

CONVERGENCE RATE ESTIMATES FOR THE LOW MACH AND ALFVÉN NUMBER THREE-SCALE SINGULAR LIMIT OF COMPRESSIBLE IDEAL MAGNETOHYDRODYNAMICS *

BIN CHENG¹, QIANGCHANG JU² AND STEVE SCHOCHET³

Abstract. Convergence rate estimates are obtained for singular limits of the compressible ideal magnetohydrodynamics equations, in which the Mach and Alfvén numbers tend to zero at different rates. The proofs use a detailed analysis of exact and approximate fast, intermediate, and slow modes together with improved estimates for the solutions and their time derivatives, and the time-integration method. When the small parameters are related by a power law the convergence rates are positive powers of the Mach number, with the power varying depending on the component and the norm. Exceptionally, the convergence rate for two components involve the ratio of the two parameters, and that rate is proven to be sharp via corrector terms. Moreover, the convergence rates for the case of a power-law relation between the small parameters tend to the two-scale convergence rate as the power tends to one. These results demonstrate that the issue of convergence rates for three-scale singular limits, which was not addressed in the authors' previous paper, is much more complicated than for the classical two-scale singular limits.

Résumé. Des estimations du taux de convergence sont obtenues pour des limites singulières des équations magnétohydrodynamiques idéales compressibles, quand les nombres de Mach et Alfvén tendent à zéro à des taux différents. Les démonstrations utilisent une analyse détaillée des modes rapides, intermédiaires, et lents exacts et approximatifs ainsi que des estimations améliorées pour les solutions et leurs dérivées temporelles, et la méthode d'intégration temporelle. Lorsque les petits paramètres sont liés par une loi de puissance, les taux de convergence sont des puissances positives du nombre de Mach, la puissance variant selon le composant et la norme. Exceptionnellement, le taux de convergence pour deux composantes implique le rapport des deux paramètres. Ces résultats démontrent que la question des taux de convergence pour les limites singulières à trois échelles est beaucoup plus compliquée que pour les limites singulières classiques à deux échelles.

1991 Mathematics Subject Classification. 35B25, 35L45, 35Q35, 76W05.

...

Keywords and phrases: rate of convergence, singular limit, magnetohydrodynamics

* Ju is supported by the NSFC (Grants No.11571046, 11471028, 11671225). Cheng and Ju are supported by the UK Royal Society "International Exchanges" scheme (Award No. IE150886). Ju and Schochet are supported by the ISF-NSFC joint research program (NSFC Grant No. 11761141008 and ISF Grant No. 2519/17). Cheng is supported by the Leverhulme Trust (Award No. RPG-2017-098) and the EPSRC (Grant No. EP/R029628/1).

¹ Department of Mathematics, University of Surrey, Guildford, GU2 7XH, United Kingdom

² Institute of Applied Physics and Computational Mathematics, P.O. Box 8009, Beijing 100088, China

³ School of Mathematical Sciences, Tel-Aviv University, Tel Aviv 69978, Israel

1. INTRODUCTION

A uniform existence theorem and a convergence theorem as the small parameters tend to zero were recently developed [CJS18] for singular limits of symmetric hyperbolic systems of the form

$$A_0(\varepsilon_M \mathbf{V}) \mathbf{V}_t + \sum_{i=1}^d A_i(\mathbf{V}) \mathbf{V}_{x_i} = \frac{1}{\varepsilon_A} \mathcal{L}_A \mathbf{V} + \frac{1}{\varepsilon_M} \mathcal{L}_M \mathbf{V}, \quad (1)$$

where ε_A and ε_M are small positive parameters and \mathcal{L}_A and \mathcal{L}_M are skew-adjoint constant-coefficient first-order differential operators. If $\frac{\varepsilon_A}{\varepsilon_M}$ tends to zero as the parameters tend to zero then systems of the form (1) have three time scales: $O(\frac{1}{\varepsilon_A})$, $O(\frac{1}{\varepsilon_M})$, and $O(1)$.

In this paper we begin the study of the rate of convergence of solutions of three-scale singular limits to corresponding solutions of their limit equations, an issue that was not considered in [CJS18], but which is significant for applications because it determines the accuracy of using the limiting dynamics to approximate the original system. The convergence rate in the general case will undoubtedly be very complicated, since in general many different limit systems are obtained for different power-law relations between the two small parameters as they both tend to zero [CJS18, §4]. As a first step, we study here the particular case of the low Mach and Alfvén number limit of the compressible ideal magnetohydrodynamics (MHD) equations in the presence of a large uniform magnetic field, which was the motivating example for our work. As we will show, that system has essentially only one limit system, although the limit

$$\mu_{\text{lim}} := \lim_{\varepsilon_A, \varepsilon_M \rightarrow 0} \mu \quad (2)$$

of the ratio

$$\mu := \frac{\varepsilon_A}{\varepsilon_M} \quad (3)$$

appears in that limit system as a parameter. Even for the MHD system the study of the convergence rate is much more intricate than for two-scale singular limits because, as described below, the bounds on the first time derivative satisfied by solutions to three-scale systems are weaker than those satisfied for two-scale systems, and the eigenspace projections of the large operator $\frac{1}{\varepsilon_A} \mathcal{L}_A + \frac{1}{\varepsilon_M} \mathcal{L}_M$ depend on the ratio of the small parameters instead of being fixed as in the two-scale case.

The asymptotic analysis of singular limits is of fundamental importance in the theory of fluid dynamics, since some continuum fluid models are limits of the Boltzmann equation, which in turn is the limit of particle dynamics [Gal19], and many continuum models are limits of other such models as one or more dimensionless parameters, such as the Mach number, Froude number, Rossby number, and Alfvén number, tend to zero. The modern theory of two-scale singular limits of continuum fluid models and PDEs in general, in which all small parameters tend to zero at the same rate, was initiated in [KM81, KM81], where smooth solutions of initial-value problems for the compressible isentropic Euler or Navier-Stokes equations were shown to converge to solutions of the corresponding incompressible equations, and a rate of convergence was demonstrated. Restrictions on the initial data in that pioneering work were later eliminated [Uka86, Asa87, Sch94, Gre97a]. Analogous results have subsequently been shown for a variety of initial and boundary conditions and for a wide variety of two-scale singular limits arising in fluid dynamics and related equations [Gre97b, CG00, DH04, Sec06, Ala08, Gal08, CM13, DHWZ13, LL18], including various forms of the MHD equations [Got90, JLL10, JLL12, JX18]. Some of these papers include a convergence rate of the type that will be discussed below. Like the results presented in this paper, the results discussed so far concern smooth solutions, but there is a parallel theory of singular limits of weak solutions [LM98, DGLM99, FN09, CGS16].

The MHD system in three spatial dimensions that we study in this paper, derived in §A from a standard formulation of that system, is

$$a(\varepsilon_M r) (\partial_t r + (\mathbf{u} \cdot \nabla) r) + \frac{a(\varepsilon_M r) \rho(\varepsilon_M r)}{\varepsilon_M} \nabla \cdot \mathbf{u} = 0 \quad (4a)$$

$$\rho(\varepsilon_M r) (\partial_t \mathbf{u} + (\mathbf{u} \cdot \nabla) \mathbf{u}) + \frac{a(\varepsilon_M r) \rho(\varepsilon_M r)}{\varepsilon_M} \nabla r + \nabla \frac{|\mathbf{b}|^2}{2} - (\mathbf{b} \cdot \nabla) \mathbf{b} = \frac{\partial_z \mathbf{b} - \nabla b_3}{\varepsilon_A}, \quad (4b)$$

$$\partial_t \mathbf{b} + (\mathbf{u} \cdot \nabla) \mathbf{b} + (\nabla \cdot \mathbf{u}) \mathbf{b} - (\mathbf{b} \cdot \nabla) \mathbf{u} = \frac{\partial_z \mathbf{u} - \mathbf{e}_z \nabla \cdot \mathbf{u}}{\varepsilon_A}, \quad (4c)$$

$$\nabla \cdot \mathbf{b} = 0, \quad (4d)$$

where

$$a(s) := \frac{p'(1+s)}{1+s}, \quad \rho(s) := 1+s. \quad (5)$$

The divergence-free condition (4d) on the magnetic field is preserved by the dynamics of (4c), and so is just a restriction on the initial data. Hence straightforward calculations show that the system (4) has the form (1), with $\mathbf{V} = (r, \mathbf{u}, \mathbf{b})$.

Our main result is a rate of convergence of solutions of the MHD system (4) as the small parameters ε_A and ε_M tend to zero. As a preliminary, we prove a uniform existence and a convergence result, including determining the limit system. Before stating these results we discuss notations, operators, initial data, and parameters that will be used in the statement of the theorem. First, we let $\|\cdot\|_k$ denote the H^k norm. Next, for any vector \mathbf{w} let its “horizontal” part w_h denote its x, y components $(\begin{smallmatrix} w_1 \\ w_2 \end{smallmatrix})$, and let $\mathcal{P}_h^{\text{div}}$ denote the *two-dimensional* Leray-Helmholtz projection onto divergence-free velocity fields in the x, y plane or 2-torus, i.e.

$$\mathcal{P}_h^{\text{div}} w_h := w_h - \nabla_h \Delta_h^{-1} \nabla_h \cdot w_h, \quad \text{where } \nabla_h := \begin{pmatrix} \partial_x \\ \partial_y \end{pmatrix}, \quad \Delta_h := \partial_x^2 + \partial_y^2. \quad (6)$$

The large terms in (4) form the “Alfvén” and “Mach” operators

$$\mathcal{L}_A \mathbf{V} := \begin{pmatrix} 0 \\ (-\nabla_h b_3 + \partial_z b_h) \\ 0 \\ (\partial_z u_h) \\ -\nabla_h \cdot u_h \end{pmatrix}, \quad \mathcal{L}_M \mathbf{V} := \begin{pmatrix} -\nabla \cdot \mathbf{u} \\ (-\nabla_h r) \\ -\partial_z r \\ 0_3 \end{pmatrix}, \quad (7)$$

where for notational convenience we have normalized the pressure law $p(\rho)$ to satisfy

$$p'(1) = 1 \quad (8)$$

by rescaling ε_M . Also, the full average $\text{av} f$ and the vertical average $\mathbf{a}_z f$ of any function f defined on the 3-torus \mathbb{T}^3 are

$$\text{av} f := \frac{\iiint f \, dx \, dy \, dz}{\iiint 1 \, dx \, dy \, dz}, \quad (\mathbf{a}_z f)(x, y) := \frac{\int f(x, y, z) \, dz}{\int 1 \, dz}. \quad (9)$$

Although the uniform existence and convergence results require that the initial data satisfy the “well-preparedness” condition

$$\|(\varepsilon_A^{-1} \mathcal{L}_A + \varepsilon_M^{-1} \mathcal{L}_M) \mathbf{V}^0\|_{n-1} \leq c, \quad (10)$$

they do not require any assumption about the rate at which the initial data converge to their limit. However, such an assumption is obviously required in order to obtain a rate of convergence of solutions of the PDE. In the convergence rate result we assume that the initial data has the form developed in §2, which is a specialization of the general form of initial data satisfying (10). Specifically, after expanding the initial data in powers of the small parameters and their ratio, the leading-order terms are assumed to be independent of the small parameters in order to avoid degrading the convergence rate. However, valid estimates for any initial data satisfying (10) below can be obtained simply by adding the size of the difference of the initial data when that difference is larger than the estimates obtained below.

When the parameter μ in (3) is fixed then (1) and (4) essentially contain only one small parameter and hence have only two time scales. Uniform existence and convergence results for initial-value problems of general systems containing one small parameter were obtained in [KM81]. Moreover, their results remain valid with only cosmetic changes to the proofs whenever $\mu_{\text{lim}} > 0$. Convergence rate theorems for both specific and general two-scale systems have been proven in [KM82, Sch88, Sch94, Che12, Che14]. We therefore focus on the more

challenging case when $\mu \rightarrow 0$, although our results will be phrased so as to remain valid when $\mu_{\text{lim}} > 0$. It will be convenient to express our convergence results using just powers of ε_M , by defining a parameter ν determined by

$$\varepsilon_A = \varepsilon_M^{1+\nu}, \quad \text{or, equivalently,} \quad \mu = \varepsilon_M^\nu, \quad \text{i.e.,} \quad \nu := \frac{\ln(\frac{1}{\varepsilon_A})}{\ln(\frac{1}{\varepsilon_M})} - 1 = \frac{\ln(\frac{1}{\mu})}{\ln(\frac{1}{\varepsilon_M})}, \quad (11)$$

where in view of the results for two-scale singular limits we will assume for notational simplicity that

$$\varepsilon_A < \varepsilon_M, \quad \text{i.e.,} \quad \mu < 1 \text{ and } \nu > 0. \quad (12)$$

To simplify the exposition we will assume that the spatial domain is periodic. The uniform existence result remains valid with the same proof when the spatial domain is \mathbb{R}^3 , while the the limit system is then identically zero because it is independent of the vertical coordinate.

Throughout this paper c and C denote positive constants that are independent of ε_A and ε_M , which may take different values in each appearance.

Theorem 1.1. *Let $n \geq 3$ be an integer. Assume that the spatial domain is \mathbb{T}^3 and that the small parameters are restricted to the region*

$$0 < \varepsilon_M \leq \varepsilon_M^0 \quad \text{and} \quad \varepsilon_A \geq c \varepsilon_M^{1+\frac{1}{n-1}}. \quad (13)$$

Assume in addition that the initial data $\mathbf{V}^0 := (r^0, \mathbf{u}^0, \mathbf{b}^0)$ for system (4), which may depend on the small parameters ε_A and ε_M , are uniformly bounded in H^n and satisfy (4d) and (10).

Uniform Existence.: *Under the above conditions there exist fixed positive T and K such that for $(\varepsilon_A, \varepsilon_M)$ satisfying (13) the solution to (4) having the initial data \mathbf{V}^0 exists for $0 \leq t \leq T$ and satisfies*

$$\sup_{0 \leq t \leq T} \left[\|\mathbf{V}\|_n + \|\mathbf{V}_t\|_0 + \sum_{j=1}^n \varepsilon_M^{j-1} \left(\min\left(\frac{\varepsilon_A}{\varepsilon_M}, 1\right) \right)^{n-1} \|\partial_t^j \mathbf{V}\|_{n-j} \right] \leq K. \quad (14)$$

Convergence and Limit.: *Assume in addition that the normalization (8) holds and that as $(\varepsilon_A, \varepsilon_M)$ satisfying (13) tend to zero, their ratio $\frac{\varepsilon_A}{\varepsilon_M}$ converges to some value μ_{lim} and the initial data \mathbf{V}^0 converges in H^n to $\bar{\mathbf{V}}^0$. Then the solution $\mathbf{V} = (r, \mathbf{u} = (u_h, u_3), \mathbf{b} = (b_h, b_3))$ of the MHD system (4) with initial data \mathbf{V}^0 converges in $C^0([0, T]; H^{n-\alpha})$ for every $\alpha > 0$. Its limit is independent of z , and is the unique solution $\bar{\mathbf{V}} = (\bar{r}, (\bar{u}_h, \bar{u}_3), (\bar{b}_h, \bar{b}_3))$ of the limit system*

$$(1 + \mu_{\text{lim}}^2) [\partial_t \bar{r} + (\bar{u}_h \cdot \nabla_h) \bar{r}] + \mu_{\text{lim}} (\bar{b}_h \cdot \nabla_h) \bar{u}_3 = 0, \quad (15a)$$

$$\mathcal{P}_h^{\text{div}} (\partial_t \bar{u}_h + (\bar{u}_h \cdot \nabla_h) \bar{u}_h - (\bar{b}_h \cdot \nabla_h) \bar{b}_h) = 0, \quad \nabla_h \cdot \bar{u}_h = 0, \quad (15b)$$

$$\partial_t \bar{u}_3 + (\bar{u}_h \cdot \nabla_h) \bar{u}_3 + \mu_{\text{lim}} (\bar{b}_h \cdot \nabla_h) \bar{r} = 0, \quad (15c)$$

$$\partial_t \bar{b}_h + (\bar{u}_h \cdot \nabla_h) \bar{b}_h - (\bar{b}_h \cdot \nabla_h) \bar{u}_h = 0, \quad \nabla_h \cdot \bar{b}_h = 0, \quad (15d)$$

$$\bar{b}_3 = \text{av} \bar{b}_3^0 - \mu_{\text{lim}} (\bar{r} - \text{av} \bar{r}^0) \quad (15e)$$

having initial data $\bar{\mathbf{V}}^0$.

Rate of Convergence.: *In addition to the original assumptions and the additional assumptions of the convergence part, assume that the initial data for the MHD system have the more specific form (44), (46), (50) and that (12) holds. Then there is a constant c independent of ε_A and ε_M such that for all $t \in [0, T]$,*

$$\|(1 - \mathcal{P}_h^{\text{div}} \mathbf{a}_z) \mathbf{u}_h\|_j + \|(1 - \mathbf{a}_z) \mathbf{b}_h\|_j + \|b_3 - \text{av} \bar{b}_3^0 + \mu (\mathbf{a}_z r - \text{av} \bar{r}^0)\|_j \leq c \varepsilon_M^{1-(j-1)\nu}, \quad j = 0, \dots, n-1, \quad (16a)$$

$$\|\mathcal{P}_h^{\text{div}} \mathbf{a}_z \mathbf{u}_h - \bar{u}_h\|_{n-2} + \|\mathbf{a}_z \mathbf{b}_h - \bar{b}_h\|_{n-2} \leq c \varepsilon_M, \quad (16b)$$

$$\|(1 - \mathbf{a}_z) r\|_j + \|(1 - \mathbf{a}_z) u_3\|_j \leq c \varepsilon_M^{1-(j-1)\nu}, \quad j = 1, \dots, n-1, \quad (16c)$$

$$\|\mathbf{a}_z r - \bar{r}\|_{n-2} + \|\mathbf{a}_z u_3 - \bar{u}_3\|_{n-2} \leq c \left[\varepsilon_M^{1-\max(n-5,0)\nu} + |\mu - \mu_{\text{lim}}| \right]. \quad (16d)$$

Moreover, there exist $O(1)$ correctors $(r^{(\text{cor})}, u_3^{(\text{cor})})$ defined in (133)–(134) such that

$$\|\mathbf{a}_z r - (\bar{r} + \frac{\mu - \mu_{\text{lim}}}{1 + \mu^2} r^{(\text{cor})})\|_{n-2} + \|\mathbf{a}_z u_3 - (\bar{u}_3 + (\mu - \mu_{\text{lim}}) u_3^{(\text{cor})})\|_{n-2} \leq c \varepsilon_M^{1-\max(n-5,0)\nu}. \quad (16e)$$

The uniform existence part of Theorem 1.1 is a special case of the corresponding general result for systems (1) stated in Lemma B.1 and proven in §B, which is an improvement of [CJS18, Theorem 3.6]. The convergence part of Theorem 1.1 will be proven in §C. The convergence-rate estimates in (16) are direct consequences of the estimates (54), (66), (92), (115), and (135) proven in §3.

Under the scaling (13) all powers of ε_M appearing in (16) are positive, so a nontrivial rate of convergence is obtained over the full range of allowed values of ν , in all the norms listed in the theorem. The corrector estimate (16e) has been included because (13)–(11) imply that estimate (16d) is much weaker than the other estimates in (16) in the main case of interest in which $\mu_{\text{lim}} = 0$. The improved estimate (16e) involving the corrector shows that (16d) is in fact sharp in this case, and gives a formula for the principal error term.

For well-prepared initial data like that considered here, the convergence rate for solutions of two-scale systems is typically first order in the single small parameter [Sch88, Sch94]. That result depends crucially on the uniform boundedness of the first time derivative of solutions being propagated for positive time, which does *not* generally hold for three-scale systems (1) [CJS18, §2]. We use the time-integration method developed in [Che12, Che14] to mitigate the effect of the lack of uniform boundedness of first time derivatives. Moreover, the estimates in (16) are $O(\varepsilon_M)$ for those components and norms for which uniform boundedness of the first time derivative holds, and tend, except for (16e), to the two-scale $O(\varepsilon_M)$ convergence rate as $\nu \rightarrow 0$, which makes (1) tend to a two-scale system. Obtaining that asymptotic consistency is only possible on account of the improvement (14) over the estimate $\varepsilon_M \|\mathbf{V}_t\|_{n-1} \leq c$ in [CJS18].

The H^j estimates (16a) and (16c) for intermediate values of j are obtained using interpolation. Those estimates are the starting point for an improved estimate for z -averages of products derived in §E, which is used in the proof of (16b), and should be useful more generally.

One of the main techniques used in the proof of (16) is partitioning the solution into fast, intermediate, and slow modes, and analyzing each mode separately. In contrast to the two-scale case (e.g. [Che14, §2]), the exact eigenspaces of the large operator $\frac{1}{\varepsilon_A} \mathcal{L}_A + \frac{1}{\varepsilon_M} \mathcal{L}_M$ having eigenvalues of sizes strictly $O(\frac{1}{\varepsilon_A})$, $O(\frac{1}{\varepsilon_M})$, and $o(\frac{1}{\varepsilon_M})$ depend on the parameter μ . For simplicity, in §2 we define the fast, intermediate, and slow modes to be the fixed projections onto the limits as $\mu \rightarrow 0$ of the exact eigenspaces. These projections are also used to determine the appropriate form of the initial data used in the convergence rate part of Theorem 1.1. However, certain estimates in §3 for the intermediate modes require the use of the exact μ -dependent modes.

The only previous convergence rate result we know for evolutionary PDEs with two parameters tending to zero independently appears in [FFGHR17], which considered the low Rossby and magnetic Reynolds number limit of the stochastically-forced viscous incompressible rotating MHD system. Compared to the hyperbolic system considered here the deterministic case without forcing [FFGHR17, p. 4444] of their system has many simplifying features, including the absence of a matrix multiplying the time derivatives, which eliminates the need for restriction (13), the presence of a closed L^2 energy estimate, and the presence of regularizing viscous terms whose diffusivity rates tend to infinity as the small parameters tend to zero, which yields a highly parabolic system that induces smoothing when the large operator is used to determine the fast components in terms of the slow component.

2. ANALYSIS OF THE LARGE OPERATOR

Following [CJS18, §4] but without treating each Fourier mode separately, let \mathbb{P}^0 denote the L^2 -orthogonal projection operator onto the nullspace of \mathcal{L}_A , and let \mathbb{P}^1 denote the L^2 -orthogonal projection operator onto the

nullspace of $\mathbb{P}^0 \mathcal{L}_M \mathbb{P}^0$. Then $R(I - \mathbb{P}^0) \subseteq N(\mathbb{P}^0 \mathcal{L}_M \mathbb{P}^0) = R(\mathbb{P}^1) = N(I - \mathbb{P}^1)$, so $R(I - \mathbb{P}^0) \perp R(I - \mathbb{P}^1)$, and hence

$$(I - \mathbb{P}^0)(I - \mathbb{P}^1) = 0 = (I - \mathbb{P}^1)(I - \mathbb{P}^0). \quad (17)$$

Expanding the factors in (17) shows that $\mathbb{P}^0 \mathbb{P}^1 = \mathbb{P}^1 \mathbb{P}^0$, which implies that $\mathbb{P} := \mathbb{P}^0 \mathbb{P}^1$ is an orthogonal projection operator satisfying $\mathbb{P}(I - \mathbb{P}^j) = 0 = (I - \mathbb{P}^j)\mathbb{P}$ for $j \in \{0, 1\}$. Moreover, (17) and the definition of \mathbb{P} yield $(I - \mathbb{P}^0) + (I - \mathbb{P}^1) + \mathbb{P} = I + (I - \mathbb{P}^0)(I - \mathbb{P}^1) = I$, which shows that the sum of the fast, intermediate, and slow modes defined by

$$\mathbf{V}^F := (I - \mathbb{P}^0)\mathbf{V}, \quad \mathbf{V}^I := (I - \mathbb{P}^1)\mathbf{V}, \quad \mathbf{V}^S := \mathbb{P}\mathbf{V} \quad (18)$$

satisfies $\mathbf{V}^F + \mathbf{V}^I + \mathbf{V}^S = \mathbf{V}$. These modes are the limits as $\mu \rightarrow 0$ of the direct sums of the eigenspaces of $\mathcal{L}_A + \mu \mathcal{L}_M$ whose eigenvalues are strictly $O(1)$, $O(\mu)$, and $o(\mu)$, respectively [Kat82], [CJS18, §4]. Moreover, since the \mathbb{P}^j are orthogonal projections onto the null spaces of constant-coefficient differential operators they commute with derivatives, and hence are also orthogonal in any H^k . Therefore estimates for the full solution obtained by combining estimates for each mode are as sharp as the component estimates, modulo constant factors.

The above results do not depend on the particular form of the operators \mathcal{L}_A and \mathcal{L}_M . We now calculate the projections and modes for the MHD system (4). For brevity, we restrict consideration to $\mathbf{V} = (r, \mathbf{u}, \mathbf{b})$ satisfying $\nabla \cdot \mathbf{b} = 0$, which causes no difficulties since we only consider initial data and solutions satisfying that constraint. Recall that $\mathcal{P}_h^{\text{div}}$, \mathbf{a}_z , and av were defined in (6) and (9).

Lemma 2.1. *Assume that the spatial domain is \mathbb{T}^3 and that $\mathbf{V} = (r, \mathbf{u}, \mathbf{b})$ where \mathbf{b} satisfies $\nabla \cdot \mathbf{b} = 0$. Then*

(1) $\mathcal{L}_A \mathbf{V} = \mathbf{0}$ iff $\partial_z u_h = \mathbf{0} = \partial_z b_h$, $\nabla_h \cdot u_h = 0$, and $b_3 = \text{av} b_3$.

(2) $\left(\frac{1}{\varepsilon_A} \mathcal{L}_A + \frac{1}{\varepsilon_M} \mathcal{L}_M\right) \mathbf{V} = \mathbf{0}$ iff

$$\partial_z \mathbf{V} = \mathbf{0}, \quad \nabla_h \cdot u_h = 0, \quad b_3 = \text{av} b_3 - \frac{\varepsilon_A}{\varepsilon_M} (r - \text{av} r). \quad (19)$$

(3) *The formulas for the projections are*

$$\mathbb{P}^0 = \begin{pmatrix} 1 & & & \\ & \begin{pmatrix} \mathcal{P}_h^{\text{div}} & \mathbf{a}_z I_{2 \times 2} \\ & 1 \end{pmatrix} & & \\ & & \begin{pmatrix} \mathbf{a}_z I_{2 \times 2} \\ \text{av} \end{pmatrix} & \\ & & & \end{pmatrix}, \quad \mathbb{P}^1 = \begin{pmatrix} \mathbf{a}_z & & & \\ & \begin{pmatrix} I_{2 \times 2} & \mathbf{a}_z \\ & \end{pmatrix} & & \\ & & & I_{3 \times 3} \\ & & & \end{pmatrix}, \quad (20)$$

$$\mathbb{P} = \begin{pmatrix} \mathbf{a}_z & & & \\ & \begin{pmatrix} \mathcal{P}_h^{\text{div}} & \mathbf{a}_z I_{2 \times 2} \\ & \mathbf{a}_z \end{pmatrix} & & \\ & & & \begin{pmatrix} \mathbf{a}_z I_{2 \times 2} \\ \text{av} \end{pmatrix} \\ & & & \end{pmatrix},$$

where all missing entries vanish.

(4) *All eigenvalues of $\mathcal{L}_A + \mu \mathcal{L}_M$ that are $o(\mu)$ are identically zero.*

(5) *Using the notations $\mathbf{V}^\ell = (r^\ell, \mathbf{u}^\ell, \mathbf{b}^\ell)$ for $\ell \in \{F, I, S\}$ and $\mathbf{w}^\ell = (w_h^\ell, w_3^\ell)$ for $w \in \{\mathbf{u}, \mathbf{b}\}$, the formulas for the fast, intermediate and slow modes are*

$$\begin{aligned} (r^F \quad \mathbf{u}^F \quad \mathbf{b}^F) &= \begin{pmatrix} 0 & \begin{pmatrix} (1 - \mathbf{a}_z \mathcal{P}_h^{\text{div}}) u_h \\ 0 \end{pmatrix} & \begin{pmatrix} (1 - \mathbf{a}_z) b_h \\ (1 - \text{av}) b_3 \end{pmatrix} \end{pmatrix}, \\ (r^I \quad \mathbf{u}^I \quad \mathbf{b}^I) &= \begin{pmatrix} (1 - \mathbf{a}_z) r & \begin{pmatrix} 0_2 \\ (1 - \mathbf{a}_z) u_3 \end{pmatrix} & \mathbf{0}_3 \end{pmatrix}, \\ (r^S \quad \mathbf{u}^S \quad \mathbf{b}^S) &= \begin{pmatrix} \mathbf{a}_z r & \begin{pmatrix} \mathcal{P}_h^{\text{div}} \mathbf{a}_z u_h \\ \mathbf{a}_z u_3 \end{pmatrix} & \begin{pmatrix} \mathbf{a}_z b_h \\ \text{av} b_3 \end{pmatrix} \end{pmatrix}. \end{aligned} \quad (21)$$

In particular,

$$\nabla \cdot \mathbf{b}^F = 0, \quad \nabla \cdot \mathbf{b}^I = 0, \quad \nabla \cdot \mathbf{b}^S = 0, \quad (22)$$

$$\mathbf{V}^S \text{ is independent of } z, \quad (23)$$

and

$$\nabla_{\mathbf{h}} \cdot \mathbf{u}_{\mathbf{h}}^S = 0 = \nabla_{\mathbf{h}} \cdot \mathbf{b}_{\mathbf{h}}^S. \quad (24)$$

(6) For any nonnegative integer j , there are constants c_1 and c_2 such that

$$\begin{aligned} c_1 \|(\mathcal{L}_A + \mu \mathcal{L}_M) \mathbf{V}\|_j &\leq \|\partial_z \mathbf{u}_{\mathbf{h}}^F\|_j + \|\nabla_{\mathbf{h}} \cdot \mathbf{u}_{\mathbf{h}}^F\|_j + \|\partial_z (b_{\mathbf{h}}^F - \mu \Delta^{-1} \nabla_{\mathbf{h}} \partial_z r^I)\|_j \\ &\quad + \|b_3^F + \mu \Delta^{-1} \Delta_{\mathbf{h}} r\|_{j+1} + \mu [\|\partial_z r^I\|_j + \|\partial_z u_3^I\|_j] \\ &\leq c_2 \|(\mathcal{L}_A + \mu \mathcal{L}_M) \mathbf{V}\|_j. \end{aligned} \quad (25)$$

Proof. Applying \mathcal{L}_A to \mathbf{V} yields $(0, (\partial_z b_{\mathbf{h}} - \nabla_{\mathbf{h}} b_3), (\frac{\partial_z \mathbf{u}_{\mathbf{h}}}{-\nabla_{\mathbf{h}} \cdot \mathbf{u}_{\mathbf{h}}}))$. Hence the last part of $\mathcal{L}_A \mathbf{V}$ vanishes iff $\partial_z \mathbf{u}_{\mathbf{h}} = \mathbf{0}$ and $\nabla_{\mathbf{h}} \cdot \mathbf{u}_{\mathbf{h}} = 0$. Taking the horizontal divergence of the second component of \mathcal{L}_A and using the fact that $\nabla \cdot \mathbf{b} = 0$ yields

$$\nabla_{\mathbf{h}} \cdot (\partial_z b_{\mathbf{h}} - \nabla_{\mathbf{h}} b_3) = -\Delta b_3. \quad (26)$$

Hence if $\mathcal{L}_A \mathbf{V} = 0$ then b_3 is a constant, i.e., $b_3 = \text{av} b_3$, which implies further that $\partial_z b_{\mathbf{h}} = \nabla_{\mathbf{h}} b_3 = \mathbf{0}$. On the other hand, when those conditions hold then each term in the second part of $\mathcal{L}_A \mathbf{V}$ vanishes.

Similarly,

$$\left(\frac{1}{\varepsilon_A} \mathcal{L}_A + \frac{1}{\varepsilon_M} \mathcal{L}_M \right) \mathbf{V} = \left(\begin{array}{c} -\frac{1}{\varepsilon_M} \nabla \cdot \mathbf{u} \\ \left(\frac{1}{\varepsilon_A} (-\nabla_{\mathbf{h}} b_3 + \partial_z b_{\mathbf{h}} - \frac{\varepsilon_A}{\varepsilon_M} \nabla_{\mathbf{h}} r) \right) \\ -\frac{1}{\varepsilon_M} \partial_z r \\ \frac{1}{\varepsilon_A} \left(\frac{\partial_z \mathbf{u}_{\mathbf{h}}}{-\nabla_{\mathbf{h}} \cdot \mathbf{u}_{\mathbf{h}}} \right) \end{array} \right). \quad (27)$$

The last component of the second part vanishes iff

$$\partial_z r = 0, \quad (28)$$

and since $\partial_z u_3 = \nabla \cdot \mathbf{u} - \nabla_{\mathbf{h}} \cdot \mathbf{u}_{\mathbf{h}}$, the first and last parts vanish iff $\nabla_{\mathbf{h}} \cdot \mathbf{u}_{\mathbf{h}} = 0$ and $\partial_z \mathbf{u} = \mathbf{0}$. Next, use (26) and (28) to write the horizontal divergence of the second part as

$$\nabla_{\mathbf{h}} \cdot \left(-\nabla_{\mathbf{h}} b_3 + \partial_z b_{\mathbf{h}} - \frac{\varepsilon_A}{\varepsilon_M} \nabla_{\mathbf{h}} r \right) = -(\Delta b_3 + \frac{\varepsilon_A}{\varepsilon_M} \Delta_{\mathbf{h}} r) = \Delta (b_3 + \frac{\varepsilon_A}{\varepsilon_M} r), \quad (29)$$

which vanishes iff the last equation of (19) holds. That equation together with (28) implies that $\partial_z b_3 = 0$. Finally, the last equation of (19) also shows that the horizontal components of the second part of (27) vanish iff $\partial_z b_{\mathbf{h}} = \mathbf{0}$.

The formula for \mathbb{P}^0 follows from the conditions for $\mathcal{L}_A \mathbf{V}$ to vanish in the first part of the lemma, and that formula together with the formula for \mathcal{L}_M in (7) implies that

$$\mathbb{P}^0 \mathcal{L}_M \mathbb{P}^0 \mathbf{V} = -(\partial_z u_3, (\frac{0}{\partial_z r}), 0). \quad (30)$$

Formula (30) implies the formula for \mathbb{P}^1 , and the formulas for the \mathbb{P}^j yield \mathbb{P} .

The conditions (19) for $(\mathcal{L}_A + \mu \mathcal{L}_M) \mathbf{V}$ to vanish differ from the conditions to belong to the null space of \mathbb{P} only by adding an $O(\mu)$ term to the formula for b_3 . Hence the rank of the restriction to any Fourier mode of $N(\mathcal{L}_A + \mu \mathcal{L}_M)$ equals the rank of the restriction to that mode of \mathbb{P} , which in turn equals the dimension of the direct sum of all eigenspaces of $\mathcal{L}_A + \mu \mathcal{L}_M$ in that Fourier mode having eigenvalues of size $o(\mu)$ [Kat82], [CJS18, §4]. Hence all such eigenvalues vanish identically.

The formulas (21) for the modes follow from their definition (18) and formula (20) for the projections. Formula (21) for \mathbf{b}^S plus the fact that $\mathbf{a}_z \mathbf{b}$ is independent of z imply that $\nabla \cdot \mathbf{b}^S = 0$, and trivially $\nabla \cdot \mathbf{b}^I = 0$, which implies the rest of (22). Each component of \mathbf{V}^S contains \mathbf{a}_z or av , so (23) holds. The left equation in (24) follows from the presence of the operator $\mathcal{P}_{\mathbf{h}}^{\text{div}}$ in the formula for $\mathbf{u}_{\mathbf{h}}^S$, while the right equation there follows from (22)–(23).

The formula for u_h^F in (21) and the formula (27) for $(\mathcal{L}_A + \mu\mathcal{L}_M)\mathbf{V}$ yield

$$\|\partial_z u_h^F\|_j = \|\partial_z u_h\|_j \leq c\|(\mathcal{L}_A + \mu\mathcal{L}_M)\mathbf{V}\|_j, \quad (31)$$

$$\|\nabla_h \cdot u_h^F\|_j = \|\nabla_h \cdot u_h\|_j \leq c\|(\mathcal{L}_A + \mu\mathcal{L}_M)\mathbf{V}\|_j. \quad (32)$$

By the ellipticity of Δ , formula (21) for b_3^F , the first identity in (29) and (27),

$$\begin{aligned} \|b_3^F + \mu\Delta^{-1}\Delta_h r\|_{j+1} &= \|\Delta^{-1}[\Delta(1 - \text{av})b_3 + \mu\Delta_h r]\|_{j+1} \leq c\|\Delta b_3 + \mu\Delta_h r\|_{j-1} \\ &= c\|\nabla_h \cdot (\nabla_h b_3 + \mu\nabla_h r - \partial_z b_h)\|_{j-1} \leq c\|(\mathcal{L}_A + \mu\mathcal{L}_M)\mathbf{V}\|_j \end{aligned} \quad (33)$$

Also, combining formula (21) for b_h^F , formula (27) for $\mathcal{L}_A + \mu\mathcal{L}_M$, and (33) yields

$$\begin{aligned} \|\partial_z(b_h^F - \mu\Delta^{-1}\nabla_h \partial_z r^I)\|_j &= \|\partial_z(b_h - \mu\Delta^{-1}\nabla_h \partial_z r)\|_j \\ &= \|\partial_z b_h - \mu\nabla_h \Delta^{-1}(\Delta r - \Delta_h r)\|_j \\ &= \|(\partial_z b_h - \mu\nabla_h r - \nabla_h b_3) + \nabla_h(b_3 + \mu\Delta^{-1}\Delta_h r)\|_j \\ &\leq \|(\mathcal{L}_A + \mu\mathcal{L}_M)\mathbf{V}\|_j + \|b_3^F + \mu\Delta^{-1}\Delta_h r\|_{j+1} \leq c\|(\mathcal{L}_A + \mu\mathcal{L}_M)\mathbf{V}\|_j. \end{aligned} \quad (34)$$

By the formula (21) defining the modes and formula (27) for $\mathcal{L}_A + \mu\mathcal{L}_M$,

$$\mu\|\partial_z r^I\|_j = \mu\|\partial_z r\|_j \leq c\|(\mathcal{L}_A + \mu\mathcal{L}_M)\mathbf{V}\|_j, \quad (35)$$

$$\mu\|\partial_z u_3^I\|_j = \mu\|\partial_z u_3\|_j = \mu\|\nabla \cdot \mathbf{u} - \nabla_h \cdot u_h\|_j \leq c\|(\mathcal{L}_A + \mu\mathcal{L}_M)\mathbf{V}\|_j \quad (36)$$

Combining (31)–(36) yields the right inequality in (25), and the expressions estimated there yield all the terms in $(\mathcal{L}_A + \mu\mathcal{L}_M)\mathbf{V}$ so the left inequality also holds. \square

Solving the PDE for $(\mathcal{L}_A + \mu\mathcal{L}_M)\mathbf{V}$ and using the bounds (14) to estimate the result yields a bound for the H^{n-1} norm of that expression, which by (25) implies “static” estimates for the fast and intermediate modes, which will be written explicitly in §3. However, we also need to obtain “dynamic” estimates for the intermediate and slow modes via differential inequalities. To do so we cannot use the limit modes defined above. The exact slow eigenspace of the zero eigenvalue of $\mathcal{L}_A + \mu\mathcal{L}_M$ was determined in Lemma 2.1. In the next lemma we obtain formulas for 7-vectors of Fourier multiplier operators \mathbf{V}_α and \mathbf{V}_β and the self-adjoint Fourier multiplier operator \mathcal{Q} , such that

$$(\mathcal{L}_A + \mu\mathcal{L}_M)\mathbf{V}_\alpha \cdot \tilde{\mathbf{V}} + \mu\mathcal{Q}\partial_z \mathbf{V}_\beta \cdot \tilde{\mathbf{V}} = 0 = (\mathcal{L}_A + \mu\mathcal{L}_M)\mathbf{V}_\beta \cdot \tilde{\mathbf{V}} + \mu\mathcal{Q}\partial_z \mathbf{V}_\alpha \cdot \tilde{\mathbf{V}} \quad (37)$$

for every vector-valued function $\tilde{\mathbf{V}}$. Equations (37) imply that for every Fourier mode (k, l, m) the linear combinations $(\mathbf{V}_\alpha \pm \mathbf{V}_\beta)e^{i(kx+ly+mz)}$ are eigenfunctions of the operator $\mathcal{L}_A + \mu\mathcal{L}_M$ with purely imaginary or zero eigenvalues $\mp im\mu\hat{\mathcal{Q}}$ where

$$\hat{\mathcal{Q}} := e^{-i(kx+ly+mz)} \left[\mathcal{Q}e^{i(kx+ly+mz)} \right], \quad (38)$$

i.e., they yield the exact μ -dependent intermediate eigenspaces. However, in §3 we will obtain dynamic estimates for $(1 - \mathbf{a}_z)\mathbf{V}_\alpha \cdot \mathbf{V}$ and $(1 - \mathbf{a}_z)\mathbf{V}_\beta \cdot \mathbf{V}$ rather than $(\mathbf{V}_\alpha \pm \mathbf{V}_\beta) \cdot \mathbf{V}$, to reduce disruption to the structure of the rest of the PDE. The operator $1 - \mathbf{a}_z$ is applied because the eigenvalues $\pm im\mu\mathcal{Q}$ are only of size μ when $m \neq 0$. Moreover, since $\nabla \cdot \mathbf{b} = 0$ we will replace \mathbf{V}_α by the variant $\mathbf{V}_\alpha^{\text{no-div}}$ defined in (39) that omits the gradient term in the magnetic field component.

Since the limits as $\mu \rightarrow 0$ of \mathbf{V}_α and \mathbf{V}_β should belong to the intermediate mode defined in (21), trying various perturbations leads to the ansatz

$$\begin{aligned}\mathbf{V}_\alpha &:= \mathbf{V}_\alpha^{\text{div}} + \mathbf{V}_\alpha^{\text{no-div}} = \begin{pmatrix} 0 \\ 0_3 \\ \mu \mathcal{A} \Delta^{-1} \partial_z \nabla \end{pmatrix} + \begin{pmatrix} 1 \\ 0_3 \\ -\mu \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} + \mu^3 \begin{pmatrix} 0 \\ 0 \\ \mathcal{B} \end{pmatrix} \end{pmatrix}, \\ \mathbf{V}_\beta &:= \begin{pmatrix} 0 \\ \left(\mu^2 \mathcal{C} \Delta^{-1} \partial_z \nabla_h \right) \\ \mathcal{D} \\ 0_3 \end{pmatrix},\end{aligned}\tag{39}$$

where the vectors have been normalized by setting the first component of \mathbf{V}_α to the identity operator 1, and factors of Δ^{-1} have been included so that if $(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})$ are all homogeneous order zero Fourier multipliers then all components of \mathbf{V}_α and \mathbf{V}_β will also be Fourier multipliers of order zero. Furthermore, in the limit as $\mu \rightarrow 0$ the operator $-\partial_z \mathcal{Q}$ should tend to the operator $-\partial_z$ appearing in $\mathbb{P}^0 \mathcal{L}_M \mathbb{P}^0$ in (30), i.e., \mathcal{Q} should tend to one.

Lemma 2.2. *The vectors \mathbf{V}_α and \mathbf{V}_β defined in (39) satisfy (37) with*

$$\mathcal{Q} = \sqrt{2 \left(1 + \mu^2 + \sqrt{(1 + \mu^2)^2 - 4\mu^2 \partial_z^2 \Delta^{-1}} \right)^{-1}}, \quad \text{i.e.,} \quad \hat{\mathcal{Q}} = \frac{\sqrt{2}}{\sqrt{1 + \mu^2 + \sqrt{(1 + \mu^2)^2 - 4\mu^2 \frac{m^2}{k^2 + l^2 + m^2}}}},\tag{40}$$

provided that

$$\mathcal{C} = -\mathcal{Q} \left(1 - \mu^2 \partial_z^2 \Delta^{-1} \mathcal{Q}^2 \right)^{-1}, \quad \mathcal{A} = -\mathcal{C} \mathcal{Q}^{-1}, \quad \mathcal{B} = \partial_z^2 \Delta^{-1} \mathcal{C} \mathcal{Q}, \quad \mathcal{D} = \mathcal{Q}^{-1}.\tag{41}$$

Proof. Substituting (39) into (37) yields

$$\begin{pmatrix} 0 \\ \left(-\mu^3 (\mathcal{B} - \mathcal{C} \mathcal{Q} \partial_z^2 \Delta^{-1}) \nabla_h \right) \\ -\mu \partial_z (1 - \mathcal{D} \mathcal{Q}) \\ 0_3 \end{pmatrix} = \mathbf{0} = \begin{pmatrix} -\mu \partial_z (\mathcal{D} - \mathcal{Q} + \mu^2 \mathcal{C} \Delta_h \Delta^{-1}) \\ 0_3 \\ \mu^2 \partial_z^2 \Delta^{-1} (\mathcal{C} + \mathcal{A} \mathcal{Q}) \nabla - \begin{pmatrix} 0 \\ 0 \\ \mu^2 \partial_z (\mathcal{C} + \mathcal{Q} - \mu^2 \mathcal{B} \mathcal{Q}) \end{pmatrix} \end{pmatrix},\tag{42}$$

which will hold provided that

$$\mathcal{C} + \mathcal{A} \mathcal{Q} = 0, \quad \mathcal{D} \mathcal{Q} = 1, \quad \mathcal{B} = \mathcal{C} \mathcal{Q} \partial_z^2 \Delta^{-1}, \quad \mathcal{C} + \mathcal{Q} = \mu^2 \mathcal{B} \mathcal{Q}, \quad \mathcal{D} - \mathcal{Q} + \mu^2 \Delta_h \Delta^{-1} \mathcal{C} = 0.\tag{43}$$

Solving the first three equations in (43) for \mathcal{A} , \mathcal{D} , and \mathcal{B} yields the formulas for those operators claimed in (41). Substituting those formulas into the fourth equation in (43) and solving the result for \mathcal{C} yields the formula for that operator in (41). Substituting the formulas obtained so far into the last equation in (43) yields $\mathcal{Q}^{-1} - \mathcal{Q} - \mu^2 \Delta_h \Delta^{-1} \mathcal{Q} (1 - \mu^2 \partial_z^2 \Delta^{-1} \mathcal{Q}^2)^{-1} = 0$, whose only solution tending to one as $\mu \rightarrow 0$ is (40). \square

We now turn to explicating the form that initial data satisfying (10) takes.

Lemma 2.3. *Initial data $\mathbf{V}^0 = (r^0, \mathbf{u}^0, \mathbf{b}^0)$ will be uniformly bounded in H^n and satisfy the constraint $\nabla \cdot \mathbf{b}^0 = 0$ and the condition (10) iff it has the form*

$$\begin{aligned}r^0 &= r^{0,S} + \varepsilon_M r^{0,I}, \quad u_h^0 = u_h^{0,S} + \varepsilon_A u_h^{0,F}, \quad u_3^0 = u_3^{0,S} + \varepsilon_M u_3^{0,I}, \\ b_h^0 &= b_h^{0,S} + \varepsilon_A b_h^{0,F}, \quad b_3^0 = a v b_3^0 - \mu (1 - a v) r^{0,S} + \varepsilon_A b_3^{0,F},\end{aligned}\tag{44}$$

where every term $w^{0,\ell}$ with $w \in \{r, \mathbf{u}_h, u_3, \mathbf{b}_h, b_3\}$ and $\ell \in \{F, I, S\}$ has the form specified for the w component of the ℓ mode in (21), and may depend on $(\varepsilon_A, \varepsilon_M)$ but satisfies

$$\begin{aligned} & \|r^{0,S}\|_n + [\|\partial_z r^{0,I}\|_{n-1} + \varepsilon_M \|r^{0,I}\|_n] \\ & + \|\mathbf{u}_h^{0,S}\|_n + [\|\partial_z \mathbf{u}_h^{0,F}\|_{n-1} + \|\nabla_{\mathbf{h}} \cdot \mathbf{u}_h^{0,F}\|_n + \varepsilon_A \|\mathbf{u}_h^{0,F}\|_n] + \|u_3^{0,S}\|_n + [\|\partial_z u_3^{0,I}\|_{n-1} + \varepsilon_M \|u_3^{0,I}\|_n] \\ & + \|\mathbf{b}_h^{0,S}\|_n + [\|\partial_z \mathbf{b}_h^{0,F}\|_{n-1} + \varepsilon_A \|\mathbf{b}_h^{0,F}\|_n] + |\text{av} b_3^0| + \|b_3^{0,F} + \Delta^{-1} \Delta_h r^{0,I}\|_n \leq c \end{aligned} \quad (45)$$

uniformly in those parameters, and

$$\nabla_{\mathbf{h}} \cdot \mathbf{b}_h^{0,S} = 0 = \nabla_{\mathbf{h}} \cdot \mathbf{b}_h^{0,F} + \partial_z b_3^{0,F}. \quad (46)$$

Proof. Since the terms in (44) are allowed to depend on ε_A and ε_M , that formula simply expresses the separation of the initial data into fast, intermediate, and slow modes, with the factors of ε_A and ε_M and the inclusion of the specific term $-\mu(1 - \text{av})r^{0,S}$ being purely for later convenience. By Lemma 2.1, the condition $\nabla \cdot \mathbf{b} = 0$ implies that each mode is divergence-free, and the conditions $\nabla \cdot \mathbf{b}^\ell = 0$ for $\ell \in \{F, I, S\}$ clearly imply that $\nabla \cdot \mathbf{b} = 0$, so for initial data of the form (44) the conditions (46) are equivalent to the assumed condition that $\nabla \cdot \mathbf{b}^0 = 0$.

Since, as shown above, the square of the H^n norm of \mathbf{V}^0 equals the sum of the squares of the H^n norms of its modes, the assumed uniform boundedness of $\|\mathbf{V}^0\|_n$ is equivalent to

$$\begin{aligned} & \|r^{0,S}\|_n + \varepsilon_M \|r^{0,I}\|_n + \|\mathbf{u}_h^{0,S}\|_n + \varepsilon_A \|\mathbf{u}_h^{0,F}\|_n + \|u_3^{0,S}\|_n + \varepsilon_M \|u_3^{0,I}\|_n \\ & + \|\mathbf{b}_h^{0,S}\|_n + \varepsilon_A \|\mathbf{b}_h^{0,F}\|_n + |\text{av} b_3^0| + \|-\mu(1 - \text{av})r^{0,S} + \varepsilon_A b_3^{0,F}\|_n \leq c. \end{aligned} \quad (47)$$

For $k = n - 1$ the sum of terms estimated in (25) is equivalent to the H^{n-1} norm of $(\mathcal{L}_A + \mu \mathcal{L}_M) \mathbf{V}$. Hence the assumed uniform boundedness of $\frac{1}{\varepsilon_A} \|(\mathcal{L}_A + \mu \mathcal{L}_M) \mathbf{V}^0\|_{n-1}$ becomes, for initial data \mathbf{V}^0 of the form (44) satisfying (46),

$$\begin{aligned} & \|\partial_z \mathbf{u}_h^{0,F}\|_{n-1} + \|\nabla_{\mathbf{h}} \cdot \mathbf{u}_h^{0,F}\|_n + \|\partial_z (\mathbf{b}_h^{0,F} - \Delta^{-1} \nabla_{\mathbf{h}} \partial_z r^{0,I})\|_{n-1} + \|\partial_z r^{0,I}\|_{n-1} \\ & + \|\partial_z u_3^{0,I}\|_{n-1} + \|-\varepsilon_M^{-1} (1 - \text{av}) r^{0,S} + b_3^{0,F} + \varepsilon_M^{-1} \Delta^{-1} \Delta_h r^{0,S} + \Delta^{-1} \Delta_h r^{0,I}\|_n \leq c. \end{aligned} \quad (48)$$

Since the expression $-\mu(1 - \text{av})r^{0,S}$ appearing in the last term in (47) can be estimated by the first term there it can be omitted from that last term, leaving $\varepsilon_A \|b_3^{0,F}\|_n$. Similarly, the expression $-\partial_z (\Delta^{-1} \Delta_h) \partial_z r^{0,I}$ in the third term of (48) can be omitted. Also the expression $-\varepsilon_M^{-1} (1 - \text{av}) r^{0,S} + \varepsilon_M^{-1} \Delta^{-1} \Delta_h r^{0,S}$ in the last term of (48) vanishes identically because $r^{0,S}$ is independent of z , leaving $\|b_3^{0,F} + \Delta^{-1} \Delta_h r^{0,I}\|_n$. The uniform bounds for that term and for $\varepsilon_M \|r^{0,I}\|_n$ from (47) imply the uniform boundedness of $\varepsilon_M \|b_3^{0,F}\|_n$, so the modified term $\varepsilon_A \|b_3^{0,F}\|_n$ for (47) can be omitted. However, since (48) only contains an $O(1)$ estimate for $\partial_z r^{0,I}$ not $r^{0,I}$ itself, it is not possible to omit the expression $\Delta^{-1} \Delta_h r^{0,I}$ from the term $\|b_3^{0,F} + \Delta^{-1} \Delta_h r^{0,I}\|_n$. Adding (47) and (48) and making the above-mentioned modifications shows that (45) is equivalent to the uniform boundedness of $\|\mathbf{V}^0\|_n + \frac{1}{\varepsilon_A} \|(\mathcal{L}_A + \mu \varepsilon_M) \mathbf{V}^0\|_{n-1}$ for initial data of the form (44). \square

We note that, by a slight notational exception, the fast part of b_3^0 in (44) is

$$(b_3^0)^F = -\mu(1 - \text{av})r^{0,S} + \varepsilon_A b_3^{0,F}, \quad (49)$$

showing that, to the leading $O(\mu)$ order, $(b_3^0)^F$ only depends on the slow part of r^0 .

Although Lemma 2.3 determines the most general initial data satisfying the conditions needed for the existence and convergence results, as noted in the introduction we need more to obtain the rate of convergence.

In order to allow the initial data to contain all modes but not interfere with the convergence rate, we will still assume that the initial data have the form (44), and that (46) holds, but will assume that

$$\text{all terms } w^{0,\ell} \text{ in (44) are uniformly bounded in } H^n \text{ and each term } w^{0,S} \text{ is fixed,} \quad (50)$$

which automatically implies that (45) holds. We allow the $w^{0,F}, w^{0,I}$ to depend on $(\varepsilon_A, \varepsilon_M)$ because we only bound the distance of the fast and intermediate modes to zero (i.e., estimate their Sobolev norms), not their distance to any non-trivial limits.

3. CONVERGENCE RATE ESTIMATES

Recall that “static” estimates are obtained by solving the PDE (4) for certain terms and bounding the norms of the result via (14), while “dynamic” estimates are obtained via energy estimates for the time evolution. We will estimate the size of the fast modes statically, the size of the intermediate modes statically and then dynamically, and the difference between the slow modes and the solution of the limit system dynamically, with earlier estimates used when deriving later ones. To optimize the use of earlier estimates we use the interpolation estimate (145) to obtain smaller estimates in weaker norms. Recall that $\mu = \frac{\varepsilon_A}{\varepsilon_M} = \varepsilon_M^\nu$ is assumed to be less than one. To see easily how the estimate to be obtained depends on the norm used, we introduce an increasing geometric sequence

$$\varepsilon_j := \varepsilon_M^{1+\nu-j\nu} \quad (51)$$

so that by (13), (11), $\varepsilon_0 = \varepsilon_A$, $\varepsilon_1 = \varepsilon_M$, $\varepsilon_j = \mu\varepsilon_{j+1}$, $\varepsilon_n \leq c$.

3.1. Static estimates

Static estimates will be obtained by solving the PDE system for $(\mathcal{L}_A + \mu\mathcal{L}_M)\mathbf{V}$ and using the uniform bounds (14) together with the standard Sobolev product and composition estimates

$$\|fg\|_j \leq c\|f\|_{n-1}\|g\|_j, \quad j = 0, \dots, n-1 \quad (52)$$

$$\|F(g)\|_{n-1} \leq C(\|g\|_{n-1}), \quad (53)$$

which will all be used henceforth without mention, plus the interpolation estimate (145). We will also need an estimate for the time integral of certain fast terms, which will be obtained similarly from the time integral of the PDE.

Theorem 3.1. *Assume that the basic conditions of Theorem 1.1 hold. Then the fast component \mathbf{V}^F satisfies the estimates*

$$\sup_{t \in [0, T]} [\|u_h^F\|_j + \|\nabla_h \cdot u_h^F\|_j + \|\partial_z u^F\|_j + \|\partial_z \mathbf{b}^F\|_j + \|b_h^F\|_j + \|(1 - \mathbf{a}_z)b_3^F\|_j + \|\mathbf{a}_z u_h^F\|_{j+1} + \|b_3^F + \mu\Delta^{-1}\Delta_h r\|_{j+1}] \leq C\varepsilon_j, \quad j = 0 \dots n-1, \quad (54)$$

$$\sup_{t \in [0, T]} \left[\left\| \int_0^t \mathbf{a}_z u_h^F dt' \right\|_n + \left\| \int_0^t u_h^F dt' \right\|_{n-1} + \left\| \int_0^t \nabla_h \cdot u_h^F dt' \right\|_{n-1} + \left\| \int_0^t \mathbf{a}_z (b_3^F + \mu\Delta^{-1}\Delta_h r^S) dt' \right\|_n \right] \leq C(T)\varepsilon_M^{1+\nu}, \quad (55)$$

and the intermediate component $(r^I, u_3^I) = ((1 - \mathbf{a}_z)r, (1 - \mathbf{a}_z)u_3)$ satisfies

$$\|(r^I, u_3^I)\|_j \leq c\|(\partial_z r^I, \partial_z u_3^I)\|_j \leq c\varepsilon_{j+1} = \frac{c\varepsilon_j}{\mu}, \quad j = 0 \dots n-1. \quad (56)$$

Note that the bounds in (55) for the time integrals of fast components are smaller than the bounds in (54) for those components themselves.

Proof. Taking the H^k norm of both sides of ε_A times (1) and using the interpolation bounds (145) to estimate the left side shows that

$$\|(\mathcal{L}_A + \mu\mathcal{L}_M)\mathbf{V}\|_j \leq c\varepsilon_M^{1+\nu-j\nu} = c\varepsilon_j. \quad (57)$$

Combining (57), (25) and the Poincaré inequality

$$\|(1 - \mathbf{a}_z)f\|_j \leq c\|\partial_z f\|_j, \quad (58)$$

yields (56). By (25), (57), and the fact that $u_3^F \equiv 0$,

$$\|\partial_z \mathbf{u}^F\|_j + \|\nabla_{\mathbf{h}} \cdot \mathbf{u}_{\mathbf{h}}^F\|_j + \|b_3^F + \mu\Delta^{-1}\Delta_h r\|_{j+1} \leq c\varepsilon_j, \quad j = 0, \dots, n-1. \quad (59)$$

Combining the definition (21) of the fast modes, the Poincaré inequality (58), the second inequality of (25), and (57) yields

$$\begin{aligned} & [\|\partial_z \mathbf{b}^F\|_j + \|\mathbf{b}_{\mathbf{h}}^F\|_j + \|(1 - \mathbf{a}_z)b_3^F\|_j] \leq c\|\partial_z \mathbf{b}^F\|_j \\ & \leq c[\|\partial_z (b_{\mathbf{h}}^F - \mu\Delta^{-1}\nabla_{\mathbf{h}}\partial_z r^I)\|_j + \mu\|\partial_z r^I\|_j + \|\partial_z (b_3^F + \mu\Delta^{-1}\Delta_h r)\|_j + \mu\|\partial_z r^I\|_j] \\ & \leq c\|(\mathcal{L}_A + \mu\mathcal{L}_M)\mathbf{V}\|_j \leq c\varepsilon_j, \quad j = 0, \dots, n-1. \end{aligned} \quad (60)$$

By the definition (21) of the fast modes

$$\mathbf{u}_{\mathbf{h}}^F = (1 - \mathbf{a}_z)\mathbf{u}_{\mathbf{h}}^F + \mathbf{a}_z \mathbf{u}_{\mathbf{h}}^F = (1 - \mathbf{a}_z)\mathbf{u}_{\mathbf{h}}^F + \nabla_{\mathbf{h}}\Delta_h^{-1} \mathbf{a}_z(\nabla_{\mathbf{h}} \cdot \mathbf{u}_{\mathbf{h}}^F). \quad (61)$$

Note that Δ_h is elliptic when applied to functions independent of z so

$$\|\nabla_{\mathbf{h}}\Delta_h^{-1} \mathbf{a}_z f\|_{j+1} \leq c\|f\|_j. \quad (62)$$

By (61), (58), (62), (25), and (57),

$$\begin{aligned} \|\mathbf{u}_{\mathbf{h}}^F\|_j + \|\mathbf{a}_z \mathbf{u}_{\mathbf{h}}^F\|_{j+1} & \leq \|(1 - \mathbf{a}_z)\mathbf{u}_{\mathbf{h}}^F\|_j + \|\nabla_{\mathbf{h}}\Delta_h^{-1} \mathbf{a}_z(\nabla_{\mathbf{h}} \cdot \mathbf{u}_{\mathbf{h}}^F)\|_{j+1} \\ & \leq c\|\partial_z \mathbf{u}_{\mathbf{h}}^F\|_j + c\|\nabla_{\mathbf{h}} \cdot \mathbf{u}_{\mathbf{h}}^F\|_j \leq c\|(\mathcal{L}_A + \mu\mathcal{L}_M)\mathbf{V}\|_j \leq c\varepsilon_j \quad j = 0, \dots, n-1. \end{aligned} \quad (63)$$

Combining (59), (60), and (63) yields (54).

The bounds on time integrals in (55) are obtained by integrating (1) with respect to time, which yields

$$\int_0^t (\mathcal{L}_A + \mu\mathcal{L}_M)\mathbf{V} = \varepsilon_A \left(A_0(\varepsilon_M \mathbf{V})\mathbf{V}|_0^t - \int_0^t (\varepsilon_M \mathbf{V}_t \cdot \nabla_{\mathbf{V}} A_0)\mathbf{V} + \int_0^t \sum_{i=1}^d A_i(\mathbf{V}) \mathbf{V}_{x_i} \right). \quad (64)$$

For the MHD system (4), only the variable r appears in the argument of A_0 , and (4a) shows that the time derivative of r is $O(\varepsilon_M^{-1})$, so the term $\varepsilon_M \mathbf{V}_t \cdot \nabla_{\mathbf{V}} A_0$ is uniformly bounded. This yields the estimate

$$\left\| \int_0^t (\mathcal{L}_A + \mu\mathcal{L}_M)\mathbf{V} dt' \right\|_{n-1} \leq c\varepsilon_A = c\varepsilon_M^{1+\nu}. \quad (65)$$

Since spatial operators commute with time integration, replacing every solution component in the second inequality of (25) with its time-integral from 0 to t and combining the result with the bound (65) yields (55) since $\mathbf{a}_z r = r^S$. \square

3.2. Intermediate system dynamic estimates

Theorem 3.2. *Assume that the conditions of the convergence part of Theorem 1.1 hold, and let $(r, \mathbf{u}, \mathbf{b})$ be the solution obtained for the MHD system (4). Then*

$$\|r^I\|_j + \|u_3^I\|_j \leq c \varepsilon_{\max(j,1)}, \quad j = 0, \dots, n-1. \quad (66)$$

The case $j = 0$ in (66) was already proven in Theorem 3.1. The remaining cases in (66) are an improvement over the corresponding cases in (56) by one factor of μ .

To prove (66) we will use the variables

$$\begin{aligned} \alpha &:= (1 - \mathbf{a}_z) \mathbf{V}_\alpha^{\text{no-div}} \cdot \mathbf{V} = (1 - \mathbf{a}_z)(r - \mu b_3 + \mu^3 \mathcal{C} \mathcal{Q} \Delta^{-1} \partial_z^2 b_3), \\ \beta &:= (1 - \mathbf{a}_z) \mathbf{V}_\beta \cdot \mathbf{V} = (1 - \mathbf{a}_z)(\mathcal{D} u_3 + \mu^2 \mathcal{C} \Delta^{-1} \partial_z \nabla_{\mathbf{h}} \cdot \mathbf{u}_{\mathbf{h}}), \end{aligned} \quad (67)$$

where the operators $\mathbf{V}_\alpha^{\text{no-div}}$, \mathbf{V}_β , \mathcal{C} , \mathcal{D} , and \mathcal{Q} were defined in (39)–(41). As a preliminary we will derive a system of PDEs satisfied by (α, β) , with remainder terms that are consistent with the desired estimate (66). The general idea is to apply each of the operators $\mathbf{V}_\alpha^{\text{no-div}}$ and \mathbf{V}_β to the PDE, note that by Lemma 2.2 the large terms of the result are $\partial_z \mathcal{Q}$ applied to the other operator, and calculate the form of the order one terms. To simplify that calculation we first simplify the original equations by moving to the right sides all terms whose H^j norms can be estimated by a constant times ε_j or the H^j norms of r^I and u_3^I using (54), (56), and (145). To do so we will use the formulas (21) and in particular their consequence

$$\mathbf{u} \cdot \nabla = (\mathbf{u}^S + \mathbf{u}^I + \mathbf{u}^F) \cdot \nabla = \mathbf{u}^S \cdot \nabla + u_3^I \partial_z + u_{\mathbf{h}}^F \cdot \nabla_{\mathbf{h}}. \quad (68)$$

Starting from the MHD equations (4), we replace the argument $\varepsilon_M r$ of ρ and a by $\varepsilon_M r^S$ or zero where possible and compensate by adding terms to the right sides of the equations, except that we retain a factor of $a(\varepsilon_M r^S)$ multiplying $\frac{1}{\varepsilon_M} \nabla_{\mathbf{h}} \cdot \mathbf{u}_{\mathbf{h}}$ because that expression will then cancel exactly when we build the time evolution equation for α , (72). In addition, we apply $1 - \mathbf{a}_z$ to the equations since that operator appears in all terms of the formulas (67) for α and β . Since slow modes are independent of z the operator $1 - \mathbf{a}_z$ can be moved past most coefficients. This yields

$$a(\varepsilon_M r^S)(\partial_t + (\mathbf{u}^S \cdot \nabla))r^I + \frac{a(\varepsilon_M r^S)\rho(\varepsilon_M r^S)}{\varepsilon_M} \partial_z u_3^I + \frac{a(\varepsilon_M r^S)}{\varepsilon_M} (1 - \mathbf{a}_z) \nabla_{\mathbf{h}} \cdot \mathbf{u}_{\mathbf{h}} = (1 - \mathbf{a}_z) R_1, \quad (69a)$$

$$\begin{aligned} \rho(\varepsilon_M r^S)(\partial_t + (\mathbf{u}^S \cdot \nabla))(1 - \mathbf{a}_z) \mathbf{u}_{\mathbf{h}} + \frac{a(\varepsilon_M r^S)\rho(\varepsilon_M r^S) - 1}{\varepsilon_M} \nabla_{\mathbf{h}} r^I \\ + (1 - \mathbf{a}_z) [\nabla_{\mathbf{h}} \frac{|\mathbf{b}|^2}{2} - (\mathbf{b} \cdot \nabla) \mathbf{b}_{\mathbf{h}}] - \frac{(1 - \mathbf{a}_z)}{\varepsilon_A} (\partial_z \mathbf{b}_{\mathbf{h}}^F - \nabla_{\mathbf{h}} \mathbf{b}_3^F - \mu \nabla_{\mathbf{h}} r^I) = (1 - \mathbf{a}_z) R_2, \end{aligned} \quad (69b)$$

$$\begin{aligned} \rho(\varepsilon_M r^S)(\partial_t + (\mathbf{u}^S \cdot \nabla))u_3^I + \frac{a(\varepsilon_M r^S)\rho(\varepsilon_M r^S)}{\varepsilon_M} \partial_z r^I - (\mathbf{b}_{\mathbf{h}}^S \cdot \nabla_{\mathbf{h}})(1 - \mathbf{a}_z) \mathbf{b}_3^F \\ = (1 - \mathbf{a}_z) R_3, \end{aligned} \quad (69c)$$

$$(\partial_t + (\mathbf{u}^S \cdot \nabla))(1 - \mathbf{a}_z) \mathbf{b}_3 - (\mathbf{b}_{\mathbf{h}}^S \cdot \nabla_{\mathbf{h}})u_3^I + (1 - \mathbf{a}_z) \frac{1}{\varepsilon_A} \nabla_{\mathbf{h}} \cdot \mathbf{u}_{\mathbf{h}} = (1 - \mathbf{a}_z) R_4, \quad (69d)$$

where the equation for $\mathbf{b}_{\mathbf{h}}$ has been omitted since it does not enter into α or β , and

$$\begin{aligned} R_1 &:= -a(\varepsilon_M r) (u_3^I \partial_z r + (u_{\mathbf{h}}^F \cdot \nabla_{\mathbf{h}})r) + (a(\varepsilon_M r^S) - a(\varepsilon_M r)) (\partial_t r + (\mathbf{u}^S \cdot \nabla)r) \\ &\quad + \frac{a(\varepsilon_M r^S)\rho(\varepsilon_M r^S) - a(\varepsilon_M r)\rho(\varepsilon_M r)}{\varepsilon_M} (\nabla_{\mathbf{h}} \cdot \mathbf{u}_{\mathbf{h}} + \partial_z u_3^I) + a(\varepsilon_M r^S) \frac{1 - \rho(\varepsilon_M r^S)}{\varepsilon_M} \nabla_{\mathbf{h}} \cdot \mathbf{u}_{\mathbf{h}}, \end{aligned} \quad (70a)$$

$$\begin{aligned} R_2 &:= -\rho(\varepsilon_M r) (u_3^I \partial_z \mathbf{u}_{\mathbf{h}} + (u_{\mathbf{h}}^F \cdot \nabla_{\mathbf{h}}) \mathbf{u}_{\mathbf{h}}) + (\rho(\varepsilon_M r^S) - \rho(\varepsilon_M r)) (\partial_t \mathbf{u}_{\mathbf{h}} + (\mathbf{u}^S \cdot \nabla) \mathbf{u}_{\mathbf{h}}) \\ &\quad + \frac{a(\varepsilon_M r^S)\rho(\varepsilon_M r^S) - a(\varepsilon_M r)\rho(\varepsilon_M r)}{\varepsilon_M} \nabla_{\mathbf{h}} r, \end{aligned} \quad (70b)$$

$$R_3 := -\rho(\varepsilon_M r) (u_3^I \partial_z u_3 + (\mathbf{u}_h^F \cdot \nabla_h) u_3) + (\rho(\varepsilon_M r^S) - \rho(\varepsilon_M r)) (\partial_t u_3 + (\mathbf{u}^S \cdot \nabla) u_3) \\ + \frac{a(\varepsilon_M r^S) \rho(\varepsilon_M r^S) - a(\varepsilon_M r) \rho(\varepsilon_M r)}{\varepsilon_M} \partial_z r - \mathbf{b}_h \cdot \partial_z \mathbf{b}_h^F + (\mathbf{b}_h^F \cdot \nabla_h) b_3^F, \quad (70c)$$

$$R_4 = - (u_3^I \partial_z b_3 + (\mathbf{u}_h^F \cdot \nabla_h) b_3) - (\nabla_h \cdot \mathbf{u}_h) b_3 + (\mathbf{b}_h^F \cdot \nabla_h) u_3. \quad (70d)$$

Wherever u_3^I appears undifferentiated in R_i , the H^j norm of the term in which it appears can be estimated by a constant times the H^j norm of u_3^I , for $j = 0, \dots, n-1$. Similarly, for any smooth function F and $j = 0, \dots, n-1$, $\|\frac{F(\varepsilon_M r^S) - F(\varepsilon_M r)}{\varepsilon_M}\|_j \leq c \|r^I\|_j$, and $\|[F(\varepsilon_M r^S) - F(\varepsilon_M r)](\partial_t \mathbf{V} + (\mathbf{u}^S \cdot \nabla) \mathbf{V})\|_j \leq c \varepsilon_M \|r^I\|_j (\|\mathbf{V}_t\|_{n-1} + \|\mathbf{V}\|_n) \leq c \|r^I\|_j$. By (54), terms containing \mathbf{u}_h^F , \mathbf{b}_h^F , $\nabla_h \cdot \mathbf{u}_h^F = \nabla_h \cdot \mathbf{u}_h$, or $\partial_z \mathbf{b}_h^F$ without further derivatives can be estimated in the H^j norm by $c \varepsilon_j$, for $0 \leq j \leq n-1$. Since these cases cover all the terms in the R_i ,

$$\sum_{i=1}^4 \|(1 - \mathbf{a}_z) R_i\|_j \leq \sum_{i=1}^4 \|R_i\|_j \leq c (\|r^I, u_3^I\|_j + \varepsilon_j), \quad j = 0, \dots, n-1. \quad (71)$$

To obtain the evolution equation for α , subtract $\mu a(\varepsilon_M r^S)$ times (69d) from (69a) and add $a(\varepsilon_M r^S) \mu^3 \mathcal{C} \mathcal{Q} \Delta^{-1} \partial_z^2$ applied to (69d) to the result. Then commute $\mathcal{C} \mathcal{Q} \Delta^{-1} \partial_z^2$ past $\mathbf{u}^S \cdot \nabla$, make the coefficient of the large terms that do not cancel be $a \rho$ everywhere while compensating via terms on the right side, force the function to which $\mu(\mathbf{b}_h^S \cdot \nabla_h)$ is applied to be $\frac{\beta}{\sqrt{1+\mu^2}}$ for reasons to be explained later and again compensate on the right side, and use the identity $1 = \mathcal{Q} \mathcal{D}$ from (41) that makes the large terms exactly involve β , as we know from (37) that they must. This yields

$$a(\varepsilon_M r^S) (\partial_t + (\mathbf{u}^S \cdot \nabla)) \alpha + \frac{\mu}{\sqrt{1+\mu^2}} (\mathbf{b}_h^S \cdot \nabla_h) \beta + \frac{a(\varepsilon_M r^S) \rho(\varepsilon_M r^S)}{\varepsilon_M} \mathcal{Q} \partial_z \beta = R_{\alpha,1} + R_{\alpha,2} + R_{\alpha,3} + R_{\alpha,4}, \quad (72)$$

where

$$R_{\alpha,1} := (1 - \mathbf{a}_z) R_1 - \mu a(\varepsilon_M r^S) (1 - \mathbf{a}_z) R_4 + \mu^3 a(\varepsilon_M r^S) \mathcal{C} \mathcal{Q} \Delta^{-1} \partial_z^2 R_4 \quad (73)$$

comes from the right sides of the modified equations (69),

$$R_{\alpha,2} := -\mu^3 a(\varepsilon_M r^S) [\mathcal{C} \mathcal{Q} \Delta^{-1} \partial_z^2, \mathbf{u}^S] \cdot \nabla (1 - \mathbf{a}_z) b_3^F \quad (74)$$

comes from commuting the operator applied to (69d) past the coefficient u_3^S ,

$$R_{\alpha,3} := -\mu \varepsilon_M \frac{a(\varepsilon_M r^S) - 1}{\varepsilon_M} (\mathbf{b}_h^S \cdot \nabla_h) u_3^I + \mu^3 a(\varepsilon_M r^S) \mathcal{C} \mathcal{Q} \Delta^{-1} \partial_z [(\mathbf{b}_h^S \cdot \nabla_h) \partial_z u_3^I] \\ + \mu^3 a(\varepsilon_M r^S) \frac{\rho(\varepsilon_M r^S) - 1}{\varepsilon_A} \mathcal{C} \mathcal{Q} \Delta^{-1} \partial_z^2 (\nabla_h \cdot \mathbf{u}_h) \quad (75)$$

comes from adding compensating terms to the right side and moving entire terms there, and

$$R_{\alpha,4} := \mu (\mathbf{b}_h^S \cdot \nabla_h) \left(\frac{\beta}{\sqrt{1+\mu^2}} - u_3^I \right) \quad (76)$$

comes from forcing the term involving $\mathbf{b}_h^S \cdot \nabla_h$ to have the desired form, and will be rearranged further later.

Similarly, to obtain the evolution equation for β , add $\mu^2 \mathcal{C} \Delta^{-1} \partial_z \nabla_h \cdot$ applied to (69b) and \mathcal{D} applied to (69c), and rearrange terms in similar fashion as for (72). Then force the function to which $\mathbf{b}_h^S \cdot \nabla_h$ is applied on the left side of the equation to be $\frac{\mu}{\sqrt{1+\mu^2}} \alpha$ for reasons to be explained later, and subtract appropriate constants from the

factors appearing inside commutators since that does not affect their value, in order to facilitate estimate the size of the resulting terms. Also, to ensure the symmetry of the resulting system for α and β , multiply the large term $\frac{\mu^2}{\varepsilon_A}(1 - \mathbf{a}_z)\mathcal{C}\Delta^{-1}\partial_z\nabla_{\mathbf{h}}\cdot(\partial_z\mathbf{b}_{\mathbf{h}} - \nabla_{\mathbf{h}}\mathbf{b}_3 - \mu\nabla_{\mathbf{h}}r)$ appearing on the left side of the equation by $a(\varepsilon_M r^S)\rho(\varepsilon_M r^S)$ and compensate by adding $a(\varepsilon_M r^S)\rho(\varepsilon_M r^S) - 1$ times that term to the right side. By (41), the resulting large term exactly involves α . This yields

$$\rho(\varepsilon_M r^S)(\partial_t + (\mathbf{u}^S \cdot \nabla))\beta + \frac{\mu}{\sqrt{1+\mu^2}}(\mathbf{b}_{\mathbf{h}}^S \cdot \nabla_{\mathbf{h}})\alpha + \frac{a(\varepsilon_M r^S)\rho(\varepsilon_M r^S)}{\varepsilon_M}\mathcal{Q}\partial_z\alpha = R_{\beta,1} + R_{\beta,2} + R_{\beta,3} + R_{\beta,4}, \quad (77)$$

where

$$R_{\beta,1} := \mathcal{D}(1 - \mathbf{a}_z)R_3 + \mu^2\mathcal{C}\Delta^{-1}\partial_z\nabla_{\mathbf{h}}\cdot(1 - \mathbf{a}_z)R_2$$

comes from the right sides of (69),

$$\begin{aligned} R_{\beta,2} &= -\mu^2\varepsilon_M\left[\frac{\mathcal{D}-1}{\mu^2}, \frac{\rho(\varepsilon_M r^S)-1}{\varepsilon_M}\right](\partial_t + (\mathbf{u}^S \cdot \nabla))u_3^I - \mu^2\rho(\varepsilon_M r^S)\left[\frac{\mathcal{D}-1}{\mu^2}, \mathbf{u}^S\right]\cdot\nabla u_3^I - \mu^2\left[\frac{\mathcal{D}-1}{\mu^2}, \frac{a(\varepsilon_M r^S)\rho(\varepsilon_M r^S)-1}{\varepsilon_M}\right]\partial_z r^I \\ &\quad - \mu^2\varepsilon_M[\mathcal{C}\Delta^{-1}\partial_z\nabla_{\mathbf{h}}, \frac{\rho(\varepsilon_M r^S)-1}{\varepsilon_M}] \cdot (\partial_t + (\mathbf{u}^S \cdot \nabla))(1 - \mathbf{a}_z)u_{\mathbf{h}}^F - \mu^2\rho(\varepsilon_M r^S)[\mathcal{C}\Delta^{-1}\partial_z\nabla_{\mathbf{h}}, \mathbf{u}^S]\nabla(1 - \mathbf{a}_z)u_{\mathbf{h}}^F \\ &\quad - \mu^2[\mathcal{C}\Delta^{-1}\partial_z\nabla_{\mathbf{h}}, \frac{a(\varepsilon_M r^S)\rho(\varepsilon_M r^S)-1}{\varepsilon_M}] \cdot \nabla_{\mathbf{h}}r^I \end{aligned}$$

comes from commuting the operators applied to (69c) and (69b) past coefficients in those equations,

$$\begin{aligned} R_{\beta,3} &:= -\mu^2\frac{a(\varepsilon_M r^S)\rho(\varepsilon_M r^S)-1}{\varepsilon_M}\mathcal{C}\Delta^{-1}\Delta_{\mathbf{h}}\partial_z r^I - \mu^2\mathcal{C}\Delta^{-1}\partial_z\nabla_{\mathbf{h}}\cdot(\nabla_{\mathbf{h}}\frac{|\mathbf{b}|^2}{2} - (\mathbf{b}\cdot\nabla)\mathbf{b}_{\mathbf{h}}) \\ &\quad - \mu^2\frac{a(\varepsilon_M r^S)\rho(\varepsilon_M r^S)-1}{\varepsilon_A}\mathcal{C}\partial_z(b_3^F + \mu\Delta^{-1}\Delta_{\mathbf{h}}r) \end{aligned}$$

comes from moving terms to the right side, balancing a term added on the left side, and using (26), and

$$R_{\beta,4} := \frac{\mu}{\sqrt{1+\mu^2}}(\mathbf{b}_{\mathbf{h}}^S \cdot \nabla_{\mathbf{h}})\alpha + \mathcal{D}(\mathbf{b}_{\mathbf{h}}^S \cdot \nabla_{\mathbf{h}})(1 - \mathbf{a}_z)b_3^F \quad (78)$$

comes from forcing the term involving $\mathbf{b}_{\mathbf{h}}^S \cdot \nabla_{\mathbf{h}}$ to have the desired form, and will be rearranged further later.

We now estimate the terms $R_{\alpha,i}$ and $R_{\beta,i}$. Since the operators applied to the R_i in $R_{\alpha,1}$ and $R_{\beta,1}$ are all bounded, (71) implies that

$$\|R_{\alpha,1}\|_j + \|R_{\beta,1}\|_j \leq c(\|(r^I, u_3^I)\|_j + \varepsilon_j), \quad j = 0, \dots, n-1. \quad (79)$$

The terms in $R_{\alpha,2}$ and $R_{\beta,2}$ all involve commutators, and the following lemma says that the commutator gains one derivative, which in many cases is a vital improvement.

Lemma 3.3. ([MS01, Lemma 2.5]) *Let $\widehat{P}(k, l, m)$ be homogeneous of degree zero and real analytic for $(k, l, m) \neq (0, 0, 0)$, and let P be the Fourier multiplier operator defined by $\widehat{P}f = \widehat{P}f$. Then for all $n \geq 3$, $f \in H^n(\mathbb{T}^3)$, $j \in 1, \dots, n$, and $g \in H^{j-1}$,*

$$\|[P, f]g\|_j \leq c\|f\|_n\|g\|_{j-1}. \quad (80)$$

The constant-coefficient pseudo-differential operators appearing in the commutators in $R_{\alpha,2}$ and $R_{\beta,2}$ are all homogeneous of degree zero and bounded uniformly in μ (in particular (88) below implies a bound on $\frac{\mathcal{D}-1}{\mu^2}$), and they are real analytic for $(k, l, m) \neq (0, 0, 0)$ since the denominators in the formulas for \mathcal{C} , \mathcal{D} , and \mathcal{Q} in (40), (41) are positive for $\mu < 1$, so they satisfy the conditions of Lemma 3.3. The expressions to which the commutators are applied have one of two forms: either they consist of a single spatial derivative applied to a fast or intermediate component that is estimated in (54) or (56), or they are some component of $(\partial_t + (\mathbf{u}^S \cdot \nabla))\mathbf{V}$. In

the former case the expression contains a factor μ^p with $p \geq 2$, so by Lemma 3.3 its H^j norm for $j = 0, \dots, n-1$ can be estimated by

$$c\mu^2 \|\nabla((1 - \mathbf{a}_z)b_3^F, u_3^I, r^I, u_h^F)\|_{\max(0, j-1)} \leq c\mu^2 \|((1 - \mathbf{a}_z)b_3^F, u_3^I, r^I, u_h^F)\|_{\max(1, j)} \leq c\mu^2 \varepsilon_{j+2} \leq c\varepsilon_j.$$

In the latter case the expression contains the factor $\mu^2 \varepsilon_M$ so by the interpolation bounds (145) its H^j norm for $j = 0, \dots, n-1$ can be estimated by $c\mu^2 \varepsilon_M (\|\mathbf{V}_t\|_{j-1} + c) \leq c\mu^2 \varepsilon_M \mu^{-j} \leq c\mu \varepsilon_j$. Hence

$$\|R_{\alpha,2}\|_j + \|R_{\beta,2}\|_j \leq c\varepsilon_j, \quad j = 0, \dots, n-1. \quad (81)$$

To estimate the terms in $R_{\alpha,3}$ and $R_{\beta,3}$ note that those terms either have a factor $\mu^2 \varepsilon_M$ that is smaller than all ε_j , have a zeroth-order pseudo-differential operator applied to $\partial_z r^I$, \mathbf{b}_z (after applying the z -derivative in the expression to some factor of \mathbf{b}), $\nabla_h \cdot \mathbf{u}_h$, or $\partial_z(b_3^F + \mu\Delta^{-1}\Delta_h r)$, all of which have their H^j norms estimated in (54) or (56), or have a zeroth-order pseudo-differential operator applied to $\Delta^{-1}\partial_z[(\mathbf{b}_h \cdot \nabla_h)\partial_z u_3^I]$. In the first case using the uniform estimate (14) shows that the H^{n-1} norm is bounded by $c\varepsilon_0$. In the middle cases each term contains a factor of μ^2 so its H^j norm is bounded by $c\mu^2 \varepsilon_{j+1} \leq c\varepsilon_j$. In the final case, since $\|\Delta^{-1}\partial_z f\|_j \leq c\|f\|_{\max(j-1,0)}$ and the term contains a factor of μ^3 , it is bounded by $c\mu^3 \|\nabla\partial_z u_3^I\|_{\max(j-1,0)} \leq c\mu^3 \|\partial_z u_3^I\|_{\max(j,1)} \leq c\mu^3 \varepsilon_{j+2} \leq c\mu \varepsilon_j$. Together, these yield

$$\|R_{\alpha,3}\|_j + \|R_{\beta,3}\|_j \leq c\varepsilon_j, \quad j = 0, \dots, n-1. \quad (82)$$

For $j \leq n-2$ the terms $R_{\alpha,4}$ and $R_{\beta,4}$ can be estimated by using the fact that (54) and (56) imply that

$$\mu\|\mathbf{V}_*^F\|_{j+1} + \mu^2\|\mathbf{V}^I\|_{j+1} \leq c(\mu\varepsilon_{j+1} + \mu^2\varepsilon_{j+2}) \leq c\varepsilon_j, \quad j = 0, \dots, n-2, \quad (83)$$

where \mathbf{V}_*^F means all components of \mathbf{V}^F except $\mathbf{a}_z b_3^F$, which is not estimated in (54). The estimate $\|R_{\alpha,4}\|_j + \|R_{\beta,4}\|_j \leq c\varepsilon_j$ can therefore be obtained for $j \leq n-2$ by using the formulas for α and β in (67) together with the facts that $\frac{\mathcal{D}-1}{\mu^2}$ is a bounded zeroth-order pseudo-differential operator and that

$$b_3^F = (b_3^F + \mu\Delta^{-1}\Delta_h r) + \mu\Delta^{-1}\partial_z^2 r^I - \mu(r - avr) \quad (84)$$

plus estimates similar to those used for $R_{\alpha,3}$ and $R_{\beta,3}$.

However, (83) is not valid for $j = n-1$ because (54) and (56) do not hold for $j = n$. We therefore rearrange $R_{\alpha,4}$ and $R_{\beta,4}$ to be linear combinations of the terms

$$\nabla_h \cdot \mathbf{u}_h, \quad \partial_z \mathbf{u}^F, \quad \partial_z \mathbf{b}, \quad \nabla(b_3^F + \Delta^{-1}\Delta_h r), \quad \partial_z r \quad \text{and} \quad \partial_z u_3 \quad (85)$$

whose H^j norms are estimated in (54) or (56) even though they involve a first derivative of \mathbf{V} ; in addition a factor of μ must be present multiplying the terms $\partial_z(r^I, u_3^I)$ to compensate for the extra factor of $\frac{1}{\mu}$ in (56). Substituting (84) into the formula for α in (67) and solving the result for r^I yields

$$(1 + \mu^2)r^I = \alpha + \mu(1 - \mathbf{a}_z)(b_3^F + \mu\Delta^{-1}\Delta_h r) + \mu^2\Delta^{-1}\partial_z^2 r^I - \mu^3\mathcal{C}\mathcal{Q}\Delta^{-1}\partial_z^2 r^I. \quad (86)$$

Applying $(1 - \mathbf{a}_z)$ to (84), which turns the final $-\mu(r - avr)$ into $-\mu r^I$, dividing (86) by $1 + \mu^2$ and substituting the result for that final r^I , and substituting the result into (78) shows that $R_{\beta,4}$ equals

$$\mu\left(\frac{1}{\sqrt{1+\mu^2}} - \frac{1}{1+\mu^2}\mathcal{D}\right)(\mathbf{b}_h \cdot \nabla_h)\alpha \quad (87)$$

plus a sum of terms involving operators of order zero applied to the expressions in (85) whose H^j norms can be bounded by $c\varepsilon_j$ using (54) and (56). To estimate (87) we use the identity

$$\mathcal{D} = \sqrt{1 + \mu^2} - \frac{4\mu^2\Delta^{-1}\partial_z^2}{\left(\sqrt{\mu^2(2-4\Delta^{-1}\partial_z^2)+\mu^4+1+\mu^2+1}\right)\left(\sqrt{2}\sqrt{\sqrt{\mu^2(2-4\Delta^{-1}\partial_z^2)+\mu^4+1+\mu^2+1+2\sqrt{\mu^2+1}}}\right)} \quad (88)$$

derived from (41), which makes the constant term in (87) cancel. The other term in (88) contains a factor of ∂_z multiplied by an operator of order -1 , the z -derivative of all the constituents of α are estimated in (54) and (56), and an overall factor of μ is present in (88), so the H^j norm of $R_{\beta,4}$ is bounded by a constant times ε_j even for $j = n - 1$. Similarly, substituting formula (67) for β into the definition (76) of $R_{\alpha,4}$ and substituting (88) into the result makes the constant term from (88) cancel. All remaining terms involve the expressions from (85) and so can be estimated by a constant times ε_j since an overall factor of μ is present. This yields

$$\|R_{\alpha,4}\|_j + \|R_{\beta,4}\|_j \leq c\varepsilon_j, \quad j = 0, \dots, n-1. \quad (89)$$

Proof of Theorem 3.2. The system (72), (77) has the form of the Klainerman-Majda two-scale theory. Moreover, (67) plus estimates similar to those above yield

$$\|\alpha - r^I\|_j + \|\beta - u_3^I\|_j \leq c\varepsilon_j, \quad j = 0, \dots, n-1. \quad (90)$$

Together with the estimates (79), (81), (82), and (89), (90) shows that the H^j norm of the right sides of those equations is bounded by a constant times $\|(\alpha, \beta)\|_j + \varepsilon_j$. Hence the standard Klainerman-Majda energy estimates show that

$$\max_{0 \leq t \leq T} \|(\alpha, \beta)\|_j \leq c(\|(\alpha(0), \beta(0))\|_j + \varepsilon_j), \quad j = 0, \dots, n-1. \quad (91)$$

The initial data (44), (50) satisfies $\|\mathbf{V}^I\|_{n-1}|_{t=0} \leq c\varepsilon_M = c\varepsilon_1$, so (90) implies that $\|(\alpha(0), \beta(0))\|_j \leq c\varepsilon_1$ and hence (91) implies that $\max_{0 \leq t \leq T} \|(\alpha, \beta)\|_j \leq c \max(\varepsilon_j, \varepsilon_1)$ for $j = 0, \dots, n-1$. Using (90) once more yields (66). \square

3.3. Equations and estimates for horizontal components of the slow mode

Like for the intermediate mode dynamic estimates, estimating the difference between the slow modes of the solution to the original system and the solution of the limit system requires PDEs for the exact zero eigenspace of the operator $\mathcal{L}_A + \mu\mathcal{L}_M$. The horizontal velocity and magnetic field slow modes belong to that eigenspace, so we will write the equations for those modes as the limit equations plus error terms, by applying the projection \mathbb{P} onto the slow horizontal modes to the original system, expanding all dependent variables into fast, intermediate, and slow modes, and moving all terms except those involving purely slow modes to the right sides of the equations. The remaining slow modes will be treated in the following subsection.

Theorem 3.4. *Assume that the conditions of the convergence part of Theorem 1.1 hold. Let $(r, \mathbf{u}, \mathbf{b})$ be the solution of the MHD system (4), and let $(\bar{r}, \bar{\mathbf{u}}_h, \bar{u}_3, \bar{\mathbf{b}}_h, \bar{b}_3)$ be the solution of the limit system (15) whose initial data is the limit $(r^{0,S}, \mathbf{u}^{0,S}, \mathbf{b}^{0,S} + \mu_{\text{lim}}(1 - \text{av})r^{0,S})$ of the initial data (44) of the original system. Then*

$$\|\mathbf{u}_h^S - \bar{\mathbf{u}}_h\|_{n-2} + \|\mathbf{b}_h^S - \bar{\mathbf{b}}_h\|_{n-2} \leq c\varepsilon_M. \quad (92)$$

Before proving Theorem 3.4 we need to derive appropriate equations. Since all slow modes contain the averaging operator \mathbf{a}_z in the z direction, it will be convenient to write the equations in conservation form, so that derivatives with respect to z disappear when \mathbf{a}_z is applied. In particular,

$$\mathbf{a}_z(\mathbf{w} \cdot \nabla f + (\nabla \cdot \mathbf{w})f) = \mathbf{a}_z \nabla \cdot (f\mathbf{w}) = \mathbf{a}_z \nabla_h \cdot (f\mathbf{w}_h) = \mathbf{a}_z(\mathbf{w}_h \cdot \nabla_h f + (\nabla_h \cdot \mathbf{w}_h)f). \quad (93)$$

Adding the continuity equation (138a) to the momentum equation (4b) yields

$$\partial_t(\rho\mathbf{u}) + \mathbf{u} \cdot \nabla(\rho\mathbf{u}) + (\nabla \cdot \mathbf{u})(\rho\mathbf{u}) + \nabla\Phi - \mathbf{b} \cdot \nabla\mathbf{b} = \varepsilon_A^{-1}(\partial_z\mathbf{b} - \nabla b_3), \quad (94)$$

where Φ is some scalar-valued function defined on \mathbb{T}^3 , i.e., periodic in x, y, z . Apply $\mathcal{P}_h^{\text{div}} \mathbf{a}_z$ to the first two components of (94) and \mathbf{a}_z to the first two components of (4c), and simplify the result by using the identity (93) not only with $\mathbf{w} = \mathbf{u}$ but also with $\mathbf{w} = \mathbf{b}$ since the constraint $\nabla \cdot \mathbf{b} = 0$ implies that $\mathbf{b} \cdot \nabla f$ equals $\nabla \cdot (f\mathbf{b})$. The

resulting equations can be further simplified by using the definitions (21) of the modes to obtain the identities $\mathbf{a}_z \mathbf{b}_h = \mathbf{b}_h^S$ and $\mathbf{a}_z(\rho \mathbf{u}_h) = \mathbf{a}_z(\rho \mathbf{u}_h^S) + \mathbf{a}_z(\rho \mathbf{u}_h^F) = (\mathbf{a}_z \rho) \mathbf{u}_h^S + \mathbf{a}_z(\rho \mathbf{u}_h^F)$, and then using the facts that $\rho = 1 + \varepsilon_M r$ and $\mathbf{a}_z \mathbf{u}_h^F$ is a gradient to obtain $\mathcal{P}_h^{\text{div}} \mathbf{a}_z(\rho \mathbf{u}_h^F) = \varepsilon_M \mathcal{P}_h^{\text{div}} \mathbf{a}_z(r \mathbf{u}_h^F)$. This yields

$$\partial_t \mathcal{P}_h^{\text{div}}((\mathbf{a}_z \rho) \mathbf{u}_h^S) + \mathcal{P}_h^{\text{div}} \mathbf{a}_z \nabla_{h^\cdot} (\rho \mathbf{u}_h \otimes \mathbf{u}_h - \mathbf{b}_h \otimes \mathbf{b}_h) = -\varepsilon_M \partial_t \mathcal{P}_h^{\text{div}} \mathbf{a}_z(r \mathbf{u}_h^F), \quad (95a)$$

$$\partial_t \mathbf{b}_h^S + \mathbf{a}_z \nabla_{h^\cdot} (\mathbf{u}_h \otimes \mathbf{b}_h - \mathbf{b}_h \otimes \mathbf{u}_h) = 0, \quad (95b)$$

where the tensor product \otimes follows the convention that $\nabla_{h^\cdot}(\mathbf{u}_h \otimes \mathbf{b}_h) = \mathbf{u}_h \nabla_{h^\cdot} \mathbf{b}_h + (\nabla_{h^\cdot} \mathbf{u}_h) \mathbf{b}_h$. Since $\partial_t \mathcal{P}_h^{\text{div}}((\mathbf{a}_z \rho) \mathbf{u}_h^S) = \partial_t \mathbf{u}_h^S + \varepsilon_M \partial_t \mathcal{P}_h^{\text{div}}((\mathbf{a}_z r) \mathbf{u}_h^S)$, (95) together with the bounds (14) and (13) and the relation (11) show that

$$\|\partial_t \mathbf{u}_h^S\|_{n-1} + \|\partial_t \mathbf{b}_h^S\|_{n-1} \leq c. \quad (96)$$

Recalling that \mathbf{u}_h and \mathbf{b}_h have no intermediate part, we separate them into their fast and slow parts in the tensor products in (95a):

$$\begin{aligned} \mathbf{a}_z \nabla_{h^\cdot} (\rho \mathbf{u}_h \otimes \mathbf{u}_h - \mathbf{b}_h \otimes \mathbf{b}_h) &= \mathbf{a}_z \nabla_{h^\cdot} (\rho \mathbf{u}_h^S \otimes \mathbf{u}_h^S - \mathbf{b}_h^S \otimes \mathbf{b}_h^S) \\ &\quad + \mathbf{a}_z \nabla_{h^\cdot} \left((\rho \mathbf{u}_h^S \otimes \mathbf{u}_h^F - \mathbf{b}_h^S \otimes \mathbf{b}_h^F) + \text{trsp} \right) + \mathbf{a}_z \nabla_{h^\cdot} (\rho \mathbf{u}_h^F \otimes \mathbf{u}_h^F - \mathbf{b}_h^F \otimes \mathbf{b}_h^F) \end{aligned} \quad (97)$$

where trsp denotes the tensor transpose. By (23)–(24), the slow parts $(\mathbf{u}_h^S, \mathbf{b}_h^S)$ are independent of z and divergence-free, so the first term on the right side of (97) simplifies to $\mathbf{u}_h^S \cdot \nabla_{h^\cdot} ((\mathbf{a}_z \rho) \mathbf{u}_h^S) - \mathbf{b}_h^S \cdot \nabla_{h^\cdot} \mathbf{b}_h^S$. Also, since \mathbf{b}_h^S is independent of z while (21) shows that $\mathbf{a}_z \mathbf{b}_h^F = 0$, the expression $-\mathbf{a}_z \nabla_{h^\cdot} (\mathbf{b}_h^S \otimes \mathbf{b}_h^F)$ appearing in the second term on the right in (97) is identically zero. Next, we can drop the $\mathcal{P}_h^{\text{div}}$ operator from (95a) at the cost of adding a term $\nabla_{h^\cdot} \theta(t, x, y)$ to that equation, since a 2-vector is in the kernel of $\mathcal{P}_h^{\text{div}}$ if and only if it is a horizontal gradient. Thus, (95a) becomes

$$\begin{aligned} \partial_t ((\mathbf{a}_z \rho) \mathbf{u}_h^S) + \mathbf{u}_h^S \cdot \nabla_{h^\cdot} ((\mathbf{a}_z \rho) \mathbf{u}_h^S) - \mathbf{b}_h^S \cdot \nabla_{h^\cdot} \mathbf{b}_h^S \\ = -\varepsilon_M \partial_t \mathbf{a}_z(r \mathbf{u}_h^F) - \mathbf{a}_z \nabla_{h^\cdot} (\rho \mathbf{u}_h^S \otimes \mathbf{u}_h^F + \text{trsp}) - \mathbf{a}_z \nabla_{h^\cdot} (\rho \mathbf{u}_h^F \otimes \mathbf{u}_h^F - \mathbf{b}_h^F \otimes \mathbf{b}_h^F) + \nabla_{h^\cdot} \theta. \end{aligned}$$

The second term on the right side is a “slow-fast” product, which can be rewritten using time-integrated variable

$$\mathbf{A}(t, \cdot) := \int_0^t \mathbf{a}_z \mathbf{u}_h^F dt', \quad (98)$$

as the time derivative of a small term plus a small term, since (55) shows that

$$\|\mathbf{A}\|_n \leq c \varepsilon_M^{1+\nu}. \quad (99)$$

For example,

$$\begin{aligned} \mathbf{a}_z(\rho \mathbf{u}_h^S \otimes \mathbf{u}_h^F) &= \mathbf{a}_z(\mathbf{u}_h^S \otimes \mathbf{u}_h^F) + \varepsilon_M \mathbf{a}_z(r \mathbf{u}_h^S \otimes \mathbf{u}_h^F) = \mathbf{u}_h^S \otimes (\mathbf{a}_z \mathbf{u}_h^F) + \varepsilon_M \mathbf{a}_z(r \mathbf{u}_h^S \otimes \mathbf{u}_h^F) \\ &= \partial_t (\mathbf{u}_h^S \otimes \mathbf{A}) - (\partial_t \mathbf{u}_h^S) \otimes \mathbf{A} + \varepsilon_M \mathbf{a}_z(r \mathbf{u}_h^S \otimes \mathbf{u}_h^F). \end{aligned}$$

Hence we obtain

$$(\partial_t + (\mathbf{u}_h^S \cdot \nabla_{h^\cdot}))((\mathbf{a}_z \rho) \mathbf{u}_h^S) - \mathbf{b}_h^S \cdot \nabla_{h^\cdot} \mathbf{b}_h^S = \partial_t \Xi_1 + \xi_1 + \nabla_{h^\cdot} \theta, \quad \nabla_{h^\cdot} \mathbf{u}_h^S = 0 \quad (100)$$

with

$$\begin{aligned} \Xi_1 &:= -\varepsilon_M \mathbf{a}_z(r \mathbf{u}_h^F) - \nabla_{h^\cdot} (\mathbf{u}_h^S \otimes \mathbf{A} + \text{trsp}), \\ \xi_1 &:= \nabla_{h^\cdot} \left\{ (\partial_t \mathbf{u}_h^S) \otimes \mathbf{A} + \text{trsp} \right\} - \varepsilon_M \mathbf{a}_z \nabla_{h^\cdot} (r \mathbf{u}_h^S \otimes \mathbf{u}_h^F + \text{trsp}) - \mathbf{a}_z \nabla_{h^\cdot} (\rho \mathbf{u}_h^F \otimes \mathbf{u}_h^F - \mathbf{b}_h^F \otimes \mathbf{b}_h^F). \end{aligned}$$

The bound (54) together with the constraint (13), the relation (11) between the parameters and the definition (51) of the ε_j implies that $\|u_h^F\|_{n-1} + \|b_h^F\|_{n-1} \leq c\varepsilon_M^\nu$. Using that estimate, the time-derivative bound (96), estimate (99) for \mathbf{A} , the formula $\rho = 1 + \varepsilon_M r$, and Corollary E.3 yields the estimates

$$\|\Xi_1\|_{n-1} + \|\xi_1\|_{n-2} \leq c\varepsilon_M^{1+\nu}. \quad (101)$$

Applying the same ideas to the slow magnetic equation (95b), and in particular noting that $\mathbf{a}_z(u_h^S \otimes b_h^F + \text{trsp}) = 0$ yields

$$(\partial_t + (u_h^S \cdot \nabla_h))b_h^S - b_h^S \cdot \nabla_h u_h^S = \partial_t \Xi_{b_h^S} + \xi_{b_h^S}, \quad \nabla_h \cdot b_h^S = 0, \quad (102)$$

with

$$\Xi_{b_h^S} := -\nabla_h \cdot [\mathbf{A} \otimes b_h^S - \text{trsp}], \quad \xi_{b_h^S} := \nabla_h \cdot [\mathbf{A} \otimes \partial_t b_h^S - \text{trsp}] - \mathbf{a}_z \cdot \nabla_h \cdot (u_h^F \otimes b_h^F - \text{trsp}). \quad (103)$$

The same bounds as for (101) show that

$$\|\Xi_{b_h^S}\|_{n-1} + \|\xi_{b_h^S}\|_{n-2} \leq c\varepsilon_M^{1+\nu}. \quad (104)$$

The equation for the time evolution of u_h^S can be further simplified using an equation for the time evolution of r^S . By the first part of (24) and the facts from (21) that $u_h^I = 0$ and $\mathbf{a}_z r = r^S$, the vertically-averaged equation (147) simplifies to

$$\partial_t r^S + u_h^S \cdot \nabla_h (r^S) = -\mathbf{a}_z \cdot \nabla_h \cdot (r u_h^F) - \varepsilon_M^{-1} \nabla_h \cdot (\mathbf{a}_z u_h^F). \quad (105)$$

Using the time-integrated variable \mathbf{A} from (98), (105) can be rewritten as

$$(\partial_t + (u_h^S \cdot \nabla_h))(\varepsilon_M r^S) = -\varepsilon_M \mathbf{a}_z \cdot \nabla_h \cdot (r u_h^F) - \partial_t \nabla_h \cdot \mathbf{A}. \quad (106)$$

Subtracting u_h^S times (106) from (100) noting that $\mathbf{a}_z \rho = 1 + \varepsilon_M r^S$, and rewriting the term $u_h^S \partial_t \nabla_h \cdot \mathbf{A}$ on the right side of the result as $\partial_t (u_h^S \nabla_h \cdot \mathbf{A}) - (\partial_t u_h^S) \nabla_h \cdot \mathbf{A}$ yields

$$(1 + \varepsilon_M r^S)(\partial_t + (u_h^S \cdot \nabla_h))u_h^S - b_h^S \cdot \nabla_h b_h^S = \partial_t \Xi_2 + \xi_2 + \nabla_h \theta_2, \quad \nabla_h \cdot u_h^S = 0, \quad (107)$$

where

$$\Xi_2 := \Xi_1 + u_h^S \nabla_h \cdot \mathbf{A}, \quad \xi_2 := \xi_1 - (\partial_t u_h^S) \nabla_h \cdot \mathbf{A} + \varepsilon_M u_h^S \mathbf{a}_z \cdot \nabla_h \cdot (r u_h^F) \quad (108)$$

also satisfy

$$\|\Xi_2\|_{n-1} + \|\xi_2\|_{n-2} \leq c\varepsilon_M^{1+\nu} \quad (109)$$

in view of the estimate (101) and the same bounds used to obtain that estimate.

Now move the $\varepsilon_M r^S$ term to the right side of (107) and replace $\partial_t \Xi_2$ there by its divergence-free part, which only changes the divergence term, to obtain

$$(\partial_t + (u_h^S \cdot \nabla_h))u_h^S - b_h^S \cdot \nabla_h b_h^S = \partial_t \Xi_{u_h^S} + \xi_{u_h^S} + \nabla_h \theta_{u_h^S}, \quad \nabla_h \cdot u_h^S = 0 \quad (110)$$

with

$$\Xi_{u_h^S} := \mathcal{P}_h^{\text{div}} \Xi_2, \quad \xi_{u_h^S} := \xi_2 - \varepsilon_M r^S (\partial_t + (u_h^S \cdot \nabla_h))u_h^S. \quad (111)$$

Using the estimate (109), the fact from (96) that $\partial_t u_h^S = O(1)$, and the fact that the projection $\mathcal{P}_h^{\text{div}}$ does not increase Sobolev norms yields

$$\|\Xi_{u_h^S}\|_{n-1} \leq c\varepsilon_M^{1+\nu}, \quad \|\xi_{u_h^S}\|_{n-2} \leq c\varepsilon_M \quad (112)$$

in view of the term that is explicitly $O(\varepsilon_M)$ in the definition (111) of $\xi_{u_h^S}$.

Proof of Theorem 3.4. The functions (\bar{u}_h, \bar{b}_h) satisfy the systems

$$(\partial_t + (\bar{u}_h \cdot \nabla_h))\bar{u}_h - (\bar{b}_h \cdot \nabla_h)\bar{b}_h = \nabla_h \Phi, \quad \nabla_h \cdot \bar{u}_h = 0, \quad (113)$$

$$(\partial_t + (\bar{u}_h \cdot \nabla_h))\bar{b}_h - (\bar{b}_h \cdot \nabla_h)\bar{u}_h = 0. \quad (114)$$

We now apply Theorem D.1 to the system (110), (102) for $u := (u_h, b_h)$ and the system (113), (114) for $U := (\bar{u}_h, \bar{b}_h)$. Assumptions (44) and (50) ensure that the difference in their initial data is $O(\varepsilon_M^{1+\nu})$. Since $\Xi_{u_h^S}$ contains $\mathcal{P}_h^{\text{div}}$, $\Xi_u := (\Xi_{u_h^S}, \Xi_{b_h^S})$ and $\Xi_U := 0$ satisfy $L\Xi_u = 0 = L\Xi_U$, where $L = (\nabla_h \cdot \ 0)$. Define $\xi_u = (\xi_{u_h^S}, \xi_{b_h^S})$ and $\xi_U = 0$. The estimates (112), (104) together with the above-mentioned estimate on the difference of the initial data then imply that the hypotheses of Theorem D.1 hold with $k = n$, $r = 1$, and $\delta = \varepsilon_M$. Hence the conclusion of that theorem yields (92). \square

3.4. Equations and estimates for remaining slow modes

The third component of equation (4b) implies that there are no large terms in the PDE for $u_3^S := \mathbf{a}_z u_3$, i.e., $(0, \begin{pmatrix} 0_2 \\ u_3^S \end{pmatrix}, 0_3)$ is a zero eigenvector of the full large operator $\mathcal{L}_A + \mu\mathcal{L}_M$. However, as shown in Lemma 2.1, the μ -dependent zero eigenvector of $\mathcal{L}_A + \mu\mathcal{L}_M$ having a nontrivial projection onto the density component is not just the slow mode r^S . Specifically, (19) implies that $\mathbf{V}_r = (\mathbf{a}_z, 0_3, \begin{pmatrix} 0_2 \\ -\mu \mathbf{a}_z \end{pmatrix})$ satisfies $(\mathcal{L}_A + \mu\mathcal{L}_M)\mathbf{V}_r \cdot \tilde{\mathbf{V}} = 0$ for all functions $\tilde{\mathbf{V}}$, which by the MHD system (4) and the skew-adjointness of $(\mathcal{L}_A + \mu\mathcal{L}_M)$ implies that the PDE satisfied by $\mathbf{a}_z(r - \mu b_3) = \mathbf{V}_r \cdot \mathbf{V}$ will contain no large terms. We therefore need to calculate the PDE system satisfied by $\mathbf{a}_z(r - \mu b_3)$ and $\mathbf{a}_z u_3$. It turns out that while the PDEs for those two variables are coupled by terms that are strictly $O(1)$, their coupling to other components of the solution contains only terms that are $o(1)$ and so can be considered as small perturbations.

Theorem 3.5. *Under the conditions of Theorem 3.4,*

$$\|r^S - \bar{r}\|_{n-2} + \|u_3^S - \bar{u}_3\|_{n-2} \leq c \left[\varepsilon_M^{1-\max(n-5,0)\nu} + |\mu - \mu_{\text{lim}}| \right]. \quad (115)$$

Proof. Writing the variables in (148) in terms of fast, intermediate, and slow components and using the facts that the slow components are independent of z , u_h^S and b_h^S have zero horizontal divergence, the vertical averages of r^I , u_3^I and b_3^I vanish, and b_3^S is constant in time as well as space transforms that equation into

$$\begin{aligned} & \partial_t(r^S - \mu \mathbf{a}_z b_3^F) + (u_h^S \cdot \nabla_h)(r^S - \mu \mathbf{a}_z b_3^F) + \mu(b_h^S \cdot \nabla_h)u_3^S \\ &= -\mathbf{a}_z \{ \nabla_h \cdot [(r - \mu b_3)u_h^F] \} - \mu \mathbf{a}_z [\nabla_h \cdot (u_3^I b_h^F)] \\ &= -\nabla_h \cdot [(r^S - \mu \mathbf{a}_z b_3) \mathbf{a}_z u_h^F] - \mathbf{a}_z \nabla_h \cdot [(r^I - \mu(1 - \mathbf{a}_z)b_3^F)u_h^F] - \mu \mathbf{a}_z [\nabla_h \cdot (u_3^I b_h^F)] \end{aligned} \quad (116)$$

Replacing u_h^S , b_h^S , and μ on the left side of (116) by their limit values, and compensating by adding terms to the right side yields

$$[\partial_t + (\bar{u}_h \cdot \nabla_h)](r^S - \mu \mathbf{a}_z b_3^F) + \mu_{\text{lim}}(\bar{b}_h \cdot \nabla_h)u_3^S = \partial_t \Xi_{r^S - \mu \mathbf{a}_z b_3^F} + \xi_{r^S - \mu \mathbf{a}_z b_3^F}, \quad (117)$$

where

$$\begin{aligned} \Xi_{r^S - \mu \mathbf{a}_z b_3^F} &:= -\nabla_h \cdot \left[(r^S - \mu \mathbf{a}_z b_3) \int_0^t \mathbf{a}_z u_h^F \right] = -\nabla_h \cdot [(r^S - \mu \mathbf{a}_z b_3)\mathbf{A}] \\ \xi_{r^S - \mu \mathbf{a}_z b_3^F} &:= \nabla_h \cdot [(\partial_t(r^S - \mu \mathbf{a}_z b_3))\mathbf{A}] - \mathbf{a}_z \nabla_h \cdot [(r^I - \mu(1 - \mathbf{a}_z)b_3^F)u_h^F] - \mu \mathbf{a}_z [\nabla_h \cdot (u_3^I b_h^F)] \\ &\quad + (\bar{u}_h - u_h^S) \cdot \nabla_h (r^S - \mu \mathbf{a}_z b_3^F) + \mu((\bar{b}_h - b_h^S) \cdot \nabla_h)u_3^S + (\mu_{\text{lim}} - \mu)(\bar{b}_h \cdot \nabla_h)u_3^S. \end{aligned} \quad (118)$$

Since $\partial_t b_3^S \equiv 0$, (116) implies a uniform H^{n-1} bound for $\partial_t(r^S - \mu \mathbf{a}_z b_3)$. Using in addition the uniform estimate (14), estimate (99) for \mathbf{A} , the estimate (92) for the convergence rate of the horizontal components, and Corollary E.3 shows that

$$\|\Xi_{r^S - \mu \mathbf{a}_z b_3^F}\|_{n-1} \leq c \varepsilon_M^{1+\nu}, \quad \|\xi_{r^S - \mu \mathbf{a}_z b_3^F}\|_{n-2} \leq c [\varepsilon_M + |\mu - \mu_{\text{lim}}|]. \quad (119)$$

The third component of (94) can be written as

$$\partial_t(\rho u_3) + \mathbf{u} \cdot \nabla(\rho u_3) + (\nabla \cdot \mathbf{u})(\rho u_3) + \partial_z \Phi - \mathbf{b}_h \cdot \nabla_h b_3 = 0.$$

Applying the vertical averaging operator \mathbf{a}_z and using (93) reduces this to

$$\partial_t(\mathbf{a}_z(\rho u_3)) + \mathbf{a}_z[\mathbf{u}_h \cdot \nabla_h(\rho u_3)] + \mathbf{a}_z[(\nabla_h \cdot \mathbf{u}_h)(\rho u_3)] - \mathbf{a}_z[\mathbf{b}_h \cdot \nabla_h b_3^F] = 0.$$

In order to treat the term $\mathbf{a}_z[\mathbf{b}_h \cdot \nabla_h b_3^F]$ we write $\mathbf{a}_z b_3^F$ as

$$\mathbf{a}_z b_3^F = -\mu \Delta^{-1} \Delta_h r^S + \mathbf{a}_z(b_3^F + \mu \Delta^{-1} \Delta_h r^S) = -\mu(r^S - \text{avr}^S) + \mathbf{a}_z(b_3^F + \mu \Delta^{-1} \Delta_h r^S) \quad (120)$$

in accordance with the expression estimated in (54). Using (120) while noting that $(\mathbf{b}_h^S \cdot \nabla_h)(\text{avr}^S) = 0$, and using the facts from (5), (23), and (21) that $\rho = 1 + \varepsilon_M r$, r^S and u_3^S are independent of z and $\mathbf{a}_z u_3^I = 0 = u_3^F$, which imply that $\mathbf{a}_z(\rho u_3^S) = (1 + \varepsilon_M r^S) u_3^S$, $\mathbf{a}_z(\rho u_3^I) = \varepsilon_M \mathbf{a}_z(r^I u_3^I)$, and $\mathbf{a}_z r = r^S$, we obtain

$$\begin{aligned} & (1 + \varepsilon_M r^S) [\partial_t + (\mathbf{u}_h^S \cdot \nabla_h)] u_3^S + u_3^S [\partial_t + (\mathbf{u}_h^S \cdot \nabla_h)] (\varepsilon_M r^S) + \mu (\mathbf{b}_h^S \cdot \nabla_h) r^S \\ &= (\mathbf{b}_h^S \cdot \nabla_h) [\mathbf{a}_z(b_3^F + \mu \Delta^{-1} \Delta_h r^S)] + \mathbf{a}_z\{(\mathbf{b}_h^F \cdot \nabla_h)(1 - \mathbf{a}_z) b_3^F\} - \varepsilon_M (\mathbf{u}_h^S \cdot \nabla_h) (\mathbf{a}_z(r^I u_3^I)) \\ & \quad - \varepsilon_M \partial_t \mathbf{a}_z(r^I u_3^I) - \nabla_h \cdot [(\partial_t \mathbf{A}) u_3^S] - \mathbf{a}_z\{\nabla_h \cdot (\mathbf{u}_h^F u_3^I)\} - \varepsilon_M \mathbf{a}_z\{\nabla_h \cdot (\mathbf{u}_h^F r u_3)\}, \end{aligned} \quad (121)$$

where the last line results from separating the various modes in $\mathbf{a}_z\{\nabla_h \cdot (\mathbf{u}_h^F \rho u_3)\}$ and using the definition of \mathbf{A} from (98). Since there are no terms of size ε_A^{-1} in the equations for the time evolution of r or u_3 , (121) implies that

$$\|\partial_t u_3^S\|_{n-1} \leq c. \quad (122)$$

Subtracting u_3^S times (106) from (121), moving the term $\varepsilon_M r^S [\partial_t + (\mathbf{u}_h^S \cdot \nabla_h)] u_3^S$ to the right side of the result, noting that the two terms involving \mathbf{A} partially cancel and rewriting the remaining term $(\partial_t \mathbf{A} \cdot \nabla_h) u_3^S$ as $\partial_t[(\mathbf{A} \cdot \nabla_h) u_3^S] - (\mathbf{A} \cdot \nabla_h) \partial_t u_3^S$ yields

$$(\partial_t + (\mathbf{u}_h^S \cdot \nabla_h)) u_3^S + \mu (\mathbf{b}_h^S \cdot \nabla_h) r^S = \partial_t \tilde{\Xi}_{u_3^S} + \tilde{\xi}_{u_3^S} + (\mathbf{b}_h^S \cdot \nabla_h) [\mathbf{a}_z(b_3^F + \mu \Delta^{-1} \Delta_h r^S)] \quad (123)$$

where

$$\begin{aligned} \tilde{\Xi}_{u_3^S} &= (\mathbf{A} \cdot \nabla_h) u_3^S - \varepsilon_M \mathbf{a}_z(r^I u_3^I), \\ \tilde{\xi}_{u_3^S} &= \mathbf{a}_z\{(\mathbf{b}_h^F \cdot \nabla_h)(1 - \mathbf{a}_z) b_3^F\} - \varepsilon_M (\mathbf{u}_h^S \cdot \nabla_h) (\mathbf{a}_z(r^I u_3^I)) - \mathbf{a}_z\{\nabla_h \cdot (\mathbf{u}_h^F u_3^I)\} - \varepsilon_M \mathbf{a}_z\{\nabla_h \cdot (\mathbf{u}_h^F r u_3)\} \\ & \quad + \varepsilon_M u_3^S \mathbf{a}_z \nabla_h \cdot (r \mathbf{u}_h^F) - (\mathbf{A} \cdot \nabla_h) (\partial_t u_3^S) - \varepsilon_M r^S [\partial_t + (\mathbf{u}_h^S \cdot \nabla_h)] u_3^S. \end{aligned}$$

As a step towards symmetrizing the system consisting of (117), (123), we want to replace r^S in the last term on the left side of (123) by $r^S - \mu \mathbf{a}_z b_3^F$, which requires adding a balancing term involving $\mathbf{a}_z b_3^F$, which must also be rewritten using (120). This leads us to write $\mu (\mathbf{b}_h^S \cdot \nabla_h) r^S$ as $k_1 (\mathbf{b}_h^S \cdot \nabla_h) (r^S - \mu \mathbf{a}_z b_3^F) + k_2 (\mathbf{b}_h^S \cdot \nabla_h) (\mathbf{a}_z b_3^F + \mu r^S)$. Equating those two expressions and comparing the coefficients of $(\mathbf{b}_h^S \cdot \nabla_h) (\mathbf{a}_z b_3^F)$ shows that $k_2 = k_1 \mu$, and then

comparing the coefficients of $(\mathbf{b}_h \cdot \nabla_h) r^S$ yields $k_1 = \frac{\mu}{1+\mu^2}$,

$$\begin{aligned} \mu(\mathbf{b}_h^S \cdot \nabla_h) r^S &= \frac{\mu}{1+\mu^2} (\mathbf{b}_h^S \cdot \nabla_h) (r^S - \mu \mathbf{a}_z b_3^F) + \frac{\mu^2}{1+\mu^2} (\mathbf{b}_h^S \cdot \nabla_h) (\mathbf{a}_z b_3^F + \mu r^S) \\ &= \frac{\mu}{1+\mu^2} (\mathbf{b}_h^S \cdot \nabla_h) (r^S - \mu \mathbf{a}_z b_3^F) + \frac{\mu^2}{1+\mu^2} (\mathbf{b}_h^S \cdot \nabla_h) [\mathbf{a}_z (b_3^F + \mu \Delta^{-1} \Delta_h r^S)], \end{aligned} \quad (124)$$

where the second equation follows as in (120). Substituting (124) into (123), moving the second term from (124) to the right side of the resulting equation and combining it with the similar term already present, and replacing u_h^S , b_h^S and μ on the left side of the result by their limiting values and compensating on the right side yields

$$[\partial_t + (\bar{\mathbf{u}}_h \cdot \nabla_h)] u_3^S + \frac{\mu_{\text{lim}}}{1+\mu_{\text{lim}}^2} (\bar{\mathbf{b}}_h \cdot \nabla_h) (r^S - \mu \mathbf{a}_z b_3^F) = \partial_t \Xi_{u_3^S} + \xi_{u_3^S} \quad (125)$$

where

$$\Xi_{u_3^S} = \tilde{\Xi}_{u_3^S} + \frac{1}{1+\mu^2} (\mathbf{b}_h^S \cdot \nabla_h) \left[\int_0^t \mathbf{a}_z (b_3^F + \mu \Delta^{-1} \Delta_h r^S) \right] \quad (126)$$

$$\begin{aligned} \xi_{u_3^S} &= \tilde{\xi}_{u_3^S} + \frac{1}{1+\mu^2} (\partial_t \mathbf{b}_h^S \cdot \nabla_h) \left[- \int_0^t \mathbf{a}_z (b_3^F + \mu \Delta^{-1} \Delta_h r^S) \right] + (\bar{\mathbf{u}}_h - \mathbf{u}_h^S) \cdot \nabla_h u_3^S \\ &\quad + \frac{\mu}{1+\mu^2} ((\bar{\mathbf{b}}_h - \mathbf{b}_h^S) \cdot \nabla_h) (\mathbf{a}_z b_3^F + \mu r^S) + \left(\frac{\mu_{\text{lim}}}{1+\mu_{\text{lim}}^2} - \frac{\mu}{1+\mu^2} \right) (\bar{\mathbf{b}}_h \cdot \nabla_h) (\mathbf{a}_z b_3^F + \mu r^S). \end{aligned} \quad (127)$$

The system consisting of (117), (125) can be symmetrized by multiplying the latter equation by $1 + \mu_{\text{lim}}^2$. The estimates used to obtain (119) together with the time-derivative estimates (96), (122) and the time-integrated estimate (55) show that

$$\|\Xi_{u_3^S}\|_{n-1} \leq c \varepsilon_M^{1+\nu}, \quad \|\xi_{u_3^S}\|_{n-2} \leq c [\varepsilon_M + |\mu - \mu_{\text{lim}}|]. \quad (128)$$

Using (15e) and the fact that $\text{avr}^{0,S}$ is a constant, the limit equations (15a) and (15c) can be rewritten as the system

$$[\partial_t + (\bar{\mathbf{u}}_h \cdot \nabla_h)] \{ (1 + \mu_{\text{lim}}^2) \bar{r} - \mu_{\text{lim}}^2 \text{avr}^{0,S} \} + \mu_{\text{lim}} (\bar{\mathbf{b}}_h \cdot \nabla_h) \bar{u}_3 = 0, \quad (129)$$

$$[\partial_t + (\bar{\mathbf{u}}_h \cdot \nabla_h)] \bar{u}_3 + \frac{\mu_{\text{lim}}}{1+\mu_{\text{lim}}^2} (\bar{\mathbf{b}}_h \cdot \nabla_h) \{ (1 + \mu_{\text{lim}}^2) \bar{r} - \mu_{\text{lim}}^2 \text{avr}^{0,S} \} = 0, \quad (130)$$

for the dependent variables $(1 + \mu_{\text{lim}}^2) \bar{r} - \mu_{\text{lim}}^2 \text{avr}^{0,S}$ and \bar{u}_3 , which has the same form as the system (117), (125) for the dependent variables $r^S - \mu \mathbf{a}_z b_3$ and u_3^S , except that the terms on the right sides are omitted. Since the evolution equation for r shows that $\text{avr} = \text{avr}^0 = \text{avr}^{0,S} + \varepsilon_M \text{avr}^{0,I} = \text{avr}^{0,S}$,

$$\begin{aligned} r^S - \mu \mathbf{a}_z b_3^F &= (1 + \mu^2) r^S - \mu^2 \text{avr} - \mu \mathbf{a}_z (b_3^F + \mu(r - \text{avr})) \\ &= [(1 + \mu^2) r^S - \mu^2 \text{avr}^{0,S}] - \mu \mathbf{a}_z (b_3^F + \mu \Delta^{-1} \Delta_h r). \end{aligned} \quad (131)$$

Hence, by (44), (50), the difference between the initial data for the two systems is bounded in H^n by a constant times $\varepsilon_M^{1+2\nu} + |\mu - \mu_{\text{lim}}|$. In view of that bound plus the estimates (119), (128) for the right sides of (117), (125), Theorem D.1 shows that

$$\|(r^S - \mu \mathbf{a}_z b_3^F) - \{ (1 + \mu_{\text{lim}}^2) \bar{r} - \mu_{\text{lim}}^2 \text{avr}^{0,S} \}\|_{n-2} + \|u_3^S - \bar{u}_3\|_{n-2} \leq c [\varepsilon_M + |\mu - \mu_{\text{lim}}|]. \quad (132)$$

By (131), the static estimate (54) with $j = n - 3$ applied to the $-\mu \mathbf{a}_z (b_3^F + \mu \Delta^{-1} \Delta_h r)$ term of (131) shows that (132) implies that (115) holds. \square

As discussed in the introduction, the term $|\mu - \mu_{\text{lim}}|$ is the dominating error term in (132) and (115) whenever $\mu_{\text{lim}} = 0$, but that term will be eliminated in Theorem 3.6 below by adding corrector terms.

Theorem 3.6. Let $(r^{(\text{cor})}, u_3^{(\text{cor})})$ be the solution of the inhomogeneous linear system

$$\partial_t r^{(\text{cor})} + (\bar{\mathbf{u}}_h \cdot \nabla_h) r^{(\text{cor})} + \mu (\bar{\mathbf{b}}_h \cdot \nabla_h) u_3^{(\text{cor})} = -(\bar{\mathbf{b}}_h \cdot \nabla_h) \bar{u}_3 - (\mu + \mu_{\text{lim}}) (\partial_t + (\bar{\mathbf{u}}_h \cdot \nabla_h)) \bar{r}, \quad (133)$$

$$\partial_t u_3^{(\text{cor})} + (\bar{\mathbf{u}}_h \cdot \nabla_h) u_3^{(\text{cor})} + \frac{\mu}{1+\mu^2} (\bar{\mathbf{b}}_h \cdot \nabla_h) r^{(\text{cor})} = -\frac{1-\mu\mu_{\text{lim}}}{(1+\mu^2)(1+\mu_{\text{lim}}^2)} (\bar{\mathbf{b}}_h \cdot \nabla_h) ((1+\mu^2)\bar{r} - \mu^2 \text{avr}^{0,S}) \quad (134)$$

having initial data zero. If the conditions of Theorem 3.4 hold then

$$\|r^S - (\bar{r} + \frac{\mu - \mu_{\text{lim}}}{1+\mu^2} r^{(\text{cor})})\|_{n-2} + \|u_3^S - (\bar{u}_3 + (\mu - \mu_{\text{lim}}) u_3^{(\text{cor})})\|_{n-2} \leq c \varepsilon_M^{1-\max(n-5,0)\nu}. \quad (135)$$

Proof. Since $(\mu - \mu_{\text{lim}})(\mu + \mu_{\text{lim}}) = \mu^2 - \mu_{\text{lim}}^2$ and $\text{avr}^{0,S}$ is a constant, adding $\mu - \mu_{\text{lim}}$ times (133) to (129) yields

$$[\partial_t + (\bar{\mathbf{u}}_h \cdot \nabla_h)] \{(1+\mu^2)\bar{r} - \mu^2 \text{avr}^{0,S} + (\mu - \mu_{\text{lim}}) r^{(\text{cor})}\} + \mu (\bar{\mathbf{b}}_h \cdot \nabla_h) (\bar{u}_3 + (\mu - \mu_{\text{lim}}) u_3^{(\text{cor})}) = 0, \quad (136)$$

Similarly, since $(\mu - \mu_{\text{lim}})$ times $\frac{1-\mu\mu_{\text{lim}}}{(1+\mu^2)(1+\mu_{\text{lim}}^2)}$ equals $\frac{\mu}{1+\mu^2} - \frac{\mu_{\text{lim}}}{1+\mu_{\text{lim}}^2}$, adding $\mu - \mu_{\text{lim}}$ times (134) to (130) yields

$$[\partial_t + (\bar{\mathbf{u}}_h \cdot \nabla_h)] \{\bar{u}_3 + (\mu - \mu_{\text{lim}}) u_3^{(\text{cor})}\} + \frac{\mu}{1+\mu^2} (\bar{\mathbf{b}}_h \cdot \nabla_h) \{(1+\mu^2)\bar{r} - \mu^2 \text{avr}^{0,S} + (\mu - \mu_{\text{lim}}) r^{(\text{cor})}\} = 0. \quad (137)$$

Equations (136)–(137) have the same form as the system (117), (125) for the dependent variables $r^S - \mu \mathbf{a}_z b_3$ and u_3^S , except that the terms on the right sides are omitted and all occurrences of μ_{lim} in the coefficients on the left sides are replaced by μ . Omitting the step of replacing μ by μ_{lim} in the derivation of (117), (125) yields those equations with all occurrences of μ_{lim} on the left sides replaced by μ and the terms of order $\mu - \mu_{\text{lim}}$ omitted from their right sides. Since the terms of order μ^2 in (136) now involve μ^2 rather than μ_{lim}^2 , as in (131) and in contrast to (129), there is also no longer a term of size $O(|\mu - \mu_{\text{lim}}|)$ in the difference in the initial data. Hence applying Theorem D.1 now yields an estimate without the term involving $|\mu - \mu_{\text{lim}}|$, and by using (131) the estimate so obtained can be written as (135). \square

APPENDIX A. DERIVATION OF THE MHD SYSTEM

Suitably scaled, the motion of an isentropic compressible, conducting, inviscid fluid is modeled by the MHD system ([Dav01, §3.8])

$$\partial_t \rho + \nabla \cdot (\rho \mathbf{u}) = 0 \quad (138a)$$

$$\partial_t (\rho \mathbf{u}) + \mathbf{u} \cdot \nabla (\rho \mathbf{u}) + (\nabla \cdot \mathbf{u}) \rho \mathbf{u} + \varepsilon_M^{-2} \nabla p(\rho) + \varepsilon_A^{-2} \mathbf{B} \times (\nabla \times \mathbf{B}) = 0, \quad (138b)$$

$$\partial_t \mathbf{B} - \nabla \times (\mathbf{u} \times \mathbf{B}) = 0, \quad (138c)$$

$$\nabla \cdot \mathbf{B} = 0. \quad (138d)$$

Here ε_M denotes the well-known Mach number, ρ is the fluid density, $p(\rho)$ is the pressure law that satisfies $p' > 0$, \mathbf{u} is the fluid velocity, and \mathbf{B} is the magnetic field. The parameter ε_A , as we call the Alfvén number in this article, is the ratio between flow velocity and speed of the magnetosonic waves; in [KM81] the Alfvén number is the reciprocal of our version.

We consider the case in which a uniform magnetic field is applied in the direction \mathbf{e}_z parallel to the z -axis, which subjects the fluid to a large Lorenz force. To reformulate the system (138) into a form to which the results of [CJS18] can be applied, we begin by rescaling the magnetic field and the density via

$$\mathbf{B} = \mathbf{e}_z + \varepsilon_A \mathbf{b}, \quad \rho = 1 + \varepsilon_M r. \quad (139)$$

Applying calculus identities for the curl, subtracting \mathbf{u} times (138a) from (138b), and multiplying (138a) by $a(\varepsilon_M r)$ from (5) yields the system (4).

APPENDIX B. IMPROVED UNIFORM BOUND

Lemma B.1. *Let $n \geq s_0 + 1$ be an integer, where $s_0 := \lfloor \frac{d}{2} \rfloor + 1$ is the Sobolev embedding exponent, i.e., the smallest integer s for which $\|f\|_{L^\infty} \leq c\|f\|_{H^s}$. Assume that the spatial domain is \mathbb{R}^d or \mathbb{T}^d and that the system (1) and its initial data \mathbf{V}^0 satisfy the following conditions:*

- (1) *the operators \mathcal{L}_A and \mathcal{L}_M are constant-coefficient differential operators of order at most one and are skew-adjoint on L^2 ,*
- (2) *the matrices A_i are smooth symmetric functions for $j \geq 0$ and the matrix A_0 is positive definite,*
- (3) *the small parameters are restricted to the region (13),*
- (4) *the initial data \mathbf{V}^0 , which may depend on the small parameters ε_A and ε_M , are uniformly bounded in H^n and satisfy the “well-preparedness” condition (10).*

Then there exist fixed positive T and K such that for $(\varepsilon_A, \varepsilon_M)$ satisfying (13) the solution to (1) having the initial data \mathbf{V}^0 exists for $0 \leq t \leq T$ and satisfies (14); in particular the solution is uniformly bounded in H^n .

Proof. Lemma B.1 differs from Theorem 3.6 of [CJS18] only by having different weights multiplying the norms of time derivatives. Hence it suffices to show that in all places in the proof of [CJS18, Theorem 3.6] where the use of the weights

$$\max_{0 \leq t \leq T} \left[\sum_{j=0}^n \varepsilon_M^j \|\partial_t^j \mathbf{V}\|_{n-j} + \|\mathbf{V}_t\|_0 \right] \leq K \quad (140)$$

was justified the use of the improved weights in (14) is also justified. There are only two such places, namely where it was shown that the weighted sum of norms is bounded at time zero and where it was shown that the small parameters scale out of the estimate for the time derivative of an appropriately weighted energy.

As noted in [CJS18, proof of Lemma 3.5], assumption (10) ensures that $\|\mathbf{V}_t|_{t=0}\|_{n-1}$ is bounded uniformly in the small parameters and the PDE (1) then yields the estimates $\|\partial_t^j \mathbf{V}|_{t=0}\|_{n-j} \leq c\varepsilon_A^{1-j}$ for $1 \leq j \leq n$. Therefore, for $1 \leq j \leq n$,

$$\varepsilon_M^{j-1} \left(\min \left(\frac{\varepsilon_A}{\varepsilon_M}, 1 \right) \right)^{n-1} \|\partial_t^j \mathbf{V}|_{t=0}\|_{n-j} \leq \varepsilon_M^{j-1} \left(\min \left(\frac{\varepsilon_A}{\varepsilon_M}, 1 \right) \right)^{n-1} \left(c\varepsilon_A^{1-j} \right) \leq c \left(\min \left(\frac{\varepsilon_A}{\varepsilon_M}, 1 \right) \right)^{n-j} \leq c,$$

which shows that the weighted sum of norms in (14) is also bounded at time zero uniformly in the small parameters.

The energy estimate both in [CJS18] and here makes use of the norms

$$\|f\|_{\ell, A_0} := \sqrt{\sum_{0 \leq |\alpha| \leq \ell} \int (D^\alpha f)^T A_0(\varepsilon_M \mathbf{V}) D^\alpha f \, dx}, \quad (141)$$

where \mathbf{V} is a solution to (1) and $D^\alpha = \partial_{x_1}^{\alpha_1} \cdots \partial_{x_d}^{\alpha_d}$. As shown in [CJS18], in order to prove a weighted energy estimate like (140) or (14) it suffices to obtain a uniform bound for

$$E := \|\mathbf{V}\|_{n, A_0}^2 + \|\mathbf{V}_t\|_{0, A_0}^2 + \sum_{j=1}^n w_j^2 \|\partial_t^j \mathbf{V}\|_{n-j, A_0}^2, \quad (142)$$

where the weights w_j are ε_M^j for the estimate (140) or

$$w_j = \varepsilon_M^{j-1} \left(\min \left(\frac{\varepsilon_A}{\varepsilon_M}, 1 \right) \right)^{n-1}, \quad 1 \leq j \leq n \quad (143)$$

for the estimate (14). Moreover, in the estimates [CJS18, (3.12), (3.24)] for $\frac{d}{dt}E$, the only facts used about the weights w_j to prove a uniform bound for E are that for some finite c that may be different in each appearance

$$\varepsilon_M \leq c w_1, \tag{144a}$$

$$\varepsilon_M w_j \leq c w_{j+1} \quad \text{for } 1 \leq j \leq n-1, \tag{144b}$$

$$w_k \leq c \prod_{j=1}^J w_{k_j} \quad \text{whenever } \sum_{j=1}^J k_j = k, \tag{144c}$$

$$\varepsilon_M w_k \leq c \prod_{j=1}^J w_{k_j} \quad \text{whenever } \sum_{j=1}^J k_j = k+1. \tag{144d}$$

Since (144d) can be obtained by substituting (144c) with k replaced by $k+1$ into (144b) with j set equal to k , it suffices to prove (144a)–(144c). The definitions (143) imply that (144b) holds provided that c there is at least one, while both (144a) and (144c) reduce to the condition $\varepsilon_M \leq c \left(\min \left(\frac{\varepsilon_A}{\varepsilon_M}, 1 \right) \right)^{n-1}$ that is equivalent to (13). \square

Combining estimate (14) with the standard Sobolev interpolation inequality $\|f\|_r \leq C_{r,s} \|f\|_s^{\frac{r}{s}} \|f\|_0^{1-\frac{r}{s}}$ for $0 \leq r \leq s$ (e.g. [Maj84, (2.32)]), yields the following result.

Corollary B.2. *When the basic conditions of Theorem 1.1 hold and $\mu \leq 1$ then*

$$\|\mathbf{V}_t\|_j \leq c \mu^{-j} = c \varepsilon_M^{-j\nu} \quad j = 0, \dots, n-1. \tag{145}$$

APPENDIX C. CONVERGENCE AND LIMIT

The convergence part of Theorem 1.1 follows from [CJS18, Theorem 4.6] when $\mu_{\text{lim}} = 0$, and from simple modifications of convergence results for two-scale singular limits when $\mu_{\text{lim}} > 0$. Since we need explicit formulas for the limit equations and will use some of the formulas derived below in §3 we indicate a direct unified proof.

Proof of the convergence part of Theorem 1.1. The uniform bounds on \mathbf{V} and \mathbf{V}_t provide compactness, which together with the uniqueness of solutions to the limit equations ensures the convergence of \mathbf{V} , in $C^0([0, T]; H^{n-\alpha})$ for any $\alpha > 0$ and weak-* in $L^\infty([0, T]; H^n)$, to a limit $\bar{\mathbf{V}}$ as $\varepsilon_A, \varepsilon_M$ tend to zero with their ratio converging to a given limit μ_{lim} , with \mathbf{V}_t converging weak-* in $L^\infty([0, T], H^{n-1})$ to $\bar{\mathbf{V}}_t$.

Multiplying (1) by ε_A or applying $\varepsilon_M \mathbb{P}^0$ to it, and taking the limit yields

$$(\mathcal{L}_A + \mu_{\text{lim}} \mathcal{L}_M) \bar{\mathbf{V}} = 0 = \mathbb{P}^0 \mathcal{L}_M \bar{\mathbf{V}}. \tag{146}$$

Identities (146) and Lemma 2.1 imply that $\bar{\mathbf{V}}$ is independent of z , the horizontal parts of its velocity and magnetic field are divergence free, and (15e) holds. If the spatial domain is \mathbb{R}^3 then $\bar{\mathbf{V}}$ must therefore vanish, so from now on that domain is \mathbb{T}^3 . By (20) and (27), $\mathbb{P} \left(\frac{1}{\varepsilon_A} \mathcal{L}_A + \frac{1}{\varepsilon_M} \mathcal{L}_M \right) = \left(\frac{1}{\varepsilon_M} \mathbf{a}_z \nabla_{\mathbf{h}} \cdot \mathbf{u}_{\mathbf{h}}, 0_3, 0_3 \right)$. Taking the limit of the equations with no large terms in \mathbb{P} applied to (4) yields (15b), (15c), (15d).

To determine the limit equation for the density, divide (4a) by $a(\varepsilon_M r)$, which puts it in conservation form, and apply \mathbf{a}_z to obtain

$$\partial_t (\mathbf{a}_z r) + \mathbf{a}_z [\nabla_{\mathbf{h}} \cdot (r \mathbf{u}_{\mathbf{h}})] + \frac{1}{\varepsilon_M} \mathbf{a}_z (\nabla_{\mathbf{h}} \cdot \mathbf{u}_{\mathbf{h}}) = 0. \tag{147}$$

To eliminate the large term in (147), write the third component of (4c) in conservation form as $\partial_t b_3 + \nabla \cdot (b_3 \mathbf{u}) - \nabla \cdot (u_3 \mathbf{b}) + \varepsilon_A^{-1} \nabla_{\mathbf{h}} \cdot \mathbf{u}_{\mathbf{h}} = 0$, apply $\mu \mathbf{a}_z$ and subtract the result from (147), which yields

$$\partial_t [\mathbf{a}_z (r - \mu b_3)] + \mathbf{a}_z \{ \nabla_{\mathbf{h}} \cdot [(r - \mu b_3) \mathbf{u}_{\mathbf{h}}] \} + \mu \mathbf{a}_z [\nabla_{\mathbf{h}} \cdot (u_3 \mathbf{b}_{\mathbf{h}})] = 0. \tag{148}$$

Taking the limit of (148), using facts that $\bar{\mathbf{V}}$ is independent of z and that $\bar{\mathbf{u}}_h$ and $\bar{\mathbf{b}}_h$ are divergence free, and substituting (15e) into the result yields (15a). \square

APPENDIX D. PERTURBATION THEOREM

The following perturbation theorem is a variant of [Che14, Lemma 3.2]), and can be proven by similar methods.

Theorem D.1. *Suppose that u and U are solutions in $C^0([0, T]; H^k)$ of*

$$A_0(u)u_t + \sum_{i=1}^d A_i(u)u_{x_i} = F + L^*v + A_0(u)\partial_t \Xi_u + \xi_u, \quad Lu = 0, \quad (149)$$

$$A_0(U)U_t + \sum_{i=1}^d A_i(U)U_{x_i} = F + L^*V + A_0(U)\partial_t \Xi_U + \xi_U, \quad LU = 0, \quad (150)$$

having the same initial value $u_0 \in H^k$, where $k \geq \lfloor \frac{d}{2} \rfloor + 2$, the matrices A_i are smooth and symmetric and A_0 is positive-definite, F is a given function of t and x , L is a first-order differential operator with constant coefficients, with L^* denoting its L^2 -adjoint, and Ξ_u, Ξ_U, ξ_u , and ξ_U satisfy

$$\begin{aligned} \|\Xi_u\|_{k-r} + \|\Xi_U\|_{k-r} + \|\xi_u\|_{k-r-1} + \|\xi_U\|_{k-r-1} &\leq c\delta \quad \text{for some } 0 \leq r \leq k-1, \\ L\Xi_u = 0 = L\Xi_U, \quad \text{and} \quad \|\partial_t \Xi_U\|_{k-r-1} &\leq c. \end{aligned}$$

Then $\max_{0 \leq t \leq T} \|u - U\|_{k-r-1} \leq c\delta$.

APPENDIX E. CALCULUS INEQUALITIES FOR VERTICAL AVERAGES

The following result is sharper than what would be obtained by the standard product estimate (e.g., [Maj84, Proposition 2.1A]), because the entire product is estimated using the $W^{1,1}$ norm rather than pulling out one factor in the L^∞ norm, and the Gagliardo-Nirenberg inequalities are used in dimension two rather than three.

Lemma E.1. *For all $j \geq 1$ there exists a constant C_j such that for $f, g \in H^j(\mathbb{T}^3)$*

$$\|\mathbf{a}_z(fg)\|_{H^{j-1}(\mathbb{T}^2)} \leq C_j (\|f\|_{H^j(\mathbb{T}^3)}\|g\|_{L^2(\mathbb{T}^3)} + \|f\|_{L^2(\mathbb{T}^3)}\|g\|_{H^j(\mathbb{T}^3)}). \quad (151)$$

Proof. We first prove (151) for $j = 1$: By the Gagliardo-Nirenberg inequality $\|h\|_{L^2(\mathbb{T}^2)} \leq c\|h\|_{W^{1,1}(\mathbb{T}^2)}$ and the Cauchy-Schwartz inequality,

$$\|\mathbf{a}_z(fg)\|_{L^2(\mathbb{T}^2)} \leq c\|\mathbf{a}_z(fg)\|_{W^{1,1}(\mathbb{T}^2)} \leq c(\|f\|_{H^1(\mathbb{T}^3)}\|g\|_{L^2(\mathbb{T}^3)} + \|f\|_{L^2(\mathbb{T}^3)}\|g\|_{H^1(\mathbb{T}^3)}).$$

Now let j be any integer greater than one. By the definition of the H^{j-1} norm, the result for the case $j = 1$, the Sobolev interpolation inequality, and Young's inequality for products $a^\sigma b^{1-\sigma} \leq a + b$ for $0 \leq \sigma \leq 1$,

$$\begin{aligned} \|\mathbf{a}_z(fg)\|_{H^{j-1}(\mathbb{T}^2)} &\leq c \sum_{|\alpha| \leq j-1} \|\mathbf{a}_z(D_{x,y}^\alpha(fg))\|_{L^2(\mathbb{T}^2)} \\ &\leq c \sum_{|\beta|+|\gamma| \leq j-1} \|\mathbf{a}_z((D_{x,y}^\beta f)(D_{x,y}^\gamma g))\|_{L^2(\mathbb{T}^2)} \\ &\leq c \sum_{|\beta|+|\gamma| \leq j-1} (\|f\|_{H^{|\beta|+1}(\mathbb{T}^3)}\|g\|_{H^{|\gamma|}(\mathbb{T}^3)} + \|f\|_{H^{|\beta|}(\mathbb{T}^3)}\|g\|_{H^{|\gamma|+1}(\mathbb{T}^3)}) \end{aligned} \quad (152)$$

$$\begin{aligned} &\leq c \sum_{0 \leq i \leq j-1} (\|f\|_{H^{i+1}} \|g\|_{L^2} + \|g\|_{H^i} \|f\|_{H^1}) + [\|f\|_{H^i} \|g\|_{H^1} + \|g\|_{H^{i+1}} \|f\|_{L^2}] \\ &\leq c (\|f\|_{H^j} \|g\|_{L^2} + \|g\|_{H^{j-1}} \|f\|_{H^1} + \|f\|_{H^{j-1}} \|g\|_{H^1} + \|g\|_{H^j} \|f\|_{L^2}) \end{aligned}$$

A second application of the Sobolev interpolation inequality followed by Young's inequality shows that each of the terms in the final line of (152) in which the H^{j-1} and H^1 norms appear is bounded by the sum of the two terms there in which the H^j and L^2 norms appear, which yields (151) for $j > 1$. \square

Corollary E.2. *Consider integer $n \geq 3$ and a geometric sequence $\{\varepsilon_j\}$ with common ratio $\frac{1}{\mu} \geq 1$ and $\varepsilon_n \leq c$. Suppose $v(x, y, z), w(x, y, z) \in H^n(\mathbb{T}^3)$ satisfy the “interpolative estimates”*

$$\|(v, w)\|_{H^j(\mathbb{T}^3)} \leq c\varepsilon_j, \quad j = 0, \dots, n-1, \quad \|(v, w)\|_{H^n(\mathbb{T}^3)} \leq c.$$

Then

$$\|\mathbf{a}_z(vw)\|_{H^{n-1}(\mathbb{T}_{x,y}^2)} \leq c\varepsilon_0, \quad \|\mathbf{a}_z(vw)\|_{H^{n-2}(\mathbb{T}_{x,y}^2)} \leq c\varepsilon_0 \mu. \quad (153)$$

Proof. By Lemma E.1, $\|\mathbf{a}_z(vw)\|_{j-1} \leq c(\|v\|_j \|w\|_0 + \|v\|_0 \|w\|_j) \leq c\varepsilon_j \varepsilon_0$. Since $\varepsilon_n \leq c$ and $\varepsilon_{n-1} \leq \mu \varepsilon_n \leq c\mu$, this implies (153). \square

In view of the uniform H^n estimate (14), the static estimates (54), (56), and the relations (13), (11) between the parameters, Corollary E.2 yields the following estimates for products of components of various modes.

Corollary E.3. *Assume that the basic conditions of Theorem 1.1 hold. Let v^F, w^F be either $(1 - \mathbf{a}_z)b_3^F$ or any component of \mathbf{V}^F except b_3^F and let v^I, w^I be any component of \mathbf{V}^I . Then,*

$$\sup_{0 \leq t \leq T} \{ \|\mathbf{a}_z(v^F w^F)\|_{n-1} + \mu \|\mathbf{a}_z(v^I v^F)\|_{n-1} + \mu^2 \|\mathbf{a}_z(v^I w^I)\|_{n-1} + \|\mathbf{a}_z \nabla \cdot (v^I u_h^F)\|_{n-2} + \varepsilon_M \|v^F\|_{n-1} \} \leq c\varepsilon_A,$$

and the estimates also hold when μ or μ^2 on the left side is replaced by ε_M .

The estimate of $\varepsilon_M \|v^F\|_{n-1}$ does not use Lemma E.1 but is included for convenience.

REFERENCES

- [Ala08] T. Alazard. A minicourse on the low Mach number limit. *Discrete Contin. Dyn. Syst. Ser. S*, 1(3):365–404, 2008.
- [Asa87] K. Asano. On the incompressible limit of the compressible Euler equation. *Japan J. Appl. Math.*, 4(3):455–488, 1987.
- [CG00] S. Cordier and E. Grenier. Quasineutral limit of an Euler-Poisson system arising from plasma physics. *Comm. Partial Differential Equations*, 25(5-6):1099–1113, 2000.
- [CGS16] R. M. Colombo, G. Guerra, and V. Schleper. The compressible to incompressible limit of one dimensional Euler equations: the non smooth case. *Arch. Ration. Mech. Anal.*, 219(2):701–718, 2016.
- [Che12] B. Cheng. Singular limits and convergence rates of compressible Euler and rotating shallow water equations. *SIAM J. Math. Anal.*, 44(2):1050–1076, 2012.
- [Che14] B. Cheng. Improved accuracy of incompressible approximation of compressible Euler equations. *SIAM J. Math. Anal.*, 46(6):3838–3864, 2014.
- [CJS18] B. Cheng, Q. Ju, and S. Schochet. Three-scale singular limits of evolutionary PDEs. *Arch. Ration. Mech. Anal.*, 229(2):601–625, 2018.
- [CM13] B. Cheng and A. Mahalov. Time-averages of fast oscillatory systems. *Discrete Contin. Dyn. Syst. Ser. S*, 6(5):1151–1162, 2013.
- [Dav01] P. A. Davidson. *An introduction to magnetohydrodynamics*. Cambridge Texts in Applied Mathematics. Cambridge University Press, Cambridge, 2001.
- [DGLM99] B. Desjardins, E. Grenier, P.-L. Lions, and N. Masmoudi. Incompressible limit for solutions of the isentropic Navier-Stokes equations with Dirichlet boundary conditions. *J. Math. Pures Appl. (9)*, 78(5):461–471, 1999.
- [DH04] A. Dutrifoy and T. Hmidi. The incompressible limit of solutions of the two-dimensional compressible Euler system with degenerating initial data. *Comm. Pure Appl. Math.*, 57(9):1159–1177, 2004.
- [DHWZ13] S. Ding, J. Huang, H. Wen, and R. Zi. Incompressible limit of the compressible nematic liquid crystal flow. *J. Funct. Anal.*, 264(7):1711–1756, 2013.

- [FFGHR17] J. Földes, S. Friedlander, N. Glatt-Holtz, and G. Richards. Asymptotic analysis for randomly forced MHD. *SIAM J. Math. Anal.*, 49(6):4440–4469, 2017.
- [FN09] E. Feireisl and A. Novotný. *Singular limits in thermodynamics of viscous fluids*. Advances in Mathematical Fluid Mechanics. Birkhäuser Verlag, Basel, 2009.
- [Gal08] I. Gallagher. A mathematical review of the analysis of the betaplane model and equatorial waves. *Discrete Contin. Dyn. Syst. Ser. S*, 1(3):461–480, 2008.
- [Gal19] I. Gallagher. From Newton to Navier-Stokes, or how to connect fluid mechanics equations from microscopic to macroscopic scales. *Bull. Amer. Math. Soc. (N.S.)*, 56(1):65–85, 2019.
- [Got90] S. Goto. Singular limit of the incompressible ideal magneto-fluid motion with respect to the Alfvén number. *Hokkaido Math. J.*, 19(1):175–187, 1990.
- [Gre97a] E. Grenier. Oscillatory perturbations of the Navier-Stokes equations. *J. Math. Pures Appl. (9)*, 76(6):477–498, 1997.
- [Gre97b] E. Grenier. Pseudo-differential energy estimates of singular perturbations. *Comm. Pure Appl. Math.*, 50(9):821–865, 1997.
- [JLL10] S. Jiang, Q. Ju, and F. Li. Incompressible limit of the compressible magnetohydrodynamic equations with periodic boundary conditions. *Comm. Math. Phys.*, 297(2):371–400, 2010.
- [JLL12] S. Jiang, Q. Ju, and F. Li. Low Mach number limit for the multi-dimensional full magnetohydrodynamic equations. *Nonlinearity*, 25(5):1351–1365, 2012.
- [JX18] Q. Ju and X. Xu. Small Alfvén number limit of the plane magnetohydrodynamic flows. *Appl. Math. Lett.*, 86:77–82, 2018.
- [Kat82] T. Kato. *A short introduction to perturbation theory for linear operators*. Springer-Verlag, New York-Berlin, 1982.
- [KM81] S. Klainerman and A. Majda. Singular limits of quasilinear hyperbolic systems with large parameters and the incompressible limit of compressible fluids. *Comm. Pure Appl. Math.*, 34(4):481–524, 1981.
- [KM82] S. Klainerman and A. Majda. Compressible and incompressible fluids. *Comm. Pure Appl. Math.*, 35(5):629–651, 1982.
- [LL18] H. Lindblad and C. Luo. A priori estimates for the compressible Euler equations for a liquid with free surface boundary and the incompressible limit. *Comm. Pure Appl. Math.*, 71(7):1273–1333, 2018.
- [LM98] P.-L. Lions and N. Masmoudi. Incompressible limit for a viscous compressible fluid. *J. Math. Pures Appl. (9)*, 77(6):585–627, 1998.
- [Maj84] A. Majda. *Compressible fluid flow and systems of conservation laws in several space variables*, volume 53 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1984.
- [MS01] G. Métivier and S. Schochet. The incompressible limit of the non-isentropic euler equations. *Arch. Ration. Mech. Anal.*, 158:61–90, 2001.
- [Sch88] S. Schochet. Asymptotics for symmetric hyperbolic systems with a large parameter. *J. Differential Equations*, 75(1):1–27, 1988.
- [Sch94] S. Schochet. Fast singular limits of hyperbolic PDEs. *J. Differential Equations*, 114(2):476–512, 1994.
- [Sec06] P. Secchi. 2D slightly compressible ideal flow in an exterior domain. *J. Math. Fluid Mech.*, 8(4):564–590, 2006.
- [Uka86] S. Ukai. The incompressible limit and the initial layer of the compressible Euler equation. *J. Math. Kyoto Univ.*, 26(2):323–331, 1986.