ACADÉMIE
DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus
Mathématique

Elias Hess-Childs, Matthew Rosenzweig and Sylvia Serfaty

Another look at regularity in transport-commutator estimates

Volume 364 (2026), p. 437-478

Online since: 27 July 2026

https://doi.org/10.5802/crmath.837

CENTRE MERSENNE

The Comptes Rendus. Mathématique *are a member of the Mersenne Center for open scientific publishing*
www.centre-mersenne.org — e-ISSN : 1778-3569

ACADÉMIE
DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus. Mathématique
2026, Vol. 364, p. 437-478
https://doi.org/10.5802/crmath.837

Research article
Mathematical physics, Partial differential equations

Another look at regularity in transport-commutator estimates

Elias Hess-Childs [a], **Matthew Rosenzweig** [a] **and Sylvia Serfaty** [b,c,d]

[a] Carnegie Mellon University, Department of Mathematical Sciences, Pittsburgh, PA, USA

[b] Sorbonne Université, CNRS, Université de Paris, Laboratoire Jacques-Louis Lions (LJLL), F-75005 Paris, France

[c] Institut Universitaire de France

[d] Courant Institute of Mathematical Sciences, New York University, USA

E-mails: ehesschi@andrew.cmu.edu, mrosenz2@andrew.cmu.edu, serfaty@cims.nyu.edu

Abstract. We are interested in how regular a transport velocity field must be in order to control Riesz-type commutators. Estimates for these commutators play a central role in the analysis of the mean-field limit and fluctuations for systems of particles with pairwise Riesz interactions through modulated energy techniques, which we start by reviewing. In these applications, the transport field is generated by the limiting PDE or by its adjoint linearized flow. Relaxing the regularity demanded of the transport field to control the commutator enlarges the class of limiting densities and fluctuation observables accessible to the method, especially at scaling-critical regularities.

Our first new result shows that the usual L^∞ assumption on the gradient of the velocity field cannot, in general, be relaxed to a BMO assumption. We construct counterexamples in all dimensions and all Riesz singularities $-2 < \mathsf{s} < \mathsf{d}$, except for the one-dimensional logarithmic endpoint $\mathsf{s} = 0$. At this exceptional endpoint, such a relaxation is possible, a fact related to the classical Coifman–Rochberg–Weiss commutator bound for the Hilbert transform. Our second result identifies a trade-off between the singularity of the interaction potential and the required regularity of the velocity field. Roughly speaking, smoother (less singular) interactions require stronger velocity control if one wants a commutator estimate in the natural energy seminorm determined by the potential. We formulate this principle for a broad class of potentials and show that, in the sub-Coulomb Riesz regime, the velocity regularity appearing in the known commutator inequality is sharp. Despite these negative findings, we show as our third result that a *defective* commutator estimate holds for almost-Lipschitz transport fields. Such a defective estimate, which is a consequence of the celebrated Brezis–Wainger–Hansson inequality, allows us to prove rates of convergence when the mean-field density belongs to the scaling-critical Sobolev space.

Keywords. Commutator estimates, Riesz interactions, Coulomb interactions, modulated energy, mean-field limits.

2020 Mathematics Subject Classification. 42B20, 82C22.

Funding. E.H.C. was supported by NSF grants DMS-2342349, DMS-2424139.
M.R. was supported by NSF grants DMS-2441170, DMS-2345533, DMS-2342349.
S.S. was supported by NSF grant DMS-2247846 and by Institut Universitaire de France.
This paper is based upon work supported by the National Science Foundation under Grant No. DMS-2424139, while the first author was in residence at the Simons Laufer Mathematical Sciences Institute in Berkeley, California, during the Fall 2025 semester.

Note. Article submitted by invitation.

Manuscript received 5 January 2026, revised 31 March 2026, accepted 19 May 2026, online since 27 July 2026.

This article is dedicated to the memory of Haïm Brezis whose life and work were a great source of inspiration

1. Introduction

1.1. *Mean-field limits*

In recent years, the study of *mean-field limits* for interacting particle systems has seen significant progress. A basic model is the first-order system[1]

$$\dot{x}_i^t = \frac{1}{N}\sum_{1\le j\le N:\, j\ne i} \mathbb{M}\nabla \mathsf{g}(x_i^t - x_j^t) - \mathsf{V}^t(x_i^t), \quad i \in [N]. \tag{1}$$

Here, g is an *interaction potential* to be specified, and V^t is a (possibly time-dependent) *external field.* $\mathbb{M}$ is a $\mathsf{d}\times\mathsf{d}$ matrix with real entries that specifies the type of dynamics, with $\mathbb{M} = -\mathbb{I}$ (gradient/dissipative) and $\mathbb{M}$ antisymmetric (Hamiltonian/conservative) being the most common choices. The mean-field limit refers to the convergence as $N\to\infty$ of the *empirical measure*

$$\mu_N^t := \frac{1}{N}\sum_{i=1}^N \delta_{x_i^t} \tag{2}$$

associated to a solution $X_N^t := (x_1^t,\dots,x_N^t)$ of the system (1). Assuming the initial points X_N^0 are such that μ_N^0 converges to a sufficiently regular measure μ^0, a formal calculation leads one to expect that for $t>0$, μ_N^t converges to the solution of the *mean-field equation*

$$\partial_t \mu^t = \operatorname{div}\big(\mu^t(\mathsf{V}^t - \mathbb{M}\nabla\mathsf{g} * \mu^t)\big). \tag{3}$$

Convergence of the empirical measure is qualitatively equivalent to *propagation of molecular chaos* (see [28,42] and references therein). This latter notion means that if $f_N^0(x_1,\dots,x_N) = \mathrm{Law}(X_N^0)$ is μ^0-chaotic (i.e. the k-point marginals $f_{N;k}^0 \rightharpoonup (\mu^0)^{\otimes k}$ as $N\to\infty$ for every fixed k), then $f_N^t = \mathrm{Law}(X_N^t)$ is μ^t-chaotic.

Mean-field limits for systems of the form (1) have a long history, beginning with regular drifts (typically, Lipschitz) and then progressing to increasingly more singular interactions over the years. In the interest of brevity, we refer to the introduction of our companion paper [46] for some guidance to the literature giving a proper review of these developments.

Instead, we focus on the much more recent treatment of singular interactions, model examples of which are the family of *log/Riesz* potentials[2] given by

$$\mathsf{g}(x) := \begin{cases} \frac{1}{\mathsf{s}}|x|^{-\mathsf{s}}, & \mathsf{s}\ne 0, \\ -\log|x|, & \mathsf{s}=0. \end{cases} \tag{4}$$

We consider any dimension $\mathsf{d}\ge 1$ and assume that $-2<\mathsf{s}<\mathsf{d}$. Up to a normalizing constant $\mathsf{c}_{\mathsf{d},\mathsf{s}}$, the potential g is the fundamental solution of the fractional Laplacian $(-\Delta)^{\frac{\mathsf{d}-\mathsf{s}}{2}} = |\nabla|^{\mathsf{d}-\mathsf{s}}$, i.e. $|\nabla|^{\mathsf{d}-\mathsf{s}}\mathsf{g} = \mathsf{c}_{\mathsf{d},\mathsf{s}}\delta_0$. The cases $0\le\mathsf{s}<\mathsf{d}$ and $-2<\mathsf{s}<0$ are respectively referred to as the *singular* and *nonsingular* regimes because of the behavior of g at the origin. The restriction $\mathsf{s}>-2$ is natural because the Riesz potential ceases to be conditionally positive definite at $\mathsf{s}=-2$.[3] The restriction

[1] More general dynamics include positive temperature, non-pairwise interactions, non-translation-invariant kernels, and second-order/kinetic models.

[2] Since the log case may be understood as the $\mathsf{s}\to 0$ limit of the Riesz case, we always mean the term "Riesz" to include the log case.

[3] We say that a kernel $k\colon \mathbb{R}^{\mathsf{d}}\times\mathbb{R}^{\mathsf{d}}\to\mathbb{R}$ is *conditionally positive definite (CPD)* iff for any signed Borel measure ρ with zero mass, it holds that $\int_{(\mathbb{R}^{\mathsf{d}})^2} k(x,y)\,d\rho(x)\,d\rho(y)\ge 0$. When the inequality is strict for any nonzero ρ, k is said to be *strictly* CPD (SCPD).

to the *potential* regime $\mathsf{s} < \mathsf{d}$, in which g is locally integrable, is to exclude the *hypersingular* case, which is not of the mean-field type considered in this paper (e.g. see [38,39]). The case $\mathsf{s} = \mathsf{d} - 2$ corresponds to the classical *Coulomb* potential from electrostatics/gravitation. Based on this threshold, it is convenient to call the interaction *sub-Coulomb* if $\mathsf{s} < \mathsf{d} - 2$ and *super-Coulomb* if $\mathsf{s} > \mathsf{d} - 2$. More generally, long-range interactions of the form (4) play an important role in physics [24], approximation theory [6], and machine learning [1,35,43,44] — just to name a few fields. Motivations for considering systems of the form (1) are extensively reviewed in [87].

Mean-field convergence for the full range $-2 < \mathsf{s} < \mathsf{d}$ has only been resolved in the last few years: the nonsingular case $-2 < \mathsf{s} < 0$ [80], the sub-Coulomb case $\mathsf{s} < \mathsf{d} - 2$, [16,41], the Coulomb/super-Coulomb case $\mathsf{d} - 2 \le \mathsf{s} < \mathsf{d}$ [5,17,25,85], and the full singular case $0 \le \mathsf{s} < \mathsf{d}$ [62]. See also [9,45,67,74] for further extensions incorporating noise. These advances are due in large part to the *modulated energy* method, which we review in the next subsection. For a more thorough review of the method, we refer to the third author's lecture notes [87, Part 2] and the second author's recent survey [72].

1.2. *Modulated energy and commutators*

Since the empirical measure μ_N^t associated with the microscopic system (1) is a weak solution of the mean-field equation (3), a natural approach for showing mean-field convergence is a weak-strong stability estimate for (3). For Riesz interactions, such stability is conveniently formulated in terms of the energy-based distance

$$\mathsf{F}_N(X_N,\mu) := \frac{1}{2}\int_{(\mathbb{R}^{\mathsf{d}})^2\setminus\triangle} \mathsf{g}(x-y)\,\mathrm{d}\Big(\frac{1}{N}\sum_{i=1}^N \delta_{x_i} - \mu\Big)(x)\,\mathrm{d}\Big(\frac{1}{N}\sum_{i=1}^N \delta_{x_i} - \mu\Big)(y), \tag{5}$$

where we excise the diagonal $\triangle := \{(x,y) \in (\mathbb{R}^{\mathsf{d}})^2 : x = y\}$ from the domain of integration in order to remove the infinite self-interaction of each particle. The *modulated energy* (5) originated in the study of the statistical mechanics of Coulomb/Riesz gases [64,81–83] and later was used in the derivation of mean-field dynamics [25,62,85] and following works. In the nonsingular case, where the diagonal contribution vanishes, the modulated energy coincides with the square of what is known in the statistics literature as *maximum mean discrepancy* (MMD) [31–33], a type of integral probability metric [61]. MMDs for varying choices of kernels are widely used in statistics and machine learning contexts as distances between probability measures (e.g. see [44,54,59]). We refer to [87, Chapter 4] for a comprehensive discussion of the modulated energy.

The modulated energy approach to mean-field convergence consists of establishing a Grönwall relation for $\mathsf{F}_N(X_N^t,\mu^t)$, where X_N^t is a solution of the microscopic system (1) and μ^t is a solution of the mean-field equation (3). An essential point in this approach is to control the derivative of F_N along a transport $v\colon \mathbb{R}^{\mathsf{d}} \to \mathbb{R}^{\mathsf{d}}$, that is

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \mathsf{F}_N\big((\mathbb{I}+tv)^{\oplus N}(X_N), (\mathbb{I}+tv)\#\mu\big) = \frac{1}{2}\int_{(\mathbb{R}^{\mathsf{d}})^2\setminus\triangle} \nabla\mathsf{g}(x-y)\cdot\big(v(x)-v(y)\big)\,\mathrm{d}(\mu_N-\mu)^{\otimes 2}(x,y), \tag{6}$$

where for any configuration $X_N = (x_1,\dots,x_N)$, μ_N denotes $\frac{1}{N}\sum_{i=1}^N \delta_{x_i}$, $\mathbb{I}\colon \mathbb{R}^{\mathsf{d}} \to \mathbb{R}^{\mathsf{d}}$ is the identity and $(\mathbb{I}+tv)^{\oplus N}(X_N) := \big(x_1+tv(x_1),\dots,x_N+tv(x_N)\big)$. For applications to the mean-field limit, v is the velocity field of the limiting evolution (3). For applications to central limit theorems (CLTs) for the fluctuations (the next-order description beyond the mean-field limit), v is the gradient of a test function evolved along the adjoint linearized mean-field flow [18,48]. These inequalities are at the core of the "transport" approach to fluctuations of canonical Gibbs ensembles [4,55,63,86] and also appear in the loop (Dyson–Schwinger) equations in the random matrix theory literature, e.g. [3,7,8]. Higher-order variations of the modulated energy are also important, a point we return to in Section 5.3.

The desired control is a functional inequality of the form

$$\big|(6)\big| \le C_1\big(\mathsf{F}_N(X_N,\mu) + C_2 N^{-\alpha}\big), \tag{7}$$

where $\alpha > 0$ depends on d and s, $C_1 > 0$ is a constant depending on d,s,v, and $C_2 > 0$ depends on $\mathsf{d},\mathsf{s},\mu$. In applications, an estimate of the form (7) turns the derivative formula (6) into a Grönwall- or more generally Osgood-type bound for $F_N(X_N^t,\mu^t)$. The coercivity of the modulated energy (see, e.g., [46, Proposition 2.15]) then implies that if the initial modulated energy vanishes, then one obtains the convergence of the empirical measure μ_N^t to the mean-field density μ^t with an explicit rate in negative-order Sobolev distances, locally uniformly in time. The method is inherently quantitative and, importantly, requires no knowledge of the microscopic dynamics in terms of individual particle trajectories — information that is difficult to obtain in practice. As alluded to above, the method works because there is a weak-strong stability principle for the mean-field equation (3) in the energy distance (see, e.g., [87, Section 6.2.2]).

The form of control (7) was first proved by Leblé and the third author in [55] in the 2D Coulomb case, then generalized to the super-Coulomb case in [85]. The proof relied on identifying a *stress-energy tensor* structure in (6) and using integration by parts. Subsequently, the second author [67] observed that the expressions (6) may also be viewed as the quadratic form of a *commutator*, akin to the famous Calderón commutator [15,21,22,84]. Indeed, ignoring the exclusion of the diagonal,

$$\int_{(\mathbb{R}^{\mathsf{d}})^2}\big(v(x)-v(y)\big)\cdot\nabla \mathsf{g}(x-y)f(x)g(y)\,\mathrm{d}x\,\mathrm{d}y = \mathsf{c}_{\mathsf{d},\mathsf{s}}\Big\langle f,\big[v\cdot,\nabla|\nabla|^{\mathsf{s}-\mathsf{d}}\big]g\Big\rangle_{L^2}, \tag{8}$$

where $\langle\cdot,\cdot\rangle_{L^2}$ denotes the L^2 inner product. For this reason, estimates of the form (7) are called *commutator estimates.*

This commutator perspective was used by the last two authors and Q. H. Nguyen [62] to show the desired functional inequality for the full singular regime as well as a broader class of Riesz-type interactions including Lennard–Jones-type potentials. The argument of [62] has also been extended to the nonsingular regime [80]. As recognized in [78], the stress-energy and commutator perspectives are in fact two sides of the same coin. These functional inequalities are crucial not only for proving CLTs for fluctuations of Riesz gases [55,63,86], but also for deriving mean-field limits [19,62,66,69,70,78,85] and large deviation principles [45]. They have also been used to show the joint classical and mean-field limits for quantum systems of particles [29], and the supercritical mean-field limits of classical [36,58,71,77] and quantum systems of particles [65,68].

Let us also mention that for gradient dynamics at positive temperature (i.e. with Brownian noise in the microscopic dynamics (1)), the modulated energy combines naturally with the previously used relative entropy [27,34,49,75] in the form of the *modulated free energy* [9–11,14,19,75,76], which is well-suited to proving entropic propagation of chaos and can even handle logarithmically attractive interactions [10,11,20]. Commutator estimates play the same essential role in the positive temperature setting.

The main problem for these functional inequalities has been to determine the optimal size of the additive error, i.e. the exponent α in (7), which a priori depends on d and s. In formulating optimality, the appropriate comparison is with the minimal size of $\mathsf{F}_N(X_N,\mu)$. When $0 \le \mathsf{s} < \mathsf{d}$, it is known that $\mathsf{F}(X_N,\mu) + \frac{\log(N\|\mu\|_{L^\infty})}{2\mathsf{d}N}\mathbf{1}_{\mathsf{s}=0} \ge -C\|\mu\|_{L^\infty}^{\frac{\mathsf{s}}{\mathsf{d}}}N^{\frac{\mathsf{s}}{\mathsf{d}}-1}$ for $C = C(\mathsf{d},\mathsf{s}) > 0$ and this lower bound is sharp [46,78].[4] In controlling the magnitude of (6), one needs a right-hand side which is also nonnegative. Thus, the best error one may hope for is of size $N^{\frac{\mathsf{s}}{\mathsf{d}}-1}$. When $-2 < \mathsf{s} < 0$, the modulated energy is nonnegative (it is a squared MMD as mentioned above), and its minimal

[4]The term $\frac{\log(N\|\mu\|_{L^\infty})}{2\mathsf{d}N}$ in the $\mathsf{s}=0$ case is *not* an additive error like $\|\mu\|_{L^\infty}^{\frac{\mathsf{s}}{\mathsf{d}}}N^{\frac{\mathsf{s}}{\mathsf{d}}-1}$, but rather a consequence of the fact that the right quantity to consider is $\mathsf{F}(X_N,\mu) + \frac{\log(N\|\mu\|_{L^\infty})}{2\mathsf{d}N}\mathbf{1}_{\mathsf{s}=0}$.

value is $\propto N^{\frac{\mathsf{s}}{\mathsf{d}}-1}$ [47]. Moreover, the functional inequality holds without additive error because the interaction is nonsingular [80].

Only very recently has the program to show the sharp $N^{\frac{\mathsf{s}}{\mathsf{d}}-1}$ error rate been completed by the authors and collaborators: first, the Coulomb case [55,71,86], then the entire (super-)Coulomb case [78], and finally the entire singular Riesz case [46]. The completion of this program has yielded the sharp rates of convergence for the mean-field limit in the modulated energy distance. It has also been crucial for studying fluctuations around the mean-field limit in and out of thermal equilibrium [48,55,63,73,79,86].

Summarizing the state of the art for these commutator estimates is the following theorem, combining the results of [46,62,74,78,80]. In particular, taking $p = \infty$ in (12), (13) below yields the announced sharp $N^{\frac{\mathsf{s}}{\mathsf{d}}-1}$ additive error.

To compactify the notation, given a point configuration X_N, reference measure μ, and vector field v, let us write

$$\mathsf{A}_1[X_N,\mu,v] := \int_{(\mathbb{R}^{\mathsf{d}})^2\setminus\triangle} \nabla \mathsf{g}(x-y)\cdot\big(v(x)-v(y)\big)\,\mathrm{d}\Big(\frac{1}{N}\sum_{i=1}^N \delta_{x_i}-\mu\Big)^{\otimes 2}(x,y). \tag{9}$$

The subscript 1 indicates that A_1 is a first-order variation/commutator, a point whose significance will become clearer in Section 5.3.

Theorem 1. *Let $-2<\mathsf{s}<\mathsf{d}$. Let $\mu \in L^1\cap L^p$ for $\frac{\mathsf{d}}{\mathsf{d}-\mathsf{s}} < p \le \infty$ with $\int_{\mathbb{R}^{\mathsf{d}}} \mathrm{d}\mu = 1$, and associated to μ, define the length scales*

$$\lambda := \big(N\|\mu\|_{L^p}\big)^{-\frac{1}{\mathsf{d}}}, \tag{10}$$

$$\kappa := \big(N^{\frac{1}{\mathsf{s}+1}}\|\mu\|_{L^p}\big)^{-\frac{1}{\mathsf{d}}}. \tag{11}$$

When $-2<\mathsf{s}\le 0$, assume that $\int_{(\mathbb{R}^{\mathsf{d}})^2}|\mathsf{g}|(x-y)\,\mathrm{d}|\mu|^{\otimes 2}<\infty$.[5] *Let v be a Lipschitz vector field. Then for any pairwise distinct point configuration $X_N \in (\mathbb{R}^{\mathsf{d}})^N$, the following hold.*

- *If $\mathsf{s} \ge \max(0,\mathsf{d}-2)$, then*

$$\big|\mathsf{A}_1[X_N,\mu,v]\big| \le C\|\nabla v\|_{L^\infty}\Big(\mathsf{F}_N(X_N,\mu) - \frac{\log\lambda}{2N}\mathbf{1}_{\mathsf{s}=0} + C_p\|\mu\|_{L^p}\lambda^{\frac{\mathsf{d}(p-1)}{p}-\mathsf{s}}\Big) \tag{12}$$

(cf. [78]).

- *If $0\le\mathsf{s}<\mathsf{d}-2$, then for any $\mathsf{a}\in(\mathsf{d},\mathsf{d}+2)$,*

$$\begin{aligned}&\big|\mathsf{A}_1[X_N,\mu,v]\big| \\ &\quad\le C\Big(\|\nabla v\|_{L^\infty} + C_{\mathsf{a}}\big\||\nabla|^{\frac{\mathsf{a}}{2}}v\big\|_{L^{\frac{2\mathsf{d}}{\mathsf{a}-2}}}\mathbf{1}_{\mathsf{a}>2}\Big)\Big(\mathsf{F}_N(X_N,\mu) - \frac{\log\lambda}{2N}\mathbf{1}_{\mathsf{s}=0} + C_p\|\mu\|_{L^p}\lambda^{\frac{\mathsf{d}(p-1)}{p}-\mathsf{s}}\Big)\end{aligned} \tag{13}$$

(cf. [46]), and for any $\frac{\mathsf{d}}{\mathsf{d}-\mathsf{s}-1}<p\le\infty$,

$$\begin{aligned}&\big|\mathsf{A}_1[X_N,\mu,v]\big| \\ &\quad\le C\Big(\|\nabla v\|_{L^\infty} + \big\||\nabla|^{\frac{\mathsf{d}-s}{2}}v\big\|_{L^{\frac{2\mathsf{d}}{\mathsf{d}-\mathsf{s}-2}}}\Big)\Big(\mathsf{F}_N(X_N,\mu) + C_p\|\mu\|_{L^p}\kappa^{\frac{\mathsf{d}(p-1)}{p}-\mathsf{s}}(1-(\log\kappa)\mathbf{1}_{\mathsf{s}=0})\Big)\end{aligned} \tag{14}$$

(cf. [62,74]).

- *If $-2<\mathsf{s}<0$, then*

$$\big|\mathsf{A}_1[X_N,\mu,v]\big| \le C\Big(\|\nabla v\|_{L^\infty} + \big\||\nabla|^{\frac{\mathsf{d}-\mathsf{s}}{2}}v\big\|_{L^{\frac{2\mathsf{d}}{\mathsf{d}-\mathsf{s}-2}}}\mathbf{1}_{\mathsf{s}<\mathsf{d}-2}\Big)\mathsf{F}_N(X_N,\mu). \tag{15}$$

(cf. [80]).

In all cases, the constant $C>0$ depends only on d and s, and the constants $C_{\mathsf{a}}, C_p>0$ additionally depend on a, p, respectively.

[5] This condition is to ensure that the modulated energy is finite. When $0<\mathsf{s}<\mathsf{d}$, it is implied by the assumption that $\mu\in L^1\cap L^p$ for $p>\frac{\mathsf{d}}{\mathsf{d}-\mathsf{s}}$. When $\mathsf{s}=0$, it is implied by a logarithmic moment assumption $\int_{\mathbb{R}^{\mathsf{d}}}\log\big(1+|x|\big)\,\mathrm{d}|\mu|<\infty$; and when $-2<\mathsf{s}<0$, it is implied by the moment assumption $\int_{\mathbb{R}^{\mathsf{d}}}|x|^{|\mathsf{s}|/2}\,\mathrm{d}|\mu|<\infty$.

Given a measure ν with density in $\mathcal{S}(\mathbb{R}^{\mathsf{d}})$ and $\int_{\mathbb{R}^{\mathsf{d}}} \mathrm{d}\nu = 1$, letting $f := \nu - \mu$, we can find a sequence of pairwise distinct point configurations X_N such that, as $N \to \infty$,

$$\mathsf{F}_N(X_N, \mu) \to \frac{1}{2}\int_{(\mathbb{R}^{\mathsf{d}})^2} \mathsf{g}(x-y) f(x) f(y)\,\mathrm{d}x\,\mathrm{d}y = \frac{\mathsf{c}_{\mathsf{d},\mathsf{s}}}{2}\|f\|^2_{\dot{H}^{\frac{\mathsf{s}-\mathsf{d}}{2}}} \tag{16}$$

and

$$\mathsf{A}_1[X_N, \mu, v] \to \frac{1}{2}\int_{(\mathbb{R}^{\mathsf{d}})^2} \nabla\mathsf{g}(x-y)\cdot\big(v(x)-v(y)\big) f(x) f(y)\,\mathrm{d}x\,\mathrm{d}y = \frac{1}{2}\Big\langle f, \big[v\cdot, \nabla|\nabla|^{\mathsf{s}-\mathsf{d}}\big] f\Big\rangle_{L^2}. \tag{17}$$

Thus, a commutator estimate with an additive error involving the singular zero-mass measure $\frac{1}{N}\sum_{i=1}^N \delta_{x_i} - \mu$ implies a commutator estimate without an additive error for the regular zero-mass measure f. More precisely, Theorem 1 implies the following standard or *unrenormalized* commutator estimate (see [62, Proposition 3.1]).

Proposition 2. *Let* $\mathsf{d} \geq 1$ *and* $-2 < \mathsf{s} < \mathsf{d}$*. Then there exists* $C > 0$*, depending only on* d *and* s*, such that for all* $f, g \in \mathcal{S}(\mathbb{R}^{\mathsf{d}})$*,*[6] *we have*

$$\left|\int_{(\mathbb{R}^{\mathsf{d}})^2} \big(v(x)-v(y)\big)\cdot\nabla\mathsf{g}(x-y) f(x) g(y)\,\mathrm{d}x\,\mathrm{d}y\right| \\ \leq C\Big(\|\nabla v\|_{L^\infty} + \big\||\nabla|^{\frac{\mathsf{d}-\mathsf{s}}{2}} v\big\|_{L^{\frac{2\mathsf{d}}{\mathsf{d}-\mathsf{s}-2}}} \mathbf{1}_{\mathsf{s}<\mathsf{d}-2}\Big)\|f\|_{\dot{H}^{\frac{\mathsf{s}-\mathsf{d}}{2}}}\|g\|_{\dot{H}^{\frac{\mathsf{s}-\mathsf{d}}{2}}}. \tag{18}$$

It is straightforward to check that the preceding right-hand side scales the same way as the left-hand side, and that both v terms on the right-hand side scale the same way as well. Moreover, the presence of the $\dot{H}^{\frac{\mathsf{s}-\mathsf{d}}{2}}$ seminorm is natural, given this corresponds to the square root of the energy associated to g.

In fact, the philosophy of [62,67,78], and to a more implicit degree the works [25,55,85,86], proceed in the reverse order: first show the unrenormalized commutator estimate of Proposition 2, then combine it with a renormalization procedure, implemented via charge smearing (i.e. regularization of the δ_{x_i}), to allow for the irregular test measure $\frac{1}{N}\sum_{i=1}^N \delta_{x_i}$. It is this renormalization step, which is the most technically involved part of the proof, that leads to the additive N-dependent errors in the singular case $0 \leq \mathsf{s} < \mathsf{d}$. For this reason, the estimates (12), (13), (14) are sometimes referred to as *renormalized* commutator estimates to distinguish them from the unrenormalized estimate (18). In the nonsingular case $-2 < \mathsf{s} < 0$, no renormalization is necessary, and one can apply a commutator estimate directly, leading to the absence of additive errors in (15).

These state-of-the-art commutator estimates are in large part the subject of the second author's recent survey [72]. We refer the reader to this work for a more extensive review, in particular more detailed — yet still concise — outlines of the proofs of these results and the evolution of techniques behind them.

1.3. *Transport regularity*

The question we take up in this paper is the optimal regularity assumption for the transport velocity v in the aforementioned commutator estimates. Compared to sharpness of the additive errors in these estimates, discussed above, this question has received little attention in the literature. As we explain below, this regularity question is not only for the sake of mathematics, but is also motivated by applications. In mean-field problems, the transport is precisely the limiting velocity $u_t = \mathsf{V}_t - \mathbb{M}\nabla\mathsf{g} * \mu_t$. Thus, every commutator assumption on v becomes a regularity assumption on the solution μ_t of the limiting PDE. Lowering the regularity demanded of v

[6] If $\mathsf{s} = 0$, then implicit is the requirement that f and g are zero-mean to avoid the low-frequency divergence in $\dot{H}^{-\frac{\mathsf{d}}{2}}(\mathbb{R}^{\mathsf{d}})$.

therefore enlarges the class of limiting solutions for which the modulated-energy method yields quantitative mean-field convergence. The regularity question is equally relevant at next order in the context of fluctuations, where one applies the same estimates to the adjoint linearized flow. Lowering the regularity assumptions on v allows for the study of fluctuations around less regular mean-field solutions. It also allows for a broader, rougher class of test functions to be considered as observables for statistics of the system.

In this direction, it is natural to ask whether the Lipschitz assumption on v may be weakened while still preserving the scaling. For example, the space of *bounded mean oscillation* (BMO) [51] is a common relaxation of L^∞, and one can ask whether $\|\nabla v\|_{L^\infty}$ may be replaced by $\|\nabla v\|_{\mathrm{BMO}}$. It is also unclear whether, in the sub-Coulomb regime $\mathsf{s} < \mathsf{d}-2$, the additional term of the form $\||\nabla|^{\frac{\mathsf{a}}{2}} v\|_{L^{\frac{2\mathsf{d}}{\mathsf{a}-2}}}$ is genuinely necessary and how small the exponent a may be taken (by Sobolev embedding, $\||\nabla|^{\frac{\mathsf{a}}{2}} v\|_{L^{\frac{2\mathsf{d}}{\mathsf{a}-2}}} \lesssim \||\nabla|^{\frac{\mathsf{a}'}{2}} v\|_{L^{\frac{2\mathsf{d}}{\mathsf{a}'-2}}}$ for $2(\mathsf{d}+1) > \mathsf{a}' \geq \mathsf{a}$). Moreover, there is a gap between our two sub-Coulomb estimates (13), (14). In the former, the N-dependent additive error is sharp but requires a stronger regularity condition on v (given that $\mathsf{a} > \mathsf{d}$ by assumption) compared to the latter, in the N-dependent error is not sharp.

It is a general principle in PDE that if an equation has a scaling invariance, then a natural function space to study the equation is one that is invariant under the equation's scaling [53]. Such *scaling-critical* or *scaling-invariant* spaces are expected to be thresholds for the well-posedness/ill-posedness of the equation along with other phenomena, such as finite-time blowup vs. global existence.

For $a, b, c > 0$, set $\tilde{\mu}(t,x) := a\mu(bt, cx)$. When $\mathsf{V}^t = 0$, the reader may check that if μ is a solution to (3), then $\tilde{\mu}$ is also a solution provided that $bc^{\mathsf{d}-2-\mathsf{s}} = a$. Fixing $b = 1$ so that the solution lifespan is preserved under rescaling, we arrive at the relation $a = c^{\mathsf{d}-2-\mathsf{s}}$. We seek (semi)norms that are invariant under the transformation $\mu \mapsto \tilde{\mu}$.

For $1 \leq p \leq \infty$, the reader may check that

$$\|\tilde{\mu}\|_{L^p} = c^{\mathsf{d}-\mathsf{s}-2-\frac{\mathsf{d}}{p}} \|\mu\|_{L^p}, \tag{19}$$

which implies that $p = \frac{\mathsf{d}}{\mathsf{d}-\mathsf{s}-2}$ is the scaling-critical exponent. It is possible to choose such a p if and only if $\mathsf{s} \leq \mathsf{d}-2$. In other words, the potential can be at most Coulomb in order for there to be a critical L^p space. More generally, letting $\dot{W}^{a,p}$ denote the homogeneous Bessel potential (fractional Sobolev) space (e.g. see [89, Section 3.3] or [30, Section 1.3]), we see that

$$\|\tilde{\mu}\|_{\dot{W}^{a,p}} = c^{\mathsf{d}-\mathsf{s}-2+a-\frac{\mathsf{d}}{p}} \|\mu\|_{\dot{W}^{a,p}}, \tag{20}$$

which shows that $\dot{W}^{a,p}$ is a critical space if and only if $a = \frac{\mathsf{d}}{p} + \mathsf{s} + 2 - \mathsf{d}$. Other scales of function spaces, such as Hölder–Zygmund spaces (e.g. see [89, Section 4]), are also interesting to consider as critical spaces. We will return to this point in the next section.

When $\mu \in \dot{W}^{\frac{\mathsf{d}}{p}+\mathsf{s}+2-\mathsf{d},p}$, for $1 < p < \infty$, some basic Fourier analysis shows that the associated mean-field velocity $v = \mathbb{M}\nabla \mathsf{g} * \mu$ is in $\dot{W}^{\frac{\mathsf{d}}{p}+1,p}$. Unfortunately, $\dot{W}^{\frac{\mathsf{d}}{p}+1,p}$ does not embed into the homogeneous Lipschitz space $\dot{W}^{1,\infty}$. If $f \in \dot{W}^{\frac{\mathsf{d}}{p}+1,p}$, then one only has

$$\|\nabla f\|_{\mathrm{BMO}} \leq C\|f\|_{\dot{W}^{\frac{\mathsf{d}}{p}+1,p}}.$$

Even when $\mathsf{s} + 2 - \mathsf{d} = k$ is an integer and taking $p = \infty$, having $\mu \in \dot{W}^{k,\infty}$ does not imply that $v \in \dot{W}^{1,\infty}$. This is problematic because all of the commutator estimates of Theorem 1 feature $\|\nabla v\|_{L^\infty}$ on the right-hand side. Thus, to show mean-field convergence when the mean-field density μ^t belongs to the critical space, it would be desirable to have a commutator estimate with a regularity assumption that is satisfied when $\mu \in \dot{W}^{\frac{\mathsf{d}}{p}+\mathsf{s}+2-\mathsf{d},p}$.

1.4. *Our contributions*

Having laid out the question of interest concerning transport regularity in commutator estimates for the modulated energy and motivated such inequalities by their role in studying mean-field and related limits, we come to the new contributions of the present paper. We describe them here at a high level. Precise statements of our main results (Theorems 3, 5, 8, 10, and Corollary 6) and detailed comments about the proofs are deferred to Section 2.

Our first contribution is to show that estimate (18) can fail if either of the vector field norms is weakened (Theorem 3). We also derive necessary conditions for analogous commutator estimates to hold for a broader class of interaction potentials (Theorem 5). In particular, the norm controlling v must involve enough derivatives to compensate for the high-frequency decay of $\widehat{\mathsf{g}}$. This demonstrates a more general principle of regularity trade-off between the interaction and transport for commutator estimates. All of these results are obtained through explicit counterexamples. Given that the renormalized commutator estimates of Theorem 1 imply unrenormalized commutator estimates, as explained in Section 1.2, it suffices to show these counterexamples when $\frac{1}{N}\sum_{i=1}^N \delta_{x_i} - \mu$ is replaced by a zero-mean Schwartz function.

Our second contribution concerns mean-field convergence when the solution μ^t of the mean-field equation belongs to the critical space $\dot{W}^{\frac{\mathsf{d}}{p}+\mathsf{s}+2-\mathsf{d},p}$ for $1 < p < \infty$. Provided that $p \le \frac{2\mathsf{d}}{\mathsf{d}-\mathsf{s}-2}$, one has

$$\left\| |\nabla|^{\frac{\mathsf{d}-\mathsf{s}}{2}} u^t \right\|_{L^{\frac{2\mathsf{d}}{\mathsf{d}-\mathsf{s}-2}}} \lesssim \|\mu^t\|_{\dot{W}^{\frac{\mathsf{d}}{p}+\mathsf{s}+2-\mathsf{d},p}}.$$

But, since in general ∇u^t is only in BMO when $\mu^t \in \dot{W}^{\frac{\mathsf{d}}{p}+\mathsf{s}+2-\mathsf{d},p}$, the results described in the previous paragraph give a negative answer to our earlier question about the possibility of a commutator estimate holding in this case, regardless of the size of the additive error. That said, the situation is not completely hopeless. As recognized in [69,70], even in this borderline regularity case, one can still obtain a Grönwall-type[7] estimate for the modulated energy provided that a *defective* commutator estimate holds.

More precisely, we say that a defective version holds if there exists a locally integrable function $\rho\colon [0,\infty) \to [0,\infty)$, called the *defect*, such that for all $\epsilon > 0$,

$$\begin{aligned}\int_{(\mathbb{R}^{\mathsf{d}})^2\setminus\triangle} \big(v(x)-v(y)\big)\cdot\nabla\mathsf{g}(x-y)\,\mathrm{d}\left(\frac{1}{N}\sum_{i=1}^N \delta_{x_i}-\mu\right)^{\otimes 2}(x,y) \\ \le \rho(\epsilon) C_v\big(\mathsf{F}_N(X_N,\mu)+C_\mu N^{-\alpha}\big) + \epsilon^\beta N^\gamma C_v'\big(\mathsf{F}_N(X_N,\mu)+C_\mu' N^{-\alpha}\big), \qquad (21)\end{aligned}$$

where $\alpha > 0$, $\beta,\gamma \ge 0$ and $C_v, C_v', C_\mu, C_\mu' > 0$ are constants depending on d,s and quantitatively on norms of v,μ, respectively.

If $\rho(\epsilon) = O(|\log\epsilon|)$ as $\epsilon \to 0^+$, it turns out that a defective commutator estimate suffices to conclude an estimate for the modulated energy that in turn implies mean-field convergence, at least up to some time $T_* > 0$ which is independent of N, but a priori is smaller than the maximal time of existence for μ^t. These works considered only the Coulomb case, where the natural scale-invariant space is L^∞, whose norm is a nonincreasing (conserved when $\mathbb{M}$ is antisymmetric) quantity when $\mathsf{V}^t = 0$. In the present paper, we generalize the results of [69,70] to the entire Riesz case without any restrictions on the interval of time for which mean-field convergence holds (Theorem 10). Importantly, in the sub-Coulomb case $\mathsf{s} < \mathsf{d}-2$, we show that if the initial mean-field density $\mu^\circ \in L^{\frac{\mathsf{d}}{\mathsf{d}-\mathsf{s}-2}}$, which is scaling-critical, then, there is a unique global solution to the mean-field equation (3) which is a (quantitative) mean-field limit of the microscopic system (1).

The main technical result is a new defective commutator estimate (Theorem 8) with a defect $\rho(\epsilon) = O\big(|\log\epsilon|^\theta\big)$ for $\theta \in (0,1)$, in contrast to the $\theta = 1$ defect present in [69,70]. This improvement

[7] More precisely, an estimate in the form of Osgood's lemma, which is a generalization of Grönwall's lemma. See [2, Section 3.1.1].

is made possible by the celebrated *Brezis–Wainger–Hansson inequality* [13,37]: if $1 < p < \infty$, $1 \le q \le \infty$, and $aq > \mathsf{d}$, then

$$\forall\, f \in \mathcal{S}(\mathbb{R}^{\mathsf{d}}), \quad \|f\|_{L^\infty} \le C\|f\|_{\dot{W}^{\frac{\mathsf{d}}{p},p}}\left(1+\log^{\frac{p-1}{p}}\left(1+\frac{\|f\|_{\dot{W}^{a,q}}}{\|f\|_{\dot{W}^{\frac{\mathsf{d}}{p},p}}}\right)\right), \tag{22}$$

where $C = C(\mathsf{d}, p, a, q) > 0$. Remark that the exponent $\frac{p-1}{p}$ is sharp; and moreover, no estimate of the form

$$\|f\|_{L^\infty} \le \varepsilon\|f\|_{\dot{W}^{\frac{\mathsf{d}}{p},p}}\left(1+\log^{\frac{p-1}{p}}\left(1+\frac{\|f\|_{\dot{W}^{a,q}}}{\|f\|_{\dot{W}^{\frac{\mathsf{d}}{p},p}}}\right)\right)+C_\varepsilon, \tag{23}$$

for any $\varepsilon > 0$, can hold.

Brezis and Wainger [13, Theorem 1] first proved the estimate (22), which is a generalization of an earlier $\mathsf{d} = p = 2$ inequality due to Brezis and Gallouët [12]. In the PDE literature, inequalities of this form are sometimes called Brezis–Gallouët inequalities or logarithmic interpolation. Independently, Hansson [37] proved an inequality expressed in potential-theoretic language that is equivalent to (22), and an alternative proof is also given in [26].

The Brezis–Wainger–Hansson inequality is a substitute for the false endpoint Sobolev embedding $\|f\|_{L^\infty} \lesssim \|f\|_{\dot{W}^{\frac{\mathsf{d}}{p},p}}$, except when $\mathsf{d} = p = 1$. Although functions in $\dot{W}^{\frac{\mathsf{d}}{p},p}$ are not bounded, they are still exponentially integrable in a quantitative sense made precise by the famous Trudinger–Moser inequality [60,90]. In some sense, the estimate (22) is dual to Trudinger–Moser (see [13, Theorem A]).

1.5. *Notation*

Finally, let us review the basic notation used throughout the paper, mostly following the conventions of [46,62,74,78].

Given nonnegative quantities A and B, we write $A \lesssim B$ if there exists a constant $C > 0$, independent of A and B, such that $A \le CB$. If $A \lesssim B$ and $B \lesssim A$, we write $A \sim B$. Throughout, C will denote a generic constant that may change from line to line.

$\mathbb{N}$ denotes the natural numbers excluding zero. For $N \in \mathbb{N}$, we abbreviate $[N] := \{1, \dots, N\}$. $\mathbb{R}_+$ denotes the positive reals. Given $x \in \mathbb{R}^{\mathsf{d}}$ and $r > 0$, $B(x,r)$ and $\partial B(x,r)$ respectively denote the ball and sphere centered at x of radius r. Given a function f, we denote its support by $\operatorname{supp} f$. The notation $\nabla^{\otimes k} f$ denotes the k-tensor field with components $(\partial_{i_1} \cdots \partial_{i_k} f)_{1 \le i_1,\dots,i_k \le \mathsf{d}}$. For $x \in \mathbb{R}$, $\langle x\rangle := \left(1+|x|^2\right)^{1/2}$ is the Japanese bracket.

$\mathcal{P}(\mathbb{R}^{\mathsf{d}})$ denotes the space of Borel probability measures on $\mathbb{R}^{\mathsf{d}}$. If μ is absolutely continuous with respect to Lebesgue measure, we shall abuse notation by letting μ denote both the measure and its Lebesgue density. When the measure is clearly understood to be Lebesgue, we shall simply write $\int_{\mathbb{R}^{\mathsf{d}}} f$ instead of $\int_{\mathbb{R}^{\mathsf{d}}} f\,\mathrm{d}x$.

We use the notation $\hat{f}$ or $\widehat{f}$ to denote the Fourier transform $\widehat{f}(\xi) := \int_{\mathbb{R}^{\mathsf{d}}} f(x)e^{-2\pi i\xi\cdot x}\,\mathrm{d}x$ and let $\langle\nabla\rangle$ and $|\nabla|$ denote the Fourier multipliers with symbols $\langle 2\pi\xi\rangle$ and $|2\pi\xi|$, respectively.

Several function spaces appear in the paper. $\mathcal{S}(\mathbb{R}^{\mathsf{d}})$ and $\mathcal{S}'(\mathbb{R}^{\mathsf{d}})$ denote the spaces of Schwartz functions and tempered distributions, respectively. For $s \in \mathbb{R}$ and $1 < p < \infty$, $W^{s,p}(\mathbb{R}^{\mathsf{d}})$ denotes the inhomogeneous Bessel potential space defined by

$$W^{s,p} := \left\{f \in \mathcal{S}' : \langle\nabla\rangle^s f \in L^p\right\}, \qquad \|f\|_{W^{s,p}} := \left\|\langle\nabla\rangle^s f\right\|_{L^p}. \tag{24}$$

When $s = k$ is an integer, $W^{s,p}$ agrees with the usual Sobolev space

$$W^{k,p} = \left\{f \in L^p : \nabla^{\otimes j} f \in L^p,\ 0 \le j \le k\right\}, \qquad \|f\|_{W^{k,p}} := \sum_{j=0}^{k} \|\nabla^{\otimes j} f\|_{L^p}, \tag{25}$$

the latter, of course, making sense for the endpoints $p = 1,\infty$ too. When $p = 2$, we use the customary notation $W^{s,2} = H^s$. The homogeneous Bessel potential space $\dot{W}^{s,p}(\mathbb{R}^{\mathsf{d}})$ is defined by

$$\dot{W}^{s,p} := \left\{f \in \mathcal{S}'/\mathscr{P} : |\nabla|^s f \in L^p\right\}, \qquad \|f\|_{\dot{W}^{s,p}} := \left\||\nabla|^s f\right\|_{L^p}, \tag{26}$$

where $\mathscr{P}(\mathbb{R}^{\mathsf{d}})$ denotes the space of polynomials on $\mathbb{R}^{\mathsf{d}}$ and $\mathcal{S}'/\mathscr{P}$ denotes the quotient space where we identify any two tempered distributions whose difference is a polynomial. This identification turns the seminorm $\|\cdot\|_{\dot{W}^{s,p}}$ into a norm. Similarly, for $s = k$ integer, $\dot{W}^{s,p}$ agrees with

$$\dot{W}^{k,p} = \left\{f \in \mathcal{S}'/\mathscr{P} : \nabla^{\otimes k} f \in L^p\right\}, \qquad \|f\|_{\dot{W}^{k,p}} := \|\nabla^{\otimes k} f\|_{L^p}. \tag{27}$$

It is a basic fact that for $s \geq 0$, $f \in W^{s,p}$ iff $f \in L^p \cap \dot{W}^{s,p}$. For $\theta > 0$, we let $\mathcal{C}^\theta(\mathbb{R}^{\mathsf{d}})$ denote the inhomogeneous Hölder–Zygmund space, which for $\theta \in (0,1]$ is defined by $f \in C_{\mathrm{loc}}$ such that

$$\|f\|_{\mathcal{C}^\theta} := \|f\|_{L^\infty} + \sup_{|h|>0} \frac{\left\|f(\cdot+h) + f(\cdot-h) - 2f(\cdot)\right\|_{L^\infty}}{|h|^\theta} < \infty, \tag{28}$$

and for $\theta > 1$, we inductively define $\|f\|_{\mathcal{C}^\theta} := \|f\|_{L^\infty} + \|\nabla f\|_{\mathcal{C}^{\theta-1}}$. $\mathcal{C}^\theta$ coincides with the Besov space $B^\theta_{\infty,\infty}$, and the norms are equivalent (see [2, Section 2.11]). For noninteger values of θ, it holds that $\mathcal{C}^\theta = C^{\lfloor\theta\rfloor,\{\theta\}}$, the space of $\lfloor\theta\rfloor$-times continuously differentiable functions such that the $\lfloor\theta\rfloor$-th derivative is $\{\theta\}$-Hölder continuous, which we endow with its usual norm. When θ is an integer, $\mathcal{C}^\theta$ is strictly larger than $W^{\theta,\infty} = C^{\theta-1,1}$. The homogeneous Hölder–Zygmund space $\dot{\mathcal{C}}^\theta$ is defined to be the space of all $f \in C_{\mathrm{loc}}$ such that

$$\|f\|_{\dot{\mathcal{C}}^\theta} := \sup_{|h|>0} \frac{\left\|f(\cdot+h) + f(\cdot-h) - 2f(\cdot)\right\|_{L^\infty}}{|h|^\theta} < \infty \tag{29}$$

when $\theta \in (0,1]$ and inductively for $\theta > 1$ as before.

As is customary in the literature, a superscript $\cdot$ denotes the homogeneous space/seminorm.

2. Main results

In this section, we present the precise statements of our main results previously advertised in Section 1.3 and provide some comments on their proofs.

2.1. *BMO inequalities*

Our first result demonstrates that, in general, the Lipschitz control on the transport field in (18) cannot be relaxed to $\|\nabla v\|_{\mathrm{BMO}}$. Theorem 3 below shows that, except for the 1D logarithmic endpoint $(\mathsf{d},\mathsf{s}) = (1,0)$, such a replacement is false. When $(\mathsf{d},\mathsf{s}) = (1,0)$, the commutator estimate *does* in fact hold.

Theorem 3.

(i) *Let* $\mathsf{d} \geq 1$ *and* $-2 < \mathsf{s} < \mathsf{d}$ *with* $(\mathsf{d},\mathsf{s}) \neq (1,0)$. *For every* $M > 0$, *there exist a* C^∞ *compactly supported vector field* v *and zero-mean Schwartz functions* f *and* g *such that*

$$\frac{\left|\int_{(\mathbb{R}^{\mathsf{d}})^2} \big(v(x) - v(y)\big)\cdot\nabla\mathsf{g}(x-y) f(x) g(y)\right|}{\left(\|\nabla v\|_{\mathrm{BMO}} + \left\||\nabla|^{\frac{\mathsf{d}-\mathsf{s}}{2}} v\right\|_{L^{\frac{2\mathsf{d}}{\mathsf{d}-\mathsf{s}-2}}} \mathbf{1}_{\mathsf{s}<\mathsf{d}-2}\right)\|f\|_{\dot{H}^{\frac{\mathsf{s}-\mathsf{d}}{2}}} \|g\|_{\dot{H}^{\frac{\mathsf{s}-\mathsf{d}}{2}}}} \geq M. \tag{30}$$

(ii) *If* $(\mathsf{d},\mathsf{s}) = (1,0)$, *then there exists* $C > 0$ *such that for all* v *with* $v' \in \mathrm{BMO}$ *and all* $f,g \in \mathcal{S}$ *with zero mean, we have*

$$\left|\int_{\mathbb{R}^2} \frac{v(x) - v(y)}{x-y} f(x) g(y)\right| \leq C\|v'\|_{\mathrm{BMO}} \|f\|_{\dot{H}^{-\frac{1}{2}}} \|g\|_{\dot{H}^{-\frac{1}{2}}}. \tag{31}$$

Letting H denote the Hilbert transform $Hf(x) := \mathrm{P.V.}\frac{1}{\pi}\int_{\mathbb{R}} \frac{f(y)}{y-x}\,\mathrm{d}y$, (31) is equivalent to

$$\big\|[v,H]\big\|_{\dot{H}^{-1/2}\to\dot{H}^{1/2}} \lesssim \|v'\|_{\mathrm{BMO}}. \tag{32}$$

Thus, (31) may be viewed as a Sobolev-space Coifman–Rochberg–Weiss-type commutator theorem [23], which in its most elementary form says that $\big\|[v,H]\big\|_{L^p\to L^p} \lesssim \|v\|_{\mathrm{BMO}}$.

The proofs of Theorem 3(i) and 3(ii) are presented in Sections 3.1 and 3.2, respectively. The first assertion of the theorem is shown through an explicit counterexample. We construct a velocity field v with $\nabla v \in \mathrm{BMO}$ and a mild logarithmic singularity at the origin and test against a family of rescaled, compactly supported, zero-mean functions f_r. The commutator grows like $\log\log(1/r)$ as $r \to 0$, while the denominator in (30) remains uniformly bounded, yielding an arbitrarily large ratio. The proof of the second assertion essentially involves appropriately integrating by parts and grouping terms into components that can be individually bounded via classical Kenig–Ponce–Vega-type commutator estimates [52,56].

2.2. *Regularity trade-off*

The BMO counterexample indicates that the Lipschitz control $\|\nabla v\|_{L^\infty}$ plays an essential role in (18), so we do not attempt to weaken it further. Our second main result quantifies the trade-off between the regularity of the interaction potential and the high-frequency regularity of the transport field required to control the commutator in the natural energy norm. Since this phenomenon is not specific to Riesz potentials, we work with a larger class of potentials for which the interaction energy defines a positive definite form.

Assumption 4 (Admissible interaction potentials). *Suppose that* $\mathsf{g}\colon \mathbb{R}^{\mathsf{d}} \setminus \{0\} \to \mathbb{R}$ *satisfies the following:*

(1) $\mathsf{g} \in C^\infty\big(\mathbb{R}^{\mathsf{d}} \setminus \{0\}\big)$*;*
(2) g *is radially symmetric*[8]
(3) g *is conditionally positive definite (see footnote 3) and* $\widehat{\mathsf{g}}$ *nonincreasing.*

Associated to g, we define the energy seminorm

$$\|f\|_{\mathsf{g}}^2 := \int_{(\mathbb{R}^{\mathsf{d}})^2} \mathsf{g}(x-y)f(x)f(y) = \int_{\mathbb{R}^{\mathsf{d}}} \widehat{\mathsf{g}}|\widehat{f}|^2, \tag{33}$$

which we note is nonnegative (possibly $+\infty$) for zero-mean f. When dealing with inhomogeneous potentials, it is convenient to introduce the notation $\mathsf{g}_t(x) := \mathsf{g}(x/t)$ for $t > 0$.

Given a function $\sigma\colon [0,\infty) \to (0,\infty]$, we denote the Fourier multiplier with symbol σ by $\sigma(|\nabla|)f := \mathcal{F}^{-1}\big(\sigma(|\xi|)\widehat{f}(\xi)\big)$. Theorem 5 below shows that if σ grows too slowly relative to $\widehat{\mathsf{g}}$ at high frequency, then the commutator cannot be controlled by the energy norm $\|\cdot\|_{\mathsf{g}_t}$ and the velocity field norm $\big\|\sigma(|\nabla|)v\big\|_{L^p}$.

Theorem 5. *Suppose that* g *satisfies Assumption 4. Fix* $t > 0$ *and* $p \in [1,\infty]$. *Assume furthermore that*

$$r^2\widehat{\mathsf{g}}(r) \to 0 \quad \text{as } r \to \infty \tag{34}$$

and $\sigma\colon [0,\infty) \to [0,\infty]$ *satisfies*

$$\lim_{r\to\infty} \sqrt{\widehat{\mathsf{g}}(tr)}\,\|\sigma\|_{L^\infty([r-1,r+1])} = 0, \quad 2 \le p \le \infty, \tag{35}$$

$$\lim_{r\to\infty} \sqrt{\widehat{\mathsf{g}}(tr)}\,\max_{n\le \mathsf{d}+1} \|\partial_r^n\sigma\|_{L^\infty([r-1,r+1])} = 0, \quad 1 \le p < 2. \tag{36}$$

[8] Throughout the paper, we use radial symmetry to abuse notation by writing $\mathsf{g}(|x|) = \mathsf{g}(x)$.

Then for every $M \geq 0$, there exist a C^∞ vector field v and zero-mean Schwartz functions f, g such that

$$\frac{\left|\int_{(\mathbb{R}^\mathsf{d})^2} \big(v(x)-v(y)\big)\cdot\nabla \mathsf{g}(x-y) f(x) g(y)\right|}{\big(\|\nabla v\|_{L^\infty} + \left\|\sigma(|\nabla|)v\right\|_{L^p}\big)\|f\|_{\mathsf{g}_t}\|g\|_{\mathsf{g}_t}} \geq M. \tag{37}$$

The decay condition (34) forces a genuine trade-off: without it, one can be in a regime where a commutator estimate holds with $\|\nabla v\|_{L^\infty}$ alone. Condition (35) means that on unit-width frequency shells $|\xi| \sim r$, the multiplier $\sigma(r)$ grows strictly slower than $\widehat{\mathsf{g}}(tr)^{-1/2}$; and (36) extends this growth condition to derivatives of σ up to high enough order. Thus, Theorem 5 may be read as saying that any commutator bound measured in $\|\cdot\|_{\mathsf{g}_t}$ forces v to possess at least "as many derivatives" as are needed to compensate for $\sqrt{\widehat{\mathsf{g}}}$ at high frequency.

Note that since $\widehat{\mathsf{g}}$ is nonincreasing, if (35) holds for some $t_* > 0$, then it holds for all $t \geq t_*$. Furthermore, if $\widehat{\mathsf{g}}(\xi)$ is comparable to a polynomial in ξ, then if (35) holds for some $t > 0$, it holds for any $t > 0$. When $1 \leq p < 2$, some regularity of the symbol is needed for $\sigma(|\nabla|)v \in L^p$, but the condition (36) can almost certainly be further weakened in terms of the number of derivatives (cf. the assumptions on the symbol in the Hörmander–Mikhlin multiplier theorem). Lastly, considering unit-width frequency shells $[r-1, r+1]$ is not essential: one could consider shells of any fixed radius mutatis mutandis.

Specializing Theorem 5 to Riesz potentials g defined by (4), one has $\widehat{\mathsf{g}}(\xi) \propto |\xi|^{\mathsf{s}-\mathsf{d}}$. The condition (34) is satisfied if $\mathsf{s} < \mathsf{d}-2$, i.e. g is sub-Coulomb. The condition (35) is equivalent to $\sup_{\tau\in[r-1,r+1]}\sigma(\tau) \ll |r|^{\frac{\mathsf{d}-\mathsf{s}}{2}}$ as $r \to \infty$. In particular, taking $\sigma(\tau) = \tau^{\frac{\mathsf{a}}{2}}$, we see that among the scaling-invariant norms $\left\||\nabla|^{\frac{\mathsf{a}}{2}} v\right\|_{L^p}$ with $p = \frac{2\mathsf{d}}{\mathsf{a}-2}$, the minimal exponent for which the commutator estimate holds is $\mathsf{a} = \mathsf{d}-\mathsf{s}$. Decreasing a, equivalently increasing p, yields a strictly weaker norm for which the commutator estimate fails.

Finally, to emphasize this point of regularity trade-off, we note that Theorem 5 implies that v must be C^∞ for the commutator estimate to hold for any interaction potential g with $\widehat{\mathsf{g}}$ decaying super-polynomially. Indeed, if for any $n \geq 0$, $\left|\widehat{\mathsf{g}}(\xi)\right| \lesssim_n |\xi|^{-n}$, then (37) holds for any $\sigma(|\nabla|) = |\nabla|^n$. In particular, if $\widehat{\mathsf{g}}$ decays sub-exponentially, then by the Paley–Wiener–Schwartz theorem, v must be at least analytic for a commutator estimate to hold. The regularity requirements are even more stringent when $\widehat{\mathsf{g}}$ has sub-Gaussian decay: if a commutator estimate were to hold, v must be super-analytic.

We summarize the above described consequences of Theorem 5 with the following corollary.

Corollary 6.

(i) *Suppose that* g *is given by* (4) *for* $\mathsf{s} < \mathsf{d}-2$. *Then for all* $2 < \mathsf{a} < \mathsf{d}-\mathsf{s}$ *and* $M > 0$, *there exist a C^∞ vector field v and zero-mean Schwartz functions f and g such that*

$$\frac{\left|\int_{(\mathbb{R}^\mathsf{d})^2} \big(v(x)-v(y)\big)\cdot\nabla \mathsf{g}(x-y) f(x) g(y)\right|}{\left(\|\nabla v\|_{L^\infty} + \left\||\nabla|^{\frac{\mathsf{a}}{2}} v\right\|_{L^{\frac{2\mathsf{d}}{\mathsf{a}-2}}}\right)\|f\|_{\dot H^{\frac{\mathsf{s}-\mathsf{d}}{2}}}\|g\|_{\dot H^{\frac{\mathsf{s}-\mathsf{d}}{2}}}} \geq M. \tag{38}$$

(ii) *Suppose that* g *is admissible and* $\widehat{\mathsf{g}}$ *has super-polynomial decay. Then for any $n, M, t > 0$, there exist a C^∞ vector field v and zero-mean Schwartz functions f and g such that*

$$\frac{\left|\int_{(\mathbb{R}^\mathsf{d})^2} \big(v(x)-v(y)\big)\cdot\nabla \mathsf{g}(x-y) f(x) g(y)\right|}{\left\|\langle\nabla\rangle^n v\right\|_{L^1}\|f\|_{\mathsf{g}_t}\|g\|_{\mathsf{g}_t}} \geq M. \tag{39}$$

(iii) *Suppose that* g *is admissible and for some* $c, m > 0$, $\left|\widehat{\mathsf{g}}(\xi)\right| \lesssim e^{-c|\xi|^m}$. *Then for any* $b < \frac{ct^m}{2}$ *and* $M > 0$, *there exist a* C^∞ *vector field* v *and zero-mean Schwartz functions* f, g *such that*

$$\frac{\left|\int_{(\mathbb{R}^{\mathsf{d}})^2} \big(v(x) - v(y)\big) \cdot \nabla \mathsf{g}(x-y) f(x) g(y) y\right|}{\|e^{b|\nabla|^m} v\|_{L^1} \|f\|_{\mathsf{g}_t} \|g\|_{\mathsf{g}_t}} \geq M. \tag{40}$$

Remark 7. Suppose g is such that there exists $t > 0$ so that $|x|\left|\nabla \mathsf{g}(x)\right| \lesssim \mathsf{g}(x/t)$. Such a condition is obviously satisfied if g is polynomial, exponential, or Gaussian. Then using $\left|v(x) - v(y)\right| \leq \|\nabla v\|_{L^\infty} |x-y|$, we obtain for $f, g \geq 0$ the bound

$$\begin{aligned}\left|\int_{(\mathbb{R}^{\mathsf{d}})^2} \big(v(x) - v(y)\big) \cdot \nabla \mathsf{g}(x-y) f(x) g(y)\right| &\lesssim \|\nabla v\|_{L^\infty} \int_{(\mathbb{R}^{\mathsf{d}})^2} \mathsf{g}\left(\frac{x-y}{t}\right) f(x) g(y) \\ &\lesssim \|\nabla v\|_{L^\infty} \|f\|_{\mathsf{g}_t} \|g\|_{\mathsf{g}_t}.\end{aligned} \tag{41}$$

This observation shows that under the assumption f, g have a definite sign, only Lipschitz regularity is needed for a commutator estimate to hold. In light of Corollary 6, a general commutator estimate for f, g without a definite sign necessarily requires much higher regularity on v.

The proof of Theorem 5 is given in Section 3.3. We construct v, f, g so that their Fourier supports sit on a few carefully chosen (and well separated) frequency shells. This forces one contribution to the commutator to be “off-resonant” and therefore small while the other contribution is “resonant” and stays bounded below. At the same time, the same frequency localization makes the velocity norm essentially depend on $\sup_{r \in [k-1, k+1]} \sigma(r)$ when $2 \leq p \leq \infty$, while $\|f\|_{\mathsf{g}_t}$ scales like $\sqrt{\widehat{\mathsf{g}}(tk)}$; after normalizing by $\|f\|_{\mathsf{g}_t}$ this produces a factor $1/\sqrt{\widehat{\mathsf{g}}(tk)}$ in the final ratio. Letting $k \to \infty$ then produces the desired ratio growth.

2.3. *Defective commutator estimates and mean-field convergence*

We now come to our last set of results on mean-field convergence when the solution μ^t of the mean-field equation (3), for g as in (4), belongs to the critical space $\dot{W}^{\frac{\mathsf{d}}{p}+\mathsf{s}+2-\mathsf{d},p}$ for $1 < p < \infty$.

Although the commutator estimate fails when v is not Lipschitz (except for $(\mathsf{d},\mathsf{s}) = (1,0)$) by Theorem 3(i), as is the case for $v = \mathbb{M}\nabla \mathsf{g} * \mu^t$ when $\mu^t \in \dot{W}^{\frac{\mathsf{d}}{p}+\mathsf{s}+2-\mathsf{d},p}$, a defective commutator estimate still holds. This is the content of Theorem 8 below. We exclude the case $\mathsf{s} = 0$ from the theorem because doing so simplifies the presentation and this case is already covered by the argument of [69].[9]

Theorem 8. *Assume that* $-2 < \mathsf{s} < \mathsf{d}$ *and* $\mathsf{s} \neq 0$. *Let* $v \in \dot{W}^{\frac{\mathsf{d}}{p}+1,p}$ *for* $1 < p < \infty$ *and* $\mu \in L^1$ *with* $\int_{\mathbb{R}^{\mathsf{d}}} d\mu = 1$ *and* $\int_{(\mathbb{R}^{\mathsf{d}})^2} |\mathsf{g}|(x-y)\, d|\mu|^{\otimes 2} < \infty$. *Further, assume the following:*

(a) *if* $-2 < \mathsf{s} \leq -1$, *then* $\int_{\mathbb{R}^{\mathsf{d}}} |x|^{|\mathsf{s}|-1} d|\mu| < \infty$*;*
(b) *if* $-1 < \mathsf{s} < 0$, *then* $\int_{\mathbb{R}^{\mathsf{d}}} |x|^r d|\mu| < \infty$ *for some* $r < \frac{|\mathsf{s}|}{2}$*;*[10]
(c) *if* $-1 \leq \mathsf{s} < \mathsf{d}-1$, *then* $\mu \in L^q$ *for some* $\frac{\mathsf{d}}{\mathsf{d}-\mathsf{s}-1} < q \leq \infty$ *(*$q = 1$ *is allowed if* $\mathsf{s} = -1$*);*
(d) *if* $\mathsf{s} \geq \mathsf{d}-1$, *then* $\mu \in \dot{C}^\theta$ *for some* $\theta > \mathsf{s}+1-\mathsf{d}$.[11]

[9] Strictly speaking, [69] only considers the case $(\mathsf{d},\mathsf{s}) = (2,0)$, but an examination of the argument shows that it applies to all log cases, replacing L^∞ by the appropriate critical space.

[10] Obviously, by interpolation, if $\int_{\mathbb{R}^{\mathsf{d}}} |x|^{r'} d|\mu| < \infty$ for some $r' \geq \frac{r}{2}$, then $\int_{\mathbb{R}^{\mathsf{d}}} |x|^r d|\mu| < \infty$ for any $0 \leq r \leq r'$, in particular for some $r < \frac{|\mathsf{s}|}{2}$.

[11] If $\mu \in L^1 \cap \dot{C}^\theta$ for some $\theta > 0$, then necessarily $\mu \in L^p$ for any $1 \leq p \leq \infty$ because of the interpolation estimate

$$\|f\|_{L^p} \leq \|f\|_{L^1}^{1-\frac{\mathsf{d}(p-1)}{p(\mathsf{d}+\theta)}} \|f\|_{\dot{C}^\theta}^{\frac{\mathsf{d}(p-1)}{p(\mathsf{d}+\theta)}}.$$

Then for any pairwise distinct configuration $X_N \in (\mathbb{R}^{\mathsf{d}})^N$ and $\epsilon > 0$, the following holds.

- *If* $\max(\mathsf{d}-2,0) \le \mathsf{s} < \mathsf{d}$*, then*

$$\begin{aligned}
&\left|\mathsf{A}_1[X_N,\mu,v]\right| \\
&\le C_p\|v\|_{\dot W^{\frac{\mathsf{d}}{p}+1,p}}\left(1+|\log\epsilon|\right)^{1-\frac{1}{p}}\left(\mathsf{F}_N(X_N,\mu)+C_q\|\mu\|_{L^q}\lambda^{\frac{\mathsf{d}(q-1)}{q}-\mathsf{s}}\right) \\
&\quad + CN^{\frac{2(\mathsf{s}+1)}{\mathsf{s}}-1}\|v\|_{\dot{\mathcal{C}}^1}\epsilon\left(\mathsf{F}_N(X_N,\mu)+C_p\|\mu\|_{L^p}\lambda^{\frac{\mathsf{d}(p-1)}{p}-\mathsf{s}}\right)^{\frac{\mathsf{s}+1}{\mathsf{s}}} \\
&\quad + \epsilon\left(1+\|\mu\|_{L^1}\right)\|v\|_{\dot{\mathcal{C}}^1}\left(C_q\|\mu\|_{L^1}^{1-\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}}\|\mu\|_{L^q}^{\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}}+C_\theta\|\mu\|_{L^1}^{1-\frac{\mathsf{s}-\mathsf{d}+1}{\theta}}\left(\|\mu\|_{\dot{\mathcal{C}}^\theta}^{\frac{\mathsf{s}-\mathsf{d}+1}{\theta}}+\epsilon^{\mathsf{d}-\mathsf{s}-1}\|\mu\|_{L^\infty}^{\frac{\mathsf{s}-\mathsf{d}+1}{\theta}}\right)\mathbf{1}_{\mathsf{s}\ge\mathsf{d}-1}\right).
\end{aligned}$$

- *If* $0 < \mathsf{s} < \mathsf{d}-2$*, then*

$$\begin{aligned}
&\left|\mathsf{A}_1[X_N,\mu,v]\right| \\
&\le \left(C_p\|v\|_{\dot W^{\frac{\mathsf{d}}{p}+1,p}}\left(1+|\log\epsilon|\right)^{1-\frac{1}{p}}+C_a\left\||\nabla|^{\frac{\mathsf{a}}{2}}v\right\|_{L^{\frac{2\mathsf{d}}{\mathsf{a}-2}}}\mathbf{1}_{\mathsf{a}>2}\right)\left(\mathsf{F}_N(X_N,\mu)+C_q\|\mu\|_{L^q}\lambda^{\frac{\mathsf{d}(q-1)}{q}-\mathsf{s}}\right) \\
&\quad + CN^{\frac{2(\mathsf{s}+1)}{\mathsf{s}}-1}\|v\|_{\dot{\mathcal{C}}^1}\epsilon\left(\mathsf{F}_N(X_N,\mu)+C_p\|\mu\|_{L^p}\lambda^{\frac{\mathsf{d}(p-1)}{p}-\mathsf{s}}\right)^{\frac{\mathsf{s}+1}{\mathsf{s}}} \\
&\quad + \epsilon\left(1+\|\mu\|_{L^1}\right)\|v\|_{\dot{\mathcal{C}}^1}C_q\|\mu\|_{L^1}^{1-\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}}\|\mu\|_{L^q}^{\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}}
\end{aligned}$$

or

$$\begin{aligned}
&\left|\mathsf{A}_1[X_N,\mu,v]\right| \\
&\le \left(C_p\|v\|_{\dot W^{\frac{\mathsf{d}}{p}+1,p}}\left(1+|\log\epsilon|\right)^{1-\frac{1}{p}}+C\left\||\nabla|^{\frac{\mathsf{d}-\mathsf{s}}{2}}v\right\|_{L^{\frac{2\mathsf{d}}{\mathsf{d}-\mathsf{s}-2}}}\right)\left(\mathsf{F}_N(X_N,\mu)+C_q\|\mu\|_{L^q}\kappa^{\frac{\mathsf{d}(q-1)}{q}-\mathsf{s}}\right) \\
&\quad + CN^{\frac{2(\mathsf{s}+1)}{\mathsf{s}}-1}\|v\|_{\dot{\mathcal{C}}^1}\epsilon\left(\mathsf{F}_N(X_N,\mu)+C_p\|\mu\|_{L^p}\lambda^{\frac{\mathsf{d}(p-1)}{p}-\mathsf{s}}\right)^{\frac{\mathsf{s}+1}{\mathsf{s}}} \\
&\quad + \epsilon\left(1+\|\mu\|_{L^1}\right)\|v\|_{\dot{\mathcal{C}}^1}C_q\|\mu\|_{L^1}^{1-\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}}\|\mu\|_{L^q}^{\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}}.
\end{aligned}$$

- *If* $-2 < \mathsf{s} < 0$*, then*

$$\begin{aligned}
&\left|\mathsf{A}_1[X_N,\mu,v]\right| \\
&\le \left(C_p\|v\|_{\dot W^{\frac{\mathsf{d}}{p}+1,p}}\left(1+|\log\epsilon|\right)^{1-\frac{1}{p}}+C\left\||\nabla|^{\frac{\mathsf{d}-\mathsf{s}}{2}}v\right\|_{L^{\frac{2\mathsf{d}}{\mathsf{d}-\mathsf{s}-2}}}\right)\mathsf{F}_N(X_N,\mu) \\
&\quad + \left(C_{\vartheta,\vartheta'}\|v\|_{\dot{\mathcal{C}}^1}\epsilon^{\vartheta'}\left(\int_{\mathbb{R}^{\mathsf{d}}}|x|^{|\mathsf{s}|-\vartheta}d\mu+\left(1+\|\mu\|_{L^1}\right)\mathsf{F}_N^{\frac{|\mathsf{s}|-\vartheta}{|\mathsf{s}|}}(X_N,\mu)\right)+C\|v\|_{\dot{\mathcal{C}}^1}\epsilon\right)\mathbf{1}_{-1<\mathsf{s}<0} \\
&\quad + C\|v\|_{\dot{\mathcal{C}}^1}\epsilon\left(\int_{\mathbb{R}^{\mathsf{d}}}|x|^{|\mathsf{s}|-1}d\mu+\left(1+\|\mu\|_{L^1}\right)\mathsf{F}_N^{\frac{|\mathsf{s}|-1}{|\mathsf{s}|}}(X_N,\mu)\right)\mathbf{1}_{-2<\mathsf{s}\le-1} \\
&\quad + \epsilon\left(1+\|\mu\|_{L^1}\right)\|v\|_{\dot{\mathcal{C}}^1}\left(\int_{\mathbb{R}^{\mathsf{d}}}|x|^{|\mathsf{s}|-1}d\mu+\mathsf{F}_N^{\frac{|\mathsf{s}|-1}{|\mathsf{s}|}}(X_N,\mu)\mathbf{1}_{-2<\mathsf{s}\le-1}+C_q\|\mu\|_{L^1}^{1-\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}}\|\mu\|_{L^q}^{\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}}\mathbf{1}_{-1\le\mathsf{s}<0}\right).
\end{aligned}$$

Above, $\vartheta \in \left[|\mathsf{s}|-r, |\mathsf{s}|\right)$, $0 < \vartheta' < \vartheta$, $C = C(\mathsf{d},\mathsf{s}) > 0$ *and* $C_\vartheta, C_p, C_q, C_a, C_\theta > 0$ *additionally depend on* $\vartheta, p, q, a, \theta$*, respectively.*

Remark 9. The endpoint $p = \infty$ also holds if we replace $\dot W^{1,\infty}$ with the Zygmund space $\dot{\mathcal{C}}^1$. The reader should remember that, per our notation (recall Section 1.5), the limit as $p \to \infty$ of $\dot W^{\frac{d}{p}+1,p}$ is not the homogeneous Lipschitz space, but rather the ill-behaved space $\left\{f \in \mathcal{S}' : \langle\nabla\rangle f \in L^\infty\right\}$, which does not coincide with the Lipschitz space, and is usually not considered in the harmonic analysis literature. See Section 5.4 for further comments on this point.

With this defective commutator estimate in hand, one can apply it to $v = \mu^t$ to obtain a Grönwall-type bound for the modulated energy (a more precise estimate is given during the proof of Theorem 10 in Section 4.4). This bound implies mean-field convergence provided that the initial modulated energy vanishes sufficiently fast as $N \to \infty$, an assumption which holds a.s. for most interesting choices of randomization of the initial particle positions (see Remark 11 below).

Theorem 10. *Let* $\mathsf{d} \geq 1$ *and* $-2 < \mathsf{s} < \mathsf{d}$ *and* $\mathsf{s} \neq 0$. *Suppose that for* $T > 0$, $\int_0^T \Big(\|\mathsf{V}^t\|_{\dot{W}^{\frac{\mathsf{d}}{p}+1,p}} + \big\||\nabla|^{\frac{\mathsf{d}-\mathsf{s}}{2}} \mathsf{V}^t\big\|_{L^{\frac{2\mathsf{d}}{\mathsf{d}-\mathsf{s}-2}}} \mathbf{1}_{\mathsf{s}<\mathsf{d}-2}\Big) \mathrm{d}t < \infty$. *Let* μ *be a solution to the mean-field equation* (3) *in* $L^\infty\big([0,T], \mathcal{P}(\mathbb{R}^\mathsf{d}) \cap \dot{W}^{\frac{\mathsf{d}}{p}+\mathsf{s}+2-\mathsf{d},p}(\mathbb{R}^\mathsf{d})\big)$ *for* $1 < p \leq \frac{2\mathsf{d}}{\mathsf{d}-\mathsf{s}-2}$. *Suppose further that* $\int_{\mathbb{R}^\mathsf{d}} |x|^{\frac{|\mathsf{s}|}{2}} \mathrm{d}\mu^t < \infty$ *for every* $t \in [0,T]$.[12]

For $q > \frac{\mathsf{d}}{\mathsf{d}-\mathsf{s}-1}$, $\vartheta \in \big(\frac{|\mathsf{s}|}{2}, |\mathsf{s}|\big)$, *and* $0 < \vartheta' < \vartheta$, *let*

$$\mathscr{E}^t := \mathsf{F}_N(X_N^t, \mu^t) + \sup_{t \leq \tau \leq T} \zeta^\tau, \tag{42}$$

where

$$\zeta^\tau := \epsilon \begin{cases} \Big(C_q \|\mu^\tau\|_{L^q} \lambda_\tau^{\frac{\mathsf{d}(q-1)}{q}-\mathsf{s}} + C_q \|\mu^\tau\|_{L^q}^{\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}} \mathbf{1}_{\mathsf{s}<\mathsf{d}-1} \\ \qquad + C\Big(\|\mu^\tau\|_{\dot{C}^{\mathsf{s}+2-\mathsf{d}}}^{\frac{\mathsf{s}+1-\mathsf{d}}{\mathsf{s}+2-\mathsf{d}}} + \epsilon^{\mathsf{d}-\mathsf{s}-1} \|\mu^\tau\|_{L^\infty}^{\frac{\mathsf{s}-\mathsf{d}+1}{\mathsf{s}+2-\mathsf{d}}} \Big) \mathbf{1}_{\mathsf{s}\geq\mathsf{d}-1} \Big), & \max(\mathsf{d}-2, 0) \leq \mathsf{s} < \mathsf{d}, \\ C_q \Big(\|\mu^\tau\|_{L^q} \kappa_\tau^{\frac{\mathsf{d}(q-1)}{q}-\mathsf{s}} + \|\mu^\tau\|_{L^q}^{\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}} \Big), & 0 < \mathsf{s} < \mathsf{d}-2, \\ C_{\vartheta,\vartheta'} \epsilon^{\vartheta'-1} \int_{\mathbb{R}^\mathsf{d}} |x|^{|\mathsf{s}|-\vartheta} \mathrm{d}\mu^\tau + C\Big(1 + C_q \|\mu^\tau\|_{L^q}^{\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}}\Big), & -1 < \mathsf{s} < 0, \\ C \int_{\mathbb{R}^\mathsf{d}} |x|^{|\mathsf{s}|-1} \mathrm{d}\mu^\tau, & -2 < \mathsf{s} \leq -1, \end{cases} \tag{43}$$

and for $\alpha, \delta > 0$,[13]

$$\epsilon := \begin{cases} \delta N^{-1-\frac{2(\mathsf{s}+1)}{\mathsf{s}}-\alpha} (\mathscr{E}^0)^{-\frac{1}{\mathsf{s}}}, & 0 < \mathsf{s} < \mathsf{d}, \\ N^{-\alpha}, & -2 < \mathsf{s} < 0. \end{cases} \tag{44}$$

There exists a $\delta > 0$ *depending only on* $\mathsf{d}, \mathsf{s}, p, \alpha, \int_0^T \|\mu^t\|_{\dot{W}^{\frac{\mathsf{d}}{p}+\mathsf{s}+2-\mathsf{d},p}} \mathrm{d}t$, *such that for any solution* X_N *of the microscopic system* (1), *the following holds for* $t \in [0,T]$:

- *if* $0 < \mathsf{s} < \mathsf{d}$,

$$\mathscr{E}^t \leq C \mathscr{E}^0 \exp\bigg(\int_0^t C_p \|u^{t'}\|_{\dot{W}^{\frac{\mathsf{d}}{p}+1,p}} \Big(1 + \big|\log(\delta N^{1-\frac{2(\mathsf{s}+1)}{\mathsf{s}}-\alpha} (\mathscr{E}^0)^{-\frac{1}{\mathsf{s}}})\big|\Big)^{1-\frac{1}{p}} \mathrm{d}t' \bigg); \tag{45}$$

- *if* $-1 < \mathsf{s} < 0$, *then*

$$(\mathscr{E}^t)^{\frac{\vartheta}{|\mathsf{s}|}} \leq \exp\bigg(\frac{C_p \vartheta}{|\mathsf{s}|} \int_0^t \|u^\tau\|_{\dot{W}^{\frac{\mathsf{d}}{p}+1,p}} (1 + |\log N^{-\alpha}|)^{1-\frac{1}{p}} \mathrm{d}\tau \bigg) \bigg((\mathscr{E}^0)^{\frac{\vartheta}{|\mathsf{s}|}} + \frac{\vartheta}{|\mathsf{s}|} \int_0^t C_{\vartheta,\vartheta'} \|u^{t'}\|_{\dot{C}^1} \epsilon^{\vartheta} \mathrm{d}t' \bigg); \tag{46}$$

- *if* $-2 < \mathsf{s} \leq -1$, *then*

$$(\mathscr{E}^t)^{\frac{1}{|\mathsf{s}|}} \leq \exp\bigg(\frac{C_p}{|\mathsf{s}|} \int_0^t \|u^\tau\|_{\dot{W}^{\frac{\mathsf{d}}{p}+1,p}} (1 + |\log N^{-\alpha}|)^{1-\frac{1}{p}} \mathrm{d}\tau \bigg) \bigg((\mathscr{E}^0)^{\frac{1}{|\mathsf{s}|}} + \frac{1}{|\mathsf{s}|} \int_0^t C \|u^{t'}\|_{\dot{C}^1} \epsilon \, \mathrm{d}t' \bigg). \tag{47}$$

Above, $C > 0$ *depends only on* d, s, *and* $C_{\vartheta,\vartheta'}, C_q, C_p$ *additionally depend on* $(\vartheta, \vartheta'), q, p$, *respectively.*

In particular, if $\mathsf{F}_N(X_N^0, \mu) = O(N^{-m})$ *as* $N \to \infty$, *for some* $m > 0$, *then* $\mu_N^t \to \mu^t$ *as* $N \to \infty$ *in* $H^{-r}(\mathbb{R}^\mathsf{d})$ *uniformly on* $[0,T]$ *for any* $r > \frac{\mathsf{d}}{2}$.

---

[12] This condition ensures that the modulated energy/MMD is finite (see [59, Example 4]) and that all our moment assumptions in the statement of the theorem are satisfied. Through a Grönwall argument, the reader may check that if μ^0 satisfies this moment condition, then μ^t does uniformly in $[0,T]$.

[13] When $0 < \mathsf{s} < \mathsf{d}$, the definition of ϵ is given implicitly. See the proof of Theorem 10 for why the implicit equation has a solution.

Remark 11. A natural way to produce initial point configurations X_N such that $\left|\mathsf{F}_N(X_N,\mu)\right| \le CN^{-m}$ for $C,m>0$ independent of N is by taking X_N to be distributed according to a *modulated Gibbs measure* (a terminology introduced in [76])

$$\mathrm{d}\mathbb{Q}_{N,\beta}(\mu)(X_N) := \frac{e^{-\beta N\mathsf{F}_N(X_N,\mu)}\,\mathrm{d}\mu^{\otimes N}(X_N)}{\mathsf{K}_{N,\beta}(\mu)}, \qquad \mathsf{K}_{N,\beta}(\mu) := \int_{(\mathbb{R}^{\mathsf{d}})^N} e^{-\beta N\mathsf{F}_N(X_N,\mu)}\,\mathrm{d}\mu^{\otimes N}(X_N), \quad (48)$$

where $\beta\in[0,\infty]$ is a parameter, which may be interpreted as inverse temperature. The extreme $\beta=0$ corresponds to μ-i.i.d. points, while the extreme $\beta=\infty$ formally corresponds to points that are uniformly distributed among the set of minimizers of $X_N \mapsto \mathsf{F}_N(X_N,\mu)$. When $\beta=0$, one has $\mathbb{E}\left[\mathsf{F}_N(X_N,\mu)\right] = CN^{-1}$ and in fact, this scaling holds with high probability. When $\beta=\infty$, one has $\min\mathsf{F}_N(X_N,\mu) = (-1)^{\mathrm{sgn}(\mathsf{s})}CN^{\frac{\mathsf{s}}{\mathsf{d}}-1}$. One can produce scaling laws in N interpolating between these two extremes by varying β. We refer to [87, Chapter 5] for further details.

Remark 12. There is a large literature on the well-posedness of equations of the form (3). We refer to the introduction of [85] for some comments in this direction. Here, we briefly mention that when $\mathsf{s}<\mathsf{d}-2$, solutions of the mean-field equation (3) are unique and global for probability density initial data in the critical space $L^{\frac{\mathsf{d}}{\mathsf{d}-\mathsf{s}-2}}$. This follows from essentially the same argument as for the Coulomb case $\mathsf{s}=\mathsf{d}-2$ with initial density in L^∞ (e.g. see [57,88,92]).

Section 4 is devoted to the proofs of Theorems 8 and 10. The defective commutator estimate, which is the workhorse, proceeds along lines similar to [69,70]. Namely, we mollify the vector field v at some scale ϵ and apply Theorem 1 with v_ϵ. We then need to separately estimate the error term $\mathsf{A}_1[X_N,\mu,v-v_\epsilon]$ in terms of $\mathsf{F}_N(X_N,\mu),\epsilon,N$. Of course, as Theorem 1 requires at least Lipschitz regularity, which is not satisfied when $v\in\dot{W}^{\frac{\mathsf{d}}{p}+1,p}$, we have to pay a price in terms of divergence as $\epsilon\to 0$ (i.e. the defect). Simply repeating the arguments from [69,70] does not yield the estimates of Theorem 8, and consequently also not yield Theorem 10. [69] is specific to the log case, which is well-suited to the log-Lipschitz regularity of vector fields in $\dot{W}^{\frac{\mathsf{d}}{p},p}$ (more generally, in the Zygmund space $\dot{\mathcal{C}}^1$). While the cruder argument in [70], devised to handle the non-logarithmic nature of the Coulomb potential when $\mathsf{d}\neq 2$, leads to factors $\propto \log N$ in front of $\mathsf{F}_N(X_N^t,\mu^t)$. Such factors are only acceptable if one restricts to short times.

As advertised in Section 1.3, the key new ingredient in our proof is the Brezis–Wainger–Hansson inequality (22). More precisely, a corollary of (22), obtained through the classical Morrey argument, is that if $1<p<\infty$, then for any $f\in\mathcal{S}(\mathbb{R}^{\mathsf{d}})$,

$$\forall\, x,y\in\mathbb{R}^{\mathsf{d}}, \quad \left|f(x)-f(y)\right| \le C\|f\|_{\dot{W}^{\frac{\mathsf{d}}{p}+1,p}}|x-y|\left(1+\left|\log|x-y|\right|\right)^{1-\frac{1}{p}}, \quad (49)$$

where $C=C(\mathsf{d},p)>0$.[14] This inequality may be rephrased as saying $L^\infty\cap\dot{W}^{\frac{\mathsf{d}}{p}+1,p}$ embeds into the log-Lipschitz space $\mathrm{Lip}_{\infty,\infty}^{(1,-1+\frac{1}{p})}$ [40,50]. In applying Theorem 1 with v_ϵ as part of the proof, we need to quantify the divergence of $\|\nabla v_\epsilon\|_{L^\infty}$ as $\epsilon\to 0^+$. In [69,70], the fact that $\dot{\mathcal{C}}^1$ functions are log-Lipschitz was used to bound $\|\nabla v_\epsilon\|_{L^\infty} \lesssim \|v\|_{\dot{\mathcal{C}}^1}\left(1+|\log\epsilon|\right)$. Note that $\|v\|_{\dot{\mathcal{C}}^1}\lesssim\|v\|_{\dot{W}^{\frac{\mathsf{d}}{p}+1,p}}$ by Sobolev embedding. The modulus of continuity bound (49) allows us to improve this bound to $\|\nabla v_\epsilon\|_{L^\infty}\lesssim\|v\|_{\dot{W}^{\frac{\mathsf{d}}{p}+1,p}}\left(1+|\log\epsilon|\right)^{1-\frac{1}{p}}$ (see Lemma 15). Combining with better estimates for the mollification error $v-v_\epsilon$ (see Lemma 19), this improvement then yields Theorem 8. We refer to Section 4.3 for the complete proof.

To obtain the result of mean-field convergence Theorem 10 from Theorem 8, we need to choose $\epsilon=\epsilon_t$ appropriately in a possibly time-dependent fashion. ϵ_t needs to be sufficiently small depending on $N,\mathsf{F}_N(X_N^t,\mu^t)$ to compensate for the divergent prefactors of N and the powers of $\mathsf{F}_N(X_N^t,\mu^t)$ with exponent >1 in the estimates of Theorem 8. Thanks to the improvement

[14] Strictly speaking, [13] shows the inequality with the inhomogeneous Sobolev norm $\|f\|_{W^{\frac{\mathsf{d}}{p}+1,p}}$ on the right-hand side. Replacing f with $f_\lambda := f(\lambda\cdot)$ and (x,y) with $\lambda^{-1}(x,y)$, then letting $\lambda\to\infty$ yields (49).

$\big(1+|\log\epsilon|\big)^{1-\frac{1}{p}}$, this is possible to do and is still suitable for closing the differential inequality for $\mathsf{F}_N(X_N^t,\mu^t)$. We refer to Section 4.4 for the details. At the risk of being repetitive, we emphasize that if $1-\frac{1}{p}$ were replaced by 1, then we would be stuck with the same short-time restriction as in [70].

Remark 13. The reader should note that Theorems 8 and 10 do not cover the case $p=\infty$ (they also do not cover $p=1$, but this is less relevant). We return to this exception in Section 5.4.

3. Counterexamples

This section is devoted to the proofs of our counterexample results, Theorems 3 and 5. We have already explained in Section 2.2 how Corollary 6 follows from Theorem 5, so we do not discuss this result further.

3.1. *Proof of Theorem 3(i)*

Let $\chi\in C_c^\infty(\mathbb{R}^{\mathsf{d}})$ such that $0\le\chi\le1$, $\operatorname{supp}\chi\subset B(0,1/2)$, and $\chi\equiv1$ on $B(0,1/4)$. Set $v:=(v_1,\dots,v_{\mathsf{d}})$, where

$$v_1(x):=-x_1\log\big(\log\big(1/|x|\big)\big)\chi(x)\qquad\text{and}\qquad v_i(x):=0,\quad 2\le i\le\mathsf{d}. \tag{50}$$

Evidently, v is compactly supported. It can also be made C^∞ by convolution with χ_ϵ, for $\epsilon=\epsilon(r)>0$, where the parameter r is as below. We omit the implementation of this step.

We claim that $v'\in\mathrm{BMO}(\mathbb{R})$ when $\mathsf{d}=1$ and $\nabla^{\otimes\mathsf{d}}v\in L^{\frac{\mathsf{d}}{\mathsf{d}-1}}(\mathbb{R}^{\mathsf{d}})$ when $\mathsf{d}\ge2$. Indeed, when $\mathsf{d}=1$

$$\frac{\mathrm{d}}{\mathrm{d}x}\Big(-x\log\big(\log\big(1/|x|\big)\big)\Big)=-\log\big(\log\big(1/|x|\big)\big)+\frac{1}{\log\big(1/|x|\big)}\in\mathrm{BMO}(\mathbb{R}), \tag{51}$$

and thus so is v since multiplication by χ preserves $\mathrm{BMO}(\mathbb{R})$. On the other hand, when $\mathsf{d}\ge2$ for all $k\ge1$ there exists $C_{\mathsf{d},k}>0$ such that

$$\forall\ |x|\le\frac{1}{2},\quad\Big|\nabla^{\otimes k}\log\big(\log\big(1/|x|\big)\big)\Big|\le\frac{C_{\mathsf{d},k}}{|x|^k\log\big(1/|x|\big)}. \tag{52}$$

With the Leibniz rule and the fact that $\operatorname{supp}\chi'\subset B\big(0,\frac14\big)^c$, the preceding estimate implies that

$$\forall\ |x|\le\frac{1}{2},\quad\big|\nabla^{\otimes\mathsf{d}}v(x)\big|\le\frac{C'_{\mathsf{d},k}}{|x|^{\mathsf{d}-1}\log\big(1/|x|\big)}. \tag{53}$$

Since χ is supported on $B\big(0,\frac12\big)$, it follows that

$$\int_{\mathbb{R}^{\mathsf{d}}}|\nabla^{\otimes\mathsf{d}}v|^{\frac{\mathsf{d}}{\mathsf{d}-1}}\le C\int_{B(0,1/2)}\frac{1}{|x|^{\mathsf{d}-1}\log\big(1/|x|\big)}=C\int_0^{1/2}\frac{1}{r\log(1/r)^{\frac{\mathsf{d}}{\mathsf{d}-1}}}<\infty, \tag{54}$$

where in the first equality, we have used a radial coordinate change. By Sobolev embedding, this also implies that $\nabla v\in\mathrm{BMO}(\mathbb{R}^{\mathsf{d}})$ and $|\nabla|^{\frac{\mathsf{d}-\mathsf{s}}{2}}v\in L^{\frac{2\mathsf{d}}{\mathsf{d}-\mathsf{s}-2}}(\mathbb{R}^{\mathsf{d}})$ for all $-2<\mathsf{s}<\mathsf{d}-2$.

Next, suppose that $f\in C_c^\infty(\mathbb{R}^{\mathsf{d}})$ is a zero-mean function supported in $B(0,1/4)$ such that

$$\int_{(\mathbb{R}^{\mathsf{d}})^2}\frac{\big|(x-y)^1\big|^2}{|x-y|^{\mathsf{s}+2}}f(x)f(y)>0. \tag{55}$$

Such an f exists since if f is symmetric under permutation of coordinates, then

$$\begin{aligned}\int_{(\mathbb{R}^{\mathsf{d}})^2}\frac{\big|(x-y)^1\big|^2}{|x-y|^{\mathsf{s}+2}}f(x)f(y)&=\mathsf{c}_{\mathsf{d},\mathsf{s}}\int_{\mathbb{R}^{\mathsf{d}}}|\xi|^{\mathsf{s}-\mathsf{d}-2}\big(|\xi|^2-(\mathsf{d}-\mathsf{s})|\xi^1|^2\big)\big|\hat{f}(\xi)\big|^2\\&=\mathsf{c}_{\mathsf{d},\mathsf{s}}\,\mathsf{s}\int_{\mathbb{R}^{\mathsf{d}}}|\xi|^{\mathsf{s}-\mathsf{d}-2}|\xi^1|^2\big|\hat{f}(\xi)\big|^2.\end{aligned} \tag{56}$$

Let $f_r(x) := r^{-\mathsf{d}} f(r^{-1}x)$ for $r > 0$, so that

$$\begin{aligned}&\int_{(\mathbb{R}^{\mathsf{d}})^2} \big(v(x)-v(y)\big)\cdot\nabla\mathsf{g}(x-y) f_r(x) f_r(y)\\ &= \int_{(\mathbb{R}^{\mathsf{d}})^2} \log\big(\log(1/|x|)\big)\frac{\big|(x-y)^1\big|^2}{|x-y|^{\mathsf{s}+2}} f_r(x) f_r(y) + \int_{(\mathbb{R}^{\mathsf{d}})^2} \log\left(\frac{\log(1/|x|)}{\log(1/|y|)}\right)\frac{y^1(x-y)^1}{|x-y|^{\mathsf{s}+2}} f_r(x) f_r(y). \qquad (57)\end{aligned}$$

Changing coordinates, the first term on the right-hand side of (57) equals

$$r^{-\mathsf{s}}\log\big(\log(1/r)\big)\int_{(\mathbb{R}^{\mathsf{d}})^2} \frac{\log\big(\log(1/|x|)-\log(r)\big)}{\log\big(\log(1/r)\big)}\frac{\big|(x-y)^1\big|^2}{|x-y|^{\mathsf{s}+2}} f(x) f(y), \qquad (58)$$

while the second term equals

$$r^{-\mathsf{s}}\int_{(\mathbb{R}^{\mathsf{d}})^2} \log\left(\frac{\log(1/|x|)-\log(r)}{\log(1/|y|)-\log(r)}\right)\frac{y^1(x-y)^1}{|x-y|^{\mathsf{s}+2}} f(x) f(y). \qquad (59)$$

By the dominated convergence theorem,

$$\lim_{r\to 0}\int_{(\mathbb{R}^{\mathsf{d}})^2} \frac{\log\big(\log(1/|x|)-\log(r)\big)}{\log\big(\log(1/r)\big)}\frac{\big|(x-y)^1\big|^2}{|x-y|^{\mathsf{s}+2}} f(x) f(y) = \int_{(\mathbb{R}^{\mathsf{d}})^2}\frac{\big|(x-y)^1\big|^2}{|x-y|^{\mathsf{s}+2}} f(x) f(y) \qquad (60)$$

and

$$\lim_{r\to 0}\int_{(\mathbb{R}^{\mathsf{d}})^2} \log\left(\frac{\log(1/|x|)-\log(r)}{\log(1/|y|)-\log(r)}\right)\frac{y^1(x-y)^1}{|x-y|^{\mathsf{s}+2}} f(x) f(y) = 0. \qquad (61)$$

Combining (57)–(61), we see that there exits $C > 0$ depending only on d,s, f such that

$$\int_{(\mathbb{R}^{\mathsf{d}})^2} \big(v(x)-v(y)\big)\cdot\nabla\mathsf{g}(x-y) f_r(x) f_r(y) \ge C^{-1} r^{-\mathsf{s}}\log\big(\log(1/r)\big) - C r^{-\mathsf{s}} \qquad (62)$$

for all sufficiently small r.

Since $\|f_r\|^2_{\dot H^{\frac{\mathsf{s}-\mathsf{d}}{2}}} = r^{-\mathsf{s}}\|f\|^2_{\dot H^{\frac{\mathsf{s}-\mathsf{d}}{2}}}$, in total we thus have found that for all sufficiently small r

$$\frac{\left|\int_{(\mathbb{R}^{\mathsf{d}})^2} \big(v(x)-v(y)\big)\cdot\nabla\mathsf{g}(x-y) f_r(x) f_r(y)\right|}{\left(\|\nabla v\|_{\mathrm{BMO}} + \mathbf{1}_{\mathsf{s}<\mathsf{d}-2}\left\||\nabla|^{\frac{\mathsf{d}-\mathsf{s}}{2}} v\right\|_{L^{\frac{2\mathsf{d}}{\mathsf{d}-\mathsf{s}-2}}}\right)\|f_r\|^2_{\dot H^{\frac{\mathsf{s}-\mathsf{d}}{2}}}} \ge C^{-1}\log\big(\log(1/r)\big) - C. \qquad (63)$$

Taking $r \to 0$ concludes the claim.

Remark 14. We note that if $\mathsf{d} = 1$ and $\mathsf{s} = 0$, the above counterexample fails since

$$\int_{(\mathbb{R}^{\mathsf{d}})^2} \frac{\big|(x-y)^1\big|^2}{|x-y|^2} f(x) f(y) = \left(\int_{\mathbb{R}^{\mathsf{d}}} f\right)^2 = 0 \qquad (64)$$

for all zero-mean functions f.

3.2. *Proof of Theorem 3(ii)*

We use an equivalent formulation of (18) for Schwartz functions with Fourier transforms supported away from 0. If F and G are Schwartz with Fourier transforms supported away from the origin, then $f := |\nabla|F$ and $g := |\nabla|G$ are both Schwartz with Fourier transforms supported away from the origin. Additionally, since $|\nabla|\mathsf{g} = \mathsf{c}_{\mathsf{d},\mathsf{s}}\delta_0$, it holds that

$$\int_{\mathbb{R}^2} \big(v(x)-v(y)\big)\cdot\nabla\mathsf{g}(x-y) f(x) g(y) = \mathsf{c}\int_{\mathbb{R}} v\cdot\big(F'|\nabla|G + G'|\nabla|F\big). \qquad (65)$$

Since $\|F\|_{\dot H^{\frac12}} = \mathsf{c}_{\mathsf{d},\mathsf{s}}\|f\|_{\dot H^{-\frac12}}$ and similarly for G and g, it suffices to show that there exists $C > 0$ such that

$$\left|\int_{\mathbb{R}} v\big(F'|\nabla|G + G'|\nabla|F\big)\right| \le C\|v'\|_{\mathrm{BMO}}\|F\|_{\dot H^{1/2}}\|G\|_{\dot H^{1/2}} \qquad (66)$$

for all $F, G \in \mathcal{S}(\mathbb{R})$ with Fourier support away from the origin.

To this end, we integrate by parts

$$\begin{aligned}\int_{\mathbb{R}} v\big(F'|\nabla|G+G'|\nabla|F\big) &= -\int_{\mathbb{R}}\Big(v'\big(F|\nabla|G+G|\nabla|F\big)+vF\partial_x|\nabla|G+vG\partial_x|\nabla|F\Big)\\ &= -\int_{\mathbb{R}}\Big(v'\big(F|\nabla|G+G|\nabla|F\big)-\partial_x|\nabla|^{\frac12}(vG)|\nabla|^{\frac12}F-\partial_x|\nabla|^{\frac12}(vF)|\nabla|^{\frac12}G\Big)\\ &= -\int_{\mathbb{R}}\Big(v'F|\nabla|G-|\nabla|^{\frac12}v'F|\nabla|^{\frac12}G\Big)-\int_{\mathbb{R}}\Big(v'G|\nabla|F-|\nabla|^{\frac12}v'G|\nabla|^{\frac12}F\Big)\\ &\quad+\int_{(\mathbb{R})}\Big(v\big(|\nabla|^{\frac12}F'|\nabla|^{\frac12}G+|\nabla|^{\frac12}G'|\nabla|^{\frac12}F\big)+3v'|\nabla|^{\frac12}F|\nabla|^{\frac12}G\Big)\\ &\quad+\int_{\mathbb{R}}\Big(T(v,F)|\nabla|^{\frac12}G+T(v,G)|\nabla|^{\frac12}F\Big),\end{aligned} \tag{67}$$

where T is the bilinear operator

$$T(v,H) := \partial_x|\nabla|^{\frac12}(vH)-|\nabla|^{\frac12}v'H-v|\nabla|^{\frac12}H'-\frac{3}{2}v'|\nabla|^{\frac12}H \tag{68}$$

defined for $H\in\mathcal{S}(\mathbb{R})$.

First, we note that after integrating by parts

$$\int_{\mathbb{R}} v\Big(|\nabla|^{\frac12}F'|\nabla|^{\frac12}G+|\nabla|^{\frac12}G'|\nabla|^{\frac12}F\Big) = -\int_{\mathbb{R}} v'|\nabla|^{\frac12}F|\nabla|^{\frac12}G. \tag{69}$$

We can thus split (67) into a sum of three terms defined by

$$\mathrm{Term}_1 := -\int_{\mathbb{R}}\Big(v'F|\nabla|G-|\nabla|^{\frac12}v'F|\nabla|^{\frac12}G-v'|\nabla|^{\frac12}F|\nabla|^{\frac12}G\Big), \tag{70}$$

$$\mathrm{Term}_2 := -\int_{\mathbb{R}}\Big(v'G|\nabla|F-|\nabla|^{\frac12}v'G|\nabla|^{\frac12}F-v'|\nabla|^{\frac12}F|\nabla|^{\frac12}G\Big), \tag{71}$$

$$\mathrm{Term}_3 := \int_{\mathbb{R}}\Big(T(v,F)|\nabla|^{\frac12}G+T(v,G)|\nabla|^{\frac12}F\Big). \tag{72}$$

We bound each of these terms separately.

We begin with Term_1. Integrating by parts, we have that

$$\mathrm{Term}_1 = -\int_{\mathbb{R}}\Big(|\nabla|^{\frac12}(v'F)-|\nabla|^{\frac12}v'F-v'|\nabla|^{\frac12}F\Big)|\nabla|^{\frac12}G. \tag{73}$$

The expression inside the brackets is a Kenig–Ponce–Vega-type commutator [52], and by [56, Remark 1.3, (1.7)] with $s=\frac12$, $f=F$, $g=v'$, we have that

$$\Big\||\nabla|^{\frac12}(v'F)-|\nabla|^{\frac12}v'F-v'|\nabla|^{\frac12}F\Big\|_{L^2}\le C\|v'\|_{\mathrm{BMO}}\|F\|_{\dot H^{1/2}} \tag{74}$$

for some $C>0$. So by Cauchy–Schwarz,

$$|\mathrm{Term}_1|\le C\|v'\|_{\mathrm{BMO}}\|F\|_{\dot H^{1/2}}\|G\|_{\dot H^{1/2}}. \tag{75}$$

Swapping F and G, we can bound Term_2 identically.

For Term_3, we note that T is a higher-order commutator and by [56, Corollary 1.4(2)] applied with $s=\frac32$, $f=H$, $g=v$, $A^s=\partial_x|\nabla|^{\frac12}$, $p=2$, $s_2=1$, $s_1=\frac12$, we obtain

$$\big\|T(v,H)\big\|_{L^2}\le C\big\||\nabla|v\big\|_{\mathrm{BMO}}\|H\|_{\dot H^{1/2}}, \tag{76}$$

for some $C>0$. Note that since the Hilbert transform is bounded on BMO, one may bound $\big\||\nabla|v\big\|_{\mathrm{BMO}}\le C\|v'\|_{\mathrm{BMO}}$. By Cauchy–Schwarz, it then follows that

$$|\mathrm{Term}_3|\le C\|v'\|_{\mathrm{BMO}}\|F\|_{\dot H^{1/2}}\|G\|_{\dot H^{1/2}}. \tag{77}$$

Combining (75) for $\mathrm{Term}_1, \mathrm{Term}_2$ and (77) for Term_3, the proof is complete.

3.3. *Proof of Theorem 5*

Let ϕ be a radially symmetric Schwartz function on $\mathbb{R}^{\mathsf{d}}$ such that the Fourier transform $\widehat{\phi}$ satisifies $0 \le \widehat{\phi} \le 1$, $\operatorname{supp}\widehat{\phi} \subset B(0,1/4)$, and $\widehat{\phi} \equiv 1$ on $B(0,1/8)$. Letting $e_1 := (1,0,\ldots,0)$ and $k > 2$, define the (real) vector field v by

$$\widehat{v}(\xi) := ie_1\big(\widehat{\phi}(\xi + ke_1) - \widehat{\phi}(\xi - ke_1)\big) \tag{78}$$

and (real) functions f, g by

$$\widehat{f}(\xi) := \widehat{\phi}\big(\xi + (k+1)e_1\big) + \widehat{\phi}\big(\xi - (k+1)e_1\big), \tag{79}$$

$$\widehat{g}(\xi) := \widehat{\phi}(\xi + e_1) + \widehat{\phi}(\xi - e_1). \tag{80}$$

Clearly, f and g are both Schwartz and have Fourier transforms supported away from 0. We then note that

$$\int_{(\mathbb{R}^{\mathsf{d}})^2} \big(v(x) - v(y)\big) \cdot \nabla \mathsf{g}(x-y) f(x) g(y) = \int_{\mathbb{R}^{\mathsf{d}}} v \cdot \nabla \mathsf{g} * fg + \int_{\mathbb{R}^{\mathsf{d}}} v \cdot \nabla \mathsf{g} * gf. \tag{81}$$

By Plancherel's theorem,

$$\int_{\mathbb{R}^{\mathsf{d}}} v \cdot \nabla \mathsf{g} * fg = \int_{(\mathbb{R}^{\mathsf{d}})^2} \widehat{v}(\xi_1) \cdot 2\pi i(\xi_1 - \xi_2)\widehat{\mathsf{g}}(\xi_1 - \xi_2)\widehat{f}(\xi_1 - \xi_2)\widehat{g}(\xi_2). \tag{82}$$

Unpacking the definitions of v, f and g, the preceding right-hand side is equal to

$$\begin{aligned}
&-\sum_{\epsilon_i \in \{\pm 1\}} \epsilon_1 \int_{(\mathbb{R}^{\mathsf{d}})^2} \widehat{\phi}(\xi_1 + \epsilon_1 k e_1) e_1 \cdot (\xi_1 - \xi_2)\widehat{\mathsf{g}}(\xi_1 - \xi_2)\widehat{\phi}\big(\xi_1 - \xi_2 + \epsilon_2(k+1)e_1\big)\widehat{\phi}(\xi_2 + \epsilon_3 e_1) \\
&\qquad = -\sum_{\epsilon_i \in \{\pm 1\}} \epsilon_1 \int_{(\mathbb{R}^{\mathsf{d}})^2} \widehat{\phi}(\zeta_1)\widehat{\phi}(\zeta_2)\widehat{\phi}\big(\zeta_1 - \zeta_2 + \big((\epsilon_3 + \epsilon_2) + (\epsilon_2 - \epsilon_1)k\big)e_1\big) \\
&\qquad\qquad \times e_1 \cdot \big(\zeta_1 - \zeta_2 + (\epsilon_3 - \epsilon_1 k)e_1\big)\widehat{\mathsf{g}}\big(\zeta_1 - \zeta_2 + (\epsilon_3 - \epsilon_1 k)e_1\big),
\end{aligned} \tag{83}$$

where the final equality follows from a change of variables. Since $\operatorname{supp}\widehat{\phi} \subset B(0,1/4)$, the integrals in the sum above are only nonzero if

$$\big|\zeta_1 - \zeta_2 + \big(\epsilon_3 + \epsilon_2 + (\epsilon_2 - \epsilon_1)k\big)e_1\big| \le \frac{1}{4} \tag{84}$$

for some ζ_i in $B(0,1/4)$. Since $k > 2$, we must have $(\epsilon_1, \epsilon_2, \epsilon_3) = \pm(1,1,-1)$. Using symmetry, we have shown that

$$\int_{\mathbb{R}^{\mathsf{d}}} v \cdot \nabla \mathsf{g} * fg = 2\int_{(\mathbb{R}^{\mathsf{d}})^2} \widehat{\phi}(\zeta_1)\widehat{\phi}(\zeta_2)\widehat{\phi}(\zeta_1 - \zeta_2)\big((\zeta_2 - \zeta_1) \cdot e_1 + 1 + k\big)\widehat{\mathsf{g}}\big(\zeta_2 - \zeta_1 + (1+k)e_1\big). \tag{85}$$

Since $|\zeta_i| \le \frac{1}{4}$ on the support of $\widehat{\phi}(\zeta_i)$ and $\widehat{\mathsf{g}}$ is decreasing by Assumption 4, we may crudely bound the right-hand side above using Young's convolution inequality to obtain

$$\left|\int_{\mathbb{R}^{\mathsf{d}}} v \cdot \nabla \mathsf{g} * fg\right| \le Ck\widehat{\mathsf{g}}(k)\|\widehat{\phi}\|_{L^{3/2}}^3, \tag{86}$$

where $C > 0$ is a constant only depending on d.

By the same reasoning as above,

$$\begin{aligned}
\int_{\mathbb{R}^{\mathsf{d}}} v \cdot \nabla \mathsf{g} * gf &= -\sum_{\epsilon_i \in \{\pm 1\}} \epsilon_1 \int_{(\mathbb{R}^{\mathsf{d}})^2} \widehat{\phi}(\zeta_1)\widehat{\phi}(\zeta_2)\widehat{\phi}\big(\zeta_1 - \zeta_2 + \big((\epsilon_2 + \epsilon_3) + (\epsilon_3 - \epsilon_1)k\big)e_1\big) \\
&\qquad \times e_1 \cdot \big(\zeta_1 - \zeta_2 + \big(\epsilon_3 + (\epsilon_3 - \epsilon_1)k\big)e_1\big)\widehat{\mathsf{g}}\big(\zeta_1 - \zeta_2 + \big(\epsilon_3 + (\epsilon_3 - \epsilon_1)k\big)e_1\big).
\end{aligned} \tag{87}$$

For an integral in the sum on the preceding right-hand side to be nonzero, it must now be the case that $(\epsilon_1, \epsilon_2, \epsilon_3) = \pm(1,-1,1)$. Using symmetry, we have thus shown that

$$\int_{\mathbb{R}^{\mathsf{d}}} v \cdot \nabla \mathsf{g} * gf = 2\int_{(\mathbb{R}^{\mathsf{d}})^2} \widehat{\phi}(\zeta_1)\widehat{\phi}(\zeta_2)\widehat{\phi}(\zeta_1 - \zeta_2)\big(1 - e_1 \cdot (\zeta_1 - \zeta_2)\big)\widehat{\mathsf{g}}(\zeta_2 - \zeta_1 + e_1). \tag{88}$$

Since $\big(1-e_1\cdot(\zeta_1-\zeta_2)\big)\geq\frac{1}{2}$ and $|\zeta_2-\zeta_1+e_1|\leq 2$ on the support of $\widehat{\phi}(\zeta_1)\widehat{\phi}(\zeta_2)$, we have found the lower bound

$$\int_{\mathbb{R}^{\mathsf{d}}} v\cdot\nabla\mathsf{g}*gf\geq\widehat{\mathsf{g}}(2)\int_{(\mathbb{R}^{\mathsf{d}})^2}\widehat{\phi}(\zeta_1)\widehat{\phi}(\zeta_2)\widehat{\phi}(\zeta_1-\zeta_2)\geq C^{-1}\widehat{\mathsf{g}}(2), \tag{89}$$

where the last inequality uses that $\widehat{\phi}=1$ on $B(0,1/8)$ and $C>0$ only depends on ϕ.

On the other hand, it is immediate from Plancherel's theorem and Fourier support considerations that

$$\|g\|_{\mathsf{g}_t}\lesssim\|\phi\|_{L^2}, \tag{90}$$

$$\|f\|_{\mathsf{g}_t}\lesssim\widehat{\mathsf{g}}(tk)^{1/2}\|\phi\|_{L^2}, \tag{91}$$

for implicit constants depending on d,g,t, while Fourier inversion shows that

$$\|\nabla v\|_{L^\infty}\lesssim k\|\phi\|_{L^1}. \tag{92}$$

On the other hand, since $\operatorname{supp}\widehat{v}\subset\big\{k-1\leq|\xi|\leq k+1\big\}$, we have $\sigma(|\xi|)\leq\sup_{r\in[k-1,k+1]}\sigma(r)$ on $\operatorname{supp}\widehat{v}$. If $2\leq p\leq\infty$, then by the Hausdorff–Young inequality and the Fourier support of v,

$$\big\|\sigma(|\nabla|)v\big\|_{L^p}\lesssim\Big(\sup_{r\in[k-1,k+1]}\sigma(r)\Big)\|\widehat{\phi}\|_{L^{\frac{p-1}{p}}}, \tag{93}$$

for any $p\in[1,\infty]$, with constants depending only on d,p. If $1\leq p<2$, then under the assumption that $\partial_x^n\sigma$ is essentially bounded away from the origin for $n\leq\mathsf{d}+1$, it follows from Young's inequality that

$$\big\|\sigma(|\nabla|)v\big\|_{L^p}\lesssim\max_{n\leq\mathsf{d}+1}\sup_{r\in[k-1,k+1]}\big|\partial_r^n\sigma(r)\big|\|\phi\|_{L^p}. \tag{94}$$

Combining (86) and (89)–(94), in total we have found that

$$\begin{aligned}&\frac{\left|\int_{(\mathbb{R}^{\mathsf{d}})^2}\big(v(x)-v(y)\big)\cdot\nabla\mathsf{g}(x-y)f(x)g(y)\right|}{\big(\|\nabla v\|_{L^\infty}+\big\|\sigma(|\nabla|)v\big\|_{L^p}\big)\|f\|_{\mathsf{g}_t}\|g\|_{\mathsf{g}_t}}\\ &\qquad\geq\frac{C^{-1}-Ck\widehat{\mathsf{g}}(k)}{C\Big(k+\mathbf{1}_{1\leq p<2}\max_{n\leq\mathsf{d}+1}\|\partial_r^n\sigma\|_{L^\infty([k-1,k+1])}+\mathbf{1}_{2\leq p\leq\infty}\|\sigma\|_{L^\infty([k-1,k+1])}\Big)\widehat{\mathsf{g}}(tk)^{1/2}},\end{aligned} \tag{95}$$

for some $C=C(\mathsf{d},\mathsf{s},\phi)>0$. By (34), we have that $k\widehat{\mathsf{g}}(k)$ and $k\widehat{\mathsf{g}}(tk)^{1/2}$ vanish as $k\to\infty$, while by hypothesis,

$$\lim_{k\to\infty}\Big(\mathbf{1}_{1\leq p<2}\max_{n\leq\mathsf{d}+1}\|\partial_r^n\sigma\|_{L^\infty([k-1,k+1])}+\mathbf{1}_{2\leq p\leq\infty}\|\sigma\|_{L^\infty([k-1,k+1])}\Big)\widehat{\mathsf{g}}(tk)^{1/2}=0. \tag{96}$$

Taking $k\to\infty$ completes the proof.

4. Defective commutator estimates and mean-field convergence

This section is devoted to the results Theorems 8 and 10 advertised in Section 2.3 about defective commutator estimates and scaling-critical modulated energy estimates, respectively. The remainder of this section is organized as follows. In Section 4.1, we show some preliminary rates of convergence for mollifiers. In Section 4.2, we recall some properties of the modulated energy that are useful in the sequel. In Section 4.3, we prove our defective commutator estimate Theorem 8, which is the primary technical result. Finally, in Section 4.4, we prove Theorem 10.

Some comments on notation. Given $\mu\in L^p$, we use throughout this section the notation λ,κ from (10), (11). When $\mu=\mu^t$, we write λ_t,κ_t to indicate the time dependence.

4.1. *Mollification*

Let $\chi \in C_c^\infty(\mathbb{R}^\mathsf{d})$ be an even bump function which is 1 on $B\big(0,\frac14\big)$, zero outside $B(0,1)$, and satisfies $1 \ge \chi \ge 0$, and $\int_{\mathbb{R}^\mathsf{d}} \chi \,\mathrm{d}x = 1$. Let $v\colon \mathbb{R}^\mathsf{d} \to \mathbb{R}^\mathsf{d}$ be a vector field, and set

$$v_\epsilon(x) := (v * \chi_\epsilon)(x), \qquad \chi_\epsilon(x) := \epsilon^{-\mathsf{d}}\chi(x/\epsilon), \tag{97}$$

where the convolution is performed componentwise. It thus holds that v_ε is C^∞ and

$$\forall\, k \ge 0, \quad \|\nabla^{\otimes k} v_\epsilon\|_{L^\infty} \lesssim_k \epsilon^{-k}\|v\|_{L^\infty}. \tag{98}$$

Additionally, if $\|\cdot\|_X$ is any translation-invariant semi-norm, then we also have $\|v_\epsilon\|_X \le \|v\|_X$.

The purpose of this subsection is to establish estimates for the mollification error $v - v_\epsilon$ as well as to estimate how the Lipschitz seminorm of v_ϵ diverges as $\epsilon \to 0$ under certain regularity assumptions on v. The main result is the following lemma.

Lemma 15. *For any integer $k \ge 0$ and reals $0 < b \le a \le 1$, there exists a constant $C = C(a) > 0$ such that*

$$\|\nabla^{\otimes k} v - \nabla^{\otimes k} v_\epsilon\|_{L^\infty} \le C\|\nabla^{\otimes k} v\|_{\dot{C}^a}\epsilon^a, \tag{99}$$

$$\|\nabla^{\otimes k} v - \nabla^{\otimes k} v_\epsilon\|_{\dot{C}^b} \le C\|\nabla^{\otimes k} v\|_{\dot{C}^a}\epsilon^{a-b}. \tag{100}$$

Suppose that there exists $\delta_k > 0$ and an increasing function $\omega_k\colon [0,\delta_k] \to [0,\infty)$ such that $\big|\nabla^{\otimes k} v(x) - \nabla^{\otimes k} v(y)\big| \le \omega_k(|x-y|)$. Then there is a constant $C > 0$ such that

$$\|\nabla^{\otimes k} v - \nabla^{\otimes k} v_\epsilon\|_{L^\infty} \le C\omega_k(2\epsilon), \qquad 0 \le \epsilon \le \frac{\delta_k}{2}, \tag{101}$$

$$\|\nabla^{\otimes k} v_\epsilon\|_{L^\infty} \le C\epsilon^{-1}\omega_{k-1}(2\epsilon), \qquad 0 < \epsilon \le \frac{\delta_{k-1}}{2}. \tag{102}$$

Proof. Observe that since χ is even and $\int_{\mathbb{R}^\mathsf{d}} \chi_\epsilon = 1$,

$$\begin{aligned} v(x) - v_\epsilon(x) &= \int_{\mathbb{R}^\mathsf{d}} \big(v(x) - v(x-y)\big)\chi_\epsilon(y)\,\mathrm{d}y \\ &= \int_{\mathbb{R}^\mathsf{d}} \big(v(x) - v(x+y)\big)\chi_\epsilon(-y)\,\mathrm{d}y \\ &= \int_{\mathbb{R}^\mathsf{d}} \big(v(x) - v(x+y)\big)\chi_\epsilon(y)\,\mathrm{d}y. \end{aligned} \tag{103}$$

Hence,

$$v(x) - v_\epsilon(x) = \frac12\int_{\mathbb{R}^\mathsf{d}} \big(2v(x) - v(x-y) - v(x+y)\big)\chi_\epsilon(y)\,\mathrm{d}y, \tag{104}$$

which implies that

$$\begin{aligned} \|v - v_\epsilon\|_{L^\infty} &\le \frac12\int_{\mathbb{R}^\mathsf{d}} \left\|\frac{\big(2v(\cdot) - v(\cdot - y) - v(\cdot + y)\big)}{|y|^a}\right\|_{L^\infty} |y|^a\chi_\epsilon(y)\,\mathrm{d}y \\ &\le \frac{\|v\|_{\dot{C}^a}}{2}\int_{\mathbb{R}^\mathsf{d}} |y|^a\chi_\epsilon(y)\,\mathrm{d}y \\ &= \frac{\epsilon^a\|v\|_{\dot{C}^a}}{2}\int_{\mathbb{R}^\mathsf{d}} |y|^a\chi(y)\,\mathrm{d}y, \end{aligned} \tag{105}$$

where the penultimate inequality follows the change of variable $y/\epsilon \mapsto y$. Replacing v by $\nabla^{\otimes k} v$ in the preceding argument, for any $k \ge 0$, then yields the desired conclusion (99).

Let $0 < b \le a \le 1$. Using the interpolation estimate $\|\cdot\|_{\dot{C}^\theta} \le \|\cdot\|_{\dot{C}^{\theta_1}}^t\|\cdot\|_{\dot{C}^{\theta_2}}^{1-t}$ for $\theta_1 \le \theta \le \theta_2$ and $t = \frac{\theta_2-\theta}{\theta_2-\theta_1}$, we have

$$\|v - v_\epsilon\|_{\dot{C}^b} \le \|v - v_\epsilon\|_{L^\infty}^{\frac{a-b}{a}}\|v - v_\epsilon\|_{\dot{C}^a}^{\frac{b}{a}} \le \big(C\|v\|_{\dot{C}^a}\epsilon^a\big)^{\frac{a-b}{a}}\big(2\|v\|_{\dot{C}^a}\big)^{\frac{b}{a}} = C'\|v\|_{\dot{C}^a}\epsilon^{a-b}, \tag{106}$$

where we have used (99) to obtain the second inequality. Replacing v by $\nabla^{\otimes k} v$ then yields (100).

Now suppose that $\left|v(x)-v(y)\right| \le \omega(|x-y|)$ for all $|x-y| \le \delta$ for an increasing $\omega\colon [0,\delta] \to [0,\infty)$. From the identity (103), it follows that if $\epsilon \le \delta/2$,

$$\|v - v_\epsilon\|_{L^\infty} \le \int_{\mathbb{R}^{\mathsf{d}}} \omega(y)\chi_\epsilon(y)\,\mathrm{d}y = \int_{B(0,2)} \omega(\epsilon y)\chi(y)\,\mathrm{d}y \le \omega(2\epsilon)\int_{B(0,2)} \chi(y)\,\mathrm{d}y, \tag{107}$$

where we have again made the change of variable $y/\epsilon \mapsto y$ and used the monotonicity of ω. Replacing v by $\nabla^{\otimes k} v$ then yields (101).

Finally, using that $\nabla\chi_\epsilon$ is odd by assumption that χ_ϵ is even, and therefore $\int_{\mathbb{R}^{\mathsf{d}}} \nabla\chi_\epsilon = 0$, we have

$$\nabla^{\otimes k} v_\epsilon(x) = \int_{\mathbb{R}^{\mathsf{d}}} \left(\nabla^{\otimes k-1} v(x-y) - \nabla^{\otimes k-1} v(x)\right) \otimes \nabla\chi_\epsilon(y)\,\mathrm{d}y. \tag{108}$$

Suppose that $\left|\nabla^{\otimes k-1} v(x) - \nabla^{\otimes k-1} v(y)\right| \le \omega_{k-1}(|x-y|)$ for $|x-y| \le \delta_{k-1}$, where $\omega_{k-1}\colon [0,\delta_{k-1}] \to [0,\infty)$ is increasing. Making the change of variable $y/\epsilon \mapsto y$,

$$\begin{aligned}
\|\nabla^{\otimes k} v_\epsilon\|_{L^\infty} &\le \frac{1}{\epsilon}\int_{\mathbb{R}^{\mathsf{d}}} \left|\nabla^{\otimes k-1} v(x-\epsilon y) - \nabla^{\otimes k-1} v(x)\right|\left|\nabla\chi(y)\right|\mathrm{d}y \\
&\le \frac{1}{\epsilon}\int_{\mathbb{R}^{\mathsf{d}}} \omega_{k-1}(\epsilon y)\left|\nabla\chi(y)\right|\mathrm{d}y \\
&\le \frac{\omega_{k-1}(2\epsilon)}{\epsilon}\int_{\mathbb{R}^{\mathsf{d}}} \left|\nabla\chi(y)\right|\mathrm{d}y,
\end{aligned} \tag{109}$$

where we use the monotonicity of ω_{k-1} to obtain the final line. This yields (102) and completes the proof. $\square$

4.2. *Properties of modulated energy*

In this subsection, we recall some properties of the modulated energy (5) that are important for the proofs of Theorems 8 and 10.

The first result is the following proposition taken from [46, Proposition 2.15] in the singular case $0 \le \mathsf{s} < \mathsf{d}$ and [59, Example 4] in the nonsingular case $-2 < \mathsf{s} < 0$, which shows the coercivity of the modulated energy in the sense that it controls a squared negative-order Sobolev norm up to $O\big(N^{-1+\frac{\mathsf{s}p}{\mathsf{d}(p-1)}}\big)$ error for densities $\mu \in L^p$. In particular, this yields that the modulated energy is almost nonnegative and that vanishing of the modulated energy implies weak convergence of the empirical measure to the target measure μ.

Proposition 16. *Let $0 \le \mathsf{s} < \mathsf{d}$ and $p > \frac{\mathsf{d}}{\mathsf{d}-\mathsf{s}}$. There exists a constant $C = C(\mathsf{d},\mathsf{s},p) > 0$ such that the following holds: for any $\mu \in L^1 \cap L^p$ with $\int_{\mathbb{R}^{\mathsf{d}}} \mathrm{d}\mu = 1$ and $\int_{(\mathbb{R}^{\mathsf{d}})^2} |\mathsf{g}|(x-y)\,\mathrm{d}|\mu|^{\otimes 2}(x,y) < \infty$ if $\mathsf{s} = 0$, $\lambda \le 1$, and any pairwise distinct configuration $X_N \in (\mathbb{R}^{\mathsf{d}})^N$, it holds that*

$$\left\|\frac{1}{N}\sum_{i=1}^N \delta_{x_i} - \mu\right\|_{H^{-r/2}}^2 \le C\left(\mathsf{F}_N(X_N,\mu) + \frac{\left(\mathsf{g}(\lambda)+C\right)}{2N}\mathbf{1}_{\mathsf{s}=0} + C\frac{\mathsf{g}(\lambda)}{2N}\mathbf{1}_{\mathsf{s}>0} + C\|\mu\|_{L^p}\lambda^{\frac{\mathsf{d}(p-1)}{p}-\mathsf{s}}\right), \tag{110}$$

If $-2 < \mathsf{s} < 0$, then for any $\mu \in L^1$ with $\int_{\mathbb{R}^{\mathsf{d}}} \mathrm{d}\mu = 1$ and $\int_{\mathbb{R}^{\mathsf{d}}} |\mathsf{g}|(x-y)\,\mathrm{d}|\mu|^{\otimes 2} < \infty$,

$$\mathsf{F}_N(X_N,\mu) = \mathsf{c}_{\mathsf{d},\mathsf{s}}\left\|\frac{1}{N}\sum_{i=1}^N \delta_{x_i} - \mu\right\|_{\dot{H}^{\frac{\mathsf{s}-\mathsf{d}}{2}}}^2. \tag{111}$$

The second result, taken from [46, Corollary 2.14], shows that the modulated energy controls the small-scale interactions.

Proposition 17. *Let $p > \frac{\mathsf{d}}{\mathsf{d}-\mathsf{s}}$ and $\mu \in L^1 \cap L^p$. Define the nearest-neighbor type distance*

$$\mathsf{r}_i := \frac{1}{4}\min\Big(\min_{1\le j\le N: j\ne i} |x_i - x_j|, \lambda\Big). \tag{112}$$

There exists $C = C(\mathsf{d},\mathsf{s},p) > 0$, such that for every $\eta \le \lambda$,

$$\mathsf{F}_N(X_N,\mu) + \frac{\big(\mathsf{g}(\eta)+C\big)}{2N}\mathbf{1}_{\mathsf{s}=0} + C\frac{\mathsf{g}(\eta)}{2N}\mathbf{1}_{\mathsf{s}>0} + C\|\mu\|_{L^p}\eta^{\frac{\mathsf{d}(p-1)}{p}-\mathsf{s}} \ge \begin{cases} \dfrac{1}{2N^2}\displaystyle\sum_{i=1}^N \mathsf{g}(4\mathsf{r}_i), & \mathsf{s}>0,\\ \dfrac{1}{2N^2}\displaystyle\sum_{i=1}^N \mathsf{g}(4\mathsf{r}_i/\eta), & \mathsf{s}=0, \end{cases} \tag{113}$$

and

$$\mathsf{F}_N(X_N,\mu) + \frac{\big(\mathsf{g}(\eta)+C\big)}{2N}\mathbf{1}_{\mathsf{s}=0} + C\frac{\mathsf{g}(\eta)}{2N}\mathbf{1}_{\mathsf{s}>0} + C\|\mu\|_{L^p}\eta^{\frac{\mathsf{d}(p-1)}{p}-\mathsf{s}} \ge \begin{cases} \dfrac{1}{2CN^2}\displaystyle\sum_{\substack{1\le i\ne j\le N\\ |x_i-x_j|\le\eta}} |x_i-x_j|^{-\mathsf{s}}, & \mathsf{s}>0,\\ \dfrac{1}{2CN^2}\displaystyle\sum_{\substack{1\le i\ne j\le N\\ |x_i-x_j|\le\eta}} -\log\Big(\frac{x_i-x_j}{\eta}\Big), & \mathsf{s}=0. \end{cases} \tag{114}$$

The final result, which is taken from [80] and is an elementary consequence of Hölder's inequality and the Fourier representation of Riesz potentials, shows that in the nonsingular case $-2<\mathsf{s}<0$, the modulated energy/squared MMD controls moments of order $<|\mathsf{s}|/2$.

Proposition 18. *Suppose that $-2<\mathsf{s}<0$, and let $0<r<\frac{|\mathsf{s}|}{2}$. There exists a constant $C = C(\mathsf{d},\mathsf{s},r) > 0$ such that for any $x_0\in\mathbb{R}^\mathsf{d}$, $\mu\in L^1$ with $\int_{\mathbb{R}^\mathsf{d}} d\mu = 1$ and $\int_{(\mathbb{R}^\mathsf{d})^2}|\mathsf{g}|(x-y)\,d|\mu|^{\otimes 2}<\infty$ and $X_N\in(\mathbb{R}^\mathsf{d})^N$, it holds that*

$$\int_{\mathbb{R}^\mathsf{d}} |x-x_0|^r\, d\Big(\frac{1}{N}\sum_{i=1}^N \delta_{x_i} - \mu\Big) \le C\big(1+\|\mu\|_{L^1}\big)^{1-\frac{2r}{|\mathsf{s}|}} \mathsf{F}_N^{\frac{r}{|\mathsf{s}|}}(X_N,\mu). \tag{115}$$

4.3. *Defective commutator estimate*

We now combine the functional inequalities of Theorem 1 with the mollification error estimates from Lemma 15 to obtain a defective commutator estimate in the form of Theorem 8 below. As commented previously, this weakening of the Lipschitz assumption is necessary in the case $v = \mathbb{M}\nabla\mathsf{g}*\mu$, where μ belongs to the critical space $\dot{W}^{\frac{\mathsf{d}}{p}+\mathsf{s}+2-\mathsf{d},p}$ for $1<p<\infty$. Below, χ_ϵ and v_ϵ are as in the previous subsections.

In what follows, introduce the commutator kernel notation

$$k_v(x,y) := \nabla\mathsf{g}(x-y)\cdot\big(v(x)-v(y)\big). \tag{116}$$

The proof of Theorem 8 follows similar lines to [69,70]. The main technical ingredient is Lemma 19 below, which controls the error from replacing v by the mollified vector field v_ϵ (cf. [69, Lemma 4.2] and [70, Lemma 4.3]). The proof of the lemma is streamlined in comparison with previous works, taking advantage of better mollification estimates in Lemma 15 and small-scale control in Corollary 17. Note that we omit the log case $\mathsf{s}=0$ in Lemma 19 as this is covered by the same argument in [69], and the omission simplifies the presentation.

Lemma 19. *Assume that $-2<\mathsf{s}<\mathsf{d}$ and $\mathsf{s}\ne 0$. Let $v\in\dot{\mathcal{C}}^1$ and $\mu\in L^1$ with $\int_{\mathbb{R}^\mathsf{d}} d\mu = 1$ and $\int_{(\mathbb{R}^\mathsf{d})^2}|\mathsf{g}|(x-y)\,d|\mu|^{\otimes 2}<\infty$. Further, assume the following:*

(a) *if $-2<\mathsf{s}\le -1$, then $\int_{\mathbb{R}^\mathsf{d}}|x|^{|\mathsf{s}|-1}\,d|\mu|<\infty$;*
(b) *if $-1<\mathsf{s}<0$, then $\int_{\mathbb{R}^\mathsf{d}}|x|^r\,d\mu<\infty$ for some $r<\frac{|\mathsf{s}|}{2}$;*
(c) *if $-1\le\mathsf{s}<\mathsf{d}-1$, then $\mu\in L^p$ for some $\frac{\mathsf{d}}{\mathsf{d}-\mathsf{s}-1}<p\le\infty$ ($p=1$ is allowed if $\mathsf{s}=-1$);*
(d) *if $\mathsf{s}\ge\mathsf{d}-1$, then $\mu\in\dot{\mathcal{C}}^\theta$ for some $\theta>\mathsf{s}+1-\mathsf{d}$.*

Then for any pairwise distinct configuration $X_N \in (\mathbb{R}^{\mathsf{d}})^N$ *and* $\epsilon \geq 0$*,*

$$\begin{aligned}
&\left|\int_{(\mathbb{R}^{\mathsf{d}})^2\setminus\triangle} k_{(v-v_\epsilon)}(x,y)\,\mathrm{d}\left(\frac{1}{N}\sum_{i=1}^N \delta_{x_i}-\mu\right)^{\otimes 2}\right| \\
&\quad\leq CN^{\frac{2(\mathsf{s}+1)}{\mathsf{s}}-1}\|v\|_{\dot{\mathcal{C}}^1}\epsilon\left(\mathsf{F}_N(X_N,\mu)+C_p\|\mu\|_{L^p}\lambda^{\frac{\mathsf{d}(p-1)}{p}-\mathsf{s}}\right)^{\frac{\mathsf{s}+1}{\mathsf{s}}}\mathbf{1}_{0<\mathsf{s}<\mathsf{d}} \\
&\qquad+\left(C_{\vartheta,\vartheta'}\|v\|_{\dot{\mathcal{C}}^1}\epsilon^{\vartheta'}\left(\int_{\mathbb{R}^{\mathsf{d}}}|x|^{|\mathsf{s}|-\vartheta}\,\mathrm{d}\mu+\left(1+\|\mu\|_{L^1}\right)\mathsf{F}_N^{\frac{|\mathsf{s}|-\vartheta}{|\mathsf{s}|}}(X_N,\mu)\right)+C\|v\|_{\dot{\mathcal{C}}^1}\epsilon\right)\mathbf{1}_{-1<\mathsf{s}<0} \\
&\qquad+C\|v\|_{\dot{\mathcal{C}}^1}\epsilon\left(\int_{\mathbb{R}^{\mathsf{d}}}|x|^{|\mathsf{s}|-1}\,\mathrm{d}\mu+\left(1+\|\mu\|_{L^1}\right)\mathsf{F}_N^{\frac{|\mathsf{s}|-1}{|\mathsf{s}|}}(X_N,\mu)\right)\mathbf{1}_{-2<\mathsf{s}\leq -1} \\
&\qquad+\epsilon\left(1+\|\mu\|_{L^1}\right)\|v\|_{\dot{\mathcal{C}}^1}\begin{cases}\int_{\mathbb{R}^{\mathsf{d}}}|x|^{|\mathsf{s}|-1}\,\mathrm{d}\mu+\mathsf{F}_N^{\frac{|\mathsf{s}|-1}{|\mathsf{s}|}}(X_N,\mu), & -2<\mathsf{s}\leq -1,\\
C_q\|\mu\|_{L^1}^{1-\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}}\|\mu\|_{L^q}^{\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}}, & -1\leq \mathsf{s}<\mathsf{d}-1,\\
C_\theta\|\mu\|_{L^1}^{1-\frac{\mathsf{s}-\mathsf{d}+1}{\theta}}\|\mu\|_{\dot{\mathcal{C}}^\theta}^{\frac{\mathsf{s}-\mathsf{d}+1}{\theta}} & \\
\qquad+C_\theta\epsilon^{\mathsf{d}-\mathsf{s}-1}\|\mu\|_{L^1}^{1-\frac{\mathsf{s}-\mathsf{d}+1}{\theta}}\|\mu\|_{L^\infty}^{\frac{\mathsf{s}-\mathsf{d}+1}{\theta}}, & \mathsf{s}\geq \mathsf{d}-1,\end{cases}
\end{aligned} \tag{117}$$

where $\vartheta \in \left[|\mathsf{s}|-r, |\mathsf{s}|\right)$, $0<\vartheta'<\vartheta$, $C=C(\mathsf{d},\mathsf{s})>0$ *and* $C_{\vartheta,\vartheta'}, C_p, C_\theta > 0$ *additionally depend on* (ϑ,ϑ'), p, θ*, respectively.*

Before proving Lemma 19, let us show how Theorem 8 follows from it and Theorem 1.

Proof of Theorem 8. Writing $v = v_\epsilon + (v-v_\epsilon)$ and applying the triangle inequality, we find that

$$\begin{aligned}
&\left|\int_{(\mathbb{R}^{\mathsf{d}})^2\setminus\triangle} k_{v}(x,y)\,\mathrm{d}\left(\frac{1}{N}\sum_{i=1}^N \delta_{x_i}-\mu\right)^{\otimes 2}\right| \\
&\quad\leq \left|\int_{(\mathbb{R}^{\mathsf{d}})^2\setminus\triangle} k_{v_\epsilon}(x,y)\,\mathrm{d}\left(\frac{1}{N}\sum_{i=1}^N \delta_{x_i}-\mu\right)^{\otimes 2}\right| + \left|\int_{(\mathbb{R}^{\mathsf{d}})^2\setminus\triangle} k_{(v-v_\epsilon)}(x,y)\,\mathrm{d}\left(\frac{1}{N}\sum_{i=1}^N \delta_{x_i}-\mu\right)^{\otimes 2}\right|.
\end{aligned} \tag{118}$$

Depending on the value of s, we apply the estimates (12), (13), (14), or (15) from Theorem 1 to the first term on the right-hand side of (118). We use (102) and the modulus of continuity bound (49) to control

$$\|\nabla v_\epsilon\|_{L^\infty} \leq C\|v\|_{\dot{W}^{\frac{\mathsf{d}}{p}+1,p}}\left(1+|\log\epsilon|\right)^{1-\frac{1}{p}}, \tag{119}$$

where $C = C(\mathsf{d},p) > 0$. If $v \in \dot{W}^{\frac{\mathsf{a}}{2},\frac{2\mathsf{d}}{\mathsf{a}-2}}$, then by commutativity of differentiation and mollification,

$$\left\||\nabla|^{\frac{\mathsf{a}}{2}} v_\epsilon\right\|_{L^{\frac{2\mathsf{d}}{\mathsf{a}-2}}} \leq \left\||\nabla|^{\frac{\mathsf{a}}{2}} v\right\|_{L^{\frac{2\mathsf{d}}{\mathsf{a}-2}}}. \tag{120}$$

Altogether, we conclude that

$$\begin{aligned}
&\left|\int_{(\mathbb{R}^{\mathsf{d}})^2\setminus\triangle} k_{v_\epsilon}(x,y)\,\mathrm{d}\left(\frac{1}{N}\sum_{i=1}^N \delta_{x_i}-\mu\right)^{\otimes 2}\right| \\
&\leq C\|v\|_{\dot{W}^{\frac{\mathsf{d}}{p}+1,p}}\left(1+|\log\epsilon|\right)^{1-\frac{1}{p}}\left(\mathsf{F}_N(X_N,\mu)+C_q\|\mu\|_{L^q}\lambda^{\frac{\mathsf{d}(q-1)}{q}-\mathsf{s}}\right)\mathbf{1}_{\mathsf{s}\geq\max(0,\mathsf{d}-2)} \\
&\quad+C\min\Bigg(\left(\|v\|_{\dot{W}^{\frac{\mathsf{d}}{p}+1,p}}\left(1+|\log\epsilon|\right)^{1-\frac{1}{p}}+\left\||\nabla|^{\frac{\mathsf{a}}{2}}v\right\|_{L^{\frac{2\mathsf{d}}{\mathsf{a}-2}}}\mathbf{1}_{\mathsf{a}>2}\right)\left(\mathsf{F}_N(X_N,\mu)+C_q\|\mu\|_{L^q}\lambda^{\frac{\mathsf{d}(q-1)}{q}-\mathsf{s}}\right), \\
&\qquad\left(\|v\|_{\dot{W}^{\frac{\mathsf{d}}{p}+1,p}}\left(1+|\log\epsilon|\right)^{1-\frac{1}{p}}+\left\||\nabla|^{\frac{\mathsf{d}-\mathsf{s}}{2}}v\right\|_{L^{\frac{2\mathsf{d}}{\mathsf{d}-\mathsf{s}-2}}}\right)\left(\mathsf{F}_N(X_N,\mu)+C_{\tilde{q}}\|\mu\|_{L^{\tilde{q}}}\kappa^{\frac{\mathsf{d}(\tilde{q}-1)}{\tilde{q}}-\mathsf{s}}\right)\Bigg)\mathbf{1}_{0\leq\mathsf{s}<\mathsf{d}-2} \\
&\quad+C\left(\|v\|_{\dot{W}^{\frac{\mathsf{d}}{p}+1,p}}\left(1+|\log\epsilon|\right)^{1-\frac{1}{p}}+\left\||\nabla|^{\frac{\mathsf{d}-\mathsf{s}}{2}}v\right\|_{L^{\frac{2\mathsf{d}}{\mathsf{d}-\mathsf{s}-2}}}\right)\mathsf{F}_N(X_N,\mu)\mathbf{1}_{-2<\mathsf{s}<0},
\end{aligned} \tag{121}$$

for $q > \frac{\mathsf{d}}{\mathsf{d}-\mathsf{s}}$ and $\widetilde{q} > \frac{\mathsf{d}}{\mathsf{d}-\mathsf{s}-1}$, where λ, κ are as in (10), (11), respectively. Applying (117) to the second term on the right-hand side of (118) and combining the resulting bound with (121) then completes the proof. $\square$

We now turn to the proof of Lemma 19.

Proof of Lemma 19. Our starting point is the decomposition

$$\int_{(\mathbb{R}^{\mathsf{d}})^2\setminus\triangle} k_{(v-v_\epsilon)}(x,y)\,\mathrm{d}\bigg(\frac{1}{N}\sum_{i=1}^N \delta_{x_i} - \mu\bigg)^{\otimes 2}(x,y) = \sum_{j=1}^3 \mathrm{Term}_j, \tag{122}$$

where

$$\mathrm{Term}_1 = \frac{1}{N^2}\sum_{1\le i\ne j\le N} k_{(v-v_\epsilon)}(x_i,x_j), \tag{123}$$

$$\mathrm{Term}_2 = -2\frac{1}{N}\sum_{i=1}^N \int_{\mathbb{R}^{\mathsf{d}}} k_{(v-v_\epsilon)}(x_i,y)\,\mathrm{d}\mu(y), \tag{124}$$

$$\mathrm{Term}_3 = \int_{\mathbb{R}^{\mathsf{d}}} k_{(v-v_\epsilon)}(x,y)\,\mathrm{d}\mu^{\otimes 2}(x,y). \tag{125}$$

We separately estimate each of $\mathrm{Term}_1, \mathrm{Term}_2, \mathrm{Term}_3$.

$\mathbf{Term_1}$. For $1 \le i \le N$, let r_i be the nearest-neighbor type distance from (112). By (99) with $a = 1$, we have the elementary bound

$$\forall\, x \ne y, \quad \big|k_{(v-v_\epsilon)}(x,y)\big| \le \frac{C\|v\|_{\dot{C}^1}\epsilon}{|x-y|^{\mathsf{s}+1}}. \tag{126}$$

We now consider cases based on the value of s.

If $0 < \mathsf{s} < \mathsf{d}$, then applying the bound (126) to each summand of Term_1,

$$\begin{aligned}
|\mathrm{Term}_1| &\le \frac{C\|v\|_{\dot{C}^1}\epsilon}{N^2}\sum_{i=1}^N\sum_{j\ne i}|x_i-x_j|^{-\mathsf{s}-1} \\
&\le \frac{C\|v\|_{\dot{C}^1}\epsilon}{N}\sum_{i=1}^N \mathsf{r}_i^{-\mathsf{s}-1} \\
&\le \frac{C\|v\|_{\dot{C}^1}\epsilon}{N}\bigg(\sum_{i=1}^N \mathsf{r}_i^{-\mathsf{s}}\bigg)^{\frac{\mathsf{s}+1}{\mathsf{s}}} \\
&\le \frac{C'(\mathsf{s}N^2)^{\frac{\mathsf{s}+1}{\mathsf{s}}}\epsilon}{N}\Big(\mathsf{F}_N(X_N,\mu) + C_p\|\mu\|_{L^p}\lambda^{\frac{\mathsf{d}(p-1)}{p}-\mathsf{s}}\Big)^{\frac{\mathsf{s}+1}{\mathsf{s}}},
\end{aligned} \tag{127}$$

where the second inequality follows from the definition of r_i, the third inequality from the embedding $\|\cdot\|_{\ell^{\mathsf{s}+1}} \le \|\cdot\|_{\ell^{\mathsf{s}}}$, and the fourth inequality from Corollary 17. Above, $C, C' > 0$ depend only on d, s and $\frac{\mathsf{d}}{\mathsf{d}-\mathsf{s}} < p \le \infty$. If $-2 < \mathsf{s} \le -1$, then we use $1 > |\mathsf{s}| - 1 \ge 0$ to instead argue that

$$\begin{aligned}
|\mathrm{Term}_1| &\le \frac{C\|v\|_{\dot{C}^1}\epsilon}{N^2}\sum_{i=1}^N\sum_{j\ne i}|x_i-x_j|^{-\mathsf{s}-1} \\
&\le \frac{C\|v\|_{\dot{C}^1}\epsilon}{N^2}\sum_{i=1}^N\sum_{j\ne i}\big(|x_i|^{|\mathsf{s}|-1} + |x_j|^{|\mathsf{s}|-1}\big) \\
&\le \frac{2C\|v\|_{\dot{C}^1}\epsilon}{N}\sum_{i=1}^N |x_i|^{|\mathsf{s}|-1} \\
&\le 2C'\|v\|_{\dot{C}^1}\epsilon\bigg(\int_{\mathbb{R}^{\mathsf{d}}}|x|^{|\mathsf{s}|-1}\,\mathrm{d}\mu + \big(1+\|\mu\|_{L^1}\big)^{1-\frac{2(|\mathsf{s}|-1)}{|\mathsf{s}|}}\mathsf{F}_N^{\frac{|\mathsf{s}|-1}{|\mathsf{s}|}}(X_N,\mu)\bigg),
\end{aligned} \tag{128}$$

where we have used $|\mathsf{s}| - 1 < \frac{|\mathsf{s}|}{2}$ and Proposition 18 to obtain the final line.

If $-1 < \mathsf{s} < 0$, then in addition to (126) we observe the bound

$$\forall\, |x-y| \le 1, \quad \left|k_{(v-v_\epsilon)}(x,y)\right| \le C\|v\|_{\dot{\mathcal{C}}^1}|x-y|^{-\mathsf{s}}\big(1+\big|\log|x-y|\big|\big). \tag{129}$$

which is a consequence of $\dot{\mathcal{C}}^1$ functions being log-Lipschitz (e.g. see [2, Proposition 2.107]). Interpolating between (126) and (129), we find that for any $\vartheta \in [0,1]$,

$$\forall\, |x-y| \le 1, \quad \left|k_{(v-v_\epsilon)}(x,y)\right| \le C\|v\|_{\dot{\mathcal{C}}^1}|x-y|^{-\mathsf{s}}\left(\frac{\epsilon}{|x-y|}\right)^{\vartheta}\big(1+\big|\log|x-y|\big|\big)^{1-\vartheta}. \tag{130}$$

Note that we may discard the log factor up to taking ϑ smaller in the exponent of ϵ. Indeed, for $1 \ge \vartheta > \vartheta'$, we have that

$$\sup_{|x-y|\le 1} |x-y|^{\vartheta-\vartheta'}\big(1+\big|\log|x-y|\big|\big)^{1-\vartheta'} \le C', \tag{131}$$

and therefore,

$$\begin{aligned}|x-y|^{-\mathsf{s}}\left(\frac{\epsilon}{|x-y|}\right)^{\vartheta'}\big(1+\big|\log|x-y|\big|\big)^{1-\vartheta'} &= \epsilon^{\vartheta'}|x-y|^{|\mathsf{s}|-\vartheta}|x-y|^{\vartheta-\vartheta'}\big(1+\big|\log|x-y|\big|\big)^{1-\vartheta'} \\ &\le C'\epsilon^{\vartheta'}|x-y|^{|\mathsf{s}|-\vartheta}.\end{aligned} \tag{132}$$

Hence, choosing $0 < \vartheta' < \vartheta < 1$ so that $r \ge |\mathsf{s}| - \vartheta > 0$, which translates to $|\mathsf{s}| - r \le \vartheta < |\mathsf{s}|$, and arguing similarly to the previous case,

$$\begin{aligned}|\mathrm{Term}_1| &\le \frac{1}{N^2}\sum_{i=1}^N \sum_{j:|x_i-x_j|\le 1}\left|k_{(v-v_\epsilon)}(x,y)\right| + \frac{1}{N^2}\sum_{i=1}^N\sum_{j:|x_i-x_j|>1}\left|k_{(v-v_\epsilon)}(x,y)\right| \\ &\le \frac{C\|v\|_{\dot{\mathcal{C}}^1}\epsilon^{\vartheta'}}{N}\sum_{i=1}^N |x_i|^{|\mathsf{s}|-\vartheta} + C\|v\|_{\dot{\mathcal{C}}^1}\epsilon \\ &\le C\|v\|_{\dot{\mathcal{C}}^1}\epsilon^{\vartheta'}\left(\int_{\mathbb{R}^\mathsf{d}}|x|^{|\mathsf{s}|-\vartheta}\,d\mu + \big(1+\|\mu\|_{L^1}\big)^{1-\frac{2(|\mathsf{s}|-\vartheta)}{|\mathsf{s}|}}\mathsf{F}_N^{\frac{|\mathsf{s}|-\vartheta}{|\mathsf{s}|}}(X_N,\mu)\right) + C\|v\|_{\dot{\mathcal{C}}^1}\epsilon.\end{aligned} \tag{133}$$

Combining the bounds (127), (128), (133) for $0 < \mathsf{s} < \mathsf{d}$, $-2 < \mathsf{s} \le -1$, and $-1 < \mathsf{s} < 0$, respectively, we conclude

$$\begin{aligned}&|\mathrm{Term}_1| \\ &\le \frac{C(N^2)^{\frac{\mathsf{s}+1}{\mathsf{s}}}\epsilon}{N}\left(\mathsf{F}_N(X_N,\mu) + C_p\|\mu\|_{L^p}\lambda^{\frac{\mathsf{d}(p-1)}{p}-\mathsf{s}}\right)^{\frac{\mathsf{s}+1}{\mathsf{s}}}\mathbf{1}_{0<\mathsf{s}<\mathsf{d}} \\ &+ C\|v\|_{\dot{\mathcal{C}}^1}\epsilon\left(\int_{\mathbb{R}^\mathsf{d}}|x|^{|\mathsf{s}|-1}\,d\mu + \big(1+\|\mu\|_{L^1}\big)^{1-\frac{2(|\mathsf{s}|-1)}{|\mathsf{s}|}}\mathsf{F}_N^{\frac{|\mathsf{s}|-1}{|\mathsf{s}|}}(X_N,\mu)\right)\mathbf{1}_{-2<\mathsf{s}\le -1} \\ &+ \left(C_{\vartheta,\vartheta'}\|v\|_{\dot{\mathcal{C}}^1}\epsilon^{\vartheta'}\left(\int_{\mathbb{R}^\mathsf{d}}|x|^{|\mathsf{s}|-\vartheta}\,d\mu + \big(1+\|\mu\|_{L^1}\big)^{1-\frac{2(|\mathsf{s}|-\vartheta)}{|\mathsf{s}|}}\mathsf{F}_N^{\frac{|\mathsf{s}|-\vartheta}{|\mathsf{s}|}}(X_N,\mu)\right) + C\|v\|_{\dot{\mathcal{C}}^1}\epsilon\right)\mathbf{1}_{-1<\mathsf{s}<0},\end{aligned} \tag{134}$$

where $C = C(\mathsf{d},\mathsf{s}) > 0$ and $C_{\vartheta,\vartheta'} > 0$ additionally depends on ϑ, ϑ'.

Term$_2$. Unpacking the definition of $k_{(v-v_\epsilon)}$,

$$\int_{\mathbb{R}^\mathsf{d}} k_{(v-v_\epsilon)}(\cdot,y)\,d\mu(y) = (v-v_\epsilon)\cdot\nabla h^\mu + \operatorname{div} h^{(v-v_\epsilon)\mu}, \tag{135}$$

where $h^f := \mathsf{g} * f$. For the first term on the right-hand side, we crudely bound

$$\begin{aligned}\left\|(v-v_\epsilon)\cdot\nabla h^\mu\right\|_{L^\infty} &\le \|v-v_\epsilon\|_{L^\infty}\|\nabla h^\mu\|_{L^\infty} \\ &\le \|v\|_{\dot{\mathcal{C}}^1}\epsilon\begin{cases} C_q\|\mu\|_{L^1}^{1-\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}}\|\mu\|_{L^q}^{\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}}, & -1 \le \mathsf{s} < \mathsf{d}-1, \\ C_\theta\|\mu\|_{L^1}^{1-\frac{\mathsf{s}-\mathsf{d}+1}{\theta}}\|\mu\|_{\dot{\mathcal{C}}^\theta}^{\frac{\mathsf{s}-\mathsf{d}+1}{\theta}}, & \mathsf{s} \ge \mathsf{d}-1,\end{cases}\end{aligned} \tag{136}$$

for $q > \frac{\mathsf{d}}{\mathsf{d}-\mathsf{s}-1}$ ($q \ge 1$ if $\mathsf{s} = -1$) and $\theta > \mathsf{s}+1-\mathsf{d}$, where in the final inequality, we have used (99) and the interpolation estimate

$$\|\nabla h^f\|_{L^\infty} \le \begin{cases} C_q \|f\|_{L^1}^{1-\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}} \|f\|_{L^q}^{\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}}, & -1 \le \mathsf{s} < \mathsf{d}-1, \\ C_\theta \|f\|_{L^1}^{1-\frac{\mathsf{s}-\mathsf{d}+1}{\theta}} \|f\|_{\dot{C}^\theta}^{\frac{\mathsf{s}-\mathsf{d}+1}{\theta}}, & \mathsf{s} \ge \mathsf{d}-1. \end{cases} \tag{137}$$

For the second term on the right-hand side of (135), we again use the interpolation estimate (137) and Hölder's inequality to obtain

$$\begin{aligned} &\left\|\operatorname{div} h^{(v-v_\epsilon)\mu}\right\|_{L^\infty} \\ &\le \begin{cases} C_q \left\|(v-v_\epsilon)\mu\right\|_{L^1}^{1-\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}} \left\|(v-v_\epsilon)\mu\right\|_{L^q}^{\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}}, & -1 \le \mathsf{s} < \mathsf{d}-1, \\ C_\theta \left\|(v-v_\epsilon)\mu\right\|_{L^1}^{1-\frac{\mathsf{s}-\mathsf{d}+1}{\theta}} \left\|(v-v_\epsilon)\mu\right\|_{\dot{C}^\theta}^{\frac{\mathsf{s}-\mathsf{d}+1}{\theta}}, & \mathsf{s} \ge \mathsf{d}-1, \end{cases} \\ &\le \begin{cases} C_q \|v-v_\epsilon\|_{L^\infty} \|\mu\|_{L^1}^{1-\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}} \|\mu\|_{L^q}^{\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}}, & -1 \le \mathsf{s} < \mathsf{d}-1, \\ C_\theta \big(\|v-v_\epsilon\|_{L^\infty} \|\mu\|_{L^1}\big)^{1-\frac{\mathsf{s}-\mathsf{d}+1}{\theta}} \big(\|v-v_\epsilon\|_{L^\infty}\|\mu\|_{\dot{C}^\theta} + \|v-v_\epsilon\|_{\dot{C}^\theta}\|\mu\|_{L^\infty}\big)^{\frac{\mathsf{s}-\mathsf{d}+1}{\theta}}, & \mathsf{s} \ge \mathsf{d}-1, \end{cases} \\ &\le \begin{cases} C_q C\epsilon \|v\|_{\dot{C}^1} \|\mu\|_{L^1}^{1-\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}} \|\mu\|_{L^q}^{\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}}, & -1 \le \mathsf{s} < \mathsf{d}-1, \\ C_\theta \big(C\epsilon\|v\|_{\dot{C}^1} \|\mu\|_{L^1}\big)^{1-\frac{\mathsf{s}-\mathsf{d}+1}{\theta}} \big(C\epsilon\|v\|_{\dot{C}^1}\|\mu\|_{\dot{C}^\theta} + C'_\theta\|v\|_{\dot{C}^1}\epsilon^{1-\theta}\|\mu\|_{L^\infty}\big)^{\frac{\mathsf{s}-\mathsf{d}+1}{\theta}}, & \mathsf{s} \ge \mathsf{d}-1, \end{cases} \end{aligned} \tag{138}$$

where we have also used the algebra property of Hölder–Zygmund spaces.

If $-2 < \mathsf{s} < -1$, then we instead use (126) to bound

$$\int_{\mathbb{R}^\mathsf{d}} \left|k_{(v-v_\epsilon)}(x,y)\right| \mathrm{d}|\mu|(y) \le C\epsilon\|v\|_{\dot{C}^1} \int_{\mathbb{R}^\mathsf{d}} |x-y|^{-\mathsf{s}-1}\,\mathrm{d}|\mu|(y). \tag{139}$$

Using Proposition 18 with $x_0 = y$

$$\begin{aligned} &\frac{1}{N}\sum_{i=1}^N \int_{\mathbb{R}^\mathsf{d}} |x_i - y|^{|\mathsf{s}|-1}\,\mathrm{d}|\mu|(y) \\ &\qquad \le C\int_{\mathbb{R}^\mathsf{d}} \left(\int_{\mathbb{R}^\mathsf{d}} |x-y|^{|\mathsf{s}|-1}\,\mathrm{d}\mu + \big(1+\|\mu\|_{L^1}\big)^{1-\frac{2(|\mathsf{s}|-1)}{|\mathsf{s}|}} \mathsf{F}_N^{\frac{|\mathsf{s}|-1}{|\mathsf{s}|}}(X_N,\mu)\right)\mathrm{d}|\mu|(y) \\ &\qquad \le C'\|\mu\|_{L^1}\left(\int_{\mathbb{R}^\mathsf{d}} |x|^{|\mathsf{s}|-1}\,\mathrm{d}|\mu| + \big(1+\|\mu\|_{L^1}\big)^{1-\frac{2(|\mathsf{s}|-1)}{|\mathsf{s}|}} \mathsf{F}_N^{\frac{|\mathsf{s}|-1}{|\mathsf{s}|}}(X_N,\mu)\right). \end{aligned} \tag{140}$$

Hence,

$$\begin{aligned} &\frac{1}{N}\sum_{i=1}^N \int_{\mathbb{R}^\mathsf{d}} \left|k_{(v-v_\epsilon)}(x_i,y)\right|\mathrm{d}|\mu|(y) \\ &\qquad \le C''\epsilon\|v\|_{\dot{C}^1}\|\mu\|_{L^1}\left(\int_{\mathbb{R}^\mathsf{d}} |x|^{|\mathsf{s}|-1}\,\mathrm{d}|\mu| + \big(1+\|\mu\|_{L^1}\big)^{1-\frac{2(|\mathsf{s}|-1)}{|\mathsf{s}|}} \mathsf{F}_N^{\frac{|\mathsf{s}|-1}{|\mathsf{s}|}}(X_N,\mu)\right). \end{aligned} \tag{141}$$

Combining the estimates (135), (136), (138), (141), simplifying and relabeling constants, we conclude that

$$|\mathrm{Term}_2| \le \epsilon\|v\|_{\dot{C}^1} \begin{cases} \|\mu\|_{L^1}\left(\int_{\mathbb{R}^\mathsf{d}} |x|^{|\mathsf{s}|-1}\,\mathrm{d}\mu + \big(1+\|\mu\|_{L^1}\big)^{\frac{2}{|\mathsf{s}|}} \mathsf{F}_N^{\frac{|\mathsf{s}|-1}{|\mathsf{s}|}}(X_N,\mu)\right), & -2 < \mathsf{s} < -1, \\ C_q\|\mu\|_{L^1}^{1-\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}} \|\mu\|_{L^q}^{\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}}, & -1 \le \mathsf{s} < \mathsf{d}-1, \\ C_\theta\|\mu\|_{L^1}^{1-\frac{\mathsf{s}-\mathsf{d}+1}{\theta}} \|\mu\|_{\dot{C}^\theta}^{\frac{\mathsf{s}-\mathsf{d}+1}{\theta}} + C_\theta\epsilon^{\mathsf{d}-\mathsf{s}-1}\|\mu\|_{L^1}^{1-\frac{\mathsf{s}-\mathsf{d}+1}{\theta}} \|\mu\|_{L^\infty}^{\frac{\mathsf{s}-\mathsf{d}+1}{\theta}}, & \mathsf{s} \ge \mathsf{d}-1, \end{cases} \tag{142}$$

for $q > \frac{\mathsf{d}}{\mathsf{d}-\mathsf{s}-1}$ ($q \ge 1$ if $\mathsf{s} = -1$) and $\theta > \mathsf{s}+1-\mathsf{d}$.

Term$_3$. Unpacking the definition of $k_{(v-v_\epsilon)}$ and desymmetrizing,

$$\mathrm{Term}_3 = \int_{\mathbb{R}^{\mathsf{d}}} (v - v_\epsilon)\cdot\nabla h^\mu \, \mathrm{d}\mu, \tag{143}$$

and by Hölder's inequality,

$$\left|\int_{\mathbb{R}^{\mathsf{d}}} (v - v_\epsilon)\cdot\nabla h^\mu \, \mathrm{d}\mu\right| \le \|v - v_\epsilon\|_{L^\infty}\|\nabla h^\mu\|_{L^\infty}\|\mu\|_{L^1}. \tag{144}$$

Using the estimate (136), we conclude that

$$|\mathrm{Term}_3| \le \epsilon\|v\|_{\dot{\mathcal{C}}^1}\|\mu\|_{L^1}\begin{cases} C_q\|\mu\|_{L^1}^{1-\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}}\|\mu\|_{L^q}^{\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}}, & -1 \le \mathsf{s} < \mathsf{d}-1, \\ C_\theta\|\mu\|_{L^1}^{1-\frac{\mathsf{s}-\mathsf{d}+1}{\theta}}\|\mu\|_{\dot{\mathcal{C}}^\theta}^{\frac{\mathsf{s}-\mathsf{d}+1}{\theta}}, & \mathsf{s} \ge \mathsf{d}-1, \end{cases} \tag{145}$$

for $q > \frac{\mathsf{d}}{\mathsf{d}-\mathsf{s}-1}$ and $\theta > \mathsf{s}+1-\mathsf{d}$.

If $-2 < \mathsf{s} < -1$, then we instead argue similarly to Term$_2$ to obtain

$$|\mathrm{Term}_3| \le C\epsilon\|v\|_{\dot{\mathcal{C}}^1}\|\mu\|_{L^1}\int_{\mathbb{R}^{\mathsf{d}}}|x|^{|\mathsf{s}|-1}\,\mathrm{d}\mu. \tag{146}$$

Combining the estimates (145), (146), we conclude that

$$|\mathrm{Term}_3| \le \epsilon\|v\|_{\dot{\mathcal{C}}^1}\|\mu\|_{L^1}\begin{cases} \int_{\mathbb{R}^{\mathsf{d}}}|x|^{|\mathsf{s}|-1}\,\mathrm{d}\mu, & -2 < -\mathsf{s} < -1, \\ C_q\|\mu\|_{L^1}^{1-\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}}\|\mu\|_{L^q}^{\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}}, & -1 \le \mathsf{s} < \mathsf{d}-1, \\ C_\theta\|\mu\|_{L^1}^{1-\frac{\mathsf{s}-\mathsf{d}+1}{\theta}}\|\mu\|_{\dot{\mathcal{C}}^\theta}^{\frac{\mathsf{s}-\mathsf{d}+1}{\theta}}, & \mathsf{s} \ge \mathsf{d}-1, \end{cases} \tag{147}$$

for $q > \frac{\mathsf{d}}{\mathsf{d}-\mathsf{s}-1}$ and $\theta > \mathsf{s}+1-\mathsf{d}$.

Conclusion of proof. Combining the estimates (134), (142), (147) for Term$_1$, Term$_2$, Term$_3$, respectively, we arrive at

$$\begin{aligned}
&\left|\int_{(\mathbb{R}^{\mathsf{d}})^2\setminus\triangle} k_{(v-v_\epsilon)}(x,y)\,\mathrm{d}\left(\frac{1}{N}\sum_{i=1}^N \delta_{x_i} - \mu\right)^{\otimes 2}\right| \\
&\le \frac{C(N^2)^{\frac{\mathsf{s}+1}{\mathsf{s}}}\|v\|_{\dot{\mathcal{C}}^1}\epsilon}{N}\left(\mathsf{F}_N(X_N,\mu) + C_p\|\mu\|_{L^p}\lambda^{\frac{\mathsf{d}(p-1)}{p}-\mathsf{s}}\right)^{\frac{\mathsf{s}+1}{\mathsf{s}}}\mathbf{1}_{0<\mathsf{s}<\mathsf{d}} \\
&\quad + C\|v\|_{\dot{\mathcal{C}}^1}\epsilon\left(1+\|\mu\|_{L^1}\right)^{\frac{2}{|\mathsf{s}|}}\left(\int_{\mathbb{R}^{\mathsf{d}}}|x|^{|\mathsf{s}|-1}\,\mathrm{d}\mu + \mathsf{F}_N^{\frac{|\mathsf{s}|-1}{|\mathsf{s}|}}(X_N,\mu)\right)\mathbf{1}_{-2<\mathsf{s}\le-1} \\
&\quad + C_{\vartheta,\vartheta'}\|v\|_{\dot{\mathcal{C}}^1}\left(\epsilon^{\vartheta'}\left(\int_{\mathbb{R}^{\mathsf{d}}}|x|^{|\mathsf{s}|-\vartheta}\,\mathrm{d}\mu + \left(1+\|\mu\|_{L^1}\right)^{\frac{2\vartheta}{|\mathsf{s}|}}\mathsf{F}_N^{\frac{|\mathsf{s}|-\vartheta}{|\mathsf{s}|}}(X_N,\mu)\right) + \epsilon\right)\mathbf{1}_{-1<\mathsf{s}<0} \\
&\quad + \epsilon\|\mu\|_{L^1}\|v\|_{\dot{\mathcal{C}}^1}\begin{cases} C\left(\int_{\mathbb{R}^{\mathsf{d}}}|x|^{|\mathsf{s}|-1}\,\mathrm{d}\mu + \left(1+\|\mu\|_{L^1}\right)^{\frac{2}{|\mathsf{s}|}}\mathsf{F}_N^{\frac{|\mathsf{s}|-1}{|\mathsf{s}|}}(X_N,\mu)\right), & -2 < \mathsf{s} \le -1, \\ C_q\|\mu\|_{L^1}^{1-\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}}\|\mu\|_{L^q}^{\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}}, & -1 \le \mathsf{s} < \mathsf{d}-1, \\ C_\theta\|\mu\|_{L^1}^{1-\frac{\mathsf{s}-\mathsf{d}+1}{\theta}}\|\mu\|_{\dot{\mathcal{C}}^\theta}^{\frac{\mathsf{s}-\mathsf{d}+1}{\theta}} + C_\theta\epsilon^{\mathsf{d}-\mathsf{s}-1}\|\mu\|_{L^1}^{1-\frac{\mathsf{s}-\mathsf{d}+1}{\theta}}\|\mu\|_{L^\infty}^{\frac{\mathsf{s}-\mathsf{d}+1}{\theta}}, & \mathsf{s} \ge \mathsf{d}-1, \end{cases}
\end{aligned} \tag{148}$$

for $\vartheta \in \left[|\mathsf{s}|-r, |\mathsf{s}|\right)$, $0 < \vartheta' < \vartheta$, $p > \frac{\mathsf{d}}{\mathsf{d}-\mathsf{s}}$, $q > \frac{\mathsf{d}}{\mathsf{d}-\mathsf{s}-1}$, and $\theta > \mathsf{s}+1-\mathsf{d}$. The constant C depends only on d,s, while the constants $C_{\vartheta,\vartheta'}, C_p, C_q, C_\theta$ additionally depend on $(\vartheta,\vartheta'), p, q, \theta$, respectively. Comparing this inequality to the statement of the lemma, we see that the proof is complete. □

4.4. *Mean-field convergence*

Using the defective commutator estimate of Theorem 8, we now show convergence of the empirical measure for the mean-field particle dynamics (1) to a (necessarily unique) solution of the limiting PDE (3) in the modulated energy distance. This proves Theorem 10. In contrast to previous works (cf. [78, Theorem 1.5] and [46, Theorem 1.5]), the rate of convergence is not the optimal $O(N^{\frac{\mathsf{s}}{\mathsf{d}}-1})$ rate. This is not surprising giving we have to pay a price for relaxing the assumption that the mean-field vector field is Lipschitz.

To close a differential inequality for $\mathsf{F}_N(X_N^t,\mu^t)$, which is the key step of the theorem, we will use the following generalization of the Grönwall–Bellman inequality.

Lemma 20. *Let $a \ge 0$. Suppose that $C_1, C_2 \colon [0,T] \to [0,\infty)$ are locally integrable, and $x \colon [0,T] \to [0,\infty)$ is absolutely continuous with*

$$\dot{x}^t \le C_1^t (x^t)^a + C_2^t x^t, \quad \text{a.e. on } [0,T]. \tag{149}$$

If $a > 1$, set $I^t := (a-1)\int_0^t C_1^s e^{(a-1)\int_0^s C_2^\tau \mathrm{d}\tau}\mathrm{d}s$. If T_* is defined as the maximal time such that $I^{T_*} \le (x^0)^{1-a}$, then

$$\forall\ t < T_*, \quad x^t \le e^{\int_0^t C_2^s \mathrm{d}s}\big((x^0)^{1-a} - I^t\big)^{-\frac{1}{a-1}}. \tag{150}$$

If $a \le 1$, then

$$\forall\ t \le T, \quad (x^t)^{1-a} \le (x^0)^{1-a} e^{(1-a)\int_0^t C_2^\tau \mathrm{d}\tau} + (1-a)\int_0^t C_1^s e^{(1-a)\int_s^t C_2^\tau \mathrm{d}\tau}\mathrm{d}s. \tag{151}$$

Proof. For the assertion (150), set $z^t := e^{-\int_0^t C_2^s \mathrm{d}s} x^t$, which satisfies

$$\dot{z}^t \le e^{-\int_0^t C_2^s \mathrm{d}s} C_1^t (x^t)^a = C_1^t e^{(a-1)\int_0^t C_2^s \mathrm{d}s}(z^t)^a. \tag{152}$$

Now let $w^t := (z^t)^{1-a}$, which is well-defined by our assumption $x^t \ge 0$. Since $1-a<0$,

$$\dot{w}^t = (1-a)(z^t)^{-a}\dot{z}^t \ge (1-a)C_1^t e^{(a-1)\int_0^t C_2^s \mathrm{d}s}. \tag{153}$$

Integrating both sides from 0 to t and then rearranging completes the proof.

For the assertion (151), set $y^t := (z^t)^{1-a}$, which satisfies the differential inequality

$$\dot{y}^t \le (1-a)C_1^t + (1-a)C_2^t y^t. \tag{154}$$

Multiplying both sides of the preceding inequality by $e^{-(1-a)\int_0^t C_2^s \mathrm{d}s}$, rearranging, then integrating yields the desired conclusion. □

With this lemma in hand, we turn to the proof of Theorem 10.

Proof of Theorem 10. The second assertion concerning convergence in H^{-r} for $r > \frac{\mathsf{d}}{2}$ follows immediately from the first assertion and Proposition 16. Therefore, we concentrate on the modulated energy.

Recall from [85, Lemma 2.1] or [75, Lemma 3.6] that $\mathsf{F}_N(X_N^t,\mu^t)$ satisfies the differential inequality

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{F}_N(X_N^t,\mu^t) \le \int_{(\mathbb{R}^{\mathsf{d}})^2\setminus\triangle} \nabla\mathsf{g}(x-y)\cdot\big(u^t(x)-u^t(y)\big)\,\mathrm{d}\Big(\frac{1}{N}\sum_{i=1}^N \delta_{x_i^t} - \mu^t\Big)^{\otimes 2}, \tag{155}$$

where $u^t := \mathsf{V}^t - \mathbb{M}\nabla\mathsf{g} * \mu^t$. To make the presentation more digestible, let us break the argument into cases based on whether s is (super-)Coulomb, sub-Coulomb, or nonsingular.

Case 1: $\mathsf{s} \ge \mathsf{d}-2$. Note that $\frac{\mathsf{d}}{p}+\mathsf{s}+2-\mathsf{d} \ge \frac{\mathsf{d}}{p}$, with strict inequality if $\mathsf{s} > \mathsf{d}-2$, and so by Sobolev embedding,

$$\|\mu^t\|_{L^q} \le C\|\mu^t\|_{\dot{W}^{\frac{\mathsf{d}}{p}+\mathsf{s}+2-\mathsf{d},p}}^{\frac{\mathsf{d}(q-1)}{q(\mathsf{s}+2)}}, \quad (q,\mathsf{s}) \ne (\infty, \mathsf{d}-2), \tag{156}$$

$$\|u^t\|_{\dot{W}^{\frac{\mathsf{d}}{p}+1,p}} \le \|\mathsf{V}^t\|_{\dot{W}^{\frac{\mathsf{d}}{p}+1,p}} + C|\mathbb{M}|\|\mu^t\|_{\dot{W}^{\frac{\mathsf{d}}{p}+\mathsf{s}+2-\mathsf{d},p}}, \tag{157}$$

$$\|\mu^t\|_{\dot{C}^{\mathsf{s}+2-\mathsf{d}}} \le C\|\mu^t\|_{\dot{W}^{\frac{\mathsf{d}}{p}+\mathsf{s}+2-\mathsf{d},p}}. \tag{158}$$

For a.e. t, we apply Theorem 8 with $v = u^t$, $q < \infty$ if $\mathsf{s} = \mathsf{d}-2$ and $q = \infty$ if $\mathsf{s} \ge \mathsf{d}-2$, and $\theta = \mathsf{s}+2-\mathsf{d}$ if $\mathsf{s} \ge \mathsf{d}-1$, to obtain

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{F}_N(X_N^t, \mu^t) \\
&\qquad \le C_p\|u^t\|_{\dot{W}^{\frac{\mathsf{d}}{p}+1,p}}\big(1+|\log\epsilon_t|\big)^{1-\frac{1}{p}}\Big(\mathsf{F}_N(X_N^t,\mu^t) + C_q\|\mu^t\|_{L^q}\lambda_t^{\frac{\mathsf{d}(q-1)}{q}-\mathsf{s}}\Big) \\
&\qquad\quad + CN^{\frac{2(\mathsf{s}+1)}{\mathsf{s}}-1}\epsilon_t\|u^t\|_{\dot{C}^1}\Big(\mathsf{F}_N(X_N^t,\mu^t) + C_q\|\mu^t\|_{L^q}\lambda_t^{\frac{\mathsf{d}(q-1)}{q}-\mathsf{s}}\Big)^{\frac{\mathsf{s}+1}{\mathsf{s}}} \\
&\qquad\quad + \epsilon_t\|u^t\|_{\dot{C}^1}\Big(C_q\|\mu^t\|_{L^q}^{\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}}\mathbf{1}_{\mathsf{s}<\mathsf{d}-1} + C_\theta\Big(\|\mu^t\|_{\dot{C}^\theta}^{\frac{\mathsf{s}-\mathsf{d}+1}{\theta}} + \epsilon_t^{\mathsf{d}-\mathsf{s}-1}\|\mu^t\|_{L^\infty}^{\frac{\mathsf{s}-\mathsf{d}+1}{\theta}}\Big)\mathbf{1}_{\mathsf{s}\ge\mathsf{d}-1}\Big),
\end{aligned} \tag{159}$$

where ϵ_t is to be chosen momentarily and λ_t corresponds to (10) with $\mu = \mu^t$. Letting

$$\begin{aligned}
\mathscr{E}^t := \mathsf{F}_N(X_N^t,\mu^t) + \sup_{t\le\tau\le T}\epsilon_\tau\Big(&C_q\|\mu^\tau\|_{L^q}\lambda_\tau^{\frac{\mathsf{d}(q-1)}{q}-\mathsf{s}} + C_q\|\mu^\tau\|_{L^q}^{\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}}\mathbf{1}_{\mathsf{s}<\mathsf{d}-1} \\
&+ C_\theta\Big(\|\mu^\tau\|_{\dot{C}^\theta}^{\frac{\mathsf{s}-\mathsf{d}+1}{\theta}} + \epsilon_\tau^{\mathsf{d}-\mathsf{s}-1}\|\mu^\tau\|_{L^\infty}^{\frac{\mathsf{s}-\mathsf{d}+1}{\theta}}\Big)\mathbf{1}_{\mathsf{s}\ge\mathsf{d}-1}\Big),
\end{aligned}$$

it follows from the inequality (159) that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathscr{E}^t \le C_1^t(\mathscr{E}^t)^{\frac{\mathsf{s}+1}{\mathsf{s}}} + C_2^t\mathscr{E}^t, \tag{160}$$

where

$$C_1^t := CN^{\frac{2(\mathsf{s}+1)}{\mathsf{s}}-1}\epsilon_t\|u^t\|_{\dot{C}^1}, \tag{161}$$

$$C_2^t := C_p\|u^t\|_{\dot{W}^{\frac{\mathsf{d}}{p}+1,p}}\big(1+|\log\epsilon_t|\big)^{1-\frac{1}{p}}. \tag{162}$$

Appealing to (150) with $a = \frac{\mathsf{s}+1}{\mathsf{s}}$, we obtain

$$\begin{aligned}
\forall\, 0 \le t < T_*, \quad \mathscr{E}^t \le \exp&\Big(\int_0^t C_p\|u^{t'}\|_{\dot{W}^{\frac{\mathsf{d}}{p}+1,p}}\big(1+|\log\epsilon_{t'}|\big)^{1-\frac{1}{p}}\mathrm{d}t'\Big) \\
&\times\Big((\mathscr{E}^0)^{-\frac{1}{\mathsf{s}}} - \frac{1}{\mathsf{s}}\int_0^t N^{\frac{2(\mathsf{s}+1)}{\mathsf{s}}-1}\epsilon_{t'}\|u^{t'}\|_{\dot{C}^1}e^{\frac{1}{\mathsf{s}}\int_0^{t'}C_p\|u^{t''}\|_{\dot{W}^{1+\frac{1}{p},p}}(1+|\log\epsilon_{t''}|)^{1-\frac{1}{p}}\mathrm{d}t''}\mathrm{d}t'\Big)^{-\mathsf{s}}.
\end{aligned} \tag{163}$$

We have quite a bit of freedom to choose ϵ_t to obtain a satisfactory estimate. For example, choose $\epsilon = \epsilon_t$ (i.e. time-independent) to satisfy the implicit equation $\epsilon = \delta N^{1-\frac{2(\mathsf{s}+1)}{\mathsf{s}}-\alpha}(\mathscr{E}^0)^{-\frac{1}{\mathsf{s}}}$ for $\delta, \alpha > 0$ fixed. This is possible to do because unpacking the definition of $\mathscr{E}^0$, the implicit equation for ϵ may be rewritten as

$$\begin{aligned}
\epsilon^{\mathsf{s}}\Big(\mathsf{F}_N(X_N^0,\mu^0) + \epsilon\sup_{0\le\tau\le T}\Big(&C_q\|\mu^\tau\|_{L^q}\lambda_\tau^{\frac{\mathsf{d}(q-1)}{q}-\mathsf{s}} + C_q\|\mu^\tau\|_{L^q}^{\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}}\mathbf{1}_{\mathsf{s}<\mathsf{d}-1} \\
&+ C_\theta\Big(\|\mu^\tau\|_{\dot{C}^\theta}^{\frac{\mathsf{s}-\mathsf{d}+1}{\theta}} + \epsilon^{\mathsf{d}-\mathsf{s}-1}\|\mu^\tau\|_{L^\infty}^{\frac{\mathsf{s}-\mathsf{d}+1}{\theta}}\Big)\mathbf{1}_{\mathsf{s}\ge\mathsf{d}-1}\Big)\Big) = \big(\delta N^{1-\frac{2(\mathsf{s}+1)}{\mathsf{s}}-\alpha}\big)^{\mathsf{s}}.
\end{aligned} \tag{164}$$

The left-hand side, which is a continuous function of ϵ, tends to zero as $\epsilon \to 0$ and to ∞ as $\epsilon \to \infty$. So, a solution exists by the intermediate value theorem.

With this choice for ϵ_t,

$$\frac{1}{\mathsf{s}}\int_0^t N^{\frac{2(\mathsf{s}+1)}{\mathsf{s}}-1}\epsilon\|u^{t'}\|_{\dot{\mathcal{C}}^1}\exp\Big(\frac{1}{\mathsf{s}}\int_0^{t'} C_p\|u^{t''}\|_{\dot{W}^{1+\frac{1}{p},p}}\big(1+|\log\epsilon|\big)^{1-\frac{1}{p}}\,dt''\Big)dt'$$
$$\leq \frac{\delta N^{-\alpha}(\mathscr{E}^0)^{-\frac{1}{\mathsf{s}}}}{\mathsf{s}}\int_0^T\|u^{t'}\|_{\dot{\mathcal{C}}^1}\exp\Big(\frac{1}{\mathsf{s}}\int_0^T C_p\|u^{t''}\|_{\dot{W}^{1+\frac{1}{p},p}}\Big(1+\big|\log(N^{1-\frac{2(\mathsf{s}+1)}{\mathsf{s}}-\alpha}\delta)\big|+\Big|\frac{1}{\mathsf{s}}\log\mathscr{E}^0\Big|\Big)^{1-\frac{1}{p}}\Big).$$

As $1-\frac{1}{p}<1$, and for any constant $C>0$, $re^{C(1+|\log r|)^{1-\frac{1}{p}}}\to 0$ as $r\to 0^+$, we see that we can choose $\delta>0$ sufficiently small depending only on $\mathsf{d},\mathsf{s},p,\alpha,\int_0^T\|u^{t''}\|_{\dot{W}^{1+\frac{\mathsf{d}}{p},p}}\,dt'',\big|\mathsf{F}_N(X_N^0,\mu^0)\big|$ such that for every $0\leq t\leq T$,

$$(\mathscr{E}^0)^{-\frac{1}{\mathsf{s}}}-\frac{1}{\mathsf{s}}\int_0^t N^{\frac{2(\mathsf{s}+1)}{\mathsf{s}}-1}\epsilon\|u^{t'}\|_{\dot{\mathcal{C}}^1}e^{\frac{1}{\mathsf{s}}\int_0^{t'}C_p\|u^{t''}\|_{\dot{W}^{1+\frac{1}{p},p}}(1+|\log\epsilon|)^{1-\frac{1}{p}}\,dt''}\,dt'\geq\frac{1}{2}(\mathscr{E}^0)^{-\frac{1}{\mathsf{s}}},\tag{165}$$

which allows us to upgrade the bound (163) to

$$\forall\, 0\leq t\leq T,\quad \mathscr{E}^t\leq 2^{\mathsf{s}}\mathscr{E}^0\exp\Big(\int_0^t C_p\|u^{t'}\|_{\dot{W}^{\frac{\mathsf{d}}{p}+1,p}}\Big(1+\big|\log(\delta N^{1-\frac{2(\mathsf{s}+1)}{\mathsf{s}}-\alpha}(\mathscr{E}^0)^{-\frac{1}{\mathsf{s}}})\big|\Big)^{1-\frac{1}{p}}\,dt'\Big).\tag{166}$$

Evidently, the preceding right-hand side vanishes as $N\to\infty$ provided that for any $C>0$, $\big|\mathsf{F}_N(X_N^0,\mu^0)\big|e^{C|\log N|^{1-\frac{1}{p}}}\to 0$ as $N\to\infty$ (e.g. $\mathsf{F}(X_N^0,\mu^0)=O(N^{-r})$ for some $r>0$). This completes the argument in the (super-)Coulomb case.

Case 2: $0<\mathsf{s}<\mathsf{d}-2$**.** We divide further into two cases based on the value of p in the assumption that $\mu^t\in\dot{W}^{\frac{\mathsf{d}}{p}+\mathsf{s}+2-\mathsf{d},p}$.

If $p<\frac{2\mathsf{d}}{\mathsf{d}-2}$, then we may choose an exponent r satisfying $\min(2,p)\leq r<\frac{2\mathsf{d}}{\mathsf{d}-2}$ so that letting $\mathsf{a}=2\big(\frac{\mathsf{d}}{r}+1\big)$, we have $\mathsf{a}\in(\mathsf{d},\mathsf{d}+2)$ and

$$\begin{aligned}\big\||\nabla|^{\frac{\mathsf{a}}{2}}u^t\big\|_{L^{\frac{2\mathsf{d}}{\mathsf{a}-2}}}&\leq\big\||\nabla|^{\frac{\mathsf{a}}{2}}\mathsf{V}^t\big\|_{L^{\frac{2\mathsf{d}}{\mathsf{a}-2}}}+C|\mathbb{M}|\big\||\nabla|^{\frac{\mathsf{a}}{2}+\mathsf{s}+1-\mathsf{d}}\mu^t\big\|_{L^{\frac{2\mathsf{d}}{\mathsf{a}-2}}}\\&=\big\||\nabla|^{\frac{\mathsf{a}}{2}}\mathsf{V}^t\big\|_{L^{\frac{2\mathsf{d}}{\mathsf{a}-2}}}+C|\mathbb{M}|\big\||\nabla|^{\frac{\mathsf{d}}{r}+\mathsf{s}+2-\mathsf{d}}\mu^t\big\|_{L^r}\\&\leq\big\||\nabla|^{\frac{\mathsf{a}}{2}}\mathsf{V}^t\big\|_{L^{\frac{2\mathsf{d}}{\mathsf{a}-2}}}+C'|\mathbb{M}|\big\||\nabla|^{\frac{\mathsf{d}}{p}+\mathsf{s}+2-\mathsf{d}}\mu^t\big\|_{L^p},\end{aligned}\tag{167}$$

where the final inequality is by Sobolev embedding. Recycling notation, let

$$\mathscr{E}^t:=\mathsf{F}_N(X_N^t,\mu^t)+\sup_{t\leq\tau\leq T}\epsilon_\tau\Big(C_q\|\mu^\tau\|_{L^q}\lambda_\tau^{\frac{\mathsf{d}(q-1)}{q}-\mathsf{s}}+C_q\|\mu^\tau\|_{L^q}^{\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}}\Big).\tag{168}$$

Applying Theorem 8, we have that (160) holds with C_1^t,C_2^t as in (161), (162), respectively. Repeating the argument from the $\mathsf{d}-2\leq\mathsf{s}<\mathsf{d}$ case then completes the proof.

If $p\geq\frac{2\mathsf{d}}{\mathsf{d}-2}$, then we cannot necessarily control the term $\big\||\nabla|^{\frac{\mathsf{a}}{2}}u^t\big\|_{L^{\frac{2\mathsf{d}}{\mathsf{a}-2}}}$, for $\mathsf{a}\in(\mathsf{d},\mathsf{d}+2)$, by $\|\mu^t\|_{\dot{W}^{\frac{\mathsf{d}}{p}+\mathsf{s}+2-\mathsf{d},p}}$ and instead have to use the suboptimal estimate only requiring $\big\||\nabla|^{\frac{\mathsf{d}-\mathsf{s}}{2}}u^t\big\|_{L^{\frac{2\mathsf{d}}{\mathsf{d}-\mathsf{s}-2}}}$. Indeed, if $r\leq\frac{2\mathsf{d}}{\mathsf{d}-\mathsf{s}-2}$, then by Sobolev embedding,

$$\begin{aligned}\big\||\nabla|^{\frac{\mathsf{d}-\mathsf{s}}{2}}u^t\big\|_{L^{\frac{2\mathsf{d}}{\mathsf{d}-\mathsf{s}-2}}}&\leq\big\||\nabla|^{\frac{\mathsf{d}-\mathsf{s}}{2}}\mathsf{V}^t\big\|_{L^{\frac{2\mathsf{d}}{\mathsf{d}-\mathsf{s}-2}}}+C\big\||\nabla|^{\frac{\mathsf{d}}{r}+1}\mathbb{M}\nabla\mathsf{g}*\mu^t\big\|_{L^r}\\&\leq\big\||\nabla|^{\frac{\mathsf{d}-\mathsf{s}}{2}}\mathsf{V}^t\big\|_{L^{\frac{2\mathsf{d}}{\mathsf{d}-\mathsf{s}-2}}}+C'|\mathbb{M}|\|\mu^t\|_{\dot{W}^{\frac{\mathsf{d}}{r}+\mathsf{s}+2-\mathsf{d},r}}\\&\leq\big\||\nabla|^{\frac{\mathsf{d}-\mathsf{s}}{2}}\mathsf{V}^t\big\|_{L^{\frac{2\mathsf{d}}{\mathsf{d}-\mathsf{s}-2}}}+C'|\mathbb{M}|\|\mu^t\|_{L^{\frac{\mathsf{d}}{\mathsf{d}-\mathsf{s}-2}}}.\end{aligned}\tag{169}$$

Since $\frac{2\mathsf{d}}{\mathsf{d}-2}\leq p\leq\frac{2\mathsf{d}}{\mathsf{d}-\mathsf{s}-2}$, we are done. Redefining

$$\mathscr{E}^t:=\mathsf{F}_N(X_N^t,\mu^t)+\sup_{t\leq\tau\leq T}\epsilon_\tau C_q\Big(\|\mu^\tau\|_{L^q}\kappa_\tau^{\frac{\mathsf{d}(q-1)}{q}-\mathsf{s}}+\|\mu^\tau\|_{L^q}^{\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}}\Big),\tag{170}$$

the inequality (160) holds with C_1^t, C_2^t as before, and repeating the same reasoning completes the proof.

Case 3: $-2 < \mathsf{s} < 0$**.** We divide into two cases based on the value of s.

If $-2 < \mathsf{s} \le -1$, then we redefine

$$\mathscr{E}^t := \mathsf{F}_N(X_N^t, \mu^t) + \sup_{t \le \tau \le T} C\epsilon_\tau \int_{\mathbb{R}^\mathsf{d}} |x|^{|\mathsf{s}|-1}\, \mathrm{d}\mu^\tau, \tag{171}$$

and we see from Theorem 8 that $\mathscr{E}^t$ obeys the differential inequality,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathscr{E}^t \le C_1^t (\mathscr{E}^t)^{\frac{|\mathsf{s}|-1}{|\mathsf{s}|}} + C_2^t \mathscr{E}^t, \tag{172}$$

where C_2^t is as above, but now $C_1^t := C\|u^t\|_{\dot{C}^1}\epsilon_t$. Applying (151) with $a = \frac{|\mathsf{s}|-1}{|\mathsf{s}|}$, we obtain

$$\forall\ t \le T, \quad (\mathscr{E}^t)^{\frac{1}{|\mathsf{s}|}} \le (\mathscr{E}^0)^{\frac{1}{|\mathsf{s}|}} \exp\left(\frac{1}{|\mathsf{s}|}\int_0^t C_2^\tau\, \mathrm{d}\tau\right) + \frac{1}{|\mathsf{s}|}\int_0^t C_1^{t'} \exp\left(\frac{1}{|\mathsf{s}|}\int_{t'}^t C_2^\tau\, \mathrm{d}\tau\right) \mathrm{d}t'. \tag{173}$$

For any $\alpha > 0$, we can choose $\epsilon_\tau = N^{-\alpha}$, and it follows that the preceding right-hand side tends to zero as $N \to \infty$ provided that for some $A > 0$, $\left|\mathsf{F}_N(X_N^0, \mu^0)\right|^{\frac{1}{|\mathsf{s}|}} e^{A(1+\log N)^{1-\frac{1}{p}}} \to 0$ as $N \to \infty$.

If $-1 < \mathsf{s} < 0$, then we instead let

$$\mathscr{E}^t := \mathsf{F}_N(X_N^t, \mu^t) + \sup_{t \le \tau \le T}\left(C_{\vartheta,\vartheta'}\epsilon_\tau^{\vartheta'} \int_{\mathbb{R}^\mathsf{d}} |x|^{|\mathsf{s}|-\vartheta}\, \mathrm{d}\mu^\tau + C\epsilon_\tau\left(1 + C_q \|\mu^\tau\|_{L^q}^{\frac{(\mathsf{s}+1)q}{\mathsf{d}(q-1)}}\right)\right), \tag{174}$$

and we see from Theorem 8 that $\mathscr{E}^t$ obeys the differential inequality,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathscr{E}^t \le C_1^t (\mathscr{E}^t)^{\frac{|\mathsf{s}|-\vartheta}{|\mathsf{s}|}} + C_2^t \mathscr{E}^t, \tag{175}$$

where C_2^t is as above, but now $C_1^t := C_\vartheta \|u^t\|_{\dot{C}^1}\epsilon_t^\vartheta$. Applying (151) with $a = \frac{|\mathsf{s}|-\vartheta}{|\mathsf{s}|}$, we find that

$$\forall\ t \le T, \quad (\mathscr{E}^t)^{\frac{\vartheta}{|\mathsf{s}|}} \le (\mathscr{E}^0)^{\frac{\vartheta}{|\mathsf{s}|}} \exp\left(\frac{\vartheta}{|\mathsf{s}|}\int_0^t C_2^\tau\, \mathrm{d}\tau\right) + \frac{\vartheta}{|\mathsf{s}|}\int_0^t C_1^{t'} \exp\left(\frac{\vartheta}{|\mathsf{s}|}\int_{t'}^t C_2^\tau\, \mathrm{d}\tau\right) \mathrm{d}t'. \tag{176}$$

Choosing $\epsilon_\tau = N^{-\alpha}$ then concludes the argument. With this last case, the proof of the theorem is complete. □

5. Remaining questions and open problems

We conclude the paper by discussing a few questions that have been so far unaddressed and posing several open problems concerning transport-commutator functional inequalities, the regularity of the transport, and the regularity of the mean-field density. We hope that this discussion will inspire future research on the subject.

5.1. *Optimal sub-Coulomb transport regularity*

As we noted in Section 1, there is an unfortunate gap between the two sub-Coulomb commutator estimates (13), (14) in Theorem 1. In the latter, the rate in N is suboptimal, but the transport regularity $\left\||\nabla|^{\frac{\mathsf{d}-\mathsf{s}}{2}} v\right\|_{L^{\frac{2\mathsf{d}}{\mathsf{d}-\mathsf{s}-2}}}$ is expected to be optimal in light of Corollary 6(i). While in the former, the rate in N is optimal, but the transport regularity $\left\||\nabla|^{\frac{\mathsf{a}}{2}} v\right\|_{L^{\frac{2\mathsf{d}}{\mathsf{a}-2}}} \mathbf{1}_{\mathsf{a}>2}$ for $\mathsf{a} \in (\mathsf{d}, \mathsf{d}+2)$, although scaling the same way as $\left\||\nabla|^{\frac{\mathsf{d}-\mathsf{s}}{2}} v\right\|_{L^{\frac{2\mathsf{d}}{\mathsf{d}-\mathsf{s}-2}}}$, is stronger by Sobolev embedding and we believe suboptimal.

The first problem we pose is to show a commutator estimate which has both the sharp error $N^{\frac{\mathsf{s}}{\mathsf{d}}-1}$ and the believed sharp transport regularity $\left\||\nabla|^{\frac{\mathsf{d}-\mathsf{s}}{2}} v\right\|_{L^{\frac{2\mathsf{d}}{\mathsf{d}-\mathsf{s}-2}}}$ when $0 \le \mathsf{s} < \mathsf{d}-2$.

The dependence $\||\nabla|^{\frac{a}{2}}v\|_{L^{\frac{2d}{a-2}}}$, for any $a \in (d, d+2)$, is an artifact of our passing through an intermediate commutator estimate for Bessel potentials in the proof of Theorem 8. We expect the latter to be the optimal regularity dependence even with the sharp $N^{\frac{s}{d}-1}$ error, but it does not seem obtainable with our current proof. The solution to this problem likely requires devising a different renormalization argument compared to [46,62], given this part of the proof is the source of the issue. As the latter cited work is the only one that gives sharp errors, let us explain more why its approach does not seem amenable to obtaining the optimal transport regularity.

The starting point for the argument of [46] is a "wavelet-type" representation of the Riesz potential g as an average of approximate identities, which may be truncated at small scale $\eta \geq 0$ to obtain

$$\forall\, x \neq 0, \quad g_\eta(x) := c_{\phi,d,s}\int_\eta^\infty t^{d-s}\phi_t(x)\frac{dt}{t}, \tag{177}$$

where ϕ is any sufficiently nice radial function with nonnegative Fourier transform, $\phi_t(x) := t^{-d}\phi(x/t)$, and $c_{\phi,d,s}$ depends on ϕ, d, s. Given suitable assumptions on ϕ (see [46, Lemma 2.2]), g_η enjoys a number of nice properties. For instance, g_η is bounded for $\eta > 0$, $g_\eta \leq g$, g_η is positive definite, and $g - g_\eta$ decays rapidly at scale η outside the ball $B(0,\eta)$. In the log case $s = 0$, the integral has to be renormalized at large scales to obtain a convergent expression.

The main technical result [46, Corollary 3.2] is a commutator estimate for the truncated potential,

$$\left|\int_{(\mathbb{R}^d)^2}\big(v(x)-v(y)\big)\cdot\nabla g_\eta(x-y)f(x)f(y)\right| \leq C(v)\int_{(\mathbb{R}^d)^2} g_\eta(x-y)f(x)f(y). \tag{178}$$

This is achieved by using the representation (177) and scaling invariance to reduce the problem to proving the estimate with g_η replaced by ϕ. Taking ϕ to be the Bessel potential $\widehat{\phi}(\xi) = \langle 2\pi\xi\rangle^{-a}$ with $a \in (d, d+2)$, this then holds as a consequence of Kato–Ponce-type commutator estimates for the inhomogeneous Fourier multiplier $\langle\nabla\rangle^a$ (the availability of such estimates motivates our choice of ϕ), but with a norm on v of the form $\||\nabla|^{a/2}v\|_{L^{\frac{2d}{a-2}}}$. The upper bound on a is not important to us, as we would like to take $a = d - s$. The lower bound is natural given $a > d$ is necessary for ϕ to be bounded, which in turn is needed for the truncated potential g_η to be bounded.

Theorem 5 suggests that approaches based on an intermediate kernel ϕ with fast Fourier decay necessarily result in an estimate with a suboptimal regularity requirement for v. The question thus remains whether (178) can be established directly, i.e. without relying on an intermediate commutator estimate for the potential ϕ.

5.2. *Localizability*

Another important question is the *localizability* of commutator estimates. To illustrate what we mean, consider the Coulomb case $s = d-2$. Integrating by parts, one has the *electric reformulation* of the energy

$$\int_{(\mathbb{R}^d)^2} g(x-y)f(x)f(y) = \frac{1}{c_{d,s}}\int_{\mathbb{R}^d}|\nabla h^f|^2, \tag{179}$$

where $h^f := g * f$. Evidently, this latter quantity may be localized to any domain $\Omega \subset \mathbb{R}^d$. Exploiting the stress-energy tensor structure the commutator possesses when $s = d-2$, one has the localized commutator estimate

$$\left|\int_{(\mathbb{R}^d)^2}\big(v(x)-v(y)\big)\cdot\nabla g(x-y)f(x)f(y)\right| \leq C\|\nabla v\|_{L^\infty}\int_{\operatorname{supp} v}|\nabla h^f|^2. \tag{180}$$

This argument can also be extended to $s > d-2$, as well as to higher-order commutator estimates and their renormalized variants [55,78,86]. Such localized estimates are important for applications to CLTs for the fluctuations of Riesz gases at mesoscales (e.g. see [87, Part 3]).

The case of a local version of the (renormalized) commutator estimate is open for the sub-Coulomb range, where the commutator no longer has the same stress-energy tensor structure as is crucial to the argument in [78]. Moreover, the proofs in [46,62] are not suitable for localized estimates given their reliance on intermediary nonlocal commutator estimates.

In the (sub-Coulomb) "local" case $\mathsf{s} = \mathsf{d} - 2k$, for integer $k \geq 2$, where g is the fundamental solution of $(-\Delta)^k$, one can still reformulate the energy,

$$\int_{(\mathbb{R}^{\mathsf{d}})^2} \mathsf{g}(x-y) f(x) f(y) = \frac{1}{\mathsf{c}_{\mathsf{d},\mathsf{s}}} \int_{\mathbb{R}^{\mathsf{d}}} |\nabla^{\otimes k} h^f|^2, \tag{181}$$

where the electric field ∇h^f is replaced by the k-tensor field $\nabla^{\otimes k} h^f$, and evidently, the latter quantity in (181) may be localized as before. Preliminary calculations indicate that there is a "higher-order" stress-energy tensor structure to the commutator in this case, which may be exploited through integration by parts to prove an unrenormalized commutator estimate. That said, the optimal regularity dependence for v remains to be determined. In the nonlocal case $\mathsf{s} \neq \mathsf{d} - 2k$, this reasoning can likely be adapted using the generalization of the Caffarelli–Silvestre extension for higher powers of the fractional Laplacian [91], analogous to the relation between the Coulomb case $\mathsf{d}-2$ and the super-Coulomb case $\mathsf{s} > \mathsf{d} - 2$.

What is less straightforward and the crux of the problem is finding a suitable renormalization scheme with which to combine the unrenormalized localized commutator estimate to obtain a localized version of the estimate (13) or (14). The two renormalization schemes available for the sub-Coulomb case, the averaging scheme in [62] and the potential truncation scheme in [46], do not seem suited to this task. The former will not produce sharp errors in N, while the latter cannot obviously be combined with the localized electric formulation described above.

5.3. *Higher-order estimates*

So far, we have only considered first-order commutator estimates corresponding to the first variation (6) of the modulated energy. Also of interest are the higher-order variations corresponding to $n \geq 2$ in

$$\begin{aligned} &\frac{\mathrm{d}^n}{\mathrm{d}t^n}\Big|_{t=0} \mathsf{F}_N\big((\mathbb{I}+tv)^{\oplus N}(X_N), (\mathbb{I}+tv)\#\mu\big) \\ &\qquad = \frac{1}{2}\int_{(\mathbb{R}^{\mathsf{d}})^2\setminus\triangle} \nabla^{\otimes n}\mathsf{g}(x-y) : \big(v(x)-v(y)\big)^{\otimes n}\, \mathrm{d}(\mu_N-\mu)^{\otimes 2} \\ &\qquad =: \mathsf{A}_n[X_N,\mu,v], \end{aligned} \tag{182}$$

where $\mathbb{I}+tv, (\mathbb{I}+tv)^{\oplus N}$ are as in Section 1.2 and $:$ denotes the inner product between tensors. As in the first-order case, one seeks to control these quantities by $C_v(\mathsf{F}_N(X_N,\mu) + C_\mu N^{-\alpha})$, and we call such control a *higher-order commutator estimate.*

Second-order (i.e. $n = 2$) estimates are important for the aforementioned transport method for studying the fluctuations of Coulomb/Riesz gases, as well as for deriving mean-field limits with multiplicative noise [67]. Estimates beyond second order are also useful, allowing to obtain finer estimates on the fluctuations of Riesz gases to treat a broader class of interactions [63] and also to compute the asymptotics of n-th order cumulants [73].

Higher-order commutator estimates were first shown at second order in [55,67] for the $\mathsf{d} = 2$ Coulomb case, then in [86] for the general Coulomb case (but with an estimate that is suboptimal in its dependence on F_N), and later in [62,80] for the full range $-2 < \mathsf{s} < \mathsf{d}$, as well as for more general Riesz-type interactions. The approach of [62] yields higher-order estimates, but these are generally not sharp in their additive error. In contrast, the estimates from [86] are sharp but only to second order for the $\mathsf{d} = 2$ Coulomb case. More importantly, their proof is quite intricate and not evidently generalizable to higher order. In [78], sharp localized estimates were shown at all orders for the (super-)Coulomb case by means of a delicate induction argument.

Although it is somewhat orthogonal to our focus on transport regularity, given its importance for studying fluctuations around the mean-field limit, we start by posing the problem of a sharp estimate for higher-order commutators in the sub-Coulomb case $\mathsf{s} < \mathsf{d} - 2$. The aforementioned work [46] by the authors proves a sharp first-order estimate, but it is not clear how to extend the method of that paper to $n \geq 2$. Desymmetrizing (182) and writing in terms of iterated commutators entails difficulties of considering commutators of commutators, not too mention the algebraic complexity as n increases. Moreover, the approach to higher orders in [62] via integrating by parts to throw derivatives onto the commutator kernel is not obviously compatible with the renormalization scheme in [46].

Even for the (super)Coulomb case, where a sharp localized estimate at any order has been shown by the last two authors in [78], the transport regularity assumptions are not fully satisfactory, and this issue is intimately linked to the regularity theory for commutators, first initiated in that work. To illustrate what we mean, we recall some facts from [78]. So as to focus on the main ideas, we limit the discussion to the Coulomb case, but we note that everything may be generalized to the super-Coulomb case.

Fixing a vector field v and given a test function f, let $\kappa^{(n),f}$ be the *n-th order commutator* defined by

$$\kappa^{(n),f}(x) := \int_{\mathbb{R}^{\mathsf{d}}} \nabla^{\otimes n}\mathsf{g}(x-y) : \big(v(x)-v(y)\big)^{\otimes n}\, \mathrm{d}f(y). \tag{183}$$

$\kappa^{(n),f}$ obeys the PDE (see [78, Section 4.1] for the calculation)

$$\begin{aligned}-\Delta\kappa^{(n),f} = \mathsf{c}_{\mathsf{d},\mathsf{s}}(-1)^n \sum_{\sigma\in\mathbb{S}_n} \partial_{i_{\sigma_1}} v^{i_1}\cdots\partial_{i_{\sigma_n}} v^{i_n} f - n\partial_i\big(\partial_i v\cdot \mathcal{v}^{(n-1),f}\big)\\ - n\partial_i v\cdot\partial_i \mathcal{v}^{(n-1),f} + n(n-1)(\partial_i v)^{\otimes 2} : \mu^{(n-2),f},\end{aligned} \tag{184}$$

where $\mathbb{S}_n$ denotes the symmetric group on $[n]$ and a repeated index denotes summation over that index. On the right-hand side, $\mathcal{v}$ is a vector field on $\mathbb{R}^{\mathsf{d}}$ that "morally" is like $\nabla\kappa^{(n),f}$, while μ (not to be confused with the mean-field density) is a symmetric matrix field on $\mathbb{R}^{\mathsf{d}}$ that morally is like $\nabla^{\otimes 2}\kappa^{(n),f}$ (see [78, (4.12), (4.14)], respectively, for the precise definitions). The right-hand side of (184) is good because it only depends on lower-order (i.e. $n-1$, $n-2$) commutators.

We believe that the right-hand side of (184) can be written in divergence form involving terms built from products of the components of ∇v and $\mathcal{v}^{(m)}$ for $m \leq n-1$. More precisely, we conjecture

$$-\Delta\kappa^{(n),f} = -n\,\mathrm{div}\big(\nabla v^i \mathcal{v}_i^{(n-1),f}\big) + \sum_{k=1}^n C_{k,n} \sum_{\sigma\in\mathbb{S}_{k+1}} (-1)^{|\sigma|}\partial_{i_{\sigma(k+1)}}\big(\partial_{i_{\sigma(1)}} v^{i_1}\cdots\partial_{i_{\sigma(k)}} v^{i_k} \mathcal{v}_{i_{k+1}}^{(n-k),f}\big), \tag{185}$$

where $C_{k,n}$ are certain combinatorial coefficients and $|\sigma|$ denotes the signature of the permutation σ. By testing (185) against $h^w := \mathsf{g} * w$, using that

$$\int_{(\mathbb{R}^{\mathsf{d}})^2} \big(v(x)-v(y)\big)\cdot\nabla\mathsf{g}(x-y) f(x) g(y) = \frac{1}{\mathsf{c}_{\mathsf{d},\mathsf{s}}} \int_{\mathbb{R}^{\mathsf{d}}} (-\Delta)\kappa^{(n),f} h^g, \tag{186}$$

and reversing the product rule, one arrives at a stress-energy tensor structure to the higher-order commutators. If such an identity (185) holds and one has a bound

$$\int_{\mathbb{R}^{\mathsf{d}}} |\mathcal{v}^{(m),f}|^2 \leq C_m\big(\|\nabla v\|_{L^\infty}\big)^{2m} \int_{\mathrm{supp}\,\nabla v} |\nabla h^f|^2, \quad m \leq n-1, \tag{187}$$

then it follows immediately from integration by parts and Cauchy–Schwarz that

$$\int_{\mathbb{R}^{\mathsf{d}}} |\nabla\kappa^{(n),f}|^2 \leq C_n\big(\|\nabla v\|_{L^\infty}\big)^{2n} \int_{\mathrm{supp}\,\nabla v} |\nabla h^f|^2. \tag{188}$$

Since the $\mathcal{v}$'s obey a recursion in terms of the $\nabla\kappa$'s (see [78, (4.13)]), the estimate (188) implies that (187) holds for $m = n$ and by induction, the estimates (188), (187) hold for any integer $n \geq 1$.

Unfortunately, we were only able to prove the identity (185) for $n \le 2$ (see [78, Section 4.3]), and the computation is already quite involved once $n = 2$. If we stick with the non-divergence form of the right-hand side of (184), then we encounter terms that have a factor of $\kappa^{(n),f}$. For example,

$$\int_{\mathbb{R}^{\mathsf{d}}} \kappa^{(n),f} \partial_i v \cdot \partial_i v^{(n-1),f} = -\int_{\mathbb{R}^{\mathsf{d}}} \partial_i \kappa^{(n),f} \partial_i v \cdot v^{(n-1),f} - \int_{\mathbb{R}^{\mathsf{d}}} \kappa^{(n),f} \partial_i^2 v \cdot v^{(n-1),f}. \tag{189}$$

One cannot simply use Cauchy–Schwarz on the second term on the right-hand side, as in general, there is no way to control $\int_{\operatorname{supp}\nabla v} |\kappa^{(n),f}|^2$ by $\int_{\operatorname{supp}\nabla v} |\nabla \kappa^{(n),f}|^2$. However, if one's primary interest is in localized estimates, then there is a way out of this issue by exploiting the localization from the start. Since $-\Delta\kappa^{(n),f}$ has zero average, we may equivalently test the equation (184) against $\kappa^{(n),f} - \overline{\kappa}^{(n),f}$, where $\overline{\kappa}^{(n),f}$ denotes the average in a ball B containing the support of v. The strategy is to integrate by parts and use Cauchy–Schwarz as before, setting up an induction argument for estimates satisfied by $\kappa^{(m),f}, v^{(m),f}, \mu^{(m),f}$ (see [78, Lemma 4.4]). Crucially, the Lebesgue measure on B satisfies a Poincaré inequality, which allows to control $\int_B |\kappa^{(n),f} - \overline{\kappa}^{(n),f}|^2$ by $\int_B |\nabla\kappa^{(n),f}|^2$.

Although this approach requires a bound for $\|\nabla^{\otimes 2} v\|_{L^\infty}$, which is a stronger demand than the Lipschitz requirement that would follow if (185) holds, this is still sufficient for applications, such as to CLTs for linear statistics of Coulomb/Riesz gases. Nevertheless, improving the regularity to $\|\nabla v\|_{L^\infty}$ would allow one to consider rougher test functions in this application, not to mention the aesthetic motivation for an estimate that is uniform in the localization.

5.4. *Scaling-critical mean-field convergence for $p = \infty$*

The last question we consider concerns Theorems 8 and 10 in the excluded case $p = \infty$. The case $p = 1$ is also excluded, but this is not particularly relevant because the critical space $\dot{W}^{\mathsf{s}+2,1}$ never corresponds to a nonincreasing/conserved quantity, unlike L^∞, which is scaling-critical in the Coulomb case.

First, the reader should recall from Remark 9 that per our notation (see Section 1.5), the limit of the space $\dot{W}^{\frac{\mathsf{d}}{p}+\mathsf{s}+2-\mathsf{d},p}$ as $p \to \infty$ is the space

$$\left\{ f \in \mathcal{S}'(\mathbb{R}^{\mathsf{d}}) : \left\| \langle\nabla\rangle^{\mathsf{s}+2-\mathsf{d}} f \right\|_{L^\infty} < \infty \right\}. \tag{190}$$

This space is not well-behaved from a harmonic analysis perspective in terms of singular integral operators/Fourier multipliers being bounded on it, and a more natural L^∞-type space is the homogeneous Hölder–Zygmund space $\dot{\mathcal{C}}^{\mathsf{s}+2-\mathsf{d}}$, which is well-behaved. In particular, we have

$$\left\| \mathbb{M}\nabla \mathsf{g} * \mu \right\|_{\dot{\mathcal{C}}^1} \lesssim |\mathbb{M}| \|\mu\|_{\dot{\mathcal{C}}^{\mathsf{s}+2-\mathsf{d}}}. \tag{191}$$

If $\mathsf{s} = \mathsf{d} - 2$ (i.e. Coulomb), then $\dot{\mathcal{C}}^0$ should be understood as the homogeneous Besov space $\dot{B}^0_{\infty,\infty}$, into which L^∞ embeds.

One can still prove a defective commutator estimate like Theorem 8 when the vector field $v \in \dot{\mathcal{C}}^1$, but the divergent prefactor $\left(1 + |\log\epsilon|\right)^{1-\frac{1}{p}}$ that comes from estimating $\|\nabla v_\epsilon\|_{L^\infty}$ is replaced by $1 + |\log\epsilon|$. In the log case $\mathsf{s} = 0$, this factor $1 + |\log\epsilon|$ is still sufficient to prove a version of Theorem 10 (see [69, Proposition 1.8]) by essentially taking $\epsilon = \epsilon_t$ to be comparable to the modulated energy $\mathsf{F}_N(X_N^t, \mu^t)$ and using that $r|\log r|$ satisfies the conditions of the Osgood lemma. Importantly, in estimating the mollification error $v - v_\epsilon$ (specifically, Term_1 in the proof of Lemma 19), the log-Lipschitz regularity of v allows one to bound for $|x_i - x_j| \le \frac{1}{2}$,

$$\left| \nabla \mathsf{g}(x_i - x_j) \cdot \left((v - v_\epsilon)(x_i) - (v - v_\epsilon)(x_j) \right) \right| \le C \|v\|_{\dot{\mathcal{C}}^1} \frac{\left|\log|x_i - x_j|\right|}{|x_i - x_j|^{\mathsf{s}}} = C \|v\|_{\dot{\mathcal{C}}^1} \mathsf{g}(x_i - x_j). \tag{192}$$

This relation is, of course, only true in the log case. When $0 < \mathsf{s} < \mathsf{d}$, one has the leftover factor $\left|\log|x_i - x_j|\right|$. In the proof of [70, Proposition 4.1], this extra log factor is handled by controlling the minimal distance between points by the N-particle energy. Taking $\epsilon \propto N^{-\alpha}$ for large enough $\alpha > 0$ then allows one to close a Grönwall relation, but only for short times depending on the size of the initial modulated energy. The reason is that one encounters the quantity $N^{-\beta} e^{Ct|\log N^{-\alpha}|}$, for some $\beta > 0$, which only vanishes as $N \to \infty$ if t is small enough (independently of N). Even replacing this part of the argument with that used to estimate Term_1 in Lemma 19 would not suffice if we only had bound $\|\nabla v_\epsilon\|_{L^\infty} \lesssim \big(1 + |\log \epsilon|\big)$. The prefactor $N^{\frac{2(\mathsf{s}+1)}{\mathsf{s}}-1}\epsilon$ on the right-hand side of (117) when $0 < \mathsf{s} < \mathsf{d}$ again requires our taking $\epsilon \lesssim N^{-\alpha}$ for large enough α, and we would run into the same short-time issue as before. Thus, what is essential is having the control $\|\nabla v_\epsilon\|_{L^\infty} \lesssim \big(1 + |\log \epsilon|\big)^\theta$ for $\theta \in (0,1)$, which is made possible by the stronger assumption that $v \in \dot{W}^{\frac{\mathsf{d}}{p}+1,p}$.

To date, mean-field convergence under the assumption that the limiting density $\mu \in L_t^\infty L_x^\infty$ has only been shown to hold on some interval of time $[0, T_*]$, despite the fact that solutions of the mean-field equation are global. Removing this short-time restriction, which we think is purely technical, is an important open problem. More generally, it would be desirable to show convergence without any short-time restriction in the super-Coulomb case when $\mu \in L_t^\infty \dot{C}_x^{\mathsf{s}+2-\mathsf{d}}$.

Acknowledgments

The second author thanks the Courant Institute of Mathematical Sciences, NYU for its hospitality, where part of the work for this project was carried out during a visit in November 2021.

Declaration of interests

The authors do not work for, advise, own shares in, or receive funds from any organization that could benefit from this article, and have declared no affiliations other than their research organizations.

References

[1] F. Altekrüger, J. Hertrich and G. Steidl, "Neural Wasserstein gradient flows for discrepancies with Riesz kernels", in *Proceedings of the 40th International Conference on Machine Learning* (A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato and J. Scarlett, eds.), Proceedings of Machine Learning Research, vol. 202, PMLR, 2023, pp. 664–690.

[2] H. Bahouri, J.-Y. Chemin and R. Danchin, *Fourier analysis and nonlinear partial differential equations*, Grundlehren der Mathematischen Wissenschaften, vol. 343, Springer, 2011.

[3] R. Bauerschmidt, P. Bourgade, M. Nikula and H.-T. Yau, "The two-dimensional Coulomb plasma: quasi-free approximation and central limit theorem", *Adv. Theor. Math. Phys.* **23** (2019), no. 4, pp. 841–1002.

[4] F. Bekerman, T. Leblé and S. Serfaty, "CLT for fluctuations of β-ensembles with general potential", *Electron. J. Probab.* **23** (2018), article no. 115 (31 pages).

[5] R. J. Berman and M. Önnheim, "Propagation of chaos for a class of first order models with singular mean field interactions", *SIAM J. Math. Anal.* **51** (2019), no. 1, pp. 159–196.

[6] S. V. Borodachov, D. P. Hardin and E. B. Saff, *Discrete energy on rectifiable sets*, Springer Monographs in Mathematics, Springer, 2019.

[7] G. Borot and A. Guionnet, "Asymptotic expansion of β matrix models in the one-cut regime", *Commun. Math. Phys.* **317** (2013), no. 2, pp. 447–483.

[8] G. Borot and A. Guionnet, "Asymptotic expansion of β matrix models in the multi-cut regime", *Forum Math. Sigma* **12** (2024), article no. e13 (93 pages).

[9] D. Bresch, P.-E. Jabin and Z. Wang, “Modulated free energy and mean field limit”, in *Séminaire Laurent Schwartz — EDP et applications. Année 2019–2020*, Institut des Hautes Études Scientifiques, 2019. Exposé no. 2 (22 pages).
[10] D. Bresch, P.-E. Jabin and Z. Wang, “On mean-field limits and quantitative estimates with a large class of singular kernels: application to the Patlak–Keller–Segel model”, *Comptes Rendus. Mathématique* **357** (2019), no. 9, pp. 708–720.
[11] D. Bresch, P.-E. Jabin and Z. Wang, “Mean field limit and quantitative estimates with singular attractive kernels”, *Duke Math. J.* **172** (2023), no. 13, pp. 2591–2641.
[12] H. Brezis and T. Gallouët, “Nonlinear Schrödinger evolution equations”, *Nonlinear Anal., Theory Methods Appl.* **4** (1980), no. 4, pp. 677–681.
[13] H. Brezis and S. Wainger, “A note on limiting cases of Sobolev embeddings and convolution inequalities”, *Commun. Partial Differ. Equations* **5** (1980), no. 7, pp. 773–789.
[14] S. Cai, X. Feng, Y. Gong and Z. Wang, “Propagation of chaos for 2D log gas on the whole space”, 2024. Online at https://arxiv.org/abs/2411.14777.
[15] A. P. Calderón, “Commutators, singular integrals on Lipschitz curves and applications”, in *Proceedings of the International Congress of Mathematicians (Helsinki, 1978)* (O. Lehto, ed.), Academia Scientiarum Fennica, 1980, pp. 85–96.
[16] J. A. Carrillo, Y.-P. Choi and M. Hauray, “The derivation of swarming models: mean-field limit and Wasserstein distances”, in *Collective dynamics from bacteria to crowds* (A. Muntean and F. Toschi, eds.), CISM International Centre for Mechanical Sciences, vol. 553, Springer, 2014, pp. 1–46.
[17] J. A. Carrillo, L. C. F. Ferreira and J. C. Precioso, “A mass-transportation approach to a one dimensional fluid mechanics model with nonlocal velocity”, *Adv. Math.* **231** (2012), no. 1, pp. 306–327.
[18] A. Cecchin and P. Nikolaev, “Convergence rate for fluctuations of mean field interacting diffusion and application to 2D viscous vortex model and Coulomb potential”, 2025. Online at https://arxiv.org/abs/2509.01266.
[19] A. Chodron de Courcel, M. Rosenzweig and S. Serfaty, “Sharp uniform-in-time mean-field convergence for singular periodic Riesz flows”, *Ann. Inst. Henri Poincaré C, Anal. Non Linéaire* **42** (2023), no. 2, pp. 391–472.
[20] A. Chodron de Courcel, M. Rosenzweig and S. Serfaty, “The attractive log gas: stability, uniqueness, and propagation of chaos”, *Commun. Am. Math. Soc.* **5** (2025), pp. 695–773.
[21] M. Christ and J.-L. Journé, “Polynomial growth estimates for multilinear singular integral operators”, *Acta Math.* **159** (1987), no. 1–2, pp. 51–80.
[22] R. R. Coifman and Y. F. Meyer, “Commutateurs d’intégrales singulières et opérateurs multilinéaires”, *Ann. Inst. Fourier* **28** (1978), no. 3, pp. 177–202.
[23] R. R. Coifman, R. Rochberg and G. Weiss, “Factorization theorems for Hardy spaces in several variables”, *Ann. Math.* **103** (1976), no. 3, pp. 611–635.
[24] T. Dauxois, S. Ruffo, E. Arimondo and M. Wilkens, “Dynamics and thermodynamics of systems with long-range interactions: an introduction”, in *Dynamics and thermodynamics of systems with long-range interactions (Les Houches, 2002)* (T. Dauxois, S. Ruffo, E. Arimondo and M. Wilkens, eds.), Lecture Notes in Physics, vol. 602, Springer, 2002, pp. 1–19.
[25] M. Duerinckx, “Mean-field limits for some Riesz interaction gradient flows”, *SIAM J. Math. Anal.* **48** (2016), no. 3, pp. 2269–2300.
[26] H. Engler, “An alternative proof of the Brezis–Wainger inequality”, *Commun. Partial Differ. Equations* **14** (1989), no. 4, pp. 541–544.
[27] X. Feng and Z. Wang, “Quantitative propagation of chaos for 2D viscous vortex model with general circulations on the whole space”, 2024. Online at https://arxiv.org/abs/2411.14266.

[28] F. Golse, "On the dynamics of large particle systems in the mean field limit", in *Macroscopic and large scale phenomena: coarse graining, mean field limits and ergodicity* (A. Muntean, J. D. M. Rademacher and A. Zagaris, eds.), Lecture Notes in Applied Mathematics and Mechanics, vol. 3, Springer, 2016, pp. 1–144.

[29] F. Golse and T. Paul, "Mean-field and classical limit for the N-body quantum dynamics with Coulomb interaction", *Commun. Pure Appl. Math.* **75** (2022), no. 6, pp. 1332–1376.

[30] L. Grafakos, *Modern Fourier analysis*, Third edition, Graduate Texts in Mathematics, vol. 250, Springer, 2014.

[31] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf and A. J. Smola, "A kernel approach to comparing distributions", in *AAAI'07: Proceedings of the 22nd National Conference on Artificial Intelligence. Volume 2* (A. Cohn, ed.), AAAI Press, 2007, pp. 1637–1641.

[32] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf and A. J. Smola, "A kernel method for the two-sample-problem", in *Advances in Neural Information Processing Systems 19* (B. Schölkopf, J. Platt and T. Hoffman, eds.), MIT Press, 2007, pp. 513–520.

[33] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf and A. J. Smola, "A kernel two-sample test", *J. Mach. Learn. Res.* **13** (2012), pp. 723–773.

[34] A. Guillin, P. Le Bris and P. Monmarché, "Uniform in time propagation of chaos for the 2D vortex model and other singular stochastic systems", *J. Eur. Math. Soc.* **27** (2024), no. 6, pp. 2359–2386.

[35] P. Hagemann, J. Hertrich, F. Altekrüger, R. Beinert, J. Chemseddine and G. Steidl, "Posterior sampling based on gradient flows of the MMD with negative distance kernel", 2023. Online at https://arxiv.org/abs/2310.03054.

[36] D. Han-Kwan and M. Iacobelli, "From Newton's second law to Euler's equations of perfect fluids", *Proc. Am. Math. Soc.* **149** (2021), no. 7, pp. 3045–3061.

[37] K. Hansson, "Imbedding theorems of Sobolev type in potential theory", *Math. Scand.* **45** (1979), no. 1, pp. 77–102.

[38] D. P. Hardin, T. Leblé, E. B. Saff and S. Serfaty, "Large deviation principles for hypersingular Riesz gases", *Constructive Approximation* **48** (2018), no. 1, pp. 61–100.

[39] D. P. Hardin, E. B. Saff, R. Shu and E. Tadmor, "Dynamics of particles on a curve with pairwise hyper-singular repulsion", *Discrete Contin. Dyn. Syst.* **41** (2021), no. 12, pp. 5509–5536.

[40] D. D. Haroske, *Envelopes and sharp embeddings of function spaces*, CRC Research Notes in Mathematics, vol. 437, Chapman & Hall/CRC, 2007.

[41] M. Hauray, "Wasserstein distances for vortices approximation of Euler-type equations", *Math. Models Methods Appl. Sci.* **19** (2009), no. 8, pp. 1357–1384.

[42] M. Hauray and S. Mischler, "On Kac's chaos and related problems", *J. Funct. Anal.* **266** (2014), no. 10, pp. 6055–6157.

[43] J. Hertrich, R. Beinert, M. Gräf and G. Steidl, "Wasserstein gradient flows of the discrepancy with distance kernel on the line", in *Scale space and variational methods in computer vision* (L. Calatroni, M. Donatelli, S. Morigi, M. Prato and M. Santacesaria, eds.), Springer, 2023, pp. 431–443.

[44] J. Hertrich, C. Wald, F. Altekrüger and P. Hagemann, "Generative sliced MMD flows with Riesz kernels", 2023. Online at https://arxiv.org/abs/2305.11463.

[45] E. Hess-Childs, "Large deviation principles for singular Riesz-type diffusive flows", 2023. Online at https://arxiv.org/abs/2312.02904.

[46] E. Hess-Childs, M. Rosenzweig and S. Serfaty, "A sharp commutator estimate for all Riesz modulated energies", 2025. Online at https://arxiv.org/abs/2511.13461.

[47] E. Hess-Childs, M. Rosenzweig and S. Serfaty, "Optimal quantization for Riesz maximum mean discrepancies". In preparation.

[48] J. Huang, M. Rosenzweig and S. Serfaty, “Fluctuations around the mean field limit for noisy singular Riesz flows”. In preparation.

[49] P.-E. Jabin and Z. Wang, “Quantitative estimates of propagation of chaos for stochastic systems with $W^{-1,\infty}$ kernels”, *Invent. Math.* **214** (2018), no. 1, pp. 523–591.

[50] S. Janson, “Generalizations of Lipschitz spaces and an application to Hardy spaces and bounded mean oscillation”, *Duke Math. J.* **47** (1980), no. 4, pp. 959–982.

[51] F. John and L. Nirenberg, “On functions of bounded mean oscillation”, *Commun. Pure Appl. Math.* **14** (1961), no. 3, pp. 415–426.

[52] C. E. Kenig, G. Ponce and L. Vega, “Well-posedness and scattering results for the generalized Korteweg–de Vries equation via the contraction principle”, *Commun. Pure Appl. Math.* **46** (1993), no. 4, pp. 527–620.

[53] S. Klainerman, “PDE as a unified subject”, in *Visions in mathematics: GAFA 2000 special volume. Part I* (N. Alon, J. Bourgain, A. Connes, M. Gromov and V. D. Milman, eds.), Birkhäuser, 2010, pp. 279–315.

[54] S. Kolouri, K. Nadjahi, S. Shahrampour and U. Simsekli, “Generalized sliced probability metrics”, in *ICASSP 2022 — 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE Press, 2022, pp. 4513–4517.

[55] T. Leblé and S. Serfaty, “Fluctuations of two dimensional Coulomb gases”, *Geom. Funct. Anal.* **28** (2018), no. 2, pp. 443–508.

[56] D. Li, “On Kato–Ponce and fractional Leibniz”, *Rev. Mat. Iberoam.* **35** (2019), no. 1, pp. 23–100.

[57] F. Lin and P. Zhang, “On the hydrodynamic limit of Ginzburg–Landau vortices”, *Discrete Contin. Dyn. Syst.* **6** (2000), no. 1, pp. 121–142.

[58] M. Ménard, “Mean-field limit derivation of a monokinetic spray model with gyroscopic effects”, *SIAM J. Math. Anal.* **56** (2024), no. 1, pp. 1068–1113.

[59] T. Modeste and C. Dombry, “Characterization of translation invariant MMD on $\mathbb{R}^d$ and connections with Wasserstein distances”, *J. Mach. Learn. Res.* **25** (2024), article no. 237 (39 pages).

[60] J. K. Moser, “A sharp form of an inequality by N. Trudinger”, *Indiana Univ. Math. J.* **20** (1970), pp. 1077–1092.

[61] A. Müller, “Integral probability metrics and their generating classes of functions”, *Adv. Appl. Probab.* **29** (1997), no. 2, pp. 429–443.

[62] Q.-H. Nguyen, M. Rosenzweig and S. Serfaty, “Mean-field limits of Riesz-type singular flows”, *Ars Inven. Anal.* **2022** (2022), article no. 4 (45 pages).

[63] L. Peilen and S. Serfaty, “Local laws and fluctuations for super-Coulombic Riesz gases”, 2025. Online at https://arxiv.org/abs/2511.18623.

[64] M. Petrache and S. Serfaty, “Next order asymptotics and renormalized energy for Riesz interactions”, *J. Inst. Math. Jussieu* **16** (2017), no. 3, pp. 501–569.

[65] I. B. Porat, “Derivation of Euler’s equations of perfect fluids from von Neumann’s equation with magnetic field”, *J. Stat. Phys.* **190** (2023), no. 7, article no. 121 (44 pages).

[66] I. B. Porat, J. A. Carrillo and P.-E. Jabin, “Singular flows with time-varying weights”, 2025. Online at https://arxiv.org/abs/2503.02276.

[67] M. Rosenzweig, “The mean-field limit of stochastic point vortex systems with multiplicative noise”, 2020. Online at https://arxiv.org/abs/2011.12180. To be published in *Commun. Pure Appl. Math.*

[68] M. Rosenzweig, “From quantum many-body systems to ideal fluids”, 2021. Online at https://arxiv.org/abs/2110.04195.

[69] M. Rosenzweig, “Mean-field convergence of point vortices to the incompressible Euler equation with vorticity in L^∞”, *Arch. Ration. Mech. Anal.* **243** (2022), no. 3, pp. 1361–1431.

[70] M. Rosenzweig, “The mean-field approximation for higher-dimensional Coulomb flows in the scaling-critical L^∞ space”, *Nonlinearity* **35** (2022), no. 6, pp. 2722–2766.
[71] M. Rosenzweig, “On the rigorous derivation of the incompressible Euler equation from Newton’s second law”, *Lett. Math. Phys.* **113** (2023), no. 1, article no. 13 (32 pages).
[72] M. Rosenzweig, “Commutators, mean-field, and supercritical mean-field limits for Coulomb/Riesz gases”, 2026. Online at https://arxiv.org/abs/2603.02169.
[73] M. Rosenzweig, “Cumulants of Riesz gases”. Online at https://tinyurl.com/rosenzweig-comm-cumulants. Unpublished note.
[74] M. Rosenzweig and S. Serfaty, “Global-in-time mean-field convergence for singular Riesz-type diffusive flows”, *Ann. Appl. Probab.* **33** (2023), no. 2, pp. 754–798.
[75] M. Rosenzweig and S. Serfaty, “Relative entropy and modulated free energy without confinement via self-similar transformation”, 2024. Online at https://arxiv.org/abs/2402.13977.
[76] M. Rosenzweig and S. Serfaty, “Modulated logarithmic Sobolev inequalities and generation of chaos”, *Ann. Fac. Sci. Toulouse, Math. (6)* **34** (2025), no. 1, pp. 107–134.
[77] M. Rosenzweig and S. Serfaty, “The lake equation as a supercritical mean-field limit”, *J. Éc. Polytech., Math.* **12** (2025), pp. 1019–1068.
[78] M. Rosenzweig and S. Serfaty, “Sharp commutator estimates of all order for Coulomb and Riesz modulated energies”, *Commun. Pure Appl. Math.* **79** (2026), no. 2, pp. 207–292.
[79] M. Rosenzweig and S. Serfaty, “Commutator estimates, Stein’s method, and the transport approach to fluctuations of Riesz gases”. In preparation.
[80] M. Rosenzweig, D. Slepčev and L. Wang, *Wasserstein gradient flow of maximum mean discrepancy with energy kernels*. In preparation.
[81] N. Rougerie and S. Serfaty, “Higher-dimensional Coulomb gases and renormalized energy functionals”, *Commun. Pure Appl. Math.* **69** (2016), no. 3, pp. 519–605.
[82] E. Sandier and S. Serfaty, “1D log gases and the renormalized energy: crystallization at vanishing temperature”, *Probab. Theory Relat. Fields* **162** (2015), no. 3–4, pp. 795–846.
[83] E. Sandier and S. Serfaty, “2D Coulomb gases and the renormalized energy”, *Ann. Probab.* **43** (2015), no. 4, pp. 2026–2083.
[84] A. Seeger, C. K. Smart and B. Street, *Multilinear singular integral forms of Christ–Journé type*, Memoirs of the American Mathematical Society, vol. 1231, American Mathematical Society, 2019.
[85] S. Serfaty, “Mean field limit for Coulomb-type flows”, *Duke Math. J.* **169** (2020), no. 15, pp. 2887–2935.
[86] S. Serfaty, “Gaussian fluctuations and free energy expansion for Coulomb gases at any temperature”, *Ann. Inst. Henri Poincaré, Probab. Stat.* **59** (2023), no. 2, pp. 1074–1142.
[87] S. Serfaty, “Lectures on Coulomb and Riesz gases”, 2024. Online at https://arxiv.org/abs/2407.21194.
[88] S. Serfaty and J. L. Vázquez, “A mean field equation as limit of nonlinear diffusions with fractional Laplacian operators”, *Calc. Var. Partial Differ. Equ.* **49** (2014), no. 3–4, pp. 1091–1120.
[89] E. M. Stein, *Singular integrals and differentiability properties of functions*, Princeton Mathematical Series, vol. 30, Princeton University Press, 1970.
[90] N. S. Trudinger, “On imbeddings into Orlicz spaces and some applications”, *J. Math. Mech.* **17** (1967), pp. 473–483.
[91] R. Yang, “On higher order extensions for the fractional Laplacian”, 2013. Online at https://arxiv.org/abs/1302.4413.
[92] V. I. Yudovich, “Non-stationary flow of an ideal incompressible liquid”, *U.S.S.R. Comput. Math. Math. Phys.* **3** (1963), no. 6, pp. 1407–1456.