HAL
open science

ASYMPTOTICS OF THE THREE DIMENSIONAL VLASOV EQUATION IN THE LARGE MAGNETIC FIELD LIMIT

Francis Filbet, Luis Miguel Miguel Rodrigues

To cite this version:

Francis Filbet, Luis Miguel Miguel Rodrigues. ASYMPTOTICS OF THE THREE DIMENSIONAL VLASOV EQUATION IN THE LARGE MAGNETIC FIELD LIMIT. 2019. hal-01933746v2

HAL Id: hal-01933746

https://hal.science/hal-01933746v2

Preprint submitted on 12 Jul 2019 (v2), last revised 24 Feb 2020 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ASYMPTOTICS OF THE THREE DIMENSIONAL VLASOV EQUATION IN THE LARGE MAGNETIC FIELD LIMIT

FRANCIS FILBET AND L. MIGUEL RODRIGUES

Abstract. We study the asymptotic behavior of solutions to the Vlasov equation in the presence of a strong external magnetic field. In particular we provide a mathematically rigorous derivation of the guiding-center approximation in the general three dimensional setting under the action of large inhomogeneous magnetic fields. First order corrections are computed and justified as well, including electric cross field, magnetic gradient and magnetic curvature drifts. We also treat long time behaviors on two specific examples, the two dimensional case in cartesian coordinates and a poloidal axi-symmetric geometry, the former for expository purposes. Algebraic manipulations that underlie concrete computations make the most of the linearity of the stiffest part of the system of characteristics instead of relying on any particular variational structure.

Keywords: Vlasov equation; guiding center approximation; gyrokinetics; asymptotic analysis.
2010 MSC: 35Q83, 78A35, 82D10, 35B40.

Contents

1. Introduction 2
2. Definitions and main results 3
2.1. From ODEs to PDEs 4
2.2. Slow variables and first-order asymptotics 5
2.3. Second-order asymptotics 7
2.4. Long-time asymptotics in a toroidal axi-symmetric geometry 9
2.5. Further comments and numerical illustrations 11
3. Two-dimensional homogeneous case 13
3.1. Uniform bounds 14
3.2. Elimination of linear terms 14
3.3. First elimination and partial asymptotics 15
3.4. Elimination of quadratic terms 17
3.5. Second elimination and full asymptotics 18
3.6. Long-time asymptotics 19
3.7. PDE counterparts 20
4. General three-dimensional case 22
4.1. Slow variables and uniform bounds 22
4.2. Elimination of linear terms 23
4.3. Elimination of quadratic terms 25
4.4. Proof of Theorem 2.2 26
4.5. Elimination of higher-order terms 27
4.6. Proof of Theorem 2.4 29
5. Proof of Theorem 2.8 32
5.1. Geometric framework 32
5.2. Uniform bounds and asymptotics 33
Appendix A. Comparison with the classical adiabatic invariant formulation 34
A.1. Long-time asymptotics in the toroidal axi-symmetric case 35

Date: July 12, 2019.
FF was supported by the EUROfusion Consortium and has received funding from the Euratom research and training programme 2014-2018 under grant agreement No 633053. The views and opinions expressed herein do not necessarily reflect those of the European Commission.

Research of LMR has received funding from the city of Rennes.

A.2. General second-order asymptotics 37
References 38

1. Introduction

Since fusion configurations involve very hot plasmas, they typically require a careful design to maintain fast moving particles inside the core of the device on sufficiently long times. In the magnetic confinement approach [Bellan, 2008, Chen, 2016, Freidberg, 2008, Hazeltine and Meiss, 2003, Miyamoto, 2006, Piel, 2010], in particular in tokamak plasmas, a strong external field is applied to confine the plasma leading to the oscillatory nature of the fast motions.

Various models are in use to describe such phenomena. In the kinetic modeling, the unknowns are the number densities of particles, $f \equiv f(t, \mathbf{x}, \mathbf{v})$ depending on time $t \geq 0$, position $\mathbf{x} \in \Omega \subset \mathbf{R}^3$ and velocity $\mathbf{v} \in \mathbf{R}^3$. Such kinetic models provide an appropriate description of turbulent transport in a fairly general context, but in fusion configurations their numerical simulations require to solve a stiff six-dimensional problem, leading to a huge computational cost. To bypass this obstacle, it is classical — see for instance [Garbet et al., 2010] — to use reduced asymptotic models that describe only the slowest part of the plasma dynamics hence effectively reducing both the stiffness of the problem and the number of variables (since fastest variables are omitted). Over the years, due to its rich and fundamental nature, the physically-based derivation of such models has grown as a — still very active — field of its own, often referred to as gyrokinetics. Besides the already mentioned general monographs [Bellan, 2008, Chen, 2016, Freidberg, 2008, Hazeltine and Meiss, 2003, Miyamoto, 2006, Piel, 2010], the reader may consult [Krommes, 2012, Brizard and Hahm, 2007, Matteo, 2017, Scott, 2017, Degond and Filbet, 2016] and references therein as more specialized entering gates to the field.

Despite considerable efforts in recent years, concerning mathematically rigorous derivations from collisionless[1] kinetic equations, the state of art is such that one must choose between linear models that neglect couplings due to self-consistent fields or nonlinear ones set in a deceptively simple geometry. See for instance the introductions and bibliographies of [Han-Kwan, 2011, Lutz, 2013, Herda, 2017] for relatively recent panoramas on the question. For instance, for the kind of problem considered here, on the nonlinear side of the literature the most significant mathematical result — which requires a careful analysis — is restricted to a two-dimensional setting with a constant magnetic field and interactions described through the Poisson equation, and yet validates only half[2] of the slow dynamics; see [Saint-Raymond, 2002], building on [Golse and Saint-Raymond, 1999] and recently revisited in [Miot, 2016].

We consider here a plasma confined by a strong unsteady inhomogeneous magnetic field without any a priori geometric constraint but, in order to allow for such a generality, we do neglect effects of self-consistent fields. The plasma is thus entirely modeled with a scalar linear kinetic equation, where the unknown is one of the number densities of particles. The approach that we follow focuses on the characteristic equations associated with the kinetic conservation law. By itself the study of those equations may follow the classical roadmap of the averaging of ordinary differential equations, as expounded in [Bogoliubov and Mitropolsky, 1961, Sanders et al., 2007]. Yet, here, beyond the body of work already required to follow this road in usual ODE problems, a careful track of the dependence of averaging estimates on initial data, living here in an unbounded phase space, is necessary so as to derive asymptotics for the solutions of the original partial differential equations problem.

To be more specific, the Lorentz force term in our original nondimensionalized kinetic equation is scaled by a large parameter, $1/\varepsilon$, where ε stands for the typical cyclotron period, *i.e.* the typical rotation period of particles about a magnetic field line (or Larmor rotation). The dynamical time scales we focus on are in any case much larger than the cyclotron period and we

[1]See for instance [Herda, 2016, Herda and Rodrigues, 2016] and references therein for an introduction to the corresponding collisional issues.

[2]The nontrivial half, however. This is possible there only because a very specific geometric cancellation uncouples part of the slow dynamics from the remaining one, which is expected to be slaved to it. See however the recent preprint [Bostan, 2018] for a more complete model, derived under more stringent assumptions.

establish asymptotic descriptions in the limit $\varepsilon \to 0$. As is classical in the field, we distinguish between short-time scales that are $\mathcal{O}(1)$ with respect to ε, and long time scales that are $\sim 1/\varepsilon$ in the limit $\varepsilon \to 0$. Correspondingly, slow dynamics refer to dynamics where typical time derivatives are at most of order $\mathcal{O}(1)$ on short-time scales, and at most of order $\mathcal{O}(\varepsilon)$ on long-time scales so that on long time scales two kinds of fast dynamics may co-exist, principal ones at typical speed of order $1/\varepsilon$ and subprincipal ones at typical speed of order 1; see for instance [Cheverry, 2017] for a description of those various oscillations in a specific class of axi-symmetric geometries, without electric field and with a magnetic field nowhere toroidal and whose angle to the toroidal direction is also independent of the poloidal angle. With this terminology in hands, our results may be roughly stated as the identification and mathematical proofs of

(1) a second-order — that is, up to $\mathcal{O}(\varepsilon^2)$ — description of the slow dynamics on short time scales but in arbitrary geometry;
(2) a first-order description of the slow dynamics on long time scales but in an axi-symmetric geometry with a magnetic field everywhere poloidal and an electric field everywhere orthogonal to the magnetic field.

The geometry of the latter is very specific and the proof of such a description is mostly carried out here to illustrate that the short-time second-order description contains all the ingredients to analyze long-time dynamics at first-order. Note that in any case, on long-time scales some restrictions are indeed necessary to ensure that sub-principally fast dynamics do not prevent long-time confinement and are of oscillatory type so that the issue of the identification of a long-time slow dynamics becomes meaningful.

A key feature of our analysis that underpins a treatment of essentially arbitrary fields is that we make no explicit use of any geometric structure, neither Hamiltonian (see for instance the pioneer work of R. G. Littlejohn [Littlejohn, 1979, Littlejohn, 1981, Littlejohn, 1983] and later [Benettin and Sempio, 1994, Frénod and Lutz, 2014]) nor Lagrangian (see [Possanner, 2018]). The main role of these structures in the averaging process is to ease the identification of terms that are asymptotically irrevelant as time-derivatives of small terms. Instead, in the present contribution this explicit identification hinges heavily on the linearity of principal oscillations. As an upset, besides generality, we gain the freedom to use change of variables that are also arbitrary and to focus on slow variables instead of carrying geometric constraints all along.

A key motivation for our methodology is that in the design of well-adapted numerical schemes, that capture the slow part of the dynamics even with discretization meshes too rough to compute stiff scales, one might correspondingly aim at large classes of schemes of arbitrary order; see for instance [Lee, 1983, Filbet and Rodrigues, 2016, Filbet and Rodrigues, 2017]. Likewise our choice of studying first characteristics instead of using directly partial differential equations techniques and our will to prove error estimates echoes the particle-in-cell methodology and its numerical analysis. Alternative PDE-based methods include most notably two-scale convergence analyses [Frénod and Sonnendrücker, 1998, Frénod and Sonnendrücker, 2000] and filtering techniques hinging on ergodic von Neumann's theorem [Bostan, 2010b, Bostan, 2010a]. Two main advantages of going through characteristics are that the limiting partial differential equation is by construction a conservation law for a density distribution and that increasing the order of description may be carried out merely by continuing the argument used to identify the leading order. We benefit from the latter to *prove* for the first time a second-order description in full generality.

2. Definitions and main results

We consider the kinetic PDE

$$\partial_t f^\varepsilon + \mathrm{div}_{\mathbf{x}}(f^\varepsilon \mathbf{v}) + \mathrm{div}_{\mathbf{v}}\left(f^\varepsilon \left(\frac{\mathbf{v}\wedge\mathbf{B}(t,\mathbf{x})}{\varepsilon} + \mathbf{E}(t,\mathbf{x})\right)\right) = 0 \tag{2.1}$$

and its characteristic flow encoded by the following ODEs

$$\begin{cases} \dfrac{\mathrm{d}\,\mathbf{x}}{\mathrm{d}\,t} = \mathbf{v}\,, \\ \dfrac{\mathrm{d}\,\mathbf{v}}{\mathrm{d}\,t} = \dfrac{\mathbf{v}\wedge\mathbf{B}(t,\mathbf{x})}{\varepsilon} + \mathbf{E}(t,\mathbf{x})\,, \end{cases} \tag{2.2}$$

where $\wedge$ denotes the standard vector product on $\mathbf{R}^3$, $\mathbf{B}$ stands for the external magnetic field, $\mathbf{E}$ for the external electric field.

As a preliminary we begin by recalling the classical link between (2.1) and (2.2) and making explicit how it can be used to analyze the slow part of the dynamics hidden in the stiff (2.1).

2.1. **From ODEs to PDEs.** Throughout the present contribution we shall use the following notational conventions. We denote $\Psi_*(\mu)$ the push-forward of μ by Ψ, which can be defined for instance when μ is a distribution and Ψ is a smooth proper map by, for any test-function φ,

$$\int_B \varphi \, \mathrm{d}\, \Psi_*(\mu) \;=\; \int_{\Psi^{-1}(B)} \varphi \circ \Psi \, \mathrm{d}\, \mu.$$

When considering flows associated with ODEs, $\Phi(t, s, \mathbf{y})$ denotes the value at time t of the solution starting from $\mathbf{y}$ at time s and the associated map is $\mathbf{y} \mapsto \Phi(t, s, \mathbf{y})$. In particular the solution to (2.1) starting from f_0 at time 0 is given at time t by $\Phi(t, 0, \cdot)_* \, (f_0)$ where Φ is the flow associated with (2.2).

For general purpose we shall state an abstract proposition, almost tautological, converting estimates on characteristics into estimates on densities. First, to enlighten the meaning of the following statement, identifying measures with their densities, we recall that the "value" at a of $\mathcal{A}_*(\mu)$ the push-forward of μ by $\mathcal{A}$ is essentially the average of μ on the level set $\mathcal{A}^{-1}(\{a\})$. Indeed for any function f at any regular value a of $\mathcal{A}$

$$\mathcal{A}_* \, (f)(a) \;=\; \int_{\mathcal{A}^{-1}(\{a\})} f(\mathbf{y}) \, \frac{\mathrm{d}\, \sigma_a(\mathbf{y})}{\sqrt{\det(\mathrm{d}\, \mathcal{A}(\mathbf{y})(\mathrm{d}\, \mathcal{A}(\mathbf{y}))^*)}} \, ,$$

where σ_a denotes the surface measure on $\mathcal{A}^{-1}(\{a\})$, d denotes the differential operator and * the adjoint operator. For instance if $\mathbf{y} = (y_1, y_2) \in \mathbf{R}^2$, then with $\mathcal{A}(\mathbf{y}) = \|\mathbf{y}\| = \sqrt{y_1^2 + y_2^2}$,

$$\mathcal{A}_* \, (f)(r) \;=\; \int_0^{2\pi} f(r \; \mathbf{e}(\theta)) \, r \; \mathrm{d}\, \theta \, ,$$

where $\mathbf{e}(\theta) = (\cos(\theta), \sin(\theta))$, whereas with $\mathcal{A}(\mathbf{y}) = \frac{1}{2}\|\mathbf{y}\|^2$,

$$\mathcal{A}_*(f)(e) \;=\; \int_0^{2\pi} f(\sqrt{2\, e} \; \mathbf{e}(\theta)) \, \mathrm{d}\, \theta \, .$$

It turns out that the correct way to "average" the stiff equation (2.1) is precisely to push f by a map $\mathcal{A}$ defining a complete[3] set of slow variables.

Proposition 2.1. *Let Φ and Φ_{slow} be flows associated with respective ODEs*

$$\frac{\mathrm{d}\, \mathbf{y}}{\mathrm{d}\, t} = \mathcal{X}(t, \mathbf{y}) \qquad \textit{and} \qquad \frac{\mathrm{d}\, \mathbf{a}}{\mathrm{d}\, t} = \mathcal{X}_{\mathrm{slow}}(t, \mathbf{a})$$

and assume that there exist time-dependent slow maps $\mathcal{A}(t, \cdot)$ and weights $\mathcal{M}(t, \cdot)$ such that for a.e. $t \geq 0$,

$$\|\mathcal{A}(t, \Phi(t, 0, \cdot)) - \Phi_{\mathrm{slow}}(t, 0, \mathcal{A}(0, \cdot))\| \;\leq\; \mathcal{M}(t, \cdot) \, .$$

Then if f solves

$$\partial_t f \; + \; \mathrm{div}_{\mathbf{y}}(\mathcal{X} \, f) \;=\; 0,$$

with initial data a measure f_0 and $F(t, \cdot) = \mathcal{A}(t, \cdot)_* \, f(t, \cdot)$ is the push-forward of f by the slow map $\mathcal{A}$ then for a.e. $t \geq 0$

$$\|F(t, \cdot) - G(t, \cdot)\|_{\dot{W}^{-1,1}} \;\leq\; \int \mathcal{M}(t, \cdot) \, \mathrm{d}\, |f_0| \, ,$$

where G solves

$$\partial_t G \; + \; \mathrm{div}_{\mathbf{a}}(\mathcal{X}_{\mathrm{as}} \, G) \;=\; 0,$$

with initial data $F_0 := \mathcal{A}(0, \cdot)_* \, f_0$.

[3]So that an uncoupled system is obtained in closed form (at the required order).

In the former we have denoted $\dot{W}^{-1,1}$ the dual of $\dot{W}^{1,\infty}$. Incidentally we observe that the distance on $\dot{W}^{-1,1}$ coincides with the 1-Wasserstein distance from optimal transportation. Explicitly

$$\|\mu\|_{\dot{W}^{-1,1}} = \sup_{\|\nabla\varphi\|_{L^\infty}\leq 1} \int \varphi \, \mathrm{d}\, \mu \, .$$

Here and throughout L^p denotes the classical Lebesgue space of index p, $W^{s,p}$ and $\dot{W}^{s,p}$ their corresponding Sobolev spaces at regularity s, respectively in inhomogeneous and homogeneous versions. Associated (semi-)norms are denoted $\|\cdot\|_{L^p}$, $\|\cdot\|_{W^{s,p}}$ and $\|\cdot\|_{\dot{W}^{s,p}}$.

Proof. This stems readily from

$$F(t,\cdot) = \mathcal{A}(t,\Phi(t,0,\cdot))_*\,(f_0)\,, \qquad G(t,\cdot) = \Phi_{\text{slow}}(t,0,\mathcal{A}(0,\cdot))_*\,(f_0)\,,$$

and

$$\|\varphi\circ\mathcal{A}(t,\Phi(t,0,\cdot)) - \varphi\circ\Phi_{\text{slow}}(t,0,\mathcal{A}(0,\cdot))\| \leq \|\nabla\varphi\|_{L^\infty}\,\mathcal{M}(t,\cdot)\,.$$

□

Note that in the foregoing statement, for readability's sake, we have deliberately left domains in time, original variables and slow variables, unspecified. However this may be straightened by classical ODE considerations, notably when fields are continuous, and locally Lipschitz in respectively $\mathbf{y}$ and $\mathbf{a}$ and either the support of f_0 is compact or involved vector-fields grow at most linearly.

2.2. **Slow variables and first-order asymptotics.** Getting back to our concrete system we begin our identification of a slow dynamics.

First, as is classical, we split the magnetic field $\mathbf{B}$ as

$$\mathbf{B}(t,\mathbf{x}) = B(t,\mathbf{x})\;\mathbf{e}_\parallel(t,\mathbf{x}),$$

with $B(t,\mathbf{x}) = \|\mathbf{B}(t,\mathbf{x})\|$. Accordingly we define, for any $\mathbf{x}\in\mathbf{R}^3$ and any time t, the linear operator $\mathbb{J}(t,\mathbf{x})$ as

$$\mathbb{J}(t,\mathbf{x})\,\mathbf{a} = \mathbf{a}\wedge\mathbf{e}_\parallel(t,\mathbf{x})\,. \tag{2.3}$$

The direction of the magnetic field plays a very special role and it is expedient to introduce for velocities an associated decomposition into parallel and orthogonal components

$$\begin{cases} v_\parallel(t,\mathbf{x},\mathbf{v}) = \langle \mathbf{v}\,,\,\mathbf{e}_\parallel(t,\mathbf{x})\rangle\,, \\ \mathbf{v}_\perp(t,\mathbf{x},\mathbf{v}) = \mathbf{v} - v_\parallel(t,\mathbf{x},\mathbf{v})\;\mathbf{e}_\parallel(t,\mathbf{x}) \end{cases}$$

and similarly for the electric field $\mathbf{E}$,

$$\begin{cases} E_\parallel(t,\mathbf{x}) = \langle \mathbf{E}(t,\mathbf{x})\,,\,\mathbf{e}_\parallel(t,\mathbf{x})\rangle\,, \\ \mathbf{E}_\perp(t,\mathbf{x}) = \mathbf{E}(t,\mathbf{x}) - E_\parallel(t,\mathbf{x},\mathbf{v})\;\mathbf{e}_\parallel(t,\mathbf{x})\,, \end{cases}$$

where hereabove $\langle\cdot,\cdot\rangle$ denotes the canonical Euclidean scalar product, and below $\|\cdot\|$ denotes the associated Euclidean norm.

From system (2.2) it is clear that at least one component out of the six-dimensional $(\mathbf{x},\mathbf{v})$ must obey a dynamics forcing oscillations of amplitude of typical size 1 and typical frequency $1/\varepsilon$. However at typical size 1 a five-dimensional slow dynamics survives. This is already suggested by the fact that one may derive from (2.2) for the slow variables $(\mathbf{x}, v_\parallel, w_\perp)$,

$$\begin{cases} \dfrac{\mathrm{d}\,\mathbf{x}}{\mathrm{d}\,t} = \mathbf{v}, \\ \dfrac{\mathrm{d}\,v_\parallel}{\mathrm{d}\,t} = E_\parallel(t,\mathbf{x}) + \langle \mathbf{v}_\perp, \partial_t\,\mathbf{e}_\parallel(t,\mathbf{x}) + \mathrm{d}_{\mathbf{x}}\,\mathbf{e}_\parallel(t,\mathbf{x})\mathbf{v}\rangle\,, \\ \dfrac{\mathrm{d}\,w_\perp}{\mathrm{d}\,t} = \langle \mathbf{E}_\perp(t,\mathbf{x}) - v_\parallel\;(\partial_t\,\mathbf{e}_\parallel(t,\mathbf{x}) + \mathrm{d}_{\mathbf{x}}\,\mathbf{e}_\parallel(t,\mathbf{x})\,\mathbf{v})\,,\mathbf{v}_\perp\rangle\,, \end{cases} \tag{2.4}$$

where $w_\perp = \frac{1}{2}\|\mathbf{v}_\perp\|^2$ and we have used the shorthand $v_\parallel(t)$ for $v_\parallel(t,\mathbf{x}(t),\mathbf{v}(t))$ and similarly for $\mathbf{v}_\perp$.

Our goal is to identify such a slow dynamics, uncoupled from fast oscillations. Roughly speaking, since $\mathbf{v}_\perp$ is expected to weakly converge to zero when ε goes to zero, at leading order

the only issue is to identify the asymptotic behavior of quadratic terms in $\mathbf{v}_\perp$ in (2.4). It turns out that those are responsible for the apparition of terms $w\ \mathrm{div}_{\mathbf{x}}\,\mathbf{e}_\parallel$ in the asymptotic model, set on a reduced phase space, where slow variables $\mathbf{Z} = (\mathbf{y}, v, w)$ live. Introducing the limiting vector field

$$\mathcal{V}_0(t,\mathbf{Z}) \;=\; \begin{pmatrix} v\ \mathbf{e}_\parallel(t,\mathbf{y}) \\ E_\parallel(t,\mathbf{y}) + w\ \mathrm{div}_{\mathbf{x}}\,\mathbf{e}_\parallel(t,\mathbf{y}) \\ -v\,w\ \mathrm{div}_{\mathbf{x}}\,\mathbf{e}_\parallel(t,\mathbf{y}) \end{pmatrix} \tag{2.5}$$

we may state our first significant result.

Theorem 2.2. *Let $\mathbf{E} \in W^{1,\infty}$ and $\mathbf{B}$ be such that $1/B \in W^{1,\infty}$ and $\mathbf{e}_\parallel \in W^{2,\infty}$. There exists a constant C depending polynomially on $\|\mathbf{E}\|_{W^{1,\infty}}$, $\|B^{-1}\|_{W^{1,\infty}}$ and $\|\,\mathbf{e}_\parallel\,\|_{W^{2,\infty}}$ such that if f^ε solves (2.1) with initial data a positive density[4] f_0, then F^ε defined by*

$$F^\varepsilon(t,\mathbf{x},v_\parallel,w_\perp) = \int_{\mathbf{S}_{t,\mathbf{x}}} f^\varepsilon(t,\mathbf{x},v_\parallel\ \mathbf{e}_\parallel(\mathbf{x}) \,+\, \sqrt{2\,w_\perp}\ \widehat{\mathbf{e}})\ \mathrm{d}\,\sigma_{t,\mathbf{x}}(\widehat{\mathbf{e}}),$$

with $\mathbf{S}_{t,\mathbf{x}} = \{\mathbf{e}_\parallel(t,\mathbf{x})\}^\perp \cap \mathbf{S}^2$ and $\sigma_{t,\mathbf{x}}$ its canonical line-measure, satisfies for a.e. $t \geq 0$

$$\|F^\varepsilon(t,\cdot) - G(t,\cdot)\|_{\dot{W}^{-1,1}} \leq C\,\varepsilon\,e^{C\,t^4} \int_{\mathbf{R}^3\times\mathbf{R}^3} e^{C\,t\,\|\mathbf{v}\|^3}\,\|\mathbf{v}\|\,(1+\|\mathbf{v}\|^2)\,f_0(\mathbf{x},\mathbf{v})\ \mathrm{d}\,\mathbf{x}\ \mathrm{d}\,\mathbf{v},$$

where G solves

$$\partial_t G \,+\, \mathrm{div}_{\mathbf{Z}}\,(\mathcal{V}_0\,G) \;=\; 0, \tag{2.6}$$

with $\mathcal{V}_0$ given by (2.5) and the initial datum G_0 is

$$G_0(\mathbf{Z}) = \int_{\mathbf{S}_{0,\mathbf{y}}} f_0(\mathbf{y}, v\ \mathbf{e}_\parallel(\mathbf{y}) \,+\, \sqrt{2\,w}\ \widehat{\mathbf{e}})\ \mathrm{d}\,\sigma_{0,\mathbf{y}}(\widehat{\mathbf{e}})\,.$$

Theorem 2.2 is proved in Section 4.4. The underlying vector field $\mathcal{V}_0$ of the asymptotic model being divergence-free, many conservation laws already come as consequences of the asymptotic model. Yet as we state below a few more may be obtained if one assumes classical extra structure on electromagnetic fields.

Proposition 2.3. *Assume that $\mathbf{E} = -\nabla_{\mathbf{x}}\phi$ where the couple (ϕ, B) does not depend on time and suppose that the confining magnetic field satisfies the Gauss' law*

$$\mathrm{div}_{\mathbf{x}}\,\mathbf{B} \;=\; 0.$$

Then solutions to the asymptotic model (2.6) satisfy

- *the conservation of energy*

$$\partial_t\left(\left(\frac{v^2}{2} + w + \phi\right)\,G\right) \,+\, \mathrm{div}_{\mathbf{Z}}\left(\left(\frac{v^2}{2} + w + \phi\right)\,\mathcal{V}_0\,G\right) \;=\; 0;$$

- *the conservation of the classical adiabatic invariant $\mu_\perp = w/B$*

$$\partial_t\left(\frac{w}{B}\,G\right) \,+\, \mathrm{div}_{\mathbf{Z}}\left(\frac{w}{B}\,\mathcal{V}_0\,G\right) \;=\; 0.$$

Proof. For the asymptotic model (2.6), the balance law for the kinetic energy is

$$\partial_t\left(\left(\frac{v^2}{2} + w\right)\,G\right) \,+\, \mathrm{div}_{\mathbf{Z}}\left(\left(\frac{v^2}{2} + w\right)\,\mathcal{V}_0\,G\right) \;=\; G\,v\,E_\parallel,$$

which is a conservation law only if $E_\parallel \equiv 0$. Then if $\mathbf{E}$ derives from a potential, $\mathbf{E} = -\nabla\phi$, the corresponding balance law for the total energy of the asymptotic model is

$$\partial_t\left(\left(\frac{v^2}{2} + w + \phi\right)\,G\right) \,+\, \mathrm{div}_{\mathbf{Z}}\left(\left(\frac{v^2}{2} + w + \phi\right)\,\mathcal{V}_0\,G\right) \;=\; G\,\partial_t\phi,$$

which reduces to the claimed conservation law when $\partial_t\phi \equiv 0$.

[4]Results would equally well hold with measures of arbitrary sign, but we stick to densities to provide nicer integral formulations for push-forwards when available, and to positive densities to remove absolute values in error bounds.

Note moreover that from

$$\frac{\operatorname{div}_{\mathbf{x}} \mathbf{B}(t,\mathbf{x})}{(B(t,\mathbf{x}))^2} \;=\; \frac{\operatorname{div}_{\mathbf{x}} \mathbf{e}_{\parallel}(t,\mathbf{x})}{B(t,\mathbf{x})} - \mathbf{e}_{\parallel}(t,\mathbf{x}) \cdot \nabla_{\mathbf{x}} \left(\frac{1}{B}\right)(t,\mathbf{x})$$

follows for the asymptotic model the balance law

$$\partial_t \left(\frac{w}{B}\, G\right) \;+\; \operatorname{div}_{\mathbf{Z}} \left(\frac{w}{B}\, \mathcal{V}_0\, G\right) \;=\; -\, G\, \frac{w}{B^2}\, \left(\partial_t B + v \operatorname{div}_{\mathbf{x}} \mathbf{B}\right),$$

which is indeed a conservation law when $\partial_t B \equiv 0$ and $\mathbf{B}$ is divergence-free. □

2.3. **Second-order asymptotics.** Though already instructive, equation (2.6) fails to capture parts of the slow dynamics that are too slow, for instance it does not describe the evolution of w/B (when $\operatorname{div}_{\mathbf{x}} \mathbf{B} \equiv 0$). One way to correct this is to derive a higher-order description of the slow dynamics.

It is at this next order that are found macroscopic velocities, including those classically known as the $\mathbf{E} \times \mathbf{B}$ drift, the curvature drift, the grad-B drift and the magnetic rotational drift, that with notation below read respectively $\mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{y})$, $v^2\, \mathbf{U}_{\mathrm{curv}}(t,\mathbf{y})$, $e\, \mathbf{U}_{\nabla B\times\mathbf{B}}(t,\mathbf{y})$, and $w\, \mathbf{U}_{\operatorname{curl} \mathbf{e}_{\parallel}}(t,\mathbf{y})$. Those have simple expressions in terms of vectors fields depending only on time t and space $\mathbf{y}$ variables, and defined themselves as

$$\left\{\begin{aligned}
&\mathbf{U}_{\mathbf{E}\times\mathbf{B}} \;:=\; \frac{\mathbb{J}\,\mathbf{E}}{B} \;=\; \frac{\mathbf{E}\wedge\mathbf{B}}{B^2}\,,\\
&\mathbf{U}_{\mathrm{curv}} \;:=\; -\frac{\mathbb{J}}{B}(\mathrm{d}_{\mathbf{x}}\, \mathbf{e}_{\parallel}\;\mathbf{e}_{\parallel}) \;=\; -\frac{1}{B^2}\;(\mathrm{d}_{\mathbf{x}}\, \mathbf{e}_{\parallel}\;\mathbf{e}_{\parallel}) \wedge \mathbf{B}\,,\\
&\mathbf{U}_{\nabla B\times\mathbf{B}} \;:=\; \mathbb{J}\nabla_{\mathbf{x}} \left(\frac{1}{B}\right) \;=\; -\frac{1}{B^3}\nabla_{\mathbf{x}} B \wedge \mathbf{B},\\
&\mathbf{U}_{\operatorname{curl} \mathbf{e}_{\parallel}} \;:=\; \frac{1}{B}\langle \operatorname{curl}_{\mathbf{x}} \mathbf{e}_{\parallel}, \mathbf{e}_{\parallel}\rangle\; \mathbf{e}_{\parallel},
\end{aligned}\right. \tag{2.7}$$

where $\mathbb{J}$ is given in (2.3). Since the direction of the magnetic field $\mathbf{e}_{\parallel}$ is allowed to depend on time, another drift is present, given by $v\, \mathbf{U}_{\partial_t}(t,\mathbf{y})$ where

$$\mathbf{U}_{\partial_t} \;:=\; -\frac{\mathbb{J}}{B}\, \partial_t\, \mathbf{e}_{\parallel} \;=\; -\frac{1}{B^2}\, \partial_t\, \mathbf{e}_{\parallel} \wedge \mathbf{B}\,. \tag{2.8}$$

Since it appears repeatedly it is convenient to introduce a piece of notation for a special combination of $\mathbf{U}_{\partial_t}$ and $\mathbf{U}_{\mathrm{curv}}$,

$$\mathbf{\Sigma}(t,\mathbf{y},v) \;=\; \mathbf{U}_{\partial_t}(t,\mathbf{y}) \;+\; v\, \mathbf{U}_{\mathrm{curv}}(t,\mathbf{y})\,. \tag{2.9}$$

With the above definitions we may write the full drift vector field $\mathbf{U}_{\mathrm{drift}}(t,\mathbf{Z})$ in the concise form

$$\mathbf{U}_{\mathrm{drift}}(t,\mathbf{Z}) \;=\; (\mathbf{U}_{\mathbf{E}\times\mathbf{B}} \;+\; v\, \mathbf{\Sigma})\,(t,\mathbf{y},v) \;+\; w\;(\mathbf{U}_{\operatorname{curl} \mathbf{e}_{\parallel}} + \mathbf{U}_{\nabla B\times\mathbf{B}})\,(t,\mathbf{y}). \tag{2.10}$$

where $\mathbf{Z} = (\mathbf{y}, v, w)$ stands for our set of slow variables in the asymptotic model. For the sake of comparison with the existing literature we observe the equivalent reformulations that may be derived from $(\mathrm{d}_{\mathbf{x}}\, \mathbf{e}_{\parallel})^*\, \mathbf{e}_{\parallel} = 0$, a consequence of $\mathbf{e}_{\parallel}$ being unitary valued,

$$\left\{\begin{aligned}
&\mathrm{d}_{\mathbf{x}}\mathbf{e}_{\parallel}\;\mathbf{e}_{\parallel} \;=\; \operatorname{curl}_{\mathbf{x}} \mathbf{e}_{\parallel} \wedge \mathbf{e}_{\parallel}\,,\\
&\mathbf{U}_{\mathrm{curv}} \;=\; -\frac{1}{B}(\operatorname{curl}_{\mathbf{x}} \mathbf{e}_{\parallel} \wedge \mathbf{e}_{\parallel}) \wedge \mathbf{e}_{\parallel}
\end{aligned}\right.$$

and observing that

$$(\operatorname{curl}_{\mathbf{x}} \mathbf{e}_{\parallel} \wedge \mathbf{e}_{\parallel}) \wedge \mathbf{e}_{\parallel} \;=\; -\operatorname{curl}_{\mathbf{x}} \mathbf{e}_{\parallel} \;+ \langle \operatorname{curl}_{\mathbf{x}} \mathbf{e}_{\parallel}, \mathbf{e}_{\parallel}\rangle\; \mathbf{e}_{\parallel}$$

we get that

$$\mathbf{U}_{\operatorname{curl} \mathbf{e}_{\parallel}} + \mathbf{U}_{\mathrm{curv}} \;=\; \frac{\operatorname{curl}_{\mathbf{x}} \mathbf{e}_{\parallel}}{B} \;=\; \frac{\operatorname{div}_{\mathbf{x}} \mathbb{J}}{B} \tag{2.11}$$

The vector-field involved in our higher-order description of the complete slow dynamics is then given by

$$\mathcal{V}^{\varepsilon} \;=\; \mathcal{V}_0 \;+\; \varepsilon\, \mathcal{V}_1,$$

where the first order contribution $\mathcal{V}_1(t,\mathbf{Z})$ is

$$\mathcal{V}_1(t,\mathbf{Z}) \;=\; \begin{pmatrix} \mathbf{U}_{\rm drift}(t,\mathbf{Z}) \\ \langle \mathbf{\Sigma}(t,\mathbf{y},v), \mathbf{E}(t,\mathbf{y})\rangle + w\ \mathrm{div}_{\mathbf{x}}\, \mathbf{\Sigma}(t,\mathbf{y},v) \\ -w\ [\,\langle \mathbf{U}_{\rm curv}(t,\mathbf{y}), \mathbf{E}(t,\mathbf{y})\rangle + \mathrm{div}_{\mathbf{x}}\,(\mathbf{U}_{\mathbf{E}\times\mathbf{B}} + v\mathbf{\Sigma})\,(t,\mathbf{y},v)\,] \end{pmatrix}. \tag{2.12}$$

The foregoing vector-field describes the dynamics of variables that are ε-corrections of $(\mathbf{x}, v_\parallel, w_\perp)$ but that are slower than those. The corrected spatial position

$$\mathbf{x}^\varepsilon_{\rm gc}(t,\mathbf{x},\mathbf{v}) \;:=\; \mathbf{x} + \varepsilon \frac{\mathbb{J}(t,\mathbf{x})\,\mathbf{v}}{B(t,\mathbf{x})} \;=\; \mathbf{x} + \varepsilon \frac{\mathbf{v}\wedge\mathbf{B}(t,\mathbf{x})}{(B(t,\mathbf{x}))^2} \tag{2.13}$$

is well-known as the guiding center position, whereas the corrected parallel velocity is given as

$$v^\varepsilon_{\rm gc}(t,\mathbf{x},\mathbf{v}) \;:=\; v_\parallel + \;\varepsilon\, \langle \mathbf{v}_\perp,\, \mathbf{\Sigma}(t,\mathbf{x},v_\parallel)\rangle + \;\frac{\varepsilon}{2B(t,\mathbf{x})}\, \langle \mathbb{J}(t,\mathbf{x})\,\mathbf{v}_\perp,\, \Re(\mathrm{d}_{\mathbf{x}}\,\mathbf{e}_\parallel(t,\mathbf{x}))\,\mathbf{v}_\perp\rangle \tag{2.14}$$

and the corrected version of the part of the kinetic energy in the plane perpendicular to the magnetic field direction is

$$\begin{cases} w^\varepsilon_{\rm gc}(t,\mathbf{x},\mathbf{v}) \;:=\; w_\perp \;\; - \varepsilon\, \langle \mathbf{v}_\perp,\, \mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{x}) + v_\parallel\, \mathbf{\Sigma}(t,\mathbf{x},v_\parallel)\rangle \\ \qquad\qquad\qquad\qquad\qquad\qquad - \dfrac{\varepsilon\, v_\parallel}{2B(t,\mathbf{x})}\, \langle \mathbb{J}(t,\mathbf{x})\,\mathbf{v}_\perp,\, \Re(\mathrm{d}_{\mathbf{x}}\,\mathbf{e}_\parallel(t,\mathbf{x}))\,\mathbf{v}_\perp\rangle\,, \end{cases} \tag{2.15}$$

where $w_\perp = \frac{1}{2}\|\mathbf{v}_\perp\|^2$, whereas $\Re$ denotes the symmetric part

$$\Re(\mathbf{A}) \;=\; \frac{1}{2}\,(\mathbf{A} + \mathbf{A}^*), \tag{2.16}$$

with $\mathbf{A}^*$ denoting the adjoint of $\mathbf{A}$. Therefore, our global sets of slower components are derived at time t from $(\mathbf{x},\mathbf{v})$ by

$$\mathbf{Z}^\varepsilon_{\rm gc}(t,\mathbf{x},\mathbf{v}) \;=\; (\mathbf{x}^\varepsilon_{\rm gc}, v^\varepsilon_{\rm gc}, w^\varepsilon_{\rm gc})(t,\mathbf{x},\mathbf{v}).$$

We can now state our main theorem.

Theorem 2.4. *Let $\mathbf{E}\in W^{2,\infty}$ and $\mathbf{B}$ be such that $1/B\in W^{2,\infty}$ and $\mathbf{e}_\parallel \in W^{3,\infty}$. There exists a constant C depending polynomially on $\|\mathbf{E}\|_{W^{2,\infty}}$, $\|B^{-1}\|_{W^{2,\infty}}$ and $\|\,\mathbf{e}_\parallel\,\|_{W^{3,\infty}}$ such that if f^ε solves (2.1) with initial data a positive density f_0, then F^ε defined by*

$$F^\varepsilon(t,\cdot) \;=\; \mathbf{Z}^\varepsilon_{\rm gc}(t,\cdot)_*\,(f^\varepsilon(t,\cdot))$$

satisfies for a.e. $t\geq 0$

$$\|F^\varepsilon(t,\cdot) - G^\varepsilon(t,\cdot)\|_{\dot W^{-1,1}} \leq C\,\varepsilon^2\, e^{C\,t^4\,(1+\varepsilon\,t)} \int_{\mathbf{R}^6} e^{C\,t\,\|\mathbf{v}\|^3\,(1+\varepsilon\,\|\mathbf{v}\|)}\,(1+\|\mathbf{v}\|^4)\,f_0(\mathbf{x},\mathbf{v})\;\mathrm{d}\,\mathbf{x}\;\mathrm{d}\,\mathbf{v},$$

where G^ε solves

$$\partial_t G^\varepsilon \;+\; \mathrm{div}_{\mathbf{Z}}\,(\mathcal{V}^\varepsilon\, G^\varepsilon) \;=\; 0, \tag{2.17}$$

with initial data G^ε_0 given by

$$G^\varepsilon_0 \;=\; \mathbf{Z}^\varepsilon_{\rm gc}(0,\cdot)_*\,(f_0)\,.$$

The proof of this asymptotic result is given in Section 4.6.

A few comments on the structure of the asymptotic model are now in order. To begin with we observe that $\mathcal{V}^\varepsilon(t,\cdot)$ is still divergence-free. This follows from Lemma 2.5 below and the fact that $\mathrm{div}_{\mathbf{x}}\,\mathrm{div}_{\mathbf{x}}(\mathbb{J}/B) = 0$ by the skew-symmetry of values of $\mathbb{J}$.

Lemma 2.5. *Consider $\mathbf{U}_{\mathrm{curl}\,\mathbf{e}_\parallel}$, $\mathbf{U}_{\rm curv}$ and $\mathbf{U}_{\nabla B\times\mathbf{B}}$ defined in (2.7). Then we have*

$$\mathrm{div}_{\mathbf{x}}\left(\frac{\mathbb{J}}{B}\right) \;=\; \mathbf{U}_{\mathrm{curl}\,\mathbf{e}_\parallel} + \mathbf{U}_{\rm curv} + \mathbf{U}_{\nabla B\times\mathbf{B}},$$

where $\mathbb{J}$ is given in (2.3) and $B = \|\mathbf{B}\|$.

Proof. Straightforward by chain rule and (2.11) since $\mathbf{U}_{\mathrm{curl}\,\mathbf{e}_\parallel} + \mathbf{U}_{\rm curv} = \mathrm{div}_{\mathbf{x}}\,(\mathbb{J})\,/B$ and $\mathbf{U}_{\nabla B\times\mathbf{B}} = \mathbb{J}\nabla_{\mathbf{x}}(1/B)$. □

Then we observe that we also have

Proposition 2.6. *Assume that* $\mathbf{E} = -\nabla_{\mathbf{x}}\phi$*, where* ϕ *does not depend on time. Then solutions to the asymptotic model* (2.17) *satisfy the conservation of energy*

$$\partial_t \left(\left(\frac{v^2}{2} + w + \phi \right) G^\varepsilon \right) + \operatorname{div}_{\mathbf{Z}} \left(\left(\frac{v^2}{2} + w + \phi \right) \mathcal{V}^\varepsilon G^\varepsilon \right) = 0;$$

Proof. If $\mathbf{E}$ derives from a potential, $\mathbf{E} = -\nabla\phi$, then one obtains the following balance law for the total energy of the second-order asymptotic model

$$\partial_t \left(\left(\frac{v^2}{2} + w + \phi \right) G^\varepsilon \right) + \operatorname{div}_{\mathbf{Z}} \left(\left(\frac{v^2}{2} + w + \phi \right) \mathcal{V}^\varepsilon G^\varepsilon \right) = G^\varepsilon \, \partial_t \phi$$

by using Lemma 2.5 and observing that

$$-\left\langle \operatorname{div}_{\mathbf{x}} \left(\frac{\mathbb{J}}{B} \right), \mathbf{E} \right\rangle + \operatorname{div}_{\mathbf{x}} \mathbf{U}_{\mathbf{E}\times\mathbf{B}} = \operatorname{Tr} \left(\frac{\mathbb{J}}{B} \, \mathrm{d}_{\mathbf{x}} \mathbf{E} \right) = 0$$

since $\mathbb{J}$ is skew-symmetric and $\mathrm{d}_{\mathbf{x}} \mathbf{E}$ is symmetric. From this stems the claimed conservation of energy when $\partial_t \phi \equiv 0$. □

2.4. **Long-time asymptotics in a toroidal axi-symmetric geometry.** Another way to unravel the dynamics of slower components is to derive asymptotics that hold on time scales of typical size $1/\varepsilon$. Yet this seems doable only if the dynamical geometry of the first asymptotic model captured by Theorem 2.2 is sufficiently confining to ensure that the motion at speed of typical size 1 is purely oscillatory and thus may be uncoupled from a dynamics evolving with macroscopic velocities of typical size ε.

Our claim is that when such conditions are satisfied the proof of Theorem 2.4, and more specifically the normal form on which it hinges (see System (4.20)), contains sufficient ingredients to identify this long-time dynamics. To support this claim we illustrate it with a consideration of one of the simplest non trivial confining geometries.

We fix now a unitary vector $\mathbf{e}_z$ and for any $\mathbf{x} \in \mathbf{R}^3$ define the coordinate of $\mathbf{x}$ along $\mathbf{e}_z$ and its distance to the axis $\mathbf{R}\,\mathbf{e}_z$

$$z(\mathbf{x}) = \langle \mathbf{e}_z, \mathbf{x} \rangle, \qquad r(\mathbf{x}) = \| \mathbf{e}_z \wedge \mathbf{x} \| .$$

We assume that for some $r_0 > 0$, where $r(\mathbf{x}) \geq r_0$, $\mathbf{B}$ and $\mathbf{E}$ are axi-symmetric, $\mathbf{B}$ is stationary and toroidal and $\mathbf{E}$ is orthogonal to $\mathbf{B}$, that is,

$$\mathbf{e}_\parallel(\mathbf{x}) = \frac{\mathbf{e}_z \wedge \mathbf{x}}{r(\mathbf{x})}, \qquad B(t,\mathbf{x}) = b(r(\mathbf{x}), z(\mathbf{x})), \qquad E_\parallel(t,\mathbf{x}) = 0,$$

and

$$\mathbf{E}_\perp(t,\mathbf{x}) = E_r(t, r(\mathbf{x}), z(\mathbf{x}))\, \mathbf{e}_r(\mathbf{x}) + E_z(t, r(\mathbf{x}), z(\mathbf{x}))\, \mathbf{e}_z$$

for some smooth b, E_r, E_z, with b non vanishing and

$$\mathbf{e}_r(\mathbf{x}) = \mathbf{e}_\parallel(\mathbf{x}) \wedge \mathbf{e}_z .$$

Under the foregoing geometric assumptions, we have both $E_\parallel \equiv 0$ and $\operatorname{div}(\mathbf{e}_\parallel) \equiv 0$ so that the only motion at speed of typical size 1 is the rotation of $\mathbf{x}$ around $\mathbf{e}_z$ at angular velocity $v_\parallel$. Since by axi-symmetric assumption the corresponding angle is easily factored out one may expect to capture a slow dynamics at typical speed ε in variables $(r, z, v_\parallel, w_\perp)$. This is the content of the next theorem. See Remark 5.4 for some hints on the relaxation of the assumptions made here for simplicity.

The involved asymptotic vector field is $\varepsilon\mathcal{W}_1$ with $\mathcal{W}_1$ defined as

$$\mathcal{W}_1(t,\mathbf{Z}) = \begin{pmatrix} -\dfrac{E_z}{b}(t,r,z) - w\,\partial_z \left(\dfrac{1}{b}\right)(r,z) \\ \dfrac{E_r}{b}(t,r,z) + \dfrac{v^2}{r\,b(r,z)} + w\,\partial_r \left(\dfrac{1}{b}\right)(r,z) \\ \dfrac{v}{r}\left(\dfrac{E_z}{b}(t,r,z) + w\,\partial_z \left(\dfrac{1}{b}\right)(r,z) \right) \\ w\left[\partial_r \left(\dfrac{E_z}{b}\right)(t,r,z) - \partial_z \left(\dfrac{E_r}{b}\right)(t,r,z) - \dfrac{v^2}{r}\,\partial_z \left(\dfrac{1}{b}\right)(r,z) \right] \end{pmatrix},$$

where the new slow variable is $\mathbf{Z} = (r, z, v, w)$.

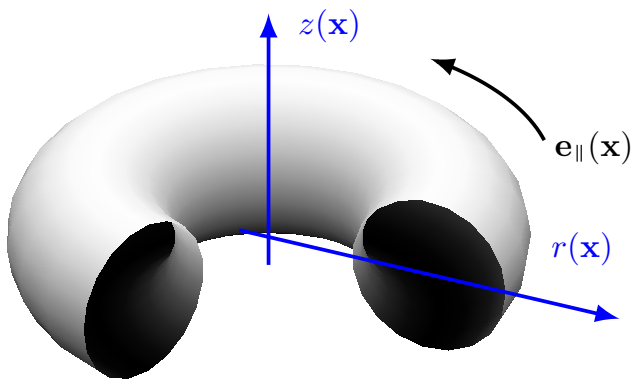

FIGURE 1. Representation of the torus local frame $(\mathbf{e}_r(\mathbf{x}), \mathbf{e}_\parallel(\mathbf{x}), \mathbf{e}_z(\mathbf{x}))$ where the magnetic field is along the unit vector field $\mathbf{e}_\parallel$ whereas the electric field $\mathbf{E}$ is orthogonal to the magnetic field $\mathbf{B}$.

Remark 2.7. *On the two first components of $\mathcal{W}_1$ one readily identifies that in the present geometry along slower variables only survive as spatial drifts, the $\mathbf{E}\times\mathbf{B}$ and grad-B drifts. This is due to the fact that here $\mathrm{curl}_{\mathbf{x}}\,\mathbf{e}_\parallel$ vanishes identically in the zone of interest.*

Theorem 2.8. *Let $\mathbf{B}$ be a stationary, axi-symmetric and toroidal magnetic field and $\mathbf{E}$ be an axi-symmetric electric field orthogonal to $\mathbf{B}$, with $(E_r, E_z, 1/b) \in W^{2,\infty}$ in the region where $r(\mathbf{x}) \geq r_0$ for some r_0. For any $r_1 > r_0$, there exist positive constants ε_0 τ_0 and C_0, $(1/\varepsilon_0, 1/\tau_0, C_0)$ depending polynomially on $1/r_0$, $1/(r_1 - r_0)$ and $\|(E_r, E_z, 1/b)\|_{W^{2,\infty}([r_0,\infty[\times\mathbf{R})}$, such that the following holds with*

$$\varepsilon_{max}(R_0) := \frac{\varepsilon_0}{1+R_0} \qquad \text{and} \qquad T_{max}(R_0) := \frac{\tau_0}{1+R_0^2}\,.$$

Consider f^ε a solution to (2.1) with initial datum a positive density f_0 supported where

$$r(\mathbf{x}) \geq r_1 \qquad \text{and} \qquad \|\mathbf{v}\| \leq R_0$$

for some $R_0 > 0$ and define F^ε as

$$F^\varepsilon(t,r,z,v,e) = \int_0^{2\pi}\int_0^{2\pi} f^\varepsilon(t, r\,\mathbf{e}_r^\theta + z\,\mathbf{e}_z, v\,\mathbf{e}_\parallel(r\,\mathbf{e}_r^\theta + z\,\mathbf{e}_z) + \sqrt{2\,w}\;\mathbf{e}_\perp^{\theta,\varphi})\; r\,\mathrm{d}\varphi\,\mathrm{d}\theta$$

with

$$\mathbf{e}_r^\theta = \cos(\theta)\,\mathbf{e}_x + \sin(\theta)\,\mathbf{e}_y\,, \qquad \mathbf{e}_\perp^{\theta,\varphi} = \cos(\varphi)\,\mathbf{e}_r^\theta + \sin(\varphi)\,\mathbf{e}_z\,,$$

where $(\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z)$ is a fixed[5] orthonormal basis. Then provided that

$$0 < \varepsilon \leq \varepsilon_{max}(R_0)\,,$$

we have for a.e. $0 \leq t \leq T_{max}(R_0)/\varepsilon$

$$\|F^\varepsilon(t,\cdot) - G^\varepsilon(t,\cdot)\|_{\dot{W}^{-1,1}} \leq C_0\,\varepsilon \int_{\mathbf{R}^3\times\mathbf{R}^3} e^{C\,\varepsilon\,t\,\|\mathbf{v}\|^4}\,(1+\|\mathbf{v}\|^3)\,f_0(\mathbf{x},\mathbf{v})\;\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{v},$$

where G^ε solves

$$\partial_t G^\varepsilon + \varepsilon\,\mathrm{div}_{\mathbf{Z}}\,(\mathcal{W}_1\,G^\varepsilon) = 0, \tag{2.18}$$

with initial datum G_0 given by

$$G_0(\mathbf{Z}) = \int_0^{2\pi}\int_0^{2\pi} f_0(\,r\,\mathbf{e}_r^\theta + z\,\mathbf{e}_z, v\,\mathbf{e}_\parallel(r\,\mathbf{e}_r^\theta + z\,\mathbf{e}_z) + \sqrt{2\,w}\;\mathbf{e}_\perp^{\theta,\varphi})\;r\,\mathrm{d}\varphi\,\mathrm{d}\theta\,.$$

Again note that averaging formulas coincide with push-forwards by the slow map $(\mathbf{x},\mathbf{v}) \mapsto (r, z, v_\parallel, w_\perp)(t,\mathbf{x},\mathbf{v})$. Furthermore, we observe that $r\,\mathcal{W}_1$ is divergence-free and provide the following analogous to Proposition 2.3.

[5] $(\mathbf{e}_x, \mathbf{e}_y)$ are somehow arbitrary and the particular choice made here does not change F^ε. In contrast we recall that $\mathbf{e}_z$ plays a special role as it directs the axis of symmetry.

Proposition 2.9. *Suppose that* $\mathbf{E}$ *derives from a stationary axi-symmetric potential*

$$\phi(\mathbf{x}) = \phi(r(\mathbf{x}), z(\mathbf{x}))\,, \qquad\qquad \mathbf{E} = -\nabla_{\mathbf{x}}\phi.$$

Then solutions to the asymptotic model (2.18) *satisfy*

- *the conservation of energy;*
- *the conservation of the classical adiabatic invariant.*

Proof. When $\mathbf{E}$ derives from an axi-symmetric potential as above, the corresponding balance law for the total energy of the asymptotic model (2.18) is

$$\partial_t\left(\left(\frac{v^2}{2} + w + \phi\right) G^\varepsilon\right) + \varepsilon\,\mathrm{div}_{\mathbf{Z}}\left(\left(\frac{v^2}{2} + w + \phi\right)\mathcal{W}_1\, G^\varepsilon\right) = G^\varepsilon\,\partial_t\phi$$

which is a conservation law if furthermore $\partial_t\phi \equiv 0$. Moreover observe that $\mathbf{B}$ is divergence-free in the present configuration and that the asymptotic model (2.18) comes with the balance law

$$\partial_t\left(\frac{w}{b}\,G^\varepsilon\right) + \varepsilon\,\mathrm{div}_{\mathbf{Z}}\left(\frac{w}{b}\,\mathcal{W}_1\,G^\varepsilon\right) = \varepsilon\,G^\varepsilon\,\frac{w}{b^2}\,(\partial_r E_z - \partial_z E_r)\,,$$

which is a conservation law if $\mathbf{E}$ is curl-free, hence in particular if $\mathbf{E}$ derives from a potential. □

2.5. **Further comments and numerical illustrations.** Though we have chosen not add this level of (mostly notational) complexity, the introduction of parameter dependencies in fields $\mathbf{B}$ and $\mathbf{E}$ would be immaterial to our analysis provided they satisfy upper bounds on $1/B$, $\mathbf{E}$ and the needed number of derivatives of $(\mathbf{B}, \mathbf{E})$. See the related Remarks 3.14 and 3.14. In particular in this context one may expand and simplify further our asymptotic systems if one assumes an expansion of $\mathbf{B}$ and $\mathbf{E}$ with respect to ϵ or an ϵ-ordering of gradients, or likewise one may perform a second asymptotic expansion with respect to another small parameter...

We believe that our leading-order slow variables $(\mathbf{x}, v_\parallel, w_\perp)$ are both simple and natural. Yet many other choices have been used in the literature, and for comparison we provide in Appendix A versions of our main results with another commonly-used choice, $(\mathbf{x}, v_\parallel, w_\perp/B)$. Once a leading-order choice has been made, the higher-order corrections added to it to reach varying order of slowness are uniquely determined provided that a normalization is chosen. All through our analysis our implicit choice is to enforce that corrections have no slow component in the sense that they have zero mean with respect to the fast angle. See the related Remarks 3.9 and 4.5.

Though we have chosen to focus on the description of the slow dynamics, the method would equally well provide a detailed description of the oscillations as slaved to the evolution of the slow variables. We stress that in most of methods relying on variational principles one needs to provide both descriptions jointly even though the oscillating part is subordinated to the slow part, as those methods proceed by performing full changes of variables in the original phase-space preserving the geometric structure under consideration. Note that in principle to be fully justified from a mathematical point of view this requires a careful tracking of how small ε must be to guarantee that performed transformations are indeed changes of variables. Here instead, with the exception of results from Section 2.4 where ε is constrained to ensure sufficient confinement on large times, our results are free of smallness constraint on the gradient of force fields $(\mathbf{E}, \mathbf{B})$ and on the initial data.

To conclude the presentation of our results, we provide the reader with some numerical simulations illustrating and hopefully making more concrete respective error bounds. Since it is simpler to visualize we restrict numerical experiments to single-particle simulations.

In the present numerical experiments, we choose the electric field equal to zero and the magnetic field $\mathbf{B}$ as

$$\mathbf{B}(t,\mathbf{x}) = -\frac{20}{\|\mathbf{x}\|^5}\begin{pmatrix} 3\,x\,z \\ 3\,y\,z \\ 2\,z^2 - x^2 - y^2 \end{pmatrix},$$

where $\mathbf{x} = (x, y, z)$. The initial data is

$$\mathbf{x}(0) = (10, 0, 0), \qquad \mathbf{v}(0) = \left(0, 6\cos\left(\frac{\pi}{3}\right), 6\sin\left(\frac{\pi}{3}\right)\right),$$

We approximate the solution of the initial system (2.2) using a fourth order Runge-Kutta scheme with a time step sufficiently small to resolve oscillations and compare it with the numerical solution obtained with the first order approximation corresponding to the characteristic curves of (2.5)-(2.6) (given in Proposition 4.6) and with the second order approximation corresponding to the characteristic curves of (2.5), (2.12) and (2.17) (given in Proposition 4.9).

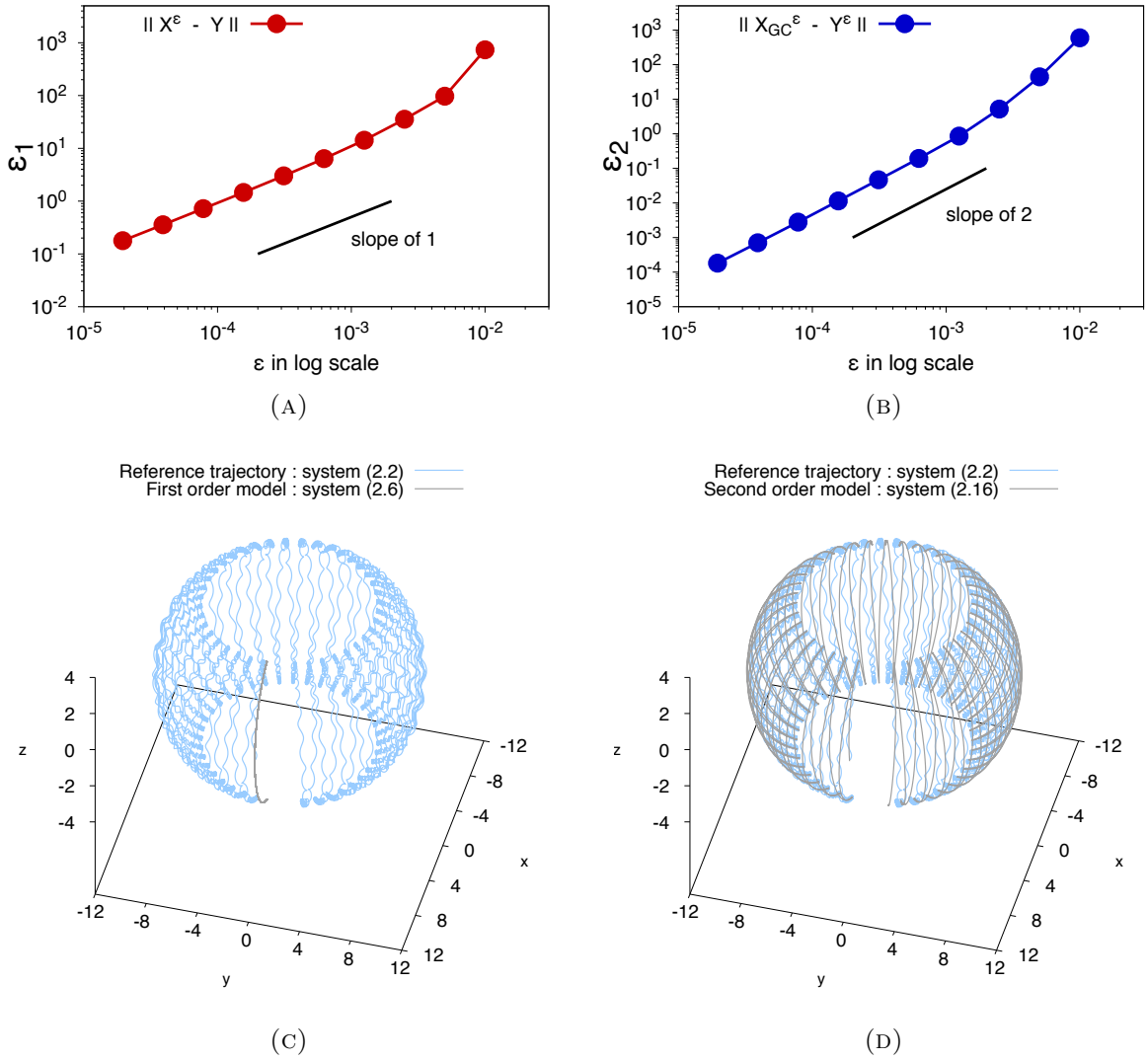

FIGURE 2. Numerical simulation: (A) error between the first order model and (2.2) (B) error between the second order model and (2.2) (C) particle trajectory obtained with the first order model and (2.2) (D) particle trajectory obtained with the second order model and (2.2).

On the one hand, Figure 2-(A) represents the error

$$\mathcal{E}_1(\varepsilon) = \int_0^{10} \|\mathbf{x}^\varepsilon(t) - \mathbf{y}(t)\| \, \mathrm{d}\, t,$$

where $\mathbf{x}^\varepsilon$ is the spatial component of the solution $(\mathbf{x}^\varepsilon, \mathbf{v}^\varepsilon)$ to (2.2) and $\mathbf{y}$ is the spatial component of $(\mathbf{y}, v, w)$ satisfying the differential system with initial data as in Proposition 4.6 (and corresponding to (2.6)), whereas Figure 2-(B) represents the error

$$\mathcal{E}_2(\varepsilon) = \int_0^{10} \|\mathbf{x}^\varepsilon_{\mathrm{gc}}(t) - \mathbf{y}^\varepsilon(t)\| dt,$$

where $\mathbf{x}^\varepsilon_{\mathrm{gc}}$ is obtained through (2.13) from $(\mathbf{x}^\varepsilon, \mathbf{v}^\varepsilon)$ solving (2.2) and $\mathbf{y}^\varepsilon$ is the spatial component of $(\mathbf{y}^\varepsilon, v^\varepsilon, w^\varepsilon)$ satisfying the Cauchy problem as in Proposition 4.9 (and corresponding to (2.17)). These numerical results illustrate the order of accuracy stated in Theorems 2.2 (first order)

and 2.4 (second order). On the other hand, we have also claimed that to capture long-times dynamics it is also crucial to include second-order terms in the asymptotic models. Theorem 2.8 provides some quantitative support to the claim, in a specific geometry. We now provide in a different configuration another, qualitative, illustration of the claim, by plotting in Figures 2-(C) and 2-(D) examples of spatial parts of particle trajectories obtained with original, first and second-order models. Here we take $\varepsilon = 10^{-3}$ and solve on $[0, 250]$. Roughly speaking, the solution to the original problem exhibits a superposition of three kinds of spatial motions, namely, with decreasing velocity, the cyclotron oscillation about magnetic field lines, an oscillation along magnetic field lines, and a slower drift, responsible of a horizontally circular displacement. By construction, both asymptotic descriptions remove the cyclotron motion. However, whereas the second-order asymptotic model seems able to reproduce the slow part of the complicate multi-scale behavior, the first-order one only describes oscillations along the magnetic field lines. Indeed, since the first order model does not include classical drifts $\mathbf{U}_{\nabla B\times\mathbf{B}}$, $\mathbf{U}_{\mathrm{curv}}$ and $\mathbf{U}_{\mathrm{curl}\,\mathbf{e}_\parallel}$, it is not adequate to follow accurately the correct trajectory on times sufficiently long to feel the effects of those.

3. Two-dimensional homogeneous case

As a warm up we begin our analysis by revisiting the two-dimensional homogeneous case. The goal is to expound the tenets of the method without being slowed down by computational complexity. For the sake of exposition, for this simple system we prove first results that are even weaker than what the method may prove but that correspond to the best that is expected from the general 3-D system without assuming special symmetry.

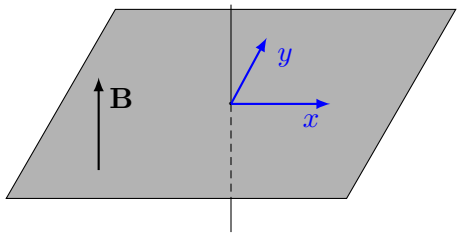

Figure 3. Representation of perpendicular plane to the magnetic field $\mathbf{B} = B(\mathbf{x})\,\mathbf{e}_z$.

Since here the parallel direction is fixed and we follow only perpendicular motions we drop temporarily $_\perp$ and $_\parallel$ indices. Thus, we consider for any $(\mathbf{x}, \mathbf{v}) \in \mathbf{R}^2 \times \mathbf{R}^2$ and $t \geq 0$,

$$\partial_t f^\varepsilon + \mathrm{div}_{\mathbf{x}}\,(f\,\mathbf{v}) + \mathrm{div}_{\mathbf{v}}\left(f\,\left(\frac{1}{\varepsilon}\,B\,\mathbb{J}\mathbf{v} + \mathbf{E}\right)\right) = 0\,. \tag{3.1}$$

Characteristics of the underlying PDE are obtained by solving

$$\begin{cases} \dfrac{\mathrm{d}\,\mathbf{x}^\varepsilon}{\mathrm{d}\,t} = \mathbf{v}^\varepsilon, \\ \dfrac{\mathrm{d}\,\mathbf{v}^\varepsilon}{\mathrm{d}\,t} = \dfrac{1}{\varepsilon}\,B\,\mathbb{J}\mathbf{v}^\varepsilon + \mathbf{E}(t, \mathbf{x}^\varepsilon), \end{cases} \tag{3.2}$$

with $B > 0$ and

$$\mathbb{J}\,(a_1, a_2) = (a_2, -a_1)\,,$$

important properties of $\mathbb{J}$ being

$$\mathbb{J}^2 = -\,\mathrm{Id}\,, \qquad \mathbb{J}^* = -\mathbb{J}\,.$$

For the sake of readability, from now on, when no confusion is possible, that is, when no asymptotic comparison is under consideration, we shall drop $^\varepsilon$ exponents on solutions.

We shall perform a series of transformations so as to extract from system (3.2) a normal form where some slow variables satisfy a system of ODEs uncoupled from fast scales up to error terms. It is worth pointing out that under stringent assumptions on fields one may expect to perform at once an infinite number of transformations and uncouple at infinite order slow variables from fast variables. We shall not pursue this line of investigation here but as a consequence one should keep in mind that variables that we designate as slow are slow only up to a certain order

and that depending on the objective at hand the level of slowness required may vary. As an example, anticipating a bit the analysis below, note that depending on the aimed conclusion one may be allowed to work directly with the spatial position $\mathbf{x}$ or need to manipulate the gyrocenter $\mathbf{x}+\varepsilon B^{-1}\mathbb{J}\mathbf{v}$, or even be compelled to use a version of those corrected by higher-order powers of ε.

3.1. **Uniform bounds.** Both to enforce that terms expected to be irrelevant are indeed irrelevant and to ensure that solutions persist on a sufficiently long time interval, uniform bounds on the solution are needed. Let us obtain them by introducing a kinetic energy variable $w(\mathbf{v}) = \frac{1}{2}\|\mathbf{v}\|^2$ and noting that system (3.2) yields

$$\begin{cases} \dfrac{\mathrm{d}\,\mathbf{x}}{\mathrm{d}\,t} = \mathbf{v}, \\ \dfrac{\mathrm{d}\,w}{\mathrm{d}\,t} = \langle \mathbf{v}, \mathbf{E}(t,\mathbf{x})\rangle, \\ \dfrac{\mathrm{d}\,\mathbf{v}}{\mathrm{d}\,t} = \dfrac{B}{\varepsilon}\mathbb{J}\mathbf{v} + \mathbf{E}(t,\mathbf{x})\,. \end{cases}$$

Remark 3.1. *We warn the reader that though we write the latter system as if w and $\mathbf{v}$ were independent variables this is mostly an algebraic trick here. In particular one should keep in mind that the system does contain some redundancy that is kept for the sake of simplicity of algebraic manipulations. In contrast an* augmented *formulation was in turn crucially used in* [Filbet and Rodrigues, 2017] *jointly with suitable numerical schemes so as to allow the discretization to disconnect the weak convergence of $\mathbf{v}$ to zero from the strong convergence of w to a non trivial limit.*

Of course here one could obtain from a Lipschitz assumption on $\mathbf{E}$ global-in-time existence and some bounds growing exponentially in time from a standard Grönwall lemma. Yet for expository reasons we show how to perform simple better estimates. Note however that, as we derive below in the long-time analysis, those are still deceptively pessimistic.

Lemma 3.2. *Solutions to* (3.2) *starting from* $(\mathbf{x}_0, \mathbf{v}_0)$ *are defined globally in time and satisfy for any* $t \geq 0$

$$\begin{cases} \|\mathbf{x}(t)\| \leq \|\mathbf{x}_0\| + t\,\|\mathbf{v}_0\| + t^2\,\|\mathbf{E}\|_{L^\infty}, \\ \|\mathbf{v}(t)\| \leq \|\mathbf{v}_0\| + 2\,t\,\|\mathbf{E}\|_{L^\infty}\,. \end{cases}$$

Proof. From the equation on w stems, for any $t \geq 0$, as long as the solution exists

$$\max_{s\in[0,t]} \|\mathbf{v}(s)\|^2 \leq \|v_0\|^2 + 2\max_{s\in[0,t]} \|\mathbf{v}(s)\|\,t\,\|\mathbf{E}\|_{L^\infty},$$

hence by solving the second-order inequality, for any $t \geq 0$, as long as the solution exists

$$\max_{s\in[0,t]} \|\mathbf{v}(s)\| \leq \sqrt{\|\mathbf{v}_0\|^2 + t^2\|\mathbf{E}\|_{L^\infty}^2} + t\,\|\mathbf{E}\|_{L^\infty}\,.$$

This yields the estimate on $\mathbf{v}$. In turn it implies the estimate on $\mathbf{x}$ by a mere integration, and jointly they prove global well-posedness by ruling out finite-time blow-up. □

3.2. **Elimination of linear terms.** We begin the uncoupling process. The thrust of the method is that the equation that forces $\mathbf{v}$ — or more exactly its argument — to evolve on fast scales also provides a way to eliminate at leading order $\mathbf{v}$ — or more exactly dependences on its argument — in slow equations. This general philosophy, that may be turned into rigorous arguments, explain why slow evolutions may be uncoupled from fast scales at any prescribed order. Explicitly elimination, at leading order, of linear terms in $\mathbf{v}$ is summarized as

Lemma 3.3. *Consider* $\mathbf{L} \in W^{1,\infty}\left(\mathbf{R}_t^+; \mathcal{L}_1(\mathbf{R}^2, \mathbf{R}^p)\right)$, $p \in \mathbf{N}^*$ *and* $(\mathbf{x}, \mathbf{v})$ *a solution to* (3.2). *Then for a.e.* $t \geq 0$, *we have*

$$\mathbf{L}(t)(\mathbf{v}(t)) = -\varepsilon\,\frac{\mathrm{d}}{\mathrm{d}\,t}\left(\mathbf{L}(t)\left(\frac{\mathbb{J}\,\mathbf{v}}{B}\right)\right) + \varepsilon\,\mathbf{L}'(t)\left(\frac{\mathbb{J}\mathbf{v}}{B}\right) + \varepsilon\,\mathbf{L}(t)(\mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{x})), \tag{3.3}$$

with

$$\mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{x}) = \frac{\mathbb{J}\,\mathbf{E}}{B}(t,\mathbf{x})\,.$$

In the former we have used the following notational convention. For any $\alpha \in \mathbf{N}$, $\mathcal{L}_\alpha(V,W)$ denotes the space of α-linear operators from V^α to W. In particular, $\mathcal{L}_1(V,W)$ is the set of linear operators from V to W.

Proof. This follows directly from

$$\mathbf{v} \,=\, -\frac{\varepsilon}{B}\frac{\mathrm{d}}{\mathrm{d}\,t}(\mathbb{J}\mathbf{v}) \,+\, \frac{\varepsilon}{B}\,\mathbb{J}\,\mathbf{E}(t,\mathbf{x})\,.$$

□

The latter term $\mathbf{U}_{\mathbf{E}\times\mathbf{B}}$ identifies the classical $\mathbf{E}\times\mathbf{B}$ velocity drift from gyrokinetic theory. The foregoing lemma singles out the prominent role played by the $\mathbf{U}_{\mathbf{E}\times\mathbf{B}}$ drift in two-dimensional gyrokinetics.

3.3. **First elimination and partial asymptotics.** By using Lemma 3.3 first with $\mathbf{L}(t)(\mathbf{v}) = \mathbf{v}$ then with $\mathbf{L}(t)(\mathbf{v}) = \langle \mathbf{E}(t,\mathbf{x}(t)),\mathbf{v}\rangle$, one derives that from system (3.2) follows for the guiding center variable,

$$\frac{\mathrm{d}}{\mathrm{d}\,t}\left[\mathbf{x}+\varepsilon\frac{\mathbb{J}\mathbf{v}}{B}\right] \,=\, \varepsilon\,\mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{x}) \tag{3.4}$$

and for the corrected kinetic energy $w(\mathbf{v}) = \frac{1}{2}\|\mathbf{v}\|^2$,

$$\frac{\mathrm{d}}{\mathrm{d}\,t}\left[w(\mathbf{v}) \,-\, \varepsilon\,\langle \mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{x}),\mathbf{v}\rangle\right] \,=\, -\varepsilon\,\langle \partial_t\mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{x}) + \mathrm{d}_{\mathbf{x}}\,\mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{x})(\mathbf{v}),\mathbf{v}\rangle\,. \tag{3.5}$$

Remark 3.4. *In the present paper the stiffer part of the fast equation is always linear in* $\mathbf{v}$*. This leads to a quite simple elimination of terms linear in* $\mathbf{v}$*. In particular, since the slow equations on* $(\mathbf{x},w)$ *are linear in* $\mathbf{v}$ *the first elimination comes almost for free. However in general each simplification increases the level of nonlinearity in* $\mathbf{v}$ *of slow equations and subsequent simplifications get more and more algebraically cumbersome.*

A specific feature of System (3.2) is that slow variables evolve with speeds of typical size $\mathcal{O}(\varepsilon)$ and not $\mathcal{O}(1)$. Therefore on time intervals $[0,T^\varepsilon_{\mathrm{obs}}]$, one hopes to validate approximation of the slow part by the solution of an uncoupled system up to error terms of size $\mathcal{O}(\varepsilon^2 T^\varepsilon_{\mathrm{obs}})$ with $T^\varepsilon_{\mathrm{obs}} = \mathcal{O}(\varepsilon^{-1})$. We first prove this claim with $T^\varepsilon_{\mathrm{obs}}$ of size 1 then refine the analysis to reach $T^\varepsilon_{\mathrm{obs}}$ of size ε^{-1}. Note that when $T^\varepsilon_{\mathrm{obs}}$ is of size ε^{-1} we aim at an error of size $\mathcal{O}(\varepsilon)$ and thus we may use directly $(\mathbf{x},w)$ as slow variables whereas when $T^\varepsilon_{\mathrm{obs}}$ is $\mathcal{O}(1)$ we aim at precision $\mathcal{O}(\varepsilon^2)$ thus we should use

$$\left(\mathbf{x}+\varepsilon\frac{\mathbb{J}\mathbf{v}}{B},\, w(\mathbf{v}) \,+\, \varepsilon\,\left\langle \mathbf{E}(t,\mathbf{x}),\frac{\mathbb{J}\mathbf{v}}{B}\right\rangle\right),$$

or a higher-order version of the latter.

Note that without further simplification the aforementioned asymptotics may not be readily derived since the equation on e still contains $\mathbf{v}$-terms at leading order. However an aspect even more peculiar to System (3.2) is that at leading order the equation for $\mathbf{x}$ uncouples not only from the argument of $\mathbf{v}$ but also from w. At this stage an asymptotic description of the slow part corresponding to $\mathbf{x}$ may be guessed without any further computation.

Proposition 3.5. *Assume* $\mathbf{E} \in L^\infty\left(\mathbf{R}^+_t; W^{1,\infty}(\mathbf{R}^2)\right)$ *and let* $(\mathbf{x}^\varepsilon,\mathbf{v}^\varepsilon)$ *be the solution to* (3.2) *starting from* $(\mathbf{x}_0,\mathbf{v}_0)$*. Then the guiding center variable (2.13) satisfies for a.e.* $t \geq 0$*,*

$$\left\|\mathbf{x}^\varepsilon_{\mathrm{gc}}(t) - \mathbf{y}^\varepsilon(t)\right\| \,\leq\, \frac{\varepsilon^2}{B^2}\|\,\mathrm{d}_{\mathbf{x}}\,\mathbf{E}\|_{L^\infty}\, e^{\frac{\varepsilon\,t}{B}\|\,\mathrm{d}_{\mathbf{x}}\,\mathbf{E}\|_{L^\infty}}\,(\,t\,\|\mathbf{v}_0\| \,+\, t^2\,\|\mathbf{E}\|_{L^\infty}\,),$$

where $\mathbf{y}^\varepsilon$ *solves*

$$\begin{cases} \dfrac{\mathrm{d}\,\mathbf{y}^\varepsilon}{\mathrm{d}\,t} \,=\, \varepsilon\,\mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{y}^\varepsilon), \\ \mathbf{y}^\varepsilon(0) = \mathbf{x}_{\mathrm{gc}}(0). \end{cases} \tag{3.6}$$

Proof. We consider $\mathbf{x}^\varepsilon_{\mathrm{gc}}$ given in (2.13), which satisfies

$$\frac{\mathrm{d}\,\mathbf{x}^\varepsilon_{\mathrm{gc}}}{\mathrm{d}\,t} \,=\, \varepsilon\,\mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{x}^\varepsilon_{\mathrm{gc}}) \,+\, \varepsilon\left[\mathbf{U}_{\mathbf{E}\times\mathbf{B}}\left(t,\mathbf{x}^\varepsilon_{\mathrm{gc}} - \varepsilon\,\frac{\mathbb{J}\mathbf{v}^\varepsilon}{B}\right) - \mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{x}^\varepsilon_{\mathrm{gc}})\right].$$

This implies for a.e. $t \geq 0$

$$\|\mathbf{x}^\varepsilon_{\rm gc}(t) - \mathbf{y}^\varepsilon(t)\| \leq \frac{\varepsilon}{B} \|\,\mathrm{d}_{\mathbf{x}}\,\mathbf{E}\|_{L^\infty} \int_0^t \|\mathbf{x}^\varepsilon_{\rm gc}(s) - \mathbf{y}^\varepsilon(s)\|\,\mathrm{d}\,s + \frac{\varepsilon^2}{B^2} \|\,\mathrm{d}_{\mathbf{x}}\,\mathbf{E}\|_{L^\infty} \int_0^t \|\mathbf{v}^\varepsilon(s)\|\,\mathrm{d}\,s\,.$$

Thus by the Grönwall lemma, for a.e. $t \geq 0$,

$$\|\mathbf{x}^\varepsilon_{\rm gc}(t) - \mathbf{y}^\varepsilon(t)\| \;\leq\; \frac{\varepsilon^2}{B^2} \|\,\mathrm{d}_{\mathbf{x}}\,\mathbf{E}\|_{L^\infty}\, e^{\frac{\varepsilon\, t}{B} \|\,\mathrm{d}_{\mathbf{x}}\,\mathbf{E}\|_{L^\infty}} \int_0^t \|\mathbf{v}^\varepsilon(s)\|\,\mathrm{d}\,s\,.$$

Then the result follows from Lemma 3.2. □

The foregoing bound is very simple but is not sharp with respect to ε. Indeed the principal part of the error term of the equation is linear in $\mathbf{v}$ thus may also be eliminated.

Remark 3.6. *The special structure of equation (3.4) is somewhat fortuitous. However the fact that the error introduced by replacing $\mathbf{x}^\varepsilon$ with its ε-correction $\mathbf{x}^\varepsilon_{\rm gc}$ may be cast away at leading order is not mere luck. It is due to our choice in (3.3) of an antiderivative*

$$\varepsilon\, \mathbf{L}(t) \left(\frac{\mathbb{J}\mathbf{v}}{B}\right),$$

that at leading order contains no slow part. Henceforth in similar cases enforcing such properties will always streamline our particular choices.

The announced further elimination yields the following refinement.

Proposition 3.7. *Assume that $\mathbf{E} \in W^{2,\infty}$. There exists a constant $C > 0$, depending polynomially on $\|\mathbf{E}\|_{W^{2,\infty}}$ and B^{-1}, such that if $(\mathbf{x}^\varepsilon, \mathbf{v}^\varepsilon)$ is a solution to (3.2) starting from $(\mathbf{x}_0, \mathbf{v}_0)$, then it satisfies for a.e. $t \geq 0$,*

$$\left\|\mathbf{x}^\varepsilon_{\rm gc}(t) - \mathbf{y}^\varepsilon(t)\right\| \;\leq\; C\,\varepsilon^3\, e^{\frac{\varepsilon\, t}{B} \|\,\mathrm{d}_{\mathbf{x}}\,\mathbf{E}\|_{L^\infty}}\, (\,1 \,+\, t\,(1 + \|\mathbf{v}_0\| \,+\, t\,\|\mathbf{E}\|_{L^\infty})\,)\,(\|\mathbf{v}_0\| \,+\, t\,\|\mathbf{E}\|_{L^\infty}),$$

where $\mathbf{x}^\varepsilon_{\rm gc}$ is as in (2.13) and $\mathbf{y}^\varepsilon$ solves (3.6).

Proof. The term to weed out is linear in $\mathbf{v}$ and by applying Lemma 3.3 with

$$\mathbf{L}(t)(\mathbf{v}) = -\,\mathrm{d}_{\mathbf{x}}\,\mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t, \mathbf{x}^\varepsilon_{\rm gc}(t)) \left(\frac{\mathbb{J}\mathbf{v}}{B}\right)$$

one obtains

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\,t}\left[\mathbf{x}^\varepsilon_{\rm gc} + \varepsilon^3\,\mathrm{d}_{\mathbf{x}}\,\mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t, \mathbf{x}^\varepsilon_{\rm gc}) \left(\frac{\mathbf{v}^\varepsilon}{B^2}\right)\right] \\
&= \varepsilon\,\mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t, \mathbf{x}^\varepsilon_{\rm gc}) - \varepsilon^3\,\mathrm{d}_{\mathbf{x}}\,\mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t, \mathbf{x}^\varepsilon_{\rm gc}) \left(\frac{\mathbb{J}\,\mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t, \mathbf{x}^\varepsilon)}{B}\right) \\
&+ \varepsilon^3\,\mathrm{d}_{\mathbf{x}}\,\partial_t \mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t, \mathbf{x}^\varepsilon_{\rm gc}) \left(\frac{\mathbf{v}^\varepsilon}{B^2}\right) \,+\, \varepsilon^3\,\mathrm{d}^2_{\mathbf{x}}\,\mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t, \mathbf{x}^\varepsilon_{\rm gc}) \left(\mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t, \mathbf{x}^\varepsilon), \frac{\mathbf{v}^\varepsilon}{B^2}\right) \\
&+ \varepsilon \left[\mathbf{U}_{\mathbf{E}\times\mathbf{B}}\left(t, \mathbf{x}^\varepsilon_{\rm gc} - \varepsilon\,\frac{\mathbb{J}\mathbf{v}^\varepsilon}{B}\right) - \mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t, \mathbf{x}^\varepsilon_{\rm gc}) + \varepsilon\,\mathrm{d}_{\mathbf{x}}\,\mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t, \mathbf{x}^\varepsilon_{\rm gc}) \left(\frac{\mathbb{J}\mathbf{v}^\varepsilon}{B}\right)\right].
\end{aligned}$$

Therefore, for a.e. $t \geq 0$, one has

$$\begin{aligned}
\|\mathbf{x}^\varepsilon_{\rm gc}(t) - \mathbf{y}^\varepsilon(t)\| \leq\; & \frac{\varepsilon}{B} \|\,\mathrm{d}_{\mathbf{x}}\,\mathbf{E}\|_{L^\infty} \int_0^t \|\mathbf{x}^\varepsilon_{\rm gc}(s) - \widetilde{\mathbf{y}}^\varepsilon(s)\|\,\mathrm{d}\,s \,+\, \frac{\varepsilon^3}{B^3}\,\|\,\mathrm{d}_{\mathbf{x}}\,\mathbf{E}\|_{L^\infty}\,(\|\mathbf{v}_0\| + \|\mathbf{v}^\varepsilon(t)\|) \\
&+ \frac{\varepsilon^3}{B^3}\,(\|\,\mathrm{d}_{\mathbf{x}}\,\partial_t \mathbf{E}\|_{L^\infty} + B^{-1}\|\,\mathrm{d}^2_{\mathbf{x}}\,\mathbf{E}\|_{L^\infty} \|\mathbf{E}\|_{L^\infty}) \int_0^t \|\mathbf{v}^\varepsilon(s)\|\,\mathrm{d}\,s \\
&+ \frac{\varepsilon^3}{2\,B^3} \|\,\mathrm{d}^2_{\mathbf{x}}\,\mathbf{E}\|_{L^\infty} \int_0^t \|\mathbf{v}^\varepsilon(s)\|^2\,\mathrm{d}\,s\,.
\end{aligned}$$

At this stage the result follows from Lemma 3.2 and the Grönwall lemma. □

One may go on by correcting $\mathbf{x}^\varepsilon_{\rm gc}$ into a "higher-order" approximation

$$\mathbf{x}^\varepsilon_{\rm ho} \;=\; \mathbf{x}^\varepsilon_{\rm gc} + \varepsilon^3\,\mathrm{d}_{\mathbf{x}}\,\mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t, \mathbf{x}^\varepsilon_{\rm gc}) \left(\frac{\mathbf{v}^\varepsilon}{B^2}\right),$$

then expanding from $\mathbf{x}^\varepsilon_{\rm ho}$ and eliminating terms involving $\mathbf{v}^\varepsilon$. But the expansion process would involve terms quadratic in $\mathbf{v}$ whose elimination brings a coupling with w^ε as may be seen from Lemma 3.8 below. Proposition 3.7 is therefore expected to be optimal with respect to ε-scaling on time intervals of length $\mathcal{O}(1)$.

3.4. **Elimination of quadratic terms.** In order to obtain asymptotics for the full set of slow variables $(\mathbf{x}, w)$ we study now the extraction of slow components from expressions that are quadratic in $\mathbf{v}$.

Lemma 3.8. *Consider* $\mathbf{A} \in W^{1,\infty}\left(\mathbf{R}^+_t;\, \mathcal{L}_2(\mathbf{R}^2, \mathbf{R}^p)\right)$, $p \in \mathbf{N}^*$ *and* $(\mathbf{x}, \mathbf{v})$ *a solution to* (3.2). *Then, for a.e.* $t \geq$, *we have*

$$\mathbf{A}(t)(\mathbf{v}(t), \mathbf{v}(t)) \;=\; w(\mathbf{v})\ \mathrm{Tr}(\mathbf{A}(t)) - \varepsilon \frac{\mathrm{d}\,\chi_{\mathbf{A}}}{\mathrm{d}\,t}(t) \;+\; \varepsilon\, \eta_{\mathbf{A}}(t),$$

where $w(\mathbf{v}) = \|\mathbf{v}\|^2/2$, *whereas* Tr *denotes the trace operator, for the canonical basis* $(\mathbf{e}_x, \mathbf{e}_y)$ *of* $\mathbf{R}^2$

$$\mathrm{Tr}(\mathbf{A}(t)) \;=\; \mathbf{A}(t)(\mathbf{e}_x, \mathbf{e}_x) \;+\; \mathbf{A}(t)(\mathbf{e}_y, \mathbf{e}_y)$$

whereas $\chi_{\mathbf{A}}$ *is given by*

$$\chi_{\mathbf{A}} \;=\; \frac{1}{2}\,\Re(\mathbf{A})\left(\mathbf{v}, B^{-1}\mathbb{J}\,\mathbf{v}\right)$$

and $\eta_{\mathbf{A}}$ *is*

$$\begin{aligned}\eta_{\mathbf{A}} \;&=\; \frac{1}{2}\,\Re(\mathbf{A})(\mathbf{v}, \mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{x})) \;+\; \frac{1}{2}\,\Re(\mathbf{A})'\left(\mathbf{v}, B^{-1}\mathbb{J}\,\mathbf{v}\right)\\ &+\; \frac{1}{2}\,\Re(\mathbf{A})\left(\mathbf{E}(t,\mathbf{x}), B^{-1}\mathbb{J}\,\mathbf{v}\right),\end{aligned}$$

with $\Re$ *denoting the symmetric part defined in (2.16).*

Remark 3.9. *Consistently with Remark 3.6, note that* $\chi_{\mathbf{A}}$ *has itself no slow component at leading order since* $\Re(\mathbf{A}(t))\,(\cdot, \mathbb{J}(\cdot))$ *is trace-free. Indeed its trace is*

$$\begin{aligned}\Re(\mathbf{A})\,(\mathbf{a}, \mathbb{J}\mathbf{a}) + \Re(\mathbf{A})\,(\mathbb{J}\mathbf{a}, \mathbb{J}\,\mathbb{J}\mathbf{a}) \;&=\; \Re(\mathbf{A})\,(\mathbf{a}, \mathbb{J}\mathbf{a}) - \Re(\mathbf{A})\,(\mathbb{J}\mathbf{a}, \mathbf{a})\\ &=\; 0,\end{aligned}$$

where $\mathbf{a}$ *is any unitary vector. In the latter to express the trace we have used that* $(\mathbf{a}, \mathbb{J}\mathbf{a})$ *form an orthonormal basis for any unitary* $\mathbf{a}$.

Proof. Note first that one may assume without loss of generality that $\mathbf{A}$ is valued in symmetric bilinear forms. Thus we assume $\Re(\mathbf{A}) = \mathbf{A}$ for the sake of notational concision. By differentiation one derives

$$\begin{aligned}2\frac{\mathrm{d}\,\chi_{\mathbf{A}}}{\mathrm{d}\,t}(t) \;&=\; \mathbf{A}'(t)\left(\mathbf{v}(t), B^{-1}\mathbb{J}\mathbf{v}(t)\right)\\ &+\; \mathbf{A}(t)\left[\frac{\mathrm{d}\,\mathbf{v}}{\mathrm{d}\,t}(t), B^{-1}\mathbb{J}\mathbf{v}(t)\right] \;+\; \mathbf{A}(t)\left[\mathbf{v}(t), B^{-1}\mathbb{J}\left(\frac{\mathrm{d}\,\mathbf{v}}{\mathrm{d}\,t}(t)\right)\right]\\ &=\; \mathbf{A}'(t)\left(\mathbf{v}(t), B^{-1}\mathbb{J}\mathbf{v}(t)\right)\\ &+\; \frac{1}{\varepsilon}\mathbf{A}(t)\left(\mathbb{J}\mathbf{v}(t), \mathbb{J}\mathbf{v}(t)\right) + \mathbf{A}(t)\left(\mathbf{E}(t,\mathbf{x}(t)), B^{-1}\mathbb{J}\mathbf{v}(t)\right)\\ &-\; \frac{1}{\varepsilon}\mathbf{A}(t)\left(\mathbf{v}(t), \mathbf{v}(t)\right) + \mathbf{A}(t)\left(\mathbf{v}(t), \mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{x}(t))\right)\end{aligned}$$

and the result follows by multiplying by $\varepsilon/2$ then adding $\mathbf{A}(t)(\mathbf{v}(t), \mathbf{v}(t))$ and using that

$$w\ \mathrm{Tr}(\mathbf{A}) \;=\; \frac{1}{2}\left(\mathbf{A}(\mathbf{v}, \mathbf{v}) + \mathbf{A}(\mathbb{J}\mathbf{v}, \mathbb{J}\mathbf{v})\right).$$

□

3.5. **Second elimination and full asymptotics.** For the sake of concision and symmetry we introduce

$$w^\varepsilon_{\rm gc} \,=\, w(\mathbf{v}^\varepsilon) \,-\, \varepsilon\, \langle \mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{x}^\varepsilon), \mathbf{v}^\varepsilon\rangle \,, \tag{3.7}$$

which corresponds to the corrected kinetic energy, a two-dimensional version of (2.15).

By applying Lemmas 3.3 with

$$\mathbf{L}(t)(\mathbf{v}) = -\,\langle \partial_t \mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{x}(t)), \mathbf{v}\rangle$$

and Lemma 3.8 with

$$\mathbf{A}(t)(\mathbf{v},\mathbf{u}) \,=\, -\,\langle \mathrm{d}_{\mathbf{x}}\, \mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{x}(t))(\mathbf{v}), \mathbf{u}\rangle \,,$$

equation (3.5) in system (3.4)-(3.5) may be turned into

$$\frac{\mathrm{d}}{\mathrm{d}\,t}\left[w^\varepsilon_{\rm gc} + \frac{\varepsilon^2}{B^2}\,\chi^\varepsilon\right] \,=\, -\varepsilon\, w(\mathbf{v}^\varepsilon)\, \mathrm{div}_{\mathbf{x}}(\mathbf{U}_{\mathbf{E}\times\mathbf{B}})(t,\mathbf{x}^\varepsilon) \,+\, \frac{\varepsilon^2}{B^2}\,\eta^\varepsilon, \tag{3.8}$$

where

$$\chi^\varepsilon(t,\mathbf{x},\mathbf{v}) \,=\, \langle \partial_t \mathbf{E}(t,\mathbf{x}), \mathbf{v}\rangle - \frac{1}{4}\,\left[\langle \mathrm{d}_{\mathbf{x}}\, \mathbf{E}(t,\mathbf{x})(\mathbf{v}), \mathbf{v}\rangle \,-\, \langle \mathrm{d}_{\mathbf{x}}\, \mathbf{E}(t,\mathbf{x})(\mathbb{J}\mathbf{v}), \mathbb{J}\mathbf{v}\rangle\right]$$

and

$$\begin{aligned}
\eta^\varepsilon(t,\mathbf{x},\mathbf{v}) \;&=\; -\,\langle \partial_t \mathbf{E}(t,\mathbf{x}), \mathbf{E}(t,\mathbf{x})\rangle - \left\langle \partial_t^2 \mathbf{E}(t,\mathbf{x}) + \mathrm{d}_{\mathbf{x}}\, \partial_t \mathbf{E}(t,\mathbf{x})(\mathbf{v}), \mathbf{v}\right\rangle \\
&+\; \frac{1}{4}\,\langle \mathrm{d}_{\mathbf{x}}\, \mathbf{E}(t,\mathbf{x})(\mathbb{J}\mathbf{E}(t,\mathbf{x})), \mathbb{J}\mathbf{v}\rangle - \frac{1}{4}\,\langle \mathrm{d}_{\mathbf{x}}\, \mathbf{E}(t,\mathbf{x})(\mathbf{v}), \mathbf{E}(t,\mathbf{x})\rangle \\
&-\; \frac{1}{4}\,\langle \mathrm{d}_{\mathbf{x}}\, \mathbf{E}(t,\mathbf{x})(\mathbf{E}(t,\mathbf{x})), \mathbf{v}\rangle + \frac{1}{4}\,\langle \mathrm{d}_{\mathbf{x}}\, \mathbf{E}(t,\mathbf{x})(\mathbb{J}\mathbf{v}), \mathbb{J}\mathbf{E}(t,\mathbf{x})\rangle \\
&-\; \frac{1}{4}\left\langle \mathrm{d}_{\mathbf{x}}\, \partial_t \mathbf{E}(t,\mathbf{x})(\mathbf{v}) + \mathrm{d}_{\mathbf{x}}^2\, \mathbf{E}(t,\mathbf{x})(\mathbf{v},\mathbf{v}), \mathbf{v}\right\rangle \\
&+\; \frac{1}{4}\left\langle \mathrm{d}_{\mathbf{x}}\, \partial_t \mathbf{E}(t,\mathbf{x})(\mathbb{J}\mathbf{v}) + \mathrm{d}_{\mathbf{x}}^2\, \mathbf{E}(t,\mathbf{x})(\mathbf{v},\mathbb{J}\mathbf{v}), \mathbb{J}\mathbf{v}\right\rangle \,.
\end{aligned}$$

Now we may complete Proposition 3.7 to obtain leading-order asymptotics for $(\mathbf{x}^\varepsilon_{\rm gc}, w^\varepsilon_{\rm gc})$.

Proposition 3.10. *Assume $\mathbf{E} \in W^{2,\infty}$. There exists a constant $C > 0$, depending polynomially on $\|\mathbf{E}\|_{W^{2,\infty}}$ and B^{-1} such that the following holds. Let $(\mathbf{x}^\varepsilon, \mathbf{v}^\varepsilon)$ be the solution to (3.2) starting from $(\mathbf{x}_0, \mathbf{v}_0)$. Then, for a.e. $t \geq 0$,*

$$\begin{cases}
\left\|\mathbf{x}^\varepsilon_{\rm gc}(t) - \mathbf{y}^\varepsilon(t)\right\| \,\leq\, C\,\varepsilon^3\, e^{C\,\varepsilon\, t}\left(1 + t\,(1 + \|\mathbf{v}_0\| \,+\, t)\right)(\|\mathbf{v}_0\| \,+\, t), \\
\left\|w^\varepsilon_{\rm gc}(t) - w^\varepsilon(t)\right\| \,\leq\, C\,\varepsilon^2\, e^{C\,\varepsilon\, t}\left(1 + (1+t)\,(\|\mathbf{v}_0\| \,+\, t) + t\,(\|\mathbf{v}_0\| \,+\, t)^2\right)(\|\mathbf{v}_0\| \,+\, t),
\end{cases}$$

where $(\mathbf{x}^\varepsilon_{\rm gc}, w^\varepsilon_{\rm gc})$ is as in (2.13) and (3.7) and $(\mathbf{y}^\varepsilon, w^\varepsilon)$ solves

$$\begin{cases}
\dfrac{\mathrm{d}\,\mathbf{y}^\varepsilon}{\mathrm{d}\,t} \,=\, \varepsilon\, \mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{y}^\varepsilon), \\
\dfrac{\mathrm{d}\,w^\varepsilon}{\mathrm{d}\,t} \,=\, -\varepsilon\, w^\varepsilon\, \mathrm{div}_{\mathbf{x}}(\mathbf{U}_{\mathbf{E}\times\mathbf{B}})(t,\mathbf{y}^\varepsilon),
\end{cases} \tag{3.9}$$

with initial data $\mathbf{y}^\varepsilon(0) = \mathbf{x}^\varepsilon_{\rm gc}(0)$ and $w^\varepsilon(0) = w^\varepsilon_{\rm gc}(0)$.

Remark 3.11. *Note that if $\mathbf{E}$ derives from a potential, that is, if $\mathbf{E}$ is curl-free, then the equation on w^ε is trivial since $\mathrm{div}_{\mathbf{x}}(\mathbf{U}_{\mathbf{E}\times\mathbf{B}}) = 0$. Yet this cancellation does not improve any convergence rate. Incidentally we point out that in this case $\mathrm{d}_{\mathbf{x}}\, \mathbf{E}$ is symmetric so that the cancellation follows at a more abstract level from computations of Remark 3.9.*

Proof. We consider $(\mathbf{x}^\varepsilon, \mathbf{v}^\varepsilon)$ the solution to (3.2) starting from $(\mathbf{x}_0, \mathbf{v}_0)$ and the corresponding $(\mathbf{y}^\varepsilon, w^\varepsilon)$, the solution to (3.9) starting from $(\mathbf{x}^\varepsilon_{\rm gc}(0), w^\varepsilon_{\rm gc}(0))$, where $(\mathbf{x}^\varepsilon_{\rm gc}, w^\varepsilon_{\rm gc})$ is as in (2.13)-(3.7). First, to ease comparisons, we recall that $w(\mathbf{v}^\varepsilon) = \|\mathbf{v}^\varepsilon\|^2/2$ (and ban temporarily the confusing

shorthands w^ε and w for $w(\mathbf{v}^\varepsilon)$) and write (3.8) as

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\,t}\left[w^\varepsilon_{\mathrm{gc}} + \frac{\varepsilon^2}{B^2}\,\chi^\varepsilon\right] \;=\; & -\varepsilon\, w^\varepsilon_{\mathrm{gc}}\,\mathrm{div}_{\mathbf{x}}\,\mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{y}^\varepsilon) + \frac{\varepsilon^2}{B^2}\,\eta^\varepsilon \\ & -\varepsilon\,\left(w(\mathbf{v}^\varepsilon) - w^\varepsilon_{\mathrm{gc}}\right)\,\mathrm{div}_{\mathbf{x}}\,\mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{y}^\varepsilon) \\ & -\varepsilon\, w(\mathbf{v}^\varepsilon)\,\left(\mathrm{div}_{\mathbf{x}}\,\mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{x}^\varepsilon) - \mathrm{div}_{\mathbf{x}}\,\mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{x}^\varepsilon_{\mathrm{gc}})\right) \\ & -\varepsilon\, w(\mathbf{v}^\varepsilon)\,\left(\mathrm{div}_{\mathbf{x}}\,\mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{x}^\varepsilon_{\mathrm{gc}}) - \mathrm{div}_{\mathbf{x}}\,\mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{y}^\varepsilon)\right).\end{aligned}$$

Then, from subtracting the latter equation to the one for w^ε in (3.9) stems for a.e. $t \geq 0$,

$$\begin{aligned}|w^\varepsilon_{\mathrm{gc}}(t) - w^\varepsilon(t)| \;\leq\; & C\,\varepsilon\int_0^t |w^\varepsilon_{\mathrm{gc}}(s) - w^\varepsilon(s)|\,\mathrm{d}\,s + C\,\varepsilon^2\,\left(|\chi^\varepsilon(0)| + |\chi^\varepsilon(t)|\right) \\ & + C\,\varepsilon^2\int_0^t |\eta^\varepsilon(s)|\,\mathrm{d}\,s + C\,\varepsilon^2\int_0^t \|\mathbf{v}^\varepsilon(s)\|\,\mathrm{d}\,s \\ & + C\,\varepsilon^2\int_0^t \|\mathbf{v}^\varepsilon(s)\|^3\,\mathrm{d}\,s + C\varepsilon\int_0^t \|\mathbf{v}^\varepsilon(s)\|^2\,\|\mathbf{x}^\varepsilon_{\mathrm{gc}}(s) - \mathbf{y}^\varepsilon(s)\|\,\mathrm{d}\,s,\end{aligned}$$

where C depends polynomially on $\|\mathbf{E}\|_{W^{2,\infty}}$ and B^{-1}. Finally the estimate on $w^\varepsilon_{\mathrm{gc}} - w^\varepsilon$ follows from Lemma 3.2, Proposition 3.5 and the Grönwall lemma. □

3.6. **Long-time asymptotics.** As aforementioned, the fact that vector fields appearing in the leading-order asymptotics seem to be $\mathcal{O}(\varepsilon)$ suggests that it should also be possible to validate asymptotics for $(\mathbf{x}^\varepsilon, w(\mathbf{v}^\varepsilon))$ on time intervals of length $\mathcal{O}(\varepsilon^{-1})$ with convergence rates $\mathcal{O}(\varepsilon)$. To carry out such achievement we need to refine bounds from Lemma 3.2.

Lemma 3.12. *There exists a universal positive constant C such that any solution to (3.2) starting from $(\mathbf{x}_0, \mathbf{v}_0)$ satisfies for any $t \geq 0$*

$$\begin{cases} \|\mathbf{x}(t)\| \;\leq\; \|\mathbf{x}_0\| + \dfrac{\varepsilon}{B}\,t\,\|\mathbf{E}\|_{L^\infty} + C\,\varepsilon\,e^{C\,\frac{\varepsilon}{B}\,t\,\|\mathbf{E}\|_{\dot{W}^{1,\infty}}}\left(1 + \|\mathbf{v}_0\| + \dfrac{\varepsilon}{B}\,\|\mathbf{E}\|_{L^\infty}\right), \\ \|\mathbf{v}(t)\| \;\leq\; C\,e^{C\,\frac{\varepsilon}{B}\,t\,\|\mathbf{E}\|_{\dot{W}^{1,\infty}}}\left(1 + \|\mathbf{v}_0\| + \dfrac{\varepsilon}{B}\,\|\mathbf{E}\|_{L^\infty}\right). \end{cases}$$

Proof. Integrating (3.5) yields for a.e. $t \geq 0$

$$\begin{aligned} w(\mathbf{v}) \;\leq\; & w(\mathbf{v}_0) + \frac{\varepsilon}{B}\|\mathbf{E}\|_{L^\infty}\|\mathbf{v}_0\| + \frac{w(\mathbf{v})}{2} + \frac{\varepsilon^2}{2B^2}\|\mathbf{E}\|^2_{L^\infty} \\ + \;\; & \frac{\varepsilon}{B}t\,\|\partial_t\mathbf{E}\|_{L^\infty} + \frac{\varepsilon}{B}(2\|\,\mathrm{d}_{\mathbf{x}}\,\mathbf{E}\|_{L^\infty} + \|\partial_t\mathbf{E}\|_{L^\infty})\int_0^t w(s)\,\mathrm{d}\,s,\end{aligned}$$

which after a few algebraic manipulations and an application of the Grönwall lemma proves the claim on $\|\mathbf{v}\|$. The bound on $\mathbf{x}$ is obtained simarly by integrating (3.4). □

We now focus on large-time asymptotics for $(\mathbf{x}, w(\mathbf{v}))$.

Proposition 3.13. *Assume $\mathbf{E} \in W^{2,\infty}$. There exists a constant $C > 0$, depending polynmially on $\|\mathbf{E}\|_{W^{2,\infty}}$ and B^{-1}, such that any solution to (3.2) starting from $(\mathbf{x}_0, \mathbf{v}_0)$ satisfies for a.e. $t \geq 0$*

$$\begin{cases} \|\mathbf{x}^\varepsilon(t) - \mathbf{y}^\varepsilon(t)\| \;\leq\; C\,\varepsilon\,e^{C\,\varepsilon\,t}\,(1 + \varepsilon + \|\mathbf{v}_0\|), \\ \|w(\mathbf{v}^\varepsilon(t)) - w^\varepsilon(t)\| \;\leq\; C\,\varepsilon\,e^{C\,\varepsilon\,t}\,(1 + \varepsilon + \|\mathbf{v}_0\|)^3, \end{cases}$$

where $w(\mathbf{v}) = \frac{1}{2}\|\mathbf{v}\|^2$ and $(\mathbf{y}^\varepsilon, w^\varepsilon)$ solves (3.9) with initial data $\mathbf{y}^\varepsilon(0) = \mathbf{x}_0$ and $w^\varepsilon(0) = \frac{1}{2}\|\mathbf{v}_0\|^2$.

Proof. The estimate on $\mathbf{x}^\varepsilon - \mathbf{y}^\varepsilon$ follows from Lemma 3.12 and the Grönwall lemma after an integration of (3.4). To proceed, we use (3.8) in the form

$$\begin{aligned}& \frac{\mathrm{d}}{\mathrm{d}\,t}\left[w(\mathbf{v}^\varepsilon) - \varepsilon\,\langle\mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{x}^\varepsilon), \mathbf{v}^\varepsilon\rangle + \frac{\varepsilon^2}{B^2}\,\chi^\varepsilon\right] \\ & = -\varepsilon\,w(\mathbf{v}^\varepsilon)\,\mathrm{div}_{\mathbf{x}}\,\mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{y}^\varepsilon) + \frac{\varepsilon^2}{B^2}\,\eta^\varepsilon - \varepsilon\,w(\mathbf{v}^\varepsilon)\,\left(\mathrm{div}_{\mathbf{x}}\,\mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{x}^\varepsilon) - \mathrm{div}_{\mathbf{x}}\,\mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{y}^\varepsilon)\right),\end{aligned}$$

thus, for a.e. $t \geq 0$,

$$\begin{aligned} |w(\mathbf{v}^\varepsilon(t)) - w^\varepsilon(t)| \quad \leq \quad & C\varepsilon \int_0^t |w(\mathbf{v}^\varepsilon(s)) - w^\varepsilon(s)| \, \mathrm{d}\, s \\ & + C\,\varepsilon\,(\|\mathbf{v}_0\| + \|\mathbf{v}^\varepsilon(t)\|) + C\varepsilon^2\,(|\chi^\varepsilon(0)| + |\chi^\varepsilon(t)|) \\ & + C\,\varepsilon^2 \int_0^t |\eta^\varepsilon(s)| \, \mathrm{d}\, s + C\varepsilon \int_0^t \|\mathbf{v}^\varepsilon(s)\|^2 \, \|\mathbf{x}^\varepsilon(s) - \mathbf{y}^\varepsilon(s)\| \, \mathrm{d}\, s, \end{aligned}$$

where C depends polynomially on $\|\mathbf{E}\|_{W^{2,\infty}}$ and B^{-1}. One may conclude again with Lemma 3.12 and the Grönwall lemma. □

Remark 3.14. *The proof also yields the analysis of dynamics involving fields depending on ε but satisfying bounds uniform with respect to ε. In particular the result may be extended without change to the case where $\mathbf{E}^\varepsilon(t,\mathbf{x}) = \mathbf{E}(\varepsilon\, t, \mathbf{x})$, $0 < \varepsilon \lesssim 1$. This somehow simpler problem is the one classically considered because then the asymptotic dynamics is essentially independent of ε at leading order since*

$$(\mathbf{y}^\varepsilon, w^\varepsilon)(t) \;=\; (\mathbf{y}, w)(\varepsilon\, t),$$

with $(\mathbf{y}, w)$ independent of ε.

Of course we may also use Lemma 3.12 to refine time dependences in Proposition 3.10 so as to fill the gap concerning what happens at leading-order for intermediate times $1 \lesssim t \lesssim \varepsilon^{-1}$. For possible external reference let us store without proof the corresponding result.

Proposition 3.15. *Assume $\mathbf{E} \in W^{2,\infty}$. Then there exists a constant $C > 0$, depending polynomially on $\|\mathbf{E}\|_{W^{2,\infty}}$ and B^{-1}, such that the following holds. Let $(\mathbf{x}^\varepsilon, \mathbf{v}^\varepsilon)$ be the solution to (3.2) starting from $(\mathbf{x}_0, \mathbf{v}_0)$. Then, for a.e. $t \geq 0$,*

$$\begin{cases} \|\mathbf{x}^\varepsilon_{\mathrm{gc}}(t) - \mathbf{y}^\varepsilon(t)\| \;\leq\; C\,\varepsilon^3\, e^{C\,\varepsilon\, t} \Big[1 + t\,(1 + \varepsilon + \|\mathbf{v}_0\|)\Big]\,(1 + \varepsilon + \|\mathbf{v}_0\|), \\ \|w^\varepsilon_{\mathrm{gc}}(t) - w^\varepsilon(t)\| \;\leq\; C\,\varepsilon^2\, e^{C\,\varepsilon\, t} \Big[1 + t\,(1 + \varepsilon + \|\mathbf{v}_0\|)\Big]\,(1 + \varepsilon + \|\mathbf{v}_0\|)^2, \end{cases}$$

where $(\mathbf{x}^\varepsilon_{\mathrm{gc}}, w^\varepsilon_{\mathrm{gc}})$ is as in (2.13) and (3.7) and $(\mathbf{y}^\varepsilon, w^\varepsilon)$ solves (3.9) with initial data $\mathbf{y}^\varepsilon(0) = \mathbf{x}^\varepsilon_{\mathrm{gc}}(0)$ and $w^\varepsilon(0) = w^\varepsilon_{\mathrm{gc}}(0)$.

3.7. **PDE counterparts.** Now let us translate the foregoing results at the PDE level.

On the reduced phase-space where $\mathbf{Z} = (\mathbf{y}, w)$ lives the relevant macroscopic velocity is $\varepsilon\mathcal{W}_1(t,\mathbf{Z})$ where

$$\mathcal{W}_1(t,\mathbf{Z}) \;=\; \begin{pmatrix} \mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{y}) \\ -w\;\mathrm{div}_{\mathbf{x}}\,\mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{y}) \end{pmatrix}, \tag{3.10}$$

which corresponds to the velocity field of system (3.9) defining the characteristic curves of the equation

$$\partial_t G^\varepsilon \;+\; \varepsilon\;\mathrm{div}_{\mathbf{Z}}\,(\mathcal{W}_1\, G^\varepsilon) \;=\; 0. \tag{3.11}$$

With this in hands we may deduce from Proposition 3.13 and Proposition 3.15 the following statement, where we have made explicit push-forwards that were easy to compute.

Theorem 3.16. *Let $\mathbf{E} \in W^{2,\infty}$. There exists a constant C depending polynomially on $\|\mathbf{E}\|_{W^{2,\infty}}$ and B^{-1} such that the following holds for any solution f^ε to (3.1) with initial datum a positive density f_0.*

(i). Long-time first-order asymptotics. *F^ε defined by*

$$F^\varepsilon(t,\mathbf{x},w) = \int_0^{2\pi} f^\varepsilon(t,\mathbf{x},\sqrt{2\,w}\;\mathbf{e}(\theta))\,\mathrm{d}\,\theta$$

satisfies for a.e. $t \geq 0$

$$\|F^\varepsilon(t,\cdot) - G^\varepsilon(t,\cdot)\|_{\dot{W}^{-1,1}} \leq C\,\varepsilon\, e^{C\,\varepsilon\, t} \int_{\mathbf{R}^2\times\mathbf{R}^2} (1 + \varepsilon + \|\mathbf{v}\|)^3\, f_0(\mathbf{x},\mathbf{v})\;\mathrm{d}\,\mathbf{x}\;\mathrm{d}\,\mathbf{v}$$

where G^ε solves (3.11) with initial datum $G_0 = F^\varepsilon(0,\cdot)$ given by

$$G_0(\mathbf{x}, w) = \int_0^{2\pi} f_0(\mathbf{x}, \sqrt{2\,w}\ \mathbf{e}(\theta))\,\mathrm{d}\,\theta\,.$$

(ii). Short-time second-order asymptotics. *The push-forwards $F^\varepsilon_{\mathrm{gc}}(t,\cdot)$ of $f^\varepsilon(t,\cdot)$ by the maps*

$$(\mathbf{x},\mathbf{v}) \mapsto \Big(\mathbf{x} + \frac{\varepsilon}{B}\,\mathbb{J}\mathbf{v}\,,\ \tfrac{1}{2}\|\mathbf{v}\|^2 \ - \ \varepsilon\ \langle \mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{x}), \mathbf{v}\rangle\Big)$$

satisfy for a.e. $t \geq 0$

$$\|F^\varepsilon_{\mathrm{gc}}(t,\cdot) - G^\varepsilon_{\mathrm{gc}}(t,\cdot)\|_{\dot{W}^{-1,1}} \leq C\,\varepsilon^2\,e^{C\,\varepsilon\,t}\,(1+t)\int_{\mathbf{R}^2\times\mathbf{R}^2}(1+\varepsilon+\|\mathbf{v}\|)^3\,f_0(\mathbf{x},\mathbf{v})\ \mathrm{d}\,\mathbf{x}\ \mathrm{d}\,\mathbf{v}$$

where $G^\varepsilon_{\mathrm{gc}}$ solves (3.11) with initial datum $G_0 = F^\varepsilon_{\mathrm{gc}}(0,\cdot)$, defined as the push-forward of f_0 by the map

$$(\mathbf{x},\mathbf{v}) \mapsto \Big(\mathbf{x} + \frac{\varepsilon}{B}\,\mathbb{J}\mathbf{v}\,,\ \tfrac{1}{2}\|\mathbf{v}\|^2 \ - \ \varepsilon\ \langle \mathbf{U}_{\mathbf{E}\times\mathbf{B}}(0,\mathbf{x}), \mathbf{v}\rangle\Big)\ .$$

Proof. The first result is a direct consequence of the abstract Proposition 2.1 and the estimates provided in Proposition 3.13 on the characteristic curves. The second one follows the same lines with the help of Proposition 3.15 instead of Proposition 3.13. □

Due to the special structure of the homogeneous two-dimensional case, with essentially the same proof one may also provide versions focusing only on the spatial variables and its ε-corrections. This involves the asymptotic equation

$$\partial_t r^\varepsilon \ + \ \varepsilon\ \mathrm{div}_{\mathbf{y}}\,(r^\varepsilon\,\mathbf{U}_{\mathbf{E}\times\mathbf{B}}) \ = \ 0\,. \tag{3.12}$$

Proposition 3.17. *Let $\mathbf{E} \in W^{2,\infty}$. There exists a constant C depending polynomially on $\|\mathbf{E}\|_{W^{2,\infty}}$ and B^{-1} such that the following holds for any solution f^ε to (3.1) with initial datum a positive density f_0.*
(i). Long-time first-order asymptotics. *ρ^ε defined by*

$$\rho^\varepsilon(t,\mathbf{x}) = \int_{\mathbf{R}^2} f^\varepsilon(t,\mathbf{x},\mathbf{v})\,\mathrm{d}\,\mathbf{v}$$

satisfies for a.e. $t \geq 0$

$$\|\rho^\varepsilon(t,\cdot) - r^\varepsilon(t,\cdot)\|_{\dot{W}^{-1,1}} \leq C\,\varepsilon\,e^{C\,\varepsilon\,t}\int_{\mathbf{R}^2\times\mathbf{R}^2}(1+\varepsilon+\|\mathbf{v}\|)\,f_0(\mathbf{x},\mathbf{v})\ \mathrm{d}\,\mathbf{x}\ \mathrm{d}\,\mathbf{v}$$

where r^ε solves (3.12) with initial datum $r_0 = \rho^\varepsilon(0,\cdot)$ given by

$$r_0(\mathbf{x}) = \int_{\mathbf{R}^2} f_0(\mathbf{x},\mathbf{v})\,\mathrm{d}\,\mathbf{v}\,.$$

(ii). Short-time third-order asymptotics. *$\rho^\varepsilon_{\mathrm{gc}}$ defined by*

$$\rho^\varepsilon_{\mathrm{gc}}(t,\mathbf{y}) = \int_{\mathbf{R}^2} f^\varepsilon\left(t, \mathbf{y} - \frac{\varepsilon}{B}\,\mathbb{J}\mathbf{v}, \mathbf{v}\right)\mathrm{d}\,\mathbf{v}$$

satisfies for a.e. $t \geq 0$

$$\|\rho^\varepsilon_{\mathrm{gc}}(t,\cdot) - r^\varepsilon_{\mathrm{gc}}(t,\cdot)\|_{\dot{W}^{-1,1}} \leq C\,\varepsilon^3\,e^{C\,\varepsilon\,t}\,(1+t)\int_{\mathbf{R}^2\times\mathbf{R}^2}(1+\varepsilon+\|\mathbf{v}\|)^2\,f_0(\mathbf{x},\mathbf{v})\ \mathrm{d}\,\mathbf{x}\ \mathrm{d}\,\mathbf{v}$$

where $r^\varepsilon_{\mathrm{gc}}$ solves (3.12) with initial datum $(r^\varepsilon_{\mathrm{gc}})_0 = \rho^\varepsilon_{\mathrm{gc}}(0,\cdot)$, given by

$$(r^\varepsilon_{\mathrm{gc}})_0(\mathbf{y}) = \int_{\mathbf{R}^2} f_0\left(\mathbf{y} - \frac{\varepsilon}{B}\,\mathbb{J}\mathbf{v}, \mathbf{v}\right)\mathrm{d}\,\mathbf{v}\,.$$

4. General three-dimensional case

We come back to the three-dimensional system

$$\left\{\begin{array}{l} \dfrac{\mathrm{d}\,\mathbf{x}}{\mathrm{d}\,t} \;=\; \mathbf{v}\,, \\[2ex] \dfrac{\mathrm{d}\,\mathbf{v}}{\mathrm{d}\,t} \;=\; \dfrac{\mathbf{v}\wedge\mathbf{B}(t,\mathbf{x})}{\varepsilon} \;+\; \mathbf{E}(t,\mathbf{x})\,, \end{array}\right. \tag{4.1}$$

and follow the pattern of the short-time analysis of Section 3. As there we do not mark ε-dependences as long as no confusion is possible.

4.1. **Slow variables and uniform bounds.** First, Lemma 3.2 stands without change in its statement or its proof.

Lemma 4.1. *Solutions to* (4.1) *starting from* $(\mathbf{x}_0,\mathbf{v}_0)$ *are defined globally in time and satisfy for any* $t\geq 0$

$$\left\{\begin{array}{l} \|\mathbf{x}(t)\| \;\leq\; \|\mathbf{x}_0\| + t\,\|\mathbf{v}_0\| + t^2\,\|\mathbf{E}\|_{L^\infty}, \\[2ex] \|\mathbf{v}(t)\| \;\leq\; \|\mathbf{v}_0\| + 2\,t\,\|\mathbf{E}\|_{L^\infty}\,. \end{array}\right.$$

Here some geometric preparation is needed to identify some set of slow variables. At leading order the fast motion is locally a rotation of $\mathbf{v}$ around $\mathbf{e}_\parallel(t,\mathbf{x})$ where we recall that $\mathbf{e}_\parallel$ is defined through

$$B(t,\mathbf{x}) = \|\mathbf{B}(t,\mathbf{x})\|\,, \qquad \mathbf{B}(t,\mathbf{x}) \;=\; B(t,\mathbf{x})\;\mathbf{e}_\parallel(t,\mathbf{x})\,.$$

As aforementioned this naturally suggests first a separation of $\mathbf{v}$ between a component aligned on $\mathbf{e}_\parallel(t,\mathbf{x})$, $v_\parallel(t,\mathbf{x},\mathbf{v})\;\mathbf{e}_\parallel(t,\mathbf{x})$, and a perpendicular component $\mathbf{v}_\perp(t,\mathbf{x},\mathbf{v})$, and second by mimicking the homogeneous case the introduction of a kinetic energy variable associated with $\mathbf{v}_\perp(t,\mathbf{x},\mathbf{v})$, $w_\perp(t,\mathbf{x},\mathbf{v}) = \frac{1}{2}\|\mathbf{v}_\perp(t,\mathbf{x},\mathbf{v})\|^2$.

We recall that the above decomposition is explicitly given as

$$\left\{\begin{array}{l} v_\parallel(t,\mathbf{x},\mathbf{v}) \;=\; \langle\mathbf{v}\,,\,\mathbf{e}_\parallel(t,\mathbf{x})\rangle, \\[2ex] \mathbf{v}_\perp(t,\mathbf{x},\mathbf{v}) = \mathbf{v} - v_\parallel(t,\mathbf{x},\mathbf{v})\;\mathbf{e}_\parallel(t,\mathbf{x}), \end{array}\right.$$

and that correspondingly we introduce the decomposition of the electric field[6] $\mathbf{E} = E_\parallel\;\mathbf{e}_\parallel + \mathbf{E}_\perp$,

$$E_\parallel(t,\mathbf{x},\mathbf{v}) = \langle\mathbf{E}(t,\mathbf{x})\,,\,\mathbf{e}_\parallel(t,\mathbf{x})\rangle\,, \qquad\qquad \mathbf{E}_\perp(t,\mathbf{x},\mathbf{v}) = \mathbf{E}(t,\mathbf{x}) - E_\parallel(t,\mathbf{x},\mathbf{v})\;\mathbf{e}_\parallel(t,\mathbf{x})\,.$$

Both to ease computations and to emphasize analogies with the two-dimensional case it is expedient to introduce, for any $\mathbf{x}\in\mathbf{R}^3$, the linear operator $\mathbb{J}(t,\mathbf{x})$ defined as

$$\mathbb{J}(t,\mathbf{x})\,\mathbf{a} \;=\; \mathbf{a}\wedge\mathbf{e}_\parallel(t,\mathbf{x})\,.$$

Going on with geometric considerations, we note that the following simple relations are of pervasive use in latter computations:

$$\left\{\begin{array}{l} \mathbb{J}(t,\mathbf{x})\,\mathbf{e}_\parallel(t,\mathbf{x}) \;=\; 0\,, \qquad \mathbf{e}_\parallel(t,\mathbf{x})\cdot\mathbb{J}(t,\mathbf{x})\mathbf{a} = 0\,, \\[2ex] \mathbb{J}(t,\mathbf{x})^2\,\mathbf{a} \;=\; -\mathbf{v}_\perp(t,\mathbf{x},\mathbf{a})\,, \qquad \mathbb{J}(t,\mathbf{x})^* = -\mathbb{J}(t,\mathbf{x})\,, \end{array}\right. \tag{4.2}$$

and $\mathbf{e}_\parallel(t,\mathbf{x})\cdot\partial_t\,\mathbf{e}_\parallel(t,\mathbf{x}) \;=\; 0$, $\mathbf{e}_\parallel(t,\mathbf{x})\cdot\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel(t,\mathbf{x})\,\mathbf{a} \;=\; 0$.

For the sake of concision, but somewhat inconsistently, from now on we shall use the shorthand $v_\parallel(t)$ for $v_\parallel(t,\mathbf{x}(t),\mathbf{v}(t))$ and similarly for $\mathbf{v}_\perp$ and $w_\perp$. We shall also identify functions of $(\mathbf{x},\mathbf{v})$ with functions of $(\mathbf{x},v_\parallel,\mathbf{v}_\perp)$. Then, we may split (4.1) as

$$\left\{\begin{array}{l} \dfrac{\mathrm{d}\,\mathbf{x}}{\mathrm{d}\,t} \;=\; \mathbf{v}, \\[2ex] \dfrac{\mathrm{d}\,v_\parallel}{\mathrm{d}\,t} \;=\; E_\parallel(t,\mathbf{x}) \;+\; \langle\mathbf{v}_\perp,\partial_t\,\mathbf{e}_\parallel(t,\mathbf{x}) + \mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel(t,\mathbf{x})\mathbf{v}\rangle\,, \\[2ex] \dfrac{\mathrm{d}\,w_\perp}{\mathrm{d}\,t} \;=\; \langle\mathbf{E}_\perp(t,\mathbf{x}) \;-\; v_\parallel\;(\partial_t\,\mathbf{e}_\parallel(t,\mathbf{x}) \;+\; \mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel(t,\mathbf{x})\,\mathbf{v})\,,\mathbf{v}_\perp\rangle\,, \end{array}\right. \tag{4.3}$$

[6] That is, $E_\parallel(t,\mathbf{x}) = v_\parallel(t,\mathbf{x},\mathbf{E}(t,\mathbf{x}))$, $\mathbf{E}_\perp(t,\mathbf{x}) = \mathbf{v}_\perp(t,\mathbf{x},\mathbf{E}(t,\mathbf{x}))$.

and

$$\frac{\mathrm{d}\,\mathbf{v}_\perp}{\mathrm{d}\,t} \;=\; \frac{B(t,\mathbf{x})}{\varepsilon}\,\mathbb{J}(t,\mathbf{x})\,\mathbf{v}_\perp \;+\; \mathbf{F}(t,\mathbf{x},\mathbf{v}), \tag{4.4}$$

where $\mathbf{v} = v_\parallel\,\mathbf{e}_\parallel + \mathbf{v}_\perp$ and the force field $\mathbf{F}$ is

$$\mathbf{F}(t,\mathbf{x},\mathbf{v}) = \mathbf{F}_0(t,\mathbf{x},v_\parallel) + \mathbf{F}_1(t,\mathbf{x},\mathbf{v}) + \mathbf{F}_2(t,\mathbf{x},\mathbf{v}),$$

with $\mathbf{F}_1$ depending linearly on $\mathbf{v}_\perp$, $\mathbf{F}_2$ quadratic in $\mathbf{v}_\perp$, explicitly

$$\left\{\begin{array}{rcl} \mathbf{F}_0(t,\mathbf{x},v_\parallel) & = & \mathbf{E}_\perp(t,\mathbf{x}) \,-\, v_\parallel\,\left(\partial_t\,\mathbf{e}_\parallel(t,\mathbf{x}) \,+\, v_\parallel\;\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel(t,\mathbf{x})\;\mathbf{e}_\parallel(t,\mathbf{x})\right)\,, \\ \mathbf{F}_1(t,\mathbf{x},\mathbf{v}) & = & -\,\langle \partial_t\,\mathbf{e}_\parallel(t,\mathbf{x}) \,+\, v_\parallel\;\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel(t,\mathbf{x})\;\mathbf{e}_\parallel(t,\mathbf{x})\,,\,\mathbf{v}_\perp\rangle\;\mathbf{e}_\parallel(\mathbf{x}) \\ & & -\,v_\parallel\;\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel(\mathbf{x})\,\mathbf{v}_\perp\,, \\ \mathbf{F}_2(t,\mathbf{x},\mathbf{v}_\perp) & = & -\,\langle \mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel(\mathbf{x})\,\mathbf{v}_\perp\,,\,\mathbf{v}_\perp\rangle\;\mathbf{e}_\parallel(\mathbf{x})\,. \end{array}\right. \tag{4.5}$$

Remark 4.2. *As already pointed out in Remark 3.1 along the analysis of the homogeneous case, it is convenient to work with a formulation containing some redundancy such as (4.3)-(4.4). Indeed, here, to suppress the apparent overdetermination one could for instance replace (4.4) with an equation for an angle of $\mathbf{v}_\perp$ but then one loses track of an important property of System (4.1): at principal order oscillations are linear in $\mathbf{v}_\perp$. In contrast, as already apparent in the homogeneous case or in the splitting of $\mathbf{F}$, all our algebraic manipulations will be organized by the degree of linearity in $\mathbf{v}_\perp$.*

4.2. **Elimination of linear terms.** A direct consequence of Lemma 4.1 and (4.3) is that $(\mathbf{x}, v_\parallel, w_\perp)$ in $W^{1,\infty}_{loc}$ and $\mathbf{v}_\perp$ in L^∞_{loc} are uniformly bounded with respect to ε. This is sufficient to extract converging sequences but not to take limits in the equations because of the nonlinearity in $\mathbf{v}_\perp$.

Instead, to proceed, we begin an uncoupling process similar to the one carried out in Section 3. Elimination, at leading order, of linear terms in $\mathbf{v}$ is summarized as

Lemma 4.3. *For any $\mathbf{L} \in W^{1,\infty}(\mathbf{R}^+_t;\,\mathcal{L}_1(\mathbf{R}^3,\mathbf{R}^p))$, $p \in \mathbf{N}^*$, solutions $(\mathbf{x},\mathbf{v})$ to (4.1) satisfy for a.e. $t \geq 0$,*

$$\mathbf{L}(t)\mathbf{v}_\perp \;=\; -\varepsilon\frac{\mathrm{d}}{\mathrm{d}\,t}\left[\mathbf{L}(t)\left(\frac{\mathbb{J}\mathbf{v}_\perp}{B}\right)\right] \;+\; \varepsilon\,\mathbf{L}'(t)\left(\frac{\mathbb{J}\mathbf{v}_\perp}{B}\right) \;+\; \varepsilon\mathbf{L}(t)\mathbf{U},$$

with the macroscopic velocity $\mathbf{U}$ given by

$$\mathbf{U}(t,\mathbf{x},\mathbf{v}) \;=\; \frac{\mathbb{J}\,\mathbf{F}}{B} + \left[\partial_t\left(\frac{\mathbb{J}}{B}\right) \;+\; \mathrm{d}_\mathbf{x}\left(\frac{\mathbb{J}}{B}\right)\mathbf{v}\right]\mathbf{v}_\perp\,. \tag{4.6}$$

Proof. Applying $\frac{\varepsilon\mathbb{J}}{B}$ to (4.4) and combining with the first line of (4.3) yields

$$\mathbf{v}_\perp \;=\; -\varepsilon\frac{\mathrm{d}}{\mathrm{d}\,t}\left[\frac{\mathbb{J}\,\mathbf{v}_\perp}{B(\mathbf{x})}\right] \;+\; \varepsilon\,\mathbf{U}(t,\mathbf{x},\mathbf{v})\,. \tag{4.7}$$

Then the result follows from the chain rule. □

Note that the macroscopic velocity $\mathbf{U}$ is split according to degree in $\mathbf{v}_\perp$ as $\mathbf{U} = \mathbf{U}_{10}+\mathbf{U}_{11}+\mathbf{U}_{12}$ where $\mathbf{U}_{10}$ contains terms which do not depend on $\mathbf{v}_\perp$,

$$\begin{array}{rcl} \mathbf{U}_{10}(t,\mathbf{x},v_\parallel) & = & \dfrac{\mathbb{J}(t,\mathbf{x})}{B(t,\mathbf{x})}\,\mathbf{F}_0(t,\mathbf{x},v_\parallel) \\ & = & \mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{x}) \;+\; v_\parallel^2\,\mathbf{U}_{\mathrm{curv}}(t,\mathbf{x}) \;+\; v_\parallel\,\mathbf{U}_{\partial_t}(t,\mathbf{x}) \\ & = & \mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{x}) \;+\; v_\parallel\,\mathbf{\Sigma}(t,\mathbf{x},v_\parallel), \end{array}$$

which corresponds to the classical drifts defined in (2.7)-(2.9), whereas $\mathbf{U}_{11}$ is given by

$$\mathbf{U}_{11}(t,\mathbf{x},v_\parallel,\mathbf{v}_\perp) \;=\; \frac{\mathbb{J}\,\mathbf{F}_1}{B}(t,\mathbf{x},v_\parallel,\mathbf{v}_\perp) \;+\; \left[\partial_t\left(\frac{\mathbb{J}}{B}\right) \;+\; v_\parallel\,\mathrm{d}_\mathbf{x}\left(\frac{\mathbb{J}}{B}\right)\mathbf{e}_\parallel\right]\mathbf{v}_\perp$$

and observing that $\mathbb{J}\,\mathbf{F}_2 = 0$, we have for $\mathbf{U}_{12}$,

$$\mathbf{U}_{12}(t,\mathbf{x},\mathbf{v}_\perp) \;=\; \left[\mathrm{d}_\mathbf{x}\left(\frac{\mathbb{J}}{B}\right)(t,\mathbf{x})\,\mathbf{v}_\perp\right]\mathbf{v}_\perp\,.$$

Let us anticipate that the partial elimination of $\mathbf{U}_{12}$ will give a contribution known as the grad-B drift and that encodes the influence of the variations of the intensity B on the spatial trajectory.

With Lemma 4.3, at leading-order in ε one may eliminate from (4.3) terms that are linear in $\mathbf{v}_\perp$. We first treat the first equation in (4.3) by applying Lemma 4.3 with $\mathbf{L}(t)\mathbf{v}_\perp = \mathbf{v}_\perp$, which reduces to (4.7). This leads to

$$\frac{\mathrm{d}\,\mathbf{x}_{\mathrm{gc}}}{\mathrm{d}\,t} = v_\parallel\ \mathbf{e}_\parallel(t,\mathbf{x}) + \varepsilon\,\mathbf{U}(t,\mathbf{x},\mathbf{v}), \tag{4.8}$$

where $\mathbf{U}$ is as in (4.6) and we have introduced the so-called guiding center already defined in (2.13).

Then we consider the second equation in (4.3) and apply Lemma 4.3 with

$$\mathbf{L}(t)\mathbf{v}_\perp = \langle \mathbf{v}_\perp, \partial_t\,\mathbf{e}_\parallel(t,\mathbf{x}(t)) + v_\parallel(t)\,\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel(t,\mathbf{x}(t))\,\mathbf{e}_\parallel(t,\mathbf{x}(t))\rangle,$$

to remove the linear part with respect to $\mathbf{v}_\perp$ in the right hand side and to derive an equation for a first correction of the parallel velocity,

$$\frac{\mathrm{d}}{\mathrm{d}\,t}\left[v_\parallel + \frac{\varepsilon}{B}\,\langle \mathbb{J}\,\mathbf{v}_\perp\,,\,\partial_t\,\mathbf{e}_\parallel + v_\parallel\ \mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\ \mathbf{e}_\parallel\rangle\right] = E_\parallel + \langle \mathbf{v}_\perp\,,\,\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\ \mathbf{v}_\perp\rangle + \varepsilon\,u_1(t,\mathbf{x},\mathbf{v}), \tag{4.9}$$

where $u_1 = u_{10} + u_{11} + u_{12} + u_{13}$ is obtained from

$$\begin{aligned} u_{10} &= \langle \mathbf{U}_{10}, \partial_t\,\mathbf{e}_\parallel + v_\parallel\ \mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\ \mathbf{e}_\parallel\rangle \\ &= \frac{1}{B}\,\langle \mathbb{J}\,\mathbf{E}_\perp\,,\,\partial_t\,\mathbf{e}_\parallel + v_\parallel\ \mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\ \mathbf{e}_\parallel\rangle \\ &= \langle \mathbf{E}\,,\,\mathbf{U}_{\partial_t} + v_\parallel\mathbf{U}_{\mathrm{curv}}\rangle = \langle \mathbf{E}\,,\,\mathbf{\Sigma}\rangle\,, \end{aligned}$$

where $\mathbf{U}_{\mathrm{curv}}$ and $\mathbf{U}_{\partial_t}$ are given in (2.7)-(2.8) and u_{11} is

$$\begin{aligned} u_{11} &= \langle \mathbf{U}_{11}, \partial_t\,\mathbf{e}_\parallel + v_\parallel\ \mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\ \mathbf{e}_\parallel\rangle + \frac{E_\parallel}{B}\,\langle \mathbb{J}\,\mathbf{v}_\perp, \mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\,\mathbf{e}_\parallel\rangle \\ &+ \frac{1}{B}\,\left\langle \mathbb{J}\mathbf{v}_\perp, \partial_t^2\,\mathbf{e}_\parallel + v_\parallel\,\partial_t(\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel)\,\mathbf{e}_\parallel + v_\parallel\ \mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\,\partial_t\,\mathbf{e}_\parallel\right\rangle \\ &+ \frac{v_\parallel}{B}\,\left\langle \mathbb{J}\mathbf{v}_\perp, \left[\partial_t(\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel) + v_\parallel\ \mathrm{d}_\mathbf{x}^2\,\mathbf{e}_\parallel\,\mathbf{e}_\parallel + v_\parallel\ (\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel)^2\right]\mathbf{e}_\parallel\right\rangle\,, \end{aligned}$$

whereas the last terms (u_{12}, u_{13}) are

$$\begin{aligned} u_{12} &= \langle \mathbf{U}_{12}(t,\mathbf{x},\mathbf{v})\,,\,\partial_t\,\mathbf{e}_\parallel + v_\parallel\ \mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\ \mathbf{e}_\parallel\rangle \\ &+ \frac{1}{B}\,\left\langle \mathbb{J}\mathbf{v}_\perp, \left[\partial_t(\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel) + v_\parallel\ \mathrm{d}_\mathbf{x}^2\,\mathbf{e}_\parallel\,\mathbf{e}_\parallel + v_\parallel\ (\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel)^2\right]\mathbf{v}_\perp\right\rangle \\ &+ \frac{1}{B}\,\langle \mathbf{v}_\perp, \partial_t\,\mathbf{e}_\parallel + v_\parallel\ \mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\ \mathbf{e}_\parallel\rangle\,\langle \mathbb{J}\mathbf{v}_\perp\,,\,\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\ \mathbf{e}_\parallel\rangle\,, \end{aligned}$$

and

$$u_{13} = \frac{1}{B}\,\langle \mathbf{v}_\perp, \mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\ \mathbf{v}_\perp\rangle\,\langle \mathbb{J}\,\mathbf{v}_\perp\,,\,\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\ \mathbf{e}_\parallel\rangle\,.$$

Finally we conclude the elimination of linear terms by reformulating the third equation in (4.3). To proceed we apply Lemma 4.3 with

$$\begin{aligned} \mathbf{L}(t)\mathbf{v}_\perp &= \langle \mathbf{v}_\perp\,,\,\mathbf{F}_0(t,\mathbf{x}(t),v_\parallel(t))\rangle \\ &= \left\langle \mathbf{v}_\perp\,,\,\mathbf{E}(t,\mathbf{x}(t)) - v_\parallel(t)\,\partial_t\,\mathbf{e}_\parallel(t,\mathbf{x}(t)) - (v_\parallel(t))^2\ \mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel(t,\mathbf{x}(t))\ \mathbf{e}_\parallel(t,\mathbf{x}(t))\right\rangle \end{aligned}$$

and naturally obtain an equation for a first correction of the kinetic energy in the perpendicular plan to the magnetic field,

$$\frac{\mathrm{d}}{\mathrm{d}\,t}\left[w_\perp + \frac{\varepsilon}{B}\,\langle \mathbb{J}\,\mathbf{v}_\perp, \mathbf{F}_0\rangle\right] = -v_\parallel\,\langle \mathbf{v}_\perp, \mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel(\mathbf{x})\mathbf{v}_\perp\rangle + \varepsilon\,d_1(t,\mathbf{x},\mathbf{v}), \tag{4.10}$$

where $d_1 = d_{11} + d_{12} + d_{13}$ is obtained from

$$\begin{aligned} d_{11} &= \langle \mathbf{U}_{11}\,,\,\mathbf{F}_0\rangle - \frac{E_\parallel}{B}\,\langle \mathbb{J}\mathbf{v}_\perp\,,\,\partial_t\,\mathbf{e}_\parallel + 2v_\parallel\,\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\,\mathbf{e}_\parallel\rangle \\ &+ \left\langle \frac{\mathbb{J}}{B}\mathbf{v}_\perp\,,\,\partial_t\mathbf{E} + v_\parallel\,\mathrm{d}_\mathbf{x}\,\mathbf{E}\,\mathbf{e}_\parallel\right\rangle \\ &- \frac{v_\parallel}{B}\,\langle \mathbb{J}\mathbf{v}_\perp, \partial_t^2\,\mathbf{e}_\parallel + v_\parallel\,\partial_t(\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel)\,\mathbf{e}_\parallel + v_\parallel\,\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\,\partial_t\,\mathbf{e}_\parallel\rangle \\ &- \frac{v_\parallel^2}{B}\,\left\langle \mathbb{J}\mathbf{v}_\perp\,,\,\left[\partial_t(\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel) + v_\parallel\,\mathrm{d}_\mathbf{x}^2\,\mathbf{e}_\parallel\,\mathbf{e}_\parallel + v_\parallel\,(\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel)^2\right]\mathbf{e}_\parallel\right\rangle\,, \end{aligned}$$

and d_{12} is given by

$$\begin{aligned} d_{12} &= \langle \mathbf{U}_{12}\,,\,\mathbf{F}_0\rangle + \left\langle \frac{\mathbb{J}}{B}\mathbf{v}_\perp\,,\,\mathrm{d}_\mathbf{x}\,\mathbf{E}\mathbf{v}_\perp\right\rangle \\ &- \frac{1}{B}\,\langle \mathbf{v}_\perp, \partial_t\,\mathbf{e}_\parallel + v_\parallel\,\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\,\mathbf{e}_\parallel\rangle\,\langle \mathbb{J}\mathbf{v}_\perp\,,\,\partial_t\,\mathbf{e}_\parallel + 2v_\parallel\,\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\,\mathbf{e}_\parallel\rangle \\ &- \frac{v_\parallel}{B}\,\left\langle \mathbb{J}\mathbf{v}_\perp\,,\,\left[\partial_t(\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel) + v_\parallel\,\mathrm{d}_\mathbf{x}^2\,\mathbf{e}_\parallel\,\mathbf{e}_\parallel + v_\parallel\,(\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel)^2\right]\mathbf{v}_\perp\right\rangle\,, \end{aligned}$$

and

$$d_{13} = -\frac{1}{B}\,\langle \mathbf{v}_\perp, \mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\,\mathbf{v}_\perp\rangle\,\langle \mathbb{J}\mathbf{v}_\perp\,,\,\partial_t\,\mathbf{e}_\parallel + 2v_\parallel\,\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\,\mathbf{e}_\parallel\rangle\,.$$

To summarize, gathering (4.8), (4.9), and (4.10), we have derived from (4.3)-(4.4) the following system of equations,

$$\left\{\begin{aligned} &\frac{\mathrm{d}\,\mathbf{x}_{\mathrm{gc}}}{\mathrm{d}\,t} = v_\parallel\,\mathbf{e}_\parallel + \varepsilon\,\mathbf{U}, \\ &\frac{\mathrm{d}}{\mathrm{d}\,t}\left[v_\parallel + \frac{\varepsilon}{B}\,\langle \mathbb{J}\,\mathbf{v}_\perp\,,\,\partial_t\,\mathbf{e}_\parallel + v_\parallel\,\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\,\mathbf{e}_\parallel\rangle\right] = E_\parallel + \langle \mathbf{v}_\perp, \mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\,\mathbf{v}_\perp\rangle + \varepsilon\,u_1, \\ &\frac{\mathrm{d}}{\mathrm{d}\,t}\left[w_\perp + \frac{\varepsilon}{B}\,\langle \mathbb{J}\,\mathbf{v}_\perp, \mathbf{F}_0\rangle\right] = -v_\parallel\,\langle \mathbf{v}_\perp, \mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\,\mathbf{v}_\perp\rangle + \varepsilon\,d_1\,. \end{aligned}\right. \tag{4.11}$$

Now to derive the leading order of an uncoupled slow dynamics it remains to analyze the contribution of the quadratic term $\langle \mathbf{v}_\perp, \mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\,\mathbf{v}_\perp\rangle$ that appears — at zeroth order with respect to ε — in the last equations of (4.11).

4.3. **Elimination of quadratic terms.** Lemma 4.3 encodes that all terms linear in $\mathbf{v}_\perp$ — the variable whose angle is oscillating at frequency $1/\varepsilon$ — are ε-small in $W^{-1,\infty}$. This is directly related to the fact that they all have zero mean with respect to the fast angle. In contrast, as in the homogeneous case, quadratic terms do produce slow contributions that are asymptotically relevant. The next result identifies what are those contributions.

To state it we introduce a notion of trace restricted to the plane orthogonal to $\mathbf{e}_\parallel$. For any $\mathbf{A} \in \mathcal{L}_2(\mathbf{R}^3, \mathbf{R}^p)$, $p \in \mathbf{N}^*$, at any point $\mathbf{x}$ and time t

$$\mathrm{Tr}_\perp^{t,\mathbf{x}}\,\mathbf{A} = \mathrm{Tr}\,\mathbf{A} - \mathbf{A}(\mathbf{e}_\parallel(t,\mathbf{x}), \mathbf{e}_\parallel(t,\mathbf{x})). \tag{4.12}$$

In particular, for any $\mathbf{a} \in \mathbf{R}^3$ orthogonal to $\mathbf{e}_\parallel(t,\mathbf{x})$, we observe that

$$\|\mathbf{a}\|^2\,\mathrm{Tr}_\perp^{t,\mathbf{x}}\,\mathbf{A} = \mathbf{A}(\mathbf{a},\mathbf{a}) + \mathbf{A}(\mathbb{J}(t,\mathbf{x})\,\mathbf{a}, \mathbb{J}(t,\mathbf{x})\,\mathbf{a})\,.$$

Since with any linear operator $\mathbf{A} \in \mathcal{L}_1(\mathbf{R}^3, \mathbf{R}^3 \otimes \mathbf{R}^p)$ one may associate a quadratic operator in $\mathcal{L}_2(\mathbf{R}^3, \mathbf{R}^p)$ by $(\mathbf{a},\mathbf{b}) \mapsto \langle \mathbf{a}, \mathbf{A}\mathbf{b}\rangle$ the above definitions may be extended to such operators by identification. We also recall that $\Re$ denotes the symmetric part.

Lemma 4.4. *For any $\mathbf{A} \in W^{1,\infty}(\mathbf{R}_t^+; \mathcal{L}_2(\mathbf{R}^3, \mathbf{R}^p))$, $p \in \mathbf{N}^*$, solutions to (4.1) satisfy at a.e. t*

$$\mathbf{A}(t)(\mathbf{v}_\perp(t), \mathbf{v}_\perp(t)) = w_\perp(t)\,\mathrm{Tr}_\perp^{t,\mathbf{x}(t)}\,(\mathbf{A}(t)) - \varepsilon\,\frac{\mathrm{d}\,\chi_\mathbf{A}}{\mathrm{d}\,t}(t) + \varepsilon\,\eta_\mathbf{A}(t),$$

where

$$\left\{\begin{aligned} \chi_\mathbf{A} &= \frac{1}{2\,B}\,\Re(\mathbf{A})\,(\mathbf{v}_\perp, \mathbb{J}\mathbf{v}_\perp)\,, \\ \eta_\mathbf{A} &= \frac{1}{2}\,\Re(\mathbf{A})(\mathbf{v}_\perp, \mathbf{U}) + \frac{1}{2\,B}\,\Re(\mathbf{A})'\,(\mathbf{v}_\perp, \mathbb{J}\mathbf{v}_\perp) + \frac{1}{2\,B}\,\Re(\mathbf{A})\,(\mathbf{F}, \mathbb{J}\mathbf{v}_\perp)\,. \end{aligned}\right.$$

Remark 4.5. *Consistently with Remarks 3.6 and 3.9, note that $\chi_{\mathbf{A}}$ has itself no slow component at leading order since $\Re(\mathbf{A}(t))\,(\cdot,\mathbb{J}(t,\mathbf{x})(\cdot))$ is trace-free on the plane orthogonal to $\mathbf{e}_\parallel(t,\mathbf{x})$.*

Proof. One may assume without loss of generality that $\mathbf{A}$ is symmetric. Then by combining (4.4) and (4.7) and using (4.2) we derive that

$$\begin{aligned}\varepsilon\frac{\mathrm{d}}{\mathrm{d}\,t}\left[\mathbf{A}\left(\mathbf{v}_\perp,\frac{\mathbb{J}\,\mathbf{v}_\perp}{B}\right)\right] &= -\mathbf{A}(\mathbf{v}_\perp,\mathbf{v}_\perp)\,+\,\mathbf{A}\left(\mathbb{J}\mathbf{v}_\perp,\mathbb{J}\mathbf{v}_\perp\right)\\ &+\ \varepsilon\left[\mathbf{A}(\mathbf{v}_\perp,\mathbf{U})\,+\,\frac{1}{B}\,\mathbf{A}'\left(\mathbf{v}_\perp,\mathbb{J}\mathbf{v}_\perp\right)\,+\,\frac{1}{B}\,\mathbf{A}\left(\mathbf{F},\mathbb{J}\mathbf{v}_\perp\right)\right].\end{aligned}$$

By multiplying the latter by 1/2 and adding $\mathbf{A}(\mathbf{v}_\perp,\mathbf{v}_\perp)$ on both sides one achieves the proof. □

4.4. **Proof of Theorem 2.2.** We have now sufficient materials to prove Theorem 2.2 on the asymptotic behavior on solutions to (4.3)-(4.4) when $\varepsilon\to 0$.

On the one hand, applying Lemma 4.4 with the quadratic form associated with $\mathrm{d}_{\mathbf{x}}\,\mathbf{e}_\parallel(t,\mathbf{x})$, one may partially eliminate quadratic terms in $\mathbf{v}_\perp$ from (4.11). As a result

$$\langle\mathbf{v}_\perp,\mathrm{d}_{\mathbf{x}}\,\mathbf{e}_\parallel\,\mathbf{v}_\perp\rangle = w_\perp\,\mathrm{div}_{\mathbf{x}}(\mathbf{e}_\parallel)\,-\,\frac{\varepsilon}{2}\frac{\mathrm{d}}{\mathrm{d}\,t}\left[\left\langle\mathbf{v}_\perp,\frac{\Re(\mathrm{d}_{\mathbf{x}}\,\mathbf{e}_\parallel)\mathbb{J}}{B}\,\mathbf{v}_\perp\right\rangle\right]+\varepsilon\,u_2, \tag{4.13}$$

where $u_2=u_{21}+u_{22}+u_{23}$ with u_{21},

$$u_{21}\,=\,\frac{1}{2}\left[\langle\mathbf{v}_\perp,\Re(\mathrm{d}_{\mathbf{x}}\,\mathbf{e}_\parallel)\mathbf{U}_{10}\rangle+\left\langle\mathbf{F}_0,\frac{\Re(\mathrm{d}_{\mathbf{x}}\,\mathbf{e}_\parallel)\mathbb{J}}{B}\,\mathbf{v}_\perp\right\rangle\right],$$

whereas u_{22} is given by

$$\begin{aligned}u_{22}\,&=\,\frac{1}{2}\left[\langle\mathbf{v}_\perp,\Re(\mathrm{d}_{\mathbf{x}}\,\mathbf{e}_\parallel)\mathbf{U}_{11}\rangle+\left\langle\mathbf{F}_1,\frac{\Re(\mathrm{d}_{\mathbf{x}}\,\mathbf{e}_\parallel)\mathbb{J}}{B}\,\mathbf{v}_\perp\right\rangle\right]\\ &+\,\frac{1}{2\,B}\left\langle\mathbf{v}_\perp,\Re\left(\partial_t(\mathrm{d}_{\mathbf{x}}\,\mathbf{e}_\parallel)+v_\parallel\,\mathrm{d}^2_{\mathbf{x}}\,\mathbf{e}_\parallel\,\mathbf{e}_\parallel\right)\mathbb{J}\mathbf{v}_\perp\right\rangle\end{aligned}$$

and u_{23}

$$\begin{aligned}u_{23}\,&=\,\frac{1}{2}\left[\langle\mathbf{v}_\perp,\Re(\mathrm{d}_{\mathbf{x}}\,\mathbf{e}_\parallel)\mathbf{U}_{12}\rangle+\left\langle\mathbf{F}_2,\frac{\Re(\mathrm{d}_{\mathbf{x}}\,\mathbf{e}_\parallel)\mathbb{J}}{B}\,\mathbf{v}_\perp\right\rangle\right]\\ &+\,\frac{1}{2\,B}\left\langle\mathbf{v}_\perp,\Re\left(\mathrm{d}^2_{\mathbf{x}}\,\mathbf{e}_\parallel\,\mathbf{v}_\perp\right)\mathbb{J}\mathbf{v}_\perp\right\rangle.\end{aligned}$$

Substituting (4.13) in the second equation of (4.11), we get an equation for the corrected parallel velocity v_{gc}, defined in (2.14), that is,

$$\frac{\mathrm{d}\,v_{\mathrm{gc}}}{\mathrm{d}\,t}\,=\,E_\parallel\,+\,w_\perp\,\mathrm{div}_{\mathbf{x}}\,\mathbf{e}_\parallel+\varepsilon\,(u_1+u_2)\,. \tag{4.14}$$

On the other hand, we proceed in the same way with the quadratic term associated with $-v_\parallel(t)\;\mathrm{d}_{\mathbf{x}}\,\mathbf{e}_\parallel(t,\mathbf{x}(t))$ to transform the third equation of (4.11) into a new equation for a correction to the part of the kinetic energy in the perpendicular plan to the magnetic field direction w_{gc}, already defined in (2.15),

$$\frac{\mathrm{d}\,w_{\mathrm{gc}}}{\mathrm{d}\,t}\,=\,-v_\parallel w_\perp\,\mathrm{div}_{\mathbf{x}}\,\mathbf{e}_\parallel+\varepsilon\,(d_1+d_2), \tag{4.15}$$

where $d_2=d_{21}+d_{22}+d_{23}+d_{24}$ with

$$\begin{cases}d_{21}\,=\,-v_\parallel\,u_{21}\,,\\ d_{22}\,=\,-v_\parallel\,u_{22}\,-\,\dfrac{E_\parallel}{2\,B}\,\langle\mathbf{v}_\perp,\Re\left(\mathrm{d}_{\mathbf{x}}\,\mathbf{e}_\parallel\right)\mathbb{J}\,\mathbf{v}_\perp\rangle\,,\\ d_{23}\,=\,-v_\parallel\,u_{23}\,-\,\dfrac{1}{2B}\,\langle\mathbf{v}_\perp,\partial_t\,\mathbf{e}_\parallel+v_\parallel\,\mathrm{d}_{\mathbf{x}}\,\mathbf{e}_\parallel\,\mathbf{e}_\parallel\rangle\,\langle\mathbf{v}_\perp,\Re\left(\mathrm{d}_{\mathbf{x}}\,\mathbf{e}_\parallel\right)\mathbb{J}\mathbf{v}_\perp\rangle\,,\end{cases}$$

and the quartic term d_{24} is

$$d_{24}\,=\,-\frac{1}{2\,B}\,\langle\mathbf{v}_\perp,\mathrm{d}_{\mathbf{x}}\,\mathbf{e}_\parallel\,\mathbf{v}_\perp\rangle\,\langle\mathbf{v}_\perp,\Re\left(\mathrm{d}_{\mathbf{x}}\,\mathbf{e}_\parallel\right)\mathbb{J}\mathbf{v}_\perp\rangle\,.$$

With (4.14) and (4.15), System (4.3)-(4.4) yields

$$
\left\{
\begin{aligned}
\frac{\mathrm{d}\,\mathbf{x}_{\mathrm{gc}}}{\mathrm{d}\,t} &= v_\parallel\,\mathbf{e}_\parallel(t,\mathbf{x}) + \varepsilon\,\mathbf{U}(t,\mathbf{x},\mathbf{v}),\\
\frac{\mathrm{d}\,v_{\mathrm{gc}}}{\mathrm{d}\,t} &= E_\parallel(t,\mathbf{x}) \,+\, w_\perp\,\operatorname{div}_{\mathbf{x}}\mathbf{e}_\parallel(t,\mathbf{x}) + \varepsilon\,(u_1+u_2)(t,\mathbf{x},\mathbf{v}),\\
\frac{\mathrm{d}\,w_{\mathrm{gc}}}{\mathrm{d}\,t} &= -v_\parallel w_\perp\,\operatorname{div}_{\mathbf{x}}\mathbf{e}_\parallel(t,\mathbf{x}) + \varepsilon\,(d_1+d_2)(t,\mathbf{x},\mathbf{v})\,.
\end{aligned}
\right.
\tag{4.16}
$$

At this juncture, the leading-order part of the slow evolution system (4.16) is already uncoupled from the fast equation (4.4). This allows to derive the following asymptotic result by mimicking the analysis of Section 3, relying this time on Lemma 4.1 to bound remainders.

Proposition 4.6. *Under the assumptions of Theorem 2.2, there exists a constant $C>0$, depending polynomially on $\|\mathbf{E}\|_{W^{1,\infty}}$, $\|B^{-1}\|_{W^{1,\infty}}$ and $\|\,\mathbf{e}_\parallel\,\|_{W^{2,\infty}}$, such that the following holds. Consider $(\mathbf{x}^\varepsilon,\mathbf{v}^\varepsilon)$ a solution to (4.1) starting from $(\mathbf{x}_0,\mathbf{v}_0)$. Then, for a.e. $t\geq 0$*

$$
\begin{aligned}
\|\mathbf{x}^\varepsilon(t)-\mathbf{y}(t)\| + \|v^\varepsilon_\parallel(t)-v(t)\| + \|w^\varepsilon_\perp(t)-w(t)\|\\
\leq C\,\varepsilon\,e^{C\,t\,(\|\mathbf{v}_0\|^3+t^3)}\,\|\mathbf{v}_0\|\,(1+\|\mathbf{v}_0\|^2),
\end{aligned}
$$

where $(\mathbf{y},v,e)$ solves

$$
\left\{
\begin{aligned}
\frac{\mathrm{d}\,\mathbf{y}}{\mathrm{d}\,t} &= v\;\mathbf{e}_\parallel(t,\mathbf{y}),\\
\frac{\mathrm{d}\,v}{\mathrm{d}\,t} &= E_\parallel(t,\mathbf{y}) \,+\, w\;\operatorname{div}_{\mathbf{x}}\mathbf{e}_\parallel(t,\mathbf{y}),\\
\frac{\mathrm{d}\,w}{\mathrm{d}\,t} &= -v\,w\;\operatorname{div}_{\mathbf{x}}\mathbf{e}_\parallel(t,\mathbf{y})\,,
\end{aligned}
\right.
$$

with initial data $\mathbf{y}(0)=\mathbf{x}_0$, $v(0)=v_{\parallel 0}$ and $w(0)=\frac{1}{2}\|\mathbf{v}_{\perp 0}\|^2$.

Finally to achieve the proof of Theorem 2.2, we simply apply Proposition 2.1 where the slow map $\mathcal{A}(t,\cdot)$ is given by $(\mathbf{x},\mathbf{v})\mapsto(\mathbf{x},v_\parallel(t,\mathbf{x},\mathbf{v}),w_\perp(t,\mathbf{x},\mathbf{v}))$ and the weights $\mathcal{M}$ are given by

$$
\mathcal{M}(t,\mathbf{x},\mathbf{v}) \;=\; C\,\varepsilon\,e^{C\,t\,(\|\mathbf{v}\|^3+t^3)}\,\|\mathbf{v}\|\,(1+\|\mathbf{v}\|^2)
$$

with C as in Proposition 4.6.

4.5. **Elimination of higher-order terms.** Though the latter result does provide some insights, in general it fails to capture leading-order dynamics of all slow variables, since some of them are slower than what can be described with a system of zeroth order in ε. A simple example is the essentially two-dimensional case where $\mathbf{e}_\parallel$ is constant and asymptotically at zeroth order only $v_\parallel$ and the parallel component of $\mathbf{y}$ are moving.

To provide a more comprehensive picture, we need a system containing terms of order ε. The purpose of Theorem 2.4 is to take into account this correction. Lemmas 4.3 and 4.4 already contains the basis to clean ε-terms of (4.16) that are of order at most 2 with respect to $\mathbf{v}_\perp$. Yet, d_1, d_2 and u_2 contain cubic terms and u_2 also exhibits a quartic term. Therefore we need to investigate how to handle those.

We first show how to eliminate cubic terms.

Lemma 4.7. *Let $\mathbf{E}\in W^{1,\infty}$ and $\mathbf{B}$ be such that $1/B\in W^{1,\infty}$ and $\mathbf{e}_\parallel\in W^{1,\infty}$. There exists a constant C depending polynomially on $\|\mathbf{E}\|_{W^{1,\infty}}$, $\|B^{-1}\|_{W^{1,\infty}}$ and $\|\,\mathbf{e}_\parallel\,\|_{W^{1,\infty}}$ such that for any $\mathbf{A}\in W^{1,\infty}\left(\mathbf{R}^+_t;\,\mathcal{L}_3(\mathbf{R}^3,\mathbf{R}^p)\right)$, $p\in\mathbf{N}^*$, solutions to (4.1) satisfy at a.e. $t\geq 0$,*

$$
\mathbf{A}(t)(\mathbf{v}_\perp(t),\mathbf{v}_\perp(t),\mathbf{v}_\perp(t)) \;=\; -\varepsilon\frac{\mathrm{d}\,\chi_{\mathbf{A}}}{\mathrm{d}\,t}(t) \,+\, \varepsilon\,\eta_{\mathbf{A}}(t),
$$

for some $(\chi_{\mathbf{A}},\eta_{\mathbf{A}})$ such that for a.e. t

$$
\left\{
\begin{aligned}
\|\chi_{\mathbf{A}}(t)\| &\leq C\,\|\mathbf{A}(t)\|\,\|\mathbf{v}_\perp(t)\|^3\,,\\
\|\eta_{\mathbf{A}}(t)\| &\leq C\,\|\mathbf{v}_\perp(t)\|^2\,\left[\,\|\mathbf{A}'(t)\|\,\|\mathbf{v}_\perp(t)\| + \|\mathbf{A}(t)\|\,(1+\|\mathbf{v}(t)\|^2)\,\right]\,.
\end{aligned}
\right.
$$

Proof. One may assume without loss of generality that $\mathbf{A}$ is symmetric-valued. Then from (4.4) and (4.7) stem

$$\begin{aligned}\varepsilon\,\frac{\mathrm{d}}{\mathrm{d}\,t}\left[\mathbf{A}\left(\frac{\mathbb{J}\,\mathbf{v}_\perp}{B},\mathbf{v}_\perp,\mathbf{v}_\perp\right)\right](t) &= -\,\mathbf{A}(\mathbf{v}_\perp,\mathbf{v}_\perp,\mathbf{v}_\perp)+2\mathbf{A}(\mathbb{J}\mathbf{v}_\perp,\mathbb{J}\mathbf{v}_\perp,\mathbf{v}_\perp)\\ &+ \frac{\varepsilon}{B}\,\mathbf{A}'\left(\mathbb{J}\,\mathbf{v}_\perp,\mathbf{v}_\perp,\mathbf{v}_\perp\right)+\varepsilon\,\mathbf{A}\left(\mathbf{U},\mathbf{v}_\perp,\mathbf{v}_\perp\right)\\ &+ \frac{2\varepsilon}{B}\,\mathbf{A}\left(\mathbb{J}\,\mathbf{v}_\perp,\mathbf{F},\mathbf{v}_\perp\right)\end{aligned}$$

and

$$\begin{aligned}\varepsilon\,\frac{\mathrm{d}}{\mathrm{d}\,t}\left[B^2\mathbf{A}\left(\frac{\mathbb{J}\,\mathbf{v}_\perp}{B},\frac{\mathbb{J}\,\mathbf{v}_\perp}{B},\frac{\mathbb{J}\,\mathbf{v}_\perp}{B}\right)\right](t) &= -3\,\mathbf{A}(\mathbb{J}\mathbf{v}_\perp,\mathbb{J}\mathbf{v}_\perp,\mathbf{v}_\perp)\\ &+ \frac{\varepsilon}{B}\,\mathbf{A}'\left(\mathbb{J}\,\mathbf{v}_\perp,\mathbb{J}\,\mathbf{v}_\perp,\mathbb{J}\,\mathbf{v}_\perp\right)+3\,\varepsilon\,\mathbf{A}\left(\mathbf{U},\mathbb{J}\,\mathbf{v}_\perp,\mathbb{J}\,\mathbf{v}_\perp\right)\\ &+ \frac{2\,\varepsilon}{B}\,\frac{\partial_t B+\mathrm{d}_{\mathbf{x}}\,B\,\mathbf{v}}{B}\,\mathbf{A}\left(\mathbb{J}\,\mathbf{v}_\perp,\mathbb{J}\,\mathbf{v}_\perp,\mathbb{J}\,\mathbf{v}_\perp\right).\end{aligned}$$

Then summing the former with 2/3 of the latter yields the result. □

To complete the uncoupling at order ε remains the task of analyzing the possible contribution of quartic terms. By using (4.4), (4.7) and the fact that $\mathbb{J}(t,\mathbf{x})^2\mathbf{a}=-\mathbf{a}$ for any $\mathbf{a}$ orthogonal to $\mathbf{e}_\parallel(t,\mathbf{x})$, it is possible to achieve this task at the level of generality considered so far. As a result one would prove that in general the elimination of quartic terms may indeed leave relevant slow terms. However, for concision's sake we choose to specialize the discussion to the specific form required by

$$d_{24} \;=\; \langle\mathbf{v}_\perp,\Re\left(\mathrm{d}_{\mathbf{x}}\,\mathbf{e}_\parallel\right)\mathbf{v}_\perp\rangle\;\langle\mathbf{v}_\perp,\Re\left(\mathrm{d}_{\mathbf{x}}\,\mathbf{e}_\parallel\right)\mathbb{J}\mathbf{v}_\perp\rangle$$

and that may be eliminated at leading-order.

Lemma 4.8. *Let $\mathbf{E}\in W^{1,\infty}$ and $\mathbf{B}$ be such that $1/B\in W^{1,\infty}$ and $\mathbf{e}_\parallel\in W^{1,\infty}$. There exists a constant C depending polynomially on $\|\mathbf{E}\|_{W^{1,\infty}}$, $\|B^{-1}\|_{W^{1,\infty}}$ and $\|\,\mathbf{e}_\parallel\,\|_{W^{1,\infty}}$ such that for any symmetric-valued $\mathbf{A}\in W^{1,\infty}\left(\mathbf{R}_t^+;\,\mathcal{L}_2(\mathbf{R}^3,\mathbf{R}^p)\right)$, $p\in\mathbf{N}^*$, solutions to (4.1) satisfy at a.e. t*

$$\mathbf{A}(t)(\mathbf{v}_\perp(t),\mathbf{v}_\perp(t))\;\times\;\mathbf{A}(t)(\mathbf{v}_\perp(t),\mathbb{J}(t,\mathbf{x}(t))\mathbf{v}_\perp(t))\;=\;-\varepsilon\frac{\mathrm{d}\,\chi_{\mathbf{A},2}}{\mathrm{d}\,t}(t)+\varepsilon\eta_{\mathbf{A},2}(t),$$

for some $(\chi_{\mathbf{A},2},\eta_{\mathbf{A},2})$ such that for a.e. t

$$\begin{cases}\|\chi_{\mathbf{A},2}(t)\| \;\le\; C\,\|\mathbf{A}(t)\|^2\,\|\mathbf{v}_\perp(t)\|^4\,,\\ \|\eta_{\mathbf{A},2}(t)\| \;\le\; C\,\|\mathbf{A}(t)\|\,\|\mathbf{v}_\perp(t)\|^3\left(\|\mathbf{A}'(t)\|\,\|\mathbf{v}_\perp(t)\|+\|\mathbf{A}(t)\|\,(1+\|\mathbf{v}(t)\|^2)\right).\end{cases}$$

Proof. We introduce $\widetilde{\mathbf{A}}(t)\;:\;(\mathbf{a},\mathbf{b})\mapsto\mathbf{A}(t)(\mathbf{a},\mathbb{J}(t,\mathbf{x}(t))\mathbf{b})$ and recall from Remark 4.5 that $\mathrm{Tr}_\perp^{t,\mathbf{x}(t)}\,\widetilde{\mathbf{A}}(t)=0$. Thus revisiting the proof of Lemma 4.4 without symmetrization yields

$$\widetilde{\mathbf{A}}(\mathbf{v}_\perp,\mathbf{v}_\perp)=-\varepsilon\frac{\mathrm{d}}{\mathrm{d}\,t}\left(\frac{1}{2\,B}\,\widetilde{\mathbf{A}}\left(\mathbf{v}_\perp,\mathbb{J}\mathbf{v}_\perp\right)\right)+\varepsilon\,\widetilde{\eta}_{\widetilde{\mathbf{A}}}$$

with

$$\widetilde{\eta}_{\widetilde{\mathbf{A}}}\;=\;\frac{1}{2}\,\widetilde{\mathbf{A}}(\mathbf{v}_\perp,\mathbf{U})\;+\;\frac{1}{2\,B}\,\widetilde{\mathbf{A}}'\left(\mathbf{v}_\perp,\mathbb{J}\mathbf{v}_\perp\right)\;+\;\frac{1}{2\,B}\,\widetilde{\mathbf{A}}\left(\mathbf{F},\mathbb{J}\mathbf{v}_\perp\right).$$

By multiplying with $\mathbf{A}(\mathbf{v}_\perp,\mathbf{v}_\perp)$, one derives the result with

$$\begin{cases}\chi_{\mathbf{A},2}\;=\;-\dfrac{1}{4\,B}\left(\mathbf{A}\left(\mathbf{v}_\perp,\mathbf{v}_\perp\right)\right)^2,\\ \eta_{\mathbf{A},2}\;=\;\mathbf{A}\left(\mathbf{v}_\perp,\mathbf{v}_\perp\right)\widetilde{\eta}_{\widetilde{\mathbf{A}}}-\dfrac{1}{4\,B}\dfrac{\partial_t B+\mathrm{d}_{\mathbf{x}}\,B\,\mathbf{v}}{B}\left(\mathbf{A}\left(\mathbf{v}_\perp,\mathbf{v}_\perp\right)\right)^2.\end{cases}$$

□

4.6. **Proof of Theorem 2.4.** To spare some pieces of notation, in the justification of the foregoing claim we shall use $\lesssim$ to denote $\leq C\times$ with C a local variable depending only and polynomially on $\|\mathbf{E}\|_{W^{2,\infty}}$, $\|B^{-1}\|_{W^{2,\infty}}$ and $\|\,\mathbf{e}_\parallel\,\|_{W^{3,\infty}}$. All along we consider $(\mathbf{x}(t),\mathbf{v}(t))$ a solution to (4.3)-(4.4). We observe that

$$\left\|\frac{\mathrm{d}\,\mathbf{x}}{\mathrm{d}\,t}\right\| \lesssim \|\mathbf{v}\|\,, \qquad \left|\frac{\mathrm{d}\,v_\parallel}{\mathrm{d}\,t}\right| \lesssim 1+\|\mathbf{v}_\perp\|\,\|\mathbf{v}\|$$

and

$$\|\mathbf{F}\| \lesssim 1+\|\mathbf{v}\|^2\,, \qquad \|\mathbf{U}\| \lesssim 1+\|\mathbf{v}\|^2\,.$$

First we apply Lemma 4.3 with the linear application $\mathbf{v}_\perp \mapsto \mathbf{U}_{11}(t,\mathbf{x}(t),v_\parallel(t),\mathbf{v}_\perp)$ and Lemma 4.4 with the quadratic function $\mathbf{v}_\perp \mapsto \mathbf{U}_{12}(t,\mathbf{x}(t),v_\parallel(t),\mathbf{v}_\perp)$. As a result there exist functions $\boldsymbol{\chi}_x^1$, $\boldsymbol{\chi}_x^2$, $\boldsymbol{\eta}_x^1$, $\boldsymbol{\eta}_x^2$ such that

$$\begin{aligned}\mathbf{U}_{11} &= -\varepsilon\frac{\mathrm{d}\,\boldsymbol{\chi}_\mathbf{x}^1}{\mathrm{d}\,t}+\varepsilon\,\boldsymbol{\eta}_\mathbf{x}^1\\ \mathbf{U}_{12} &= w_\perp\,\mathrm{Tr}_\perp^{t,\mathbf{x}(t)}\left(\mathrm{d}_\mathbf{x}\left(\frac{\mathbb{J}}{B}\right)\right)-\varepsilon\,\frac{\mathrm{d}\,\boldsymbol{\chi}_\mathbf{x}^2}{\mathrm{d}\,t}+\varepsilon\,\boldsymbol{\eta}_\mathbf{x}^2\end{aligned}$$

so that, with $\boldsymbol{\chi}_\mathbf{x}=\boldsymbol{\chi}_\mathbf{x}^1+\boldsymbol{\chi}_\mathbf{x}^2$ and $\boldsymbol{\eta}_\mathbf{x}=\boldsymbol{\eta}_\mathbf{x}^1+\boldsymbol{\eta}_\mathbf{x}^2$

$$\mathbf{U} = \mathbf{U}_{10}+w_\perp\,\mathrm{Tr}_\perp^{t,\mathbf{x}(t)}\left(\mathrm{d}_\mathbf{x}\left(\frac{\mathbb{J}}{B}\right)\right)-\varepsilon\frac{\mathrm{d}\,\boldsymbol{\chi}_\mathbf{x}}{\mathrm{d}\,t}+\varepsilon\boldsymbol{\eta}_\mathbf{x}$$

with

$$\|\boldsymbol{\chi}_\mathbf{x}\| \lesssim \|\mathbf{v}_\perp\|\,\|\mathbf{v}\|\,, \qquad \|\boldsymbol{\eta}_\mathbf{x}\| \lesssim \|\mathbf{v}\|\,(1+\|\mathbf{v}\|^2)\,. \tag{4.17}$$

Before going on we make $\mathrm{Tr}_\perp^{t,\mathbf{x}(t)}\left(\mathrm{d}_\mathbf{x}\left(\frac{\mathbb{J}}{B}\right)\right)$ more explicit. First, by differentiating $\frac{\mathbb{J}}{B}\,\mathbf{e}_\parallel=0$ we get

$$0=\mathrm{d}_\mathbf{x}\left[\left(\frac{\mathbb{J}}{B}\right)\mathbf{e}_\parallel\right]\,\mathbf{e}_\parallel = \left[\mathrm{d}_\mathbf{x}\left(\frac{\mathbb{J}}{B}\right)\mathbf{e}_\parallel\right]\,\mathbf{e}_\parallel+\frac{\mathbb{J}}{B}\,\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\,\mathbf{e}_\parallel,$$

thus, with $\mathbf{U}_{\mathrm{curv}}$ defined as in (2.7),

$$\left[\mathrm{d}_\mathbf{x}\left(\frac{\mathbb{J}}{B}\right)\mathbf{e}_\parallel\right]\,\mathbf{e}_\parallel = -\frac{\mathbb{J}}{B}\,\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\,\mathbf{e}_\parallel = \mathbf{U}_{\mathrm{curv}}\,.$$

Therefore, recalling definition (4.12) and Lemma 2.5,

$$\mathrm{Tr}_\perp^{t,\mathbf{x}(t)}\left(\mathrm{d}_\mathbf{x}\left(\frac{\mathbb{J}}{B}\right)\right) = \mathrm{div}_\mathbf{x}\left(\frac{\mathbb{J}}{B}\right)-\mathbf{U}_{\mathrm{curv}} = \mathbf{U}_{\mathrm{curl}\,\mathbf{e}_\parallel}+\mathbf{U}_{\nabla B\times\mathbf{B}}.$$

In particular the equation on $\mathbf{x}_{\mathrm{gc}}$ takes the form

$$\frac{\mathrm{d}}{\mathrm{d}\,t}\left[\mathbf{x}_{\mathrm{gc}}+\varepsilon^2\boldsymbol{\chi}_x\right] = v_\parallel\,\mathbf{e}_\parallel(t,\mathbf{x})+\varepsilon^2\,\boldsymbol{\eta}_x+\varepsilon\,\mathbf{U}_{\mathrm{drift}}(t,\mathbf{x},v_\parallel,w_\perp),$$

where $\mathbf{U}_{\mathrm{drift}}$ is defined in (2.10) and $(\boldsymbol{\chi}_x,\boldsymbol{\eta}_x)$ satisfies (4.17).

Likewise we may clean up the second equation of (4.16). After some calculations, with arguments identical to those used hereabove, we obtain

$$\begin{aligned}u_1 &= \langle\mathbf{E},\boldsymbol{\Sigma}\rangle-\varepsilon\frac{\mathrm{d}\,\chi_\parallel^1}{\mathrm{d}\,t}+\varepsilon\eta_\parallel^1\\ &+ w_\perp\left[\left\langle\mathrm{div}_\mathbf{x}\left(\frac{\mathbb{J}}{B}\right),\partial_t\,\mathbf{e}_\parallel+v_\parallel\,\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\,\mathbf{e}_\parallel\right\rangle+\left\langle\frac{\mathbb{J}}{B}\,\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\,\mathbf{e}_\parallel,\partial_t\,\mathbf{e}_\parallel\right\rangle\right]\\ &- w_\perp\,\mathrm{Tr}\left(\frac{\mathbb{J}}{B}\left(\partial_t(\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel)+v_\parallel\,\mathrm{d}_\mathbf{x}^2\,\mathbf{e}_\parallel\,\mathbf{e}_\parallel+v_\parallel\,(\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel)^2\right)\right)\\ &- w_\perp\,\mathrm{Tr}\left[\frac{\mathbb{J}}{B}\,(\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\,\mathbf{e}_\parallel\,(\partial_t\,\mathbf{e}_\parallel)^*)\right].\end{aligned}$$

and

$$u_2 = -\varepsilon\frac{\mathrm{d}\,\chi_\parallel^2}{\mathrm{d}\,t}+\varepsilon\,\eta_\parallel^2$$

with $\chi_\parallel=\chi_\parallel^1+\chi_\parallel^2$ and $\eta_\parallel=\eta_\parallel^1+\eta_\parallel^2$ satisfying

$$|\chi_\parallel| \lesssim \|\mathbf{v}_\perp\|\,(1+\|\mathbf{v}\|^2)\,, \qquad |\eta_\parallel| \lesssim 1+\|\mathbf{v}\|^4\,. \tag{4.18}$$

In the computations aforementioned we stress that we have made extensive use of relations $\mathbb{J}\,\mathbf{e}_\parallel = 0$, $(\mathrm{d}_\mathbf{x}\,\mathbb{J}\,\mathbf{e}_\parallel)\,\mathbf{e}_\parallel = -\mathbb{J}(\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel)\,\mathbf{e}_\parallel$, $\langle \mathbf{a}, \mathbb{J}\,\mathbf{a}\rangle = 0$ for any $\mathbf{a} \in \mathbf{R}^3$, and $(\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel)^*\,\mathbf{e}_\parallel = 0$. In particular we point out that, by the skew-symmetry of values of $\mathrm{d}_\mathbf{x}\mathbb{J}\ \mathbf{e}_\parallel$ and $\partial_t\mathbb{J}$

$$\begin{aligned}\mathrm{Tr}_\perp\left[\Re(\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel)\,\mathrm{d}_\mathbf{x}\,\mathbb{J}\ \mathbf{e}_\parallel\right] &= \mathrm{Tr}\left[\Re(\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel)\,\mathrm{d}_\mathbf{x}\,\mathbb{J}\,\mathbf{e}_\parallel\right] - \frac{1}{2}\langle \mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\,\mathbf{e}_\parallel, [\mathrm{d}_\mathbf{x}\,\mathbb{J}\,\mathbf{e}_\parallel]\,\mathbf{e}_\parallel\rangle \\ &= \frac{1}{2}\langle \mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\,\mathbf{e}_\parallel, \mathbb{J}\,\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\,\mathbf{e}_\parallel\rangle = 0\end{aligned}$$

and

$$\begin{aligned}\mathrm{Tr}_\perp\left[\Re(\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel)\partial_t\mathbb{J}\right] &= \mathrm{Tr}\left[\Re(\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel)\partial_t\mathbb{J}\right] - \frac{1}{2}\langle \mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\,\mathbf{e}_\parallel, \partial_t(\mathbb{J})\,\mathbf{e}_\parallel\rangle \\ &= \frac{1}{2}\langle \mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\,\mathbf{e}_\parallel, \mathbb{J}\partial_t\,\mathbf{e}_\parallel\rangle\,.\end{aligned}$$

To simplify further the expression of u_1 we observe that

$$\mathrm{Tr}\left[\frac{\mathbb{J}}{B}\left(\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\ \mathbf{e}_\parallel\,(\partial_t\,\mathbf{e}_\parallel)^*\right)\right] = \left\langle \frac{\mathbb{J}}{B}\,\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\,\mathbf{e}_\parallel, \partial_t\,\mathbf{e}_\parallel\right\rangle$$

and, by the skew-symmetry of values of $\mathbb{J}$, that

$$\left\langle \mathrm{div}_\mathbf{x}\left(\frac{\mathbb{J}}{B}\right), \partial_t\,\mathbf{e}_\parallel\right\rangle - \mathrm{Tr}\left(\frac{\mathbb{J}}{B}\partial_t(\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel)\right) = -\,\mathrm{div}_\mathbf{x}\left(\frac{\mathbb{J}}{B}\partial_t\,\mathbf{e}_\parallel\right) = \mathrm{div}_\mathbf{x}\,\mathbf{U}_{\partial_t}$$

$$\left\langle \mathrm{div}_\mathbf{x}\left(\frac{\mathbb{J}}{B}\right), \mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\,\mathbf{e}_\parallel\right\rangle - \mathrm{Tr}\left(\frac{\mathbb{J}}{B}\left(\mathrm{d}^2_\mathbf{x}\,\mathbf{e}_\parallel\ \mathbf{e}_\parallel + (\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel)^2\right)\right) = -\,\mathrm{div}_\mathbf{x}\left(\frac{\mathbb{J}}{B}\,\mathrm{d}_\mathbf{x}\,\mathbf{e}_\parallel\ \mathbf{e}_\parallel\right) = \mathrm{div}_\mathbf{x}\,\mathbf{U}_{\mathrm{curv}}\,.$$

As a result

$$u_1 = -\varepsilon\,\frac{\mathrm{d}\,\chi^1_\parallel}{\mathrm{d}\,t} + \varepsilon\,\eta^1_\parallel + \langle \mathbf{\Sigma}\,,\,\mathbf{E}\rangle + w_\perp\,\mathrm{div}_\mathbf{x}\,\mathbf{\Sigma}$$

with $(\chi^1_\parallel, \eta^1_\parallel)$ as before.

Therefore gathering the expressions for u_1 and u_2, we derive

$$\frac{\mathrm{d}}{\mathrm{d}\,t}\left[v_{\mathrm{gc}} + \varepsilon^2\chi_\parallel\right] = E_\parallel + w_\perp\ \mathrm{div}_\mathbf{x}\,\mathbf{e}_\parallel + \varepsilon\left[\langle \mathbf{\Sigma}\,,\,\mathbf{E}\rangle + w_\perp\ \mathrm{div}_\mathbf{x}\,\mathbf{\Sigma}\right] + \varepsilon^2\,\eta_\parallel$$

with $(\chi_\parallel, \eta_\parallel)$ satisfying (4.18).

Finally we treat the last equation of (4.16) in the same manner. This leads to

$$\begin{aligned} d_1 &= -\varepsilon\frac{\mathrm{d}\,\chi^1_\perp}{\mathrm{d}\,t} + \varepsilon\eta^1_\perp - v_\parallel w_\perp\,\mathrm{div}_\mathbf{x}\,\mathbf{\Sigma} \\ &+ w_\perp\left[\langle \mathbf{U}_{\mathrm{curl}\,\mathbf{e}_\parallel} + \mathbf{U}_{\nabla B\times\mathbf{B}}\,,\,\mathbf{E}\rangle - \mathrm{Tr}\left(\frac{\mathbb{J}}{B}\,\mathrm{d}_\mathbf{x}\,\mathbf{E}\right)\right], \\ d_2 &= -\varepsilon\frac{\mathrm{d}\,\chi^2_\perp}{\mathrm{d}\,t} + \varepsilon\,\eta^2_\perp, \end{aligned}$$

with $\chi_\perp = \chi^1_\perp + \chi^2_\perp$ and $\eta_\perp = \eta^1_\perp + \eta^2_\perp$ satisfying

$$|\chi_\perp| \lesssim \|\mathbf{v}_\perp\|\,(1 + \|\mathbf{v}\|^3)\,, \qquad |\eta_\perp| \lesssim 1 + \|\mathbf{v}\|^5\,. \tag{4.19}$$

Now to simplify the expression for d_1 we observe that from Lemma 2.5 stems

$$\langle \mathbf{U}_{\mathrm{curl}\,\mathbf{e}_\parallel} + \mathbf{U}_{\mathrm{curv}} + \mathbf{U}_{\nabla B\times\mathbf{B}}\,,\,\mathbf{E}\rangle - \mathrm{Tr}\left(\frac{\mathbb{J}}{B}\,\mathrm{d}_\mathbf{x}\,\mathbf{E}\right) = -\,\mathrm{div}_\mathbf{x}\left(\frac{\mathbb{J}\,\mathbf{E}}{B}\right) = -\,\mathrm{div}_\mathbf{x}\,\mathbf{U}_{\mathbf{E}\times\mathbf{B}},$$

so that

$$d_1 = -\varepsilon\frac{\mathrm{d}\,\chi^1_\perp}{\mathrm{d}\,t} + \varepsilon\eta^1_\perp - w_\perp\left[v_\parallel\ \mathrm{div}_\mathbf{x}\,\mathbf{\Sigma} + \langle \mathbf{U}_{\mathrm{curv}}\,,\,\mathbf{E}\rangle + \mathrm{div}_\mathbf{x}\,\mathbf{U}_{\mathbf{E}\times\mathbf{B}}\right].$$

The upshot is

$$\frac{\mathrm{d}}{\mathrm{d}\,t}\left[w_{\mathrm{gc}} + \varepsilon^2\,\chi_\perp\right] = -v_\parallel w_\perp\,\mathrm{div}_\mathbf{x}\,\mathbf{e}_\parallel - \varepsilon w_\perp\left[\mathrm{div}_\mathbf{x}\,(\mathbf{U}_{\mathbf{E}\times\mathbf{B}} + v_\parallel\mathbf{\Sigma}) + \langle \mathbf{U}_{\mathrm{curv}}, \mathbf{E}\rangle\right] + \varepsilon^2\,\eta_\perp$$

with $(\chi_\perp, \eta_\perp)$ satisfying (4.19).

Altogether we have derived

$$
(4.20)\qquad \begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}\,t}\left[\mathbf{x}_{\mathrm{gc}}+\varepsilon^2\boldsymbol{\chi}_{\mathbf{x}}\right] &= v_\parallel\,\mathbf{e}_\parallel(t,\mathbf{x})\,+\,\varepsilon\,\mathbf{U}_{\mathrm{drift}}(t,\mathbf{x},v_\parallel,w_\perp)\,+\,\varepsilon^2\,\boldsymbol{\eta}_{\mathbf{x}},\\ \dfrac{\mathrm{d}}{\mathrm{d}\,t}\left[v_{\mathrm{gc}}\,+\,\varepsilon^2\chi_\parallel\right] &= E_\parallel(t,\mathbf{x})\,+\,w_\perp\,\operatorname{div}_{\mathbf{x}}\mathbf{e}_\parallel(t,\mathbf{x})\,+\,\varepsilon^2\,\eta_\parallel\\ &\quad+\varepsilon\,\left(\langle\boldsymbol{\Sigma},\mathbf{E}\rangle\,+\,w_\perp\,\operatorname{div}_{\mathbf{x}}\boldsymbol{\Sigma}\right)(t,\mathbf{x},v_\parallel)\,,\\ \dfrac{\mathrm{d}}{\mathrm{d}\,t}\left[w_{\mathrm{gc}}\,+\,\varepsilon^2\chi_\perp\right] &= -v_\parallel\,w_\perp\,\operatorname{div}_{\mathbf{x}}\mathbf{e}_\parallel(t,\mathbf{x})\,+\,\varepsilon^2\,\eta_\perp\\ &\quad-\varepsilon\,w_\perp\,\left(\operatorname{div}_{\mathbf{x}}(\mathbf{U}_{\mathbf{E}\times\mathbf{B}}+v_\parallel\boldsymbol{\Sigma})+\langle\mathbf{U}_{\mathrm{curv}}\,,\,\mathbf{E}\rangle\right)(t,\mathbf{x},v_\parallel), \end{cases}
$$

with error bounds (4.17)-(4.18)-(4.19). Now we want to write (4.20) in terms of $(\mathbf{x}_{\mathrm{gc}},v_{\mathrm{gc}},w_{\mathrm{gc}})$ plus remainders. As in Section 3.5 corrections — of size ε^2 — stemming from terms of size ε may be considered directly as error terms. Yet here some terms of size 1 are present and to deal with corrections arising from those we follow a different path: first linearize them — a process that produces errors of size ε^2 that can be handled directly — then remove the terms of size ε introduced in this way by using Lemmas 4.3 and 4.4 and the fact that $(\mathbf{x},v_\parallel,w_\perp)$ differs from $(\mathbf{x}_{\mathrm{gc}},v_{\mathrm{gc}},w_{\mathrm{gc}})$ by terms that are either linear in $\mathbf{v}_\perp$ or quadratic in $\mathbf{v}_\perp$ but trace-free in the plane orthogonal to $\mathbf{e}_\parallel(t,\mathbf{x})$, as follows from Remark 4.5. Besides aforementioned estimates this elimination also requires

$$
\left|\frac{\mathrm{d}\,w_\perp}{\mathrm{d}\,t}\right| \lesssim \|\mathbf{v}_\perp\|\,(1+|v_\parallel|\,\|\mathbf{v}\|)
$$

and results in new functions $(\widehat{\boldsymbol{\chi}}_{\mathbf{x}},\widehat{\chi}_\parallel,\widehat{\chi}_\perp)$, and $(\widehat{\boldsymbol{\eta}}_{\mathbf{x}},\widehat{\eta}_\parallel,\widehat{\eta}_\perp)$ such that,

$$
(4.21)\qquad \begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}\,t}\left[\mathbf{x}_{\mathrm{gc}}+\varepsilon^2\,\widehat{\boldsymbol{\chi}}_{\mathbf{x}}\right] &= v_{\mathrm{gc}}\;\mathbf{e}_\parallel(t,\mathbf{x}_{\mathrm{gc}})\,+\,\varepsilon\mathbf{U}_{\mathrm{drift}}(t,\mathbf{x}_{\mathrm{gc}},v_{\mathrm{gc}},w_{\mathrm{gc}})\,+\,\varepsilon^2\,\widehat{\boldsymbol{\eta}}_{\mathbf{x}}\,,\\ \dfrac{\mathrm{d}}{\mathrm{d}\,t}\left[v_{\mathrm{gc}}+\varepsilon^2\,\widehat{\chi}_\parallel\right] &= E_\parallel(t,\mathbf{x}_{\mathrm{gc}})\,+\,w_{\mathrm{gc}}\operatorname{div}_{\mathbf{x}}\mathbf{e}_\parallel(t,\mathbf{x}_{\mathrm{gc}})\,+\,\varepsilon^2\,\widehat{\eta}_\parallel\\ &\quad+\varepsilon\,\left(\langle\boldsymbol{\Sigma},\mathbf{E}\rangle\,+\,w_{\mathrm{gc}}\operatorname{div}_{\mathbf{x}}\boldsymbol{\Sigma}\right)(t,\mathbf{x}_{\mathrm{gc}},v_{\mathrm{gc}})\,,\\ \dfrac{\mathrm{d}}{\mathrm{d}\,t}\left[w_{\mathrm{gc}}+\varepsilon^2\,\widehat{\chi}_\perp\right] &= -v_{\mathrm{gc}}\,w_{\mathrm{gc}}\;\operatorname{div}_{\mathbf{x}}\mathbf{e}_\parallel(t,\mathbf{x}_{\mathrm{gc}})\,+\,\varepsilon^2\,\widehat{\eta}_\perp\\ &\quad-\varepsilon\,w_{\mathrm{gc}}\,\left(\operatorname{div}_{\mathbf{x}}(\mathbf{U}_{\mathbf{E}\times\mathbf{B}}+v_{\mathrm{gc}}\boldsymbol{\Sigma})+\langle\mathbf{U}_{\mathrm{curv}}\,,\,\mathbf{E}\rangle\right)(t,\mathbf{x}_{\mathrm{gc}},v_{\mathrm{gc}}), \end{cases}
$$

with

$$
\|\widehat{\boldsymbol{\chi}}_{\mathbf{x}}\| \lesssim \|\mathbf{v}_\perp\|\,\|\mathbf{v}\|\,,\qquad |\widehat{\chi}_\parallel| \lesssim \|\mathbf{v}_\perp\|\,(1+\|\mathbf{v}\|^2)\,,\qquad |\widehat{\chi}_\perp| \lesssim \|\mathbf{v}_\perp\|\,(1+\|\mathbf{v}\|^3)
$$

and

$$
\|\widehat{\boldsymbol{\eta}}_{\mathbf{x}}\| \lesssim \|\mathbf{v}\|\,(1+\|\mathbf{v}\|^2)\,,\qquad |\widehat{\eta}_\parallel| \lesssim 1+\|\mathbf{v}\|^4\,,\qquad |\widehat{\eta}_\perp| \lesssim 1+\|\mathbf{v}\|^5\,.
$$

At this stage arguing as in Section 3.5 we prove the following

Proposition 4.9. *Under the assumptions of Theorem 2.4, there exists a constant $C>0$, depending polynomially on $\|\mathbf{E}\|_{W^{2,\infty}}$, $\|B^{-1}\|_{W^{2,\infty}}$ and $\|\,\mathbf{e}_\parallel\,\|_{W^{3,\infty}}$ such that the following holds. Let $(\mathbf{x}^\varepsilon,\mathbf{v}^\varepsilon)$ be the solution to (4.1) starting from $(\mathbf{x}_0,\mathbf{v}_0)$ and $\mathbf{Z}^\varepsilon_{\mathrm{gc}}=(\mathbf{x}^\varepsilon_{\mathrm{gc}},v^\varepsilon_{\mathrm{gc}},w^\varepsilon_{\mathrm{gc}})$ be deduced from it through (2.13)-(2.15). Then, for a.e. $t\geq 0$*

$$
\left\|\mathbf{Z}^\varepsilon_{\mathrm{gc}}(t)-\mathbf{Z}^\varepsilon(t)\right\| \;\leq\; C\,\varepsilon^2\,e^{C\,t\,(\|\mathbf{v}_0\|^3+t^3)(1+\varepsilon\,(\|\mathbf{v}_0\|+t))}\,\|\mathbf{v}_0\|\,(1+\|\mathbf{v}_0\|^3),
$$

where $\mathbf{Z}^\varepsilon=(\mathbf{y}^\varepsilon,v^\varepsilon,w^\varepsilon)$ solves

$$
\begin{cases} \dfrac{\mathrm{d}\,\mathbf{y}^\varepsilon}{\mathrm{d}\,t} &= v^\varepsilon\;\mathbf{e}_\parallel(t,\mathbf{y}^\varepsilon)\,+\,\varepsilon\,\mathbf{U}_{\mathrm{drift}}(t,\mathbf{y}^\varepsilon,v^\varepsilon,w^\varepsilon),\\ \dfrac{\mathrm{d}\,v^\varepsilon}{\mathrm{d}\,t} &= E_\parallel(t,\mathbf{y}^\varepsilon)\,+\,w^\varepsilon\;\operatorname{div}_{\mathbf{x}}\mathbf{e}_\parallel(t,\mathbf{y}^\varepsilon)\;+\,\varepsilon\,\left(\langle\boldsymbol{\Sigma},\mathbf{E}\rangle+w^\varepsilon\operatorname{div}_{\mathbf{x}}\boldsymbol{\Sigma}\right)(t,\mathbf{y}^\varepsilon,v^\varepsilon),\\ \dfrac{\mathrm{d}\,w^\varepsilon}{\mathrm{d}\,t} &= -v^\varepsilon\,w^\varepsilon\;\operatorname{div}_{\mathbf{x}}\mathbf{e}_\parallel(t,\mathbf{y}^\varepsilon)-\,\varepsilon\,w^\varepsilon\,\left(\operatorname{div}_{\mathbf{x}}(\mathbf{U}_{\mathbf{E}\times\mathbf{B}}+v^\varepsilon\boldsymbol{\Sigma})+\langle\mathbf{U}_{\mathrm{curv}},\mathbf{E}\rangle\right)(t,\mathbf{y}^\varepsilon,v^\varepsilon), \end{cases}
$$

with $\mathbf{Z}^\varepsilon(0)=\mathbf{Z}^\varepsilon_{\mathrm{gc}}(0)$.

We may then use Proposition 2.1 to derive Theorem 2.4 from Proposition 4.9.

5. Proof of Theorem 2.8

We now want to provide a three-dimensional analogous to Section 3.6, that is, a description of a long-time slow dynamics. Yet the presence of terms of order 1 in the (short) time asymptotics prevents this from happening unless those terms generates a confined purely oscillatory dynamics in some components and one focuses on the remaining ones.

This requires a special form of geometry of magnetic field lines. We introduce now an example of such a configuration.

5.1. **Geometric framework.** Let us fix a unitary vector $\mathbf{e}_z \in \mathbf{S}^2$ to define an axis of symmetry. For vectors $\mathbf{x} \notin \mathbf{R}\,\mathbf{e}_z$, it is expedient to introduce

$$z(\mathbf{x}) = \langle \mathbf{x}, \mathbf{e}_z \rangle\,, \qquad r(\mathbf{x}) = \|\,\mathbf{e}_z \wedge \mathbf{x}\|\,, \qquad \mathbf{e}_r(\mathbf{x}) = \frac{\mathbf{x} - z(\mathbf{x})\,\mathbf{e}_z}{r(\mathbf{x})}\,.$$

Note that then by construction $\mathbf{e}_r(\mathbf{x})$ is unitary, orthogonal to $\mathbf{e}_z$ and

$$\mathbf{x} \;=\; r(\mathbf{x})\;\mathbf{e}_r(\mathbf{x}) + z(\mathbf{x})\;\mathbf{e}_z\;.$$

We now assume that far from the axis $\mathbf{R}\,\mathbf{e}_z$ the magnetic field is stationary, toroidal, axi-symmetric and non vanishing, that is (up to a change of $\mathbf{e}_z$ with $-\,\mathbf{e}_z$), for some $r_0 > 0$, when $r(\mathbf{x}) \geq r_0$

$$\mathbf{e}_\parallel(\mathbf{x}) \;=\; \frac{\mathbf{e}_z \wedge \mathbf{x}}{r(\mathbf{x})} \qquad\qquad \text{and} \qquad\qquad B(\mathbf{x}) \;=\; b(r(\mathbf{x}), z(\mathbf{x}))$$

for some function b with $1/b \in L^\infty([r_0, +\infty[\,\times \mathbf{R})$. Note that the first equality already ensures $\mathrm{div}_{\mathbf{x}}(\mathbf{e}_\parallel)(\mathbf{x}) = 0$ when $r(\mathbf{x}) \geq r_0$ so that the second one is actually equivalent to the natural condition $\mathrm{div}_{\mathbf{x}}\,\mathbf{B} \equiv 0$.

In this context straightforward computations yield when $r(\mathbf{x}) \geq r_0$

$$\mathrm{d}_{\mathbf{x}}\, r(\mathbf{x}) = \langle \mathbf{e}_r(\mathbf{x}),\, \cdot\,\rangle\,, \qquad \mathrm{d}_{\mathbf{x}}\,\mathbf{e}_r(\mathbf{x}) = \frac{\mathbf{e}_\parallel(\mathbf{x})}{r(\mathbf{x})}\,\langle \mathbf{e}_\parallel(\mathbf{x}),\, \cdot\,\rangle\,, \qquad \mathrm{d}_{\mathbf{x}}\,\mathbf{e}_\parallel(\mathbf{x}) = -\frac{\mathbf{e}_r(\mathbf{x})}{r(\mathbf{x})}\,\langle \mathbf{e}_\parallel(\mathbf{x}),\, \cdot\,\rangle$$

and for any $\mathbf{a} \in \mathbf{R}^3$,

$$\mathrm{d}_{\mathbf{x}}\,\mathbb{J}(\mathbf{x})\,\mathbf{a} \;=\; \frac{\langle \mathbf{e}_\parallel(\mathbf{x}), \mathbf{a}\rangle}{r(\mathbf{x})}\,(\mathbf{e}_z\,\langle \mathbf{e}_\parallel(\mathbf{x}),\, \cdot\,\rangle - \mathbf{e}_\parallel(\mathbf{x})\,\langle \mathbf{e}_z,\, \cdot\,\rangle)$$

and

$$\mathrm{d}^2_{\mathbf{x}}\,\mathbf{e}_\parallel(\mathbf{x})(\mathbf{e}_\parallel(\mathbf{x}),\, \cdot\,) \;=\; \frac{\mathbf{e}_r(\mathbf{x})}{(r(\mathbf{x}))^2}\,\langle \mathbf{e}_r(\mathbf{x}),\, \cdot\,\rangle - \frac{\mathbf{e}_\parallel(\mathbf{x})}{(r(\mathbf{x}))^2}\,\langle \mathbf{e}_\parallel(\mathbf{x}),\, \cdot\,\rangle\,,$$

so that in particular

$$\mathrm{div}_{\mathbf{x}}\left(\frac{\mathbb{J}}{B}\right)(\mathbf{x}) \;=\; \partial_r\left(\frac{1}{b}\right)(r(\mathbf{x}), z(\mathbf{x}))\;\mathbf{e}_z - \partial_z\left(\frac{1}{b}\right)(r(\mathbf{x}), z(\mathbf{x}))\;\mathbf{e}_r(\mathbf{x})$$

and

$$\mathrm{div}_{\mathbf{x}}\,\mathbf{\Sigma}(\mathbf{x}) \;=\; \frac{1}{r(\mathbf{x})}\,\partial_z\left(\frac{1}{b}\right)(r(\mathbf{x}), z(\mathbf{x}))\,,$$

whereas the drifts $\mathbf{F}$ and $\mathbf{U}$ are given by

$$\mathbf{F}(t, \mathbf{x}, \mathbf{v}) \;=\; \mathbf{E}_\perp(t, \mathbf{x}) \;+\; \frac{v_\parallel}{r(\mathbf{x})}\,\langle \mathbf{e}_r(\mathbf{x}), \mathbf{v}_\perp\rangle\;\mathbf{e}_\parallel(t, \mathbf{x}) \;+\; \frac{v_\parallel^2}{r(\mathbf{x})}\;\mathbf{e}_r(\mathbf{x})$$

and

$$\begin{aligned}\mathbf{U}(t, \mathbf{x}, \mathbf{v}) \;&=\; \mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t, \mathbf{x}) + \frac{v_\parallel}{r(\mathbf{x})\,b(r(\mathbf{x}), z(\mathbf{x}))}\,(v_\parallel\,\mathbf{e}_z - \langle \mathbf{e}_z, \mathbf{v}_\perp\rangle\;\mathbf{e}_\parallel(\mathbf{x}))\\ &+\; \left(\partial_r\left(\frac{1}{b}\right)(r(\mathbf{x}), z(\mathbf{x}))\,\langle \mathbf{e}_r(\mathbf{x}), \mathbf{v}_\perp\rangle + \partial_z\left(\frac{1}{b}\right)(r(\mathbf{x}), z(\mathbf{x}))\,\langle \mathbf{e}_z, \mathbf{v}_\perp\rangle\right)\mathbb{J}(\mathbf{x})\mathbf{v}_\perp\,.\end{aligned}$$

We assume moreover that $E_\parallel(t, \mathbf{x}) \equiv 0$ when $r(\mathbf{x}) \geq r_0$ and that $\mathbf{E}$ is axi-symmetric, hence

$$\mathbf{E}_\perp(t, \mathbf{x}) \;=\; E_r(t, r(\mathbf{x}), z(\mathbf{x}))\;\mathbf{e}_r(\mathbf{x}) \;+\; E_z(t, r(\mathbf{x}), z(\mathbf{x}))\;\mathbf{e}_z,$$

for some E_r and E_z. With these notational conventions, we may derive from System (4.20)

$$
(5.1)\qquad \begin{cases}
\dfrac{\mathrm{d}}{\mathrm{d}\,t}\Big[r-\dfrac{\varepsilon}{b}\,\langle \mathbf{e}_z,\mathbf{v}_\perp\rangle+\varepsilon^2\langle \mathbf{e}_r,\boldsymbol{\chi}_x\rangle\Big] \\
\qquad\qquad = \varepsilon\left(-\dfrac{E_z}{b}-w_\perp\,\partial_z\left(\dfrac{1}{b}\right)\right)+\varepsilon^2\left(\langle \mathbf{e}_r,\boldsymbol{\eta}_x\rangle+\dfrac{v_\parallel}{r}\langle \mathbf{e}_\parallel,\boldsymbol{\chi}_x\rangle\right), \\
\dfrac{\mathrm{d}}{\mathrm{d}\,t}\Big[z+\dfrac{\varepsilon}{b}\,\langle \mathbf{e}_r,\mathbf{v}_\perp\rangle+\varepsilon^2\langle \mathbf{e}_z,\boldsymbol{\chi}_x\rangle\Big] \\
\qquad\qquad = \varepsilon\left(\dfrac{E_r}{b}+\dfrac{v_\parallel^2}{r\,b}+w_\perp\,\partial_r\left(\dfrac{1}{b}\right)\right)+\varepsilon^2\langle \mathbf{e}_z,\boldsymbol{\eta}_x\rangle, \\
\dfrac{\mathrm{d}}{\mathrm{d}\,t}\Big[v_\parallel+\dfrac{\varepsilon\,v_\parallel}{r\,b}\,\langle \mathbf{e}_z,\mathbf{v}_\perp\rangle+\varepsilon^2\chi_\parallel\Big] \\
\qquad\qquad = \dfrac{\varepsilon\,v_\parallel}{r}\left(\dfrac{E_z}{b}+w_\perp\partial_z\left(\dfrac{1}{b}\right)\right)+\varepsilon^2\eta_\parallel\,, \\
\dfrac{\mathrm{d}}{\mathrm{d}\,t}\Big[w_\perp-\varepsilon\,\left(\langle \mathbf{U}_{\mathbf{E}\times\mathbf{B}},\mathbf{v}_\perp\rangle+\dfrac{v_\parallel^2}{r\,b}\,\langle \mathbf{e}_z,\mathbf{v}_\perp\rangle\right)+\varepsilon^2\chi_\perp\Big] \\
\qquad\qquad = \varepsilon\,w_\perp\left(\partial_r\left(\dfrac{E_z}{b}\right)-\partial_z\left(\dfrac{E_r}{b}\right)-\dfrac{v_\parallel^2}{r}\partial_z\left(\dfrac{1}{b}\right)\right)+\varepsilon^2\eta_\perp\,.
\end{cases}
$$

5.2. **Uniform bounds and asymptotics.** As follows from the proof of the following proposition the assumption $E_\parallel(t,\mathbf{x})\equiv 0$ is sufficient by itself to improve the dependence on time of uniform bounds on velocity.

Proposition 5.1. *Under the assumptions of Theorem 2.4, for any $r_1>r_0$, there exist positive ε_0, τ_0 and C_0, $(1/\varepsilon_0,1/\tau_0,C_0)$ depending polynomially on $1/r_0$, $1/(r_1-r_0)$ and $\|(E_r,E_z,1/b)\|_{W^{2,\infty}([r_0,\infty[\times\mathbf{R})}$ such that any $(\mathbf{x},\mathbf{v})$ solution to (4.1) starting from $(\mathbf{x}_0,\mathbf{v}_0)$ with*

$$
r(\mathbf{x}_0)\geq r_1\,,\qquad\qquad 0<\varepsilon\,\leq\,\frac{\varepsilon_0}{1+\|\mathbf{v}_0\|}
$$

satisfies for a.e. $0\leq t\leq[\tau_0\,(1+\|\mathbf{v}_0\|^2)^{-1}]/\varepsilon$

$$
r(\mathbf{x}(t))\geq r_0\qquad \textit{and}\qquad \|\mathbf{v}(t)\|\leq C_0\,(\|\mathbf{v}_0\|+\varepsilon\,t)\,.
$$

Proof. We start with

$$
\frac{\mathrm{d}}{\mathrm{d}\,t}\Big[\frac{1}{2}v_\parallel^2+w_\perp\Big]\;=\;v_\parallel\langle \mathbf{E}_\perp(t,\mathbf{x}),\mathbf{v}_\perp\rangle\,,
$$

then thanks to Lemma 4.3 we derive

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\,t}\left[\frac{1}{2}v_\parallel^2+w_\perp-\varepsilon v_\parallel\langle \mathbf{U}_{\mathbf{E}\times\mathbf{B}}(t,\mathbf{x}),\mathbf{v}_\perp\rangle\right]\\
&=\varepsilon\,v_\parallel\left(\langle\partial_t\mathbf{E}_\perp(t,\mathbf{x})+\mathrm{d}_{\mathbf{x}}\,\mathbf{E}_\perp(t,\mathbf{x})\mathbf{v},\mathbf{v}_\perp\rangle+\langle \mathbf{E}_\perp(t,\mathbf{x}),\mathbf{U}(t,\mathbf{x},\mathbf{v})\rangle\right)\,.
\end{aligned}
$$

Therefore as long as $r(\mathbf{x})\geq r_0$

$$
\max_{s\in[0,t]}\|\mathbf{v}(s)\|^2\leq\|\mathbf{v}_0\|^2(1+C\,\varepsilon)+C\,\varepsilon\,\max_{s\in[0,t]}\|\mathbf{v}(s)\|^2+C\,\varepsilon\,t\,\max_{s\in[0,t]}\|\mathbf{v}(s)\|\,(1+\max_{s\in[0,t]}\|\mathbf{v}(s)\|^2),
$$

for some C depending on $1/r_0$ and $\|(E_r,E_z,1/b)\|_{W^{1,\infty}([r_0,\infty[\times\mathbf{R})}$. Thus as long as $r(\mathbf{x})\geq r_0$ provided

$$
0<\varepsilon\leq\varepsilon_0\qquad\text{and}\qquad 0\leq t\leq\frac{\tau_0}{1+\|\mathbf{v}_0\|}\,\frac{1}{\varepsilon}\,,
$$

we derive

$$
\|\mathbf{v}(t)\|\leq C_0\,(\|\mathbf{v}_0\|+\varepsilon\,t),
$$

with ε_0 and τ_0 sufficiently small and C_0 sufficiently large depending on $\|(E_r,E_z,1/b)\|_{W^{1,\infty}([r_0,\infty[\times\mathbf{R})}$ and $1/r_0$. Then using this bound in an integrated version of the first equation of System (5.1) achieves the proof provided we strengthen the constraint on times to

$$
0\leq t\leq\frac{\tau_0}{1+\|\mathbf{v}_0\|^2}\,\frac{1}{\varepsilon}\,,
$$

with τ_0 sufficiently small. □

From the foregoing Proposition and System (5.1) we deduce the following proposition.

Proposition 5.2. *Under the assumption of Theorem 2.8, there exist positive constants ε_0, τ_0 and C_0, $(1/\varepsilon_0, 1/\tau_0, C_0)$ depending polynomially on $1/r_0$, $1/(r_1-r_0)$ and $\|(E_r, E_z, 1/b)\|_{W^{2,\infty}([r_0,\infty[\times\mathbf{R})}$, such that the following holds with*

$$\varepsilon_{max}(\|\mathbf{v}_0\|) := \frac{\varepsilon_0}{1+\|\mathbf{v}_0\|} \qquad \text{and} \qquad T_{max}(\|\mathbf{v}_0\|) := \frac{\tau_0}{1+\|\mathbf{v}_0\|^2}\,.$$

Let $(\mathbf{x}^\varepsilon, \mathbf{v}^\varepsilon)$ be a solution to (4.1) starting from $(\mathbf{x}_0, \mathbf{v}_0)$ satisfying $r(\mathbf{x}_0) \geq r_1$. Then provided that

$$0 < \varepsilon \leq \varepsilon_{max}(\|\mathbf{v}_0\|)\,,$$

$\mathbf{Z}^\varepsilon_{\rm gc} = (r(\mathbf{x}^\varepsilon), z(\mathbf{x}^\varepsilon), v_\parallel(\mathbf{x}^\varepsilon, \mathbf{v}^\varepsilon), w_\perp(\mathbf{v}^\varepsilon))$ satisfies for a.e. $0 \leq t \leq T_{max}(\|\mathbf{v}_0\|)/\varepsilon$

$$\left\|\mathbf{Z}^\varepsilon_{\rm gc}(t) - \mathbf{Z}^\varepsilon(t)\right\| \leq C\,\varepsilon\, e^{C\,\varepsilon\, t\,\|\mathbf{v}_0\|^4}\,(1+\|\mathbf{v}_0\|^3),$$

where $\mathbf{Z}^\varepsilon = (r^\varepsilon, z^\varepsilon, v^\varepsilon, w^\varepsilon)$ solves

$$\begin{cases} \dfrac{\mathrm{d}\,r^\varepsilon}{\mathrm{d}\,t} = \varepsilon\left(-\dfrac{E_z(t,r^\varepsilon,z^\varepsilon)}{b(r^\varepsilon,z^\varepsilon)} - w^\varepsilon\,\partial_z\left(\dfrac{1}{b}\right)(r^\varepsilon,z^\varepsilon)\right)\,, \\ \dfrac{\mathrm{d}\,z^\varepsilon}{\mathrm{d}\,t} = \varepsilon\left(\dfrac{E_r(t,r^\varepsilon,z^\varepsilon)}{b(r^\varepsilon,z^\varepsilon)} + \dfrac{(v^\varepsilon)^2}{r\,b(r^\varepsilon,z^\varepsilon)} + w^\varepsilon\,\partial_r\left(\dfrac{1}{b}\right)(r^\varepsilon,z^\varepsilon)\right)\,, \\ \dfrac{\mathrm{d}\,v^\varepsilon}{\mathrm{d}\,t} = \varepsilon\,\dfrac{v^\varepsilon}{r^\varepsilon}\left(\dfrac{E_z(t,r^\varepsilon,z^\varepsilon)}{b(r^\varepsilon,z^\varepsilon)} + w^\varepsilon\partial_z\left(\dfrac{1}{b}\right)(r^\varepsilon,z^\varepsilon)\right)\,, \\ \dfrac{\mathrm{d}\,w^\varepsilon}{\mathrm{d}\,t} = \varepsilon\,w^\varepsilon\left(\partial_r\left(\dfrac{E_z}{b}\right)(t,r^\varepsilon,z^\varepsilon) - \partial_z\left(\dfrac{E_r}{b}\right)(t,r^\varepsilon,z^\varepsilon) - \dfrac{(v^\varepsilon)^2}{r^\varepsilon}\partial_z\left(\dfrac{1}{b}\right)(r^\varepsilon,z^\varepsilon)\right)\,, \end{cases}$$

with $\mathbf{Z}^\varepsilon(0) = \mathbf{Z}^\varepsilon_{\rm gc}(0)$.

Finally from Proposition 5.2, we derive Theorem 2.8 through Proposition 2.1.

Remark 5.3. *As in Remark 3.14, we stress that the proof also yields the analysis of dynamics involving fields depending on ε but satisfying bounds uniform with respect to ε. In particular the result may be extended without change to the case where $\mathbf{E}^\varepsilon(t,\mathbf{x}) = \mathbf{E}(\varepsilon\,t,\mathbf{x})$, $0 < \varepsilon \lesssim 1$. In this somehow simpler case the asymptotic dynamics is essentially independent of ε at leading order since*

$$\mathbf{Z}^\varepsilon(t) = \mathbf{Z}(\varepsilon\,t),$$

with $\mathbf{Z}$ independent of ε.

Remark 5.4. *Though we have chosen not to delve into this here as it would have lead us too far beyond our scope, one may remove the assumption that $\mathbf{E}$ is axi-symmetric and still obtain a similar result provided one stays away from $v_\parallel = 0$. It would follow from an analysis similar to the one expounded here but using instead of (4.4) an equation encoding rotation of $\mathbf{x}$ around $\mathbf{e}_z$ at speed $v_\parallel$.*

Appendix A. Comparison with the classical adiabatic invariant formulation

In the present Section, for the sake of comparison with part of the physical literature, we derive counterparts to our three-dimensional results expressed in terms of slow variables

$$\left(\mathbf{x}, v_\parallel(t,\mathbf{x},\mathbf{v}), \frac{w_\perp(t,\mathbf{x},\mathbf{v})}{B(t,\mathbf{x})}\right)$$

and corrections thereof. To do so we explicitly introduce the function $\mu_\perp$ defined by

$$\mu_\perp(t,\mathbf{x},\mathbf{v}) = \frac{w_\perp(t,\mathbf{x},\mathbf{v})}{B(t,\mathbf{x})} = \frac{\|\mathbf{v}_\perp(t,\mathbf{x},\mathbf{v})\|^2}{2\,B(t,\mathbf{x})}\,.$$

A.1. **Long-time asymptotics in the toroidal axi-symmetric case.** Since this is slightly less computationally demanding, we provide first a counterpart to Theorem 2.8. Our starting point is System (5.1). From it we derive

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\,t}\Big[\mu_\perp - \varepsilon\,\Big(\mu_\perp\,\Big\langle \partial_r\left(\frac{1}{b}\right)\mathbf{e}_z - \partial_z\left(\frac{1}{b}\right)\mathbf{e}_r, \mathbf{v}_\perp\Big\rangle + \frac{1}{b}\,\langle \mathbf{U}_{\mathbf{E}\times\mathbf{B}}, \mathbf{v}_\perp\rangle + \frac{v_\parallel^2}{r\,b^2}\,\langle \mathbf{e}_z, \mathbf{v}_\perp\rangle\Big) + \varepsilon^2\chi_\mu\Big] \\
&= \varepsilon\,\frac{\mu_\perp}{b}\,(\partial_r E_z - \partial_z E_r) + \varepsilon^2\,\eta_\mu \\
&- \varepsilon\,\Big(\langle \mathbf{U}_{\mathbf{E}\times\mathbf{B}}, \mathbf{v}_\perp\rangle + \frac{v_\parallel^2}{r\,b}\,\langle \mathbf{e}_z, \mathbf{v}_\perp\rangle\Big)\,\Big\langle \partial_r\left(\frac{1}{b}\right)\mathbf{e}_r + \partial_z\left(\frac{1}{b}\right)\mathbf{e}_z, \mathbf{v}_\perp\Big\rangle \\
&- \varepsilon\mu_\perp\,\langle \mathbf{e}_z, \mathbf{v}_\perp\rangle\,\Big(\partial_r^2\left(\frac{1}{b}\right)\,\langle \mathbf{e}_r, \mathbf{v}_\perp\rangle + \partial_{r\,z}^2\left(\frac{1}{b}\right)\,\langle \mathbf{e}_z, \mathbf{v}_\perp\rangle\Big) \\
&+ \varepsilon\mu_\perp\,\langle \mathbf{e}_r, \mathbf{v}_\perp\rangle\,\Big(\partial_{r\,z}^2\left(\frac{1}{b}\right)\,\langle \mathbf{e}_r, \mathbf{v}_\perp\rangle + \partial_z^2\left(\frac{1}{b}\right)\,\langle \mathbf{e}_z, \mathbf{v}_\perp\rangle\Big) \\
&+ \frac{\varepsilon}{b}\,\Big\langle \partial_z\left(\frac{1}{b}\right)\mathbf{e}_r - \partial_r\left(\frac{1}{b}\right)\mathbf{e}_z, \mathbf{v}_\perp\Big\rangle\,\langle \mathbf{E}_\perp(t,\mathbf{x}) \,-\, v_\parallel\,\mathrm{d}_{\mathbf{x}}\,\mathbf{e}_\parallel(t,\mathbf{x})\,\mathbf{v}, \mathbf{v}_\perp\rangle
\end{aligned}
$$

with

$$
\chi_\mu \,=\, \Big\langle \partial_r\left(\frac{1}{b}\right)\mathbf{e}_r + \partial_z\left(\frac{1}{b}\right)\mathbf{e}_z, \boldsymbol{\chi}_x\Big\rangle + \frac{\chi_\perp}{b}
$$

and

$$
\begin{aligned}
\eta_\mu = \,& w_\perp\,\Big(\Big\langle \partial_r\left(\frac{1}{b}\right)\mathbf{e}_r + \partial_z\left(\frac{1}{b}\right)\mathbf{e}_z, \boldsymbol{\eta}_x\Big\rangle + \frac{v_\parallel}{r}\partial_r\left(\frac{1}{b}\right)\langle \mathbf{e}_\parallel, \boldsymbol{\chi}_x\rangle\Big) + \frac{\eta_\perp}{b} \\
&+ \Big\langle \partial_r\left(\frac{1}{b}\right)\mathbf{e}_r + \partial_z\left(\frac{1}{b}\right)\mathbf{e}_z, \boldsymbol{\chi}_x\Big\rangle\,\langle \mathbf{E}_\perp(t,\mathbf{x}) \,-\, v_\parallel\,\mathrm{d}_{\mathbf{x}}\,\mathbf{e}_\parallel(t,\mathbf{x})\,\mathbf{v}, \mathbf{v}_\perp\rangle \\
&+ w_\perp\,\langle \mathbf{e}_r, \boldsymbol{\chi}_x\rangle\,\Big(\partial_r^2\left(\frac{1}{b}\right)\,\langle \mathbf{e}_r, \mathbf{v}_\perp\rangle + \partial_{r\,z}^2\left(\frac{1}{b}\right)\,\langle \mathbf{e}_z, \mathbf{v}_\perp\rangle\Big) \\
&+ w_\perp\,\langle \mathbf{e}_z, \boldsymbol{\chi}_x\rangle\,\Big(\partial_{r\,z}^2\left(\frac{1}{b}\right)\,\langle \mathbf{e}_r, \mathbf{v}_\perp\rangle + \partial_z^2\left(\frac{1}{b}\right)\,\langle \mathbf{e}_z, \mathbf{v}_\perp\rangle\Big) \\
&+ \chi_\perp\,\Big\langle \partial_r\left(\frac{1}{b}\right)\mathbf{e}_r + \partial_z\left(\frac{1}{b}\right)\mathbf{e}_z, \mathbf{v}_\perp\Big\rangle\,.
\end{aligned}
$$

Now we note that the extra $\mathcal{O}(\varepsilon)$-terms in the right-hand side of the foregoing system are either third-order with respect to $\mathbf{v}_\perp$, or second-order but trace-free in the plan orthogonal to $\mathbf{e}_\parallel$ (when suitably paired). As a result they may be eliminated, leaving

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\,t}\Big[\mu_\perp - \varepsilon\,\Big(\mu_\perp\,\Big\langle \partial_r\left(\frac{1}{b}\right)\mathbf{e}_z - \partial_z\left(\frac{1}{b}\right)\mathbf{e}_r, \mathbf{v}_\perp\Big\rangle + \frac{1}{b}\,\langle \mathbf{U}_{\mathbf{E}\times\mathbf{B}}, \mathbf{v}_\perp\rangle + \frac{v_\parallel^2}{r\,b^2}\,\langle \mathbf{e}_z, \mathbf{v}_\perp\rangle\Big) + \varepsilon^2\widehat{\chi}_\mu\Big] \\
&= \varepsilon\,\frac{\mu_\perp}{b}\,(\partial_r E_z - \partial_z E_r) + \varepsilon^2\,\widehat{\eta}_\mu
\end{aligned}
$$

with

$$
|\widehat{\chi}_\mu| \lesssim \|\mathbf{v}_\perp\|\,(1 + \|\mathbf{v}\|^3)\,, \qquad |\widehat{\eta}_\mu| \lesssim 1 + \|\mathbf{v}\|^5\,. \tag{A.1}
$$

Therefore, the involved asymptotic vector field is now $\varepsilon \mathcal{Z}_1$ with $\mathcal{Z}_1$ defined as

$$\mathcal{Z}_1(t,\mathbf{Z}) \;=\; \begin{pmatrix} -\dfrac{E_z(t,r,z)}{b(r,z)} + \dfrac{\mu}{b(r,z)}\,\partial_z b(r,z) \\ \dfrac{E_r(t,r,z)}{b(r,z)} + \dfrac{v^2}{r\,b(r,z)} - \dfrac{\mu}{b(r,z)}\,\partial_r b(r,z) \\ \dfrac{v}{r}\left(\dfrac{E_z(t,r,z)}{b(r,z)} - \dfrac{\mu}{b(r,z)}\,\partial_z b(r,z)\right) \\ \dfrac{\mu}{b(r,z)}\left(\partial_r E_z - \partial_z E_r\right)(t,r,z) \end{pmatrix},$$

where the slow variable is now $\mathbf{Z} = (r,z,v,\mu)$. Note that $r\,b\,\mathcal{Z}_1$ is divergence-free.

Theorem A.1. *Let $\mathbf{B}$ be a stationary, axi-symmetric and toroidal magnetic field and $\mathbf{E}$ be an axi-symmetric electric field orthogonal to $\mathbf{B}$, with $(E_r, E_z, 1/b) \in W^{2,\infty}$ in the region where $r(\mathbf{x}) \geq r_0$ for some r_0. For any $r_1 > r_0$, there exist positive constants ε_0, τ_0 and C_0, $(1/\varepsilon_0, 1/\tau_0, C_0)$ depending polynomially on $1/r_0$, $1/(r_1 - r_0)$ and $\|(E_r, E_z, 1/b)\|_{W^{2,\infty}([r_0,\infty[\times\mathbf{R})}$, such that the following holds with*

$$\varepsilon_{max}(R_0) := \frac{\varepsilon_0}{1+R_0} \qquad \text{and} \qquad T_{max}(R_0) := \frac{\tau_0}{1+R_0^2}\,.$$

Consider f^ε a solution to (2.1) with initial datum a positive density f_0 supported where

$$r(\mathbf{x}) \geq r_1 \qquad \text{and} \qquad \|\mathbf{v}\| \leq R_0$$

for some $R_0 > 0$ and define F^ε as

$$F^\varepsilon(t,r,z,v,\mu) = \int_0^{2\pi}\!\!\int_0^{2\pi} f^\varepsilon(t, r\,\mathbf{e}_r^\theta + z\,\mathbf{e}_z, v\,\mathbf{e}_\parallel(r\,\mathbf{e}_r^\theta + z\,\mathbf{e}_z) + \sqrt{2\,\mu\,b(r,z)}\;\mathbf{e}_\perp^{\theta,\varphi})\; r\,b(r,z)\;\mathrm{d}\varphi\;\mathrm{d}\theta$$

with

$$\mathbf{e}_r^\theta \;=\; \cos(\theta)\,\mathbf{e}_x + \sin(\theta)\,\mathbf{e}_y\,, \qquad\qquad \mathbf{e}_\perp^{\theta,\varphi} \;=\; \cos(\varphi)\,\mathbf{e}_r^\theta + \sin(\varphi)\,\mathbf{e}_z\,,$$

where $(\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z)$ is a fixed orthonormal basis. Then provided that

$$0 < \varepsilon \leq \varepsilon_{max}(R_0)\,,$$

we have for a.e. $0 \leq t \leq T_{max}(R_0)/\varepsilon$

$$\|F^\varepsilon(t,\cdot) - G^\varepsilon(t,\cdot)\|_{\dot{W}^{-1,1}} \leq C_0\,\varepsilon \int_{\mathbf{R}^3\times\mathbf{R}^3} e^{C\,\varepsilon\,t\,\|\mathbf{v}\|^4}\,(1+\|\mathbf{v}\|^3)\,f_0(\mathbf{x},\mathbf{v})\;\mathrm{d}\mathbf{x}\;\mathrm{d}\mathbf{v},$$

where G^ε solves

$$\partial_t G^\varepsilon \;+\; \varepsilon\,\mathrm{div}_{\mathbf{Z}}\,(G^\varepsilon\,\mathcal{Z}_1) \;=\; 0, \tag{A.2}$$

with initial datum G_0 given by

$$G_0(\mathbf{Z}) \;=\; \int_0^{2\pi}\!\!\int_0^{2\pi} f_0(\,r\,\mathbf{e}_r^\theta + z\,\mathbf{e}_z, v\,\mathbf{e}_\parallel(r\,\mathbf{e}_r^\theta + z\,\mathbf{e}_z) + \sqrt{2\,\mu\,b(r,z)}\;\mathbf{e}_\perp^{\theta,\varphi})\;r\,b(r,z)\;\mathrm{d}\varphi\;\mathrm{d}\theta\,.$$

Incidentally we observe that of course it is easier to derive the balance law

$$\partial_t(\mu G) \;+\; \varepsilon\,\mathrm{div}_{\mathbf{Z}}\,(\mu\,G^\varepsilon\,\mathcal{Z}_1) \;=\; \frac{\mu}{b}\,G\,(\partial_r E_z - \partial_z E_r)$$

for the asymptotic equation (A.2) than the corresponding result for the original formulation. Yet the energy balance law is in turn less straightforward to derive.

At the level of description considered in this section since going from variables $(r,z,v_\parallel,w_\perp)$ to $(r,z,v_\parallel,\mu_\perp)$ is quite simple we could have deduced Theorem A.1 from Theorem 2.8. To give two hints in this direction, note that if we denote F_e^ε and F_μ^ε the averaged densities respectively from Theorem 2.8 and Theorem A.1, then $F_\mu^\varepsilon(t,r,z,v,\mu)\;\mathrm{d}r\;\mathrm{d}z\;\mathrm{d}v\;\mathrm{d}\mu$ is the push-forward of $F_e^\varepsilon(t,r,z,v,w)\;\mathrm{d}r\;\mathrm{d}z\;\mathrm{d}v\;\mathrm{d}w$ by the map $(r,z,v,w) \mapsto (r,z,v,w/b(r,z))$. Likewise, with the same convention, it may be checked that $G_\mu^\varepsilon(t,r,z,v,\mu)\;\mathrm{d}r\;\mathrm{d}z\;\mathrm{d}v\;\mathrm{d}\mu$ is obtained from $G_e^\varepsilon(t,r,z,v,w)\;\mathrm{d}r\;\mathrm{d}z\;\mathrm{d}v\;\mathrm{d}w$ in the same way, by using that

$$\mathcal{Z}_1^\mu(t,r,z,v,\mu) \;=\; \left[\frac{\mathcal{W}_1^e}{b} - \frac{w}{b^2}\,(\partial_r b\,\mathcal{W}_1^r + \partial_z b\,\mathcal{W}_1^z)\right](t,r,z,v,\mu\,b(r,z))$$

where superscripts denote components. However we have chosen not to follow this path and instead to come back to the normal form (5.1) since the strategy under discussion is more cumbersome when higher-order corrections are taken into account.

A.2. **General second-order asymptotics.** We now provide a counterpart to Theorem 2.4. Our starting point is System (4.20). From it stems

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\,t}\Big[\mu_\perp - \varepsilon\frac{\mu_\perp}{B^2}\,\mathrm{d}_{\mathbf{x}}B\;\mathbb{J}\,\mathbf{v}_\perp - \frac{\varepsilon}{B}\,\langle\mathbf{v}_\perp,\,\mathbf{U}_{\mathbf{E}\times\mathbf{B}} + v_\parallel\boldsymbol{\Sigma}\rangle - \frac{\varepsilon\,v_\parallel}{2B^2}\,\langle\mathbb{J}\,\mathbf{v}_\perp,\,\Re(\mathrm{d}_{\mathbf{x}}\,\mathbf{e}_\parallel)\,\mathbf{v}_\perp\rangle + \varepsilon^2\chi_\mu\Big]\\
&\qquad = -\frac{\mu_\perp}{B}\,(\partial_t B + v_\parallel\,\mathrm{div}_{\mathbf{x}}\,\mathbf{B}) - \varepsilon\,\frac{\mu_\perp}{B}\,\mathrm{div}_{\mathbf{x}}\,(B\,\mathbf{U}_{\mathbf{E}\times\mathbf{B}} + v_\parallel\,B\,\boldsymbol{\Sigma})\\
&\qquad - \varepsilon\,\mu_\perp\,\langle\mathbf{U}_{\mathrm{curv}}\,,\,\mathbf{E}\rangle - \varepsilon\,\mu_\perp^2\,\mathrm{d}_{\mathbf{x}}B\;\mathbf{U}_{\mathrm{curl}\,\mathbf{e}_\parallel} + \varepsilon^2\eta_\mu\\
&\qquad + \varepsilon\mu_\perp\,\mathrm{d}_{\mathbf{x}}\,\partial_t\left(\frac{1}{B}\right)\;\mathbb{J}\,\mathbf{v}_\perp + \varepsilon\mu_\perp\,\mathrm{d}_{\mathbf{x}}^2\left(\frac{1}{B}\right)(\mathbf{v},\,\mathbb{J}\,\mathbf{v}_\perp)\\
&\qquad - \frac{\varepsilon}{B^3}\,\mathrm{d}_{\mathbf{x}}B\;\mathbb{J}\mathbf{v}_\perp\;\langle\mathbf{E}_\perp(t,\mathbf{x}) - v_\parallel\;(\partial_t\,\mathbf{e}_\parallel(t,\mathbf{x}) + \mathrm{d}_{\mathbf{x}}\,\mathbf{e}_\parallel(t,\mathbf{x})\,\mathbf{v})\,,\mathbf{v}_\perp\rangle\\
&\qquad - \varepsilon\left(\partial_t\left(\frac{1}{B}\right) + \mathrm{d}_{\mathbf{x}}\left(\frac{1}{B}\right)\mathbf{v}\right)\Big[\,\langle\mathbf{v}_\perp,\,\mathbf{U}_{\mathbf{E}\times\mathbf{B}} + v_\parallel\,\boldsymbol{\Sigma}\,\rangle + \frac{v_\parallel}{2B}\,\langle\mathbb{J}\,\mathbf{v}_\perp,\,\Re(\mathrm{d}_{\mathbf{x}}\,\mathbf{e}_\parallel)\,\mathbf{v}_\perp\rangle\,\Big]
\end{aligned}$$

with

$$\chi_\mu = -\frac{\mu_\perp}{B}\,\mathrm{d}_{\mathbf{x}}B\;\boldsymbol{\chi}_{\mathbf{x}} + \frac{\chi_\perp}{B}$$

$$\begin{aligned}
\eta_\mu = &-\frac{\mu_\perp}{B}\,\mathrm{d}_{\mathbf{x}}B\;\boldsymbol{\eta}_{\mathbf{x}} + \frac{\eta_\perp}{B} + w_\perp\,\mathrm{d}_{\mathbf{x}}\left[\partial_t\left(\frac{1}{B}\right)\right]\boldsymbol{\chi}_{\mathbf{x}} + w_\perp\,\mathrm{d}_{\mathbf{x}}^2\left(\frac{1}{B}\right)(\mathbf{v},\,\boldsymbol{\chi}_{\mathbf{x}}) - \frac{\chi_\perp}{B^2}\,(\partial_t B + \mathrm{d}_{\mathbf{x}}B\;\mathbf{v})\\
&- \frac{1}{B^2}\,\mathrm{d}_{\mathbf{x}}B\;\boldsymbol{\chi}_{\mathbf{x}}\;\langle\mathbf{E}_\perp(t,\mathbf{x}) - v_\parallel\;(\partial_t\,\mathbf{e}_\parallel(t,\mathbf{x}) + \mathrm{d}_{\mathbf{x}}\,\mathbf{e}_\parallel(t,\mathbf{x})\,\mathbf{v})\,,\mathbf{v}_\perp\rangle\;.
\end{aligned}$$

The last three lines of the foregoing system may be discarded as being linear or cubic in $\mathbf{v}_\perp$, or quadratic in $\mathbf{v}_\perp$ but with zero trace in the plane orthogonal to $\mathbf{v}_\perp$. This leads to

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\,t}\Big[\mu_\perp - \varepsilon\frac{\mu_\perp}{B^2}\,\mathrm{d}_{\mathbf{x}}B\;\mathbb{J}\,\mathbf{v}_\perp - \frac{\varepsilon}{B}\,\langle\mathbf{v}_\perp,\,\mathbf{U}_{\mathbf{E}\times\mathbf{B}} + v_\parallel\boldsymbol{\Sigma}\rangle - \frac{\varepsilon\,v_\parallel}{2B^2}\,\langle\mathbb{J}\,\mathbf{v}_\perp,\,\Re(\mathrm{d}_{\mathbf{x}}\,\mathbf{e}_\parallel)\,\mathbf{v}_\perp\rangle + \varepsilon^2\widehat{\chi}_\mu\Big]\\
&\qquad = -\frac{\mu_\perp}{B}\,(\partial_t B + v_\parallel\,\mathrm{div}_{\mathbf{x}}\,\mathbf{B}) - \varepsilon\,\frac{\mu_\perp}{B}\,\mathrm{div}_{\mathbf{x}}\,(B\,\mathbf{U}_{\mathbf{E}\times\mathbf{B}} + v_\parallel\,B\,\boldsymbol{\Sigma})\\
&\qquad - \varepsilon\,\mu_\perp\,\langle\mathbf{U}_{\mathrm{curv}}\,,\,\mathbf{E}\rangle - \varepsilon\,\mu_\perp^2\,\mathrm{d}_{\mathbf{x}}B\;\mathbf{U}_{\mathrm{curl}\,\mathbf{e}_\parallel} + \varepsilon^2\widehat{\eta}_\mu
\end{aligned}$$

with

$$|\widehat{\chi}_\mu| \lesssim \|\mathbf{v}_\perp\|\,(1 + \|\mathbf{v}\|^3)\,, \qquad\qquad |\widehat{\eta}_\mu| \lesssim 1 + \|\mathbf{v}\|^5\,.$$

At this stage one may follow the final lines of the proof of Theorem 2.4, by replacing in the zeroth and first-order terms of the right-hand side $(\mathbf{x}, v_\parallel, \mu_\perp)$ with ε-corrections $(\mathbf{x}_{\mathrm{gc}}, v_{\mathrm{gc}}, \mu_{\mathrm{gc}})$ up to $\mathcal{O}(\varepsilon^2)$-terms that may be added to $\varepsilon^2\widehat{\eta}_\mu$ and ε-terms that may be removed by the by-now familiar elimination process, resulting in another harmless modification of $\widehat{\chi}_\mu$ and $\widehat{\eta}_\mu$.

To state the resulting theorem, we introduce

$$\mu_{\mathrm{gc}} = \frac{w_\perp}{B} - \varepsilon\frac{w_\perp}{B^3}\,\mathrm{d}_{\mathbf{x}}B\;\mathbb{J}\,\mathbf{v}_\perp - \frac{\varepsilon}{B}\,\langle\mathbf{v}_\perp,\,\mathbf{U}_{\mathbf{E}\times\mathbf{B}} + v_\parallel\boldsymbol{\Sigma}\rangle - \frac{\varepsilon\,v_\parallel}{2B^2}\,\langle\mathbb{J}\,\mathbf{v}_\perp,\,\Re(\mathrm{d}_{\mathbf{x}}\,\mathbf{e}_\parallel)\,\mathbf{v}_\perp\rangle \tag{A.3}$$

and

$$\mathcal{Y}^\varepsilon = \mathcal{Y}_0 + \varepsilon\mathcal{Y}_1$$

with

$$\mathcal{Y}_0(t,\mathbf{Z}) \;=\; \begin{pmatrix} v\;\mathbf{e}_\parallel(t,\mathbf{y}) \\ E_\parallel(t,\mathbf{y}) + \mu\, B(t,\mathbf{y})\;\mathrm{div}_{\mathbf{x}}\,\mathbf{e}_\parallel(t,\mathbf{y}) \\ -\dfrac{\mu}{B(t,\mathbf{y})}\,(\partial_t B + v\,\mathrm{div}_{\mathbf{x}}\,\mathbf{B})\,(t,\mathbf{y}) \end{pmatrix}$$

and

$$\mathcal{Y}_1(t,\mathbf{Z}) \;=\; \begin{pmatrix} \mathbf{U}_{\mathrm{drift}}(t,\mathbf{y},v,\mu\,B(t,\mathbf{y})) \\ \langle \mathbf{\Sigma}(t,\mathbf{y},v),\mathbf{E}(t,\mathbf{y})\rangle + \mu\,B(t,\mathbf{y})\;\mathrm{div}_{\mathbf{x}}\,\mathbf{\Sigma}(t,\mathbf{y},v) \\ \Big[-\dfrac{\mu}{B}\;\mathrm{div}_{\mathbf{x}}\,(B\,\mathbf{U}_{\mathbf{E}\times\mathbf{B}} + v\,B\,\mathbf{\Sigma}) - \;\mu\;\langle \mathbf{U}_{\mathrm{curv}}\,,\,\mathbf{E}\rangle - \;\mu^2\,\mathrm{d}_{\mathbf{x}}B\;\mathbf{U}_{\mathrm{curl}\,\mathbf{e}_\parallel}\Big]\,(t,\mathbf{y},v) \end{pmatrix}$$

defining vector fields on the reduced phase-space where $\mathbf{Z} = (\mathbf{y}, v, \mu)$ lives.

Theorem A.2. *Let $\mathbf{E} \in W^{2,\infty}$ and $\mathbf{B}$ be such that $1/B \in W^{2,\infty}$ and $\mathbf{e}_\parallel \in W^{3,\infty}$. There exists a constant C depending polynomially on $\|\mathbf{E}\|_{W^{2,\infty}}$, $\|B^{-1}\|_{W^{2,\infty}}$ and $\|\,\mathbf{e}_\parallel\,\|_{W^{3,\infty}}$ such that if f^ε solves (2.1) with initial data a positive density f_0, then F^ε defined so that $F^\varepsilon(t,\cdot)$ is the push-forward of $f^\varepsilon(t,\cdot)$ by the map $(\mathbf{x},\mathbf{v}) \mapsto (\mathbf{x}_{\mathrm{gc}}, v_{\mathrm{gc}}, \mu_{\mathrm{gc}})(t,\mathbf{x},\mathbf{v})$ satisfies for a.e. $t \geq 0$*

$$\|F^\varepsilon(t,\cdot) - G^\varepsilon(t,\cdot)\|_{\dot{W}^{-1,1}} \leq C\,\varepsilon^2\,e^{C\,t^4\,(1+\varepsilon\,t)} \int_{\mathbf{R}^6} e^{C\,t\,\|\mathbf{v}\|^3\,(1+\varepsilon\,\|\mathbf{v}\|)}\,(1+\|\mathbf{v}\|^4)\,f_0(\mathbf{x},\mathbf{v})\;\mathrm{d}\,\mathbf{x}\;\mathrm{d}\,\mathbf{v},$$

where G^ε solves

$$\partial_t G^\varepsilon \;+\; \mathrm{div}_{\mathbf{Z}}\,(\mathcal{Y}^\varepsilon\,G^\varepsilon) \;=\; 0, \tag{A.4}$$

with initial data $G_0^\varepsilon = F^\varepsilon(0,\cdot)$.

References

[Bellan, 2008] Bellan, P. M. (2008). *Fundamentals of plasma physics*. Cambridge University Press.

[Benettin and Sempio, 1994] Benettin, G. and Sempio, P. (1994). Adiabatic invariants and trapping of a point charge in a strong nonuniform magnetic field. *Nonlinearity*, 7(1):281–303.

[Bogoliubov and Mitropolsky, 1961] Bogoliubov, N. N. and Mitropolsky, Y. A. (1961). *Asymptotic methods in the theory of non-linear oscillations*. Translated from the second revised Russian edition. International Monographs on Advanced Mathematics and Physics. Hindustan Publishing Corp., Delhi, Gordon and Breach Science Publishers, New York.

[Bostan, 2010a] Bostan, M. (2010a). Gyrokinetic Vlasov equation in three dimensional setting. Second order approximation. *Multiscale Model. Simul.*, 8(5):1923–1957.

[Bostan, 2010b] Bostan, M. (2010b). Transport equations with disparate advection fields. Application to the gyrokinetic models in plasma physics. *J. Differential Equations*, 249(7):1620–1663.

[Bostan, 2018] Bostan, M. (2018). Asymptotic behavior for the Vlasov-Poisson equations with strong external magnetic field. straight magnetic field lines. *HAL preprint*, hal-01683869.

[Brizard and Hahm, 2007] Brizard, A. J. and Hahm, T. S. (2007). Foundations of nonlinear gyrokinetic theory. *Rev. Modern Phys.*, 79(2):421–468.

[Chen, 2016] Chen, F. (2016). *Introduction to Plasma Physics and Controlled Fusion*. Springer, third edition.

[Cheverry, 2017] Cheverry, C. (2017). Anomalous transport. *J. Differential Equations*, 262(3):2987–3033.

[Degond and Filbet, 2016] Degond, P. and Filbet, F. (2016). On the Asymptotic Limit of the Three Dimensional Vlasov–Poisson System for Large Magnetic Field: Formal Derivation. *J. Stat. Phys.*, 165(4):765–784.

[Filbet and Rodrigues, 2016] Filbet, F. and Rodrigues, L. M. (2016). Asymptotically stable particle-in-cell methods for the Vlasov-Poisson system with a strong external magnetic field. *SIAM J. Numer. Anal.*, 54(2):1120–1146.

[Filbet and Rodrigues, 2017] Filbet, F. and Rodrigues, L. M. (2017). Asymptotically preserving particle-in-cell methods for inhomogeneous strongly magnetized plasmas. *SIAM J. Numer. Anal.*, 55(5):2416–2443.

[Freidberg, 2008] Freidberg, J. (2008). *Plasma Physics and Fusion Energy*. Cambridge University Press.

[Frénod and Lutz, 2014] Frénod, E. and Lutz, M. (2014). On the geometrical gyro-kinetic theory. *Kinet. Relat. Models*, 7(4):621–659.

[Frénod and Sonnendrücker, 1998] Frénod, E. and Sonnendrücker, É. (1998). Homogenization of the Vlasov equation and of the Vlasov-Poisson system with a strong external magnetic field. *Asymptot. Anal.*, 18(3-4):193–213.

[Frénod and Sonnendrücker, 2000] Frénod, E. and Sonnendrücker, É. (2000). Long time behavior of the two-dimensional Vlasov equation with a strong external magnetic field. *Math. Models Methods Appl. Sci.*, 10(4):539–553.

[Garbet et al., 2010] Garbet, X., Idomura, Y., Villard, L., and Watanabe, T. H. (2010). Gyrokinetic simulations of turbulent transport. *Nuclear Fusion*, 50:043002.

[Golse and Saint-Raymond, 1999] Golse, F. and Saint-Raymond, L. (1999). The Vlasov-Poisson system with strong magnetic field. *J. Math. Pures Appl. (9)*, 78(8):791–817.

[Han-Kwan, 2011] Han-Kwan, D. (2011). *Contribution à l'étude mathématique des plasmas fortement magnétisés*. PhD thesis, Université Pierre et Marie Curie-Paris VI.

[Hazeltine and Meiss, 2003] Hazeltine, R. and Meiss, J. (2003). *Plasma Confinement*. Dover Publications.

[Herda, 2016] Herda, M. (2016). On massless electron limit for a multispecies kinetic system with external magnetic field. *J. Differential Equations*, 260(11):7861–7891.

[Herda, 2017] Herda, M. (2017). *Analyse asymptotique et numérique de quelques modèles pour le transport de particules chargées*. PhD thesis, Université Claude Bernard Lyon 1.

[Herda and Rodrigues, 2016] Herda, M. and Rodrigues, L. M. (2016). Anisotropic Boltzmann-Gibbs dynamics of strongly magnetized Vlasov-Fokker-Planck equations. *arXiv preprint arXiv:1610.05138*.

[Krommes, 2012] Krommes, J. A. (2012). The gyrokinetic description of microturbulence in magnetized plasmas. In *Annual review of fluid mechanics. Volume 44, 2012*, volume 44 of *Annu. Rev. Fluid Mech.*, pages 175–201. Annual Reviews, Palo Alto, CA.

[Lee, 1983] Lee, W. (1983). Gyrokinetic approach in particle simulation. *Phys. Fluids*, 26(2):556–562.

[Littlejohn, 1979] Littlejohn, R. G. (1979). A guiding center hamiltonian : A new approach. *J. Math. Phys.*, 20:2445–2458.

[Littlejohn, 1981] Littlejohn, R. G. (1981). Hamiltonian formulation of guiding center motion. *Phys. Fluids*, 24:1730–1749.

[Littlejohn, 1983] Littlejohn, R. G. (1983). Variational principles of guiding center motion. *J. Plasma Physics*, 29:111–124.

[Lutz, 2013] Lutz, M. (2013). *Étude mathématique et numérique d'un modèle gyrocinétique incluant des effets électromagnétiques pour la simulation d'un plasma de Tokamak*. PhD thesis, Université de Strasbourg.

[Matteo, 2017] Matteo, V. F. (2017). *Gyrokinetic theory for particle transport in fusion plasmas*. PhD thesis, Università di Roma Tre.

[Miot, 2016] Miot, É. (2016). On the gyrokinetic limit for the two-dimensional Vlasov-Poisson system. *arXiv preprint*, arXiv-1603.04502.

[Miyamoto, 2006] Miyamoto, K. (2006). *Plasma physics and controlled nuclear fusion*, volume 38 of *Springer Series on Atomic, Optical, and Plasma Physics*. Springer-Verlag Berlin-Heidelberg.

[Piel, 2010] Piel, A. (2010). *Plasma Physics: An Introduction to Laboratory, Space, and Fusion Plasmas*. Springer Berlin Heidelberg.

[Possanner, 2018] Possanner, S. (2018). Gyrokinetics from variational averaging: Existence and error bounds. *J. Math. Phys.*, 59(8):082702, 34.

[Saint-Raymond, 2002] Saint-Raymond, L. (2002). Control of large velocities in the two-dimensional gyrokinetic approximation. *J. Math. Pures Appl. (9)*, 81(4):379–399.

[Sanders et al., 2007] Sanders, J. A., Verhulst, F., and Murdock, J. (2007). *Averaging methods in nonlinear dynamical systems*, volume 59 of *Applied Mathematical Sciences*. Springer, New York, second edition.

[Scott, 2017] Scott, B. D. (2017). Gyrokinetic Field Theory as a Gauge Transform or: gyrokinetic theory without Lie transforms. *arXiv preprint*, arXiv-1708.06265.

Université de Toulouse III & IUF, UMR5219, Institut de Mathématiques de Toulouse, 118, route de Narbonne; F-31062 Toulouse Cedex, FRANCE
Email address: `francis.filbet@math.univ-toulouse.fr`

Univ Rennes & IUF, CNRS, IRMAR - UMR 6625, F-35000 Rennes, FRANCE
Email address: `luis-miguel.rodrigues@univ-rennes1.fr`