

INSTANTANEOUS REGULARISATION VIA HARNACK-TYPE METHODS FOR A DEGENERATE NON-LOCAL EQUATION

SIMON M. SCHULZ

ABSTRACT. We prove an instantaneous smoothing effect for a degenerate non-local evolution equation (originally derived as the many-particle limit of a system of active Brownian particles). The underlying mechanism is that, by deriving a Harnack-type inequality for an averaged quantity, we are able to show that the equation becomes strongly parabolic for positive times, and thus self-regularising. This paper is a significant improvement of a previous work, where additional (weak) non-degeneracy assumptions were imposed on the initial data. Herein, we completely remove these assumptions, and prove—in addition to the regularity bootstrap—quantitative estimates on the aforementioned averaged quantity by means of an approach *à la* Bombieri–Giusti.

CONTENTS

1. Introduction	1
2. Definitions and Background	3
3. Main Theorem	7
4. The Bombieri–Giusti Lemma	7
5. Moser-type Local L^∞ Estimates	9
6. Logarithmic Estimates	15
7. Quantitative Instantaneous Separation from $\{\rho = 1\}$	17
8. Instantaneous Smoothing	21
Appendix A. Proofs of Technical Lemmas	39
References	40

1. INTRODUCTION

The purpose of this paper is to present a refined regularity theory for weak solutions $f(t, x, \theta)$ of the degenerate non-local evolution equation

$$\partial_t f + \operatorname{div}((1 - \rho)f\mathbf{e}(\theta)) = \operatorname{div}((1 - \rho)\nabla f + f\nabla\rho) + \partial_\theta^2 f, \quad (1.1)$$

where $\mathbf{e}(\theta) = (\cos \theta, \sin \theta)$, and the (non-local) averaged quantity ρ is defined by

$$\rho(t, x) := \int_{\mathbb{T}} f(t, x, \theta) \, d\theta, \quad (1.2)$$

where $\mathbb{T} \cong [0, 2\pi)$ is the torus. The notations div and ∇ represent the divergence and gradient operators with respect to the x variable only. The initial value problem for (1.1) is posed with periodic boundary conditions both in the space variable $x \in \mathbb{T}^2$ and the angle variable $\theta \in \mathbb{T}$. In the paper, we denote the *space-angle coordinate* by $\xi = (x, \theta) \in \mathbb{T}^3$.

Equation (1.1) was derived in [7], and corresponds to the many-particle limit of a system of active Brownian particles (*cf.* [9, 19])—these models appear naturally when studying collective behaviour in the biological and social sciences (*cf.* [22, 24]). A global existence result for (1.1) was obtained in [8], by means of a non-local *boundedness-by-entropy method*, where one rewrites (1.1) as an evolution on the first variation of a suitable non-local functional (which itself can be interpreted as a perturbation of a Wasserstein-type gradient flow). The weak solutions obtained therein are dual to test functions belonging to $L_t^6 W_{x,\theta}^{1,6}$ (thus, in principle, they are endowed with rather little regularity), and satisfy $0 \leq \rho \leq 1$ a.e.; hence (1.1) is of *degenerate parabolic* type, with degeneracy at $\rho = 1$.

2020 *Mathematics Subject Classification.* 35B65, 35K65, 35K55, 35Q92, 35A15.

Key words and phrases. Degenerate non-local equations, Harnack inequality, infinite speed of propagation, Bombieri–Giusti Lemma, active particle models.

Recently, a rather general regularity theory was established in [1]; following ideas presented for a simpler active model in [3]. Therein, the authors proved that weak solutions of (1.1) are in fact smooth for all positive times, provided certain (weak) non-degeneracy assumptions are imposed on the initial data. More precisely, provided one has suitable control on the degenerate zone $\{\rho = 1\}$ (where (1.1) loses strict parabolicity) at the initial time, then one can show that (1.1) becomes self-regularising. In this work, we completely remove these non-degeneracy assumptions on the initial data, thereby showing that—surprisingly, given the degenerate/non-local appearance of the PDE—all weak solutions (as defined in [8]) instantaneously become smooth. The present paper is therefore a significant improvement of the previous contribution [1], and (in some sense) it concludes the regularity study of equation (1.1); all weak solutions are classical solutions for positive times. The questions of uniqueness/non-uniqueness and convergence to stationary states are still largely open and will be the focus of future investigations.

While (1.1) appears to be fraught with degeneracies and non-local effects, it is actually very simple to (heuristically) explain its underlying self-regularising mechanism. Indeed, the main obstacle for regularity is that the *self-diffusive* term $\operatorname{div}((1-\rho)\nabla f)$ is degenerate when $\rho = 1$. Consequently, in the vicinity of $\{\rho = 1\}$, the smoothing effects provided by this term cannot compensate for the losses of regularity produced by the non-local drift and the *infinite cross-diffusion* term $\operatorname{div}(f\nabla\rho)$; this latter term formally comprises $f\Delta\rho$, which includes competing diffusive effects from all angles $\theta \in [0, 2\pi)$. However, upon integrating (1.1) in the angle variable θ , one observes that the equation for ρ is strongly parabolic:

$$\partial_t \rho + \operatorname{div}((1-\rho)\mathbf{p}) = \Delta \rho, \quad (1.3)$$

where the *polarisation* $\mathbf{p}$ is a non-local “moment” quantity determined from f by

$$\mathbf{p}(t, x) := \int_{\mathbb{T}} f(t, x, \theta) \mathbf{e}(\theta) \, d\theta. \quad (1.4)$$

In turn, this latter equation is endowed with an *instantaneous separation from* $\{\rho = 1\}$ property, meaning that $0 < \rho < 1$ for all positive times. A quantitative description of this *infinite speed of propagation* effect is embodied by a Harnack-type inequality for $1 - \rho$ (which we prove in §7 of this manuscript). Once this quantitative estimate is established, one deduces that (1.1) is strongly parabolic for all positive times, and the strategy of [1, §6-7] (see also [3, §4]) can be adapted to show smoothness of the weak solution. While a lower bound on $1 - \rho$ had been obtained in [1] by an ad-hoc method, the authors had not clearly identified that it could be derived from a Harnack inequality, which led them to impose the aforementioned unnecessary non-degeneracy conditions on the initial data.

Herein, we prove the desired Harnack-type inequality directly, instead of invoking an abstract result from the literature—we explain our reason for doing so in the paragraph that follows. Before jumping to this, we remind the reader that our aim, ultimately, is to obtain a *quantitative* form of the instantaneous separation of ρ from the degenerate state $\{\rho = 1\}$; cf. Theorem 1. This phenomenon is *expected* from the strong parabolicity of (1.3), but obtaining an explicit (uniform-in- x) lower bound on $1 - \rho$ near the initial time is *a priori* not obvious.

Although (1.3) may be interpreted as a linear drift-diffusion equation for ρ —and is thus endowed with a Harnack inequality—the constants appearing in its corresponding Harnack inequality depend quantitatively on the drift $\mathbf{p}$. In particular, when applying the estimate on cylinders approaching the initial time, one must carefully track the dependence of the Harnack constant on the rescaled drift coefficient—*rescaled* because the equation (1.3) is not exactly invariant under the usual parabolic rescaling $(t, x) \mapsto (t_0 + \lambda^2 t, x_0 + \lambda x)$. Since this dependence is seldom made explicit in Harnack inequalities found in reference texts (cf. e.g. [17, Theorem 6.27]), we adopt a more hands-on approach and prove the required estimate (cf. Lemma 7.2) by means of an argument in the spirit of Bombieri–Giusti [5, 18].

Once the desired uniform lower bound on $1 - \rho$ is established (and hence the strong parabolicity of (1.1)) for positive times, the regularity bootstrap essentially begins with one key observation on the structure of (1.1): provided one justifies *passing from divergence to non-divergence form* in (1.1), a cancellation in the mixed terms $\nabla f \cdot \nabla \rho$ occurs, yielding

$$\partial_t f + \operatorname{div}((1 - \rho)f\mathbf{e}(\theta)) = (1 - \rho)\Delta f + f\Delta\rho + \partial_\theta^2 f.$$

By formally testing the above with Δf (and then $\partial_\theta^2 f$), one is able to close the usual H^2 -type estimate. Then, through repeated applications of difference quotients in time and Schauder-type results, we are able to successively take additional time and space derivatives in (1.1) and show that all derivatives bounded, thereby establishing smoothness. The rigorous passage from the divergence to non-divergence form had already been written in [1], and relies on a Galerkin scheme (see Lemma 8.10 herein). We briefly highlight that the use of Galerkin's method as a tool for regularity (while classical) has recently proven to be very useful in settings where the method of difference quotients (in space) is not justifiable, *e.g.* in the context of elliptic equations in metric measure spaces (*cf.* [20]).

The rest of the paper is organised as follows. In §2 we recall the definition of weak solution as well as the global existence result obtained in [8]. We also recall some basic regularity results that are used throughout the paper. In §3 we state our main theorem. §4 is concerned with a classical result known as the *Bombieri–Giusti Lemma*, which is the cornerstone of the approach “without the John–Nirenberg Lemma” (*cf.* [18]) to obtaining Harnack inequalities; to make the paper self-contained, we write a statement and proof which is tailored to our setting. In §5–6 (which are respectively concerned with Moser-type L^∞ bounds and estimates on the sub-/super-level sets of the logarithm) we verify that $1 - \rho$ and its reciprocal satisfy the assumptions of the Bombieri–Giusti Lemma of §4. Having shown that this is the case, we obtain the sought quantitative lower bound on $1 - \rho$ in §7. §8 is then concerned with the regularity bootstrap for (1.1), which (although entirely self-contained herein) relies heavily on the ideas already introduced in [1, §6–7].

2. DEFINITIONS AND BACKGROUND

In this part of the manuscript, we recall the definition of weak solution of (1.1) introduced in the prior works [8, 1], and summarise the main existence result as well as some background notions that will appear in later sections.

2.1. Definitions. We define the following function space. Note that, therein, the space containing $\partial_t f$ is precisely the dual of $L_{\text{loc}}^6([0, \infty); W^{1,6}(\mathbb{T}^3))$.

Definition 2.1 (Function space X). The function space X is defined to be the family of functions $f \in L_{\text{loc}}^3([0, \infty); L^3(\mathbb{T}^3))$ which satisfy, with $\rho := \int_{\mathbb{T}} f \, d\theta$, the following:

- (i) $f \geq 0$ a.e. $(0, \infty) \times \mathbb{T}^3$;
- (ii) $\partial_\theta \sqrt{f} \in L_{\text{loc}}^2([0, \infty); L^2(\mathbb{T}^3))$ and $\partial_t f \in L_{\text{loc}}^{6/5}([0, \infty); W^{-1,6/5}(\mathbb{T}^3))$;
- (iii) $0 \leq \rho \leq 1$ a.e. $(0, \infty) \times \mathbb{T}^2$, $\partial_t \rho \in L_{\text{loc}}^2(0, \infty; H^{-1}(\mathbb{T}^2))$, $\nabla \sqrt{1 - \rho} \in L_{\text{loc}}^2(0, \infty; L^2(\mathbb{T}^2))$;
- (iv) $\sqrt{1 - \rho} \nabla \sqrt{f} \in L_{\text{loc}}^2([0, \infty); L^2(\mathbb{T}^3))$.

We recall the definition of weak solution (*cf.* [1, Definition 2.1]). For succinctness, we define the test functions $\mathcal{A} := \{\varphi \in L_{\text{loc}}^6([0, \infty); W^{1,6}(\mathbb{T}^3)) : \partial_t \varphi \in L_{\text{loc}}^{3/2}([0, \infty); L^{3/2}(\mathbb{T}^3))\}$.

Definition 2.2 (Weak solution). We say that $f \in X$ is a *weak solution* of (1.1) if, for all $\varphi \in \mathcal{A}$, it holds: for a.e. $t_1, t_2 \in (0, \infty)$

$$\begin{aligned} & \int_{t_1}^{t_2} \int_{\mathbb{T}^3} f \partial_t \varphi \, d\xi \, dt + \int_{t_1}^{t_2} \int_{\mathbb{T}^3} (1 - \rho) f \mathbf{e}(\theta) \cdot \nabla \varphi \, d\xi \, dt \\ &= \int_{t_1}^{t_2} \int_{\mathbb{T}^3} \left(((1 - \rho) \nabla f + f \nabla \rho) \cdot \nabla \varphi + \partial_\theta f \partial_\theta \varphi \right) d\xi \, dt + \int_{\mathbb{T}^3} f \varphi|_{t_2} \, d\xi - \int_{\mathbb{T}^3} f \varphi|_{t_1} \, d\xi. \end{aligned} \tag{2.1}$$

Furthermore, we say that f is a weak solution *with initial data* f_0 if it is a weak solution, and (2.1) holds with $t_1 = 0$ and the final integral replaced by $\int_{\mathbb{T}^3} f_0(\xi) \varphi(0, \xi) d\xi$.

For the convenience of the reader, we recall the notation $\xi = (x, \theta)$ in the integrals above. It is also clear that, by considering the periodic extension of f to all of $\mathbb{R}^3$, the definition of weak solution given above also implies that (1.1) holds in the sense of distributions dual to the class of test functions $C_c^\infty([0, \infty) \times \mathbb{R}^3)$; cf. [8, Lemma 4.2].

2.2. Background notions.

2.2.1. Global existence and attainment of initial data. The next remark summarises the main existence result obtained in [8].

Remark 2.3 (Global existence). Recall that, given initial data f_0 which satisfies

$$0 \leq f_0 \in L^q(\mathbb{T}^3) \quad \text{for some } q \in (1, \infty], \quad 0 \leq \rho_0 \leq 1 \quad \text{a.e. in } \mathbb{T}^2, \quad (2.2)$$

then there exists a weak solution of (1.1) with initial data f_0 (in the precise sense of Definition 2.2), which satisfies

$$\lim_{t \rightarrow 0^+} \|f(t) - f_0\|_{W^{-1, 6/5}(\mathbb{T}^3)} = 0; \quad (2.3)$$

this was proved in [8, Theorem 1]. Throughout the manuscript, we say that f_0 is *admissible initial data* if it satisfies (2.2), and whenever initial data f_0 is mentioned, the letter ρ_0 always denotes $\rho_0 = \int_{\mathbb{T}} f_0 d\theta$, i.e. the initial data for ρ .

For the convenience of the reader, we explain in detail the distributional convergence to the initial data (2.3) in the remark that follows.

Remark 2.4 (Attainment of initial data). Let f be a weak solution with initial data f_0 as per Definition 2.2. Then, for all $\varphi \in W^{1,6}(\mathbb{T}^3)$ depending only on ξ , it holds a.e. $t > 0$

$$\begin{aligned} \int_{\mathbb{T}^3} (f(t, \xi) - f_0(\xi)) \varphi(\xi) d\xi &= \int_0^t \int_{\mathbb{T}^3} (1 - \rho) f \mathbf{e}(\theta) \cdot \nabla \varphi d\xi ds \\ &\quad - \int_0^t \int_{\mathbb{T}^3} \left(((1 - \rho) \nabla f + f \nabla \rho) \cdot \nabla \varphi + \partial_\theta f \partial_\theta \varphi \right) d\xi ds. \end{aligned} \quad (2.4)$$

The boundedness of the relevant terms from Definition 2.2 (cf. [8, Eq. (1.5)]) yields (2.3):

$$\lim_{t \rightarrow 0^+} \|f(t) - f_0\|_{(W^{1,6}(\mathbb{T}^3))'} = \lim_{t \rightarrow 0^+} \sup_{\|\varphi\|_{W^{1,6} \leq 1}} \left| \int_{\mathbb{T}^3} (f(t) - f_0) \varphi d\xi \right| = 0. \quad (2.5)$$

Next is a remark about the attainment of ρ -initial data, which was not proved before.

Remark 2.5 (Convergence to initial data for ρ). Suppose that f is a weak solution of (1.1) with admissible initial data f_0 . Then, it holds

$$\lim_{t \rightarrow 0^+} \|\rho(t) - \rho_0\|_{L^2(\mathbb{T}^2)} = 0. \quad (2.6)$$

Note that $\rho \in C([0, \infty); L^2(\mathbb{T}^2))$ is immediate from $\partial_t \rho \in L_{\text{loc}}^2([0, \infty); H^{-1}(\mathbb{T}^2))$, $\rho \in L_{\text{loc}}^2([0, \infty); H^1(\mathbb{T}^2))$ and the Lions–Magenes Lemma [21, Ch.3, §1.4 Lem.1.2], but this *does not directly imply* (2.6). Nevertheless, this convergence is true, cf. Appendix A.1.

2.2.2. Background regularity notions. We recall some elementary boundedness results, which hold for any weak solution of (1.1) produced by admissible initial data (2.2).

Recall the equation (1.3) for ρ , where the drift term depends on the polarisation $\mathbf{p}$ given by (1.4), $\mathbf{p} = \int_{\mathbb{T}} f \mathbf{e}(\theta) d\theta$. Since $f \geq 0$ and $0 \leq \rho \leq 1$ a.e. from Definition 2.1–2.2, it holds

$$|\mathbf{p}(t, x)| \leq \int_{\mathbb{T}} |f(t, x, \theta)| d\theta = \rho(t, x) \leq 1 \implies \|\mathbf{p}\|_{L^\infty((0, \infty) \times \mathbb{T}^2)} \leq 1; \quad (2.7)$$

this boundedness holds for all *angular moments* of f , *e.g.* $\mathbf{P} = \int_{\mathbb{T}} f \mathbf{e}(\theta) \otimes \mathbf{e}(\theta) d\theta$ satisfies $\|\mathbf{P}\|_{L^\infty} < \infty$. Furthermore, simply by multiplying the equation (1.1) for f with $\mathbf{e}(\theta)$ and integrating in θ , we uncover the equation for $\mathbf{p}$:

$$\partial_t \mathbf{p} + \operatorname{div}((1 - \rho)\mathbf{P}) = \operatorname{div}((1 - \rho)\nabla \mathbf{p} + \mathbf{p} \otimes \nabla \rho) - \mathbf{p}, \quad (2.8)$$

which, in coordinate form with summation convention, reads as $\partial_t p_i + \partial_j((1 - \rho)P_{ij}) = \partial_j((1 - \rho)\partial_j p_i + p_i \partial_j \rho) - p_i$. We also recall that the drift in (1.3) satisfies

$$\operatorname{div}((1 - \rho)\mathbf{p}) \in L_{\operatorname{loc}}^{3/2}([0, \infty); L^{3/2}(\mathbb{T}^2)); \quad (2.9)$$

indeed, $\operatorname{div}((1 - \rho)\mathbf{p}) = -\nabla \rho \cdot \mathbf{p} + (1 - \rho) \operatorname{div} \mathbf{p}$, with $\nabla \rho \cdot \mathbf{p} \in L_{\operatorname{loc}}^2((0, \infty); L^2(\mathbb{T}^2))$ from Definition 2.1 and (2.7), and (for all $T > 0$)

$$\begin{aligned} \int_0^T \int_{\mathbb{T}^2} |(1 - \rho)\nabla \mathbf{p}|^{3/2} dx dt &= \int_0^T \int_{\mathbb{T}^2} (1 - \rho)^{3/2} \left| \int_{\mathbb{T}} \nabla f \otimes \mathbf{e}(\theta) d\theta \right|^{3/2} dx dt \\ &\leq \sqrt{2\pi} \int_0^T \int_{\mathbb{T}^3} \left(\sqrt{1 - \rho} \frac{|\nabla f|}{\sqrt{f}} \right)^{3/2} (\sqrt{f})^{3/2} d\xi dt \\ &\leq 4\sqrt{\pi} \|f\|_{L^3([0, T] \times \mathbb{T}^3)}^{3/4} \|\sqrt{1 - \rho} \nabla \sqrt{f}\|_{L^2([0, T] \times \mathbb{T}^3)}^{3/2}, \end{aligned} \quad (2.10)$$

where we used Jensen's inequality and $0 \leq 1 - \rho \leq 1$ to get the second line and Hölder's inequality with the exponent $4/3$ to get the final line. By (1.3) and (2.9), we get

$$\partial_t \rho - \Delta \rho = -\operatorname{div}((1 - \rho)\mathbf{p}) \in L_{\operatorname{loc}}^{3/2}([0, \infty); L^{3/2}(\mathbb{T}^2)),$$

whence, by the Calderón–Zygmund Theorem for the heat equation (*cf.* [16, Theorem 3.1]),

$$\partial_t \rho, \nabla^2 \rho \in L^{3/2}((t, T) \times \mathbb{T}^2) \quad \forall 0 < t < T; \quad (2.11)$$

we will use (2.11) as a starting point for our analysis, in §5 of the present work.

We finish this subsection with four well-known standard results which will be used repeatedly in §8. The first is an interpolation lemma that follows from the Gagliardo–Nirenberg inequality, the second and third are Schauder-type results for non-divergence form parabolic equations (in Sobolev spaces, and Hölder spaces, respectively), and the fourth is a Gehring Lemma for divergence form parabolic equations.

Lemma 2.6 (§1 Proposition 3.2, [10]). *Let $d \in \mathbb{N}$, $\Omega \subset \mathbb{R}^d$ have piecewise smooth boundary, $I \subseteq [0, \infty)$ any bounded interval, and $p, m \geq 1$. Then, there exists a positive constant $C = C(d, p, m, \partial\Omega, \Omega)$, independent of I , such that, for all $v \in L^\infty(I; L^m(\Omega)) \cap L^p(I; W^{1,p}(\Omega))$,*

$$\|v\|_{L^q(I \times \Omega)} \leq C(1 + |I|)^{\frac{1}{q}} \left(\|v\|_{L^\infty(I; L^m(\Omega))} + \|v\|_{L^p(I; W^{1,p}(\Omega))} \right) \quad \text{with } q = p(1 + \frac{m}{d}).$$

Consequently, we have, for any bounded interval $I \subseteq [0, \infty)$, the continuous embeddings

$$\begin{aligned} L^\infty(I; L^2(\mathbb{T}^2)) \cap L^2(I; H^1(\mathbb{T}^2)) &\hookrightarrow L^4(I \times \mathbb{T}^2), \\ L^\infty(I; L^2(\mathbb{T}^3)) \cap L^2(I; H^1(\mathbb{T}^3)) &\hookrightarrow L^{10/3}(I \times \mathbb{T}^3). \end{aligned}$$

The next two results are well-known. The first is [23, Theorem 1.9] (*cf.* [11, Lemma 5.14] and [13, Theorem 2.4]), and the second is [23, Theorem 1.19] (*cf.* [15, Theorem 8.11.1]). We recall the Hölder spaces and their parabolic analogues: for $\alpha \in (0, 1)$ and $Q \subseteq \mathbb{R}^{1+d}$,

$$C^\alpha(Q) = \left\{ v \in C(\overline{Q}) : \|v\|_{C^\alpha(Q)} < \infty \right\}, \quad \|v\|_{C^\alpha(Q)} = \|v\|_{L^\infty(Q)} + [v]_{\alpha; Q},$$

where $[v]_{\alpha; Q} = \sup_{z_1, z_2 \in Q, z_1 \neq z_2} \frac{|v(z_1) - v(z_2)|}{|z_1 - z_2|^\alpha}$; recall that $(C^\alpha, \|\cdot\|_{C^\alpha})$ is a Banach space, and

$$C^{\alpha/2, \alpha}(Q) = \left\{ v \in C(\overline{Q}) : \|v\|_{C^{\alpha/2, \alpha}(Q)} < \infty \right\}, \quad \|v\|_{C^{\alpha/2, \alpha}(Q)} = \|v\|_{L^\infty(Q)} + [v]_{\alpha/2, \alpha; Q},$$

where $[v]_{\alpha/2, \alpha; Q} = \sup_{(t_1, x_1) \neq (t_2, x_2)} \frac{|v(t_1, x_1) - v(t_2, x_2)|}{|t_1 - t_2|^{\alpha/2} + |x_1 - x_2|^\alpha}$. Also define

$$C^{1+\alpha/2, 2+\alpha}(Q) = \left\{ \partial_t v, \nabla^2 v \in C^{\alpha/2, \alpha}(Q) \right\},$$

with the norm $\|v\|_{C^{1+\alpha/2, 2+\alpha}(Q)} = \|v\|_{L^\infty(Q)} + \|\nabla v\|_{L^\infty(Q)} + \|\partial_t v\|_{C^{\alpha/2, \alpha}(Q)} + \|\nabla^2 v\|_{C^{\alpha/2, \alpha}(Q)}$.

Lemma 2.7 (Schauder estimate: Sobolev spaces). *Let $I \subseteq [0, \infty)$ be any compact interval, $p \in [2, \infty)$, $\alpha \in (0, 1)$, $c > 1$, $d \in \mathbb{N}$. Assume $A \in [C^\alpha(I \times \mathbb{T}^d)]^{d \times d}$ is symmetric and*

$$c^{-1}|w|^2 \leq A(t, x)w \cdot w \leq c|w|^2 \quad \forall w \in \mathbb{R}^d, \quad \forall (t, x) \in I \times \mathbb{T}^d.$$

Suppose that $v \in L^2(I; H^2(\mathbb{T}^d))$, with $\partial_t v \in L^2(I \times \mathbb{T}^d)$, satisfies, for $F \in L^p(I \times \mathbb{T}^d)$,

$$\partial_t v - A : \nabla^2 v = F.$$

Then, for any $I' \subset\subset I$ compactly contained in I , it holds $\partial_t v, \nabla^2 v \in L^p(I' \times \mathbb{T}^d)$, and

$$\|\partial_t v\|_{L^p(I' \times \mathbb{T}^d)} + \|\nabla_\xi^2 v\|_{L^p(I' \times \mathbb{T}^d)} \leq C \left(\|F\|_{L^p(I \times \mathbb{T}^d)} + \|v\|_{L^2(I \times \mathbb{T}^d)} \right), \quad (2.12)$$

where C depends only on $|I \setminus I'|$, I , c , $\|A\|_{C^\alpha}$, α , d .

Lemma 2.8 (Schauder estimate: Hölder spaces). *Let $I \subseteq [0, \infty)$ be any compact interval, $\alpha \in (0, 1)$, $c > 1$, and $d \in \mathbb{N}$. Assume the matrix $A \in [C^\alpha(I \times \mathbb{T}^d)]^{d \times d}$ is symmetric and*

$$c^{-1}|w|^2 \leq A(t, x)w \cdot w \leq c|w|^2 \quad \forall w \in \mathbb{R}^d, \quad \forall (t, x) \in I \times \mathbb{T}^d.$$

Suppose that $v \in L^2(I; H^2(\mathbb{T}^d))$, with $\partial_t v \in L^2(I \times \mathbb{T}^d)$, satisfies, for $F \in C^{\alpha/2, \alpha}(I \times \mathbb{T}^d)$,

$$\partial_t v - A : \nabla^2 v = F.$$

Then, for any $I' \subset\subset I$ compactly contained in I , it holds $v \in C^{1+\alpha/2, 2+\alpha}(I' \times \mathbb{T}^d)$, and

$$\|v\|_{C^{1+\alpha/2, 2+\alpha}(I' \times \mathbb{T}^d)} \leq C \left(\|F\|_{C^{\alpha/2, \alpha}(I \times \mathbb{T}^d)} + \|v\|_{L^2(I \times \mathbb{T}^d)} \right), \quad (2.13)$$

where C depends only on $|I \setminus I'|$, I , c , $\|A\|_{C^\alpha}$, α , d .

The final result of this subsection is a consequence of (and a non-quantitative version of) [4, Theorem 8.2] and is a classical Gehring-type “improvement of integrability” result for solutions of strongly parabolic *systems* in divergence form with *measurable* coefficients—we emphasise that, unlike in Lemmas 2.7–2.8, the entries of the matrix A are *not* required to be uniformly continuous, and A need not be symmetric. In what follows, for $p \geq 1$,

$$p_* := \frac{(d+2)p}{d+2+p}, \quad (2.14)$$

where d is the spatial dimension of the domain.

Lemma 2.9 (Gehring-type Lemma). *Let $I \subseteq [0, \infty)$ be any compact interval, $c > 1$, $N, d \in \mathbb{N}$, and suppose g, G are vector fields satisfying*

$$g \in [L^{p_*}(I \times \mathbb{T}^d)]^N, \quad G \in [L^p(I \times \mathbb{T}^d)]^{N \times d}$$

for some $p > 2$. Assume the tensor $A \in [L^\infty(I \times \mathbb{T}^d)]^{N^2 d^2}$ satisfies the ellipticity condition

$$c^{-1}|w|^2 \leq \underbrace{A(t, x)w \cdot w}_{\sum_{\alpha, \beta, i, j} A(t, x)_{ij}^{\alpha\beta} w_i^\alpha w_j^\beta} \leq c|w|^2 \quad \forall w \in \mathbb{R}^{d \times N}, \quad \forall (t, x) \in I \times \mathbb{T}^d,$$

and suppose that $v \in [L^\infty(I \times \mathbb{T}^d) \cap L^2(I; H^1(\mathbb{T}^d))]^N$, with $\partial_t v \in [L^2(I; H^{-1}(\mathbb{T}^d))]^N$, satisfies the linear divergence form equation

$$\partial_t v - \operatorname{div}(A \nabla v) = g + \operatorname{div} G.$$

Then, there exists $\delta > 0$ depending only on $c, I, p, d, \|A\|_{L^\infty}$, such that, for any $I' \subset\subset I$ compactly contained in I , it holds $\nabla v \in L^{2+\delta}(I' \times \mathbb{T}^d)$.

2.3. Notations. Throughout the paper, for $r > 0$ and $d = 2, 3$, we have the following notation for the parabolic cylinders:

$$Q_r(t) = (t - r, t + r) \times \mathbb{T}^d, \quad Q_r^+(t) = (t, t + r) \times \mathbb{T}^d, \quad Q_r^-(t) = (t - r, t) \times \mathbb{T}^d; \quad (2.15)$$

and we simply write Q_r to denote $Q_r(0)$ and analogously for $Q_r^\pm = Q_r^\pm(0)$. We also use the notations $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$ for real numbers a, b . We write $L_{t,\text{loc}}^p W_x^{k,q}$ (resp. $L_{t,\text{loc}}^p W_\xi^{k,q}$) as shorthand for $L^p((t, T); W^{k,q}(\mathbb{T}^d))$ ($d = 2, 3$, resp.) for all $(t, T) \subset\subset (0, \infty)$, where $\subset\subset$ means *compactly contained*. In the bootstrap argument of §8.3, we use the notation $D^m v$ with $m \in \mathbb{N}$ to denote the tensor of all derivatives of order m in space-angle (*i.e.* excluding the time variable). Throughout §5–7, we write $u = 1 - \rho$.

3. MAIN THEOREM

Our main theorem is the following. It may be interpreted as a significant improvement of [1, Theorem 1], where one no longer requires stronger assumptions on the initial data (*cf.* [1, Definition 3.1]) to deduce the smoothness of weak solutions. The result shows that all weak solutions are classical solutions, provided we exclude “saturated” initial data satisfying $\rho_0 \equiv 1$ a.e. The underlying mechanism (as previously mentioned) is that a strictly positive lower bound on $1 - \rho$ makes (1.1) strongly parabolic, and thus self-regularising.

Theorem 1 (Interior non-degeneracy and smoothness of weak solutions). *Let f be a weak solution of (1.1) with admissible initial data f_0 satisfying (2.2). Then, there exists a constant $C > 1$ (independent of t, T , the initial data, and the solution) such that*

$$\operatorname{ess\,inf}_{(t,T) \times \mathbb{T}^2} (1 - \rho) \geq C^{-1} \|1 - \rho_0\|_{L^1(\mathbb{T}^2)} e^{-C(T+t^{-10})} \quad (3.1)$$

for all $0 < t < T < \infty$. Moreover, if $\|1 - \rho_0\|_{L^1(\mathbb{T}^2)} > 0$, then $f \in C^\infty((0, \infty) \times \mathbb{T}^3)$.

Of course, we do not claim optimality in the exponents on the right-hand side of (3.1). Note that Theorem 1 allows for degenerate choices of initial data with $\rho_0 = 1$ on a Borel set of positive measure (provided it is not of full measure), and we remind the reader that the terminology “ f is a weak solution with initial data f_0 ” was explained in Definition 2.2.

Remark 3.1 (Limiting behaviour as $T \rightarrow \infty$). Note that the lower bound provided by (3.1) vanishes in the limit as $T \rightarrow \infty$, thus leading to (potentially) degenerate diffusion at infinite times. This observation is compatible with the segregated (and hence non-smooth) stationary states that were numerically obtained in [7]; *cf.* Figures 4 and 5 therein.

4. THE BOMBIERI–GIUSTI LEMMA

The following lemma is fundamental when proving Harnack inequalities (it will be used subsequently to prove the sought lower bound on $1 - \rho$ in §7, which then implies that (1.1) is strongly parabolic for positive times). It applies to any non-negative measurable function, not necessarily the solution of a particular equation. The result is classical and the underlying argument is well-known to specialists (*cf.* [5, 18]); of course we do not claim novelty, but we include a proof tailored to our setting (with slightly different statement and proof to—but closely following—the one in [6, §3.2]) to make the paper self-contained.

Lemma 4.1 (Bombieri–Giusti Lemma). *Let $\beta > 0$, $\tilde{r} \in (0, 1]$, and $\{\mathcal{Q}_r\}_{r \in (0, \tilde{r}]}$ be a family of nested subsets of $\mathbb{R}^{d+1}$ with $\mathcal{Q}_r \subset \mathcal{Q}_R$ for all $0 < r < R \leq \tilde{r}$ and $|\mathcal{Q}_1| < \infty$. Assume that $\varphi \in L^1(\mathcal{Q}_1)$ is non-negative and satisfies:*

- (i) *there exists a positive constant C_0 independent of $\beta, r, R, \tilde{r}, p$ such that*

$$\|\varphi\|_{L^\infty(\mathcal{Q}_r)} \leq \left(\frac{C_0}{(R-r)^\beta} \right)^{\frac{1}{p}} \|\varphi\|_{L^p(\mathcal{Q}_R)} \quad \text{for all } 0 < r < R \leq \tilde{r} \text{ and all } p \in (0, 1); \quad (4.1)$$

(ii) there exists a positive constant $C_1 > e^{-1}$ such that, for all $r \in (0, \tilde{r})$,

$$|\{(t, y) \in \mathcal{Q}_r : \log \varphi(t, y) > s\}| \leq \frac{C_1}{s} \quad \text{for all } s > 0. \quad (4.2)$$

Then, there exists a positive constant $C_2 = C_2(C_0, C_1, \beta)$, independent of $r, \tilde{r}$, such that

$$\|\varphi\|_{L^\infty(\mathcal{Q}_r)} \leq \exp(C_2(\tilde{r} - r)^{-2\beta}) \quad \text{for all } r \in (0, \tilde{r}). \quad (4.3)$$

We emphasise that C_0, C_1 are both independent of $p, r, \tilde{r}, R$ in the assumptions (4.1)–(4.2); and hence C_2 also is. Furthermore, the sets $\mathcal{Q}_r$ may be more general subsets than the parabolic cylinders of (2.15). We highlight that the content of the result is not only the local boundedness of φ *per se*, but the precise dependence on $r, \tilde{r}$ given by (4.3).

Proof. Define the increasing function $r \mapsto \Phi(r) := \text{ess sup}_{\mathcal{Q}_r} \log \varphi$. This function is well-defined since φ is non-negative; however, *a priori* it may be infinite. We will prove that the assumptions of the lemma imply that, for any $r \in (0, \tilde{r})$, either:

- (a) $\Phi(r) \leq 2C_1$, in which case there is nothing to prove and we set $C_2 = 2C_1$;
- (b) or, $\Phi(r) > 2C_1$, in which case it holds

$$\Phi(r) \leq \frac{3}{4}\Phi(R) + \frac{C}{(R - r)^{2\beta}} \quad \forall R \in (r, \tilde{r}], \quad (4.4)$$

where the positive constant $C = C(C_0, C_1)$ is independent of $r, R, \tilde{r}$.

The rest of the proof is divided into two steps.

1. *Proof of (4.3) in case (b).* We prove the result (4.3) in the case where (b) holds. Suppose that case (b) holds at a given $r \in (0, \tilde{r})$. Define the increasing sequence of radii $\{r_n\}_{n \geq 0}$:

$$r_n = \tilde{r} - \frac{\tilde{r} - r}{1 + n} \quad \forall n \in \mathbb{N} \cup \{0\} \implies r_{n+1} - r_n = \frac{\tilde{r} - r}{(n+2)(n+1)},$$

whence $r = r_0 < r_1 < \dots < r_n \nearrow \tilde{r}$. Note that, if there exists $n \in \mathbb{N}$ such that $\Phi(r_n) \leq 2C_1$, then since Φ is increasing, we get $\Phi(r) = \Phi(r_0) \leq \Phi(r_n) \leq 2C_1$, and we are back to case (a), where there is nothing to prove. Thus, henceforth we assume $\Phi(r_n) > 2C_1$ for all $n \in \mathbb{N}$, whence (4.4) holds with $r = r_n, R = r_{n+1}$ ($n \in \mathbb{N}_{\geq 0}$). By induction on (4.4),

$$\Phi(r) = \Phi(r_0) \leq \left(\frac{3}{4}\right)^n \Phi(r_n) + C \sum_{j=0}^{n-1} \left(\frac{3}{4}\right)^j \frac{1}{(r_{j+1} - r_j)^{2\beta}}, \quad (4.5)$$

and, using the assumption (4.1) with $R = \tilde{r}$, the first term on the right-hand side decays:

$$\begin{aligned} \left(\frac{3}{4}\right)^n \Phi(r_n) &= \left(\frac{3}{4}\right)^n \log \|\varphi\|_{L^\infty(\mathcal{Q}_{r_n})} \leq \left(\frac{3}{4}\right)^n \log \left(\left(\frac{C_0}{(\tilde{r} - r_n)^\beta} \right)^{\frac{1}{p}} \|\varphi\|_{L^p(\mathcal{Q}_1)} \right) \\ &= \left(\frac{3}{4}\right)^n \left(\frac{1}{p} \log C_0 + \log \|\varphi\|_{L^p(\mathcal{Q}_1)} \right) + \frac{\beta}{p} \left(\frac{3}{4}\right)^n \log \left(\frac{1+n}{\tilde{r} - r} \right), \end{aligned}$$

so $\left(\frac{3}{4}\right)^n \Phi(r_n) \rightarrow 0$, for any fixed $p \in (0, 1)$; note that $\|\varphi\|_{L^p(\mathcal{Q}_1)} \leq |\mathcal{Q}_1|^{1/p-1} \|\varphi\|_{L^1(\mathcal{Q}_1)} < \infty$ for $p \in (0, 1)$ by Hölder's inequality. By letting $n \rightarrow \infty$ in (4.5), we recover

$$\Phi(r) \leq C \sum_{j=0}^{\infty} \left(\frac{3}{4}\right)^j \frac{1}{(r_{j+1} - r_j)^{2\beta}} = \underbrace{C \left(\sum_{j=0}^{\infty} \left(\frac{3}{4}\right)^j (j+2)^{2\beta} (j+1)^{2\beta} \right)}_{=: C'} (\tilde{r} - r)^{-2\beta},$$

where C' is independent of $r, \tilde{r}$; note that the infinite sum manifestly converges for all $\beta > 0$ by the d'Alembert ratio test. It then follows (recalling the non-negativity of φ) that

$$\|\varphi\|_{L^\infty(\mathcal{Q}_r)} = \text{ess sup}_{\mathcal{Q}_r} \exp \log \varphi = e^{\Phi(r)} \leq e^{C'(\tilde{r} - r)^{-2\beta}}.$$

We set $C_2 = C'$, which depends only on C_0, C_1, β , and the proof of (4.3) is complete.

2.i. *Proof of (4.4): L^p estimate for the logarithm.* Our goal in this step is to obtain (for φ the function in the statement of Lemma 4.1 with $r \in (0, \tilde{r})$ and $\Phi(r)$ satisfying case (b) at the beginning of the proof) the following estimate:

$$\|\varphi\|_{L^{p_R}(\mathcal{Q}_R)}^{p_R} \leq C_3 e^{\frac{1}{2}p_R\Phi(R)} \quad \forall R \in (r, \tilde{r}], \quad (4.6)$$

where the positive constant C_3 is independent of $p_R, \beta, r, \tilde{r}, R$, and p_R is prescribed by

$$p_R = \frac{2}{\Phi(R)} \log \frac{\Phi(R)}{2C_1}. \quad (4.7)$$

Note that $p_R \in (0, 1)$ since the assumption $\Phi(r) > 2C_1$ with $C_1 > e^{-1}$ implies $\Phi(R) \geq \Phi(r) > 2C_1$ since Φ is increasing, and (since $\sup_{s \in (1, \infty)} s^{-1} \log s = e^{-1}$)

$$0 < \frac{2}{\Phi(R)} \log \frac{\Phi(R)}{2C_1} = \frac{1}{C_1} \frac{2C_1}{\Phi(R)} \log \frac{\Phi(R)}{2C_1} < e \sup_{s \in (1, \infty)} s^{-1} \log s = 1, \quad (4.8)$$

Define $A_R := \{(t, y) \in \mathcal{Q}_R : \log \varphi(t, y) > \frac{1}{2}\Phi(R)\}$ and $B_R := \{(t, y) \in \mathcal{Q}_R : \log \varphi(t, y) \leq \frac{1}{2}\Phi(R)\}$. For all $p > 0$, it holds $\text{ess sup}_{\mathcal{Q}_R} \varphi^p = \text{ess sup}_{\mathcal{Q}_R} \exp(p \log \varphi) = e^{p\Phi(R)}$, whence

$$\|\varphi\|_{L^p(\mathcal{Q}_R)}^p = \iint_{A_R} \varphi^p + \iint_{B_R} \varphi^p \leq |A_R| e^{p\Phi(R)} + |B_R| e^{\frac{p}{2}\Phi(R)} \leq \frac{2C_1}{\Phi(R)} e^{p\Phi(R)} + |\mathcal{Q}_1| e^{\frac{p}{2}\Phi(R)}, \quad (4.9)$$

where we used (4.2) for the first term, and bluntly bounded $|B_R| \leq |\mathcal{Q}_R| \leq |\mathcal{Q}_1|$. We emphasise that the above estimate holds for all possible choices of $p > 0$ and $R \in (r, \tilde{r}]$; including if we choose p to depend on R as per (4.7). Next, we estimate the first term on the right-hand side of (4.9) by the final term. Setting $p = p_R$ as per (4.7), we get

$$\frac{2C_1}{\Phi(R)} e^{p_R\Phi(R)} = e^{\frac{p_R}{2}\Phi(R)},$$

and we deduce from (4.9) that, defining the positive constant $C_3 = 1 + |\mathcal{Q}_1|$ independent of $p, d, \beta, r, R, \tilde{r}$, there holds (4.6) with p_R prescribed by (4.7).

2.ii. *Proof of (4.4): estimate on $\Phi(r)$.* Using (4.1) with the choice $p = p_R \in (0, 1)$ and (4.6), the monotonicity of $\log$ implies, since $\Phi(r) = \frac{1}{p_R} \log(\|\varphi\|_{L^\infty(\mathcal{Q}_r)}^{p_R})$,

$$\begin{aligned} \Phi(r) &\leq \frac{1}{p_R} \log \left(\frac{C_0}{(R-r)^\beta} C_3 e^{\frac{1}{2}p_R\Phi(R)} \right) \leq \frac{1}{2}\Phi(R) + \frac{1}{p_R} \log \left(\frac{C_0 C_3}{(R-r)^\beta} \right) \\ &= \frac{1}{2}\Phi(R) \left(1 + \frac{\log \left(\frac{C_0 C_3}{(R-r)^\beta} \right)}{\log \left(\frac{\Phi(R)}{2C_1} \right)} \right), \end{aligned} \quad (4.10)$$

where we used the explicit representation (4.7) for p_R to obtain the final line. We divide the analysis into two cases. The first case is when it holds: $\log(\frac{C_0 C_3}{(R-r)^\beta}) / \log(\frac{\Phi(R)}{2C_1}) \leq \frac{1}{2}$, in which case we deduce from (4.10) that it holds $\Phi(r) \leq \frac{3}{4}\Phi(R)$, and hence (4.4) is satisfied. On the other hand, if instead it holds $\log(\frac{C_0 C_3}{(R-r)^\beta}) / \log(\frac{\Phi(R)}{2C_1}) > \frac{1}{2}$, then (using again $\Phi(R) \geq \Phi(r) > 2C_1$) we deduce $\Phi(r) \leq \Phi(R) < \frac{2C_1 C_0^2 C_3^2}{(R-r)^{2\beta}}$, where we used that Φ is an increasing function of the radius to write the first inequality. In either case, (4.4) holds with $C(C_0, C_1) = 2C_1 C_0^2 (1 + |\mathcal{Q}_1|)^2$ using the definition of C_3 . The proof is complete. $\square$

5. MOSER-TYPE LOCAL L^∞ ESTIMATES

In this section, we derive the usual Moser-type local L^∞ estimates on the quantity $(1-\rho)$ for which we wish to show infinite speed of propagation; here, *local* means local-in-time, *i.e.* away from $t = 0$. This will consist in proving Caccioppoli-type inequalities for this quantity (and its reciprocal $(1-\rho)^{-1}$) and estimates on the measure of the super-/sub-level sets of its logarithm; in essence, we wish to show that $(1-\rho)$ satisfies the assumptions of

the Bombieri–Giusti Lemma 4.1. We follow the presentation in [6, §3.2], which is based on Moser’s iteration, as opposed to the De Giorgi method used in *e.g.* [1, Proposition 5.2].

The main focus of the subsections that follow is the quantity $u := 1 - \rho$, for which we wish to show an instantaneous positivity result. We rewrite (1.3) in terms of u :

$$\partial_t u - \operatorname{div}(\mathbf{p}u) = \Delta u. \quad (5.1)$$

Since $0 \leq \rho \leq 1$ a.e., we have $0 \leq u \leq 1$ a.e. Moreover, the regularity (2.11) implies

$$\partial_t u, \nabla^2 u \in L^{3/2}((t, T) \times \mathbb{T}^2) \quad \forall 0 < t < T. \quad (5.2)$$

5.1. Moser lower bound. The main estimate of this subsection is the following (Lemma 5.1). For clarity of presentation, the analysis is performed for weak solutions of (5.1) studied on the parabolic half-cylinder Q_r^- centered at $t = 0$ (*cf.* (2.15)); however, by translation these estimates hold on any cylinder of the form $Q_r^-(t) = (t - r, t) \times \mathbb{T}^2$.

As per (5.2), we assume that the weak solutions u of (5.1) satisfy:

$$0 \leq u \leq 1 \quad \text{a.e.}, \quad \partial_t u, \nabla^2 u \in L_{\text{loc}}^{3/2}(Q), \quad (5.3)$$

where Q is any open cylinder which is compactly contained within the domain of definition of the solution (which we do not specify explicitly, for succinctness).

Lemma 5.1 (L^∞ bound $(u + \varepsilon)^{-1}$). *Let u be a weak solution of (5.1) satisfying (5.3), $\varepsilon, p \in (0, 1)$, $0 < r < R \leq 1$. There exists a constant C independent of p, r, R, ε such that*

$$\|(u + \varepsilon)^{-1}\|_{L^\infty(Q_r^-)} \leq \left(\frac{C}{(R - r)^2} \right)^{\frac{1}{p}} \|(u + \varepsilon)^{-1}\|_{L^p(Q_R^-)}. \quad (5.4)$$

Proof. The proof is divided into three steps.

1. *Caccioppoli-type inequality.* Let $q > 0$ and, for all $\varepsilon \in (0, 1)$, define $w_\varepsilon := (u + \varepsilon)^{-q/2}$. Let $0 < r < R \leq 1$. Our objective in this step is to show (for C independent of q, r, R, ε):

$$\operatorname{ess\,sup}_{s \in (-r, 0)} \|w_\varepsilon(s)\|_{L^2(\mathbb{T}^2)}^2 + \int_{-r}^0 \|\nabla w_\varepsilon(s)\|_{L^2(\mathbb{T}^2)}^2 \, ds \leq \frac{C(1 + q^2)}{R - r} \int_{-R}^0 \|w_\varepsilon(s)\|_{L^2(\mathbb{T}^2)}^2 \, ds. \quad (5.5)$$

To this end, let $\psi \in C_c^\infty((-R, 0])$ be such that

$$0 \leq \psi \leq 1 \text{ on } [-1, 0], \quad \psi|_{[-1, -R]} = 0, \quad \psi|_{[-r, 0]} = 1, \quad \|\partial_t \psi\|_{L^\infty((0, 1))} \leq \frac{C}{R - r}, \quad (5.6)$$

for some C independent of r, R ; ψ is only a function of t . Recalling the discussion that follows (5.1) and (5.3), we have $w_\varepsilon \in L^\infty((-1, 1) \times \mathbb{T}^2) \cap L^2((-1, 1); H^1(\mathbb{T}^2))$; furthermore,

$$\phi := -q(u + \varepsilon)^{-q-1} \psi^2 = -q w_\varepsilon^{2+\frac{2}{q}} \psi^2$$

also belongs to this space, whence it is an admissible test function. We therefore insert ϕ as a test function into the weak formulation of (5.1) and get

$$\langle \partial_t u, \phi \rangle = -q \langle \partial_t (w_\varepsilon^{-\frac{2}{q}}), w_\varepsilon^{2+\frac{2}{q}} \psi^2 \rangle = 2 \langle w_\varepsilon \partial_t w_\varepsilon, \psi^2 \rangle = \frac{d}{dt} \int_{\mathbb{T}^2} |\psi w_\varepsilon|^2 \, dx - 2 \int_{\mathbb{T}^2} \psi \partial_t \psi w_\varepsilon^2 \, dx,$$

where we used (5.3) to justify the Sobolev chain rule (*cf.* [12, §5.10]); we emphasise that we are *not* using the Lions–Magenes Lemma here. Similarly, it holds

$$\begin{aligned} -\langle \operatorname{div}(\mathbf{p}u), \phi \rangle &= -2(q+1) \int_{\mathbb{T}^2} \psi^2 \mathbf{p} (w_\varepsilon^{-\frac{2}{q}} - \varepsilon) \cdot w_\varepsilon^{1+\frac{2}{q}} \nabla w_\varepsilon \, dx \\ &= -2(q+1) \int_{\mathbb{T}^2} \psi^2 w_\varepsilon \mathbf{p} \cdot \nabla w_\varepsilon \, dx + 2(q+1) \varepsilon \int_{\mathbb{T}^2} \psi^2 w_\varepsilon^{1+\frac{2}{q}} \mathbf{p} \cdot \nabla w_\varepsilon \, dx, \end{aligned}$$

and $\langle \Delta u, \phi \rangle = q \int_{\mathbb{T}^2} \psi^2 \nabla(w_\varepsilon^{-\frac{2}{q}}) \cdot \nabla(w_\varepsilon^{2+\frac{2}{q}}) dx = -\frac{4}{q}(q+1) \int_{\mathbb{T}^2} |\nabla(\psi w_\varepsilon)|^2 dx$. Observe that

$$\left| \varepsilon \int_{\mathbb{T}^2} \psi^2 w_\varepsilon^{1+\frac{2}{q}} \mathbf{p} \cdot \nabla w_\varepsilon dx \right| \leq \varepsilon \int_{\mathbb{T}^2} \underbrace{w_\varepsilon^{\frac{2}{q}}}_{\leq \varepsilon^{-1}} \psi w_\varepsilon |\nabla(\psi w_\varepsilon)| dx \leq \int_{\mathbb{T}^2} \psi w_\varepsilon |\nabla(\psi w_\varepsilon)| dx,$$

where we also used $|\mathbf{p}| \leq 1$ by (2.7), hence, using also Young's inequality, we obtain

$$\begin{aligned} \frac{d}{dt} \|\psi w_\varepsilon\|_{L^2(\mathbb{T}^2)}^2 + 4(1+q^{-1}) \|\nabla(\psi w_\varepsilon)\|_{L^2(\mathbb{T}^2)}^2 &\leq 4(q+1) \int \psi w_\varepsilon |\nabla(\psi w_\varepsilon)| dx + 2 \int \psi |\partial_t \psi| w_\varepsilon^2 dx \\ &\leq 2q^2(1+q^{-1}) \|\psi w_\varepsilon\|_{L^2(\mathbb{T}^2)}^2 + 2(1+q^{-1}) \|\nabla(\psi w_\varepsilon)\|_{L^2(\mathbb{T}^2)}^2 + \frac{2C}{R-r} \|w_\varepsilon\|_{L^2(\mathbb{T}^2)}^2. \end{aligned}$$

By bounding the first term on the right-hand side bluntly using $0 \leq \psi \leq 1$, it follows that

$$\frac{d}{dt} \|\psi w_\varepsilon\|_{L^2(\mathbb{T}^2)}^2 + (1+q^{-1}) \|\nabla(\psi w_\varepsilon)\|_{L^2(\mathbb{T}^2)}^2 \leq C \left(q^2(1+q^{-1}) + \frac{1}{R-r} \right) \|w_\varepsilon\|_{L^2(\mathbb{T}^2)}^2,$$

where C is independent of q, r, R, ε . Integrating in time over $[-R, t]$, for any $t \in (-r, 0)$,

$$\|w_\varepsilon(t)\|_{L^2(\mathbb{T}^2)}^2 + \int_{-R}^t \int_{\mathbb{T}^2} |\nabla(\psi w_\varepsilon)|^2 dx ds \leq \frac{C(1+q^2)}{R-r} \int_{-R}^t \|w_\varepsilon(s)\|_{L^2(\mathbb{T}^2)}^2 ds,$$

where we used that $\psi(t) = 1$ for such t and $\psi(-R) = 0$ from (5.6), $1 \leq 1+q^{-1}$, and Young's inequality to write $2q \leq 1+q^2$. Using also the fact that ψ does not depend on x and $\psi \equiv 1$ in $[-r, 0]$ to control the second term on the left-hand side of the previous inequality, as well as $0 < R-r < 1$, we deduce (5.5) as desired.

2. *Recursive estimate.* Our goal in this step is to obtain (for C independent of q, r, R, ε):

$$\|(u+\varepsilon)^{-1}\|_{L^{2q}(Q_r^-)} \leq \left(\frac{C(1+q^2)}{R-r} \right)^{\frac{1}{q}} \|(u+\varepsilon)^{-1}\|_{L^q(Q_R^-)}, \quad (5.7)$$

for all choices of $0 < r < R \leq 1$ with $R-r < 1$.

The standard interpolation inequality of Lemma 2.6 implies

$$\|w_\varepsilon\|_{L^4(Q_r^-)} \leq C \left(\|w_\varepsilon\|_{L^\infty(-r, 0; L^2(\mathbb{T}^2))} + \|\nabla w_\varepsilon\|_{L^2((-r, 0) \times \mathbb{T}^2)} \right),$$

where the positive constant C is independent of r, R, q, ε . By inserting into (5.5), we get

$$\|w_\varepsilon\|_{L^4(Q_r^-)} \leq C \left(\frac{1+q^2}{R-r} \right)^{\frac{1}{2}} \|w_\varepsilon\|_{L^2(Q_R^-)},$$

which is a *reverse Hölder inequality*, and we obtain (5.7) by rewriting the above in terms of $u + \varepsilon$, where the constant C has changed but is still independent of q, r, R, ε .

3. *Moser's iterations.* As per the statement, let $0 < r < R \leq 1$ be fixed. Define the radii

$$r_0 = R, \quad r_n := R - (R-r) \sum_{k=1}^n 2^{-k} \quad \forall n \in \mathbb{N}; \quad (5.8)$$

note that the sequence is monotonically decreasing, and $r_n \rightarrow r$ as $n \rightarrow \infty$. Furthermore,

$$r_n - r_{n+1} = (R-r)2^{-n-1} \quad \forall n \in \mathbb{N}; \quad (5.9)$$

note $0 < r_{n+1} < r_n \leq 1$ and $0 < r_n - r_{n+1} < 1$ for all n . Also define the powers $q_n := p2^n$ for all n , where we note $q_0 = p$ and $q_n \rightarrow \infty$; where $p \in (0, 1)$ is given in the statement.

By applying the result of (5.7) with $R = r_n$, $r = r_{n+1}$, $q = q_n$, we deduce

$$\|(u+\varepsilon)^{-1}\|_{L^{p2^{n+1}}(Q_{r_{n+1}}^-)} \leq \underbrace{\left(\frac{C(1+p^2 2^{2n})}{r_n - r_{n+1}} \right)^{\frac{1}{p2^n}}}_{=: \kappa_n} \|(u+\varepsilon)^{-1}\|_{L^{p2^n}(Q_{r_n}^-)} \quad \forall n \in \mathbb{N}. \quad (5.10)$$

Note that, for all n sufficiently small such that $p2^n \leq 1$, it holds $\kappa_n \leq \left(\frac{C2^{n+2}}{R-r}\right)^{\frac{1}{p2^n}} = \left(\frac{4C}{R-r}\right)^{\frac{1}{p2^n}} 2^{\frac{n}{p2^n}}$, while for all n sufficiently large such that $p2^n \geq 1$, we obtain $\kappa_n \leq \left(\frac{Cp^2 2^{3n+2}}{R-r}\right)^{\frac{1}{p2^n}} = \left(\frac{4Cp^2}{R-r}\right)^{\frac{1}{p2^n}} 2^{\frac{3n}{p2^n}}$, whence, since $p \in (0, 1)$, we deduce in either case that

$$\kappa_n \leq \left(\frac{C}{R-r}\right)^{\frac{1}{p2^n}} 2^{\frac{3n}{p2^n}} \quad \forall n \in \mathbb{N},$$

where C has changed, but remains independent of p, r, R, ε, n . By iteration on (5.10),

$$\|(u+\varepsilon)^{-1}\|_{L^{p2^n}(Q_{r_n}^-)} \leq \underbrace{\left(\prod_{k=0}^{n-1} \left(\frac{C}{R-r}\right)^{\frac{1}{p2^k}} 2^{\frac{3k}{p2^k}}\right)}_{=: A_n} \|(u+\varepsilon)^{-1}\|_{L^p(Q_R^-)} \quad \forall n \geq 1. \quad (5.11)$$

We study the convergence of the coefficients A_n . We have

$$\begin{aligned} \log A_n &= \sum_{k=0}^{n-1} \log \left[\left(\frac{C}{R-r}\right)^{\frac{1}{p2^k}} 2^{\frac{3k}{p2^k}} \right] = \frac{1}{p} \log \left(\frac{C}{R-r}\right) \sum_{k=0}^{n-1} 2^{-k} + \frac{3}{p} \log 2 \sum_{k=0}^{n-1} k 2^{-k} \\ &\longrightarrow \frac{2}{p} \log \left(\frac{C}{R-r}\right) + \frac{1}{p} \log \bar{C}, \end{aligned}$$

where $\bar{C} = 2^{3 \sum_{n \geq 0} n 2^{-n}}$ is independent of p, r, R, ε . It follows that $A_n \nearrow A = \left(\frac{C}{(R-r)^2}\right)^{\frac{1}{p}}$, for some C independent of p, r, R, ε . Passing to the limit $n \rightarrow \infty$ in (5.11) yields (5.4). $\square$

5.2. Moser upper bound. The following result is not sharp; especially in terms of the quintic power of $R-r$ in the denominator, however it is sufficient for our purposes. Again, for clarity of exposition, the estimates are written for weak solutions of (5.1) studied on the parabolic cylinders Q_r centered at $t = 0$ (cf. (2.15)); however, by translation they hold on any cylinder of the form $Q_r(t) = (t-r, t+r) \times \mathbb{T}^2$.

Lemma 5.2 (L^∞ bound $u + \varepsilon$). *Let u be a weak solution of (5.1) satisfying (5.3), $\varepsilon, p \in (0, 1)$, $0 < r < R \leq 1$. There exists a constant C independent of p, r, R, ε such that*

$$\|u + \varepsilon\|_{L^\infty(Q_r)} \leq \left(\frac{C}{(R-r)^5}\right)^{\frac{1}{p}} \|u + \varepsilon\|_{L^p(Q_R)}. \quad (5.12)$$

Proof. The proof is divided into several steps.

1. *Caccioppoli-type inequality.* Define $w_\varepsilon := (u + \varepsilon)^{q/2}$ for $q > 0$ such that $q \neq 1$, and

$$\delta := |1 - q^{-1}|. \quad (5.13)$$

Our objective in this step is to prove

$$\operatorname{ess\,sup}_{s \in (-r, r)} \|w_\varepsilon(s)\|_{L^2(\mathbb{T}^2)}^2 + \int_{-r}^r \|\nabla w_\varepsilon(s)\|_{L^2(\mathbb{T}^2)}^2 \, ds \leq C \left(\frac{1+q^2+\delta^{-1}}{R-r}\right) \int_{-R}^R \|w_\varepsilon(s)\|_{L^2(\mathbb{T}^2)}^2 \, ds. \quad (5.14)$$

Let $\psi \in C_c^\infty((-R, R))$ be such that

$$0 \leq \psi \leq 1, \quad \psi|_{[-r, r]} = 1, \quad \psi|_{[-1, 1] \setminus [-R, R]} = 0, \quad \|\partial_t \psi\|_{L^\infty((-1, 1))} \leq \frac{C}{R-r}, \quad (5.15)$$

for some constant C independent of r, R . As before, ψ is only a function of t , and again recalling the discussion that follows (5.1) and (5.3), $\phi := q(u + \varepsilon)^{q-1} \psi^2 = q w_\varepsilon^{2-\frac{2}{q}} \psi^2$ is an admissible test function. We insert ϕ into the weak formulation of (5.1) and get

$$\langle \partial_t u, \phi \rangle = q \langle \partial_t (w_\varepsilon^{\frac{2}{q}}), w_\varepsilon^{2-\frac{2}{q}} \psi^2 \rangle = 2 \langle w_\varepsilon \partial_t w_\varepsilon, \psi^2 \rangle = \frac{d}{dt} \int_{\mathbb{T}^2} |\psi w_\varepsilon|^2 \, dx - 2 \int_{\mathbb{T}^2} \psi \partial_t \psi w_\varepsilon^2 \, dx,$$

where we use (5.3), as well as $-\langle \operatorname{div}(\mathbf{p}u), \phi \rangle = 2(q-1) \int_{\mathbb{T}^2} \psi^2 \mathbf{p}(w_\varepsilon^{\frac{2}{q}} - \varepsilon) \cdot w_\varepsilon^{1-\frac{2}{q}} \nabla w_\varepsilon \, dx$, and $\langle \Delta u, \phi \rangle = -q \int_{\mathbb{T}^2} \psi^2 \nabla(w_\varepsilon^{\frac{2}{q}}) \cdot \nabla(w_\varepsilon^{2-\frac{2}{q}}) \, dx = 4(q^{-1} - 1) \int_{\mathbb{T}^2} |\nabla(\psi w_\varepsilon)|^2 \, dx$. In turn, we get

$$\begin{aligned} & -\frac{d}{dt} \int_{\mathbb{T}^2} |\psi w_\varepsilon|^2 \, dx + 4(q^{-1} - 1) \int_{\mathbb{T}^2} |\nabla(\psi w_\varepsilon)|^2 \, dx \\ & = 2(q-1) \int_{\mathbb{T}^2} \psi^2 w_\varepsilon \mathbf{p} \cdot \nabla w_\varepsilon \, dx - 2(q-1) \varepsilon \int_{\mathbb{T}^2} \psi^2 w_\varepsilon^{1-\frac{2}{q}} \mathbf{p} \cdot \nabla w_\varepsilon \, dx - 2 \int_{\mathbb{T}^2} \psi \partial_t \psi w_\varepsilon^2 \, dx; \end{aligned} \quad (5.16)$$

the second term on the right-hand side is bounded by (using also (2.7) to control $\mathbf{p}$)

$$\left| \varepsilon \int_{\mathbb{T}^2} \psi^2 w_\varepsilon^{1-\frac{2}{q}} \mathbf{p} \cdot \nabla w_\varepsilon \, dx \right| \leq \varepsilon \int_{\mathbb{T}^2} \underbrace{w_\varepsilon^{-\frac{2}{q}}}_{\leq \varepsilon^{-1}} \psi w_\varepsilon |\nabla(\psi w_\varepsilon)| \, dx \leq \int_{\mathbb{T}^2} \psi w_\varepsilon |\nabla(\psi w_\varepsilon)| \, dx.$$

The rest of the proof of (5.14) is divided into two substeps (depending on the value of q).

1.i. *Parabolic estimate: case $q \in (0, 1)$.* In this case $q^{-1} - 1 > 0$, and (5.16) yields

$$\begin{aligned} & -\frac{d}{dt} \|\psi w_\varepsilon\|_{L^2(\mathbb{T}^2)}^2 + 4(q^{-1} - 1) \|\nabla(\psi w_\varepsilon)\|_{L^2(\mathbb{T}^2)}^2 \leq 4(1-q) \int_{\mathbb{T}^2} \psi w_\varepsilon |\nabla(\psi w_\varepsilon)| \, dx + 2 \int_{\mathbb{T}^2} \psi |\partial_t \psi| w_\varepsilon^2 \, dx \\ & \leq 2q^2(q^{-1} - 1) \|\psi w_\varepsilon\|_{L^2(\mathbb{T}^2)}^2 + 2(q^{-1} - 1) \|\nabla(\psi w_\varepsilon)\|_{L^2(\mathbb{T}^2)}^2 + \frac{2C}{R-r} \|w_\varepsilon\|_{L^2(\mathbb{T}^2)}^2, \end{aligned}$$

where we used Young's inequality in the final line. Using $0 \leq \psi \leq 1$, it follows that

$$-\frac{d}{dt} \|\psi w_\varepsilon\|_{L^2(\mathbb{T}^2)}^2 + (q^{-1} - 1) \|\nabla(\psi w_\varepsilon)\|_{L^2(\mathbb{T}^2)}^2 \leq C \left(q^2(q^{-1} - 1) + \frac{1}{R-r} \right) \|w_\varepsilon\|_{L^2(\mathbb{T}^2)}^2,$$

where the positive constant C is independent of q, r, R, ε . Note that since $q \in (0, 1)$ we have $q^2(q^{-1} - 1) \leq q^2 q^{-1} = q < 1$. By rewriting the left-hand side in terms of δ and integrating in time over the interval $[t, R]$, for any $t \in (-r, r)$, we get

$$\|w_\varepsilon(t)\|_{L^2(\mathbb{T}^2)}^2 + \delta \int_t^R \int_{\mathbb{T}^2} |\nabla(\psi w_\varepsilon)|^2 \, dx \, ds \leq \frac{C}{R-r} \int_t^R \|w_\varepsilon(s)\|_{L^2(\mathbb{T}^2)}^2 \, ds,$$

where we used that $\psi(t) = 1$ for such t , $\psi(R) = 0$ from (5.15), and $0 < R - r < 1$. Thus, by considering each term on the left-hand side separately, we deduce for $q \in (0, 1)$:

$$\|w_\varepsilon(t)\|_{L^2(\mathbb{T}^2)}^2 + \int_t^R \int_{\mathbb{T}^2} |\nabla(\psi w_\varepsilon)|^2 \, dx \, ds \leq C \left(\frac{1 + \delta^{-1}}{R-r} \right) \int_t^R \|w_\varepsilon(s)\|_{L^2(\mathbb{T}^2)}^2 \, ds.$$

One deduces (5.14) easily for this range of q .

1.ii. *Parabolic estimate: case $q > 1$.* For $q > 1$, applying Young's inequality to (5.16) gives

$$\begin{aligned} & \frac{d}{dt} \int_{\mathbb{T}^2} |\psi w_\varepsilon|^2 \, dx + 4(1 - q^{-1}) \int_{\mathbb{T}^2} |\nabla(\psi w_\varepsilon)|^2 \, dx \\ & \leq 2q^2(1 - q^{-1}) \|\psi w_\varepsilon\|_{L^2(\mathbb{T}^2)}^2 + 2(1 - q^{-1}) \|\nabla(\psi w_\varepsilon)\|_{L^2(\mathbb{T}^2)}^2 + \frac{2C}{R-r} \|w_\varepsilon\|_{L^2(\mathbb{T}^2)}^2, \end{aligned}$$

and thus, rewriting in terms of δ , noting that $\delta \in (0, 1)$ in this case, and $q^2(1 - q^{-1}) = q^2\delta$,

$$\delta \frac{d}{dt} \|\psi w_\varepsilon\|_{L^2(\mathbb{T}^2)}^2 + \delta \|\nabla(\psi w_\varepsilon)\|_{L^2(\mathbb{T}^2)}^2 \leq C \left(q^2\delta + \frac{1}{R-r} \right) \|w_\varepsilon\|_{L^2(\mathbb{T}^2)}^2,$$

where the constant C is independent of q, r, R, ε . Dividing by δ and integrating in time over $[-R, t]$ for $t \in (-r, r)$, arguing as per the end of Step 1.i, we again obtain (5.14).

2. *Recursive estimate.* Using $\|w_\varepsilon\|_{L^4(Q_r)} \leq C(\|w_\varepsilon\|_{L^\infty(-r, r; L^2(\mathbb{T}^2))} + \|\nabla w_\varepsilon\|_{L^2(Q_r)})$ from Lemma 2.6 as per Step 2 of the proof of Lemma 5.1 and arguing identically, we get

$$\|u + \varepsilon\|_{L^{2q}(Q_r)} \leq \left(\frac{C(1 + q^2 + \delta^{-1})}{R-r} \right)^{\frac{1}{q}} \|u + \varepsilon\|_{L^q(Q_R)}, \quad (5.17)$$

where C is independent of q, r, R, ε . Next, as per Step 2 of the proof of [6, Lemma 3.9], set

$$\gamma := 5/3,$$

and using the Hölder inequality, we get $\|u + \varepsilon\|_{L^{\gamma q}(Q_r)} \leq |Q_r|^{(1-\frac{\gamma}{2})\frac{1}{\gamma q}} \|u + \varepsilon\|_{L^{2q}(Q_r)} \leq |Q_1|^{\frac{1}{10q}} \|u + \varepsilon\|_{L^{2q}(Q_r)}$, which, using (5.17), yields the *reverse Hölder inequality*

$$\|u + \varepsilon\|_{L^{\gamma q}(Q_r)} \leq \left(\frac{C(1 + q^2 + \delta^{-1})}{R-r} \right)^{\frac{1}{q}} \|u + \varepsilon\|_{L^q(Q_R)}, \quad (5.18)$$

where C is independent of q, r, R, ε . The reason for introducing this exponent γ is to avoid $q_n = 1$ for some n for the sequence of powers in the iteration (cf. next step of the proof).

3. *Moser's iterations.* Let $0 < r < R \leq 1$ be fixed as in the statement, and recall from (5.8) the sequence of radii $r_0 = R$, $r_n = R - (R - r) \sum_{k=1}^n 2^{-k}$, which satisfies $r_n - r_{n+1} = (R - r)2^{-n-1}$ as per (5.9). For the sequence of powers, we set

$$n_0 := 1 + \left\lfloor \frac{\log(\frac{1+\gamma}{2p})}{\log \gamma} \right\rfloor, \quad q_0 := \frac{1+\gamma}{2\gamma^{n_0}},$$

where p is given in the statement, and $q_n := q_0 \gamma^n$ for all n . The definition of n_0 implies

$$\frac{\log(\frac{1+\gamma}{2p})}{\log \gamma} < n_0 \leq 1 + \frac{\log(\frac{1+\gamma}{2p})}{\log \gamma} \implies \frac{1+\gamma}{2p} < \gamma^{n_0} \leq \gamma \frac{1+\gamma}{2p},$$

whence $q_0 < p \leq \gamma q_0$ and

$$\gamma^{n_0} q_0 = \frac{1+\gamma}{2} = \frac{4}{3} > 1 > \frac{4}{5} = \frac{1+\gamma}{2\gamma} = \gamma^{n_0-1} q_0,$$

whence $|q_n - 1| \geq \min\{q_{n_0} - 1, 1 - q_{n_0-1}\} = \frac{1}{5}$. Defining $\delta_n := |1 - q_n^{-1}|$ as per (5.13), it holds

$$\delta_n^{-1} = \frac{q_n}{|q_n - 1|} \leq 5q_n \leq \frac{5}{2}(1 + q_n^2) \quad \forall n.$$

As such, by iterating (5.18) along the sequence of radii $\{r_n\}$ and powers $\{q_n\}$, we get

$$\|u + \varepsilon\|_{L^{q_0 \gamma^{n+1}}(Q_{r_{n+1}})} \leq \underbrace{\left(\frac{C(1 + q_0^2 \gamma^{2n})}{r_n - r_{n+1}} \right)^{\frac{1}{q_0 \gamma^n}}}_{=: \kappa_n} \|u + \varepsilon\|_{L^{q_0 \gamma^n}(Q_{r_n})} \quad \forall n \in \mathbb{N}. \quad (5.19)$$

As per Step 3 of the proof of Lemma 5.1, we study the boundedness of κ_n : for all n sufficiently small such that $q_0 \gamma^n \leq 1$, it holds $\kappa_n \leq \left(\frac{C 2^{n+2}}{R-r} \right)^{\frac{1}{q_0 \gamma^n}} \leq \left(\frac{4C}{R-r} \right)^{\frac{1}{q_0 \gamma^n}} 2^{\frac{n}{q_0 \gamma^n}}$, while for all n sufficiently large such that $q_0 \gamma^n \geq 1$, we have $\kappa_n \leq \left(\frac{C q_0^2 \gamma^{2n} 2^{n+2}}{R-r} \right)^{\frac{1}{q_0 \gamma^n}} \leq \left(\frac{4C q_0^2}{R-r} \right)^{\frac{1}{q_0 \gamma^n}} 2^{\frac{3n}{q_0 \gamma^n}} \leq \left(\frac{4C}{R-r} \right)^{\frac{1}{q_0 \gamma^n}} 2^{\frac{3n}{q_0 \gamma^n}}$, where we also used $0 < q_0 < p < 1$ and $\gamma < 2$. In either case, we obtain

$$\kappa_n \leq \left(\frac{C}{R-r} \right)^{\frac{1}{q_0 \gamma^n}} 2^{\frac{3n}{q_0 \gamma^n}} \quad \forall n \in \mathbb{N},$$

where C is independent of $q_0, p, r, R, \varepsilon, n$. Returning to (5.19) and iterating,

$$\|u + \varepsilon\|_{L^{q_0 \gamma^{n+1}}(Q_{r_{n+1}})} \leq \underbrace{\left(\prod_{k=0}^n \left(\frac{C}{R-r} \right)^{\frac{1}{q_0 \gamma^k}} 2^{\frac{3k}{q_0 \gamma^k}} \right)}_{=: B_n} \|u + \varepsilon\|_{L^{q_0}(Q_R)},$$

i.e., using Hölder's inequality and $|Q_1|^{\frac{1}{q_0} - \frac{1}{p}} \leq (1 + |Q_1|)^{\frac{1}{q_0} - \frac{1}{p}}$, we get

$$\|u + \varepsilon\|_{L^{q_0 \gamma^{n+1}}(Q_{r_{n+1}})} \leq B_n |Q_1|^{\frac{1}{q_0} - \frac{1}{p}} \|u + \varepsilon\|_{L^p(Q_R)} \leq B_n (1 + |Q_1|)^{\frac{1}{q_0}} \|u + \varepsilon\|_{L^p(Q_R)}. \quad (5.20)$$

Analogously to Step 3 of the proof of Lemma 5.1, we study the convergence of $\{B_n\}_n$:

$$\log B_n = \frac{1}{q_0} \log \left(\frac{C}{R-r} \right) \sum_{k=0}^n \gamma^{-k} + \frac{3}{q_0} \log 2 \sum_{k=0}^n k \gamma^{-k} \nearrow \frac{1}{q_0} \frac{\gamma}{\gamma-1} \log \left(\frac{C}{R-r} \right) + \frac{1}{q_0} \log \bar{C},$$

with $\log \bar{C} = 3 \log 2 \sum_{n \geq 0} n \gamma^{-n}$, and thus using the continuity of the logarithm, it holds

$$B_n \nearrow B = \left(\frac{C}{(R-r)^{\frac{\gamma}{\gamma-1}}} \right)^{\frac{1}{q_0}},$$

for some C independent of q_0, R, r and can be assumed greater than 1. Returning to (5.20),

$$\|u + \varepsilon\|_{L^\infty(Q_r)} \leq \left(\frac{C}{(R-r)^{\frac{\gamma}{\gamma-1}}} \right)^{\frac{1}{q_0}} \|u + \varepsilon\|_{L^p(Q_R)} \leq \left(\frac{C}{(R-r)^{\frac{\gamma^2}{\gamma-1}}} \right)^{\frac{1}{p}} \|u + \varepsilon\|_{L^p(Q_R)}$$

using that $p \leq \gamma q_0$ and $C \geq 1 > R - r$, where C is independent of p, r, R, ε but is allowed to depend on γ . The result follows by bluntly bounding $(R-r)^{-\frac{\gamma^2}{\gamma-1}} = (R-r)^{-25/6} \leq (R-r)^{-5}$, using again that $R-r < 1$. $\square$

6. LOGARITHMIC ESTIMATES

The main result of this subsection is the following estimate on the measure of the super-/sub-level sets of the logarithm of weak solutions of (5.1). Our strategy closely follows the exposition in [6, Lemma 3.11], though our proof is quite different on account of being tailored to the space-periodic setting (as such, we do not need to localise in the x variable).

Lemma 6.1 (Super-/sub-level set estimate of the logarithm). *Let u be a weak solution of (5.1) satisfying (5.3), $\varepsilon, r \in (0, 1)$. Define*

$$w_\varepsilon := -\log(u + \varepsilon), \quad a_\varepsilon := \int_{\mathbb{T}^2} \log(u(0, x) + \varepsilon) \, dx \quad (6.1)$$

There exists a positive constant C independent of ε, r such that, for all $s > 0$, it holds

$$\left| \left\{ (t, x) \in Q_r^+ : w_\varepsilon(t, x) + a_\varepsilon > s \right\} \right| \leq \frac{C}{s}, \quad \left| \left\{ (t, x) \in Q_r^- : w_\varepsilon(t, x) + a_\varepsilon < -s \right\} \right| \leq \frac{C}{s}. \quad (6.2)$$

As per the previous subsections, for clarity of presentation our analysis is performed for weak solutions of (5.1) studied on the parabolic half-cylinder $Q_r^\pm$ centered at $t = 0$ (cf. (2.15)); however, by translation these estimates hold on any cylinder of the form $Q_r^\pm(t_0)$, with the quantity a_ε in (6.1) modified to $\int_{\mathbb{T}^2} \log(u(t_0, x) + \varepsilon) \, dx$.

Proof. The proof is divided into three steps.

1. *A differential inequality for the mean.* Recall (5.3): this regularity implies that (in view of the chain rule for Sobolev functions, cf. e.g. [12, §5.10]) the formula

$$\frac{d}{dt} \int_{\mathbb{T}^2} -\log(u + \varepsilon) \, dx = \langle \partial_t u, -(u + \varepsilon)^{-1} \rangle$$

is justified a.e. t . By testing (5.1) against $-(u + \varepsilon)^{-1}$, we get

$$\int_{\mathbb{T}^2} w_\varepsilon(t_2, x) \, dx - \int_{\mathbb{T}^2} w_\varepsilon(t_1, x) \, dx + \int_{t_1}^{t_2} \int_{\mathbb{T}^2} |\nabla w_\varepsilon|^2 \, dx \, d\tau = \int_{t_1}^{t_2} \int_{\mathbb{T}^2} \frac{u}{(u + \varepsilon)} \mathbf{p} \cdot \nabla w_\varepsilon \, dx \, d\tau.$$

for a.e. $-1 < t_1 < t_2 < 1$. Thus, by applying Young's inequality and using $\|\mathbf{p}\|_{L^\infty} \leq 1$,

$$\int_{\mathbb{T}^2} w_\varepsilon(t_2, x) \, dx - \int_{\mathbb{T}^2} w_\varepsilon(t_1, x) \, dx + \frac{1}{2} \int_{t_1}^{t_2} \int_{\mathbb{T}^2} |\nabla w_\varepsilon|^2 \, dx \, dt \leq C_*(t_2 - t_1), \quad (6.3)$$

where $C_* = \frac{1}{2}|\mathbb{T}^2|$ is independent of t_1, t_2, ε . Poincaré–Wirtinger’s inequality in $\mathbb{T}^2$ implies

$$\int_{t_1}^{t_2} \int_{\mathbb{T}^2} |w_\varepsilon(\tau, x) - \bar{w}_\varepsilon(\tau)|^2 dx d\tau \leq C_P^2 \int_{t_1}^{t_2} \int_{\mathbb{T}^2} |\nabla w_\varepsilon|^2 dx d\tau,$$

where the Poincaré constant $C_P > 0$ is independent of t_1, t_2, ε , and $\bar{w}_\varepsilon(\tau) := \int_{\mathbb{T}^2} w_\varepsilon(\tau, x) dx$ is the spatial mean. By dividing (6.3) by $|\mathbb{T}^2|$, the previous inequality implies

$$\frac{\bar{w}^\varepsilon(t_2) - \bar{w}^\varepsilon(t_1)}{t_2 - t_1} + \frac{(1/2C_P^2)}{t_2 - t_1} \int_{t_1}^{t_2} \int_{\mathbb{T}^2} |w_\varepsilon(\tau, x) - \bar{w}_\varepsilon(\tau)|^2 dx d\tau \leq C_*,$$

where the positive constant C_* is independent of t_1, t_2, ε . By letting $t_2 \rightarrow t_1$ and using the Lebesgue Differentiation Theorem, we recover the *differential inequality for the mean*:

$$\frac{d}{dt} \bar{w}_\varepsilon(t) + \frac{1}{2C_P^2} \int_{\mathbb{T}^2} |w_\varepsilon(t, x) - \bar{w}_\varepsilon(t)|^2 dx \leq C_* \quad \text{a.e. } t \in (-1, 1). \quad (6.4)$$

2. *Super-level set estimates.* Our aim in this step is to adapt the strategy of Step 1 to recover a *monotonically decreasing* quantity from the differential inequality (6.4); *i.e.* replacing the right-hand side C_* with 0 instead. To this end, we consider the function:

$$\underline{w}(t, x) := w_\varepsilon(t, x) - \bar{w}_\varepsilon(0) - C_* t, \quad (6.5)$$

where C_* is the positive constant from (6.4). We also set

$$W(t) := \bar{w}_\varepsilon(t) - \bar{w}_\varepsilon(0) - C_* t; \quad (6.6)$$

note that $W(0) = 0$, $|\underline{w}(t, x) - W(t)| = |w_\varepsilon(t, x) - \bar{w}_\varepsilon(t)|$ and $\frac{d}{dt} W(t) = \frac{d}{dt} \bar{w}_\varepsilon(t) - C_*$. Substituting these computations in (6.4), we deduce

$$\frac{d}{dt} W(t) + \frac{1}{2C_P^2} \int_{\mathbb{T}^2} |\underline{w}(t, x) - W(t)|^2 dx \leq 0 \implies W(t) \leq W(0) = 0 \quad \text{a.e. } t \in (0, 1). \quad (6.7)$$

Next, for all $s > 0, t \in (0, 1)$, define the super-level set $\Sigma_s(t)$ and $\sigma_s(t)$ by

$$\Sigma_s(t) := \{x \in \mathbb{T}^2 : \underline{w}(t, x) > s\}, \quad \sigma_s(t) := \{x \in \mathbb{T}^2 : w_\varepsilon(t, x) - \bar{w}_\varepsilon(0) > s\}; \quad (6.8)$$

our strategy will be to use estimates on $\Sigma_s(t)$ to control the measure of $\sigma_s(t)$. Indeed, for all $t \in (0, 1)$, we verify directly the inclusions

$$\Sigma_s(t) \subset \sigma_s(t) \quad \forall s > 0, \quad \sigma_{2s}(t) \subset \Sigma_s(t) \quad \forall s \geq C_*. \quad (6.9)$$

Using (6.7), direct computation (using that $W(t) \leq 0$) yields

$$\underline{w}(t, x) - W(t) > s - W(t) \geq s > 0 \quad \text{for all } x \in \Sigma_s(t). \quad (6.10)$$

Furthermore, using also $\frac{d}{dt} W(t) = -\frac{d}{dt}(s - W(t))$ and (6.7), we get

$$\frac{1}{2C_P^2 |\mathbb{T}^2|} \int_{\Sigma_s(t)} (s - W(t))^2 dx \leq \frac{d}{dt}(s - W(t)) \implies \frac{d}{dt}(s - W(t)) \geq \frac{1}{2C_P^2 |\mathbb{T}^2|} |\Sigma_s(t)| (s - W(t))^2.$$

We integrate the final inequality in t over the interval $[0, r]$ and get

$$|\{(t, x) \in Q_r^+ : \underline{w}(t, x) > s\}| = \int_0^r |\Sigma_s(t)| dt \leq \frac{2C_P^2 |\mathbb{T}^2|}{s - W(0)} - \frac{2C_P^2 |\mathbb{T}^2|}{s - W(r)} \leq \frac{2C_P^2 |\mathbb{T}^2|}{s},$$

where we used $W(0) = 0, s - W(r) > 0$. By the above and (6.9), we get

$$|\{(t, x) \in Q_r^+ : w_\varepsilon(t, x) - \bar{w}_\varepsilon(0) > s\}| \leq |\{(t, x) : \underline{w}(\tau, x) > \frac{s}{2}\}| \leq \frac{4C_P^2 |\mathbb{T}^2|}{s} \quad \forall s \geq C_*, \quad (6.11)$$

while, for $0 < s < C_*$, we estimate bluntly by the measure of the whole space

$$|\{(t, x) \in Q_r^+ : w_\varepsilon(t, x) - \bar{w}_\varepsilon(0) > s\}| \leq |Q_1^+| \frac{s}{s} < \frac{|Q_1| C_*}{s} \quad \forall s \in (0, C_*). \quad (6.12)$$

By putting together the bounds of (6.11)–(6.12), we obtain the super-level set estimate

$$|\{(t, x) \in Q_r^+ : w_\varepsilon(t, x) - \bar{w}_\varepsilon(0) > s\}| \leq \frac{C}{s} \quad \forall s > 0,$$

where the positive constant $C = \max\{|Q_1|C_*, 4C_P^2|\mathbb{T}^2|\}$ is independent of ε, r . Rewriting in terms of a_ε as per (6.1) yields the first part of (6.2).

3. *Sub-level set estimates.* As per the final step of [6, Lemma 3.11], rewriting only Step 2 of the proof with $-t$ instead of t yields the second part of (6.2). We write out several details for the benefit of the reader. In this latter case, it holds $W(t) \geq 0$ for all $t \in (-1, 0)$ with W as per (6.6), and the inclusions of (6.9) are again satisfied, but with re-defining the sets in (6.8) as $\Sigma_s(t) = \{x : \underline{w}(t, x) < -s\}$ and $\sigma_s(t) = \{x : w_\varepsilon(t, x) - \bar{w}_\varepsilon(0) < -s\}$, and $\underline{w}$ still defined as per (6.5). Then, we notice that for all $x \in \Sigma_s(t)$, it holds

$$W(t) - \underline{w}(t, x) > W(t) + s \geq s > 0,$$

and thus, arguing as before, we obtain from the differential inequality of (6.7) that

$$\frac{d}{dt}W(t) + \frac{1}{2C_P^2|\mathbb{T}^2|} \int_{\Sigma_s(t)} (W(t) + s)^2 dx \leq 0 \implies \frac{1}{2C_P^2|\mathbb{T}^2|} |\Sigma_s(t)| \leq \frac{d}{dt}(W(t) + s)^{-1},$$

and thus $|\{(t, x) \in Q_r^- : \underline{w}(t, x) < -s\}| = \int_{-r}^0 |\Sigma_s(t)| dt \leq \frac{2C_P^2|\mathbb{T}^2|}{W(0)+s} - \frac{2C_P^2|\mathbb{T}^2|}{W(-r)+s} \leq \frac{2C_P^2|\mathbb{T}^2|}{s}$, and the required conclusion follows as per the end of Step 2 of the proof. $\square$

7. QUANTITATIVE INSTANTANEOUS SEPARATION FROM $\{\rho = 1\}$

The main result of this section is the following; it is concerned with the non-degeneracy of $1 - \rho$, which is at the heart of the regularity argument §8.

Proposition 7.1 (Quantitative instantaneous non-degeneracy). *Let f be a weak solution of (1.1) with admissible initial data f_0 , $\rho = \int_{\mathbb{T}} f d\theta$. Then, there exists a positive constant $C > 1$ (independent of t, T , the initial data, and the solution) such that*

$$\operatorname{ess\,inf}_{(t,T) \times \mathbb{T}^2} (1 - \rho) \geq C^{-1} \|1 - \rho_0\|_{L^1(\mathbb{T}^2)} e^{-C(T + (1 \wedge t)^{-10})}, \quad \forall 0 < t < T < \infty. \quad (7.1)$$

We highlight that the lower bound (3.1) in Theorem 3.1 follows immediately from (7.1), after enlarging the constant C appropriately since $1 \geq e^{-Ct^{-10}}$; changing the factor C^{-1} in (7.1) to $C^{-1}e^{-C}$ suffices. Hence, once Proposition 7.1 is established, so is (3.1).

7.1. Harnack estimate for small times. The following is a precursor to Proposition 7.1 concerned only with small times. Note the cylinders in (7.2) are separated by a *gap*.

Lemma 7.2 (Harnack estimate for small times). *Let u be a weak solution of (5.1) on $(0, \infty) \times \mathbb{T}^2$ satisfying (5.3). Then, there exists $C > 0$, independent of t and u , such that*

$$\operatorname{ess\,inf}_{(\frac{5}{4}t, \frac{3}{2}t) \times \mathbb{T}^2} u \geq e^{-Ct^{-10}} \operatorname{ess\,sup}_{(\frac{1}{4}t, \frac{3}{4}t) \times \mathbb{T}^2} u \quad \forall t \in (0, 1). \quad (7.2)$$

Proof. The proof is divided into several steps. We begin by fixing $t_0 \in (0, 1)$, setting $a_\varepsilon = \int_{\mathbb{T}^2} \log(u(t_0, x) + \varepsilon) dx$ from (6.1) and defining the non-negative measurable functions

$$v_\varepsilon^-(t, x) := e^{a_\varepsilon} (u(t, x) + \varepsilon)^{-1}, \quad v_\varepsilon^+(t, x) := e^{-a_\varepsilon} (u(t, x) + \varepsilon).$$

1. *Upper bound on the supremum.* By translating the relevant cylinders and multiplying by e^{-a_ε} (which is justified since $e^{-a_\varepsilon} u(\frac{t_0}{2} + t, x)$ is also a solution of (5.1) when u is), the conclusion of Lemma 5.2 holds in the form: for all $\varepsilon, p \in (0, 1)$, and $0 < r < R \leq \frac{t_0}{2}$, there exists a positive constant C independent of $p, r, R, \varepsilon, t_0$ such that

$$\|v_\varepsilon^+\|_{L^\infty(Q_r(\frac{t_0}{2}))} \leq \left(\frac{C}{(R-r)^5} \right)^{\frac{1}{p}} \|v_\varepsilon^+\|_{L^p(Q_R(\frac{t_0}{2}))}.$$

It therefore follows that v_ε^+ satisfies the first assumption of the Bombieri–Giusti Lemma 4.1. Furthermore, for all $s > 0$ and all $r < t_0/2$, using the notation $w_\varepsilon(t, x) = -\log(u(t, x) + \varepsilon)$,

$$\begin{aligned} |\{(t, x) \in Q_r(\frac{t_0}{2}) : \log v_\varepsilon^+(t, x) > s\}| &= |\{(t, x) \in Q_r(\frac{t_0}{2}) : w_\varepsilon(t, x) + a_\varepsilon < -s\}| \\ &\leq |\{(t, x) \in Q_{t_0}^-(t_0) : w_\varepsilon(t, x) + a_\varepsilon < -s\}| \leq \frac{C}{s}, \end{aligned}$$

by applying Lemma 6.1 with the choice $r = t_0$ on the translated cylinder at time $t = t_0$ (note that a_ε corresponds precisely to $\int_{\mathbb{T}^2} \log(u(t_0, x) + \varepsilon) dx$ so that Lemma 6.1 applies with the cylinder centered at t_0), where the positive constant C is independent of ε, s, r, t_0 , and where we used the identification of the cylinders $Q_r(\frac{t_0}{2}) = (\frac{t_0}{2} - r, \frac{t_0}{2} + r) \times \mathbb{T}^2 \subset (0, t_0) \times \mathbb{T}^2 = Q_{t_0}^-(t_0)$ for all $r < t_0/2$. The second condition of the Bombieri–Giusti Lemma 4.1 is therefore also verified for v_ε^+ for the family of nested sets $\{Q_r = Q_r(\frac{t_0}{2})\}_{r < t_0/2}$, whence an application of this result (taking $\beta = 5$ and $\tilde{r} = t_0/2$) implies $\|v_\varepsilon^+\|_{L^\infty(Q_r(t_0))} \leq e^{\bar{C}_1(\frac{t_0}{2}-r)^{-10}}$ for all $0 < r < \frac{t_0}{2}$, where the positive constant $\bar{C}_1$ is independent of ε, r, t_0 . Thus,

$$\varepsilon + \operatorname{ess\,sup}_{Q_r(\frac{t_0}{2})} u \leq e^{a_\varepsilon} e^{\bar{C}_1(\frac{t_0}{2}-r)^{-10}} \quad \forall 0 < r < \frac{t_0}{2}. \quad (7.3)$$

2. *Lower bound on the infimum.* Similarly, by multiplying the conclusion of Lemma 5.1 with e^{a_ε} and centering the cylinders at $t = 3t_0/2$ by translation, we get

$$\|v_\varepsilon^-\|_{L^\infty(Q_r^-(\frac{3}{2}t_0))} \leq \left(\frac{C}{(R-r)^2} \right)^{\frac{1}{p}} \|v_\varepsilon^-\|_{L^p(Q_R^-(\frac{3}{2}t_0))} \quad \forall 0 < r < R \leq \frac{3}{2}t_0,$$

for all $p \in (0, 1)$, with C again independent of $\varepsilon, r, R, p, t_0$. Meanwhile, another application of Lemma 6.1 (this time with the choice $r = t_0/2$ and having translated the cylinder to $t = t_0$) implies for all $s > 0$ and all $r < t_0/2$, with the notation $w_\varepsilon(t, x) = -\log(u(t, x) + \varepsilon)$,

$$\begin{aligned} |\{(t, x) \in Q_r^-(\frac{3}{2}t_0) : \log v_\varepsilon^-(t, x) > s\}| &= |\{(t, x) \in Q_r^-(\frac{3}{2}t_0) : w_\varepsilon(t, x) + a_\varepsilon > s\}| \\ &\leq |\{(t, x) \in Q_{t_0}^+(t_0) : w_\varepsilon(t, x) + a_\varepsilon > s\}| \leq \frac{C}{s}, \end{aligned}$$

where the positive constant C is independent of ε, s, r, t_0 , where we used the identification of the cylinders $Q_r^-(\frac{3}{2}t_0) = (\frac{3}{2}t_0 - r, \frac{3}{2}t_0) \times \mathbb{T}^2 \subset (t_0, \frac{3}{2}t_0) = Q_{t_0/2}^+(t_0)$ for all $r < t_0/2$. The second condition of the Bombieri–Giusti Lemma 4.1 is thus also fulfilled by v_ε^- for the family of nested sets $\{Q_r = Q_r^-(\frac{3}{2}t_0)\}_{r < t_0/2}$. Applying this result with the choices $\beta = 2$ and $\tilde{r} = t_0/2$ implies $\|v_\varepsilon^-\|_{L^\infty(Q_r^-(\frac{3}{2}t_0))} \leq e^{\bar{C}_2(\frac{t_0}{2}-r)^{-4}}$ for all $0 < r < \frac{t_0}{2}$, where the positive constant $\bar{C}_2$ is independent of ε, r, t_0 , from which we deduce

$$\varepsilon + \operatorname{ess\,inf}_{Q_r^-(\frac{3}{2}t_0)} u \geq e^{a_\varepsilon} e^{-\bar{C}_2(\frac{t_0}{2}-r)^{-4}} \quad \forall 0 < r < \frac{t_0}{2}. \quad (7.4)$$

3. *Harnack inequality.* Choosing $r = t_0/4$ in (7.3) and (7.4), we get (with $C'_1 = 4^{10}\bar{C}_1$, $C'_2 = 4^4\bar{C}_2$), the bound $(\varepsilon + \operatorname{ess\,inf}_{Q_{t_0/4}^-(\frac{3}{2}t_0)} u) e^{C'_2 t_0^{-4}} \geq e^{a_\varepsilon} \geq (\varepsilon + \operatorname{ess\,sup}_{Q_{t_0/4}(\frac{t_0}{2})} u) e^{-C'_1 t_0^{-10}}$. Setting $C = C'_1 + C'_2$, which is independent of ε, t_0 , we deduce by letting $\varepsilon \rightarrow 0^+$

$$\operatorname{ess\,inf}_{Q_{t_0/4}^-(\frac{3}{2}t_0)} u \geq e^{-C t_0^{-10}} \operatorname{ess\,sup}_{Q_{t_0/4}(\frac{t_0}{2})} u,$$

where we also used $0 < t_0 < 1$. Rewriting the cylinders and setting $t = t_0$ yields (7.2). $\square$

We readily deduce the following corollary, where the previous lower bound (in terms of the supremum) in the Harnack inequality is replaced by the L^1 norm of the initial data.

Corollary 7.3. *Let f be a weak solution of (1.1) with admissible initial data f_0 , $\rho = \int_{\mathbb{T}} f d\theta$, and $u = 1 - \rho$. Then, there exists a positive constant C (independent of t , the initial data, and the solution), such that*

$$\operatorname{ess\,inf}_{(\frac{5}{4}t, \frac{3}{2}t) \times \mathbb{T}^2} u \geq C^{-1} \|1 - \rho_0\|_{L^1(\mathbb{T}^2)} e^{-Ct^{-10}} \quad \forall t \in (0, 1). \quad (7.5)$$

Proof. Recall that (5.1) conserves its initial mass, i.e. directly from (A.1) with $\varphi = 1$, we have $\|1 - \rho\|_{L^1(\mathbb{T}^2)} = \int_{\mathbb{T}^2} u dx = \int_{\mathbb{T}^2} u_0 dx = \|1 - \rho_0\|_{L^1(\mathbb{T}^2)}$ a.e. $t > 0$. It therefore follows, using also Lemma 7.2, that for a.e. $t \in (0, 1)$ it holds $\|1 - \rho_0\|_{L^1(\mathbb{T}^2)} = \int_{\mathbb{T}^2} u(\frac{t}{2}, x) dx \leq |\mathbb{T}|^2 \operatorname{ess\,sup}_{(\frac{1}{4}t, \frac{3}{4}t) \times \mathbb{T}^2} u \leq |\mathbb{T}|^2 e^{Ct^{-10}} \operatorname{ess\,inf}_{(\frac{5}{4}t, \frac{3}{2}t) \times \mathbb{T}^2} u$, and the result follows. $\square$

7.2. Propagation of non-degeneracy. In this portion of the manuscript, we prove Proposition 7.1. We record firstly the following propagation (in time) of lower-boundedness for $1 - \rho$, which closely follows [1, Lemmas 5.3–5.4].

Lemma 7.4 (Propagation of boundedness away from $\{\rho = 1\}$). *Let f be a weak solution of (1.1) with admissible initial data f_0 , $\rho = \int_{\mathbb{T}} f d\theta$, and $u = 1 - \rho$. Fix any Lebesgue point $t_0 \in (0, 1)$ of $t \mapsto \operatorname{ess\,inf}_{x \in \mathbb{T}^2} u(t, x)$. Then, there exists a strictly positive constant C (independent of t_0, t , the initial data, and the solution) such that*

$$\operatorname{ess\,inf}_{(t_0, t) \times \mathbb{T}^2} u \geq C^{-1} e^{-Ct} \operatorname{ess\,inf}_{x \in \mathbb{T}^2} u(t_0, x) \quad \forall t > t_0.$$

This result relies on an iteration procedure which is different from the (Moser-style) ones presented in §5. In [1, Lemma 5.4], the final estimate was closed using a technical lemma borrowed from [14]. In order to make the paper self-contained, and because we require a more accurate quantitative estimate than that of [14, Lemma 3.2], we record the following iterative lemma (the proof is postponed to Appendix A).

Lemma 7.5 (Quantitative iteration lemma). *Let $\{A_k\}_{k \geq 0}$ be a sequence of non-negative differentiable functions satisfying the differential recurrence inequality on $[0, t]$:*

$$\frac{dA_k}{dt} + A_k \leq C_*^k (A_{k-1})^2 \quad \forall k \geq 1, \quad (7.6)$$

where $C_* \geq 2$ is independent of k . Let $n_k = 2^k$, and suppose there exists $M_1, M_2 > 0$ such that $\sup_{k \geq 0} A_k(0)^{1/n_k} \leq M_1$ and $\|A_0\|_{L^\infty([0, t])} \leq M_2$. Then, with $\beta = 1 + \sum_{j=1}^\infty (j+1)2^{-j}$,

$$\sup_{k \geq 0} \|A_k\|_{L^\infty([0, t])}^{1/n_k} \leq C_*^\beta (M_1 + M_2). \quad (7.7)$$

We highlight that the constant C_* in the previous lemma has no relation to the constant bearing the same name in the proof of Lemma 6.1. We are ready to prove Lemma 7.4.

Proof of Lemma 7.4. Fix $t_0 \in (0, 1)$ as in the statement. By Corollary 7.3, it holds $\operatorname{ess\,inf}_{x \in \mathbb{T}^2} u(t_0, x) > 0$ provided $\|1 - \rho_0\|_{L^1(\mathbb{T}^2)} > 0$, which we assume. In other words, defining $v := 1/u$ and $v_0 = v(t_0, \cdot)$,

$$\|v_0\|_{L^\infty(\mathbb{T}^2)} = \left(\operatorname{ess\,inf}_{x \in \mathbb{T}^2} u(t_0, x) \right)^{-1} < \infty. \quad (7.8)$$

We will now rewrite the equation (5.1) as an evolution on v , and then show by an iteration procedure that v is bounded for all subsequent times by a quantity depending on $\|v_0\|_{L^\infty}$.

1. *Equation for v .* By formal manipulations, we expect the equation for v to be $\partial_t v + \operatorname{div}(\mathbf{p}v) - 2\mathbf{p} \cdot \nabla v = \Delta v - 2 \frac{|\nabla v|^2}{v}$, i.e., we expect v to be a weak subsolution of

$$\partial_t v + \operatorname{div}(\mathbf{p}v) - 2\mathbf{p} \cdot \nabla v = \Delta v. \quad (7.9)$$

Note however that *a priori* it is not even clear that $v \in L_{\text{loc}}^2(0, \infty; H^1(\mathbb{T}^2))$ for $t \geq t_0$; the purpose of this step of the proof is to verify this rigorously.

Equation (5.1) is only satisfied in the weak sense dual to $L^2_{\text{loc}}(0, \infty; H^1(\mathbb{T}^2))$. To this end, we cannot test (5.1) directly with functions of v itself, but instead we consider the regularised version $v_\varepsilon := 1/(u + \varepsilon) \in L^2_{\text{loc}}(0, \infty; H^1(\mathbb{T}^2)) \cap L^\infty((0, \infty) \times \mathbb{T}^2)$ by Definition 2.2 (cf. also (5.3)). We test (5.1) against the admissible test function $-v_\varepsilon^3$ (since, formally, $\frac{1}{2}\partial_t(v_\varepsilon^2) = v_\varepsilon\partial_tv_\varepsilon = -(u + \varepsilon)^3\partial_tu = -v_\varepsilon^3\partial_tu$), and get

$$\langle \partial_t u, -v_\varepsilon^3 \rangle - 3 \int_{\mathbb{T}^2} v_\varepsilon^2 \nabla v_\varepsilon \cdot \mathbf{p} u \, dx = 3 \int_{\mathbb{T}^2} v_\varepsilon^2 \nabla v_\varepsilon \cdot \nabla u \, dx,$$

using (5.2) and the chain/product rules for Sobolev functions belonging to $H^1(\mathbb{T}^2)$ for a.e. t , cf. e.g. [12, §5.10]. Then, writing $u = v_\varepsilon^{-1} - \varepsilon$ and again using the chain rule for Sobolev functions to write $\nabla u = -v_\varepsilon^{-2} \nabla v_\varepsilon$ (here we also use that $v_\varepsilon^{-2} = (u + \varepsilon)^2 \leq (1 + \varepsilon)^2 \in L^\infty$),

$$\langle \partial_t u, -v_\varepsilon^3 \rangle - 3 \int_{\mathbb{T}^2} v_\varepsilon \nabla v_\varepsilon \cdot \mathbf{p} \, dx + 3\varepsilon \int_{\mathbb{T}^2} v_\varepsilon^2 \nabla v_\varepsilon \cdot \mathbf{p} \, dx = -3 \int_{\mathbb{T}^2} |\nabla v_\varepsilon|^2 \, dx.$$

Using that $(1 + \varepsilon)^{-1} \leq v_\varepsilon \leq \varepsilon^{-1} \in L^\infty$, $v_\varepsilon \in L^2_{\text{loc}}(0, \infty; H^1(\mathbb{T}^2))$, and $\partial_t u \in L^{3/2}_{\text{loc}}([t_0, \infty) \times \mathbb{T}^2)$ from (5.2), the chain rule for Sobolev functions and the Lions–Magenes Lemma [21, Ch.3, §1.4 Lem.1.2] yield $\frac{1}{2} \frac{d}{dt} \|v_\varepsilon\|_{L^2(\mathbb{T}^2)}^2 = \langle \partial_t v_\varepsilon, v_\varepsilon \rangle = -\langle v_\varepsilon^2 \partial_t u, v_\varepsilon \rangle = \langle \partial_t u, -v_\varepsilon^3 \rangle$, whence (using the boundedness of $\mathbf{p}$ (2.7), $\varepsilon v_\varepsilon^2 \leq v_\varepsilon$, and Hölder’s inequality)

$$\frac{1}{2} \frac{d}{dt} \|v_\varepsilon\|_{L^2(\mathbb{T}^2)}^2 + 3 \|\nabla v_\varepsilon\|_{L^2(\mathbb{T}^2)}^2 \leq 6 \|v_\varepsilon\|_{L^2(\mathbb{T}^2)} \|\nabla v_\varepsilon\|_{L^2(\mathbb{T}^2)}.$$

Applying Young’s inequality and then Grönwall’s Lemma (integrating over $[t_0, t]$) implies

$$\text{ess sup}_{s \in [t_0, t]} \|v_\varepsilon(s, \cdot)\|_{L^2(\mathbb{T}^2)}^2 + \|\nabla v_\varepsilon\|_{L^2((t_0, t) \times \mathbb{T}^2)}^2 \leq C \|v_\varepsilon(t_0, \cdot)\|_{L^2(\Omega)}^2 e^{Ct} \leq C \|v_0\|_{L^\infty}^2 e^{Ct}, \quad (7.10)$$

for a.e. $t > t_0$, where we used (7.8) and $v_\varepsilon \leq v$, and the constant C changed in the final inequality (but is independent of t_0, t and the initial data). Letting $\varepsilon \rightarrow 0^+$ and using the lower semicontinuity of the norm, we get $v \in L^2_{\text{loc}}([t_0, \infty); H^1(\mathbb{T}^2)) \cap L^\infty_{\text{loc}}([t_0, \infty); L^2(\mathbb{T}^2))$.

It remains to write down the equation for v . By the aforementioned applications of the chain rule, it holds $\partial_t v_\varepsilon = -v_\varepsilon^2 \partial_t u$ with $\partial_t u$ integrable (5.3) and $\nabla v_\varepsilon = -v_\varepsilon^2 \nabla u$. It follows (using the previous computations) that, in the weak sense dual to $L^2_{\text{loc}}([t_0, \infty); H^1(\mathbb{T}^2))$,

$$\partial_t v_\varepsilon + \text{div}(\mathbf{p} v_\varepsilon (1 - \varepsilon v_\varepsilon)) - 2\mathbf{p} \cdot \nabla v_\varepsilon (1 - \varepsilon v_\varepsilon) = \Delta v_\varepsilon - 2 \frac{|\nabla v_\varepsilon|^2}{v_\varepsilon}. \quad (7.11)$$

Letting $\varepsilon \rightarrow 0^+$, using the Dominated Convergence Theorem with $v_\varepsilon = \frac{v}{1 + \varepsilon v} \rightarrow v$ a.e. and the regularity $v \in L^2_{\text{loc}}([t_0, \infty); H^1(\mathbb{T}^2))$, we get that v is a weak subsolution of (7.9).

2. *Boundedness of v .* We argue very closely to [1, Lemma 5.4] using the equation (7.11). We test the equation (7.11) with v_ε^{n-1} for $n \geq 2$ and obtain

$$\frac{d}{dt} \int_{\mathbb{T}^2} |v_\varepsilon^{\frac{n}{2}}|^2 \, dx + 4(1 + \frac{1}{n}) \int_{\mathbb{T}^2} |\nabla(v_\varepsilon^{\frac{n}{2}})|^2 \, dx = 2(n+1) \int_{\mathbb{T}^2} (1 - \varepsilon v_\varepsilon) v_\varepsilon^{\frac{n}{2}} \nabla(v_\varepsilon^{\frac{n}{2}}) \cdot \mathbf{p} \, dx,$$

whence applying Young’s inequality (noting also that $0 < v_\varepsilon \leq \varepsilon^{-1}$ so that $|1 - \varepsilon v_\varepsilon| \leq 1$),

$$\frac{d}{dt} \|v_\varepsilon^{\frac{n}{2}}\|_{L^2(\mathbb{T}^2)}^2 + 3 \|\nabla(v_\varepsilon^{\frac{n}{2}})\|_{L^2(\mathbb{T}^2)}^2 \leq 2n^2 \|v_\varepsilon^{\frac{n}{2}}\|_{L^2(\mathbb{T}^2)}^2. \quad (7.12)$$

Next, an application of the Gagliardo–Nirenberg inequality implies

$$\|v_\varepsilon^{\frac{n}{2}}\|_{L^2(\mathbb{T}^2)} \leq C_{GN} \left(\|\nabla(v_\varepsilon^{\frac{n}{2}})\|_{L^2(\mathbb{T}^2)}^{1/2} \|v_\varepsilon^{\frac{n}{2}}\|_{L^1(\mathbb{T}^2)}^{1/2} + \|v_\varepsilon^{\frac{n}{2}}\|_{L^1(\mathbb{T}^2)} \right);$$

note that, by squaring both sides and using Young’s inequality, the above also gives $\|\nabla v_\varepsilon^{\frac{n}{2}}\|_{L^2(\mathbb{T}^2)}^2 \geq C_1 \|v_\varepsilon^{\frac{n}{2}}\|_{L^2(\mathbb{T}^2)}^2 - C_2 \|v_\varepsilon^{\frac{n}{2}}\|_{L^1(\mathbb{T}^2)}^2$ for some positive constants C_1, C_2 depending only on C_{GN} . In turn, substituting into the right-hand side of (7.12) and using Young’s

inequality again yields $\frac{d}{dt} \|v_\varepsilon^{\frac{n}{2}}\|_{L^2(\mathbb{T}^2)}^2 + 2\|\nabla(v_\varepsilon^{\frac{n}{2}})\|_{L^2(\mathbb{T}^2)}^2 \leq (4n^4 C_{GN}^4 + 4n^2 C_{GN}^2) \|v_\varepsilon^{\frac{n}{2}}\|_{L^1(\mathbb{T}^2)}^2$,
i.e., using also $\|v_\varepsilon^{\frac{n}{2}}\|_{L^2(\mathbb{T}^2)}^2 = \|v_\varepsilon^n\|_{L^1(\mathbb{T}^2)}$,

$$\frac{d}{dt} \|v_\varepsilon^n\|_{L^1(\mathbb{T}^2)} + \|v_\varepsilon^n\|_{L^1(\mathbb{T}^2)} \leq Cn^4 \|v_\varepsilon^{\frac{n}{2}}\|_{L^1(\mathbb{T}^2)}^2,$$

where the positive constant C depends only on C_{GN} , is independent of ε, n, t_0 , and can be assumed greater than 1. By setting $n_k := 2^k$ and $A_k(t) := \|v_\varepsilon^{n_k}(t + t_0, \cdot)\|_{L^1(\mathbb{T}^2)}$, we obtain, with $C_* := 16C$ which can be assumed to verify $C_* \geq 2$, the recurrence inequality

$$\frac{dA_k}{dt} + A_k \leq C_*^k (A_{k-1})^2,$$

which is precisely (7.6). Using the estimate (7.10) and

$$\begin{aligned} \sup_k A_k(0)^{1/n_k} &= \sup_k \|v_\varepsilon(t_0)\|_{L^{n_k}(\mathbb{T}^2)} = \|v_\varepsilon(t_0)\|_{L^\infty(\mathbb{T}^2)} \leq \|v_0\|_{L^\infty(\mathbb{T}^2)} =: M_1, \\ \|A_0\|_{L^\infty([t_0, t])} &= \operatorname{ess\,sup}_{s \in [t_0, t]} \|v_\varepsilon(s)\|_{L^1(\mathbb{T}^2)} \leq |\mathbb{T}^2|^{\frac{1}{2}} \operatorname{ess\,sup}_{s \in [t_0, t]} \|v_\varepsilon(s)\|_{L^2(\mathbb{T}^2)} \leq |\mathbb{T}^2|^{\frac{1}{2}} \|v_0\|_{L^\infty} e^{Ct} =: M_2, \end{aligned}$$

where we also used Hölder's inequality in the second line (and where the constant C has changed but remains independent of ε, t_0, t , and the initial data), Lemma 7.5 yields

$$\|v_\varepsilon\|_{L^\infty([t_0, t] \times \mathbb{T}^2)} \leq \sup_k \|A_k\|_{L^\infty([0, t])}^{1/n_k} \leq C(M_1 + M_2) \leq C\|v_0\|_{L^\infty} e^{Ct},$$

for some positive constant C independent of ε, t_0, t , the initial data, and the solution. As a result, by letting $\varepsilon \rightarrow 0^+$ and using that $v_\varepsilon = \frac{v}{1+\varepsilon v}$ so that $v_\varepsilon \rightarrow v$ a.e., we get

$$\operatorname{ess\,inf}_{(t_0, t)} u \geq C^{-1} e^{-Ct} \|v_0\|_{L^\infty}^{-1} \quad \forall t > t_0,$$

which, by rewriting in terms of $u(t_0, \cdot)$ as per (7.8), yields the required result. $\square$

Proposition 7.1 now follows immediately from Corollary 7.3 and Lemma 7.4, up to modifying the constant C ; we omit the details for clarity of presentation.

8. INSTANTANEOUS SMOOTHING

Having obtained the strictly positive lower bound (3.1) on $1 - \rho$ (in Proposition 7.1), the proof of the instantaneous smoothing effect encapsulated in Theorem 1 now follows by the same strategy as §6-7 of [1]. For completeness, and to make the present work self-contained, we write the arguments in detail. We also incorporate the corrections to [1] provided in [2], and provide a refined treatment with clearer proofs. Finally, we remark that the higher regularity bootstrap (§8.3) is written differently to the corresponding section in [1] (§7 therein); indeed, here we appeal to Schauder theory in Hölder continuous spaces, while in [1, §7] this bootstrap was written in terms of Sobolev spaces.

Throughout this section, f will always denote a weak solution of (1.1) with admissible initial data, $\rho = \int_{\mathbb{T}} f \, d\theta$, $\mathbf{p} = \int_{\mathbb{T}} f \mathbf{e}(\theta) \, d\theta$. Furthermore, given $0 < t < T < \infty$, the letter $c_{t,T}$ will always be used to denote positive lower bound of (3.1), given explicitly by

$$0 < c_{t,T} := C^{-1} \|1 - \rho_0\|_{L^1(\mathbb{T}^2)} e^{-C(T+t^{-10})}, \quad (8.1)$$

where we recall that C above is independent of t, T , the initial data, and the solution.

8.1. Local boundedness of f . The goal of this section is to prove the following result.

Proposition 8.1 (f is locally bounded). *For all $0 < t < T < \infty$, it holds $f \in L^\infty((t, T) \times \mathbb{T}^3)$.*

We divide the analysis leading up to the proof of Proposition 8.1 into several subsections.

8.1.1. *Initial upgrade of regularity for $\mathbf{p}$.* Our first result is concerned with the H^1 -type estimate for $\mathbf{p}$. We remind the reader that (2.8) holds in duality with $L^3_{\text{loc}}([0, \infty); W^{1,3}(\mathbb{T}^2))$; this can be seen easily from the boundedness $(1-\rho)\nabla\mathbf{p} \in L^{3/2}_{\text{loc}}([0, \infty); L^{3/2}(\mathbb{T}^2))$ from (2.10), which yields $\partial_t\mathbf{p} \in L^{3/2}_{\text{loc}}([0, \infty); W^{-1,3/2}(\mathbb{T}^2))$ from reading (2.8) in the sense of distributions.

Lemma 8.2 (H^1 estimate for $\mathbf{p}$, L^4 bound on $\nabla\rho$). *For all $0 < t < T < \infty$, it holds*

$$\nabla\mathbf{p} \in L^2((t, T) \times \mathbb{T}^2), \quad \partial_t\mathbf{p} \in L^2((t, T); H^{-1}(\mathbb{T}^2)), \quad \nabla\rho \in L^4((t, T) \times \mathbb{T}^2).$$

We note in passing that a version of this result was obtained in [1, Lemmas 4.2-4.3], and that a minor correction should be made therein: one should assume the lower bound $\text{ess inf}_{(t,T) \times \mathbb{T}^2} (1-\rho) \geq c_{t,T} > 0$ in the statement.

Proof. We cannot test (2.8) with $\mathbf{p}$, since it is not known to be an admissible test function. Instead, we consider smoothed approximations of $\mathbf{p}$ and argue similarly to [1, Lemma 4.2].

1. *Approximation.* Let η be the usual Friedrichs bump function, and $\{\eta^\varepsilon\}_{\varepsilon>0}$ the usual sequence of Friedrichs mollifiers in $\mathbb{R}^2$; we only smooth in the x variable. Let $\mathbf{p}^\varepsilon = \eta^\varepsilon * \mathbf{p}$, and define $D = (1-\rho)\mathbf{p}$ and its approximation $D^\varepsilon = \eta^\varepsilon * D$. We insert $\eta^\varepsilon(s, x-y)$ as test function in (2.8) and integrate in y over $\mathbb{R}^2$ —this is the same as mollifying (2.8)—and get

$$\partial_t\mathbf{p}^\varepsilon + \text{div } D^\varepsilon = \text{div}((1-\rho)\nabla\mathbf{p}^\varepsilon + \mathbf{p}^\varepsilon \otimes \nabla\rho) - \mathbf{p}^\varepsilon + \text{div}(E^\varepsilon + F^\varepsilon), \quad (8.2)$$

where $E^\varepsilon = \eta^\varepsilon * ((1-\rho)\nabla\mathbf{p}) - (1-\rho)\eta^\varepsilon * \nabla\mathbf{p}$ and $F^\varepsilon = \eta^\varepsilon * (\mathbf{p} \otimes \nabla\rho) - (\eta^\varepsilon * \mathbf{p}) \otimes \nabla\rho$.

2. *Error control.* We show that for some C_T independent of ε, t (but which depends on T),

$$\|D^\varepsilon\|_{L^\infty([0,T] \times \mathbb{T}^2)} + \|E^\varepsilon\|_{L^2([0,T] \times \mathbb{T}^2)} + \|F^\varepsilon\|_{L^2([0,T] \times \mathbb{T}^2)} \leq C_T. \quad (8.3)$$

To begin with, for a.e. $s \in (t, T)$ it holds $\|D^\varepsilon(s)\|_{L^\infty(\mathbb{T}^2)} \leq \|D(s)\|_{L^\infty(\mathbb{T}^2)} \leq \|\mathbf{p}\|_{L^\infty} \leq C$, for some positive constant C independent of ε, t, T , using standard properties of the mollifier in the first inequality. The first part of (8.3) follows. Next, we estimate F^ε : directly from basic properties of the mollifier sequence, we have for a.e. $s \in (t, T)$

$$\|\eta^\varepsilon * (\mathbf{p} \otimes \nabla\rho)(s)\|_{L^2(\mathbb{T}^2)} \leq C\|(\mathbf{p} \otimes \nabla\rho)(s)\|_{L^2(\mathbb{T}^2)} \leq C\|\mathbf{p}\|_{L^\infty}\|\nabla\rho(s)\|_{L^2(\Omega)},$$

for some positive constant C independent of ε, t, T , and analogously

$$\|(\eta^\varepsilon * \mathbf{p}) \otimes \nabla\rho(s)\|_{L^2(\mathbb{T}^2)} \leq \|(\eta^\varepsilon * \mathbf{p})(s)\|_{L^\infty(\mathbb{T}^2)}\|\nabla\rho(s)\|_{L^2(\mathbb{T}^2)} \leq \|\mathbf{p}\|_{L^\infty}\|\nabla\rho(s)\|_{L^2(\mathbb{T}^2)},$$

thus, by integrating in time, there is a positive $C_T = C\|\nabla\rho\|_{L^2([0,T] \times \mathbb{T}^2)}$ independent of ε, t such that $\|F^\varepsilon\|_{L^2([0,T] \times \mathbb{T}^2)} \leq C_T$, whence the final term in (8.3) is bounded.

The final error term E^ε is written $E^\varepsilon(s, x) = \int_{\mathbb{R}^2} \eta^\varepsilon(x-y)[\rho(s, x) - \rho(s, y)]\nabla\mathbf{p}(t, y) dy$, and—by integrating by parts—admits the two-term decomposition

$$E^\varepsilon(s, x) = \underbrace{\int_{\mathbb{R}^2} \mathbf{p}(s, y) \otimes \nabla\eta^\varepsilon(x-y)[\rho(s, x) - \rho(s, y)] dy}_{=: E_1^\varepsilon(s, x)} + \underbrace{\int_{\mathbb{R}^2} \eta^\varepsilon(x-y)\mathbf{p}(s, y) \otimes \nabla\rho(s, y) dy}_{=: E_2^\varepsilon(s, x)}.$$

Using standard properties of mollifiers, $\|E_2^\varepsilon(s)\|_{L^2(\mathbb{T}^2)} \leq C\|\nabla\rho(s)\|_{L^2(\mathbb{T}^2)}\|\mathbf{p}\|_{L^\infty}$, thus

$$\|E_2^\varepsilon\|_{L^2([0,T] \times \mathbb{T}^2)} \leq C\|\nabla\rho\|_{L^2([0,T] \times \mathbb{T}^2)}, \quad (8.4)$$

for some C independent of ε, t, T . Meanwhile, for a.e. $x \in \mathbb{R}^2$ and $y \in \text{supp } \eta^\varepsilon(x - \cdot)$,

$$\begin{aligned} |\rho(s, x) - \rho(s, y)| &= \left| \int_0^1 \frac{d}{d\sigma} \rho(s, \sigma x + (1-\sigma)y) d\sigma \right| = \left| \int_0^1 \nabla\rho(s, \sigma x + (1-\sigma)y) \cdot (x - y) d\sigma \right| \\ &\leq \varepsilon \int_0^1 |\nabla\rho(s, \sigma x + (1-\sigma)y)| d\sigma, \end{aligned}$$

using $\text{supp } \eta^\varepsilon = B(0, \varepsilon)$ to bound $|x - y| \leq \varepsilon$. The boundedness of $\mathbf{p}$ then implies

$$\begin{aligned} |E_1^\varepsilon(s, x)| &\leq \varepsilon^{-2} \int_{B(x, \varepsilon)} |\nabla \eta(\frac{x-y}{\varepsilon})| \left(\int_0^1 |\nabla \rho(s, \sigma x + (1-\sigma)y)| d\sigma \right) dy \\ &= \int_{B(0,1)} |\nabla \eta(w)| \left(\int_0^1 |\nabla \rho(s, x - (1-\sigma)\varepsilon w)| d\sigma \right) dw, \end{aligned}$$

and, by Jensen's inequality, we get $|E_1^\varepsilon(s, x)|^2 \leq C \int_{B(0,1)} \int_0^1 |\nabla \rho(s, x - (1-\sigma)\varepsilon w)|^2 d\sigma dw$, for some positive constant C independent of ε, t, T . The Fubini–Tonelli Theorem then implies, using the periodicity of $\nabla \rho$ and translation invariance of Lebesgue measure,

$$\|E_1^\varepsilon(s)\|_{L^2(\mathbb{T}^2)}^2 \leq C \int_{B(0,1)} \int_0^1 \int_{\mathbb{T}^2} |\nabla \rho(s, x)|^2 dx d\sigma dw \leq C \|\nabla \rho(s)\|_{L^2(\mathbb{T}^2)}^2.$$

By integrating in time, and combining with (8.4), we deduce $\|E^\varepsilon\|_{L^2([0,T] \times \mathbb{T}^2)} \leq CT$.

3. H^1 bound on $\mathbf{p}$. $\mathbf{p}^\varepsilon \in L^\infty(t, T; C^\infty(\mathbb{T}^2))$ is an admissible test function for (8.2). Furthermore, it is immediate from (8.2) that $\partial_t \mathbf{p}^\varepsilon \in L^2((t, T); H^{-1}(\mathbb{T}^2))$ whence the Lions–Magenes Lemma applies and $\langle \partial_t \mathbf{p}^\varepsilon, \mathbf{p}^\varepsilon \rangle = \frac{1}{2} \frac{d}{dt} \|\mathbf{p}^\varepsilon\|_{L^2}^2$. By (8.3) and Young's inequality,

$$\|\mathbf{p}^\varepsilon(s)\|_{L^2(\mathbb{T}^2)}^2 + c_{t,T} \int_t^s \int_{\mathbb{T}^2} |\nabla \mathbf{p}^\varepsilon|^2 dx ds' \leq \|\mathbf{p}^\varepsilon(t)\|_{L^2(\mathbb{T}^2)}^2 + \frac{C_T^2}{c_{t,T}} + 2c_{t,T}^{-1} \|\nabla \rho\|_{L^2([0,s] \times \mathbb{T}^2)}^2,$$

for all $t < s \leq T$, using also $\|\mathbf{p}^\varepsilon(t)\|_{L^\infty(\mathbb{T}^2)} \leq C \|\mathbf{p}(t)\|_{L^\infty(\mathbb{T}^2)} \leq C$. We deduce

$$\|\mathbf{p}^\varepsilon\|_{L^\infty((t,T); L^2(\mathbb{T}^2))}^2 + \|\nabla \mathbf{p}^\varepsilon\|_{L^2((t,T) \times \mathbb{T}^2)}^2 \leq C'_{t,T}, \quad (8.5)$$

for some $C'_{t,T}$ independent of ε . We let $\varepsilon \rightarrow 0$, and the required boundedness of $\nabla \mathbf{p}$ follows from the lower semicontinuity of the norms. Finally, $\partial_t \mathbf{p} \in L^2((t, T); H^{-1}(\mathbb{T}^2))$ from (2.8).

4. L^4 bound on $\nabla \rho$. It now follows from equation (1.3) that $\partial_t \rho - \Delta \rho = -\text{div}((1-\rho)\mathbf{p}) \in L^2((t', T') \times \mathbb{T}^2)$ for all $0 < t' < T'$, thus the Calderón–Zygmund Theorem yields

$$\partial_t \rho, \nabla^2 \rho \in L^2((t, T) \times \mathbb{T}^2) \quad \forall 0 < t < T < \infty. \quad (8.6)$$

Thus, we may rigorously perform the standard H^2 estimate on (1.3) by testing against $\Delta \rho$, which yields $\frac{1}{2} \frac{d}{dt} \|\nabla \rho\|_{L^2(\mathbb{T}^2)}^2 + \frac{1}{2} \|\nabla^2 \rho\|_{L^2(\mathbb{T}^2)}^2 \leq \frac{1}{2} \|\text{div}((1-\rho)\mathbf{p})\|_{L^2(\mathbb{T}^2)}^2$, i.e.,

$$\nabla \rho \in L^\infty((t, T); L^2(\mathbb{T}^2)) \quad \forall 0 < t < T < \infty. \quad (8.7)$$

It follows from (8.7) and (8.6) that $\nabla \rho$ satisfies the assumptions of the interpolation Lemma 2.6, which yields $\nabla \rho \in L^4((t, T) \times \mathbb{T}^2)$ for all $0 < t < T < \infty$. $\square$

Note furthermore that, by integrating (1.1) against $\mathbf{e}(\theta) \otimes \mathbf{e}(\theta)$ in θ , we also get the governing equation for $\mathbf{P}$ the drift coefficient in (2.8):

$$\partial_t \mathbf{P} + \text{div}((1-\rho)\Pi) = \text{div}((1-\rho)\nabla \mathbf{P} + \mathbf{P} \otimes \nabla \rho) + \tilde{\Pi}, \quad (8.8)$$

where $\Pi = \int_{\mathbb{T}} f(\mathbf{e}(\theta) \otimes \mathbf{e}(\theta) \otimes \mathbf{e}(\theta)) d\theta$ is a tensor which again satisfies

$$|\Pi(t, x)| \leq 2^3 \int_{\mathbb{T}} |f(t, x, \theta)| d\theta = 2^3 \rho(t, x) \leq 2^3 \implies \Pi \in L^\infty((0, \infty) \times \mathbb{T}^2),$$

and the same holds for $\tilde{\Pi} = \int_{\mathbb{T}} f \partial_\theta^2 (\mathbf{e}(\theta) \otimes \mathbf{e}(\theta)) d\theta$. In turn, by applying the same strategy as in Lemma 8.2 to the equation (8.8), we have proved the following.

Lemma 8.3 (H^1 estimate for $\mathbf{P}$). *For all $0 < t < T < \infty$, it holds $\nabla \mathbf{P} \in L^2((t, T) \times \mathbb{T}^2)$.*

Having established Lemma 8.2, it follows that (for positive times) the equation (2.8) for $\mathbf{p}$ satisfies the assumptions of the Gehring-type lemma (Lemma 2.9) when rewritten as

$$\partial_t \mathbf{p} - \text{div}((1-\rho)\nabla \mathbf{p}) = -\mathbf{p} + \text{div}(\mathbf{p} \otimes \nabla \rho - (1-\rho)\mathbf{p});$$

indeed, we set $A = (1 - \rho)\text{Id} \in L^\infty$, $c = c_{t,T}^{-1}$ provided by (8.1) for the ellipticity constant, $g = -\mathbf{p} \in L^\infty$, and $G = \mathbf{p} \otimes \nabla \rho - (1 - \rho)\mathbf{P} \in L^4((t, T) \times \mathbb{T}^2)$ so that $p = 4$ in Lemma 2.9. We emphasise that this result only requires the tensor A to have *measurable* coefficients—no continuity is needed. The *Gehring-type* Lemma 2.9 therefore yields: for all $0 < t < T$

$$\nabla \mathbf{p} \in L^{2+\delta}((t, T) \times \mathbb{T}^2), \quad (8.9)$$

for some $\delta > 0$ which depends only on $t, T, c_{t,T}$. In turn, for all $0 < t < T < \infty$ there exists some $\delta > 0$ depending only on $t, T, c_{t,T}$ such that $\text{div}((1 - \rho)\mathbf{p}) \in L^{2+\delta}((t, T) \times \mathbb{T}^2)$ —note that we have already shown in Lemma 8.2 that $\nabla \rho \in L_{t,\text{loc}}^4 L_x^4$. Consequently, Lemma 2.7 (or, analogously, the Calderón–Zygmund Theorem) and (1.3) yield: for all $0 < t < T$

$$\partial_t \rho, \nabla^2 \rho \in L^{2+\delta}((t, T) \times \mathbb{T}^2), \quad (8.10)$$

for some $\delta > 0$ depending only on $t, T, c_{t,T}$; note that this δ is *a priori* different to the one in (8.9) for the same choice of t, T , since the result of Lemma 2.7 is *local*. Noting that $\mathbb{T}^2$ is of dimension $2 < 2 + \delta$ (no matter the value of $\delta > 0$ depending on the precise parabolic cylinder), Morrey’s embedding yields: for all $0 < t < T$ we have

$$\|\nabla \rho(s)\|_{L^\infty(\mathbb{T}^2)} \leq C_S \|\nabla \rho(s)\|_{W^{1,2+\delta}(\mathbb{T}^2)} \quad \forall s \in (t, T),$$

where the Sobolev constant C_S depends only on $\mathbb{T}^2$ and δ (which depends only on $t, T, c_{t,T}$); thus, C_S depends only on $t, T, c_{t,T}$. Raising to the power $2 + \delta$ and integrating in s yields

$$\|\nabla \rho\|_{L^{2+\delta}((t,T);L^\infty(\mathbb{T}^2))} \leq C_S \|\nabla \rho\|_{L^{2+\delta}((t,T);W^{1,2+\delta}(\mathbb{T}^2))}.$$

Thus, by (8.10): for all $0 < t < T$, there exists $\delta > 0$ depending only on $t, T, c_{t,T}$ such that

$$\nabla \rho \in L^{2+\delta}((t, T); L^\infty(\mathbb{T}^2)). \quad (8.11)$$

Putting together (8.9)–(8.11), we have proved the following corollary of Lemma 8.2.

Corollary 8.4. *For all $0 < t < T < \infty$, $\exists \delta > 0$ depending only on $t, T, c_{t,T}$, such that*

$$\nabla \mathbf{p} \in L^{2+\delta}((t, T) \times \mathbb{T}^2), \quad \partial_t \rho, \nabla^2 \rho \in L^{2+\delta}((t, T) \times \mathbb{T}^2), \quad \nabla \rho \in L^{2+\delta}((t, T); L^\infty(\mathbb{T}^2)).$$

8.1.2. H^2 -type estimate for $\mathbf{p}$. The next step in our analysis involves a H^2 -type estimate on $\mathbf{p}$. The underlying idea is that, *provided one can justify changing from the divergence form to non-divergence form in (2.8)*, a crucial cancellation of mixed derivative terms occurs:

$$\partial_t \mathbf{p} + \text{div}((1 - \rho)\mathbf{P}) = (1 - \rho)\Delta \mathbf{p} + \mathbf{p}\Delta \rho - \mathbf{p}.$$

By testing the above with $\Delta \mathbf{p}$, which *a priori* is not justified since we *do not* know $\Delta \mathbf{p} \in L^2$, we formally get from Young’s inequality

$$\frac{d}{dt} \|\nabla \mathbf{p}\|_{L^2}^2 + \frac{c_{t,T}}{2} \|\Delta \mathbf{p}\|_{L^2}^2 + \|\nabla \mathbf{p}\|_{L^2}^2 \leq c_{t,T}^{-1} \left(\|\text{div}((1 - \rho)\mathbf{P})\|_{L^2}^2 + \|\Delta \rho\|_{L^2}^2 \right), \quad (8.12)$$

where we used $\nabla \mathbf{P} \in L^2((t, T) \times \mathbb{T}^2)$ (Lemma 8.3) so that the penultimate term is bounded. Integrating in time gives $\nabla \mathbf{p} \in L^\infty((t, T); L^2(\mathbb{T}^2)) \cap L^2((t, T); H^1(\mathbb{T}^2))$, as desired.

In order to justify the aforementioned passage from divergence form to non-divergence form, we approximate $\mathbf{p}$ using a Galerkin scheme. That is, we will construct x -smooth approximations $\mathbf{p}^n$ (which are linear combinations of eigenfunctions of Δ_x with periodic boundary conditions in $\mathbb{T}^2$) which converge to $\mathbf{p}$ as $n \rightarrow \infty$. The previous formal computation is then justified for each $\mathbf{p}^n$, whence we will obtain an estimate akin to (8.12)—since this estimate will turn out to be independent of n , it also holds for $\mathbf{p}$ itself. We stress that this Galerkin scheme is used only as a *tool for regularity*—not existence (already known).

At this stage of the analysis, we *fix an arbitrary pair of times* $0 < t < T < \infty$ such that t is a Lebesgue point of $t \mapsto (\mathbf{p}(t, \cdot), f(t, \cdot))$, and we denote by δ the *specific* $\delta > 0$ depending on $t, T, c_{t,T}$ such that Corollary 8.4 holds, and denote by q_δ the Hölder exponent:

$$\frac{1}{2} = \frac{1}{q_\delta} + \frac{1}{2 + \delta}; \quad (8.13)$$

where $q_\delta \in (1, \infty)$ depends only on $t, T, c_{t,T}$. In the lines that follow we introduce the required notions to write the aforementioned Galerkin scheme. Define the *ansatz*

$$\mathbf{p}^n(s, x) := \sum_{j=1}^n \boldsymbol{\alpha}_j^n(s) \varphi_j(x) \quad \forall (s, x) \in [t, T] \times \mathbb{T}^2, \quad (8.14)$$

where $\{\varphi_j\}_j$ is the smooth orthonormal basis of $L^2(\mathbb{T}^2)$ provided by the eigenfunctions of the Laplacian on $\mathbb{T}^2$ (with periodic boundary conditions), *i.e.*, with $0 < \lambda_1 < \lambda_2 < \dots$,

$$-\Delta \varphi_j = \lambda_j^2 \varphi_j \quad j \in \mathbb{N}, \quad (8.15)$$

and the coefficient vectors $\{\boldsymbol{\alpha}_j^n(s)\}_{j=1}^n \subset \mathbb{R}^2$ are to be determined, subject to initial data

$$\boldsymbol{\alpha}_j^n(t) = \langle \mathbf{p}(t), \varphi_j \rangle \quad \forall j = 1, \dots, n, \quad \forall n \in \mathbb{N}, \quad (8.16)$$

at the *initial time* $s = t$, where $\langle \cdot, \cdot \rangle$ denotes the $L^2(\mathbb{T}^2)$ inner product. By construction,

$$\lim_{n \rightarrow \infty} \|\mathbf{p}(t) - \mathbf{p}^n(t)\|_{L^2(\mathbb{T}^2)} = 0. \quad (8.17)$$

Furthermore, since $\mathbf{p}(t) \in H^1(\mathbb{T}^2)$ by Lemma 8.2, integration by parts and (8.15) imply

$$\langle \mathbf{p}(t), \varphi_j \rangle = -\lambda_j^{-2} \int_{\mathbb{T}^2} \mathbf{p}(t, x) \Delta \varphi_j(x) dx = \lambda_j^{-2} \int_{\mathbb{T}^2} \nabla \mathbf{p}(t, x) \cdot \nabla \varphi_j(x) dx = \lambda_j^{-1} \langle \mathbf{p}(t), \lambda_j^{-1} \varphi_j \rangle_{H^1},$$

where $\langle \cdot, \cdot \rangle_{H^1}$ denotes the inner product on $H^1(\mathbb{T}^2)$. The standard spectral theory of self-adjoint operators in Hilbert spaces implies $\{\lambda_j^{-1} \varphi_j\}_j$ is an orthonormal basis of $H^1(\mathbb{T}^2)$: $\mathbf{p}^n(t) = \sum_{j=1}^n \langle \mathbf{p}(t), \lambda_j^{-1} \varphi_j \rangle_{H^1} \lambda_j^{-1} \varphi_j \rightarrow \mathbf{p}(t)$ strongly in $H^1(\mathbb{T}^2)$, *i.e.*, with t the *initial time*

$$\lim_{n \rightarrow \infty} \|\mathbf{p}(t) - \mathbf{p}^n(t)\|_{H^1(\mathbb{T}^2)} = 0. \quad (8.18)$$

Moreover, $\mathbf{p}^n(s) \in C^\infty(\mathbb{T}^2)$ for all n, s , being a linear combination of eigenfunctions of Δ_x .

The next result shows that there exists $\mathbf{p}^n$ of the form (8.14) satisfying the required Galerkin approximation: for simplicity, since existence of $f, \rho, \mathbf{p}$ is already known, we construct $\mathbf{p}^n$ to be the solution of a *linear* equation.

Proposition 8.5 (Galerkin scheme $\mathbf{p}$). *For all $n \in \mathbb{N}$, $\exists \mathbf{p}^n \in AC([t, T]; C^\infty(\mathbb{T}^2))$ satisfying*

$$\begin{cases} \partial_t \mathbf{p}^n - \operatorname{div}((1 - \rho) \nabla \mathbf{p}^n + \mathbf{p}^n \otimes \nabla \rho) + \mathbf{p}^n = -\operatorname{div}((1 - \rho) \mathbf{P}), \\ \mathbf{p}^n(t) = \sum_{j=1}^n \langle \mathbf{p}(t), \varphi_j \rangle \varphi_j, \end{cases} \quad (8.19)$$

in the weak sense dual to $\operatorname{span}\{\varphi_j\}_{j=1}^n$ for all $s \in [t, T]$, with the initial data achieved in the strong L^2 -sense. Furthermore, $\mathbf{p}^n \rightharpoonup \mathbf{p}$ weakly in $L^2((t, T); H^2(\mathbb{T}^2))$, and

$$\sup_n \|\mathbf{p}^n\|_{L^\infty((t, T); L^2(\mathbb{T}^2))} + \sup_n \|\nabla \mathbf{p}^n\|_{L^2((t, T) \times \mathbb{T}^2)} + \sup_n \|\nabla^2 \mathbf{p}^n\|_{L^2((t, T) \times \mathbb{T}^2)} < \infty. \quad (8.20)$$

Proof. We divide the proof into several steps.

1. *Existence of $\mathbf{p}^n$.* Exploiting the orthonormality of the eigenfunctions and testing (8.19) with φ_j , we see that (8.19) is satisfied by the ansatz (8.14) in duality with $\operatorname{span}\{\varphi_j\}_{j=1}^n$ if

$$\frac{d\boldsymbol{\alpha}_j^n}{dt} + \sum_{k=1}^n \boldsymbol{\alpha}_k^n \left(\int_{\mathbb{T}^2} \left((1 - \rho) \nabla \varphi_k + \varphi_k \nabla \rho \right) \cdot \nabla \varphi_j dx \right) + \boldsymbol{\alpha}_j^n = \int_{\mathbb{T}^2} (1 - \rho) \mathbf{P} \nabla \varphi_j dx.$$

The above reads as a linear system of ODEs, of the form $\boldsymbol{\alpha}'(s) + A(s) \boldsymbol{\alpha}(s) = B(s)$, where the tensors A, B are sufficiently regular that the standard well-posedness theory applies—no assumptions on the invertibility/definiteness of A are needed. Specifically, $A(s), B(s)$ are *locally integrable* in the time variable s , thus an application of Carathéodory's Theorem implies that there exists $\mathbf{p}^n \in AC([t, T]; C^\infty(\mathbb{T}^2))$ solving (8.19) in duality with $\operatorname{span}\{\varphi_j\}_{j=1}^n$ (of the form (8.14)), with the initial data achieved in $L^2(\mathbb{T}^2)$ at $s = t$.

2. *Estimates independent of n .* By testing (8.19) with $\mathbf{p}^n$ (this is allowed since $\mathbf{p}^n(s) \in \text{span}\{\varphi_j\}_{j=1}^n$), we get the usual H^1 -type estimate $\frac{1}{2} \frac{d}{dt} \|\mathbf{p}^n\|_{L^2(\mathbb{T}^2)}^2 + \frac{c_{t,T}}{2} \|\nabla \mathbf{p}^n\|_{L^2(\mathbb{T}^2)}^2 + \|\mathbf{p}^n\|_{L^2(\mathbb{T}^2)}^2 \leq \frac{1}{c_{t,T}} (\|(1-\rho)\mathbf{P}\|_{L^2(\mathbb{T}^2)}^2 + \|\mathbf{p}^n |\nabla \rho|\|_{L^2(\mathbb{T}^2)}^2)$, where we used the lower bound on $1-\rho$ provided by $c_{t,T}$ and the Lions–Magenes Lemma (directly from the equation $\partial_t \mathbf{p}^n \in L^2((t,T); \text{span}\{\varphi_j\}_{j=1}^n)$ since $\text{span}\{\varphi_j\}_{j=1}^n$ is a finite-dimensional Hilbert space). Integrating in time yields (for all $t < s \leq T$)

$$\begin{aligned} \|\mathbf{p}^n(s)\|_{L^2(\mathbb{T}^2)}^2 + c_{t,T} \|\nabla \mathbf{p}^n\|_{L^2((t,s) \times \mathbb{T}^2)}^2 &\leq \|\mathbf{p}^n(t)\|_{L^2(\mathbb{T}^2)}^2 + c_{t,T}^{-1} \|(1-\rho)\mathbf{P}\|_{L^2((t,T) \times \mathbb{T}^2)}^2 \\ &\quad + c_{t,T}^{-1} \int_t^s \|\nabla \rho(s')\|_{L^\infty(\mathbb{T}^2)}^2 \|\mathbf{p}^n(s')\|_{L^2(\mathbb{T}^2)}^2 ds', \end{aligned}$$

where we dropped the positive term $\|\mathbf{p}^n\|_{L^2((t,T) \times \mathbb{T}^2)}^2$ on the left-hand side. Due to the strong convergence of initial data (8.17), we see that the first two terms on the right-hand side are bounded independently of n by some $C_{t,T}$ depending only on $t, T, c_{t,T}$. We get

$$\|\mathbf{p}^n(s)\|_{L^2(\mathbb{T}^2)}^2 + \|\nabla \mathbf{p}^n\|_{L^2((t,s) \times \mathbb{T}^2)}^2 \leq C_{t,T} \left(1 + \int_t^s \|\nabla \rho(s')\|_{L^\infty(\mathbb{T}^2)}^2 \|\mathbf{p}^n(s')\|_{L^2(\mathbb{T}^2)}^2 ds' \right),$$

and an application Grönwall's Lemma implies (for a different constant $C_{t,T}$)

$$\|\mathbf{p}^n\|_{L^\infty((t,T); L^2(\mathbb{T}^2))}^2 + c_{t,T} \|\nabla \mathbf{p}^n\|_{L^2((t,T) \times \mathbb{T}^2)}^2 \leq C_{t,T} \exp \left(C_{t,T} \int_t^T \|\nabla \rho(s')\|_{L^\infty(\mathbb{T}^2)}^2 ds' \right),$$

where the integral is bounded since $\nabla \rho \in L^{2+\delta}((t,T); L^\infty(\mathbb{T}^2))$ for some $\delta > 0$ depending on $t, T, c_{t,T}$ by Corollary 8.4, and the boundedness of the first two terms in (8.20) follows.

Next, we perform the H^2 -type estimate by testing (8.19) with $\Delta \mathbf{p}^n$ (this is allowed since $\Delta \mathbf{p}^n \in \text{span}\{\varphi_j\}_{j=1}^n$ due to (8.15)). Since $\mathbf{p}^n(s) \in C^\infty(\mathbb{T}^2)$, we rigorously pass to the non-divergence form of (8.19). Similarly to the computation (8.12), we get

$$\begin{aligned} \frac{d}{dt} \|\nabla \mathbf{p}\|_{L^2(\mathbb{T}^2)}^2 + \frac{c_{t,T}}{2} \|\Delta \mathbf{p}\|_{L^2(\mathbb{T}^2)}^2 &\leq c_{t,T}^{-1} \left(\|\text{div}((1-\rho)\mathbf{P})\|_{L^2(\mathbb{T}^2)}^2 + 2 \int_{\mathbb{T}^2} |\mathbf{p}^n|^2 |\Delta \rho|^2 dx \right) \\ &\leq c_{t,T}^{-1} \left(\|\text{div}((1-\rho)\mathbf{P})\|_{L^2(\mathbb{T}^2)}^2 + 2 \|\mathbf{p}^n(s)\|_{L^{q_\delta}(\mathbb{T}^2)}^2 \|\Delta \rho(s)\|_{L^{2+\delta}(\mathbb{T}^2)}^2 \right), \end{aligned}$$

using the Lions–Magenes Lemma in the finite-dimensional Hilbert space $\{\varphi_j\}_{j=1}^n$, where the final term is present because *a priori* we do not know $\mathbf{p}^n \in L^\infty$, unlike $\mathbf{p}$; we recall that q_δ defined in (8.13) depends on $t, T, c_{t,T}$. Next, the Sobolev inequality implies $\|\mathbf{p}^n(s)\|_{L^{q_\delta}(\mathbb{T}^2)} \leq C_S \|\mathbf{p}^n(s)\|_{H^1(\mathbb{T}^2)}$ for some C_S independent of n but which depends on $t, T, c_{t,T}$ (since q_δ does). Integrating in s and using $\nabla \mathbf{P} \in L_{\text{loc}}^2$ from Lemma 8.3, we get

$$\begin{aligned} \|\nabla \mathbf{p}^n(s)\|_{L^2(\mathbb{T}^2)}^2 + c_{t,T} \|\Delta \mathbf{p}^n\|_{L^2((t,s) \times \mathbb{T}^2)}^2 &\leq \|\nabla \mathbf{p}^n(t)\|_{L^2(\mathbb{T}^2)}^2 + c_{t,T}^{-1} \|\text{div}((1-\rho)\mathbf{P})\|_{L^2}^2 \\ &\quad + 4C_S^2 c_{t,T}^{-1} \|\mathbf{p}^n\|_{L^\infty((t,T); L^2(\mathbb{T}^2))}^2 \|\Delta \rho\|_{L^2((t,T); L^{2+\delta}(\mathbb{T}^2))}^2 \\ &\quad + 4C_S^2 c_{t,T}^{-1} \int_t^s \|\Delta \rho(s')\|_{L^{2+\delta}(\mathbb{T}^2)}^2 \|\nabla \mathbf{p}^n(s')\|_{L^2(\mathbb{T}^2)}^2 ds' \end{aligned}$$

for all $t < s \leq T$. The strong convergence to the initial data (8.18), the uniform bound on $\|\mathbf{p}^n\|_{L^\infty((t,T); L^2(\mathbb{T}^2))}$, and Corollary 8.4 implies that there exists $C_{t,T}$ independent of n such that the first three terms on the previous right-hand side are bounded by $C_{t,T}$. We deduce

$$\|\nabla \mathbf{p}^n(s)\|_{L^2(\mathbb{T}^2)}^2 + c_{t,T} \|\Delta \mathbf{p}^n\|_{L^2((t,s) \times \mathbb{T}^2)}^2 \leq C_{t,T} \left(1 + \int_t^s \|\Delta \rho(s')\|_{L^{2+\delta}(\mathbb{T}^2)}^2 \|\nabla \mathbf{p}^n(s')\|_{L^2(\mathbb{T}^2)}^2 ds' \right),$$

for some $C_{t,T}$ independent of n , whence an application of Grönwall's Lemma implies

$$\|\nabla \mathbf{p}^n\|_{L^\infty((t,T); L^2(\mathbb{T}^2))}^2 + c_{t,T} \|\Delta \mathbf{p}^n\|_{L^2((t,T) \times \mathbb{T}^2)}^2 \leq C_{t,T} \exp \left(C_{t,T} \int_t^T \|\Delta \rho(s')\|_{L^{2+\delta}(\mathbb{T}^2)}^2 ds' \right),$$

for a different $C_{t,T}$; the final integral is bounded due to Corollary 8.4. Thus (8.20) follows.

3. *Convergence to $\mathbf{p}$.* The uniform bounds (8.20) imply that $\{\mathbf{p}^n\}_n$ is a uniformly bounded sequence in $L^2((t, T); H^2(\mathbb{T}^2))$. As such, by Alaoglu's Theorem, it admits a weakly convergent limit $\mathbf{p}^\infty \in L^2((t, T); H^2(\mathbb{T}^2))$ for a subsequence which we do not relabel. By passing to the limit in the *linear* equation (8.19), we see that $\mathbf{p}^\infty$ satisfies

$$\begin{cases} \partial_t \mathbf{p}^\infty - \operatorname{div}((1 - \rho) \nabla \mathbf{p}^\infty + \mathbf{p}^\infty \otimes \nabla \rho) + \mathbf{p}^\infty = -\operatorname{div}((1 - \rho) \mathbf{P}), \\ \mathbf{p}^\infty(t) = \mathbf{p}(t), \end{cases} \quad (8.21)$$

where we used that $\mathbf{p} = \sum_{j=1}^\infty \langle \mathbf{p}(t), \varphi_j \rangle \varphi_j$ in $L^2(\mathbb{T}^2)$ from (8.18). The equation (8.21) holds in duality with $L^2((t, T); \operatorname{span}\{\varphi_j\}_{j=1}^\infty)$, and hence, by density, in duality with $L^2((t, T); H^1(\mathbb{T}^2))$; we also deduce directly from (8.21) that $\partial_t \mathbf{p}^\infty \in L^2((t, T); H^{-1}(\mathbb{T}^2))$. Next, we define the difference $\mathbf{w} := \mathbf{p}^\infty - \mathbf{p}$. On account of $\mathbf{p}, \mathbf{p}^\infty \in L^2((t, T); H^1(\mathbb{T}^2))$, from Lemma 8.2, we have that $\mathbf{w}$ satisfies

$$\partial_t \mathbf{w} - \operatorname{div}((1 - \rho) \nabla \mathbf{w} + \mathbf{w} \otimes \nabla \rho) + \mathbf{w} = 0,$$

with zero initial condition, and $\mathbf{w}$ is itself an admissible test function for this equation with $\partial_t \mathbf{w} \in L^2((t, T); H^{-1}(\mathbb{T}^2))$. The Lions–Magenes Lemma applies, and testing with $\mathbf{w}$ gives

$$\frac{1}{2} \frac{d}{dt} \|\mathbf{w}\|_{L^2(\mathbb{T}^2)}^2 + \frac{c_{t,T}}{2} \|\nabla \mathbf{w}\|_{L^2(\mathbb{T}^2)}^2 + \|\mathbf{w}\|_{L^2(\mathbb{T}^2)}^2 \leq \frac{1}{2c_{t,T}} \|\mathbf{w} \otimes \nabla \rho\|_{L^2}^2 \leq \frac{1}{2c_{t,T}} \|\nabla \rho\|_{L^\infty}^2 \|\mathbf{w}\|_{L^2}^2.$$

Writing $Y(s) := \|\mathbf{w}(s)\|_{L^2(\mathbb{T}^2)}^2$, this implies $Y'(s) \leq c_{t,T}^{-1} \|\nabla \rho(s)\|_{L^\infty(\mathbb{T}^2)}^2 Y(s)$, and hence

$$Y(s) \leq Y(t) \exp \left(c_{t,T}^{-1} \int_t^s \|\nabla \rho(s')\|_{L^\infty(\mathbb{T}^2)}^2 ds' \right) = 0 \quad \forall s \in [t, T],$$

since $Y(t) = 0$, where the integral is finite since $\nabla \rho \in L^{2+\delta}((t, T); L^\infty(\mathbb{T}^2))$ for some $\delta > 0$ depending on $t, T, c_{t,T}$ (Corollary 8.4). Hence $\mathbf{p} = \mathbf{p}^\infty \in L^2((t, T); H^2(\mathbb{T}^2))$. $\square$

Step 2 of the proof of Proposition 8.5 also gives $\nabla \mathbf{p} \in L^\infty((t, T); L^2(\mathbb{T}^2))$. Using the interpolation Lemma 2.6, and recalling from the lines above (8.13) that t, T were fixed arbitrarily, we have proved the following.

Corollary 8.6. *For all $0 < t < T < \infty$, it holds $\mathbf{p} \in L^2((t, T); H^2(\mathbb{T}^2))$ and $\nabla \mathbf{p} \in L^\infty((t, T); L^2(\mathbb{T}^2)) \cap L^4((t, T) \times \mathbb{T}^2)$.*

In turn, it follows from Lemma 8.2 and Corollary 8.6 that $\operatorname{div}((1 - \rho) \mathbf{p}) \in L^4((t, T) \times \mathbb{T}^2)$, whence the Calderón–Zygmund Theorem and (1.3) imply $\partial_t \rho, \nabla^2 \rho \in L^4((t, T) \times \mathbb{T}^2)$. We deduce that $\rho \in W^{1,4}((t, T) \times \mathbb{T}^2)$, whence Morrey's Embedding implies $\rho \in C^{1/4}([t, T] \times \mathbb{T}^2)$. Also, an application of the Gagliardo–Nirenberg inequality gives

$$\|\nabla \rho(s)\|_{L^8(\mathbb{T}^2)} \leq C \|\nabla^2 \rho(s)\|_{L^4(\mathbb{T}^2)}^{1/2} \|\rho(s)\|_{L^\infty(\mathbb{T}^2)}^{1/2} + C \|\rho(s)\|_{L^1(\mathbb{T}^2)}, \quad (8.22)$$

for some constant C depending only on $\mathbb{T}^2$. Integrating in time, we get $\|\nabla \rho\|_{L^8((t, T) \times \mathbb{T}^2)}^8 \leq C \|\nabla^2 \rho\|_{L^4((t, T) \times \mathbb{T}^2)}^4 \|\rho\|_{L^\infty}^4 + C(T - t)$. We have shown the following.

Corollary 8.7. *For all $0 < t < T < \infty$, it holds*

$$\partial_t \rho, \nabla^2 \rho \in L^4((t, T) \times \mathbb{T}^2), \quad \nabla \rho \in L^8((t, T) \times \mathbb{T}^2), \quad \rho \in C^{1/4}([t, T] \times \mathbb{T}^2).$$

8.1.3. *Boundedness of f by Moser iterations.* The goal of this subsection is to prove Proposition 8.1, which we shall show by means of Moser iterations akin to those of §5. In order to use this strategy, we must first verify that the equation (1.1) can be written as a problem in the standard parabolic space $L_{t,\operatorname{loc}}^2 H_\xi^1$ with $\partial_t f \in L_{t,\operatorname{loc}}^2 H_\xi^{-1}$.

Lemma 8.8 (f belongs to the parabolic space). *For all $0 < t < T < \infty$, it holds $f \in L^\infty((t, T); L^2(\mathbb{T}^3)) \cap L^2((t, T); H^1(\mathbb{T}^3)) \cap L^{10/3}((t, T) \times \mathbb{T}^3)$ and $\partial_t f \in L^2((t, T); H^{-1}(\mathbb{T}^3))$.*

Proof. Our objective is, as usual, to test the equation (1.1) against f , which gives the result formally. However, *a priori* f is not regular enough to be a test function. To this end, similarly to Lemma 8.2, we introduce the ξ -mollifier sequence $\{\eta^\varepsilon\}_\varepsilon$ in $\mathbb{R}^3$ constructed from the Friedrichs bump function. Define $D = (1 - \rho)f\mathbf{e}(\theta)$ and $D^\varepsilon = \eta^\varepsilon * D$. Inserting $\eta^\varepsilon(s, \xi - y)$ into the weak formulation of (1.1) and integrating in y , we get

$$\partial_t f^\varepsilon + \operatorname{div} D^\varepsilon = \operatorname{div}((1 - \rho)\nabla f^\varepsilon + f^\varepsilon \nabla \rho) + \partial_\theta^2 f^\varepsilon + \operatorname{div}(E^\varepsilon + F^\varepsilon), \quad (8.23)$$

where $E^\varepsilon = \eta^\varepsilon * ((1 - \rho)\nabla f) - (1 - \rho)\eta^\varepsilon * \nabla f$ and $F^\varepsilon = \eta^\varepsilon * (f\nabla \rho) - (\eta^\varepsilon * f)\nabla \rho$.

1. *Control of error terms.* Using that $f \in L^3_{t,\xi}$ from Definition 2.1, standard properties of the mollifier imply that $\|D^\varepsilon\|_{L^3([t,T] \times \mathbb{T}^3)} \leq C\|D\|_{L^3([t,T] \times \mathbb{T}^3)} \leq C\|f\|_{L^3([t,T] \times \mathbb{T}^3)}$, where the constant C is independent of t, T, ε . Similarly, for the term F^ε , we have $\|\eta^\varepsilon * (f\nabla \rho)(s)\|_{L^2(\mathbb{T}^3)} \leq C\|f\nabla \rho(s)\|_{L^2(\mathbb{T}^3)} \leq C\|f(s)\|_{L^3(\mathbb{T}^2)}\|\nabla \rho(s)\|_{L^6(\mathbb{T}^2)}$, whence squaring and using Young's inequality with the exponent $3/2$ yields $\|\eta^\varepsilon * (f\nabla \rho)\|_{L^2([t,T] \times \mathbb{T}^3)}^2 \leq C(\|f\|_{L^3([t,T] \times \mathbb{T}^2)}^3 + \|\nabla \rho\|_{L^6([t,T] \times \mathbb{T}^2)}^6)$, where the constant C is independent of t, T, ε ; the final term is bounded in view of Corollary 8.7. The second term in F^ε is controlled analogously, whence we deduce $\|D^\varepsilon\|_{L^2([t,T] \times \mathbb{T}^2)} + \|F^\varepsilon\|_{L^2([t,T] \times \mathbb{T}^2)} \leq C_{t,T}$ for some $C_{t,T}$ depending only on $t, T, c_{t,T}$. It remains to estimate E^ε , which, like in Lemma 8.2, we split into two parts. Standard computations (cf. Step 2 of the proof of Lemma 8.2) give

$$E^\varepsilon(s, \xi) = \underbrace{\int_{\mathbb{R}^3} f(s, y) \nabla \eta^\varepsilon(\xi - y) [\rho(s, \xi) - \rho(s, y)] dy}_{=: E_1^\varepsilon(s, \xi)} + \underbrace{\int_{\mathbb{R}^3} \eta^\varepsilon(\xi - y) f(s, y) \nabla \rho(s, y) dy}_{=: E_2^\varepsilon(s, \xi)}.$$

We have $\|E_2^\varepsilon(s)\|_{L^2(\mathbb{T}^3)} \leq C\|f\nabla \rho(s)\|_{L^2(\mathbb{T}^2)}$ using standard properties of the mollifier (with C independent of t, T, ε), which yields $\|E_2^\varepsilon\|_{L^2([t,T] \times \mathbb{T}^3)} \leq C_{t,T}$ using the same approach as for F^ε . For E_1^ε , an identical strategy to that in Step 2 of the proof of Lemma 8.2 yields

$$|E_1^\varepsilon(s, \xi)| \leq \int_{B(0,1)} f(s, \xi - \varepsilon w) |\nabla \eta(w)| \left(\int_0^1 |\nabla \rho(s, \xi - (1 - \sigma)\varepsilon w)| d\sigma \right) dw.$$

By squaring, Jensen's inequality, integrating in ξ and using Fubini's Theorem, we get

$$\|E_1^\varepsilon(s)\|_{L^2(\mathbb{T}^3)}^2 \leq C \int_{B(0,1)} \int_0^1 \left(\int_{\mathbb{T}^3} |f(s, \xi - \varepsilon w)|^2 |\nabla \rho(s, \xi - (1 - \sigma)\varepsilon w)|^2 d\xi \right) d\sigma dw,$$

where the constant C depends only on $\|\nabla \eta\|_{L^\infty}$. Using Hölder's inequality (with exponent $3/2$) and the translation invariance of Lebesgue measure, we deduce

$$\|E_1^\varepsilon(s)\|_{L^2(\mathbb{T}^3)}^2 \leq C\|f(s)\|_{L^{3/2}(\mathbb{T}^3)}^2 \|\nabla \rho(s)\|_{L^3(\mathbb{T}^2)}^2 = C\|f(s)\|_{L^3(\mathbb{T}^3)}^2 \|\nabla \rho(s)\|_{L^6(\mathbb{T}^2)}^2.$$

Young's inequality yields $\|E_1^\varepsilon\|_{L^2([t,T] \times \mathbb{T}^3)}^2 \leq C(\|f\|_{L^3}^3 + \|\nabla \rho\|_{L^6}^6)$. We have shown

$$\|D^\varepsilon\|_{L^2([t,T] \times \mathbb{T}^2)}^2 + \|E^\varepsilon\|_{L^2([t,T] \times \mathbb{T}^2)}^2 + \|F^\varepsilon\|_{L^2([t,T] \times \mathbb{T}^2)}^2 \leq C_{t,T} \quad (8.24)$$

for some $C_{t,T}$ depending only on $t, T, c_{t,T}$, and independent of ε .

2. *H^1 bound.* Note that it is clear from the previous bounds and the equation (8.23) that $\partial_t f^\varepsilon \in L^2((t, T); H^{-1}(\mathbb{T}^3))$, and hence the use of the Lions–Magenes Lemma is justified. Testing (8.23) with f^ε , we get, using (8.24), the lower bound on $1 - \rho$, and Young's inequality,

$$\frac{1}{2} \frac{d}{dt} \|f^\varepsilon\|_{L^2(\mathbb{T}^3)}^2 + \frac{c_{t,T}}{4} \|\nabla_\xi f^\varepsilon\|_{L^2(\mathbb{T}^3)}^2 \leq \frac{C_{t,T}}{2c_{t,T}} + c_{t,T}^{-1} \|f^\varepsilon\|_{L^3(\mathbb{T}^3)}^2 \|\nabla \rho\|_{L^6(\mathbb{T}^2)}^2.$$

By integrating in time, using that $\|f^\varepsilon(t)\|_{L^2(\mathbb{T}^3)} \leq C\|f(t)\|_{L^2(\mathbb{T}^3)} \leq C'\|f(t)\|_{L^3(\mathbb{T}^3)}$, which is finite by Definition 2.1, we find that $\|f^\varepsilon\|_{L^\infty((t,T); L^2(\mathbb{T}^3))}^2 + \|\nabla_\xi f^\varepsilon\|_{L^2((t,T) \times \mathbb{T}^3)}^2$ is bounded independently of ε . The result then follows from the lower semicontinuity of the norms, by letting $\varepsilon \rightarrow 0^+$; the equation (1.1) then yields $\partial_t f \in L^2((t, T); H^{-1}(\mathbb{T}^3))$. $\square$

We are ready to give the proof of Proposition 8.1; the argument is different from the one presented in [1, Proposition 6.7], as we use Moser's iterations—not those of De Giorgi. Note that the integrability of the derivatives of f is not sufficient (at this stage) to obtain f bounded directly by Morrey's embedding, which highlights the necessity of this argument.

Proof of Proposition 8.1. Let $R \leq 1, s \in [t, T]$ be such that $Q_R(s) \subset [t, T] \times \mathbb{T}^3$. We show there exist $C_{t,T}$ (depending only on $t, T, c_{t,T}$) and $\beta > 0$ (independent t, T, s, R), such that

$$\|f\|_{L^\infty(Q_r(s))} \leq \frac{C_{t,T}}{(R-r)^\beta} \quad \forall 0 < r < R, \quad (8.25)$$

from which the conclusion of the proposition follows immediately by a covering argument; recall that we are only interested in the *qualitative* statement $f \in L_{t,\text{loc}}^\infty L_\xi^\infty$. Without loss of generality (arguing by translation), we may assume $s = 0$ for our cylinders: herein,

$$\text{ess inf}_{Q_R} (1-\rho) \geq c_{t,T}, \quad \|\Delta \rho\|_{L^4(Q_R)} \leq \|\Delta \rho\|_{L^4((t,T) \times \mathbb{T}^2)}, \quad (8.26)$$

and the term involving $\Delta \rho$ is finite by Corollary 8.7.

1. *Caccioppoli inequality.* Let $M > 0, q \geq 2, 0 < r < R \leq 1$. In this step, we prove

$$\|(f \wedge M)^{\frac{q}{2}}\|_{L^\infty((-r,r; L^2(\mathbb{T}^3)))}^2 + \|\nabla_\xi(f \wedge M)^{\frac{q}{2}}\|_{L^2(Q_r)}^2 \leq \frac{\kappa q^2}{R-r} \|f^{\frac{q}{2}}\|_{L^{8/3}(Q_R)}^2, \quad (8.27)$$

where $\kappa := C c_{t,T}^{-2} (1 + \|\Delta \rho\|_{L^4((t,T) \times \mathbb{T}^3)}^2)$ and the constant C is independent of $q, M, c_{t,T}, R, r$ and all other norms of f and its averaged quantities (*i.e.* $\rho, \mathbf{p}$, *etc.*). The final term on the right-hand side of (8.27) is *a priori* only finite for $2 \leq q \leq 5/2$ by Lemma 8.8; nevertheless, the inequality remains true for all $q \geq 2$, with the possibility of this term being infinite.

Let $\psi \in C_c^\infty((-R, R))$ be a localising function (depending only on time) such that

$$0 \leq \psi \leq 1, \quad \psi|_{[-r,r]} = 1, \quad \psi|_{[-1,1] \setminus [-R,R]} = 0, \quad \|\partial_t \psi\|_{L^\infty((-1,1))} \leq \frac{C}{R-r},$$

for some constant C independent of r, R . We wish to test (1.1) with $q(f \wedge M)^{q-1} \psi^2 \in L_{t,\text{loc}}^2 H_\xi^1$ by Lemma 8.8, since we *formally expect* $\frac{d}{dt} \|(f \wedge M)^{\frac{q}{2}}\|_{L_\xi^2}^2 = \langle \partial_t f, q(f \wedge M)^{q-1} \rangle$. However, since we only know $\partial_t f \in L_{t,\text{loc}}^2 H_\xi^{-1}$, there is no obvious non-linear chain rule for the pairing $\langle \partial_t f, q(f \wedge M)^{q-1} \rangle$; the Lions–Magenes Lemma only applies when $q = 2$.

A common strategy is to use *Steklov averages* to bypass this: for all $h \in (0, 1)$, define

$$f^h(s, \xi) := \frac{1}{h} \int_s^{s+h} f(s', \xi) ds' \implies \partial_t f^h(s, \xi) = \frac{f(s+h, \xi) - f(s, \xi)}{h} \quad \text{a.e.} \quad (8.28)$$

We then obtain the evolution on f^h from (1.1) by applying $\frac{1}{h} \int_s^{s+h}$ to the equation, *i.e.*,

$$\partial_t f^h + \text{div}([(1-\rho)f]^h \mathbf{e}(\theta)) = \text{div}([(1-\rho)\nabla f + f\nabla \rho]^h) + \partial_\theta^2 f^h. \quad (8.29)$$

Now we test (8.29) with $q(f^h \wedge M)^{q-1} \psi^2$. By (8.28), $\partial_t f^h \in L_{\text{loc}}^3([0, \infty); L^3(\mathbb{T}^3))$, whence the usual chain rule for Sobolev functions applies (*cf.* [12, §5.10]), and we get

$$\langle \partial_t f^h, q(f^h \wedge M)^{q-1} \psi^2 \rangle = \frac{d}{dt} \int_{\mathbb{T}^3} \Phi_{M,q}(f^h) \psi^2 d\xi - 2 \int_{\mathbb{T}^3} \psi \partial_t \psi \Phi_{M,q}(f^h) d\xi,$$

where $\Phi_{M,q}(s) = \int_0^s q(r \wedge M)^{q-1} dr$. Observe that (for $q \geq 2$) this function satisfies $\frac{d}{ds}(s \wedge M)^q = q(s \wedge M)^{q-1} \mathbf{1}_{\{0 \leq s \leq M\}} \leq q(s \wedge M)^{q-1} = \Phi'_{M,q}(s) \leq q s^{q-1} = \frac{d}{ds} s^q$, whence

$$(s \wedge M)^q \leq \Phi_{M,q}(s) \leq s^q \quad \forall s \geq 0. \quad (8.30)$$

Integrating in time over $[-R, s]$ for $s \in (-r, r)$, for which $\psi(s) = 1$, $\psi(-R) = 0$, we get

$$\begin{aligned} & \int_{\mathbb{T}^3} \Phi_{M,q}(f^h(s)) + q(q-1) \int_{-R}^s \int_{\mathbb{T}^3} \psi^2(f^h \wedge M)^{q-2} \left([(1-\rho)\nabla f + f\nabla\rho]^h \cdot \nabla f^h + |\partial_\theta f^h|^2 \right) \mathbb{1}_{\{f^h < M\}} \\ &= 2 \int_{-R}^s \int_{\mathbb{T}^3} \psi \partial_t \psi \Phi_{M,q}(f^h) + q(q-1) \int_{-R}^s \int_{\mathbb{T}^3} \psi^2[(1-\rho)f]^h \mathbf{e}(\theta) \cdot (f^h \wedge M)^{q-2} \nabla f^h \mathbb{1}_{\{f^h < M\}}. \end{aligned} \quad (8.31)$$

Standard results on Steklov averages (e.g. [10, §1.3 Lemma 3.2]) imply $\nabla_\xi f^h \rightarrow \nabla_\xi f$ and $[(1-\rho)\nabla f + f\nabla\rho]^h \rightarrow (1-\rho)\nabla f + f\nabla\rho$ strongly L^2 (by Lemma 8.8, Corollary 8.7), hence

$$\left([(1-\rho)\nabla f + f\nabla\rho]^h \cdot \nabla f^h + |\partial_\theta f^h|^2 \right) \rightarrow \left([(1-\rho)\nabla f + f\nabla\rho] \cdot \nabla f + |\partial_\theta f|^2 \right) \quad \text{strongly in } L^1. \quad (8.32)$$

Also, $f^h \rightarrow f$ strongly in L^3 and a.e. (for a subsequence). Hence, $(f^h \wedge M)^{q-2} \mathbb{1}_{\{f^h < M\}} \rightarrow (f \wedge M)^{q-2} \mathbb{1}_{\{f < M\}}$ a.e. and, since the term is bounded by M^{q-2} , the Dominated Convergence Theorem implies $\iint \psi^2(f^h \wedge M)^{q-2} \mathbb{1}_{\{f^h < M\}} \varphi \, d\xi \, dt \rightarrow \iint \psi^2(f \wedge M)^{q-2} \mathbb{1}_{\{f < M\}} \varphi \, d\xi \, dt$ for all $\varphi \in L^1$, i.e.,

$$\psi^2(f^h \wedge M)^{q-2} \mathbb{1}_{\{f^h < M\}} \xrightarrow{*} \psi^2(f \wedge M)^{q-2} \mathbb{1}_{\{f < M\}} \quad \text{weakly-* in } L^\infty. \quad (8.33)$$

Combining (8.32) and (8.33) as a weak-strong product implies

$$\begin{aligned} & \lim_{h \rightarrow 0^+} \int_{-R}^s \int_{\mathbb{T}^3} \psi^2(f^h \wedge M)^{q-2} \left([(1-\rho)\nabla f + f\nabla\rho]^h \cdot \nabla f^h + |\partial_\theta f^h|^2 \right) \mathbb{1}_{\{f^h < M\}} \\ &= \int_{-R}^s \int_{\mathbb{T}^3} \psi^2(f \wedge M)^{q-2} \left([(1-\rho)\nabla f + f\nabla\rho] \cdot \nabla f + |\partial_\theta f|^2 \right) \mathbb{1}_{\{f < M\}}, \end{aligned}$$

and an analogous argument yields the expected convergence of the final term in (8.31). Meanwhile, directly from the definition of $\Phi_{M,q}(s) = \int_0^s q(r \wedge M)^{q-1} \, dr$, it holds

$$\begin{aligned} \Phi_{M,q}(f^h) &= (f^h)^q \mathbb{1}_{\{f^h < M\}} + [M^q + qM^{q-1}(f^h - M)] \mathbb{1}_{\{f^h \geq M\}} \\ &= [(f^h)^q - [M^q + qM^{q-1}(f^h - M)]] \mathbb{1}_{\{f^h < M\}} + [M^q + qM^{q-1}(f^h - M)], \end{aligned}$$

and we have that the first term on the right-hand side is uniformly bounded in h by $2(1+q)M^q$ and converges a.e. to $[f^q - [M^q + qM^{q-1}(f - M)]] \mathbb{1}_{\{f < M\}}$ (for a subsequence) while the second term converges strongly to $M^q + qM^{q-1}(f - M)$ in $L^3(\mathbb{T}^3)$ by standard results on Steklov averages. Using also the Dominated Convergence Theorem, we deduce

$$\lim_{h \rightarrow 0^+} 2 \int_{-R}^s \int_{\mathbb{T}^3} \psi \partial_t \psi \Phi_{M,q}(f^h) = 2 \int_{-R}^s \int_{\mathbb{T}^3} \psi \partial_t \psi \Phi_{M,q}(f).$$

Meanwhile, Fatou's Lemma implies $\int_{\mathbb{T}^3} \Phi_{M,q}(f(s)) \leq \liminf_{h \rightarrow 0^+} \int_{\mathbb{T}^3} \Phi_{M,q}(f^h(s))$. Thus, by letting $h \rightarrow 0^+$, using also the lower bound (8.26), we deduce from (8.31) that (a.e. s)

$$\begin{aligned} & \int_{\mathbb{T}^3} \Phi_{M,q}(f(s)) + q(q-1)c_{t,T} \int_{-R}^s \int_{\mathbb{T}^3} \psi^2|(f \wedge M)^{\frac{q}{2}-1} \nabla_\xi f|^2 \mathbb{1}_{\{f < M\}} \\ & \leq q(q-1) \left| \int_{-R}^s \int_{\mathbb{T}^3} \psi^2(f \wedge M)^{q-1} \nabla f \mathbb{1}_{\{f < M\}} \cdot \nabla \rho \right| \\ & \quad + q(q-1) \int_{-R}^s \int_{\mathbb{T}^3} \psi^2 |(f \wedge M)^{q-1} \nabla f \mathbb{1}_{\{f < M\}}| + \frac{C}{R-r} \int_{-R}^s \int_{\mathbb{T}^3} \Phi_{M,q}(f). \end{aligned}$$

Using the lower bound (8.30), $0 \leq \psi \leq 1$, and integration by parts to get $\Delta\rho$, we deduce

$$\begin{aligned} & \int_{\mathbb{T}^3} |(f(s) \wedge M)^{\frac{q}{2}}|^2 + 4(1-q^{-1})c_{t,T} \int_{-R}^s \int_{\mathbb{T}^3} |\nabla_\xi [\psi(f \wedge M)^{\frac{q}{2}}]|^2 \\ & \leq (q-1) \iint_{Q(R)} |(f \wedge M)^{\frac{q}{2}}|^2 |\Delta\rho| + 2(q-1) \int_{-R}^s \int_{\mathbb{T}^3} (f \wedge M)^{\frac{q}{2}} |\nabla [\psi(f \wedge M)^{\frac{q}{2}}]| + \frac{C}{R-r} \iint_{Q(R)} \Phi_{M,q}(f), \end{aligned}$$

for a.e. $s \in (-r, r)$. Using (8.26), Young's inequality, and the upper bound (8.30), it holds

$$\begin{aligned} & \| (f \wedge M)^{\frac{q}{2}} \|_{L^\infty((-r, r); L^2(\mathbb{T}^3))}^2 + 3(1 - q^{-1}) c_{t, T} \| \nabla_\xi [(f \wedge M)^{\frac{q}{2}}] \|_{L^2(Q_r)}^2 \\ & \leq (q - 1) \| | (f \wedge M)^{\frac{q}{2}} |^2 \|_{L^{\frac{4}{3}}(Q_R)} \| \Delta \rho \|_{L^4(Q_R)} + \frac{q(q - 1)}{c_{t, T}} \| (f \wedge M)^{\frac{q}{2}} \|_{L^2(Q_R)}^2 + \frac{C}{R - r} \| f^{\frac{q}{2}} \|_{L^2(Q_R)}^2; \end{aligned}$$

note the final term is *a priori* only finite for $2 \leq q \leq 5/2$ by Lemma 8.8; nevertheless, the inequality remains true for all $q \geq 2$, with the possibility of this term being infinite. Using $1 - q^{-1} \geq 1/2$ and Hölder's inequality to control the L^2 norm by $L^{8/3}$ we deduce (8.27).

2. *Moser iteration.* Lemma 2.6 and (8.27) give $\| (f \wedge M)^{\frac{q}{2}} \|_{L^{10/3}(Q_r)}^2 \leq \frac{C \kappa q^2}{R - r} \| f^{\frac{q}{2}} \|_{L^{8/3}(Q_R)}^2$, and thus, by letting $M \rightarrow \infty$ and the Monotone Convergence Theorem, it holds

$$\| f \|_{L^{5q/3}(Q_r)} \leq \left(\frac{C \kappa q^2}{R - r} \right)^{1/q} \| f \|_{L^{4q/3}(Q_R)}, \quad (8.34)$$

where we recall C, κ are independent of q, r, R ; the above holds for all choices of $0 < r < R \leq 1$ and all $q \geq 2$. Next, define the powers $q_n := 2(\frac{5}{4})^n$ for $n \geq 0$ and the radii

$$r_0 = R, \quad r_n = R - (R - r) \gamma^{-1} \sum_{j=1}^n \left(\frac{5}{4} \right)^{-j}$$

where $\gamma = \sum_{n \geq 1} (5/4)^{-n}$, so that $R = r_0 > r_1 > \dots > r_n \searrow r$. Note that $r_n - r_{n+1} = (R - r) \gamma^{-1} (5/4)^{-n-1}$. Iterating (8.34) along this sequence (setting $q = q_n$, $R = r_n$, $r = r_{n+1}$), it holds—by setting $L := 5C\kappa\gamma$ and $\alpha := 5/4$ —the estimate

$$\| f \|_{L^{\frac{8}{3}\alpha^{n+1}}(Q_{r_{n+1}})} \leq \left(\frac{L\alpha^{3n}}{R - r} \right)^{\frac{1}{2\alpha^n}} \| f \|_{L^{\frac{8}{3}\alpha^n}(Q_{r_n})} \leq \left(\prod_{j=0}^n \left(\frac{L\alpha^{3j}}{R - r} \right)^{\frac{1}{2\alpha^j}} \right) \| f \|_{L^{\frac{8}{3}}(Q_R)}. \quad (8.35)$$

We study the convergence of the product. We have

$$\log \left(\prod_{j=0}^n \left(\frac{L\alpha^{3j}}{R - r} \right)^{\frac{1}{2\alpha^j}} \right) = \frac{\alpha}{2} \log \left(\frac{L}{R - r} \right) \sum_{j=1}^{n+1} \alpha^{-j} + \frac{3}{2} \log \alpha \sum_{j=0}^n j \alpha^{-j} \nearrow \frac{\alpha\gamma}{2} \log \left(\frac{L}{R - r} \right) + \bar{C}$$

where $\bar{C}$ depends only on α . Returning to (8.35) and letting $n \rightarrow \infty$, it holds

$$\| f \|_{L^\infty(Q_r)} \leq C \left(\frac{L}{R - r} \right)^{\frac{5\gamma}{8}} \| f \|_{L^{8/3}(Q_R)},$$

where C depends only on α . Using $\| f \|_{L^{8/3}(Q_R)} \leq C_{t, T} \| f \|_{L^{10/3}((t, T) \times \mathbb{T}^3)}$ and Lemma 8.8, and recalling the definitions of L, κ from (8.27), we get (8.25). $\square$

8.2. **Control of L^p norms of $\partial_t f, \nabla^2 f$.** Having shown that $f \in L_{t, \text{loc}}^\infty L_\xi^\infty$ in Proposition 8.1, our next aim is to show the following.

Proposition 8.9 (Higher integrability $\partial_t f, \nabla_\xi^2 f$). *Let $0 < t < T < \infty$. Then, $\partial_t f, \nabla_\xi^2 f \in L^p((t, T) \times \mathbb{T}^3)$ for all $1 \leq p < \infty$.*

In order to prove this proposition, we first need to perform a H^2 -type estimate on f —in the same spirit as Proposition 8.5 for $\mathbf{p}$. This is the content of the next result.

Lemma 8.10 (H^2 estimate for f). *Let $0 < t < T < \infty$. Then, $\partial_t f, \nabla_\xi^2 f \in L^2((t, T) \times \mathbb{T}^3)$ and $\nabla_\xi f \in L^\infty((t, T); L^2(\mathbb{T}^3)) \cap L^4((t, T) \times \mathbb{T}^3)$.*

Proof. As in §8.1.2, the underlying notion is that, provided the passage from divergence form (1.1) to non-divergence form is justified, then (1.1) reads

$$\partial_t f + \operatorname{div}((1 - \rho) f \mathbf{e}(\theta)) = (1 - \rho) \Delta f + f \Delta \rho + \partial_\theta^2 f, \quad (8.36)$$

whence testing with Δf (and $\partial_\theta^2 f$) gives the conclusion. To do this rigorously, we employ Galerkin's method. Without loss of generality t is a Lebesgue point of $t \mapsto (f(t), \nabla_\xi f(t))$.

1. *Galerkin approximations.* We recall the notations of §8.1.2, with $\{\varphi_j\}_j$ the orthonormal basis of $L^2(\mathbb{T}^2)$ of eigenfunctions of the Laplacian, and define the *ansatz*

$$f^n(s, x, \theta) := \sum_{j=1}^n \alpha_j^n(s, \theta) \varphi_j(x),$$

where the scalar coefficients $\{\alpha_j^n\}_{j=1}^n$ are determined from the linear system of PDEs

$$\partial_t \alpha_j^n - \partial_\theta^2 \alpha_j^n + \sum_{k=1}^n \alpha_k^n \left(\int_{\mathbb{T}^2} ((1-\rho) \nabla \varphi_k + \varphi_k \nabla \rho) \cdot \nabla \varphi_j \, dx \right) = \int_{\mathbb{T}^2} (1-\rho) f \mathbf{e}(\theta) \cdot \nabla \varphi_j \, dx, \quad (8.37)$$

with initial data $\alpha_j^n(t, \theta) = \langle f(t, \cdot, \theta), \varphi_j \rangle$, chosen such that f^n solves the *linear* equation

$$\begin{cases} \partial_t f^n + \operatorname{div}((1-\rho) f \mathbf{e}(\theta)) = \operatorname{div}((1-\rho) \nabla f^n + f^n \nabla \rho) + \partial_\theta^2 f^n, \\ f^n(t, \cdot, \theta) = \sum_{j=1}^n \langle f(t, \cdot, \theta), \varphi_j \rangle \varphi_j \end{cases} \quad (8.38)$$

in the weak sense dual to $L^2((t, T); H^1(\mathbb{T}); \operatorname{span}\{\varphi_j\}_{j=1}^n)$. The standard theory for strongly parabolic linear systems (recall (8.1)) implies the existence $\alpha_j^n \in L^2((t, T); H^1(\mathbb{T}))$ with $\partial_t \alpha_j^n \in L^2((t, T); H^{-1}(\mathbb{T}))$ solving (8.37), and thus of f^n solving (8.38) in the aforementioned weak sense. Additionally, we have (as $n \rightarrow \infty$)

$$\|f^n(t, \cdot, \theta)\|_{L^2(\mathbb{T}^2)}^2 = \sum_{j=1}^n \langle f(t, \cdot, \theta), \varphi_j \rangle^2 \longrightarrow \|f(t, \cdot, \theta)\|_{L^2(\mathbb{T}^2)}^2 \quad \text{a.e. } \theta.$$

Integrating in θ gives $\|f^n(t)\|_{L^2(\mathbb{T}^3)}^2 = \int_{\mathbb{T}} \sum_{j=1}^n \langle f(t, \cdot, \theta), \varphi_j \rangle^2 \, d\theta$, so the Monotone Convergence Theorem (the integrands form an increasing non-negative sequence) implies

$$\lim_{n \rightarrow \infty} \|f^n(t)\|_{L^2(\mathbb{T}^3)}^2 = \int_{\mathbb{T}} \lim_{n \rightarrow \infty} \sum_{j=1}^n \langle f(t, \cdot, \theta), \varphi_j \rangle^2 \, d\theta = \int_{\mathbb{T}} \|f(t, \cdot, \theta)\|_{L^2(\mathbb{T}^2)}^2 \, d\theta.$$

Analogously, by similar manipulations as (8.18), this convergence holds for $\|\nabla_\xi f(t)\|_{L^2}$:

$$\lim_{n \rightarrow \infty} \|f^n(t)\|_{L^2(\mathbb{T}^3)}^2 = \|f(t)\|_{L^2(\mathbb{T}^3)}^2, \quad \lim_{n \rightarrow \infty} \|\nabla_\xi f^n(t)\|_{L^2(\mathbb{T}^3)}^2 = \|\nabla_\xi f(t, \cdot)\|_{L^2(\mathbb{T}^3)}^2. \quad (8.39)$$

2. *H^1 estimate.* We begin with the standard H^1 estimate, by testing (8.38) with f^n itself; this is allowed since f^n is written as a linear combination of the $\{\varphi_j\}$. Using the Lions–Magenes Lemma over the finite-dimensional Hilbert space $\operatorname{span}\{\varphi_j\}_{j=1}^n$, and (8.1),

$$\begin{aligned} \frac{1}{2} \frac{d}{dt} \|f^n\|_{L^2(\mathbb{T}^3)}^2 + \frac{c_{t,T}}{2} \|\nabla_\xi f^n\|_{L^2(\mathbb{T}^3)}^2 &\leq \frac{1}{2c_{t,T}} \|(1-\rho) f \mathbf{e}(\theta)\|_{L^2(\mathbb{T}^3)}^2 - \int_{\mathbb{T}^3} f^n \nabla f^n \cdot \nabla \rho \, d\xi \\ &\leq \frac{1}{2c_{t,T}} \|f\|_{L^2(\mathbb{T}^3)}^2 + \frac{1}{c_{t,T}} \int_{\mathbb{T}^3} |f^n|^2 |\nabla \rho|^2 \, d\xi + \frac{c_{t,T}}{4} \|\nabla_\xi f^n\|_{L^2(\mathbb{T}^3)}^2, \end{aligned}$$

whence $\frac{1}{2} \frac{d}{dt} \|f^n\|_{L^2(\mathbb{T}^3)}^2 + \frac{c_{t,T}}{4} \|\nabla_\xi f^n\|_{L^2(\mathbb{T}^3)}^2 \leq \frac{1}{2c_{t,T}} \|f\|_{L^2(\mathbb{T}^3)}^2 + \frac{1}{c_{t,T}} \|f^n\|_{L^2(\mathbb{T}^3)}^2 \|\nabla \rho\|_{L^\infty(\mathbb{T}^3)}^2$, and integrating in time yields (for all $t < s \leq T$)

$$\begin{aligned} \|f^n(s)\|_{L^2(\mathbb{T}^3)}^2 + c_{t,T} \|\nabla_\xi f^n\|_{L^2((t,s) \times \mathbb{T}^3)}^2 &\leq \|f^n(t)\|_{L^2(\mathbb{T}^3)}^2 + c_{t,T}^{-1} \|f\|_{L^2((t,T) \times \mathbb{T}^3)}^2 \\ &\quad + 2c_{t,T}^{-1} \int_t^s \|\nabla \rho(s')\|_{L^\infty(\mathbb{T}^3)}^2 \|f^n(s')\|_{L^2(\mathbb{T}^3)}^2 \, ds' \\ &\leq C_{t,T} \left(1 + \int_t^s \|\nabla \rho(s')\|_{L^\infty(\mathbb{T}^3)}^2 \|f^n(s')\|_{L^2(\mathbb{T}^3)}^2 \, ds' \right), \end{aligned}$$

where we used the convergence of initial data (8.39) to establish the boundedness-in- n of the right-hand side, and $C_{t,T}$ depends only on $t, T, c_{t,T}$. By Grönwall's Lemma, we deduce

$$\|f^n\|_{L^\infty((t,T);L^2(\mathbb{T}^3))}^2 \leq C_{t,T} \exp \left(C_{t,T} \int_t^T \|\nabla \rho(s)\|_{L^\infty(\mathbb{T}^3)}^2 ds \right),$$

where we notice that the final term is bounded by Corollary 8.4. Returning to the previous estimate, we deduce that there exists $C_{t,T}$ depending only on $t, T, c_{t,T}$ (not n) such that

$$\|f^n\|_{L^\infty((t,T);L^2(\mathbb{T}^3))}^2 + \|\nabla_\xi f^n\|_{L^2((t,T) \times \mathbb{T}^3)}^2 \leq C_{t,T}. \quad (8.40)$$

3. H^2 estimate. We pass from the divergence form to non-divergence form of (8.38); this is allowed since each f^n is smooth in x (being a linear combination of eigenfunctions of the Laplacian). Testing with Δf^n , which is allowed since $\Delta f^n(t, \cdot, \theta) \in \text{span}\{\varphi_j\}_{j=1}^n$ a.e. t, θ ,

$$\begin{aligned} \frac{1}{2} \frac{d}{dt} \|\nabla f^n\|_{L^2(\mathbb{T}^3)}^2 + \frac{c_{t,T}}{2} \|\Delta f^n\|_{L^2(\mathbb{T}^3)}^2 + \|\nabla \partial_\theta f^n\|_{L^2(\mathbb{T}^3)}^2 \\ \leq \frac{1}{2c_{t,T}} \|\text{div}((1-\rho)f\mathbf{e}(\theta))\|_{L^2(\mathbb{T}^3)}^2 - \int_{\mathbb{T}^3} f^n \Delta f^n \Delta \rho d\xi, \end{aligned}$$

so that, by Young's inequality and integrating in time, for all $t < s \leq T$,

$$\begin{aligned} \|\nabla f^n(s)\|_{L^2(\mathbb{T}^3)}^2 + \frac{c_{t,T}}{2} \|\Delta f^n\|_{L^2((t,s) \times \mathbb{T}^3)}^2 + \|\nabla \partial_\theta f^n\|_{L^2((t,s) \times \mathbb{T}^3)}^2 \\ \leq \|\nabla f^n(t)\|_{L^2(\mathbb{T}^3)}^2 + c_{t,T}^{-1} \|\text{div}((1-\rho)f\mathbf{e}(\theta))\|_{L^2((t,T) \times \mathbb{T}^3)}^2 + 2c_{t,T}^{-1} \int_t^s \|f^n \Delta \rho\|_{L^2(\mathbb{T}^3)}^2 ds'. \end{aligned}$$

From the convergence of the initial data (8.39), Lemma 8.8, and Proposition 8.1, we deduce that there exists $C_{t,T}$ depending only on t, T such that

$$\begin{aligned} \|\nabla f^n(s)\|_{L^2(\mathbb{T}^3)}^2 + \frac{c_{t,T}}{2} \|\Delta f^n\|_{L^2((t,s) \times \mathbb{T}^3)}^2 + \|\nabla \partial_\theta f^n\|_{L^2((t,T) \times \mathbb{T}^3)}^2 \\ \leq C_{t,T} \left(1 + \int_t^s \int_{\mathbb{T}} \|(f^n \Delta \rho)(s', \theta')\|_{L^2(\mathbb{T}^2)}^2 d\theta' ds' \right), \end{aligned} \quad (8.41)$$

note that we have written the θ integral explicitly in the final term. Using Hölder's inequality, this final term may be estimated as (since ρ is independent of θ)

$$\begin{aligned} \int_t^s \int_{\mathbb{T}} \|(f^n \Delta \rho)(s', \theta')\|_{L^2(\mathbb{T}^2)}^2 d\theta' ds' &\leq \int_t^s \int_{\mathbb{T}} \|f^n(s', \theta')\|_{L^4(\mathbb{T}^2)}^2 \|\Delta \rho(s', \theta')\|_{L^4(\mathbb{T}^2)}^2 d\theta' ds' \\ &\leq C \int_t^s \|\Delta \rho(s')\|_{L^4(\mathbb{T}^2)}^2 \left(\int_{\mathbb{T}} \|f^n(s', \theta')\|_{H^1(\mathbb{T}^2)}^2 d\theta' \right) ds' \end{aligned}$$

where we used the Sobolev inequality in $\mathbb{T}^2$ (i.e. with respect to only x), and the constant C depends only on $\mathbb{T}^2$ and the exponent 4. Returning to (8.41), we get

$$\begin{aligned} \|\nabla f^n(s)\|_{L^2(\mathbb{T}^3)}^2 + \frac{c_{t,T}}{2} \|\Delta f^n\|_{L^2((t,s) \times \mathbb{T}^3)}^2 + \|\nabla \partial_\theta f^n\|_{L^2((t,s) \times \mathbb{T}^3)}^2 \\ \leq C_{t,T} \left(1 + \int_t^s \|\Delta \rho(s')\|_{L^4(\mathbb{T}^2)}^2 \left(\|f^n(s')\|_{L^2(\mathbb{T}^3)}^2 + \|\nabla f^n(s')\|_{L^2(\mathbb{T}^3)}^2 \right) ds' \right) \\ \leq C_{t,T} \left(1 + \|f^n\|_{L^\infty((t,T);L^2(\mathbb{T}^3))}^2 \|\Delta \rho\|_{L^4((t,T) \times \mathbb{T}^3)}^2 + \int_t^s \|\Delta \rho(s')\|_{L^4(\mathbb{T}^2)}^2 \|\nabla f^n(s')\|_{L^2(\mathbb{T}^3)}^2 ds' \right), \end{aligned}$$

i.e., $\|\nabla f^n(s)\|_{L^2}^2 \leq C_{t,T} (1 + \int_t^s \|\Delta \rho(s')\|_{L^4}^2 \|\nabla f^n(s')\|_{L^2}^2 ds')$, and Grönwall's Lemma implies

$$\|\nabla f^n\|_{L^\infty((t,T);L^2(\mathbb{T}^3))}^2 \leq C_{t,T} \exp \left(\int_t^T \|\Delta \rho(s)\|_{L^4(\mathbb{T}^2)}^2 ds \right),$$

and the final integral is bounded on account of $\Delta\rho \in L^4((t, T) \times \mathbb{T}^3)$ by Corollary 8.7. Returning to the previous estimate, we deduce that, for some $C_{t,T}$ independent of n ,

$$\|\nabla f^n\|_{L^\infty((t,T);L^2(\mathbb{T}^3))}^2 + \|\Delta f^n\|_{L^2((t,T) \times \mathbb{T}^3)}^2 + \|\nabla \partial_\theta f^n\|_{L^2((t,T) \times \mathbb{T}^3)}^2 \leq C_{t,T}. \quad (8.42)$$

4. *Convergence and conclusion.* An analogous strategy to Step 3, but testing this time with $\partial_\theta^2 f^n$ (again this belongs to the span of the eigenfunctions for a.e. t, θ) shows the uniform-in- n boundedness of $\|\partial_\theta^2 f^n\|_{L^2}$. Putting together with the bounds (8.40) and (8.42), we get that $f^n \rightharpoonup f^\infty$ weakly in $L^2((t, T); H^2(\mathbb{T}^3))$, and an argument analogous to Step 3 of the proof of Proposition 8.5 shows that $f^\infty = f$. Hence estimate (8.42) also holds for f .

Finally, using Proposition 8.1 and Gagliardo–Nirenberg inequality (for C depending only on $\mathbb{T}^3$), $\|\nabla_\xi f(s)\|_{L^4(\mathbb{T}^3)} \leq C(\|\nabla_\xi^2 f(s)\|_{L^2(\mathbb{T}^3)}^{\frac{1}{2}} \|f(s)\|_{L^\infty(\mathbb{T}^3)}^{\frac{1}{2}} + \|f(s)\|_{L^1(\mathbb{T}^3)})$, and thus $\|\nabla_\xi f\|_{L^4((t,T) \times \mathbb{T}^3)}^4 \leq C(\|\nabla_\xi^2 f\|_{L^2((t,T) \times \mathbb{T}^3)}^2 \|f\|_{L^\infty((t,T) \times \mathbb{T}^3)}^2 + (T-t))$. $\square$

We prove Proposition 8.9, by an inductive argument similar to [1, Proposition 7.1].

Proof of Proposition 8.9. The proof is divided into two steps.

1. *Initial upgrade of integrability.* By Lemma 8.10, the ξ -derivatives of f are endowed with sufficient integrability to rigorously pass to the non-divergence form of the equation (8.36), which holds weakly in duality with $L^2((t, T) \times \mathbb{T}^3)$. By defining the matrix

$$A := \begin{pmatrix} (1-\rho)\text{Id}_2 & 0 \\ 0 & 1 \end{pmatrix}, \quad (8.43)$$

which is symmetric, $1/4$ -Hölder continuous by Corollary 8.7, and strongly elliptic by the lower bound on $1-\rho$ provided by (8.1), we may recast (8.36) in the form

$$\partial_t f - A : \nabla_\xi^2 f = f \Delta \rho + f \mathbf{e}(\theta) \cdot \nabla \rho - (1-\rho) \nabla f \cdot \mathbf{e}(\theta), \quad (8.44)$$

and thus $\partial_t f - A : \nabla_\xi^2 f \in L^4((t, T) \times \mathbb{T}^3)$ by Proposition 8.1, Corollary 8.7, and Lemma 8.10. In turn, the Schauder result of Lemma 2.7 applies, and we deduce $\partial_t f, \nabla_\xi^2 f \in L^4((t, T) \times \mathbb{T}^3)$ —technically, the estimate holds for any time interval $(t', T') \subset \subset (t, T)$, but by resizing the original (t, T) , we obtain the result on any such interval. By arguing as per Step 4 of the proof of Lemma 8.10, the Gagliardo–Nirenberg inequality implies $\|\nabla_\xi f(s)\|_{L^8(\mathbb{T}^3)}^8 \leq C(\|\nabla_\xi^2 f(s)\|_{L^4(\mathbb{T}^3)}^4 \|f(s)\|_{L^\infty(\mathbb{T}^3)}^4 + \|f(s)\|_{L^1(\mathbb{T}^3)}^4)$, hence $\nabla_\xi f \in L^8((t, T) \times \mathbb{T}^3)$.

2. *Inductive argument.* We will show that, for all $m \geq 1$ and all $(t, T) \subset \subset (0, \infty)$, it holds $\partial_t f, \nabla_\xi^2 f \in L^{2^m}((t, T) \times \mathbb{T}^3)$ and $\nabla_\xi f \in L^{2^{m+1}}((t, T) \times \mathbb{T}^3)$. We have already shown the result for $m = 1, 2$. Suppose the result is true for $m \in \mathbb{N}$: we study the case $m+1$. Then, by (1.3), it holds $\partial_t \rho - \Delta \rho = \rho \operatorname{div} \mathbf{p} - \nabla \rho \cdot \mathbf{p} \in L^{2^{m+1}}((t, T) \times \mathbb{T}^2)$ by the induction hypothesis. Thus, the Calderón–Zygmund Theorem implies that (up to resizing the original time interval), it holds $\partial_t \rho, \nabla^2 \rho \in L^{2^{m+1}}((t, T) \times \mathbb{T}^2)$. Whence, returning to (8.44) and using also the boundedness of f from Proposition 8.1, it holds $\partial_t f - A : \nabla_\xi^2 f \in L^{2^{m+1}}((t, T) \times \mathbb{T}^3)$. The Schauder Lemma 2.7 implies (up to resizing the original time interval) that $\partial_t f, \nabla_\xi^2 f \in L^{2^{m+1}}((t, T) \times \mathbb{T}^3)$, and arguing by the Gagliardo–Nirenberg inequality as per the end of Step 1 (*cf.* Step 4 of Lemma 8.10) shows $\|\nabla_\xi f\|_{L^{2^{m+2}}((t,T) \times \mathbb{T}^3)}^{2^{m+2}} \leq C(\|\nabla_\xi^2 f\|_{L^{2^{m+1}}((t,T) \times \mathbb{T}^3)}^{2^{m+1}} \|f\|_{L^\infty((t,T) \times \mathbb{T}^3)}^{2^{m+1}} + (T-t))$. By induction, the result holds for all m , and thus the conclusion of the lemma follows by interpolation in Lebesgue spaces. $\square$

8.3. Higher regularity bootstrap. We argue differently to the strategy of §7 of [1]—here, we shall rely on Schauder estimates in Hölder continuous spaces instead of repeated uses of the Schauder type result in L^p (Lemma 2.7). Ultimately, the goal will be to argue using Hölder continuity of the right-hand side of (8.44) and a bootstrap argument. For

this, we need to show that $\Delta\rho$ and ∇f are continuous, and we will do this by obtaining sufficient integrability of their derivatives and Morrey's embedding.

In what follows, we use the shorthand notation $\dot{f} = \partial_t f$, $\dot{\rho} = \partial_t \rho$, $\dot{\mathbf{p}} = \partial_t \mathbf{p}$. Furthermore, we extend $f(t)$ smoothly in $(-\infty, 0]$ so that it is defined for a.e. $t \in \mathbb{R}$. For $h > 0$, we define the *difference quotient* $D_h f(t, \xi) := \frac{f(t+h, \xi) - f(t, \xi)}{h}$ for a.e. t, ξ , and similarly $D_h \rho(t, x) := \frac{\rho(t+h, x) - \rho(t, x)}{h}$ and $D_h \mathbf{p}(t, x) := \frac{\mathbf{p}(t+h, x) - \mathbf{p}(t, x)}{h}$; observe that $D_h \rho = \int_{\mathbb{T}} D_h f \, d\theta$ and $D_h \mathbf{p} = \int_{\mathbb{T}} D_h f \mathbf{e}(\theta) \, d\theta$ by linearity. We write $\tau_h \rho(t, x) = \rho(t+h, x)$, and observe that for all $h \in (0, 1)$ and $0 < t < T < \infty$, it holds $1 \geq 1 - \tau_h \rho \geq c_{t, T+1}$ by (8.1).

Lemma 8.11 ($\dot{\rho}$ equation). *For all $0 < t < T < \infty$, it holds $\dot{\rho} \in L^\infty((t, T); L^2(\mathbb{T}^2))$, $\nabla \dot{\rho} \in L^2((t, T) \times \mathbb{T}^2)$, $\partial_t \dot{\rho} \in L^2((t, T); H^{-1}(\mathbb{T}^2))$, and*

$$\partial_t \dot{\rho} + \operatorname{div}((1 - \rho)\dot{\mathbf{p}} - \mathbf{p}\dot{\rho}) = \Delta \dot{\rho} \quad \text{weakly in } L^2((t, T); H^1(\mathbb{T}^2)). \quad (8.45)$$

Proof. Firstly, observe that $\dot{\rho}, \dot{\mathbf{p}} \in L^p((t, T) \times \mathbb{T}^2)$ for all $p \in [1, \infty)$ by Proposition 8.9, so that (8.45) has meaning in the sense of distributions—we need to show that it follows from (1.3), and we do so by means of a difference quotient argument. From (1.3) and linearity:

$$\partial_t D_h \rho + \operatorname{div}((1 - \tau_h \rho) D_h \mathbf{p} - \mathbf{p} D_h \rho) = \Delta D_h \rho \quad \text{weakly in } L^2(0, \infty; H^1(\mathbb{T}^2));$$

recall $\tau_h \rho(t) = \rho(t+h)$. We test this with $D_h \rho$, whence the Lions–Magenes Lemma gives

$$\frac{1}{2} \frac{d}{dt} \|D_h \rho\|_{L^2(\mathbb{T}^2)}^2 + \|\nabla D_h \rho\|_{L^2(\mathbb{T}^2)}^2 = \int_{\mathbb{T}^2} (1 - \tau_h \rho) \nabla D_h \rho \cdot D_h \mathbf{p} \, dx - \int_{\mathbb{T}^2} D_h \rho \mathbf{p} \cdot \nabla D_h \rho \, dx.$$

Young's inequality, the boundedness of $\mathbf{p}$, and integrating in time imply $\|D_h \rho(s)\|_{L^2(\mathbb{T}^2)}^2 + \|\nabla D_h \rho\|_{L^2((t, s) \times \mathbb{T}^2)}^2 \leq \|D_h \rho(t)\|_{L^2(\mathbb{T}^2)}^2 + 2\|D_h \mathbf{p}\|_{L^2((t, T) \times \mathbb{T}^2)}^2 + 2\|D_h \rho\|_{L^2((t, T) \times \mathbb{T}^2)}^2$ for a.e. $t < s \leq T$. Standard properties of difference quotients (cf. e.g. [12, §5.8.2]) yield

$$\begin{aligned} & \|D_h \rho\|_{L^\infty((t, T); L^2(\mathbb{T}^2))}^2 + \|\nabla D_h \rho\|_{L^2((t, T) \times \mathbb{T}^2)}^2 \\ & \leq C \left(\|\dot{\rho}(t)\|_{L^2(\mathbb{T}^2)}^2 + \|\dot{\mathbf{p}}\|_{L^2((t, T) \times \mathbb{T}^2)}^2 + \|\dot{\rho}\|_{L^2((t, T) \times \mathbb{T}^2)}^2 \right), \end{aligned}$$

where the constant C is independent of h . Standard results on difference quotients imply that (by letting $h \rightarrow 0^+$) it holds $\dot{\rho} \in L^\infty((t, T) \times \mathbb{T}^2)$, $\nabla \dot{\rho} \in L^2((t, T) \times \mathbb{T}^2)$ as claimed. In turn, having established this regularity, we may rigorously pass a full time derivative ∂_t onto (1.3), and we deduce (8.45). The boundedness $\partial_t \dot{\rho} \in L^2((t, T); H^{-1}(\mathbb{T}^2))$ follows from this equation and the boundedness of the other terms. $\square$

We need to upgrade the regularity of $\dot{\rho}$ further, as doing so will enable us to deduce continuity of $\Delta\rho$ via (1.3), and thus show that the right-hand side of (8.44) is Hölder continuous, whence we can initiate the bootstrap argument in Hölder spaces. In order to do this, we argue using the equation for $\dot{f}$, which we derive in the next lemma.

Lemma 8.12 ($\dot{f}$ equation). *For all $0 < t < T < \infty$, it holds $\dot{f} \in L^\infty((t, T); L^2(\mathbb{T}^3))$, $\nabla_\xi \dot{f} \in L^2((t, T) \times \mathbb{T}^3)$, $\partial_t \dot{f} \in L^2((t, T); H^{-1}(\mathbb{T}^3))$, and moreover, weakly in $L^2((t, T); H^1(\mathbb{T}^3))$:*

$$\partial_t \dot{f} + \operatorname{div}(((1 - \rho)\dot{f} - f\dot{\rho})\mathbf{e}(\theta)) = \operatorname{div}((1 - \rho)\nabla \dot{f} + \dot{f}\nabla \rho - \dot{\rho}\nabla f + f\nabla \dot{\rho}) + \partial_\theta^2 \dot{f}. \quad (8.46)$$

Proof. As per the proof of Lemma 8.11, we obtain directly from (1.1) and linearity:

$$\begin{aligned} & \partial_t D_h f + \operatorname{div}[(D_h f(1 - \tau_h \rho) - f D_h \rho)\mathbf{e}(\theta)] \\ & = \operatorname{div}[(1 - \tau_h \rho)\nabla D_h f - D_h \rho \nabla f + \tau_h f \nabla D_h \rho + D_h f \nabla \rho] + \partial_\theta^2 D_h f; \end{aligned}$$

recall $\tau_h f = f(t+h)$. Testing with $D_h f$, we get (using the Lions–Magenes Lemma)

$$\begin{aligned} & \frac{1}{2} \frac{d}{dt} \|D_h f\|_{L^2(\mathbb{T}^3)}^2 + c_{t, T+1} \|\nabla_\xi D_h f\|_{L^2(\mathbb{T}^3)}^2 \\ & \leq \left(\|D_h f\|_{L^2} + \|f D_h \rho\|_{L^2} + \|D_h \rho \nabla f\|_{L^2} + \|\tau_h f \nabla D_h \rho\|_{L^2} + \|D_h f \nabla \rho\|_{L^2} \right) \|\nabla D_h f\|_{L^2}, \end{aligned}$$

whence Young's inequality and standard properties of difference quotients yield

$$\begin{aligned} & \frac{1}{2} \frac{d}{dt} \|D_h f\|_{L^2(\mathbb{T}^3)}^2 + \frac{c_{t,T+1}}{2} \|\nabla_\xi D_h f\|_{L^2(\mathbb{T}^3)}^2 \\ & \leq C c_{t,T+1}^{-1} \left(\|\dot{f}\|_{L^2}^2 + \|f\|_{L^\infty}^2 \|\dot{\rho}\|_{L^2}^2 + \|\dot{\rho}\|_{L^4}^2 \|\nabla f\|_{L^4}^2 + \|\tau_h f\|_{L^\infty}^2 \|\nabla \dot{\rho}\|_{L^2}^2 + \|\dot{f}\|_{L^4}^2 \|\nabla \rho\|_{L^4}^2 \right) \end{aligned}$$

where C is independent of s, h . Using the boundedness of the terms on the right-hand side (Propositions 8.1, 8.9, Lemma 8.11) $\|D_h f\|_{L^\infty((t,T);L^2(\mathbb{T}^3))}^2 + \|\nabla_\xi D_h f\|_{L^2((t,T)\times\mathbb{T}^3)}^2 \leq C_{t,T}$ for some constant depending only on $t, T, c_{t,T+1}$. As such, $\dot{f} \in L^\infty((t,T);L^2(\mathbb{T}^3)) \cap L^2((t,T);H^1(\mathbb{T}^3))$. This regularity is sufficient to pass a time derivative (in the sense of distributions) in (1.1), and we deduce (8.46) $\square$

Lemma 8.12 implies that $\dot{\mathbf{p}}$ is endowed with the same regularity as $\dot{f}$, and

$$\partial_t \dot{\mathbf{p}} + \operatorname{div}((1-\rho)\dot{\mathbf{P}} - \mathbf{P}\dot{\rho}) = \operatorname{div}((1-\rho)\nabla \dot{\mathbf{p}} + \dot{\mathbf{p}} \otimes \nabla \rho - \dot{\rho} \nabla \mathbf{p} + \mathbf{p} \otimes \nabla \dot{\rho}) - \dot{\mathbf{p}}. \quad (8.47)$$

Returning to (8.45), we get $\partial_t \dot{\rho} - \Delta \dot{\rho} = -\operatorname{div}((1-\rho)\dot{\mathbf{p}} - \mathbf{p}\dot{\rho}) \in L_{t,\text{loc}}^2 L_x^2$, and the Calderón–Zygmund Theorem implies $\partial_t \dot{\rho}, \nabla^2 \dot{\rho} \in L^2((t,T)\times\mathbb{T}^2)$ for all $0 < t < T < \infty$, from which the usual H^2 estimate gives $\nabla \dot{\rho} \in L^\infty((t,T);L^2(\mathbb{T}^2))$ and by interpolation (Lemma 2.6) implies $\nabla \dot{\rho} \in L^4((t,T)\times\mathbb{T}^2)$. We have shown the following.

Corollary 8.13. *For all $0 < t < T < \infty$, $\partial_t \dot{\rho}, \nabla^2 \dot{\rho} \in L^2((t,T)\times\mathbb{T}^2)$, $\nabla \dot{\rho} \in L^4((t,T)\times\mathbb{T}^2)$.*

We need to upgrade the integrability of $\partial_t \dot{\rho}$ a little more: indeed, in order to use Morrey's embedding and deduce that $\dot{\rho}$ is Hölder continuous, we need $\dot{\rho} \in W_{t,\text{loc}}^{1,p} W_x^{1,p}$ for $p > 3$, with 3 being the dimension of (t, x) . To this end, we perform the H^2 estimate on $\dot{\mathbf{p}}$ from (8.47).

Lemma 8.14 (H^2 estimate for $\dot{\mathbf{p}}$). *For all $0 < t < T < \infty$, it holds*

$$\nabla \dot{\mathbf{p}} \in L^\infty((t,T);L^2(\mathbb{T}^2)) \cap L^2((t,T);H^1(\mathbb{T}^2)), \quad \nabla \dot{\mathbf{p}} \in L^4((t,T)\times\mathbb{T}^2).$$

Proof. We argue analogously to Proposition 8.5. Following this strategy to the letter, we obtain the existence of $\dot{\mathbf{p}}^n \in AC([t,T];\operatorname{span}\{\varphi_j\}_{j=1}^n)$ (φ_j eigenfunctions of Δ_x) solving

$$\begin{cases} \partial_t \dot{\mathbf{p}}^n - \operatorname{div}((1-\rho)\nabla \dot{\mathbf{p}}^n + \dot{\mathbf{p}}^n \otimes \nabla \rho) + \dot{\mathbf{p}}^n = \operatorname{div}(\mathbf{P}\dot{\rho} - (1-\rho)\dot{\mathbf{P}} + \mathbf{p} \otimes \nabla \dot{\rho} - \dot{\rho} \nabla \mathbf{p}), \\ \dot{\mathbf{p}}^n(t) = \sum_{j=1}^n \langle \dot{\mathbf{p}}(t), \varphi_j \rangle \varphi_j. \end{cases} \quad (8.48)$$

Testing the above with $\dot{\mathbf{p}}^n$ and Young's inequality we get (a.e. $t < s \leq T$)

$$\begin{aligned} \|\dot{\mathbf{p}}^n(s)\|_{L^2}^2 + \|\nabla \dot{\mathbf{p}}^n\|_{L^2((t,s)\times\mathbb{T}^2)}^2 & \leq C_{t,T} \left(\|\dot{\mathbf{p}}(t)\|_{L^2}^2 + \|\dot{\mathbf{p}}^n\|_{L^2}^2 \|\nabla \rho\|_{L^2}^2 + \|\dot{\rho}\|_{L^2}^2 \right. \\ & \quad \left. + \|\dot{\mathbf{P}}\|_{L^2}^2 + \|\mathbf{p}\|_{L^2}^2 \|\nabla \dot{\rho}\|_{L^2}^2 + \|\dot{\rho}\|_{L^2}^2 \|\nabla \mathbf{p}\|_{L^2}^2 \right) \end{aligned}$$

for $C_{t,T}$ depending only on $t, T, c_{t,T}$; we use the lower bound on $1-\rho$ from (8.1). Hence,

$$\|\dot{\mathbf{p}}^n(s)\|_{L^2(\mathbb{T}^2)} + \|\nabla \dot{\mathbf{p}}^n\|_{L^2((t,s)\times\mathbb{T}^2)} \leq C_{t,T} \left(1 + \int_t^s \|\nabla \rho(s')\|_{L^\infty(\mathbb{T}^2)}^2 \|\dot{\mathbf{p}}^n(s')\|_{L^2(\mathbb{T}^2)}^2 ds' \right),$$

and we conclude that $\|\dot{\mathbf{p}}^n\|_{L^\infty((t,T);L^2(\mathbb{T}^2))} + \|\nabla \dot{\mathbf{p}}^n\|_{L^2((t,T)\times\mathbb{T}^2)} \leq C_{t,T}$ using Grönwall's Lemma as per the proof of Proposition 8.5. Furthermore, given that $\dot{\mathbf{p}}^n(t) \in C^\infty(\mathbb{T}^2)$ for each n (being a linear combination of eigenfunctions of Δ_x), we rigorously pass to the non-divergence form of (8.48). Testing with $\Delta \dot{\mathbf{p}}^n$, we obtain the uniform bound $\sup_n (\|\nabla \dot{\mathbf{p}}^n\|_{L^\infty((t,T);L^2(\mathbb{T}^2))} + \|\Delta \dot{\mathbf{p}}^n\|_{L^2((t,T)\times\mathbb{T}^2)}) \leq C_{t,T}$; we skip the details since they are analogous (and simpler) to those in Proposition 8.5. An argument akin to Step 3 in Proposition 8.5 shows $\dot{\mathbf{p}}^n \rightharpoonup \dot{\mathbf{p}}$ weakly in $L^2((t,T);H^2(\mathbb{T}^2))$, and the result follows. $\square$

Using the result of Lemma 8.14 and (8.45), we get $\partial_t \dot{\rho} - \Delta \dot{\rho} \in L^4((t,T)\times\mathbb{T}^2)$, whence the Calderón–Zygmund Theorem implies $\partial_t \dot{\rho}, \nabla^2 \dot{\rho} \in L_{t,\text{loc}}^4 L_x^4$ and the following corollary.

Corollary 8.15. *For all $0 < t < T$, $\partial_t \dot{\rho}, \nabla^2 \dot{\rho} \in L^4((t, T) \times \mathbb{T}^2)$, and $\dot{\rho}, \Delta \rho \in C^{1/4}([t, T] \times \mathbb{T}^2)$ as well as $\nabla \dot{\rho} \in L^8((t, T) \times \mathbb{T}^2)$.*

Proof. From $\partial_t \dot{\rho}, \nabla^2 \dot{\rho} \in L^4_{t,\text{loc}} L^4_x$, we get $\dot{\rho} \in W^{1,4}([t, T] \times \mathbb{T}^2)$ for all $0 < t < T < \infty$, whence Morrey's embedding implies $\dot{\rho} \in C^{1/4}([t, T] \times \mathbb{T}^3)$. Expanding $\text{div}((1 - \rho)\mathbf{p}) = -\mathbf{p} \cdot \nabla \rho + (1 - \rho) \text{div } \mathbf{p}$, we get $\text{div}((1 - \rho)\mathbf{p}) \in W^{1,4}_{t,\text{loc}} W^{1,4}_x \hookrightarrow C^{1/4}_{t,\text{loc},x}$ using Proposition 8.9, Corollary 8.13, and Lemma 8.14. By (1.3), we deduce $\Delta \rho = \partial_t \rho + \text{div}((1 - \rho)\mathbf{p}) \in C^{1/4}([t, T] \times \mathbb{T}^2)$. Finally, since $C^{1/4}([t, T] \times \mathbb{T}^2) \hookrightarrow L^\infty([t, T] \times \mathbb{T}^2)$, Gagliardo–Nirenberg's inequality implies

$$\|\nabla \dot{\rho}(s)\|_{L^8(\mathbb{T}^2)} \leq C \left(\|\nabla^2 \dot{\rho}(s)\|_{L^4(\mathbb{T}^2)}^{1/2} \|\dot{\rho}(s)\|_{L^\infty(\mathbb{T}^2)}^{1/2} + \|\dot{\rho}(s)\|_{L^1(\mathbb{T}^2)} \right),$$

where C only depends on $\mathbb{T}^2$, and we then deduce $\nabla \dot{\rho} \in L^8_{t,\text{loc}} L^8_x$. $\square$

In order to initiate the bootstrap in Hölder spaces, it remains to show that ∇f is Hölder continuous; once this is shown, the entire right-hand side of (8.44) is Hölder continuous.

Lemma 8.16 (Hölder continuity of ∇f). *For all $0 < t < T < \infty$, $\nabla_\xi f \in C^{1/5}([t, T] \times \mathbb{T}^3)$.*

Proof. The results of Lemma 8.12 and Corollary 8.15 imply, by returning to (8.46), that

$$\partial_t \dot{f} - \text{div}((1 - \rho)\nabla \dot{f} + \dot{f}\nabla \rho) - \partial_\theta^2 \dot{f} = F, \quad (8.49)$$

with $F \in L^2_{t,\text{loc}} L^2_\xi$. By replicating the Galerkin approach of Lemma 8.10, we perform the H^2 estimate for $\dot{f}$; we omit the details for concision, which follow the formal computation associated with testing (8.49) with $\Delta \dot{f}$ after passing to non-divergence form. We get $\partial_t \dot{f}, \nabla_\xi^2 \dot{f} \in L^2_{t,\text{loc}} L^2_\xi$, as well as $\nabla_\xi \dot{f} \in L^\infty_{t,\text{loc}} L^2_\xi \cap L^2_{t,\text{loc}} H^1_\xi$. By Lemma 2.6, $\nabla_\xi \dot{f} \in L^{10/3}_{t,\text{loc}} L^{10/3}_\xi$. We now rigorously pass to the non-divergence form of (8.46) and observe that $\partial_t \dot{f} - A : \nabla_\xi^2 \dot{f} \in L^{10/3}_{t,\text{loc}} L^{10/3}_\xi$ where A is as per (8.43). Applying the Schauder result in Sobolev spaces (Lemma 2.7) implies $\partial_t \dot{f}, \nabla_\xi^2 \dot{f} \in L^{10/3}_{t,\text{loc}} L^{10/3}_\xi$. By Gagliardo–Nirenberg's inequality, we get

$$\|\nabla_\xi \dot{f}(s)\|_{L^{50/9}(\mathbb{T}^3)} \leq C \left(\|\nabla_\xi^2 \dot{f}(s)\|_{L^{10/3}(\mathbb{T}^3)}^{3/5} \|\nabla_\xi \dot{f}(s)\|_{L^2(\mathbb{T}^3)}^{2/5} + \|\nabla_\xi \dot{f}(s)\|_{L^2(\mathbb{T}^3)} \right),$$

with C depending only on $\mathbb{T}^3$. Raising to the power $50/9$ and integrating in time, we get $\nabla_\xi \dot{f} \in L^{50/9}((t, T) \times \mathbb{T}^3)$ for all $0 < t < T$. Since $50/9 > 5$, we deduce also by Proposition 8.9 that $\nabla_\xi \dot{f} \in W^{1,5}_{t,\text{loc}} W^{1,5}_\xi$, whence by Morrey's embedding $\nabla_\xi f \in C^{1/5}([t, T] \times \mathbb{T}^3)$. $\square$

We are ready to give the proof of Theorem 1, which concludes this work.

Proof of Theorem 1. The lower bound (3.1) was proved in Proposition 7.1: it suffices to show the interior smoothness of weak solutions. The proof is divided into several steps: our underlying strategy is to repeatedly apply additional ξ -derivatives to (1.1) and x -derivatives to (1.3) and deduce smoothness using the Schauder result in Hölder spaces (Lemma 2.8).

1. *Improved regularity of ρ .* Our aim in this step is to prove the following: for all $0 < t < T$, there exists $\beta > 0$ depending on $t, T, c_{t,T}$ such that

$$D\rho \in C^{1+\beta/2, 2+\beta}([t, T] \times \mathbb{T}^2), \quad \text{i.e.} \quad \partial_t D\rho, D^3 \rho \in C^{\beta/2, \beta}([t, T] \times \mathbb{T}^2). \quad (8.50)$$

Recall from Corollary 8.15 and Lemma 8.16 that $\Delta \rho, \nabla f \in C^{1/5}([t, T] \times \mathbb{T}^3)$ for all $0 < t < T$. As such, the entire right-hand side of (8.44) is Hölder continuous, as is the matrix A from (8.43); in fact for all $0 < t < T$ they belong to $C^{\alpha/2, \alpha}([t, T] \times \mathbb{T}^3)$ for some $\alpha \in (0, 1)$. As such (by modifying the size of the time interval—we omit the details for concision), the Schauder result of Lemma 2.8 implies

$$f \in C^{1+\alpha/2, 2+\alpha}([t, T] \times \mathbb{T}^3) \quad \forall 0 < t < T; \quad (8.51)$$

note that this level of regularity is already enough to deduce that f is a classical solution of (1.1). We return to the equation (1.3) for ρ . We mollify the entire equation (in x only)

and get $\partial_t \rho^\varepsilon - \Delta \rho^\varepsilon = F^\varepsilon := -\operatorname{div}((1-\rho)\mathbf{p}) * \eta_\varepsilon$, where $\rho^\varepsilon = \rho * \eta_\varepsilon$ for $\eta_\varepsilon = \varepsilon^{-2}\eta(\cdot/\varepsilon)$ with η the usual Friedrichs bump function in $\mathbb{R}^2$. By applying an x_j derivative, we get

$$\partial_t \rho_{x_j}^\varepsilon - \Delta \rho_{x_j}^\varepsilon = F_{x_j}^\varepsilon,$$

and note that $F_{x_j}^\varepsilon = F_{x_j} * \eta_\varepsilon = (\mathbf{p}_{x_j} \cdot \nabla \rho + \mathbf{p} \cdot \nabla \rho_{x_j} + \rho_{x_j} \nabla \mathbf{p} - (1-\rho) \operatorname{div} \mathbf{p}_{x_j}) * \eta_\varepsilon$. The regularity of f from (8.51) implies that, for all $0 < t < T$, there exists $\beta \in (0, 1)$ depending on $t, T, c_{t,T}$ such that $F_{x_j} \in C^{\beta/2, \beta}([t, T] \times \mathbb{T}^2)$. Standard properties of the mollifier sequence then imply that $\|F_{x_j}^\varepsilon\|_{C^{\beta/2, \beta}([t, T] \times \mathbb{T}^2)} \leq C \|F_{x_j}\|_{C^{\beta/2, \beta}([t, T] \times \mathbb{T}^2)}$ where the constant C is independent of t, T, ε . Hence, another application of Lemma 2.8 implies $\|\rho_{x_j}^\varepsilon\|_{C^{1+\beta/2, 2+\beta}([t', T'] \times \mathbb{T}^2)} \leq C(\|F_{x_j}\|_{C^{\beta/2, \beta}([t, T] \times \mathbb{T}^2)} + \|\nabla \rho\|_{L^2([t, T] \times \mathbb{T}^2)})$ for $[t', T'] \subset \subset [t, T]$ with C depending on t, t', T, T' . In turn, an application of the Ascoli–Arzelà Theorem implies the strong convergence of $\rho_{x_j}^\varepsilon \rightarrow \rho_{x_j}$ in the norm of $C^{1+\tilde{\beta}/2, 2+\tilde{\beta}}([t', T'] \times \mathbb{T}^2)$ for all $0 < \tilde{\beta} < \beta$, and (8.50) follows (for a different β depending on $t, T, c_{t,T}$).

2. *Improved regularity for f .* The regularity (8.51) implies that we may rigorously pass a ξ_j derivative in the divergence form of (1.1) to deduce that

$$\begin{aligned} \partial_t f_{\xi_j} + \operatorname{div}((1-\rho)f_{\xi_j}\mathbf{e}(\theta) - \rho_{\xi_j}f\mathbf{e}(\theta) + (1-\rho)f\mathbf{e}_{\xi_j}(\theta)) \\ = \operatorname{div}((1-\rho)\nabla f_{\xi_j} + f_{\xi_j}\nabla \rho - \rho_{\xi_j}\nabla f + f\nabla \rho_{\xi_j}) + \partial_\theta^2 f_{\xi_j} \end{aligned} \quad (8.52)$$

holds (in divergence form) in the sense dual to $L^1((t, T); W^{1,1}(\mathbb{T}^3))$. Our objective is then to show that $\nabla_\xi^2 f_{\xi_j} \in L_{t,\text{loc}}^2 L_\xi^2$ so that we may pass to the non-divergence form of this equation. To do this rigorously, we perform the same Galerkin method as in Lemma 8.10 to deduce $f_{\xi_j} \in L^2((t, T); H^2(\mathbb{T}^3))$; we do not write the details in full for clarity of presentation. The sought boundedness can be seen from the formal computation (testing (8.52) with Δf_{ξ_j} after having passed to non-divergence form)

$$\begin{aligned} \frac{1}{2} \frac{d}{dt} \|\nabla f_{\xi_j}\|_{L^2}^2 + c_{t,T} \|\Delta f_{\xi_j}\|_{L^2}^2 \leq \|\Delta f_{\xi_j}\|_{L^2} \Big(\|f_{\xi_j} \Delta \rho\|_{L^2} + \|\rho_{\xi_j} \Delta f\|_{L^2} + \|f \Delta \rho_{\xi_j}\|_{L^2} + \|\nabla f_{\xi_j}\|_{L^2} \\ + \|f_{\xi_j} \nabla \rho\|_{L^2} + \|f \nabla \rho_{\xi_j}\|_{L^2} + \|\rho_{\xi_j} \nabla f\|_{L^2} + \|\nabla f\|_{L^2} + \|f \nabla \rho\|_{L^2} \Big), \end{aligned}$$

where we note from (8.51)–(8.50) that all terms on the right-hand side are bounded in L^∞ . Applying Young's inequality yields $\Delta f_{\xi_j} \in L^2((t, T); L^2(\mathbb{T}^3))$, and repeating the same argument by testing with $\partial_\theta^2 f_{\xi_j}$ yields $\nabla_\xi^2 f_{\xi_j} \in L^2((t, T); L^2(\mathbb{T}^3))$.

Next, we show that: for all $0 < t < T$ there exists $\beta > 0$ depending on t, T such that

$$D_\xi f \in C^{1+\beta/2, 2+\beta}([t, T] \times \mathbb{T}^3). \quad (8.53)$$

Having shown $\nabla_\xi^2 f_{\xi_j} \in L_{t,\text{loc}}^2 L_\xi^2$, we return to (8.52) and rigorously pass to the non-divergence form of the equation to get

$$\partial_t f_{\xi_j} - A : \nabla_\xi^2 f_{\xi_j} = A_{\xi_j} : \nabla_\xi^2 f + G_{\xi_j},$$

where $G = f \Delta \rho + f \mathbf{e}(\theta) \cdot \nabla \rho - (1-\rho) \nabla f \cdot \mathbf{e}(\theta)$ is such that $D_\xi G \in C^\beta([t, T] \times \mathbb{T}^3)$ for some $\beta > 0$ (depending on $t, T, c_{t,T}$) by (8.50)–(8.51); note that the equation holds in the sense of L^2 functions. Furthermore, the results (8.51) and (8.50) imply that the entire right-hand side belongs to $C^\beta([t, T] \times \mathbb{T}^3)$ for some $\beta > 0$ depending on $t, T, c_{t,T}$. In turn, we directly apply the Schauder result of Lemma 2.8 and deduce that $f_{\xi_j} \in C^{1+\beta/2, 2+\beta}([t', T'] \times \mathbb{T}^3)$ for the same β on any $[t', T'] \subset \subset [t, T]$ compactly contained therein, and (8.53) follows.

3. *Conclusion.* We iterate the previous two steps: having shown that $D_\xi^m f$ is Hölder continuous, we apply a further ∂_{x_j} to (1.3) to deduce Hölder continuity of $D^{m+1} \rho$, and then returning to (1.1) and applying an additional ∂_{ξ_j} we deduce $D_\xi^{m+1} f$ is Hölder continuous. Using the equation to exchange t -derivatives for ξ -derivatives, we also obtain the Hölder continuity of all time derivatives, and mixed t, ξ derivatives. The proof is complete. $\square$

APPENDIX A. PROOFS OF TECHNICAL LEMMAS

A.1. Proof of convergence to initial data for ρ . We prove (2.6) directly from the definition of weak solution and the regularity of ρ from Definition 2.1.

Proof of (2.6). Inserting test functions $\varphi \in W^{1,6}(\mathbb{T}^2)$ independent of t, θ in (2.1), we obtain

$$\int_{\mathbb{T}^2} (\rho(t, x) - \rho_0(x)) \varphi(x) dx = \int_0^t \int_{\mathbb{T}^2} (1 - \rho) \mathbf{p} \cdot \nabla \varphi dx ds - \int_0^t \int_{\mathbb{T}^2} \nabla \rho \cdot \nabla \varphi dx ds; \quad (\text{A.1})$$

and, by a standard density argument, the above holds for all $\varphi \in H^1(\mathbb{T}^2)$. In turn, by bounding the relevant terms and arguing as per (2.5) in Remark 2.4, we deduce

$$\lim_{t \rightarrow 0^+} \|\rho(t) - \rho_0\|_{H^{-1}(\mathbb{T}^2)} = 0. \quad (\text{A.2})$$

Selecting t a Lebesgue point and inserting $\varphi = \rho(t) \in H^1(\mathbb{T}^2)$ in (A.1), we get (using Hölder's inequality) $\limsup_{t \rightarrow 0} \|\rho(t)\|_{L^2(\mathbb{T}^2)}^2 \leq \|\rho_0\|_{L^2(\mathbb{T}^2)} \limsup_{t \rightarrow 0} \|\rho(t)\|_{L^2(\mathbb{T}^2)}$; this $\limsup$ is finite since $0 \leq \rho \leq 1$ and $\mathbb{T}^2$ is bounded. Thus,

$$\limsup_{t \rightarrow 0^+} \|\rho(t)\|_{L^2(\mathbb{T}^2)} \leq \|\rho_0\|_{L^2(\mathbb{T}^2)}. \quad (\text{A.3})$$

Meanwhile, using $0 \leq \rho \leq 1$ we deduce that $\rho(t) \rightharpoonup \rho_0$ weakly in $L^2(\mathbb{T}^2)$; here we also use (A.2) to identify the limit as ρ_0 . The weak lower semicontinuity of the norm then implies

$$\|\rho_0\|_{L^2(\mathbb{T}^2)} \leq \liminf_{t \rightarrow 0^+} \|\rho(t)\|_{L^2(\mathbb{T}^2)} \leq \limsup_{t \rightarrow 0^+} \|\rho(t)\|_{L^2(\mathbb{T}^2)} \leq \|\rho_0\|_{L^2(\mathbb{T}^2)}$$

by (A.3). Thus $\|\rho(t)\|_{L^2} \rightarrow \|\rho_0\|_{L^2}$ and $\rho(t) \rightharpoonup \rho_0$ in L^2 , whence (2.6) follows. $\square$

A.2. Proof of Lemma 7.5. Our proof is similar to that of [14, Lemma 3.2], but simpler, as our setting is less general.

Proof of Lemma 7.5. Multiply (7.6) by the integrating factor e^t , and get (for all $0 \leq s \leq t$)

$$A_k(s) \leq e^{-s} A_k(0) + e^{-s} C_*^k \int_0^s e^\tau A_{k-1}(\tau)^2 d\tau.$$

By the assumptions on the initial values, we deduce (by bluntly bounding several terms), $A_k(s) \leq M_1^{n_k} + C_*^k \|A_{k-1}\|_{L^\infty([0,t])}^2 (1 - e^{-s})$, and thus

$$\|A_k\|_{L^\infty([0,t])} \leq M_1^{n_k} + C_*^k \|A_{k-1}\|_{L^\infty([0,t])}^2. \quad (\text{A.4})$$

1. *Inductive estimate.* Our objective in this step is to prove by induction that

$$\|A_k\|_{L^\infty([0,t])} \leq C_*^{\alpha_k} (M_1 + M_2)^{n_k} \quad (\text{A.5})$$

for all integer $k \geq 0$, where α_k satisfies the recurrence relation

$$\alpha_k = 2\alpha_{k-1} + k + 1 \quad (\alpha_0 := 1). \quad (\text{A.6})$$

Indeed, for $k = 0$, we simply have $\|A_0\|_{L^\infty([0,t])} \leq M_2$ whence (A.5) is trivially satisfied. Then, suppose (A.5) holds for the case $k - 1$: we show that it holds for k . Indeed, by (A.4) and using $2n_{k-1} = n_k$, it holds $\|A_k\|_{L^\infty([0,t])} \leq M_1^{n_k} + C_*^{k+2\alpha_{k-1}} (M_1 + M_2)^{n_k} \leq (1 + C_*^{k+2\alpha_{k-1}}) (M_1 + M_2)^{n_k}$, and thus

$$\|A_k\|_{L^\infty([0,t])} \leq 2C_*^{k+2\alpha_{k-1}} (M_1 + M_2)^{n_k} \leq C_*^{1+k+2\alpha_{k-1}} (M_1 + M_2)^{n_k},$$

using $2 \leq C_*$, and (A.5) follows directly from the definition (A.6) of α_k .

2. *Uniform boundedness.* We have $\|A_k\|_{L^\infty([0,t])}^{1/n_k} \leq C_*^{\alpha_k/n_k} (M_1 + M_2)$ and thus it suffices to show that α_k/n_k is bounded. Dividing (A.6) by $n_k = 2^k$ and rewriting $n_k = 2n_{k-1}$,

$$\frac{\alpha_k}{n_k} = \frac{\alpha_{k-1}}{n_{k-1}} + \frac{k+1}{2^k}$$

for all integer $k \geq 1$, from which we deduce (using $\alpha_0/n_0 = 1$) $\frac{\alpha_k}{n_k} = 1 + \sum_{j=1}^k (j+1)2^{-j}$. As a result, $\sup_k \frac{\alpha_k}{n_k} = 1 + \sum_{j=1}^{\infty} (j+1)2^{-j}$, where the final sum is manifestly convergent by the d'Alembert ratio test, and the aforementioned quotient is bounded, as required. We record, by setting $\beta := 1 + \sum_{j=1}^{\infty} (j+1)2^{-j}$, the estimate $\sup_k \|A_k\|_{L^\infty([0,t])}^{1/n_k} \leq C_*^{\sup_k \alpha_k/n_k} (M_1 + M_2) = C_*^\beta (M_1 + M_2)$, where we used again that $C_* \geq 2$, and the proof is complete. $\square$

Acknowledgements. The author thanks Luca Alasio and Nikita Simonov for useful discussions, the hospitality of CRM De Giorgi (SNS Pisa) as part of the Research in Pairs Programme, and the financial support from PEPS JC JC 2026 (Insmi CNRS).

REFERENCES

- [1] L. C. B. ALASIO AND S. M. SCHULZ, Regularity and Uniqueness for a Model of Active Particles with Angle-Averaged Diffusions, *NoDEA Nonlinear Differ. Equ. Appl.* **32** (2025) 70.
- [2] L. C. B. ALASIO AND S. M. SCHULZ, Erratum to “Regularity and Uniqueness for a Model of Active Particles with Angle-Averaged Diffusions”, *NoDEA Nonlinear Differ. Equ. Appl.* **32** (2025) 70.
- [3] L. C. B. ALASIO, J. GUERAND AND S. M. SCHULZ, Regularity and trend to equilibrium for a non-local advection-diffusion model of active particles, *Kinetic and Related Models* **18** (2025) 426-462.
- [4] P. AUSCHER, S. BORTZ, M. EGERT, AND O. SAARI, On regularity of weak solutions to linear parabolic systems with measurable coefficients, *J. Math. Pures et Appl.* **121** (2019) 216-243.
- [5] E. BOMBIERI AND E. GIUSTI Harnack’s inequality for elliptic differential equations on minimal surfaces, *Inventiones Mathematicae* **15** (1972) 24-46.
- [6] M. BONFORTE, J. DOLBEAULT, B. NAZARET AND N. SIMONOV, Stability in Gagliardo-Nirenberg-Sobolev inequalities: flows, regularity and the entropy method, to appear in *Memoirs of the Amer. Math. Soc.* (2026).
- [7] M. BRUNA, M. BURGER, A. ESPOSITO AND S. M. SCHULZ, Phase separation in systems of interacting active Brownian particles, *SIAM J. Appl. Math.* **82** (2022) 1635-1660.
- [8] M. BURGER AND S. M. SCHULZ, Well-posedness and stationary states for a crowded active Brownian system with size-exclusion, *Discrete Cont. Dyn. Sys.* **45** (2025) 2882-2915.
- [9] M. E. CATES AND J. TAILLEUR, When are active Brownian particles and run-and-tumble particles equivalent? Consequences for motility-induced phase separation, *EPL* **101** (2013) 20010.
- [10] E. DiBENEDETTO, *Degenerate Parabolic Equations*, Springer Science & Business Media, 2012.
- [11] H. DONG AND S. KIM, Partial Schauder Estimates for Second-Order Elliptic and Parabolic Equations: A Revisit, *International Mathematics Research Notices*, Vol. 2019, (2019) 2085–2136.
- [12] L. C. EVANS, *Partial Differential Equations*, 2nd edition, Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2010.
- [13] D. KIM, Parabolic equations with measurable coefficients II, *J. Math. Anal. Appl.*, **334** (2007) 534–548.
- [14] I. KIM AND Y. P. ZHANG, Regularity properties of degenerate diffusion equations with drifts, *SIAM J. Math. Anal.* **50** (2018) 4371–4406.
- [15] N. V. KRYLOV, *Lectures on Elliptic and Parabolic Equations in Hölder Spaces*, Graduate Studies in Mathematics **12**, AMS Press, Providence, RI, 1996.
- [16] J. LEWENSTEIN-SANPERA AND X. ROS-OTON, L^p estimates for the Laplacian via blow-up, *J. Differ. Equ.* **441** (2025) 113478.
- [17] G. LIEBERMAN, *Second Order Parabolic Differential Equations*, World Scientific Publishing Co., Inc., River Edge, NJ, xii+439 pp, 1996.
- [18] J. MOSER, On a pointwise estimate for parabolic differential equations, *Communications on Pure and Applied Mathematics* **24** (1971) 727–740.
- [19] P. ROMANCZUK, W. EBELING, B. LINDNER AND L. SCHIMANSKY-GEIER, Active Brownian particles, *Eur. Phys. J. Special Topics* **202** (2012) 1-162.
- [20] S. M. SCHULZ AND I. Y. VIOLO, Regularity for quasilinear elliptic equations in metric measure spaces, arXiv preprint (2025) [arXiv:2510.27564](https://arxiv.org/abs/2510.27564).
- [21] R. TEMAM, *Navier-Stokes equations. Theory and numerical analysis*, Studies in Mathematics and its Applications, Vol. 2, North-Holland Publishing Co., Amsterdam-New York-Oxford, 1977.
- [22] T. VICSEK AND A. ZAFEIRIS, Collective motion, *Phys. Rep.* **517** (2012) 71-140.
- [23] M. WANG, *Nonlinear Second Order Parabolic Equations*, CRC Press, 2021.
- [24] J. YEOMANS, The hydrodynamics of active systems, *Rivista del Nuovo Cimento* **40** (2017) 1-31.

(S. M. Schulz) LABORATOIRE DE MATHÉMATIQUES DE VERSAILLES, UMR 8100 CNRS, UVSQ, UNIVERSITÉ PARIS-SACLAY, 45 AV. DES ÉTATS-UNIS, 78000 VERSAILLES CEDEX, FRANCE
 Email address: simon.schulz@uvsq.fr