_s^{i,N,M}))dW_s^j \rangle \right].$$

We use Burkholder Davis Gundy inequality, however care is needed since the terminal time is a stopping time. It turns out the usual upper bound still holds (see for example
[25, pg. 226], hence we obtain by using Young’s inequality,

$$
E \left[ \sup_{0 \leq t \leq T \wedge \theta_m} \int_0^t \left( 2(X_{s}^{i,N} - \hat{X}_{s}^{i,N,M}, (\sigma(s, X_{s}^{i,N}) - \sigma(\kappa(s), \hat{X}_{\kappa(s)}^{i,N,M}))dW_{s}^{i} \right) \right] \\
\leq C E \left[ \left( \int_0^{T \wedge \theta_m} |X_{s}^{i,N} - \hat{X}_{s}^{i,N,M}|^2 \sum_{a=1}^l |\sigma_a(s, X_{s}^{i,N}) - \sigma_a(\kappa(s), \hat{X}_{\kappa(s)}^{i,N,M})|^2 ds \right)^{1/2} \right] \\
\leq \frac{1}{2} E \left[ \sup_{0 \leq s \leq T \wedge \theta_m} |X_{s}^{i,N} - \hat{X}_{s}^{i,N,M}|^2 \right] \\
+ C E \left[ \int_0^{T \wedge \theta_m} \sum_{a=1}^l |\sigma_a(s, X_{s}^{i,N}) - \sigma_a(\kappa(s), \hat{X}_{\kappa(s)}^{i,N,M})|^2 ds \right].
$$

Taking supremum and expectations of our original difference and using these bounds we obtain the following inequality

$$
\frac{1}{2} E \left[ \sup_{0 \leq t \leq T \wedge \theta_m} |X_{l \wedge \theta_m}^{i,N} - \hat{X}_{l \wedge \theta_m}^{i,N,M}|^2 \right] \\
\leq E \left[ \int_0^{T \wedge \theta_m} C (|X_{s}^{i,N} - \hat{X}_{s}^{i,N,M}|^2 + |X_{s}^{i,N} - \hat{X}_{s}^{i,N,M}|) \\
+ (1 + |\hat{X}_{s}^{i,N,M}|^2q + |\hat{X}_{\kappa(s)}^{i,N,M}|^2q) |\hat{X}_{s}^{i,N,M} - \hat{X}_{\kappa(s)}^{i,N,M}|^2 \\
+ h \left( 1 + |\hat{X}_{s}^{i,N,M}|^2q + |\hat{X}_{\kappa(s)}^{i,N,M}|^2q \right) |\hat{X}_{\kappa(s)}^{i,N,M} - \hat{X}_{\kappa(s)}^{i,N,M}|^2 \right] \\
+ C \sum_{a=1}^l \left( h + |X_{s}^{i,N} - \hat{X}_{s}^{i,N,M}|^2 + |X_{s}^{i,N,M} - \hat{X}_{\kappa(s)}^{i,N,M}|^2 \right) ds \right].
$$

The goal is to use a Grönwall type inequality, hence we want to bring the expectation inside the integral, collecting common terms and arguing as previously we obtain,

$$
E \left[ \sup_{0 \leq t \leq T \wedge \theta_m} |X_{l \wedge \theta_m}^{i,N} - \hat{X}_{l \wedge \theta_m}^{i,N,M}|^2 \right] \\
\leq C(hT + \int_0^T E \left[ \sup_{0 \leq t \leq s} |X_{r \wedge \theta_m}^{i,N} - \hat{X}_{r \wedge \theta_m}^{i,N,M}|^2 \right] + E \left[ \sup_{0 \leq r \leq s} |X_{r \wedge \theta_m}^{i,N} - \hat{X}_{r \wedge \theta_m}^{i,N,M}| \right] \\
+ E \left[ \left( 1 + |\hat{X}_{s}^{i,N,M}|^2q + |\hat{X}_{\kappa(s)}^{i,N,M}|^2q \right) |\hat{X}_{s}^{i,N,M} - \hat{X}_{\kappa(s)}^{i,N,M}|^2 \right] \mathbb{1}_{(s \leq \theta_m)} \\
+ E \left[ \left( 1 + |\hat{X}_{s}^{i,N,M}|^2q + |\hat{X}_{\kappa(s)}^{i,N,M}|^2q \right) |\hat{X}_{\kappa(s)}^{i,N,M} - \hat{X}_{\kappa(s)}^{i,N,M}|^2 \right] \mathbb{1}_{(s \leq \theta_m)} ds \right).$$

Noting \( \mathbb{1}_{(\cdot)} = \mathbb{1}_{(\cdot)}^2 \), we obtain via Cauchy-Schwarz inequality,

$$
E \left[ \left( 1 + |\hat{X}_{s}^{i,N,M}|^2q + |\hat{X}_{\kappa(s)}^{i,N,M}|^2q \right) |\hat{X}_{s}^{i,N,M} - \hat{X}_{\kappa(s)}^{i,N,M}|^2 \mathbb{1}_{(s \leq \theta_m)} \right] \\
\leq C(m) E \left[ |\hat{X}_{s}^{i,N,M} - \hat{X}_{\kappa(s)}^{i,N,M}|^4 \mathbb{1}_{(s \leq \theta_m)} \right].
$$

Noting that

$$
|\hat{X}_{s}^{i,N,M} - \hat{X}_{\kappa(s)}^{i,N,M}| \\
\leq |b \left( X_{\kappa(s)}^{i,N,M}, \mu_{(\kappa(s)-h)}X_{\kappa(s)}^{N,M} \right) |h + |\sigma \left( \kappa(s), \hat{X}_{\kappa(s)}^{i,N,M} \right) (W_{s}^{i} - W_{r}^{i})|,
$$
which implies,
\[
\mathbb{E}\left[|X^{i,N}_s - \hat{X}^{i,N}_s|^4 \mathbb{1}_{\{s \leq \theta_m\}}\right] \\
\leq C h \mathbb{E}\left[(1 + |X^{i,N}_s|^4 \mathbb{1}_{\{s \leq \theta_m\}}\right] \\
+ C \mathbb{E}\left[(1 + |X^{i,N}_s|^4 \mathbb{1}_{\{s \leq \theta_m\}}\right] \mathbb{E}\left[(W_i - W^{i}_s)^4\right] \leq C(m) h^2,
\]
where we used Lemma 5.6 to obtain the final inequality (note by assumption \(X_0 \in L^4(q+1)(\mathbb{R}^d)\)). Arguing in the exact same fashion also yields,
\[
\mathbb{E}\left[(1 + |\hat{X}^{i,N}_s|^2 + |X^{i,N}_s|^2)\right] \leq C(m) h^2.
\]
Substituting these bounds then implies,
\[
\mathbb{E}\left[\sup_{0 \leq t \leq T \wedge \theta'_m} |X^{i,N}_{t \wedge \theta'_m} - \hat{X}^{i,N}_{t \wedge \theta'_m}|^2\right] \\
\leq C(m) h + C \int_0^T \mathbb{E}\left[\sup_{0 \leq r \leq s} |X^{i,N}_{t \wedge \theta'_m} - \hat{X}^{i,N}_{t \wedge \theta'_m}|^2\right] + \mathbb{E}\left[\sup_{0 \leq r \leq s} |X^{i,N}_{t \wedge \theta'_m} - \hat{X}^{i,N}_{t \wedge \theta'_m}|^2\right] dt \leq (C(m)^{1/2} h^{1/2})^2 \exp(C T \int_0^T \exp(C t) dt) \leq C(m) h,
\]
which gives the result we set out to show.

We now can prove our main implicit scheme result.

**Proposition 3.7.** Let us define the error term as \(E_r(T)^i = X^{i,N}_T - \hat{X}^{i,N}_T\) and also let us note a more general version of Young’s inequality,
\[
x^s y \leq \frac{\delta s}{2} x^2 + \frac{2 - s}{2\delta s/(2-s)} y^{2/(2-s)}, \quad \forall \ x, \ y, \ \delta > 0.
\]
Hence,
\[
\mathbb{E}|X^{i,N}_T - \hat{X}^{i,N}_T|^s \leq 2^{s-1} \left(\mathbb{E}|X^{i,N}_T - \hat{X}^{i,N}_T|^s \mathbb{1}_{\{\tau_m^* > T, \ \eta_m^* > T\}}\right) \\
+ \mathbb{E}|X^{i,N}_T - \hat{X}^{i,N}_T|^s \mathbb{1}_{\{\tau_m^* > T, \ \eta_m^* > T\}} + \frac{\delta s}{2} \mathbb{E}[|E_r(T)^i|^2] + \frac{2 - s}{2\delta s/(2-s)} \mathbb{E}[\mathbb{1}_{\{\tau_m^* \leq T \text{ or } \eta_m^* \leq T\}}].
\]
From Lemma 5.8 we obtain,
\[
\mathbb{E}|\hat{X}^{i,N}_T - \hat{X}^{i,N}_T|^s \mathbb{1}_{\{\tau_m^* > T, \ \eta_m^* > T\}} \leq C(m, s) h^s.
\]
Also let us note,
\[
\mathbb{E}[|E_r(T)^i|^2] \leq 2 \mathbb{E}[|X^{i,N}_T|^2 + |\hat{X}^{i,N}_T|^2] \leq 2 C,
\]
where we have used Propositions 5.4 and 5.7. Hence for any \( \epsilon > 0 \), we can choose \( \delta \) such that,

\[
\frac{\delta s}{2} \mathbb{E}[|E^i|] \leq \frac{\epsilon}{3}.
\]

By subadditivity of measures, \( \mathbb{E} \left[ \mathbb{1}_{\{ \tau^i_m \leq T \text{ or } \eta^i_m \leq T \}} \right] \leq \mathbb{P}(\tau^i_m \leq T) + \mathbb{P}(\eta^i_m \leq T) \) and then Proposition 5.4, there exists \( m^* \) (dependent on \( \delta \)), such that for \( m \geq m^* \),

\[
\frac{2 - s}{2 \delta s/(2-s)} \mathbb{P}(\tau^i_m \leq T) \leq \frac{\epsilon}{3}.
\]

Then, noting by Lemma 5.11,

\[
\mathbb{E} \left[ \sup_{0 \leq t \leq T} |\hat{X}^{i,N,M}_{t \land \theta^i_m} - X^{i,N}_{t \land \theta^i_m}|^2 \right] \leq C(m)h.
\]

By Lemma 5.9, by taking \( h \) small enough for any \( \epsilon > 0 \), \( \mathbb{P}(\eta^i_m < T) \leq \epsilon \). Hence, for any \( \delta \) and \( m \), we can take \( h \) small enough such that,

\[
2^{s-1} \left( \mathbb{E} \left[ |X^{i,N}_{\tau^i_m} - \hat{X}^{i,N,M}_{\tau^i_m}| \mathbb{1}_{\{ \tau^i_m > T, \eta^i_m > T \}} \right] + \mathbb{E} \left[ |\hat{X}^{i,N}_{\tau^i_m} - \hat{X}^{i,N,M}_{\tau^i_m}| \mathbb{1}_{\{ \tau^i_m > T, \eta^i_m > T \}} \right] \right) + \frac{2 - s}{2 \delta s/(2-s)} \mathbb{P}(\eta^i_m \leq T) \leq \frac{\epsilon}{3},
\]

hence we can take \( \epsilon \to 0 \) by taking \( h \to 0 \), which completes the result. \( \square \)

REFERENCES

[1] 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, The Journal of Mathematical Neuroscience, 2 (2012), p. 10.

[2] F. Bolley, J. A. Cañizo, and J. A. Carrillo, Stochastic mean-field limit: non-lipschitz forces and swarming, Math Models Methods Appl Sci, 21 (2011).

[3] 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”, The Journal of Mathematical Neuroscience (JMN), 5 (2015), p. 19.

[4] M. Bossy and D. Talay, A stochastic particle method for the McKean-Vlasov and the Burgers equation, Mathematics of Computation of the American Mathematical Society, 66 (1997), pp. 157–192.

[5] A. Budhiraja and W.-T. Fan, Uniform in time interacting particle approximations for nonlinear equations of Patlak-Keller-Segel type, Electron. J. Probab., 22 (2017), pp. Paper No. 8, 37.

[6] R. Carmona, Lectures of BSDEs, Stochastic Control, and Stochastic Differential Games with Financial Applications, SIAM, 2016.

[7] R. Carmona and F. Delarue, Probabilistic Theory of Mean Field Games with Applications I, vol. 84 of Probability Theory and Stochastic Modelling, Springer International Publishing, 1st ed., 2017.

[8] J.-F. Chassagneux, A. Jacquier, and I. Mihaylov, An explicit Euler scheme with strong rate of convergence for stochastic SDEs with non-Lipschitz coefficients, SIAM Journal on Financial Mathematics, 7 (2016), pp. 993–1021.

[9] G. dos Reis, W. Salkeld, and J. Tugaut, Freidlin-Wentzell LDPs in path space for McKean-Vlasov equations and the functional iterated logarithm law. arXiv:1708.04961, To appear in Annals of Applied probability, 2017.

[10] G. dos Reis, G. Smith, and P. Tankov, Importance sampling for McKean-Vlasov SDEs, arXiv:1803.09220, (2018).

[11] W. Dreyer, P. Friz, F. Gajewski, C. Guhlke, and M. Maurelli, Stochastic model for LFP-electrodes, tech. report, WIAS preprint no. 2329, 2016.

[12] W. Dreyer, M. Gaberscek, C. Guhlke, R. Huth, and J. Jamnik, Phase transition in a rechargeable lithium battery, European Journal of Applied Mathematics, 22 (2011), pp. 267–290.
[13] W. Fang and M. B. Giles, Adaptive Euler-Maruyama method for SDEs with non-globally Lipschitz drift: Part I, finite time interval, arXiv preprint arXiv:1609.08101, (2016).

[14] E. Gobet and S. Pagliarani, Analytical approximations of non-linear SDEs of McKean-Vlasov type, Journal of Mathematical Analysis and Applications, (2018).

[15] D. J. Higham, X. Mao, and A. M. Stuart, Strong convergence of Euler-type methods for nonlinear stochastic differential equations, SIAM Journal on Numerical Analysis, 40 (2002), pp. 1041–1063.

[16] M. Hutzenthaler, A. Jentzen, and P. E. Kloeden, Strong and weak divergence in finite time of Euler’s method for stochastic differential equations with non-globally Lipschitz continuous coefficients, in Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, vol. 467, The Royal Society, 2011, pp. 1563–1576.

[17] M. Hutzenthaler, A. Jentzen, and P. E. Kloeden, Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients, The Annals of Applied Probability, 22 (2012), pp. 1611–1641.

[18] R. W. Keener, Theoretical statistics: Topics for a core course, Springer, 2011.

[19] A. Kohatsu-Higa and S. Ogawa, Weak rate of convergence for an Euler scheme of nonlinear SDE’s, Monte Carlo Methods and Applications, 3 (1997), pp. 327–345.

[20] A. Lionnet, G. dos Reis, and L. Szpruch, Time discretization of FBSDE with polynomial growth drivers and reaction-diffusion PDEs, Ann. Appl. Probab., 25 (2015), pp. 2563–2625.

[21] F. Malrieu, Convergence to equilibrium for granular media equations and their Euler schemes, Ann. Appl. Probab., 13 (2003), pp. 540–560.

[22] X. Mao and L. Szpruch, Strong convergence and stability of implicit numerical methods for stochastic differential equations with non-globally Lipschitz continuous coefficients, Journal of Computational and Applied Mathematics, 238 (2013), pp. 14–28.

[23] G. Milstein and M. V. Tretyakov, Numerical integration of stochastic differential equations with nonglobally Lipschitz coefficients, SIAM journal on numerical analysis, 43 (2005), pp. 1139–1154.

[24] D. S. Mitrinovic, J. Pečaric, and A. M. Fink, Inequalities involving functions and their integrals and derivatives, vol. 53, Springer Science & Business Media, 2012.

[25] P. E. Protter, Stochastic differential equations, Springer, 2005.

[26] S. T. Rachev and L. Rüschendorf, Mass transportation problems. Vol. II, Probability and its Applications (New York), Springer-Verlag, New York, 1998. Applications.

[27] S. Sabanis, A note on tamed Euler approximations, Electron. Comm. Probab., 18 (2013), pp. no. 47, 10.

[28] D. N. Tan, A stochastic Ginzburg-Landau equation with impulsive effects, Physica A: Statistical Mechanics and its Applications, 392 (2013), pp. 1962–1971.

[29] C. Villani, Optimal transport: old and new, vol. 338, Springer Science & Business Media, 2008.

[30] E. Zeidler, Nonlinear functional analysis and its applications. II/B, Springer-Verlag, New York, 1990. Nonlinear monotone operators, Translated from the German by the author and Leo F. Boron.