Introduction to ionosphere-magnetosphere interactions

Pekka Janhunen

June 13, 1996

Contents

Chapter 1

Some useful formulas

Vectors in 3D space:

$$
\begin{aligned}
\mathbf{A}\times(\nabla\times\mathbf{A}) &= \frac{1}{2}\nabla(\mathbf{A}^2) - \mathbf{A}\cdot\nabla\mathbf{A} \\
\nabla\cdot(\mathbf{A}\times\mathbf{B}) &= (\nabla\times\mathbf{A})\cdot\mathbf{B} - (\nabla\times\mathbf{B})\cdot\mathbf{A} \\
\nabla\times(\mathbf{A}\times\mathbf{B}) &= \mathbf{A}\nabla\cdot\mathbf{B} - \mathbf{B}\nabla\cdot\mathbf{A} - \mathbf{A}\cdot\nabla\mathbf{B} + \mathbf{B}\cdot\nabla\mathbf{A}
\end{aligned}
$$

All formulas involving the cross product can be derived easily using the Levi-Civita symbol ϵ_{ijk}:

$$\nabla\times\mathbf{A} = \hat{\mathbf{e}}_i \epsilon_{ijk}\partial_j A_k \tag{1.1}$$

where $\hat{\mathbf{e}}_i$ is the ith ($i = 1, 2, 3$) unit vector, plus the following identity:

$$\epsilon_{ijk}\epsilon_{nmk} = \delta_{in}\delta_{jm} - \delta_{im}\delta_{jn}, \tag{1.2}$$

where a summation over repeated indices is implicitly assumed and the Kronecker delta δ_{ij} is 1 for $i = j$ and 0 otherwise. The Levi-Civita tensor ϵ_{ijk} is totally antisymmetric, and $\epsilon_{ijk} = +1$ for (ijk) an even permutation of (123) and $\epsilon_{ijk} = -1$ for (ijk) an odd permutation of (123). If any two i, j, k indices are equal, ϵ vanishes.

Gauss' theorem (Divergence theorem):

$$\int_V \nabla\cdot\mathbf{A} = \oint_{\partial V} d\mathbf{S}\cdot\mathbf{A} \tag{1.3}$$

Here V is any volume, ∂V its boundary, and $d\mathbf{S}$ is the outward vector area element of the bounding surface.

Stokes' theorem:

$$\int_S d\mathbf{S}\cdot\nabla\times\mathbf{A} = \oint_{\partial S} d\mathbf{l}\cdot\mathbf{A} \tag{1.4}$$

where S is a(n open) surface and ∂S its boundary loop.

These theorems are but special cases of a general theorem for differential forms,

$$\int_M \mathrm{d}\omega = \int_{\partial M} \omega \tag{1.5}$$

where ω is a differential form, M is a N-dimensional manifold and ∂M its $N-1$-dimensional boundary.

Chapter 2

Introduction

These lecture notes give an overview of MHD processes in Earth's magnetosphere. The emphasis is on ionosphere-magnetosphere coupling. Before ionosphere-magnetosphere coupling can be discussed, however, ionospheric equations and magnetospheric MHD must be understood.

Prerequisities include Maxwell's equations, basic mathematical capabilities (vector analysis) and some familiarity with basic plasma physics.

The style of these notes may differ substantially from a 'usual' treatment. I have avoided material and analytic methods which are related to some coordinate system or some particular geometry only. On the other hand, somewhat more advanced topics, e.g. regarding the MHD eigenmodes, are included.

The reason for the new style is the availability of powerful numerical and symbolic software. It is no longer necessary to derive complicated formulas in special geometries by hand, except in some special cases. The goal is, in addition to teaching magnetospheric/ionospheric physics, to prepare the student towards large-scale computer simulations, even though simulation techniques are outside the scope of this course.

The lectures contain exercises (HOME TASKS), which are an integral part of the course. Good practical mastery of vector algebra is necessary (and enough) for most home tasks. However, some of the exercises require the use of a symbolic manipulation program such as MACSYMA, Maple or Mathematica (in alphabetical order).

Chapter 3

MHD

3.1 Introduction

The magnetohydrodynamic (MHD) model of plasma is very important for practical purposes. It is the model that lets us deal with familiar concepts such as flow velocity, pressure and temperature. The name, magneto*hydro*dynamics, is unfortunately misleading: much better name would be magnetogasdynamics because plasma is usually compressible.

One can derive the MHD equations in several ways. The most fundamental way is to take velocity moments of the Boltzmann (or Vlasov) equation and ignore higher-than-2 order moments. This is not enough, however, one has to make several approximations in addition to get the MHD system. Another way is to simply assume that the plasma is magnetized gas, i.e., gas which feels the Ampere force density $\mathbf{j}\times\mathbf{B}$.

Here we shall "derive" the MHD equations using the simpler approach.

3.2 Euler equations

Ordinary gasdynamic equations read

$$\begin{aligned} \frac{\partial \rho}{\partial t} &= -\nabla \cdot (\rho \mathbf{v}) \\ \rho \frac{d\mathbf{v}}{dt} &= -\nabla P + \nu \rho \nabla^2 \mathbf{v} \\ \frac{d}{dt}\left(P\rho^{-\gamma}\right) &= 0 \end{aligned} \tag{3.1}$$

where $d/dt \equiv \partial/\partial t + \mathbf{v}\cdot\nabla$ is the convective (co-moving) time derivative. Equation (3.1) contains five equations in five unknowns, density ρ, pressure P and three velocity components v_x, v_y, v_z. The last equation is the energy equation, which has been written for an adiabatic process. The classical value of the adiabatic constant γ is $5/3$ for three-dimensional flow of monoatomic gas. If the viscosity ν is nonzero, the system is called the Navier-Stokes equations. If the viscosity is zero, the system (3.1) is called the Euler equations.

The energy equation can also be written in the form

$$\begin{aligned} \frac{d}{dt}\left(P\rho^{-\gamma}\right) &= \frac{dP}{dt}\rho^{-\gamma} - \gamma P\rho^{-\gamma-1}\frac{d\rho}{dt} \\ &= \left(\frac{dP}{dt}\right)\rho^{-\gamma} + \gamma P\rho^{-\gamma-1}\rho\nabla\cdot\mathbf{v} \\ &= 0 \end{aligned}$$

$$\Rightarrow \frac{dP}{dt} = -\gamma P \nabla\cdot\mathbf{v} \tag{3.2}$$

From this form we see e.g. what happens if $\gamma = 1$, which should correspond to the isothermal case ($P\rho^{-\gamma} = P/\rho = \text{const} \Rightarrow T = \text{const}$ according to the ideal gas equation of state). Indeed this is the case, since for $\gamma = 1$ Equation (3.2) has the same form as the mass continuity equation, which implies that the pressure and the density satisfy the same equation. Therefore they must be proportional to each other, which means isothermality.

One can write Equation (3.1) in so-called conservation form. A time-dependent PDE (partial differential equation) is in conservation form if it is written as

$$\frac{\partial u_a}{\partial t} = -\nabla \cdot \mathbf{F}_a \tag{3.3}$$

where the fluxes $\mathbf{F}_a$ may depend on $\{u_a\}_{a=1}^n$ and their derivatives, and on space and time. Here n is the number of dynamic variables, $n = 5$ in case of fluid dynamics (gas dynamics). The reason Equation (3.3) is called conservation form is as follows. Integrate Equation (3.3) over a spatial volume V:

$$\begin{aligned}\int_V dV \frac{\partial u_a}{\partial t} &= -\int_V dV \nabla \cdot \mathbf{F}_a \\ &= -\oint_{\partial V} d\mathbf{S} \cdot \mathbf{F}_a\end{aligned}$$

$$\Rightarrow \frac{d}{dt}\int_V dV u_a = -\oint_{\partial V} d\mathbf{S} \cdot \mathbf{F}_a \tag{3.4}$$

If we consider an isolated system and if the integration volume V covers the whole system, the right-hand side (RHS) vanishes. [Why? Because all variables tend to zero outside an isolated system]. In this case the left-hand side (LHS) is also zero, and this implies that the quantity

$$C_a = \int_V dV u_a \tag{3.5}$$

stays constant in time (is a constant of motion). Thus, C_a is a conserved quantity, because it is temporarally constant in an isolated system. The RHS of Equation (3.4) can also be easily interpreted. It represents the flux, or flow, or u_a through the system boundary ∂V, in case V does not enclose an isolated system. Thus, we have noted the result that **whenever the system PDEs can be written in conservation form (3.3), the corresponding dynamic variables u_a are conserved quantities, and the corresponding RHS quantities (F_a) are the corresponding fluxes.**

The Euler equations in conservation form are (HT)

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot \mathbf{p} \tag{3.6}$$

$$\frac{\partial \mathbf{p}}{\partial t} = -\nabla \cdot \left(\frac{\mathbf{pp}}{\rho} + P\mathtt{I}\right) \tag{3.7}$$

$$\frac{\partial u}{\partial t} = -\nabla \cdot \left[(u + P)\frac{\mathbf{p}}{\rho}\right] \tag{3.8}$$

Here $\mathtt{I}$ is the unit dyad (tensor, matrix) and the direct product of two vectors has been denoted by $\mathbf{pp}$. The momentum flux $\mathbf{p} = \rho\mathbf{v}$ is now one of the dynamical variables, instead of the velocity $\mathbf{v}$. The connection between the total energy density u and pressure P is

$$P = (\gamma - 1)\left(u - \frac{\mathbf{p}^2}{2\rho}\right), \tag{3.9}$$

or

$$u = \frac{P}{\gamma - 1} + \frac{\mathbf{p}^2}{2\rho} \tag{3.10}$$

The thermal energy density is given by $P/(\gamma-1)$. [If $\gamma = 5/3$ this becomes $(3/2)P$. Recall that $P = nk_BT$ for ideal gas where n is the number density, k_B is Boltzmann's constant and T is the temperature. On the other hand, $(1/2)k_BT$ is the average energy per particle and per one degree of freedom. In 3D gas we have 3 degrees of freedom, thus the total kinetic energy per particle is $(3/2)k_BT \Rightarrow$ the associated energy density is $(3/2)nk_BT = (3/2)P$ as it should. The general classical expression for γ is $\gamma = (D+2)/D$, where D is the number of degrees of freedom.] The term $\mathbf{p}^2/(2\rho)$ is, of course, the kinetic energy associated with the bulk motion of the gas; thus u is indeed the total (thermal + bulk kinetic) energy density.

It is not difficult (HT) to establish the equivalence of Equation (3.1) and equations (3.6) through (3.8).

3.3 Derivation of MHD equations

To obtain the MHD equations we need to add the magnetic field to the Euler equations. More generally, we should couple together Maxwell's equations and Euler equations. In classical electrodynamics the setup of the problem is usually such that one tries to solve Maxwell's equations with known charge and current densities. In plasma physics the problem is more difficult because the charge and current densities are not known. Especially the charge density remains unknown. For the MHD equations to be valid at all, we need to assume that the plasma is quasineutral, i.e. that we consider scale lengths much larger than the Debye length. In this limit we cannot compute the charge density using $\rho_e = e(n_i - n_e)$ (or something similar) because n_i and n_e are so close together that their difference is numerical garbage! Therefore, Gauss' law is of no use in MHD. It can only be used afterwards to compute the charge density from the electric field, but it is not involved in the dynamical evolution of the system.

Consistent with the approximation of quasineutrality, which always holds in sufficiently large scales in a plasma, is that we can usually neglect the displacement current in Ampere's law. Thus all that is left from Maxwell's equations is:

$$\frac{\partial \mathbf{B}}{\partial t} = -\nabla\times\mathbf{E} \tag{3.11}$$

$$\nabla\cdot\mathbf{B} = 0 \tag{3.12}$$

$$\nabla\times\mathbf{B} = \mu_0\mathbf{j} \tag{3.13}$$

Ampere's law [Equation (3.13)] can be used as a definition of current density in MHD.

We are now ready to begin constructing the MHD equation system using the Euler equations [Equation (3.1)] as a starting point. One additional term is clear. In the momentum equation [Equation (3.7)] we have to add the Lorentz force. The Lorentz force on a particle is equal to $q\mathbf{v}\times\mathbf{B}$. If the number density of αth species is n_α, the total Lorentz force on a fluid element $d^3\mathbf{x}$ is

$$\begin{aligned} d\mathbf{F} &= \sum_\alpha q_\alpha \mathbf{v}_\alpha \times \mathbf{B}\, dN_\alpha \\ &= \left(\sum_\alpha n_\alpha q_\alpha \mathbf{v}_\alpha\right) \times \mathbf{B} d^3\mathbf{x} \\ &= \mathbf{j}\times\mathbf{B} d^3\mathbf{x}, \end{aligned}$$

thus the Lorentz force density is $\mathbf{j}\times\mathbf{B}$. Adding this term to the momentum equation, Equation (3.7) we obtain

$$\frac{\partial \mathbf{p}}{\partial t} = -\nabla\cdot\left(\frac{\mathbf{pp}}{\rho} + P\mathtt{I}\right) + \mathbf{j}\times\mathbf{B}. \tag{3.14}$$

Now, it is easy to show (HT) that

$$\mathbf{j}\times\mathbf{B} = -\frac{1}{\mu_0}\mathbf{B}\times(\nabla\times\mathbf{B}) = \frac{1}{\mu_0}\nabla\cdot(\mathbf{BB}) - \nabla\left(\frac{B^2}{2\mu_0}\right). \tag{3.15}$$

Thus we can write the momentum equation (3.14) in the form

$$\frac{\partial\mathbf{p}}{\partial t} = -\nabla\cdot\left(\frac{\mathbf{pp}}{\rho} + \left(P + \frac{\mathbf{B}^2}{2\mu_0}\right)\mathtt{I} - \frac{1}{\mu_0}\mathbf{BB}\right) \tag{3.16}$$

which is now in conservation form. We can identify the quantity $(P+B^2/(2\mu_0))\mathtt{I}-\frac{1}{\mu_0}\mathbf{BB}$ as a generalized pressure tensor.

The addition of the $\mathbf{j}\times\mathbf{B}$ in the momentum equation (where $\mathbf{j}\equiv(\nabla\times\mathbf{B})/\mu_0$) is in fact the only required modification to Euler's equations to get the MHD equations. The energy equation, when written in form (3.2), does not need modification. However, what is still missing is an evolution equation for the magnetic field. Faraday's law (3.11) could be used for this purpose, if we know how to compute the electric field. But this is easy, since the plasma, being ionized gas, cannot sustain an electric field for practical purposes. Thus, $\mathbf{E}=0$ in the rest frame of the plasma. Since the plasma is moving with velocity $\mathbf{v}$, this becomes $\mathbf{E}=0-\mathbf{v}\times\mathbf{B}$ in the laboratory frame (recall the Lorentz transformation of the electromagnetic field in nonrelativistic limit). Thus, in ideal MHD we can take $\mathbf{E}=\mathbf{B}\times\mathbf{v}$, and Faraday's law gives

$$\frac{\partial\mathbf{B}}{\partial t} = \nabla\times(\mathbf{v}\times\mathbf{B}) \tag{3.17}$$

We have now "derived" the ideal MHD equations:

$$\begin{aligned}
\frac{\partial\rho}{\partial t} &= -\nabla\cdot(\rho\mathbf{v}) \\
\rho\frac{d\mathbf{v}}{dt} &= \mathbf{j}\times\mathbf{B} - \nabla P \\
\frac{dP}{dt} &= -\gamma P\nabla\cdot\mathbf{v} \\
\frac{\partial\mathbf{B}}{\partial t} &= \nabla\times(\mathbf{v}\times\mathbf{B})
\end{aligned} \tag{3.18}$$

where $\mathbf{j}\equiv(\nabla\times\mathbf{B})/\mu_0$ and $d/dt\equiv\partial/\partial t+\mathbf{v}\cdot\nabla$. The system (3.18) is the simplest representation of the MHD equations, and it is not in conservation form. The dynamic variables of system (3.18) $(\rho,\mathbf{v},P,\mathbf{B})$ are called *primitive variables.*

The third Maxwell's equation, $\nabla\cdot\mathbf{B}=0$, must be always satisfied, of course. It is enough to have $\nabla\cdot\mathbf{B}=0$ at $t=0$, however, because Faraday's law then guarantees that the magnetic field remains divergenceless for all future times. Sometimes this nice property can be even carried over to the discretized realm of numerical simulations.

Let us consider little more closely the generalized pressure tensor, $\mathtt{T}=(P+\mathbf{B}^2/(2\mu_0))\mathtt{I}-\mathbf{BB}/\mu_0$. If we choose an orthogonal coordinate system where the x axis is locally parallel to the magnetic field, the tensor $\mathtt{T}$ assumes the form

$$\mathtt{T} = \begin{bmatrix} P-\frac{\mathbf{B}^2}{2\mu_0} & 0 & 0 \\ 0 & P+\frac{\mathbf{B}^2}{2\mu_0} & 0 \\ 0 & 0 & P+\frac{\mathbf{B}^2}{2\mu_0} \end{bmatrix} \tag{3.19}$$

Thus the pressure tensor is diagonal in the parallel/perpendicular basis. Furthermore, the effective pressure is smaller in the parallel direction ($T_{xx}=P-\mathbf{B}^2/(2\mu_0)$) and larger in the perpendicular directions ($T_{yy}=T_{zz}=P+\mathbf{B}^2/(2\mu_0)$). This means that field lines have "tension": it takes energy to displace or stretch them. This is a bit hard to visualize. The reverse situation, where the parallel pressure would be higher, would be much easier to visualize, because it would lead to spontaneous wiggling of the field line when it would try to maxime its parallel length. This kind of situation can actually arise, but it requires that the fluid pressure P it itself a tensor and such that its parallel component exceeds the magnetic counterterm. The associated instability is called the firehose instability. To analyze it, we need to go beyond ideal MHD, however.

It is left as an exercise (HT) to show that the conservation form of the MHD equations is

$$
\begin{aligned}
\frac{\partial \rho}{\partial t} &= -\nabla\cdot\mathbf{p} \\
\frac{\partial \mathbf{p}}{\partial t} &= -\nabla\cdot\left[\frac{\mathbf{p}\mathbf{p}}{\rho}+\left(P+\frac{\mathbf{B}^2}{2\mu_0}\right)\mathtt{I}-\frac{1}{\mu_0}\mathbf{B}\mathbf{B}\right] \\
\frac{\partial u}{\partial t} &= -\nabla\cdot\left[\left(u+P-\frac{\mathbf{B}^2}{2\mu_0}\right)\frac{\mathbf{p}}{\rho}+\frac{1}{\mu_0\rho}\mathbf{B}\times(\mathbf{p}\times\mathbf{B})\right] \\
\frac{\partial \mathbf{B}}{\partial t} &= \nabla\times\left(\frac{\mathbf{p}}{\rho}\times\mathbf{B}\right)
\end{aligned}
\tag{3.20}
$$

where the dynamic (conserved) variables are now $(\rho, \mathbf{p}=\rho\mathbf{v}, u, \mathbf{B})$ and

$$
u \equiv \frac{1}{\gamma-1}P+\frac{\mathbf{p}^2}{2\rho}+\frac{\mathbf{B}^2}{2\mu_0} \tag{3.21}
$$

is now the total (thermal + bulk kinetic + magnetic) energy density. Of course, the total energy density must now contain the magnetic component.

3.4 The frozen-in theorem

Using Faraday's law, Equation (3.11), as an evolution equation for the magnetic field,

$$
\frac{\partial \mathbf{B}}{\partial t} = \nabla\times(\mathbf{v}\times\mathbf{B}) \tag{3.22}
$$

has some fundamental consequences. The first one is the famous "frozen-in" theorem. We shall now derive this beautiful and important result. We start by establishing the following mathematical theorem:

Theorem *Let* $\mathbf{B}$ *be a divergence-free vector field* ($\nabla\cdot\mathbf{B}=0$)*. Let* $\mathbf{v}$ *be a vector field (the velocity). Let* S *be an open surface moving with velocity* $\mathbf{v}$*, and let* $\Phi\equiv\int_S d\mathbf{S}\cdot\mathbf{B}$ *be the flux of* $\mathbf{B}$ *through the surface. Then*

$$
\frac{d\Phi}{dt} = \int_S d\mathbf{S}\cdot\left[\frac{\partial \mathbf{B}}{\partial t}-\nabla\times(\mathbf{v}\times\mathbf{B})\right] \tag{3.23}
$$

Proof. Compute $d\Phi/dt$ as $(\Phi(t')-\Phi(t))/\delta t$ where $t' = t+\delta t$. Assume first that $\mathbf{v}\cdot d\mathbf{S}\geq 0$ everywhere along the surface. During δt the surface S sweeps an infinitesimal volume slice V. According to the divergence theorem, the flux of $\mathbf{B}$ through ∂V is zero because $\mathbf{B}$ is divergenceless. On the other hand, the flux through ∂V can be written as

$$
0 = \int_V d\mathbf{S}\cdot\mathbf{B} = \left(\int_{S'}-\int_S+\int_{\text{strip}}\right)d\mathbf{S}\cdot\mathbf{B} \tag{3.24}
$$

where $S' = S(t')$ and "strip" is a closed strip of infinitesimal width $|\mathbf{v}|\delta t$ which is stretched between ∂S and $\partial S'$. The outer normal area element of the strip can be written as a cross product,

$$
d\mathbf{S} = d\mathbf{l}\times\mathbf{v}\delta t \tag{3.25}
$$

(Here we needed the assumption $\mathbf{v}\cdot d\mathbf{S}\geq 0$.) Now,

$$
\begin{aligned}
\int_S d\mathbf{S}\cdot\mathbf{B} &= \Phi(\mathbf{B};t) \\
\int_{S'} d\mathbf{S}\cdot\mathbf{B} &= \Phi(\mathbf{B};t').
\end{aligned}
$$

Thus we obtain

$$\begin{aligned} 0 &= \Phi(\mathbf{B};t') - \Phi(\mathbf{B};t) + \delta t \int_{\partial S} (d\mathbf{l}\times\mathbf{v})\cdot\mathbf{B} \\ &= \Phi(\mathbf{B};t') - \Phi(\mathbf{B};t) + \delta t \int_{\partial S} d\mathbf{l}\cdot\mathbf{v}\times\mathbf{B}, \end{aligned}$$

Notice that because the strip has infinitesimal width, we could transform the flux through it to a line integral over ∂S. Thus we finally obtain

$$\begin{aligned} \frac{d\Phi}{dt} &= \frac{\Phi(\mathbf{B}';t') - \Phi(\mathbf{B};t)}{\delta t} \\ &= \frac{\Phi\left(\mathbf{B} + \delta t \frac{\partial \mathbf{B}}{\partial t};t'\right) - \Phi(\mathbf{B};t)}{\delta t} \\ &= \frac{\Phi(\mathbf{B};t') + \delta t \Phi(\frac{\partial \mathbf{B}}{\partial t};t') - \Phi(\mathbf{B};t)}{\delta t} \\ &= \Phi(\frac{\partial \mathbf{B}}{\partial t};t') + \frac{\Phi(\mathbf{B};t') - \Phi(\mathbf{B};t)}{\delta t} \\ &= \Phi(\frac{\partial \mathbf{B}}{\partial t};t) - \int_{\partial S} d\mathbf{l}\cdot\mathbf{v}\times\mathbf{B} \\ &= \int_S d\mathbf{S}\left(\frac{\partial \mathbf{B}}{\partial t} - \nabla\times(\cdot\mathbf{v}\times\mathbf{B})\right) \end{aligned}$$

where we used Stokes' theorem in the last step. This completes the proof for $\mathbf{v}\cdot d\mathbf{S} \geq 0$. If $\mathbf{v}\cdot d\mathbf{S} < 0$, there will be a sign change in all three terms of Equation (3.24), and thus the same result follows. Finally, if $\mathbf{v}\cdot d\mathbf{S}$ has both signs, we must divide the surface S in pieces where $\mathbf{v}\cdot d\mathbf{S}$ does not change sign. The theorem can then be applied to each piece, Q.E.D.

Combining the above mathematical theorem with Faraday's law, we obtain the famous **frozen-in theorem for ideal MHD**: *the magnetic flux through any co-moving surface stays constant in time.* It is often said that the magnetic field lines are **frozen in** the plasma in ideal MHD. However, one must realize that **speaking of motion of field lines is inherently dangerous**, because in order to define (draw) field lines one has to choose their starting points. If one requires that the density of field lines is inversely proportional to the field strength, the choice of starting points for field line plotting is no longer completely arbitrary, but it is still not unique. To define what it means that a field line "moves" from A to B in time Δt one has to choose the field line starting point at time t and $t + \Delta t$. One natural possibility is to let the starting points wander with the plasma flow, but then the frozen-in statement starts looking like circular reasoning!

One can see clearly the ambiguities of the frozen-in theorem if one consider plasma flow in constant magnetic field. If field lines are "frozen in the plasma", they should move with the plasma flow. However, one could equally well argue that field lines of a constant magnetic field should move nowhere! Of course, this ambiguity has no physical content but is only due to the loose speak of field line motion. For these reasons it is best to stick with the original (and proved) form of the theorem: nothing is said about field lines, but only that the magnetic flux through any co-moving surface stays constant in time.

The frozen-in theorem is a natural, geometric consequence of Faraday's law and vanishing of rest-frame electric field due to infinite conductivity.

Another theorem which is closely related is the following: *Fluid elements which lie on a common magnetic field line at $t = 0$ continue to do so for all $t > 0$, if ideal MHD is valid.* The proof is also related to the frozen-in theorem, see e.g. the book by *Boyd and Sanderson* (HT).

Home Tasks for MHD

1. Derive the conservative form of Euler's equations, that is, equations (3.6) through (3.8).

2. Verify the conservative form of the MHD equations: Show that system (3.18) follows if (3.20) is in force.

3. What is the conserved flux in Faraday's law, $\partial_t \mathbf{B} = \nabla\times(\mathbf{v}\times\mathbf{B})$, or, more generally, $\partial_t \mathbf{B} = -\nabla\times\mathbf{E}$?

4a. Compute the time derivative of a volume integral of a quantity n,

$$\frac{d}{dt}\int_{V(t)} dV n = ?$$

where the volume $V(t)$ moves at velocity $\mathbf{v}$. Apply to the situation where n is the density and $\mathbf{v}$ is the plasma flow velocity field. Interpretation?

4b. Generalize the theorem given in this chapter to the case of nonzero $\nabla\cdot\mathbf{B}$. That is, compute

$$\frac{d}{dt}\int_{S(t)} d\mathbf{S}\cdot\mathbf{B}$$

where $\mathbf{B}$ is now a general vector field and surface S is moving according to velocity field $\mathbf{v}$.

5a. A quantity X which satisfies $dX/dt = 0$ is sometimes called a passive scalar. It is constant along streamlines. Show that if a quantity a satisfies

$$\frac{\partial a}{\partial t} = -\nabla\cdot(a\mathbf{v}),$$

the ratio ρ/a is a passive scalar.

5b. Show that if the term $-\mathbf{v}\nabla\cdot\mathbf{B}$ is added to the RHS of Faraday's law, then $\nabla\cdot\mathbf{B}/\rho$ is propagated as a passive scalar.

6. Prove the second theorem mentioned in this chapter: Fluid elements which lie on a common magnetic field line at $t=0$ continue to do so for all $t>0$, if ideal MHD is valid.

Chapter 4

MHD waves

4.1 Small perturbations of a flux conservative system

Consider a system of equations which is in flux conservative form,

$$\frac{\partial u_a}{\partial t} = -\nabla \cdot \mathbf{F}_a(u) \tag{4.1}$$

where $\mathbf{F}(u)$ is an algebraic function of u. Here a labels the dynamic variables; $a = 1..8$ in case of ideal MHD. Assume that $u^{(0)}$ is a stationary solution and develop $u = u^{(0)} + u^{(1)} \exp[i(\mathbf{k} \cdot \mathbf{x} - \omega t)]$. Here $u^{(1)}$ is a small vector of perturbation amplitudes. We obtain

$$\mathbf{F}_a(u) = \mathbf{F}_a(u^{(0)}) + \sum_b u_b^{(1)} \exp[i(\mathbf{k} \cdot \mathbf{x} - \omega t)] \frac{\partial \mathbf{F}_a(u^{(0)})}{\partial u_b^{(0)}} \tag{4.2}$$

Substituting into Equation (4.1) and noting that $\nabla \cdot \mathbf{F}^{(0)}$ vanishes yields

$$-i\omega u_a^{(1)} = -i\mathbf{k} \cdot \sum_b \mathbf{A}_{ab} u_b^{(1)} \tag{4.3}$$

where $\mathbf{A}_{ab} \equiv \partial \mathbf{F}_a / \partial u_b^{(0)}$ is the flux Jacobian matrix evaluated at the unperturbed state $u^{(0)}$. Writing this in another form

$$\sum_b \hat{\mathbf{k}} \cdot \mathbf{A}_{ab} u_b^{(1)} = \frac{\omega}{k} u_a^{(1)} \tag{4.4}$$

reveals that we are dealing with an eigenvalue problem: the phase velocity $v_{ph} \equiv \omega/k$ must be an eigenvalue of the matrix $(\hat{\mathbf{k}} \cdot \mathbf{A}_{ab})$. Notice that we denote spatial vectors by boldface letters while reserving index notation $a, b, \ldots$ for the dynamical variable space (internal space) objects. If we need to use index notation in 3-space we use the indices $i, j, k, \ldots$. Some authors use boldface for internal space objects and index notation for 3-space, so be careful with the literature here. In case of MHD, the quantity $\mathbf{A}_{ab}$ has a total of $3 \times 8 \times 8 = 196$ components. The presence of the unit vector $\hat{\mathbf{k}}$ in Equation (4.4) means that the wave propagation properties in general depend on the direction, i.e. the medium defined by the stationary solution is not isotropic.

Solving for the eigenvalues and eigenvectors of a given flux conservative system is useful in many respects. In some physical situations we may indeed observe waves which are small perturbations of a simple stationary state (this was, of course, the original purpose of the derivation). In numerical simulations we need to know what is the largest phase speed in the system, because the simulation time step must be smaller than the fastest wave travel time across the smallest grid spacing. The largest eigenvalue gives us exactly this information. In more advance numerical simulations we need to perform the decomposition into fundamental wave modes explicitly in order to know what is the "upwind" direction for each wave mode. This challenging task calls for solving the eigenvalue problem locally at every grid point at every time step, and doing it fast because it has to be repeated billions of times.

4.2 MHD eigenmodes

Direct analytic computation of the MHD eigenmodes is a very difficult task. Just the starting point, the flux Jacobian 8×8 matrix, is a large expression. Even modern symbolic algebra program run into serious difficulties if one tries to solve the eigenprogram in a straightforward way (HT). However, the set of eigenvalues can be written down in surprisingly compact way; the eigenvectors are a bit more complicated. Of course, it is much easier to solve the eigenproblem numerically: finding the eigenvalues and eigenvectors of a 8-by-8 general matrix is an easy task for the EISPACK or LAPACK library, for example. It is even easier to verify numerically that the analytically produced eigenvalues and eigenvectors are really that — eigenvalues and eigenvectors.

The MHD eigenvalues are given by

$$\lambda_e = v_k \qquad \text{Entropy wave} \tag{4.5}$$

$$\lambda_A = v_k \pm v_A|\cos\alpha| \qquad \text{Alfven wave} \tag{4.6}$$

$$\lambda_f = v_k \pm v_f \qquad \text{Fast wave} \tag{4.7}$$

$$\lambda_s = v_k \pm v_s \qquad \text{Slow wave} \tag{4.8}$$

$$\lambda_0 = 0 \qquad \text{Null wave} \tag{4.9}$$

where v_A is the scalar Alfvén speed,

$$v_A = \frac{|\mathbf{B}|}{\sqrt{\mu_0 \rho}}, \tag{4.10}$$

v_f is the so-called fast wave speed,

$$v_f^2 = \frac{1}{2}\left[c_s^2 + v_A^2 + \sqrt{\left(c_s^2 + v_A^2\right)^2 - 4c_s^2 v_A^2 \cos^2\alpha}\right] \tag{4.11}$$

and v_s is the so-called slow wave speed,

$$v_s^2 = \frac{1}{2}\left[c_s^2 + v_A^2 - \sqrt{\left(c_s^2 + v_A^2\right)^2 - 4c_s^2 v_A^2 \cos^2\alpha}\right]. \tag{4.12}$$

Notice that the expressions for v_f and v_s differ only by the sign of the square root subexpression (the discriminant). In these expressions c_s is the fluid sound speed, $c_s^2 = \gamma P/\rho$ and α is the angle between the wave vector $\mathbf{k}$ and the magnetic field $\mathbf{B}$: $\cos\alpha = \hat{\mathbf{k}} \cdot \hat{\mathbf{b}}$. v_k is the projection of the fluid velocity $\mathbf{v}$ along the wave vector, $v_k = \hat{\mathbf{k}} \cdot \mathbf{v}$.

Let us consider various limiting cases. First, if we put $\mathbf{B}$ to zero we should get the ordinary gasdynamic waves. Is this so? The Alfvén velocity v_A is zero then, and thus v_s vanishes. On the other hand v_f becomes equal to c_s. Thus the slow wave and the Alfvén wave both degenerate with the entropy wave. The fast wave becomes the ordinary sound wave.

For parallel waves ($\mathbf{k}||\mathbf{B}$) $\cos\alpha = 1$ and the fast and slow wave speeds become $v_f = \max(c_s, v_A)$ and $v_s = \min(c_s, v_A)$. Thus one of magnetosonic modes (the fast and slow waves together are called magnetosonic waves) degenerates with the Alfvén mode and the other one becomes the ordinary sound wave.

For perpendicular waves $\cos\alpha = 0$. In this case we get $v_f = \sqrt{c_s^2 + v_A^2}$ and $v_s = 0$. Thus the slow wave and the Alfvén wave both degenerate with the entropy wave, but the fast magnetosonic wave remains.

The null wave and the entropy wave may seem strange at first. The null wave is related to the fact that $\nabla \cdot \mathbf{B}$ does not change due to Faraday's law. It is not a real physical wave. The entropy wave is also found in fluid dynamics. It represents features that propagate with

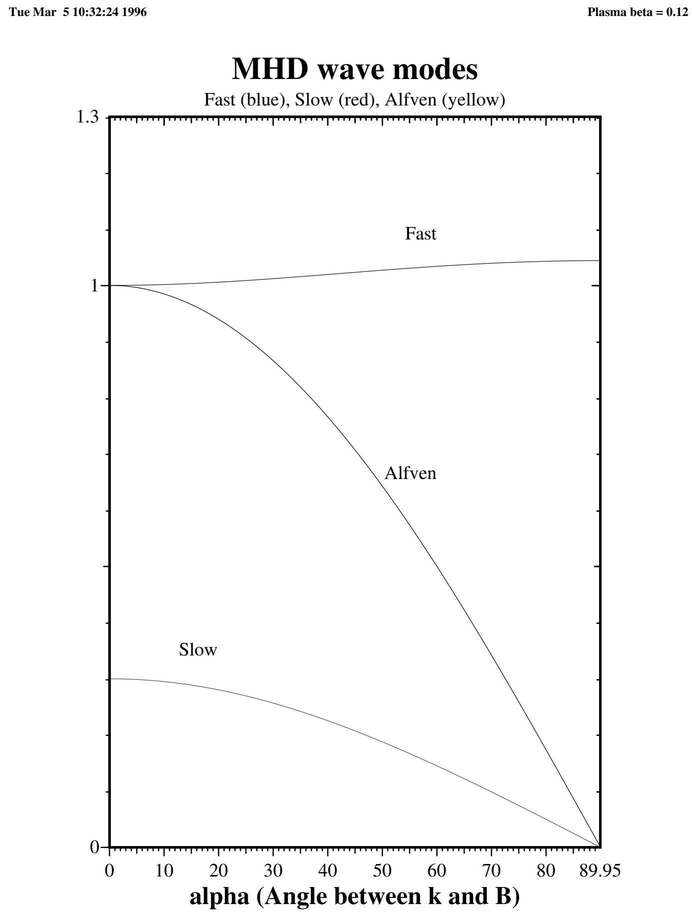

Figure 4.1: MHD wave speeds in low-beta plasma (near-Earth region), relative to v_A

the fluid velocity. All everyday fluid phenomena excluding sound waves, e.g. weather fronts, are related with the entropy wave. The same is true in the magnetosphere: the long-term evolution is characterized by the entropy wave, although we do not usually think of such processes in terms of "waves".

Figure 4.1 displays the wave speeds relative to v_A in a low-beta plasma. In low-beta plasma the Alfvén speed v_A is significantly higher than the fluid sound speed c_s. The background fluid velocity has been put to zero for simplicity. The near-Earth magnetosphere contains low-beta plasma because of the strong dipole field.

Figure 4.2 shows the wave modes in conditions where the fluid sound speed and Alfvén speed as exactly equal. These are the typical conditions in the solar wind at 1 AU. (However, the solar wind plasma can move to both low and high beta regimes in the parameter space – unit beta represents the *average* situation.)

Figure 4.3 shows the wave modes in high-beta plasma. The magnetotail current sheet is high-beta plasma. Notice that in high-beta plasma the slow wave and the Alfvén wave are nearly indistinguishable.

Remember that in every case and for all propagation angles, the fast wave is the fastest and the slow wave is the slowest. The Alfvén wave is in between. Degeneracies occur for purely parallel and perpendicular propagation.

Notice that the fluid sound speed c_s is only a symbol appearing in our equations. None of the MHD eigenmodes propagates with exactly this velocity, except in the purely parallel direction.

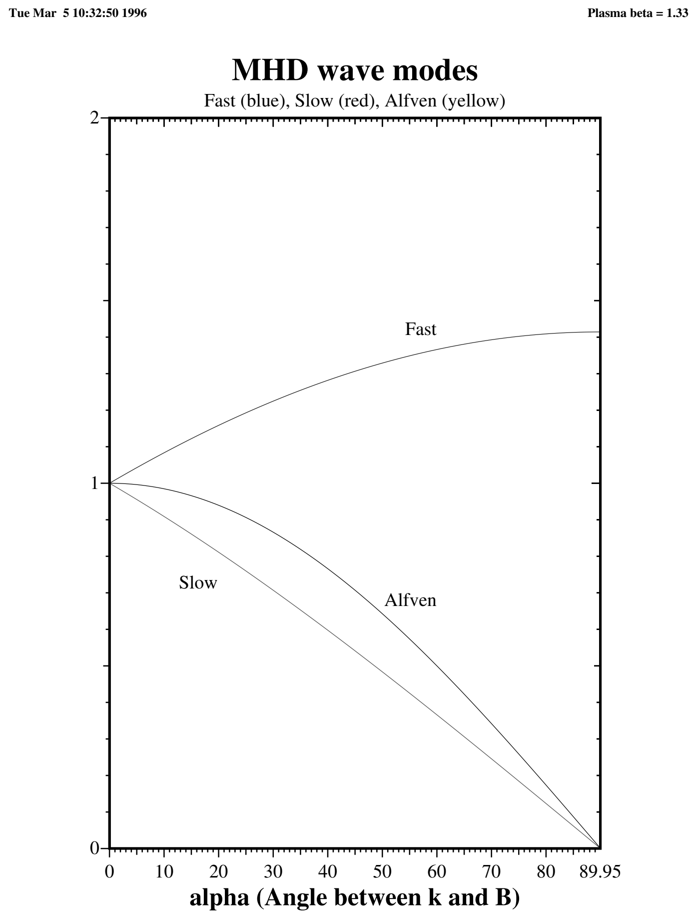

Figure 4.2: MHD wave speeds in plasma with beta close to unity (solar wind)

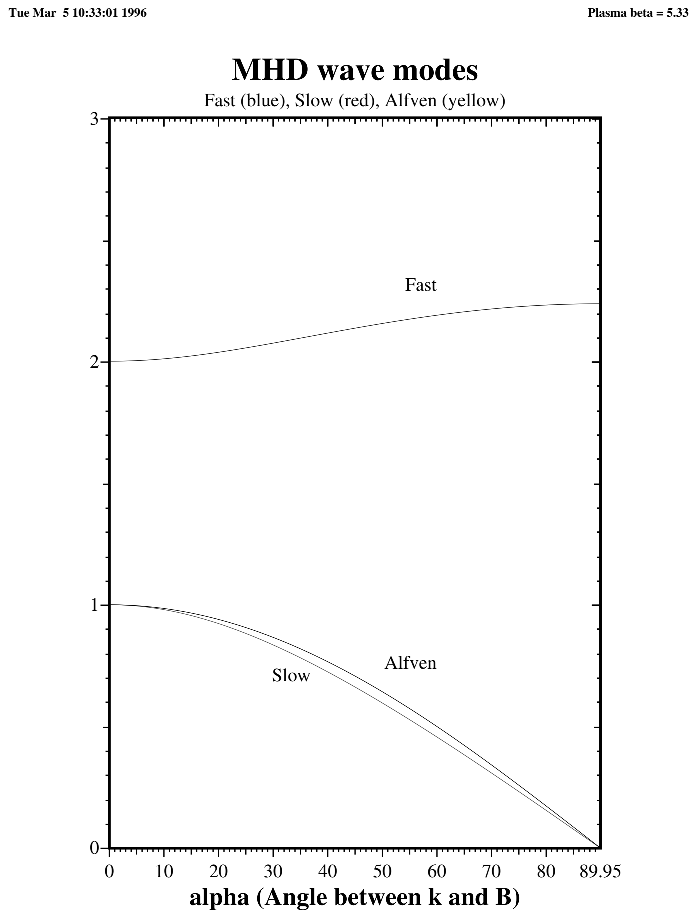

Figure 4.3: MHD wave speeds in high-beta plasma (magnetotail)

4.3 Left and right eigenvectors

In quantum mechanics etc. matrices and operators are always Hermitian. The eigenvalues are therefore real, and the eigenvectors can be made mutually orthonormal. The flux Jacobian matrix, being a matrix of partial derivatives, is usually not symmetric. The present discussion follows closely the treatment found in *Numerical Recipes*.

Let A be a general, nonsymmetric matrix. The "usual" eigenproblem is

$$Ar = \lambda r. \tag{4.13}$$

The eigenvectors r are called *right eigenvectors*. *Left eigenvectors* l of A are, on the other hand, solutions to the following eigenproblem:

$$lA = \lambda l. \tag{4.14}$$

By taking the transpose of Equation (4.14) we see that every left eigenvector is the transpose of a right eigenvector *of the transpose of* A. Because $\det A^T = \det A$, the left and right *eigenvalues* of A are identical.

If the matrix A were symmetric, then the left and right eigenvectors are just transposes of each other (they have the same numerical values as components). In the general case, we have the following calculation. Let R be the matrix formed by columns from the right eigenvectors, that is, $R_{ij} = (r_j)_i$. Let L be the matrix formed by rows from the left eigenvectors, $L_{ij} = (l_i)_j$. Then Equation (4.13) and Equation (4.14) can be rewritten as

$$\begin{aligned}
(AR)_{ij} &= \sum_k A_{ik} R_{kj} = \sum_k A_{ik} (r_j)_k \\
&= (Ar_j)_i = (\lambda_j r_j)_i = \lambda_j R_{ij} \\
&= \sum_k R_{ik} \delta_{kj} \lambda_j = \sum_k R_{ik} \Lambda_{kj} \\
&= (R\Lambda)_{ij} \Rightarrow AR = R\Lambda.
\end{aligned} \tag{4.15}$$

where $\Lambda = diag(\lambda_1, ..., \lambda_n)$. Similarly we can show that

$$LA = \Lambda L. \tag{4.16}$$

Multiplying Equation (4.15) from the left by L and multiplying Equation (4.16) from the right by R and subtracting gives

$$(LR)\Lambda = \Lambda(LR). \tag{4.17}$$

Thus the matrix of dot products of the left and right eigenvectors commutes with the diagonal matrix of eigenvalues. But the only matrices that commute with a diagonal matrix of distinct elements are themselves diagonal. Thus, if the eigenvalues are nondegenerate, each left eigenvector is orthogonal to all right eigenvectors except its corresponding one, and vice versa. By choice of normalization we can make the vectors orthonormal, i.e.

$$l_i \cdot r_j = \delta_{ij} \qquad \text{or} \qquad LR = \mathtt{I}. \tag{4.18}$$

Thus L and R are inverse matrices of each other. We therefore obtain the following representation of the matrix A:

$$A = R\Lambda R^{-1} = R\Lambda L. \tag{4.19}$$

In advanced numerical simulations, the reason why we want to make this decomposition is that we want to use so-called *upwind* differencing schemes. For scalar problems this just means that spatial discretization is not symmetric but weights the upstream direction. For PDE systems, however, each eigenmode has its own up- and downstream direction. (For example, there are two Alfvén eigenmodes, one propagating parallel and the other one antiparallel to the background magnetic field.) Finding out which direction is the proper upwind direction for each eigenmode boils down to the computation of $|A|$, which is *defined* by

$$|A| \equiv R|\Lambda|L. \tag{4.20}$$

Thus, after transforming to the eigenbasis of A the identification of the proper upwind direction is trivial. Even though the MHD equations are nonlinear, the separation to linear eigenmodes works rather well because the equations are not linearized with respect to some global zeroth order solution. Instead, the expansion is carried out at every grid cell locally.

In the following we give the MHD right eigenvectors. These correspond to the situation where the waves propagate in the x direction only. This is not an essential restriction because all components of the vector fields are included. The variable ordering is $(\rho, p_x, p_y, p_z, u, B_y, B_z)$, seven components alltogether. The B_x component has been dropped. If included, the corresponding component in the right eigenvectors would be zero all the time. Notice that we have seven, not eight, eigenvectors: one entropy wave eigenvector, two eigenvectors for $+x$ and $-x$ propagating Alfvén waves, two eigenvectors for fast waves, and two eigenvectors for slow waves.

$$r^{\text{entropy}} = \begin{bmatrix} 1 \\ v_x \\ v_y \\ v_z \\ \frac{1}{2}\mathbf{v}^2 \\ 0 \\ 0 \end{bmatrix} \tag{4.21}$$

$$r^{\text{Alfven}} = \begin{bmatrix} 0 \\ 0 \\ \mp\tilde{v}_z^A \,\text{sign}(B_x) \\ \pm\tilde{v}_y^A \,\text{sign}(B_x) \\ \mp(\tilde{v}_z^A v_y - \tilde{v}_y^A v_z)\text{sign}(Bx) \\ v_z^A\sqrt{\mu_0/\rho} \\ -v_y^A\sqrt{\mu_0/\rho} \end{bmatrix} \tag{4.22}$$

$$r^{\text{fast}} = \begin{bmatrix} a_f \\ a_f(v_x \pm v_f) \\ a_f v_y \mp a_s \tilde{v}_y^A v_f v_x^A \\ a_f v_z \mp a_s \tilde{v}_z^A v_f v_x^A \\ \frac{1}{2}a_f\mathbf{v}^2 + \frac{a_f v_f^2}{\gamma-1} \pm a_f v_f v_x \mp a_s v_f(\tilde{v}_y^A v_y + \tilde{v}_z^A v_z)v_x^A + \frac{\gamma-2}{\gamma-1}a_f(v_f^2 - c_s^2) \\ a_s\tilde{v}_y^A v_f^2\sqrt{\mu_0/\rho} \\ a_s\tilde{v}_z^A v_f^2\sqrt{\mu_0/\rho} \end{bmatrix} \tag{4.23}$$

$$r^{\text{slow}} = \begin{bmatrix} a_s \\ a_s(v_x \pm v_s) \\ a_s v_y \pm a_f\tilde{v}_y^A \text{sign}(B_x)c_s/v_f \\ a_s v_z \pm a_f\tilde{v}_z^A \text{sign}(B_x)c_s/v_f \\ \frac{1}{2}a_s\mathbf{v}^2 + \frac{a_s v_s^2}{\gamma-1} \pm a_s v_s v_x \pm a_f(\tilde{v}_y^A v_y + \tilde{v}_z^A v_z)\text{sign}(B_x)c_s/v_f + \frac{\gamma-2}{\gamma-1}a_s(v_a^2 - c_s^2) \\ -a_f\tilde{v}_y^A\sqrt{\mu_0/\rho}(c_s/v_f)^2 \\ -a_f\tilde{v}_z^A\sqrt{\mu_0/\rho}(c_s/v_f)^2 \end{bmatrix}. \tag{4.24}$$

Here we have introduced the following abbreviations:

$$v_x^A \equiv \frac{B_x}{\sqrt{\mu_0\rho}} \tag{4.25}$$

$$v_y^A \equiv \frac{B_y}{\sqrt{\mu_0\rho}} \tag{4.26}$$

$$v_z^A \equiv \frac{B_z}{\sqrt{\mu_0 \rho}} \tag{4.27}$$

$$\tilde{v}_y^A \equiv \frac{v_y^A}{\sqrt{(v_y^A)^2 + (v_z^A)^2}} \tag{4.28}$$

$$\tilde{v}_z^A \equiv \frac{v_z^A}{\sqrt{(v_y^A)^2 + (v_z^A)^2}}. \tag{4.29}$$

$$a_f \equiv \sqrt{\frac{v_f^2 - (v_x^A)^2}{v_f^2 - v_s^2}} \tag{4.30}$$

$$a_s \equiv \frac{1}{v_f}\sqrt{\frac{v_f^2 - c_s^2}{v_f^2 - v_s^2}} \tag{4.31}$$

Notice that the dimensionality of all v quantities is velocity. The a_f parameter is dimensionless. The a_s parameter has dimensionality of inverse velocity (sorry, unfortunate notation).

We now look at the above right eigenvectors in the case when the background plasma velocity $\mathbf{v}$ vanishes. This assumption will considerably simplify the expressions and can be always fulfilled by going into the co-moving coordinate frame:

$$r^{\text{entropy}} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{4.32}$$

$$r^{\text{Alfven}} = \begin{bmatrix} 0 \\ 0 \\ \mp\tilde{v}_z^A \operatorname{sign}(B_x) \\ \pm\tilde{v}_y^A \operatorname{sign}(B_x) \\ 0 \\ v_z^A \sqrt{\mu_0/\rho} \\ -v_y^A \sqrt{\mu_0/\rho} \end{bmatrix} \tag{4.33}$$

$$r^{\text{fast}} = \begin{bmatrix} a_f \\ \pm a_f v_f \\ \mp a_s \tilde{v}_y^A v_f v_x^A \\ \mp a_s \tilde{v}_z^A v_f v_x^A \\ \frac{a_f v_f^2}{\gamma-1} + \frac{\gamma-2}{\gamma-1} a_f (v_f^2 - c_s^2) \\ a_s \tilde{v}_y^A v_f^2 \sqrt{\mu_0/\rho} \\ a_s \tilde{v}_z^A v_f^2 \sqrt{\mu_0/\rho} \end{bmatrix} \tag{4.34}$$

$$r^{\text{slow}} = \begin{bmatrix} a_s \\ \pm a_s v_s \\ \pm a_f \tilde{v}_y^A \operatorname{sign}(B_x) c_s/v_f \\ \pm a_f \tilde{v}_z^A \operatorname{sign}(B_x) c_s/v_f \\ \frac{a_s v_s^2}{\gamma-1} + \frac{\gamma-2}{\gamma-1} a_s (v_a^2 - c_s^2) \\ -a_f \tilde{v}_y^A \sqrt{\mu_0/\rho} (c_s/v_f)^2 \\ -a_f \tilde{v}_z^A \sqrt{\mu_0/\rho} (c_s/v_f)^2 \end{bmatrix}. \tag{4.35}$$

From these equations we can see that the entropy wave changes only the plasma density. The Alfvén wave does not modify the density or the total energy density u; it also leaves the parallel velocity intact. The magnetosonic modes modify all quantities. Thus the Alfvén waves are transverse oscillations of the perpendicular components of the velocity and magnetic field. The magnetosonic modes are the most complicated ones. The perturbations of the gas pressure and the magnetic pressure add up in the fast wave. In the slow wave they are in anti-phase.

It is not very easy to study the behavior of the above right eigenvectors in the case of purely parallel and purely perpendicular propagation, because indeterminate forms 0/0 occur then.

Here we have given the expressions only for the right eigenvectors. The left eigenvector expressions are more or less similar. If one wants to achieve the normalization $l_i \cdot r_j = \delta_{ij}$, the normalization constant for the slow eigenvectors becomes very complicated, however, and thus the slow eigenvector normalization is best done numerically afterwards.

4.4 Shocks and other discontinuities in MHD

In gasdynamics, a shock (discontinuity) forms as soon as supersonic velocity differences occur. This is because information cannot travel faster than c_s in such a case and when a fast flow hits an obstacle, a bow shock forms.

Earth's magnetosphere also has a bow shock. The solar wind velocity exceeds the fastest MHD wave mode (that is, v_f); thus a shock forms, and the formation mechanism is similar to gasdynamics. Although the solar wind velocity varies rather lot, it apparently always is larger than v_f. On the sunward side of the bow shock the flow is completely undistorted by the obstacle (the magnetosphere). Between the bow shock and the magnetopause (which can be regarded as the obstacle surface in this case) the flow turns around the obstacle and slows down. At the same time, the density, the temperature and the magnetic field increase. As the density and the temperature increase, the gas sound velocity c_s increases. Thus the flow loses its supersonic character at some point (the fluid velocity and the sound velocity become equal). Increasing the density descreases the Alfvén speed which tends to keep the flow in the super-Alfvénic regime, but the pileup of the magnetic field more than compensates this effect.

Inside the magnetosphere the situation is more complicated. Because we have three kinds of eigenmodes in MHD, we can have at least three kinds of shocks. For example, the flow velocity might be larger than the slow velocity but smaller than the Alf̌en and fast velocities. It is probable, however, that the coupling along the magnetic field makes these "shocks" more irrelevant, because the flow usually cannot stay supersonic at every point along a field line. For example, the flow speed in the neutral sheet might exceed the local Alf̌en speed, but if we follow the field line at lower altitudes, the local Alf̌en speed eventually must become so large that it exceeds the local flow velocity. (This is even more so because the $\mathbf{E}\times\mathbf{B}$ velocity also gets reduce.) This question appears to be an interesting one, but for some reason it has been studied hardly at all. Overall, the role of shocks in magnetospheric physics remains an obscure one.

Weather fronts are familiar examples of what happens when fluids of different internal properties (temperature, humidity) are in contact. These structures are not shocks. However, in the large scale they can still be approximated as discontinuities. Similar structures probably exist in the magnetosphere, too. For example, auroral arcs could possibly be manifestations of such phenomena.

One uses the concept of *Rankine-Hugoniot* relations with shocks, which are actually quite simple. These relations are just the integrated forms of the flux-conservative equations, which state that the mass, momentum and energy are conserved across a shock or any other kind of discontinuity.

The most important thing to remember about MHD waves is the qualitative behavior of the phase velocities of the wave modes.

Home Tasks for MHD waves

1. Verify the MHD eigenvalues [Eqs. (4.5) through (4.8)] using a symbolic algebra program. Any typos?

2. This home task requires a symbolic algebra program, too. Write a general function which computes and verifies the eigenvalues and the right and left eigenvectors of any given PDE system. Apply your code to Euler equations. Identify the resulting eigenmodes. Finally, try to apply your program to the full MHD case. What happens? (Something, for sure, must happen!)

3. Derive the Alfvén eigenmode (eigenvalue and eigenvector) for parallel propagation. Start from the MHD equations. This task can be done by pencil and paper.

Chapter 5

Ionospheric electrodynamics

5.1 Pedersen and Hall conductivities

The two-fluid equations, written for ions and electrons, are

$$m_s(\frac{\partial \mathbf{v}_s}{\partial t} + \mathbf{v}_s \cdot \nabla \mathbf{v}_s) = q_s(\mathbf{E} + \mathbf{v}_s \times \mathbf{B}) - \frac{\nabla P_s}{n_s} - \nu_{sn} m_s \mathbf{v}_s$$

where $s = i, e$ and ν_{en} and ν_{in} are the electron-neutral and ion-neutral collision frequencies, respectively. The pressures P_s are equal to $n_s k_B T_s$. The ion density n_i can be assumed equal to the electron density n_e to great accuracy (the Debye length is of the order of centimeters in the ionospheric E-region).

In stationary state and assuming homogeneous plasma $\nabla P_s = 0$, Equation (5.1) can be solved for the velocities (HT). The current density $\mathbf{j} = en(\mathbf{v}_i - \mathbf{v}_e)$ becomes linearly proportional to the electric field,

$$\mathbf{j} = \sigma \cdot \mathbf{E} \tag{5.1}$$

where σ is the conductivity tensor. In vector form,

$$\mathbf{j} = \sigma_P \mathbf{E}_\perp + \sigma_H \hat{\mathbf{B}} \times \mathbf{E} + \sigma_\parallel \mathbf{E}_\parallel \tag{5.2}$$

Here the Pedersen, Hall and parallel conductivities σ_P,σ_H,$\sigma_\parallel$ are defined by the well-known relations

$$\sigma_P = \sum_i \frac{en_i}{B} \left(\frac{\nu_{en}\Omega_e}{\Omega_e^2 + \nu_{en}^2} + \frac{\nu_{in}\Omega_i}{\Omega_i^2 + \nu_{in}^2} \right) \tag{5.3}$$

$$\sigma_H = \sum_i \frac{en_i}{B} \left(\frac{\Omega_e^2}{\Omega_e^2 + \nu_{en}^2} - \frac{\Omega_i^2}{\Omega_i^2 + \nu_{in}^2} \right) \tag{5.4}$$

$$\sigma_\parallel = \frac{e^2 n_e}{m_e(\nu_{en} + \nu_{in})} + \sum_i \frac{e^2 n_i}{m_i \nu_{in}} \tag{5.5}$$

These equations have been written for the case of several ion species i, such that $n_e = \sum_i n_i$.

We have thus far assumed that the neutral fluid is at rest. If a neutral wind is present, a simple Lorentz-transformation of the $\mathbf{E}$ field gives us the updated form of the Ohm's law:

$$\mathbf{j} = \sigma \cdot (\mathbf{E} + \mathbf{v}_n \times \mathbf{B}) \tag{5.6}$$

It should be noted that the ionospheric Ohm's law, as it is called, is conceptually very different from the MHD Ohm's law, which can be written as $\mathbf{j} = \sigma(\mathbf{E} + \mathbf{v} \times \mathbf{B})$ with certain assumptions. In the ionospheric Ohm's law it is the *neutral* gas velocity that appears in the Lorentz-transformation. In MHD, it is the plasma velocity. The question may arise if there is any smooth connection between the two. The answer is no, because when we move from the ionosphere to the magnetosphere the neutral collisions become less and less significant. Indeed, both σ_P and σ_H go to zero when neutral collisions are removed. The point is that we cannot compute $\mathbf{j}$ as $en(\mathbf{v}_i - \mathbf{v}_e)$ in MHD. The expressions for $\mathbf{v}_e$ and $\mathbf{v}_i$, however, give the result that the both ions and electrons move with $\mathbf{E}\times\mathbf{B}$ drift if there are no neutral collisions – as it should.

5.2 The electrostatic approximation

In the ionosphere it is a good approximation to assume a potential electric field, $\nabla\times\mathbf{E} = 0$ and $\mathbf{E} = -\nabla\Phi$. This can be justified as follows. The "curl–freeness" of the $\mathbf{E}$ field can be tested by estimating

$$\frac{|\nabla\times\mathbf{E}|}{E/L} \tag{5.7}$$

where L is a typical horizontal length scale for $\mathbf{E}$ variations. From Faraday's law,

$$|\nabla\times\mathbf{E}| = \left|\frac{\partial\mathbf{B}_1}{\partial t}\right| \tag{5.8}$$

where $\mathbf{B}_1$ is the variation geomagnetic field. (The main field $\mathbf{B}_0$ is constant in time, and $\mathbf{B} = \mathbf{B}_0 + \mathbf{B}_1$.) Thus we obtain

$$\frac{|\nabla\times\mathbf{E}|}{E/L} = \frac{\left|\frac{\partial\mathbf{B}_1}{\partial t}\right|}{E/L} \sim \frac{1}{T}\frac{B_1}{B_0}\frac{B_0 L}{E} = \frac{B_1}{B_0}\frac{L}{TV} \tag{5.9}$$

where T is a typical time scale and V is the magnitude of the $\mathbf{E}\times\mathbf{B}$ velocity. If we now assume that the spatial and temporal scales L and T are related by

$$\frac{L}{T} \sim V = \frac{E}{B_0} \tag{5.10}$$

then it follows that

$$\frac{|\nabla\times\mathbf{E}|}{E/L} \sim \frac{B_1}{B_0} \ll 1, \tag{5.11}$$

i.e., that $\mathbf{E}$ is approximately electrostatic since the variation field B_1 is much smaller than the main field B_0. Notice that the result is not necessarily valid if L/T much exceeds the $\mathbf{E}\times\mathbf{B}$ velocity. In practice this means that we cannot completely rule out the possibility of inductive ionospheric electric fields during very intense geomagnetic storms.

5.3 Elliptic PDE for the potential

Another approximation that can be safely made in the ionosphere is that the charge continuity equation can be written as $\nabla\cdot\mathbf{j} = 0$ (HT). In other words, the $\partial\rho/\partial t$ term can be neglected. Thus, the ionospheric Ohm's law [Equation (5.2)] implies

$$\nabla\cdot(\sigma\cdot\nabla\Phi) = 0 \tag{5.12}$$

Equation (5.12) is a second-order elliptic boundary value problem for the determination of Φ. If the conductivity tensor were constant (which it is never because the parallel conductivity is much larger then the other conductivities), it would reduce to Laplace's equation.

Equation (5.12) must be supplemented by boundary conditions. Suitable boundary conditions for elliptic problems include Dirichlet conditions (value of Φ specified at the boundary) and Neumann conditions (value of normal derivative of Φ specified). Solving the 3D equation is difficult for several reasons, however. Therefore one usually integrates it with respect to height. At least close to polar regions, where magnetic field lines are nearly vertical, this is a good approximation, if we are considering horizontal scales which are larger than the height of the main current-carrying layer.

Integrating Equation (5.12) with respect to height is simple. The result is (HT)

$$-\nabla \cdot (\Sigma \cdot \nabla \Phi) = j_{\|} \sin I \tag{5.13}$$

where I is the inclination angle of the magnetic field. Usually we can approximate $\sin I = 1$ in polar regions (outside polar regions the equation is not valid anyway).

Equation (5.13) is again an elliptic differential equation for the determination of Φ. The field-aligned current (FAC) is a source term. If Equation (5.13) is considered in a limited domain on the ionospheric plane, it must be supplemented by boundary conditions. (If the whole ionosphere is considered, the boundary conditions are natural, but there are still some conceptual problems of how to formulate them. We may return to these problems later if we have time.)

If the conductivity is horizontally homogeneous, Equation (5.13) reduces to Poisson's equation (HT),

$$-\nabla^2 \Phi = \frac{j_{\|}}{\Sigma_P}. \tag{5.14}$$

(assuming $\sin I = 1$). Thus the FAC plays the role of charge density in ionospheric electrodynamics. Even though this picture is not strictly valid for inhomogeneous conductivities, it is often accurate enough for qualitative considerations. Downward FAC means upward moving electrons (which generally means a low conductivity region). Downward FAC also means positive effective charge density, and thus a maximum of the potential Φ. On the other hand, upward FAC means downward moving electrons (a high conductivity region because the high-energy precipitating electrons create secondary electrons). Upward FAC also means negative charge density, and a local minimum of Φ.

Home Tasks for Ionospheric Electrodynamics

1. Derive the ionospheric Ohm'w law (Eqs. (5.3) through (5.5)) from the two-fluid equations.

2. Make it believable that one can write the continuity equation of electric charge in the form $\nabla \cdot \mathbf{j} = 0$ in the ionosphere. Any restrictions to spatial and temporal scales?

3. A fluid is called incompressible if the density is globally constant. For example, water is incompressible to a good approximation. Show that incompressible fluid satisfies $\nabla \cdot \mathbf{v} = 0$. Plasma is often called incompressible if it satisfies $\nabla \cdot \mathbf{v} = 0$. Show that ionospheric plasma is nearly incompressible. What can be said about the density in this case?

4. Derive Equation (5.13). Show that it reduces to Poisson's equation in the case of homogeneous conductivities.

Chapter 6

2D ionosphere-magnetosphere coupling

In the previous chapter we saw that the field-aligned current (FAC) acts as a source term (effective 2D charge density) for the ionospheric electrodynamics. Once the FAC and the height-integrated Pedersen and Hall conductivities are known, the electric potential can be solved from Equation (5.13). Thereafter, all other electrodynamic quantities (electric field, horizontal current) can be computed easily.

In this chapter we will consider two-dimensional ionosphere-magnetosphere coupling, where the magnetospheric plasma is assumed to be incompressible and to be confined in a two-dimensional layer which is perpendicular to the magnetic field. Such an approach is not a fully realistic description, but it is well suited for various analytic considerations which can be used to illustrate the main ideas of ionosphere-magnetosphere coupling. A three-dimensional realistic approach is presented later, which unfortunately is not so easy to understand analytically.

6.1 When the current is FAC

The height-integrated conductivity functions can often be estimated from models etc., but the FAC is difficult to known without knowing something about the source region: the magnetosphere. However, before considering this question, let us think why the current that comes from the magnetosphere should be necessarily field-aligned. The reason for this is the MHD force balance equation $\nabla P = \mathbf{j}\times\mathbf{B}$. Assume first that $\mathbf{j}$ is not field-aligned at all, $\mathbf{j} \equiv \mathbf{j}_\perp$. Because $\mathbf{j} = \nabla\times\mathbf{B}_1/\mu_0$, where $\mathbf{B}_1$ is the perturbation magnetic field (i.e., field without dipole field), we obtain after replacing the gradients by a length scale L,

$$\frac{P}{L} \sim \frac{B_1 B_0}{\mu_0 L}. \tag{6.1}$$

Thus

$$B_1 \sim \frac{\mu_0 P}{B_0}. \tag{6.2}$$

Recalling the definition of the plasma beta parameter, $\beta = P/(B^2/(2\mu_0))$, this is just

$$B_1 \sim \beta B_0. \tag{6.3}$$

Thus we get a limit for the perturbation magnetic field. Above ionosphere B_0 is approximately 10000 nT and the pressure P is generally smaller than 1 nPa. Thus β is smaller than 10^{-5}, which would limit the magnetic perturbations to less than 1 nT. As the real magnetic perturbations are much larger, often more than 100 nT, we have a contradiction with our original assumption that the current was perpendicular. We can then conclude that the only possibility is that the majority of the current is field-aligned near ionospheric heights.

Put another way, the current must be field-aligned because otherwise it would generate a $\mathbf{j}\times\mathbf{B}$ force far to strong to be cancelled by any pressure gradient in the system.

However, this result holds only at relatively low heights. When the distance from the Earth's center is more than $6R_E$, roughly speaking, then an appreciable portion of the current may be perpendicular. Thus the current which leaves or enters the ionosphere field-aligned need not stay field-aligned all the way to the equatorial plane (for example).

Thus we have shown that if current loops (current systems) traverse both the ionosphere and the magnetosphere, the current must be field-aligned when entering and leaving the ionosphere. Within the ionosphere itself the previous arguments do not hold, of course, and the current is mainly perpendicular (Pedersen and Hall current).

When the plasma beta is low, the most efficient MHD mode for energy transfer is the Alfvén wave. Other MHD eigenmodes depend on pressure and density variations, and since pressure cannot become negative, there is a limit for the amplitude of these waves. For Alfvén waves, however, such a limit does not exist because only the plasma velocity and the magnetic field take part in the oscillation.

We have thus shown that **in a low-beta plasma the most important currents flow in the field-aligned direction, the most important waves are Alfvén waves, and the magnetic field configuration is approximately a force-free one, satisfying $\mathbf{j}\times\mathbf{B} \sim 0$.**

6.2 2D coupling equations

We now derive a simple, even simplistic, model for ionosphere-magnetosphere interaction. Assume that the magnetospheric plasma is (a) incompressible, (b) two-dimensional in strong background magnetic field. The MHD momentum equation reads

$$\rho\left(\partial_t \mathbf{v} + \mathbf{v}\cdot\nabla\mathbf{v}\right) = \mathbf{j}\times\mathbf{B} - \nabla P. \tag{6.4}$$

Taking the curl gives (HT)

$$\frac{\partial\omega}{\partial t} = \frac{1}{B}\hat{z}\cdot\nabla\omega\times\nabla\Phi - \frac{B}{\rho}\nabla\cdot\mathbf{j}_\perp \tag{6.5}$$

where we have also used the ideal MHD Ohm's law $\mathbf{v} = \mathbf{E}\times\mathbf{B}/B^2$. ω is the vorticity function, $\omega = \hat{z}\cdot\nabla\times\mathbf{v}$. We have chosen our coordinates so that z points to the direction of the background magnetic field $\mathbf{B}$, and all dynamics is in xy plane. Further, assuming that the magnetospheric plasma is confined to a layer of thickness H enables us to write $\nabla\cdot\mathbf{j}_\perp = j_{\|m}/H$. We use the subscript m to denote magnetospheric variables and the subscript i for ionospheric ones. Because of flux tube geometry we have (HT)

$$j_{\|i} = \frac{B_i}{B_m} j_{\|m}. \tag{6.6}$$

Again using the ideal Ohm's law it is easy to show (HT) that

$$\omega = \frac{1}{B_m}\nabla_m^2\Phi_m. \tag{6.7}$$

We assume a linear relationship between the magnetospheric potential Φ_m and the ionospheric potential Φ_i:

$$j_{\|i} = K(\Phi_i - \Phi_m). \tag{6.8}$$

This relation (which will be discussed later in more detail) is based on observations. The parameter K is called the Lyons-Evans-Lundin constant, or the LEL-constant. A basic numerical value for it is $K \sim 10^{-9}$ mho m^{-2} (mho $\equiv$ 1/ohm).

Putting everything together, we obtain the ionosphere-magnetosphere coupling equations:

$$\frac{\partial\omega}{\partial t} = \frac{1}{B_m}\hat{z}\cdot\nabla_m\omega\times\nabla_m\Phi_m + \frac{B_m^2 K}{\rho H B_i}(\Phi_m - \Phi_i) \tag{6.9}$$

$$\nabla_m^2 \Phi_m = B_m \omega \tag{6.10}$$

$$-\Sigma_P \nabla_i^2 \Phi_i = K(\Phi_m - \Phi_i) \tag{6.11}$$

where we have assumed homogeneous ionospheric conductivities for simplicity. (This is by no means a necessary assumption, neither is it usually valid!)

6.3 The resistive scale length λ_i

Equation (6.11) has one very interesting feature. Written another way, it reads

$$-(\nabla_i^2 - \frac{K}{\Sigma_P})\Phi_i = \frac{K}{\Sigma_P}\Phi_m \tag{6.12}$$

which is an inhomogeneous Helmholtz' equation with source term proportional to Φ_m. If we introduce the parameter $k_0 = \sqrt{K/\Sigma_P}$, which has the dimension of inverse length, we obtain

$$-(\nabla_i^2 - k_0^2)\Phi_i = s \equiv \frac{K}{\Sigma_P}\Phi_m. \tag{6.13}$$

This equation has constant coefficients, so we can use Fourier analysis with success. In Fourier space, $\nabla_i^2 \to -k^2$; thus

$$\Phi_{ik} = \frac{s_k}{k^2 + k_0^2} \tag{6.14}$$

When k is larger than k_0, this equation dictates $\Phi_{ik} \sim 1/k^2$. When k is smaller than k_0, however, we get $\Phi_{ik} \sim$ const.. Thus the system acts as a low-pass filter: small wavelengths (large k) are damped. In other words, small-scale variations in the source term (the magnetospheric potential Φ_m) do not cause small-scale variations in the ionospheric potential Φ_i but are effectively filtered away. The limiting wavelength is equal to $\lambda_i = 1/k_0$,

$$\lambda_i = \sqrt{\frac{\Sigma_P}{K}}. \tag{6.15}$$

The typical value for this so-called resistive scale length λ_i is 100 km. Roughly speaking, in scale sizes larger than about 100 km, the ionospheric and magnetospheric potentials are equal and one may ignore the field-aligned conductance (i.e., one can assume that $K = \infty$). On the contrary, on smaller-scale sizes the ionospheric potential is a smoothed-out version of the magnetospheric one.

The eastward-drifting Ω bands are probably features which are associated with the resistive scale length λ_i.

The resistive scale length is, of course, variable, but due to the square root dependence it does not depend too much on the Pedersen conductivity, for example. Furthermore, the Pedersen conductivity is the one that varies less; the Hall conductivity varies more. One can understand the formula for λ_i roughly as follows (Figure 6.1): If two cable-like field-aligned currents arrive to and leave from the ionosphere, the horizontal current in between them will not move as a straight line, but divide itself over a larger area. The area width and length will be approximately equal. When the separation between the cable-like FACs is equal to λ_i, the resistance of the field line (which is equal to $1/(KA)$ where A is the cross-sectional area of the cable-like current) is equal to the resistance of the horizontal ionosphere. Thus, if the ionosphere is made more conducting (Σ_P is increased), the resistive scale length also increases because the current can afford to travel a longer distance in a better-conducting ionosphere. If the field-line conductance is increased, the total resistance is smaller and λ_i is decreased.

Equations (6.9) through (6.11) can be perhaps more easily analyzed if we eliminate the all variables except one. Using the Fourier-transformed quantities, this is easy. Assuming isotropic and constant magnetosphere-ionosphere mapping, we have $\nabla_i = M\nabla_m$, where

$$M \equiv \sqrt{\frac{B_i}{B_e}}. \tag{6.16}$$

The constancy of the mapping holds if the region under study is sufficiently small, less than 1000 km in the ionosphere, say. To Fourier-analyze, we take $\nabla_m \to i\mathbf{k}$. We obtain

$$\begin{aligned}
\Phi_{ik} &= \frac{K\Phi_{mk}}{K + k^2 M^2 \Sigma_P} = \frac{\Phi_{mk}}{1 + k^2\lambda_m^2} \\
\lambda_i &\equiv \sqrt{\frac{\Sigma_P}{K}} \\
\lambda_m &\equiv M\lambda_i \\
\Phi_{mk} - \Phi_{ik} &= \frac{k^2\lambda_m^2\Phi_{mk}}{1 + k^2\lambda_m^2} \\
\omega_k &= \frac{-1}{B_m}k^2\Phi_{mk}
\end{aligned}$$

By eliminating Φ_i and ω and assigning to Equation (6.9) we obtain

$$-k^2\frac{\partial\Phi_{mk}}{\partial t} = \frac{1}{B_m}\left[\hat{z}\cdot(\nabla\nabla^2\Phi_m)\times\nabla\Phi_m\right]_k + \frac{B_m^2 K}{\rho H}\frac{1}{M^2}\frac{k^2\lambda_m^2}{1 + k^2\lambda_m^2}\Phi_{mk}. \tag{6.17}$$

Or, we can write it as a pair of equations

$$\begin{aligned}
\frac{\partial\omega_k}{\partial t} &= \frac{1}{B_m}\left[\hat{z}\cdot\nabla\omega\times\nabla\Phi_m\right]_k - A\frac{\lambda_m^2}{1 + k^2\lambda_m^2}\omega_k \\
\omega &= \frac{1}{B_m}\nabla_m^2\Phi_m
\end{aligned} \tag{6.18}$$

where the constant A has dimensionality $\mathrm{m}^{-2}\,\mathrm{s}^{-1}$:

$$A \equiv \frac{B_m^2 K}{\rho H}\frac{1}{M^2} = \frac{B_m^2 K}{\rho H}\frac{B_m}{B_i} \tag{6.19}$$

These equations are analogous to 2D incompressible hydrodynamics, except for the ionospheric coupling term. The coupling term vanishes ($A \to 0$) in the following cases: (1) $B_m \to 0$. In this case there is no coupling because there are no field lines. (2) $K \to 0$. Now there is no coupling because the field lines are so resistive that the current does not close through the ionosphere. (3) $\rho H \to \infty$. In this case there is so much plasma in the magnetosphere that the plasma is effectively unmagnetized. Likewise, the coupling term dominates if the field lines are very good conductors or if there is only little plasma in the magnetosphere. The parameter λ_m is the magnetospheric image of the ionospheric resistive scale length λ_i. For small enough wavelengths ($k\lambda_m \gg 1$) the coupling term is always small and the magnetospheric plasma behaves as a normal fluid.

Home Tasks for 2D Ionosphere-Magnetosphere Coupling

1. Try to estimate the location of $\beta = 1$ surface in the nightside magnetosphere. Here β is the plasma beta parameter, defined by $\beta = 2\mu_0 P/\mathbf{B}^2$. Explain how the MHD waves and electric current systems are different on the earthward and tailward sides of this zone.

2. Given a 'realistic' 3D magnetic field model of the magnetosphere (e.g. the Tsyganenko-89 model) but no satellite plasma measurements, how would you estimate the plasma pressure in the near-Earth region?

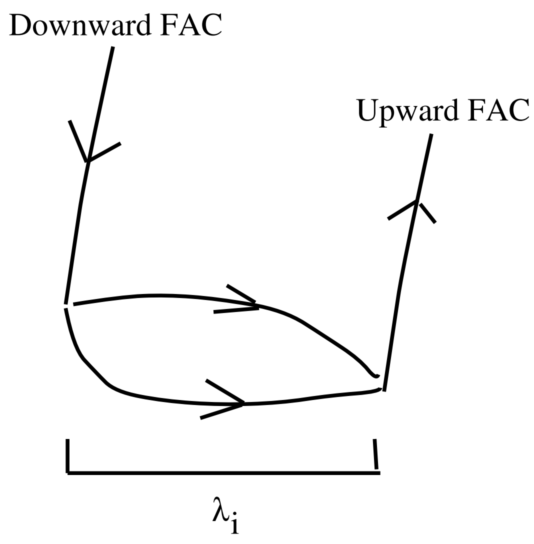

Figure 6.1: About the role of the resistive scale length λ_i.

3. The ring current is basically a toroidal current loop that flow around the Earth at about 3-4 R_E distance, centered at the equatorial plane. The current flows from east to west. Changes in the ring current magnitude are continuously monitored by equatorial ground magnetometers. Develop a relationship between the ring current intensity variations and the (maximum) plasma pressure in the near-Earth region. Ignore possible variations in shape which may be associated with the intensity variations. Also ignore the counterstreaming component of the ring current, which is found at the earthward edge of the main ring current. Use some suitable analytic representation for the ring current cross-sectional distribution.

4. To do this home task, you need the Tsyganenko-89 or Tsyganenko-87 model subroutines. Compute the pressure distribution along the nightside GSM x axis (i.e., the intersection of the noon-midnight meridional plane and the equatorial plane) using the method of HT 2. What is the maximum plasma pressure? How does it depend on the inputs of the model (the K_p index)? Review HT 1, i.e. estimate the plasma beta parameter, using this new estimate for the plasma pressure. Compute the plasma beta along the nightside GSM x axis, assuming the T89 or T87 model is completely valid.

5. Derive the 2D ionosphere-magnetosphere coupling equations [Eqs.(6.9) through (6.11)] in detail. What assumptions are needed?

6. Show that 2D incompressible fluid is described by equations similar to (6.18) with $A = 0$. (*Hint*: use the streamfunction-vorticity formulation.)

Chapter 7

3D Ionosphere-magnetosphere coupling

In this chapter we present three-dimensional ionosphere-magnetosphere coupling equations. This approach has the potential for a fully realistic global modeling (taking into account limitations of temporal and spatial scales, of course). The drawback is that the equations are not easy to analyze analytically.

7.1 Origin of field-aligned current

In this section we derive a formula for ionospheric field-aligned current.

The MHD momentum equation reads.

$$\rho \frac{d\mathbf{v}}{dt} = \mathbf{j} \times \mathbf{B} - \nabla P. \tag{7.1}$$

Let us solve this equation for the perpendicular current density $\mathbf{j}_\perp$:

$$\mathbf{j}_\perp = \frac{\mathbf{B}}{B^2} \times \left(\nabla P + \rho \frac{d\mathbf{v}}{dt} \right). \tag{7.2}$$

Expanding $\mathbf{j} = \mathbf{j}_\perp + \mathbf{j}_\parallel$, we find

$$0 = \nabla \cdot \mathbf{j} \Rightarrow \nabla \cdot \mathbf{j}_\parallel = -\nabla \cdot \mathbf{j}_\perp. \tag{7.3}$$

On the other hand,

$$\begin{aligned} \nabla \cdot \mathbf{j}_\parallel &= \nabla \cdot \left(\mathbf{B} \frac{j_\parallel}{B} \right) \\ &= \mathbf{B} \cdot \nabla \left(\frac{j_\parallel}{B} \right) \end{aligned}$$

because $\nabla \cdot \mathbf{B} = 0$. Let us define the field-aligned coordinate s by

$$\mathbf{B} \cdot \nabla = B \frac{d}{ds} \tag{7.4}$$

Thus we obtain

$$B \frac{d}{ds} \left(\frac{j_\parallel}{B} \right) = \mathbf{B} \cdot \nabla \left(\frac{j_\parallel}{B} \right) = \nabla \cdot \mathbf{j}_\parallel = -\nabla \cdot \mathbf{j}_\perp. \tag{7.5}$$

Integrating over s we obtain

$$\Delta\left(\frac{j_{\parallel}}{B}\right) = -\int_1^2 ds\frac{\nabla\cdot\mathbf{j}_{\perp}}{B} \tag{7.6}$$

where Δ denotes the change when moving from point 1 to 2. Taking point 1 to be on the equatorial plane where by symmetry $j_{\parallel} = 0$, we obtain

$$j_{\parallel}^{(2)} = -B^{(2)}\int_1^2 ds\frac{\nabla\cdot\mathbf{j}_{\perp}}{B}. \tag{7.7}$$

If point 2 is at the ionosphere, this can be written (denoting ionospheric quantities by i and equatorial plane quantities by e) as

$$j_{\parallel}^{i} = -B^{i}\int_e^i ds\frac{\nabla\cdot\mathbf{j}_{\perp}}{B}. \tag{7.8}$$

Substituting from Equation (7.2) we have a general formula for ionospheric FAC in terms of MHD variables:

$$j_{\parallel}^{i} = -B^{i}\int_e^i \frac{ds}{B}\nabla\cdot\left[\frac{\mathbf{B}}{B^2}\times\left(\nabla P+\rho\frac{d\mathbf{v}}{dt}\right)\right] \tag{7.9}$$

While this formula can be further manipulated to even simpler forms, the above form clearly shows that each term in the MHD momentum equation may contribute to the ionospheric FAC. The first term is associated with pressure gradients; it is believed that the Region-2 currents are mainly due to this term. The Region-1 current system, on the other hand, is believed to be associated with the second term, which is associated with plasma inertia. Notice that these *inertial FACs* may exist also in stationary state, because $d\mathbf{v}/dt = \mathbf{v}\cdot\nabla\mathbf{v}$ then.

Denoting $\nabla P + \rho d\mathbf{v}/dt \equiv \mathbf{a}$ we can manipulate further (HT; notice that $\mathbf{j}\cdot\mathbf{a} = 0$):

$$\nabla\cdot\left(\frac{\mathbf{B}\times\mathbf{a}}{B^2}\right) = \mathbf{B}\cdot\mathbf{a}\times\nabla\left(\frac{1}{B^2}\right) - \frac{\mathbf{B}}{B^2}\cdot\nabla\times\left(\rho\frac{d\mathbf{v}}{dt}\right) \tag{7.10}$$

Thus our formula (7.9) for the ionospheric FAC can also be written in the form

$$j_{\parallel}^{i} = -B^{i}\int_e^i ds\left[\hat{\mathbf{b}}\cdot\mathbf{a}\times\nabla\left(\frac{1}{B^2}\right) - \frac{\mathbf{B}}{B^3}\cdot\nabla\times\left(\rho\frac{d\mathbf{v}}{dt}\right)\right] \tag{7.11}$$

7.2 Integrated form of the ionospheric FAC formula

In the case of isotropic pressure and in the absence of other terms than $\mathbf{j}\times\mathbf{B}$ and ∇P in the MHD momentum equation, the ionospheric FAC formula can be written in a yet simpler form (Vasyliunas). We again start from the momentum balance equation,

$$\mathbf{j}\times\mathbf{B} = \nabla P \tag{7.12}$$

which implies

$$\mathbf{j}_{\perp} = \frac{\mathbf{B}\times\nabla P}{B^2}. \tag{7.13}$$

Consider a flux tube of small cross-sectional area (Figure 7.1)). (We will make the area infinitesimal at the end of the calculation.) Assume that the flux tube extends from the equatorial plane to the ionosphere. Because $\nabla\cdot\mathbf{j} = 0$, by Gauss' law the flux of $\mathbf{j}$ over the surface of the flux tube is zero: This implies

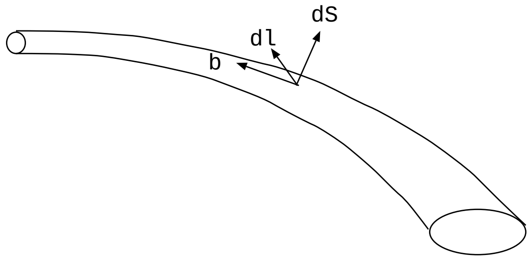

Figure 7.1: Flux tube for the derivation of the integral form of the ionospheric FAC formula

$$I_{\|} = -\int_{\text{sides}} d\mathbf{S}\cdot\mathbf{j}_{\perp} \tag{7.14}$$

where $I_{\|}$ is the current leaving the flux tube and entering the ionosphere (which is field-aligned current, of course). The current through the equatorial plane end of the tube is zero due to symmetry (no FAC flows there), thus the corresponding term is missing from the above equation. Only perpendicular current enters the integrand because $d\mathbf{S}\cdot\mathbf{B} = 0$.

We write the vector surface element of the flux tube as

$$d\mathbf{S} = d\mathbf{l}\times\hat{\mathbf{b}}ds. \tag{7.15}$$

where l is a coordinate that rotates in the positive direction around the field line along the surface of the flux tube. The vectors $(d\mathbf{l}, \hat{\mathbf{b}}, d\mathbf{S})$ form a right-handed system. We obtain

$$\begin{aligned} I_{\|} &= -\int d\mathbf{S}\cdot\mathbf{j}_{\perp} \\ &= -\int\int \frac{1}{B} d\mathbf{l}\times\hat{\mathbf{b}}ds \cdot \hat{\mathbf{b}}\times\nabla P \\ &= -\int\int d\mathbf{l}\cdot\hat{\mathbf{b}}\times(\hat{\mathbf{b}}\times\nabla P)\frac{ds}{B} \\ &= \int\int d\mathbf{l}\cdot\nabla P\frac{ds}{B} \end{aligned}$$

where we used the fact that $\hat{\mathbf{b}}\cdot\nabla P = 0$. Now, the quantity $d\mathbf{l}\cdot\nabla P$ is actually equal to $dl(\partial P/\partial l)$ (the directional derivative). Since P does not depend on s, $\partial P/\partial l$ can be taken outside of the s integral, and we get

$$I_{\|} = \int dl\frac{\partial P}{\partial l}\int\frac{ds}{B}. \tag{7.16}$$

Let us denote $\int ds/B \equiv V$, the "flux tube volume". V is a function of the equatorial plane coordinates only. Notice that when V is evaluated at different positions in the equatorial plane, both the magnetic field and the field line over which the line integral is taken change. Anyway, we obtain

$$\begin{aligned} I_{\|} &= \int d\mathbf{l}\cdot V\nabla P \\ &= \int d\mathbf{S}\cdot\nabla\times(V\nabla P) \\ &= \int d\mathbf{S}\cdot\nabla V\times\nabla P \end{aligned}$$

where we used Stokes' theorem. Here the surface integral is over the cross-sectional area of the flux tube in the equatorial plane. This is almost our final result. Now we make the flux tube area infinitesimal, whence we get the ionospheric FAC density

$$j_{\|}^{\text{iono}} = \frac{B_i}{B_e}\hat{\mathbf{z}} \cdot \nabla V \times \nabla P. \qquad (7.17)$$

where B_i is the ionospheric magnetic field, B_e is the equatorial plane magnetic field, and $\hat{\mathbf{z}}$ is a unit vector in the direction of $\mathbf{B}$ in the equatorial plane. The factor B_i/B_e comes from the fact that the current density is inversely proportional to the flux tube cross sectional area, which is in turn inversely proportional to B.

Equation (7.17) does not include the inertial currents. Thus it is useful for describing the Region-2 current system only.

7.3 Global ionosphere-magnetosphere coupling simulations

Global MHD is a powerful simulation technique for studying the large-scale behavior of the magnetosphere. However, to be really useful it must be properly coupled to the ionosphere. The solar wind-magnetosphere interaction is easier to model at least in principle because the same set of MHD equations can be used for both systems.

The geometry of global MHD models is typically a large rectangular box extending from $-200 R_E to + 30 R_E$ in x and from -50 to $+50 R_E$ in y and z. Here x, y, z are the GSM coordinates: x points to the Sun; the projection of z and the dipole axis is required to coincide in the plane perpendicular to x (it points to the north); and y completes the right-hand-side system (it points from dawn to dusk, from east to west). The Earth's surroundings is a spherical hole in the simulation box, with radius of typically 3-5 R_E. The reason why the MHD equations are not used all the way down to the ionosphere is that as the background dipole field grows, the Alfvén speeds becomes very large and the Courant (Courant-Friedrichs-Lewy) condition which dictates that the timestep must be less than the fastest wave travel time across a grid cell forces us to use a very short timestep. Thus the near-Earth region $(1 - 3R_E)$ is excluded from the MHD simulations. It is modelled as a more or less passive medium which mediates in the IM-interaction.

7.4 Implementation of IM-coupling in simulations

The implementation of ionosphere-magnetosphere coupling in simulations is a bit problematic. If we take the electrostatic approximation, restricting ourselves to time scales that are longer ("much longer") than the Alfvén travel time from the inner boundary of the MHD equations (3-5 R_E) to the ionosphere. This time is not very long (HT), thus the electrostatic approximation is generally usable.

The magnetosphere and the ionosphere form a self-consistent system. The field-aligned current that leaves/enters the magnetosphere must be equal to the field-aligned current that enters/leaves the ionosphere. The ionospheric and magnetospheric electric fields (which are potential fields in the electrostatic approximation) must map to each other in a controlled way, possibly taking into account the ionosphere-magnetosphere potential drop.

From the ionospheric point of view, we could compute the coupling in (at least) two ways: (1) Assume that the magnetosphere gives us the electric potential Φ_m. From this we can compute Φ_i, by assuming $j_{\|} = K(\Phi_m - \Phi_i)$ and using $-\nabla \cdot (\Sigma \cdot \Phi_i) = j_{\|}$. Then we can easily compute $j_{\|}$ and give it to the magnetosphere. (2) Assume that the magnetosphere gives us the field-aligned current $j_{\|}$. Then it is possible to solve Φ_i from the ionospheric current continuity equation alone. Then, Φ_m can be computed from $\Phi_m = \Phi_i + j_{\|}/K$. Thus we give Φ_m to the magnetosphere.

It turns out that while approach (1) can be succesfully used in simple 2D IM-coupling studies, where the magnetosphere is modelled with incompressible fluid, it seems not be possible to generalize it to the full MHD equations. This is mainly because it is very difficult or even impossible to formulate the MHD simulation in such a way that the field-aligned current can be given as a boundary condition. It is also very difficult to compute the electric potential Φ_m at a boundary surface of an MHD box, because the electric field in MHD is by definition not electrostatic.

Therefore, approach (2) is used in global MHD-ionosphere simulations. The field-aligned current can be computed easily from the MHD variables at the boundary by using its definition $j_{\|} = \hat{\mathbf{B}}\cdot\nabla\times\mathbf{B}/\mu_0$. Likewise, when the ionosphere gives the electric potential Φ_m to the magnetosphere, it is possible to compute the perpendicular velocity $\mathbf{v}_\perp = -\nabla\Phi_m\times\mathbf{B}/B^2$ (assuming ideal MHD). The velocity, being one of the primary MHD variables, can be given as a boundary condition far more easily than the field-aligned current, which is a derivative of the primary variables.

Home Tasks for 3D Ionosphere-Magnetosphere Coupling

1. Estimate the Alfvén travel time (a) from the 3.7 R_E shell to the ionosphere, (b) from the equatorial plane at $x = -7R_E$ to the ionosphere in dipole field.

2. Derive Equation (7.10).

3. Show that the generalization of Equation (7.2) for anisotropic field-aligned pressure tensor $\mathtt{P} = P_\perp \mathtt{I} + (P_\| - P_\perp)\hat{\mathbf{b}}\hat{\mathbf{b}}$ is

$$\mathbf{j}_\perp = \frac{\mathbf{B}}{B^2}\times\left(\nabla P_\perp + \rho\frac{d\mathbf{v}}{dt}\right) + \frac{P_\| - P_\perp}{B}\left(\nabla\times\hat{\mathbf{b}}\right)_\perp \tag{7.18}$$

Write down the corresponding ionospheric field-aligned current.

4. Generalize Equation (7.9) to the case of nonzero viscosity in the MHD momentum equation.

5. (Difficult) Try to derive the integrated form of the ionospheric FAC formula (Equation 7.17) by starting from the non-integrated version (Equation 7.9 or 7.11).

6. Calculate the "flux tube volume" $\int ds/B$ in dipole magnetic field.

7. (a) Write the ionospheric elliptic equation (Equation 5.13) in spherical coordinates. (b) Study the character of the equation (elliptic/hyperbolic/parabolic). Does it change character from point to point? (c) Discretize the equation by using second-order accurate formulas (centered differences, see e.g. Abramowitz and Stegun). Write it in the form

$$\Phi_{i,j} + a_{+0}\Phi_{i+1,j} + a_{-0}\Phi_{i-1,j} + a_{0-}\Phi_{i,j-1} + a_{0+}\Phi_{i,j+1} = s_{i,j} \tag{7.19}$$

Write down explicit expressions for the a coefficients and the source term s.

8. The ratio of equatorial plane versus ionospheric cross sectional area of flux tubes grows as we move farther down the magnetotail (Why?). Thus, there is more plasma that can potentially precipitate to a given ionospheric area in the far tail than there is in the near-Earth tail. Calculate this effect quantitatively, i.e., show how the ionospheric precipitation depends on the flux tube cross section ratio, if plasma density and temperature are held constant.

9. (Rather laborious but straightforward.) Assume that the electric potential on the ionospheric plane is know, that is, the function $\Phi(\theta,\varphi)$ is known. Assuming a dipolar magnetic field (with dipole axis oriented parallel to the z axis for simplicity), derive expressions for the electric field (E_x, E_y, E_z) at an arbitrary position (x, y, z) in the magnetosphere in the stationary state.

Chapter 8

MHD from Boltzmann's equation

The MHD equations were "derived" earlier by adding terms to Euler equations using simple physical arguments. This approach is successful, but it leaves in doubt to what extent the associated Ohm's law, $\mathbf{E}+\mathbf{v}\times\mathbf{B} = 0$ is valid. Indeed, the generalization of this equation turns out to be such that it would be very difficult indeed to derive it using intuitive arguments. For this reason, a systematic approach for deriving the MHD equations is needed.

We shall proceed in two phases. First we derive the so-called two-fluid equations using Boltzmann's equation as starting point. Then we manipulate these electron and ion equations to yield the MHD equations.

Boltzmann's equation for species s is

$$\partial_t f_s + \mathbf{v} \cdot \partial_{\mathbf{x}} f_s + \mathbf{a} \cdot \partial_{\mathbf{v}} f_s = (\partial_t f_s)_c \tag{8.1}$$

where $\mathbf{a} = (q_s/m_s)(\mathbf{E} + \mathbf{v}\times\mathbf{B})$ is the acceleration and $(\partial_t f_s)_c$ is the rate of change of the distribution function $f_s(\mathbf{x}, \mathbf{v}, t)$ due to binary collisions. The derivation of Boltzmann's equation is outside our scope here. However, we note that it is easy to returns Boltzmann's equation in mind by noting that the LHS is actually the total time derivative of $f(\mathbf{x}(t), \mathbf{v}(t), t)$, where $(\mathbf{x}(t), \mathbf{v}(t))$ defines a particle's orbit.

8.1 Continuity equation

We temporarily suppress the species index s to simplify notation. Integrating Equation (8.1) over velocity space yields

$$\partial_t n + \nabla \cdot (n\mathbf{u}) = 0 \tag{8.2}$$

where

$$n \equiv \int d\mathbf{v} f \tag{8.3}$$

is the number density and

$$\mathbf{u} \equiv \frac{1}{n} \int d\mathbf{v}\mathbf{v} f \tag{8.4}$$

is the average velocity. (In this section we denote the fluid velocity by $\mathbf{u}$, not $\mathbf{v}$, to prevent confusion with the velocity space variable $\mathbf{v}$.) To obtain Equation (8.2) we need to prove that

$$\int d\mathbf{v}\mathbf{a} \cdot \partial_{\mathbf{v}} f = 0, \tag{8.5}$$

$$\int d\mathbf{v}(\partial_t f)_c = 0. \tag{8.6}$$

Equation (8.5) follows easily if we note that

$$\begin{aligned}\int d\mathbf{v}\mathbf{v}\times\mathbf{B}\cdot\partial_{\mathbf{v}} f &= -\mathbf{B}\cdot\int d\mathbf{v}\mathbf{v}\times\partial_{\mathbf{v}} f \\ &= -\mathbf{B}\cdot\int d\mathbf{v}\left[-\partial_{\mathbf{v}}\times(\mathbf{v}f) + f\partial_{\mathbf{v}}\times\mathbf{v}\right] \\ &= 0+0\end{aligned}$$

Equation (8.6) is a natural requirement that any collision model must obey, because elastic collision can change only the particle velocities, not the particle density.

If there is plasma production or recombination, as there is in cometary environments and in the ionosphere, then Equation (8.2) is, of course, not valid. For example in the ionosphere it takes the form

$$\partial_t n + \nabla\cdot(n\mathbf{u}) = q - \alpha n^2 \tag{8.7}$$

where q is the ionization production rate and α is the recombination coefficient.

To summarize, we have now derived the continuity equation (8.2) for single particle species from Boltzmann's equation (8.1) by integrating it over velocity space (taking the zeroth order moment equation).

We can multiply Boltzmann's equation by successive powers of $\mathbf{v}$ and integrate over $\mathbf{v}$. In this way, an infinite chain of moment equations can be produced. Each new equation contains new unknown moments which are expressed in terms of still higher-order moments. Thus the chain of equations never terminates, unless we terminate it using some approximations (the closure scheme). When deriving these equations we must assume that $f(\mathbf{v})$ goes to zero sufficiently fast as $v\to\infty$. If all the equations in the infinite chain are considered, then $f(\mathbf{v})$ must vanish exponentially. In the following we use only the first few moment equations, but even in this case the condition for $f(\mathbf{v})\to 0$ is rather stringent: for example $f(v)\sim 1/v^4$ is *not sufficient*. Sometimes observations suggest that the velocity distribution in space plasmas is polynomial rather than exponential. If this is really the case, we have a problem with the moment equations.

8.2 Momentum equation

We now multiply Boltzmann's equation (8.1) by $\mathbf{v}$ and again integrate over velocities. We obtain

$$\partial_t(n\mathbf{u}) + \nabla\cdot\int d\mathbf{v}\mathbf{v}\mathbf{v}f + \int d\mathbf{v}\mathbf{v}\mathbf{a}\cdot\partial_{\mathbf{v}} f = \int d\mathbf{v}\mathbf{v}(\partial_t f)_c. \tag{8.8}$$

At this stage let us define the pressure tensor

$$\mathtt{P} \equiv m\int d\mathbf{v}(\mathbf{v}-\mathbf{u})(\mathbf{v}-\mathbf{u})f. \tag{8.9}$$

The quantity $\int d\mathbf{v}\mathbf{v}\mathbf{v}f$ can be expanded as follows,

$$\int d\mathbf{v}\mathbf{v}\mathbf{v}f = \frac{1}{m}\mathtt{P} + n\mathbf{u}\mathbf{u} \tag{8.10}$$

because the cross terms vanish due to the definition of $\mathbf{u}$. The acceleration term of Equation (8.8) can be simplified as

$$
\begin{aligned}
\int d\mathbf{v}\mathbf{v}\mathbf{a}\cdot\partial_{\mathbf{v}} f &= \frac{q}{m}\left(E_i\int d\mathbf{v}\mathbf{v}\partial_{v_i} f + \int d\mathbf{v}\mathbf{v}\epsilon_{ijk}v_j B_k \partial_{v_i} f\right)\\
&= \frac{q}{m}\left(-E_i\int d\mathbf{v}\hat{e}_i f + \epsilon_{ijk}B_k\int d\mathbf{v}\mathbf{v}v_j\partial_{v_i} f\right)\\
&= \frac{q}{m}\left(-\mathbf{E}n - \epsilon_{ijk}B_k\int d\mathbf{v} f(\delta_{ij}\mathbf{v} + v_j\hat{e}_i)\right)\\
&= -\frac{q}{m}\mathbf{E}n - \frac{q}{m}\hat{e}_i\epsilon_{ijk}B_k n u_j\\
&= -\frac{q}{m}(\mathbf{E}+\mathbf{u}\times\mathbf{B})\,n.
\end{aligned}
$$

Thus we obtain from Equation (8.8)

$$
\partial_t(n\mathbf{u}) + \frac{1}{m}\nabla\cdot\mathsf{P} + \nabla\cdot(n\mathbf{u}\mathbf{u}) = \frac{q}{m}n(\mathbf{E}+\mathbf{u}\times\mathbf{B}) + \int d\mathbf{v}\mathbf{v}(\partial_t f)_c. \tag{8.11}
$$

This equation is very similar to the conservation form of Euler's momentum equation. It is customary, however, to use the continuity equation (8.2) to obtain the equation

$$
(\partial_t + \mathbf{u}\cdot\nabla)\mathbf{u} = \frac{q}{m}(\mathbf{E}+\mathbf{u}\times\mathbf{B}) - \frac{1}{nm}\nabla\cdot\mathsf{P} + \frac{1}{n}\int d\mathbf{v}\mathbf{v}(\partial_t f)_c. \tag{8.12}
$$

Notice, however, that this equation is valid only in the case where we do not have any ionization production or recombination. If we have, the production and loss terms would also enter into play through the use of the continuity equation!

8.3 One-fluid momentum equation

Writing Equation (8.12) for ions and electrons yields

$$
\begin{aligned}
(\partial_t + \mathbf{u}_i\cdot\nabla)\mathbf{u}_i &= \frac{e}{m_i}(\mathbf{E}+\mathbf{u}_i\times\mathbf{B}) - \frac{1}{nm_i}\nabla\cdot\mathsf{P}_i - \nu_{ie}(\mathbf{u}_i-\mathbf{u}_e) &(8.13)\\
(\partial_t + \mathbf{u}_e\cdot\nabla)\mathbf{u}_e &= -\frac{e}{m_e}(\mathbf{E}+\mathbf{u}_e\times\mathbf{B}) - \frac{1}{nm_e}\nabla\cdot\mathsf{P}_e - \nu_{ei}(\mathbf{u}_e-\mathbf{u}_i). &(8.14)
\end{aligned}
$$

Here we have made the approximation $n_e \approx n_i \equiv n$, that is, we have assumed that the length scales of interest are much larger than the Debye length. This is usually well satisfied in space plasmas (in other words, other approximations break down easier than this one), but in case of any doubt one can easily check its validity, of course.

We have employed the Krook model for the collision terms. The electron-ion collision frequency ν_{ei} is hugely different from the ion-electron collision frequency ν_{ie}; they obey the relation $\nu_{ei} = (m_i/m_e)\nu_{ie}$.

Let us define the center-of-mass velocity $\mathbf{u}$ by

$$
\mathbf{u} \equiv \frac{m_e\mathbf{u}_e + m_i\mathbf{u}_i}{m_i + m_e} \tag{8.15}
$$

The current density is

$$
\mathbf{j} = en(\mathbf{u}_i - \mathbf{u}_e). \tag{8.16}
$$

Solving equations (8.15) and (8.16) for $\mathbf{u}_e, \mathbf{u}_i$ yields

$$
\mathbf{u}_e = \mathbf{u} - \frac{\mathbf{j}}{en} \tag{8.17}
$$

$$
\mathbf{u}_i = \mathbf{u} + \frac{m_e}{m_i}\frac{\mathbf{j}}{en} \tag{8.18}
$$

where we have used $m_e/m_i \ll 1$. In fact, we could assume $\mathbf{u}_i \approx \mathbf{u}$ from now on without changing much of our final results, but as this is not quite clear a priori, we will not do it.

Multiplying Equation (8.13) by m_i and multiplying Equation (8.14) by m_e and summing gives

$$(m_i + m_e)\partial_t \mathbf{u} + m_e \mathbf{u}_e \cdot \nabla \mathbf{u}_e + m_i \mathbf{u}_i \cdot \nabla \mathbf{u}_i = (e - e)\mathbf{E} + \frac{1}{n}\mathbf{j}\times\mathbf{B} - \frac{1}{n}\nabla \cdot (\mathsf{P}_e + \mathsf{P}_i) \tag{8.19}$$

It is easy to show that (HT)

$$m_e \mathbf{u}_e \cdot \nabla \mathbf{u}_e + m_i \mathbf{u}_i \cdot \nabla \mathbf{u}_i = (m_i + m_e)\mathbf{u} \cdot \nabla \mathbf{u} + \frac{m_e}{e^2 n}\mathbf{j} \cdot \nabla \left(\frac{\mathbf{j}}{n}\right). \tag{8.20}$$

Thus we obtain the one-fluid momentum equation

$$(m_i + m_e)(\partial_t + \mathbf{u} \cdot \nabla)\mathbf{u} + \frac{m_e}{e^2 n}\mathbf{j} \cdot \nabla \left(\frac{\mathbf{j}}{n}\right) = \frac{\mathbf{j}\times\mathbf{B} - \nabla \cdot (\mathsf{P}_e + \mathsf{P}_i)}{n} \tag{8.21}$$

Notice that the collision terms cancel from Equation (8.21).

8.4 Generalized Ohm's law

We have now used only half the information contained in the electron and ion momentum equations, because we have just summed these equations to obtain the one-fluid momentum equation (8.21). We now proceed by effectively forming the difference of these equations. Multiply Equation (8.13) by en and multiplying Equation (8.14) by $-en$ and summing gives

$$\begin{aligned} &en\partial_t(\mathbf{u}_i - \mathbf{u}_e) + en(\mathbf{u}_i \cdot \nabla \mathbf{u}_i - \mathbf{u}_e \cdot \nabla \mathbf{u}_e) = \\ &ne^2\left(\frac{1}{m_e} + \frac{1}{m_i}\right)\mathbf{E} + ne^2\left(\frac{\mathbf{u}_i}{m_i} + \frac{\mathbf{u}e}{m_e}\right)\times\mathbf{B} + \frac{e}{m_e}\nabla \cdot \mathsf{P}_e - \frac{e}{m_i}\nabla \cdot \mathsf{P}_i \\ &-\nu_{ei}\mathbf{j} \end{aligned} \tag{8.22}$$

We can drop two terms from this by utilizing $1/m_i \ll 1/m_e$ and $|(e/m_i)\nabla \cdot \mathsf{P}_i| \ll |(e/m_e)\nabla \cdot \mathsf{P}_e|$.

Using equations (8.17) and (8.18) and doing some algebra (HT) gives

$$\begin{aligned} &\partial_t \mathbf{j} - \mathbf{j}\frac{\partial_t n}{n} + en\left(\frac{\mathbf{j}}{en} \cdot \nabla \mathbf{u} + \mathbf{u} \cdot \nabla \left(\frac{\mathbf{j}}{en}\right)\right) - \mathbf{j} \cdot \nabla \left(\frac{\mathbf{j}}{en}\right) = \\ &\frac{ne^2}{m_e}(\mathbf{E} + \mathbf{u}\times\mathbf{B}) - \frac{e}{m_e}\mathbf{j}\times\mathbf{B} + \frac{e}{m_e}\nabla \cdot \mathsf{P}_e - \nu_{ei}\mathbf{j} \end{aligned} \tag{8.23}$$

Rearranging this a bit we get

$$\frac{m_e}{ne^2}\left\{\partial_t \mathbf{j} + \nabla \cdot (\mathbf{u}\mathbf{j} + \mathbf{j}\mathbf{u}) - \mathbf{j} \cdot \nabla \left(\frac{\mathbf{j}}{en}\right)\right\} = \mathbf{E} + \left(\mathbf{u} - \frac{\mathbf{j}}{en}\right)\times\mathbf{B} + \frac{1}{en}\nabla \cdot \mathsf{P}_e. \tag{8.24}$$

This is the so-called generalized Ohm's law.

The electron and ion pressure tensors appearing in equations (8.21) and (8.24) are defined by

$$\mathsf{P}_e \equiv m_e \int d\mathbf{v}(\mathbf{v} - \mathbf{u}_e)(\mathbf{v} - \mathbf{u}_e) f_e \tag{8.25}$$

$$\mathsf{P}_i \equiv m_i \int d\mathbf{v}(\mathbf{v} - \mathbf{u}_i)(\mathbf{v} - \mathbf{u}_i) f_i \tag{8.26}$$

The one-fluid momentum equation (8.21) and the generalized Ohm's law (8.24) can be somewhat simplified if we write them in terms of *center-of-mass pressures* $\mathtt{P}_e^{\mathrm{CM}}$ and $\mathtt{P}_i^{\mathrm{CM}}$, which are defined by

$$\mathtt{P}_e^{\mathrm{CM}} \equiv m_e \int d\mathbf{v}(\mathbf{v}-\mathbf{u})(\mathbf{v}-\mathbf{u})f_e \tag{8.27}$$

$$\mathtt{P}_i^{\mathrm{CM}} \equiv m_i \int d\mathbf{v}(\mathbf{v}-\mathbf{u})(\mathbf{v}-\mathbf{u})f_i \tag{8.28}$$

with $\mathbf{u}$ of course defined by Equation (8.15).

The connection between the divergences of the two pressure tensors is (HT)

$$\nabla\cdot\mathtt{P}_e = \nabla\cdot\mathtt{P}_e^{\mathrm{CM}} - \frac{m_e}{e^2}\mathbf{j}\cdot\nabla\left(\frac{\mathbf{j}}{n}\right) \tag{8.29}$$

$$\nabla\cdot\mathtt{P}_i = \nabla\cdot\mathtt{P}_i^{\mathrm{CM}} - \frac{m_e^2}{m_i e^2}\mathbf{j}\cdot\nabla\left(\frac{\mathbf{j}}{n}\right) \tag{8.30}$$

With these relations, the one-fluid momentum equation and the generalized Ohm's law assume the form

$$(m_i+m_e)(\partial_t+\mathbf{u}\cdot\nabla)\mathbf{u} = \frac{1}{n}\left(\mathbf{j}\times\mathbf{B} - \nabla\cdot(\mathtt{P}_e^{\mathrm{CM}}+\mathtt{P}_i^{\mathrm{CM}})\right) \tag{8.31}$$

$$\frac{m_e}{ne^2}\left\{\partial_t\mathbf{j} + \nabla\cdot(\mathbf{uj}+\mathbf{ju})\right\} = \mathbf{E} + \left(\mathbf{u}-\frac{\mathbf{j}}{en}\right)\times\mathbf{B} + \frac{1}{en}\nabla\cdot\mathtt{P}_e^{\mathrm{CM}} - \frac{m_e\nu_{ei}}{ne^2}\mathbf{j} \tag{8.32}$$

Equation (8.31) is the familiar MHD momentum equation, except that the pressure is now a tensor. Equation (8.32), on the other hand, is much more complicated than the ideal MHD Ohm'w law $\mathbf{E}+\mathbf{u}\times\mathbf{B}=0$. The LHS of Equation (8.32), however, can usually be neglected (notice that it is proportional to m_e) in space plasma physics (HT). This leaves us

$$\mathbf{E} + \left(\mathbf{u}-\frac{\mathbf{j}}{en}\right)\times\mathbf{B} + \frac{1}{en}\nabla\cdot\mathtt{P}_e^{\mathrm{CM}} - \frac{m_e\nu_{ei}}{ne^2}\mathbf{j} = 0 \tag{8.33}$$

Notice that $\mathbf{u}-\mathbf{j}/(en)$ is essentially the electron velocity $\mathbf{u}_e$, thus

$$\mathbf{E} + \mathbf{u}_e\times\mathbf{B} + \frac{1}{en}\nabla\cdot\mathtt{P}_e^{\mathrm{CM}} - \frac{m_e\nu_{ei}}{ne^2}\mathbf{j} = 0 \tag{8.34}$$

Thus we have two corrections to the ideal Ohm's law: (a) the magnetic field is frozen not to the center-of-mass plasma flow but to the electron flow, (b) there is also an additional term proportional to the divergence of the electron pressure. This equation is also simple to derive directly from the electron momentum equation (8.14) by ignoring the electron mass (at the same time, $\mathtt{P}_e^{\mathrm{CM}}\to\mathtt{P}_e$).

The term $-\mathbf{j}\times\mathbf{B}/(en)$ appearing in Equation (8.33) is called the Hall term. It must not be confused with the Hall term of ionospheric electrodynamics, which is somewhat similar but has different physical origin. Magnetohydrodynamics augmented with the Hall term is sometimes called Hall-effect magnetohydrodynamics (HMHD). The Hall term is easy to include in MHD equations. The electron pressure term is much more difficult because the electron partial pressure is not known self-consistently in the one-fluid approach. Fortunately, in Earth's magnetosphere the electron temperature is usually significantly smaller than the ion temperature and thus the electron pressure term can be often neglected even if the Hall term is retained without being too inconsistent.

Usually the difference between the center-of-mass pressure and the ordinary pressure is not significant (HT). Satellite particle detectors measure the ordinary pressures, because they deal with one particle population at a time.

The electron-ion collision frequency ν_{ei} is given by

$$\nu_{ei} = \frac{v}{l_F} \tag{8.35}$$

where l_F is the electron mean free path and v is an average electron velocity (which may be taken to be the thermal velocity, $v = \sqrt{3k_BT_e/m_e}$). The mean free path is given by

$$l_F = \frac{1}{n\sigma_{ei}} \tag{8.36}$$

where σ_{ei} is the electron-ion shielded Coulomb collision collision cross section. The cross section is given by

$$\sigma_{ei} = 8\pi\left(\frac{e^2}{4\pi\epsilon_0 m_e v^2}\right)^2 \ln\Lambda \tag{8.37}$$

where Λ is the so-called Coulomb logarithm,

$$\Lambda \equiv \frac{3}{2}\frac{1}{\sqrt{\pi n}}\left(\frac{4\pi\epsilon_0 k_B T_e}{e^2}\right)^{3/2}. \tag{8.38}$$

Putting in numbers, one finds that the electron-ion collision frequency ν_{ei} is truly negligible in magnetospheric plasma. One can introduce the plasma conductivity σ (unfortunately the cross section and conductivity have the same symbol),

$$\sigma \equiv \frac{ne^2}{m_e\nu_{ei}} \tag{8.39}$$

thus generalized Ohm's law can be written

$$\frac{m_e}{ne^2}\{\partial_t\mathbf{j} + \nabla\cdot(\mathbf{uj}+\mathbf{ju})\} = \mathbf{E} + \left(\mathbf{u} - \frac{\mathbf{j}}{en}\right)\times\mathbf{B} + \frac{1}{en}\nabla\cdot\mathsf{P}_e^{\mathrm{CM}} - \frac{\mathbf{j}}{\sigma} \tag{8.40}$$

The classical conductivity value is truly infinite in space plasmas. Thus the corresponding term can be dropped from Ohm's law. In some MHD simulations one writes Ohm's law in the form $\mathbf{E} = \mathbf{v}\times\mathbf{B} = \mathbf{j}/\sigma$, which can be justified only by assuming wave-particle collisions which are many orders of magnitude more frequent than classical collisions between plasma particles.

The formulas given for ν_{ei} and the Coulomb logarithm may well contain an inaccuracy of order unity, but this does not change the qualitative conclusion that classicaly, magnetospheric plasma is collisionless.

Home Tasks for Boltzmann-MHD

1a. Derive Equation (8.20).

1b. Derive Equation (8.23).

2. Derive Equation (8.29) and Equation (8.30). Give conditions under which one can ignore the difference between center-of-mass and ordinary pressure tensors for electrons and ions, respectively.

3. Study the magnitude of different terms of the generalized Ohm's law [Equation (8.32)] in (a) the solar wind at 1 AU, where $B = 5$ nT, $n = 7$ cm^{-3}, $v = 400$ km/s, $T = 10^5$ K, (b) in the midtail region where $B = 2$ nT and $n = 1$ cm^{-3} and current scale length is 1 R_E, (c) in the near-Earth region where $B = 50$ nT and the other parameters are the same. Use reasonable values for the missing parameters. Also try to give some more general rules like ("this term can be neglected in all cases except ...").

Chapter 9

Instabilities in fluids

Neutral and magnetized fluids can exhibit various instabilities which lead to chaotic, turbulent or explosive processes. An *instability* as we understand it here is a phenomenon in which a smooth, stationary or almost stationary state becomes unstable for small amplitude perturbations. As a result, these seed perturbations start growing. At first the growth rate can be determined from the linearized equations, because the perturbations amplitude is small. At a later stage the instability is saturated or shut off by nonlinear processes. Typically these processes are such that the original reason for instability is gradually removed, unless some external process can maintain it. These processes can usually not be analyzed quantitatively using the linearized equations.

9.1 The Rayleigh-Taylor instability

The original Rayleigh-Taylor instability shows up when a heavier fluid is placed above a lighter fluid in a gravitational field. The boundary between the fluids will be unstable for small perturbations. The easiest way to see this is to consider the potential energy of two fluid elements A and B such that A is below B and thus A consists of the lighter fluid. Let us consider only the gravitational potential energy. The initial potential energy of A and B is

$$W_0 = \rho_A V g h_A + \rho_B V g h_B \tag{9.1}$$

where $h_{A,B}$ are the vertical positions of A and B, V is the volume of the fluid element (which is the same for both A and B) and g is the gravitational acceleration. Now consider a new situation where the fluid elements are swapped. Their volume V does not change in this process, but now element B is below and A is above. The potential energy expression can be obtained from W_0 by interchanging h_A and h_B,

$$W_{\text{new}} = \rho_A V g h_B + \rho_B V g h_A . \tag{9.2}$$

If $\Delta W \equiv W_{\text{new}} - W_0 < 0$ the system is unstable with respect to the interchange of A and B. If $\Delta W > 0$ it is stable and if $\Delta W = 0$ it is metastable. We obtain

$$\Delta W = Vg(\rho_A - \rho_B)(h_B - h_A) < 0 \tag{9.3}$$

because $\rho_A < \rho_B$ and $h_B > h_A$. Thus such interchanges will occur spontaneously, and the system is always unstable.

Rayleigh-Taylor type instabilities are very common in nature. For example, the formation of *cumulus* clouds (Figure 9.1) is due to this instability. The sun warms up the ground which warms up the nearby air. The warmer air expands to maintain pressure equilibrium, which results in a situation where we have lighter fluid below a heavier (colder) one. However, the warming must be fast enough so that the resulting expansion overwhelms the background exponential decrease of the density with height. This cumulus situation, although easily understood physically, is rather complicated from the analytic viewpoint because the background density and temperature profiles must be considered.

Figure 9.1: Cumulus clouds. These are produced by the Rayleigh-Taylor instability.

The cumulus example is also a good example of various nonlinear processes that can affect the later development of the instability. The cumulus cloud forms only if the convection cell reaches a critical altitude, where its temperature drops below the dewpoint. This can happen because the moisture follows the fluid motion as a passive scalar. The critical altitude is typically about 2 km during summertime. During clear weather the altitude is higher because the temperature drops more slowly with altitude. Once the cloud has started to form, a new process sets in. The condensation of water vapour brings extra energy to the system, which strengthens the upward movement. Thus, those convection cells that reach the critical altitude will get an extra boost. For this reason, cumulus clouds that remain very small for all their lifetime are rare.

If the background temperature profile of the atmosphere has some interesting shape, the development of the cumulus clouds may take different routes. For example, if the temperature starts to rise at some altitude, the clouds cannot rise above this altitude. This happens at the tropopause at latest. On the other hand, if the temperature falls rapidly with height at some height range, for example 3-5 km, then the cumulus clouds that reach this altitude become strong and may develop into *cumulonimbus* or thunderstorm. Even the thunderstorms, however, are limited by the tropopause, where the temperature starts rising again. The shape of the temperature profile determines the probability of a convection cell to develop into a thunder cell.

These familiar wheather phenomena demonstrate clearly some important aspects of fluid instabilities. The onset of the instability requires some external energy input (in this case, the solar radiation). If only one physical process is involved, the instability does not usually grow to anything spectacular. If the convection cells never reach the critical altitude (for example because the air is very dry) they will just continuously mix up the lower layers of the troposphere so that the temperature gradient introduced by solar radiation remains limited due to mixing. In this case the nonlinearly saturated state is turbulent. We require some other physical processes (condensation heat) to reach cloud formation. Yet another external factors (nontrivial shape of the background temperature profile at higher altitudes) are required to elevate some of the convection cells to thunderstorm altitudes. We can also see that even phenomena that are rather simple to describe in physical terms may be very complicated to treat analytically; we did not go beyond the simplest case of two fluids with a discontinuous (i.e., nonphysical), boundary.

In a plasma the Rayleigh-Taylor instability can arise in even more manyfold ways than in neutral fluid dynamics. The role of the gravitational field can be taken by a background

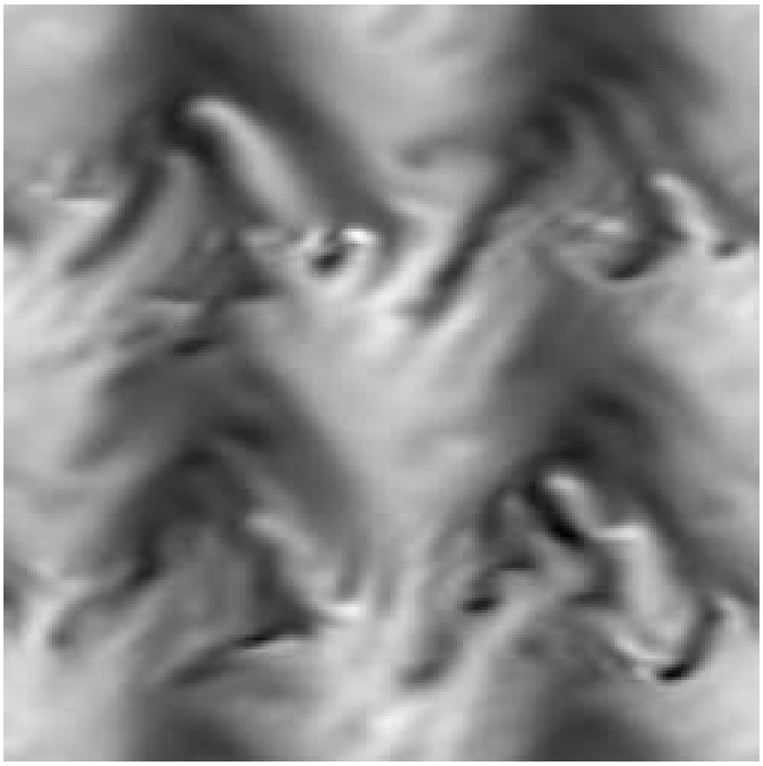

Figure 9.2: Gradient drift instability in the ionospheric E-region. The initial state contained large-scaled sinusoidal waves which contained enough electron density gradients so that secondary gradient drift waves were produced. These waves are rather close to being Farley-Buneman (both gradient drift and Farley-Buneman waves are obtained from the same set of equations, and a simple linear dispersion relation can be written which contains both instabilities). Farley-Buneman waves are discussed in connection with the two-stream instability below. – The quantity shown is the electron density and the size of the box is 100 meters across. The maximum variation in electron density is a few percent. The simulation plane is perpendicular to the magnetic field. – Notice the fingerlike structures which are typical to gradient drift instabilities.

electric field. We may also have a stabilizing background magnetic field.

In the **ionospheric E-region** a Rayleigh-Taylor type instability is at work (Figure 9.2) whenever we have a strong enough electron density gradient perpendicular to the magnetic field and a simultaneous velocity difference between the electron and ion fluids. This kind of velocity difference is possible because the ions are stuck with the neutrals whereas the electrons obey the $\mathbf{E}\times\mathbf{B}$ drift (the electron Larmor frequency is much larger than the electron-neutral collision frequency, but for the ions the situation is reversed.) This version of the instability is usually called the **gradient drift instability**. In the equatorial electrojet it is a very common phenomenon because there is a permanent strong electron density gradient (vertical) perpendicular to the background magnetic field, which is horizontal. In the auroral region the gradient drift instability can arise if there is a steep enough horizontal electron density gradient, e.g. due to auroral arcs.

In the **equatorial ionospheric F-region** the Rayleigh-Taylor instability can also arise. During daytime the electron density is high throughout the F-region. After sunset the electron density starts to decay according to the equation

$$\frac{\partial n_e}{\partial t} = q - \alpha n_e^2 \tag{9.4}$$

where q is the ionization production rate (~ 0 during nighttime) and α is the recombination coefficient, which is larger at lower altitudes. Therefore the electron density starts to erode

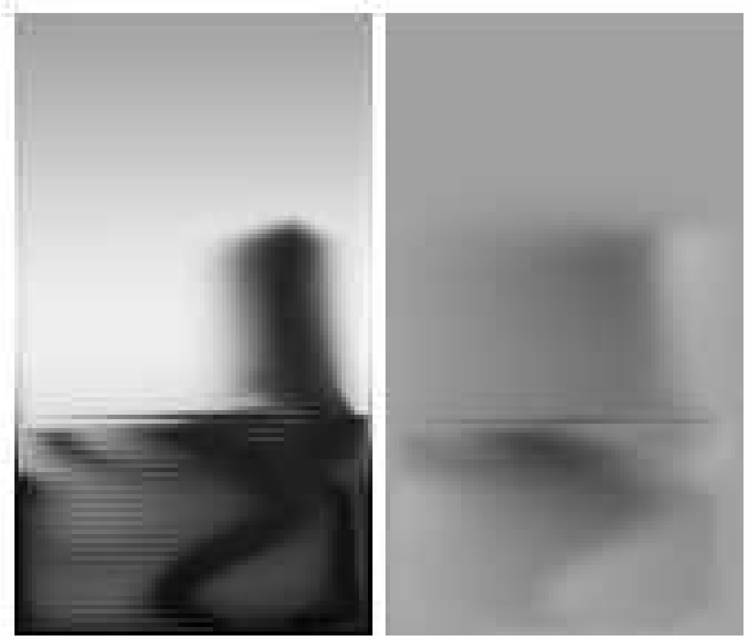

Figure 9.3: Simulation of equatorial F-region bubbles. Left panel is the electron density, right panel is the potential. The width of the box is few tens of kilometers, the height is few hundred.

from the bottomside F-region but remains high on the topside F. The resulting vertical electron density gradient becomes unstable to Rayleigh-Taylor. The resulting phenomenon is rising bubbles of low density plasma (Figure 9.3) which are known are "equatorial spread-F" because of their messy trace in F-region radars. The electron density inside the bubble is usually 1-2 *orders of magnitude* lower than in the background plasma. The boundaries of the bubbles may be unstable to smaller-scale instabilities, and the bubbles may bifurcate into two. These bubbles are essentially two-dimensional very elongated structures (the direction of the background **B** field is the homogeneous direction). It is also believed that the instability is somewhat regulated by the Pedersen conductivity of the ionospheric E-region footpoint of the field line containing the bubble.

9.2 The Kelvin-Helmholtz instability

The Kelvin-Helmholtz instability arises when we have enough *velocity shear* in the system. Velocity shear means that the velocity field has gradient perpendicular to the direction of the velocity. For example, $\mathbf{v} = (v_x(y), 0, 0)$ has velocity shear but $\mathbf{v} = (v_x(x), 0, 0)$ does not.

The Kelvin-Helmholtz instability can arise in wheather phenomena, but it is usually not a pure Kelvin-Helmholtz. A rather common situation is the breaking of the winter inversion phenomenon by wind. In the wintertime inversion the temperature is rising with altitude in some altitude range, usually below 1 km. This system is strongly Rayleigh-Taylor *stable*, because there is heavier fluid under a lighter one. However, a strong enough wind can break the inversion and mix it up. (This is also the reason why nightfrost in the autumn is prevented by wind.) The wind first blows in the upper (warmer) layer. The lower layer does not move at first, which results in a velocity shear. If the shear is strong enough it becomes unstable, resulting in turbulence and mixing.

The Kelvin-Helmholtz instability is believed to be at work at the magnetopause, because there is a strong velocity difference between the solar wind and magnetospheric plasma. However, it is as yet a bit unclear how common phenomenon this is and exactly what kind of perturbations are produced. The magnetic field tries to stabilize the system. The Kelvin-Helmholtz instability *may* also be at work in some auroral phenomena, e.g. some discrete

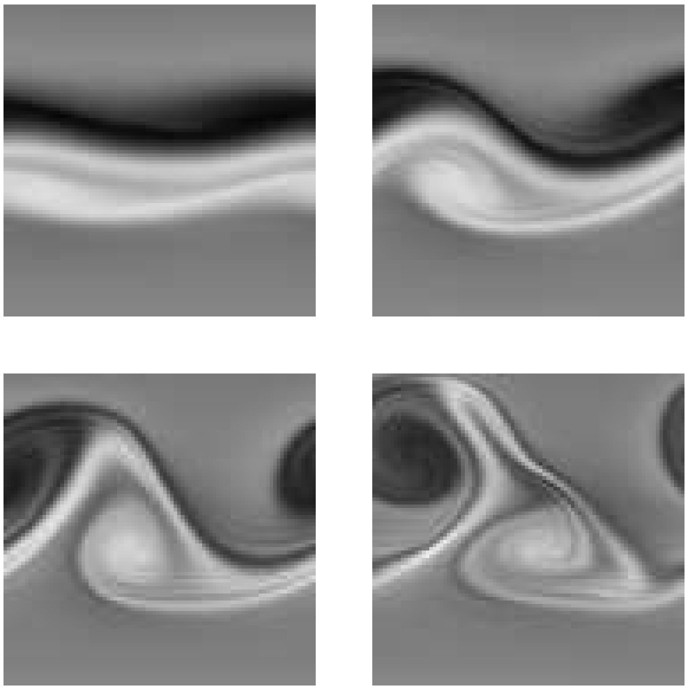

Figure 9.4: Development of a two-dimensional incompressible Kelvin-Helmholtz instability in a 2D ionosphere-magnetosphere coupling model. At the center there are two opposite directed jets, between which there is enough velocity shear to excite the instability.

arcs (Figure 9.4).

If the viscosity is nonzero, the Kelvin-Helmholtz instability arises for smaller velocity shear because the fluids tend to drag each other, resulting in irreversible heating of the fluid. When the instability has set in, the resulting small-scale turbulent pattern acts as an effective viscosity term if the large scales are considered. E.g., when making a numerical simulation one may want to model the small-scale features (which are smaller than the grid spacing) by some turbulent viscosity coefficient. In the large scale the net effect is a nonlinear viscosity: the instability produces small-scale turbulence which acts as effective viscosity which affects the instability.

9.3 The two-stream instability

The two-stream instability requires kinetic or two-fluid treatment so it is not strictly speaking an MHD instability. If there is a large enough velocity difference between the ions and the electrons, instability arises. In fully ionized plasma where $T_e = T_i$ the velocity difference must be larger than (some constant factor of order unity times) the electron thermal velocity for the instability to arise. In ionospheric E-region plasma, where plasma-neutral collisions are much more important than plasma-plasma collisions, it sufficies that the velocity difference exceeds the *ion* thermal velocity (again, up to some constant factor of order unity). In the ionospheric E-region the neutral collision modified two-stream instability is usually called the *Farley-Buneman* instability to honor its finders (Farley, Buneman, 1963) or just the **modified two-stream instability**. In the magnetosphere the two-stream instability can probably not (HT).

9.4 Ballooning and interchange instabilities

The **ballooning mode instability** is related to the Rayleigh-Taylor instability, but now the centrifugal force in a rotating plasma device, or the centrifugal force of particles bouncing

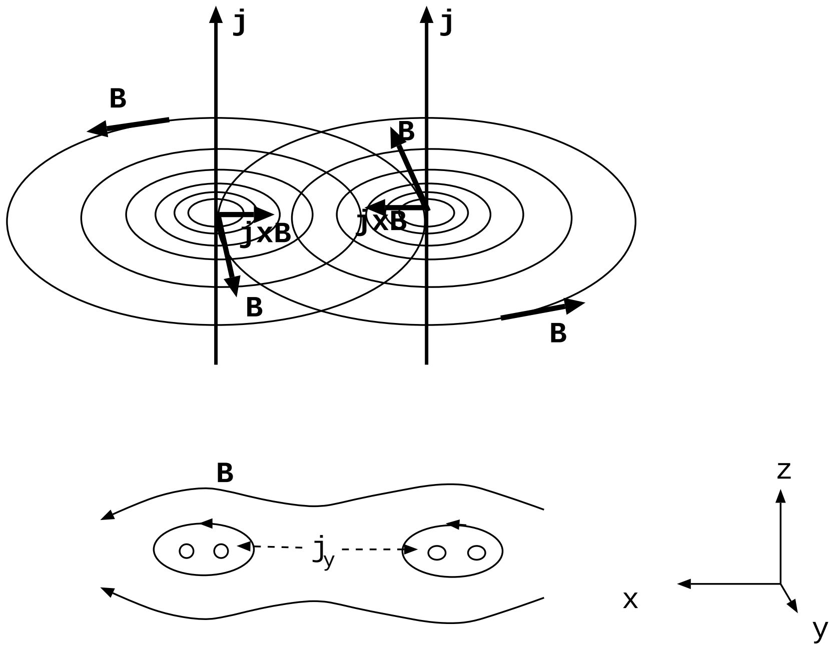

Figure 9.5: Heuristic explanation of tearing mode instability: Wires with parallel currents attract each other.

in a curved (e.g. dipolar) field, is also taken into account, in addition two pressure gradient and gravity field. The name *interchange instability* is often used to describe a subclass of ballooning instabilities which do not modify the magnetic field. A centrifugal ballooning instability is believed to be important in rapidly corotating magnetospheres such as Jovian. A pressure gradient based ballooning instability (which could also be called Rayleigh-Taylor) has been considered as one of the candidate instabilities responsible for auroral substorms.

9.5 Tearing instabilities

MHD current sheets (e.g. the geomagnetic tail) can be unstable to the tearing mode instabilities, which results in the formation of magnetic islands and associated wavy transverse structure of the current sheet. In the geomagnetic tail the cross-tail current is in the y direction. In the initial state nothing depends on y and the x depence is very weak or nonexistent, but the z depence is strong. After tearing mode instability has set in, the current has formed islands so that it is still y-directed but it has a periodic or other kind of structure in the x direction. The heuristic explanation for the tearing instability is that parallel current-carrying wires attract each other (Figure 9.5) and thus a current sheet which can be thought to consist of many wires is inherently unstable. The mathematical analysis of the tearing mode is very complicated. It is one of the candidates for *the* substorm instability.

Home Tasks for Instabilities

1. Show that it is difficult to get the two-stream instability in the magnetosphere.

2. Show that it is not difficult to get the two-stream instability in the ionospheric E-region.

Chapter 10

Magnetospheric physics

All celestial bodies are immersed in some sort of external plasma flow. Inside the solar system, all bodies face the solar wind. Solar systems are immersed in the interstellar plasma, which in general moves with respect to the central star. Here we will focus on the magnetospheres which form on planets. The solar wind is always magnetized, thus we usually cannot neglect the magnetic field even if the planet does not have an internal magnetic field.

Thus, magnetospheres are extremely common in the Universe.

10.1 Planetary magnetospheres

The Moon does not have an internal magnetic field, nor does it have an atmosphere. Thus the solar wind will collide with the surface. The solar wind plasma is absorbed by the surface. Thus there is a long taillike void region in the nightside. As a zeroth approximation the situation can be described as a fluid dynamical problem, i.e. without the magnetic field.

Mercury probably has a very weak internal field. It is believed that the internal field is just strong enough to protect the surface from a direct impact of the solar wind. However, under some circumstances it is possible that the solar wind reaches the dayside planetary surface.

Venus does not have an internal field. However, it has a dense ionosphere. The result is a moon-like plasma environment, except that the solar wind plasma is not as effectively absorbed by a gaseous ionosphere as by a rigid surface. On the contrary, the ionospheric gases may escape to the solar wind and mass-load it, because their atomic weight is much heavier than that of hydrogen. Nevertheless a taillike void region appears on the nightside (it is not completely void, of course).

Mars probably has a weak internal field. It also has an ionosphere, so the result may be somewhat similar with Venus. The role of the internal field is still unclear.

10.2 Earth's magnetosphere

The Earth, like Jupiter, has a strong internal field, which causes a large magnetosphere to form. As a zeroth approximation, Earth's magnetosphere is a long tubelike void object, with the tube radius equal to about 20 R_E. The most pronounced effect of the internal field is just to increase the size of the "head" of the magnetosphere tremendously.

The size of the "head" of the magnetosphere (i.e., the distance of the subsolar point from the center of the Earth) can be determined approximately from the MHD pressure balance equation:

$$1/2\rho v^2 = \frac{B^2}{2\mu_o} \tag{10.1}$$

Inside the magnetopause the only pressure is the magnetic pressure. Outside it is only the solar wind dynamic pressure (kinetic energy density). The thermal pressure of the solar wind is small compared with the kinetic energy density. This equation yields the following formula for the subsolar distance:

$$r_{\text{subsolar}} = R_E \left(\frac{B_e^2}{\mu_o \rho v^2} \right)^{1/6} \tag{10.2}$$

where $B_e = 30000$ nT is the equatorial field strength, v is the solar wind speed (often close to 400 km/s) and ρ is the solar wind density (often close to $5 - 1010^6$ protons/m^3). Substituting the number gives the subsolar distance of 8-10 R_E in average conditions.

The solar wind by itself is rather variable. The density and the velocity can easily vary tens of percents. The magnetic field varies even more. Even though magnetic fields of about 5 nT are the most commonly quoted, at least for the Z component, values of 20-30 nT are not exceptional. Thus, the solar wind beta parameter also varies. Its "average" value is close to unity, but it can vary quite a lot in both magnetized and non-magnetized directions. The only thing which apparently always remains valid is the fully supersonic nature of the solar wind at 1 AU: the solar wind speed is always larger than the local fast MHD wave phase speed.

10.3 Reconnection

Processes which change the magnetic field topology are called *reconnection* of field lines. Reconnection can potentially take place if there are oppositely directed magnetic fields on two sides of a current sheet. All current sheets do not share this property. Therefore the situation is qualitatively dissimilar with the Kelvin-Helmholtz instability: in K-H it was not essential that the velocities were in opposite directions; the magnitude of the velocity difference was all that mattered. However, we cannot transfer away the magnetic field by a (nonrelativistic) coordinate transformation; thus current sheets which change the sign of the magnetic field are candidates for reconnection whereas current sheets which merely change the magnitude of the B field are not.

In Earth's magnetosphere reconnection can happen in mainly two places: (1) On the dayside magnetopause if the solar wind magnetic field (interplanetary magnetic field, IMF) is locally opposite to the magnetospheric magnetic field. (2) In the nightside tail current sheet. We shall first consider the dayside reconnection.

Dayside reconnection can possibly occur if IMF is opposite to the magnetospheric field, which is qualitatively similar to the dipole field. At the subsolar point this means that IMF B_z must be negative. On other places also the X and Y components of the IMF affect the decision whether or not reconnection is possible.

A two-dimensional picture of reconnection (Figure 10.1) makes the onset of the process apparent but it does not tell us what happens when conditions remain suitable for reconnection for a longer time. Obviously the dayside magnetopause cannot erode very far because the dipole field quickly becomes overwhelmingly strong when we approach the Earth. How can the process continue? The answer is probably that reconnection gives rise to plasma motion in both Z and Y directions. The net result is enhanced circulation of the plasma inside the magnetosphere. Because the magnetic flux through a co-moving loop remains constant in time, the circulation will involve both the magnetic field and the plasma. In other words, reconnection acts as an effective friction between the solar wind plasma and the magnetospheric plasma. It is an open question whether or not the magnetic field exhibits substantial time variations or not in such a situation if the solar wind is smooth (in other words, is steady reconnection a meaningful concept).

Reconnection will also yield into some plasma input to the magnetosphere, because the higher-density solar wind plasma easily flows inside the magnetosphere through the reconnected flux tubes. Thus, dayside reconnection is a source of mass, momentum and energy, as far as magnetosphere is concerned.

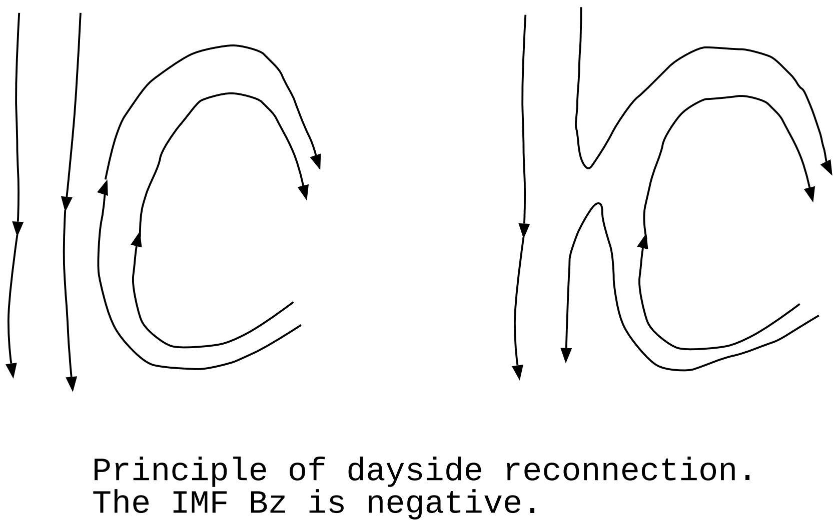

Figure 10.1: Principle of dayside reconnection. The left panel shows the situation before reconnection when IMF B_z has turned negative. The right panel is the situation after reconnection has taken place.

10.4 Substorms

Dayside reconnection and the associated circulation of plasma gives rise to lengthening of the magnetotail and the thinning of the tail current sheet. Stretching of the tail also involves increasing the neutral sheet plasma pressure and density. This is to make the total pressure (magnetic + thermal) approximately constant in the tail: stretching means increased B field in the lobes which must be balanced by an increased thermal pressure in the neutral sheet, where the magnetic field is weak and the plasma beta is high. Where does the extra plasma come to the neutral sheet? It is originally solar wind plasma that has entered the magnetosphere through dayside reconnection.

Stretching of the tail increases the total magnetic energy of the tail. Thus, at least part of the energy that has entered the magnetosphere by dayside reconnection is stored in the magnetotail as B^2 energy, while part of the extra plasma is stored in the neutral sheet. (HT: where does the momentum go?)

Magnetic substorm is an explosive process that releases some of the magnetic energy of the tail. Probably also the neutral sheet plasma energy is released, but the energy stored in the neutral sheet plasma is smaller than the magnetic energy of the lobes because the lobes are larger in spatial volume. In the process the near-Earth field gets a dipolar form (dipolarization). It is also possible that the rest of the magnetotail is cut off, resulting in a tailward propagating plasmoid (an approximately spherical plasma body with self-connected looplike magnetic field). However, it is unclear whether the formation of a plasmoid (for which there is observational evidence) is related to substorm onsets.

The magnetic substorm is sometimes described qualitatively in terms of Klimas' "dripping faucet" analog model. The water drop at the nozzle of the faucet represents the energy stored in the magnetotail. When it becomes large enough, it will drop. The leaking rate of the faucet represents the dayside reconnection (input of mass, momentum and energy); it is usually time dependent. When the drop is hanging on the faucet, it may be released any time because of some external disturbance (e.g. a shock front in solar wind). Even without external trigger the drop will eventually be released when it gets large enough. The surface tension of the drop is the analog of the magnetic tension of the field lines.

The dipolarization of the magnetic field is accompanied by a collapse of the neutral sheet. This is at least partly responsible for the enhanced precipitation in the ionosphere. In the fluid picture, the flux tube volumes shrink, which raises the pressure and the temperature.

The triggering mechanism of the substorm is an open question. It is also not sure whether there is only one mechanism or several ones operating in succession, as was the case in the thunderstorm example of the last chapter. Candidate instabilities include the tearing mode instability of the thin current sheet and a ballooning mode instability of the near-Earth region. Other mechanisms have also been proposed. Some of the "substorm theories" do not try specify the onset instability mechanism in detail at all.

10.4.1 Role of the ionosphere

It is clear that the ionosphere has some sort of impact on the substorm process. Whether the impact is decisive or not clear. However, quantitative models of ionosphere-magnetosphere coupling all indicate that if the ionospheric conductivity is zero or infinite makes a huge difference. If the ionospheric conductivity would be infinite, the ionospheric electric field would be zero. That would virtually hinder any circulation of plasma in the magnetosphere, and possibly also the preconditions of auroral substorms.

The ionospheric conductivity (more accurately speaking, both Pedersen and Hall conductivities must be considered) is about the only means the ionosphere can affect the substorm process. Other means are of the chemical type (outflow of oxygen). Then a question arises whether it is high or low ionospheric conductivity that is more favorable for substorm onsets. The answer depends on the nature of the onset instability. A current-related instability such as the tearing mode instability (in the framework of the tail current disruption model) benefits from high ionospheric conductivity, because the instability wants to circulate part of the magnetospheric current through the ionosphere. This will happen more easily if the ionospheric conductivity is high. On the other hand, an electric field-related instability such as the ballooning mode instability (which only involves changes in the plasma velocity, not in the magnetic field) is more easily excited if the ionospheric conductivity is low. A lower conductivity will mean higher electric fields, and thus higher plasma velocities, because $v \sim E/B$. An infinitely conducting ionosphere would not allow any interchange instability of the flux tubes to take place, because the flux tubes would sit tightly in the ionosphere.

If strong electric fields persist for an hour or so in the ionosphere, the neutral atmosphere in the E region will start to co-move with the $\mathbf{E}\times\mathbf{B}$ velocity because ion-neutral collisions transfer momentum between the two species. Once a neutral wind has been set up by this mechanism, it takes equally much time to stop it. It has been proposed that sometimes a substorm might be triggered by a strong neutral wind, which is a remnant of a previous substorm. If there is any unexpected correlation in the solar wind - geomagnetic index relationships on the few hour time delay range, one possible reason is this ionospheric "flywheel" effect. Neutral winds affect the plasma flow in the same way as the corotation electric field. One can think of Earth's rotation as a neutral wind of a particular simple geometry.

The outflow of ionospheric oxygen due to upward field-aligned currents has also been proposed as a possible factor assisting some onset-related instabilities. The oxygen outflow will increase the magnetospheric plasma mass density substantially. It depends on the instability whether or not a higher mass density is stabilizing of destabilizing. It can hardly be denied that some oxygen outflow will occur during the expansion phase of a substorm, but whether it takes place quickly enough to affect the onset itself is not clear.

10.4.2 Miscellaneous notes about substorms

- The main energy storage is the lobe magnetic field, thus any substorm theory must involve a substantial change in the magnetic field. (For example, the ballooning mode theory by itself is thus not sufficient.) The signature of a profound field change (dipolarization) is also very evident from satellite magnetometer data.
- The ionospheric conductivity has an important role in regulating the plasma flow in the perpendicular (w.r.t $\mathbf{B}$) direction. If the ionospheric conductivity were infinite, no plasma circulation in the magnetosphere would be possible.

- The auroral onset usually (always?) happens at the most equatorward arc, i.e., rather close to the Earth, maybe sometimes even inside the geosynchronous orbit ($6.6 R_E$).
- Inside geosynchronous orbit the particle distribution functions are complicated. There are the dense but cold plasmaspheric population and the dilute and very hot radiation belt particle populations. Thus it is not clear that MHD works.