Key Notes in Plasma Physics

Hongyang Zhou

2/8/23

Table of contents

Welcome

This website is generated by Quarto. The content is also available as **PDF**.

1 About

These are my notes about plasma physics from multiple courses and textbooks, built via Books.jl and is made possible by the Julia programming language (Bezanson et al. 2017) and pandoc.

Thanks to Prof. F. F. Chen, Y. Y. Lau, Yi Li, Yuming Wang, Richard Fitzpatrick, Paul Bellan, and Gábor Tóth through my journey of learning about plasmas. The backbone of the notes is (F. F. Chen 2016). I also learned a lot from Richard Fitzpatrick's online Plasma Physics Course Notes. Y. Y. Lau's graduate course notes also have many references to (Bellan 2008). (Kilpua and Koskinen 2017) is a faily short and concise introduction that contains many cute practical notes.

The future is about nuclear fusion: what do you think?

2 Introduction

- What is plasma?

Quasi-neutral ionized gases. The fourth state of matter: solid, liquid, gas, **plasma**.

- Where is plasma?
  - Astrophysics & geophysics
    - Dynamo
    - Shock
    - Reconnection
    - Particle acceleration
  - Engineering & application
    - Controlled fusion
    - Power conversion
    - Reentry of intercontinental ballistic missiles and spacecrafts
    - Plasma jet as a new propulsion for space vehicles
    - Global warming
- Methodologies applied in plasma physics
  - Kinetic theory
    - Microscopic point of view
    - Positive, negative & neutral particles
    - EM force, collisions
    - Ordinary gas .vs. Plasma
      - binary collision
      - free between collisions

$$
\begin{aligned}
&nf(\mathbf{x}+\mathrm{d}\mathbf{x},\mathbf{p}+\mathrm{d}\mathbf{p},t+\mathrm{d}t)\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{p}-nf(\mathbf{x},\mathbf{p},t)\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{p}\\
=&\frac{\partial(nf(\mathbf{x},\mathbf{p},t))}{\partial t}\Big|_{\text{collision}}\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{p}\mathrm{d}t\\
=&\frac{\partial(nf)}{\partial t}+\frac{\partial nf}{\partial \mathbf{x}}\cdot\frac{\mathbf{p}}{m}+\frac{nf}{\partial \mathbf{p}}\cdot\mathbf{F}
\end{aligned}
$$

This is also called the *Fokker-Planck* term.

- Theory of continuum
  - Macroscopic point of view, bulk motion
  - Fundamental equations from conservation laws
  - Electro-Magneto-Hydro-Dynamics (EMHD)

Definition: Plasma = *ionized gas*

In plasma physics we usually use energy units "eV" instead of temperature. Given Boltzmann constant $k_B = 1.38 \times 10^{-23} J/K$,

$$1\text{eV} = (11600\text{K}) \times k_B = 1.602 \times 10^{-19}\text{J}([V] = [J]/[C])$$

There are generally two kinds of plasma:

1. High temperature(> 1keV): fusion plasma
2. Low temperature($\sim 1\text{eV} - 100\text{eV}$): plasma processing, space propulsion

At the beginning, it is important to get the idea of size. Starting from two basic length scales:

- $a_0 \equiv$ atomic scale ($\sim 10^{-10}$m)
- n = number density ($\sim 1/\text{cm}^3$ to $10^{20}/\text{cm}^3$)

Physically, n represents the number of particles in a unit volume, from which we know $n^{-1/3}$ has unit of length. It means a typical length for a single particle. Ionized state implies that $k_B T \gg E_i =$ ionization potential energy$\sim \frac{e^2}{4\pi\epsilon_0 a_0} \approx 14eV$).

Gaseous state implies that $n^{-1/3} \gg a_0$.

$$n^{-1/3} \gg a_0 \sim \frac{e^2}{4\pi\epsilon_0 E_i} \gg \frac{e^2}{4\pi\epsilon_0 (k_B T)}$$
$$4\pi n^{-1/3} n\Big(\frac{\epsilon_0 k_B T}{e^2 n}\Big) \gg 1$$

$n^{2/3}$ has unit of length^{-2}, indicating that the other term should be in the unit of length^2. This is derived using dimensional analysis, which gives

$$\lambda_D = \sqrt{\frac{\epsilon_0 k_B T}{e^2 n}} \equiv \text{Debye length} = 740(\text{cm})\sqrt{\frac{T(\text{eV})}{n(\text{cm}^{-3})}}$$

$$(4\pi)^{3/2} n {\lambda_D}^3 \gg 1$$
$$n{\lambda_D}^3 \gg 1 \Rightarrow \text{Large number of particles in a sphere}$$

This is necessarily satisfied for plasma under previous definitions.

Indication from the definition of plasma:

1. There are large number of particles in a Debye sphere.

Ex.1 laboratory plasma $10^8\text{cm}^{-3} \to 10^{20}\text{cm}^{-3}$. Tokamak $n = 10^{14}\text{cm}^{-3}$, $k_B T = 10\text{keV} = 10^4\text{eV}$, $n^{-1/3} = 2.15 \times 10^{-5}\text{cm}$, $\lambda_D = 7.4 \times 10^{-3}\text{cm} \gg n^{-1/3}$, $n{\lambda_D}^3 \gg 1$.

$L \gg \lambda_D \gg n^{-1/3}$, where L is the scale of the physical system ($\sim$ 1mm to 10m).

Put it in another way, a criterion for an ionized gas to be a plasma is that it be dense enough that λ_D is much smaller than L.

Ex.2 Interstellar plasma $n = 1\text{cm}^{-3}, T = 1eV$, $\lambda_D = 740\text{cm}\sqrt{\frac{1}{1}} = 740\text{cm} \gg n^{-1/3}$, $n{\lambda_D}^3 \gg 1$.

2. Particle kinetic energy $\gg$ average Coulombic interaction energy.

$$
\begin{aligned}
KE &\sim k_B T \\
PE &\sim \frac{e^2}{4\pi\epsilon_0 (n^{-1/3})}
\end{aligned}
$$

From the definition we can easily know that

$$
k_B T \gg \frac{e^2}{4\pi\epsilon_0 n^{-1/3}} \Rightarrow KE \gg PE
$$

This implies that particles are hard to be deflected by their neighbours.

3. Particles are scattered mostly by accumulative, multiple small angle collisions rather than by large angle collisions. This naturally follows the above energy discussion, and will be explained in further detail later.

4. Collective interactions (the basics of all instabilities) dominate over Coulombic interactions between individual particles. This has an analogy to economy: one has nearly no effect on the whole, but the whole has huge effect on individuals. EM fields stand as a bridge between particles.

2.1 Occurrence of Plasma in Nature

It is now believed that the universe is made of 69% dark energy, 27% dark matter, and 1% normal matter. All that we can see in the sky is the part of normal matter that is in the plasma state, emitting radiation. Plasma in physics, not to be confused with blood plasma, is an "ionized" gas in which at least one of the electrons in an atom has been stripped free, leaving a positively charged nucleus, called an ion. Sometimes plasma is called the "fourth state of matter". When a solid is heated, it becomes a liquid. Heating a liquid turns it into a gas. Upon further heating, the gas is ionized into a plasma. Since a plasma is made of ions and electrons, which are charged, electric fields are rampant everywhere, and particles "collide" not just when they bump into one another, but even at a distance where they can feel their electric fields. Hydrodynamics, which describes the flow of water through pipes, say, or the flow around boats in yacht races, or the behavior of airplane wings, is already a complicated subject. Adding the electric fields of a plasma greatly expands the range of possible motions, especially in the presence of magnetic fields.

Plasma usually exists only in a vacuum. Otherwise, air will cool the plasma so that the ions and electrons will recombine into normal neutral atoms. In the laboratory, we need to pump the air out of a vacuum chamber. In the vacuum of space, however, much of the gas is in the plasma state, and we can see it. Stellar interiors and atmospheres, gaseous nebulas, and entire galaxies can be seen because they are in the plasma state. On earth, however, our atmosphere limits our experience with plasmas to a few examples: the flash of a lightning bolt, the soft glow of the Aurora Borealis, the light of a fluorescent tube, or the pixels of a plasma TV. We live in a small part of the universe where plasmas do not occur naturally; otherwise, we would not be alive.

The reason for this can be seen from the Saha equation, which tells us the amount of ionization to be expected in a gas in thermal equilibrium:

$$\frac{n_i}{n_n} \approx 2.4 \times 10^{21} \frac{T^{3/2}}{n_i} e^{-\frac{U_i}{k_B T}} \tag{2.1}$$

Here n_i and n_n are, respectively, the density (number per m^3) of ionized atoms and of neutral atoms, T is the gas temperature in Kelvin, k_B is Boltzmann's constant, and U_i is the ionization energy of the gas — that is, the number of joules required to remove the outermost electron from an atom. For ordinary air at room temperature, we may take $n_n \approx 3 \times 10^{25}\,\mathrm{m}^{-3}$, $T \approx 300\,\mathrm{K}$, and $U_i = 14.5\,\mathrm{eV}$ (for nitrogen), where $1\,\mathrm{eV} = 1.6 \times 10^{19}\,\mathrm{J}$. The fractional ionization $n_i/(n_n + n_i) \approx n_i/n_n$ is ridiculously low:

$$\frac{n_i}{n_n} \approx 10^{-122}$$

As the temperature is raised, the degree of ionization remains low until U_i is only a few times $k_B T$. Then n_i/n_e rises abruptly, and the gas is in a plasma state. Further increase in temperature makes n_n less than n_i, and the plasma eventually becomes fully ionized. This is the reason plasmas exist in astronomical bodies with temperatures of millions of degrees, but not on the earth. Life could not easily coexist with a plasma — at least, plasma of the type we are talking about. The natural occurrence of plasmas at high temperatures is the reason for the designation "the fourth state of matter".

Although we do not intend to emphasize the Saha equation, we should point out its physical meaning. Atoms in a gas have a spread of thermal energies, and an atom is ionized when, by chance, it suffers a collision of high enough energy to knock out an electron. In a cold gas, such energetic collisions occur infrequently, since an atom must be accelerated to much higher than the average energy by a series of "favorable" collisions. The exponential factor in Equation 2.1 expresses the fact that the number of fast atoms falls exponentially with $U_i/k_B T$. Once an atom is ionized, it remains charged until it meets an electron; it then very likely recombines with the electron to become neutral again. The recombination rate clearly depends on the density of electrons, which we can take as equal to n_i. ni. The equilibrium ion fraction, therefore, should decrease with n_i; and this is the reason for the factor ni^{-1} on the right-hand side of Equation 2.1. The plasma in the interstellar medium owes its existence to the low value of n_i (about 1 per cm^3), and hence the low recombination rate.

2.2 Definition of Plasma

Any ionized gas cannot be called a plasma, of course; there is always some small degree of ionization in any gas. A useful definition is as follows:

A plasma is a quasineutral gas of charged and neutral particles which exhibits collective behavior.

We must now define "*quasineutral*" and "*collective behavior*". The meaning of quasineutrality will be made clear in Sect. 1.4 LINK?. What is meant by "collective behavior" is as follows.

Consider the forces acting on a molecule of, say, ordinary air. Since the molecule is neutral, there is no net electromagnetic force on it, and the force of gravity is negligible. The molecule moves undisturbed until it makes a collision with another molecule, and these collisions control the particle's motion. A macroscopic force applied to a neutral gas, such as from a loudspeaker generating sound waves, is transmitted to the individual atoms by collisions. The situation is totally different in a plasma, which has charged particles. As these charges move around, they can generate local concentrations of positive or negative charge, which give rise to electric fields. Motion of charges also generates currents, and hence magnetic fields. These fields affect the motion of other charged particles far away.

Let us consider the effect on each other of two slightly charged regions of plasma separated by a distance r. ADD FIGURE! The Coulomb force between A and B diminishes as $1/r^2$.

However, for a given solid angle (that is, $\Delta r/r$ = constant), the volume of plasma in B that can affect A increases as r^3. Therefore, elements of plasma exert a force on one another even at large distances. It is this long-ranged Coulomb force that gives the plasma a large repertoire of possible motions and enriches the field of study known as plasma physics. In fact, the most interesting results concern so-called "collisionless" plasmas, in which the long-range electromagnetic forces are so much larger than the forces due to ordinary local collisions that the latter can be neglected altogether. By "collective behavior" we mean motions that depend not only on local conditions but on the state of the plasma in remote regions as well.

The word "plasma" seems to be a misnomer. It comes from the Greek , o , o, which means something molded or fabricated. Because of collective behavior, a plasma does not tend to conform to external influences; rather, it often behaves as if it had a mind of its own.

2.3 Concept of Temperature

Before proceeding further, it is important to review and extend our physical notions of "temperature". A gas in thermal equilibrium has particles of all velocities, and the most probable distribution of these velocities is known as the Maxwellian distribution. For simplicity, consider a gas in which the particles can move only in one dimension. (This is not entirely frivolous; a strong magnetic field, for instance, can constrain electrons to move only along the field lines.) The one-dimensional Maxwellian distribution is given by

$$f(v) = A \exp\left(-\frac{1}{2}mv^2/k_B T\right) \tag{2.2}$$

where $f\mathrm{d}u$ is the number of particles per m^3 with velocity between u and $u + \mathrm{d}u$, $\frac{1}{2}mu^2$ is the kinetic energy, and k_B is Boltzmann's constant. The density n, or number of particles per m^3, is given by

$$n = \int_{-\infty}^{\infty} f(v)\mathrm{d}v$$

The constant A is related to the density n by

$$A = n\left(\frac{m}{2\pi k_B T}\right)^{1/2}$$

The width of the distribution is characterized by the constant T, which we call the temperature. To see the exact meaning of T, we can compute the average kinetic energy of particles in this distribution:

$$E_{\mathrm{av}} = \frac{\int_{-\infty}^{\infty} \frac{1}{2} m v^2 f(v) \mathrm{d}v}{\int_{-\infty}^{\infty} f(v) \mathrm{d}v} \tag{2.3}$$

Defining $y = v/v_{\mathrm{th}}$ and

$$v_{\mathrm{th}} = (2k_B T/m)^{1/2} \tag{2.4}$$

we can write Equation 2.2 as

$$f(v) = A \exp(-v^2/v_{\mathrm{th}}^2)$$

and Equation 2.3 as

$$E_{\mathrm{av}} = \frac{\frac{1}{2} m A v_{\mathrm{th}}^3 \int_{-\infty}^{\infty} [\exp(-y^2)] y^2 \mathrm{d}y}{A v_{\mathrm{th}} \int_{-\infty}^{\infty} [\exp(-y^2)] \mathrm{d}y}$$

The integral in the numerator is integrable by parts:

$$\begin{aligned} \int_{-\infty}^{\infty} y \cdot [\exp(-y^2)] y \mathrm{d}y &= \left[-\frac{1}{2} [\exp(-y^2)] y \right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} -\frac{1}{2} [\exp(-y^2)] \mathrm{d}y \\ &= \frac{1}{2} \int_{-\infty}^{\infty} [\exp(-y^2)] \mathrm{d}y \end{aligned}$$

Canceling the integrals, we have

$$E_{\mathrm{av}} = \frac{\frac{1}{2} m A v_{\mathrm{th}}^3 \frac{1}{2}}{A v_{\mathrm{th}}} = \frac{1}{4} m v_{\mathrm{th}}^2 = \frac{1}{2} k_B T$$

Thus the average kinetic energy is $\frac{1}{2} k_B T$.

It is easy to extend this result to three dimensions. Maxwell's distribution is then

$$f(v_x, v_y, v_z) = A_3 \exp\left[-\frac{1}{2} m (v_x^2 + v_y^2 + v_z^2)/k_B T \right]$$

where

$$A_3 = n \left(\frac{m}{2\pi k_B T} \right)^{3/2}$$

Because a Maxwellian distribution is isotropic (i.e. the form is symmetric in v_x, v_y, and v_z), we can separate each dimension. The average kinetic energy is then 3 times the single dimension result

$$E_{\text{av}} = \frac{3}{2} k_B T$$

The general result is that E_{av} equals $\frac{1}{2} k_B T$ per degree of freedom.

Since T and E_{av} are so closely related, it is customary in plasma physics to give temperatures in units of energy. To avoid confusion on the number of dimensions involved, it is not E_{av} but the energy corresponding to $k_B T$ that is used to denote the temperature. For $k_B T = 1\,\text{eV} = 1.6 \times 10^{-19}\,\text{J}$, we have

$$T = \frac{q}{k_B} = 11600$$

Thus the conversion factor is

$$1\,\text{eV} = 11600\,\text{K}$$

By a 2-eV plasma we mean that $k_B T = 2\,\text{eV}$, or $E_{\text{av}} = 3\,\text{eV}$ in three dimensions.

It is interesting that a plasma can have several temperatures at the same time. It often happens that the ions and the electrons have separate Maxwellian distributions with different temperatures T_i and T_e. This can come about because the collision rate among ions or among electrons themselves is larger than the rate of collisions between an ion and an electron. Then each species can be in its own thermal equilibrium, but the plasma may not last long enough for the two temperatures to equalize. When there is a magnetic field $\mathbf{B}$, even a single species, say ions, can have two temperatures. This is because the forces acting on an ion along $\mathbf{B}$ are different from those acting perpendicular to $\mathbf{B}$ (due to the Lorentz force). The components of velocity perpendicular to $\mathbf{B}$ and parallel to $\mathbf{B}$ may then belong to different Maxwellian distributions with temperatures $T_\perp$ and $T_\parallel$.

Before leaving our review of the notion of temperature, we should dispel the popular misconception that high temperature necessarily means a lot of heat. People are usually amazed to learn that the electron temperature inside a fluorescent light bulb is about 20000 K. "My, it doesn't feel that hot!" Of course, the heat capacity must also be taken into account. The density of electrons inside a fluorescent tube is much less than that of a gas at atmospheric pressure, and the total amount of heat transferred to the wall by electrons striking it at their thermal velocities is not that great. For example, the temperature of a cigarette ash is high enough to cause a burn, but the total amount of heat involved is not. Many laboratory plasmas have temperatures of the order of $1,000,000\,\text{K}$ ($100\,\text{eV}$), but at densities of only $10^{18} - 10^{19}$ per m^3, the heating of the walls is not a serious consideration.

2.4 Debye Shielding

(Problem 4 on P27 (Bellan 2008))

A fundamental characteristic of the behavior of plasma is its ability to shield out electric potentials that are applied to it. Suppose we tried to put an electric field inside a plasma by inserting two charged balls connected to a battery (Fig. 1.3 ADD IT!). The balls would attract particles of the opposite charge, and almost immediately a cloud of ions would surround the negative ball and a cloud of electrons would surround the positive ball. (We assume that a layer of dielectric keeps the plasma from actually recombining on the surface, or that the battery is large enough to maintain the potential in spite of this.) If the plasma were cold and there were no thermal motions, there would be just as many charges in the cloud as in the ball, the shielding would be perfect, and no electric field would be present in the body of the plasma outside of the clouds. On the other hand, if the temperature is finite, those particles that are at the edge of the cloud, where the electric field is weak, have enough thermal energy to escape from the electrostatic potential well. The "edge" of the cloud then occurs at the radius where the potential energy is approximately equal to the thermal energy k_BT of the particles, and the shielding is not complete. Potentials of the order of k_BT/e can leak into the plasma and cause finite electric fields to exist there.

Let us compute the approximate thickness of such a charge cloud. Let us put a point charge q_T into a quasi-neutral gas of charged particles. The gas is in *local thermodynamic equilibrium* (LTE) with ion and electron temperatures T_e and T_i (They are not necessarily the same, and in fact, quite different in nature.). (ADD FIGURE!) We wish to compute $\phi(x)$. For simplicity, we assume that the ion-electron mass ratio M/m is infinite, so that the ions do not move but form a uniform background of positive charge. To be more precise, we can say that M/m is large enough that the inertia of the ions prevents them from moving significantly on the time scale of the experiment.

Maxwell distribution:

$$f(\mathbf{v}) = n_0\Big(\frac{m}{2\pi kT}\Big)^{3/2}\exp\Big(-\frac{mv^2}{2kT}\Big)$$

In the presence of a potential energy $q\phi$m the distribution function follows the Boltzmann distribution:

$$f(\mathbf{x},\mathbf{v}) = n_0\Big(\frac{m}{2\pi kT}\Big)^{3/2}\exp\Big(-\frac{\frac{1}{2}mv^2+q\phi(\mathbf{x})}{kT}\Big)$$

It would not be worthwhile to prove this here. What this equation says is intuitively obvious: There are fewer particles at places where the potential energy is large, since not all particles have enough energy to get there.

Note that the concept of temperature is valid in LTE. From the Boltzmann equation,

$$\begin{aligned}\frac{\mathrm{d}f}{\mathrm{d}t} &= \frac{\partial f}{\partial t} + \nabla_x \cdot (f\mathbf{v}) + \nabla_v \cdot (f\mathbf{a}) = 0 \\ &\Rightarrow \frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla_x f + \frac{\mathbf{F}}{m} \cdot \nabla_v f = 0\end{aligned}$$

If we neglect the time derivatives, we obtain

$$(\mathbf{v}_j \cdot \nabla) f(\mathbf{r}, \mathbf{v}_j) + [-\frac{1}{m_j}\nabla(q_j\phi) \cdot \nabla_v] f(\mathbf{r}, \mathbf{v}_j) = 0 \tag{2.5}$$

The LTE distribution is in a Maxwellian shape,

$$f(\mathbf{r}, \mathbf{v}_j) = n_j(\mathbf{r})\Big(\frac{m}{2\pi k_B T}\Big)^{3/2} \exp\Big(-\frac{mv^2}{2k_B T}\Big)$$

Substitute into Equation 2.5, we have

$$\begin{aligned}(\mathbf{v}_j \cdot \nabla) n_j + \frac{\mathbf{v}_j \cdot \nabla(q\phi)}{k_B T_j} n_j = 0 \\ \Rightarrow n_j(\mathbf{r}) = n_{j0} \exp\Big[-\frac{q_j \phi_j(\mathbf{r})}{k_B T_j}\Big]\end{aligned}$$

Then we get the Boltzmann distribution of each species with a separate temperature for each.

Consider a uniform plasma with $n_i = n_{i0} e^{-e\phi/k_B T_i}$, $n_e = n_{e0} e^{e\phi/k_B T_e}$, and with the quasi-neutral assumption, $n_{e0} = n_{i0} = n_0$. Introduce a point charge q_T in the initially neutralized plasma. Poisson equation gives

$$\nabla^2 \phi = -\frac{\rho}{\epsilon_0} = -\frac{1}{\epsilon_0}\Big[e(n_e - n_i) + q_T \delta(\mathbf{r})\Big]$$

Due to symmetry, $\phi = \phi(r)$. When the temperature is high, $q\phi(r) \ll kT$, the exponent is small and we can expand in a Taylor series and leave only the first order term. In spherical coordinates,

$$
\begin{aligned}
\nabla^2\phi &= \frac{1}{r^2}\frac{\partial}{\partial r}\Big(r^2\frac{\partial\phi}{\partial r}\Big)\\
\frac{1}{r^2}\frac{\partial}{\partial r}\Big(r^2\frac{\partial\phi}{\partial r}\Big)+\frac{e}{\epsilon_0}\Big(\underbrace{n_{e0}e^{e\phi/k_BT_e}-n_{i0}e^{-e\phi/k_BT_i}}_{\approx n_0\left[e\phi/k_BT_e+e\phi/k_BT_i\right]}\Big) &= -\frac{q_T}{\epsilon_0}\delta(\mathbf{r})\\
\frac{1}{r^2}\frac{\partial}{\partial r}\Big(r^2\frac{\partial\phi}{\partial r}\Big)-\underbrace{\frac{e^2n_0}{\epsilon_0}\Big(\frac{1}{k_BT_e}+\frac{1}{k_BT_i}\Big)}_{\frac{1}{\lambda_D{}^2}\equiv\frac{1}{\lambda_{Di}{}^2}+\frac{1}{\lambda_{De}{}^2}}\phi &= -\frac{q_T}{\epsilon_0}\delta(\mathbf{r})\\
\frac{1}{r^2}\frac{\partial}{\partial r}\Big(r^2\frac{\partial\phi}{\partial r}\Big)-\frac{1}{\lambda_D{}^2}\phi &= -\frac{q_T}{\epsilon_0}\delta(\mathbf{r})
\end{aligned}
$$

where

$$
\lambda_D = \sqrt{\frac{\epsilon_0 k_B T^*}{n_0 e^2}} \tag{2.6}
$$

is the Debye length, $T^* = \left(\frac{1}{T_e}+\frac{1}{T_i}\right)^{-1}$. The Debye length, is a measure of the shielding distance or thickness of the sheath. Note that larger density indicates better shielding effects, as one would expect, since each layer of plasma contains more electrons. Furthermore, λ_D increases with increasing k_BT^*. Without thermal agitation, the charge cloud would collapse to an infinitely thin layer. Finally, most commonly it is the *electron* temperature ($T^* \approx T_e$) which is used in the definition of λ_D because the electrons, being more mobile than the ions, generally do the shielding by moving so as to create a surplus or deficit of negative charge. Only in special situations is this not true.

For $r > 0$ (away from the charge),

$$
\frac{1}{r^2}\frac{\partial}{\partial r}\Big(r^2\frac{\partial\phi}{\partial r}\Big)-\frac{1}{\lambda_D{}^2}\phi = 0
$$

Let $\phi(r) = \frac{F(r)}{r}$, we have

$$
\begin{aligned}
\frac{\mathrm{d}^2F}{\mathrm{d}r^2}-\frac{F}{\lambda_D{}^2} &= 0\\
\Rightarrow F &= Ae^{-r/\lambda_D}+\cancel{Be^{r/\lambda_D}}(r>0)\\
&= Ae^{-r/\lambda_D}
\end{aligned}
$$

Anticipate when $r\to 0$, $\phi(r)\to\frac{q_T}{4\pi\epsilon_0 r}\approx\frac{A}{r}$. So

$$\phi(r) = \frac{q_t}{4\pi\epsilon_0 r} e^{-r/\lambda_D}$$

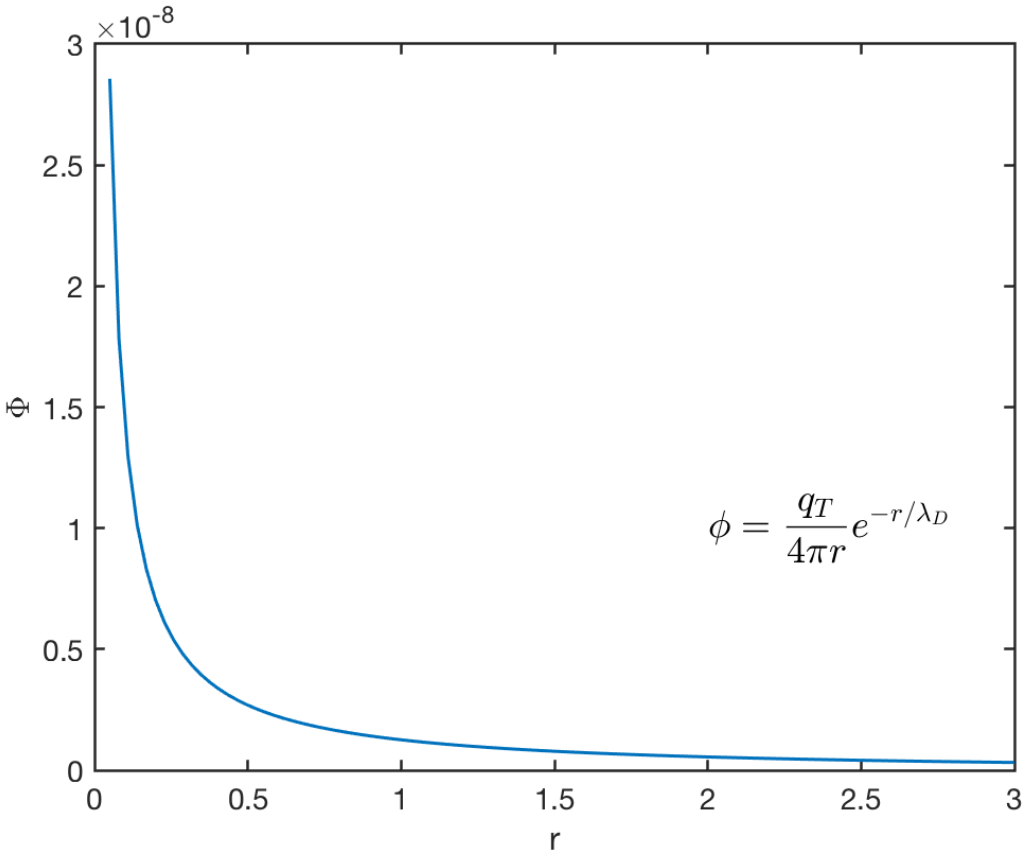

Figure 2.1: Electric potential.

The electric potential drops very quickly (Figure 2.1). In other words, if you insert a charge into a neutral plasma, its potential will only effect within a small range. Recall the Gauss's law,

$$\oint \mathbf{E} \cdot \mathrm{d}\mathbf{A} = \frac{Q_T}{\epsilon_0} \rightarrow \text{ as } R \rightarrow \infty, \text{ E goes to zero.}$$

The following are useful forms of Equation 2.6:

$$\lambda_D = 69(T_e/n)^{1/2}\,\mathrm{m}, \quad T_e \text{ in K}$$
$$\lambda_D = 7430(k_B T_e/n)^{1/2}\,\mathrm{m}, \quad k_B T_e \text{ in eV}$$

We are now in a position to define “quasineutrality.” If the dimensions L of a system are much larger than λ_D, then whenever local concentrations of charge arise or external potentials are introduced into the system, these are shielded out in a distance short compared with L, leaving the bulk of the plasma free of large electric potentials or fields. Outside of the sheath on the wall or on an obstacle, $\nabla^2\phi$ is very small, and n_i is equal to n_e. It takes only a small charge imbalance to give rise to potentials of the order of k_BT/e. The plasma is “quasineutral”; that is, neutral enough so that one can take $n_i \simeq n_e \simeq n$, where n is a common density called the *plasma density*, but not so neutral that all the interesting electromagnetic forces vanish. Plasma may be considered neutral over a length scale larger than Debye length.

Additional comments:

1. Plasma acts like dielectron.
2. There is a plasma sheath near to where materials contact, inside which charge neutrality is violated. This will be discussed in Section 13.1.

2.5 Criteria for Plasmas

We have given two conditions that an ionized gas must satisfy to be called a plasma. A third condition has to do with collisions. The weakly ionized gas in an airplane’s jet exhaust, for example, does not qualify as a plasma because the charged particles collide so frequently with neutral atoms that their motion is controlled by ordinary hydrodynamic forces rather than by electromagnetic forces. If ω is the frequency of typical plasma oscillations and τ is the mean time between collisions with neutral atoms, we require $\omega\tau > 1$ for the gas to behave like a plasma rather than a neutral gas.

1. $\lambda_D \ll L$
2. $N_D \gg 1$
3. $\omega\tau > 1$

Often in lectures people think the condition 1 and 2 are equivalent; they are NOT exactly. The argument would be that condition 1 is valid when the distribution is Maxwellian, which happens to be the equilibrium distribution for large number of particles; condition 2 is the statistical priori for that to be valid.

2.6 Plasma Frequency

Besides the length scale, we are also interested in time scale. The plasma frequency,

$$\omega_p = \sqrt{\frac{n\, e^2}{\epsilon_0\, m}}$$

is the most fundamental time-scale in plasma physics. Clearly, there is a different plasma frequency for each species. However, the relatively fast electron frequency is, by far, the most important, and references to "the plasma frequency" in most textbooks invariably mean the electron plasma frequency.

ω_p corresponds to the typical electrostatic oscillation frequency of a given species in response to a small charge separation. For instance, consider a one-dimensional situation in which a slab consisting entirely of one charge species is displaced from its quasi-neutral position by an infinitesimal distance δx. The resulting charge density which develops on the leading face of the slab is $\sigma = e\,n\,\delta x$. An equal and opposite charge density develops on the opposite face. The x-directed electric field generated inside the slab is of magnitude $E_x = -\sigma/\epsilon_0 = -e\,n\,\delta x/\epsilon_0$. Thus, Newton's law applied to an individual particle inside the slab yields

$$m\,\frac{\mathrm{d}^2\delta x}{\mathrm{d}t^2} = e\,E_x = -m\,{\omega_p}^2\,\delta x$$

giving $\delta x = \delta x_0\,\cos\omega_p t$.

Plasma frequency is closely associated with plasma oscillation and waves. A more thorough derivation is given in Section 7.1. Note that plasma oscillations will only be observed if the plasma system is studied over time periods τ longer than the plasma period $\tau_p \equiv 1/\omega_p$, and if external actions change the system at a rate no faster than ω_p. In the opposite case, one is clearly studying something other than plasma physics (e.g., nuclear reactions), and the system cannot not usefully be considered to be a plasma. Likewise, observations over length-scales L shorter than the distance $v_t\,\tau_p$ traveled by a typical plasma particle during a plasma period will also not detect plasma behaviour. In this case, particles will exit the system before completing a plasma oscillation. This distance, which is the spatial equivalent to τ_p, is called the *Debye length*, and takes the form

$$\lambda_D \equiv \sqrt{T/m}\;\omega_p^{-1}.$$

Note that

$$\lambda_D = \sqrt{\frac{\epsilon_0\,T}{n\,e^2}}$$

is independent of mass, and therefore generally comparable for different species. Clearly, our idealized system can only usefully be considered to be a plasma provided that

$$\frac{\lambda_D}{L} \ll 1 \tag{2.7}$$

and

$$\frac{\tau_p}{\tau} \ll 1$$

Here, τ and L represent the typical time-scale and length-scale of the process under investigation. It should be noted that, despite the conventional requirement Equation 2.7, plasma physics is capable of considering structures on the Debye scale. The most important example of this is the Debye sheath: i.e., the boundary layer which surrounds a plasma confined by a material surface.

2.7 Spatial Scales

Ion inertial length: the scale at which ions decouple from electrons and the magnetic field becomes frozen into the electron fluid rather than the bulk plasma.

$$d_i = \frac{c}{\omega_{pi}}$$

Electron inertial length: the scale over which the electron distribution function can change substantially. ???

$$r_e = \frac{v_A}{\omega_{pe}}$$

Gyroradius: also known as Larmor radius, the scale of gyration around the magnetic field

$$r_L = \frac{mv_{th}}{qB}$$

2.8 How to Study Plasma Physics

There is no universal theory for plasma physics due to the multi-scale problems. All useful equations are derived under certain scaling approximations —

3 Math

3.1 Vector Identities

Some useful vector identities are listed below:

$$
\begin{aligned}
\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) &= \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) + \mathbf{C}(\mathbf{A} \cdot \mathbf{B}) \\
(\mathbf{A} \times \mathbf{B}) \times \mathbf{C} &= \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) + \mathbf{A}(\mathbf{B} \cdot \mathbf{C}) \\
\nabla \times \nabla f &= 0 \\
\nabla \cdot (\nabla \times \mathbf{A}) &= 0 \\
\nabla \cdot (f\mathbf{A}) &= (\nabla f) \cdot \mathbf{A} + f \nabla \cdot \mathbf{A} \\
\nabla \times (f\mathbf{A}) &= (\nabla f) \times \mathbf{A} + f \nabla \times \mathbf{A} \\
\nabla \cdot (\mathbf{A} \times \mathbf{B}) &= \mathbf{B} \cdot (\nabla \times \mathbf{A}) - \mathbf{A} \cdot (\nabla \times \mathbf{B}) \\
\nabla(\mathbf{A} \cdot \mathbf{B}) &= (\mathbf{B} \cdot \nabla)\mathbf{A} + (\mathbf{A} \cdot \nabla)\mathbf{B} + \mathbf{B} \times (\nabla \times \mathbf{A}) + \mathbf{A} \times (\nabla \times \mathbf{B}) \\
\nabla \cdot (\mathbf{A}\mathbf{B}) &= (\mathbf{A} \cdot \nabla)\mathbf{B} + \mathbf{B}(\nabla \cdot \mathbf{A}) \\
\nabla \times (\mathbf{A} \times \mathbf{B}) &= (\mathbf{B} \cdot \nabla)\mathbf{A} - (\mathbf{A} \cdot \nabla)\mathbf{B} - \mathbf{B}(\nabla \cdot \mathbf{A}) + \mathbf{A}(\nabla \cdot \mathbf{B}) \\
\nabla \times (\nabla \times \mathbf{A}) &= \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}
\end{aligned}
$$

In most cases Einsten notation shall be used to derive all the identities:

$$y = \sum_{i=1}^{3} c_i x^i = c_1 x^1 + c_2 x^2 + c_3 x^3$$

is simplified by the convention to

$$y = c_i x^i$$

The upper indices are not exponents but are indices of coordinates, coefficients or basis vectors. That is, in this context x^2 should be understood as the second component of x rather than the square of x (this can occasionally lead to ambiguity). The upper index position in x^i is because, typically, an index occurs once in an upper (superscript) and once in a lower (subscript) position in a term.

A basis gives such a form (via the dual basis), hence when working on $\mathbf{R}^n$ with a Euclidean metric and a fixed orthonormal basis, one has the option to work with only subscripts. However, if one changes coordinates, the way that coefficients change depends on the variance of the object, and one cannot ignore the distinction. (See this wiki link.)

Commonly used identities:

$$\mathbf{A} \cdot \mathbf{B} = A_i B_i$$
$$\mathbf{A} \times \mathbf{B} = \epsilon_{ijk} A_j B_k$$
$$\nabla \cdot \mathbf{A} = \partial_i A_i$$
$$\nabla \times \mathbf{A} = \epsilon_{ijk} \partial_j A_k$$
$$\epsilon_{ijk}\epsilon_{imn} = \delta_{jm}\delta_{kn} - \delta_{jn}\delta_{km}$$

3.2 Differential of Line Segments

The differential of line segments in a fluid is:

$$\frac{\mathrm{d}\mathbf{l}}{\mathrm{d}t} = \mathbf{l} \cdot \nabla \mathbf{u} \tag{3.1}$$

Proof. Let the two line segments $\mathbf{l}_1$ and $\mathbf{l}_2$ be expressed as starting point vectors and end point vectors:

$$\mathbf{l}_1 = \mathbf{r}_2 - \mathbf{r}_1$$
$$\mathbf{l}_2 = \mathbf{r}_2' - \mathbf{r}_1'$$

After time Δt, $\mathbf{l}_1 \to \mathbf{l}_2$,

$$\mathbf{l}_2 = \mathbf{r}_2' - \mathbf{r}_1' = (\mathbf{r}_2 + \mathbf{v}(r_2)\Delta t) - (\mathbf{r}_1 + \mathbf{v}(\mathbf{r}_1)\Delta t) = \mathbf{l}_1 + (\mathbf{v}(\mathbf{r}_2) - \mathbf{v}(\mathbf{r}_1))\Delta t$$

The velocity difference can be written as

$$\mathbf{v}(x_2, y_2, z_2) - \mathbf{v}(x_1, y_1, z_1) = \frac{\partial \mathbf{v}}{\partial x}(x_2 - x_1) + \frac{\partial \mathbf{v}}{\partial y}(y_2 - y_1) + \frac{\partial \mathbf{v}}{\partial z}(z_2 - z_1) = (\mathbf{l} \cdot \nabla)\mathbf{v}$$

So

$$\mathbf{l}_2 = \mathbf{l}_1 + \mathbf{l} \cdot \nabla \mathbf{v} \cdot \Delta t \Rightarrow \frac{\mathbf{l}_2 - \mathbf{l}_1}{\Delta t} = \mathbf{l} \cdot \nabla \mathbf{v}$$

3.3 Lagrangian vs Eulerian Descriptions

The Lagrangian operator $\mathrm{d}/\mathrm{d}t$ is defined to be the *convective derivative*

$$\frac{\mathrm{d}}{\mathrm{d}t} = \frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla \tag{3.2}$$

which characterizes the temporal rate of change seen by an observer moving with the velocity $\mathbf{u}$. An everyday example of the convective term would be the apparent temporal increase in density of automobiles seen by a motorcyclist who enters a traffic jam of stationary vehicles and is not impeded by the traffic jam.

The convective derivative is sometimes written as $\mathrm{D}/\mathrm{D}t$, which is used to emphasize the fact that the this specific convective derivative is defined using the center-of-mass velocity. Note that $(\mathbf{u} \cdot \nabla)$ is a *scalar* differential operator.

3.4 Complex Analysis

In complex analysis, the following statements are equivalent:

1. $f(z)$ is an analytic function of z in some neighbourhood of z_0.
2. $\oint_C f(z)dz = 0$ for every closed contour C that lies in that neighbourhood of z_0.
3. $df(z)/dz$ exists at $z = z_0$.
4. $f(z)$ has a convergent Taylor expansion about $z = z_0$.
5. The n^{th} derivative $\mathrm{d}^n f(z)/dz^n$ exists at $z = z_0$ for all n.
6. The *Cauchy-Riemann condition* is satisfied at $z = z_0$. Take u and v to be the real and imaginary parts respectively of a complex-valued function of a single complex variable $z = x + iy$, $f(x + iy) = u(x, y) + iv(x, y)$.

$$\begin{aligned}
\frac{\partial u}{\partial x} &= \frac{\partial v}{\partial y} \\
\frac{\partial u}{\partial y} &= -\frac{\partial v}{\partial x}
\end{aligned}$$

An idea of *analytic continuation* is introduced here. In practice, an analytic function is usually defined by means of some mathematical expression — such as a polynomial, an infinite series, or an integral. Ordinarily, there is some region within which the expression is meaningful and does yield an analytic function. Outside this region, the expression may cease to be meaningful, and the question then arises as to whether or not there is any way of extending the definition of the function so that this "extended" function is analytic over a larger region. A simple example is given as follows.

Ex.1 Polynomial series

$$f(z)=\sum_{n=0}^{\infty} z^n$$

which describes an analytic function for $|z|<1$ but which diverges for $|z|>1$. However, the function

$$g(z)=\frac{1}{1-z}$$

is analytic over the whole plane (except at $z=1$), and it coincides with $f(z)$ inside the unit circle.

Such a function $g(z)$, which coincides with a given analytic $f(z)$ over that region for which $f(z)$ is defined by which also is analytic over some extension of that region, is said to be an *analytic continuation* of $f(z)$. It is useful to think of $f(z)$ and $g(z)$ as being the same function and to consider that the formula defining $f(z)$ failed to provide the values of $f(z)$ in the extended region because of some defect in the mode of description rather than because of some limitation inherent in $f(z)$ itself. [c.f. G.F.Carrier, M.Krook and C.E.Pearson, Functions of a Complex Variable, McGraw-Hill (1966), p.63]

Ex.2 Laplace transform

$$\mathscr{L}[1]=\int_0^{\infty} \mathrm{d}t 1\cdot e^{i\omega t}=-\frac{1}{i\omega},\text{ if }\Im(\omega)>0$$

If you have a pure real frequency ω, then when you integrate v over the real axis, at $v=\omega/k$ you will encounter a singular point. Actually, this integration is convergent if and only if $\Im(\omega)>0$. $-\frac{1}{i\omega}$ is the *analytic continuation* of $f(\omega)$ for all complex ω except $\omega=0$.

To calculate the integral around singular points, we may take advantage of the Cauchy integral formula and the residual theorem.

Theorem 2.1 Cauchy integral

Let C_ϵ be a circular arc of radius ϵ, centered at α, with subtended angle θ_0 in counterclockwise direction. Let $f(z)$ be an analytic function on C_ϵ+inside C_ϵ. Then

$$\lim_{\epsilon\to 0}\int_{C_\epsilon}\frac{f(z)dz}{z-\alpha}=i\theta_0 f(\alpha)$$

Proof: On C_ϵ, $z=\alpha+\epsilon e^{i\theta}$, $dz=i\epsilon e^{i\theta}\mathrm{d}\theta$.

$$LHS = \lim_{\epsilon \to 0} \int_{C_\epsilon} \frac{f(\alpha + \epsilon e^{i\theta}) i\epsilon e^{i\theta} \mathrm{d}\theta}{\epsilon e^{i\theta}} = i\theta_0 f(\alpha).$$

Theorem 2.2 Residue

Let $f(z)$ be an analytic function on a closed contour C+inside C. If point α is inside C, we have

$$f(\alpha) = \frac{1}{2\pi i} \oint_c \frac{f(z)dz}{z - \alpha}$$

Proof:

$\frac{f(z)}{z-\alpha}$ is analytic within region bounded by $C + L_1 - C_\epsilon + L_2$, where L_1 and L_2 are two paths that connects/breaks C and C_ϵ. Therefore

$$\int_C + \int_{L_1} - \int_{C_\epsilon} + \int_{L_2} \frac{f(z)}{z - \alpha} dz = 0$$
$$\Rightarrow \oint_C \frac{f(z)}{z - \alpha} dz = \oint_{C_\epsilon} \frac{f(z)}{z - \alpha} dz = 2\pi i f(\alpha) \text{ as } L_1 \to -L_2, \epsilon \to 0.$$

There is also a purely algebraic proof available.

Note that the value of $f(z)$ on C determines value of $f(\alpha)$ for all α within C. This has a close relation to the potential theory. Actually, what Cauchy-Riemann condition says physically is that the potential flow is both irrotational and incompressible!

Theorem 2.3 Residual theorem

Let $f(z)$ be an analytic function on C+inside C. If point α is inside C, we have

$$\oint_c \frac{f(z)dz}{z - \alpha} = 2\pi i f(\alpha) \equiv 2\pi i \mathrm{Res}\Big[\frac{f(z)}{z - x}; z = \alpha\Big] \tag{3.3}$$

Khan Academy has a nice video on this. Applying this powerful theorem, we can calculate many integrals analytically which contain singular points.

Ex.3

$$f(\omega) = \int_{-\infty}^{\infty} dv \frac{e^{iv}}{v - \omega} = 2\pi i e^{i\omega},\ \Im(\omega) > 0$$

Pick a semi-circle contour C_R in the upper plane of complex v. Let C be a closed contour of a line along the real axis $\Re(v)$ and the semi-circle C_R. e^{iv} is analytic along and inside C, so

$$f(\omega) = \Big(\oint_C - \int_{C_R}\Big)\frac{dve^{iv}}{v-\omega} = 2\pi i e^{i\omega} - \int_{C_R}\frac{dve^{iv}}{v-\omega} = 2\pi i e^{i\omega} \text{ as } R\to\infty$$

$$\Big(e^{iv} = e^{i(v_r+iv_i)} = e^{iv_r}e^{-v_i}, v_i>0; v-\omega\to\infty\Big)$$

$2\pi i e^{i\omega}$ is the *analytic continuation* of $f(\omega)$ for all ω. Analytic continuation is achieved if we deform the contour integration in the complex v plane.

Ex.4

$$\frac{\epsilon(\omega)}{\epsilon_0} = 1 - \frac{{\omega_{pe}}^2}{k^2}\chi(\omega)$$

where

$$\begin{aligned}\chi(\omega) &= \int_{-\infty}^{\infty} dv\frac{\partial g/\partial v}{v-\omega/k},\ Im(\omega)>0\\ &= \int_L dv\frac{\partial g(v)/\partial v}{v-\omega/k}, \text{for all complex } \omega, \text{ as long as } L \text{ lies below } \omega\end{aligned}$$

Landau integral: pick a trajectory *under* the singular point in the complex plane to achieve the integration.

FIGURE NEEDED!

Let C_ϵ be a small semi-circle under ω/k. Then

$$\begin{aligned}\frac{\epsilon(\omega)}{\epsilon_0} &= 1 - \frac{{\omega_{pe}}^2}{k^2}\Big[P\int_{-\infty}^{\infty} dv\frac{\partial g(v)\partial v}{v-\omega/k} + \int_{C_\epsilon} dv\frac{\partial g(v)\partial v}{v-\omega/k}\Big]\\ &= 1 - \frac{{\omega_{pe}}^2}{k^2}\Big[P\int_{-\infty}^{\infty} dv\frac{\partial g(v)\partial v}{v-\omega/k} + i\pi\frac{\partial g(v)}{\partial v}|_{v=\frac{\omega}{k}}\Big]\end{aligned}$$

where P denotes the principle value integral. This is the same as Equation 8.31 that will be discussed in Section 8.5.

3.5 Helmholtz's Theorem

Helmholtz's theorem, also known as the fundamental theorem of vector calculus, states that any sufficiently smooth, rapidly decaying vector field in three dimensions can be resolved into the sum of an irrotational (curl-free) vector field and a solenoidal (divergence-free) vector field.

Let $\mathbf{F}$ be a vector field on a bounded domain $V \subseteq \mathbf{R}^3$, which is twice continuously differentiable, and let S be the surface that encloses the domain V. Then $\mathbf{F}$ can be decomposed into a curl-free component and a divergence-free component:

$$\mathbf{F} = \nabla\Phi + \nabla \times \mathbf{A}$$

where

$$\begin{aligned}
\Phi(\mathbf{r}) &= \frac{1}{4\pi}\int_V \frac{\nabla' \cdot \mathbf{F}'(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}dV' - \frac{1}{4\pi}\oint_s \hat{\mathbf{n}}' \cdot \frac{\mathbf{F}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}dS' \\
\mathbf{A} &= \frac{1}{4\pi}\int_V \frac{\nabla' \times \mathbf{F}'(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}dV' - \frac{1}{4\pi}\oint_s \hat{\mathbf{n}}' \times \frac{\mathbf{F}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}dS'
\end{aligned}$$

and ∇' is the nabla operator with respect to $\mathbf{r}'$, not $\mathbf{r}$.

If $V = \mathbf{R}^3$ and is therefore unbounded, and $\mathbf{F}$ vanishes faster than $1/r$ as $r \to \infty$, then the second component of both scalar and vector potential are zero. That is,

$$\begin{aligned}
\Phi(\mathbf{r}) &= \frac{1}{4\pi}\int_V \frac{\nabla' \cdot \mathbf{F}'(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}dV' \\
\mathbf{A} &= \frac{1}{4\pi}\int_V \frac{\nabla' \times \mathbf{F}'(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}dV'
\end{aligned}$$

3.6 Toroidal and Poloidal Decomposition

The earliest use of these terms cited by the Oxford English Dictionary (OED) is by Walter M. Elsasser (1946) in the context of the generation of the Earth's magnetic field by currents in the core, with "toroidal" being parallel to lines of latitude and "poloidal" being in the direction of the magnetic field (i.e. towards the poles).

The OED also records the later usage of these terms in the context of toroidally confined plasmas, as encountered in magnetic confinement fusion. In the plasma context, the toroidal direction is the long way around the torus, the corresponding coordinate being denoted by z in the slab approximation or ζ or ϕ in magnetic coordinates; the poloidal direction is the

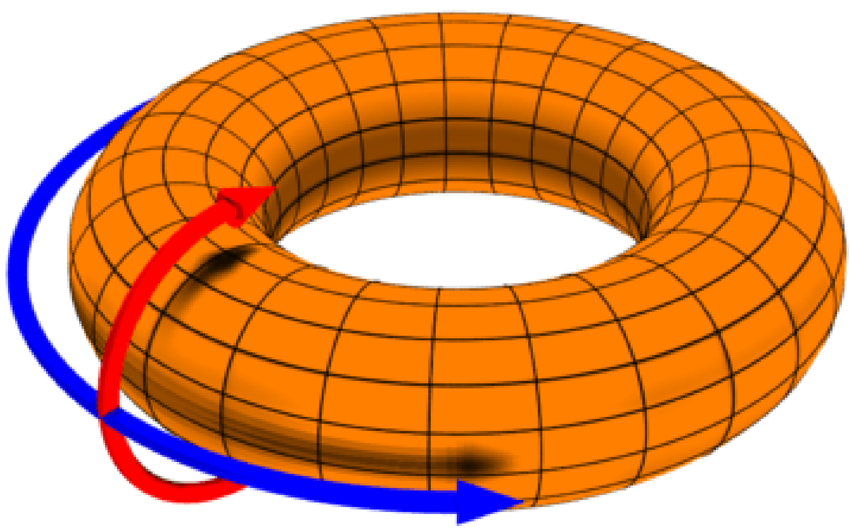

Figure 3.1: Toroidal-Poloidal Decomposition.

short way around the torus, the corresponding coordinate being denoted by y in the slab approximation or θ in magnetic coordinates. (The third direction, normal to the magnetic surfaces, is often called the "radial direction", denoted by x in the slab approximation and variously ψ, χ, r, ρ, or s in magnetic coordinates.)

In vector calculus, a topic in pure and applied mathematics, a poloidal–toroidal decomposition is a restricted form of the Helmholtz decomposition. It is often used in the spherical coordinates analysis of solenoidal vector fields, for example, magnetic fields and incompressible fluids.

For a three-dimensional vector field $\mathbf{F}$ with zero divergence

$$\nabla \cdot \mathbf{F} = 0$$

This $\mathbf{F}$ can be expressed as the sum of a toroidal field $\mathbf{T}$ and poloidal vector field $\mathbf{P}$

$$\mathbf{F} = \mathbf{T} + \mathbf{P}$$

where $\hat{r}$ is a radial vector in spherical coordinates (r, θ, ϕ). The toroidal field is obtained from a scalar field, $\psi(r, \theta, \phi)$ as the following curl,

$$\mathbf{T} = \nabla \times (\hat{r}\Psi(\mathbf{r}))$$

and the poloidal field is derived from another scalar field $\phi(r, \theta, \phi)$ as a twice-iterated curl,

$$\mathbf{P} = \nabla \times \nabla \times (\hat{r}\Phi(\mathbf{r}))$$

This decomposition is symmetric in that the curl of a toroidal field is poloidal, and the curl of a poloidal field is toroidal. The poloidal–toroidal decomposition is unique if it is required that the average of the scalar fields Ψ and Φ vanishes on every sphere of radius r.

3.7 Magnetic Dipole Field

If there's no current,

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} = 0 \Rightarrow B = -\nabla V$$

The divergence free condition for magnetic field then yields a laplace equation

$$\Delta V = 0$$

for the Gauss potential V. The solution of Laplace equation within a certain volume has been first described by Gauss and reads in complex notation

$$V(\mathbf{r},t) = Re\sum_{l=0}^{\infty}\sum_{m=0}^{l}\Big\{R_0(\frac{R_0}{r})^{l+1}G_{lm}^*(t)Y_{lm}^*(\theta,\phi) + R_0(\frac{r}{R_0})^{l}C_{lm}^*(t)Y_{lm}^*(\theta,\phi)\Big\}$$

where R_0 is a constant coefficient, and the complex internal Gauss coefficients

$$G_{lm}^*(t) = G_{lm}(t) - iH_{lm}(t)$$

and external coefficients

$$C_{lm}^*(t) = C_{lm}(t) - iD_{lm}(t)$$

adopting all time variability as well as the complex spherical harmonics

$$Y_{lm}^*(\theta,\phi) = e_{lm}^{im\phi}(\cos\theta)$$

with associated Legendre polynomials $P_{lm}(\cos\theta)$. e.g. $P_{00}(\cos\theta) = 1,\ P_{10} = \cos\theta,\ P_{11} = \sin\theta$.

Ex 1. Planet dipole field

Let the radius of the planet be R (the same as the coefficient). Then the Gauss potential for the dipole reads

$$V = \frac{R^3}{r^2}[G_{10}\cos\theta + G_{11}\cos\phi\sin\theta + H_{11}\sin\phi\sin\theta]$$

Let dipole moment be $\mathbf{m} = (G_{11}, H_{11}, G_{10})$ in Cartesian coordinates, we have

$$\mathbf{B} = -\nabla V = -\nabla\Big[\frac{R^3\mathbf{m}\cdot\mathbf{r}}{r^3}\Big]$$

This is why in literature we often see dipole moments in the unit of [nT]: it is equivalent to the equatorial field strength B_{eq} at the planet's surface. If we specify the dipole moment M in [Tm^3], then $B_{\mathrm{eq}} = M/R_{\mathrm{E}}^3$.

When we expand the gradient operator, we have

$$\mathbf{B} = \frac{\mu_0}{4\pi}\frac{3(\mathbf{m}\cdot\hat{r})\hat{r} - \mathbf{m}}{r^3}$$

In spherical coordinates (r, θ, φ) aligned with the dipole moment,

$$\mathbf{B} = \frac{\mu_0}{4\pi}\frac{m}{r^3}(-2\cos\theta, -\sin\theta,\, 0) \tag{3.4}$$

It is usually convenient to work in terms of the latitude, $\vartheta = \pi/2 - \theta$, rather than the polar angle, θ. An individual magnetic field-line satisfies the equation

$$r = r_{\text{eq}}\cos^2\vartheta$$

where r_{eq} is the radial distance to the field line in the equatorial plane ($\vartheta = 0^\circ$). It is conventional to label field lines using the L-shell parameter, $L = r_{\text{eq}}/R_E$. Here, $R_E = 6.37 \times 10^6\,\text{m}$ is the Earth's radius. Thus, the variation of the magnetic field-strength along a field line characterized by a given L-value is

$$B = \frac{B_E}{L^3}\frac{(1 + 3\sin^2\vartheta)^{1/2}}{\cos^6\vartheta}$$

where $B_E = \mu_0 M_E/(4\pi R_E^3) = 3.11 \times 10^{-5}\,\text{T}$ is the equatorial magnetic field-strength on the Earth's surface.

In Cartesian representation,

$$\mathbf{B} = \frac{\mu_0}{4\pi}\frac{1}{r^5}\begin{pmatrix} (3x^2 - r^2) & 3xy & 3xz \\ 3yx & (3y^2 - r^2) & 3yz \\ 3zx & 3zy & (3z^2 - r^2) \end{pmatrix}\mathbf{m} \tag{3.5}$$

Note that (x, y, z) and r are all normalized to the radius of the planet R.

3.8 Green's Function

The Vlasov theory and the 2-fluid theory only tell us when instability will happen, but neither tell us the severity of instability. To know the magnitude of instability mathematically, we can introduce Green's function

$$G(x,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikx - i\omega(k)t}dk$$

where t is the beam's pulse length and x is the propagation distance. At $t = 0$, we have

$$
\begin{aligned}
G(x,0) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikx}dk = \frac{1}{2i\pi x}e^{ikx}\Big|_{k=-\infty}^{k=\infty} \\
&= \lim_{k\to\infty}\frac{1}{2i\pi x}\big[\cos kx + i\sin kx - \cos kx + i\sin kx\big] \\
&= \lim_{k\to\infty}\frac{1}{\pi}\frac{\sin kx}{x} = \delta(x)
\end{aligned}
$$

where $\delta(x)$ is the δ-function.

The integral $\int_0^\infty \frac{\sin x}{x}\mathrm{d}x$ is called the Dirichlet integral. It is a very famous and important generalized improper integral. Here at least you can see that

$$
\int_{-\infty}^{\infty} G(x,0) = \frac{1}{\pi}\int_0^\infty \frac{\sin kx}{x}\mathrm{d}x = 1 \quad \text{and} \quad G(0,0) = \infty
$$

3.9 Linearization

In linear wave analysis, we need to check how small perturbations evolve in the PDEs. For plasma physics, it is usually very useful to transform into a magnetic field guided coordinates, with the notation $\|$ being parallel to the background magnetic field and $\perp$ being perpendicular to the background magnetic field $\mathbf{B}$. Besides the perturbation of plasma moments (i.e. density, velocity, pressure, etc.), we also need the perturbations to the magnitude of the magnetic field B and the unit vector $\hat{\mathbf{b}}$. Linearzing $B^2 = \mathbf{B}\cdot\mathbf{B}$, we find

$$
\delta B = \hat{\mathbf{b}}\cdot\delta\mathbf{B} = \delta B_\| \tag{3.6}
$$

Linearzing $\hat{\mathbf{b}} = \mathbf{B}/B$, we obtain

$$
\delta\hat{\mathbf{b}} = \delta\Big(\frac{\mathbf{B}}{B}\Big) = \frac{\delta\mathbf{B}B - \delta B\mathbf{B}}{B^2} = \frac{\delta\mathbf{B}}{B} - \frac{\delta\mathbf{B}_\|}{B} = \frac{\delta\mathbf{B}_\perp}{B} \tag{3.7}
$$

The divergence-free of magnetic field gives

$$
\mathbf{k}\cdot\delta\mathbf{B} = k_\|\delta B_\| + k_\perp\delta B_\perp = 0
$$

Thus

$$
\mathbf{k}\cdot\hat{\mathbf{b}} = \mathbf{k}\cdot\frac{\delta\mathbf{B}_\perp}{B} = \frac{k_\perp\delta B_\perp}{B} = -k_\|\frac{\delta B_\|}{B} = -k_\|\frac{\delta B}{B} \tag{3.8}
$$

These seemingly trivial relations have profound implications in physics. Equation 3.6 tells us that the perturbation of magnetic field magnitude has only contribution from the parallel component, which is why in satellite observations people only look at parallel perturbations for compressional wave modes. Equation 3.7 tells us that the perturbation in the unit vector is only related to the perpendicular fluctuations.

3.10 Wave Equations

The waves in plasma physics is governed by second order ODEs. Here we list some second order ODEs that has been studied mostly in plasma physics.

- Schrödinger Equation:

$$\frac{\mathrm{d}^2\varphi}{\mathrm{d}x^2}+\frac{2m}{\hbar^2}[E-U(x)]\varphi=0$$

Schrödinger Equation appears in the nonlinear wave studies (Chapter 13).

- Shear Alfvén wave:

$$\frac{\mathrm{d}}{\mathrm{d}x}\Big\{\rho_0[\omega^2-(\mathbf{k}\cdot\mathbf{v}_A(x))^2]\frac{dE}{\mathrm{d}x}\Big\}-k^2\rho_0[\omega^2-(\mathbf{k}\cdot\mathbf{v}_A)^2-g\frac{1}{\rho_0}\frac{\mathrm{d}\rho_0}{\mathrm{d}x}]E=0$$

The Shear Alfvén wave equation appears in nonuniform ideal MHD (Equation 7.24, Equation 12.8).

- EM waves in non-magnetized plasma, O mode:

$$\frac{\mathrm{d}^2E}{\mathrm{d}x^2}+\frac{\omega^2}{c^2}\Big[1-\frac{\omega_{pe}(x)^2}{\omega^2}\Big]E=0 \tag{3.9}$$

- Electron cyclotron resonance heating (ECRH):

$$\frac{\mathrm{d}^2E}{\mathrm{d}x^2}+\frac{\omega^2}{c^2}\Big[1-\frac{\omega_{pe}(x)^2}{\omega(\omega-\Omega_e(x))}\Big]E=0$$

In general, a second order ODE

$$\frac{\mathrm{d}^2u(x)}{\mathrm{d}x^2}+a_1(x)\frac{du(x)}{\mathrm{d}x}+a_2(x)u(x)=0$$

can be rewritten to get rid of the first derivative. Let

$$u(x) = E(x)e^{-\frac{1}{2}\int^x a_1(x)\mathrm{d}x}$$

we have

$$\frac{\mathrm{d}^2 E(x)}{\mathrm{d}x^2} + k^2(x)E(x) = 0 \tag{3.10}$$

where

$$k^2(x) = a_2(x) - \frac{{a_1}^2}{4} - \frac{1}{2}\frac{\mathrm{d}a_1(x)}{\mathrm{d}x}$$

Note that the lower bound of the integral is left on purpose to account for a constant. We will concentrate at special points, i.e. zeros (cutoff) and poles (resonance) of $k^2(x) \equiv \frac{\omega^2}{c^2}n^2(x)$.

First, we will introduce *Wentzel–Kramers–Brillouin-Jeffreys* (WKBJ) solution to Equation 3.10:

$$E(x) \sim \frac{1}{\sqrt{k(x)}}e^{\pm i\int^x k(x')\mathrm{d}x'}.$$

Proof.

For simplicity, here we use a simple harmonic oscillation analogy. Consider

$$\frac{\mathrm{d}^2 x(t)}{\mathrm{d}t^2} + \Omega^2(t)x(t) = 0$$

Assume $\Omega \gg 1$, and T is the time scale over which Ω changes appreciably. We would anticipate solutions like

$$\begin{aligned}
x(t) &\sim e^{\pm i\Phi(t)} \\
\dot{x}(t) &\sim \pm i\dot{\Phi}(t)x(t) \\
\ddot{x}(t) &\sim -\dot{\Phi}^2(t)x(t)\cancel{\pm i\ddot{\Phi}(t)x(t)} \\
\Rightarrow \dot{\Phi}(t) &= \Omega(t),\ \text{ or } \Phi(t) = \int^t \Omega(t')\mathrm{d}t' + \text{const.}
\end{aligned}$$

Write $x(t) = A(t)\sin[\Phi(t)]$, where $A(t)$ is slowly varying in time, $\dot{A}(t) \ll \Omega A$, which is almost a periodic motion. From adiabatic theory in classical mechanics, $\oint p dq \simeq \text{const.}$, we have

$$\begin{aligned}
\oint m v_x \mathrm{d}x &\simeq \text{const.} \\
\oint m\dot{x}\dot{x}\mathrm{d}t &\simeq \text{const.}
\end{aligned}$$

Then in a period $2\pi/\Omega$,

$$\frac{2\pi}{\Omega}\oint m\dot{x}^2 dt = \frac{2\pi}{\Omega}\oint mA^2\Omega^2\cos^2\Phi \mathrm{d}t \simeq \text{const.}$$

which leads to

$$mA^2\Omega \simeq \text{const.},\ A\sim\frac{1}{\sqrt{\Omega}}$$

Thus the general form of solution can be written as

$$x(t)\sim\frac{1}{\sqrt{\Omega}}e^{\pm i\Phi(t)}\sim\frac{1}{\sqrt{\Omega}}e^{\pm i\int^t \Omega(t')\mathrm{d}t'+\text{const.}}$$

Note:

1. There is no lower limit to the integral, because it is like adding a constant.
2. This solution is valid if $k\cdot L > O(3)$, where L is the length scale over which $k^2(x)$ changes appreciably. (???)Formally the condition should be $kL \gg 1$. Apparently near resonance ($k\rightarrow\infty$), the condition breaks down, then how do we reconcile the solution? There are more discussions coming up and applications in Chapter 7 and Chapter 12.

3.10.1 Airy Function

We want to develop a general method for cut-off and resonance. Away from the turning point x_t,

$$E_{\text{WKBJ}}(x)\sim\frac{1}{\sqrt{k(x)}}e^{\pm i\int^x k(x)\mathrm{d}x}$$

Near $x=x_t$, we can use a linear approximation for $k^2(x)$ (first term in the Taylor expansion),

$$k^2(x)\approx -{k_0}^2 a(x-x_t)$$

Then

$$\begin{aligned}\frac{\mathrm{d}^2E}{\mathrm{d}x^2}+k^2(x)E=0\\ \frac{\mathrm{d}^2E}{\mathrm{d}x^2}-{k_0}^2a(x-x_t)E=0\end{aligned}$$

Let $\frac{x-x_t}{l}=\zeta$, we have

$$\frac{\mathrm{d}^2 E}{\mathrm{d}\zeta^2} - l^2 {k_0}^2 a l \zeta E(\zeta) = 0$$

If we choose l s.t. $l^3 {k_0}^2 a = 1$ (non-dimensional treatment), we obtain

$$\frac{\mathrm{d}^2 E}{\mathrm{d}\zeta^2} - \zeta E(\zeta) = 0 \tag{3.11}$$

which is equivalent to the case where $k^2(\zeta) = -\zeta$ that shows the linear approximation. Equation 3.11, known as the *Airy equation* or *Stokes equation*, is the simplest second-order linear differential equation with a turning point (a point where the character of the solutions changes from oscillatory to exponential). From the WKBJ theory, we get the solution

$$E_{\mathrm{WKBJ}} \sim \frac{1}{\sqrt{k(x)}} e^{\pm i \int^x k(x')\mathrm{d}x'} = \begin{cases} \zeta^{-1/4} e^{\mp \frac{2}{3}\zeta^{3/2}} & \text{if } \zeta > 0 \\ (-\zeta)^{-1/4} e^{\pm i \frac{2}{3}(-\zeta)^{3/2}} & \text{if } \zeta < 0 \end{cases}$$

Note that the solution blows up at $\zeta = 0$ (turning point) miserably. For $\zeta > 0$, one solution is exponentially decay while the other is exponentially growing; for $\zeta < 0$, the two solutions are oscillatory. Solutions can also be found as series in ascending powers of ζ by the standard method. Assume that a solution is $E = a_0 + a_1\zeta + a_2\zeta^2 + ...$ Substitute this in Stokes Equation and equate powers of ζ will give relations between the constants a_0, a_1, a_2,etc., and lead finally to

$$\begin{aligned} E = & a_0 \Big\{ 1 + \frac{\zeta^3}{3\cdot 2} + \frac{\zeta^6}{6\cdot 5\cdot 3\cdot 2} + \frac{\zeta^9}{9\cdot 8\cdot 6\cdot 5\cdot 3\cdot 2} + ... \Big\} \\ & + a_1 \Big\{ \zeta + \frac{\zeta^4}{4\cdot 3} + \frac{\zeta^7}{7\cdot 6\cdot 4\cdot 3} + \frac{\zeta^{10}}{10\cdot 9\cdot 7\cdot 6\cdot 4\cdot 3} + ... \Big\} \end{aligned}$$

which contains the two arbitrary constants a_0 and a_1, and is therefore the most general solution. The series are convergent for all ζ, which confirms that every solution of Stokes Equation is finite, continuous and single valued.

This form is usually not easy to interpret in physical sense. Besides this, we can find a more useful solution to Equation 3.11 using the integral representation. An equivalent but maybe more intuitive approach is to solve Equation 3.11 with Fourier transform; a coefficient 2π naturally appears. Both approaches reach the same results.

Write

$$E(\zeta) = \int_a^b \mathrm{d}t e^{t\zeta} f(t)$$

where the integral represents a path in the complex t plane from a to b. Then

$$\begin{aligned}\frac{\mathrm{d}^2E}{\mathrm{d}\zeta^2} &= \int_a^b \mathrm{d}t\ t^2 e^{t\zeta} f(t) \\ \zeta E(\zeta) &= \int_a^b \mathrm{d}t\zeta e^{t\zeta} f(t) = \int_a^b \mathrm{d}t \frac{\mathrm{d}}{\mathrm{d}t}(e^{t\zeta}) f(t) \\ &= e^{t\zeta} f(t)\Big|_a^b - \int_a^b dt e^{t\zeta} f'(t)\end{aligned}$$

The limits a, b are be chosen so that the first term vanishes at both limits. Then Equation 3.11 is satisfied if

$$\begin{cases} e^{t\zeta} f(t)\Big|_a^b = 0 \\ t^2 f(t) = -\frac{\mathrm{d}f(t)}{\mathrm{d}t} \Rightarrow f(t) = Ae^{-\frac{1}{3}t^3} \end{cases}$$

where A is a constant. The solution is now written as

$$E(\zeta) = \int_a^b \mathrm{d}t A e^{t\zeta - \frac{1}{3}t^3}$$

The limits a and b must therefore be chosen so that $e^{t\zeta - \frac{1}{3}t^3}$ is zero for both. (Note that ζ is a constant.) This is only possible if $t = \infty$ and the real part of t^3 is positive:

$$\begin{aligned}\Re(t^3) > 0 &\Leftrightarrow 2\pi n - \frac{1}{2}\pi < \arg t^3 < 2\pi n + \frac{1}{2}\pi \\ &\Leftrightarrow 2\pi n - \frac{1}{2}\pi < 3\arg t < 2\pi n + \frac{1}{2}\pi\end{aligned}$$

where n is an integer. Figure 3.2 is a diagram of the complex t-plane, and a and b must each be at ∞ in one of the shaded sectors. They cannot both be in the same sector, for then the integral would be zero. Hence the contour may be chosen in three ways, as shown by the three curves C_1, C_2, C_3, and the solution would be

$$E(\zeta) = \int_{C_1, C_2, C_3} \mathrm{d}t\ e^{t\zeta - \frac{1}{3}t^3}$$

This might appear at first to give three independent solutions, but the contour C_1 and be distorted so as to coincide with the two contours $C_2 + C_3$, so that

$$\int_{C_1} = \int_{C_2} + \int_{C_3}$$

and therefore there are only two independent solutions.

Proof.

Note that:

1. $e^A \neq 0 \ \forall |A| < \infty$;
2. $e^A = 0 \Leftrightarrow A \to \infty \,\text{and}\, \Re(A) < 0$.

Therefore we have

$$e^{-\frac{1}{3}t^3} \to 0 \,\text{as}\, t \to \infty \,\text{and}\, \Re\Big(\frac{1}{3}t^3\Big) > 0$$

In polar coordinates, let $t = |t|e^{i\theta}$. Then

$$t^3 = |t|^3 e^{i3\theta} = |t|^3\big(\cos 3\theta + i \sin 3\theta\big)$$

$$\Re\Big(\frac{1}{3}t^3\Big) > 0 \iff \cos 3\theta > 0 \iff 3\theta \in (-\frac{\pi}{2}, \frac{\pi}{2}),\ (\frac{3\pi}{2}, \frac{5\pi}{2}),\ (-\frac{3\pi}{2}, -\frac{5}{2}\pi)$$

Jeffreys (1956) defines two special Airy functions $Ai(x), Bi(x)$ as follows

$$\begin{aligned} Ai(\zeta) &= \frac{1}{2\pi i}\int_{C_1} \mathrm{d}t e^{-\frac{1}{3}t^3 + \zeta t} \\ Bi(\zeta) &= \frac{1}{2\pi}\Big[\int_{C_2} + \int_{C_3}\Big]\mathrm{d}t e^{-\frac{1}{3}t^3 + \zeta t} \end{aligned} \tag{3.12}$$

Obviously he took the Fourier transform such that a coefficient 2π naturally appears. In Equation 3.12, the contour C_1 can be distorted so as to coincide with the imaginary t-axis for almost its whole length. It must be displaced slightly to the left of this axis at its ends to remain in the shaded region at infinity. Let $t = is$. Then the Airy function of the first kind in Equation 3.12 becomes

$$\begin{aligned} Ai(\zeta) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i(\zeta s + \frac{1}{3}s^3)} ds \\ &= \frac{1}{\pi}\int_0^{\infty} \cos\Big(\zeta s + \frac{1}{3}s^3\Big) ds \end{aligned}$$

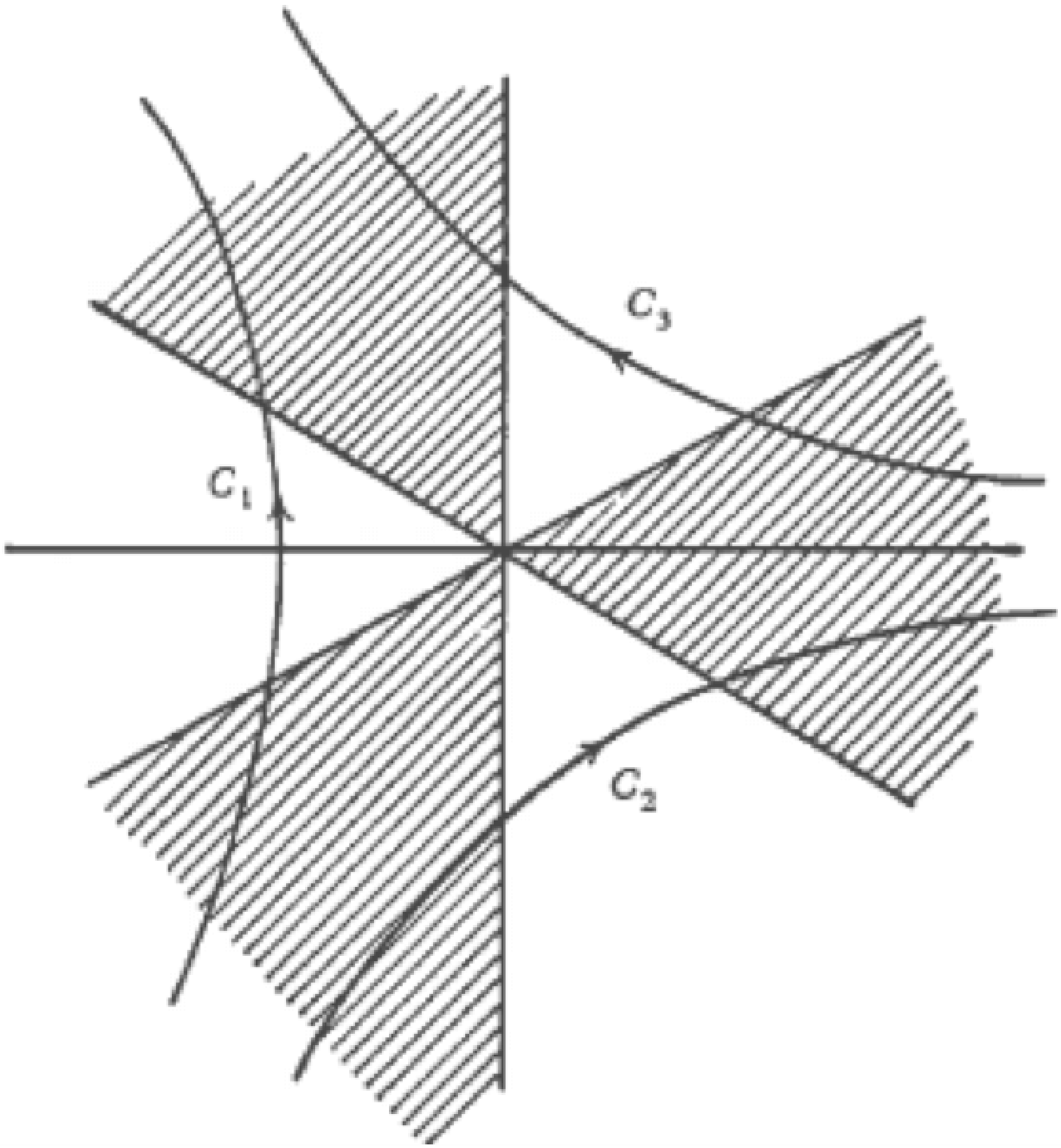

Figure 3.2: Solution to Stokes Equation in complex t-plane.

It is known as the Airy integral, which is the solution for $y \to 0$ as $x \to \infty$. The other linearly independent solution, the Airy function of the second kind, denoted $Bi(x)$, is defined as the solution with the same amplitude of oscillation as $Ai(x)$ as $x \to -\infty$ which differs in phase by $\pi/2$:

$$Bi(\zeta) = \frac{1}{\pi} \int_0^\infty \left[e^{-\frac{s^3}{3} + s\zeta} + \sin\left(\zeta s + \frac{1}{3}s^3\right) \right] ds$$

$Ai(x)$ and $Bi(x)$ are shown in **?@fig-airy**.

```
KeyNotes.plot_airy()
```

As an interesting experiment, we can check if $Ai(x)$ is recovered from solving the second order ODE numerically:

```
@sco KeyNotes.plot_airy_ode()
```

Here we start from the right boundary and move towards the left.

Now, to find approximations for Airy functions, we use the method of *steepest descent*. This approximation is based on the assumption that major contribution to the integral is from near the saddle point. As an example of saddle point, consider a complex function $\exp f(z) = e^{-z^2}$, where $z = x + iy$.

- If $z = \text{real} = x$, $e^{-z^2} = e^{-x^2}$;
- If $z = \text{imag} = iy$, $e^{-z^2} = e^{y^2}$.

The procedure goes as follows:

1. Detour path of integral s.t. it passes through the saddle point of the integral, along the direction of steepest descent.
2. Obtain major contribution by integrating the Gaussian function.

$$I = \int_C \mathrm{d}t e^{f(t)}, \quad f(t) = f(t_s) + (t - t_s)f'(t_s) + \frac{1}{2}(t - t_s)^2 f''(t_s) + \ldots$$

where $f'(t_s) = 0$ at the saddle point $t = t_s$ simplifies I to the integral of a Gaussian function.

For $Ai(\zeta)$,

$$\begin{aligned} f(t) &= t\zeta - \frac{1}{3}t^3 \\ f'(t_s) &= \zeta - {t_s}^2 = 0 \\ f''(t_s) &= -2t_s \end{aligned}$$

Consider $\zeta > 0$ (where the solution is either exponentially decaying or growing.) Then

$$t_{s1} = -\sqrt{\zeta}, \quad t_{s2} = \sqrt{\zeta}$$
$$f(t_{s1}) = -\frac{2}{3}\zeta^{3/2}, \quad f''(t_{s1}) = 2\zeta^{1/2}$$

So

$$\begin{aligned} Ai(\zeta) &= \frac{1}{2\pi i}\int_{C_1} \mathrm{d}t\; e^{-\frac{2}{3}\zeta^{3/2}+\zeta^{1/2}(t-t_s)^2} + ... \\ &\approx \frac{1}{2\pi i} e^{-\frac{2}{3}\zeta^{3/2}} \int_{C_1} \mathrm{d}t\; e^{\sqrt{\zeta}(t-t_s)^2} \end{aligned}$$

Let $e^{\sqrt{\zeta}(t-t_s)^2} = e^{-\rho^2}$, where $\rho =$ real is the direction of steepest descent, and $\rho =$ imag is the direction of steepest ascent.

$$\begin{aligned} &i\rho = \pm\zeta^{1/4}(t-t_s) \\ &\mathrm{d}t = \frac{i\mathrm{d}\rho}{\zeta^{1/4}} \Rightarrow \ \mathrm{d}t \text{ is purely imaginary along steepest descent.} \end{aligned}$$

Then for $\zeta > 0$

$$\begin{aligned} Ai(\zeta) &= \frac{1}{2\pi i} e^{-\frac{2}{3}\zeta^{3/2}} \int_{-\infty}^{\infty} \frac{id\rho}{\zeta^{1/4}} e^{-\rho^2} \\ &= \frac{1}{2\sqrt{\pi}\zeta^{1/4}} e^{-\frac{2}{3}\zeta^{3/2}} \text{ as } \zeta \to \infty \end{aligned}$$

The ratio of accuracy is shown in the following table (LABEL???). In practice, it is already a very good approximation when $\zeta > 3$. (think of $kL \gg 3$ for WKBJ solution! ???)

ζ	1	2	3	6	∞
ratio	0.934	0.972	0.983	0.993	1

Now, consider $\zeta < 0$. When $\zeta \to -\infty$, we anticipate oscillating behavior of $Ai(\zeta)$. $\zeta = -|\zeta|$,

$$
\begin{cases}
f(t) & = \zeta t - \frac{1}{3}t^3 = -t|\zeta| - \frac{1}{3}t^3 \\
f'(t) & = -|\zeta| - t^2 = 0 \\
f''(t) & = -2t
\end{cases}
$$

$$
\begin{aligned}
\Rightarrow f(t_{s1}) &= i\frac{2}{3}|\zeta|^{3/2} \\
f''(t_{s1}) &= 2i\sqrt{|\zeta|}
\end{aligned}
$$

so the contribution from t_{s1} is

$$
\frac{1}{2\pi i} e^{i\frac{2}{3}|\zeta|^{3/2}} \int_{t_{s1}} e^{\frac{1}{2}(2i\sqrt{|\zeta|})(t-t_s)^2+\dots}
$$

Let $-\rho^2 = i\sqrt{|\zeta|}(t-t_{s1})^2$, differentiate on both sides to get

$$
\mathrm{d}t = \pm\frac{e^{i\pi/4}}{|\zeta|^{1/4}}\mathrm{d}\rho
$$

Again, $\rho =$ real is the direction of steepest descent at $t = t_{s1}$. Do the same to t_{s2}, then by summing them up we have for $\zeta < 0$,

$$
\begin{aligned}
Ai(\zeta) &\approx \frac{1}{2\pi i}\Big[e^{i\frac{2}{3}|\zeta|^{3/2}} \int_{-\infty}^{\infty} \frac{e^{i\pi/4}}{|\zeta|^{1/4}}\mathrm{d}\rho e^{-\rho^2} - e^{-i\frac{2}{3}|\zeta|^{3/2}} \int_{-\infty}^{\infty} \frac{e^{-i\pi/4}}{|\zeta|^{1/4}}\mathrm{d}\rho e^{-\rho^2}\Big] \\
&= \frac{1}{2\pi i}\Big[\frac{e^{i\frac{2}{3}|\zeta|^{3/2}+i\pi/4}}{|\zeta|^{1/4}}\sqrt{\pi} - \frac{e^{-i\frac{2}{3}|\zeta|^{3/2}-i\pi/4}}{|\zeta|^{1/4}}\sqrt{\pi}\Big]
\end{aligned}
\tag{3.13}
$$

An interesting application of the gradient descent method is to find Stirling's approximation. Mathematically, we can proof that

$$
n! = \int_0^\infty \mathrm{d}t\ e^{-t}t^n = \int_0^\infty \mathrm{d}t\ e^{-t+n\ln t} \equiv \int_0^\infty \mathrm{d}t\ e^{f(t)}
$$

and by following the steepest descent method,

$$
\begin{aligned}
f(t) &= -t + n\ln t \\
f'(t) &= -1 + n\frac{1}{t} = 0 \\
f''(t) &= -n\frac{1}{t^2}
\end{aligned}
$$

we can find the approximation Stirling formula

$$
\begin{aligned}
n! &\approx \int_0^{\infty} \mathrm{d}t\; e^{-n+n\ln n-\frac{1}{2n}(t-t_s)^2} \\
&= e^{-n}n^n \int_0^{\infty} \mathrm{d}t\; e^{-\frac{1}{2n}(t-t_s)^2} \\
&= \sqrt{2n\pi}e^{-n}n^n
\end{aligned}
$$

The following table (LABEL???) show the goodness of approximation of Stirling formula. In practice, it is already a very good approximation when $n > 3$.

ζ	1	2	3	10	∞
ratio	0.922	0.9595	0.973	0.9917	1

See (Budden 1961) *Chapter 15: The Airy Integral Function, And The Stokes Phenomenon* for more details.

3.10.2 Uniformly Valid WKBJ Solution Across the Turning Point

In this section, we present the WKBJ solution that is uniformly valid everywhere, even at the turning point.

Consider the standard equation,

$$
\frac{\mathrm{d}^2E(x)}{\mathrm{d}x^2} + k^2(x)E(x) = 0
$$

Let $x = 0$ be the turning point, i.e. we assume that $k^2(x)$ is a monotonically decreasing function of x with $k(0) = 0$. For the region $x > 0$, we first identify the exponentially decaying factor of the Airy function, $Ai(\zeta)$, to be the phase integral in the WKBJ solution:

$$
-\frac{2}{3}\zeta^{3/2} = i\int_0^x k(x')\mathrm{d}x'
$$

Note that this yields $\zeta = \zeta(x)$, *a known function of* x (I confuse myself later about ζ and x...). The branch for $\zeta(x)$ is to be chosen so that for $x > 0$, ζ is real and positive.

We next verify that the uniformly valid solution to the standard equation is simply

$$
E(x) = \frac{1}{\sqrt{\zeta'(x)}}Ai(\zeta) \tag{3.14}
$$

where the prime denotes a derivative. For large values of ζ, we can use the asymptotic formula of Ai Equation 3.13, and notice that

$$-\zeta^{1/2}\zeta' = ik(x)$$
$$\zeta'(x) = -i\frac{k(x)}{\zeta^{1/2}}$$

we can see that Equation 3.14 reduces to the standard WKBJ solutions for large values of ζ

$$E(x) = \frac{1}{\sqrt{\zeta'(x)}}Ai(\zeta) \sim \frac{1}{\sqrt{k(x)}}e^{-\frac{2}{3}\zeta^{3/2}} = \frac{1}{\sqrt{k(x)}}e^{i\int_0^x k(x')\mathrm{d}x'}$$

We can also show that Equation 3.14 is valid for small values of ζ, i.e. near the turning point at $x = 0$. (Hint: Near $x = 0$, $k^2(x)$ may be approximated by a linear function of x. This linear approximation then yields $\zeta(x)$ as a linear function of x according to Equation 3.14.)

Ex. Choose a smooth plasma density profile which monotonically increases with x s.t.

$$\frac{\omega_{pe}^2(x)}{\omega^2} = 1 + \tanh x$$

and launch a wave of frequency ω from $x = -\infty$, the vacuum region, toward the positive x-direction with $\frac{\omega^2}{c^2} = 10\mathrm{m}^{-2}$. (This is like launching a wave from low B region into high B region.) Numerically integrate the wave equation Equation 3.9,

$$\frac{\mathrm{d}^2E}{\mathrm{d}x^2} + \frac{\omega^2}{c^2}\Big[1 - \frac{\omega_{pe}(x)^2}{\omega^2}\Big]E = 0$$

from some large positive values of x, we get the results in Figure 3.3.

We know that away from x_t, WKBJ solution works. To the left of x_t (with $\zeta < -3$), Equation 3.13 gives

$$Ai(\zeta) = \frac{1}{2i\sqrt{\pi}|\zeta|^{1/4}}\Big[e^{i\frac{2}{3}|\zeta|^{3/2}+i\frac{\pi}{4}} - e^{-\frac{2}{3}|\zeta|^{3/2}-i\frac{\pi}{4}}\Big]$$

Choose the branch s.t. $\zeta x > 0$ if $x > x_t$; $\zeta(x) < 0$ if $x < x_t$.

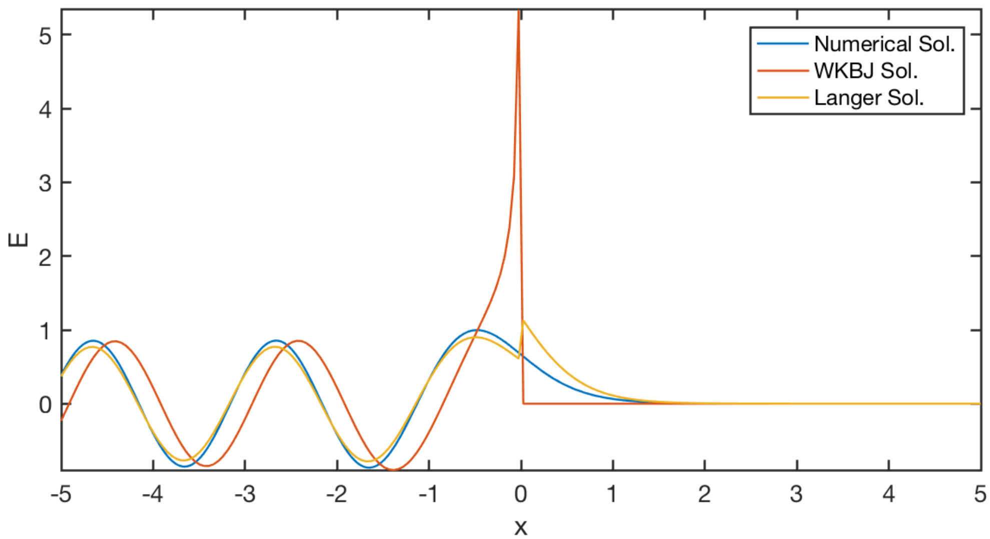

Figure 3.3: Comparison between the WKBJ solution, uniformly valid solution and numerical integral solution for the O-mode with monotonically increasing density with x.

$$\begin{aligned}
E(x) &= \frac{1}{\sqrt{\zeta'(x)}} Ai(\zeta) \\
&= \frac{C_3}{\sqrt{k(x)/|\zeta|^{1/2}}} \frac{1}{|\zeta|^{1/4}} \Big\{ e^{i\frac{2}{3}|\zeta|^{3/2} + i\frac{\pi}{4}} - c.c \Big\} \\
&= \frac{C_3}{\sqrt{k(x)}} \Big\{ e^{i(-\int_{x_t}^{x} k(x')\mathrm{d}x') + i\frac{\pi}{4}} - c.c \Big\} \\
&= \frac{C_3}{\sqrt{k(x)}} \Big\{ e^{-i\int_{x_t}^{a} k(x')\mathrm{d}x' - i\int_{a}^{x} k(x')\mathrm{d}x' + i\frac{\pi}{4}} - c.c \Big\} \\
&= \frac{C_4}{\sqrt{k(x)}} \Big\{ e^{-i\int_{a}^{x} k(x')\mathrm{d}x'} - e^{i\int_{x_t}^{a} k(x')\mathrm{d}x'} e^{2i\int_{x_t}^{a} k(x')\mathrm{d}x' - i\frac{\pi}{2}} \Big\} \\
&= \frac{C_4}{\sqrt{k(x)}} \Big\{ e^{-i\int_{a}^{x} k\mathrm{d}x'} + R \cdot e^{i\int_{a}^{x} k\mathrm{d}x'} \Big\},
\end{aligned}$$

where

$$C_4 = C_3 \cdot e^{-i\int_{x_t}^{a} k\mathrm{d}x' + i\frac{\pi}{4}}$$

and

$$R = -e^{2i\int_{x_t}^{a} k\mathrm{d}x' - i\frac{\pi}{2}} = ie^{2i\int_{x_t}^{a} k\mathrm{d}x'} = ie^{-2i\int_{a}^{x_t} k\mathrm{d}x'}$$

is the reflection coefficient at $x = a$.

3.10.3 Stokes Phenomenon

In complex analysis the Stokes phenomenon is that the asymptotic behavior of functions can differ in different regions of the complex plane, and that these differences can be described in a quantitative way.

Ex. For the simple wave equation

$$\frac{\mathrm{d}^2 E}{dz^2} - E = 0$$

the solution can be given in various forms

$$E = e^z, e^{-z}, \cosh z, \sinh z$$
$$E = \cosh z = \frac{1}{2}(e^z + e^{-z}) \Rightarrow \begin{cases} E \sim \frac{1}{2}e^z \; z \to \infty \\ E \sim \frac{1}{2}e^{-z} \; z \to -\infty \end{cases}$$
$$E = \sinh z = \frac{1}{2}(e^z - e^{-z}) \Rightarrow \begin{cases} E \sim \frac{1}{2}e^z \; z \to \infty \\ E \sim \frac{-1}{2}e^{-z} \; z \to -\infty \end{cases}$$

Note that if a solution is exponentially growing in one direction, its asymptotic solution can contain an arbitrary amount of the exponentially decaying solution; that is, specifying an asymptotic growing solution in one direction cannot completely specify the solution in the entire complex plane.

The two linearly independent approximate solutions to Airy Equation 3.11 are the Airy function approximations from WKBJ method:

$$\frac{\mathrm{d}^2 E}{\mathrm{d}\zeta^2} - \zeta E = 0$$
$$\Rightarrow E(\zeta) = \begin{cases} Ai(\zeta) \sim \frac{1}{2\sqrt{\pi}\zeta^{1/4}} e^{-\frac{2}{3}\zeta^{3/2}}, & \zeta > 0 \\ Bi(\zeta) \sim \frac{1}{2\sqrt{\pi}\zeta^{1/4}} e^{\frac{2}{3}\zeta^{3/2}}, & \zeta > 0 \end{cases}$$

which is very accurate for $\zeta > 3$ (see previous section).

Stokes found that you can add an arbitrary amount of $Ai(\zeta)$ to $Bi(\zeta)$ without changing the behaviour of solution. ($Ai(\zeta) < O(\zeta^{-1})$???)

We want to find ζ s.t. E_{WKBJ} is purely growing/decaying exponentially:

$$\zeta^{3/2}\text{is purely real} \iff (|\zeta|e^{i\theta})^{3/2}\text{is purely real} \iff \sin\frac{3}{2}\theta = 0,\ \theta = 0, \pm\frac{2}{3}\pi, \pm\frac{4}{3}\pi$$

The lines in the complex plane on which WKBJ solution is purely growing/decaying exponentially are called *Stokes lines* (Figure 3.4). It is accompanied with *anti-Stokes lines* (in the opposite direction to Stokes lines), on which WKBJ solution is purely oscillatory. The exponentially growing solution on Stokes lines is called the *dominant solution*; the decaying solution on Stokes lines is called the *sub-dominant/recessive solution*. The sub-dominant solution will always becomes dominant in a neighboring Stokes line. However, the inverse is not true. (It may contain an amount of sub-dominant solution.) Each term changes from dominant to subdominant, or the reverse, when ζ crosses an anti-Stokes line???

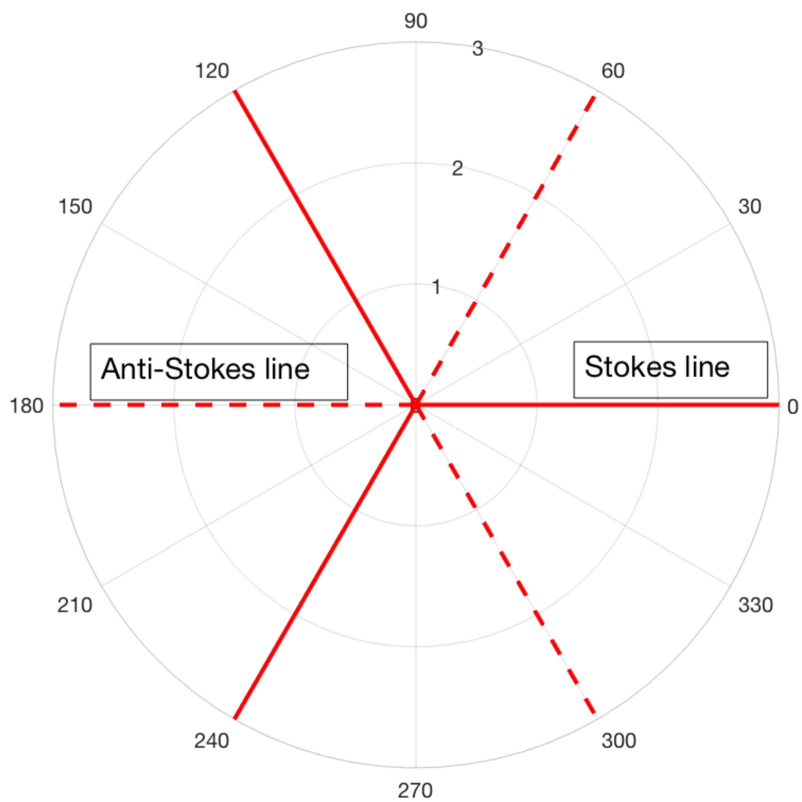

Figure 3.4: Stokes lines and anti-Stokes lines for WKBJ solution of Airy Equation.

Even though we say “arbitrary”, the analytic solution in the whole complex plane possess a limit on that amount. The next question would be: how much exponentially decaying solution can you add to the exponentially growing solution? (Note that the asympotic series is also a divergent series: more terms don’t lead to high resolution accuracy! I have questions on this part???)

For the asymptotic approximation solution of Airy Equation 3.11, we need to define $\arg(\zeta)$

properly to make it single value. Let us choose the branch cut $-\pi \leq \arg \zeta < \pi$. The complex plane is demonstrated in Figure 3.5.

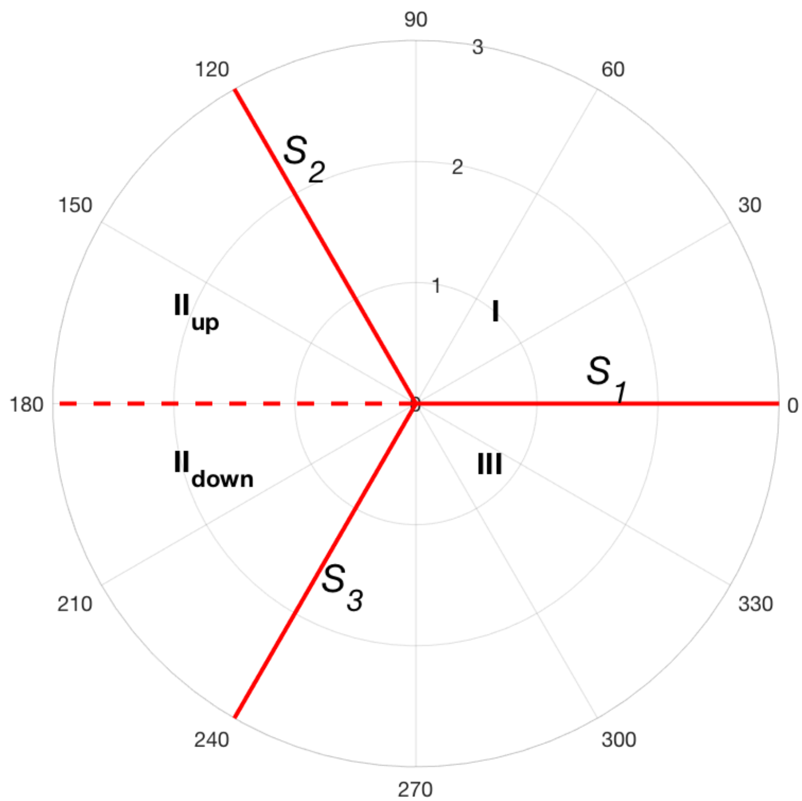

Figure 3.5: Marked Stokes lines and regions for Airy solution.

Let us start from region I. The solution in region I is

$$E_I \sim \zeta^{-1/4}[A_1 e^{-\frac{2}{3}\zeta^{3/2}} + B_1 e^{\frac{2}{3}\zeta^{3/2}}]$$

On S_1, $e^{-\frac{2}{3}\zeta^{3/2}}$ is sub-dominant, $e^{\frac{2}{3}\zeta^{3/2}}$ is dominant. (The former would be dominant on neighboring Stokes lines S_2 and S_3.) Crossing S_2 into region II_{up}, we have

$$E_{\text{II}_{\text{up}}} \sim \zeta^{-1/4}\Big[\underbrace{A_1 e^{-\frac{2}{3}\zeta^{3/2}}}_{\text{dominant}} + \underbrace{(\lambda_1 A_1 + B_1)e^{\frac{2}{3}\zeta^{3/2}}}_{\text{sub-dominant}}\Big]$$

where λ_2 is the Stokes constant on S_2. The constant in the subdominant term has changed by $\lambda_2 A_1$ because the differential equation is linear, and it cannot depend on B_1. Otherwise it would be unaltered if we added to the solution in region I any multiple of the solution in which $A_1 = 0$. (See (Budden 1961) 15.13.) Crossing the branch cut to region $\mathrm{II}_{\mathrm{down}}$, $\zeta^{\mathrm{up}} = \zeta^{\mathrm{down}} e^{i2\pi}$.

$$\begin{aligned} \zeta_{\mathrm{up}}^{-1/4} &= \zeta_{\mathrm{down}}^{-1/4} e^{i2\pi(-1/4)} = i\zeta_{\mathrm{down}}^{1/4} \\ \zeta_{\mathrm{up}}^{3/2} &= \zeta_{\mathrm{down}}^{3/2} e^{i2\pi(3/2)} = -\zeta_{\mathrm{down}}^{3/2} \end{aligned}$$

so

$$E_{\mathrm{II}_{\mathrm{down}}} = -i\zeta_{\mathrm{II}_{\mathrm{down}}}^{1/4} \left[A_1 e^{\frac{2}{3}\zeta_{\mathrm{II}_{\mathrm{down}}}^{3/2}} + (\lambda_2 A_1 + B_1) e^{-\frac{2}{3}\zeta_{\mathrm{II}_{\mathrm{down}}}^{3/2}} \right]$$

We can also go clockwise from I to III, crossing S_1,

$$E_{III} \sim \zeta^{-1/4} \Big[\underbrace{B_1 e^{\frac{2}{3}\zeta^{3/2}}}_{\text{dominant}} + \underbrace{(-\lambda_1 B_1 + A_1) e^{-\frac{2}{3}\zeta^{3/2}}}_{\text{sub-dominant}} \Big]$$

where λ_1 is the Stokes constant on S_1, and the minus sign indicates the clockwise direction.

Crossing S_3 from III to $\mathrm{II}_{\mathrm{down}}$,

$$E_{\mathrm{II}_{\mathrm{down}}} \sim \zeta^{-1/4} \Big[\underbrace{(-\lambda_1 B_1 + A_1) e^{-\frac{2}{3}\zeta^{3/2}}}_{\text{dominant}} + \underbrace{[(-\lambda_3)(-\lambda_1 B_1 + A_1) + B_1] e^{\frac{2}{3}\zeta^{3/2}}}_{\text{sub-dominant}} \Big]$$

where λ_3 is the Stokes constant on S_3, and the minus sign indicates the clockwise direction.

Since the solution to Airy Equation is analytic, the solutions in region $\mathrm{II}_{\mathrm{down}}$ obtained from two directions must be equal. Therefore

$$\begin{aligned} -\lambda_1 B_1 + A_1 &= -i(\lambda_2 A_1 + B_1) \\ -\lambda_3(-\lambda_1 B_1 + A_1) + B_1 &= -iA_1 \end{aligned}$$

$$\begin{pmatrix} i\lambda_2 + 1 & -\lambda_1 + i \\ -\lambda_3 + i & \lambda_3\lambda_1 + 1 \end{pmatrix} \begin{pmatrix} A_1 \\ B_1 \end{pmatrix} = 0$$
$$\Rightarrow \lambda_1 = \lambda_2 = \lambda_3 = i$$

All three Stokes constants are i. For the Airy equation, the amount of exponentially decaying solution you can have going from one Stokes line to another is restricted to i in counter-clockwise direction.

3.10.4 Application of Stokes Lines

*Ex.*1 Reflection Coefficient from Stokes Constant ???

*Ex.*2 O-mode

For EM O-mode wave in a non-magnetized plasma, the governing Equation 3.9 is rewritten here

$$\frac{\mathrm{d}^2 E}{dz^2} + k^2 E = 0, \quad k^2 = \frac{\omega^2}{c^2}\Big[1 - \frac{{\omega_p}^2(z)}{\omega^2}\Big]$$

Let the turning point be $z = z_t$. Near z_t, let $k^2(z) = -p^2(z)(z - z_t)$ (Figure 3.3), where $p(z)$ is real and positive for all real z. Then

$$k = \pm ip(z)(z - z_t)^{1/2}$$

In the following we need to choose sign s.t. k is real and positive on the anti-Stokes line AS_1. (The figure of complex z plane is the same as Figure 3.5 except that the center point is $z = z_t$. AS_1 is the dashed line.)

To make z single value, define $-\pi \leq \arg(z - z_t) < \pi$. On AS_1,

$$\begin{aligned} k &= \pm ip(z)|z - z_t|^{1/2} e^{i\frac{\pi}{2}} \\ &= \mp|p(z)(z - z_t)|^{1/2} \\ &= \mp|k| \end{aligned}$$

We choose the second sign for right propagating wave, which has

$$k = -ip(z)(z - z_t)^{1/2}$$

The sign is important here because it determines the propagation direction.

For z on S_1, $\arg(z - z_t) = 0$.

$$k = -ip(z)|z - z_t|^{1/2} e^{i\frac{1}{2}0} = -i|k|$$

In region I (between S_1 and S_2)

$$E_I \sim \frac{1}{\sqrt{k}} e^{-i\omega t \pm i\int_{z_t}^{z} kdz'} \sim \frac{1}{\sqrt{k}} e^{-i\omega t \pm \int_{z_t}^{z} |k|dz'}$$

For a decaying (subdominant) solution we must use minus (lower) sign, so

$$E_I \sim E_{S_1} \sim \frac{1}{\sqrt{k}} e^{-i\omega t - \int_{z_t}^{z} k dz'}$$

Crossing S_2 into region II_{up},

$$E_{S_2} \sim \frac{1}{\sqrt{k}} e^{-i\omega t} \Big[\underbrace{e^{-i\int_{z_t}^{z} k dz'}}_{\text{dominant on } S_2} + \underbrace{ie^{i\int_{z_t}^{z} k dz'}}_{\text{sub-dominant on } S_2} \Big] = E_{AS_1}$$

where the first term represents the reflected wave, and the second term represents the incident wave. Since the magnitude of the incident and reflected wave is the same, the reflection coefficient should be 1, and the absorption coefficient should be 0. (In fact, there is a transmitted wave that is exponentially decay. WKBJ method cannot resolve this exponentially small value.)

*Ex.*3 Bohr-Sommerfield Quantization Rule

This is the classical potential well problem, where the two boundaries are $z = z_1$ and $z = z_2$.

Schrödinger equation reads

$$\frac{\mathrm{d}^2\Psi}{dz^2} - \frac{2m}{\hbar^2}[E - V(z)]\Psi = 0$$

Imagine there is an potential well between z_1 and z_2 shown in Figure 3.6: what are the allowable energy state?

In this case, $k^2 = -\frac{2m}{\hbar^2}[E - V(z)]$. Wave travelling to the right has positive $k(z)$, while wave travelling to the left has negative $k(z)$. Following the discussion of reflection coefficient,

$$\begin{aligned} R &= ie^{-2i\int_0^{z_2} k(z')dz'} \\ R' &= ie^{+2i\int_0^{z_1} k(z')dz'} \end{aligned}$$

For waves bouncing back and forth inside the well, we must have

$$\begin{aligned} RR' &= 1 \\ \Rightarrow (-1)e^{-2i\int_{z_1}^{z_2} kdz'} &= 1 \\ \Rightarrow \int_{z_1}^{z_2} k(z')dz' &= \pi(n + \frac{1}{2}),\ n = 0, 1, 2, 3, ... \end{aligned}$$

which is called the *Bohr-Sommerfield quantization rule*.

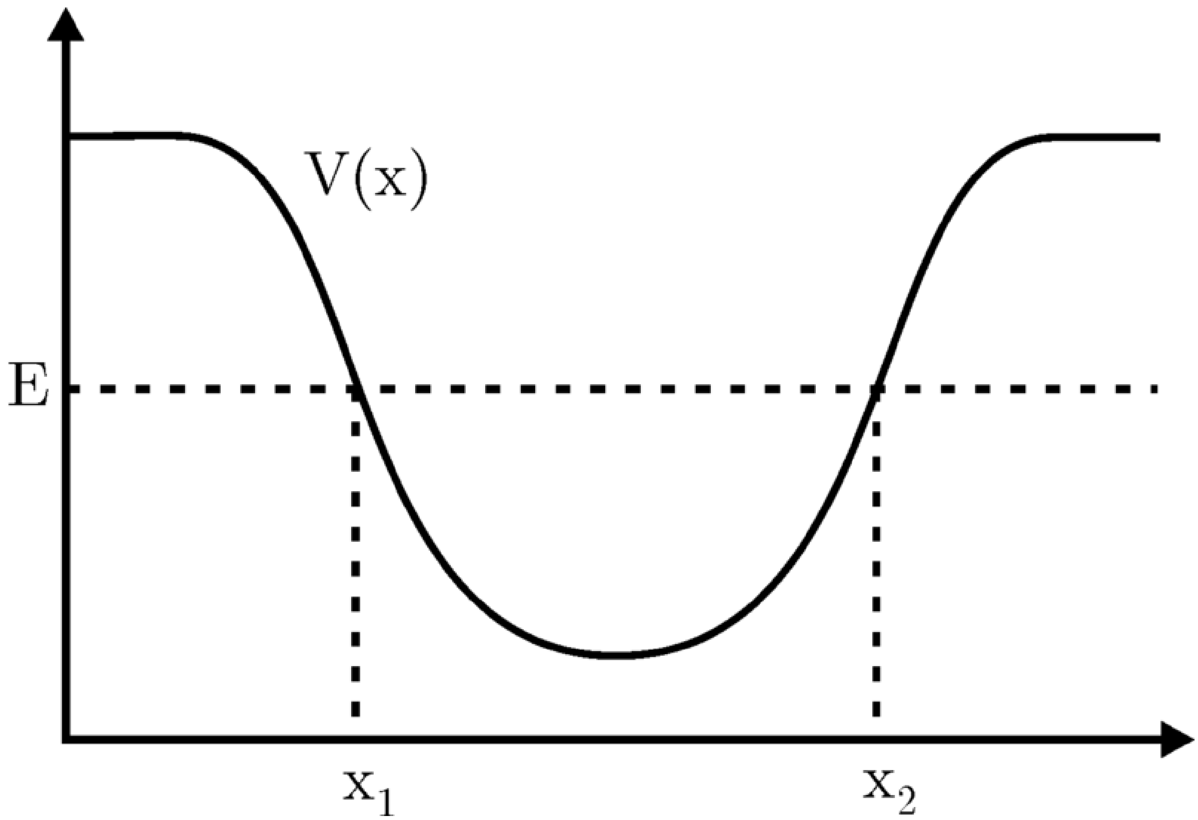

Figure 3.6: Potential well.

Another way to do this problem is by recognizing the Stokes and anti-Stokes lines in the complex z plane and match the solution in the whole domain. Figure needed!!! As the opposite in HW8.1%k2 .vs. z , Stokes line

$$k^2 = -\frac{2m}{\hbar^2}[E - V(z)] \equiv -p^2(z)(z - z_1)(z - z_2)$$

where for large z, we must choose a minus sign in front of $p(z)$ for $p(z)$ to be real and positive. So

$$k = \pm ip(z)(z - z_1)^{1/2}(z - z_2)^{1/2}$$

Then for a single-value solution, pick

$$-\pi \leq \arg(z - z_1) < \pi$$
$$-\frac{\pi}{2} \leq \arg(z - z_2) < \frac{3\pi}{2}$$

For z on AS_1,

$$k(z) = \pm ip(z)|z - z_1|^{1/2}e^{i\frac{1}{2}\cdot 0} \cdot |z - z_2|^{1/2}e^{i\frac{1}{2}\cdot\pi} = \pm i|k|e^{i\frac{1}{2}\pi} = \mp|k|$$

For a real and positive k (right-traveling wave), we must choose the lower sign $+$, so

$$k = -ip(z)(z - z_1)^{1/2}(z - z_2)^{1/2}$$

For z on S_1,

$$k = -ip(z)|z - z_1|^{1/2}|z - z_2|^{1/2} = -i|k|$$

For z on S_4,

$$k = -ip(z)|z - z_1|^{1/2}e^{i\frac{1}{2}\pi}|z - z_2|^{1/2}e^{i\frac{1}{2}\pi} = i|k|$$

So

$$k(z) = \begin{cases} -i|k| & z > z_2 \\ |k| & z_1 < z < z_2 \\ i|k| & z < z_1 \end{cases} \tag{3.15}$$

On S_1,

$$\Psi_{E_{S_1}} \sim \frac{1}{\sqrt{k}} e^{\pm i \int_{z_2}^{z} k dz'} = \frac{1}{\sqrt{k}} e^{-i \int_{z_2}^{z} k dz'}$$

such that this is sub-dominant on S_1 $(k = -i|k|)$.

Crossing S_2 into region I, we have

$$\Psi_I \sim \frac{1}{\sqrt{k}} \Big\{ \underbrace{e^{-i \int_{z_2}^{z} k dz'}}_{\text{dominant on } S_2} + i \underbrace{e^{+i \int_{z_2}^{z} k dz'}}_{\text{subdominant on } S_2} \Big\} \tag{3.16}$$

On S_4,

$$\Psi_{S_4} \sim \frac{1}{\sqrt{k}} e^{\pm i \int_{z_1}^{z} k dz'} \sim \frac{1}{\sqrt{k}} e^{-i \int_{z_1}^{z} k dz'},$$

where minus sign is chosen s.t. it is subdominant on S_4.

Crossing S_3 into region I in the clockwise direction, we have

$$\begin{aligned} \Psi_I &\sim \Big\{ \underbrace{e^{-i \int_{z_1}^{z} k dz'}}_{\text{dominant on } S_3} - i \underbrace{e^{+i \int_{z_1}^{z} k dz'}}_{\text{subdominant on } S_3} \Big\} \\ &= \frac{1}{\sqrt{k}} \Big\{ e^{-i \int_{z_2}^{z} k dz'} \cdot e^{-i \int_{z_1}^{z_2} k dz'} - i e^{i \int_{z_2}^{z} k dz'} \cdot e^{i \int_{z_1}^{z_2} k dz'} \Big\} \end{aligned} \tag{3.17}$$

Because the solution is analytic, Equation 3.16 and Equation 3.17 should be equal, therefore

$$\begin{aligned} &\frac{1}{e^{-i \int_{z_1}^{z_2} k dz'}} = \frac{i}{-i e^{i \int_{z_1}^{z_2} k dz'}} \\ \Rightarrow &e^{2i \int_{z_1}^{z_2} k dz'} = -1 = e^{i(\pi + 2n\pi)}, \quad n = 0, 1, 2, 3, ... \\ \Rightarrow &\int_{z_1}^{z_2} k dz' = (\frac{1}{2} + n)\pi, \quad n = 0, 1, 2, 3, ... \end{aligned}$$

*Ex.*4 Tunneling Problem

Instead of a potential well, now we consider another classical tunneling problem, where $E > V$ if $z_1 < z < z_2$.

$$k^2(z) = \pm p^2(z)(z - z_1)(z - z_2) = p^2(z)(z - z_1)(z - z_2)$$

where the plus sign is chosen s.t. $p(z)$ is real and positive for all real z, and then

$$k = p(z)(z - z_1)^{1/2}(z - z_2)^{1/2}$$

Define arguments as follows

$$0 \leq \arg z - z_2 < 2\pi, \quad -\pi < \arg z - z_1 \leq \pi$$

to remove the ambiguity in the branch in the expression of k.

Since $z = z_1$, $z = z_2$ are simple turning points (with different slope in linear approximation), the Stokes and anti-Stokes lines are shown in Figure needed!!!

The branch cut defined as above give the following values of k on the real z axis:

$$k(z) = \begin{cases} |k| & z > z_2 \\ i|k| & z_1 < z < z_2 \\ -|k| & z < z_1 \end{cases} \tag{3.18}$$

On AS_1, the solution

$$E \sim \frac{1}{\sqrt{k}} e^{-i\omega t + i\int_{z_2}^{z} k dz'} \tag{3.19}$$

represents the outgoing wave propagating to $z = \infty$.

On S_2, solution Equation 3.19 behaves like $E \sim e^{\int_b^z |k| dz'}$ which is dominant on S_2 with respect to $z = z_2$. Thus, on S_1, it is subdominant.

On crossing S_1 from AS_1, solution Equation 3.19 remains valid. In fact, it is valid in the region bounded by S_1, S_2 and S_3. Note that it is subdominant on S_2 with respect to $z = z_1$ (because we just showed that it is dominant on S_2 with respect to $z = z_2$). Thus, we rewrite the solution as

$$E \sim \frac{1}{\sqrt{k}} e^{-i\omega t + i\int_{z_2}^{z_1} k dz'} \cdot \underbrace{e^{i\int_{z_1}^{z} k dz'}}_{\text{subdominant on } \; S_2 w.r.t \; z=a}$$

which becomes dominant on S_3. Upon crossing S_3, this solution becomes

$$E \sim \frac{1}{\sqrt{k}} e^{-i\omega t + i\int_{z_2}^{z_1} k dz'} [\underbrace{e^{i\int_{z_1}^{z} k dz'}}_{\text{dominant on } S_3} + \underbrace{i e^{-i\int_{z_1}^{z} k dz'}}_{\text{subdominant on } S_3}]$$

where we pick up the Stokes constant i and the subdominant solution, represented by the last term of the solution.

Now, this is the solution on AS_4 $(z < z_1)$. Referring to Equation 3.18, we see that the first term of the solution represents the reflected wave, and the second term the incident wave from $z = -\infty$. Since the incident wave and reflected wave have equal amplitude, we have

$$|R|^2 = 1$$

The transmitted wave has unit amplitude (see Eq.(3.19)). The incident wave has amplitude $e^{i\int_a^b kdz'}(i)$. Thus,

$$|T|^2 = \frac{1}{\left|e^{i\int_{z_1}^{z_2} kdz'}(i)\right|^2} = \frac{1}{\left|e^{i\int_{z_1}^{z_2} i|k|dz'}(i)\right|^2} = e^{-2\int_{z_1}^{z_2}|k|dz}$$

where z_1, z_2 are far apart, i.e. $|T|^2$ is exponentially small.

Note: $|R|$ and $|T|$ above are not valid if z_1, z_2 are close to each other. The correct transmission coefficient, for general valies of z_1, z_2 is

$$|T|^2 = \frac{e^{-2\int_{z_1}^{z_2}|k|dz}}{1 + e^{-2\int_{z_1}^{z_2}|k|dz}}$$

and the reflection coefficient is

$$|R|^2 = 1 - |T|^2$$

*Ex.*5 Two waves: one is launched from $z = +\infty$, incident onto a resonant layer at $z = a(a > 0)$, and the other launched from $z = -\infty$, incident onto a absorption layer at $z = 0$, with the model equation

$$\frac{\mathrm{d}^2 E(z)}{dz} + {k_0}^2\Big(\frac{z}{z-a}\Big)E(z) = 0$$

This looks like a similar case for ECRH.

$$k^2(z) = {k_0}^2\Big(\frac{z}{z-a}\Big) \Rightarrow k(z) = k_0 z^{1/2}(z-a)^{-1/2}$$

where k_0 is real and positive.

Define arguments as follows

$$-\frac{3\pi}{2} \leq \arg z < \frac{\pi}{2}, \quad -\frac{3\pi}{2} < \arg z - a \leq \frac{\pi}{2}$$

to remove the ambiguity in the branch in the expression of k.

For z on AS_1,

$$k = k_0|z|^{1/2}e^{-\frac{1}{2}0}|z-a|^{-1/2}e^{i\frac{1}{2}0} = |k|$$

For z on S_1,

$$k = k_0|z|^{1/2}e^{-\frac{1}{2}0}|z-a|^{-1/2}e^{i\frac{1}{2}\pi} = i|k|$$

For z on AS_2,

$$k = k_0|z|^{1/2}e^{i\frac{1}{2}(-\pi)}|z-a|^{-1/2}e^{i\frac{-1}{2}(-\pi)} = |k|$$

Then the branch cut defined as above give the following values of k on the real z axis:

$$k(z) = \begin{cases} |k| & z > a \\ i|k| & 0 < z < a \\ |k| & z < 0 \end{cases} \tag{3.20}$$

For wave launched from large magnetic field side $z = +\infty$ going to $z = -\infty$ on AS_3, the solution on AS_3 is (Note that even though z is negative, it is the upper limit of the integral and k is positive. It took me a long time to get the right sign here for a left propagating wave.)

$$E \sim \frac{1}{\sqrt{k}} e^{-i\omega t - i\int_0^z k dz'}$$

This solution behaves like

$$E \sim \frac{1}{\sqrt{k}} e^{-i\omega t - i\int_0^z i|k| dz'} \sim \frac{1}{\sqrt{k}} e^{-i\omega t + \int_0^z |k| dz'} \tag{3.21}$$

on S_1. Thus, it is dominant on S_1 with respect to $z = 0$; it is subdominant on the neighboring Stokes line S_2. Thus solution Equation 3.21 remains valid upon crosing Stokes line S_2. It is the solution in region I, bounded by S_1 and S_2. Since it is dominant on S_1 with respect to $z = 0$, it is subdominant on S_1 with respect to $z = a$. So we can write it as

$$E \sim \frac{1}{\sqrt{k}} e^{-i\omega t - i\int_0^a k dz'} \cdot \underbrace{e^{-i\int_a^z k dz'}}_{\text{subdominant on } S_1 \ w.r.t. \ z=a}$$

which is also the incident wave from $z = 0$. This is the solution in region I as well as on AS_1 ($k = |k|$). So it is the incident wave solution from $z = \infty$. The reflection coefficient must be 0!

The transmisson coefficient is then

$$|T|^2 = \Big|\frac{E(z=-\infty)}{E(z=\infty)}\Big|^2 = \frac{1}{|e^{-i\int_0^a kdz}|^2}$$
$$= e^{-2\int_0^a |k|dz} = e^{-2\int_0^a k_0\sqrt{\frac{z}{a-z}}dz}$$

Let $z = a\sin^2\theta$, we can finally have

$$|T|^2 = e^{-k_0 a\pi}$$

The fraction of power absorbed by resonance is

$$1 - |R|^2 - |T|^2 = 1 - e^{-k_0 a\pi}$$

Note: these expressions of $|T|^2,\ |R|^2$ are valid even if a is small.

For wave launched from low magnetic field side $z = -\infty$,

$$E_{AS_1} \sim \frac{1}{\sqrt{k}} e^{-i\omega t + i\int_a^z kdz'}$$

On S_1, a subdominant (decaying) solution with respect to $z = 0$ is

$$E_{S_1} \sim \frac{1}{\sqrt{k}} e^{-i\omega t + i\int_0^z i|k|dz'}$$

In regions I bounded by $S_1,\ S_2$ and AS_1,

$$E_I \sim E_{S_1}\frac{1}{\sqrt{k}} e^{i\int_a^0 kdz'} \cdot e^{i\int_0^z kdz'}$$

E_{S_1} is dominant on S_1 with respect to $z = a$, but is subdominant on S_1 with respect to $z = 0$.

Crossing S_2 into AS_2,

$$E_{AS_2} \sim \frac{1}{\sqrt{k}} e^{-i\omega t} \cdot e^{i\int_a^0 kdz'}\Big[\underbrace{e^{i\int_0^z kdz'}}_{\text{dominant on } S_3} - i \underbrace{e^{-i\int_0^z kdz'}}_{\text{subdominant on } S_3}\Big],$$

where the first term represents the reflected wave, and the second term represents the incident wave.

The transmission coefficient is

$$
\begin{aligned}
T &= \Big|\frac{E(z=\infty)}{E(z=-\infty)}\Big| \\
&= \Big|\frac{\sqrt{k(-\infty)}}{\sqrt{k(\infty)}}\Big|\Big|\frac{e^{i\int_a^\infty kdz}}{e^{i\int_a^0 kdz}\cdot e^{i\int_0^{-\infty} kdz}}\Big| \\
&= \frac{1}{\big|e^{i\int_a^0 i|k|dz}\big|} = e^{-\int_0^a |k|dz}
\end{aligned}
$$

Then the transmission coefficient in power is

$$
T^2 = e^{-2\int_0^a k\sqrt{|\frac{z}{z-a}|}dz'} = e^{-k_0 a\pi}
$$

where the last equivalence requires some transformation tricks. If $a \sim \lambda, k_0 = 2\pi/\lambda$, the transmitted power fraction is on the order of $e^{-20} \sim 10^{-9}$.

The reflection coefficient is zero and that the transmission coefficient (in power) is $T = e^{-k_0 a\pi}$. Thus, the fraction $1-T$ of the incident power is absorbed by that resonant layer.

3.10.5 "Exact" WKBJ Solution

- Shielding or tunneling

If you have a high density region between x_1 and x_2,

$$
k^2(x) = \frac{\omega^2}{c^2}[1-\frac{{\omega_p}^2(x)}{\omega^2}]
$$

The transmission coefficient in power is

$$
|T|^2 = \frac{e^{-2\int_{x_1}^{x_2}|k|\mathrm{d}x}}{1+e^{-2\int_{x_1}^{x_2}|k|\mathrm{d}x}}
$$

and the reflection coefficient in power is

$$
|R|^2 = \frac{1}{1+e^{-2\int_{x_1}^{x_2}|k|\mathrm{d}x}}
$$

Note:

- $|R|^2 + |T|^2 = 1$
- When $x_1 = x_2$, i.e. $\omega^2 = {\omega_p}^2$(peak), $|R|^2 = |T|^2 = \frac{1}{2}$.
- ECRH

$$\frac{\mathrm{d}^2 E}{dz^2} + {k_0}^2 \Big(\frac{z}{z-a}\Big) E = 0$$

where

$$k^2(z) = {k_0}^2 \Big(\frac{z}{z-a}\Big)$$

(I) Incident wave from $z = -\infty$ (from low B_0 side)

$$\begin{aligned} |R| &= 1 - e^{-\pi k_0 a} \\ |T| &= e^{-\frac{1}{2}\pi k_0 a} \\ |R|^2 + |T|^2 &= 1 - e^{-\pi k_0 a} + e^{-2\pi k_0 a} < 1 \end{aligned}$$

(II) Incident wave from $z = \infty$

$$\begin{aligned} |R| &= 0 \\ |T| &= e^{-\pi k_0 a/2} \\ |R|^2 + |T|^2 &= e^{-\pi k_0 a} < 1 \end{aligned}$$

which means almost perfect absorption at $z = a$.

Note: These formulas are valid even if $a \to 0$, in which limit $|R| \to 0$, $|T| \to 1$, for both (I) and (II), as expected from the model equation.

Singularity in refractive index n can lead to absorption, even without collision.

3.11 Normalization

In physics we frequently have to deal with unit conversions. A solid understanding of the unit system and normalization methods paves the path for understanding the scales behind systems.

3.12 Theoretical Mechanics

Newtonian mechanics is great, but theoretical mechanics is marvellous. It does not tell you anything new, but let you think of the same thing from a different prospective. Theoretical mechanics takes advantage of a system's constraints to solve problems. The constraints limit the *degrees of freedom* the system can have, and can be used to reduce the number of coordinates needed to solve for the motion. The formalism is well suited to arbitrary choices of coordinates, known in the context as *generalized coordinates*. The kinetic and potential energies of the system are expressed using these generalized coordinates or momenta, and the equations of motion can be readily set up, thus analytical mechanics allows numerous mechanical problems to be solved with greater efficiency than fully vectorial methods. *It does not always work for non-conservative forces or dissipative forces like friction, in which case one may revert to Newtonian mechanics.*

3.12.1 Generalized Coordinates

In Newtonian mechanics, one customarily uses all three Cartesian coordinates, or other 3D coordinate system, to refer to a body's position during its motion. In physical systems, however, some structure or other system usually constrains the body's motion from taking certain directions and pathways. So a full set of Cartesian coordinates is often unneeded, as the constraints determine the evolving relations among the coordinates, which relations can be modeled by equations corresponding to the constraints. In the Lagrangian and Hamiltonian formalisms, the constraints are incorporated into the motion's geometry, reducing the number of coordinates to the minimum needed to model the motion. These are known as generalized coordinates, denoted $q_i\,(i = 1, 2, 3...)$.

3.12.2 D'Alembert's Principle

This principle is the analogy of Newton's second law in theoretical mechanics. It states that infinitesimal virtual work done by a force across reversible displacements is zero, which is the work done by a force consistent with ideal constraints of the system. The idea of a constraint is useful — since this limits what the system can do, and can provide steps to solving for the motion of the system. The equation for D'Alembert's principle is:

$$\delta W = \mathbf{Q} \cdot \mathrm{d}\mathbf{q} = 0$$

where

$$\mathbf{Q} = (Q_1, Q_2, ..., Q_N)$$

are the generalized forces (script Q instead of ordinary Q is used here to prevent conflict with canonical transformations below) and $\mathbf{q}$ are the generalized coordinates. This leads to the generalized form of Newton's laws in the language of theoretical mechanics:

$$\mathbf{Q} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial T}{\partial \dot{\mathbf{q}}}\right) - \frac{\partial T}{\partial \mathbf{q}}$$

where T is the total kinetic energy of the system, and the notation

$$\frac{\partial}{\partial \mathbf{q}} = \left(\frac{\partial}{\partial q_1}, \frac{\partial}{\partial q_2}, ..., \frac{\partial}{\partial q_N}\right)$$

3.12.3 Lagrangian Mechanics

Lagrangian and Euler–Lagrange equations

The introduction of generalized coordinates and the fundamental Lagrangian function:

$$L(\mathbf{q}, \dot{\mathbf{q}}, t) = T(\mathbf{q}, \dot{\mathbf{q}}, t) - V(\mathbf{q}, \dot{\mathbf{q}}, t)$$

where T is the total kinetic energy and V is the total potential energy of the entire system, then either following the calculus of variations or using the above formula — lead to the Euler-Lagrange equations;

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{\mathbf{q}}}\right) = \frac{\partial L}{\partial \mathbf{q}}$$

which are a set of N second-order ordinary differential equations, one for each $q_i(t)$. This formulation identifies the actual path followed by the motion as a selection of the path over which the time integral of kinetic energy is least, assuming the total energy to be fixed, and imposing no conditions on the time of transit.

3.12.4 Hamiltonian Mechanics

Hamiltonian and Hamilton's equations

The Legendre transformation of the Lagrangian replaces the generalized coordinates and velocities $(\mathbf{q}, \dot{\mathbf{q}})$ with $(\mathbf{q}, \mathbf{p})$; the generalized coordinates and the generalized momenta conjugate to the generalized coordinates:

$$\mathbf{p} = \frac{\partial L}{\partial \dot{\mathbf{q}}} = \left(\frac{\partial L}{\partial \dot{q}_1}, \frac{\partial L}{\partial \dot{q}_2}, , ..., \frac{\partial L}{\partial \dot{q}_N}\right) = (p_1, p_2, ..., p_N)$$

and introduces the Hamiltonian (which is in terms of generalized coordinates and momenta):

$$H(\mathbf{q}, \mathbf{p}, t) = \mathbf{p} \cdot \dot{\mathbf{q}} - L(\mathbf{q}, \dot{\mathbf{q}}, t)$$

This leads to the Hamilton's equations:

$$\dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{q}}, \quad \dot{\mathbf{q}} = \frac{\partial H}{\partial \mathbf{p}}$$

which are now a set of 2N first-order ordinary differential equations, one for each $q_i(t)$ and $p_i(t)$. Another result from the Legendre transformation relates the time derivatives of the Lagrangian and Hamiltonian:

$$\frac{dH}{\mathrm{d}t} = -\frac{\partial L}{\partial t}$$

which is often considered one of Hamilton's equations of motion additionally to the others. The generalized momenta can be written in terms of the generalized forces in the same way as Newton's second law:

$$\dot{\mathbf{p}} = \mathbf{Q}$$

One interesting application of the Hamilton's equation in plasma physics is the prove of Vlasov equation (See Giulia's notes)

$$\frac{df(\mathbf{q}, \mathbf{p}, t)}{\mathrm{d}t} = 0$$

You may also find the application in deriving the gyrokinetic equations.

3.12.5 Hamilton-Lagrange formalism vs. Lorentz equation

Two mathematically equivalent formalisms describe charged particle dynamics, namely

1. the Lorentz equation
$$m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$
2. Hamiltonian-Lagrangian dynamics.

The two formalisms are complementary: the Lorentz equation is intuitive and suitable for approximate methods, whereas the more abstract Hamiltonian-Lagrangian formalism exploits time and space symmetries. A brief review of the Hamiltonian-Lagrangian formalism follows, emphasizing aspects relevant to dynamics of charged particles.

The starting point is to postulate the existence of a function L, called the Lagrangian, which:

1. contains *all* information about the particle dynamics for a given situation;
2. depends only on generalized coordinates $Q_i(t), \dot{Q}_i(t)$ appropriate to the problem;
3. possibly has an explicit dependence on time t.

If such a function $L(Q_i(t), \dot{Q}_i(t), t)$ exists, then information on particle dynamics is retrieved by manipulation of the *action integral*

$$S = \int_{t_1}^{t_2} L(Q_i(t), \dot{Q}_i(t), t)\mathrm{d}t$$

This manipulation is based on d'Alembert's principle of least action. According to this principle, one considers the infinity of possible trajectories a particle could follow to get from its initial position $Q_i(t_1)$ to its final position $Q_i(t_2)$, and postulates that the trajectory actually followed is the one that results in the lowest value of S. Thus, the value of S must be minimized (note that S here is action, and not entropy). Minimizing S does not give the actual trajectory directly, but rather gives equations of motion, which can be solved to give the actual trajectory.

Minimizing S is accomplished by considering an arbitrary nearby alternative trajectory $Q_i(t) + \delta Q_i(t)$ having the same beginning and end points as the actual trajectory, i.e., $\delta Q_i(t_1) = Q_i(t_2) = 0$. In order to make the variational argument more precise, δQ_i is expressed as

$$\delta Q_i(t) = \epsilon \eta_i(t)$$

where ϵ is an arbitrarily adjustable scalar assumed to be small so that $\epsilon^2 < \epsilon$ and $\eta_i(t)$ is a function of t that vanishes when $t = t_1$ or $t = t_2$ but is otherwise arbitrary. Calculating δS to second order in ϵ gives

$$\begin{aligned}
\delta S &= \int_{t_1}^{t_2} L(Q_i + \delta Q_i, \dot{Q}_i + \delta \dot{Q}_i, t)\mathrm{d}t - \int_{t_1}^{t_2} L(Q_i(t), \dot{Q}_i(t), t)\mathrm{d}t \\
&= \int_{t_1}^{t_2} L(Q_i + \epsilon\eta_i, \dot{Q}_i + \epsilon\dot{\eta}_i, t)\mathrm{d}t - \int_{t_1}^{t_2} L(Q_i(t), \dot{Q}_i(t), t)\mathrm{d}t \\
&= \int_{t_1}^{t_2} \left(\epsilon\eta_i \frac{\partial L}{\partial Q_i} + \frac{(\epsilon\eta_i)^2}{2} \frac{\partial^2 L}{\partial Q_i^2} + \epsilon\dot{\eta}_i \frac{\partial L}{\partial \dot{Q}_i} + \frac{(\epsilon\dot{\eta}_i)^2}{2} \frac{\partial^2 L}{\partial \dot{Q}_i^2} \right) \mathrm{d}t
\end{aligned}$$

Suppose the trajectory $Q_i(t)$ is the one that minimizes S. Any other trajectory must lead to a higher value of S and so δS must be positive for any finite value of ϵ. If ϵ is chosen to be sufficiently small, then the absolute values of the terms of order ϵ^2 above will be smaller than the absolute values of the terms linear in ϵ. The sign of ϵ could then be chosen to make δS nagative, but this would violate the requirement that δS must be positive. The only way out of this dilemma is to insist that the sum of the terms linear in ϵ vanishes so $\delta S \sim \epsilon^2$ and is therefore always positive as required. Insisting that the sum of terms linera in ϵ vanishes implies

$$0 = \int_{t_1}^{t_2} \left(\eta_i \frac{\partial L}{\partial Q_i} + \dot{\eta}_i \frac{\partial L}{\partial \dot{Q}_i} \right) \mathrm{d}t$$

Using $\dot{\eta}_i = \mathrm{d}\eta_i/\mathrm{d}t$ the above expression may be integrated by parts to obtain

$$\begin{aligned} 0 &= \int_{t_1}^{t_2} \left(\eta_i \frac{\partial L}{\partial Q_i} + \dot{\eta}_i \frac{\partial L}{\partial \dot{Q}_i} \right) \mathrm{d}t \\ &= \left[\eta_i \frac{\partial L}{\partial \dot{Q}_i} \right]_{t_1}^{t_2} + \int_{t_1}^{t_2} \left[\eta_i \frac{\partial L}{\partial Q_i} - \eta_i \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial L}{\partial \dot{Q}_i} \right) \right] \mathrm{d}t \end{aligned}$$

Since $\eta_i(t_{1,2}) = 0$, the integrated term vanishes and since η_i was an arbitrary function of t, the coefficient of i in the integrand must vanish, yielding Lagrange's equation

$$\frac{\mathrm{d}P_i}{\mathrm{d}t} = \frac{\partial L}{\partial Q_i} \tag{3.22}$$

where the *canonical momentum* P_i is defined as

$$P_i \equiv \frac{\partial L}{\partial \dot{Q}_i} \tag{3.23}$$

Lagrange's equation shows that if L does not depend on a particular generalized coordinate Q_j, then $\mathrm{d}P_j/\mathrm{d}t = 0$, in which case the canonical momentum P_j is a *constant of motion*; the coordinate Q_j is called a *cyclic* or *ignorable* coordinate. This is a very powerful and profound result. Saying that the Lagrangian function does not depend on a coordinate is equivalent to saying that the system is *symmetric* in that coordinate or translationally invariant with respect to that coordinate. The quantities P_j and Q_j are called conjugate and action has the dimensions of the product of these quantities.

Hamilton extended this formalism by introducing a new function related to the Lagrangian. This new function, called the Hamiltonian, provides further useful information and is defined as

$$H \equiv \left(\sum_i P_i \dot{Q}_i \right) - L \tag{3.24}$$

Partial derivative of H with respect to P_i and to Q_i give Hamilton's equations

$$\dot{Q}_i = \frac{\partial H}{\partial P_i}, \quad \dot{P}_i = -\frac{\partial H}{\partial Q_i}$$

which are equations of motion having a close relation to phase-space concepts. The time derivative of the Hamiltonian is

$$\frac{\mathrm{d}H}{\mathrm{d}t} = \sum_i \frac{\mathrm{d}P_i}{\mathrm{d}t} \dot{Q}_i + \sum_i P_i \frac{\mathrm{d}\dot{Q}_i}{\mathrm{d}t} - \left(\sum_i \frac{\partial L}{\partial Q_i} \dot{Q}_i + \sum_i \frac{\partial L}{\partial \dot{Q}_i} \frac{\mathrm{d}\dot{Q}_i}{\mathrm{d}t} + \frac{\partial L}{\partial t} \right) = -\frac{\partial L}{\partial t} \tag{3.25}$$

This shows that if Ldoes not explicitly depend on time, the Hamiltonian H is a *constant of the motion*. As will be shown later, H corresponds to the energy of the system, so if $\partial L/\partial t = 0$, the energy is a constant of the motion. Thus, energy is conjugate to time in analogy to canonical momentum being conjugate to position (note that energy $\times$ time also has the units of action). If the Lagrangian does not explicitly depend on time, then the system can be thought of as being "symmetric" with respect to time, or "translationally" invariant with respect to time.

The Lagrangian for a charged particle in an electromagnetic field is

$$L = \frac{mv^2}{2} + q\mathbf{v} \cdot \mathbf{A}(\mathbf{x}, t) - q\phi(\mathbf{x}, t) \tag{3.26}$$

The validity of Equation 3.26 will now be established by showing that it generates the Lorentz equation when inserted into Lagrange's equation. Since no symmetry is assumed, there is no reason to use any special coordinate system and so ordinary Cartesian coordinates will be used as the canonical coordinates, in which case Equation 3.23 gives the canonical momentum as

$$\mathbf{P} = m\mathbf{v} + q\mathbf{A}(\mathbf{x}, t)$$

The left-hand side of Equation 3.22 becomes

$$\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}t} = m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} + q\left(\frac{\partial \mathbf{A}}{\partial t} + \mathbf{v} \cdot \nabla\mathbf{A}\right)$$

while the right-hand side of Equation 3.22 becomes

$$\begin{aligned}\frac{\partial L}{\partial \mathbf{x}} &= q\nabla(\mathbf{v} \cdot \mathbf{A}) - q\nabla\phi = q(\mathbf{v} \cdot \nabla\mathbf{A} + \mathbf{v} \times \nabla \times \mathbf{A}) - q\nabla\phi \\ &= q(\mathbf{v} \cdot \nabla\mathbf{A} + \mathbf{v} \times \mathbf{B}) - q\nabla\phi\end{aligned}$$

Equating the above two expressions gives the Lorentz equation, where the electric field is defined as $\mathbf{E} = -\partial\mathbf{A}/\partial t - \nabla\phi$ in accordance with Faraday's law. This proves that Equation 3.26 is mathematically equivalent to the Lorentz equation when used with the principle of least action.

The Hamiltonian associated with this Lagrangian is, in Cartesian coordinates,

$$\begin{aligned}H &= \mathbf{P} \cdot \mathbf{v} - L = \frac{mv^2}{2} + q\phi \\ &= \frac{\left(\mathbf{P} - q\mathbf{A}(\mathbf{x}, t)\right)^2}{2m} + q\phi(\mathbf{x}, t)\end{aligned} \tag{3.27}$$

where the last line is the form more suitable for use with Hamilton's equations, i.e., $H = H(\mathbf{x}, \mathbf{P}, t)$. Equation 3.27 also shows that H is, as promised, the particle energy. If generalized coordinates are used, the energy can be written in a general form as $E = H(Q, P, t)$. Equation 3.25 showed that even though both Q and P depend on time, the energy depends on time only if H explicitly depends on time. Thus, in a situation where H does not explicitly depend on time, the energy would have the form $E = H(Q(t), P(t)) = \text{const}$.

It is important to realize that both canonical momentum and energy depend on the reference frame. For example, a bullet fired in an airplane in the direction opposite to the airplane motion, and with a speed equal to the airplane's speed, has a large energy as measured in the airplane frame, but zero energy as measured by an observer on the ground. A more subtle example (of importance to later analysis of waves and Landau damping) occurs when $\mathbf{A}$ and/or ϕ has a wave-like dependence, e.g., $\phi(\mathbf{x}, t) = \phi(\mathbf{x} - \mathbf{v}_{\text{ph}} t)$, where $\mathbf{v}_{\text{ph}}$ is the wave phase velocity. This potential is time-dependent in the lab frame and so the associated Lagrangian has an explicit dependence on time in the lab frame, which implies that *energy is not a constant of the motion in the lab frame*. In contrast, ϕ is time-independent in the wave frame and so the energy is a *constant of the motion* in the wave frame. Existence of a constant of the motion reduces the complexity of the system of equations and typically makes it possible to integrate at least one equation in closed form. Thus, it is advantageous to analyze the system in the frame having the most constants of the motion.

3.13 Wavelet Analysis

Wavelet Transform (WT) decomposes a function into a set of wavelets. A Wavelet is a wave-like oscillation that is localized in time. Two basic properties of a wavelet are *scale* and *location*.

Wavelet is a predecessor of Fourier Tranform (FT), which provides better results when dealing with changing background. It is the de-facto method for modern wave analysis. The key advantages of WT compared with FT are:

- Fewer "hard" parameters to tune in WT. For example, if you want to make a spectrogram with FT, you need to specify the size of local DFTs. In WT you do not need to worry about this; instead the validity of the result can be shown by the cone of influence. In a sense, FT makes immediate conversion from time to frequency domain, while WT let you choose the intermediate steps you wish for.
- Flexible forms of wavelets to choose from. In practice, if you have any prior knowledge to the signal you want to identify, you can find for an appropriate wavelet that is close to that shape, which gives better fitting compared to the sinuisoidal functions in FT.

Check the notes from my blog and the references therein.

4 Single-Particle Motions

What makes plasmas particularly difficult to analyze is the fact that the densities fall in an intermediate range. Fluids like water are so dense that the motions of individual molecules do not have to be considered. Collisions dominate, and the simple equations of ordinary fluid dynamics suffice. At the other extreme in very low-density devices, only single-particle trajectories need to be considered; collective effects are often unimportant. Plasma behaves sometimes like fluids, and sometimes like a collection of individual particles. The first step in learning how to deal with this schizophrenic personality is to understand how single particles behave in electric and magnetic fields.

Single particle motion in neutral gases is trivial – particles move in straight lines until they hit other particles or the wall. Because of this simplicity, there is no point in keeping track of the details of single particle motion in a neutral gas and instead a statistical averaging of this motion suffices; this averaging shows that neutral gases have Maxwellian velocity distributions and are in a local thermodynamic equilibrium. In contrast, plasma particles are nearly collisionless and typically have complex trajectories that are strongly affected by both electric and magnetic fields.

As what will be shown in Section 8.2.5, the velocity distribution in a plasma will become Maxwellian when enough collisions have occurred to maximize the entropy. However, since collisions occur infrequently in hot plasmas, many important phenomena have time scales shorter than the time required for the plasma velocity distribution to become Maxwellian. A collisionless model is thus required to characterize these fast phenomena. In these situations randomization does not occur, entropy is conserved, the distribution function need not be Maxwellian, and the plasma is not in thermodynamic equilibrium. Thermodynamic concepts therefore do not apply, and the plasma is instead characterized by concepts from classical mechanics such as momentum or energy conservation of individual particles. In these collisionless situations the complex details of single particle dynamics are not washed out by collisions but instead persist and influence the macroscopic scale. As an example, the cyclotron resonance of a single particle can be important at the macroscopic scale in a collisionless plasma. This chapter examines various aspects of single particle motion and shows how these aspects can influence the macroscopic properties of a plasma.

Furthermore, study of single particle dynamics has a very direct relevance to Vlasov theory because, as shown in Section 8.3.1, any function constructed from constants of single particle motion is a valid solution of the collisionless Vlasov equation. Thus, knowledge of single particle dynamics provides a “repertoire” of constants of the motions from which solutions to the Vlasov equation suitable for various situations can be constructed.

Finally, the study of single particle motion develops valuable intuition regarding wave–particle interactions and identifies certain unusual situations, such as stochastic or non-adiabatic particle motion, that are beyond the descriptive capability of fluid models.

Here we assume that the EM fields are prescribed and not affected by the charged particles. The associated approach is often called *test particle*. Alternatively, we can say that it is possible to deduce intuitive and quite accurate analytic solutions for the drift of charged particles in arbitrarily complicated electric and magnetic fields provided the field are *slowly changing in both space and time*(this requirement is essentially the slowness requirement for adiabatic invariance introduced in Section 4.7). Drift solutions are obtained by solving the Lorentz equation

$$m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = q\left(\mathbf{E} + \mathbf{v}\times\mathbf{B}\right)$$

iteratively, taking advantage of the assumed separation of scales between fast and slow motions.

4.1 Uniform E and B Fields

The convention here is that velocity $\mathbf{v}$ represents single particle velocity as you will see through most of this chapter.

4.1.1 E=0

In this case, a charged particle has a simple cyclotron gyration. The equation of motion is

$$m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = q\mathbf{v}\times\mathbf{B}$$

Taking $\hat{z}$ to be the direction of $\mathbf{B}$ ($\mathbf{B} = B\hat{z}$), we have

$$\begin{aligned}
m\dot{v}_x = qBv_y,\quad m\dot{v}_y = -qBv_x,\quad m\dot{v}_z = 0,\\
\ddot{v}_x = \frac{qB}{m}\dot{v}_y = -\Big(\frac{qB}{m}\Big)^2 v_x\\
\ddot{v}_y = \frac{qB}{m}\dot{v}_x = -\Big(\frac{qB}{m}\Big)^2 v_y
\end{aligned}$$

This describes a simple harmonic oscillator at the *cyclotron frequency*, which we define to be

$$\omega_c \equiv \frac{|q|B}{m}$$

By the convention we have chosen, ω_c is always nonnegative. The unit of ω_c in SI units is rad/s. To convert it to hertz, we need to divide it by 2π. The solution of velocity is then

$$v_{x,y} = v_\perp \exp(\pm i\omega_c t + i\delta_{x,y})$$

The $\pm$ denote the sign of q. We may choose the phase δ so that

$$v_x = v_\perp e^{i\omega_c t} = \dot{x}$$

where $v_\perp$ is a positive constant denoting the speed in the plane perpendicular to $\mathbf{B}$. Then

$$v_y = \frac{m}{qB}\dot{v}_x = \pm\frac{1}{\omega_c}\dot{v}_x = \pm i v_\perp e^{i\omega_c t} = \dot{y}$$

Integrating once again, we have

$$\begin{aligned} x - x_0 &= -i\frac{v_\perp}{\omega_c}e^{i\omega_c t} \\ y - y_0 &= \pm i\frac{v_\perp}{\omega_c}e^{i\omega_c t} \end{aligned}$$

We define the *Larmor radius* to be

$$r_L \equiv \frac{v_\perp}{\omega_c} = \frac{mv_\perp}{|q|B} \tag{4.1}$$

Note that 2π does not appear in Equation 4.1!

Taking the real part of the positions, we have

$$\begin{aligned} x - x_0 &= r_L \sin\omega_c t \\ y - y_0 &= \pm r_L \cos\omega_c t \end{aligned}$$

This describes a circular orbit about a *guiding center* (x_0, y_0) which is fixed. The direction of the gyration is always such that the magnetic field generated by the charged particle is opposite to the externally imposed field. Plasma particles, therefore, tend to *reduce* the magnetic field, and plasmas are *diamagnetic*. In addition to this motion, there is an arbitrary velocity v_z along $\mathbf{B}$ which is not affected by $\mathbf{B}$. The trajectory of a charged particle in space is, in general, a helix.

4.1.2 Finite E

If now we allow an electric field to be present, the motion will be found to be the sum of two motions: the usual circular Larmor gyration plus a drift of the guiding center. The assumed spatial uniformity and time-independence of the fields represent the extreme limit of assuming that the fields are slowly changing in space and time. We may choose **E** to lie in the x-z plane so that $E_y = 0$. As before, the z component of velocity is unrelated to the transverse components and can be treated separately. The equations of motion is now

$$m\dot{\mathbf{v}} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

whose z component is

$$\dot{v_z} = \frac{q}{m} E_z$$

or

$$v_z = \frac{qE_z}{m} t + v_{z0}$$

This is a straightforward acceleration along **B**. The transverse components are

$$\begin{aligned}
\dot{v_x} &= \frac{q}{m} E_x \pm \omega_c v_y \\
\dot{v_y} &= 0 \mp \omega_c v_x
\end{aligned}$$

Differentiating, we have (for constant **E**)

$$\begin{aligned}
\ddot{v}_x &= -\omega_c^2 v_x \\
\ddot{v}_y &= \mp\omega_c \Big(\frac{q}{m} E_x \pm \omega_c v_y \Big) = -\omega_c^2 \Big(v_y + \frac{E_x}{B} \Big)
\end{aligned}$$

We can write this as

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} \Big(v_y + \frac{E_x}{B} \Big) = -\omega_c^2 \Big(v_y + \frac{E_x}{B} \Big)$$

so that it reduces to the previous case if we replace v_y by $v_y + (E_x/B)$. The velocity solution is then replaced by

$$\begin{aligned}
v_x &= v_\perp e^{i\omega_c t} \\
v_y &= \pm v_\perp e^{i\omega_c t} - \frac{E_x}{B}
\end{aligned}$$

The Larmor motion is the same as before, but there is superimposed a drift $\mathbf{v}_{gc}$ of the guiding center in the $-y$ direction (for $E_x > 0$).

To obtain a general formula for $\mathbf{v}_{gc}$, we can solve the momentum equation in vector form. We may omit the $m\mathrm{d}\mathbf{v}/\mathrm{d}t$ term, since this term gives only the circular motion at ω_c, which we already know about. Then the momentum equation becomes

$$\mathbf{E} + \mathbf{v} \times \mathbf{B} = 0$$

Taking the cross product with $\mathbf{B}$, we have

$$\mathbf{E} \times \mathbf{B} = \mathbf{B} \times (\mathbf{v} \times \mathbf{B}) = \mathbf{v}B^2 - \mathbf{B}(\mathbf{v} \cdot \mathbf{B})$$

The transverse components with respect to $\mathbf{B}$ of this equation are

$$\mathbf{v}_{gc} = \mathbf{E} \times \mathbf{B}/B^2 \equiv \mathbf{v}_E$$

We define this to be $\mathbf{v}_E$, the electric field drift of the guiding center. In magnitude, this drift is

$$v_E = \frac{E(\mathrm{V/m})}{B(\mathrm{tesla})} \frac{\mathrm{m}}{\mathrm{sec}}$$

It is important to note that $\mathbf{v}_E$ is independent of q, m, and $v_\perp$. The reason is obvious from the following physical picture. In the first half-cycle of the ion's orbit, it gains energy from the electric field and increases in $v_\perp$ and, hence, in r_L. In the second half-cycle, it loses energy and decreases in r_L. This difference in r_L on the left and right sides of the orbit causes the drift v_E. A negative electron gyrates in the opposite direction but also gains energy in the opposite direction; it ends up drifting in the same direction as an ion. For particles of the same velocity but different mass, the lighter one will have smaller r_L and hence drift less per cycle. However, its gyration frequency is also larger, and the two effects exactly cancel. Two particles of the same mass but different energy would have the same ω_c. The slower one will have smaller r_L and hence gain less energy from $\mathbf{E}$ in a half-cycle. However, for less energetic particles the fractional change in r_L for a given change in energy is larger, and these two effects cancel. Another way of interpreting this behavior is to recall that according to the theory of special relativity the electric field $\mathbf{E}'$ observed in a frame moving with velocity $\mathbf{u}$ is $\mathbf{E}' = \mathbf{E} + \mathbf{u} \times \mathbf{B}$ and so $\mathbf{v}_E$ is simply a statement that a particle drifts in such a way to ensure that the electric field seen in its own frame vanishes.

The three-dimensional orbit in space for a charged particle with a nonzero initial velocity is therefore a slanted helix with changing pitch.

4.1.3 Gravitational Field

The foregoing result can be applied to other forces by replacing $q\mathbf{E}$ in the equation of motion by a general force $\mathbf{F}$. The guiding center drift caused by $\mathbf{F}$ is then

$$\mathbf{v}_f = \frac{1}{q} \frac{\mathbf{F} \times \mathbf{B}}{B^2}$$

In particular, if $\mathbf{F}$ is the force of gravity $m\mathbf{g}$, there is a drift

$$\mathbf{v}_g = \frac{m}{q} \frac{\mathbf{g} \times \mathbf{B}}{B^2} \tag{4.2}$$

This is similar to the drift $\mathbf{v}_E$ in that it is perpendicular to both the force and $\mathbf{B}$, but it differs in one important respect. The drift $\mathbf{v}_g$ changes sign with the particle's charge. Under a gravitational force, ions and electrons drift in opposite directions, so there is a net current density in the plasma given by

$$\mathbf{j} = n(m_i + m_e)\frac{\mathbf{g} \times \mathbf{B}}{B^2}$$

The physical reason for this drift is again the change in Larmor radius as the particle gains and loses energy in the gravitational field. Now the electrons gyrate in the opposite sense to the ions, but the force on them is in the same direction, so the drift is in the opposite direction. The magnitude of $\mathbf{v}_g$ is usually negligible, but when the lines of force (i.e. magnetic field lines) are curved, there is an effective gravitational force due to centrifugal force. This force, which is not negligible, is independent of mass; this is why we did not stress the m dependence of the drift here. Centrifugal force is the basis of a plasma instability called the "gravitational" instability, which has nothing to do with real gravity.

And now it shall be clear that the $\mathbf{E} \times \mathbf{B}$ drift analysis can be easily generalized to describe the effect on a charged particle of any force orthogonal to $\mathbf{B}$ by simply making the replacement $\mathbf{E} \to \mathbf{F}/q$ in the Lorentz equation. Thus, any spatially uniform, temporally constant force orthogonal to $\mathbf{B}$ will cause a drift

$$\mathbf{v}_F = \frac{\mathbf{F} \times \mathbf{B}}{qB^2}$$

which leads to two counter-intuitive and important conclusions:

1. A steady-state electric field perpendicular to a magnetic field does not drive currents in a plasma, but instead causes a bulk motion of the entire plasma across the magnetic field with the velocity $\mathbf{v}_E$.
2. A steady-state force (e.g., gravity, centrifugal force, etc.) perpendicular to the magnetic field causes oppositely directed motions for electrons and ions and so drives a cross-field current

$$\mathbf{j}_F = \sum_\sigma n_\sigma \frac{\mathbf{F} \times \mathbf{B}}{B^2}$$

4.2 Nonuniform B Field

Now that the concept of a guiding center drift is firmly established, we can discuss the motion of particles in inhomogeneous fields — $\mathbf{E}$ and $\mathbf{B}$ fields which vary in space or time. For uniform fields we were able to obtain exact expressions for the guiding center drifts. As soon as we introduce inhomogeneity, the problem becomes too complicated to solve exactly. To get an approximate answer, it is customary to expand in the small ratio r_L/L, where L is the scale length of the inhomogeneity. This type of theory, called *orbit theory*, can become extremely involved. We shall examine only the simplest cases, where only one inhomogeneity occurs at a time.

4.2.1 B B: Grad-B Drift

Here the magnetic field lines are straight, but their density increases, say, in the y direction. We can anticipate the result by using our simple physical picture. The gradient in |B| causes the Larmor radius to be larger at the bottom of the orbit than at the top, and this should lead to a drift, in opposite directions for ions and electrons, perpendicular to both B and ∇B. The drift velocity should obviously be proportional to r_L/L and to $v_\perp$.

Consider the Lorentz force $\mathbf{F} = q\mathbf{v} \times \mathbf{B}$, averaged over a gyration. Clearly, $\bar{F}_x = 0$, since the particle spends as much time moving up as down. We wish to calculate $\bar{F}_y$, in an approximate fashion, by using the *undisturbed orbit* of the particle to find the average. The undisturbed orbit is given by the solution in the first section for a uniform **B** field. Taking the real part of the solution for v_x and y, we have

$$F_y = -qv_x B_z(y) = -qv_\perp(\cos\omega_c t)\Big[B_0 \pm r_L(\cos\omega_c t)\frac{\partial B}{\partial y}\Big]$$

where we have made a Taylor expansion of **B** field about the point $x_0 = 0, y_0 = 0$

$$\mathbf{B} = \mathbf{B}_0 + (\mathbf{r}\cdot\nabla)\mathbf{B} + ...$$
$$Bz = B_0 + y(\partial B_z/\partial y) + ...$$

This expansion of course requires $r_L/L \ll 1$, where L is the length scale of $\partial Bz/\partial y$. The first term above averages to zero in a gyration, and the average of $\cos^2\omega_c t$ is 1/2, so that

$$\bar{F}_y = \mp q v_\perp r_L \frac{1}{2}\frac{\partial B}{\partial y}$$

The guiding center drift velocity is then

$$\mathbf{v}_{gc} = \frac{1}{q}\frac{\mathbf{F}\times\mathbf{B}}{B^2} = \frac{1}{q}\frac{\bar{F}_y}{|B|}\hat{x} = \mp\frac{v_\perp r_L}{B}\frac{1}{2}\frac{\partial B}{\partial y}\hat{x}$$

where we have used the formula shown previously. Since the choice of the y axis was arbitrary, this can be generalized to

$$\mathbf{v}_{\nabla B} = \pm\frac{1}{2}v_\perp r_L\frac{\mathbf{B}\times\nabla B}{B^2}$$

This has all the dependences we expected from the physical picture; only the factor $\frac{1}{2}$ (arising from the averaging) was not predicted. Note that the $\pm$ stands for the sign of the charge, and lightface B stands for $|B|$. The quantity $\mathbf{v}_{\nabla B}$ is called the *grad-B drift*; it is in opposite directions for ions and electrons and causes a current transverse to **B**. An exact calculation of $\mathbf{v}_{\nabla B}$ would require using the exact orbit, including the drift, in the averaging process.

4.2.2 Curved B: Curvature Drift

Here we assume the magnetic field lines to be curved with a constant radius of curvature R_c, and we take $|B|$ to be constant. Such a field does not obey Maxwell's equations in a vacuum, so in practice the grad-B drift will always be added to the effect derived here. A guiding center drift arises from the centrifugal force felt by the particles as they move along the field lines in their thermal motion. If $v_{\|}^2$ denotes the average square of the component of random velocity along $\mathbf{B}$, and $\mathbf{R}_c$ denotes the radius of curvature vector from the center to the curve, the average centrifugal force is

$$\mathbf{F}_{cf} = \frac{mv_{\|}^2}{R_c}\hat{r} = mv_{\|}^2 \frac{\mathbf{R}_c}{R_c^2}$$

According to the guiding center drift formula, this gives rise to a drift

$$\mathbf{v}_R = \frac{1}{q}\frac{\mathbf{F}_{cf} \times \mathbf{B}}{B^2} = \frac{mv_{\|}^2}{qB^2}\frac{\mathbf{R}_c \times \mathbf{B}}{R_c^2}$$

The drift $\mathbf{v}_R$ is called the *curvature drift*.

We must now compute the grad-B drift which accompanies this when the decrease of $|B|$ with radius is taken into account. In a vacuum, we have $\nabla \times \mathbf{B} = 0$. In the cylindrical coordinates, $\nabla \times \mathbf{B}$ only has a z component, since $\mathbf{B}$ has only a θ component and ∇B only an r component. We then have

$$(\nabla \times \mathbf{B})_z = \frac{1}{r}\frac{\partial}{\partial r}(rB_\theta) = 0,\ B \propto \frac{1}{r}$$

Thus

$$|B| \propto \frac{1}{R_c},\ \frac{\nabla B}{B} = -\frac{\mathbf{R}_c}{R_c^2}$$

Using the expression of the grad-B drift, we have

$$\mathbf{v}_{\nabla B} = \mp\frac{1}{2}\frac{v_\perp r_L}{B^2}\mathbf{B} \times \frac{\mathbf{R}_c}{R_c^2} = \pm\frac{1}{2}\frac{v_\perp^2}{\omega_c}\frac{\mathbf{R}_c \times \mathbf{B}}{R_c^2 B} = \frac{1}{2}\frac{m}{q}v_\perp^2\frac{\mathbf{R}_c \times \mathbf{B}}{R_c^2 B^2}$$

Adding this to $\mathbf{v}_R$, we have the total drift in a curved vacuum field:

$$\mathbf{v}_R + \mathbf{v}_{\nabla B} = \frac{m}{q}\frac{\mathbf{R}_c \times \mathbf{B}}{R_c^2 B^2}\left(v_{\|}^2 + \frac{1}{2}v_\perp^2\right)$$

It is unfortunate that these drifts add. This means that if one bends a magnetic field into a torus for the purpose of confining a thermonuclear plasma, the particles will drift out of the torus no matter how one juggles the temperatures and magnetic fields.

For a Maxwellian distribution, $\bar{v_\parallel^2}$ and $\frac{1}{2}\bar{v_\perp^2}$ are each equal to k_BT/m, since $v_\perp$ involves two degrees of freedom. Then the average curved-field drift can be written as

$$\bar{\mathbf{v}}_{R+\nabla B} = \pm\frac{v_{th}^2}{R_c\omega_c}\hat{y} = \pm\frac{\bar{r}_L}{R_c}v_{th}\hat{y}$$

where $\hat{y}$ here is the direction of $\widehat{R}_c \times \mathbf{B}$. This shows that $\bar{\mathbf{v}}_{R+\nabla B}$ depends on the charge of the species but not on its mass.

4.2.3 B B: Magnetic Mirrors

Now we consider a magnetic field which is pointed primarily in the z direction and whose magnitude varies in the z direction. Let the field be axisymmetric, with $B_\theta = 0$ and $\partial/\partial\theta = 0$. Since the magnetic field lines converge and diverge, there is necessarily a component B_r. We wish to show that this gives rise to a force which can trap a particle in a magnetic field.

We can obtain B_r from $\nabla\cdot\mathbf{B} = 0$:

$$\frac{1}{r}\frac{\partial}{\partial r}(rB_r) + \frac{\partial B_z}{\partial z} = 0$$

If $\partial\mathbf{B}_z/\partial z$ is given at $r = 0$ and does not vary much with r, we have approximately

$$\begin{aligned} rB_r &= -\int_0^r r\frac{\partial B_z}{\partial z}dr \simeq -\frac{1}{2}r^2\left[\frac{\partial\mathbf{B}_z}{\partial z}\right]_{r=0} \\ B_r &= -\frac{1}{2}r\left[\frac{\partial\mathbf{B}_z}{\partial z}\right]_{r=0} \end{aligned}$$

The variation of $|B|$ with r causes a grad-B drift of guiding centers about the axis of symmetry, but there is no radial grad-B drift, because $\partial B/\partial\theta = 0$. The components of the Lorentz force are

$$\begin{aligned} F_r &= q(\underbrace{v_\theta B_z}_{1} - v_z\cancel{B_\theta}) \\ F_\theta &= q(\underbrace{-v_rB_z}_{2} + \underbrace{v_zB_r}_{3}) \\ F_z &= q(v_r\cancel{B_\theta} - \underbrace{v_\theta B_r}_{4}) \end{aligned}$$

Two terms vanish if $B_\theta = 0$, and terms 1 and 2 give rise to the usual Larmor gyration. Term 3 vanishes on the axis; when it does not vanish, this azimuthal force causes a drift in the radial direction. This drift merely makes the guiding centers follow the magnetic field lines. Term 4 is the one we are interested in. Using the expression of B_r, we have

$$F_z = \frac{1}{2}qv_\theta r_L\frac{\partial B_z}{\partial z}$$

We must now average over one gyration. For simplicity, consider a particle whose guiding center lies on the axis. Then v_θ is a constant during a gyration; depending on the sign of q, v_θ is $\mp v_\perp$. Since $r = r_L$, the average force is

$$\bar{F}_z = \mp\frac{1}{2}qv_\perp r_L\frac{\partial B_z}{\partial z} = \mp\frac{1}{2}q\frac{v_\perp^2}{\omega_c}\frac{\partial B_z}{\partial z} = -\frac{1}{2}\frac{mv_\perp^2}{B}\frac{\partial B_z}{\partial z}$$

We define the *magnetic moment* of the gyrating particle to be

$$\mu \equiv \frac{1}{2}mv_\perp^2/B$$

so that

$$\bar{F}_z = -\mu\frac{\partial B_z}{\partial z}$$

This is a specific example of the force on a diamagnetic particle, which in general can be written

$$\mathbf{F}_\parallel = -\mu\frac{\partial B}{\partial \mathbf{s}} = -\mu\nabla_\parallel B$$

where d**s** is a line element along **B**. Note that the definition of magnetic moment here is the same as the usual definition for the magnetic moment of a current loop with area A and current I: $\mu = IA$. In the case of a singly charged ion, I is generated by a charge e coming around $\omega_c/2\pi$ times a second: $I = e\omega_c/2\pi$. The area A is $\pi r_L^2 = \pi r_L^2/\omega_c^2$. Thus

$$\mu = \frac{e\omega_c}{2\pi}\frac{\pi r_L^2}{\omega_c^2} = \frac{1}{2}\frac{v_\perp^2 e}{\omega_c} = \frac{1}{2}\frac{mv_\perp^2}{B}$$

As the particle moves into regions of stronger or weaker B, its Larmor radius changes, but remains *invariant*. To prove this, consider the component of the equation of motion along **B**:

$$m\frac{\mathrm{d}v_\parallel}{\mathrm{d}t} = -\mu\frac{\partial B}{\partial s}$$

Multiplying by $v_\parallel$ on the left and its equivalent $ds/\mathrm{d}t$ on the right, we have

$$mv_\parallel\frac{\mathrm{d}v_\parallel}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\Big(\frac{1}{2}mv_\parallel^2\Big) = -\mu\frac{\partial B}{\partial s}\frac{ds}{\mathrm{d}t} = -\mu\frac{\mathrm{d}B}{\mathrm{d}t}$$

Here $\mathrm{d}B/\mathrm{d}t$ is the variation of B as seen by the particle; B itself is constant. The particle's energy must be conserved, so we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big(\frac{1}{2}mv_\parallel^2 + \frac{1}{2}mv_\perp^2\Big) = \frac{\mathrm{d}}{\mathrm{d}t}\Big(\frac{1}{2}mv_\parallel^2 + \mu B\Big) = 0$$

With the previous equation this becomes

$$-\mu\frac{\mathrm{d}B}{\mathrm{d}t}+\frac{\mathrm{d}}{\mathrm{d}t}(\mu B)=0$$

so that

$$\frac{\mathrm{d}\mu}{\mathrm{d}t}=0$$

The invariance of μ is the basis for one of the primary schemes for plasma confinement: the *magnetic mirror*. As a particle moves from a weak-field region to a strong-field region in the course of its thermal motion, it sees an increasing B, and therefore its $v_\perp$ must increase in order to keep constant. Since its total energy must remain constant, $v_\parallel$ must necessarily decrease. If B is high enough in the "throat" of the mirror, $v_\parallel$ eventually becomes zero; and the particle is "reflected" back to the weak-field region. It is, of course, the force $\mathbf{F}_\parallel$ which causes the reflection. The nonuniform field of a simple pair of coil form two magnetic mirrors between which a plasma can be trapped. This effect works on both ions and electrons.

The trapping is not perfect, however. For instance, a particle with $v_\perp = 0$ will have no magnetic moment and will not feel any force along $\mathbf{B}$. A particle with small $v_\perp/v_\parallel$ at the midplane ($B = B_0$) will also escape if the maximum field B_m is not large enough. For given B_0 and B_m, which particles will escape? A particle with $v_\perp = v_{\perp 0}$ and $v_\parallel = v_{\parallel 0}$ at the midplane will have $v_\perp = v'_\perp$ and $v_\parallel = 0$ at its turning point. Let the field be B' there. Then the invariants of μ yields

$$\frac{1}{2}\frac{mv_{\perp 0}^2}{B_0}=\frac{1}{2}\frac{m{v'_{\perp 0}}^2}{B'}$$

Conservation of energy requires

$${v'_{\perp 0}}^2=v_{\perp 0}^2+v_{\parallel 0}^2\equiv v_0^2$$

Combining the above two equations, we find

$$\frac{B_0}{B'}=\frac{v_{\perp 0}^2}{{v'_\perp}^2}\equiv\sin^2\theta$$

where θ is the pitch angle of the orbit in the weak-field region. Particles with smaller will mirror in regions of higher B. If is too small, B' exceeds B_m; and the particles does not mirror at all. Replacing B' by B_m, we see that the smallest θ of a confined particle is given by

$$\sin^2\theta_m=\frac{B_0}{B_m}\equiv\frac{1}{R_m}$$

where R_m is the *mirror ratio*. It defines the boundary of a region in velocity space in the shape of a cone, called a *loss cone*. Particles lying within the loss cone are not confined. Consequently, a mirror-confined plasma is never isotropic. Note that the loss cone is independent of q or

m. Without collisions, both ions and electrons are equally well confined. When collisions occur, particles are lost when they change their pitch angle in a collision and are scattered into the loss cone. Generally, electrons are lost more easily because they have a higher collision frequency.

The magnetic mirror was first proposed by Enrico Fermi as a mechanism for the acceleration of cosmic rays. Protons bouncing between magnetic mirrors approaching each other at high velocity could gain energy at each bounce. How such mirrors could arise is another story. A further example of the mirror effect is the confinement of particles in the Van Allen belts. The magnetic field of the earth, being strong at the poles and weak at the equator, forms a natural mirror with rather large R_m.

4.3 Nonuniform E Field

Now we let the magnetic field be uniform and the electric field be nonuniform. For simplicity, we assume $\mathbf{E}$ to be in the x direction and to vary sinusoidally in the x direction:

$$\mathbf{E} \equiv E_0(\cos kx)\hat{x}$$

This field distribution has a wavelength $\lambda = 2\pi/k$, and is the result of a sinusoidal distribution of charges, which we need not specify. In practice, such a charge distribution can arise in a plasma during a wave motion. The equation of motion is

$$m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = q[\mathbf{E}(x) + \mathbf{v} \times \mathbf{B}]$$

whose transverse components are

$$\begin{aligned} \dot{v}_x &= \frac{qB}{m}v_y + \frac{q}{m}E_x(x) \\ \dot{v}_y &= -\frac{qB}{m}v_x \end{aligned}$$

$$\begin{aligned} \ddot{v}_x &= -\omega_c^2 v_x \pm \omega_c \frac{\dot{E}_x}{B} \\ \ddot{v}_y &= -\omega_c^2 v_y - \omega_c^2 \frac{E_x(x)}{B} \end{aligned}$$

Here $E_x(x)$ is the electric field at the position of the particle. To evaluate this, we need to know the particle's orbit, which we are trying to solve for in the first place. If the electric field is weak, we may, as an approximation, use the *undisturbed orbit* to evaluate $E_x(x)$. The orbit in the absence of the E field was given before

$$x = x_0 + r_L \sin\omega_c t$$

so we have

$$\ddot{v}_y = -\omega_c^2 v_y - \omega_c^2 \frac{E_0}{B} \cos k(x_0 + r_L \sin \omega_c t)$$

Anticipating the result, we look for a solution which is the sum of a gyration at ω_c and a steady drift $\mathbf{v}_E$. Since we are interested in finding an expression for v_E, we take out the gyratory motion by averaging over a cycle. The v_x component then gives $\bar{v}_x = 0$.[1] In the v_y component, the oscillating term $\ddot{v}_y$ clearly averages to zero, and we have

$$\bar{\ddot{v}}_y = 0 = -\omega_c^2 \bar{v}_y - \omega_c^2 \frac{E_0}{B} \overline{\cos k(x_0 + r_L \sin \omega_c t)}$$

Expanding the cosine, we have

$$\cos k(x_0 + r_L \sin \omega_c t) = \cos(kx_0) \cos(kr_L \sin \omega_c t) - \sin(kx_0) \sin(kr_L \sin \omega_c t)$$

It will suffice to treat the small Larmor radius case, $kr_L \ll 1$. The Taylor expansions allow us to write

$$\cos k(x_0 + r_L \sin \omega_c t) \approx \cos(kx_0)(1 - \frac{1}{2}k^2 r_L^2 \sin^2 \omega_c t) - \sin(kx_0) kr_L \sin \omega_c t$$

The last term vanishes upon averaging over time, and it gives

$$\bar{v}_y = -\frac{E_0}{B} \cos(kx_0)(1 - \frac{1}{4}k^2 r_L^2) = -\frac{E_x(x_0)}{B}(1 - \frac{1}{4}k^2 r_L^2)$$

Thus the usual $\mathbf{E} \times \mathbf{B}$ drift is modified by the inhomogeneity to read

$$\mathbf{v}_E = \frac{\mathbf{E} \times \mathbf{B}}{B^2}(1 - \frac{1}{4}k^2 r_L^2)$$

The physical reason for this is easy to see. An ion with its guiding center at a maximum of $\mathbf{E}$ actually spends a good deal of its time in regions of weaker $\mathbf{E}$. Its average drift, therefore, is less than E/B evaluated at the guiding center. In a linearly varying $\mathbf{E}$ field, the ion would be in a stronger field on one side of the orbit and in a field weaker by the same amount on the other side; the correction to $\mathbf{v}_E$ then cancels out. From this it is clear that the correction term depends on the *second derivative* of $\mathbf{E}$. For the sinusoidal distribution we assumed, the second derivative is always negative with respect to $\mathbf{E}$. For an arbitrary variation of $\mathbf{E}$, we need only replace ik by ∇ and write the drift as

$$\mathbf{v}_E = \left(1 + \frac{1}{4} r_L^2 \nabla^2\right) \frac{\mathbf{E} \times \mathbf{B}}{B^2} \tag{4.3}$$

[1] I am kind of lost here. Probably the logic here is that first we assume the solution has an additional steady drift, so we let $\ddot{v}_y = 0$? Check Y.Y's note.

The second term is called the *finite-Larmor-radius effect.* What is the significance of this correction? Since r_L is much larger for ions than for electrons, $\mathbf{v}_E$ is no longer independent of species. If a density clump occurs in a plasma, an electric field can cause the ions and electrons to separate, generating another electric field. If there is a feedback mechanism that causes the second electric field to enhance the first one, $\mathbf{E}$ grows indefinitely, and the plasma is unstable. Such an instability, called a *drift instability*, is one type of plasma instabilities. The grad-B drift, of course, is also a finite-Larmor-radius effect and also causes charges to separate. However, $\mathbf{v}_{\nabla B} \propto k r_L$ whereas the correction term above is proportional to $k^2 r_L^2$. The nonuniform-E-field effect, therefore, is important at relatively large k, or small scale lengths of the inhomogeneity. For this reason, drift instabilities belong to a more general class called *microinstabilities.*

4.4 Time-Varying E Field

Let us now take $\mathbf{E}$ and $\mathbf{B}$ to be uniform in space but varying in time. First, consider the case in which $\mathbf{E}$ alone varies sinusoidally in time, and let it lie along the x axis:

$$\mathbf{E} = E_0 e^{i\omega t}\hat{x}$$

Since $\dot{E}_x = i\omega E_x$, we can write the velocity x-component equation as

$$\ddot{v}_x = -\omega_c^2\Big(v_x \mp \frac{i\omega}{\omega_c}\frac{\tilde{E}_x}{B}\Big)$$

Let us define

$$\tilde{v}_p \equiv \pm\frac{i\omega}{\omega_c}\frac{\tilde{E}_x}{B}$$
$$\tilde{v}_E \equiv \frac{\tilde{E}_x}{B}$$

where the tilde has been added merely to emphasize that the drift is oscillating. The plus (minus) sign, as usual, denotes positive (negative) q. Now the x and y velocity component equations can be written as

$$\ddot{v}_x = -\omega_c^2(v_x - \tilde{v}_p)$$
$$\ddot{v}_y = -\omega_c^2(v_y - \tilde{v}_E)$$

By analogy with the derivation in the case of a uniform EM field, we try a solution which is the sum of a drift and a gyratory motion:

$$v_x = v_\perp e^{i\omega_c t} + \tilde{v}_p$$
$$v_y = \pm i v_\perp e^{i\omega_c t} + \tilde{v}_E$$

If we now differentiate twice with respect to time, we find

$$\ddot{v}_x = -\omega_c^2 v_x + (\omega_c^2 - \omega^2)\tilde{v}_p$$
$$\ddot{v}_y = -\omega_c^2 v_y + (\omega_c^2 - \omega^2)\tilde{v}_E$$

This is **not** the same as the previous expressions right above unless $\omega^2 \ll \omega_c^2$. If we now make the assumption that **E** varies slowly, so that $\omega^2 \ll \omega_c^2$, then there we have the approximate solution.

The solution of velocities in x and y tells us that the guiding center motion has two components. The y component, perpendicular to **B** and **E**, is the usual $\mathbf{E}\times\mathbf{B}$ drift, except that v_E now oscillates slowly at the frequency ω. The x component, a new drift *along the direction of* **E**, is called the *polarization drift*. By replacing $i\omega$ with $\partial/\partial t$, we can generalize the expression of v_p and define the polarization drift as

$$\mathbf{v}_p = \pm\frac{1}{\omega_c B}\frac{\mathrm{d}\mathbf{E}}{\mathrm{d}t}$$

Since $\mathbf{v}_p$ is in opposite directions for ions and electrons, there is a *polarization current*; for Z = 1, this is

$$\mathbf{j}_p = ne(v_{ip} - v_{ep}) = \frac{ne}{eB^2}(M+m)\frac{\mathrm{d}\mathbf{E}}{\mathrm{d}t} = \frac{\rho}{B^2}\frac{\mathrm{d}\mathbf{E}}{\mathrm{d}t} \tag{4.4}$$

where ρ is the mass density.

The physical reason for the polarization current is simple. Consider an ion at rest in a magnetic field. If a field **E** is suddenly applied, the first thing the ion does is to move in the direction of **E**. Only after picking up a velocity **v** does the ion feel a Lorentz force $e\mathbf{v}\times\mathbf{B}$ and begin to move perpendicular to both fields. If **E** is now kept constant, there is no further $\mathbf{v}_p$ drift but only a $\mathbf{v}_E$ drift. However, if **E** is reversed, there is again a momentary drift, this time to the left. Thus $\mathbf{v}_p$ is a startup drift due to inertia and occurs only in the first half-cycle of each gyration during which **E** changes. Consequently, $\mathbf{v}_p$ goes to zero when $\omega/\omega_c \ll 1$.

The polarization effect in a plasma is similar to that in a solid dielectric, where $\mathbf{D} = \epsilon_0\mathbf{E} + \mathbf{P}$. The dipoles in a plasma are ions and electrons separated by a distance r_L. But since ions and electrons can move around to preserve quasineutrality, the application of a steady **E** field does not result in a polarization field **P**. However, if **E** oscillates, an oscillating current $\mathbf{j}_p$ results from the lag due to the ion inertia.

It is obvious that in this case the $\mathbf{E}\times\mathbf{B}$ drift speed depends on time and the guiding center coordinate system is non-inertial. The polarization drift arises from the fictuous (inertial) force $-m\mathrm{d}\mathbf{v}_E/\mathrm{d}t$. It means that the guiding center locations will not change due to the polarization drift!

Let us demonstrate this in a single configuration. Assume $\mathbf{B} = B\hat{z}$, $\mathbf{E} = Et\hat{y}$. Let θ be the angle in the perpendicular plane:

$$\mathbf{v}_\perp = v_\perp \hat{e}_\perp = v_\perp (\cos\theta\hat{x} + \sin\theta\hat{y})$$

The guiding center is defined as

$$\mathbf{X} = \mathbf{x} - \mathbf{r}_L = \mathbf{x} - \mathbf{v}_\perp/\omega_c = \mathbf{x} - \frac{v_\perp}{\omega_c}\hat{\theta}$$

where $\hat{\theta} = \mathbf{b} \times \hat{e}_\perp$. The equation of motion

$$m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$
$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{v}$$

in the guiding center coordinates the two selected scalar equations can be written as

$$m\frac{\mathrm{d}v_x}{\mathrm{d}t} = qE_x + qv_y B$$
$$\frac{dX_y}{\mathrm{d}t} = v_y - \frac{1}{\omega_c}\frac{\mathrm{d}v_x}{\mathrm{d}t}$$

such that

$$\frac{dX_y}{\mathrm{d}t} = 0$$

4.5 Time-Varying B Field

Finally, we allow the magnetic field to vary in time. Since the Lorentz force is always perpendicular to $\mathbf{v}$, a magnetic field itself cannot impart energy to a charged particle. However, associated with $\mathbf{B}$ is an electric field given by

$$\nabla \times \mathbf{E} = -\dot{\mathbf{B}}$$

and this can accelerate the particles. We can no longer assume the fields to be completely uniform. Let $\mathbf{v}_\perp = \mathrm{d}\mathbf{l}/\mathrm{d}t$ be the transverse velocity, $\mathbf{l}$ being the element of the path along a particle trajectory (with $v_\parallel$ neglected). Taking the scalar product of the equation of motion with $\mathbf{v}_\perp$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big(\frac{1}{2}mv_\perp^2\Big) = q\mathbf{E}\cdot\mathbf{v}_\perp = q\mathbf{E}\cdot\frac{\mathrm{d}\mathbf{l}}{\mathrm{d}t}$$

The change in one gyration is obtained by integrating over one period:

$$\delta\Big(\frac{1}{2}mv_\perp^2\Big) = \int_0^{2\pi/\omega_c} q\mathbf{E}\cdot\frac{\mathrm{d}\mathbf{l}}{\mathrm{d}t}\mathrm{d}t$$

If the field changes slowly, we can replace the time integral by a line integral over the unperturbed orbit:

$$\delta\Big(\frac{1}{2}mv_\perp^2\Big) = \oint q\mathbf{E}\cdot \mathrm{d}\mathbf{l} = q\int_s (\nabla\times\mathbf{E})\cdot \mathrm{d}\mathbf{S} = -q\int_s \dot{\mathbf{B}}\cdot \mathrm{d}\mathbf{S}$$

Here $\mathbf{S}$ is the surface enclosed by the Larmor orbit and has a direction given by the right-hand rule when the fingers point in the direction of $\mathbf{v}$. Since the plasma is diamagnetic, we have $\mathbf{B}\cdot \mathrm{d}\mathbf{S} < 0$ for ions and $\mathbf{B}\cdot \mathrm{d}\mathbf{S} > 0$ for electrons. Then

$$\delta\Big(\frac{1}{2}mv_\perp^2\Big) = \pm q\dot{B}\pi r_L^2 = \pm q\pi\dot{B}]fracv_\perp^2\omega_c\frac{m}{\pm qB} = \frac{\frac{1}{2}mv_\perp^2}{B}\cdot\frac{2\pi\dot{B}}{\omega_c}$$

The quantity $\frac{2\pi\dot{B}}{\omega_c} = \frac{\dot{B}}{\omega_c}$ is just the change δB during one period of gyration. Thus

$$\delta\Big(\frac{1}{2}mv_\perp^2\Big) = \mu\delta B$$

Since the left-hand-side is $\delta(\mu B)$, we have the desired result

$$\delta\mu = 0$$

The magnetic moment is invariant in slowly varying magnetic fields.

As the B field varies in strength, the Larmor orbits expand and contract, and the particles lose and gain transverse energy. This exchange of energy between the particles and the field is described very simply by the invariant of magnetic moment. The invariance of μ allows us to prove easily the following well-known theorem: *The magnetic flux through a Larmor orbit is constant.*

The flux Φ is given by BS, with $S = \pi r_L^2$. Thus

$$\Phi = B\pi\frac{v_\perp^2}{\omega_c^2} = B\pi\frac{v_\perp^2 m^2}{q^2B^2} = \frac{2\pi m}{q^2}\frac{\frac{1}{2}mv_\perp^2}{B} = \frac{2\pi m}{q^2}\mu$$

Therefore, Φ is constant if μ is constant.

This property is used in a method of plasma heating known as *adiabatic compression*. A plasma is injected into a sequence of magnetic mirrors and by keep increasing the magnetic field in subsequent mirrors we can increase the plasma velocities.

4.6 Summary of Guiding Center Drifts

General force:

$$\mathbf{v}_f = \frac{1}{q}\frac{\mathbf{F}\times\mathbf{B}}{B^2}$$

Electric field:

$$\mathbf{v}_E = \frac{\mathbf{E}\times\mathbf{B}}{B^2}$$

Gravitational field:

$$\mathbf{v}_g = \frac{m}{q}\frac{\mathbf{g}\times\mathbf{B}}{B^2}$$

Nonuniform electric field:

$$\mathbf{v}_E = \Big(1+\frac{1}{4}r_L^2\nabla^2\Big)\frac{\mathbf{E}\times\mathbf{B}}{B^2}$$

Nonuniform magnetic field:

Grad-B:

$$\mathbf{v}_{\nabla B} = \pm\frac{1}{2}v_\perp r_L\frac{\mathbf{B}\times\nabla B}{B^2}$$

Curvature drift:

$$\mathbf{v}_R = \frac{mv_\parallel^2}{q}\frac{\mathbf{R}_c\times\mathbf{B}}{R_c^2B^2}$$

Curved vacuum field:

$$\mathbf{v}_R+\mathbf{v}_{\nabla B} = \frac{m}{q}\Big(v_\parallel^2+\frac{1}{2}v_\perp^2\Big)\frac{\mathbf{R}_c\times\mathbf{B}}{R_c^2B^2}$$

Polarization drift:

$$\mathbf{v}_p = \pm\frac{1}{\omega_c B}\frac{\mathrm{d}\mathbf{E}}{\mathrm{d}t}$$

A non-relativistic equation for the guiding center is given by Northrop in 1963:

$$\frac{\mathrm{d}\mathbf{R}}{\mathrm{d}t} = v_\parallel\hat{b}+\frac{\hat{b}}{B}\times\Big[-\mathbf{E}+\frac{m}{q}\Big(v_\parallel\frac{\mathrm{d}\hat{b}}{\mathrm{d}t}+\frac{\mathrm{d}\mathbf{v}_{E\times B}}{\mathrm{d}t}\Big)+\frac{\mu}{q}\nabla B\Big]$$

and the parallel motion can be described by (???)

$$\frac{\mathrm{d}(mv_\parallel)}{\mathrm{d}t} = m\mathbf{u}_q E_\parallel - \mu\hat{b}\cdot\nabla B$$

By applying the guiding center approximation, we reduce the number of independent variables from 6 (x,y,z,v_x,v_y,v_z) to 5 $(x,y,z,v_\perp,v_\parallel)$.

4.7 Adiabatic Invariants

It is well known in classical mechanics that whenever a system has a periodic motion, the action integral $\oint pdq$ taken over a period is a constant of the motion. Here p and q are the generalized momentum and coordinate which repeat themselves in the motion. If a slow change is made in the system, so that the motion is not quite periodic, the constant of the motion does not change and is then called an *adiabatic invariant*. By slow here we mean slow compared with the period of motion, so that the integral $\oint pdq$ is well defined even though it is strictly no longer an integral over a closed path. Adiabatic invariants play an important role in plasma physics; they allow us to obtain simple answers in many instances involving complicated motions. There are three adiabatic invariants, each corresponding to a different type of periodic motion.

4.7.1 Adiabatic Invariant of a Pendulum

Perfect symmetry is never attained in reality. This leads to the practical question of how constants of the motion behave when space and/or time symmetries are "good", but not perfect. Does the utility of constants of the motion collapse abruptly when the slightest non-symmetrical blemish rears its ugly head, does the utility decay gracefully, or does something completely different happen? To answer these questions, we begin by considering the problem of a small-amplitude pendulum having a time-dependent, but slowly changing resonant frequency $\omega(t)$. Since $\omega^2 = g/l$, the time-dependence of the frequency might result from either a slow change in the gravitational acceleration g or else from a slow change in the pendulum length l. In both cases the pendulum equation of motion will be

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + \omega^2(t)x = 0 \tag{4.5}$$

This equation cannot be solved exactly for arbitrary $\omega(t)$ but if a modest restriction is put on $\omega(t)$ the equation can be solved *approximately* using the WKB method (Wentzel 1926, Kramers 1926, Brillouin 1926). This method is based on the hypothesis that the solution for a time-dependent frequency is likely to be a generalization of the constant-frequency solution

$$x = \mathrm{Re}\left[A\exp(i\omega t)\right]$$

where this generalization is postulated to be of the form

$$x = \mathrm{Re}\left[A(t)e^{i\int^t \omega(t')\mathrm{d}t'}\right] \tag{4.6}$$

Here $A(t)$ is an amplitude function determined as follows: calculate the first derivative we get

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \mathrm{Re}\left[i\omega A e^{i\int^t \omega(t')\mathrm{d}t'} + \frac{\mathrm{d}A}{\mathrm{d}t}e^{i\int^t \omega(t')\mathrm{d}t'}\right]$$

then the second derivative

$$\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} = \mathrm{Re}\left[\left(i\frac{\mathrm{d}\omega}{\mathrm{d}t}A + 2i\omega\frac{\mathrm{d}A}{\mathrm{d}t} - \omega^2 A + \frac{\mathrm{d}^2 A}{\mathrm{d}t^2}\right) e^{i\int^t \omega(t')\mathrm{d}t'}\right]$$

and insert this last result into Equation 4.5 which reduces to

$$i\frac{\mathrm{d}\omega}{\mathrm{d}t}A + 2i\omega\frac{\mathrm{d}A}{\mathrm{d}t} + \frac{\mathrm{d}^2 A}{\mathrm{d}t^2} = 0 \tag{4.7}$$

since the terms involving ω^2 cancel exactly. To proceed further, we make an assumption — the validity of which is to be checked later — that the time dependence of $\mathrm{d}A/\mathrm{d}t$ is *sufficiently* slow to allow dropping the last term in Equation 4.7 relative to the middle term. The two terms that remain in Equation 4.7 can then be rearranged as

$$\frac{1}{\omega}\frac{\mathrm{d}\omega}{\mathrm{d}t} = -\frac{2}{A}\frac{\mathrm{d}A}{\mathrm{d}t}$$

which has the exact solution

$$A(t) \sim \frac{1}{\sqrt{\omega(t)}} \tag{4.8}$$

The assumption of slowness is thus at least self-consistent, for if t is indeed slowly changing, Equation 4.8 shows that At will also be slowly changing and the dropping of the last term in Equation 4.7 is justified. The slowness requirement can be quantified by assuming that the frequency has an exponential dependence

$$\omega(t) = \omega_0 e^{\alpha t}$$

Thus,

$$\alpha = \frac{1}{\omega}\frac{\mathrm{d}\omega}{\mathrm{d}t}$$

is a measure of how fast the frequency is changing compared to the frequency itself. Hence, dropping the last term in Equation 4.7 is legitimate if

$$\alpha = \frac{1}{\omega}\frac{\mathrm{d}\omega}{\mathrm{d}t} \ll 4\omega_0 \tag{4.9}$$

In other words, if Equation 4.9 is satisfied, then the fractional change of the pendulum period per period is small.

Equation 4.8 indicates that when ω is time-dependent, the pendulum amplitude is not constant and so the pendulum energy is *not* conserved. It turns out that what is conserved is the *action integral*

$$S = \oint v\mathrm{d}x \tag{4.10}$$

where the integration is over one period of oscillation. This integral can also be written in terms of time as

$$S = \int_{t_0}^{t_0+\tau} v\frac{\mathrm{d}x}{\mathrm{d}t}\mathrm{d}t$$

where t_0 is a time when x is at an instantaneous maximum and τ, the period of a complete cycle, is defined as the interval between two successive times when $\mathrm{d}x/\mathrm{d}t = 0$ and $\mathrm{d}^2/\mathrm{d}t^2$ has the same sign (e.g., for a pendulum, t_0 would be a time when the pendulum has swung all the way to the right and so is reversing its velocity while τ is the time one has to wait for this to happen again). To show that action is conserved, Equation 4.10 can be integrated by parts as

$$\begin{aligned} S &= \int_{t_0}^{t_0+\tau} \left[\frac{\mathrm{d}}{\mathrm{d}t}\left(x\frac{\mathrm{d}x}{\mathrm{d}t}\right) - x\frac{\mathrm{d}^2x}{\mathrm{d}t^2}\right]\mathrm{d}t \\ &= \left[x\frac{\mathrm{d}x}{\mathrm{d}t}\right]_{t_0}^{t_0+\tau} - \int_{t_0}^{t_0+\tau} x\frac{\mathrm{d}^2x}{\mathrm{d}t^2}\mathrm{d}t \\ &= \int_{t_0}^{t_0+\tau} \omega^2x^2\mathrm{d}t \end{aligned} \tag{4.11}$$

where the integrated term has vanished by virtue of the definitions of t_0 and τ, and Equation 4.5 has been used to substitute $\mathrm{d}^2x/\mathrm{d}t^2$. Equation 4.6 and Equation 4.8 can be combined to give

$$x(t) = x(t_0)\sqrt{\frac{\omega(t_0)}{\omega(t)}}\cos\left(\int_{t_0}^{t}\omega(t')\mathrm{d}t'\right)$$

so Equation 4.11 becomes

$$\begin{aligned} S &= \int_{t_0}^{t_0+\tau} \omega(t')^2\left[x(t_0)\sqrt{\frac{\omega(t_0)}{\omega(t')}}\cos\left(\int_{t_0}^{t'}\omega(t'')\mathrm{d}t''\right)\right]\mathrm{d}t' \\ &= [x(t_0)]^2\omega(t_0)\int_{t_0}^{t_0+\tau}\omega(t')\cos^2\left(\int_{t_0}^{t'}\omega(t''\mathrm{d}t')\right)\mathrm{d}t' \\ &= [x(t_0)]^2\omega(t_0)\int_0^{2\pi}\mathrm{d}\xi\cos^2\xi = \pi[x(t_0)]^2\omega(t_0) = \text{const.} \end{aligned} \tag{4.12}$$

where $\xi = \int_{t_0}^{t'}\omega(t'')\mathrm{d}t''$ and $\mathrm{d}\xi = \omega(t')\mathrm{d}t'$. Equation 4.10 shows that S is the area in phase-space enclosed by the trajectory $[x(t), v(t)]$ and Equation 4.12 shows that for a slowly changing pendulum frequency, *this area is a constant of the motion*. Since the average energy of the pendulum scales as $\sim [\omega(t)x(t)]^2$, we see from Equation 4.8 that the ratio

$$\frac{\text{energy}}{\text{frequency}} \sim \omega(t)x^2(t) \sim S \sim \text{const.} \tag{4.13}$$

The ratio in Equation 4.13 is the classical equivalent of the quantum number N of a simple harmonic oscillator because in quantum mechanics the energy E of a simple harmonic oscillator is related to the frequency by the relation $E/\hbar\omega = N + 1/2$.

This analysis clearly applies to any dynamical system having an equation of motion of the form of Equation 4.5. Hence, if the dynamics of plasma particles happens to be of this form, then S can be added to our repertoire of constants of the motion.

4.7.2 Extension of WKB method to general adiabatic invariant

Action has the dimensions of (canonical momentum) $\times$ (canonical coordinate) so we may anticipate that for general Hamiltonian systems, the action integral given in Equation 4.10 is not an invariant because v is not, in general, proportional to P. We postulate that the general form for the action integral is

$$S = \oint P \mathrm{d}Q \tag{4.14}$$

where the integral is over one period of the periodic motion and P, Q are the relevant canonical momentum-coordinate conjugate pair. The proof of adiabatic invariance used for Equation 4.10 does not work directly for Equation 4.14; we now present a slightly more involved proof to show that Equation 4.14 is indeed the more general form of adiabatic invariant.

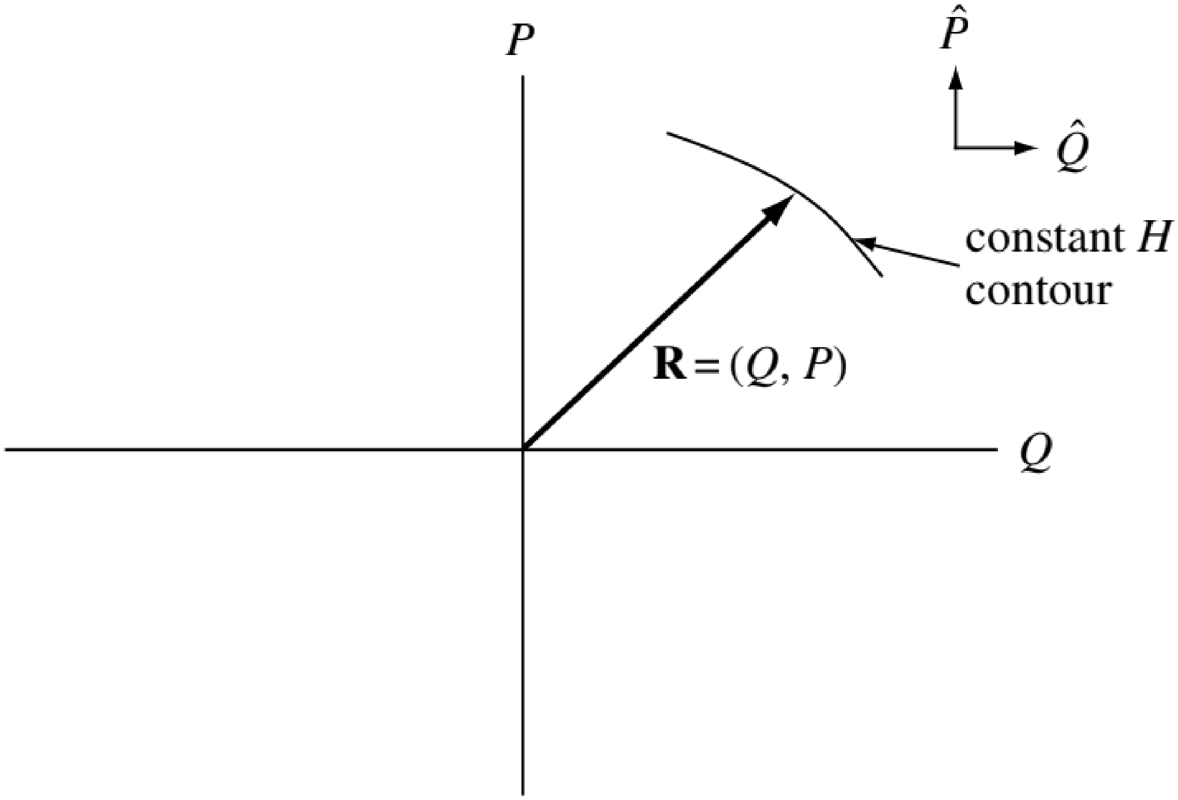

Figure 4.1: Canonical coordinate-momentum plane.

Let us define the radius vector in the $Q-P$ plane to be $\mathbf{R} = (Q, P)$ and define unit vectors in the Q and P directions by $\hat{Q}$ and $\hat{P}$; these definitions are shown in Figure 4.1. Furthermore, we define the z direction as being normal to the $Q-P$ plane; thus, the unit vector $\hat{z} = \hat{Q} \times \hat{P}$. Hamilton's equations ($\dot{P} = -\partial H/\partial Q, \dot{Q} = \partial H/\partial P$) may be written in vector form as

$$\frac{\mathrm{d}\mathbf{R}}{\mathrm{d}t} = -\hat{z} \times \nabla H \tag{4.15}$$

where

$$\nabla = \hat{Q}\frac{\partial}{\partial Q} + \hat{P}\frac{\partial}{\partial P}$$

is the gradient operator in the $Q-P$ plane. Equation 4.15 shows that the phase-space "velocity" $\mathrm{d}\mathbf{R}/\mathrm{d}t$ is *orthogonal* ot ∇H. Hence, $\mathbf{R}$ *stays on a level contour of* H. If H is a constant, then, in order for the motion to be periodic, the path along this level contour must circle around and join itself, like a road of constant elevation around the rim of a mountain (or a crater). If H is not constant, but slowly changing in time, the contour will circle around and nearly join itself.

Equation 4.15 can be inverted by crossing it with $\hat{z}$ to give

$$\nabla H = \hat{z} \times \frac{\mathrm{d}\mathbf{R}}{\mathrm{d}t} \tag{4.16}$$

For periodic and near-periodic motions, $\mathrm{d}\mathbf{R}/\mathrm{d}t$ is always in the same sense (always clockwise or always counterclockwise). Thus, Equation 4.16 shows that an observer following the path would always see that H is increasing on the left-hand side of the path and decreasing on the right-hand side (or vice versa). For clarity, the origin of the $Q-P$ plane is redefined to be at a local maximum or minimum of H. Hence, near the extremum H must have the Taylor expansion

$$H(P, Q) = H_{\text{extremum}} + \frac{P^2}{2}\left[\frac{\partial^2 H}{\partial P^2}\right]_{P=0,Q=0} + \frac{Q^2}{2}\left[\frac{\partial^2 H}{\partial Q^2}\right]_{P=0,Q=0}$$

where the second order derivatives are either both positive (valley) or both negative (hill). Since H is assumed to have a slow dependence on time, these second derivatives will be time-dependent so that H has the form

$$H = \alpha(t)\frac{P^2}{2} + \beta(t)\frac{Q^2}{2} \tag{4.17}$$

where $\alpha(t)$ and $\beta(t)$ have the same sign. The term H_{extremum} has been dropped because it is just an additive constant to the energy and does not affect Hamilton's equations. From Equation 4.15 the direction of rotation of $\mathbf{R}$ is seen to be counterclockwise if the extremum of H is a hill, and clockwise if a valley.

Hamilton's equations operating on Equation 4.17 give

$$\frac{\mathrm{d}P}{\mathrm{d}t} = -\beta Q, \quad \frac{\mathrm{d}Q}{\mathrm{d}t} = \alpha P \tag{4.18}$$

These equations do not directly generate the simple harmonic oscillator equation because of the time dependence of α, β. However, if we define the auxiliary variable

$$\tau = \int^t \beta(t')\mathrm{d}t'$$

then

$$\frac{\mathrm{d}}{\mathrm{d}t} = \frac{\mathrm{d}\tau}{\mathrm{d}t}\frac{\mathrm{d}}{\mathrm{d}\tau} = \beta\frac{\mathrm{d}}{\mathrm{d}\tau}$$

so Equation 4.18 becomes

$$\frac{\mathrm{d}P}{\mathrm{d}\tau} = -Q, \quad \frac{\mathrm{d}Q}{\mathrm{d}\tau} = \frac{\alpha}{\beta}P \tag{4.19}$$

Substituting for Q in the right-hand equation using the left-hand equation gives

$$\frac{\mathrm{d}^2P}{\mathrm{d}\tau^2} + \frac{\alpha}{\beta}P = 0$$

which is a simple harmonic oscillator with $\omega^2(\tau) = \alpha(\tau)/\beta(\tau)$. The action integral may be rewritten as

$$S = \oint P\frac{\mathrm{d}Q}{\mathrm{d}\tau}\mathrm{d}\tau$$

where the integral is over one period of the motion. Using Equation 4.19 and following the same procedure as was used with Equation 4.12, this becomes

$$S = \oint P^2\frac{\alpha}{\beta}\mathrm{d}\tau = \lambda^2\int\left[\left(\frac{\alpha(\tau')}{\beta(\tau')}\right)^{1/2}\cos^2\left(\int^{\tau'}(\alpha/\beta)^{1/2}\mathrm{d}\tau''\right)\right]\mathrm{d}\tau'$$

where λ is a constant dependent on initial conditions. By introducing the orbit phase $\phi = \int^\tau(\alpha/\beta)^{1/2}\mathrm{d}\tau$, the above becomes

$$S = \lambda^2\int_0^{2\pi}\mathrm{d}\phi\cos^2\phi = \text{const.}$$

Thus, the general action integral is indeed an adiabatic invariant. This proof is of course only valid in the vicinity of an extremum of H, i.e., only where H can be adequately represented by Equation 4.17.

4.7.3 Proof for the general adiabatic invariant

We now develop a proof for the general adiabatic invariant. This proof is not restricted to small oscillations (i.e., being near an extremum of H) as was the previous discussion. Let the

Hamiltonian depend on time via a slowly changing parameter $\lambda(t)$, so that $H = H(P, Q, \lambda(t))$. From Equation 3.27 the energy is given by

$$E(t) = H(P, Q, \lambda(t)) \tag{4.20}$$

and, in principle, this relation can be inverted to give $P = P(E(t), Q, \lambda(t))$. Suppose a particle is executing nearly periodic motion in the $Q-P$ plane. We define the turning point Q_{tp} as a position where $\mathrm{d}Q/\mathrm{d}t = 0$. Since Q is oscillating, there will be a turning point associated with Q having its maximum value and a turning point associated with Q having its minimum value. From now on let us only consider turning points where Q has its maximum value, that is, we only consider the turning points on the right-hand side of the nearly periodic trajectories in the $Q-P$ plane shown in Figure 4.2.

If the motion is periodic, then the turning point for the $N+1^{\mathrm{th}}$ period will be the same as the turning point for the N^{th} period, but if the motion is only nearly periodic, there will be a slight difference as shown in Figure 4.2. This difference can be characterized by making the turning point a function of time so $Q_{tp} = Q_{tp}(t)$. This function is only defined for the times when $\mathrm{d}Q/\mathrm{d}t = 0$. When the motion is not exactly periodic, this turning point is such that $Q_{tp}(t+\tau) \neq Q_{tp}(t)$, where τ is the time interval required for the particle to go from the first turning point to the next turning point. The action integral is over one entire period of oscillation starting from a right-hand turning point and then going to the next right-hand turning point and so can be written as

$$S = \oint P\mathrm{d}Q = \int_{Q_{tp}(t)}^{Q_{tp}(t+\tau)} P\mathrm{d}Q$$

From Equation 3.27 it is seen that P/m is not, in general, the velocity and so the velocity $\mathrm{d}Q/\mathrm{d}t$ is not, in general, proportional to P (???). Thus, the turning points are not necessarily at the locations where P vanishes, and in fact P need not change sign during a period. However, S still corresponds to the area of phase-space enclosed by one period of the phase-space trajectory.

We can now calculate

$$\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}t} &= \frac{\mathrm{d}}{\mathrm{d}t}\int_{Q_{tp}(t)}^{Q_{tp}(t+\tau)} P(E(t), Q, \lambda(t))\mathrm{d}Q \\
&= \left[P\frac{\mathrm{d}Q}{\mathrm{d}t}\right]_{Q_{tp}(t)}^{Q_{tp}(t+\tau)} + \int_{Q_{tp}(t)}^{Q_{tp}(t+\tau)} \left(\frac{\partial P}{\partial t}_Q\right)\mathrm{d}Q \\
&= \int_{Q_{tp}(t)}^{Q_{tp}(t+\tau)} \left[\left(\frac{\partial P}{\partial E}\right)_{Q,\lambda}\frac{\mathrm{d}E}{\mathrm{d}t} + \left(\frac{\partial P}{\partial \lambda}\right)_{Q,E}\frac{\mathrm{d}\lambda}{\mathrm{d}t}\right]\mathrm{d}Q
\end{aligned} \tag{4.21}$$

Because $\mathrm{d}Q/\mathrm{d}t = 0$ at the turning point, the integrated term vanishes and so there is no contribution from motion of the turning point. From Equation 4.20 we have

$$1 = \frac{\partial H}{\partial P}\left(\frac{\partial P}{\partial E}\right)_{Q,\lambda}$$

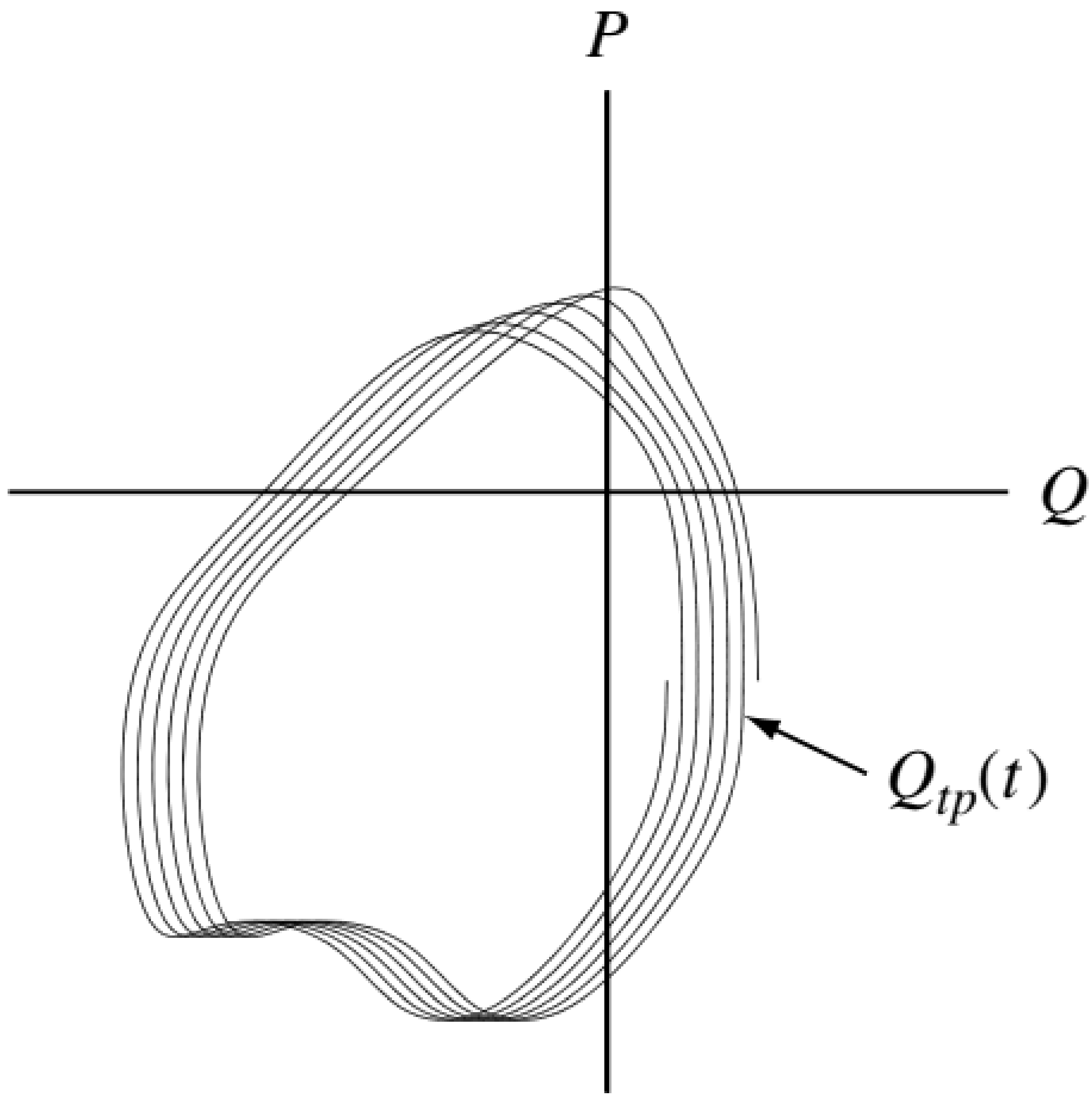

Figure 4.2: Nearly periodic-phase space trajectory for slowly changing Hamiltonian. The turning point $Q_{tp}(t)$ is where Q is at its maximum.

and

$$0 = \frac{\partial H}{\partial P}\left(\frac{\partial P}{\partial \lambda}\right)_{Q,E} + \frac{\partial H}{\partial \lambda}$$

so Equation 4.21 becomes

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \oint \left(\frac{\partial H}{\partial P}\right)^{-1}\left[\frac{\mathrm{d}E}{\mathrm{d}t} - \frac{\partial H}{\partial \lambda}\frac{\mathrm{d}\lambda}{\mathrm{d}t}\right]\mathrm{d}Q \tag{4.22}$$

From Equation 4.20 we have

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{\partial H}{\partial P}\frac{\mathrm{d}P}{\mathrm{d}t} + \frac{\partial H}{\partial Q}\frac{\mathrm{d}Q}{\mathrm{d}t} + \frac{\partial H}{\partial \lambda}\frac{\mathrm{d}\lambda}{\mathrm{d}t} = \frac{\partial H}{\partial \lambda}\frac{\mathrm{d}\lambda}{\mathrm{d}t} \tag{4.23}$$

since the first two terms canceled due to Hamilton's equations. Substitution of Equation 4.23 into Equation 4.22 gives $\mathrm{d}S/\mathrm{d}t = 0$, completing the proof of adiabatic invariance. No assumption has been made here that P, Q are close to the values associated with an extremum of H.

This proof seems too neat, because it has established adiabatic invariance simply by careful use of the chain rule, and by taking partial derivatives. However, this observation reveals the underlying essence of adiabaticity, namely it is the differentiability of H, P with respect to λ from one period to the next and the Hamilton nature of the system, which together provide the conditions for the adiabatic invariant to exist. If the motion had been such that after one cycle the motion had changed so drastically that taking a derivative of H or P with respect to would not make sense, then the adiabatic invariant would not exist.

4.7.4 The First Adiabatic Invariant

We have already met the quantity

$$\mu = mv_\perp^2/2B$$

and have proved its invariance in spatially and temporally varying **B** fields. The periodic motion involved, of course, is the Larmor gyration. If we take p to be angular momentum $mv_\perp r$ and dq to be the coordinate $\mathrm{d}\theta$, the action integral becomes

$$\oint p dq = \oint mv_\perp r \mathrm{d}\theta = 2\pi r_L mv_\perp = 2\pi\frac{mv_\perp^2}{\omega_c} = 4\pi\frac{m}{|q|}\mu$$

Thus   is a constant of the motion as long as q/m is not changed. We have proved the invariance of   only with the implicit assumption $\omega/\omega_c \ll 1$, where ω is a frequency characterizing the rate of change of **B** as seen by the particle. A proof exists, however, that   is invariant even when $\omega/\omega_c \leq q$. In theorists' language,   is invariant "to all orders in an expansion in / c ." What this means in practice is that   remains much more nearly constant than **B** does during one period of gyration.

It is just as important to know when an adiabatic invariant does *not* exist as to know when it does. Adiabatic invariance of   is violated when   is not small compared with ω_c . We give three examples of this.

- *Magnetic Pumping.* If the strength of **B** in a mirror confinement system is varied sinusoidally, the particles' $v_\perp$ would oscillate; but there would be no gain of energy in the long run. However, if the particles make collisions, the invariance of   is violated, and the plasma can be heated. In particular, a particle making a collision during the compression phase can transfer part of its gyration energy into $v_\parallel$ energy, and this is not taken out again in the expansion phase.
- *Cyclotron Heating.* Now imagine that the **B** field is oscillated at the frequency ω_c. The induced electric field will then rotate in phase with some of the particles and accelerate their Larmor motion continuously. The condition $\omega \ll \omega_c$ is violated,   is not conserved, and the plasma can be heated.
- *Magnetic Cusps.* If the current in one of the coils in a simple magnetic mirror system is reversed, a magnetic cusp is formed. This configuration has, in addition to the usual mirrors, a spindle-cusp mirror extending over 360^o in azimuth. A plasma confined in a cusp device is supposed to have better stability properties than that in an ordinary mirror. Unfortunately, the loss-cone losses are larger because of the additional loss region; and the particle motion is nonadiabatic. Since the **B** field vanishes at the center of symmetry, ω_c is zero there; and   is not preserved. The local Larmor radius near the center is larger than the device. Because of this, the adiabatic invariant   does not guarantee that particles outside a loss cone will stay outside after passing through the nonadiabatic region. Fortunately, there is in this case another invariant: the canonical angular momentum $p_\theta = mrv_\theta - erA_\theta$. This ensures that there will be a population of particles trapped indefinitely until they make a collision.

4.7.5 The Second Adiabatic Invariant

Consider a particle trapped between two magnetic mirrors: it bounces between them and therefore has a periodic motion at the "bounce frequency". A constant of this motion is given by $\oint mv_\parallel ds$, where ds is an element of path length (of the guiding center) along a field line. However, since the guiding center drifts across field lines, the motion is not exactly periodic, and the constant of the motion becomes an adiabatic invariant. This is called the *longitudinal invariant* J and is defined for a half-cycle between the two turning points

$$J = \int_a^b v_\parallel ds$$

We shall prove that J is invariant in a static, nonuniform B field; the result is also true for a slowly time-varying B field.

Before embarking on this somewhat lengthy proof, let us consider an example of the type of problem in which a theorem on the invariance of J would be useful. As we have already seen, the earth's magnetic field mirror-traps charged particles, which slowly drift in longitude around the earth. If the magnetic field were perfectly symmetric, the particle would eventually drift back to the same magnetic field line. However, the actual field is distorted by such effects as the solar wind. In that case, will a particle ever come back to the same magnetic field line? Since the particle's energy is conserved and is equal to $\frac{1}{2}mv_\perp^2$ at the turning point, the invariance of indicates that $|B|$ remains the same at the turning point. However, upon drifting back to the same longitude, a particle may find itself on another magnetic field line at a different altitude. This cannot happen if J is conserved. J determines the length of the field lines between turning points, and no two lines have the same length between points with the same $|B|$. Consequently, the particle returns to the same magnetic field line even in a slightly asymmetric field.

To prove the invariance of J, we first consider the invariance of $v_\parallel \delta s$, where δs is a segment of the path along $\mathbf{B}$. Because of guiding center drifts, a particle on s will find itself on another magnetic field line $\delta s'$ after a time Δt. The length of $\delta s'$ is defined by passing planes perpendicular to $\mathbf{B}$ through the end points of δs. The length of δs is obviously proportional to the radius of curvature:

$$\frac{\delta s}{R_c} = \frac{\delta s'}{R_c'}$$

so that

$$\frac{\delta s' - \delta s}{\Delta t \delta s} = \frac{R_c' - R_c}{\Delta t R_c}$$

The "radial" component of $\mathbf{v}_{gc}$ is just

$$\mathbf{v}_{gc} \cdot \frac{\mathbf{R}_c}{R_c} = \frac{R_c' - R_c}{\Delta t}$$

The guiding center drift in curved vacuum field is

$$\mathbf{v}_{gc} = \mathbf{v}_R + \mathbf{v}_{\nabla B} = \pm \frac{1}{2} v_\perp r_L \frac{\mathbf{B} \times \nabla B}{B^2} + \frac{m v_\parallel^2}{q} \frac{\mathbf{R}_c \times \mathbf{B}}{R_c^2 B^2}$$

The curvature drift has no component along $\mathbf{R}_c$. Using the above three equations, we have

$$\frac{1}{\delta s} \frac{\mathrm{d}}{\mathrm{d}t} \delta s = \mathbf{v}_{gs} \cdot \frac{\mathbf{R}_c}{R_c^2} = \frac{1}{2} \frac{m}{q} \frac{v_\perp^2}{B^3} (\mathbf{B} \times \nabla B) \cdot \frac{\mathbf{R}_c}{R_c^2}$$

This is the rate of change of δs as seen by the particle. We must now get the rate of change of $v_\parallel$ as seen by the particle. The parallel and perpendicular energies are defined by

$$W \equiv \frac{1}{2}mv_{\parallel}^2 + \frac{1}{2}mv_{\parallel}^2 = \frac{1}{2}mv_{\parallel}^2 + \mu B \equiv W_{\parallel} + W_{\perp}$$

Thus $v_{\parallel}$ can be written

$$v_{\parallel} = \left[\frac{2}{m}(W - \mu B)\right]^{1/2}$$

Here W and   are constant, and only B varies. Therefore,

$$\frac{\dot{v}_{\parallel}}{v_{\parallel}} = -\frac{1}{2}\frac{\mu \dot{B}}{W - \mu B} = -\frac{1}{2}\frac{\mu \dot{B}}{W_{\parallel}} = -\frac{\mu \dot{B}}{mv_{\parallel}^2}$$

Since B was assumed static, $\dot{B}$ is not zero only because of the guiding center motion:

$$\dot{B} = \frac{\mathrm{d}B}{\mathrm{d}\mathbf{r}} \cdot \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = \mathbf{v}_{gc} \cdot \nabla B = \frac{mv_{\parallel}^2}{q}\frac{\mathbf{R}_c \times \mathbf{B}}{R_c^2 B^2} \cdot \nabla B$$

Now we have

$$\frac{\dot{v}_{\parallel}}{v_{\parallel}} = -\frac{\mu}{q}\frac{(\mathbf{R}_c \times \mathbf{B}) \cdot \nabla B}{R_c^2 B^2} = -\frac{1}{2}\frac{m}{q}\frac{v_{\perp}^2}{B}\frac{(\mathbf{R}_c \times \mathbf{B}) \cdot \nabla B}{R_c^2 B^2}$$

The fractional change in $v_{\parallel}\delta s$ is

$$\frac{1}{v_{\parallel}\delta s}\frac{\mathrm{d}}{\mathrm{d}t}(v_{\parallel}\delta s) = \frac{1}{\delta s}\frac{\mathrm{d}\delta s}{\mathrm{d}t} + \frac{1}{v_{\parallel}}\frac{\mathrm{d}v_{\parallel}}{\mathrm{d}t}$$

From the previous derivations we see that the these two terms cancel, so that

$$v_{\parallel}\delta s = \text{constant}$$

This is not exactly the same as saying that J is constant, however. In taking the integral of $v_{\parallel}\delta s$, between the turning points, it may be that the turning points on $\delta s'$ do not coincide with the intersections of the perpendicular planes. However, any error in J arising from such a discrepancy is negligible because near the turning points, $v_{\parallel}$ is nearly zero. Consequently, we have proved

$$J = \int_a^b v_{\parallel} ds = \text{constant}$$

An example of the violation of J invariance is given by a plasma heating scheme called *transit-time magnetic pumping*. Suppose an oscillating current is applied to the coils of a mirror system so that the mirrors alternately approach and withdraw from each other near the bounce frequency. Those particles that have the right bounce frequency will always see an approaching mirror and will therefore gain $v_{\parallel}$. J is not conserved in this case because the change of **B** occurs on a time scale not long compared with the bounce time.

4.7.6 The Third Adiabatic Invariant

Referring again to the earth dipole case, we see that the slow drift of a guiding center around the earth constitutes a third type of periodic motion. The adiabatic invariant connected with this turns out to be the total magnetic flux Φ enclosed by the drift surface,

$$\Phi = \int \mathbf{B} \cdot \mathrm{d}\mathbf{S} \tag{4.24}$$

It is almost obvious that, as $\mathbf{B}$ varies, the particle will stay on a surface such that the total number of magnetic field lines enclosed remains constant. This invariant, Φ, has few applications because most fluctuations of $\mathbf{B}$ occur on a time scale short compared with the drift period. As an example of the violation of Φ invariance, we can cite some recent work on the excitation of ultra-low frequency magnetohydrodynamic waves in the ionosphere. These waves have a long period comparable to the drift time of a particle around the earth. The particles can therefore encounter the wave in the same phase each time around. If the phase is right, the wave can be excited by the conversion of particle drift energy to wave energy.

Here are some time scales to remember for a 1 MeV proton in the Earth's radiation belt. The drift period is about 16 mins, the bounce period is about 0.22 s, and the gyroperiod is about 0.22 ms. These numbers are from Solène Lejosne's calculator.

In radiation belt studies, it is typical to consider a quantity that is inversely proportional to the magnetic flux:

$$L^* = \frac{2\pi M}{|\phi| R_E} = \frac{2\pi B_{eq} R_E^2}{|\phi|} \tag{4.25}$$

where M is the magnetic moment of the Earth's dipole field, R_E is the radius of the Earth, and B_{eq} is the magnetic field strength at the equatorial surface. This presentation of the third invariant is termed L^*, or Roederer's L parameter, which is the radial distance to the equatorial point where the particle would be found if all nondipolar perturbations in the geomagnetic field were adiabatically turned off leaving only the dipole field (Roederer, 1970). It should be noted that in a nondipolar field L^* does *not* represent a spatial coordinate, instead L^* is a property of a stably trapped particle (Roederer and Lejosne, 2018). That is, in absence of acceleration or loss processes, L^* defines a surface with constant phase space density. If a particle encounter a boundary of the radiation belt system (i.e. the magnetopause or atmosphere) during its drift leading to its loss from the system, the particle does not have an L^* value. The particle is said to be in the drift loss cone.

Therefore, the L and L^* parameters coincide in a dipole field, but in a nondipolar field, these parameters are different. Typically, L is used with particle fluxes whereas L^* is better suited for phase space density.

4.8 Ring Current

One typical application of the single particle motion is the study of ring current. Beyond the lowest order of motion (i.e. gyration combined with parallel drift) of charged particles in the magnetosphere, let us examine the higher order corrections. For the case of non-time-varying fields, and a weak electric field, these corrections consist of a combination of $\mathbf{E} \times \mathbf{B}$ drift, grad-B drift, and curvature drift:

$$\mathbf{v}_{1\perp} = \frac{\mathbf{E} \times \mathbf{B}}{B^2} + \frac{\mu}{m\Omega}\hat{b} \times \nabla B + \frac{v_\parallel^2}{\Omega}\hat{b} \times (\hat{b} \cdot \nabla)\hat{b}$$

Let us neglect $\mathbf{E} \times \mathbf{B}$ drift, since this motion merely gives rise to the convection of plasma within the magnetosphere, without generating a current. By contrast, there is a net current associated with grad-B drift and curvature drift. In the limit in which this current does not strongly modify the ambient magnetic field (i.e., $\nabla \times \mathbf{B} \simeq \mathbf{0}$), which is certainly the situation in the Earth's magnetosphere, we can write

$$(\mathbf{B} \cdot \nabla)\,\mathbf{B} = -\mathbf{B} \times (\nabla \times \mathbf{B}) \simeq \frac{\nabla_\perp B}{B}$$

It follows that the higher order drifts can be combined to give

$$\mathbf{v}_{1\perp} = \frac{({v_\perp}^2/2 + {v_\parallel}^2)}{\Omega\, B}\,\mathbf{B} \times \nabla B$$

For the dipole field Equation 3.4, the above expression yields

$$\mathbf{v}_{1\perp} \simeq -\mathrm{sgn}(\Omega)\,\frac{6\,\mathcal{E}L^2}{eB_E R_E}(1 - B/2B_m)\frac{\cos^5\vartheta(1+\sin^2\vartheta)}{(1+3\sin^2\vartheta)^2}\,\hat{\varphi} \tag{4.26}$$

where $\mathcal{E}$ is the energy of the particle, B_E is the equatorial magnetic field-strength on the Earth's surface, and B_m is the magnetic field-strength at the mirror points. Note that the drift is in the azimuthal direction. A positive drift velocity corresponds to eastward motion, whereas a negative velocity corresponds to westward motion. It is clear that, in addition to their gyromotion and periodic bouncing motion along field lines, charged particles trapped in the magnetosphere also slowly precess around the Earth. The ions drift westwards and the electrons drift eastwards, giving rise to a net westward current circulating around the Earth. This current is known as the *ring current.*

Although the perturbations to the Earth's magnetic field induced by the ring current are small, they are still detectable. In fact, the ring current causes a slight reduction in the Earth's magnetic field in equatorial regions. The size of this reduction is a good measure of the number of charged particles contained in the Van Allen belts (Chapter 20). During the development of so-called geomagnetic storms, charged particles are injected into the Van Allen belts from the outer magnetosphere, giving rise to a sharp increase in the ring current, and a corresponding decrease in the Earth's equatorial magnetic field. These particles eventually precipitate out of the magnetosphere into the upper atmosphere at high latitudes, giving rise to intense auroral activity, serious interference in electromagnetic communications, and, in extreme cases, disruption of electric power grids. The ring current induced reduction in the Earth's magnetic field is measured by the so-called *Dst* index, which is based on hourly averages of the northward horizontal component of the terrestrial magnetic field recorded at four low-latitude observatories; Honolulu (Hawaii), San Juan (Puerto Rico), Hermanus (South Africa), and Kakioka (Japan). A reduction in the Dst index by 600 nT corresponds to a 2% reduction in the terrestrial magnetic field at the equator.

According to Equation 4.26, the precessional drift velocity of charged particles in the magnetosphere is a rapidly decreasing function of increasing latitude (i.e., most of the ring current is concentrated in the equatorial plane). Since particles typically complete many bounce orbits during a full rotation around the Earth, it is convenient to average Equation 4.26 over a bounce period to obtain the *average* drift velocity. This averaging can only be performed numerically. The final answer is well approximated by

$$\langle v_d \rangle \simeq \frac{6\,\mathcal{E}\,L^2}{e\,B_E\,R_E}\,(0.35 + 0.15\,\sin\alpha_{\rm eq}) \tag{4.27}$$

The average drift period (i.e., the time required to perform a complete rotation around the Earth) is simply

$$\langle \tau_d \rangle = \frac{2\pi\,L\,R_E}{\langle v_d \rangle} \simeq \frac{\pi e B_E R_E^2}{3\,\mathcal{E}\,L}\,(0.35 + 0.15\,\sin\alpha_{\rm eq})^{-1}$$

Thus, the drift period for protons and electrons is

$$\langle \tau_d \rangle_p = \langle \tau_d \rangle_e \simeq \frac{1.05}{\mathcal{E}(\text{MeV})L}\,(1 + 0.43\,\sin\alpha_{\rm eq})^{-1}\ \text{hours}$$

Note that MeV energy electrons and ions precess around the Earth with about the same velocity, only in opposite directions, because there is no explicit mass dependence in Equation 4.27. It typically takes ~ 20 mins to perform a full rotation. The drift period only depends weakly on the equatorial pitch angle $\alpha_{\rm eq}$, as is the case for the bounce period. Somewhat paradoxically, the drift period is shorter on more distant L-shells. Note, of course, that particles only get a chance to complete a full rotation around the Earth if the inner magnetosphere remains quiescent on time-scales of order an hour, which is, by no means, always the case.

Note, finally, that, since the rest mass of an electron is 0.51 MeV, most of the above formulae require relativistic correction when applied to MeV energy electrons. Fortunately, however, there is no such problem for protons, whose rest mass energy is 0.94 GeV.

4.9 Particle Acceleration

- *Fermi acceleration* is the acceleration that charged particles undergo when being repeatedly reflected, usually by a magnetic mirror. This is sometimes referred to as *diffusive shock acceleration*, which is actually a subclass of Fermi acceleration. There are two types of Fermi acceleration: *first-order Fermi acceleration* (in shocks) and *second-order Fermi acceleration* (in the environment of moving magnetized gas clouds). In both cases the environment has to be collisionless in order for the mechanism to be effective. This is because Fermi acceleration only applies to particles with energies exceeding the thermal energies, and frequent collisions with surrounding particles will cause severe energy loss and as a result no acceleration will occur. See more in wiki.

- *Betatron acceleration* refers to situations in which the magnetic field strength increases slowly in time (compared with a gyroperiod), so that μ remains constant, but the particle kinetic energy is not constant due to the presence of electric fields (associated with the time-varying magnetic field). Then, the perpendicular energy is increased due to constancy of μ.

Particle accelerations are important consequences of magnetic reconnection and shocks. In magnetic reconnection processes, the energy is rapidly converted from magnetic field to particles in the forms of flows, bulk ("thermal") heating, and high-energy ("nonthermal") tails. (Ji et al. 2022) describes the current understanding of electron acceleration during magnetic reconnection in the non-relativistic and relativistic regimes. First-principles numerical simulations suggest power-law nonthermal particle populations are more prevalent in systems with low β and moderate guide (perpendicular to the reconnection plane) magnetic fields, with acceleration efficiency enhanced in 3D compared to simplified 2D calculations. Multiple acceleration mechanisms have been studied. In the localized diffusion region near an X-line without a guide field, particles may become demagnetized, prone to be directly accelerated by the reconnection electric field as they undergo so-called *Speiser* or *meandering orbits*. There is recent laboratory evidence for electron acceleration by this mechanism in action during magnetically driven reconnection at low-β using laser-powered capacitor coils. There is also numerical evidence that a substantial fraction of dissipated magnetic energy is carried by particles accelerated by this mechanism when they are demagnetized and escape from plasmoids in 3D.

Over larger scales when the particles remain magnetized, there are three basic particle acceleration processes that dominate in collisionless reconnection: *Fermi acceleration* which takes place as particles stream along and drift in relaxing curved magnetic field lines, relatively *localized electric fields* $E_\parallel$ parallel to the magnetic field directly accelerate particles, and *betatron heating* which occurs as particles drift into regions of stronger magnetic field while conserving the first adiabatic moment $\mu = mv_\perp^2/B$. These mechanisms are described by guiding center equations of motion for the particles, and the energization rate in the non-relativistic limit is given approximately by the following equation:

$$
\frac{\mathrm{d}\mathcal{E}}{dt} = qE_{\parallel}v_{\parallel} + \mu\frac{\mathrm{d}B}{\mathrm{d}t} + q\mathbf{E}\cdot\mathbf{u}_c + \frac{1}{2}m\frac{\mathrm{d}}{\mathrm{d}t}|\mathbf{u}_E|^2 \tag{4.28}
$$

where $\mathcal{E}$ is the energy of a particle, $\mathbf{u}_E$ is the $\mathbf{E}\times\mathbf{B}$ drift velocity, $\frac{\mathrm{d}}{\mathrm{d}t} = \frac{\partial}{\partial t} + \mathbf{u}_E\cdot\nabla$, $E_{\parallel}$ is the parallel electric field, and $v_{\parallel}$ is the drift-corrected guiding center parallel velocity. For slowly varying fields, $\mathbf{u}_c$ is the curvature drift of particles and reduces to $\mathbf{u}_c \sim (mv_{\parallel}^2/qB)(\hat{b}\times\boldsymbol{\kappa})$, where $\boldsymbol{\kappa} = \hat{b}\cdot\nabla\hat{b}$ is the magnetic field curvature and $\hat{b}$ is the unit vector in the direction of $\mathbf{B}$. The terms on the right-hand side represent energy gain by the parallel electric field, betatron heating, Fermi acceleration, and the polarization drift respectively. The first three mechanisms are sketched in **?@fig-particle_acceleration_reconnection**. The polarization drift is particularly important for lower-energy ions, and this terms accounts for the kinetic ennergy gain in the large-scale reconnection outflow jets. Although Equation 4.28 is valid in the non-relativistic limit, the same basic acceleration mechanisms are responsible for energizing particles in more extreme relativistic contexts relevant to astrophysical systems.

Particles gain energy when their drift induced by the curvature of the magnetic field lines is aligned with the reconnection electric field, which is a form of Fermi heating. Fermi acceleration may resolve two major challenges for heating models. First, it may operate over large volumes in the presence of many magnetic islands or plasmoids, which may be sampled by electrons in the presence of 3D ergodic field lines. There are several space observations of energetic electrons near magnetic islands or plasmoids. Second, the Fermi process naturally generates power laws because a particle undergoing Fermi acceleration gains energy at a rate proportional to its energy, $\delta\mathcal{E}\propto\mathcal{E}$. In the non-relativistic limit, adiabatic particle energization by Fermi acceleration along with the other heating mechanisms for magnetized particles leads to a Parker-like transport equation for the distribution $f(v_{\parallel}, v_{\perp})$ of particles within a magnetic flux tube including time-varying plasma density n and magnetic field $\mathbf{B}$:

$$
\frac{\partial f}{\partial t} + \frac{\dot{B}}{B}v_{\perp}^2\frac{\partial f}{\partial v_{\perp}^2} + \left(\frac{\dot{n}}{n} - \frac{\dot{B}}{B}\right)v_{\parallel}\frac{\partial f}{\partial v_{\parallel}} = \frac{D\{f\}}{\tau_{\text{diff}}} - \frac{f}{\tau_{\text{esc}}} + \frac{f_{\text{inj}}}{\tau_{\text{inj}}} \tag{4.29}
$$

where $\frac{D\{f\}}{\tau_{\text{diff}}}$ includes diffusive processes such as pitch-angle or energy scattering, τ_{esc} is a typical time for particles to escape the reconnection geometry, and $f_{\text{inj}}/\tau_{\text{inj}}$ is a source of newly injected particles. This transport equation leads to power-law solutions under certain assumptions, though the Fermi process becomes much less effective at generating power-law nonthermal tails in the presence of a significant guide magnetic field component. With certain assumptions, it is possible to solve the Parker equation using the velocity and magnetic fields from MHD simulations to obtain predictions of the particle acceleration, but this approach does not include the self-consistent feedback.

4.10 Test Particle Model

The test particle method is not self-consistent, because we only describe the effect of the fields onto particles, but not vice versa. This reduces the PDEs to ODEs which is much simpler to solve. Despite of this, we use the trajectories of test particles to infer approximate kinetic properties of the system. You will see many examples of understanding more complicated scenario using the drifts derived from test particle motions, especially in terms of stability analysis.

There are four methods in test particle modelling:

1. Trajectory sampling. It solves individual particle trajectories, but it may not be trivial to select the samples.
2. Forward Monte Carlo. We inject particles in source regions where f is known, and follow them until they reach the regions of interest. This is similar to PIC, except that the fields are not self-consistent.
3. *Forward Liouville.*
4. *Backward Liouville.* It also makes use of the Liouville's theorem for f, but there is no sampling, neither in $\mathbf{x}$ nor in $\mathbf{v}$ space, implying no statistical errors. The procedure starts by choosing a given point $\mathbf{x}$ in space …

5 Plasmas as Fluid

In a plasma the situation is much more complicated than that single particle orbits; the **E** and **B** fields are not prescribed but are determined by the positions and motions of the charges themselves. One must solve a self-consistent problem; that is, find a set of particle trajectories and field patterns such that the particles will generate the fields as they move along their orbits and the fields will cause the particles to move in those exact orbits. And this must be done in a time-varying situation. It sounds very hard, but it is not.

We have seen that a typical plasma density might be 10^{18} ion–electron pairs per m^3. If each of these particles follows a complicated trajectory and it is necessary to follow each of these, predicting the plasma's behavior would be a hopeless task. Fortunately, this is not usually necessary because, surprisingly, the majority-perhaps as much as 80%-of plasma phenomena observed in real experiments can be explained by a rather crude model. This model is that used in fluid mechanics, in which the identity of the individual particle is neglected, and only the motion of fluid elements is taken into account. Of course, in the case of plasmas, the fluid contains electrical charges. In an ordinary fluid, frequent collisions between particles keep the particles in a fluid element moving together. It is surprising that such a model works for plasmas, which generally have infrequent collisions. But we shall see that there is a reason for this.

A more refined treatment-the kinetic theory of plasmas-requires more mathematical calculation. An introduction to kinetic theory is given in Chapter 8.

In some plasma problems, neither fluid theory nor kinetic theory is sufficient to describe the plasma's behavior. Then one has to fall back on the tedious process of following the individual trajectories. Brute-force computer simulation can play an important role in filling the gap between theory and experiment in those instances where even kinetic theory cannot come close to explaining what is observed.

5.1 Definitions

Variable	Description
m_s	mass
q_s	charge
n_s	number density

Variable	Description
ρ_s	mass density
σ_s, ρ^*	charge density ($= n_s q_s$)
T_s	temperature
p_s	scalar pressure
$\mathbf{u}_s$	flow velocity
$\mathbf{j}_s$	current density ($= \rho_s^* \mathbf{u}_s$)
e_s	internal energy
ϕ_s	potential energy
ϵ_s	total energy ($= e_s + \frac{\mathbf{u_s}^2}{2} + \phi_s$)

where s denotes the species (e.g. $\mathrm{H}^+, \mathrm{O}^+$). Do not confuse σ here with conductivity. Then the total quantities without subscripts can be written as

$$
\begin{aligned}
&n = \sum_s n_s \\
&\rho = \sum_s \rho_s = \sum_s m_s n_s \\
&p = \sum_s p_s \\
&T = \sum_s \frac{n_s}{n} T_s \\
&\mathbf{u} = \frac{1}{\rho} \sum_s \rho_s \mathbf{u}_s \\
&\mathbf{v}_s = \mathbf{u}_s - \mathbf{u}, \quad \text{relative velocity of the } s^{th} \text{ species} \\
&\sigma = \sum_s \sigma_s = \sum_s q_s n_s \\
&\mathbf{j} = \sum_s \mathbf{j}_s = \sum_s \sigma_s \mathbf{u}_s = \sum_s \sigma_s \mathbf{v}_s + \mathbf{u} \sum_s \sigma_s = \sum_s \sigma_s \mathbf{v}_s + \sigma \mathbf{u} = \mathbf{j}_{cd} + \mathbf{j}_{cv} \\
&\quad \text{where } \mathbf{j}_{cd} = \sum_s \sigma_s \mathbf{v}_s \text{ is the conduction current density} \\
&\qquad\qquad \mathbf{j}_{cv} = \sigma \mathbf{u} \text{ is the convection current density} \\
&\epsilon = \frac{1}{\rho} \sum_s \rho_s \epsilon_s = e + \frac{u^2}{2} + \phi \text{ (internal + kinetic + potential)}
\end{aligned}
$$

It can be easily verified that

$$\sum_s \rho_s \mathbf{v}_s = 0$$

In this note we tried to use $\mathbf{v}$ for all species/particle-based velocities, and $\mathbf{u}$ for all bulk velocities.

We have 5 independent unknown quantities for each species: $n_j, \mathbf{u}_j, p_j$. The pressure can be replaced by temperature T_j. Additionally, we have the 6 EM field quantities: $\mathbf{B}, \mathbf{E}$. For a 2-fluid despcription with ions and electrons, altogether we have $5 * 2 + 6 = 16$ unknowns, so we need 16 equations to determine the system.

5.2 Relation of Plasma to Ordinary Electromagnetics

5.2.1 Maxwell's Equations

In vacuum:

$$\begin{aligned} \epsilon_0 \nabla \cdot \mathbf{E} &= \sigma \\ \nabla \times \mathbf{E} &= -\dot{\mathbf{B}} \\ \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{B} &= \mu_0 (\mathbf{j} + \epsilon_0 \dot{\mathbf{E}}) \end{aligned} \tag{5.1}$$

In a medium:

$$\begin{aligned} \nabla \cdot \mathbf{D} &= \sigma \\ \nabla \times \mathbf{E} &= -\dot{\mathbf{B}} \\ \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{H} &= \mathbf{j} + \dot{\mathbf{D}} \\ \mathbf{D} &= \epsilon \mathbf{E} \\ \mathbf{B} &= \mu \mathbf{H} \end{aligned} \tag{5.2}$$

σ and $\mathbf{j}$ stand for the "free" charge and current densities. The "bound" charge and current densities arising from polarization and magnetization of the medium are included in the definition of the quantities $\mathbf{D}$ and $\mathbf{H}$ in terms of ϵ and μ. In a plasma, the ions and electrons comprising the plasma are the equivalent of the "bound" charges and currents. Since these charges move in a complicated way, it is impractical to try to lump their effects into two constants ϵ and μ. Consequently, in plasma physics, one generally works with the vacuum equations, in which σ and $\mathbf{j}$ include *all* the charges and currents, both external and internal.

Note that we have used $\mathbf{E}$ and $\mathbf{B}$ in the vacuum equations rather than their counterparts $\mathbf{D}$ and $\mathbf{H}$, which are related by the constants ϵ_0 and μ_0. This is because the forces $q\mathbf{E}$ and $\mathbf{j} \times \mathbf{B}$ depend on $\mathbf{E}$ and $\mathbf{B}$ rather than $\mathbf{D}$ and $\mathbf{H}$, and it is not necessary to introduce the latter quantities as long as one is dealing with the vacuum equations.

5.2.2 Classical Treatment of Magnetic Materials

Since each gyrating particle has a magnetic moment, it would seem that the logical thing to do would be to consider a plasma as a magnetic material with a permeability μ_m. ((We have put a subscript m on the permeability to distinguish it from the adiabatic invariant μ.) To see why this is *not* done in practice, let us review the way magnetic materials are usually treated.

The ferromagnetic domains, say, of a piece of iron have magnetic moments μ_i, giving rise to a bulk magnetization

$$\mathbf{M} = \frac{1}{V}\sum_i \mu_i$$

per unit volume. This has the same effect as a bound current density equal to

$$\mathbf{j}_b = \nabla \times \mathbf{M}$$

In the vacuum Ampère's law, we must include in $\mathbf{j}$ both this current and the "free", or externally applied, current $\mathbf{j}_f$:

$$\mu_0^{-1}\nabla \times \mathbf{B} = \mathbf{j}_f + \mathbf{j}_b + \epsilon_0\dot{\mathbf{E}}$$

We wish to write this equation in the simple form

$$\nabla \times \mathbf{H} = \mathbf{j}_f + \epsilon_0\dot{\mathbf{E}}$$

by including $\mathbf{j}_b$ in the definition of $\mathbf{H}$. This can be done if we let

$$\mathbf{H} = \mu_0^{-1}\mathbf{B} - \mathbf{M}$$

To get a simple relation between $\mathbf{B}$ and $\mathbf{H}$, we assume $\mathbf{M}$ to be proportional to $\mathbf{B}$ or $\mathbf{H}$:

$$\mathbf{M} = \chi_m\mathbf{H}$$

The constant χ_m is the magnetic susceptibility. We now have

$$\mathbf{B} = \mu_0(1 + \chi_m)\mathbf{H} \equiv \mu_m\mathbf{H}$$

This simple relation between $\mathbf{B}$ and $\mathbf{H}$ is possible because of the linear relation between $\mathbf{M}$ and $\mathbf{H}$.

In a plasma with a magnetic field, each particle has a magnetic moment μ_α, and the quantity $\mathbf{M}$ is the sum of all these μ_α's in 1 m^3. But now we have

$$\mu_\alpha = \frac{mv_{\perp\alpha}^2}{2B} \propto \frac{1}{B} \quad \mathbf{M} \propto \frac{1}{B}$$

The relation between $\mathbf{M}$ and $\mathbf{H}$ (or $\mathbf{B}$) is no longer linear, and we cannot write $\mathbf{B} = \mu_m \mathbf{H}$ with μ_m constant. It is therefore not useful to consider a plasma as a magnetic medium.

5.2.3 Classical Treatment of Dielectrics

The polarization $\mathbf{P}$ per unit volume is the sum over all the individual moments $\mathbf{p}_i$ of the electric dipoles. This gives rise to a bound charge density

$$\sigma_b = -\nabla \cdot \mathbf{P} \tag{5.3}$$

In the vacuum equation, we must include both the bound charge and the free charge:

$$\epsilon_0 \nabla \cdot \mathbf{E} = \sigma_f + \sigma_b$$

We wish to write this in the simple form

$$\nabla \cdot \mathbf{D} = \sigma_f$$

by including σ_b in the definition of $\mathbf{D}$. This can be done by letting

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} \equiv \epsilon \mathbf{E}$$

If $\mathbf{P}$ is linearly proportional to $\mathbf{E}$,

$$\mathbf{P} = \epsilon_0 \chi_e \mathbf{E}$$

then ϵ is a constant given by

$$\epsilon = (1 + \chi_e)\epsilon_0$$

There is no a priori reason why a relation like the above cannot be valid in a plasma, so we may proceed to try to get an expression for ϵ in a plasma.

5.2.4 The Dielectric Constant of a Plasma

We have seen in Section 4.4 that a fluctuating $\mathbf{E}$ field gives rise to a polarization current $\mathbf{j}_p$. This leads, in turn, to a polarization charge given by the equation of continuity:

$$\frac{\partial \sigma_p}{\partial t} + \nabla \cdot \mathbf{j}_p = 0$$

This is the equivalent of Equation 5.3, except that, as we noted before, a polarization effect does not arise in a plasma unless the electric field is time varying. Since we have an explicit expression for $\mathbf{j}_p$ but not for σ_p, it is easier to work with the Ampère's law:

$$\nabla \times \mathbf{B} = \mu_0(\mathbf{j}_f + \mathbf{j}_p + \epsilon\dot{\mathbf{E}})$$

We wish to write this in the form

$$\nabla \times \mathbf{B} = \mu_0(\mathbf{j}_f + \epsilon\dot{\mathbf{E}})$$

This can be done if we let

$$\epsilon = \epsilon_0 + \frac{j_p}{\dot{E}}$$

From Equation 4.4 for $\mathbf{j}_p$, we have

$$\epsilon = \epsilon_0 + \frac{\rho}{B^2} \quad \text{or} \quad \epsilon_R \equiv \frac{\epsilon}{\epsilon_0} = 1 + \frac{\mu_0 \rho c^2}{B^2} \tag{5.4}$$

This is the *low-frequency plasma dielectric constant for transverse motions*. The qualifications are necessary because our expression for $\mathbf{j}_p$ is valid only for $\omega^2 \ll \omega_c^2$ and for $\mathbf{E}$ perpendicular to $\mathbf{B}$. The general expression for ϵ, of course, is very complicated and hardly fits on one page.

Note that as $\rho \to 0$, ϵ_R approaches its vacuum value, unity, as it should. As $B \to \infty$, ϵ_R also approaches unity. This is because the polarization drift $\mathbf{v}_p$ then vanishes, and the particles do not move in response to the transverse electric field. In a usual laboratory plasma, the second term in Equation 5.4 is large compared with unity. For instance, if $n = 10^{16}\,\mathrm{m}^{-3}$ and $B = 0.1\,\mathrm{T}$ we have (for hydrogen)

$$\frac{\mu_0 \rho c^2}{B^2} = \frac{(4\pi \times 10^{-7})(10^{16})(1.67 \times 10^{-27})(3 \times 10^8)^2}{(0.1)^2} = 189$$

This means that the electric fields due to the particles in the plasma greatly alter the fields applied externally. A plasma with large ϵ shields out alternating fields, just as a plasma with small λ_D shields out dc fields.

5.3 Fluid Equations

In (F. F. Chen 2016), this is introduced in the diffusion chapter 5.7, which is a bit weird. In (Bellan 2008), this is introduced in chapter 2 where Vlasov equation is first derived and then follows the simplifications which lead to 2-fluid and MHD.

5.3.1 Equation of Continuity

The integral form of mass conservation for each species is

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_V \rho_s dx^3 = 0$$

The conservation of matter requires that the total number of particles N_s in a volume V can change only if there is a net flux of particles across the surface S bounding that volume. Since the particle flux density is $n_s\mathbf{u}_s$, we have, by the divergence theorem,

$$\frac{\partial N_s}{\partial t} = \int_V \frac{\partial n_s}{\partial t} dV = -\oint n\mathbf{u}_s \cdot \mathrm{d}\mathbf{S} = -\int_V \nabla \cdot (n_s\mathbf{u}_s) dV$$

Since this must hold for any volume V, the integrands must be equal:

$$\frac{\partial n_s}{\partial t} + \nabla \cdot (n_s\mathbf{u}_s) = 0 \tag{5.5}$$

There is one such *equation of continuity* for each species. Any sources or sinks of particles are to be added to the right-hand side.

5.3.2 Momentum Equation

Maxwell's equations tell us what **E** and **B** are for a given state of the plasma. To solve the self-consistent problem, we must also have an equation giving the plasma's response to given **E** and **B**. In the fluid approximation, we consider the plasma to be composed of two or more interpenetrating fluids, one for each species. In the simplest case, when there is only one species of ion, we shall need two equations of motion, one for the positively charged ion fluid and one for the negatively charged electron fluid. In a partially ionized gas, we shall also need an equation for the fluid of neutral atoms. The neutral fluid will interact with the ions and electrons only through collisions. The ion and electron fluids will interact with each other even in the absence of collisions, because of the **E** and **B** fields they generate.

The equation of motion for a single particle is

$$m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \tag{5.6}$$

Assume first that there are no collisions and no thermal motions. Then all the particles in a fluid element move together, and the average velocity $\mathbf{u}$ of the particles in the element is the same as the individual particle velocity $\mathbf{v}$. The fluid equation is obtained simply by multiplying Equation 5.6 by the density n:

$$mn\frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t} = qn(\mathbf{E} + \mathbf{u} \times \mathbf{B}) \tag{5.7}$$

This is, however, not a convenient form to use. In Equation 5.6, the time derivative is to be taken *at the position of the particles*. On the other hand, we wish to have an equation for fluid elements *fixed in space*, because it would be impractical to do otherwise. Consider a drop of cream in a cup of coffee as a fluid element. As the coffee is stirred, the drop distorts into a filament and finally disperses all over the cup, losing its identity. A fluid element at a fixed spot in the cup, however, retains its identity although particles continually go in and out of it.

To make the transformation to variables in a fixed frame, consider $\mathbf{G}(x,t)$ to be any property of a fluid in one-dimensional x space. The change of $\mathbf{G}$ with time *in a frame moving with the fluid* is the sum of two terms:

$$\frac{\mathrm{d}\mathbf{G}(x,t)}{\mathrm{d}t} = \frac{\partial \mathbf{G}}{\partial t} + \frac{\partial \mathbf{G}}{\partial x}\frac{\partial x}{\partial t} = \frac{\partial \mathbf{G}}{\partial t} + u_x\frac{\partial \mathbf{G}}{\partial x}$$

The first term on the right represents the change of $\mathbf{G}$ at a fixed point in space, and the second term represents the change of G as the observer moves with the fluid into a region in which $\mathbf{G}$ is different. In three dimensions, this generalizes to

$$\frac{\mathrm{d}\mathbf{G}}{\mathrm{d}t} = \frac{\partial \mathbf{G}}{\partial t} + (\mathbf{u}\cdot\nabla)\mathbf{G}$$

This is called the *convective derivative* (Equation 3.2).

In the case of a plasma, we take $\mathbf{G}$ to be the fluid velocity $\mathbf{u}$ and write Equation 5.7 as

$$mn\Big[\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u}\cdot\nabla)\mathbf{u}\Big] = qn(\mathbf{E} + \mathbf{u} \times \mathbf{B})$$

where $\partial\mathbf{u}/\partial t$ is the time derivative in a fixed frame.

Stress Tensor –> scalar pressure

$$mn\Big[\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u}\cdot\nabla)\mathbf{u}\Big] = qn(\mathbf{E} + \mathbf{u} \times \mathbf{B}) - \nabla p \tag{5.8}$$

What we have shown here is only a special case: the transfer of x momentum by motion in the x direction; and we have assumed that the fluid is isotropic, so that the same result holds in the y and z directions. But it is also possible to transfer y momentum by motion in the x direction, for instance. This shear stress cannot be represented by a scalar p but must be given by a tensor $\mathbf{P}$, the stress tensor, whose components $P_{ij} = mn\overline{v_i v_j}$ specify both the direction of motion and the component of momentum involved. In the general case the term ∇p is replaced by $-\nabla \cdot \mathbf{P}$.

When the distribution function is an isotropic Maxwellian, $\mathbf{P}$ is written

$$\mathbf{P} = \begin{pmatrix} p & 0 & 0 \\ 0 & p & 0 \\ 0 & 0 & p \end{pmatrix}$$

$-\nabla \cdot \mathbf{P} = \nabla p$. A plasma could have two temperatures $T_\perp$ and $T_\parallel$ in the presence of a magnetic field. In that case, there would be two pressures $T_\perp$ and $T_\parallel$ in the presence of a magnetic field. In that case, there would be two pressure $p_\perp = nk_BT_\perp$ and $p_\parallel = nk_BT_\parallel$. The stress tensor is then

$$\mathbf{P} = \begin{pmatrix} p_\perp & 0 & 0 \\ 0 & p_\perp & 0 \\ 0 & 0 & p_\parallel \end{pmatrix}$$

where the coordinate of the third row or column is the direction of $\mathbf{B}$. This is still diagonal and shows isotropy in a plane perpendicular to $\mathbf{B}$.

In an ordinary fluid, the off-diagonal elements of $\mathbf{P}$ are usually associated with viscosity. When particles make collisions, they come off with an average velocity in the direction of the fluid velocity $\mathbf{u}$ at the point where they made their last collision. This momentum is transferred to another fluid element upon the next collision. This tends to equalize $\mathbf{u}$ at different points, and the resulting resistance to shear flow is what we intuitively think of as viscosity. The longer the mean free path, the farther momentum is carried, and the larger is the viscosity. In a plasma there is a similar effect which occurs even in the absence of collisions. The Larmor gyration of particles (particularly ions) brings them into different parts of the plasma and tends to equalize the fluid velocities there. The Larmor radius rather than the mean free path sets the scale of this kind of collisionless viscosity. It is a *finite-Larmor-radius effect* which occurs in addition to collisional viscosity and is closely related to the $\mathbf{v}_E$ drift in a nonuniform $\mathbf{E}$ field (Equation 4.3).

5.3.2.1 Collisions

If there is a neutral gas, the charged fluid will exchange momentum with it through collisions. The momentum lost per collision will be proportional to the relative velocity $\mathbf{u} - \mathbf{u}_0$, where $\mathbf{u}_0$ is the velocity of the neutral fluid. If τ, the mean free time between collisions, is approximately constant, the resulting force term can be roughly written as $-mn(\mathbf{u}-\mathbf{u}_0)/\tau$. The equation of motion can be generalized to include anisotropic pressure and neutral collisions as follows:

$$mn\Big[\frac{\partial \mathbf{u}}{\partial t}+(\mathbf{u}\cdot\nabla)\mathbf{u}\Big]=qn(\mathbf{E}+\mathbf{u}\times\mathbf{B})-\nabla\cdot\overleftrightarrow{P}-\frac{mn(\mathbf{u}-\mathbf{u}_0)}{\tau} \tag{5.9}$$

This can also be written as (including the pressure term and other forces)

$$\rho\frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t}=(\rho^*\mathbf{E}+\mathbf{j}\times\mathbf{B})-\nabla\cdot\overleftrightarrow{P}+\mathbf{f}_n$$

or, in an equivalent conservative form,

$$\frac{\partial(\rho\mathbf{u})}{\partial t}+\nabla\cdot(\rho\mathbf{u}\mathbf{u})=(\rho^*\mathbf{E}+\mathbf{j}\times\mathbf{B})-\nabla\cdot\overleftrightarrow{P}+\mathbf{f}_n$$

Collisions between charged particles have not been included; these will be discussed in sec-collision ADD IT!.

5.3.2.2 Comparison with Ordinary Hydrodynamics

Ordinary fluids obey the Navier–Stokes equation

$$\rho\Big[\frac{\partial \mathbf{u}}{\partial t}+(\mathbf{u}\cdot\nabla)\mathbf{u}\Big]=-\nabla p+\rho\nu\nabla^2\mathbf{u} \tag{5.10}$$

This is the same as Equation 5.9 except for the absence of electromagnetic forces and collisions between species (there being only one species). The viscosity term $\rho\nu\nabla^2\mathbf{u}$, where ν is the kinematic viscosity coefficient, is just the collisional part of $\nabla\cdot\overleftrightarrow{P}-\nabla p$ in the absence of magnetic fields. Equation 5.10 describes a fluid in which there are frequent collisions between particles. Equation 5.9, on the other hand, was derived without any explicit statement of the collision rate. Since the two equations are identical except for the $\mathbf{E}$ and $\mathbf{B}$ terms, can Equation 5.9 really describe a plasma species? The answer is a guarded yes, and the reasons for this will tell us the limitations of the fluid theory. This is extremely important to clarify — there are still quite many people think that collision is assumed for the fluid theory, therefore do not believe that MHD can be used to describe plasmas.

In the derivation of Equation 5.9, we did actually assume implicitly that there were many collisions. This assumption came in the derivation of the pressure tensor (Section 8.4) when we took the velocity distribution to be Maxwellian. Such a distribution generally comes about as the result of frequent collisions. However, this assumption was used only to take the average of v^2. Any other distribution with the same average would give us the same answer. The fluid theory, therefore, is not very sensitive to deviations from the Maxwellian distribution, but there are instances in which these deviations are important. Kinetic theory must then be used.

There is also an empirical observation by Irving Langmuir which helps the fluid theory. In working with the electrostatic probes which bear his name, Langmuir discovered that the electron distribution function was far more nearly Maxwellian than could be accounted for by the collision rate. This phenomenon, called *Langmuir's paradox*, has been attributed at times to high-frequency oscillations. There has been no satisfactory resolution of the paradox, but this seems to be one of the few instances in plasma physics where nature works in our favor.

Another reason the fluid model works for plasmas is that the magnetic field, when there is one, can play the role of collisions in a certain sense. When a particle is accelerated, say by an $\mathbf{E}$ field, it would continuously increase in velocity if it were allowed to free-stream. When there are frequent collisions, the particle comes to a limiting velocity proportional to $\mathbf{E}$. The electrons in a copper wire, for instance, drift together with a velocity $\mathbf{v} = \mu\mathbf{E}$, where μ is the mobility. A magnetic field also limits free-streaming by forcing particles to gyrate in Larmor orbits. The electrons in a plasma also drift together with a velocity proportional to $\mathbf{E}$, namely, $\mathbf{v}_E = \mathbf{E}\times\mathbf{B}/B^2$. In this sense, a collisionless plasma behaves like a collisional fluid. Of course, particles do free-stream *along* the magnetic field, and the fluid picture is not particularly suitable for motions in that direction. *For motions perpendicular to* $\mathbf{B}$*, the fluid theory is a good approximation.*

5.3.3 Equation of State

One more relation is needed to close the system of equations. A complete description requires the equation of energy. However, the simplest way is to use the thermodynamic equation of state relating p to n:

$$p = C\rho^\gamma \tag{5.11}$$

where C is a constant and γ is the ratio of specific heats C_p/C_ν. The term ∇p is therefore given by

$$\frac{\nabla p}{p} = \gamma\frac{\nabla n}{n} \tag{5.12}$$

For isothermal compression, we have

$$\nabla p = \nabla(nk_BT) = k_BT\nabla n$$

so that, clearly, $\gamma = 1$. For adiabatic compression, k_BT will also change, giving γ a value larger than one. If N is the number of degrees of freedom, γ is given by

$$\gamma = (2+N)/N$$

The validity of the equation of state requires that heat flow be negligible; that is, that thermal conductivity be low. Again, this is more likely to be true in directions perpendicular to $\mathbf{B}$ than parallel to it. Fortunately, most basic phenomena can be described adequately by the crude assumption of the equation of state.

5.4 Two-Fluid Model

Besides the Vlasov theory (Chapter 8), we can apply the simpler but equally powerful 2-fluid model, in which electrons and ions are treated as two different fluids. Depending on the form of the pressure term, we have the 5/6/10-moment equations.

5.4.1 Five-moment

Five-moment refers to $(\rho, \mathbf{u}, p)$. Under either adiabatic or isothermal assumptions, we can get rid of the pressure/energy equation and replace it with an equation of state. For simplicity, let the plasma have only two species: ions and electrons; extension to more species is trivial. The charge and current densities are then given by

$$
\begin{aligned}
\sigma &= n_i q_i + n_e q_e \\
\mathbf{j} &= n_i q_i \mathbf{v}_i + n_e q_e \mathbf{v}_e
\end{aligned}
$$

We shall neglect collisions and viscosity. The complete equation set is:

$$
\begin{aligned}
\epsilon_0 \nabla\cdot\mathbf{E} &= n_i q_i + n_e q_e \\
\nabla\times\mathbf{E} &= -\dot{\mathbf{B}} \\
\nabla\cdot\mathbf{B} &= 0 \\
\mu_0^{-1}\nabla\times\mathbf{B} &= n_i q_i \mathbf{u}_i + n_e q_e \mathbf{u}_e + \epsilon_0\dot{\mathbf{E}} \\
m_j n_j\Big[\frac{\partial \mathbf{u}_j}{\partial t} + (\mathbf{u}_j\cdot\nabla)\mathbf{u}_j\Big] &= q_j n_j(\mathbf{E} + \mathbf{u}_j\times\mathbf{B}) - \nabla p_j \quad j = i, e \\
\frac{\partial n_j}{\partial t} + \nabla\cdot(n_j\mathbf{u}_j) &= 0 \quad j = i, e \\
p_j &= C_j n_j^\gamma \quad j = i, e
\end{aligned}
\tag{5.13}
$$

where C_j is a constant and γ is the adiabatic index with the approximations in Table 5.2.

Table 5.2: Equation of state approximations.

Regime	Equation of state	Name
$V_{\mathrm{ph}} \gg v_{ts}$	$p_s \sim n_s^\gamma$	adiabatic

Regime	Equation of state	Name
$V_{\mathrm{ph}} \ll v_{tx}$	$p_s = n_s k_B T_s$, $T_s = \text{const.}$	isothermal

There are 16 scalar unknowns: $n_i, n_e, p_i, p_e, \mathbf{u}_i, \mathbf{u}_e, \mathbf{E}$, and $\mathbf{B}$. There are apparently 18 scalar equations if we count each vector equation as three scalar equations. However, two of Maxwell's equations are superfluous, since the two of the equations can be recovered from the divergences of the other two. The simultaneous solution of this set of 16 equations in 16 unknowns gives a self-consistent set of fields and motions in the fluid approximation.

5.4.2 Six-moment

For monatomic gases, the six-moment equations for all charged fluids (indexed by s) can be written as:

$$\begin{aligned}
\frac{\partial \rho_s}{\partial t} + \nabla\cdot(\rho_s \mathbf{u}_s) &= 0 \\
\frac{\partial(\rho_s \mathbf{u}_s)}{\partial t} + \nabla\cdot\left[\rho_s \mathbf{u}_s \mathbf{u}_s + p_{s\perp}\boldsymbol{I} + (p_{s\parallel} - p_{s\perp})\hat{b}\hat{b}\right] &= \frac{q_s}{m_s}\rho_s(\mathbf{E} + \mathbf{u}_s \times \mathbf{B}) \\
\frac{\partial p_{s\parallel}}{\partial t} + \nabla\cdot(p_{s\parallel}\mathbf{u}_s) &= -2p_{s\parallel}\hat{b}\cdot(\hat{b}\cdot\nabla)\mathbf{u}_s \\
\frac{\partial p_{s\perp}}{\partial t} + \nabla\cdot(p_{s\perp}\mathbf{u}_s) &= -p_{s\perp}(\nabla\cdot\mathbf{u}_s) + p_{s\perp}\hat{b}\cdot(\hat{b}\cdot\nabla)\mathbf{u}_s
\end{aligned} \tag{5.14}$$

The two pressure equations in Equation 5.14 can be combined to give the equation for the average pressure $p = (2p_\perp + p_\parallel)/3$:

$$\frac{\partial p_s}{\partial t} + \nabla\cdot(p_s \mathbf{u}_s) = (p_s - p_{s\parallel})\hat{b}\cdot(\hat{b}\cdot\nabla)\mathbf{u}_s - \left(p_s - \frac{p_{s\parallel}}{3}\right)\nabla\cdot\mathbf{u}_s \tag{5.15}$$

Alternatively, we can solve for the hydrodynamic energy density $e = \frac{n\mathbf{u}^2}{2} + \frac{3}{2}p$ for each species:

$$\frac{\partial e_s}{\partial t} + \nabla\cdot\left[\mathbf{u}_s(e_s + p_s) + \mathbf{u}_s\cdot(p_{s\parallel} - p_{s\perp})\hat{b}\hat{b}\right] = \frac{q_s}{m_s}\rho_s \mathbf{u}_s\cdot\mathbf{E}$$

which is more of a conservative form and can be beneficial to get better jump conditions across shock waves. Note, however, that the parallel pressure equation is still solved with the adiabatic assumption, so non-adiabatic heating is not properly captured. In addition, the magnetic energy is not included into the energy density, so the jump conditions are only approximate (I think this is more of a numerical consideration). In general, there can be many more source terms on the right hand sides of the above equations corresponding to gravity, charge exchange, chemical reactions, collisions, etc.

The electric field $\mathbf{E}$ and magnetic field $\mathbf{B}$ are obtained from Maxwell's equations:

$$
\begin{aligned}
\frac{\partial \mathbf{B}}{\partial t} + \nabla \times \mathbf{E} = 0 \\
\frac{\partial \mathbf{E}}{\partial t} - c^2 \nabla \times \mathbf{B} = -c^2 \mu_0 \mathbf{j} \\
\nabla \cdot \mathbf{E} = \frac{\rho_c}{\epsilon_0} \\
\nabla \cdot \mathbf{B} = 0
\end{aligned}
$$

where $\rho_c = \sum_s (q_s/m_s)\rho_s$ is the total charge density and $\mathbf{j} = \sum_s (q_s/m_s)\rho_s \mathbf{u}_s$ is the current density.

The last two equations are constraints on the initial conditions; these are not guaranteed to hold numerically. Classical tricks involves using hyperbolic/parabolic cleaning or a facially-collocated Yee-type mesh.

5.4.3 Five/Ten-moment

(Wang et al. 2020)

- Continuity equation for each species

$$
\frac{\partial (m_s n_s)}{\partial t} + \frac{\partial (m_s n_s u_{j,s})}{\partial x_j} = 0
$$

- Momentum equation for each species

$$
n_s m_s \frac{\mathrm{d}\mathbf{u}_s}{\mathrm{d}t} = n_s q_s (\mathbf{E} + \mathbf{u} \times \mathbf{B}) - \nabla \cdot \overleftrightarrow{P}_s - \mathbf{R}_{s\alpha}
$$

or in Einstein's notation,

$$
\frac{\partial (m_s n_s u_{i,s})}{\partial t} = n_s q_s (E_i + \epsilon_{ijk} u_{j,s} B_k) - \frac{\partial \mathcal{P}_{ij,s}}{\partial x_j} - R_{i,s\alpha}
$$

where ϵ_{ijk} is the Levi-Civita symbol. The moments are defined as

$$
\begin{aligned}
n_s(\mathbf{x}) &\equiv \int f_s \mathrm{d}\mathbf{v} \\
u_{i,s}(\mathbf{x}) &\equiv \frac{1}{n_s(\mathbf{x})} \int v_i f_s \mathrm{d}\mathbf{v} \\
\mathcal{P}_{ij,s}(\mathbf{x}) &\equiv m_s \int v_i v_j f_s \mathrm{d}\mathbf{v}
\end{aligned}
$$

with $f_s(\mathbf{x}, \mathbf{v}, t)$ being the phase space distribution function. We will neglect the subscript s hereinafter for convenience. For completeness, $\mathcal{P}_{ij}$ relates to the more familiar thermal pressure tensor

$$P_{ij} \equiv m \int (v_i - u_i)(v_i - u_i) f \mathrm{d}\mathbf{v}$$

by

$$\mathcal{P}_{ij} = P_{ij} + nmu_i u_j$$

For simplicity, non-ideal effects like viscous dissipation are neglected. The electric and magnetic fields $\mathbf{E}$ and $\mathbf{B}$ are evolved using Maxwell equations

$$\begin{aligned}
\frac{\partial \mathbf{B}}{\partial t} + \nabla \times \mathbf{E} &= 0 \\
\frac{\partial \mathbf{E}}{\partial t} - c^2 \nabla \times \mathbf{B} &= -\frac{1}{\epsilon_0} \sum_s q_s n_s \mathbf{u}_s \\
\nabla \cdot \mathbf{B} &= 0 \\
\nabla \cdot \mathbf{E} &= \epsilon_0^{-1} \sum_s n_s q_s
\end{aligned}$$

with $c = 1/\sqrt{(}\mu_0 \epsilon_0)$ being the speed of light.

To close the system, the second order moment $\mathcal{P}_{ij}$ or P_{ij} must be specified. For example, a cold fluid closure simply sets $P_{ij} = 0$, while an isothermal equation of state (EOS) assumes that the temperature is constant. Or, assuming zero heat flux and that the pressure tensor is isotropic, we can write an adiabatic EOS for $P_{ij} = pI_{ij}$,

$$\frac{\partial \epsilon}{\partial t} + \nabla \cdot [(p + \epsilon)\mathbf{u}] = nq\mathbf{u} \cdot \mathbf{E}$$

where

$$\epsilon \equiv \frac{p}{\gamma - 1} + \frac{1}{2} mn\mathbf{u}^2$$

is the total fluid (thermal + kinetic) energy and γ is the adiabatic index, setting to 5/3 for a fully ionized plasma. For a plasma with S species ($s = 1, ..., S$) this system is closed and has a total of $5S + 6$ equations, and are here referred to as the *five-moment* model. More general models can be obtained by retaining the evolution equations for all six components of the pressure tensor in the so-called *ten-moment* model

$$\frac{\mathcal{P}_{ij,s}}{\partial t} + \frac{\partial \mathcal{Q}_{ijm,s}}{\partial x_m} = n_s q_s u_{i,s} E_j + \frac{q_s}{m_s} \epsilon_{iml} \mathcal{P}_{mj,s} B_l \tag{5.16}$$

where the third moment

$$\mathcal{Q}_{ijm,s}(\mathbf{x}) \equiv m_s \int v_i v_j v_m f_s \mathrm{d}\mathbf{v}$$

relates to the heat flux tensor defined in the fluid frame

$$Q_{ijm} \equiv m \int (v_i - u_i)(v_j - u_j)(v_m - u_m) \mathrm{d}\mathbf{v}$$

by

$$\mathcal{Q}_{ijm} = Q_{ijm} + u_i \mathcal{P}_{jm} - 2nm u_i u_j u_m$$

Again, the equations here must be closed by some approximation for the heat-flux tensor. Another option is to include evolution equations for even higher order moments, e.g., the ten independent components of the heat-flux tensor.

Theoretically, the multifluid-Maxwell equations approach the Hall magnetohydrodynamics (MHD) under asymptotic limits of vanishing electron mass ($m_e \to 0$) and infinite speed of light ($c \to \infty$). All waves and effects within the two-fluid picture are retained, for example, the light wave, electron and ion inertial effects like the ion cyclotron wave and whistler wave. Particularly, through properly devised heat-flux closures, the ten-moment model could partially capture nonlocal kinetic effects like Landau damping, in a manner similar to the gyrokinetic models.

Although multifluid-Maxwell models provide a more complete description of the plasma than reduced, asymptotic models like MHD, they are less frequently used. The reason for this is the fast kinetic scales involved. Retaining the electron inertia adds plasma-frequency and cyclotron time-scale, while non-neutrality adds Debye length spatial-scales. Further, inclusion of the displacement currents means that EM waves must be resolved when using an explicit scheme. Fortunately, the restrictions due to kinetic scales are introduced *only through the non-hyperbolic source terms* of the momentum equation, the Ampère's law, and the pressure equation. Therefore we may eliminate these restrictions by *updating the source term separately either exactly* or *using an implicit algorithm* (Note: BATSRUS applies the point-implicit scheme.). This allows larger time steps and leads to significant speedup, especially with realistic electron/ion mass ratios. The speed of light constraint still exists, however, can be greatly relaxed, using reduced values for the speed of light and/or sub-cycling Maxwell equations. Of course, an implicit Maxwell solver, or a reduced set of electromagnetic equations like the Darwin approximation (???), can also relax the time-step restrictions.

5.4.4 Characteristic wave speeds

The fastest wave speed in the 5/6/10-moment equations is the speed of light c. Theoretically, the characteristic wave speeds shall be consistent and fallback to MHD in the isotropic 5-moment case.

Numerically, however, using c in the numerical fluxes makes the scheme rather diffusive (?). The discretization time step is limited by the Courant-Friedrichs-Lewy (CFL) condition based on the speed of light. In practice, we can reduce the speed of light to a value that is a factor of 2-3 faster than the fastest flow and fast wave speed to speed up the simulation.

5.5 Generalized Ohm's Law

DERIVATIONS TO BE ADDED...

The generalized Ohm's law can be derived from two-fluid equations:

$$\mathbf{E} = -\mathbf{u}\times\mathbf{B} + \eta\mathbf{j} + \frac{1}{en}\mathbf{j}\times\mathbf{B} - \frac{1}{en}\nabla\cdot\overleftrightarrow{P}_e + \frac{m_e}{ne^2}\Big[\frac{\partial\mathbf{j}}{\partial t} + \nabla\cdot(\mathbf{ju}+\mathbf{uj})\Big] \tag{5.17}$$

where the first term on the right-hand side is the *convection term*, the second term is the *resistivity term* (*conductive term*), the third term is called the *Hall term*, the fourth term is the *electron pressure term*, and the fifth term is called the *electron inertia term*, since it is proportional to the mass of electrons.

Note that both $\mathbf{u}$ and $\mathbf{j}$ are the first-order moments, with $\mathbf{u}$ being the (weighted) sum of the first-order moment of electrons and ions while $\mathbf{j}$ being the difference between them. The generalized Ohm's law is actually the difference between the electrons' and ions' first-order moment equations. *The generalized Ohm's law is an equation that governs the time evolution of* $\mathbf{j}$. Also note that Ampère's law, with the displacement current retained, is an equation governing the time evolution of $\mathbf{E}$. However, in the approximation of the resistive MHD, the time derivative terms $\partial\mathbf{E}/\partial t$ and $\partial\mathbf{j}/\partial t$ are ignored in Ampère's law and Ohm's law, respectively. In this approximation, Ohm's law is directly solved to determine $\mathbf{E}$ and Ampère's law is directly solved to determine $\mathbf{j}$. [Introduction to plasma physics: with space and laboratory applications, D. A. Gurnett and A. Bhattacharjee.]

$$\underbrace{\mathbf{E}}_{a} + \underbrace{\mathbf{u}\times\mathbf{B}}_{b} - \underbrace{\frac{1}{\sigma}}_{c}\mathbf{j} - \underbrace{\frac{\mathbf{j}\times\mathbf{B}}{en_e}}_{d} + \underbrace{\frac{1}{n_e e}\nabla(n_e k_B T_e)}_{e} = 0$$

Denote each term above with a to e. From the single-fluid MHD momentum equation, let ω be the oscillation frequency of the perturbed velocity, and variable in scalar form be the characteristic magnitude of that quantity,

$$\begin{aligned}\rho(\frac{\partial\mathbf{u}}{\partial t} + \mathbf{u}\cdot\nabla\mathbf{u}) &= \mathbf{j}\times\mathbf{B} - \nabla p + \rho\mathbf{g} \\ &\Rightarrow \omega\rho U \sim JB\end{aligned} \tag{5.18}$$

Using and the assumptions of MHD, the relations between each term in generalized Ohm's law are

$$\begin{aligned}\frac{b}{d} &= \frac{UB}{JB/en_e} = \frac{en_eUB}{\omega\rho U} = \frac{en_eB}{\omega n_e m_i} = \frac{\Omega_i}{\omega} \gg 1 \\ \frac{c}{d} &= \frac{J/\sigma}{JB/en_e} = \frac{en_e}{B\sigma} = \frac{en_e}{Be^2n_e/\nu_{ei}m_e} = \frac{\nu_{ei}}{\Omega_e} \ll 1 \\ \frac{b}{d} &= \frac{U}{J/en_e} \ll 1,\ \text{if } \mathbf{J} \text{ is carried by } \mathbf{u}_e\end{aligned}$$

which generates direct contradiction with the MHD assumption that $\mathbf{u} \ll \mathbf{u}_e$ if currents are mostly carried by electrons. (WHAT ABOUT THE PRESSURE TERM SCALING? The gradient implies that it is related to the system size; the pressure implies that it is also related to thermal motion?)

There is an interesting point about the electron pressure term. If we assume an isotropic plasma with $n_i = n_e$ and an adiabatic process $P/n^\gamma = C$ where C is a constant and γ is the adiabatic index, we have

$$\begin{aligned}\mathbf{E}_{\nabla P_e} &= \frac{\nabla P_e}{n} = \frac{\nabla P_i}{n} = \frac{\nabla (C n^\gamma)}{n} \\ &= C\gamma n^{\gamma-2}\nabla n = C\frac{\gamma}{\gamma-1}\nabla n^{\gamma-1}\end{aligned} \tag{5.19}$$

Equation 5.19 indicates that the associated electric field is a potential field that only relates to density.

5.6 Magnetohydrodynamics

Particle motion in the two-fluid system was described by the individual species' mean velocities $\mathbf{u}_e, \mathbf{u}_i$ and by the pressures $\overleftrightarrow{P}_e, \overleftrightarrow{P}_i$, which provide an accounting for the mean square of the random deviation of the velocity from its average value. Magnetohydrodynamics is an alternate description of the plasma where, instead of using $\mathbf{u}_e, \mathbf{u}_i$ to describe mean motion, two new velocity variables that are a linear combination of $\mathbf{u}_e, \mathbf{u}_i$ are used. As will be seen below, this means a slightly different definition for pressure must also be used.

The new velocity-like variables are

1. the current density
$$\mathbf{j} = \sum_s n_s q_s \mathbf{u}_s$$
which is essentially the relative velocity between ions and electrons, and
2. the center-of-mass velocity
$$\mathbf{u} = \frac{1}{\rho}\sum_s m_s n_s \mathbf{u}_s$$
where
$$\rho = \sum_s m_s n_s$$
is the total mass density.

Magnetohydrodynamics is primarily concerned with *low-frequency*, *long-wavelength*, *magnetic* behavior of the plasma. There are three basic assumptions in MHD:

$$
\begin{aligned}
\epsilon\omega/4\pi\sigma &\ll 1, \\
(v/c)^2 &\ll 1, \\
\lambda/L &\ll 1,
\end{aligned}
$$

where ω is the plasma frequency, v is the plasma bulk speed, λ is the average free distance, and L is the system characteristic length.

Standard orderings of ideal MHD can also be written as

$$
\epsilon \sim \omega/\Omega_c \sim k\rho
$$

where the plasma varies on frequency scales ω small compared to the gyrofrequency Ω_c, and varies on spatial scales $1/k$ long compared to the gyroradius ρ. (Here, ϵ means a "small" value.) Thus it covers phenomenon related to compressional and shear Alfvén waves and instabilities, ion acoustic waves, and ion and electron kinetic effects such as Landau damping. However, it does not include drift-waves or other micro-instabilities because they result from finite Larmor radius (FLR) effects which vanish in the usual MHD ordering.

Single fluid MHD is somehow inconsistent since there is only one velocity. The definition of current using velocity cannot be applied, and the current can only be given by Ampère's law (without the displacement current), $\mathbf{J} = \nabla \times \mathbf{B}/\mu_0$.

Multiplying Equation 5.5 by m_s and summing over species gives the MHD continuity equation

$$
\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0 \tag{5.20}
$$

To obtain an MHD equation of motion we take the first moment of the Vlasov equation and multiply the single species momentum equation by m_s and sum over species,

$$
\frac{\partial}{\partial t}\sum_s m_s \int \mathbf{v} f_s \mathrm{d}\mathbf{v} + \frac{\partial}{\partial \mathbf{x}} \cdot \sum_s \int m_s \mathbf{v}\mathbf{v} f_s \mathrm{d}\mathbf{v} + \sum_s q_s \int \mathbf{v} \frac{\partial}{\partial \mathbf{v}} \cdot [(\mathbf{E} + \mathbf{v} \times \mathbf{B}) f_s] = 0 \tag{5.21}
$$

The right-hand side is zero since $\mathbf{R}_{ei} + \mathbf{R}_{ie} = 0$, i.e., the total plasma cannot exert drag on itself. We now define random velocities relative to $\mathbf{u}$ (rather than to $\mathbf{u}_s$ as was the case for the two-fluid equations) so that the second term can be written as

$$
\sum_s \int m_s \mathbf{v}\mathbf{v} f_s \mathrm{d}\mathbf{v} = \sum_s \int m_s (\mathbf{v}' + \mathbf{u})(\mathbf{v}' + \mathbf{u}) f_s \mathrm{d}\mathbf{v} = \sum_s \int m_s \mathbf{v}'\mathbf{v}' f_s \mathrm{d}\mathbf{v} + \rho \mathbf{u}\mathbf{u} \tag{5.22}
$$

where $\sum_s \int m_s \mathbf{v}' f \mathrm{d}\mathbf{v} = 0$ has been used to eliminate terms linear in $\mathbf{v}'$. The MHD pressure tensor is now defined in terms of the random velocities relative to $\mathbf{u}$ and is given by

$$
\overleftrightarrow{P}^{\mathrm{MHD}} = \sum_s \int m_s \mathbf{v}'\mathbf{v}' f_s \mathrm{d}\mathbf{v} \tag{5.23}
$$

We insert Equation 5.22 and Equation 5.23 in Equation 5.21, integrate by parts on the acceleration term, and perform the summation over species to obtain the MHD equation of motion

$$\frac{\partial(\rho\mathbf{u})}{\partial t}+\nabla\cdot(\rho\mathbf{u}\mathbf{u})=\left(\sum_s n_s q_s\right)\mathbf{E}+\mathbf{j}\times\mathbf{B}-\nabla\cdot\overleftrightarrow{P}^{\mathrm{MHD}} \tag{5.24}$$

MHD is typically used to describe phenomena having spatial scales large enough for the plasma to be essentially neutral, i.e., $\sum_s n_s q_s = 0$ so that the $\mathbf{E}$ term can be dropped. Just as in the two-fluid situation, the left-hand side of Equation 5.24 contains a factor times the MHD continuity equation, since the left-hand side can be expanded as

$$\begin{aligned}\frac{\partial(\rho\mathbf{u})}{\partial t}+\nabla\cdot(\rho\mathbf{u}\mathbf{u}) &= \left[\cancel{\frac{\partial\rho}{\partial t}+\nabla\cdot(\rho\mathbf{u})}\right]\mathbf{u}+\rho\frac{\partial\mathbf{u}}{\partial t}+\rho\mathbf{u}\cdot\nabla\mathbf{u}\\ &= \rho\left(\rho\frac{\partial\mathbf{u}}{\partial t}+\rho\mathbf{u}\cdot\nabla\mathbf{u}\right)\end{aligned}$$

This leads to the standard form for the MHD equation of motion,

$$\rho\frac{\mathrm{D}\mathbf{u}}{\mathrm{D}t}=\mathbf{j}\times\mathbf{B}-\nabla\cdot\overleftrightarrow{P}^{\mathrm{MHD}} \tag{5.25}$$

where

$$\frac{\mathrm{D}}{\mathrm{D}t}\equiv\frac{\partial}{\partial t}+\mathbf{u}\cdot\nabla$$

is the convective derivative defined using the center-of-mass velocity. This notation is used to emphasize the difference from single species treatment. Scalar approximations of the MHD pressure tensor will be postponed until after discussing the MHD Ohm's law and its implications.

5.6.1 MHD Ohm's law

Equation 5.25 provides one equation relating $\mathbf{j}$ and $\mathbf{u}$; let us now find the other one. In order to do this, consider the two-fluid electron equation of motion,

$$m_e\frac{\mathrm{d}\mathbf{u}_e}{\mathrm{d}t}=-e(\mathbf{E}+\mathbf{u}_e\times\mathbf{B})-\frac{1}{n_e}\nabla(n_e k_B T_e)-\nu_{ei}m_e(\mathbf{u}_e-\mathbf{u}_i) \tag{5.26}$$

In MHD we are interested in low-frequency phenomena with large spatial scales. If the characteristic time scale of the phenomenon is long compared to the electron cyclotron motion, then the electron inertia term $m_e\mathrm{d}\mathbf{u}_e/\mathrm{d}t$ can be dropped since it is small compared to the magnetic force term $-e(\mathbf{u}_e\times\mathbf{B})$. This assumption is reasonable for velocities perpendicular to $\mathbf{B}$, but can be a poor approximation for the velocity component parallel to $\mathbf{B}$, since parallel velocities do not provide a magnetic force. Since $\mathbf{u}_e - \mathbf{u}_i = -\mathbf{j}/n_e e$ and $\mathbf{u}_i \simeq \mathbf{u}$, Equation 5.26 reduces to the generalized Ohm's law

$$\mathbf{E} + \mathbf{u} \times \mathbf{B} - \frac{1}{n_e e}\mathbf{j} \times \mathbf{B} + \frac{1}{n_e e}\nabla(n_e k_B T_e) = \eta\mathbf{j} \tag{5.27}$$

(What if we consider ion species with charge larger than 1?) The term $-\mathbf{j} \times \mathbf{B}/n_e e$ on the left-hand side is called the *Hall term* and can be neglected in either of the following two cases:

1. The pressure term in the MHD equation of motion, Equation 5.25 is negligible compared to the other two terms, which therefore must balance, giving

$$|\mathbf{j}| \sim \omega\rho|\mathbf{u}|/|\mathbf{B}|$$

Here $\omega \sim \mathrm{D}/\mathrm{D}t$ is the characteristic frequency of the phenomenon. In this case comparison of the Hall term with the U×B term shows that the Hall term is small by a factor $\sim \omega/\omega_{ci}$, where $\omega_{ci} = q_i B/m_i$ is the ion cyclotron frequency. Thus dropping the Hall term is justified for phenomena having characteristic frequencies small compared to ω_{ci}.

2. The electron-ion collision frequency is large compared to the electron cyclotron frequency $\omega_{ce} = q_e B/m_e$, in which case the Hall term may be dropped since it is small by a factor ω_{ce}/ν_{ei} compared to the right-hand side resistive term $\eta\mathbf{j} = (m_e\nu_{ei}/n_e e^2)\mathbf{j}$.

5.6.2 Ideal MHD

If the Hall term is dropped, the system is called ideal MHD. Typically, Equation 5.27 will not be used directly; instead its curl will be used to provide the induction equation

$$-\frac{\partial\mathbf{B}}{\partial t} + \nabla \times (\mathbf{u} \times \mathbf{B}) - \frac{1}{n_e e}\nabla n_e \times \nabla k_B T_e = \nabla \times \left(\frac{\eta}{\mu_0}\nabla \times \mathbf{B}\right)$$

Usually the density gradient is parallel to the temperature gradient so that the thermal electromotive force term $(n_e e)^{-1}\nabla n_e \times \nabla k_B T_e$ can be dropped, or the thermal term is often simply ignored, in which case the induction equation reduces to

$$-\frac{\partial\mathbf{B}}{\partial t} + \nabla \times (\mathbf{u} \times \mathbf{B}) = \nabla \times \left(\frac{\eta}{\mu_0}\nabla \times \mathbf{B}\right) \tag{5.28}$$

or written in the form of Ohm's law

$$\mathbf{E} + \mathbf{u} \times \mathbf{B} = \eta\mathbf{j} \tag{5.29}$$

If the resistive term $\eta\mathbf{j}$ is so small as to be negligible compared to the other terms in Equation 5.29, then the plasma is said to be *ideal* or *perfectly conducting*. From the Lorentz transformation of electromagnetic theory we realize that

$$\mathbf{E} + \mathbf{u} \times \mathbf{B} = \mathbf{E}'$$

where $\mathbf{E}'$ is the electric field observed in the frame moving with velocity $\mathbf{u}$. This implies that the magnetic flux in ideal plasmas is time-invariant in the frame moving with velocity $\mathbf{u}$, because otherwise Faraday's law would imply the existence of an electric field in the moving frame. The frozen-in flux concept is the essential "bed-rock" concept underlying ideal MHD. While this concept is often an excellent approximation, it must be kept in mind that the concept becomes invalid in situations when any one of the electron inertia, electron pressure, or Hall terms becomes important and leads to different, more complex behavior.

The frozen-in flux concept is frequently expressed in a slightly different form as a frozen-in field concept, i.e., it is often said that magnetic field lines are frozen into the plasma so that the plasma and magnetic field lines move together as an ensemble. While this point of view can be quite intuitive and useful, it contains some ambiguity because an ideal plasma can actually move across magnetic field lines in certain situations. These situations are such that magnetic flux is preserved within the plasma even though it is moving across field lines.

A formal proof of the frozen-in flux property will now be established by direct calculation of the rate of change of the magnetic flux through a surface $S(t)$ bounded by a material line $C(t)$, i.e., a closed contour that moves with the plasma. This magnetic flux is

$$\Phi(t) = \int_{S(t)} \mathbf{B}(\mathbf{x}, t) \cdot \mathrm{d}\mathbf{s}$$

and hte flux changes with respect to time due to either (i) the explicit time dependence of $\mathbf{B}(t)$ or (ii) changes in the surface $S(t)$ resulting from plasma motion. The rate of change of flux is thus

$$\frac{\mathrm{D}\Phi}{\mathrm{D}t} = \lim_{\delta t \to 0} \left(\frac{\int_{S(t+\delta t)} \mathbf{B}(\mathbf{x}, t+\delta t) \cdot \mathrm{d}\mathbf{s} - \int_{S(t)} \mathbf{B}(\mathbf{x}, t) \cdot \mathrm{d}\mathbf{s}}{\delta t} \right)$$

The displacement of a segment $\mathrm{d}\mathbf{l}$ of the bounding contour C is $\mathbf{u}\delta t$, where $\mathbf{u}$ is the velocity of this segment. The incremental change in surface area due to this displacement is $\Delta S = \mathbf{u}\delta t \times \mathrm{d}\mathbf{l}$. The rate of change of flux can thus be expressed as

$$\begin{aligned}
\frac{\mathrm{D}\Phi}{\mathrm{D}t} &= \lim_{\delta t \to 0} \left(\frac{\int_{S(t+\delta t)} \left(\mathbf{B} + \delta t \frac{\partial \mathbf{B}}{\partial t}\right) \cdot \mathrm{d}\mathbf{s} - \int_{S(t)} \mathbf{B} \cdot \mathrm{d}\mathbf{s}}{\delta t} \right) \\
&= \lim_{\delta t \to 0} \left(\frac{\int_{S(t)} \left(\mathbf{B} + \delta t \frac{\partial \mathbf{B}}{\partial t}\right) \cdot \mathrm{d}\mathbf{s} + \oint_C \mathbf{B} \cdot \mathbf{u}\delta t \times \mathrm{d}\mathbf{l} - \int_{S(t)} \mathbf{B} \cdot \mathrm{d}\mathbf{s}}{\delta t} \right) \\
&= \int_{S(t)} \frac{\partial \mathbf{B}}{\partial t} \cdot \mathrm{d}\mathbf{s} + \oint_C \mathbf{B} \cdot \mathbf{u} \times \mathrm{d}\mathbf{l} \\
&= \int_{S(t)} \frac{\partial \mathbf{B}}{\partial t} \cdot \mathrm{d}\mathbf{s} + \oint_C \mathrm{d}\mathbf{l} \cdot (\mathbf{B} \cdot \mathbf{u}) \\
&= \int_{S(t)} \left[\frac{\partial \mathbf{B}}{\partial t} + \nabla \times (\mathbf{B} \times \mathbf{u}) \right] \cdot \mathrm{d}\mathbf{S}
\end{aligned}$$

Thus, if

$$\frac{\partial \mathbf{B}}{\partial t} + \nabla \times (\mathbf{B} \times \mathbf{u}) = 0 \tag{5.30}$$

then

$$\frac{\mathrm{D}\Phi}{\mathrm{D}t} = 0$$

so that the magnetic flux linked by any closed material line is constant. Therefore, magnetic flux is frozen into an ideal plasma because Equation 5.28 reduces to Equation 5.30 if $\eta = 0$. Equation 5.30 is called the ideal MHD induction equation.

5.6.2.1 Equation of state

Double adiabatic laws

A procedure analogous to that which led to Equation 5.11 or Equation 8.28 gives the MHD adiabatic relation

$$\frac{p^{\mathrm{MHD}}}{\rho^{\gamma}} = \mathrm{const.}$$

where again $\gamma = (N+2)/N$ and N is the number of dimensions of the system. It was shown in the previous section that magnetic flux is conserved in the plasma frame. This means that, as shown in Figure 5.1, a tube of plasma initially occupying the same volume as a magnetic flux tube is constrained to evolve in such a way that $\int \mathbf{B} \cdot \mathrm{d}\mathbf{s}$ stays constant over the plasma tube cross-section. For a flux tube of infinitesimal cross-section, the magnetic field is approximately uniform over the cross-section and we may write this as $BA = \mathrm{const}$, where A is the cross-sectional area.

Let us define two temperatures for this magnetized plasma, namely $T_\perp$, the temperature corresponding to motions perpendicular to the magnetic field, and $T_\parallel$, the temperature corresponding to motions parallel to the magnetic field. If for some reason (e.g., anisotropic heating or compression) the temperature develops an anisotropy such that $T_\perp \neq T_\parallel$ and if collisions are infrequent, this anisotropy will persist for a long time, since collisions are the means by which the two temperatures equilibrate. Thus, rather than assuming that the MHD pressure is fully isotropic, we consider the less restrictive situation where the MHD pressure tensor is given by

$$\overleftrightarrow{P}^{\mathrm{MHD}} = \begin{bmatrix} p_\perp & 0 & 0 \\ 0 & p_\perp & 0 \\ 0 & 0 & p_\parallel \end{bmatrix} = p_\perp \mathbf{I} + (p_\parallel - p_\perp)\hat{b}\hat{b}$$

The first two coordinates (x, y-like) in the above matrix refer to the directions perpendicular to the local magnetic field $\mathbf{B}$ and the third coordinate (z-like) refers to the direction parallel to $\mathbf{B}$. The tensor expression on the right-hand side is equivalent (here $\mathbf{I}$ is the unit tensor) but allows for arbitrary, curvilinear geometry. We now develop separate adiabatic relations for the perpendicular and parallel directions:

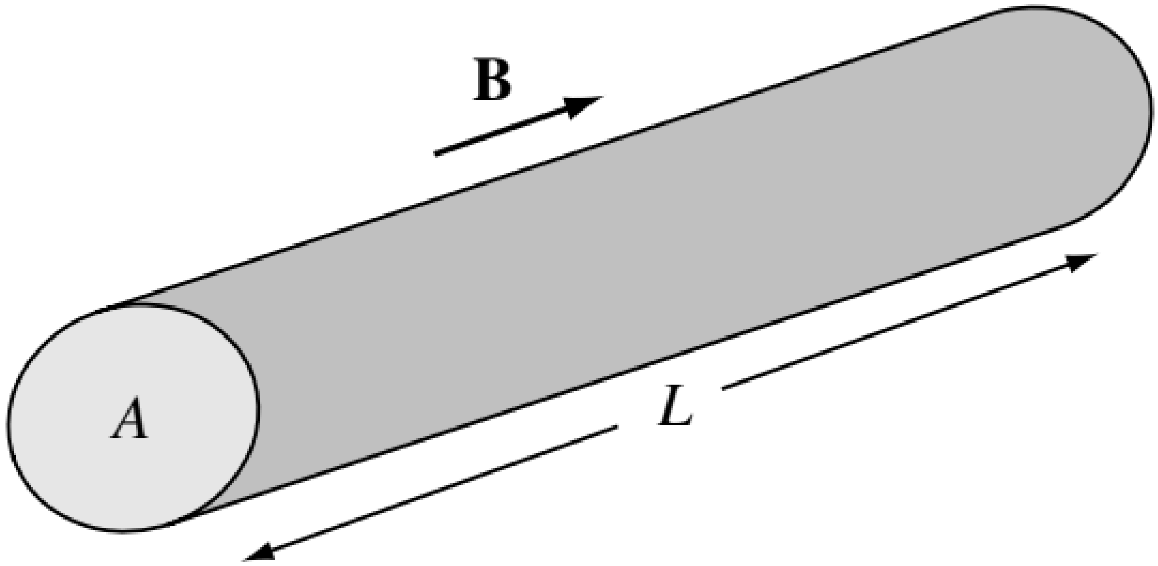

Figure 5.1: Magnetic flux tube with flux $\Phi = BA$.

- Parallel direction: here the number of dimensions is $N = 1$ so that $\gamma = 3$ and so the adiabatic law gives
$$\frac{p_{\parallel}^{1D}}{\rho_{1D}^{3}} = \text{const.} \tag{5.31}$$
where ρ_{1D} is the *one-dimensional* mass density; i.e., $\rho_{1D} \sim 1/L$, where L is the length along the flux tube in Figure 5.1. The three-dimensional mass density, which has been used implicitly until now, has the proportionality $\rho \sim 1/LA$, where A is the cross-section of the flux tube; similarly the three-dimensional pressure has the proportionality $p_{\parallel} \sim \rho T_{\parallel}$. However, we must be careful to realize that $p_{\parallel}^{1D} \sim \rho_{1D} T_{\parallel}$ so, using $BA = \text{const.}$, Equation 5.31 can be recast as
$$\text{const.} = \frac{p_{\parallel}^{1D}}{\rho_{1D}^{3}} \sim \frac{\rho_{1D} T_{\parallel}}{\rho_{1D}^{3}} \sim T_{\parallel} \rho_{1D}^{2} \sim \left(\frac{1}{LA}\right) T_{\parallel} (LA)^3 B^2$$
or
$$\frac{p_{\parallel} B^2}{\rho^3} = \text{const.} \tag{5.32}$$

- Perpendicular direction: here the number of dimensions is $N = 2$ so that $\gamma = 2$ and the adiabatic law gives
$$\frac{p_{\perp}^{2D}}{\rho_{2D}^{2}} = \text{const.} \tag{5.33}$$

where $p_\perp^{2D}$ is the 2-D perpendicular pressure, and has dimensions of energy per unit area, while ρ_{2D} is the 2-D mass density and has dimensions of mass per unit area. Thus, $\rho_{2D} \sim 1/A$ so $p_\perp^{2D} \sim \rho_{2D} T_\perp \sim T_\perp/A$, in which case Equation 5.33 can be recast as

$$\text{const.} = \frac{p_\parallel B^2}{\rho^3} = T_\perp A \sim \left(\frac{1}{LA}\right) T_\perp \frac{LA}{B}$$

or

$$\frac{p_\perp}{\rho B} = \text{const.} \tag{5.34}$$

Equation 5.32 and Equation 5.34 are called the double adiabatic or CGL laws after Chew, Goldberger, and Low (Chew, Goldberger, and Low 1956) who first developed these laws.

Single adiabatic law

If collisions are sufficiently frequent to equilibrate the perpendicular and parallel temperatures, then the pressure tensor becomes fully isotropic and the dimensionality of the system is $N = 3$ so that $\gamma = 5/3$. There is now just one pressure and one temperature and the adiabatic relation becomes

$$\frac{p}{\rho^{5/3}} = \text{const.} \tag{5.35}$$

5.6.2.2 MHD approximations for Maxwell's equations

The various assumptions contained in MHD lead to a simplifying approximation of Maxwell's equations. In particular, the assumption of charge neutrality in MHD makes Poisson's equation superfluous because Poisson's equation prescribes the relationship between non-neutrality and the electrostatic component of the electric field. The assumption of charge neutrality also has implications for the current density. To see this, the two-fluid continuity equation is multiplied by q_s and then summed over species to obtain the charge conservation equation

$$\frac{\partial}{\partial t}\left(\sum_s n_s q_s\right) + \nabla \cdot \mathbf{j} = 0 \tag{5.36}$$

Thus, charge neutrality implies

$$\nabla \cdot \mathbf{j} = 0$$

Let us now consider Ampère's law

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j} + \mu_0 \epsilon_0 \dot{\mathbf{E}}$$

Taking the divergence gives

$$\nabla \cdot$$

which is equivalent to Equation 5.36 if Poisson's equation is invoked.

Finally, MHD is restricted to phenomena having characteristic velocities V_{ph} slow compared to the speed of light in vacuum, $c = (\mu_0\epsilon_0)^{-1/2}$. Again, t_{char} is assumed to represent the characteristic time scale for a given phenomenon and l_{char} is assumed to represent the corresponding characteristic length scale so that $V_{\mathrm{ph}} \sim l_{\mathrm{char}}/t_{\mathrm{char}}$. Faraday's equation gives the scaling

$$\nabla \times \mathbf{E} = -\dot{\mathbf{B}} \to E \sim B l_{\mathrm{char}}/t_{\mathrm{char}}$$

On comparing the magnitude of the displacement current term in Ampère's law to the left-hand side it is seen that

$$\frac{\mu_0\epsilon_0|\dot{\mathbf{E}}|}{|\nabla \times \mathbf{B}|} \sim \frac{c^{-2}E/t_{\mathrm{char}}}{B/l_{\mathrm{char}}} \sim \left(\frac{V_{\mathrm{ph}}}{c}\right)^2$$

Thus, if $V_{\mathrm{ph}} \ll c$ the displacement current term can be dropped from Amperè's law resulting in the so-called "pre-Maxwell" form

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j} \tag{5.37}$$

The divergence of Equation 5.37 gives $\nabla \cdot \mathbf{j} = 0$ so it is unnecessary to specify it separately.

5.6.2.3 Complete equation set

In the most common sense, ideal MHD involves equations of compressible, adiabatic and inviscid fluid:

$$\begin{aligned}
\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) &= 0 \\
\frac{\partial \rho \mathbf{u}}{\partial t} + \nabla \cdot \left(\rho \mathbf{u}\mathbf{u} - \frac{1}{\mu_0}\mathbf{B}\mathbf{B} + p^*\right) &= 0 \\
\frac{\partial \mathcal{E}}{\partial t} + \nabla \cdot \left[(\mathcal{E} + p^*)\mathbf{u} - \frac{1}{\mu_0}\mathbf{B}(\mathbf{B}\cdot\mathbf{u})\right] &= 0 \\
\frac{\partial \mathbf{B}}{\partial t} - \nabla \times (\mathbf{u} \times \mathbf{B}) &= 0 \\
p^* = p + \frac{\mathbf{B}\cdot\mathbf{B}}{2\mu_0} & \\
\mathcal{E} = \frac{p}{\gamma - 1} + \frac{\rho \mathbf{u}\cdot\mathbf{u}}{2} + \frac{\mathbf{B}\cdot\mathbf{B}}{2\mu_0} &
\end{aligned} \tag{5.38}$$

where ρ is the mass density, $\mathbf{u}$ the velocity, $\mathcal{E}$ the total energy density, $\mathbf{B}$ the magnetic field, p the thermal pressure, and γ the adiabatic index (ratio of specific heats). Microscopic dissipations of any kind (viscosity, resistivity, or conduction) are not included in ideal MHD. Note that there is only one constant μ_0 appeared in Equation 5.38. The introduction of temperature comes with a new constant $R \equiv 2k_B/m$:

$$p = \rho\, R\, T$$

Many often in simulations, the dimensionless units are applied. Equation 5.38 under the dimensionless units can be written as

$$\begin{aligned}
\frac{\partial \rho_*}{\partial t_*} + \nabla_* \cdot (\rho_* \mathbf{u}_*) = 0 \\
\frac{\partial \rho_* \mathbf{u}_*}{\partial t_*} + \nabla_* \cdot (\rho_* \mathbf{u}_* \mathbf{u}_* - \mathbf{B}_* \mathbf{B}_* + p_*^*) = 0 \\
\frac{\partial \mathcal{E}_*}{\partial t_*} + \nabla_* \cdot [(\mathcal{E}_* + p_*^*)\mathbf{u}_* - \mathbf{B}_*(\mathbf{B}_* \cdot \mathbf{u}_*)] = 0 \\
\frac{\partial \mathbf{B}_*}{\partial t_*} - \nabla_* \times (\mathbf{u}_* \times \mathbf{B}_*) = 0 \\
p_*^* = p_* + \frac{\mathbf{B}_* \cdot \mathbf{B}_*}{2} \\
\mathcal{E}_* = \frac{p_*}{\gamma - 1} + \frac{\rho_* \mathbf{u}_* \cdot \mathbf{u}_*}{2} + \frac{\mathbf{B}_* \cdot \mathbf{B}_*}{2}
\end{aligned} \tag{5.39}$$

where the subscripts $*$ denotes the variables in dimensionless units, and the adiabatic equation of state is given as

$$p_* = \rho_* T_*$$

In more theoretical derivations, for simplicity the energy equation is substituted with either single adiabatic law Equation 5.35 or double adiabatic law Equation 5.32 and Equation 5.34. If we drop the isotropic assumption and instead use separate pressure equations for the parallel and perpendicular components as in Equation 5.14, we get the anisotropic MHD equations.

These equations provide a self-consistent description of phenomena that satisfy all the various assumptions we have made, namely:

1. The plasma is charge-neutral since characteristic lengths are much longer than a Debye length;
2. The characteristic velocity of the phenomenon under consideration is slow compared to the speed of light;
3. The pressure and density gradients are parallel, so there is no electrothermal EMF;
4. The time scale is long compared to both the electron and ion cyclotron periods.

Even though these assumptions are self-consistent, they may not accurately portray a real plasma and so MHD models, while intuitively appealing, must be used with caution.

Finally, it is worth mentioning that MHD plasmas can be categorized yet another way, namely according to the relative importance of the magnetic force $\mathbf{j} \times \mathbf{B}$ compared to the hydrodynamic force $\nabla \cdot \overleftrightarrow{P}$. If the magnetic force is negligible compared to the hydrodynamic force, then there is not much point in using MHD because in this case the system of equations is simply classical hydrodynamics. Thus, the only non-trivial MHD situations are where the magnetic and hydrodynamic forces are of comparable importance or where the magnetic force is much more important than the hydrodynamic force. Using Equation 5.37, the nominal ratio of the hydrodynamic force to the magnetic force is defined as

$$\beta = \frac{p}{B^2/2\mu_0} \sim \frac{\nabla \cdot \overleftrightarrow{P}}{\mathbf{j} \times \mathbf{B}} \sim \frac{p/L}{B^2/2\mu_0 L}$$

The characteristic gradient scale length L is assumed to be comparable for both types of forces and so cancels out in the comparison. Low-β plasmas are those where $B^2/2\mu_0$ is much larger than p so the hydrodynamic force is negligible compared to the magnetic force, whereas $\beta = \mathcal{O}(1)$ plasmas are those where the magnetic and hydrodynamic forces are comparable. The different β regime has important implications on the validity of many approximations seen in later chapters.

5.6.3 Hall MHD

As an extension to the ideal/resistive MHD model, Hall MHD decouples the electron and ion motions by retrieving the Hall term in the generalized Ohm's law

$$\mathbf{E} = -\mathbf{u} \times \mathbf{B} + \eta \mathbf{J} + \frac{1}{en}\mathbf{J} \times \mathbf{B} - \frac{\nabla p_e}{ne}$$

where the first term on the right-hand-side is the convective term, the second is the resistive term, the third is the Hall term, and the fourth is the electron pressure gradient term. The electron pressure scalar is simplified from the electron pressure tensor, which is obtained from an independent electron pressure equation.

The Hall MHD equations are

$$\begin{aligned}
\frac{\partial \rho}{\partial t} &= -\nabla \cdot (\rho \mathbf{u}) \\
\frac{\partial (\rho \mathbf{u})}{\partial t} &= -\nabla \cdot \left(\rho \mathbf{u}\mathbf{u} + (p + p_e)\bar{\bar{I}} + \frac{B^2}{2\mu_0}\bar{\bar{I}} - \frac{\mathbf{B}\mathbf{B}}{\mu_0} \right) \\
\frac{\partial e}{\partial t} &= -\nabla \cdot \left[(\epsilon + p)\,\mathbf{u} + (\epsilon_e + p_e)\mathbf{u}_e + \mathbf{u}_e \cdot \left(\frac{\mathbf{B}^2}{\mu_0}\bar{\bar{I}} - \frac{\mathbf{B}\mathbf{B}}{\mu_0} \right) - \mathbf{B} \times \eta \mathbf{j} \right] \\
\frac{\partial \mathbf{B}}{\partial t} &= -\nabla \times \left[\mathbf{u}_e \times \mathbf{B} + \eta \mathbf{j} - \frac{\nabla p_e}{ne} \right] \\
\frac{\partial p_e}{\partial t} + \nabla \cdot (p_e \mathbf{u}_e) &= -(\gamma - 1) p_e \nabla \cdot \mathbf{u}_e
\end{aligned} \tag{5.40}$$

where $\bar{\bar{I}}$ is the identity matrix, ρ is the mass density, $\mathbf{u}$ is the plasma bulk velocity, $\mathbf{B}$ is the magnetic field, p_e is the electron pressure, p is the ion thermal pressure, $\mathbf{j} = \nabla\times\mathbf{B}/\mu_0$ is the current density, $\epsilon = \rho\mathbf{u}^2/2$ is the hydrodynamic energy density, and $\epsilon_e = \rho\mathbf{u}_e^2/2$ is the electron hydrodynamic energy density. The Hall velocity is defined as

$$\mathbf{v}_H = -\frac{\mathbf{j}}{ne}$$

and the electron bulk velocity is given by

$$\mathbf{u}_e = \mathbf{u} + \mathbf{v}_H$$

The total energy density is (I may have a coefficient error here for p_e!)

$$e = \frac{1}{2}\rho\mathbf{u}^2 + \frac{1}{\gamma-1}(p+p_e) + \frac{\mathbf{B}^2}{2\mu_0}$$

where γ is the adiabatic index. Note that in Hall MHD only $(\rho, \mathbf{u}, \mathbf{B}, p, p_e)$ are unknowns; all others are derived quantities.

5.7 Fluid Drifts Perpendicular to B

Since a fluid element is composed of many individual particles, one would expect the fluid to have drifts perpendicular to $\mathbf{B}$ if the individual guiding centers have such drifts. However, since the ∇p term appears only in the fluid equations, there is a drift associated with it which the fluid elements have but the particles do not have. For each species, we have an equation of motion

$$mn\Big[\underbrace{\frac{\partial\mathbf{u}}{\partial t}}_{1} + \underbrace{(\mathbf{u}\cdot\nabla)\mathbf{u}}_{2}\Big] = qn(\mathbf{E} + \underbrace{\mathbf{u}\times\mathbf{B}}_{3}) - \nabla p \tag{5.41}$$

Consider the ratio of term 1 to term 3:

$$\frac{\text{term 1}}{\text{term 3}} \approx \left|\frac{mni\omega v_\perp}{qnv_\perp B}\right| \approx \frac{\omega}{\omega_c}$$

Here we have taken $\partial/\partial t = i\omega$ and are concerned only with $\mathbf{v}_\perp$. For drifts slow compared with the time scale associated with ω_c, we may neglect term 1. We shall also neglect the $(\mathbf{u}\cdot\nabla)\mathbf{u}$ term and show a *posteriori* that this is all right. Let $\mathbf{E}$ and $\mathbf{B}$ be uniform, but let n and p have a gradient. This is the usual situation in a magnetically confined plasma column (Fig. 3.4 ADD FIGURE!). Taking the cross product of Equation 5.41 with B, we have (neglecting the left-hand side)

$$
\begin{aligned}
0 &= qn[\mathbf{E}\times\mathbf{B} + (\mathbf{u}_\perp\times\mathbf{B})\times\mathbf{B}] - \nabla p\times\mathbf{B} \\
&= qn[\mathbf{E}\times\mathbf{B} + \mathbf{B}\cancel{(\mathbf{u}_\perp\cdot\mathbf{B})} - \mathbf{u}_\perp\mathbf{B}^2] - \nabla p\times\mathbf{B} \\
&= qn[\mathbf{E}\times\mathbf{B} - \mathbf{u}_\perp\mathbf{B}^2] - \nabla p\times\mathbf{B}
\end{aligned}
$$

Therefore,

$$
\mathbf{u}_\perp = \frac{\mathbf{E}\times\mathbf{B}}{B^2} - \frac{\nabla p\times\mathbf{B}}{qnB^2} \equiv \mathbf{u}_E + \mathbf{u}_D
$$

where

$$
\mathbf{u}_E \equiv \frac{\mathbf{E}\times\mathbf{B}}{B^2} \quad \mathbf{E}\times\mathbf{B}\,\text{drift}
$$

$$
\mathbf{u}_D \equiv -\frac{\nabla p\times\mathbf{B}}{qnB^2} \quad \text{Diamagnetic drift} \tag{5.42}
$$

The drift $\mathbf{u}_E$ is the same as for guiding centers, but there is now a new drift $\mathbf{u}_D$, called the diamagnetic drift. Since $\mathbf{u}_D$ is perpendicular to the direction of the gradient, our neglect of $(\mathbf{u}\cdot\nabla)\mathbf{u}$ is justified if $\mathbf{E}=0$. If $\mathbf{E}=-\nabla\phi\neq 0$ (i.e. a potential field), $(\mathbf{u}\cdot\nabla)\mathbf{u}$ is still zero if $\nabla\phi$ and ∇p are in the same direction; otherwise, there could be a more complicated solution involving $(\mathbf{u}\cdot\nabla)\mathbf{u}$.

With the help of Equation 5.12, we can write the diamagnetic drift as

$$
\mathbf{u}_D = \pm\frac{\gamma k_B T}{eB}\frac{\hat{z}\times\nabla n}{n} \tag{5.43}
$$

In particular, for an isothermal plasma in the geometry of Fig.3.4 (ADD IT!), in which $\nabla n = \partial n \partial r\hat{r} = n'\hat{r}$ $(n' < 0)$, we have the following formulas familiar to experimentalists who have worked with Q-machines

$$
\begin{aligned}
\mathbf{u}_{Di} &= \frac{k_B T_i}{eB}\frac{n'}{n}\hat{\theta} \\
\mathbf{u}_{De} &= -\frac{k_B T_e}{eB}\frac{n'}{n}\hat{\theta}
\end{aligned}
$$

The physical reason for this drift can be seen from Figure 5.2. Here we have drawn the orbits of ions gyrating in a magnetic field. There is a density gradient toward the left, as indicated by the density of orbits. Through any fixed volume element there are more ions moving downward than upward, since the downward-moving ions come from a region of higher density. There is, therefore, a fluid drift perpendicular to ∇n and $\mathbf{B}$, *even though the guiding centers are stationary.* The diamagnetic drift reverses sign with q because the direction of gyration reverses. The magnitude of $\mathbf{u}_D$ does not depend on mass because the $m^{-1/2}$ dependence of the velocity is cancelled by the $m^{1/2}$ dependence of the Larmor radius—less of the density gradient is sampled during a gyration if the mass is small.

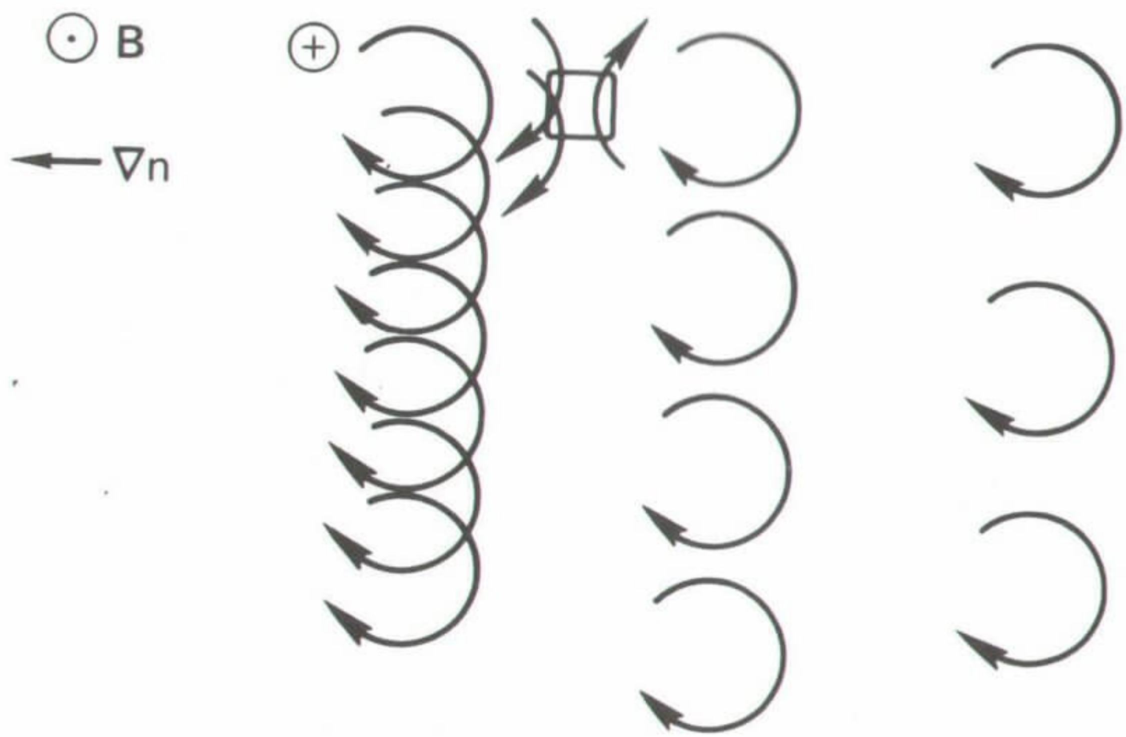

Figure 5.2: Origin of the diamagnetic drift.

Since ions and electrons drift in opposite directions, there is a diamagnetic current. For $\gamma = Z = 1$, this is given by

$$\mathbf{j}_D = ne(\mathbf{u}_{Di} + \mathbf{u}_{De}) = \frac{\mathbf{B}}{B^2} \times \nabla (p_i + p_e) = (k_B T_i + k_B T_e) \frac{\mathbf{B} \times \nabla n}{B^2} \tag{5.44}$$

In the particle picture, one would not expect to measure a current if the guiding centers do not drift. In the fluid picture, the current $\mathbf{j}_D$ flows wherever there is a pressure gradient. These two viewpoints can be reconciled if one considers that all experiments must be carried out in a finite-sized plasma. Suppose the plasma were in a rigid box (Fig. 3.6 ADD IT!). If one were to calculate the current from the single-particle picture, one would have to take into account the particles at the edges which have cycloidal paths. Since there are more particles on the left than on the right, there is a net current downward, in agreement with the fluid picture.

The reader may not be satisfied with this explanation because it was necessary to specify reflecting walls. If the walls were absorbing or if they were removed, one would find that electric fields would develop because more of one species—the one with larger Larmor radius—would collected than the other. Then the guiding centers would drift, and the simplicity of the model would be lost. Alternatively, one could imagine trying to measure the diamagnetic current with a current probe (Fig. 3.7 ADD IT!). This is just a transformer with a core of magnetic material. The primary winding is the plasma current threading the core, and the secondary is a multiturn winding all around the core. Let the whole thing be infinitesimally thin, so it does not intercept any particles. It is clear from Fig. 3.7 that a net upward current would be measured, there being higher density on the left than on the right, so that the diamagnetic current is a real current. From this example, one can see that it can be quite tricky to work with the single-particle picture. The fluid theory usually gives the right results when applied straightforwardly, even though it contains "fictitious" drifts like the diamagnetic drift.

What about the grad-B and curvature drifts which appeared in the single-particle picture? The curvature drift also exists in the fluid picture, since the centrifugal force is felt by all the particles in a fluid element as they move around a bend in the magnetic field. A term $\bar{F}_{cf} = \overline{nmv_\parallel^2}/R_c = nk_B T_\parallel / R_c$ has to be added to the right-hand side of the fluid equation of motion. This is equivalent to a gravitational force mng, with $g = k_B T_\parallel / mR_c$, and leads to a drift $\mathbf{u}_g = (m/q)(\mathbf{g} \times \mathbf{B})/B^2$, as in the particle picture (Equation 4.2).

The grad-B drift, however, does not exist for fluids. It can be shown on thermodynamic grounds that a magnetic field does not affect a Maxwellian distribution. This is because the Lorentz force is perpendicular to $\mathbf{v}$ and cannot change the energy of any particle. The most probable distribution $f(\mathbf{v})$ in the absence of $\mathbf{B}$ is also the most probable distribution in the presence of $\mathbf{B}$. If $f(\mathbf{v})$ remains Maxwellian in a nonuniform $\mathbf{B}$ field, and there is no density gradient, then the net momentum carried into any fixed fluid element is zero. There is no fluid drift even though the individual guiding centers have drifts; the particle drift in any fixed fluid element cancel out. To see this pictorially, consider the orbits of two particles moving through

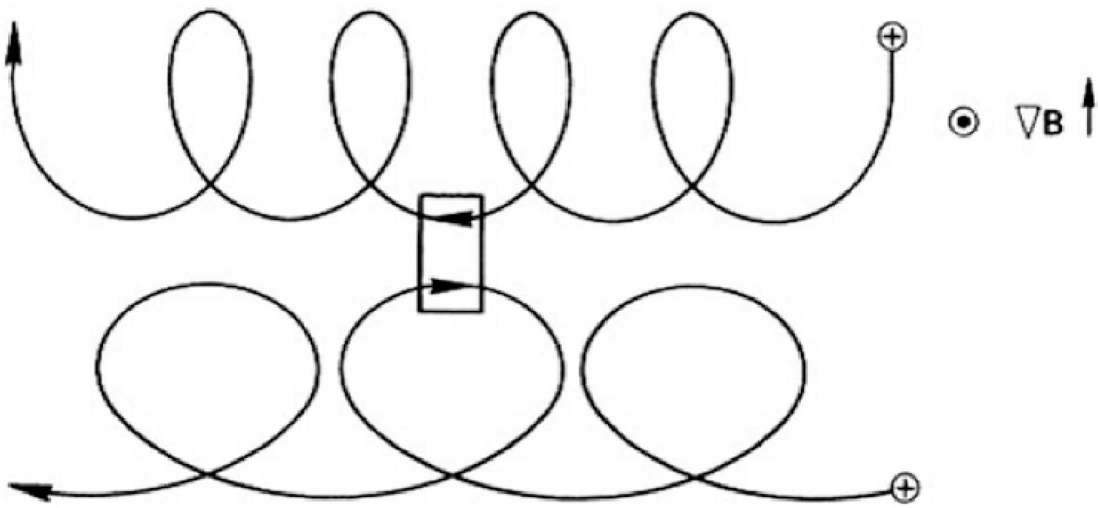

Figure 5.3: In a nonuniform B field the guiding centers drift but the fluid elements do not.

a fluid element in a nonuniform **B** field (Figure 5.3). Since there is no **E** field, the Larmor radius changes only because of the gradient in B; there is no acceleration, and the particle energy remains constant during the motion. If the two particles have the same energy, they will have the same velocity and Larmor radius while inside the fluid element. There is thus a perfect cancellation between particle pairs when their velocities are added to give the fluid velocity.

When there is a nonuniform **E** field, it is not easy to reconcile the fluid and particle pictures. Then the finite-Larmor-radius effect of Section 4.3 causes both a guiding center drift and a fluid drift, but these are not the same; in fact, they have opposite signs! The particle drift was calculated in Chapter 4, and the fluid drift can be calculated from the off-diagonal elements of **P**. It is extremely difficult to explain how the finite-Larmor-radius effects differ. A simple picture like Fig. 3.6 will not work because one has to take into account subtle points like the following: in the absence of a density gradient, the density of guiding centers is not the same as the density of particles! (???)

(I need to think carefully about these pictures.)

5.8 Fluid Drifts Parallel to B

The z component of the fluid equation of motions is

$$mn\left[\frac{\partial u_z}{\partial t} + (\mathbf{u}\cdot\nabla)u_z\right] = qnE_z - \frac{\partial \phi}{\partial z} \tag{5.45}$$

The convective term can often be neglected because it is much smaller than the $\partial u_z/\partial t$ term. We shall avoid complicated arguments here and simply consider cases in which u_z is spatially uniform. Using Equation 5.45, we have

$$\frac{\partial u_z}{\partial t} = \frac{q}{m}E_z - \frac{\gamma k_B T}{mn}\frac{\partial n}{\partial z} \tag{5.46}$$

This shows that the fluid is accelerated along **B** under the combined electrostatic and pressure gradient forces. A particularly important result is obtained by applying Equation 5.46 to massless electrons. Taking the limit $m \to 0$ and specifying $q = -e$ and $\mathbf{E} = -\nabla\phi$, we have

$$qE_z = e\frac{\partial \phi}{\partial z} = \frac{\gamma k_B T_e}{n}\frac{\partial n}{\partial z}$$

Electrons are so mobile that their heat conductivity is almost infinite. We may then assume isothermal electrons and take $\gamma = 1$. Integrating, we have

$$e\phi = k_B T_e \ln n + C$$

or

$$n = n_0 \exp(e\phi/k_B T_e) \tag{5.47}$$

This is just the *Boltzmann relation* for electrons.

What this means physically is that electrons, being light, are very mobile and would be accelerated to high energies very quickly if there were a net force on them. Since electrons cannot leave a region en masse without leaving behind a large ion charge, the electrostatic and pressure gradient forces on the electrons must be closely in balance. This condition leads to the Boltzmann relation. Note that Equation 5.47 *applies to each magnetic field lines separately.* Different field lines may be charged to different potentials arbitrarily unless a mechanism is provided for the electrons to move across **B**. The conductors on which field lines terminate can provide such a mechanism, and the experimentalist has to take these end effects into account carefully.

Figure 5.4 shows graphically what occurs when there is a local density clump in the plasma. Let the density gradient be toward the center of the diagram, and suppose $k_B T$ is constant. There is then a pressure gradient toward the center. Since the plasma is quasineutral, the gradient exists for both the electron and ion fluids. Consider the pressure gradient force $\mathbf{F}_p$ on the electron fluid. It drives the mobile electrons away from the center, leaving the ions behind. The resulting positive charge generates a field **E** whose force $\mathbf{F}_E$ on the electrons opposes $\mathbf{F}_p$. Only when $\mathbf{F}_E = -\mathbf{F}_p$ is a steady state achieved. If **B** is a constant, **E** is an electrostatic field $\mathbf{E} = -\nabla\phi$, and ϕ must be large at the center, where n is large. This is just what Equation 5.47 tells us. The deviation from stric neutrality adjusts itself so that there is just enough charge to set up the **E** field required to balance the forces on the electrons.

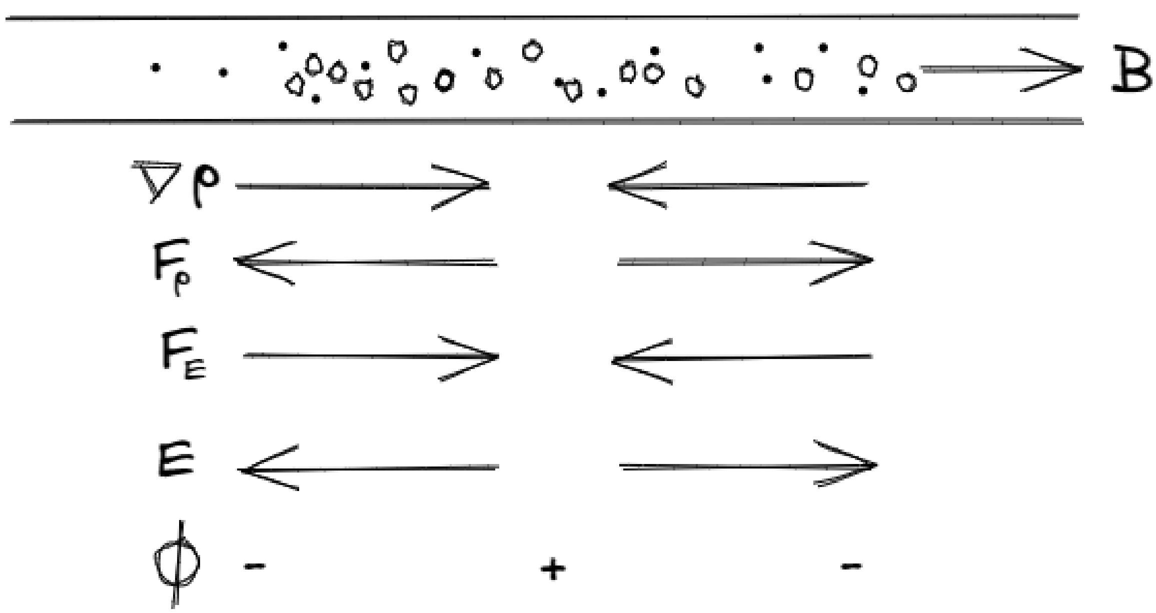

Figure 5.4: Physical reason for the Boltzmann relation between density and potential.

5.9 The Plasma Approximation

The previous example reveals an important characteristic of plasmas that has wide application. We are used to solving for $\mathbf{E}$ from Poisson's equation when we are given the charge density σ. In a plasma, the opposite procedure is generally used. $\mathbf{E}$ is found from the equations of motion, and Poisson's equation is used only to find σ. The reason is that a plasma has an overriding tendency to remain neutral. If the ions move, the electrons will follow. $\mathbf{E}$ must adjust itself so that the orbits of the electrons and ions preserve neutrality. The charge density is of secondary importance; it will adjust itself so that Poisson's equation is satisfied. This is true, of course, only for low-frequency motions in which the electron inertia is not a factor.

In a plasma, it is usually possible to assume $n_i = n_e$ and $\nabla \cdot \mathbf{E} \neq 0$ at the same time. We shall call this the *plasma approximation*. It is a fundamental trait of plasmas, one which is difficult for the novice to understand. *Do not use Poisson's equation to obtain* $\mathbf{E}$ *unless it is unavoidable*! In the set of fluid equations Equation 5.13, we may now eliminate Poisson's equation and also eliminate one of the unknowns by setting $n_i = n_e = n$.

The *plasma approximation* is almost the same as the condition of quasineutrality discussed earlier but has a more exact meaning. Whereas quasineutrality refers to a general tendency for a plasma to be neutral in its state of rest, the plasma approximation is a mathematical shortcut that one can use even for wave motions. As long as these motions are slow enough that both ions and electrons have time to move, it is a good approximation to replace Poisson's equation by the equation $n_i = n_e$. Of course, if only one species can move and the other cannot follow, such as in high-frequency electron waves, then the plasma approximation is not valid, and **E** must be found from Maxwell's Equation 5.1 rather than from the ion and electron equations of motion. Ampère's law with the displacement current retained gives the time evolution of **E**:

$$\frac{\mathrm{d}\mathbf{E}}{\mathrm{d}t} = \frac{1}{\epsilon_0}\left(\frac{\nabla \times \mathbf{B}}{\mu_0} - \mathbf{j}\right)$$

We shall return to the question of the validity of the plasma approximation when we come to the theory of ion waves. At that time, it will become clear why we had to use Poisson's equation in the derivation of Debye shielding. For electron waves, we can also show that with a simple "modified" electron test particle model that updates the electric field while keeping the magnetic field constant, we can retain the plasma oscillation and electron Berstein wave.

6 Diffusion and Resisvity

6.1 Diffusion and Mobility in Weakly Ionized Gases

The infinite, homogeneous plasmas assumed in the previous chapter for the equilibrium conditions are, of course, highly idealized. Any realistic plasma will have a density gradient, and the plasma will tend to diffuse toward regions of low density. For example, the central problem in controlled thermonuclear reactions is to impede the rate of diffusion by using a magnetic field. Before tackling the magnetic field problem, however, we shall consider the case of diffusion in the absence of magnetic fields. A further simplification results if we assume that the plasma is weakly ionized, so that charge particles collide primarily with neutral atoms rather than with one another. The case of a fully ionized plasma is deferred to a later section, since it results in a nonlinear equation for which there are few simple illustrative solutions. In any case, partially ionized gases are not rare: High-pressure arcs and ionospheric plasmas fall into this category, and most of the early work on gas discharges involved fractional ionizations between 10^3 and 10^6, when collisions with neutral atoms are dominant.

The picture, then, is of a nonuniform distribution of ions and electrons in a dense background of neutrals (Fig. 5.1 ADD IT!). As the plasma spreads out as a result of pressure-gradient and electric field forces, the individual particles undergo a random walk, colliding frequently with the neutral atoms. We begin with a brief review of definitions from atomic theory.

6.2 Collision Parameters

When an electron, say, collides with a neutral atom, it may lose any fraction of its initial momentum, depending on the angle at which it rebounds. In a head-on collision with a heavy atom, the electron can lose twice its initial momentum, since its velocity reverses sign after the collision. The probability of momentum loss can be expressed in terms of the equivalent cross section σ that the atoms would have if they were perfect absorbers of momentum.

In Fig. 5.2, electrons are incident upon a slab of area A and thickness $\mathrm{d}x$ containing n_n neutral atoms per m^3. The atoms are imagined to be opaque spheres of cross-sectional area σ; that is, every time an electron comes within the area blocked by the atom, the electron loses all of its momentum. The number of atoms in the slab is $n_n A\mathrm{d}x$. The fraction of the slab blocked by atoms is $n_n A\sigma\mathrm{d}x/A = n_n\sigma\mathrm{d}x$. If a flux Γ of electrons is incident on the slab, the flux emerging on the other side is $\Gamma' = \Gamma(1 - n_n \sigma \mathrm{d}x)$. Thus the change of Γ with distance is $\mathrm{d}\Gamma/\mathrm{d}x = -n_n \sigma \Gamma$, or

$$\Gamma = \Gamma_0 e^{-n_n \sigma x} \equiv \Gamma_0 e^{-x/\lambda_m}$$

In a distance λ_m, the flux would be decreased to $1/e$ of its initial value. The quantity λ_m is the *mean free path* for collisions:

$$\lambda_m = 1/n_n \sigma \tag{6.1}$$

After traveling a distance λ_m, a particle will have had a good probability of making a collision. The mean time between collisions, for particles of velocity v, is given by

$$\tau = \lambda_m / v$$

and the mean frequency of collisions is

$$\tau^{-1} = v/\lambda_m = n_n \sigma v \tag{6.2}$$

If we now average over particles of all velocities v in a Maxwellian distribution, we have what is generally called the collision frequency ν:

$$\nu = n_n \overline{\sigma v} \tag{6.3}$$

6.3 Diffusion Parameters

The fluid equation of motion including collisions is, for any species s,

$$mn \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = mn \left[\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{v} \right] = q_s n \mathbf{E} - \nabla p - mn\nu\mathbf{v}$$

The averaging process used to compute ν is such as to make the last equation correct; we need not be concerned with the details of this computation. The quantity ν must, however, be assumed to be a constant in order for the equation to be useful. We shall consider a steady state in which $\partial \mathbf{v}/\partial t = 0$. If $\mathbf{v}$ is sufficiently small (or ν sufficiently large), a fluid element will not move into regions of different $\mathbf{E}$ and ∇p in a collision time, and the convective derivative $\mathrm{d}\mathbf{v}/\mathrm{d}t$ will also vanish. Setting the left-hand side to zero, we have, for an isothermal plasma,

$$\mathbf{v} = \frac{1}{mn\nu}(q_s n\mathbf{E} - k_B T\nabla n) = \frac{q_s}{m\nu}\mathbf{E} - \frac{k_B T}{mv}\frac{\nabla n}{n}$$

The coefficients above are called the *mobility* and the *diffusion coefficient*:

$$\mu \equiv |q|/m\nu \quad \text{Mobility} \tag{6.4}$$

$$D \equiv k_B T/m\nu \quad \text{Diffusion coefficient} \tag{6.5}$$

These will be different for each species. Note that D is measured in m^2/s. The transport coefficients μ and D are connected by the *Einstein relation*:

$$\mu = |q|D/k_B T \tag{6.6}$$

With the help of these definitions, the flux Γ_s of the sth species can be written

$$\Gamma_s = n\mathbf{v}_s = \pm\mu_s n\mathbf{E} - D_s\nabla n \tag{6.7}$$

Fick's law of diffusion is a special case of this, occurring when either $\mathbf{E} = 0$ or the particles are uncharged, so that $\mu = 0$:

$$\Gamma = -D\nabla n \quad \text{Fick's law}$$

This equation merely expresses the fact that diffusion is a random-walk process, in which a net flux from dense regions to less dense regions occurs simply because more particles start in the dense region. This flux is obviously proportional to the gradient of the density. In plasmas, Fick's law is not necessarily obeyed. Because of the possibility of organized motions (plasma waves), a plasma may spread out in a manner which is not truly random.

6.4 Decay of a Plasma by Diffusion

6.4.1 Ambipolar Diffusion

We now consider how a plasma created in a container decays by diffusion to the walls. Once ions and electrons reach the wall, they recombine there. The density near the wall, therefore, is essentially zero. The fluid equations of motion and continuity govern the plasma behavior; but if the decay is slow, we need only keep the time derivative in the continuity equation. The time derivative in the equation of motion, will be negligible if the collision frequency ν is large. We thus have

$$\frac{\partial n}{\partial t} + \nabla \cdot \Gamma_s = 0 \tag{6.8}$$

with Γ_s given by Equation 6.7. It is clear that if Γ_i and Γ_e were not equal, a serious charge imbalance would soon arise. If the plasma is much larger than a Debye length, it must be quasineutral; and one would expect that the rates of diffusion of ions and electrons would somehow adjust themselves so that the two species leave at the same rate. How this happens is easy to see. The electrons, being lighter, have higher thermal velocities and tend to leave the plasma first. A positive charge is left behind, and an electric field is set up of such a polarity as to retard the loss of electrons and accelerate the loss of ions. The required **E** field is found by setting $\Gamma_i = \Gamma_e = \Gamma$. From @Equation 6.7, we can write

$$\Gamma = \mu_i n\mathbf{E} - D_i \nabla n = -\mu_e n\mathbf{E} - D_e \nabla n$$

$$\mathbf{E} = \frac{D_i - D_e}{\mu_i + \mu_e} \frac{\nabla n}{n}$$

The common flux Γ is then given by

$$\begin{aligned} \Gamma &= \mu_i \frac{D_i - D_e}{\mu_i + \mu_e} \nabla n - D_i \nabla n \\ &= \frac{\mu_i D_i - \mu_i D_e - \mu_i D_i - \mu_e D_i}{\mu_i + \mu_e} \nabla n \\ &= -\frac{\mu_i D_e + \mu_e D_i}{\mu_i + \mu_e} \nabla n \end{aligned}$$

This is Fick's law with a new diffusion coefficient

$$D_a \equiv \frac{\mu_i D_e + \mu_e D_i}{\mu_i + \mu_e} \tag{6.9}$$

called the *ambipolar diffusion coefficient*. If this is constant, the continuity equation above becomes simply

$$\frac{\partial n}{\partial t} = D_a \nabla^2 n \tag{6.10}$$

The magnitude of Da can be estimated if we take $\mu_e \gg \mu_i$. That this is true can be seen from Equation 6.4. Since ν is proportional to the thermal velocity, which is proportional to $m^{-1/2}$, μ is proportional to $m^{-1/2}$. Equation 6.4 and Equation 6.9 then give

$$D_a \approx D_i + \frac{\mu_i}{\mu_e} D_e = D_i + \frac{T_e}{T_i} D_i$$

For $T_e = T_i$, we have

$$D_a \approx 2D_i$$

The effect of the ambipolar electric field is to enhance the diffusion of ions by a factor of two, and the diffusion rate of the two species together is primarily controlled by the slower species.

6.4.2 Diffusion in a Slab

The diffusion Equation 6.10 can easily be solved by the method of separation of variables. We let

$$n(\mathbf{r}, t) = T(t)S(\mathbf{r})$$

whereupon Equation 6.10, with the subscript on D_a understood, becomes

$$\begin{aligned} S\frac{\mathrm{d}T}{\mathrm{d}t} &= DT\nabla^2 S \\ \frac{1}{T}\frac{\mathrm{d}T}{\mathrm{d}t} &= \frac{D}{S}\nabla^2 S \end{aligned}$$

Since the left side is a function of time alone and the right side a function of space alone, they must both be equal to the same constant, which we shall call $-1/\tau$. The function T then obeys the equation

$$\frac{\mathrm{d}T}{\mathrm{d}t} = -\frac{T}{\tau} \tag{6.11}$$

with the solution

$$T = T_0 e^{-t/\tau}$$

The spatial part S obeys the equation

$$\nabla^2 S = -\frac{1}{D\tau} S \tag{6.12}$$

In slab geometry, this becomes

$$\frac{\mathrm{d}^2 S}{\mathrm{d}x^2} = -\frac{1}{D\tau} S \tag{6.13}$$

with the solution

$$S = A \cos \frac{x}{(D\tau)^{1/2}} + B \sin \frac{x}{(D\tau)^{1/2}} \tag{6.14}$$

We would expect the density to be nearly zero at the walls (Fig. 5.3 ADD IT!) and to have one or more peaks in between. The simplest solution is that with a single maximum. By symmetry, we can reject the odd (sine) term in Equation 6.14. The boundary conditions $S = 0$ at $x = \pm L$ then requires

$$\frac{L}{(D\tau)^{1/2}} = \frac{\pi}{2}$$

or

$$\tau = \left(\frac{2L}{\pi}\right)^2 \frac{1}{D}$$

Combining all the equations above, we have

$$n = n_0 e^{-t/\tau} \cos \frac{\pi x}{2L}$$

This is called the *lowest diffusion mode*. The density distribution is a cosine, and the peak density decays exponentially with time. The time constant τ increases with L and varies inversely with D, as one would expect.

There are, of course, higher diffusion modes with more than one peak. Suppose the initial density distribution is as shown by the top curve in Fig. 5.4 ADD IT. Such an arbitrary distribution can be expanded in a Fourier series:

$$n = n_0 \Big(\sum_l a_l \cos \frac{(l + \frac{1}{2})\pi x}{L} \sum_m b_m \sin \frac{m\pi x}{L} \Big)$$

We have chosen the indices so that the boundary condition at $x = \pm L$ is automatically satisfied. To treat the time dependence, we can try a solution of the form

$$n = n_0 \Big(\sum_l a_l e^{-t/\tau_l} \cos \frac{(l+\frac{1}{2})\pi x}{L} \sum_m b_m e^{-t/\tau_m} \sin \frac{m\pi x}{L} \Big)$$

Substituting this into the diffusion Equation 6.10, we see that each cosine term yields a relation of the form

$$-\frac{1}{\tau_l} = -D\Big[\Big(l + \frac{\pi}{2}\Big)\frac{\pi}{L}\Big]^2$$

and similarly for the sine terms. Thus the decay time constant for the lth mode is given by

$$\tau_l = \Big[\frac{L}{(l+\frac{1}{2})\pi}\Big]^2 \frac{1}{D}$$

The fine-grained structure of the density distribution, corresponding to large l numbers, decays faster, with a smaller time constant τ_l. The plasma decay will proceed as indicated in Fig. 5.4 ADD IT. First, the fine structure will be washed out by diffusion. Then the lowest diffusion mode, the simple cosine distribution of Fig. 5.3 ADD IT, will be reached. Finally, the peak density continues to decay while the plasma density profile retains the same shape.

I WANT TO DO A SIMULATION ON THIS! TRY METHODOFLINES.JL?

6.4.3 Diffusion in a Cylinder

The spatial part of the diffusion equation, eq-diffusion_spatial, reads, in cylindrical geometry,

$$\frac{\mathrm{d}^2 S}{dr^2} + \frac{1}{r}\frac{dS}{dr} + \frac{1}{D\tau}S = 0 \tag{6.15}$$

This differs from Equation 6.13 by the addition of the middle term, which merely accouts for the change in coordinates. The need for the extra term is illustrated simply in Fig. 5.5 ADD IT. If a slice of plasma in (a) is moved toward larger x without being allowed to expand, the density would remain constant. On the other hand, if a shell of plasma in (b) is moved toward larger r with the shell thickness kept constant, the density would necessarily decrease as $1/r$. Consequently, one would expect the solution to Equation 6.15 to be like a damped cosine (Fig. 5.6 ADD IT). This function is called a *Bessel function of order zero*, and eq-diffusion_spatial_cylinder is called Bessel's equation (of order zero). Instead of the symbol cos, it is given the symbol J_0. The function $J0(r/[D\tau]^{1/2})$ is a solution to eq-diffusion_spatial_cylinder, just as $\cos[x/(\mathrm{d}\tau)^{1/2}]$ is a solution to eq-diffusion_spatial. Both $\cos kx$ and $J_0(kr)$ are expressible in terms of infinite series.

FIGURE: Motion of a plasma slab in rectilinear and cylindrical geometry, illustrating the difference between a cosine and a Bessel function.

To satisfy the boundary condition $n = 0$ at $r = a$, we must set $a/(D\tau)^{1/2}$ equal to the first zero of J_0; namely, 2.4. This yields the decay time constant τ. The plasma again decays exponentially, since the temporal part of the diffusion equation, Equation 6.11, is unchanged. We have described the lowest diffusion mode in a cylinder. Higher diffusion modes, with more than one maximum in the cylinder, will be given in terms of Bessel functions of higher order, in direct analogy to the case of slab geometry.

6.5 Steady State Solutions

In many experiments, a plasma is maintained in a steady state by continuous ionization or injection of plasma to offset the losses. To calculate the density profile in this case, we must add a source term to the equation of continuity:

$$\frac{\partial n}{\partial t} - D\nabla^2 n = Q(\mathbf{r})$$

The sign is chosen so that when Q is positive, it represents a source and contributes to positive $\partial n/\partial t$. In steady state, we set $\partial n/\partial t = 0$ and are left with a Poisson-type equation for $n(\mathbf{r})$.

6.5.1 Constant Ionization Function

In many weakly ionized gases, ionization is produced by energetic electrons in the tail of the Maxwellian distribution. In this case, the source term Q is proportional to the electron density n. Setting $Q = Zn$, where Z is the "ionization function", we have

$$\nabla^2 n = -(Z/D)n$$

This is the same equation as that for S, Equation 6.12. Consequently, the density profile is a cosine or Bessel function, as in the case of a decaying plasma, only in this case the density remains constant. The plasma is maintained against diffusion losses by whatever heat source keeps the electron temperature at its constant value and by a small influx of neutral atoms to replenish those that are ionized.

6.5.2 Plane Source

We next consider what profile would be obtained in slab geometry if there is a localized source on the plane $x = 0$. Such a source might be, for instance, a slit-collimated beam of ultraviolet light strong enough to ionize the neutral gas. The steady state diffusion equation is then

$$\frac{\mathrm{d}^n}{\mathrm{d}x^2} = -\frac{Q}{D}\delta(0)$$

Except at $x = 0$, the density must satisfy $\partial^2 n/\partial x^2 = 0$. This obviously has the solution (Fig. 5.7 ADD IT)

$$n = n_0\left(1 - \frac{|x|}{L}\right)$$

The plasma has a linear profile. The discontinuity in slope at the source is characteristic of δ-function sources.

6.5.3 Line Source

Finally, we consider a cylindrical plasma with a source located on the axis. Such a source might, for instance, be a beam of energetic electrons producing ionization along the axis. Except at $r = 0$, the density must satisfy

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial n}{\partial r}\right) = 0$$

The solution that vanishes at $r = a$ is

$$n = n_0 \ln(a/r)$$

The density becomes infinite at $r = 0$ (Fig. 5.8); it is not possible to determine the density near the axis accurately without considering the finite width of the source.

6.6 Recombination

When an ion and an electron collide, particularly at low relative velocity, they have a finite probability of recombining into a neutral atom. To conserve momentum, a third body must be present. If this third body is an emitted photon, the process is called *radiative recombination*. If it is a particle, the process is called *three-body recombination*. The loss of plasma by recombination can be represented by a negative source term in the equation of continuity. It is clear that this term will be proportional to $n_e n_i = n^2$. In the absence of the diffusion terms, the equation of continuity then becomes

$$\frac{\partial n}{\partial t} = -\alpha n^2$$

The constant of proportionality α is called the recombination coefficient and has units of m^3/s. This is a nonlinear equation for n. This means that the straightforward method for satisfying initial and boundary conditions by linear superposition of solutions is not available. Fortunately, it is such a simple nonlinear equation that the solution can be found by inspection. It is

$$\frac{1}{n(\mathbf{r}, t)} = \frac{1}{n_0(\mathbf{r})} + \alpha t$$

where $n_0(\mathbf{r})$ is the initial density distribution. After the density has fallen far below its initial value, it decays *reciprocally* with time, $n \propto 1/\alpha t$. This is a fundamentally different behavior from the case of diffusion, in which the time variation is exponential.

Figure 5.9 (ADD IT!) shows the results of measurements of the density decay in the afterglow of a weakly ionized H plasma. When the density is high, recombination, which is proportional to n^2, is dominant, and the density decays reciprocally. After the density has reached a low value, diffusion becomes dominant, and the decay is thenceforth exponential.

6.7 Diffusion Across a Magnetic Field

The rate of plasma loss by diffusion can be decreased by a magnetic field; this is the problem of confinement in controlled fusion research. Consider a weakly ionized plasma in a magnetic field (fig-gyration_collision ADD IT!). Charged particles will move along $\mathbf{B}$ by diffusion and mobility according to Equation 6.7, since $\mathbf{B}$ does not affect motion in the parallel direction. Thus we have, for each species,

$$\Gamma_z = \pm \mu n E_z - D \frac{\partial n}{\partial z}$$

If there were no collisions, particles would not diffuse at all in the perpendicular direction — they would continue to gyrate about the same magnetic field line. There are, of course, particle drifts across **B** because of electric fields or gradients in **B**, but these can be arranged to be parallel to the walls. For instance, in a perfectly symmetric cylinder (fig-drift_cylinder ADD IT!), the gradients are all in the radial direction, so that the guiding center drifts are in the azimuthal direction. The drifts would then be harmless.

FIGURE: A charged particle in a magnetic field will gyrate about the same line of force until it makes a collision. #fig-gyration_collision

FIGURE: Particle drifts in a cylindrically symmetric plasma column do not lead to losses.#fig-drift_cylinder

When there are collisions, particles migrate across **B** to the walls along the gradients. They do this by a random-walk process (fig-diffusion_collision_neutral ADD IT!). When an ion, say, collides with a neutral atom, the ion leaves the collision traveling in a different direction. It continues to gyrate about the magnetic field in the same direction, but its phase of gyration is changed discontinuously. (The Larmor radius may also change, but let us suppose that the ion does not gain or lose energy on the average.)

FIGURE: Diffusion of gyrating particles by collisions with neutral atoms.#fig-diffusion_collision_neutral

The guiding center, therefore, shifts position in a collision and undergoes a random walk. The particles will diffuse in the direction opposite ∇n. The step length in the random walk is no longer λ_m, as in magnetic-field-free diffusion, but has instead the magnitude of the Larmor radius r_L. Diffusion across **B** can therefore be slowed down by decreasing r_L; that is, by increasing B.

To see how this comes about, we write the perpendicular component of the fluid equation of motion for either species as follows:

6.7.1 Ambipolar Diffusion Across B

Because the diffusion and mobility coefficients are anisotropic in the presence of a magnetic field, the problem of ambipolar diffusion is not as straightforward as in the $B = 0$ case. Consider the particle fluxes perpendicular to **B** (Fig. 5.13 ADD IT!). Ordinarily, since $\Gamma_{e\perp}$ is smaller than $\Gamma_{i\perp}$, a transverse electric field would be set up so as to aid electron diffusion and retard ion diffusion. However, this electric field can be short-circuited by an imbalance of the fluxes along **B**. That is, the negative charge resulting from $\Gamma_{e\perp} < \Gamma_{i\perp}$ can be dissipated by electrons escaping along the field lines. Although the total diffusion must be ambipolar, the perpendicular part of the losses need not be ambipolar. The ions can diffuse out primarily radially, while the electrons diffuse out primarily along **B**. Whether or not this in fact happens depends on the particular experiment. In short plasma columns with the field lines terminating on conducting plates, one would expect the ambipolar electric field to be short-circuited out. Each species then diffuses radially at a different rate. In long, thin plasma columns terminated by insulating plates, one would expect the radial diffusion to be ambipolar because escape along **B** is arduous.

Mathematically, the problem is to solve simultaneously the continuity Equation 6.8 for ions and electrons. It is not the fluxes $\ _j$, but the divergence $\nabla \cdot \ _j$ which must be set equal to each other. Separating $\nabla \cdot \ _j$ into perpendicular and parallel components, we have

$$
\begin{aligned}
\nabla \cdot \ _i &= \nabla_\perp \cdot (\mu_i n \mathbf{E}_\perp - D_{i\perp} \nabla n) + \frac{\partial}{\partial z}\Big(\mu_i n E_z - D_i \frac{\partial n}{\partial z}\Big) \\
\nabla \cdot \ _e &= \nabla_\perp \cdot (\mu_e n \mathbf{E}_\perp - D_{e\perp} \nabla n) + \frac{\partial}{\partial z}\Big(-\mu_e n E_z - D_e \frac{\partial n}{\partial z}\Big)
\end{aligned}
$$

The equation resulting from setting $\nabla \cdot \ _i = \nabla \cdot \ _e$ cannot easily be separated into one-dimensional equations. Furthermore, the answer depends sensitively on the boundary conditions at the ends of the field lines. Unless the plasma is so long that parallel diffusion can be neglected altogether, there is no simple answer to the problem of ambipolar diffusion across a magnetic field.

6.8 Collisions in Fully Ionized Plasmas

When the plasma is composed of ions and electrons alone, all collisions are Coulomb collisions between charged particles. However, there is a distinct difference between

- collisions between like particles (ion–ion or electron–electron collisions) and
- collisions between unlike particles (ion–electron or electron–ion collisions).

Consider two identical particles colliding (Fig. 5.16 ADD IT!). If it is a head-on collision, the particles emerge with their velocities reversed; they simply interchange their orbits, and the two guiding centers remain in the same places. The result is the same as in a glancing collision, in which the trajectories are hardly disturbed. The worst that can happen is a 90^o collision, in which the velocities are changed 90^o in direction. The orbits after collision will then be the dashed circles, and the guiding centers will have shifted. However, it is clear that the "center of mass" of the two guiding centers remains stationary. For this reason, collisions between like particles give rise to very little diffusion. This situation is to be contrasted with the case of ions colliding with neutral atoms. In that case, the final velocity of the neutral is of no concern, and the ion random-walks away from its initial position. In the case of ion–ion collisions, however, there is a detailed balance in each collision; for each ion that moves outward, there is another that moves inward as a result of the collision.

When two particles of opposite charge collide, however, the situation is entirely different (Fig. 5.17 ADD IT!). The worst case is now the 180^o collision, in which the particles emerge with their velocities reversed. Since they must continue to gyrate about the magnetic field lines in the proper sense, both guiding centers will move in the same direction. *Unlike-particle* *collisions give rise to diffusion.* The physical picture is somewhat different for ions and electrons because of the disparity in mass. The electrons bounce off the nearly stationary ions and random-walk in the usual fashion. The ions are slightly jostled in each collision and move about as a result of frequent bombardment by electrons. Nonetheless, because of the conservation of momentum in each collision, the rates of diffusion are the same for ions and electrons, as we shall show.

6.8.1 Plasma Resistivity

The fluid equations of motion including the effects of charged-particle collisions may be written as follows

$$\begin{aligned} m_i n \frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t} &= en(\mathbf{E} + \mathbf{v}_i \times \mathbf{B}) - \nabla p_i - \nabla \cdot \boldsymbol{\pi}_i + \mathbf{P}_{ie} \\ m_e n \frac{\mathrm{d}\mathbf{v}_e}{\mathrm{d}t} &= -en(\mathbf{E} + \mathbf{v}_e \times \mathbf{B}) - \nabla p_e - \nabla \cdot \boldsymbol{\pi}_e + \mathbf{P}_{ei} \end{aligned} \tag{6.16}$$

The terms $\mathbf{P}_{ie}$ and $\mathbf{P}_{ei}$ represent, respectively, the momentum gain of the ion fluid caused by collisions with electrons, and vice versa. The stress tensor $\mathbf{P}_j$ has been split into the isotropic part p_j and the anisotropic viscosity tensor $\boldsymbol{\pi}_j$. Like-particle collisions, which give rise to stresses within each fluid individually, are contained in $\boldsymbol{\pi}_j$. Since these collisions do not give rise to much diffusion, we shall ignore the terms $\nabla \cdot \boldsymbol{\pi}_j$. As for the terms $\mathbf{P}_{ie}$ and $\mathbf{P}_{ei}$, which represent the friction between the two fluids, the conservation of momentum requires

$$\mathbf{P}_{ie} = -\mathbf{P}_{ei}$$

We can write $\mathbf{P}_{ei}$ in terms of the collision frequency in the usual manner:

$$\mathbf{P}_{ei} = mn(\mathbf{v}_i - \mathbf{v}_e)\nu_{ei}$$

and similarly for $\mathbf{P}_{ie}$. Since the collisions are Coulomb collisions, one would expect $\mathbf{P}_{ei}$ to be proportional to the Coulomb force, which is proportional to e^2 (for singly-charged ions). Furthermore, $\mathbf{P}_{ei}$ must be proportional to the density of electrons n_e and to the density of scattering centers n_i, which, of course, is equal to n_e. Finally, $\mathbf{P}_{ei}$ should be proportional to the relative velocity of the two fluids. On physical grounds, then, we can write $\mathbf{P}_{ei}$ as

$$\mathbf{P}_{ei} = \eta e^2 n^2 (\mathbf{v}_i - \mathbf{v}_e) \tag{6.17}$$

where η is a constant of proportionality. Comparing this with the last equation, we see that

$$\nu_{ei} = \frac{ne^2}{m}\eta \tag{6.18}$$

The constant η is the *specific resistivity* of the plasma; that this jibes with the usual meaning of resistivity will become clear shortly.

6.8.2 Mechanics of Coulomb Collisions

When an electron collides with a neutral atom, no force is felt until the electron is close to the atom on the scale of atomic dimensions; the collisions are like billiard-ball collisions. When an electron collides with an ion, the electron is gradually deflected by the long-range Coulomb field of the ion. Nonetheless, one can derive an effective cross section for this kind of collision. It will suffice for our purposes to give an order-of-magnitude estimate of the cross section. In Fig. 5.18 ADD IT!, an electron of velocity $\mathbf{v}$ approaches a fixed ion of charge e. In the absence of Coulomb forces, the electron would have a distance of closest approach r_0, called the impact parameter. In the presence of a Coulomb attraction, the electron will be deflected by an angle χ, which is related to r_0. The Coulomb force is

FIGURE:Orbit of an electron making a Coulomb collision with an ion.#fig-coulomb_collision}

$$F = -\frac{e^2}{4\pi\epsilon_0 r^2}$$

This force is felt during the time the electron is in the vicinity of the ion; this time is roughly

$$T \approx r_0/v$$

The change in the electron's momentum is therefore approximately

$$\Delta(mv) = |FT| \approx \frac{e^2}{4\pi\epsilon_0 r_0 v}$$

We wish to estimate the cross section for large-angle collisions, in which $\chi \geq 90^o$. For a 90^o collision, the change in mv is of the order of mv itself. Thus

$$\Delta(mv) \simeq mv \simeq e^2/4\pi\epsilon_0 r_0 v, \quad r_0 = e^2/4\pi\epsilon_0 mv^2 \tag{6.19}$$

The cross section is then

$$\sigma = \pi r_0^2 = e^4/16\pi\epsilon_0^2 m^2 v^4$$

The collision frequency is, therefore,

$$\nu_{ei} = n\sigma v = ne^4/16\pi\epsilon_0^2 m^2 v^3 \tag{6.20}$$

and the resistivity is

$$\eta = \frac{m}{ne^2}\nu_{ei} = \frac{e^2}{16\pi\epsilon_0^2 m v^3} \tag{6.21}$$

For a Maxwellian distribution of electrons, we may replace v^2 by $k_B T_e/m$ for our order-of-magnitude estimate:

$$\eta \approx \frac{\pi e^2 m^{1/2}}{(4\pi\epsilon_0)^2 (k_B T_e)^{3/2}} \tag{6.22}$$

This is the resistivity based on large-angle collisions alone. In practice, because of the long range of the Coulomb force, small-angle collisions are much more frequent, and the cumulative effect of many small-angle deflections turns out to be larger than the effect of large-angle collisions. It was shown by Spitzer that Equation 6.22 should be multiplied by a factor $\ln\Lambda$:

$$\eta \approx \frac{\pi e^2 m^{1/2}}{(4\pi\epsilon_0)^2 (k_B T_e)^{3/2}} \ln\Lambda \tag{6.23}$$

where

$$\Lambda = \overline{\lambda_D/r_0} = 12\pi n\lambda_D^3$$

This factor represents the maximum impact parameter, in units of r_0 as given by Equation 6.19, averaged over a Maxwellian distribution. The maximum impact parameter is taken to be λ_D because Debye shielding suppresses the Coulomb field at larger distances. Although Λ depends on n and $k_B T_e$, its logarithm is insensitive to the exact values of the plasma parameters. Typical values of $\ln\Lambda$ are given below (λ_D defined in Section 2.4).

$k_B T_e$ [eV]	n [m^{-3}]	$\ln\Lambda$	
0.2	10^{15}	10.7	(Q-machine)
2	10^{17}	11.8	(lab plasma)
100	10^{19}	15.4	(typical torus)
10^4	10^{21}	20.0	(fusion reactor)
10^3	10^{27}	9.6	(laser plasma)

It is evident that $\ln \Lambda$ varies only a factor of two as the plasma parameters range over many orders of magnitude. For most purposes, it will be sufficiently accurate to let $\ln \Lambda = 10$ regardless of the type of plasma involved.

6.8.3 Physical Meaning of

Let us suppose that an electric field $\mathbf{E}$ exists in a plasma and that the current that it drives is all carried by the electrons, which are much more mobile than the ions. Let $B = 0$ and $k_B T_e = 0$, so that $\nabla \cdot \mathbf{P}_e = 0$. Then, in steady state, the electron equation of motion from Equation 6.16 reduces to

$$en\mathbf{E} = \mathbf{P}_{ei} \tag{6.24}$$

Since $\mathbf{j} = en(\mathbf{v}_i - \mathbf{v}_e)$, Equation 6.17 can be written

$$\mathbf{P}_{ei} = \eta en\mathbf{j}$$

so that Equation 6.24 becomes

$$\mathbf{E} = \eta\mathbf{j} \tag{6.25}$$

This is simply Ohm's law, and the constant η is just the specific resistivity. The expression for η in a plasma, as given by Equation 6.22 or Equation 6.23, has several features which should be pointed out.

1. In Equation 6.23, we see that η is *independent of density* (except for the weak dependence in $\ln \Lambda$). This is a rather surprising result, since it means that if a field $\mathbf{E}$ is applied to a plasma, the current $\mathbf{j}$, as given by Equation 6.25, is independent of the number of charge carriers. The reason is that although j increases with n_e, the frictional drag against the ions increases with n_i. Since $n_e = n_i$ these two effects cancel. This cancellation can be seen in Equation 6.20 and Equation 6.21. The collision frequency ν_{ei} is indeed proportional to n, but the factor n cancels out in η. A fully ionized plasma behaves quite differently from a weakly ionized one in this respect. In a weakly ionized plasma, we have $\mathbf{j} = -ne\mathbf{v}_e$, $\mathbf{v}_e = -\mu_e\mathbf{E}$, so that $\mathbf{j} = ne\mu_e\mathbf{E}$. Since the electron mobility μ_e depends only on the density of *neutrals*, the current is proportional to the plasma density n.

2. Equation 6.23 shows that η is proportional to $T_e^{-3/2}$. As a plasma is heated, the Coulomb cross section decreases, and the resistivity drops rather rapidly with increasing temperature. Plasmas at thermonuclear temperatures (tens of keV) are essentially collisionless; this is the reason so much theoretical research is done on collisionless plasmas. Of course, there must always be some collisions; otherwise, there wouldn't be any fusion reactions

either. An easy way to heat a plasma is simply to pass a current through it. The I^2R or ηj^2 losses then turn up as an increase in electron temperature. This is called *ohmic heating*. The $(k_BT_e)^{-3/2}$ dependence of η, however, does not allow this method to be used up to thermonuclear temperatures. The plasma becomes such a good conductor at temperatures above 1 keV that ohmic heating is a very slow process in that range.

3. Equation 6.20 shows that ν_{ei} varies as v^{-3}. The fast electrons in the tail of the velocity distribution make very few collisions. The current is therefore carried mainly by these electrons rather than by the bulk of the electrons in the main body of the distribution. The strong dependence on v has another interesting consequence. If an electric field is suddenly applied to a plasma, a phenomenon known as *electron runaway* can occur. A few electrons which happen to be moving fast in the direction of $-\mathbf{E}$ when the field is applied will have gained so much energy before encountering an ion that they can make only a glancing collision. This allows them to pick up more energy from the electric field and decrease their collision cross section even further. If E is large enough, the cross section falls so fast that these runaway electrons never make a collision. They form an accelerated electron beam detached from the main body of the distribution.

6.8.4 Numerical Values of

Exact computations of η which take into account the ion recoil in each collision and are properly averaged over the electron distribution were first given by Spitzer. The following result for hydrogen is sometimes called the Spitzer resistivity:

$$\eta_{\parallel} = 5.2 \times 10^{-5} \frac{Z \ln \Lambda}{T^{3/2}[\text{eV}]} \, [\text{ohm} \cdot \text{m}] \tag{6.26}$$

Here Z is the ion charge number, which we have taken to be 1 elsewhere. Since the dependence on m_i is weak, these values can also be used for other gases. The subscript $\parallel$ means that this value of η is to be used for motions parallel to $\mathbf{B}$. For motions perpendicular to $\mathbf{B}$, one should use $\eta_{\perp}$ given by

$$\eta_{\perp} = 2.0\eta_{\parallel}$$

This does not mean that conductivity along $\mathbf{B}$ is only two times better than conductivity across $\mathbf{B}$. A factor like $\omega_c^2\tau^2$ still has to be taken into account. The factor 2.0 comes from a difference in weighting of the various velocities in the electron distribution. In perpendicular motions, the slow electrons, which have small Larmor radii, contribute more to the resistivity than in parallel motions.

For $k_BT_e = 100\,\text{eV}$, Equation 6.26 yields

$$\eta = 5 \times 10^{-7}\,[\text{ohm} \cdot \text{m}]$$

This is to be compared with various metallic conductors:

material	$\eta\,[\text{ohm} \cdot \text{m}]$
copper	2×10^{-8}
stainless steel	7×10^{-7}
mercury	1×10^{6}

A 100-eV plasma, therefore, has a conductivity like that of stainless steel.

6.8.5 Pulsed Currents

When a steady-state current is drawn between two electrodes aligned along the magnetic field, electrons are the dominant current carrier, and sheaths are set up at the cathode to limit the current to the set value. When the current is pulsed, however, it takes time to set up the current distribution. It was shown by Stenzel and Urrutia that this time is controlled by whistler waves (R-waves), which must travel the length of the device to communicate the voltage information.

6.8.6 Collisions Between Species

Y.Y talked about this during his lecture that the collision rate among ions or electrons is much larger than the collision rate between ions and electrons. I need to go back to the notes. See also StackExchange Q&A.

6.8.7 Conductivity Tensor

Resistivity is the inverse of conductivity. Consider the effect of collisions as friction in the equation of motion:

$$m\frac{\mathrm{d}\mathbf{u}_e}{\mathrm{d}t} = q(\mathbf{E} + \mathbf{u}_e \times \mathbf{B}) - m\nu_c(\mathbf{u}_e - \mathbf{u}')$$

where $\mathbf{u}'$ is the velocity of the collision targets.

- In unmagnetized plasma, let $\mathbf{B} = 0$.

Assume a steady state with cold electrons ($\mathbf{u}_e = \mathbf{v}_e$) and non-moving targets $\mathbf{u}' = 0$, we can easily have

$$\mathbf{E} = -\frac{m_e \nu_c}{e}\mathbf{u}_e$$

Since the electron current is defined as

$$\mathbf{j} = -en_e\mathbf{u}_e$$

we have from the last two equations

$$\mathbf{j} = \frac{n_e e^2}{m_e \nu_c}\mathbf{E} \equiv \sigma\mathbf{E} \tag{6.27}$$

where σ is the classical conductivity.

- In magnetized plasma, let $\mathbf{B} = b\hat{z}$.

Similarly to the unmagnetized case,

$$0 = -e(\mathbf{E} + \mathbf{u}_e \times \mathbf{B}) - \nu_c m_e \mathbf{u}_e$$

In order to avoid confusion, let σ_0 denote the classical conductivity from Equation 6.27. Using the definition of current density and writing in scalar forms, we now have

$$\begin{aligned}
j_x &= \sigma_0 E_x - \frac{\Omega_e}{\nu_c} j_y \\
j_y &= \sigma_0 E_y - \frac{\Omega_e}{\nu_c} j_x \\
j_z &= \sigma_0 E_z
\end{aligned}$$

the solution of which gives

$$\begin{aligned}
j_x &= \frac{\nu_c^2}{\nu_c^2 + \Omega_e^2}\sigma_0 E_x - \frac{\Omega_e \nu_c}{\nu_c^2 + \Omega_e^2} E_y \\
j_y &= \frac{\nu_c^2}{\nu_c^2 + \Omega_e^2}\sigma_0 E_y - \frac{\Omega_e \nu_c}{\nu_c^2 + \Omega_e^2} E_x \\
j_z &= \sigma_0 E_z
\end{aligned}$$

or in matrix form,

$$
\begin{pmatrix} j_x \\ j_y \\ j_z \end{pmatrix} = \begin{pmatrix} \sigma_P & -\sigma_H & 0 \\ \sigma_H & \sigma_P & 0 \\ 0 & 0 & \sigma_\parallel \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix}
$$

where

$$
\begin{aligned}
\sigma_P &= \frac{\nu_c^2}{\nu_c^2 + \Omega_e^2}\sigma_0 & \text{Pedersen conductivity} \quad & (\perp \mathbf{B} \& \parallel \mathbf{E}) \\
\sigma_H &= \frac{\Omega_e \nu_c}{\nu_c^2 + \Omega_e^2}\sigma_0 & \text{Hall conductivity} \quad & (\perp \mathbf{B} \& \perp \mathbf{E}) \\
\sigma_\parallel = \sigma_0 &= \frac{n_e e^2}{m_e \nu_c} & \text{parallel conductivity} \quad & (\parallel \mathbf{B})
\end{aligned}
$$

Magnetic field-aligned currents (FACs) are very important in plasma physics. From **?@fig-conductivity**, we see that $\Omega_e \simeq \nu_c$ is the most complicated regime where the three conductivities are comparable, and $\Omega_e \gg \nu_c$ is the most anisotropic regime as particles are tightly bound in the perpendicular direction but free to move along $\mathbf{B}$.

```
KeyNotes.plot_conductivity()
```

The study of conductivity is most relevant in the Earth's ionosphere, where $\Omega_e \simeq \nu_c$ happens at about 100 km above from the ground, and σ_0 is on the order of 100 S/m. Ionosphere forms within the atmosphere through the mechanisms of electron impact ionization and photoionization and in turn contribute in a crucial way to the plasma interactions. The ionosphere is electrically conductive and thus carry part of a current system that is continued outside the ionosphere by plasma currents. The properties of the ionospheric currents are usually described in terms of the *electrical conductivity* of the ionosphere. Because a strong background magnetic field threads the ionosphere, the conductivities are strongly anisotropic. In the direction of the magnetic field, the conductivity is referred to as the *parallel* or *Birkeland* conductivity. The Birkeland conductivity $\sigma_\parallel$ is so high that it short-circuits the field-aligned component of the ionospheric electric field, and normally electrical equipotentials align with the ionospheric magnetic field lines for quasi-static ($> 1\,\mathrm{min}$) large-scale electrodynamic conditions. Perpendicular to the magnetic field direction, the conductivity has components parallel to the electric field σ_P and perpendicular to the electric field σ_H.

Because the electric field vanishes in the field-aligned direction through the ionosphere ($\mathbf{E}_\parallel = 0$), it is meaningful to integrate the transverse σ_P and σ_H along the magnetic field direction to evaluate the Pedersen and Hall conductances Σ_P and Σ_H, respectively. For any $\mathbf{E}$,

$$
\boldsymbol{\Sigma} \cdot \mathbf{E} = \Sigma_P \mathbf{E} + \Sigma_H \hat{b} \times \mathbf{E}
$$

The continuity of currents gives

$$
\begin{aligned}
\nabla \cdot \mathbf{j} + \frac{\partial \rho^*}{\partial t} &= 0 \\
\nabla \cdot (\boldsymbol{\sigma} \cdot \mathbf{E}) &= -\frac{\partial \rho^*}{\partial t} \\
\int \nabla \cdot (\boldsymbol{\sigma} \cdot \mathbf{E}) dz &= -\int \frac{\partial \rho^*}{\partial t} dz \\
\nabla \cdot (\int \boldsymbol{\sigma} dz \cdot \mathbf{E}) &= -\rho^* v_{\parallel} \\
\nabla \cdot (\boldsymbol{\Sigma} \cdot \mathbf{E}) &= -j_{\parallel} \\
\nabla \cdot (\boldsymbol{\Sigma} \cdot \nabla \Phi) &= j_{\parallel}
\end{aligned}
\tag{6.28}
$$

$\mathbf{E} = -\nabla\Phi$, and $j_{\parallel}$ is defined such that upward ($+z$) is considered positive. Here we assume that $\mathbf{B}$ is aligned with $\hat{z}$; a more realistic geometry requires the consideration of inclination angle. See Equation 19.1 and note the sign difference between the definitions of $j_{\parallel}$.

The forces that act on a flux tube in the ionosphere can be expressed in terms of these conductances and the force balance expressions can consequently be used to derive other properties of the ionospheric flow such as its velocity. (???)

Mass loading contributes to the electrodynamics of the plasma interaction in ways that are analogous to the contributions of the ionospheric conductivities. Pickup ionization conserves momentum. An increase of the ion density results in a reduction of the flow velocity and its associated electric field. Thus in developing the theory of the interaction, pickup effects can be directly incorporated into the above conductances and the interaction region extends well above the atmospheric exobase.

The Pedersen current (aligned with the electric field) exerts forces that slow the flow in the ionosphere. The Hall current flows perpendicular to the electric and magnetic field, thereby breaking the symmetry of the interaction. The Hall current results in a rotation of the flow away from the corotation direction and produces ionospheric asymmetries.

6.9 Diffusion of Fully Ionized Plasmas

The magnetic induction equation can be derived from Faraday's law and Ohm's law. If magnetic diffusivity η is constant, we can write it in the form

$$
\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{u} \times \mathbf{B}) + \eta \nabla^2 \mathbf{B}
$$

We define *magnetic Reynold number* R_m as the ratio of the convection term and the diffusion term:

$$R_m \equiv \frac{\nabla \times (\mathbf{u} \times \mathbf{B})}{\eta \nabla^2 \mathbf{B}} \sim \frac{uBL_0^2}{\eta L_0 B} = \frac{uL_0}{\eta}$$

There are two extreme cases of R_m. One thing that you should keep in mind is that this is no longer ideal MHD because $\mathbf{E} + \mathbf{u} \times \mathbf{B} \neq 0$.

6.9.1 Small R_m condition

If $R_m \ll 1$, the convection part can be ignored, (e.g. $\mathbf{u} \approx 0$)

$$\frac{\partial \mathbf{B}}{\partial t} = \eta \nabla^2 \mathbf{B}$$

This is a diffusion equation, meaning that unlike in ideal MHD, plasma no longer ties to the magnetic field. The diffusion time scale τ_0 can be derived as follows:

$$\frac{B}{\tau_0} = \eta \frac{B}{{L_0}^2} \Rightarrow \tau_0 = \frac{{L_0}^2}{\eta} = \mu \sigma {L_0}^2$$

σ is the electrical conductivity, which is a derived parameter from Ohm's law. For fully ionized plasma, $\sigma = \frac{n_i e^2 \tau_{ei}}{m_e}$.

The magnetic diffusivity can be written as

$$\eta = \frac{1}{\mu\sigma} = \frac{m_e}{\mu n_i e^2 \tau_{ei}} = \frac{m_e n_e \ln\Lambda}{\mu n_i e^2 \cdot 5.45 \times 10^5 T^{3/2}} = 5.2 \times 10^7 \frac{\ln\Lambda}{T^{3/2}}$$

where $\ln\Lambda$ is the Coulomb logarithm (generally $10 \sim 20$ which is insensitive to n and T, Section 6.8.2).

Examples:

1. Evolution of solar coronal magnetic field and the energy release of solar flares in solar corona.

$T = 10^7 K$, $n = 10^{14} m^{-3}$, $\Rightarrow \eta = 1\ m^2 s^{-1}$

For supergranules, $L_0 = 10^7 m = 10 Mm$, ($R_s = 700 Mm$), the diffusion time scale $\tau_d = 10^{14} s \approx 3 \times 10^6$yr, which is very very long.

For flare, the observed diffusion time scale is $\tau_d = 1s \Rightarrow L_0 = 1m$. The magnetic reconnection (in solar corona) diffusion time scale is about $\tau_d = 100s \Rightarrow L_0 = 10m$. The viewing angle at $1AU$ to the sun is about 2000 arcsec, where 1 arcsecond corresponds to 700km on the sun. Right now the best resolution we have is about $0.2 \sim 0.5$ arcsecond, which is about 100km. This means that we are three orders of magnitude beyond the resolution we need to observe electron diffusion region on the sun!

2. Requirement of dynamo

If there is no such thing as a magnetic dynamo, the geomagnetic field will have been already gone. In the core of Earth, $\eta \approx 1 m^2 s^{-1}$, radius of Earth $L_0 = 6 \times 10^6 m \Rightarrow \tau_d \approx 4 \times 10^{12} s \approx 10^6$yr. However, the age of our Earth is about 10^9yr $\gg 10^6$yr. This indicates that somthing is generating magnetic field inside the core.

3. Current sheet

Consider an ideal configuration of magnetic field,

$$\mathbf{B} = B(x,t)\hat{y},$$

where

$$B(x,0) = \begin{cases} +B_0 & x > 0 \\ -B_0 & x < 0 \end{cases}$$

Then the magnetic diffusion equation can be written as

$$\frac{\partial B}{\partial t} = \eta \frac{\partial^2 B}{\partial x^2}$$

There is an analytic solution

$$B(x,t) = B_0\ \mathrm{erf}(\xi)$$

where

$$\xi = \frac{x}{\sqrt{4\eta t}}$$

and

$$\mathrm{erf}(\xi) = \frac{2}{\sqrt{\pi}} \int_0^{\xi} e^{-z^2} dz$$

is the *error function*. It has some basic properties as follows:

$$\mathrm{erf}(\xi) \Rightarrow \begin{cases} 1 & ,\xi \to +\infty \\ -1 & ,\xi \to -\infty \end{cases}$$

$$\mathrm{erf}(\xi) = \frac{2}{\sqrt{\pi}}\left(\xi - \frac{\xi^3}{3} + \frac{\xi^5}{10} - \frac{\xi^7}{42} + ...\right) \approx \frac{2\xi}{\sqrt{\pi}},\ |\xi| \ll 1$$

The shape is shown in Figure 6.1.

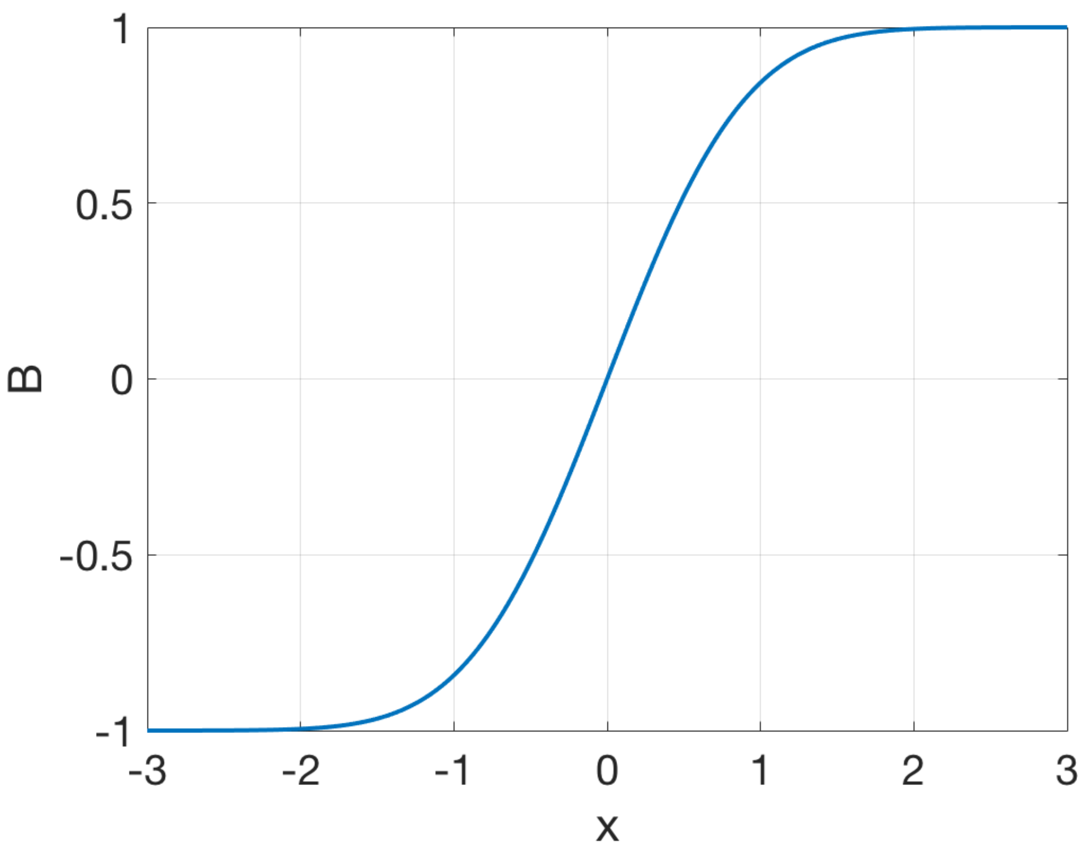

Figure 6.1: Error function.

$$|B(x,t)| \approx \begin{cases} \frac{x}{\sqrt{\pi\eta t}}B_0 & ,|\xi| \ll 1 \text{ or } |x| \ll \sqrt{\pi\eta t} \\ B_0 & ,|\xi| \gg 1 \text{ or } |x| \gg \sqrt{\pi\eta t} \\ 0.995B_0 & ,|\xi| = 2 \text{ or } |x| = 2\sqrt{\pi\eta t} \end{cases}$$

If you want to draw this time-variant magnetic field at two different times t_1 and t_2 ($t_1 < t_2$), you should have less dense lines at t_2 than t_1. An example plot is present in Figure 6.2.

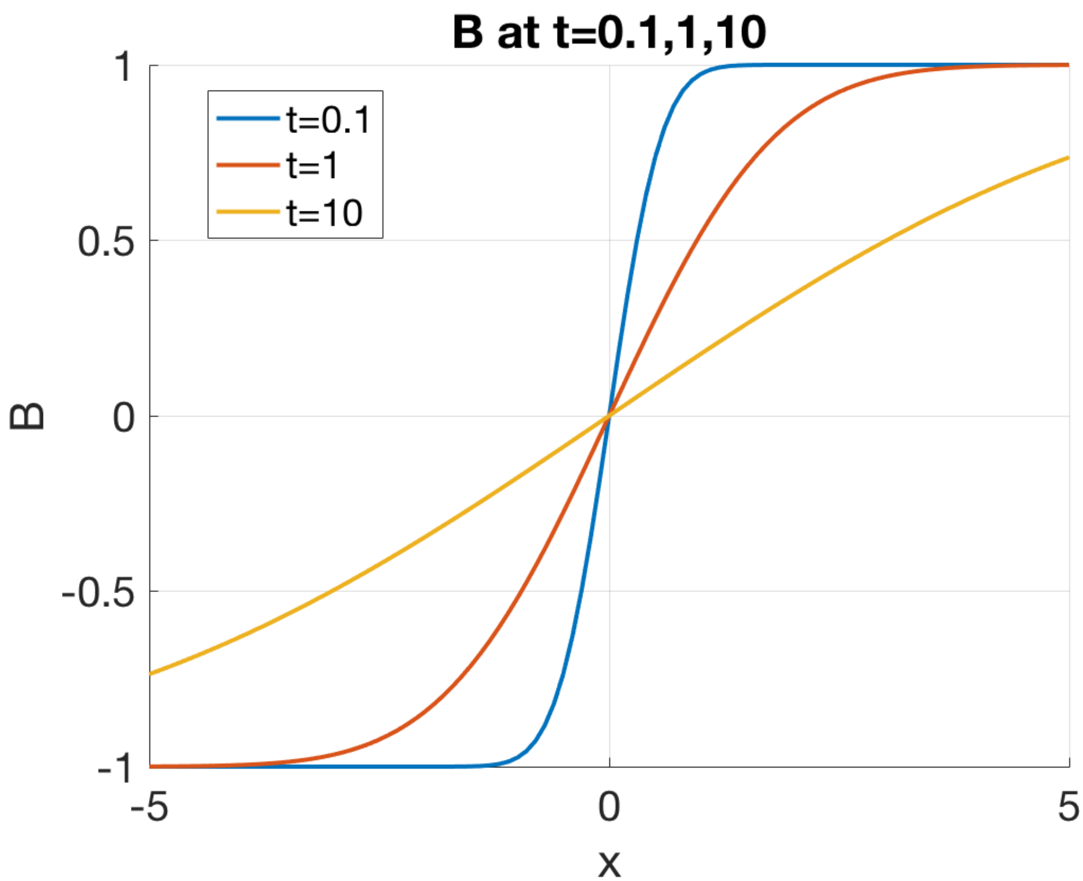

Figure 6.2: Magnetic field at $t = 0.1, 1, 10$, respectively.

Conclusion:

1. In a certain amount of time, B decreases in the near region and stays the same in the far region.
2.

$$
\begin{aligned}
j_z &= \frac{1}{\mu}\frac{\mathrm{d}B}{\mathrm{d}x}, \\
\delta = 2\sqrt{4\eta t}, &\ \text{(characteristic breadth)} \\
\frac{\mathrm{d}\delta}{\mathrm{d}t} = 2\sqrt{\frac{\eta}{t}}, &\,\text{(characteristic speed)}
\end{aligned}
$$

3.

$$
j = \int_{-\infty}^{\infty} j_z \mathrm{d}x = \frac{1}{\mu}\int_{-\infty}^{\infty} \mathrm{d}B = \frac{2B_0}{\mu} = \text{const.}
$$

The current sheet gets thickened with time, but the total current in z remains the same.

4. Dissipation of magnetic energy, $W_B = \frac{B^2}{2\mu}$???

$$
\frac{\partial}{\partial t}\int_{-\infty}^{\infty}(W_0 - W_B)\mathrm{d}x = -\frac{\partial}{\partial t}\int_{-\infty}^{\infty}\frac{B(x)^2}{2\mu}\mathrm{d}x
$$

6.9.2 Large R_m condition

If $R_m \ll 1$, the diffusion part can be ignored,

$$
\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{u} \times \mathbf{B})
$$

When the conductivity is large, the diffusion coefficient becomes very small, and frozen-in flux phenomenon happens.

1. Frozen-flux theorem

The magnetic field lines are frozen in the plasma flow. The magnetic flux through a surface moving with the plasma is conserved,

$$
\frac{\mathrm{d}\Phi}{\mathrm{d}t} = 0,
$$

where

$$
\Phi = \int \mathbf{B} \cdot \mathrm{d}\mathbf{A}.
$$

ADD FIGURES!!!

Proof.

Consider a surface S with boundary Γ. the magnetic flux through S is

$$\Phi = \iint_s \mathbf{B} \cdot \mathrm{d}\mathbf{A}$$

Then the time derivative of flux can be written as

$$\begin{aligned}
\dot{\Phi} &= \iint_s \dot{\mathbf{B}} \cdot \mathrm{d}\mathbf{A} + \iint_s \mathbf{B} \cdot \mathrm{d}\dot{\mathbf{A}} \\
&= \iint_s (-\nabla \times \mathbf{E}) \cdot \mathrm{d}\mathbf{A} + \iint_s \mathbf{B} \cdot \mathrm{d}\dot{\mathbf{A}} \\
&= \iint_s \nabla \times (\mathbf{u} \times \mathbf{B}) \cdot \mathrm{d}\mathbf{l} + \iint_s \mathbf{B} \cdot \mathrm{d}\dot{\mathbf{A}} \\
&= \oint_\Gamma \mathbf{u} \times \mathbf{B} \cdot \mathrm{d}\mathbf{l} + \iint_s \mathbf{B} \cdot \mathrm{d}\dot{\mathbf{A}}
\end{aligned}$$

The change of area can be expressed as

$$\begin{aligned}
\Delta\mathbf{A} &= (\mathbf{u}\Delta t) \times \mathrm{d}\mathbf{l}, \\
\mathrm{d}\dot{\mathbf{A}} &= \mathbf{u} \times \mathrm{d}\mathbf{l}.
\end{aligned}$$

Substituting into the above equation, we have

$$\dot{\Phi} = \oint_\Gamma \mathbf{u} \times \mathbf{B} \cdot \mathrm{d}\mathbf{l} + \oint_\Gamma \mathbf{B} \cdot (\mathbf{u} \times \mathrm{d}\mathbf{l}) = 0$$

Keep in mind we are still under MHD approximation

$$\frac{\partial|\mathbf{B}|}{\partial t} \ll \Omega_i |\mathbf{B}|$$

This looks like the conservation of magnetic moment.

2. Stretching of magnetic field lines

$$\frac{d\mathbf{B}}{\mathrm{d}t} = \frac{\partial\mathbf{B}}{\partial t} + (\mathbf{u} \cdot \nabla)\mathbf{B} = \nabla \times (\mathbf{u} \times \mathbf{B}) + (\mathbf{u} \cdot \nabla)\mathbf{B} = (\mathbf{B} \cdot \nabla)\mathbf{u} - \mathbf{B}(\nabla \cdot \mathbf{u})$$

where the first term on the right is the shear motion and the second term is the expansion related change.

Combined with the continuity equation,

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} + \rho\nabla\cdot\mathbf{u} = 0$$

we can get

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big(\frac{\mathbf{B}}{\rho}\Big) = \Big(\frac{\mathbf{B}}{\rho}\cdot\nabla\Big)\mathbf{u} \tag{6.29}$$

Proof.

$$\begin{aligned}\frac{\mathrm{d}\mathbf{B}}{\mathrm{d}t} &= (\mathbf{B}\cdot\nabla)\mathbf{u} + \frac{\mathbf{B}}{\rho}\frac{\mathrm{d}\rho}{\mathrm{d}t} \\ \frac{1}{\rho}\frac{\mathrm{d}\mathbf{B}}{\mathrm{d}t} - \frac{\mathbf{B}}{\rho^2}\frac{\mathrm{d}\rho}{\mathrm{d}t} &= \Big(\frac{\mathbf{B}}{\rho}\cdot\nabla\Big)\mathbf{u}\end{aligned}$$

Compare Equation 6.29 and Equation 3.1, we have

$$\frac{\mathbf{B}}{\rho} \propto \mathbf{l}$$

i.e. $\mathbf{B}/\rho$ and $\mathbf{l}$ has the same form of variation. If density does not change much, the magnetic field strength will be proportional to l.

6.10 Solutions of the Diffusion Equation

6.10.1 Time Dependence

6.10.2 Time-Independent Solutions

6.11 Bohm Diffusion and Neoclassical Diffusion

7 Waves

7.1 Plasma Oscillations

If the electrons in a plasma are displaced from a uniform background of ions, electric fields will be built up in such a direction as to restore the neutrality of the plasma by pulling the electrons back to their original positions. Because of their inertia, the electrons will overshoot and oscillate around their equilibrium positions with a characteristic frequency known as the *plasma frequency*. This oscillation is so fast that the massive ions do not have time to respond to the oscillating field and may be considered as fixed. In Fig. 4.2 (ADD IT!), the open rectangles represent typical elements of the ion fluid, and the darkened rectangles the alternately displaced elements of the electron fluid. The resulting charge bunching causes a spatially periodic **E** field, which tends to restore the electrons to their neutral positions.

We shall derive an expression for the plasma frequency ω_p in the simplest case, making the following assumptions:

1. There is no magnetic field;
2. there are no thermal motions ($k_B T = 0$);
3. the ions are fixed in space in a uniform distribution;
4. the plasma is infinite in extent; and
5. the electron motions occur only in the x direction. As a consequence of the last assumption, we have

$$\nabla = \hat{x}\partial x,\ \mathbf{E} = E\hat{x},\ \nabla \times \mathbf{E} = 0,\ \mathbf{E} = -\nabla\phi$$

There is, therefore, no fluctuating magnetic field; this is an electrostatic oscillation.

The electron equations of continuity and motion are

$$\frac{\partial n_e}{\partial t} + \nabla \cdot (n_e \mathbf{u}_e) = 0$$
$$m n_e \Big[\frac{\partial \mathbf{u}_e}{\partial t} + (\mathbf{u}_e \cdot \nabla)\mathbf{u}_e\Big] = -e n_e \mathbf{E}$$

The only Maxwell equation we shall need is the one that does not involve **B**: Poisson's equation. This case is an exception to the general rule of Section 5.9 that Poisson's equation cannot be used to find $\mathbf{E}$. This is a high-frequency oscillation; electron inertia is important, and the deviation from neutrality is the main effect in this particular case. Consequently, we write

$$\epsilon_0 \nabla \cdot \mathbf{E} = \epsilon_0 \partial \mathbf{E}/\partial x = e(n_i - n_e)$$

The last three equations together can be easily solved by the procedure of *linearization*. By this we mean that the amplitude of oscillation is small, and terms containing higher powers of amplitude factors can be neglected. We first separate the dependent variables into two parts: an "equilibrium" part indicated by a subscript 0, and a "perturbation" part indicated by a subscript 1:

$$n_e = n_0 + n_1 \quad \mathbf{u}_e = \mathbf{u}_0 + \mathbf{u}_1 \quad \mathbf{E}_e = \mathbf{E}_0 + \mathbf{E}_1$$

The equilibrium quantities express the state of the plasma in the absence of the oscillation. Since we have assumed a uniform neutral plasma at rest before the electrons are displaced, we have

$$\nabla n_0 = \mathbf{u}_0 = \mathbf{E}_0 = 0$$
$$\frac{\partial n_0}{\partial t} = \frac{\partial \mathbf{u}_0}{\partial t} = \frac{\partial \mathbf{E}_0}{\partial t} = 0$$

The momentum equation now becomes

$$m\frac{\partial \mathbf{u}_1}{\partial t} = -e\mathbf{E}$$

The term $(\mathbf{u}_1 \cdot \nabla)\mathbf{u}_1$ is seen to be quadratic in an amplitude quantity, and we shall linearize by neglecting it. The *linear theory* is valid as long as $|u_1|$ is small enough that such quadratic terms are indeed negligible. Similarly, the continuity equation becomes

$$\frac{\partial n_1}{\partial t} + n_0 \nabla \cdot \mathbf{u}_1 = 0$$

In Poisson's equation, we note that $n_{i0} = n_{e0}$ in equilibrium and that $n_{i1} = 0$ by the assumption of fixed ions, so we have

$$\epsilon_0 \partial \mathbf{E}/\partial x = -en_1$$

The oscillating quantities are assumed to behave sinusoidally:

$$\mathbf{n}_1 = n_1 e^{i(kx-\omega t)}$$
$$\mathbf{u}_1 = \mathbf{u}_1 e^{i(kx-\omega t)}\hat{x}$$
$$\mathbf{E}_1 = \mathbf{E}_1 e^{i(kx-\omega t)}\hat{x}$$

The time derivative $\partial/\partial t$ can therefore be replaced by $-i\omega$, and the gradient ∇ by $ik\hat{x}$. Now the linearized equations become

$$-im\omega n_1 = -n_0 iku_1$$
$$-i\omega u_1 = -eE_1$$
$$-ik\epsilon_0 E_1 = -en_1$$

Eliminating n_1 and E_1, we have

$$-m\omega u_1 = -i\frac{n_0 e^2}{\epsilon_0 \omega} u_1$$

If u_1 does not vanish, we must have

$$\omega^2 = \frac{n_0 e^2}{\epsilon_0 m}$$

The *plasma frequency* is therefore

$$\omega_p = \sqrt{\frac{n_0 e^2}{\epsilon_0 m}} \quad \text{rad/s} \tag{7.1}$$

Numerically, one can use the approximate formula

$$\omega_p/2\pi = f_p \approx 9\sqrt{n} \quad \text{m}^{-3}$$

This frequency, depending only on the plasma density, is one of the fundamental parameters of a plasma. Because of the smallness of m, the plasma frequency is usually very high. For instance, in a plasma of density $n = 10^{18}\,\text{m}^{-3}$, we have

$$f_p \approx 9(10^{18})^{1/2} = 9 \times 10^9\,\text{s}^{-1} = 9\,\text{GHz}$$

Radiation at f_p normally lies in the microwave range. We can compare this with another electron frequency: ω_c. A useful numerical formula is

$$f_{ce} \simeq 28\,\mathrm{GHz/T}$$

Thus if $B = 0.32$ T and $n = 10^{18}\,\mathrm{m}^{-3}$, the cyclotron frequency is approximately equal to the plasma frequency for electrons.

Equation 7.1 tells us that if a plasma oscillation is to occur at all, it must have a frequency depending only on n. In particular, ω does not depend on k, so the group velocity $\mathrm{d}\omega/dk$ is zero. The disturbance does not propagate. How this can happen can be made clear with a mechanical analogy (Fig. 4.3 fig-independent_springs). Imagine a number of heavy balls suspended by springs equally spaced in a line. If all the springs are identical, each ball will oscillate vertically with the same frequency. If the balls are started in the proper phases relative to one another, they can be made to form a wave propagating in either direction. The frequency will be fixed by the springs, but the wavelength can be chosen arbitrarily. The two undisturbed balls at the ends will not be affected, and the initial disturbance does not propagate. Either traveling waves or standing waves can be created, as in the case of a stretched rope. Waves on a rope, however, must propagate because each segment is connected to neighboring segments.

FIGURE: Synthesis of a wave from an assembly of independent oscillators.

This analogy is not quite accurate, because plasma oscillations have motions in the direction of $\mathbf{k}$ rather than transverse to $\mathbf{k}$. However, as long as electrons do not collide with ions or with each other, they can still be pictured as independent oscillators moving horizontally (in fig-independent_springs). But what about the electric field? Won't that extend past the region of initial disturbance and set neighboring layers of plasma into oscillation? In our simple example, it will not, because the electric field due to equal numbers of positive and negative infinite plane charge sheets is zero. In any finite system, however, plasma oscillations will propagate. In Fig. 4.4 ADD IT!, the positive and negative (shaded) regions of a plane plasma oscillation are confined in a cylindrical tube. The fringing electric field causes a coupling of the disturbance to adjacent layers, and the oscillation does not stay localized.

7.2 Dielectric Function

Usually when the permittivity of a material is function of space or frequency, it is call dielectric function. The dielectric constant ϵ is a quantity which appears in electrostatic when people describe how a material screens an external time-independent electric field. When they begin to study how a material screens an external time-dependent electric field $\mathbf{E} \propto e^{-i\omega t}$ in electrodynamic sense they found that the number ϵ depends on frequency, so one gets $\epsilon(\omega)$. It would be stupid to call a quantity, which essentially depends on frequency, just "dielectric constant", therefore one calls it "dielectric function". Further studies showed that ϵ depends not only on the frequency but also on the wave-vector of the field, $\mathbf{E} \propto e^{-i\omega t + ikx}$, so one gets the dielectric function $\epsilon = \epsilon(k, \omega)$.

7.3 Classification of EM Waves in Uniform Plasma

$$
\begin{cases} \mathbf{k} \parallel \mathbf{B}_0 & \text{Parallel Propagation,} \\ \mathbf{k} \perp \mathbf{B}_0 & \text{Perpendicular Propagation} \end{cases}
$$

$$
\begin{cases} \mathbf{k} \parallel \mathbf{E}_1 & \text{Longitudinal Waves,} \\ \mathbf{k} \perp \mathbf{E}_1 & \text{Transverse Waves} \end{cases}
$$

$$
\begin{cases} \mathbf{B}_1 = 0 & \text{Electrostatic Waves,} \\ \mathbf{B}_1 \neq 0 & \text{Electromagnetic Waves} \end{cases}
$$

Note:

1. Wave is longitudinal $\iff$ Wave is electrostatic
2. Wave is transverse $\implies$ Wave is electromagnetic
3. Wave is electromagnetic $\not\Longrightarrow$ Wave is transverse. You can always add a component of $\mathbf{E}_1$ parallel to $\mathbf{k}$ without changing $\mathbf{B}_1$.

7.4 ES vs. EM Waves

A practical way to distinguish ES and EM waves is to check $\nabla \times \mathbf{E}$ and $\nabla \cdot \mathbf{E}$, where $\mathbf{E}$ is the electric field of the wave: * If the curvature is relatively small and the divergence is relatively large, then it is likely to be ES. * Otherwise it is likely to be EM.

As we will see in Section 7.5, the dielectric function is defined in Equation 7.4. From other perspectives, the dielectric function shows up in the Ampère's law as well as the Poisson's equation

$$
\begin{aligned}
\nabla \times \mathbf{B} = \mu_0 \mathbf{j} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \equiv \mu_0 \boldsymbol{\epsilon} \cdot \frac{\partial \mathbf{E}}{\partial t} \\
\nabla \cdot (\ _{\mathbf{0}}\mathbf{E_1}) + q_j n_j \equiv \nabla \cdot (\boldsymbol{\epsilon} \cdot \mathbf{E}_1) = 0
\end{aligned}
$$

Let us consider waves in an *isotropic* plasma. For isotropic plasmas, the dielectric tensor $\boldsymbol{\epsilon}$ shrinks to a scalar ϵ. For cold plasma (static ion background), the dielectric function is

$$
\frac{\epsilon}{\epsilon_0} = 1 - \frac{{\omega_{pe}}^2}{\omega^2}
$$

For electrostatic (ES) waves, let $\epsilon = 0$, we have

$$
\omega = \pm \omega_{pe}
$$

For electromagnetic (EM) waves, from Maxwell's equations we have

$$
\begin{aligned}
\nabla\times\mathbf{E} &= -\frac{\partial\mathbf{B}}{\partial t},\\
\nabla\times\mathbf{B} &= \mu_0\mathbf{j}+\mu_0\epsilon_0\frac{\partial\mathbf{E}}{\partial t}\equiv\mu_0\epsilon\frac{\partial\mathbf{E}}{\partial t}.
\end{aligned}
$$

With $\nabla\rightarrow i\mathbf{k},\ \partial/\partial t\rightarrow -i\omega$, we can get the dispersion relation

$$
\begin{aligned}
i\mathbf{k}\times\mathbf{E} &= i\omega\mathbf{B}\\
i\mathbf{k}\times\mathbf{B} &= -i\mu_0\epsilon\omega\mathbf{E}\\
\Rightarrow k^2\mathbf{E}-\cancel{(\mathbf{k}\cdot\mathbf{E})\mathbf{k}} &= \omega^2\mu_0\epsilon\mathbf{E}.
\end{aligned}
$$

If $\mathbf{k}\perp\mathbf{E}$, by substituting the dielectric function inside we have

$$
\begin{aligned}
k^2=\omega^2\epsilon\mu_0 &= \omega^2\epsilon_0\mu_0\Big[1-\frac{{\omega_{pe}}^2}{\omega^2}\Big]\\
\Rightarrow \omega^2 &= k^2c^2+{\omega_{pe}}^2.
\end{aligned}
$$

For both waves, $\nabla\cdot(\epsilon\mathbf{E}_1)=0\Rightarrow i\epsilon(\mathbf{k}\cdot\mathbf{E}_1)=0$ is always valid. However, for electrostatic wave, $\mathbf{E}_1=-\nabla\phi_1=-i\mathbf{k}\phi_1\Rightarrow\mathbf{k}\parallel\mathbf{E}_1\Rightarrow\epsilon=0$, while for EM wave, usually $\mathbf{k}\perp\mathbf{E}_1$ ($\mathbf{k}\perp\mathbf{E}_1\Rightarrow$ EM wave, but EM waves do not necessarily need to be transverse. You can always add a component of $\mathbf{E}_1$ parallel to $\mathbf{k}$ without changing $\mathbf{B}_1$), ϵ does not need to be zero. Therefore, *getting the dispersion relation by setting ϵ to 0 is only valid for isotropic ES waves.* For EM waves, there's a systematic way to get all the dispersion relations starting from dielectric function, explained in detail in Section 7.5. Here we just have a simple summary of the steps.

From Maxwell's equation for the perturbed field,

$$
\begin{aligned}
\nabla\times\mathbf{E}_1 &= -\mu_0\frac{\partial\mathbf{H}_1}{\partial t}\\
\nabla\times\mathbf{H}_1 &= \mathbf{J}_1+\epsilon_0\frac{\partial\mathbf{E}_1}{\partial t}
\end{aligned}
$$

where we have assumed

$$
\begin{Bmatrix}\mathbf{E}_1(\mathbf{x},t)\\ \mathbf{H}_1(\mathbf{x},t)\end{Bmatrix}=\Re\begin{Bmatrix}\tilde{\mathbf{E}}_1e^{i\mathbf{k}\cdot\mathbf{x}-i\omega t}\\ \tilde{\mathbf{H}}_1e^{i\mathbf{k}\cdot\mathbf{x}-i\omega t}\end{Bmatrix}
$$

It quickly follows that

$$
\begin{aligned}
\mathbf{k}\times\mathbf{E}_1 &= \mu_0\omega\mathbf{H}_1 \\
i\mathbf{k}\times\mathbf{H}_1 &= i\mathbf{k}\times\left(\frac{\mathbf{k}\times\mathbf{E}_1}{\mu_0\omega}\right) = \mathbf{J}_1 - \epsilon_0 i\omega\mathbf{E}_1
\end{aligned}
$$

Then there comes the wave equation

$$
\mathbf{k}\times(\mathbf{k}\times\mathbf{E}_1) = \mathbf{k}(\mathbf{k}\cdot\mathbf{E}_1) - k^2\mathbf{E}_1 = -i\omega\mu_0\mathbf{J}_1 - \frac{\omega^2}{c^2}\mathbf{E}_1 \equiv -\frac{\omega^2}{c^2}\frac{\epsilon}{\epsilon_0}\mathbf{E}_1
$$

If we can express the total current density as a function of perturbed electric field, $\mathbf{J}_1 = \mathbf{J}_1(\mathbf{E}_1)$, from MHD, 2-fluid, or Vlasov model combining with the property of the media, we can obtain the expression for the dielectric function ϵ. With some effort, we get

$$
\mathbf{A}\begin{pmatrix} E_{1x} \\ E_{1y} \\ E_{1z} \end{pmatrix} = 0
$$

from which the condition for non-trivial solutions leads to

$$
\det A = 0 \Rightarrow \begin{cases} \text{eigenvalue for } \omega = \omega(\mathbf{k}) \\ \text{eigenvectors} \Rightarrow \text{polarization of E field} \end{cases}
$$

7.5 Cold Uniform Plasma

As long as $T_e = T_i = 0$, the linear plasma waves can easily be generalized to an arbitrary number of charged particle species and an arbitrary angle of propagation θ relative to the magnetic field. Waves that depend on finite T, such as ion acoustic waves, are not included in this treatment. The derivations go back to late 1920s when Appleton and Wilhelm Altar first calculated the cold plasma dispersion relation (CPDR).

First, we define the dielectric tensor of a plasma as follows. The fourth Maxwell equation is

$$
\nabla\times\mathbf{B} = \mu_0(\mathbf{j} + \epsilon_0\dot{\mathbf{E}})
$$

where $\mathbf{j}$ is the plasma current due to the motion of the various charged particle species s, with density n_s, charge q_s, and velocity $\mathbf{v}_s$:

$$
\mathbf{j} = \sum_s n_s q_s \mathbf{v}_s \tag{7.2}
$$

Considering the plasma to be a dielectric with internal currents $\mathbf{j}$, we may write Equation 7.2 as

$$\nabla \times \mathbf{B} = \mu_0 \dot{\mathbf{D}}$$

where

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \frac{i}{\omega}\mathbf{j} \tag{7.3}$$

is the **electric displacement field** or **electric induction**. It accounts for the effects of *bound charge within materials* (i.e. plasma). Here we have assumed an $\exp(-i\omega t)$ dependence for all plasma motions. Let the current $\mathbf{j}$ be proportional to $\mathbf{E}$ but not necessarily in the same direction (because of the magnetic field $B_0\hat{\mathbf{z}}$); we may then define a conductivity tensor $\boldsymbol{\sigma}$ by the relation

$$\mathbf{j} = \boldsymbol{\sigma} \cdot \mathbf{E}$$

Equation 7.3 becomes

$$\mathbf{D} = \epsilon\Big(\mathbf{I} + \frac{i}{\epsilon_0 \omega}\boldsymbol{\sigma}\Big) \cdot \mathbf{E} = \boldsymbol{\epsilon} \cdot \mathbf{E} \tag{7.4}$$

Thus the effective dielectric constant of the plasma is the tensor

$$\boldsymbol{\epsilon} = \epsilon_0(\mathbf{I} + i\boldsymbol{\sigma}/\epsilon_0\omega)$$

where $\mathbf{I}$ is the unit tensor. In electromagnetism, a **dielectric** is an electrical insulator that can be polarised by an applied electric field. When a dielectric material is placed in an electric field, electric charges do not flow through the material as they do in an electrical conductor, because they have no loosely bound, or free, electrons that may drift through the material, but instead they shift, only slightly, from their average equilibrium positions, causing dielectric polarisation.

To evaluate $\boldsymbol{\sigma}$, we use the linearized fluid equation of motion for species s, neglecting the collision and pressure terms:

$$m_s \frac{\partial \mathbf{v}_s}{\partial t} = q_s(\mathbf{E} + \mathbf{v}_s \times \mathbf{B}_0) \tag{7.5}$$

Defining the cyclotron and plasma frequencies for each species as

$$\omega_{cs} \equiv \left|\frac{q_s B_0}{m_s}\right|, \quad \omega_{ps}^2 \equiv \left|\frac{n_0 q_s^2}{\epsilon_0 m_s}\right|$$

we can separate Equation 7.5 into x, y, and z components and solve for $\mathbf{v}_s$, obtaining

$$\begin{aligned} v_{xs} &= \frac{iq_s}{m_s\omega}\frac{E_x \pm i(\omega_{cs}/\omega)E_y}{1-(\omega_{cs}/\omega)^2} \\ v_{ys} &= \frac{iq_s}{m_s\omega}\frac{E_y \mp i(\omega_{cs}/\omega)E_x}{1-(\omega_{cs}/\omega)^2} \\ v_{zs} &= \frac{iq_s}{m_s\omega}E_z \end{aligned} \tag{7.6}$$

where $\pm$ stands for the sign of q_s. The plasma current is

$$\mathbf{j} = \sum_s n_{0s} q_s \mathbf{v}_s$$

so that

$$\begin{aligned} \frac{i}{\epsilon_0\omega} j_x &= \sum_s \frac{in_{0s}}{\epsilon_0\omega}\frac{iq_s^2}{m_s\omega}\frac{E_x \pm i(\omega_{cs}/\omega)E_y}{1-(\omega_{cs}/\omega)} \\ &= \sum_s -\frac{\omega_{ps}^2}{\omega^2}\frac{E_x \pm i(\omega_{cs}/\omega)E_y}{1-(\omega_{cs}/\omega)} \end{aligned} \tag{7.7}$$

Using the identities

$$\begin{aligned} \frac{1}{1-(\omega_{cs}/\omega)^2} &= \frac{1}{2}\Big[\frac{\omega}{\omega \mp \omega_{cs}} + \frac{\omega}{\omega \pm \omega_{cs}}\Big] \\ \pm\frac{\omega_{cs}/\omega}{1-(\omega_{cs}/\omega)^2} &= \frac{1}{2}\Big[\frac{\omega}{\omega \mp \omega_{cs}} - \frac{\omega}{\omega \pm \omega_{cs}}\Big], \end{aligned}$$

we can write Equation 7.7 as follows:

$$\begin{aligned} \frac{1}{\epsilon_0\omega} j_x = -\frac{1}{2}\sum_s \frac{\omega_{ps}^2}{\omega^2}\Big[\Big(\frac{\omega}{\omega \pm \omega_{cs}} + \frac{\omega}{\omega \mp \omega_{cs}}\Big)E_x \\ + \Big(\frac{\omega}{\omega \mp \omega_{cs}} + \frac{\omega}{\omega \pm \omega_{cs}}\Big)iE_y\Big] \end{aligned} \tag{7.8}$$

Similarly, the y and z components are

$$
\begin{aligned}
\frac{1}{\epsilon_0\omega} j_y = -\frac{1}{2}\sum_s \frac{\omega_{ps}^2}{\omega^2}\Big[\Big(\frac{\omega}{\omega\pm\omega_{cs}} + \frac{\omega}{\omega\mp\omega_{cs}}\Big) iE_x \\
+ \Big(\frac{\omega}{\omega\mp\omega_{cs}} + \frac{\omega}{\omega\pm\omega_{cs}}\Big)E_y\Big]
\end{aligned}
\tag{7.9}
$$

$$
\frac{i}{\epsilon_0\omega} j_z = -\sum_s \frac{\omega_{ps}^2}{\omega^2} E_z \tag{7.10}
$$

Use of Equation 7.8 in Equation 7.3 gives

$$
\begin{aligned}
\frac{1}{\epsilon_0} D_x = E_x - \frac{1}{2}\sum_s \frac{\omega_{ps}^2}{\omega^2}\Big[\Big(\frac{\omega}{\omega\pm\omega_{cs}} + \frac{\omega}{\omega\mp\omega_{cs}}) E_x \\
+ \Big(\frac{\omega}{\omega\mp\omega_{cs}} + \frac{\omega}{\omega\pm\omega_{cs}}) iE_y\Big]
\end{aligned}
\tag{7.11}
$$

We define the convenient abbreviations

$$
\begin{aligned}
& R \equiv 1 - \sum_s \frac{\omega_{ps}^2}{\omega^2}\Big(\frac{\omega}{\omega\pm\omega_{cs}}\Big) \\
& L \equiv 1 - \sum_s \frac{\omega_{ps}^2}{\omega^2}\Big(\frac{\omega}{\omega\mp\omega_{cs}}\Big) \\
& S \equiv \frac{1}{2}(R+L) \quad D \equiv \frac{1}{2}(R-L)^* \\
& P \equiv 1 - \sum_s \frac{\omega_{ps}^2}{\omega^2}
\end{aligned}
\tag{7.12}
$$

where “R” stands for right, “L” stands for left, “S” stands for sum, “D” stands for difference, and “P” stands for plasma. Do not confuse D with the electric displacement field **D**. Using these in Equation 7.11 and proceeding similarly with the y and z components, we obtain

$$
\begin{aligned}
\epsilon_0^{-1} D_x &= SE_x - iDE_y \\
\epsilon_0^{-1} D_y &= iDE_x + SE_y \\
\epsilon_0^{-1} D_z &= PE_z
\end{aligned}
$$

Comparing with Equation 7.4, we see that

$$
\boldsymbol{\epsilon} = \epsilon_0 \begin{pmatrix} S & -iD & 0 \\ iD & S & 0 \\ 0 & 0 & P \end{pmatrix} \equiv \epsilon_0 \boldsymbol{\epsilon}_R \tag{7.13}
$$

We next derive the wave equation by taking the curl of the equation $\nabla \times \mathbf{E} = -\dot{\mathbf{B}}$ and substituting $\nabla \times \mathbf{B} = \mu_0 \boldsymbol{\epsilon} \cdot \dot{\mathbf{E}}$, obtaining

$$\nabla \times \nabla \times \mathbf{E} = -\mu_0 \epsilon_0 (\boldsymbol{\epsilon}_R \cdot \ddot{\mathbf{E}}) = -\frac{1}{c^2} \boldsymbol{\epsilon}_R \cdot \ddot{\mathbf{E}} \tag{7.14}$$

Assuming an $\exp(i\mathbf{k} \cdot \mathbf{r})$ spatial dependence of $\mathbf{E}$ and defining a vector index of refraction

$$\mathbf{n} = \frac{c}{\omega} \mathbf{k}$$

we can write Equation 7.14 as

$$\mathbf{n} \times (\mathbf{n} \times \mathbf{E}) + \boldsymbol{\epsilon}_R \cdot \mathbf{E} = 0 \tag{7.15}$$

The uniform plasma is isotropic in the $x-y$ plane, so we may choose the y axis so that $k_y = 0$, without loss of generality. If θ is the angle between $\mathbf{k}$ and $\mathbf{B}_0$, we then have

$$n_x = n \sin\theta \quad n_z = n \cos\theta \quad n_y = 0$$

The next step is to separate Equation 7.15 into components, using the elements of $\boldsymbol{\epsilon}_R$ given in Equation 7.13. This procedure readily yields

$$\mathbf{R} \cdot \mathbf{E} \equiv \begin{pmatrix} S - n^2 \cos\theta & -iD & n^2 \sin\theta \cos\theta \\ iD & S - n^2 & 0 \\ n^2 \sin\theta \cos\theta & 0 & P - n^2 \sin^2\theta \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} = 0 \tag{7.16}$$

From this it is clear that the E_x, E_y components are coupled to E_z only if one deviates from the principal angles $\theta = 0, 90^o$.

Equation 7.16 is a set of three simultaneous, homogeneous equations; the condition for the existence of a solution is that the determinant of $\mathbf{R}$ vanish: $||\mathbf{R}|| = 0$. We then obtain

$$\begin{aligned} &(iD)^2 (P - n^2 \sin^2\theta) + (S - n^2) \\ &\quad \times [(S - n^2 \cos^2\theta)(P - n^2 \sin^2\theta) - n^4 \sin^2\theta \cos^2\theta] = 0 \end{aligned} \tag{7.17}$$

By replacing $\cos^2\theta$ by $1 - \sin^2\theta$, we can solve for $\sin^2\theta$, obtaining

$$\sin^2\theta = \frac{-P(n^4 - 2Sn^2 + RL)}{n^4(S - P) + n^2(PS - RL)}$$

We have used the identity $S^2 - D^2 = RL$. Similarly,

$$\cos^2\theta = \frac{Sn^4 - (PS + RL)n^2 + PRL}{n^4(S - P) + n^2(PS - RL)}$$

Dividing the last two equations, we obtain

$$\tan^2\theta = \frac{-P(n^4 - 2Sn^2 + RL)}{Sn^4 - (PS + RL)n^2 + PRL}$$

Since $2S = R + L$, the numerator and denominator can be factored to give the cold-plasma dispersion relation

$$\tan^2\theta = \frac{-P(n^2 - R)(n^2 - L)}{(Sn^2 - RL)(n^2 - P)} \tag{7.18}$$

7.5.1 Wave Modes

The principal modes of cold plasma waves can be recovered by setting $\theta = 0^0$ and 90^o. When $\theta = 0^o$,

$$P(n^2 - R)(n^2 - L) = 0$$

There are three roots:

- $P = 0$ (Langmuir wave)
- $n^2 = R$ (R wave)
- $n^2 = L$ (L wave)

When $\theta = 90^o$,

$$(Sn^2 - RL)(n^2 - P) = 0$$

There are two roots:

- $n^2 = RL/S$ (extraordinary wave)
- $n^2 = P$ (ordinary wave)

By inserting the definitions of Equation 7.12, one can verify that these are identical to the dispersion relations given in separate derivations, with the addition of corrections due to ion motions.

7.5.2 Resonances

The resonances can be found by letting n go to ∞. We then have

$$\tan^2\theta_{res} = -P/S$$

This shows that the resonance frequencies depend on angle θ.

- If $\theta = 0^o$, the possible solutions are $P = 0$ and $S = \infty$. The former is the plasma resonance $\omega = \omega_p$, while the latter occurs when either $R = \infty$ (i.e. $\omega = \omega_{ce}$, electron cyclotron resonance) or $L = \infty$ (i.e. $\omega = \omega_{ci}$, ion cyclotron resonance).
- If $\theta = 90^o$, the possible solutions are $P = \infty$ or $S = 0$. The former cannot occur for finite ω_p and ω, and the latter yields the upper and lower hybrid frequencies, as well as the two-ion hybrid frequency when there is more than one ion species.

7.5.3 Cutoffs

The cutoffs can be found by setting $n = 0$ in Equation 7.18. Again using $S^2 - D^2 = RL$, we find that the condition for cutoff is independent of θ:

$$PRL = 0$$

- The conditions $R = 0$ and $L = 0$ yield the ω_R and ω_L cutoff frequencies, with the addition of ion corrections.

For R-waves, since $\omega_{pi}^2 \ll \omega_{pe}^2, \omega_{ci} \ll \omega_{ce}$, the cutoff frequency can be approximated by

$$\begin{aligned}
1 - \frac{\omega_{pe}^2}{\omega(\omega-\omega_{ce})} - \frac{\omega_{pi}^2}{\omega(\omega+\omega_{ci})} &= 0 \\
1 = \frac{\omega_{pe}^2\left[\omega\left(1+\cancel{\frac{\omega_{pi}^2}{\omega_{pe}^2}}\right)+\omega_{ci}-\cancel{\frac{\omega_{pi}^2}{\omega_{pe}^2}}\omega_{ce}\right]}{\omega_{ce}\omega(\omega-\omega_{ce})\left(\frac{\omega}{\omega_{ce}}+\cancel{\frac{\omega_{ci}}{\omega_{ce}}}\right)} \\
1 &= \frac{\omega_{pe}^2(\omega+\omega_{ci})}{\omega^2(\omega-\omega_{ce})} \\
\omega^3 - \omega_{ce}\omega^2 - \omega_{pe}^2\omega - \omega_{pe}^2\omega_{ci} &= 0
\end{aligned}$$

Here somehow we can ignore $\omega_{pe}^2\omega_{ci}$ (I DON'T KNOW WHY???) and obtain the positive solution

$$\omega_{R=0} \approx \frac{\omega_{ce}}{2}\left[1+\sqrt{1+4\omega_{pe}^2/\omega_{ce}^2}\right] \tag{7.19}$$

In the low density limit, $\omega_p \ll \omega_c$, $(1+x)^{1/2} \approx 1 + x/2$ when $x \to 0$,

$$\omega_{R=0} \approx \omega_{ce}(1 + \omega_{pe}^2/\omega_{ce}^2)$$

In the high density limit, $\omega_p \gg \omega_c$,

$$\omega_{R=0} \approx \omega_{pe} + \omega_{ce}/2$$

Similarly for L-waves, the cutoff frequency can be approximated by

$$\omega_{L=0} \approx \frac{\omega_{ce}}{2}\Big[-1 + \sqrt{1 + 4\omega_{pe}^2/\omega_{ce}^2}\Big] \tag{7.20}$$

In the low density limit, $\omega_p \ll \omega_c$,

$$\omega_{L=0} \approx \omega_{pe}^2/\omega_{ce}$$

In the high density limit, $\omega_p \gg \omega_c$,

$$\omega_{L=0} \approx \omega_{pe} - \omega_{ce}/2$$

- The condition $P = 0$ is seen to correspond to cutoff as well as to resonance. This degeneracy is due to our neglect of thermal motions. Actually, $P = 0$ (or $\omega = \omega_p$) is a resonance for longitudinal waves and a cutoff for transverse waves.

7.5.4 Polarizations

The information contained in Equation 7.18 is summarized in the Clemmow–Mullaly–Allis (CMA) diagram. One further result, not in the diagram, can be obtained easily from this formulation. The middle line of Equation 7.16 reads

$$iDE_x + (S - n^2)E_y = 0$$

Thus the polarization in the plane perpendicular to $\mathbf{B}_0$ is given by

$$\frac{iE_x}{E_y} = \frac{n^2 - S}{D}$$

From this it is easily seen that waves are linearly polarized at resonance ($n^2 = \infty$) and circularly polarized at cutoff ($n^2 = 0$, $R = 0$ or $L = 0$; thus $S = \pm D$).

7.5.5 Low Frequency Limit

It is very useful to obtain the circularly polarized wave dispersion relation in the MHD regime.

The R-wave corresponds to electron. When $\omega \ll \omega_{ce}$,

$$
\begin{aligned}
n^2 = R - 1 - \frac{\omega_{pi}^2}{\omega(\omega+\omega_{ci})} - \frac{\omega_{pe}^2}{\omega(\omega-\omega_{ce})} \\
k^2c^2 = \omega^2\Big(1 + \frac{\omega_{pe}^2}{\omega\omega_{ce}} - \frac{\omega_{pi}^2}{\omega\omega_{ci}}\frac{\omega_{ci}}{\omega_{ci}+\omega}\Big) = \omega^2\Big[1 + \frac{\omega_{pe}^2}{\omega_{ce}(\omega_{ci}+\omega)}\Big] \\
k^2c^2 = \omega^2\Big(1 + \frac{c^2}{v_A^2}\frac{\omega_{ci}}{\omega_{ci}+\omega}\Big)
\end{aligned}
$$

For $\omega_{ci} \ll \omega \ll \omega_{ce}$, we can make further simplification:

$$
k^2c^2 = \omega^2\Big(1 + \frac{\omega_{pe}^2}{\omega\omega_{ce}}\Big)
$$

This is the whistler wave, with group velocity $v_g = \partial\omega/\partial k \propto \sqrt{\omega}$. It means that high frequency waves transpose energy faster than low frequency waves. In other words, one will hear high frequency components earlier than low frequency components, creating a “whistler effect”. This was discovered during the first world war, and the theoretical explanation came out in the 1950s. Also note that since whistler wave travels along the field line, in near-Earth space we have signals traveling from the south hemisphere to the north hemisphere within this frequency regime. Here is an observation example from Palmer station, Antarctica. For $\omega \ll \omega_{ci}$, Alfvén wave is recovered.

See more in Section 7.15.

The L-wave corresponds to ion. When $\omega < \omega_{ci}$,

$$
\begin{aligned}
k^2c^2 &= \omega^2\Big[1 + \frac{\omega_{pi}^2}{\omega_{ci}(\omega_{ci}-\omega)}\Big] \\
&= \omega^2\Big(1 + \frac{c^2}{v_A^2}\frac{\omega_{ci}}{\omega_{ci}-\omega}\Big)
\end{aligned}
$$

For $\omega \lesssim \omega_{ci}$, we get the ion-cyclotron wave; for $\omega \ll \omega_{ci}$, Alfvén wave is recovered.

?@fig-dispersion_parallel shows the dispersion relations for L/R waves in a rough scale (ACTUALLY THE SCALES ARE SO BAD…). Above the cut-off frequencies ($\omega_{R=0}$ and $\omega_{L=0}$) the solution to the wave dispersion equation is called the *free-space mode*. Below electron and ion cyclotron frequencies the waves are called the *cyclotron modes*. At low frequencies

($\omega \to 0$) L- and R-modes merge and the dispersion becomes that of the shear Alfvén wave $n^2 \to c^2/v_A^2$.

```
KeyNotes.plot_dispersion_parallel()
```

The dispersion curve for a R-wave propagating parallel to the equilibrium magnetic field is sketched in **?@fig-dispersion__R__wave**. The continuation of the Alfvén wave above the ion cyclotron frequency is called the *electron cyclotron wave*, or sometimes the *whistler wave*. The latter terminology is prevalent in ionospheric and space plasma physics contexts. The phase speed is mostly super-Alfvénic except near the electron gyrofrequency. The wave which propagates above the cutoff frequency, ω_1, is a standard right-handed circularly polarized electromagnetic wave, somewhat modified by the presence of the plasma. Note that the low-frequency branch of the dispersion curve differs fundamentally from the high-frequency branch, because the former branch corresponds to a wave which can only propagate through the plasma in the presence of an equilibrium magnetic field, whereas the high-frequency branch corresponds to a wave which can propagate in the absence of an equilibrium field.

For a L-wave, similar considerations to the above give a dispersion curve of the form sketched in **?@fig-dispersion__L__wave**. In this case, n^2 goes to infinity at the ion cyclotron frequency, Ω_i, corresponding to the so-called *ion cyclotron resonance* (at $L \to \infty$). At this resonance, the rotating electric field associated with a left-handed wave resonates with the gyromotion of the ions, allowing wave energy to be converted into perpendicular kinetic energy of the ions. There is a band of frequencies, lying above the ion cyclotron frequency, in which the left-handed wave does not propagate. At very high frequencies a propagating mode exists, which is basically a standard left-handed circularly polarized electromagnetic wave, somewhat modified by the presence of the plasma.

As before, the lower branch in **?@fig-dispersion__L__wave** describes a wave that can only propagate in the presence of an equilibrium magnetic field, whereas the upper branch describes a wave that can propagate in the absence an equilibrium field. The continuation of the Alfvén wave to just below the ion cyclotron frequency is generally called the *ion cyclotron wave*. Note that the phase speed is always sub-Alfvénic.

7.5.6 Faraday Rotation

A linearly polarized plane wave can be expressed as a sum of left- and right-hand circularly polarized waves (R- and L-modes having equal amplitudes, E_0). If we assume that at $z=0$, the wave is linearly polarized along the x-axis, and that the wave vector $\mathbf{k}$ and the background magnetic field $\mathbf{B}_0$ are along the z-axis, we can write

$$\mathbf{E} = E_0[(e^{ik_R z} + e^{ik_L z})\hat{x} + (e^{ik_R z} - e^{ik_L z})\hat{y}]e^{-i\omega t}$$

The ratio of the E_x and E_y components is

$$\frac{E_x}{E_y} = \cot\left(\frac{k_L - k_R}{2} z\right)$$

Hence, due to different phase speeds of R- and L-modes the linrealy polarized wave that is travelling along a magnetic field will experience the rotation of its plane of polarization. This is called *Faraday rotation*. The magnitude of the rotation depends on the density and magnetic field of the plasma. Considering frequencies above the plasma frequency one can show that the rate of change in the rotation angle ϕ with the distance travelled (assumed here to be in the z-direction) is

$$\frac{\mathrm{d}\phi}{\mathrm{d}z} = \frac{-e^3}{2m_e^2 \epsilon_0 c \omega^2} n_e B_0$$

and the total rotation from the source to the observer is

$$\phi = \frac{-e^3}{2m_e^2 \epsilon_0 c \omega^2} \int_0^d n_e \mathbf{B} \cdot \mathrm{d}\mathbf{s}$$

where $\mathrm{d}\mathbf{s}$ is along the wave propagation path. The total rotation thus depends on both the dnesity and magnetic field of the medium.

Faraday rotation is an important diagnostic tool both in laboratories and in astronomy. It can be used to obtain information of the magnetic field of the cosmic plasma. Note that density has to be known using other methods. On the other hand, if the magnetic field is known, Faraday rotation can give information of the density.

7.5.7 Perpendicular Wave Propagation

Let us now consider wave propagation, at arbitrary frequencies, perpendicular to the equilibrium magnetic field, i.e. $\theta = 90^o$.

The cutoff frequencies, at which n^2 goes to zero, are the roots of $R = 0$ and $L = 0$ according to $n^2 = LR/S$. In fact, we have already solved these equations in the previous sections (recall that cutoff frequencies do not depend on θ). There are two cutoff frequencies, $\omega_{R=0}$ and $\omega_{L=0}$, which are specified by Equation 7.19 and Equation 7.20, respectively.

Let us, next, search for the resonant frequencies, at which n^2 goes to infinity. According to the previous discussions, the resonant frequencies are solutions of

$$S = 1 - \frac{\omega_{pe}^2}{\omega^2 - \Omega_e^2} - \frac{\omega_{pi}^2}{\omega^2 - \Omega_i^2} = 0 \tag{7.21}$$

The roots of this equation can be obtained as follows. First, we note that if the first two terms are equated to zero, we obtain $\omega = \omega_{\mathrm{UH}}$, where

$$\omega_{\mathrm{UH}} \equiv \sqrt{\omega_{pe}^2 + \Omega_e^2} \tag{7.22}$$

If this frequency is substituted into the third term, the result is far less than unity. We conclude that ω_{UH} is a good approximation to one of the roots of Equation 7.21. To obtain the second root, we make use of the fact that the product of the square of the roots is

$$\Omega_e^2\,\Omega_i^2 + \omega_{pe}^2\,\Omega_i^2 + \omega_{pi}^2\Omega_e^2 \simeq \Omega_e^2\Omega_i^2 + \omega_{pi}^2\,\Omega_e^2$$

We, thus, obtain $\omega = \omega_{\mathrm{LH}}$, where

$$\omega_{\mathrm{LH}} \equiv \sqrt{\frac{\Omega_e^2\Omega_i^2 + \omega_{pi}^2\Omega_e^2}{\omega_{pe}^2 + \Omega_e^2}}$$

The first resonant frequency, ω_{UH}, is greater than the electron cyclotron or plasma frequencies, and is called the *upper hybrid frequency*. The second resonant frequency, ω_{LH}, lies between the electron and ion cyclotron frequencies, and is called the *lower hybrid frequency*. (F. F. Chen 2016) gave some nice explanations of the physical origins of these frequencies by looking at the electrostatic electron/ion waves perpendicular to $\mathbf{B}$. At low frequencies, the mode in question reverts to the compressional-Alfvén wave discussed previously. Note that the shear-Alfvén wave does not propagate perpendicular to the magnetic field.

Using the above information, and the easily demonstrated fact that

$$\omega_{\mathrm{LH}} < \omega_{L=0} < \omega_{\mathrm{UH}} < \omega_{R=0}$$

we can deduce that the dispersion curve for the mode in question takes the form sketched in **?@fig-dispersion__perp__wave**. The lowest frequency branch corresponds to the compressional-Alfvén wave. The other two branches constitute the *extraordinary*, or X-, wave. The upper branch is basically a linearly polarized (in the y-direction) electromagnetic wave, somewhat modified by the presence of the plasma. This branch corresponds to a wave which propagates in the absence of an equilibrium magnetic field. The lowest branch corresponds to a wave which does not propagate in the absence of an equilibrium field. Finally, the middle branch corresponds to a wave which converts into an electrostatic plasma wave in the absence of an equilibrium magnetic field.

Wave propagation at oblique angles is generally more complicated than propagation parallel or perpendicular to the equilibrium magnetic field, but does not involve any new physical effects.

7.6 MHD Waves

7.6.1 Cold MHD

By ignoring pressure, gravity, viscosity and rotation, we have

$$\begin{aligned} \rho\frac{\partial \mathbf{u}}{\partial t} &= \mathbf{j}\times\mathbf{B}_0 \\ \mathbf{E} &= -\mathbf{u}\times\mathbf{B}_0 \\ \nabla\times\mathbf{E} &= -\frac{\partial \mathbf{B}_1}{\partial t} \\ \nabla\cdot\mathbf{B}_1 &= 0 \\ \nabla\times\mathbf{B}_1 &= \mu_0\mathbf{j} \end{aligned} \tag{7.23}$$

As usual in wave analysis, $\mathbf{u},\mathbf{j},\mathbf{E}$ are treated as perturbations. The MHD wave equation for the electric field can then be obtained,

$$\begin{aligned} \dot{\mathbf{E}} &= -\dot{\mathbf{u}}\times\mathbf{B}_0 = -\frac{1}{\rho}(\mathbf{j}\times\mathbf{B})\times\mathbf{B}_0 = -\frac{1}{\mu_0\rho}[(\nabla\times\mathbf{B}_1)\times\mathbf{B}_0]\times\mathbf{B}_0 \\ \ddot{\mathbf{E}} &= [(\nabla\times(\nabla\times\mathbf{E}))\times\mathbf{V}_A]\times\mathbf{V}_A \end{aligned}$$

where $\mathbf{V}_A = \mathbf{B}_0/\sqrt{\mu_0\rho}$ is the Alfvén velocity, or if we mutate the triad cross terms,

$$\ddot{\mathbf{E}} = \mathbf{V}_A\times[\mathbf{V}_A\times\nabla\times(\nabla\times\mathbf{E})] \tag{7.24}$$

Alternatively, we can also get the MHD wave equation for the magnetic field:

$$\begin{aligned} &\begin{cases} \dot{\mathbf{B}}_1 = \nabla\times(\mathbf{u}\times\mathbf{B}_0) \\ (\nabla\times\mathbf{B}_1)\times\mathbf{B}_0 = \mu_0\mathbf{j}\times\mathbf{B}_0 = \mu_0\rho\dot{\mathbf{u}} \end{cases} \\ \Rightarrow \ddot{\mathbf{B}}_1 &= \nabla\times\left[\left(\frac{1}{\mu_0\rho}(\nabla\times\mathbf{B}_1)\times\mathbf{B}_0\right)\times\mathbf{B}_0\right] \end{aligned}$$

or

$$\ddot{\mathbf{B}}_1 = \nabla\times\left[((\nabla\times\mathbf{B}_1)\times\mathbf{V}_A)\times\mathbf{V}_A\right] \tag{7.25}$$

We will see soon that in cold MHD the slow mode ceases to exist, and the fast mode moves at Alfvén speed, such that along the magnetic field line, we only have a single wave mode.

7.6.2 Hot MHD

$$
\begin{aligned}
\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) &= 0 \\
\rho \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} &= -\nabla p + \mathbf{j} \times \mathbf{B} \\
\mathbf{j} &= \frac{1}{\mu_0} \nabla \times \mathbf{B} \\
\frac{\mathrm{d}}{\mathrm{d}t}\left(p\rho^{-\gamma}\right) &= 0 \\
\frac{\partial \mathbf{B}}{\partial t} &= -\nabla \times \mathbf{E} \\
\mathbf{E} &= -\mathbf{v} \times \mathbf{B}
\end{aligned}
$$

$\dot{\mathbf{E}}$ is ignored because we only consider low frequency waves. We assume no background flow, $\mathbf{u}_0 = 0$, so the current is purely caused by perturbed velocity $\mathbf{u}_1$. Performing linearization and plane wave decomposition:

$$
\begin{aligned}
-i\omega\rho_1 + i\rho_0 \mathbf{k} \cdot \mathbf{v} &= 0 \\
-i\omega\rho_0 \mathbf{v} &= -i\mathbf{k}p_1 + \mathbf{j} \times \mathbf{B}_0 \\
\mathbf{j} &= \frac{1}{\mu_0} i\mathbf{k} \times \mathbf{B}_1 \\
p_1/p_0 - \gamma\rho_1/\rho_0 &= 0 \\
-i\omega\mathbf{B}_1 &= i\mathbf{k} \times (\mathbf{v} \times \mathbf{B}_0)
\end{aligned}
$$

Let $\mathbf{B}_0 = B_0\hat{z}$. The linearized equations can be further simplified:

$$
\begin{aligned}
-i\omega\rho_0\mathbf{v} &= -i\mathbf{k}(\gamma p_0 \frac{\mathbf{k}\cdot\mathbf{v}}{\omega}) + \left[\frac{1}{\mu_0} i\mathbf{k} \times (-\frac{\mathbf{k}\times(\mathbf{v}\times\mathbf{B}_0)}{\omega})\right] \times \mathbf{B}_0 \\
\omega^2\mathbf{v} - {V_s}^2\mathbf{k}(\mathbf{k}\cdot\mathbf{v}) - {V_A}^2[\mathbf{k}\times(\mathbf{k}\times(\mathbf{v}\times\hat{z}))] \times \hat{z} &= 0
\end{aligned}
$$

where $V_s = \sqrt{\frac{\gamma p_0}{\rho_0}}$ is the sound speed, and $V_A = \sqrt{\frac{{B_0}^2}{\mu_0\rho_0}}$ is the Alfvén speed. If we write $\mathbf{V}_A = \mathbf{B}_0/\sqrt{\mu_0\rho_0}$, this can also be written as

$$
\omega^2\mathbf{v} - {V_s}^2\mathbf{k}(\mathbf{k}\cdot\mathbf{v}) - \mathbf{k}\times\mathbf{k}\times(\mathbf{v}\times\mathbf{V}_A)\times\mathbf{V}_A = 0
$$

Due to the symmetry in the perpendicular x-y plane, for simplicity, we assume the wave vector $\mathbf{k}$ lies in the x-z plane with an angle w.r.t. the z axis θ:

$$
\mathbf{k} = k_x\hat{x} + k_z\hat{z} = k_x\hat{x} + k_\parallel\hat{z} = k\sin\theta\hat{x} + k\cos\theta\hat{z}
$$

Now it can be written as

$$
\begin{pmatrix}
-\omega^2/k^2 + v_A^2 + v_s^2\sin^2\theta & 0 & v_s^2\sin\theta\cos\theta \\
0 & -\omega^2/k^2 + v_A^2\cos^2\theta & 0 \\
v_s^2\sin\theta\cos\theta & 0 & -\omega^2/k^2 + v_s^2\cos^2\theta
\end{pmatrix}
\begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} = 0 \tag{7.26}
$$

7.6.3 Alfvén Wave

For any nonzero v_y, the y-component of Equation 7.26 gives

$$\omega^2 = k^2 v_A^2 \cos^2\theta = k_\parallel^2 v_A^2$$

which is known as the *Alfvén wave*, in a uniform plasma immersed in a uniform background magnetic field with phase speed

$$v_p = v_A \cos\theta$$

The group velocity and hence energy propagation is always parallel to $\mathbf{B}$ regardless of the direction of $\mathbf{k}$, and for this reason this mode is also know as the *guided* mode. This property, of course, has the direct bearing on the feature of Alfvén wave resonant absorption.

Given the velocity perturbation $\mathbf{v}_1 = (0, v_y, 0)$, $-i\omega\rho_1 + \rho_0\mathbf{k}\cdot\mathbf{v} = 0$, $\omega\mathbf{B}_1 + \mathbf{k}\times(\mathbf{v}\times\mathbf{B}_0) = 0$, the other perturbations are given as

$$\begin{aligned}
\rho_1 &= 0 \\
p_1 &= 0 \\
\mathbf{E} &= -B_0 v_y \hat{x} \\
\mathbf{B}_1 &= \frac{\mathbf{k}}{\omega}\times\mathbf{E} = -\frac{k_z B_0 v_y}{\omega}\hat{y} = -\frac{\mathbf{v}}{\omega/k_\parallel}B_0 \\
\mathbf{j} &= \frac{1}{\mu_0}\nabla\times\mathbf{B}_1 = \frac{i\mathbf{k}\times\mathbf{B}_1}{\mu_0}
\end{aligned} \tag{7.27}$$

SAW has a wave vector $\mathbf{k}$ in the XZ-plane. $\mathbf{E}$ shall oscillate in the X-direction; $\mathbf{B}$ shall oscillate in the Y-direction. The electric current of the wave $\mathbf{j}$ lies in the XZ-plane. The timescale of the variations of the wave fields is much longer than the ion gyroperiod Ω_i^{-1}. In both the perpendicular and parallel directions, the spatial scale of the waves $1/k$ are much larger than ion motion scale r_{iL}. The wave carries a Poynting flux $S = \mathbf{E}\times\mathbf{B}_1$ strictly parallel to $\mathbf{B}_0$. The ratio of the wave electric field to the wave magnetic field $|\mathbf{E}|/|\mathbf{B}_1|$ is exactly one Alfvén speed V_A.

$\mathbf{E}$ (or $\mathbf{B}_1$) in Equation 7.27 shows that the Alfvén wave in a uniform plasma is a linearly polarized wave if $\mathbf{v}_1 = (0, v_y, 0)$. If instead we set $\mathbf{k} = (0,0,k)$ $(\theta = 0^o)$ and $\mathbf{v}_1 = (v_x, v_y, 0)$, then $\mathbf{k}\cdot\mathbf{v} = 0 \Rightarrow \rho_1 = 0$ but $\mathbf{E} = -\mathbf{v}\times\mathbf{B}_0 = kv_y\hat{x} - kv_x\hat{y}$ can be a circularly polarized wave or else. Correlated $\mathbf{B}$ and $\mathbf{v}$ corresponds to waves propagating anti-parallel to the $\mathbf{B}_0$, and anti-correlated $\mathbf{B}$ and $\mathbf{v}$ corresponds to waves propagating parallel to the $\mathbf{B}_0$. The resultant magnetic field then exhibits shear, thus the Alfvén wave is called the *shear Alfvén wave* (SAW). An animation of SAW is shown in **?@fig-alfven_wave**. It is useful to remember that the magnetic perturbation is always parallel to the velocity perturbation if $\mathbf{k}$ is parallel to $\mathbf{B}_0$ and antiparallel if $\mathbf{k}$ is antiparallel to $\mathbf{B}_0$.

To understand what happens physically in an Alfvén wave, recall that this is an electromagnetic wave with a fluctuating magnetic field $\mathbf{B}_1$ given by

$$\nabla \times \mathbf{E}_1 = -\dot{\mathbf{B}}_1 \quad E_x = (\omega/k)B_y \tag{7.28}$$

The small component B_y, when added to $\mathbf{B}_0$, gives the magnetic field lines a sinuisoidal ripple, shown exaggerated in Figure 7.1. At the point shown, B_y is in the positive y direction, so, according to Equation 7.28, E_x is in the positive x direction if ω/k is in the z direction. The electric field E_x gives the plasma an $\mathbf{E}_1 \times \mathbf{B}_0$ drift in the negative y direction. Since we have taken the limit $\omega^2 \ll \Omega_c$, both ions and electrons will have the same drift v_y, obtained from Section 7.5 the component v_y under $T_i = 0$:

$$\begin{aligned} v_{ix} &= \frac{iq}{m\omega}\left(1 - \frac{\Omega_c^2}{\omega^2}\right)^{-1} E_1 \\ v_{iy} &= \frac{q}{m\omega}\frac{\Omega_c}{\omega}\left(1 - \frac{\Omega_c^2}{\omega^2}\right)^{-1} E_1 \end{aligned} \tag{7.29}$$

Thus, the fluid moves up and down in the y direction. The magnitude of this velocity is $|E_x/B_0|$. Since the ripple in the field is moving by at the phase velocity ω/k, the magnetic field is also moving downward at the point indicated in Figure 7.1. The downward velocity of the magnetic field lines is $(\omega/k)|B_y/B_0|$, which, according Equation 7.28, is just equal to the fluid velocity $|E_x/B_0|$. Thus, the fluid and the field lines oscillate together as if the particles were stuck to the lines. The magnetic field lines act as if they were mass-loaded strings under tension, and an Alfvén wave can be regarded as the propagating disturbance occurring when the strings are plucked. This concept of plasma frozen to the field lines and moving with them is a useful one for understanding many low-frequency plasma phenomena. It can be shown that this notion is an accurate one as long as there is no electric field along $\mathbf{B}$.

It remains for us to see what sustains the electric field E_x which we presupposed was there. As $\mathbf{E}_1$ fluctuates, the ions' inertia causes them to lag behind the electrons, and there is a polarization drift $\mathbf{v}_p$ in the direction of $\mathbf{E}_1$. This drift v_{ix} is given by Equation 7.29 and causes a current $\mathbf{j}_1$ to flow in the x direction. The resulting $\mathbf{j}_1 \times \mathbf{B}_0$ force on the fluid is in the y direction and is 90^o out of phase with the velocity $\mathbf{v}_1$. This force perpetuates the oscillation in the same way as in any oscillator where the force is out of phase with the velocity. It is, of course, always the ion inertia that causes an overshoot and a sustained oscillation, but in a plasma the momentum is transferred in a complicated way via the electromagnetic forces.

In a more realistic geometry for experiments, $\mathbf{E}_1$ would be in the radial direction and $\mathbf{v}_1$ in the azimuthal direction. The motion of the plasma is then incompressible. This is the reason the ∇p term in the equation of motion could be neglected.

In a non-uniform plasma, SAW attains the interesting property of a continuous spectrum. To illustrate this feature, let us consider the simplified slab model of a cold plasma with a non-uniform density, $\rho = \rho(x)$, and a uniform $\mathbf{B}_0 = B_0\hat{z}$. Assuming at $t = 0$ a localized initial

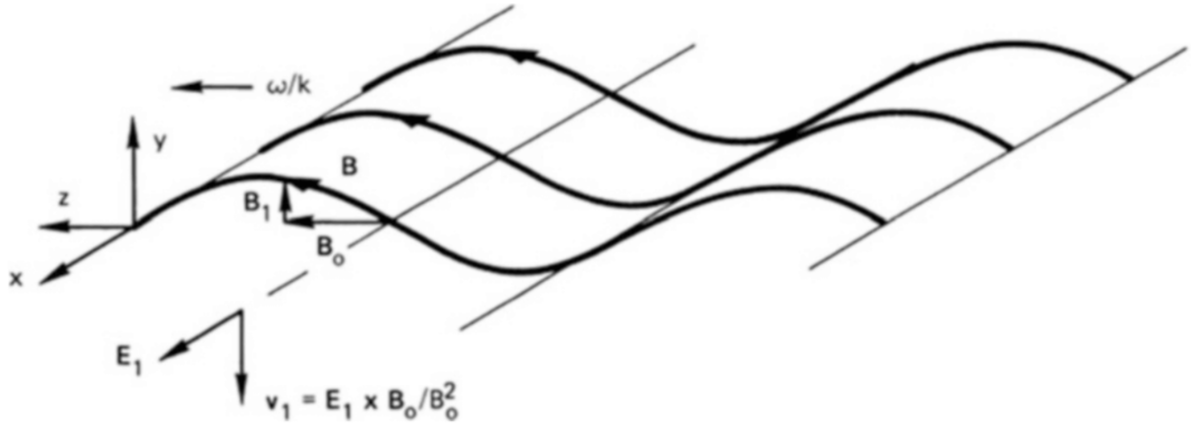

Figure 7.1: Relation among the oscillating quantities in an Alfvén wave and the (exaggerated) distortion of the lines of force.

perturbation $\mathbf{B}_{1y}(x, t = 0) = \exp(-x^2/\Delta x^2)$, $|k_y \Delta_x| \ll 1$, and $\partial \mathbf{B}_{1y}/\partial t = 0$, the perturbation then evolves according to the following wave equation (Equation 7.25, $B_{1z} = 0$ so no coupling between the fast mode and Alfvén mode):

$$[\partial_t^2 + \omega_A^2(x)]B_{1y}(x, t) = 0$$

Here $\omega_A^2(x) = k_z^2 v_A^2(x)$ and the solution is

$$B_{1y}(x, t) = \hat{B}_{1y}(x, 0) \cos[\omega_A(x)t] \tag{7.30}$$

Equation 7.30 shows that every point in x oscillates at a different frequency, $\omega_A(x)$. With a continuously varying $\omega_A(x)$; the wave frequency, thus, constitutes a continuous spectrum. While the above result is based on a model with a 1D non-uniformity in x, this general feature of SAW continuous spectrum also holds in magnetized plasmas with 2D or 3D non-uniformities. A good example is geomagnetic pulsations in the Earth's magnetosphere observed by Engebretson shown in Figure 1 of (L. Chen, Zonca, and Lin 2021).

Equation 7.30 also indicates an unique and important property of SAW continuous spectrum: the spatial structure evolves with time. Specifically, the wave number in the non-uniformity direction is, time asymptotically, given by:

$$|k_x| = \left|\frac{\partial \ln B_{1y}}{\partial x}\right| \simeq \left|\frac{\mathrm{d}\omega_A(x)}{\mathrm{d}x}\right| t \tag{7.31}$$

That $|k_x|$ increases with t is significant, since it implies that any initially long-scale perturbations will evolve into short scales. This point is illustrated in Figure ??? (CAN I PERFORM

THE SIMULATION?); showing the evolution of a smooth B_{1y} at $t = 0$ to a spatially fast varying B_{1y} at a later t.

Another consequence of $|k_x|$ increasing with t is the temporal decay of B_{1x}. From $\nabla \cdot \mathbf{B}_1 \simeq \nabla_\perp \cdot \mathbf{B}_{1\perp} = 0$, we can readily derive that, for $|\omega'_A t| \gg |k_y|$:

$$B_{1x}(x,t) \simeq \frac{k_y}{\omega'_A(x)t}\hat{B}_{1y}(x,0)e^{-i\omega_A(x)t}\Big[1 + \mathcal{O}\Big(\frac{k_y}{|\omega'_A t| + ...}\Big)\Big]$$

That is, B_{1x} decays temporally due to the phase mixing of increasingly more rapidly varying neighboring perturbations.

Noting that, as $t \to \infty$, $|k_x| \to \infty$, it thus suggests that the perturbation will develop singular structures toward the steady state. As we will see in the field line resonance Chapter 12, the singularity is reached at the Alfvén resonant point x_r, where $\omega^2 = \omega_A^2(x_r)$ along with a finite resonant wave-energy absorption rate. Note that at the isolated extrema of the SAW continuum, $|\omega'_A| = 0$, phase mixing vanishes; consequently, perturbation remains regular and experiences no damping via resonant absorption. This feature has important implications to Alfvén instabilities in laboratory plasmas.

Space plasmas support a variety of waves, but for heating the plasma and accelerating the electrons and ions, Alfvén waves are a predominant source. Near the Sun, Alfvén waves are excited and propagate outward. They exchange their energy with particles to accelerate them in the form of solar wind and heat the electrons. When perpendicular wavelength of the wave becomes comparable to the ion gyro-radius ($k_\perp r_{Li} \sim 1$) or inertial length ($k_\perp d_i \sim 1$) in the case of kinetic Alfvén waves (KAWs, Section 7.7) or inertial Alfvén waves (IAWs), respectively, then the wave in both the limits has a *nonzero parallel electric field* component which is responsible for the acceleration of the particles via the Landau mechanism. This is also consistent with the generalized Ohm's law Equation 5.17: only when we go beyond Hall MHD can $\mathbf{E}_\parallel$ be nonzero. (WHAT ABOUT $\eta\mathbf{j}$?) KAW relates to $\frac{\nabla P_e}{ne}$ term, while IAW relates to $\frac{\partial \mathbf{j}}{\partial t}$ term. The key interest is in $E_\parallel$. KAWs and IAWs have significance not only in space plasmas but also in laboratory plasma such as in fusion reactors.

Alfvén wave has very high saturation level, meaning that it takes a long time for the wave to reach the nonlinear phase. (???)

SAW in a Slab

We now look deeper into the properties of Alfvén waves in a nonuniform magnetized plasma slab that carries a current flowing along an externally imposed magnetic field $B_{0z}\hat{z}$, where B_{0z} is assumed to be a constant. First, we formulate the governing equation for the slab geometry, under the ideal MHD condition. Then we show that Alfvén waves are always neutrally stable, with important indication at the end.

The presence of an equilibrium current density $\mathbf{J}_0 = \hat{z}J_0(x)$ produces a local magnetic field of the form

$$\mathbf{B}_0 = \hat{z}B_{0z} + \hat{y}B_{0y}(x)$$

The Ampère's law gives

$$\nabla \times \mathbf{B}_0 = \mu_0 \mathbf{J}_0 \Rightarrow \frac{\partial B_{0y}}{\partial x} - \frac{\partial B_{0x}}{\partial y} = \frac{\partial B_{0y}}{\partial x} = \mu_0 J_0(x) \tag{7.32}$$

From the force balance equation,

$$\mathbf{J}_0 \times \mathbf{B}_0 = \nabla P_0 \tag{7.33}$$

Substituting Equation 7.32 into Equation 7.33, we get

$$\frac{B_{0y}(x)^2}{2\mu_0} + P_0(x) = \text{const.}$$

Designate all perturbation quantities with a subscript 1, and assume $e^{-i\omega t + ik_y y + ik_z z}$ dependence for all perturbations (nonuniform in the x-direction, thus no sinuisoidal wave assumption). From linearized Faraday's law and Ohm's law in ideal MHD,

$$\begin{aligned} -\nabla \times \mathbf{E}_1 = \nabla \times (\mathbf{v}_1 \times \mathbf{B}_0) &= \frac{\partial \mathbf{B}_1}{\partial t} \\ -i\omega \mathbf{B}_1 = \mathbf{v}_1 \cancel{(\nabla \cdot \mathbf{B}_0)} - \mathbf{B}_0 \cancel{(\nabla \cdot \mathbf{v}_1)} + (\mathbf{B}_0 \cdot \nabla)\mathbf{v}_1 &- (\mathbf{v}_1 \cdot \nabla)\mathbf{B}_0 \end{aligned}$$

where we have assumed the plasma is incompressible. Replace ∇ with $i\mathbf{k}$, $\mathbf{v}_1 = i\omega \boldsymbol{\xi}_1$ and take the x-component, we get

$$B_{1x} = i(\mathbf{k} \cdot \mathbf{B}_0)\xi_{1x} \tag{7.34}$$

where $\mathbf{k} = \hat{y}k_y + \hat{z}k_z$.

The MHD force law can be linearized to

$$\rho_0 \frac{\partial \mathbf{v}_1}{\partial t} = -\nabla\Big(p_1 + \frac{\mathbf{B}_0 \cdot \mathbf{B}_1}{\mu_0}\Big) + \frac{1}{\mu_0}[(\mathbf{B}_0 \cdot \nabla)\mathbf{B}_1 + (\mathbf{B}_1 \cdot \nabla)\mathbf{B}_0] \tag{7.35}$$

Since the plasma is incompressible, $\nabla \cdot \boldsymbol{v}_1 = 0$, $\dot{\boldsymbol{\xi}} = \mathbf{v}_1 \Rightarrow \nabla \cdot \boldsymbol{\xi} = 0$. In addition, $\nabla \cdot \mathbf{B}_1 = 0$. We then have

$$\begin{aligned} \mathbf{k} \cdot \boldsymbol{\xi}_{1yz} &= i\frac{\partial \xi_{1x}}{\partial x} \\ \mathbf{k} \cdot \boldsymbol{B}_{1yz} &= i\frac{\partial B_{1x}}{\partial x} \end{aligned}$$

where $\boldsymbol{\xi}_{1yz} = (0, \xi_{1y}, \xi_{1z})$, $\mathbf{B}_{1yz} = (0, B_{1y}, B_{1z})$. The x-component of Equation 7.35 gives

$$-\rho_0\omega^2\xi_{1x} = -\frac{\partial}{\partial x}\Big(p_1 + \frac{\mathbf{B}_0 \cdot \mathbf{B}_1}{\mu_0}\Big) + \frac{1}{\mu_0}[(\mathbf{B}_0 \cdot \nabla)B_{1x}] \tag{7.36}$$

The dot product of Equation 7.35 with $\mathbf{k}$ gives

$$
\begin{aligned}
-\rho_0\omega^2\mathbf{k}\cdot\boldsymbol{\xi}_{1yz} &= ik^2\Big(p_1+\frac{\mathbf{B}_0\cdot\mathbf{B}_1}{\mu_0}\Big)+\frac{1}{\mu_0}[i(\mathbf{k}\cdot\mathbf{B}_0)(\mathbf{k}\cdot\mathbf{B}_{1yz})+B_{1x}\frac{\partial}{\partial x}(\mathbf{k}\cdot\mathbf{B}_0)]\\
-\rho_0\omega^2 i\frac{\partial\xi_{1x}}{\partial x} &= ik^2\Big(p_1+\frac{\mathbf{B}_0\cdot\mathbf{B}_1}{\mu_0}\Big)+\frac{1}{\mu_0}[i(\mathbf{k}\cdot\mathbf{B}_0)(i\frac{\partial B_{1x}}{\partial x})+B_{1x}\frac{\partial}{\partial x}(\mathbf{k}\cdot\mathbf{B}_0)]
\end{aligned}
\tag{7.37}
$$

Finally, canceling $p_1+\frac{\mathbf{B}_0\cdot\mathbf{B}_1}{\mu_0}$ from Equation 7.36 and Equation 7.37 and substituting B_{1x} from Equation 7.34, we obtain the governing equation

$$
\frac{\partial}{\partial x}\Big\{\rho_0[\omega^2-(\mathbf{k}\cdot\mathbf{v}_A)^2]\frac{\partial\xi_{1x}}{\partial x}\Big\}-k^2\rho_0[\omega^2-(\mathbf{k}\cdot\mathbf{v}_A)^2]\xi_{1x}=0 \tag{7.38}
$$

where $k^2={k_y}^2+{k_z}^2$, $\mathbf{v}_A=v_A\mathbf{B}_0/B_0$, and $v_A=B_0/\sqrt{\mu_0\rho_0}$ is the local Alfvén speed. This is the governing equation of shear Alfvén waves in a slab geometry derived by Hasegawa and Liu Chen in the 1970s, which is readily compared with Eq.(10.33) in (Bellan 2008).

It is easy to show that this governing equation always yields *neutrally stable* solutions of SAWs, i.e. $\omega_i=\Im(\omega)=0$. Multiply it by ξ_{1x}^*, and integrate the resultant equation to get

$$
\int_{-\infty}^{\infty}\mathrm{d}x\rho_0[\omega^2-(\mathbf{k}\cdot\mathbf{v}_A)^2]\Big[\Big|\frac{\mathrm{d}\xi_{1x}}{\mathrm{d}x}\Big|^2+k^2|\xi_{1x}|^2\Big]=0
$$

where we have assumed that ξ_{1x} vanishes on the boundary. This gives

$$
\omega^2=\frac{\int_{-\infty}^{\infty}\mathrm{d}x\rho_0(\mathbf{k}\cdot\mathbf{v}_A)^2\Big[\Big|\frac{\mathrm{d}\xi_{1x}}{\mathrm{d}x}\Big|^2+k^2|\xi_{1x}|^2\Big]}{\int_{-\infty}^{\infty}\rho_0[|\frac{\mathrm{d}\xi_{1x}}{\mathrm{d}x}|^2+k^2|\xi_{1x}|^2]\mathrm{d}x}\geq 0
$$

SAWs are the dominant low frequency waves in a current carrying plasma. The neutrally stable modes studies above can be destabilized by unfavorable curvature, and such modes are called *ballooning modes* (Section 9.7.4). They may also be destabilized by a finite electrical resistivity, and these are *tearing modes* (Section 9.7.5). Their interaction with fusion-generated alpha particles are a major issue in all magnetic fusion schemes. Finally, since the governing equation exhibits a singularity when $\omega=\mathbf{k}\cdot\mathbf{v}_A$, this singularity represents resonance absorption, which forms the basis of *Alfvén wave heating* (i.e. field line resonance, Chapter 12). This singularity also give rise to the so called "Alfvén continuum spectrum" mentioned above.

Note that the governing equation is valid even if B_{0z} is an arbitrary function of x. If in addition, an external gravity $\mathbf{g}=\hat{x}g$ in the x-direction is present, the governing equation is modified simply by inserting the term $-(g/\rho_0)\mathrm{d}\rho_0/\mathrm{d}x$ in the second square bracket, and the equation is identical to Eq.(10.33) of Bellan. This is the most general equation which describes the *magneto-Rayleigh-Taylor instability* (MRT) in Cartesian geometry using the incompressible, ideal MHD model.

7.6.4 Fast and Slow Wave

The x-z components of Equation 7.26 give

$$
\begin{aligned}
(\omega^2 - k^2{v_A}^2 - {k_x}^2{v_s}^2)v_x - k_x k_z {v_x}^2 v_z = 0 \\
(\omega^2 - {k_z}^2{v_s}^2)v_z - k_x k_z {v_s}^2 v_x = 0
\end{aligned}
$$

The dispersion relation is given by the determinant being 0,

$$
\begin{aligned}
\omega^4 - k^2({v_A}^2 + {v_s}^2)\omega^2 + {k_z}^2{v_s}^2k^2{v_A}^2 = 0 \\
\frac{\omega^2}{k^2} = \frac{1}{2}({v_A}^2 + {v_s}^2) \pm \frac{1}{2}\sqrt{({v_A}^2 + {v_s}^2)^2 - 4{v_s}^2{v_A}^2\cos^2\theta}
\end{aligned}
\tag{7.39}
$$

"+" corresponds to the fast mode, or magnetosonic mode, and "-" corresponds to the slow mode. The Friedrich graph is very useful in interpreting Equation 7.39.

Given the velocity perturbation $\mathbf{v}_1 = (v_x, 0, v_z)$, the other perturbations are given as

$$
\begin{aligned}
\rho_1 &= \frac{\rho_0}{\omega}\mathbf{k}\cdot\mathbf{v} \\
p_1 &= \gamma p_0 \frac{\rho_1}{\rho_0} = \frac{\gamma p_0}{\omega}\mathbf{k}\cdot\mathbf{v} \\
\mathbf{E} &= B_0 v_x \hat{y} \\
\mathbf{B}_1 &= \frac{\mathbf{k}}{\omega}\times\mathbf{E} = \frac{(\mathbf{k}\cdot\mathbf{v})\mathbf{B}_0 - (\mathbf{k}\cdot\mathbf{B}_0)\mathbf{v}}{\omega} = -\frac{B_0 v_x k_z}{\omega}\hat{x} + \frac{B_0 v_x k_x}{\omega}\hat{z} \\
\mathbf{j} &= \frac{1}{\mu_0}\nabla\times\mathbf{B}_1 = \frac{i\mathbf{k}\times\mathbf{B}_1}{\mu_0}
\end{aligned}
\tag{7.40}
$$

$\mathbf{E}$ shows that the fast/slow modes in a uniform plasma are also linearly polarized. When we have $\theta = 90^o$ (**?@fig-fast_wave**), the magnetic perturbation can be simplified to

$$
\mathbf{B}_1 = \frac{v}{\omega/k_\perp}\mathbf{B}_0
$$

which means that the perturbed magnetic field is always aligned with the background magnetic field.

The distinction between the fast and slow waves can be further understood by comparing the signs of the wave induced fluctuations in the plasma and magnetic pressures: p and $\mathbf{B}_0\cdot\mathbf{B}/\mu_0$, respectively. It follows from Equation 7.40 that

$$
\frac{\mathbf{B}_0\cdot\mathbf{B}}{\mu_0} = \frac{\mathbf{k}\cdot\mathbf{B}_0^2 - (\mathbf{k}\cdot\mathbf{B}_0)(\mathbf{B}_0\cdot\mathbf{v})}{\mu_0\,\omega}
$$

The z-component of the perturbed momentum equation yields

$$\omega\rho_0 v_z = k\cos\theta p$$

Combining the above, we have

$$\frac{\mathbf{B}_0\cdot\mathbf{B}}{\mu_0} = \frac{V_A^2}{V_s^2}\left(1-\frac{k^2V_s^2\cos^2\theta}{\omega^2}\right)p \tag{7.41}$$

Hence, p and $\mathbf{B}_0\cdot\mathbf{B}/\mu_0$ have the same sign if $V_{\text{ph}} = \omega/k > V_s\,\cos\theta$, and the opposite sign if $V_{\text{ph}} < V_S\,\cos\theta$. It is straightforward to show that $V_+ > V_s\,\cos\theta$, and $V_- < V_s\,\cos\theta$. Thus, we conclude that in the fast magnetosonic wave the pressure and magnetic energy fluctuations reinforce one another, whereas the fluctuations oppose one another in the slow magnetosonic wave.

The temperature perturbation, based on the equation of state, can be derived as follows:

$$\begin{aligned}
T_1 = T - T_0 &= \frac{p_0+p_1}{(n_0+n_1)k_B} - \frac{p_0}{n_0k_B}\\
&= p_0\left(1+\gamma\frac{n_1}{n_0}\right)\frac{1}{(n_0+n_1)k_B} - \frac{p_0}{n_0k_B}\\
&= p_0\frac{1}{k_B}\left[\left(1+\gamma\frac{n_1}{n_0}\right)\frac{1}{n_0+n_1}-\frac{1}{n_0}\right]\\
&= p_0\frac{1}{k_B}\left[\frac{(\gamma-1)n_1}{n_0(n_0+n_1)}\right]\\
&= p_0\frac{\gamma-1}{k_B}\left[\frac{\rho_1}{\rho_0(\rho_0+\rho_1)}\right]\\
&= p_0\frac{\gamma-1}{k_B}\frac{\mathbf{k}\cdot\mathbf{v}}{\rho_0(\omega+\mathbf{k}\cdot\mathbf{v})}\\
&= T_0\frac{\gamma-1}{m}\frac{\mathbf{k}\cdot\mathbf{v}}{\omega+\mathbf{k}\cdot\mathbf{v}}
\end{aligned} \tag{7.42}$$

Equation 7.42 shows that temperature does not follow the same sinusoidal shape as the perturbed velocity.

Special attention is required for the extreme cases. When $\theta = 90^o$, the fast wave phase speed $v_{\text{ph}} = \sqrt{(V_A^2+V_s^2)}$ while the slow wave phase speed $v_{\text{ph}} = 0$. Slow waves cannot propagate perpendicular to the magnetic field. However, this is still a valid solution with all the slow mode perturbation properties, and you will see in Section 9.11.2 and Section 10.2 that this is named *mirror mode*. When $\theta = 0^o$, the fast wave phase speed $v_{\text{ph}} = \max(V_A, V_s)$ while the slow wave phase speed $v_{\text{ph}} = \min(V_A, V_s)$.

It is important to understand the nature of fast and slow waves under the *cold plasma limit*, i.e. in the low-β regime where $V_A \gg V_s$ (cold, strong B). If $V_s \to 0$, the slow mode ceases to exit ($V_{\text{ph}} \to 0$) and the phase speed of the fast mode becomes $v_p \approx v_A$. This is often called the *compressional Alfvén wave*, which is actually a fast wave in the low-β limit. If β is low but we still have a finite sound speed, the dispersion relation for the slow wave reduces to

$$\omega \simeq k V_s \cos\theta$$

Thus, in low-β plasmas the slow wave is a sound wave modified by the presence of the magnetic field.

7.7 Kinetic Alfvén Wave

The solution of Equation 7.31 exhibits singularities as $t \to \infty$ naturally suggests that the microscopic length-scale physics neglected in the ideal MHD fluid description should be included in the long-time-scale dynamics of SAWs. For low-frequency SAWs, one can readily recognize the relevant perpendicular (to $\mathbf{B}_0$) microscopic scales are either the ion Larmor radius, $r_{iL} = v_{ti}/\omega_{ci}$ with v_{ti} and ω_{ci} being, respectively, the ion thermal speed and ion cyclotron frequency, and/or $r_{eL} = v_{te}/\omega_{ce}$ with v_{te} being the electron thermal speed. Including the effects of finite r_{iL} and/or r_{eL} in the SAW dynamics then led to the discovery of the so-called *kinetic Alfvén wave* (KAW) [Hasegawa and Chen].

In KAWs, parallel electric field $E_\parallel$ can be developed and facilitate particle heating, acceleration, and transport. It has been found in the plasma sheet, at the plasma sheet boundary layer (PSBL), and in the inner magnetosphere.

KAW differ from SAW because the short wavelength requires a significant $E_\parallel$ to maintain charge neutrality due to ion density perturbations caused by the ion polarization drift. When $v_{te} > v_A$, the parallel electric field counteracts electron pressure that would push the electrons away from the ion density perturbations. When $v_{te} < v_A$, the electric field must overcome the electron inertia that prevents the electrons from responding rapidly to the ion density perturbations. $E_\parallel$ associated with small-scale KAWs may efficiently accelerate particles on the magnetic field lines. (Chaston+ 2009) presented observations in the magnetotail from the Cluster spacecraft showing that KAWs radiate outward from the X-line with outward energy fluxes equivalent to that contained in the outstreaming ions. Wave-particle energy exchange between KAWs and plasmas near the dayside magnetopause has been confirmed by MMS observations (Gershman+ 2017).

We consider a KAW with a wave vector $\mathbf{k}$ in the XZ-plane, same as in the SAW case. In many aspects, KAW is similar to SAW: $\mathbf{E}$ shall oscillate in the X-direction; $\mathbf{B}$ shall oscillate in the Y-direction. The electric current of the wave $i\mathbf{k}\times\mathbf{B}_1/\mu_0$ is still in the XZ-plane. The timescale of the variations of the wave field of KAW is much longer than the ion gyroperiod. However, unlike SAW, KAW has

- a perpendicular scale $k_\perp^{-1}$ that is comparable to the particle kinetic scale r_{iL}.

- a very oblique wave vector $k_\perp \gg k_\parallel$ so that the wave is not strongly affected by the Landau damping.
- $k_\perp^{-1} \ll k_\parallel^{-1}$. The finite-Larmor-radius effect starts to become important as the perpendicular wavelength is comparable to the ion gyromotion: Ions can not follow the $\mathbf{E}\times\mathbf{B}$ drift in the electric fields of KAW, because they encounter significantly different electric field in the different phases of the gyromotion. Electrons are still frozen-in in the presence of the wave field. The difference in the ion and electron motion in the perpendicular direction introduces charge separation and coupling to the electrostatic mode. Because the wave electric field E_x is mainly parallel to the $\mathbf{k}$, $\nabla\times\mathbf{E} = i\mathbf{k}\times\mathbf{E}$ is small and $\nabla\cdot\mathbf{E} = i\mathbf{k}\cdot\mathbf{E}$ is relatively large for KAW. Accordingly, the perpendicular wave electric field E_x is mostly electrostatic in KAW.
- Because of charge separation in KAW, electrons need to move along the magnetic field to preserve the charge neutrality. Associated with the parallel motion of electron, a small wave electric field $E_\parallel$ is established, the existence of which is a distinct feature of KAW. The parallel motion of electrons creates a field-aligned current $j_\parallel$ of KAW. From Ampère's law $i\mathbf{k}\times\mathbf{B}_1 = \mu_0\mathbf{j}$, the field-aligned current $\mathbf{j}_z$ produces a wave magnetic field $\mathbf{B}_{1y}$. As a result, KAW is in fact an EM wave. The ratio of the wave electric field to the wave magnetic field is $V_A\sqrt{1+k_x^2 r_{iL}^2}$ (Stasiewicz et al. 2000). The kinetic correction $k_x^2 r_{iL}^2$ introduces a deviation of the ratio E_x/B_y from one V_A as in the SAWs.

In principle, the KAW eigenmode solution can be obtained from the linearized two-fluid equations (Dai 2009).

While SAW satisfy $\omega = k_\parallel v_A$, the dispersion relation of KAW can be written as (Johnson & Cheng, 1997)

$$\omega^2 = k_\parallel^2 v_A^2 \Big[\frac{1}{1 - I_0(k_\perp^2 r_{iL}^2)e^{-k_\perp^2 r_{iL}^2}} + \frac{T_e}{T_i}\Big] k_\perp^2 r_{iL}^2 \tag{7.43}$$

where I_0 is the modified Bessel function. Using a Padé approximation $I_0(x)e^{-x} \approx 1/(1+x)$ when $x = k_\perp^2 r_{iL}^2 \sim \mathcal{O}(1)$, the relation can be simplified to

$$\omega^2 = k_\parallel^2 v_A^2 \Big[1 + \Big(1 + \frac{T_e}{T_i} k_\perp^2 r_{iL}^2\Big)\Big] = k_\parallel^2 v_A^2 [1 + k_\perp^2 (r_{iL}^2 + r_{sL}^2)] \tag{7.44}$$

where $r_{sL} = \sqrt{T_e/m_i}/\omega_{ci}$. This term is kind of strange, because it looks like a Larmor radius for ion with electron temperature. We can see from Equation 7.44 that the phase speed for KAW is larger than v_A.

While SAWs do not have ion density perturbations, KAWs do (???).

The polarizations of KAW can be expressed as

$$\left|\frac{\delta \mathbf{E}_\perp}{\delta \mathbf{B}_\perp}\right| = v_A(1+k_\perp^2 r_{iL}^2)[1+k_\perp^2(r_{iL}^2+r_{sL}^2)]^{-1/2}$$

where $\delta \mathbf{E}_\perp \perp \delta \mathbf{B}_\perp$. The parallel electric field is

$$\left|\frac{\delta E_\parallel}{\delta E_\perp}\right| = k_\parallel k_\perp \frac{r_{sL}^2}{1+k_\perp^2 r_{iL}^2}$$

It has been shown that KAWs can be generated via magnetic reconnection, mode conversion, and phase mixing.

(Gurram, Egedal, and Daughton 2021) used a 2D VPIC simulation to show a transition from KAWs to SAWs from the immediate vicinity of the reconnection region to the exhaust. They checked two important quantities we have seen above:

- the transition in wave speeds from super Alfvénic near the X-point to order of the Alfvén speed deeper into the exhaust ($\sim 60 d_i$ from the X-point);
- the transition of $k_\perp d_i$ from larger than unity to smaller than unity.

It would be very intuitive to compare animations between MHD Alfvén waves and kinetic Alfvén waves. In a typical MHD Alfvén wave, the particles (yellow) move freely along the magnetic field lines (blue). In a kinetic Alfvén wave, some particles become trapped in the weak spots of the wave's magnetic field and ride along with the wave as it moves through space.

Almost half a century after the discovery of KAW, (L. Chen, Zonca, and Lin 2021) demonstrated that the proper treatment of this wave requires gyrokinetics (Chapter 11). However, under some circumstances ($\beta \gg 1$?) the two-fluid theory can recover most if not all of the KAW physics. I need to go over the derivations!

7.8 Particle Motions

In Fourier transform space, integrating the velocity equations to obtain the coordinates is done by simply dividing by $-i\omega$. For a simple case where $E_y = E_z = 0$ so that E has only an x-component, we find from Equation 7.6

$$\begin{aligned}
x_s &= -\frac{q_s E_x}{m_s(\omega^2-\omega_j^2)} \\
y_s &= \frac{\pm\omega_{cs}}{i\omega} x_s
\end{aligned}$$

so that in general, the trajectory is elliptical. For $\omega \ll \omega_{cs}$, we find $x_s \ll y_s$, so the motion is principally across both the $\mathbf{E}$ and $\mathbf{B}_0$ directions. However, for $\omega \gg \omega_{cj}, x_j \gg y_j$ the motion is principally parallel to the electric field. In this latter case, we would call the particles *unmagnetized*, since the magnetic influence is small. Since it is possible for the wave frequency to be well above the ion cyclotron frequency at the same time it is well below the electron cyclotron frequency, it is possible for ions to be effectively unmagnetized while electrons are magnetized. From the discussion here, it is also clear that in the MHD low frequency regime the $\mathbf{E} \times \mathbf{B}$ drift is important, whereas for high frequency regimes (i.e. unmagnetized), the electric field influence is more important.

When $\omega \simeq \omega_{cs}$, then the linear solutions exhibit resonance effects with large amplitudes, and at resonance, the radius increases uniformly in time and no steady-state solution exists. In this vicinity, we expect the cold plasma approximation to fail and either thermal, inhomogeneous, or nonlinear effects to dominate the dynamics.

7.9 Cold Nonuniform Plasma

Waves and dispersion relations in a uniform plasma is generally nice and easy. However, more interesting and realistic waves shall be found in nonuniform plasmas.

7.9.1 Simple EM Wave

Let us start off by examining a very simple case. Consider a plane electromagnetic wave, of frequency ω, propagating along the z-axis in an unmagnetized plasma whose refractive index, n, is a function of z. We assume that the wave normal is initially aligned along the z-axis, and, furthermore, that the wave starts off polarized in the y-direction. It is easily demonstrated that the wave normal subsequently remains aligned along the z-axis, and also that the polarization state of the wave does not change. Thus, the wave is fully described by

$$E_y(z,t) \equiv E_y(z)\, \exp(-i\omega t)$$

and

$$B_x(z,t) \equiv B_x(z)\, \exp(-i\omega t)$$

It can easily be shown (???) that $E_y(z)$ and $B_x(z)$ satisfy the differential equations

$$\frac{\mathrm{d}^2 E_y}{\mathrm{d}z^2} + {k_0}^2\, n^2\, E_y = 0 \tag{7.45}$$

and

$$\frac{\mathrm{d}\,cB_x}{\mathrm{d}z} = -i\,k_0\,n^2\,E_y \tag{7.46}$$

respectively. Here, $k_0 = \omega/c$ is the wave-number in free space. Of course, the actual wave-number is $k = k_0\,n$.

The solution to Equation 7.45 for the case of a homogeneous plasma, for which n is constant, is straightforward:

$$E_y = A\,e^{\,i\,\phi(z)} \tag{7.47}$$

where A is a constant, and

$$\phi = \pm k_0\,n\,z \tag{7.48}$$

The solution Equation 7.47 represents a wave of constant amplitude, A, and phase, $\phi(z)$. According to Equation 7.48, there are, in fact, two independent waves which can propagate through the plasma. The upper sign corresponds to a wave which propagates in the $+z$-direction, whereas the lower sign corresponds to a wave which propagates in the $-z$-direction. Both waves propagate with the constant phase velocity c/n.

In general, if $n = n(z)$ then the solution of Equation 7.45 does not remotely resemble the wave-like solution Equation 7.47. However, in the limit in which $n(z)$ is a "slowly varying" function of z (exactly how slowly varying is something which will be established later on), we expect to recover wave-like solutions. Let us suppose that $n(z)$ is indeed a "slowly varying" function, and let us try substituting the wave solution Equation 7.47 into Equation 7.45. We obtain

$$\left(\frac{\mathrm{d}\phi}{\mathrm{d}z}\right)^2 = {k_0}^2 n^2 + i\,\frac{\mathrm{d}^2\phi}{\mathrm{d}z^2} \tag{7.49}$$

This is a non-linear differential equation which, in general, is very difficult to solve. However, we note that if n is a constant then $\mathrm{d}^2\phi/\mathrm{d}z^2 = 0$. It is, therefore, reasonable to suppose that if $n(z)$ is a "slowly varying" function then the last term on the right-hand side of the above equation can be regarded as being small. Thus, to a first approximation Equation 7.49 yields

$$\frac{\mathrm{d}\phi}{\mathrm{d}z} \simeq \pm k_0\,n$$

and

$$\frac{\mathrm{d}^2\phi}{\mathrm{d}z^2} \simeq \pm k_0 \frac{dn}{\mathrm{d}z} \tag{7.50}$$

It is clear from a comparison of Equation 7.49 and Equation 7.50 that $n(z)$ can be regarded as a "slowly varying" function of z as long as its variation length-scale is far longer than the wavelength of the wave. In other words, provided that $(dn/\mathrm{d}z)/(k_0\, n^2) \ll 1$.

The second approximation to the solution is obtained by substituting Equation 7.50 into the right-hand side of Equation 7.49:

$$\frac{\mathrm{d}\phi}{\mathrm{d}z} \simeq \pm\Big(k_0{}^2 n^2 \pm i\, k_0 \frac{dn}{\mathrm{d}z}\Big)^{1/2}$$

This gives

$$\frac{\mathrm{d}\phi}{\mathrm{d}z} \simeq \pm k_0\, n\Big(1 \pm \frac{i}{k_0 n}\Big)^{1/2} \simeq \pm k_0\, n + \frac{i}{2n}\frac{dn}{\mathrm{d}z}$$

where a binomial expansion has been used. The above expression can be integrated to give

$$\phi \sim \pm k_0 \int^z n\, \mathrm{d}z + i\, \log(n^{1/2}) \tag{7.51}$$

Substitution of Equation 7.51 into Equation 7.47 yields the final result

$$E_y \simeq A\, n^{-1/2}\, \exp(\pm i\, k_0 \int^z n\, \mathrm{d}z) \tag{7.52}$$

It follows from Equation 7.46 that

$$cB_x \simeq \mp A\, n^{1/2} \exp\Big(\pm\, i\, k_0 \int^z n\mathrm{d}z\Big) - \frac{iA}{2k_0 n^{3/2}}\frac{dn}{\mathrm{d}z} \exp\Big(\pm\, i\, k_0 \int^z n\mathrm{d}z\Big) \tag{7.53}$$

Note that the second term is small compared to the first, and can usually be neglected.

Let us test to what extent Equation 7.52 is a good solution of Equation 7.45 by substituting this expression into the left-hand side of the equation. The result is

$$\frac{A}{n^{1/2}}\Big[\frac{3}{4}\Big(\frac{1}{n}\frac{dn}{\mathrm{d}z}\Big)^2 - \frac{1}{2n}\frac{\mathrm{d}^2 n}{\mathrm{d}z^2}\Big] \exp\Big(\pm\, i\, k_0 \int^z n\, \mathrm{d}z\Big)$$

This must be small compared with either term on the left-hand side of Equation 7.45. Hence, the condition for Equation 7.52 to be a good solution of Equation 7.45 becomes

$$\frac{1}{k_0^2}\left|\frac{3}{4}\left(\frac{1}{n^2}\frac{dn}{\mathrm{d}z}\right)^2 - \frac{1}{2n^3}\frac{\mathrm{d}^2 n}{\mathrm{d}z^2}\right| \ll 1 \tag{7.54}$$

The solutions Equation 7.52 and Equation 7.53 (without the second term) are most commonly referred to as the *WKB* solutions, in honour of G. Wentzel, H.A. Kramers, and L. Brillouin, who are credited with independently discovering these solutions (in a quantum mechanical context) in 1926. Actually, H. Jeffries wrote a paper on the WKB solutions (in a wave propagation context) in 1923. Hence, some people call them the WKBJ solutions (or even the JWKB solutions). To be strictly accurate, the WKB solutions were first discussed by Liouville and Green in 1837, and again by Rayleigh in 1912. The advance in science discovery is always a collective achievement.

Recall, that when a propagating wave is normally incident on an interface, where the refractive index suddenly changes (for instance, when a light wave propagating through air is normally incident on a glass slab), there is generally significant reflection of the wave. However, according to the WKB solutions, when a propagating wave is normally incident on a medium in which the refractive index changes slowly along the direction of propagation of the wave then the wave is not reflected at all. This is true even if the refractive index varies very substantially along the path of propagation of the wave, as long as it varies slowly. The WKB solutions imply that as the wave propagates through the medium its wave-length gradually changes. In fact, the wave-length at position z is approximately $\lambda(z) = 2\pi/k_0\, n(z)$. The WKB solutions also imply that the amplitude of the wave gradually changes as it propagates. The amplitude of the electric field component is inversely proportional to $n^{1/2}$, whereas the amplitude of the magnetic field component is directly proportional to $n^{1/2}$. Note, however, that the energy flux in the z-direction, given by the the Poynting vector $-(E_y B_x^* + E_y^* B_x)/(4\mu_0)$, remains constant (assuming that n is predominately real).

Of course, the WKB solutions are only approximations. In reality, a wave propagating into a medium in which the refractive index is a slowly varying function of position is subject to a small amount of reflection. However, it is easily demonstrated that the ratio of the reflected amplitude to the incident amplitude is of order $(dn/\mathrm{d}z)/(k_0\, n^2)$. Thus, as long as the refractive index varies on a much longer length-scale than the wavelength of the radiation, the reflected wave is negligibly small. This conclusion remains valid as long as the inequality Equation 7.54 is satisfied. This inequality obviously breaks down in the vicinity of a point where $n^2 = 0$. We would, therefore, expect strong reflection of the incident wave from such a point. Furthermore, the WKB solutions also break down at a point where $n^2 \to \infty$, since the amplitude of B_x becomes infinite.

7.9.2 Electron Cyclotron Resonance Heating

Let us look at the problem of electron cyclotron resonance heating. The resonance condition is $R = \infty$; the governing equation has the form

$$\frac{\mathrm{d}^2 E}{\mathrm{d}z^2} + \frac{\omega^2}{c^2}\left[1 - \frac{\omega_{pe}(z)^2}{\omega(\omega - \omega_{ce}(z))}\right]E = 0$$

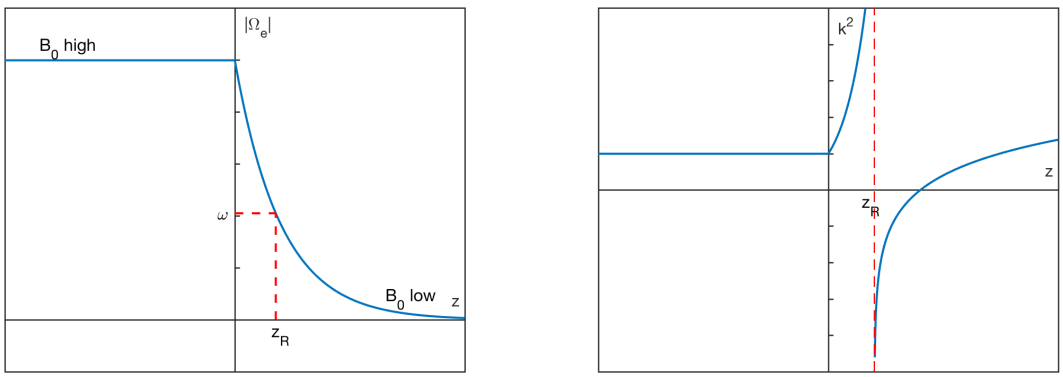

Figure 7.2: Plasma frequency and wave number as functions of z in a non-uniform plasma for the electron cyclotron resonance mode.

We use z here to remind ourselves of the fact that the wave property changes along the field line. Imagine a wave $\sim e^{i(kz-\omega t)}$ incident into a plasma with fixed density n_0 and varying magnetic field $B_0(z)$ as shown in Figure 7.2(a). At $z = z_R$, $\omega_{ce}(z) = \omega_{ce}(z_R) = \omega$. Then we can draw $k^2(z)$ as a function of z as in Figure 7.2(b). There is a pole at $z = z_R$, which indicates resonance since $k^2 \to \infty$. There is also a zero on the right of z_R.

Close to $z = z_R$, we have $k^2(z) \sim \frac{\text{const.}}{z - z_R}$. If we write

$$\Omega_e(z) = -A'(z - z_R) + \omega, \quad \zeta = z - z_R$$

then (ignore the constants)

$$\frac{\mathrm{d}^2 E}{\mathrm{d}z^2} - \frac{\text{const.}}{z - z_R}E = 0 \Rightarrow \frac{\mathrm{d}^2 E}{d\zeta^2} - \frac{1}{\zeta}E = 0$$

Now let's stare at Figure 7.2(b) for a few seconds. If there is a wave from left to right, there will be a resonance at $z = z_R$; but if there is a wave from right to left, then the wave will be attenuated before it reaches $z = z_R$ because there is a zero ahead. This means that waves from different origin will have different behaviors!

7.9.3 O-mode

Next consider EM waves in non-magnetized plasma of ordinary O-mode:

$$k^2(x) = \frac{\omega^2}{c^2} n^2 = \frac{\omega^2}{c^2}\left[1 - \frac{\omega_{pe}(x)^2}{\omega^2}\right] \quad \text{if } n \text{ is non-uniform}$$

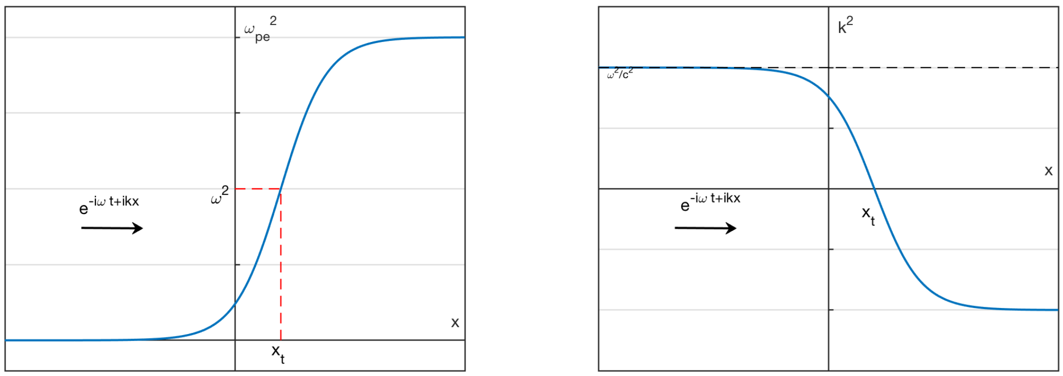

Figure 7.3: Plasma frequency and wave number as functions of x in a non-uniform plasma for O mode.

Note that there is no energy dissipation, because it is collisionless. Imagine a wave $\sim e^{i(kx-\omega t)}$ incident into a plasma with density $n_0(x)$ shown in Figure 7.3. We encounter a cutoff at $x = x_t$, $\omega_{pe}(x_t) = \omega$. Therefore we can draw $k^2(x)$ as a function of x as in Figure 7.3. Then the properties of the wave can be categorized into two regimes:

$$\begin{aligned} k^2 < 0 \Rightarrow \ & E \sim e^{\pm ikx} \text{ propagating} \\ k^2 > 0 \Rightarrow \ & E \sim e^{\pm |k| x} \text{ attenuating} \end{aligned}$$

A natural question comes up from this picture: *how does the propagating wave transform into attentuating wave*? Stokes solved this during his honey moon, which is now known as the Stokes phenomenon (Section 3.10.3).

7.10 Warm Uniform Plasma

The analysis of cold plasma waves, although very complicated already, leaves out all of the physics that relates to finite temperature effects. These effects may be included in varying degrees of approximation, and it is intructive to include at this stage only the simplest thermal correction terms through the inclusion of a finite pressure term. For the sake of simplicity, we only consider the 1D case. The approach is very similar to cold plasma situation, except that we have *pressure* included in the equation, and we also need to specify the relation of pressure and temperature through the equation of state.

7.10.1 Two-fluid approach

Equilibrium:

$$
\begin{aligned}
&m_i = \infty,\ T_i = 0,\ v_{i,0} = 0,\ n_{e0} = n_{i0} = n_0,\ E_0 = 0,\ B_0 = 0 \\
&P_0 = const.,\ v_{e0} = 0
\end{aligned}
$$

The equation of motion for electron is

$$
\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v}\cdot\nabla\mathbf{v} = -\frac{e}{m_e}\mathbf{E} - \frac{\nabla P}{\rho}
$$

Assume

$$
\begin{aligned}
P &= P_0\Big(\frac{n}{n_0}\Big)^\gamma \qquad\qquad \text{adiabatic} \\
P &= nk_BT_e \qquad \text{isothermal}(\gamma = 1)
\end{aligned}
$$

Define $P_1 = \Big(\partial P/\partial\rho\Big)\rho_1 \equiv {v_{th}}^2\rho_1 = {v_{th}}^2 n_1 m = k_BTn_1$. Decompose the primitive variables into equilibrium and perturbation components:

$$
\begin{aligned}
n &= n_0 + n_1 = n_0 + \tilde{n}_1 e^{-i\omega t+ikx} \\
v &= \cancel{v_0} + v_1 = \tilde{v}_1 e^{-i\omega t+ikx} \\
P &= P_0 + P_1 = P_0 + \tilde{P}_1 e^{-i\omega t+ikx}
\end{aligned}
$$

Substituting into the equation of motion and keeping only first order terms, we get

$$
\begin{aligned}
\frac{\partial v_1}{\partial t} &= -\frac{e}{m_e}(E_1) - \frac{\nabla(P_1)}{n_0 m_e} \\
-i\omega v_1 &= -\frac{e}{m_e}E_1 - \frac{ik{v_{th}}^2 n_1}{n_0}
\end{aligned}
$$

The continuity equation gives

$$
\begin{aligned}
\frac{\partial n}{\partial t} + \frac{\partial}{\partial x}\Big(nv\Big) &= 0 \\
\Rightarrow v_1 &= \frac{\omega n_1}{kn_0}
\end{aligned}
$$

Substituting this into the linearized equation of motion, we get

$$n_1 = \frac{kn_0eE_1}{im_e(\omega^2 - k^2{v_{th}}^2)}$$

which is the density perturbation in response to E_1. Following the same approach as before, we can easily get the dielectric function:

$$\frac{\epsilon}{\epsilon_0} = 1 - \frac{{\omega_{pe}}^2}{\omega^2 - k^2{v_{th}}^2}$$

Let $\epsilon = 0$, we get the dispersion relation for warm plasma

$$\omega^2 = k^2{v_{th}}^2 + {\omega_{pe}}^2$$

This is called the *Bohm-Gross dispersion relation* (BGDR). (Swanson 2012) has shown that by keeping higher order terms,

$$\omega^2 = \omega_{pe}^2\left[1 + \frac{3}{2}\left(\frac{kv_{th}}{\omega_{pe}}\right)^2 + \frac{15}{4}\left(\frac{kv_{th}}{\omega_{pe}}\right)^4\right]$$

so it is apparent that *the moment expansion is an expansion in the ratio of the thermal velocity to the phase velocity.* The BGDR now resolves the ambiguity in cold plasma theory and supports the notion that $P = 0$ in cold plasma is a cutoff rather than a resonance, since this dispersion relation describes a wave with a cutoff at ω_{pe} that propagates near the electron thermal speed for high frequencies.

We see from this example that the fluid equations which are based on moment expansions are valid as long as the phase velocity is large compared to the thermal speed, $v_p \gg v_{th}$.

7.10.2 Vlasov approach

Assume Maxwellian distribution for electrons:

$$g_e(v) = \frac{1}{\sqrt{2\pi}}\frac{1}{v_{th,e}}e^{-v^2/2{v_{th,e}}^2}$$

Assuming $\frac{\omega}{k} \gg v_{th,e}$, i.e. the phase speed is much larger than the characteristic thermal speed, we can do Taylor series expansion

$$\frac{1}{(v-\omega/k)^2} = \frac{1}{(\omega/k)^2}\frac{1}{(1-kv/\omega)^2} \approx \frac{k^2}{\omega^2}\Big[1 + \frac{2kv}{\omega} + \frac{3k^2v^2}{\omega^2} + ...\Big]$$

Then the dielectric function is

$$\begin{aligned}\frac{\epsilon}{\epsilon_0} &= 1 - \frac{{\omega_{pe}}^2}{k^2}\int_{-\infty}^{\infty} dv g(v)\frac{k^2}{\omega^2}\Big[1 + \frac{2kv}{\omega} + \frac{3k^2v^2}{\omega^2} + ...\Big] \\ &\doteq 1 - \frac{{\omega_{pe}}^2}{\omega^2}\Big[1 + \frac{3k^2{v_{th}}^2}{\omega^2}\Big].\end{aligned}$$

Let $\epsilon = 0$, we can get the dispersion relation for warm plasma,

$$\begin{aligned}&\text{lowest order: } \omega = \pm\omega_{pe} \\ &\text{first order:} 1 - \frac{{\omega_{pe}}^2}{\omega^2}\Big(1 + \frac{3k^2{v_{th}}^2}{{\omega_{pe}}^2}\Big) = 0 \Rightarrow \omega^2 = {\omega_{pe}}^2 + 3k^2{v_{th,e}}^2\end{aligned}$$

Note here we insert 0^{th} order solution to 1^{st} order equation to get the next level approximation. Comparing with the results from 2-fluid theory, we see that the expression is very similar except a discrepancy in the coefficient. This is owing to the fact that we do not specify γ in the equation of state. Actually, there are still ambiguities and debates about the exact equation of state. I wonder if I can get the exact coefficient under some assumption.

The opposite limit case: $\frac{\omega}{k} \ll v_{th}$. From 2-fluid theory,

$$\frac{\epsilon}{\epsilon_0} \approxeq 1 + \frac{{\omega_{pe}}^2}{k^2{v_{th,e}}^2}$$

From Vlasov theory,

$$\begin{aligned}\frac{\epsilon}{\epsilon_0} &= 1 - \frac{{\omega_{pe}}^2}{k^2}\int_{-\infty}^{\infty} dv\frac{\partial g/\partial v}{v - \omega/k} = 1 - \frac{{\omega_{pe}}^2}{k^2}\int_{-\infty}^{\infty}\frac{1}{v - \cancel{\omega/k}}\frac{1}{\sqrt{2\pi}}\frac{1}{v_{th,e}}\Big(\frac{-v}{{v_{th,e}}^2}\Big)e^{-v^2/2{v_{th,e}}^2}dv \\ &= 1 + \frac{{\omega_{pe}}^2}{k^2}\int_{-\infty}^{\infty}\frac{1}{v_{th,e}^2}\frac{1}{\sqrt{2\pi}}e^{-v^2/2{v_{th,e}}^2}dv \\ &= 1 + \frac{{\omega_{pe}}^2}{k^2{v_{th,e}}^2}\end{aligned}$$

So we can see in the two limit cases that they "almost" give the same results!

If we include ion motion in the 2-fluid theory ($n_{1i} \neq 0, T_i \neq 0, m_i \neq \infty$), applying the linear superposition property, we have

$$\frac{\epsilon}{\epsilon_0} = \underbrace{1}_{\nabla\cdot(\epsilon_0\mathbf{E}_1)} - \underbrace{\frac{{\omega_{pe}}^2}{\omega^2 - k^2{v_{th,e}}^2}}_{en_{1e}} - \underbrace{\frac{{\omega_{pi}}^2}{\omega^2 - k^2{v_{th,i}}^2}}_{en_{1i}}$$

7.10.3 Ion-acoustic wave

Assume $v_{th,e} \gg \frac{\omega}{k} \gg v_{th,i}$, we have the simplified dielectric function

$$\frac{\epsilon}{\epsilon_0} = 1 + \frac{{\omega_{pe}}^2}{k^2 {v_{th,e}}^2} - \frac{{\omega_{pi}}^2}{\omega^2}$$

Assume quasi-neutrality condition: $n_{1e} \approx n_{1i}$, s.t.

$$\nabla \cdot (\epsilon_0 \mathbf{E}_1) = e(n_{1e} - n_{1i}) \approx 0$$

so we can ignore the "1" in the dielectric function. Let $\epsilon = 0$, we get

$$\begin{aligned}\frac{{\omega_{pe}}^2}{k^2 {v_{th,e}}^2} - \frac{{\omega_{pi}}^2}{\omega^2} &= 0 \\ \Rightarrow \omega^2 = k^2 {v_{th,e}}^2 \frac{m_e}{m_i},\ \frac{\omega}{k} &= \sqrt{\frac{k_B T_e}{m_i}} \equiv c_s\end{aligned}$$

Physically, electron sees the electric field created by ions due to ion plasma oscillation. Since electrons move much faster than ions, this electric field is nearly electrostatic for electrons. As a result, electron just follows the ion motion.

There are some other ways to get the ion-acoustic wave. If $\omega \ll \omega_{pe}$, we can treat this wave as electrostatic wave for electron. The distribution for electron number density is

$$\begin{aligned}n_e = n_0 e^{e\phi/k_B T_e} &\approxeq n_{0e} + n_{1e} \approxeq n_0[1 + \frac{e\phi}{k_B T_e} + ...] \\ \Rightarrow n_{1e} = n_0 \frac{e\phi}{k_B T_e} &= \frac{n_0 e E_1}{-ikm_e {v_{th,e}}^2}\end{aligned}$$

Note that here the tilde signs $\sim$ of the variables are neglected for convenience without ambiguity.

Do we need to calculate n_{1i}??? Yes. MORE to do here!

Also, from the linearized equation of motion for electron,

$$\begin{aligned}\frac{\partial v_{e1}}{\partial t} &= -\frac{e}{m_e} E_1 - \frac{\nabla P_1}{n_0 m_e} \\ -i\omega v_{e1} = -\frac{e}{m_e} E_1 - \frac{ikP_1}{n_0 m_e} &= -\frac{e}{m_e} E_1 - \frac{ikn_{1e} k_B T_e}{n_0 m_e}\end{aligned}$$

In the $\omega \rightarrow 0$ limit, $LHS \doteq 0$, we get

$$n_{1e} = \frac{n_0 e E_1}{-ikm_e {v_{th,e}}^2}$$

Then again we get the dielectric function through Poisson's equation.

7.11 Electrostatic Wave in a Magnetized Plasma

Now we continue to discuss the property of electrostatic waves with background magnetic field.

First let us introduce a useful result for continuity equation. Assuming $n_0 = n_0(\mathbf{x})$, $\mathbf{v}_0 = \mathbf{v}_0(\mathbf{x})$ in equilibrium, $\mathbf{x}_1 = \mathbf{x}_1(\mathbf{x}, t)$ is the perturbation in displacement. We can show that the linearized continuity equation has an equivalent form:

$$\frac{\partial n_1(\mathbf{x}, t)}{\partial t} + \nabla \cdot [n_0(\mathbf{x})\mathbf{v}_1(\mathbf{x}, t) + n_1(\mathbf{x}, t)\mathbf{v}_0(\mathbf{x}, t)] = 0 \Leftrightarrow n_1(\mathbf{x}, t) = -\nabla \cdot [n_0(\mathbf{x})\mathbf{x}_1(\mathbf{x}, t)]$$

The proof is related to mass conservation shown as follows. Intuitively, you can think of this as the degree of condensation only depends on displacement, not on how you get there (speed).

In 1D,

$$\begin{aligned} n_0 = n_0(x_0) &= \text{unperturbed density} \\ x = x_0 + x_1(x_0, t) &= \text{instantanenous position} \\ n(x_0, t) = n_0(x_0) + n_1(x_0, t) &= \text{total density} \end{aligned}$$

At time t, $[x_0, x_0 + \mathrm{d}x] \rightarrow [x_0 + x_1(x_0, t), x_0 + \mathrm{d}x + x_1(x_0 + \mathrm{d}x, t)]$. Due to mass conservation, we have (This looks like the derivation of deformation in fluid dynamics.)

$$\begin{aligned} n_0(x_0)\mathrm{d}x &= [n_0(x_0 + x_1) + n_1(x_0 + x_1, t)] \cdot [x_0 + \mathrm{d}x + x_1(x_0 + \mathrm{d}x, t) - (x_0 + x_1(x_0, t))] \\ &\approx \Big[n_0(x_0) + x_1 \frac{\partial n_0(x_0)}{\partial x_0} + n_1(x_0, t)\Big]\mathrm{d}x \Big[1 + \frac{\partial x_1(x_0, t)}{\partial x_0}\Big] \end{aligned}$$

$$
\begin{aligned}
&\Rightarrow n_0(x_0) = n_0(x_0) + n_0(x_0)\frac{\partial x_1(x_0,t)}{\partial x_0} + x_1\frac{\partial n_0(x_0)}{\partial x_0} + n_1(x_0,t)\\
&\Rightarrow n_1(x_0,t) = -n_0(x_0)\frac{\partial x_1(x_0,t)}{\partial x_0} - x_1\frac{\partial n_0}{\partial x_0} = -\frac{\partial}{\partial x_0}[n_0(x_0)x_1(x_0,t)]\\
&\Rightarrow n_1(x,t) = -\frac{\partial}{\partial x}[n_0(x)x_1(x,t)]
\end{aligned}
$$

The simplest equilibrium state in a constant magnetized plasma is

$$
\begin{gathered}
n_{i0} = n_{e0} = n_0,\ \mathbf{E}_0 = 0,\ \mathbf{B}_0 = B_0\hat{z}\\
\mathbf{v}_{e0} = \mathbf{v}_{i0} = 0,\ T_e = 0,\ T_i = 0,\ m_i = \infty
\end{gathered}
$$

Now introduce an electrostatic perturbation ($\mathbf{E}_1 = -\nabla\phi_1$)

$$
\mathbf{E}_1 = \tilde{\mathbf{E}}_1 e^{-i\omega t + i\mathbf{k}\cdot\mathbf{x}} = -i\mathbf{k}\tilde{\phi}_1 e^{-i\omega t + i\mathbf{k}\cdot\mathbf{x}}
$$

we can confirm that this is indeed an electrostatic perturbation since $\mathbf{E}_1 \parallel \mathbf{k}$ and $\mathbf{B}_1 = 0$.

Case 1: $\mathbf{k} = k_z\hat{z} \parallel \mathbf{B}_0$, i.e. parallel propagation. Then $\mathbf{E}_1 = \hat{z}E_{1z}e^{-i\omega t + ik_z z}$. This is the same as if there is no magnetic field, so the dielectric function is

$$
\frac{\epsilon}{\epsilon_0} = 1 - \frac{{\omega_{pe}}^2}{\omega^2}
$$

Case 2: $\mathbf{k} \perp \mathbf{B}_0$, i.e. perpendicular propagation. Without loss of generality, let $\mathbf{k} = k_x\hat{x}$. Then

$$
\mathbf{E}_1 = \hat{x}\tilde{E}_{1x}e^{-i\omega t + ik_x x}
$$

The equations of motion (cold plasma) in the x and y direction are

$$
\begin{aligned}
\ddot{x}_1 &= -\frac{e}{m_e}[E_{1x} + \dot{y}_1 B_0]\\
\ddot{y}_1 &= -\frac{e}{m_e}[-\dot{x}_1 B_0]\\
\Rightarrow \dot{y}_1 &= \frac{eB_0}{m_e}x_1 = |\Omega_e|x_1\\
\ddot{x}_1 &= -\frac{e}{m_e}[E_{1x} + |\Omega_e|\cdot y_1 B_0] = -\frac{e}{m_e}E_{1x} - {\Omega_e}^2 x_1,\\
\Rightarrow x_1 &= \frac{-\frac{e}{m_e}E_{1x}}{-\omega^2 + {\Omega_e}^2}
\end{aligned}
$$

Then we have the perturbed density in response to the perturbed electric field E_{1x}:

$$n_1 = -n_0 \nabla \cdot \mathbf{x}_1 = -n_0 ik_x x_1 = -n_0 ik_x \frac{-\frac{e}{m_e} E_{1x}}{-\omega^2 + {\Omega_e}^2}$$

From Poisson's equation, we get the dielectric function (the same method as before):

$$\frac{\epsilon}{\epsilon_0} = 1 - \frac{{\omega_{pe}}^2}{\omega^2 - {\Omega_e}^2}$$

Let $\epsilon = 0$, we have

$$\omega = \sqrt{{\omega_{pe}}^2 + {\Omega_e}^2} \equiv \omega_{UH}$$

which is called the *upper hybrid* frequency. This is the highest characteristic frequency in plasma. This upper hybrid wave is a havoc to some beam generator devices as it appears near the electron collector.

What if ions are included? Similar to previous derivations and notice that we are still within the range of linear theory, we have

$$\frac{\epsilon}{\epsilon_0} = 1 - \frac{{\omega_{pe}}^2}{\omega^2 - {\Omega_e}^2} - \frac{{\omega_{pi}}^2}{\omega^2 - {\Omega_i}^2}$$

For $\Omega_i \ll \omega \ll \Omega_e$ with quasi-neutrality condition $n_{1e} \approx n_{1i}$, we can have a simplified dispersion relation by letting $\epsilon = 0$:

$$\begin{aligned} \frac{{\omega_{pe}}^2}{{\Omega_e}^2} &= \frac{{\omega_{pi}}^2}{\omega^2} \\ \Rightarrow \omega &= \sqrt{{\omega_{pi}}^2 \frac{{\Omega_e}^2}{{\omega_{pe}}^2}} = \sqrt{|\Omega_e \Omega_i|} \equiv \omega_{LH} \end{aligned}$$

which gives us the *low hybrid wave* frequency. It equals to the geometric mean of the two cyclotron frequencies. Actually, this can be obtained from pure plasma motion argument. Recall that $\Omega_i \ll \omega \ll \Omega_e$ means that for electrons the plasma seems to be nonmagnetized, so they moves only under the electric field,

$$v_{1ex} = \frac{-eE_{1x}}{m_e(-i\omega)}$$

On the other hand, for ions the magnetic field is strong while electric field still exists, so it experiences polarization drift along the direction of perturbed electric field,

$$v_{1ix} = \frac{1}{\Omega_i}\frac{\partial}{\partial t}\Big(\frac{E_{1x}}{B_0}\Big)$$

Under quasi-neutrality condition, $v_{1ix} = v_{1ex}$, so we have

$$\frac{-eE_{1x}}{m_e(-i\omega)} = \frac{1}{\Omega_i}\frac{\partial}{\partial t}\Big(\frac{E_{1x}}{B_0}\Big) \Rightarrow \omega = \sqrt{|\Omega_e\Omega_i|}$$

If we consider warm plasma for 1D, there is an additional pressure term in the momentum equation. The continuity equation together with Poisson's equation give the relation of perturbed displacement and electric field:

$$\begin{aligned}
&n_1 = -\nabla\cdot(n_0\mathbf{x}_1) = -ik_x n_0 x_1 \\
&\nabla\cdot(\epsilon_0\mathbf{E}_1) = ik_x\epsilon_0 E_1 = -en_{1e} = eik_x n_0 x_1 \\
\Rightarrow &E_1 = \frac{en_0x_1}{\epsilon_0}
\end{aligned}$$

Substituting into the momentum equation, we get

$$\begin{aligned}
\ddot{x}_1 + {\Omega_e}^2 x_1 &= -\frac{e}{m_e}E_1 - \frac{\nabla P_1}{n_0 m_e} \\
&= -{\omega_{pe}}^2 x_1 - {k_x}^2 {v_{th,e}}^2 x_1
\end{aligned}$$

where $v_{th,e} = \sqrt{k_BT_e/m_e}$. This gives us (You can gain a sense of the equivalent force law from the dispersion relation.)

$$\omega^2 = {\omega_{pe}}^2 + {\Omega_e}^2 + {k_x}^2 {v_{th,e}}^2$$

This is also equivalent to the dielectric function

$$\frac{\epsilon}{\epsilon_0} = 1 - \frac{{\omega_{pe}}^2}{\omega^2 - {\Omega_e}^2 - {k_x}^2{v_{th,e}}^2}$$

7.12 Animation of Waves

7.13 CMA Diagram

The Clemmow-Mullaly-Allis (CMA) diagram classifies all EM + ES waves (including ions) in a cold magnetized plasma. However it is no longer useful in hot plasma waves.

- Resonances and cutoffs
- Mode conversions

7.14 Wave-Particle Interactions

Charged particles can exchange the energy with the wave electric field both parallel and perpendicular to the mean magnetic field (*Landau resonance*). Figure 7.4 left and middle panels show these resonance types and the associated parts of the distribution function schematically. The resonance parameters for the Landau and cyclotron resonances are

$$\xi_{\mathrm{L}} = \frac{\omega}{k_{\parallel} v_{\mathrm{th},s}} \tag{7.55}$$

$$\xi_{\mathrm{C}} = \frac{\omega - \Omega_s}{k_{\parallel} v_(\mathrm{th}, s)} \tag{7.56}$$

Here the frequency ω is measured in the plasma rest frame. Ω_s denotes the cyclotron frequency of particle species s (ion species and electrons), $k_{\parallel}$ the parallel component of the wavevector, and $v_{\mathrm{th},s}$ the particle thermal speed of species s. In general, the resonance parameter can be defined for arbitrary harmonics of the cyclotron frequency $m = 0, \pm 1, \pm 2, ...$:

$$\xi^{(m)} = \frac{\omega - m\Omega_s}{k_{\parallel} v_{\mathrm{th},s}} \tag{7.57}$$

Note that the resonance parameters above are defined for a Maxwellian plasma. A correction is needed when treating a non-Maxwellian plasma to find the suitable velocity-space gradient for the resonance. The resonance is efficient when the parameter ξ_{L} or ξ_{C} is on the order of unity. Strictly speaking, the wave damping (or particle acceleration) is most efficient, typically for $1 < \xi < 5$. The upper limit is not exact, but the resonance becomes gradually inefficient at larger values of ξ. For $\xi < 1$ the particle motion is slower than the wave propagation and the particles do not have a sufficient time for exchanging energy with the wave electric field. For $\xi > 5$ there are increasingly fewer particles with higher velocities for the resonance (higher than the thermal speed).

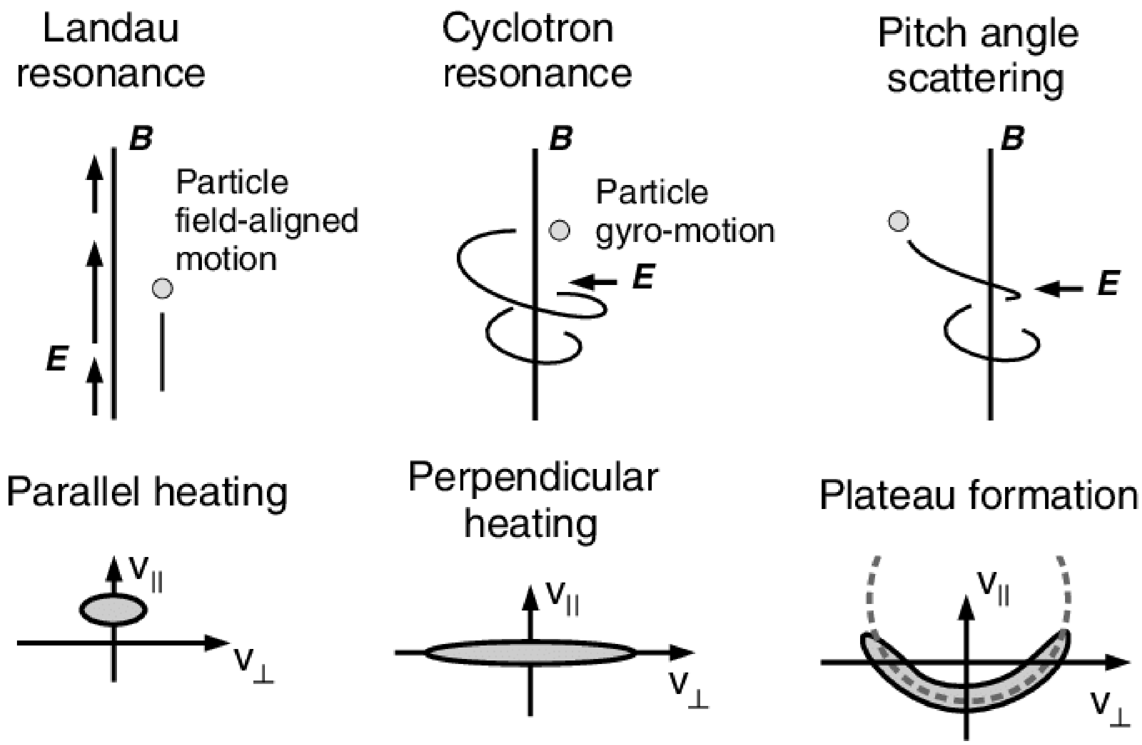

Figure 7.4: Wave-particle interactions and the associated part of the velocity distribution functions. Courtesy of Yasuhito Narita.

7.14.1 Pitch Angle Scattering

Charged particles can be scattered by the wave electric and magnetic fields incoherently, and the scattering deforms the velocity distribution function along the co-centric contours centered at the wave phase speed (Figure 7.4, right panel). The reason for the deformation is that the particle kinetic energy K_{wv} (per unit mass) does not change in the co-moving frame with the apparent wave phase speed in the parallel direction to the mean magnetic field:

$$K_{wv} = \frac{1}{2}\left[v_\perp^2 + \left(v_\parallel - \frac{\omega}{k_\parallel}\right)^2\right] = \text{const.}$$

The co-centric deformation of the distribution function achieves a quasi-linear equilibrium in that the velocity–space gradient becomes zero (plateau formation) in the pitch angle directions. For example, in the solar wind ions are found to be resonating with obliquely propagating Alfvén/ion cyclotron waves. Note that the relevant phase speed is $\omega/k_\parallel$, and is different from the true phase speed ω/k. The perpendicular component of the wavevector $k_\perp$ does not play a role in pitch angle scattering.

7.15 Whistler Wave

Whistler waves are a type of electron-scale plasma waves that contribute to electron scattering, acceleration, and energy transport. They are excited by the electron temperature anisotropy ($T_{e\perp}/T_{e\parallel} > 1$???) and shape the electron velocity distribution function through wave-particle interactions. (Ren+ 2020) Therefore, properties of whistler waves relate closely to electron-scale physics in magnetic reconnection. According to (Zhao 2017), the whistler wave can be described by *electron magnetohydrodynamics (EMHD?)* on timescales of $2\pi/\Omega_{ce} < t < 2\pi/\Omega LH$ and spatial scales of $L < \lambda_i$, where the ions are assumed as a motionless neutralizing background. When magnetic field lines are fully frozen into the electron fluid, the whistler wave follows the dispersion relation $\omega = \Omega_{ce}\lambda_e^2 k^2 \cos\theta$. The dispersion relation becomes $\omega = \Omega_{ce}\lambda_e^2 k^2 \cos\theta/(1+\lambda_e^2 k^2)$ as the electron inertia is contained in Ohm's law. Zhao further extended the model by taking the electron thermal pressure and the displacement current into the EMHD model.

We first see the derivations of whistler wave dispersion relation in Section 7.5 at the low frequency MHD limit. Another way to derive the whistler mode dispersion relation, which is probably easier, is to include the Hall term from the generalized Ohm's law:

$$\mathbf{E} = -\mathbf{U}\times\mathbf{B} + \frac{1}{ne}\mathbf{J}\times\mathbf{B}$$

Using Ampère's law and retaining only the Hall term leads to the equation

$$\dot{\mathbf{B}} = -\frac{1}{\mu_0 ne}\nabla \times [(\nabla \times \mathbf{B}) \times \mathbf{B}]$$

Performing linearization and assuming that the magnetic field is parallel to the z-axis and its perturbation is only in x and y, the last equation becomes

$$\omega B_{1x} = -i\frac{k_z^2 B_0}{\mu_0 ne}B_{1y}$$
$$\omega B_{1y} = i\frac{k_z^2 B_0}{\mu_0 ne}B_{1x}$$

which easily yields

$$\omega = \frac{B_0}{\mu_0 ne}k_z^2$$

The dispersion property of whistler waves makes it a problem for hybrid simulations (Section 24.2).

8 Kinetic Theory

8.1 Phase Space

Consider a particle moving in a one-dimensional space and let the position of the particle be $x = x(t)$ and the velocity of the particle be $v = v(t)$. A way to visualize the x and v trajectories simultaneously is to plot these trajectories parametrically on a two-dimensional graph, where the horizontal coordinate is given by $x(t)$ and the vertical coordinate is given by $v(t)$. This x-v plane is called *phase-space*. The trajectory (or orbit) of several particles can be represented as a set of curves in phase-space as shown in Figure 8.1. Examples of a few qualitatively different phase-space orbits are shown in Figure 8.1.

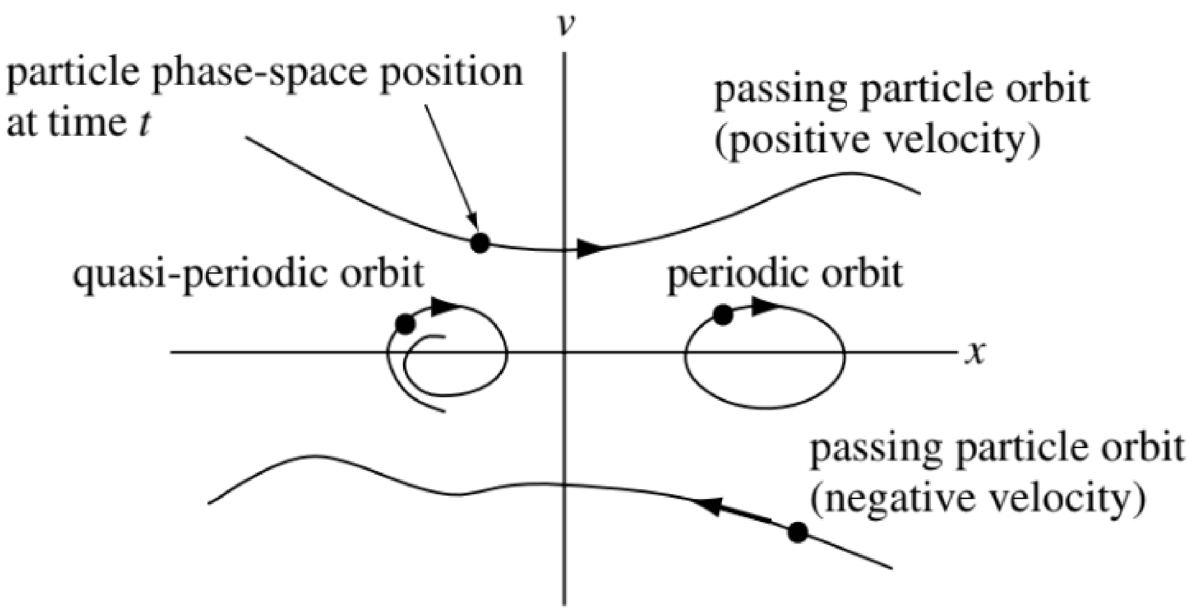

Figure 8.1: Phase-space showing different types of possible particle orbits.

Particles in the upper half-plane always move to the right, since they have a positive velocity, while those in the lower half-plane always move to the left. Particles having exact periodic motion (e.g., $x = A\cos t$, $v = -A\sin t$) alternate between moving to the right and the left and so describe an ellipse in phase-space. Particles with nearly periodic (quasi-periodic) motions will have near-ellipses or spiral orbits. A particle that does not reverse direction is called a passing particle, while a particle confined to a certain region of phase-space (e.g., a particle with periodic motion) is called a trapped particle.

8.2 Distribution Function

The fluid theory we have been using so far is the simplest description of a plasma; it is indeed fortunate that this approximation is sufficiently accurate to describe the majority of observed phenomena. There are some phenomena, however, for which a fluid treatment is inadequate. At any given time, each particle has a specific position and velocity. We can therefore characterize the instantaneous configuration of a large number of particles by specifying the density of particles at each point x, v in phase-space. The function prescribing the instantaneous density of particles in phase-space is called the *distribution function* and is denoted by $f(x, v, t)$. Thus, $f(x, v, t)\mathrm{d}x\mathrm{d}v$ is the number of particles at time t having positions in the range between x and $x + dx$ and velocities in the range between v and $v + dv$. As time progresses, the particle motion and acceleration causes the number of particles in these x and v ranges to change and so f will change. This temporal evolution of f gives a description of the system more detailed than a fluid description, but less detailed than following the trajectory of each individual particle. Using the evolution of f to characterize the system does not keep track of the trajectories of individual particles, but rather characterizes classes of particles having the same x, v.

In fluid theory, the dependent variables are functions of only four independent variables: x, y, z, and t. This is possible because the velocity distribution of each species is assumed to be Maxwellian everywhere and can therefore be uniquely specified by only one number, the temperature T. Since collisions can be rare in high-temperature plasmas, deviations from thermal equilibrium can be maintained for relatively long times. As an example, consider two velocity distributions $f_1(v_x)$ and $f_2(v_x)$ in a one-dimensional system (Figure 8.2). These two distributions will have entirely different behaviors, but as long as the areas under the curves are the same, fluid theory does not distinguish between them.

When we consider velocity distributions in 3D, we have seven independent variables: $f = f(\mathbf{r}, \mathbf{v}, t)$. By $f(\mathbf{r}, \mathbf{v}, t)$, we mean that the number of particles per meter cubed at position $\mathbf{r}$ and time t with velocity components between v_x and $v_x + dv_x$, v_y and $v_y + dv_y$, and v_z and $v_z + dv_z$ is

$$f(x, y, z, v_x, v_y, v_z, t)\mathrm{d}v_x\mathrm{d}v_y\mathrm{d}v_z$$

8.2.1 Moments of the distribution function

Let us count the particles in the shaded vertical strip in Figure 8.3. The number of particles in this strip is the number of particles lying between x and $x + \mathrm{d}x$, where x is the location of the left-hand side of the strip and $x + \mathrm{d}x$ is the location of the right-hand side. The number of particles in the strip is equivalently defined as $n(x, t)\mathrm{d}x$, where $n(x)$ is the density of particles at x. Thus we see that $\int f(x, v)\mathrm{d}v = n(x)$ the transition from a phase-space description (i.e.,

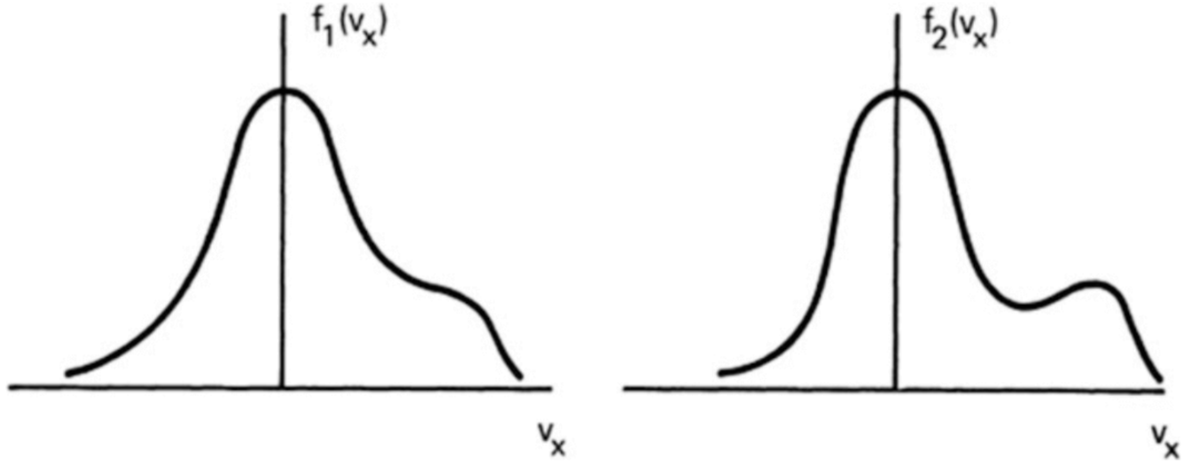

Figure 8.2: Examples of non-Maxwellian distribution functions

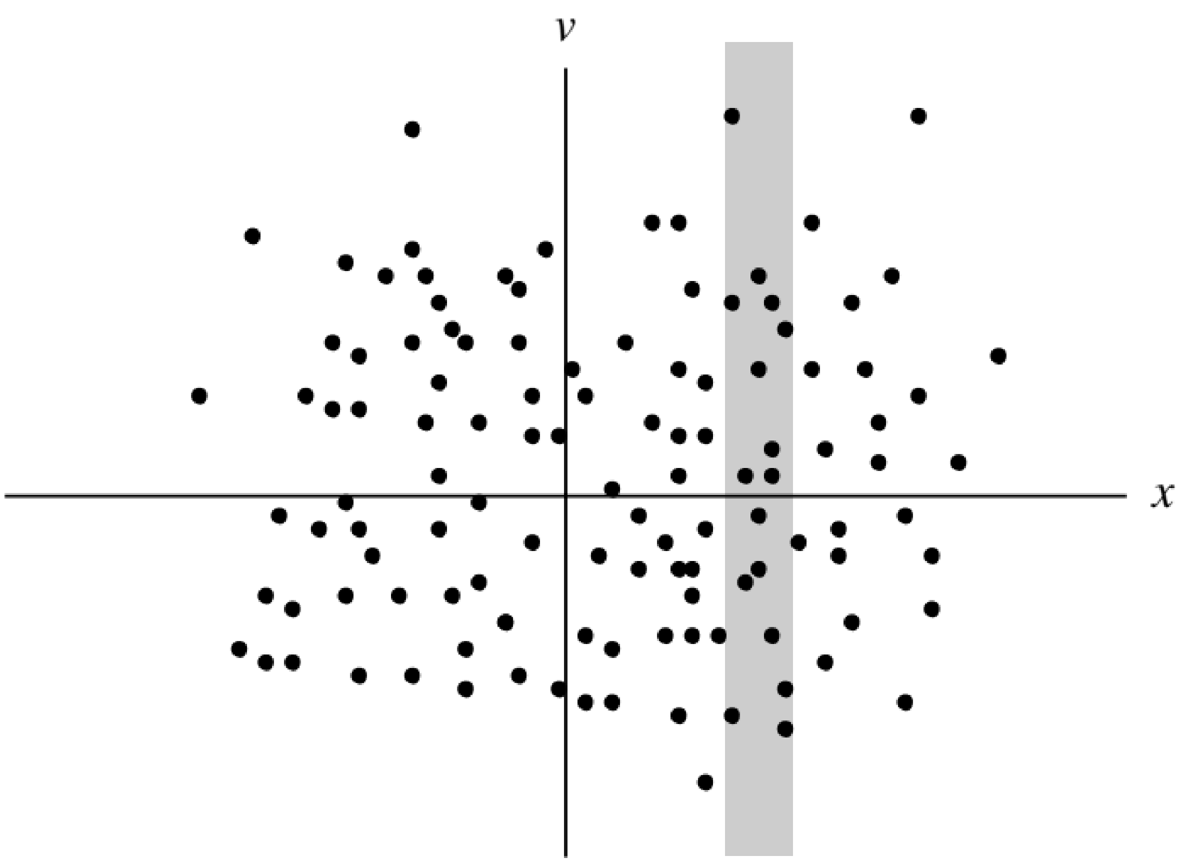

Figure 8.3: Moments give weighted averages of the particles in the shaded vertical strip.

x, v are independent variables) to a conventional description (i.e., only x is an independent variable) involves "integrating out" the velocity dependence to obtain a quantity (e.g., density) depending only on position. Since the number of particles is finite, and since f is a positive quantity, f must vanish as $v \to \pm\infty$.

In three-dimension, the density is now a function of four scalar variables, $n = n(\mathbf{r}, t)$, which is the integral of the distribution function over the velocity space:

$$\begin{aligned} n(\mathbf{r}, t) &= \int_{-\infty}^{\infty} \mathrm{d}v_x \int_{-\infty}^{\infty} \mathrm{d}v_y \int_{-\infty}^{\infty} \mathrm{d}v_z f(\mathbf{r}, \mathbf{v}, t) \\ &= \int_{-\infty}^{\infty} f(\mathbf{r}, \mathbf{v}, t) \mathrm{d}^3 v \\ &= \int_{-\infty}^{\infty} f(\mathbf{r}, \mathbf{v}, t) \mathrm{d}\mathbf{v} \end{aligned} \tag{8.1}$$

Note that $\mathrm{d}\mathbf{v}$ is not a vector; it stands for a three-dimensional volume element in velocity space. If f is normalized so that

$$\int_{-\infty}^{\infty} \hat{f}(\mathbf{r}, \mathbf{v}, t) \mathrm{d}\mathbf{v} = 1 \tag{8.2}$$

Thus

$$\hat{f}(\mathbf{r}, \mathbf{v}, t) = f(\mathbf{r}, \mathbf{v}, t) / n(\mathbf{r}, t) \tag{8.3}$$

is the probability that a randomly selected particle at position $\mathbf{r}$ has the velocity $\mathbf{v}$ at time t. Using this point of view, we see that averaging over the velocities of all particles at x gives the mean velocity

$$u(\mathrm{x}, t) = \frac{\int \mathbf{v} f(\mathbf{x}, \mathbf{v}, t) \mathrm{d}\mathbf{v}}{n(\mathbf{x}, t)} \tag{8.4}$$

Similarly, multiplying $\hat{f}$ by $mv^2/2$ and integrating over velocity will give an expression for the mean kinetic energy of all the particles. This procedure of multiplying f by various powers of $\mathbf{v}$ and then integrating over velocity is called *taking moments of the distribution function*.

Note that $\hat{f}$ is still a function of seven variables, since the shape of the distribution, as well as the density, can change with space and time. From Equation 8.2, it is clear that $\hat{f}$ has the dimensions $(\mathrm{m/s})^{-3}$; and consequently, from Equation 8.3, f has the dimensions $\mathrm{s}^3\mathrm{m}^{-6}$.

8.2.2 Maxwellian Distribution

A particularly important distribution function is the Maxwellian:

$$\hat{f}_m = \left(\frac{m}{2\pi k_B T} \right)^{3/2} \exp\left(-\frac{v^2}{v_{th}^2} \right) \tag{8.5}$$

where

$$v \equiv (v_x^2 + v_y^2 + v_z^2)^{1/2} \quad \text{and} \quad v_{th} \equiv (2k_B T/m)^{1/2}$$

This is the normalized form where $\hat{f}_m$ is equivalent to probability: by using the definite integral

$$\int_{-\infty}^{\infty} \exp(-x^2)dx = \sqrt{\pi}$$

one easily verifies that the integral of $\hat{f}_m$ over $\mathrm{d}v_x \mathrm{d}v_y \mathrm{d}v_z$ is unity.

A common question to ask is: why do we see Maxwellian/Gaussian/normal distribution ubiquitously in nature? Well, this is related to the central limit theorem in statistics: in many situations, when independent random variables are summed up, their properly normalized sum tends toward a normal distribution even if the original variables themselves are not normally distributed (e.g. a biased coin which give 95% head and 5% tail). In statistical physics, this is related to the fact that a Maxwellian distribution represents the state of a system with the highest entropy under the constraint of energy conservation.

There are several average velocities of a Maxwellian distribution that are commonly used. The root-mean-square velocity is given by

$$(\overline{v^2})^{1/2} = (3k_B T/m)^{1/2}$$

The average magnitude of the velocity $|v|$, or simply $\bar{v}$, is found as follows:

$$\bar{v} = \int_{-\infty}^{\infty} v\hat{f}(\mathbf{v})d^3v$$

Since $\hat{f}_m$ is isotropic, the integral is most easily done in spherical coordinates in v space. Since the volume element of each spherical shell is $4\pi v^2 \mathrm{d}v$, we have

$$\begin{aligned}
\bar{v} &= (m/2\pi k_B T)^{3/2} \int_0^{\infty} v[\exp(-v^2/v_{th}^2)] 4\pi v^2 \mathrm{d}v \\
&= (\pi v_{th}^2)^{-3/2} 4\pi v_{th}^4 \int_0^{\infty} [\exp(-y^2)] y^3 dy
\end{aligned}$$

The definite integral has a value 1/2, found by integration by parts. Thus

$$\bar{v} = 2\sqrt{\pi} v_{th} = 2(2k_B T/\pi m)^{1/2}$$

The velocity component in a *single direction*, say v_x, has a different average. Of course, $\bar{v}_x$ vanishes for an isotropic distribution; but $|\bar{v}_x|$ does not:

$$|\bar{v}_x| = \int |v_x| \hat{f}_m(\mathbf{v}) \mathrm{d}^3 v = \pi^{-1/2} v_{th} = (2k_B T/\pi m)^{1/2}$$

To summarize: for a Maxwellian,

$$
\begin{aligned}
v_{rms} &= (3k_B T/m)^{1/2} \\
|\bar{v}| &= 2(2k_B T/\pi m)^{1/2} \\
|\bar{v}_x| &= (2k_B T/\pi m)^{1/2} \\
\bar{v}_x &= 0
\end{aligned}
$$

For an isotropic distribution like a Maxwellian, we can define another function $g(v)$ which is a function of the scalar magnitude of $\mathbf{v}$ such that

$$
\int_0^\infty g(v)\mathrm{d}v = \int_{-\infty}^\infty f(\mathbf{v})\mathrm{d}v
$$

For a Maxwellian, we see that

$$
g(v) = 4\pi n(m/2\pi k_B T)^{3/2} v^2 \exp(-v^2/v_{th}^2) \tag{8.6}
$$

?@fig-f_g shows the difference between $g(v)$ and a one-dimensional Maxwellian distribution $f(v_x)$. Although $f(v_x)$ is maximum for $v_x = 0$, $g(v)$ is zero for $v = 0$.

This is just a consequence of the vanishing of the volume in phase space for $v = 0$. Sometimes $g(v)$ is carelessly denoted by $f(v)$, as distinct from $f(\mathbf{v})$; but $g(v)$ is a different function of its argument than $f(\mathbf{v})$ is of its argument. From eq-g_dist, it is clear that $g(v)$ has dimensions s/m^4.

ADD EXAMPLE DISTRIBUTIONS!

- isotropic
- anisotropic (pancake)

$$
f(v_\perp, v_\parallel) = \frac{n}{T_\perp T_\parallel^{1/2}} \Big(\frac{m}{2\pi k_B}\Big)^{3/2} \exp\Big(-\frac{mv_\perp^2}{2k_B T_\perp} - \frac{m(v_\parallel - v_{0\parallel})^2}{2k_B T_\parallel}\Big)
$$

- beam
- crescent shape

It is often convenient to present the distribution function as a function of energy instead of velocity. If all energy is kinetic, the energy is simply obtained from $W = mv^2/2$. In the case the particles are in the external electric potential field $U = -q\varphi$ the total energy of particles is $W = mv^2/2 + U$ and the Maxwellian distribution is

$$
f(v) = n\Big(\frac{m}{2\pi k_B T}\Big)^{3/2} \exp\Big(-\frac{W}{k_B T}\Big)
$$

This can be written as the energy distribution (???):

$$g(W) = 4\pi\Big[\frac{2(W-U)}{m^3}\Big]^{1/2} f(v)$$

The normalization factor is determined by requiring that the integration of the energy distribution over all energies gives the density.

Velocity and energy distribution functions cannot be measured directly. Instead, the observed quantity is *the particle flux to the detector*. Particle flux is defined as the number density of particles multiplied by the velocity component normal to the surface. We define the *differential flux* of particles traversing a unit area per unit time, unit solid angle (in spherical coordinates the differential solid angle is $d\Omega = \sin\theta d\theta d\phi$) and unit energy as $J(W, \Omega, \alpha, \mathbf{r}, t)$. ($\alpha$ is species?) The units of J are normally given as $(\mathrm{m^2 srseV})^{-1}$. Note that in literature cm is often used instead of m and, depending on the actual energy range considered, electron volts are often replaced by keV, MeV, or GeV.

Let us next define how differential particle flux and distribution function are related to each other. We can write the number density in a differential velocity element (in spherical coordinates $d^3v = v^2 dv d\Omega$) as ($dn = f(\alpha, \mathbf{r}, t)v^2 dv d\Omega$). By multiplying this with v we obtain another expression for the differential flux $f(\alpha, \mathbf{r}, t)v^3 dv d\Omega$. Comparing with our eailier definition of the differential flux we obtain

$$J(W, \Omega, \alpha, \mathbf{r}, t) dW d\Omega = f(\alpha, \mathbf{r}, t) v^3 dv d\Omega$$

Since $\mathrm{d}W = mvdv$ we can write the relationship between the differential flux and the distribution function as

$$J(W, \Omega, \alpha, \mathbf{r}, t) = \frac{v^2}{m} f \tag{8.7}$$

One application of the differential flux is the *particle precipitation flux*. With the idea of loss lone, we have a cone of particles that moves along the field lines and can propagate down to the ionosphere, and each shell from v to $v + \mathrm{d}v$ corresponds to a specific energy range. This is something we can measure close to the ground and use to infer the plasma properties in the magnetosphere.

8.2.3 Kappa Distribution

The Maxwellian distribution is probably the most studied one theoretically, but may not be the most commonly observed distribution in a collisionless space plasma system. In recent years, another distribution named Kappa distribution has gained more attention.

Distribution functions are often nearly Maxwellian at low energies, but they decrease more slowly at high energies. At higher energies the distribution is described better by a power law than by an exponential decay of the Maxwell distribution. Such a behavior is not surprising if we remember that the Coulomb collision frequency decreases with increasing temperature as $T^{-3/2}$ (Equation 6.23). Hence, it takes longer time for fast particles to reach Maxwellian distirbution than for slow particles. The kappa distribution has the form

$$f_\kappa(W) = n\Big(\frac{m}{2\pi\kappa W_0}\Big)^{3/2}\frac{\Gamma(\kappa+1)}{\Gamma(\kappa-1/2)}\Big(1+\frac{W}{\kappa W_0}\Big)^{-(\kappa+1)}$$

Here W_0 is the energy at the peak of the particle flux and Γ is the gamma function. When $\kappa \gg 1$ the kappa distribution approaches the Maxwellian distribution. When κ is smaller but > 1 the distribution has a high-energy tail. A thorough review is given by (Livadiotis and McComas 2013).

8.2.4 Entropy of a distribution

Collisions cause the distribution function to tend towards a simple final state characterized by having the maximum entropy for the given constraints (e.g. fixed total energy). To see this, we provide a brief discussion of entropy and show how it relates to a distribution function.

Suppose we throw two dice, labeled A and B, and let R denote the result of a throw. Thus R ranges from 2 through 12. The complete set of (A,B) combinations that gives these R's is listed in Table 8.1:

Table 8.1: All possible combinations of rolling two dices.

R	(A, B)
2	(1,1)
3	(1,2),(2,1)
4	(1,3),(2,2),(3,1)
5	(1,4),(2,3),(3,2),(4,1)
6	(1,5),(2,4),(3,3),(4,2),(5,1)
7	(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)
8	(2,6),(3,5),(4,4),(5,3),(6,2)
9	(3,6),(4,5),(5,4),(6,3)
10	(4,6),(5,5),(6,4)
11	(5,6),(6,5)
12	(6,6)

There are six (A, B) pairs that give $R = 7$, but only one pair for $R = 2$ and only one pair for $R = 12$. Thus, there are six microscopic states (distinct (A, B) pairs) corresponding to $R = 7$ but only one microscopic state corresponding to each of $R = 2$ or $R = 12$. Thus, we know more about the microscopic state of the system if $R = 2$ or $R = 12$ than if $R = 7$. We define the entropy S to be the *natural logarithm of the number of microscopic states corresponding to a given macroscopic state.* Thus, for the dice, the entropy would be the natural logarithm of the number of (A, B) pairs that correspond to a given R. The entropy for $R = 2$ or $R = 12$ would be zero since $S = \ln(1) = 0$, while the entropy for $R = 7$ would be $S = \ln(6)$ since there were six different ways of obtaining $R = 7$.

If the dice were to be thrown a statistically large number of times the most likely result for any throw is $R = 7$ this is the macroscopic state with the largest number of microscopic states. Since any of the possible microscopic states is an equally likely outcome, the most likely macroscopic state after a large number of dice throws is the macroscopic state with the highest entropy.

Now consider a situation more closely related to the concept of a distribution function. In order to do this we first pose the following simple problem: suppose we have a pegboard with $\mathcal{N}$ holes, labeled h_1, h_2, ..., $h_{\mathcal{N}}$ and we also have $\mathcal{N}$ pegs labeled by p_1, p_2, ..., $p_{\mathcal{N}}$. What is the number of ways of putting all the pegs in all the holes? Starting with hole h_1, we have a choice of $\mathcal{N}$ different pegs, but when we get to hole h_2 there are now only $\mathcal{N} - 1$ pegs remaining so that there are now only $\mathcal{N} - 1$ choices. Using this argument for subsequent holes, we see there are $\mathcal{N}!$ ways of putting all the pegs in all the holes.

Let us complicate things further. Suppose that we arrange the holes in $\mathcal{M}$ groups. Say group G_1 has the first 10 holes, group G_2 has the next 19 holes, group G_3 has the next 4 holes and so on, up to group $\mathcal{M}$. We will use f to denote the number holes in a group, thus $f(1) = 10$, $f(2) = 19$, $f(3) = 4$, etc. The number of ways arranging pegs within a group is just the factorial of the number of pegs in the group, e.g., the number of ways of arranging the pegs within group 1 is just 10! and so in general the number of ways of arranging the pegs in the jth group is $[f(j)]!$.

Let us denote C as the number of ways of putting all the pegs in all the groups *without* caring about the internal arrangement within groups. The number of ways of putting the pegs in all the groups caring about the internal arrangements in all the groups is $C \times f(1)! \times f(2) \times f(3)! \times ...f(4)!$, but this is just the number of ways of putting all the pegs in all the holes, i.e.,

$$C \times f(1)! \times f(2) \times f(3)! \times ...f(4)! = \mathcal{N}!$$

or

$$C = \frac{\mathcal{N}!}{C \times f(1)! \times f(2) \times f(3)! \times ...f(4)!}$$

Now C is just the number of microscopic states corresponding to the macroscopic state of the prescribed grouping $f(1) = 10$, $f(2) = 19$, $f(3) = 4$, etc. so the entropy is just $S = \ln C$ or

$$\begin{aligned} S &= \ln\left(\frac{\mathcal{N}!}{C \times f(1)! \times f(2) \times f(3)! \times ...f(4)!}\right) \\ &= \ln \mathcal{N}! - \ln f(1)! - \ln f(2)! - ... - \ln f(\mathcal{M})! \end{aligned}$$

Stirling's formula shows that the large-argument asymptotic limit of the factorial function is

$$\lim_{k\to\infty} \ln k! = k \ln k - k$$

Noting that $f(1) + f(2) + ... + f(\mathcal{M}) = \mathcal{N}$, the entropy becomes

$$\begin{aligned} S &= \mathcal{N} \ln \mathcal{N} - f(1) \ln f(1) - f(2) \ln f(2) - ... - f(\mathcal{M}) \ln f(\mathcal{M}) \\ &= \mathcal{N} \ln \mathcal{N} - \sum_{j=1}^{\mathcal{M}} \ln f(j) \end{aligned}$$

The constant $\mathcal{N} \ln \mathcal{N}$ is often dropped, giving

$$S = -\sum_{j=1}^{\mathcal{M}} f(j) \ln f(j)$$

If j is made into a continuous variable, say $j \to v$ so that $f(v)\mathrm{d}v$ is the number of items in the group labeled by v then the entropy can be written as

$$S = -\int \mathrm{d}v f(v) \ln f(v)$$

By now, it is obvious that f could be the velocity distribution function, in which case $f(v)\mathrm{d}v$ is just the number of particles in the group having velocity between v and $v + \mathrm{d}v$. Since the peg groups correspond to different velocity range coordinates, having more dimensions just means having more groups and so for three dimensions the entropy generalizes to

$$S = -\int \mathrm{d}\mathbf{v} f(\mathbf{v}) \ln f(\mathbf{v})$$

If the distribution function depends on position as well, this corresponds to still more peg groups, and so a distribution function that depends on both velocity and position will have the entropy

$$S = -\int \mathrm{d}\mathbf{x} \int \mathrm{d}\mathbf{v} f(\mathbf{x}, \mathbf{v}) \ln f(\mathbf{x}, \mathbf{v}) \tag{8.8}$$

8.2.5 Effect of collisions on entropy

The highest entropy state is the most likely state of the system because the highest entropy state has the highest number of microscopic states corresponding to the macroscopic state. Collisions (or other forms of randomization) will take some initial prescribed microscopic state and scramble the phase-space positions of the particles, thereby transforming the system to a different microscopic state. This new state could in principle be any microscopic state, but is most likely to be a member of the class of microscopic states belonging to the highest entropy macroscopic state. Thus, any randomization process such as collisions will cause the system to evolve towards the macroscopic state having the maximum entropy.

An important shortcoming of this argument is that it neglects any conservation relations that have to be satisfied. To see this, note that the expression for entropy could be maximized if all the particles are put in one group, in which case $C = \mathcal{N}!$ which is the largest possible value for C. Thus, the maximum entropy configuration of $\mathcal{N}$ plasma particles corresponds to all the particles having the same velocity. However, this would assign a specific energy to the system, which would in general differ from the energy of the initial microstate. This maximum entropy state is therefore not accessible in isolated systems, because energy would not be conserved if the system changed from its initial microstate to the maximum entropy state.

Thus, a qualification must be added to the argument. Randomizing processes will scramble the system to attain the state of maximum entropy consistent with any constraints placed on the system. Examples of such constraints would be the requirements that the total system energy and the total number of particles must be conserved. We therefore reformulate the problem as: given an isolated system with $\mathcal{N}$ particles in a fixed volume V and initial average energy per particle $\langle E \rangle$, what is the maximum entropy state consistent with conservation of energy and conservation of number of particles? This is a variational problem because the goal is to maximize S subject to the constraint that both $\mathcal{N}$ and $\mathcal{N}\langle E \rangle$ are fixed. The method of *Lagrange multipliers* can then be used to take into account these constraints. Using this method the variational problem becomes

$$\delta S - \lambda_1 \delta \mathcal{N} - \lambda_2 \delta(\mathcal{N}\langle E \rangle) = 0$$

where λ_1 and λ_2 are as-yet undetermined Lagrange multipliers. The number of particles is

$$\mathcal{N} = V \int f \mathrm{d}v$$

The energy of an individual particle is $E = mv^2/2$, where v is the velocity measured in the rest frame of the center of mass of the entire collection of $\mathcal{N}$ particles. Thus, the total kinetic energy of all the particles in this rest frame is

$$\mathcal{N}\langle E \rangle = V \int \frac{mv^2}{2} f(v) \mathrm{d}v$$

and so the variational problem becomes

$$\delta \int \mathrm{d}v \left(f \ln f - \lambda_1 V f - \lambda_2 V \frac{mv^2}{2} f \right) = 0$$

Incorporating the volume V into the Lagrange multipliers, and factoring out the coefficient δf, this becomes

$$\int \mathrm{d}v \delta f \left(1 + \ln f - \lambda_1 - \lambda_2 \frac{mv^2}{2} \right) = 0$$

Since δf is arbitrary, the integrand must vanish, giving

$$\ln f = \lambda_2 \frac{mv^2}{2} - \lambda_1$$

where the "1" has been incorporated into λ_1.

The maximum entropy distribution function of an isolated, energy and particle conserving system is therefore

$$f = \lambda_1 \exp\left(-\lambda_2 mv^2/2\right)$$

which is the Maxwellian distribution function. We will often assume that the plasma is locally Maxwellian so that $\lambda_1 = \lambda_1(\mathbf{x}, t), \lambda_2 = \lambda_2(\mathbf{x}, t)$. We define the temperature to be

$$k_B T_s(\mathbf{x}, t) = \frac{1}{\lambda_2(\mathbf{x}, t)}$$

The normalization factor is set to be

$$\lambda_1(\mathbf{x}, t) = n_s(\mathbf{x}, t) \left(\frac{m_s}{2\pi k_B T_s(\mathbf{x}, t)} \right)^{N/2}$$

where N is the dimensionality (1, 2 or 3) so that $\int f_s(\mathbf{x}, \mathbf{v}, t) \mathrm{d}^N \mathbf{v} = n_s(\mathbf{x}, t)$. Because the kinetic energy of individual particles was defined in terms of velocities measured in the rest frame of the center of mass of the complete system of particles, if this center of mass is moving in the lab frame with a velocity $\mathbf{u}_s$, then in the lab frame the Maxwellian will have the form

$$f_s(\mathbf{x}, \mathbf{v}, t) = n_s \left(\frac{m_s}{2\pi k_B T_s} \right)^{N/2} \exp\left(-\frac{m_s (\mathbf{v} - \mathbf{u}_s)^2}{2 k_B T_s} \right) \tag{8.9}$$

Equation 8.9 is equivalent to Equation 8.5 times number density in 3D.

8.3 Equations of Kinetic Theory

8.3.1 Vlasov equation

Now consider the rate of change of the number of particles inside a small box in phase-space, such as is shown in Figure 8.4. Defining $a(x, v, t)$ to be the acceleration of a particle, it is seen that the particle flux in the horizontal direction is fv and the particle flux in the vertical direction is fa. Thus, the particle fluxes into the four sides of the box are:

1. flux into left side of box is $f(x, v, t) v \mathrm{d}v$
2. flux into right side of box is $-f(x + \mathrm{d}x, v, t) v \mathrm{d}v$
3. flux into bottom of box is $f(x, v, t) a(x, v, t) \mathrm{d}x$

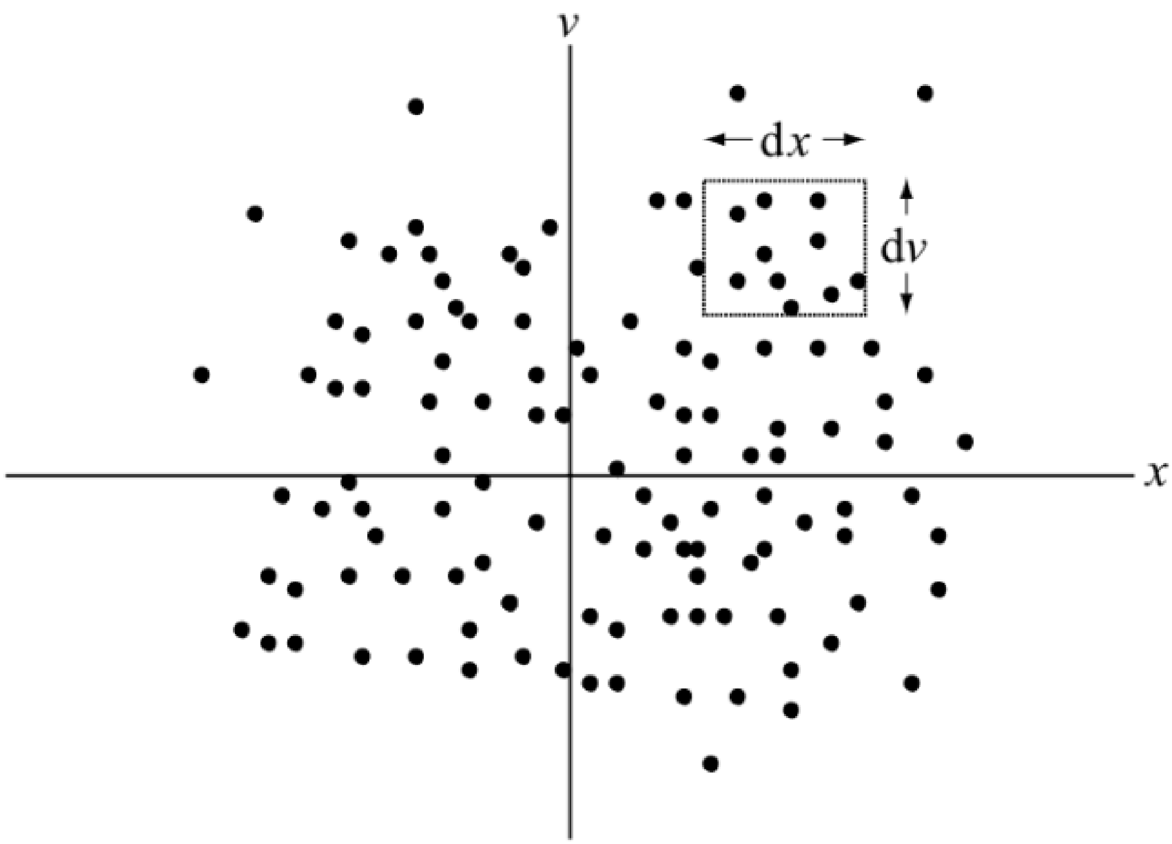

Figure 8.4: A box within phase-space having width dx and height dv.

4. flux into top of box is $-f(x, v+dv, t)a(x, v+dv, t)dx$

The number of particles in the box is $f(x, v, t)\mathrm{d}x\mathrm{d}v$ so that the rate of change of the number of particles in the box is

$$\begin{aligned}\frac{\partial f(x,v,t)}{\partial t}\mathrm{d}x\mathrm{d}v = &-f(x+\mathrm{d}x, v, t)v\mathrm{d}v + f(x,v,t)v\mathrm{d}v \\ &- f(x, v+\mathrm{d}v, t)a(x, v+\mathrm{d}v, t)\mathrm{d}x \\ &+ f(x,v,t)a(x,v,t)\mathrm{d}x\end{aligned}$$

or, on Taylor expanding the quantities on the right-hand side, we obtain the one-dimensional *Vlasov equation*,

$$\frac{\partial f}{\partial t} + v\frac{\partial f}{\partial x} + \frac{\partial}{\partial v}(af) = 0 \tag{8.10}$$

It is straightforward to generalize Equation 8.10 to three dimensions and so obtain the three-dimensional Vlasov equation,

$$\frac{\partial f}{\partial t} + \mathbf{v}\cdot\nabla f + \frac{\partial}{\partial \mathbf{v}}\cdot(\mathbf{a}f) = 0 \tag{8.11}$$

The symbol ∇ stands, as usual, for the gradient in (x, y, z) space. The symbol $\partial/\partial\mathbf{v}$ or $\nabla_{\mathbf{v}}$ stands for the gradient in velocity space:

$$\nabla_{\mathbf{v}} = \frac{\partial}{\partial \mathbf{v}} = \hat{x}\frac{\partial}{\partial v_x} + \hat{y}\frac{\partial}{\partial v_y} + \hat{z}\frac{\partial}{\partial v_z}$$

Because $\mathbf{x}, \mathbf{v}$ are independent quantities in phase-space, the spatial derivative term has the commutation property,

$$\mathbf{v}\cdot\frac{\partial f}{\partial \mathbf{x}} = \frac{\partial}{\partial \mathbf{x}}\cdot(\mathbf{v}f)$$

The particle acceleration is given by the Lorentz force

$$\mathbf{a} = \frac{q}{m}(\mathbf{E} + \mathbf{v}\times\mathbf{B})$$

Because $(\mathbf{v}\times\mathbf{B})_i = v_jB_k - v_kB_j$ is independent of v_i, the term $\partial(\mathbf{v}\times\mathbf{B})_i/\partial v_i$ vanishes so that even though the Lorentz acceleration $\mathbf{a}$ is velocity-dependent, it nevertheless commutes with the vector velocity derivative as

$$\mathbf{a}\cdot\frac{\partial f}{\partial \mathbf{v}} = \frac{\partial}{\partial \mathbf{v}}\cdot(\mathbf{a}f)$$

Because of this commutation property the Vlasov equation can also be written as

$$\frac{\partial f}{\partial t} + \mathbf{v}\cdot\frac{\partial f}{\partial \mathbf{x}} + \mathbf{a}\cdot\frac{\partial f}{\partial \mathbf{v}} = 0 \tag{8.12}$$

If we “sit on top of” a particle that has a phase-space trajectory $\mathbf{x} = \mathbf{x}(t)$, $\mathbf{v} = \mathbf{v}(t)$ and measure the distribution function as we move along with the particle, the observed rate of change of the distribution function will be $\mathrm{d}f(\mathbf{x}(t), \mathbf{v}(t), t)/\mathrm{d}t$, where the $\mathrm{d}/\mathrm{d}t$ means that the derivative is measured in the moving frame. Because $\mathrm{d}\mathbf{x}/\mathrm{d}t = \mathbf{v}$ and $\mathrm{d}\mathbf{v}/\mathrm{d}t = \mathbf{a}$, this observed rate of change is

$$\left(\frac{\mathrm{d}f(\mathbf{x}(t), \mathbf{v}(t), t)}{\mathrm{d}t}\right)_{\mathrm{orbit}} = \frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{x}} + \mathbf{a} \cdot \frac{\partial f}{\partial \mathbf{v}} = 0$$

Thus, the distribution function f as measured when moving along a particle trajectory (orbit) is constant. This gives a powerful method for finding solutions to the Vlasov equation. Since f is a constant when measured in a frame following an orbit, we can choose f to depend on any quantity that is constant along the orbit (Jeans 1915, Watson 1956).

For example, if the energy E of particles is constant along their orbits then $f = f(E)$ is a solution to the Vlasov equation. On the other hand, if both the energy and the momentum $\mathbf{p}$ are constant along particle orbits, then any distribution function with the functional dependence $f = f(E, \mathbf{p})$ is a solution to the Vlasov equation. Depending on the situation at hand, the energy and/or momentum may or may not be constant along an orbit and so whether or not $f = f(E, \mathbf{p})$ is a solution to the Vlasov equation depends on the specific problem under consideration. However, there always exists at least one constant of the motion for any trajectory because, just like every human being has an invariant birthday, the initial conditions of a particle trajectory are invariant along this trajectory. As a simple example, consider a situation where there is no electromagnetic field so that $\mathbf{a} = 0$, in which case the particle trajectories are simply $\mathbf{x}(t) = \mathbf{x}_0 + \mathbf{v}_0(t)$, $\mathbf{v}(t) = \mathbf{v}_0$, where $\mathbf{x}_0, \mathbf{v}_0$ are the initial position and velocity. Let us check to see whether $f(\mathbf{x}_0)$ is a solution to the Vlasov equation. By writing $\mathbf{x}_0 = \mathbf{x}(t) - \mathbf{v}_0 t$ so $f(\mathbf{x}_0) = f(\mathbf{x}(t) - \mathbf{v}_0 t)$ we observe that indeed $f = f(\mathbf{x}_0)$ is a solution, since

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{x}} + \mathbf{a} \cdot \frac{\partial f}{\partial \mathbf{v}} = \mathbf{v}_0 \cdot \frac{\partial f}{\partial (\mathbf{x} - \mathbf{x}_0)} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{x}} = \mathbf{v}_0 \cdot \frac{\partial f}{\partial \mathbf{x}} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{x}} = 0$$

8.3.2 Boltzmann equation

It was shown in ??? that the cumulative effect of grazing collisions dominates the cumulative effect of the more infrequently occurring large-angle collisions. In order to see how collisions affect the Vlasov equation, let us now temporarily imagine that the grazing collisions are replaced by an equivalent sequence of abrupt large scattering angle encounters as shown in Figure 8.5. Two particles involved in a collision do not significantly change their positions during the course of a collision, but they do substantially change their velocities. For example, a particle making a head-on collision with an equal mass stationary particle will stop after the collision, while the target particle will assume the velocity of the incident particle. If we draw the detailed phase-space trajectories characterized by a collision between two particles we see that each particle has a sudden change in its vertical coordinate (i.e., velocity) but no change in its horizontal coordinate (i.e., position). The collision-induced velocity jump occurs very fast so that if the phase-space trajectories were recorded with a "movie camera" having insufficient framing rate to catch the details of the jump, the resulting movie would show particles being spontaneously created or annihilated within given volumes of phase-space (e.g., within the boxes shown in Figure 8.5).

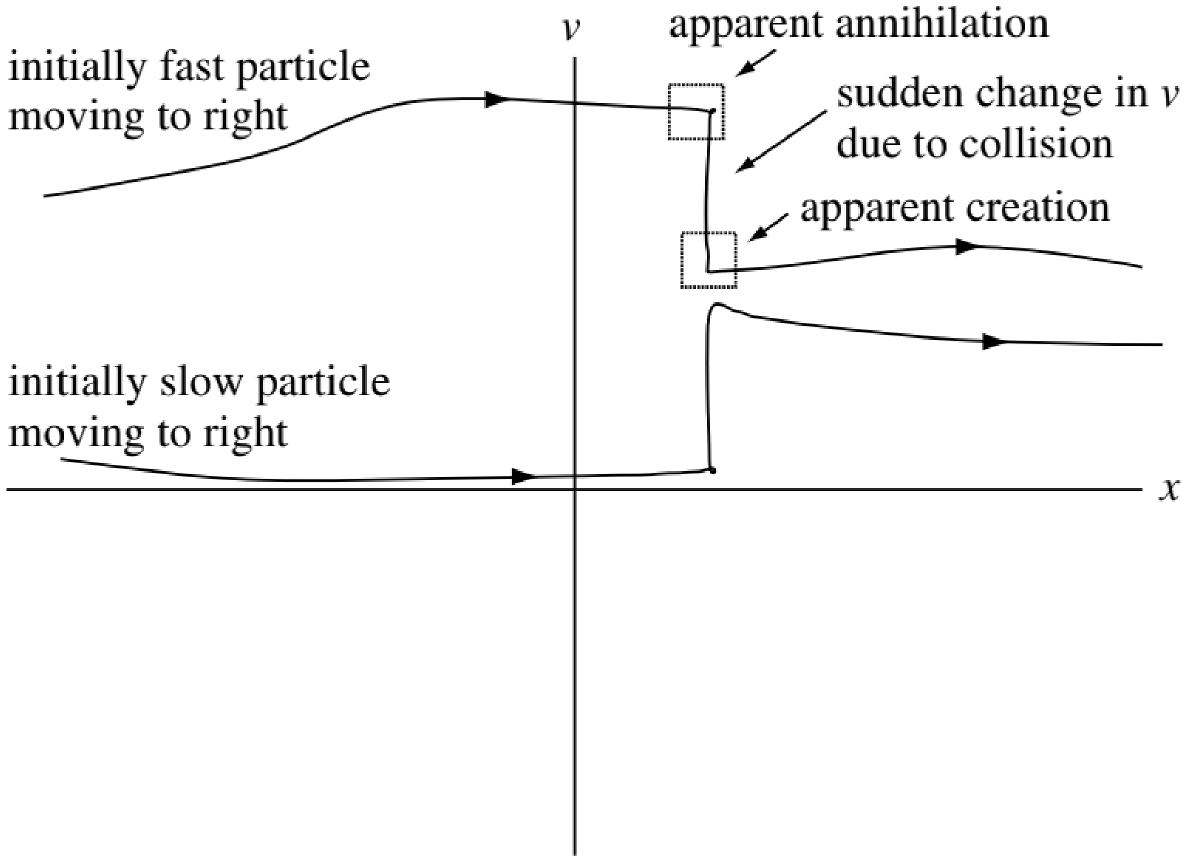

Figure 8.5: Detailed view of collisions causing "jumps" in phase-space.

The details of these individual jumps in phase-space are complicated and yet of little interest since all we really want to know is the cumulative effect of many collisions. It is therefore both efficient and sufficient to follow the trajectories on the slow time scale while accounting for the apparent "creation" or "annihilation" of particles by inserting a collision operator on the right-hand side of the Vlasov equation. In the example shown here it is seen that when a particle is apparently "created" in one box, another particle must be simultaneously "annihilated" in another box at the same x coordinate but a different v coordinate (of course, what is actually happening is that a single particle is suddenly moving from one box to the other). This coupling of the annihilation and creation rates in different boxes constrains the form of the collision operator. We will not attempt to derive collision operators in this chapter but will simply discuss the constraints on these operators. From a more formal point of view, collisions are characterized by constrained sources and sinks for particles in phase-space and inclusion of collisions in the Vlasov equation causes the Vlasov equation to assume the form

$$\frac{\partial f_s}{\partial t} + \frac{\partial}{\partial \mathbf{x}} \cdot (\mathbf{v} f_s) + \frac{\partial}{\partial \mathbf{v}} \cdot (\mathbf{a} f_s) = \sum_\alpha C_{s\alpha}(f_s) \tag{8.13}$$

where $C_{s\alpha}(f_s)$ is the rate of change of f_s due to collisions of species s with species α[1]. This is called the *Boltzmann equation*.

The constraints that must be satisfied by the collision operator $C_{s\alpha}(f_s)$ are as follows:

- Conservation of particles - Collisions cannot change the total number of particles at a particular location so
$$\int \mathrm{d}\mathbf{v} C_{s\alpha}(f_s) = 0 \tag{8.14}$$

- Conservation of momentum - Collisions between particles of the same species cannot change the total momentum of that species so
$$\int \mathrm{d}\mathbf{v} m_s \mathbf{v} C_{ss}(f_s) = 0 \tag{8.15}$$
while collisions between different species must conserve the total momentum of both species together so
$$\int \mathrm{d}\mathbf{v} m_i \mathbf{v} C_{ie}(f_i) + \int \mathrm{d}\mathbf{v} m_e \mathbf{v} C_{ei}(f_e) = 0 \tag{8.16}$$

- Conservation of energy - Collisions between particles of the same species cannot change the total energy of that species so
$$\int \mathrm{d}\mathbf{v} m_s \mathbf{v}^2 C_{ss}(f_s) = 0 \tag{8.17}$$
while collisions between different species must conserve the total energy of both species together so
$$\int \mathrm{d}\mathbf{v} m_i \mathbf{v}^2 C_{ie}(f_i) + \int \mathrm{d}\mathbf{v} m_e \mathbf{v}^2 C_{ei}(f_e) = 0 \tag{8.18}$$

In a sufficiently hot plasma, collisions can be neglected. If, furthermore, the force $\mathbf{F}$ is entirely electromagnetic, Equation 8.13 takes the special form of Equation 8.12. Because of its comparative simplicity, this is the equation most commonly studied in kinetic theory. When there are collisions with neutral atoms, the collision term in Equation 8.13 can be approximated by

$$\left(\frac{\partial f}{\partial t}\right)_c = \frac{f_n - f}{\tau}$$

[1]For simulations, the effect of numerical diffusion may be treated as one type of "collision".

where f_n is the distribution function of the neutral atoms, and τ is a constant collision time. This is called the *Krook collision term*. It is the kinetic generalization of the collision term in Eq. (5.5) in (F. F. Chen 2016). When there are Coulomb collisions, Equation 8.13 can be approximated by

$$\frac{\mathrm{d}f}{\mathrm{d}t} = -\frac{\partial}{\partial \mathbf{v}} \cdot (f \langle \Delta \mathbf{v} \rangle) \frac{1}{2} \frac{\partial^2}{\partial \mathbf{v} \partial \mathbf{v}} : (f \langle \Delta \mathbf{v} \Delta \mathbf{v} \rangle) \tag{8.19}$$

This is called the *Fokker-Planck equation*; it takes into account binary Coulomb collisions only. Here, $\Delta \mathbf{v}$ is the change of velocity in a collision, and Equation 8.19 is a shorthand way of writing a rather complicated expression. The colon operator $\mathbf{ab} : \mathbf{cd} = a_i b_j c_i d_j$.

The fact that $\mathrm{d}f/\mathrm{d}t$ is constant in the absence of collisions means that particles follow the contours of constant f as they move around in phase space. As an example of how these contours can be used, consider the beam-plasma instability of Section ???. In the unperturbed plasma, the electrons all have velocity v_0, and the contour of constant f is a straight line. The function $f(x, v_x)$ is a wall rising out of the plane of the paper at $v_x = v_0$. The electrons move along the trajectory shown. When a wave develops, the electric field $\mathbf{E}_1$ causes electrons to suffer changes in v_x as they stream along. The trajectory then develops a sinusoidal ripple (**?@fig-beam_e_dist_1d** B). This ripple travels at the phase velocity, not the particle velocity. Particles stay on the curve as they move relative to the wave. If $\mathbf{E}_1$ becomes very large as the wave grows, and if there are a few collisions, some electrons will be trapped in the electrostatic potential of the wave. In coordinate space, the wave potential appears as in **?@fig-wave_potential**. In phase space, $f(x, v_x)$ will have peaks wherever there is a potential trough (**?@fig-contour_dist**). Since the contours of f are also electron trajectories, one sees that some electrons move in closed orbits in phase space; these are just the trapped electrons.

Electron trapping is a nonlinear phenomenon which cannot be treated by straightforward solution of the Vlasov equation. However, electron trajectories can be followed on a computer, and the results are often presented in the form of a plot like **?@fig-contour_dist**.

ADD A TWO STREAM INSTABILITY PHASE ANIMATION!

8.4 Derivation of the Fluid Equations

Instead of just taking moments of the distribution function f itself, moments will now be taken of the entire Vlasov equation to obtain a set of partial differential equations relating the mean quantities $n(\mathbf{x}), \mathbf{u}(\mathbf{x})$, etc. We begin by integrating the Vlasov equation, Equation 8.12, over velocity for each species. This first and simplest step in the procedure is called taking the "zeroth" moment, since the operation of multiplying by unity can be considered as multiplying the entire Vlasov equation by $\mathbf{v}$ raised to the power zero. Multiplying the Vlasov equation by unity and then integrating over velocity gives

$$\int \left[\frac{\partial f_s}{\partial t} + \frac{\partial}{\partial \mathbf{x}} \cdot (\mathbf{v} f_s) + \frac{\partial}{\partial \mathbf{v}} \cdot (\mathbf{a} f_s) \right] \mathrm{d}\mathbf{v} = \sum_\alpha \int C_{s\alpha}(f_s) \mathrm{d}\mathbf{v}$$

The velocity integral commutes with both the time and space derivatives on the left-hand side because $\mathbf{x}, \mathbf{v}$ and t are independent variables, while the third term on the left-hand side is the volume integral of a divergence in velocity space. Gauss' theorem (i.e., $\int_V \mathrm{d}\mathbf{x}\nabla\cdot\mathbf{Q} = \int_A \mathrm{d}\mathbf{s}\cdot\mathbf{Q}$) gives f_s evaluated on a surface at $v = \infty$. However, because $f_s \to 0$ as $v \to \infty$, this surface integral in velocity space vanishes. Inserting Equation 8.1, Equation 8.4, and Equation 8.14 into the above, we have the species continuity equation

$$\frac{\partial n_s}{\partial t} + \nabla\cdot(n_s\mathbf{u}_s) = 0 \tag{8.20}$$

Now let us multiply Equation 8.12 by $\mathbf{v}$ and integrate over velocity to take the "first moment",

$$\int \mathbf{v}\left[\frac{\partial f_s}{\partial t} + \frac{\partial}{\partial \mathbf{x}}\cdot(\mathbf{v}f_s) + \frac{\partial}{\partial \mathbf{v}}\cdot(\mathbf{a}f_s)\right]\mathrm{d}\mathbf{v} = \sum_\alpha \int \mathbf{v}C_{s\alpha}(f_s)\mathrm{d}\mathbf{v}$$

This may be rearranged in a more tractable form by:

1. pulling both the time and space derivatives out of the velocity integral,
2. writing $\mathbf{v} = \mathbf{v}'(\mathbf{x}, t) + \mathbf{u}(\mathbf{x}, t)$, where $\mathbf{v}'(\mathbf{x}, t)$ is the *random* part of a given velocity, i.e., that part of the velocity that differes from the mean (note that $\mathbf{v}$ is independent of both $\mathbf{x}$ and t but $\mathbf{v}'$ is not; also $\mathrm{d}\mathbf{v} = \mathrm{d}\mathbf{v}'$ since $\mathbf{u}$ is independent of $\mathbf{v}$),
3. integrating by parts in 3-D velocity space on the acceleration term and using

$$\partial_i \mathbf{v}_j = \delta_{ij}$$

After performing these manipulations, the first moment of the Vlasov equation becomes

$$\begin{aligned}\frac{\partial(n_s\mathbf{u}_s)}{\partial t} + \frac{\partial}{\partial \mathbf{x}}\cdot\int\left(\mathbf{v}'\mathbf{v}' + \mathbf{v}'\mathbf{u}_s + \mathbf{u}_s\mathbf{v}' + \mathbf{u}_s\mathbf{u}_s\right)f_s\mathrm{d}\mathbf{v}'\\ -\frac{q_s}{m_s}\int\left(\mathbf{E} + \mathbf{v}\times\mathbf{B}\right)f_s\mathrm{d}\mathbf{v}' = -\frac{1}{m_s}\mathbf{R}_{s\alpha}\end{aligned} \tag{8.21}$$

where $\mathbf{R}_{s\alpha}$ is the net frictional drag force due to collisions of species s with species α. Note that $\mathbf{R}_{ss} = 0$, since a species cannot exert a net drag force on itself. The frictional terms have the form

$$\mathbf{R}_{ei} = \nu_{ei}m_e n_e(\mathbf{u}_e - \mathbf{u}_i)$$
$$\mathbf{R}_{ie} = \nu_{ie}m_i n_i(\mathbf{u}_i - \mathbf{u}_e)$$

so that in the ion frame the drag on electrons is simply the total electron momentum $m_e n_e \mathbf{u}_e$ measured in this frame multiplied by the rate ν_{ei} at which this momentum is destroyed by collisions with ions. This form for frictional drag force has the following properties: (i) $\mathbf{R}_{ei} + \mathbf{R}_{ie} = 0$, showing that the plasma cannot exert a frictional drag force on itself, (ii) friction causes the faster species to be slowed down by the slower species, and (iii) there is no friction between species if both have the same mean velocity.

Equation 8.21 can be further simplified by factoring $\mathbf{u}$ out of the velocity integrals and recalling that by definition $\int \mathbf{v}' f_s \mathrm{d}\mathbf{v}' = 0$. Thus Equation 8.21 reduces to

$$m_s \left[\frac{\partial (n_s \mathbf{u}_s)}{\partial t} + \frac{\partial}{\partial \mathbf{x}} \cdot (n_s \mathbf{u}_s \mathbf{u}_s) \right] = n_s q_s (\mathbf{E} + \mathbf{u}_s \times \mathbf{B}) - \frac{\partial}{\partial \mathbf{x}} \cdot \overleftrightarrow{P}_s - \mathbf{R}_{s\alpha}$$

where the *pressure tensor* is defined by

$$\overleftrightarrow{P}_s \equiv m_s \int \mathbf{v}' \mathbf{v}' f_s \mathrm{d}\mathbf{v}' \tag{8.22}$$

If f_s is an isotropic function of $\mathbf{v}'$, then the off-diagonal terms in $\overleftrightarrow{P}_s$ vanish and the three diagonal terms are identical. In this case, it is useful to define the diagonal terms to be the *scalar pressure* p_s, i.e.,

$$\begin{aligned} p_s &= m_s \int v'_x v'_x f_s \mathrm{d}\mathbf{v}' = m_s \int v'_y v'_y f_s \mathrm{d}\mathbf{v}' = m_s \int v'_z v'_z f_s \mathrm{d}\mathbf{v}' \\ &= \frac{m_s}{3} \int \mathbf{v}' \cdot \mathbf{v}' f_s \mathrm{d}\mathbf{v}' \end{aligned} \tag{8.23}$$

Equation 8.23 defines pressure for a three-dimensional isotropic system. However, we will often deal with systems of reduced dimensionality, i.e., systems with just one or two dimensions. Equation 8.23 can therefore be generalized to these other cases by introducing the general N-dimensional definition for scalar pressure

$$p_s = \frac{m_s}{N} \int \mathbf{v}' \cdot \mathbf{v}' f_s \mathrm{d}^N \mathbf{v}' \tag{8.24}$$

where $\mathbf{v}'$ is the N-dimensional random velocity.

The scalar pressure has a simple relation to the generalized Maxwellian as seen by recasting Equation 8.24 as

$$\begin{aligned} p_s &= \frac{m_s}{N} \int \mathbf{v}' \cdot \mathbf{v}' f_s \mathrm{d}^N \mathbf{v}' \\ &= \frac{n_s m_s}{N} \left(\frac{m_s}{2\pi k_B T_s} \right)^{N/2} \int \mathbf{v}'^2 \exp\left(-\frac{m_s \mathbf{v}'^2}{2k_B T_s} \right) \mathrm{d}^N \mathbf{v}' \\ &= -\frac{n_s m_s}{N} \left(\frac{\alpha}{\pi} \right)^{N/2} \frac{\mathrm{d}}{\mathrm{d}\alpha} \int e^{-\alpha v'^2} \mathrm{d}^N \mathbf{v}', \; \alpha \equiv m_s / 2k_B T_s \\ &= -\frac{n_s m_s}{N} \left(\frac{\alpha}{\pi} \right)^{N/2} \frac{\mathrm{d}}{\mathrm{d}\alpha} \left(\frac{\alpha}{\pi} \right)^{-N/2} \\ &= n_s k_B T_s \end{aligned}$$

which is just the ideal gas law. Thus, the definitions that have been proposed for pressure and temperature are consistent with everyday notions for these quantities.

It is important to emphasize that assuming isotropy is done largely for mathematical convenience and that in real systems the distribution function is often quite anisotropic. Collisions, being randomizing, drive the distribution function towards isotropy, while competing processes simultaneously drive it towards anisotropy. Thus, each situation must be considered individually in order to determine whether there is sufficient collisionality to make f isotropic. Because fully ionized hot plasmas often have insufficient collisions to make f isotropic, the often-used assumption of isotropy is an oversimplification, which may or may not be acceptable depending on the phenomenon under consideration.

On expanding the derivatives on the left-hand side of Equation 8.21, it is seen that two of the terms combine to give $\mathbf{u}$ times Equation 8.20. After removing this embedded continuity equation, E@eq-vlasov-1st-moment reduces to

$$n_s m_s \frac{\mathrm{d}\mathbf{u}_s}{\mathrm{d}t} = n_s q_s(\mathbf{E} + \mathbf{u}_s \times \mathbf{B}) - \nabla p_s - \mathbf{R}_{s\alpha}$$

where the operator $\mathrm{d}/\mathrm{d}t$ is the convective derivative (Equation 3.2).

At this point in the procedure it becomes evident that a certain pattern recurs for each successive moment of the Vlasov equation. When we took the zeroth moment, an equation for the density $\int f \mathrm{d}\mathbf{v}$ resulted, but this also introduced a term involving the next higher moment, namely the mean velocity $\sim \int \mathbf{v} f \mathrm{d}\mathbf{v}$. Then, when we took the first moment to get an equation for the velocity, an equation was obtained containing a term involving the next higher moment, namely the pressure $\sim \int \mathbf{vv} f \mathrm{d}\mathbf{v}$. Thus, each time we take a moment of the Vlasov equation, an equation for the moment we want is obtained, but because of the $\mathbf{v} \cdot \nabla f$ term in the Vlasov equation, a next higher moment also appears. Thus, moment-taking never leads to a closed system of equations; there will always be a "loose end", a highest moment for which there is no determining equation. Some sort of ad hoc closure procedure must always be invoked to terminate this chain (as seen below, typical closures involve invoking adiabatic or isothermal assumptions). Another feature of taking moments is that each higher moment embeds terms that contain complete lower moment equations multiplied by some factor. Algebraic manipulation can identify these lower moment equations and eliminate them to give a simplified higher moment equation.

Let us now take the second moment of the Vlasov equation. Unlike the zeroth and first moments, the dimensionality of the system now enters explicitly so the more general pressure definition given by Equation 8.24 will be used. Multiplying the Vlasov equation by $m_s v^2/2$ and integrating over velocity gives

$$\left\{ \begin{array}{c} \frac{\partial}{\partial t} \int \frac{m_s v^2}{2} f_s \mathrm{d}^N\mathbf{v} \\ +\frac{\partial}{\partial \mathbf{x}} \cdot \int \frac{m_s v^2}{2} \mathbf{v} f_s \mathrm{d}^N\mathbf{v} \\ +q_s \int \frac{v^2}{2} \frac{\partial}{\partial \mathbf{v}} \cdot (\mathbf{E} + \mathbf{v} \times \mathbf{B}) f_s \mathrm{d}^N\mathbf{v} \end{array} \right\} = \sum_\alpha \int m_s \frac{v^2}{2} C_{s\alpha} f_s \mathrm{d}^N\mathbf{v} \tag{8.25}$$

We consider each term of this equation separately as follows:

1. The time derivative term becomes

$$\frac{\partial}{\partial t}\int \frac{m_s v^2}{2} f_s \mathrm{d}^N\mathbf{v} = \frac{\partial}{\partial t}\int \frac{m_s(\mathbf{v}' + \mathbf{u}_s)^2}{2} f_s \mathrm{d}^N\mathbf{v}' = \frac{\partial}{\partial t}\left(\frac{Np_s}{2} + \frac{m_s n_s u_s^2}{2}\right)$$

2. Again using $\mathbf{v} = \mathbf{v}' + \mathbf{u}_s$, the space derivative term becomes

$$\frac{\partial}{\partial \mathbf{x}} \cdot \int \frac{m_s v^2}{2} \mathbf{v} f_s \mathrm{d}^N\mathbf{v} = \nabla \cdot \left(\mathbf{Q}_s + \frac{2+N}{2} p_s \mathbf{u}_s + \frac{m_s n_s u_s^2}{2}\mathbf{u}_s\right)$$

   where

$$\mathbf{Q}_s = \int \frac{m_s {v'}^2}{2} \mathbf{v}' f_s \mathrm{d}^N\mathbf{v}$$

   is called the *heat flux*.

3. On integrating by parts, the acceleration term becomes

$$q_s \int \frac{v^2}{2}\frac{\partial}{\partial \mathbf{v}} \cdot (\mathbf{E} + \mathbf{v} \times \mathbf{B}) f_s \mathrm{d}^N\mathbf{v} = -q_s \int \mathbf{v} \cdot \mathbf{E} f_s \mathrm{d}\mathbf{v} = -q_s n_s \mathbf{u}_s \cdot \mathbf{E}$$

4. The collision term becomes (using Equation 8.17)

$$\sum_\alpha \int m_s \frac{v^2}{2} C_{s\alpha} f_s \mathrm{d}^N\mathbf{v} = \int_{s\neq\alpha} m_s \frac{v^2}{2} C_{s\alpha} f_s \mathrm{d}\mathbf{v} = -\left(\frac{\partial W}{\partial t}\right)_{\mathrm{Es}}$$

   where $(\partial W/\partial t)_{\mathrm{Es}}$ is the rate at which species s collisionally transfers energy to species α.

Combining the above four relations, Equation 8.25 becomes

$$\begin{aligned}\frac{\partial}{\partial t}\left(\frac{Np_s}{2} + \frac{m_s n_s u_s^2}{2}\right) + \nabla \cdot \left(\mathbf{Q}_s + \frac{2+N}{2} p_s \mathbf{u}_s + \frac{m_s n_s u_s^2}{2}\mathbf{u}_s\right) - q_s n_s \mathbf{u}_s \cdot \mathbf{E} \\ = -\left(\frac{\partial W}{\partial t}\right)_{\mathrm{E\,s}}\end{aligned} \tag{8.26}$$

This equation can be simplified by invoking two mathmatical identities, the first being

$$\frac{\partial}{\partial t}\left(\frac{m_s n_s u_s^2}{2}\right) + \nabla \cdot \left(\frac{m_s n_s u_s^2}{2}\mathbf{u}_s\right) = n_s\left(\frac{\partial}{\partial t} + \mathbf{u}_s \cdot \nabla\right)\frac{m_s u_s^2}{2} = n_s \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{m_s u_s^2}{2}\right)$$

The second identity is obtained by dotting the equation of motion with $\mathbf{u}_s$:

$$\begin{aligned} n_s m_s \left[\frac{\partial}{\partial t}\left(\frac{u_s^2}{2}\right) + \mathbf{u}_s \cdot \left(\nabla\left(\frac{u_s^2}{2}\right) - \mathbf{u}_s \times \nabla \times \mathbf{u}_s\right)\right] \\ = n_s q_s \mathbf{u}_s \cdot \mathbf{E} - \mathbf{u}_s \cdot \nabla p_s - \mathbf{R}_{s\alpha} \cdot \mathbf{u}_s\end{aligned}$$

or

$$n_s \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{m_s u_s^2}{2}\right) = n_s q_s \mathbf{u}_s \cdot \mathbf{E} - \mathbf{u}_s \cdot \nabla p_s - \mathbf{R}_{s\alpha} \cdot \mathbf{u}_s$$

Inserting these two into Equation 8.26 gives the energy evolution equation

$$\frac{N}{2}\frac{\mathrm{d}p_s}{\mathrm{d}t} + \frac{2+N}{2} p_s \nabla \cdot \mathbf{u}_s = -\nabla \cdot \mathbf{Q}_s + \mathbf{R}_{s\alpha} \cdot \mathbf{u}_s - \left(\frac{\partial W}{\partial t}\right)_{\mathrm{Es}} \tag{8.27}$$

The first term on the right-hand side represents the heat flux, the second term gives the frictional heating of species s due to frictional drag on species α, while the last term on the right-hand side gives the rate at which species s collisionally transfers energy to other species. Although Equation 8.27 is complicated, two important limiting situations become evident if we let t_{char} be the characteristic time scale for a given phenomenon and l_{char} be its characteristic length scale. A characteristic velocity $V_{\text{ph}} \sim l_{\text{char}}/t_{\text{char}}$ may then be defined for the phenomenon and so, replacing temporal derivatives by t_{char}^{-1} and spatial derivatives by l_{char}^{-1} in Equation 8.27, it is seen that the two limiting situations are:

1. *Isothermal limit* - The heat flux term dominates all other terms, in which case the temperature becomes spatially uniform. This occurs if (i) $v_{Ts} \gg V_{\text{ph}}$ since the ratio of the left-hand side terms to the heat flux term is $\sim V_{\text{ph}}/v_{Ts}$ and (ii) the collisional terms are small enough to be ignored.
2. *Adiabatic limit* - The heat flux terms and the collisional terms are small enough to be ignored compared to the left-hand side terms; this occurs when $V_{\text{ph}} \gg v_{Ts}$. Adiabatic is a Greek word meaning “impassable”, and is used here to denote that no heat is flowing, i.e., the volume under consideration is thermally isolated from the outside world.

Both of these limits make it possible to avoid solving for $\mathbf{Q}_s$, which involves the third moment, and so both the adiabatic and isothermal limit provide a closure to the moment equations.

The energy equation may be greatly simplified in the adiabatic limit by re-arranging the continuity equation to give

$$\nabla \cdot \mathbf{u}_s = -\frac{1}{n_s}\frac{\mathrm{d}n_s}{\mathrm{d}t}$$

and then substituting this expression into the left-hand side of Equation 8.27 to obtain

$$\frac{1}{p_s}\frac{\mathrm{d}p_s}{\mathrm{d}t} = \frac{\gamma}{n_s}\frac{\mathrm{d}n_s}{\mathrm{d}t} \tag{8.28}$$

where

$$\gamma = \frac{N+2}{N}$$

Equation 8.28 implies

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{p_s}{n_s^{\gamma}}\right) = 0$$

so p_s/n_s^γ is a constant in the frame of the moving plasma. This constitutes a derivation of adiabaticity based on geometry and statistical mechanics rather than on thermodynamic arguments.

The energy equation derivation can also be found in An introductory guide to fluid models with anisotropic temperatures.

As a short summary, the fluid equations we have been using are simply moments of the Boltzmann equation. One interesting observation from Equation 8.12 is that in order to recover the hydrodynamic equations, the acceleration term $\mathbf{a}\cdot\partial_\mathbf{v} f$ is not required: the pressure gradient that drives the flow arises from the thermal motion embedded in the advection term $\mathbf{v}\cdot\partial_\mathbf{x} f$.

We must be careful not to become overconfident regarding the descriptive power of the fluid point of view because weaknesses exist in this point of view. For example, as discussed above neither the adiabatic nor the isothermal approximation is appropriate when $V_{\text{ph}} \sim v_{Ts}$. The fluid description breaks down in this situation and the Vlasov description must be used in this situation. Furthermore, the distribution function is Maxwellian only if there are sufficient collisions or some other randomizing process. Because of the weak collisionality of a plasma, this is very often not the case. In particular, since the collision frequency scales as v^{-3} (Equation 6.20), fast particles take much longer to become Maxwellian than slow particles. It is not at all unusual for a real plasma to be in a state where the low-velocity particles have reached a Maxwellian distribution whereas the fast particles form a non-Maxwellian tail.

8.5 Plasma Oscillations and Landau Damping

As an elementary illustration of the use of the Vlasov equation, we shall derive the dispersion relation for electron plasma oscillations, which is originally treated from the fluid point of view. This derivation will require a knowledge of contour integration.

In zeroth order, we assume a uniform plasma with a distribution $f_0(\mathbf{v})$, and we let $\mathbf{B}_0 = \mathbf{E}_0 = 0$. In first order, we denote the perturbation in $f(\mathbf{r},\mathbf{v},t)$ by $f_1(\mathbf{r},\mathbf{v},t)$:

$$f(\mathbf{r},\mathbf{v},t) = f_0(\mathbf{v}) + f_1(\mathbf{r},\mathbf{v},t)$$

Since $\mathbf{v}$ is now an independent variable and is not to be linearized, the first-order Vlasov equation for electron is

$$\frac{\partial f_1}{\partial t} + \mathbf{v}\cdot\nabla f_1 - \frac{e}{m}\mathbf{E}_1\cdot\frac{\partial f_0}{\partial \mathbf{v}} = 0 \tag{8.29}$$

As before, we assume the ions are massive and fixed and that the waves are plane waves in the x direction $f_1 \propto e^{i(kx-\omega t)}$. Then the linearized Vlasov equation becomes

$$
\begin{aligned}
-i\omega f_1 + ikv_x fx &= \frac{e}{m}E_x\frac{\partial f_0}{\partial v_x} \\
f_1 &= \frac{ieE_x}{m}\frac{\partial f_0/\partial v_x}{\omega - kv_x}
\end{aligned}
$$

Poisson's equation gives

$$
\epsilon_0 \nabla \cdot \mathbf{E}_1 = ik\epsilon_0 E_x = -en_1 = -e\int\int\int f_1 d^3v
$$

Substituting for f_1 and dividing by $ik\epsilon_0 E_x$, we have

$$
1 = -\frac{e^2}{km\epsilon_0}\int\int\int \frac{\partial f_0/\partial v_x}{\omega - kv_x} d^3v
$$

A factor n_0 can be factored out if we replace f_0 by a normalized function $\hat{f}_0$:

$$
1 = -\frac{\omega_p^2}{k}\int_{-\infty}^{\infty} dv_z \int_{-\infty}^{\infty} dv_y \int_{-\infty}^{\infty} \frac{\partial \hat{f}_0(v_x, v_y, v_z)/\partial v_x}{\omega - kv_x} dv_x
$$

If f_0 is a Maxwellian or some other factorable distribution, the integration over v_y and v_z can be carried out easily. What remains is the one-dimensional distribution $\hat{f}_0(v_x)$. For instance, a one-dimensional Maxwellian distribution is

$$
\hat{f}_m(v_x) = \sqrt{\frac{m}{2\pi k_B T}} e^{\frac{-mv_x^2}{2k_B T}}
$$

Since we are dealing with a one-dimensional problem we may drop the subscript x, begin careful not to confuse v (which is really v_x) with the total velocity v used earlier:

$$
1 = \frac{\omega_p^2}{k^2}\int_{-\infty}^{\infty} \frac{\partial \hat{f}_0/\partial v}{v - \omega/k} \mathrm{d}v \tag{8.30}
$$

Here, $\hat{f}_0$ is understood to be a one-dimensional distribution function, the integrations over v_y and v_z having been made. This equation holds for any equilibrium distribution $\hat{f}_0(v)$.

The integral in this equation is not straightforward to evaluate because of the singularity at $v = \omega/k$. One might think that the singularity would be of no concern, because in practice ω is almost always never real; waves are usually slightly damped by collisions or are amplified by some instability mechanisms. Since the velocity v is a real quantity, the denominator never vanishes. Landau was the first to treat this equation properly. He found that even though the singularity lies off the path of integration, its presence introduces an important modification to the plasma wave dispersion relation — an effect not predicted by the fluid theory.

Consider an initial value problem in which the plasma is given a sinusoidal perturbation, and therefore k is real. If the perturbation grows or decays, ω will be complex. This integral must be treated as a contour integral in the complex v plane. Possible contours are shown for (a) an unstable wave, with $\Im(\omega) > 0$, and (b) a dampled wave, with $\Im(\omega) < 0$. Normally, one would evaluate the line integral along the real v axis by the residue theorem:

$$\int_{C_1} G \mathrm{d}v + \int_{C_2} G \mathrm{d}v = 2\pi i R(\omega/k)$$

where G is the integrand, C_1 is the path along the real axis, C_2 is the semicircle at infinity, and $R(\omega/k)$ is the residue at ω/k. This works if the integral over C_2 vanishes. Unfortunately, this does not happen for a Maxwellian distribution, which contains the factor

$$\exp(-v^2/v_{th}^2)$$

This factor becomes large for $v \to \pm i\infty$, and the contribution from C_2 cannot be neglected. Landau showed that when the problem is properly treated as an initial value problem the correct contour to use is the curve C_1 passing below the singularity. This integral must in general be evaluated numerically.

Although an exact analysis of this problem is complicated, we can obtain an approximate dispersion relation for the case of large phase velocity and weak damping. In this case, the pole at ω/k lies near the real v axis. The contour prescribed by Landau is then a straight line along the $\Re(v)$ axis with a small semicircle around the pole. In going around the pole, one obtains $2\pi i$ time half the residue there. Then Equation 8.30 becomes

$$1 = \frac{\omega_p^2}{k^2}\Big[P\int_{-\infty}^{\infty}\frac{\partial \hat{f}_0/\partial v}{v-(\omega/k)}\mathrm{d}v + i\pi\frac{\partial \hat{f}_0}{\partial v}\Big|_{v=\omega/k}\Big] \tag{8.31}$$

where P stands for the Cauchy principal value. To evaluate this, we integrate along the real v axis but stop just before encountering the pole. If the phase velocity $v_\phi = \omega/k$ is sufficiently large, as we assume, there will not be much contribution from the neglected part of the contour, since both $\hat{f}_0$ and $\partial\hat{f}_0/\partial v$ are very small there. The integral above can be evaluated by integration by parts:

$$\int_{-\infty}^{\infty}\frac{\partial \hat{f}_0}{\partial v}\frac{\mathrm{d}v}{v-v_\phi} = \Big[\frac{\hat{f}_0}{v-v_\phi}\Big]_{-\infty}^{\infty} - \int_{-\infty}^{\infty}\frac{-\hat{f}_0\mathrm{d}v}{(v-v_\phi)^2} = \int_{-\infty}^{\infty}\frac{\hat{f}_0\mathrm{d}v}{(v-v_\phi)^2}$$

Since this is just an average of $(v-v_\phi)^{-2}$ over the distribution, the real part of the dispersion relation can be written

$$1 = \frac{\omega_p^2}{k^2}\overline{(v-v_\phi)^{-2}}$$

Since $v_\phi \gg v$ has been assumed, we can expand $(v-v_\phi)^{-2}$:

$$(v-v_\phi)^{-2} = v_\phi^{-2}\Big(1-\frac{v}{v_\phi}\Big)^{-2} = v_\phi^{-2}\Big(1+\frac{2v}{v_\phi}+\frac{3v^2}{v_\phi^2}+\frac{4v^3}{v_\phi^3}+...\Big)$$

The odd terms vanish upon taking the average, and we have

$$\overline{(v-v_\phi)^{-2}} \approx v_\phi^{-2}\Big(1+\frac{3\overline{v^2}}{v_\phi^2}\Big)$$

We now let $\hat{f}_0$ be Maxwellian and evaluate $\overline{v^2}$. Remembering that v here is an abbreviation for v_x, we can write

$$\frac{1}{2}m\overline{v_x^2} = \frac{1}{2}k_BT_e$$

there being only one degree of freedom. The dispersion relation then becomes

$$\begin{aligned}1 &= \frac{\omega_p^2}{k^2}\frac{k^2}{\omega^2}\Big(1+3\frac{k^2}{\omega^2}\frac{k_BT_e}{m}\Big)\\ \omega^2 &= \omega_p^2 + \frac{\omega_p^2}{\omega^2}\frac{3k_BT_e}{m}k^2\end{aligned}$$

If the thermal correction is small (i.e. the second term on the right-hand side is small, such that $\omega \approx \omega_p$), we may replace ω^2 by ω_p^2 in the second term. We then have

$$\omega^2 = \omega_p^2 + \frac{3k_BT_e}{m}k^2$$

which is the same as that been obtained from the fluid equations with $\gamma = 3$.

We now return to the imaginary term in the dispersion relation. In evaluating this small term, it will be sufficiently accurate to neglect the thermal correction to the real part of ω and let $\omega^2 \approx \omega_p^2$. From the evaluation of the real part above we see that the principle value of the integral is approximately k^2/ω^2. The dispersion relation now becomes

$$1 = \frac{\omega_p^2}{\omega^2} + i\pi \frac{\omega_p^2}{k^2} \frac{\partial \hat{f}_0}{\partial v}\bigg|_{v=v_\phi}$$

$$\omega^2 \Big(1 - i\pi \frac{\omega_p^2}{k^2} \frac{\partial \hat{f}_0}{\partial v}\bigg|_{v=v_\phi} \Big) = \omega_p^2$$

Treating the imaginary term as small, we can bring it to the right-hand side and take the square root by Taylar series expansion. We then obtain

$$\omega = \omega_p \Big(1 + i\frac{\pi}{2} \frac{\omega_p^2}{k^2} \frac{\partial \hat{f}_0}{\partial v}\bigg|_{v=v_\phi} \Big) \tag{8.32}$$

If $\hat{f}_0$ is a one-dimensional Maxwellian, we have

$$\frac{\partial \hat{f}_0}{\partial v} = (\pi v_{th}^2)^{-1/2} \Big(\frac{-2v}{v_{th}^2} \Big) \exp\Big(\frac{-v^2}{v_{th}^2} \Big) = -\frac{2v}{\sqrt{\pi} v_{th}} \exp\Big(\frac{-v^2}{v_{th}^2} \Big)$$

We may approximate v_ϕ by ω_p/k in the coefficient, but in the exponent we must keep the thermal correction in the real part of the dispersion relation. The damping is then given by

$$\begin{aligned}
\Im(\omega) &= -\frac{\pi}{2} \frac{\omega_p^2}{k^2} \frac{2\omega_p}{k\sqrt{\pi}} \frac{1}{v_{th}^3} \exp\Big(\frac{-\omega^2}{k^2 v_{th}^2} \Big) \\
&= -\sqrt{\pi}\omega_p \Big(\frac{\omega_p}{k v_{th}} \Big)^3 \exp\Big(\frac{-\omega_p^2}{k^2 v_{th}^2} \Big) \exp\Big(\frac{-3}{2} \Big) \\
\Im\Big(\frac{\omega}{\omega_p} \Big) &= -0.22\sqrt{\pi} \Big(\frac{\omega_p}{k v_{th}} \Big)^3 \exp\Big(\frac{-1}{2k^2 \lambda_D^2} \Big)
\end{aligned}$$

Since $\Im(\omega)$ is negative, there is a collisionless damping of plasma waves; this is called *Landau damping.* As is evident from the expression, this damping is extremely small for small $k\lambda_D$, but becomes important for $k\lambda_D = \mathcal{O}(1)$. This effect is connected with f_1, the distortion of the distribution function caused by the wave.

8.6 The Meaning of Landau Damping

The theoretical discovery of wave damping without energy dissipation by collisions is perhaps the most astounding result of plasma physics research. That this is a real effect has been demonstrated in the laboratory. Although a simple physical explanation for this damping is now available, it is a triumph of applied mathematics that this unexpected effect was first discovered purely mathematically in the course of a careful analysis of a contour integral. Landau damping is a characteristic of collisionless plasmas, but it may also have application in other fields. For instance, in the kinetic treatment of galaxy formation, stars can be considered as atoms of a plasma interacting via gravitational rather than electromagnetic forces. Instabilities of the gas of stars can cause spiral arms to form, but this process is limited by Landau damping.

To see what is responsible for Landau damping, we first notice that $\Im(\omega)$ arises from the pole at $v = v_\phi$. Consequently, the effect is connected with those particles in the distribution that have a velocity nearly equal to the phase velocity — the "resonant particles". These particles travel along with the wave and do not see a rapidly fluctuating electric field: they can, therefore, exchange energy with the wave effectively. The easiest way to understand this exchange of energy is to picture a surfer trying to catch an ocean wave. (Warning: this picture is only for directing our thinking along the right lines; it does not correctly explain the damping.) If the surfboard is not moving, it merely bobs up and down as the wave goes by and does not gain any energy on the average. Similarly, a boat propelled much faster than the wave cannot exchange much energy with the wave. However, if the surfboard has almost the same velocity as the wave, it can be caught and pushed along by the wave; this is, after all, the main purpose of the exercise. In that case, the surfboard gains energy, and therefore the wave must lose energy and is damped. On the other hand, if the surfboard should be moving slightly faster than the wave, it would push on the wave as it moves uphill; then the wave could gain energy. In a plasma, there are electrons both faster and slower than the wave. A Maxwellian distribution, however, has more slow electrons than fast ones. Consequently, there are more particles taking energy from the wave than vice versa, and the wave is damped. As particles with $v \approx v_\phi$ are trapped in the wave, $f(v)$ is flattened near the phase velocity. This distortion is $f_1(v)$ which we calculated. As seen in Fig (ADD IT!), the perturbed distribution function contains the same number of particles but has gained total energy (at the expense of the wave).

From this discussion, one can surmise that if $f_0(v)$ contained more fast particles than slow particles, a wave can be excited. Indeed, from the expression of ω above, it is apparent that $\Im(\omega)$ is positive if $\partial \hat{f}_0/\partial v$ is positive at $v = v_\phi$. Such a distribution is shown in Fig.7-19 (ADD IT!). Waves with v_ϕ in the region of positive slope will be unstable, gaining energy at the expense of the particles. This is just the finite-temperature analogy of the two stream instability. When there are two cold ($k_BT = 0$) electron streams in motion, $f_0(v)$ consists of two δ-functions. This is clearly unstable because $\partial f_0/\partial v$ is infinite; and indeed, we found the instability from fluid theory. When the streams have fnite temperature, kinetic theory tells us that the relative densities and temperatures of the two stream must be such as to have a region of positive $\partial f_0(v)/\partial v$ between them; more precisely, the total distribution function must have a minimum for instability.

The physical picture of a surfer catching waves is very appealing, but it is not precise enough to give us a real understanding of Landau damping. There are actually two kinds of Landau damping. Both kinds are independent of dissipative collisional mechanisms. If a particle is caught in the potential well of a wave, the phenomenon is called "trapping". As in the case of a surfer, particles can indeed gain or lose energy in trapping. However, trapping does not lie within the purview of the linear theory. That this is true can be seen from the equation of motion

$$m\ddot{x} = qE(x)$$

If one evaluates $E(x)$ by inserting the exact value of x, the equation would be nonlinear, since $E(x)$ is somehting like $\sin kx$. What is done in linear theory is to use for x the unperturbed orbit; i.e. $x = x_0 + v_0 t$. Then this becomes linear. This approximation, however, is no longer valid when a particle is trapped. When it encounters a potential hill large enough to reflect it, its velocity and position are, of course, greatly affected by the wave and are not close to their unperturbed values. In fluid theory, the equation of motion is

$$m\left[\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v}\cdot\nabla)\mathbf{v}\right] = q\mathbf{E}(x)$$

Here, $\mathbf{E}(x)$ is to be evaluated in the laboratory frame, which is easy; but to make up for it, there is the $(\mathbf{v}\cdot\nabla)\mathbf{v}$ term. The neglect of $(\mathbf{v}_1\cdot\nabla)\mathbf{v}_1$ in linear theory amounts to the same thing as using unperturbed orbits. In kinetic theory, the nonlinear term that is neglected is, from the first-order Vlasov Equation 8.29,

$$\frac{q}{m}E_1\frac{\partial f_1}{\partial v}$$

When the particles are trapped, they reverse their direction of travel relative to the wave, so the distribution function $f(v)$ is greatly disturbed near $v = \omega/k$. This means that $\partial f_1/\partial v$ is comparable to $\partial f_0/\partial v$, and the term above is not negligible. Hence, trapping is not in the linear theory.

When a wave grows to a large amplitude, collisionless damping with trapping does occur. One then finds that the wave does not decay monotonically; rather, the amplitude fluctuates during the decay as the trapped particles bounce back and forth in the potential wells. This is *nonlinear* Landau damping. Since the result before was derived from a *linear* theory, it must arise from a different physical effect. The question is: can untrapped electrons moving close to the phase velocity of the wave exchange energy with the wave? Before giving the answer, let us examine the energy of such electrons.

8.6.1 The Kinetic Energy of a Beam of Electrons

We may divide the electron distribution $f_0(v)$ into a large number of monoenergetic beams. Consider one of these beams: it has unperturbed velocity u and density n_u. The velocity u may lie near v_ϕ, so that this beam may consist of resonant electrons. We now turn on a plasma oscillation $E(x,t)$ and consider the kinetic energy of the beam as it moves through the crests and troughs of the wave. The wave is caused by a self-consistent motion of all the beams together. If n_u is small enough (the number of beams large enough), the beam being examined has a negligible effect on the wave and may be considered as moving in a given field $E(x,t)$. Let

$$E = E_0 \sin(kx - \omega t) = -d\phi/\mathrm{dt}$$
$$\phi = (E_0/k)\cos(kx - \omega t)$$

The linearized fluid equation for the beam is

$$m\Big(\frac{\partial v_1}{\partial t} + u\frac{\partial v_1}{\partial x}\Big) = -eE_0 \sin(kx - \omega t)$$

A possible solution is

$$v_1 = -\frac{eE_0}{m}\frac{\cos(kx-\omega t)}{\omega - ku}$$

This is the velocity modulation caused by the wave as the beam electrons move past. To conserve particle flux, there is a corresponding oscillation in density, given by the linearized continuity equation:

$$\frac{\partial n_1}{\partial t} + u\frac{\partial n_1}{\partial x} = -n_u\frac{\partial v_1}{\partial x}$$

Since v_1 is proportional to $\cos(kx-\omega t)$, we can try $n_1 = \bar{n}_1\cos(kx - \omega t)$. Substitution of this into the above yields

$$n_1 = -n_u\frac{eE_0k}{m}\frac{\cos(kx-\omega t)}{(\omega - ku)^2}$$

(n_1 and v_1 can be shown in a series of phase relation plots as in Fig.7-21)(ADD IT!) one wavelength of E and of the potential $-e\phi$ seen by the beam electrons.

We may now compute the kinetic energy W_k of the beam:

$$
\begin{aligned}
W_k &= \frac{1}{2}m(n_u + n_1)(u + v_1)^2 \\
&= \frac{1}{2}m(n_u u^2 + n_u v_1^2 + 2un_1 v_1 + n_1 u^2 + 2n_u u v_1 + n_1 v_1^2)
\end{aligned}
$$

The last three terms contain odd powers of oscillating quantites, so they will vanish when we average over a wavelength. The change in W_k due to the wave is found by subtracting the first term, which is the original energy. The average energy change is then

$$
\langle \Delta W_k \rangle = \frac{1}{2}m \left\langle n_u v_1^2 + 2un_1 v_1 \right\rangle
$$

From the form of v_1, we have

$$
n_u \left\langle v_1^2 \right\rangle = \frac{1}{2} n_u \frac{e^2 E_0^2}{m^2(\omega - ku)^2}
$$

the factor $\frac{1}{2}$ representing $\left\langle \cos^2(kx - \omega t) \right\rangle$. Similarly, from the form of n_1,

$$
2u \left\langle n_1 v_1 \right\rangle = n_u \frac{e^2 E_0^2 ku}{m^2(\omega - ku)^3}
$$

Consequently,

$$
\begin{aligned}
\langle \Delta W_k \rangle &= \frac{1}{4} m n_u \frac{e^2 E_0^2}{m^2(\omega - ku)^2} \left[1 + \frac{2ku}{\omega - ku} \right] \\
&= \frac{n_u}{4} \frac{e^2 E_0^2}{m} \frac{\omega + ku}{m^2(\omega - ku)^3}
\end{aligned}
$$

This result shows that $\langle \Delta W_k \rangle$ depends on the frame of the observer and that it does not change secularly with time. Consider the picture of a frictionless block sliding over a washboard-like surface. (ADD FIGURE!) In the frame of the washboard, $\langle \Delta W_k \rangle$ is proportional to $-(ku)^{-2}$, as seen by taking $\omega = 0$. It is intuitively clear that (1) $\langle \Delta W_k \rangle$ is negative, since the block spends more time at the peaks than at the valleys, and (2) the block does not gain or lose energy on the average, *once the oscillation is started* (no time-dependence). Now if one goes into a frame in which the washboard is movign with a steady velocity ω/k (a velocity unaffected by the motion of the block, since we have assumed that n_u is negligibly small compared with the density of the whole plasma), it is still true that the block does not gain or lose energy on the average, once the oscillation is started. But the above equation tells us that $\langle \Delta W_k \rangle$ depends on the velocity ω/k, and hence on the frame of the observer. In particular, it shows that a beam has less energy in the presence of the wave than in its absence if $\omega - ku < 0$ or $u > u_\phi$, and it has more energy if $\omega - ku > 0$ or $u < u_\phi$. The reason for this can be traced back to the phase relation between n_1 and v_1. As Fig.7-23 (ADD IT!) shows, W_k is a parabolic function of v. As v oscillates between $u - |v_1|$ and $u + |v_1|$, W_k will attain an average value larger than the equilibrium value W_{k0}, provided that the particle spends an equal amount of time in each half of the oscillation. This effect is the meaning of the first term $\frac{1}{2}m\left\langle n_u v_1^2\right\rangle$, which is positive definite. The second term $\frac{1}{2}m\left\langle 2un_1v_1\right\rangle$ is a correction due to the fact that the particle does not distribution its time equally. In Fig.7-21 (ADD IT!), one sees that both electrons a and b spend more time at the top the potential hill than at the bottom, but electron a reaches that point after a period of deceleration, so that v_1 is negative there, while electron b reaches that point after a period of acceleration (to the right), so that v_1 is positive there. This effect causes $\langle W_k\rangle$ to change sign at $u = v_\phi$.

8.6.2 The Effect of Initial Conditions

The result we have just derived, however, still has nothing to do with linear Landau damping. Damping requires a continuous increase of W_k at the expense of wave energy, but we have found that $\langle \Delta W_k\rangle$ for untrapped particles is contant in time. If neither the untrapped particles nor the trapped particles are responsible for linear Landau damping, what is? The answer can be gleaned form the following observation: if $\langle \Delta W_k\rangle$ is positive, say, there must have been a time when it was increasing. Indeed, there are particles in the original distribution which have velocities so close to v_ϕ that at time t they have not yet gone a half wavelength relative to the wave. For these particles, one cannot take the average $\langle \Delta W_k\rangle$. These particles can absorb energy from the wave and are properly called the "resonant" particles. As time goes on, the number of resonant electrons decreases, since an increasing number will have shifted more than $\frac{1}{2}\lambda$ from their orginal positions. The damping rate, however, can stay constant, since the amplitude is now smaller, and it takes fewer electrons to maintain a constant damping rate.

The effect of the initial conditions is most easily seen form a phase-space diagram (Fig.7-24)(ADD IT with Luxor?).

8.7 A Physical Derivation of Landau Damping

We are now in a position to derive the Landau damping rate recourse to contour integration. Although Landau's derivation of collisionless damping was short and neat, it was not clear that it concerned a physically observable phenomenon until J. M. Dawson gave the longer, intuitive derivation which is paraphrased in this section. As before, we divide the plasma up into beams of velocity u and density n_u, and examine their motion in a wave

$$E = E_1 \sin(kx - \omega t)$$

From the derivations in the previous section, the velocity of each beam is

$$v_1 = -\frac{eE_1}{m}\frac{\cos(kx - \omega t)}{\omega - ku}$$

This solution satisfies the equation of motion, but it does not satisfy the initial condition $v_1 = 0$ at $t = 0$. It is clear that this initial condition must be imposed; otherwise, v_1 would be very large in the vicinity of $u = \omega/k$, and the plasma would be in a specially prepared state initially. We can fix up the solution to satisfy the initial condition by adding an arbitrary function of $kx - kut$. The composite solution would still satisfy the equation of motion because the operator on the left-hand side, when applied to $f(kx - kut)$, gives zero. Obviously, to get $v_1 = 0$ at $t = 0$, the function $f(kx - kut)$ must be taken to be $-\cos(kx - kut)$. Thus we have, instead of the expression above,

$$v_1 = \frac{-eE_1}{m}\frac{\cos(kx - \omega t) - \cos(kx - kut)}{\omega - ku}$$

Next, we must solve the equation of continuity for n_1, again subject to the initial condition $n_1 = 0$ at $t = 0$. Since we are now much cleverer than before, we may try a solution of the form

$$n_1 = \bar{n}_1[\cos(kx - \omega t) - \cos(kx - kut)]$$

Inserting this into the equation of continuity and using the expression for v_1 above, we find

$$\bar{n}_1 \sin(kx - \omega t) = -n_u \frac{eE_1 k}{m}\frac{\sin(kx - \omega t) - \sin(kx - kut)}{(\omega - ku)^2}$$

Apparently, we were not clever enough, since the $\sin(kx - \omega t)$ factor does not cancel. To get a term of the form $\sin(kx - kut)$, which came from the added term in v_1, we can add a term of the form $At \sin(kx - kut)$ to n_1. This term obviously vanishes at $t = 0$, and it will give the $\sin(kx - kut)$ term when the operator on the left-hand side of the equation of continuity operates on the t factor. When the operator operates on the $\sin(kx - kut)$ factor, it yields zero. The coefficient A must be proportional to $(\omega - ku)^{-1}$ in order to match the same factor in $\partial v_1/\partial x$. Thus we take

$$\begin{aligned} n_1 = &- n_u \frac{eE_1 k}{m}\frac{1}{(\omega - ku)^2} \\ &\times [\cos(kx - \omega t) - \cos(kx - kut) - (\omega - ku)t \sin(kx - kut)] \end{aligned}$$

This clearly vanishes at $t = 0$, and one can easily verify that it satisfies the equation of continuity.

These expressions for v_1 and n_1 allow us now to calculate the work done by the wave on each beam. The force acting on a unit volume of each beam is

$$F_u = -eE_q \sin(kx - \omega t)(n_u + n_1)$$

and therefore its energy changes at the rate

$$\frac{\mathrm{d}W}{\mathrm{d}t} = F_u(u + v_1) = -eE_1 \sin(kx - \omega t)(\underbrace{n_u u}_{1} + \underbrace{n_u v_1}_{2} + \underbrace{n_1 u}_{3} + \underbrace{n_1 v_1}_{4})$$

We now take the spatial average over a wavelength. The first term vanishes because $n_u u$ is a constant. The fourth term can be neglected because it is second order, but in any case it can be shown to have zero average. The second and third terms can be evaluated with the help of identities

$$\begin{aligned}
\langle \sin(kx - \omega t)\cos(kx - kut)\rangle &= -\frac{1}{2}\sin(\omega t - kut) \\
\langle \sin(kx - \omega t)\sin(kx - kut)\rangle &= \frac{1}{2}\cos(\omega t - kut)
\end{aligned}$$

The result is then

$$\left\langle \frac{\mathrm{d}W}{\mathrm{d}t} \right\rangle_u = \frac{e^2 E_1^2}{2m} n_u \Big[\frac{\sin(\omega t - kut)}{\omega - ku} \\ + ku \frac{\sin(\omega t - kut) - (\omega - ku)t\cos(\omega t - kut)}{(\omega - ku)^2}\Big]$$

The total work done on the particles is found by summing over all the beams:

$$\sum_u \left\langle \frac{\mathrm{d}W}{\mathrm{d}t} \right\rangle_u = \int \frac{f_0(u)}{n_u} \left\langle \frac{\mathrm{d}W}{\mathrm{d}t} \right\rangle_u dy = n_0 \int \frac{\hat{f}_0(u)}{n_u} \left\langle \frac{\mathrm{d}W}{\mathrm{d}t} \right\rangle_u du$$

Inserting the expression of $\left\langle \frac{\mathrm{d}W}{\mathrm{d}t} \right\rangle_u$ and using the definition of ω_p, we then find for the rate of change of kinetic energy

$$
\begin{aligned}
\left\langle \frac{dW_k}{\mathrm{d}t} \right\rangle =& \frac{\epsilon_0 E_1^2}{2}\omega_p^2 \Big[\int \hat{f}_0(u) \frac{\sin(\omega t - kut)}{\omega - ku} du \\
&+ \int \hat{f}_0(u) \frac{\sin(\omega t - kut) - (\omega - ku)t\cos(\omega t - kut)}{(\omega - ku)^2} kudu \Big] \\
=& \frac{1}{2}\epsilon_0 E_1^2 \omega_p^2 \int_{-\infty}^{\infty} \hat{f}_0(u) du \Big\{ \frac{\sin(\omega t - kut)}{\omega - ku} + u\frac{d}{du}\Big[\frac{\sin(\omega t - kut)}{\omega - ku}\Big]\Big\} \\
=& \frac{1}{2}\epsilon_0 E_1^2 \omega_p^2 \int_{-\infty}^{\infty} \hat{f}_0(u) du \frac{d}{du}\Big[u\frac{\sin(\omega t - kut)}{\omega - ku}\Big]
\end{aligned}
$$

This is to be set equal to the rate of loss of wave energy density W_w. The wave energy consists of two parts. The first part is the energy density of the electrostatic field:

$$
\langle W_E \rangle = \frac{\epsilon \langle E^2 \rangle}{2} = \frac{\epsilon E_1^2}{4}
$$

The second part is the kinetic energy of oscillation of the particles. If we again divide the plasma up into beams, the energy per beam is given as before

$$
\langle \Delta W_k \rangle_u = \frac{1}{4}\frac{n_u}{m}\frac{e^2 E_1^2}{(\omega - ku)^2}\Big[1 + \frac{2ku}{\omega - ku}\Big]
$$

In deriving this result, we did not use the correct initial conditions, which are important for the resonant particles; however, the latter contribute very little to the total energy of the wave. Summing over the beams, we have

$$
\langle \Delta W_k \rangle = \frac{1}{4}\frac{e^2 E_1^2}{m}\int_{-\infty}^{\infty} \frac{f_0(u)}{(\omega - ku)^2}\Big[1 + \frac{2ku}{\omega - ku}\Big] du
$$

The second term in the brackets can be neglected in the limit $\omega/k \gg v_{th}$, which we shall take in order to compare with our previous results. (???) The dispersion relation is found by Poisson's equation:

$$
k\epsilon_0 E_1 \cos(kx - \omega t) = -e\sum_u n_1
$$

Using the expression for n_1 in the previous section with the wrong initial conditions, we have

$$
1 = \frac{e^2}{\epsilon_0 m}\sum_u \frac{n_u}{(\omega - ku)^2} = \frac{e^2}{\epsilon_0 m}\int_{-\infty}^{\infty} \frac{f_0(u) du}{(\omega - ku)^2}
$$

Comparing this with the expression of $\langle \Delta W_k \rangle$, we find

$$\langle \Delta W_k \rangle = \frac{1}{4}\frac{e^2 E_1^2}{m}\frac{\epsilon_0 m}{e^2} = \frac{\epsilon_0 E_1^2}{4} = \langle W_E \rangle$$

Thus

$$W_w = \frac{\epsilon E_1^2}{2}$$

The rate of change of wave energy density W_w is given by $-\left\langle \frac{dW_k}{\mathrm{d}t} \right\rangle$:

$$\frac{dW_w}{\mathrm{d}t} = -W_w \omega_p^2 \int_{-\infty}^{\infty} \hat{f}_0(u) du \frac{d}{du}\Big[u \frac{\sin(\omega t - kut)}{\omega - ku} \Big]$$

Integration by parts gives

$$\begin{aligned} \frac{dW_w}{\mathrm{d}t} =& W_w \omega_p^2 \Big\{ \Big[u \hat{f}_0(u) \frac{\sin(\omega - ku)t}{\omega - ku} \Big]_{-\infty}^{\infty} \\ & - \int_{-\infty}^{\infty} u \frac{d\hat{f}_0}{du} \frac{\sin(\omega - ku)t}{\omega - ku} du \Big\} \end{aligned}$$

The integrated part vanishes for well-behaved functions $\hat{f}_0(u)$, and we have

$$\frac{dW_w}{\mathrm{d}t} = W_w \frac{\omega}{k} \omega_p^2 \int_{-\infty}^{\infty} \frac{d\hat{f}_0}{du} \Big[\frac{\sin(\omega - ku)t}{\omega - ku} \Big] du$$

where u has been set equal to ω/k (a constant), since only velociies very close to this will contirbute to the integral. In fact, for sufficiently large t, the square bracket can be approximated by a δ-function:

$$\delta\Big(u - \frac{\omega}{k} \Big) = \frac{k}{\pi} \lim_{t \to \infty} \Big[\frac{\sin(\omega - ku)t}{\omega - ku} \Big]$$

(The original form is $\delta(x) = \lim_{\epsilon \to 0} \frac{\sin(x/\epsilon)}{\pi x}$, where the function on the right is called sinc.)

Thus

$$\frac{dW_w}{\mathrm{d}t} = W_w \frac{\omega}{k} \omega_p^2 \frac{\pi}{k} \frac{\omega}{k} \hat{f}_0\Big(\frac{\omega}{k} \Big) = W_w \pi \omega \frac{\omega_p^2}{k^2} \hat{f}_0'\Big(\frac{\omega}{k} \Big)$$

Since $\Im(\omega)$ is the growth rate of E_1, and W_w is proportional to E_1^2, we must have

$$\frac{dW_w}{\mathrm{d}t} = 2\Im(\omega) W_w$$

Hence

$$\Im(\omega) = \frac{\pi}{2}\omega\frac{\omega_p^2}{k^2}\hat{f}_0'\left(\frac{\omega}{k}\right)$$

in agreement with the previous result for $\omega = \omega_p$.

8.7.1 The Resonant Particles

We are now in a position to see precisely which are the resonant particles that contribute to linear Landau damping. Fig.7-25(Sinc function, ADD IT!) gives a plot of the factor multiplying $\hat{f}_0'(u)$ in the integrand. We see that the largest contribution comes from particles with $|\omega - ku| < \pi/t$, or $|v - v_\phi| < \pi/k = \lambda/2$; i.e. those particles in the initial distribution that have not yet traveled a half-wavelength relative to the wave. The width of the central peak narrows with time, as expected. The subsidiary peaks in the "diffraction pattern" of Fig.7-25 come form particles that have traveled into neighboring half-wavelengths of the wave potential. These particles rapidly become spread out in phase, so that they contribute little on the average; the initial distribution is forgotten. Note that the width of the central peak is independent of the initial amplitude of the wave; hence, the resonant particles may include both trapped and untrapped particles. This phenomenon is unrelated to particle trapping.

8.7.2 Two Paradoxes Resolved

Fig.7-25 shows that the integrand sinc function is an even function of $\omega - ku$, so that particles going both faster than the wave and slower than the wave add to Landau damping. This is the physical picture we found in Fig.7-24. On the other hand, the slope of the curve of Fig.7-25, which represents the factor in the integrand of one previous equation (WE NEED NUMBERING!), is an odd function of $\omega - ku$; and one would infer from this that particles traveling faster than the wave give energy to it, while those traveling slower than the wave take energy from it. The two descriptions differ by an integration by parts. Both descriptions are correct; which is to be chosen depends on whether one wishes to have $\hat{f}_0(u)$ or $\hat{f}_0(u)'$ in the integrand.

A second paradox concerns the question of Galilean invariance. If we take the view that damping requires there be fewer particles traveling faster than the wave than slower, there is no problem as long as one is in the frame in which the plasma is at rest. However, if one goes into another frame moving with a velocity V, there would appear to be more particles moving faster than the wave than slower, and one would expect the wave to grow instead of decay. This paradox is removed by reinserting the second term in Eq.(???) $\frac{2ku}{\omega-ku}$ which we neglected. As shown in the previous section, this term can make $\langle \Delta W_k \rangle$ negative. Indeed, in the moving frame, the second term is not negligible, $\langle \Delta W_k \rangle$ is negative, and the wave appears to have negative energy (that is, there is more energy in the quiescent, drifting Maxwellian distribution than in the presence of an oscillation). The wave "grows", but adding energy to a negative energy wave makes its amplitude decrease.

8.8 BGK and Van Kampen Modes

We have seen that Landau damping is directly connected to the requirement that $f_0(v)$ be initially uniform in space. (It tends to make negative slopes to zero.) On the other hand, one can generate undamped electron waves if $f(v, t = 0)$ is mode to be constant along the particle trajectories initially. (???) It is easy to from Fig.7-24 that the particles will neither gain nor lose energy, on the average, if the plasma is initially prepared so that the density is constant along each trajectory. Such a wave is called a BGK mode, since it was I. B. Bernstein, J. M. Greene, and M. D. Kruskal who first showed that undamped waves of arbitrary ω, k, amplitude, and waveform were possible. The crucial parameter to adjust in tailoring $f(v, t = 0)$ to form a BGK mode is the relative number of trapped and untrappped particles. If we take the small-amplitude limit of a BGK mode, we obtain what is called a Van Kampen mode. In this limit, only the particles with $v = v_\phi$ are trapped. We can change the number of trapped particles by adding to $f(v, t = 0)$ a term proportional to $\delta(v - v_\phi)$. Examination of Fig.7-24 will show that adding particles along the line $v = v_\phi$ will not cause damping — at a later time, there are just as many particles gaining energy as losing energy. In fact, by choosing distributions with δ-functions at other values of v_ϕ, one can generate undamped Van Kampen modes of arbitrary v_ϕ. Such sungular initial conditions are, however, not physical. To get a smoothly varying $f(v, t = 0)$, one must sum over Van Kampen modes with a distribution of $v_\phi s$. Although each mode is undamped, the total perturbation will show Landau damping because the various modes get out of phase with one another. (???)

8.9 Ion Landau Damping

Electrons are not the only resonant particles. If a wave has a slow enough phase velocity to match the thermal velocity of ions, ion Landau damping can occur. The ion acoustic wave, for instance, is greatly affected by Landau damping. Recall from the fluid theory that the dispersion relation for ion waves is

$$\frac{\omega}{k} = v_s = \Big(\frac{k_B T_e + \gamma_i k_B T_i}{m_i}\Big)^{1/2}$$

If $T_e \le T_i$, the phase velocity lies in the region where $f_{0i}(v)$ has a negative slope, as shown in Fig.7-30(A)(ADD IT!!!). Consequently, ion waves are heavily Landau-damped if $T_e \le T_i$. Ion waves are observable only if $T_e \gg T_i$(Fig.7-30(B)), so that the phase velocity lies far in the tail of the ion velocity distribution. A clever way to introduce Landau damping in a controlled manner was employed by Alexeff, Jones, and Montgomery. A weakly damped ion wave was created in a heavy-ion plasma (such as xenon) with $T_e \gg T_i$. A small amount of a light atom (helium) was then added. Since the helium had about the temperature as the xenon but had much smaller mass, its distribution function was much broader, as shown by the dashed curve in Fig.7-30(B). The resonant helium ions then caused the wave to damp.

8.9.1 The Plasma Dispersion Function

To introduce some of the standard terminlogy of kinetic theory, we now calculate the ion Landau damping of ion acoustic waves in the absence of magnetic fields. Ions and electrons follow the Vlasov equation and have perturbations of the form $f_1 \propto \exp(ikx - i\omega t)$ indicating plane waves propagating in the x direction. The solution for f_1 is given by the linearized momentum equation with appropriate modifications:

$$f_{1j} = -\frac{iq_j E}{m_j} \frac{\partial f_{oj}/\partial v_j}{\omega - kv_j}$$

where E and v_j stand for E_x, v_{xj}; and the jth species has charge q_j, mass m_j, and particle velocity v_j. The density perturbation of the jth species is given by

$$n_{1j} = \int_{-\infty}^{\infty} f_{1j}(v_j) dv_j = -\frac{iq_j E}{m_j} \int_{-\infty}^{\infty} \frac{\partial f_{oj}/\partial v_j}{\omega - kv_j} dv_j$$

Let the equilibrium distributions f_{0j} be one-dimensional Maxwellians:

$$f_{0j} = \frac{n_{0j}}{v_{thj}\pi^{1/2}} e^{-v_j^2/v_{th}^2}, \quad v_{thj} \equiv (2k_B T_j/m_j)^{1/2}$$

Introducing the dummy integration variable $s = v_j/v_{thj}$, we can write n_{1j} as

$$n_{1j} = \frac{iq_j E n_{0j}}{km_j v_{thj}^2} \frac{1}{\pi^{1/2}} \int_{-\infty}^{\infty} \frac{(d/ds)(e^{-s^2})}{s - \zeta_j} ds$$

where

$$\zeta \equiv \omega/kv_{thj}$$

We now define the *plasma dispersion function* $Z(\zeta)$:

$$Z(\zeta) = \frac{1}{\pi^{1/2}} \int_{-\infty}^{\infty} \frac{e^{-s^2}}{s-\zeta} ds \quad \Im(\zeta) > 0 \tag{8.33}$$

(Why positive imaginary part???)This is a contour integral, as explained in previous sections, and analytic continuation to the lower half plane must be used if $\Im(\zeta) < 0$. $Z(\zeta)$ is a complex function of a complex argument (since ω or k usually has an imaginary part). In cases where $Z(\zeta)$ cannot be approximated by an asymptotic formula, one can do it numerically.

To express n_{1j} in terms of $Z(\zeta)$, we take the derivative with respect to ζ:

$$Z'(\zeta) = \frac{1}{\pi^{1/2}} \int_{-\infty}^{\infty} \frac{e^{-s^2}}{(s-\zeta)^2} ds$$

Integration by parts yields

$$Z'(\zeta) = \frac{1}{\pi^{1/2}} \left[\frac{-e^{-s^2}}{s-\zeta} \right]_{-\infty}^{\infty} + \frac{1}{\pi^{1/2}} \int_{-\infty}^{\infty} \frac{(d/ds)(e^{-s^2})}{s-\zeta} ds$$

The first term vanishes, as it must for any well-behaved distribution function. Now we have

$$n_{1j} = \frac{iq_j E n_{0j}}{k m_j v_{thj}^2} Z'(\zeta_j)$$

Poisson's equation is

$$\epsilon \nabla \cdot \mathbf{E} = ik\epsilon_0 E = \sum_j q_j n_{1j}$$

Combining the last two equations, separating out the electron term explicitly, and defining

$$\Omega_{pj} \equiv (n_{oj} Z_j^2 e^2 / \epsilon_0 m_j)^{1/2}$$

We obtain the dispersion relation

$$k^2 = \frac{\omega_p^2}{v_{the}^2} Z'(\zeta_e) + \sum_j \frac{\Omega_{pj}^2}{v_{thj}^2} Z'(\zeta_j)$$

Electron plasma waves can be obtained by setting $\Omega_{pj} = 0$ (infinitely massive ions). Defining

$$k_D^2 = 2\omega_p^2 / v_{the}^2 = \lambda_D^{-2}$$

we then obtain

$$k/k_D^2 = \frac{1}{2} Z'(\zeta_e)$$

which is the same as Equation 8.30 when f_{0e} is Maxwellian.

8.9.2 Ion Waves and Their Damping

To obtain ion waves, go back to the plasma dispersion relation and use the fact that their phase velocity ω/k is much smaller than v_{the}; hence ζ_e is small, and we can expand $Z(\zeta_e)$ in a power series:

$$Z(\zeta_e) = i\sqrt{\pi}e^{-\zeta_e^2} - 2\zeta_e(1 - \frac{2}{3}\zeta_e^2 + ...)$$

The imaginary term comes from the residue at a pole lying near the real s axis and represents electron Landau damping. For $\zeta_e \ll 1$, the derivative of the above gives

$$Z'(\zeta_e) = -2i\sqrt{\pi}\zeta_e e^{-\zeta_e^2} - 2 + ... \simeq -2$$

Electron Landau damping can usually be neglected in ion waves because the slope of $f_e(v)$ is small near its peak. Replacing $Z'(\zeta_e)$ by -2 in the dispersion relation gives

$$\lambda_D \sum_j \frac{\Omega_{pj}^2}{v_{thj}^2} Z'(\zeta_j) = 1 + k^2\lambda_D^2 \simeq 1$$

The term $k^2\lambda_D^2$ represents the deviation from quasineutrality. $(1/k \sim L, k^2\lambda_D^2 \sim \lambda_D^2/L^2 \ll 1$ where L is the system length scale.)

We now specialize tothe case of a single ion species. Since $n_{0e} = Z_i n_{0i}$, the coefficient in the equation above is

$$\lambda_D \frac{\Omega_p^2}{v_{thi}^2} = \frac{\epsilon k_B T_e}{n_{0e} e^2} \frac{n_{0i} Z^2 e^2}{\epsilon m_i} \frac{m_i}{2 k_B T_i} = \frac{1}{2} \frac{Z T_e}{T_i}$$

For $k_2\lambda_D^2 \ll 1$, the dispersion relation becomes

$$Z'(\frac{\omega}{kv_{thi}}) = \frac{2T_i}{ZT_e}$$

Solving this equation is a nontrivial problem. Suppose we take real k and complex ω to study damping in time. Then the real and imaginary parts of ω must be adjusted so that $\Im(Z') = 0$ and $\Re(Z') = 2T_i/ZT_e$. There are in general many possible roots ω that satisfy this, all of them having $\Im(\omega) < 0$. The least damped, dominant root is the one having the smallest $|\Im(\omega)|$. Damping in space is usually treated by taking ω real and k complex. Again we get a series of roots k with $\Im(k) > 0$, representing spatial damping. However, the dominant root does not correspond to the same value of ζ_i as in the complex ω case. It turns out that the spatial problem has to be treated with special attention to the excitation mechanism at the boundaries and with more careful treatment of the electron term $Z'(\zeta_e)$.

To obtain an analytic result, we consider the limit $\zeta_i \gg 1$, corresponding to large temperature ratio $\theta \equiv ZT_e/T_i$. (???) The asymptotic expression for $Z'(\zeta_i)$ is

$$Z'(\zeta_i) = -2i\sqrt{\pi}\zeta_i e^{-\zeta_i^2} + \zeta_i^{-2} + \frac{3}{2}\zeta_i^{-4} + ...$$

(I think this can be found from the plasma handbook; it can also be found here) If the damping is small, we cna neglect the Landau term in the first approximation. The equation becomes

$$\frac{1}{\zeta_i^2}\Big(1 + \frac{3}{2}\frac{1}{\zeta_i^2}\Big) = \frac{2}{\theta}$$

Since θ is assumed large, ζ_i^2 is large; and we can approximate ζ_i^2 by $\theta/2$ in the second term. Thus

$$\frac{1}{\zeta_i^2}\Big(1 + \frac{3}{\theta}\Big) = \frac{2}{\theta}, \quad \zeta_i^2 = \frac{3}{2} + \frac{\theta}{2}$$

or

$$\frac{\omega^2}{k^2} = \frac{2k_BT_i}{m_i}\Big(\frac{3}{2} + \frac{ZT_e}{2T_i}\Big) = \frac{Zk_BT_e + 3k_BT_i}{m_i}$$

This is the ion wave dispersion relation with $\gamma_i = 3$, generalized to arbitrary Z.

We now substitue the above approximations back into the dispersion relation retaining the Landau term:

$$
\begin{aligned}
\frac{1}{\zeta_i^2}\Big(1+\frac{3}{\theta}\Big) - 2i\sqrt{\pi}\zeta_i e^{-\zeta_i^2} &= \frac{2}{\theta} \\
\frac{1}{\zeta_i^2}\Big(1+\frac{3}{\theta}\Big) &= \frac{2}{\theta}(1+i\sqrt{\pi}\theta\zeta_i e^{-\zeta_i^2}) \\
\zeta_i^2 &= \Big(\frac{3+\theta}{2}\Big)^{1/2}(1+i\sqrt{\pi}\theta\zeta_i e^{-\zeta_i^2})^{-1}
\end{aligned}
$$

Expanding the square root, we have

$$
\zeta_i \simeq \Big(\frac{3+\theta}{2}\Big)^{1/2}\Big(1-\frac{1}{2}i\sqrt{\pi}\theta\zeta_i e^{-\zeta_i^2}\Big)
$$

The approximate damping rate is found by using the above approximation in the imaginary term:

$$
-\frac{\Im(\zeta_i)}{\Re(\zeta_i)} = \frac{\Im(\omega)}{\Re(\omega)} = \Big(\frac{\pi}{8}\Big)^{1/2}\theta(3+\theta)^{1/2}e^{-(3+\theta)/2}
$$

This asymptotic expression, accurate for large θ, shows an exponential decrease in damping with increasing θ. When θ falls below 10, this expression becomes inacuurate, and the damping must be computed from the original expression which employs the Z-function. For the experimentally interesting region $1 < \theta < 10$, the following simple formula is ananalytic fit to the exact solution:

$$
-\Im(\omega)/\Re(\omega) = 1.1\theta^{7/4}\exp(-\theta^2)
$$

What happens when collisions are added to ion Landau damping? Surprisingly little. Ion-electron collisions are weak because the ion and electron fluids move almost in unison, creating little friction between them. Ion-ion collisions (ion viscosity) can damp ion acoustic waves, but we know that sound waves in air can propagate well in spite of the dominance of collisions. Actually, collisions spoil the particle resonances that cause Landau damping, and one finds that the total damping is *less* than the Landau damping unless the collision rate is extremely large. In summary, ion Landau damping is almost always the dominant process with ion waves, and this varies exponentially with the ratio ZT_e/T_i.

8.10 Kinetic Effects in a Magnetic Field

When either the dc magnetic field $\mathbf{B}_0$ or the oscillating magnetic field $\mathbf{B}_1$ is finite, the $\mathbf{v} \times \mathbf{B}$ term in the Vlasov equation for a collisionless plasma must be included. The linearized equation is then replaced by

$$\frac{\partial f_1}{\partial t} + \mathbf{v} \cdot \nabla f_1 + \frac{q}{m}(\mathbf{v} \times \mathbf{B}_0) \cdot \frac{\partial f_1}{\partial \mathbf{v}} = -\frac{q}{m}(\mathbf{E}_1 + \mathbf{v} \times \mathbf{B}_1) \cdot \frac{\partial f_0}{\partial \mathbf{v}}$$

Resonant particles moving along $\mathbf{B}_0$ still cause Landau damping if $\omega/k \simeq v_{th}$, but two new kinetic effects now appera which are connected with the velocity component $\mathbf{v}_\perp$ perpendicular to $\mathbf{B}_0$. One of these is cyclotron damping, which will be discussed later; the other is the generation of cyclotron harmonics, leading to the possibility of the oscillation commonly called Bernstein waves.

Harmonics of the cyclotron frequency are generated when the particles' circular Larmor orbits are distorted by the wave fields $\mathbf{E}_1$ and $\mathbf{B}_1$. These finite-r_L effects are neglected in ordinary fluid theory but can be taken into account to order $k^2 r_L^2$ by including the viscosity. A kinetic treatment can be accurate even for $k^2 r_L^2 = \mathcal{O}(1)$. To understand how harmonics aris, consider the motion of a particle in an electric field:

$$\mathbf{E} = E_x e^{i(kx - \omega t)} \hat{\mathbf{x}}$$

The equation of motion is

$$\ddot{x} + \omega_c^2 x = \frac{1}{m} E_x e^{i(kx - \omega t)}$$

If kr_L is not small, the exponent varies from one side of the orbit to the other. We can approximate kx by substituting the undisturbed orbit $x = r_L \sin \omega_c t$

$$\ddot{x} + \omega_c^2 x = \frac{1}{m} E_x e^{i(kr_L \sin \omega_c t - \omega t)}$$

The generating function for the Bessel function $J_n(z)$ is

$$e^{z(t - 1/t)/2} = \sum_{n=-\infty}^{\infty} t^n J_n(z)$$

Letting $z = kr_L$ and $t = \exp(i\omega_c t)$, we obtain

$$e^{ikr_L \sin \omega_c t} = \sum_{n=-\infty}^{\infty} J_n(kr_L) e^{in\omega_c t}$$

$$\ddot{x} + \omega_c^2 x = \frac{q}{m} E_x \sum_{n=-\infty}^{\infty} J_n(kr_L) e^{i(\omega - n\omega_c)t}$$

The following solution can be verified by direct substitution:

$$x = \frac{q}{m} E_x \sum_{n=-\infty}^{\infty} \frac{J_n(kr_L) e^{i(\omega - n\omega_c)t}}{\omega_c^2 - (\omega - n\omega_c)^2}$$

This shows that the motion has frequency components differing from the driving frequency by multiples of ω_c, and that the amplitudes of these components are proportional to $J_n(kr_L)/[\omega_c^2 - (\omega - n\omega_c)^2]$. When the denominator vanishes, the amplitude becomes large. This happens when $\omega - n\omega_c = \pm\omega_c$, or $\omega = (n \pm 1)\omega_c$, $n = 0, \pm 1, \pm 2, ...$; that is, when the field $\mathbf{E}(x,t)$ resonates with any harmonic of ω_c. In the fluid limit $kr_L \to 0$, $J_n(kr_L)$ can be approximated by $(kr_L/2)^n/n!$, which approaches 0 for all n except $n = 0$. For $n = 0$, the coefficient becomes $(\omega_c^2 - \omega^2)^{-1}$, which is the fluid result containing only the fundamental cyclotron frequency.

8.10.1 The Hot Plasma Dielectric Tensor

After Fourier analysis of $f_1(\mathbf{r}, \mathbf{v}, t)$ in space and time, the linearized Vlasov equation can be solved for a Maxwellian distribution $f_0(\mathbf{v})$, and the resulting expression $f_1(\mathbf{k}, \mathbf{v}, \omega)$ can be used to calculate the density and current of each species. The result is usually expressed in the form of an equivalent dielectric tensor $\vec{\epsilon}$, such that the dispersion vector $\mathbf{D} = \vec{\epsilon} \cdot \mathbf{E}$ can be used in the Maxwell's equations to calculate dispersion relations for various waves. The algebra is horrendous and therefore omitted. We quote only a restricted result valid for nonrelativistic plasmas with isotropic pressure $T_\perp = T_\parallel$ and no zero-order drifts $\mathbf{v}_{0j}$; these restrictions are easily removed, but he general formulas are too cluttered for our purposes. We further assume $\mathbf{k} = k_x \hat{\mathbf{x}} + k_z \hat{\mathbf{z}}$, with $\hat{\mathbf{z}}$ being the direction of $\mathbf{B}_0$; no generality is lost by setting k_y equal to zero, since the plasma is isotropic in the plane perpendicular to $\mathbf{B}_0$. The elements of $\vec{\epsilon}_R = \vec{\epsilon}/\epsilon_0$ are then

$$\epsilon_{xx} =$$
$$\epsilon_{yy} =$$
$$\epsilon_{zz} =$$
$$\epsilon_{xz} =$$
$$\epsilon_{yz} =$$
$$\epsilon_{zx} =$$

where $Z(\zeta)$ is the plasma dispersion function, $I_n(b)$ is the nth order Bessel function of imaginary argument, and the other symbols are defined by

$$\begin{aligned}
\omega_{ps}^2 &= n_{0s} Z_s^2 e^2/\epsilon_0 m_s \\
\zeta_s &= (\omega + n\omega_{cs})/k_z v_{ths} \\
\omega_{cs} &= |Z_s e B_0/m_s| \\
v_{ths}^2 &= 2k_B T_s/m_s \\
b_s &= \frac{1}{2}k_\perp^2 r_{Ls} = k_x^2 k_B T_s/m_s\omega_{cs}^2
\end{aligned}$$

The first sum is over species s, with the understanding that ω_p, b, ζ_0, and ζ_n all depend on s, and that the $\pm$ stands for the sign of the charge. The second sum is over the harmonic number n. The primes indicate differentiation with respect to the argument.

As foreseen, there appera Bessel functions of finite-r_L parameter b. (The change from $J_n(b)$ to $I_n(b)$ occurs in the integration over velocities.) The elements of $\vec{\epsilon}$ involving motion along $\hat{\mathbf{z}}$ contain $Z'(\zeta_n)$, which gives rise to Landau damping when $n = 0$ and $\omega/k_z \simeq v_{th}$. The $n \neq 0$ terms now make possible another collisionless damping mechanism, cyclotron damping, which occurs when $(\omega_c \pm n\omega_c)/k_z \simeq v_{th}$.

8.10.2 Bernstein Waves

Electrostatic waves propagating at right angles to $\mathbf{B}_0$ at harmonics of the cyclotron frequency are called Bernstein waves. The dispersion relation can be found by using the dielectric elements give in the previous section in Poisson's equation $\nabla \cdot \vec{\epsilon} \cdot \mathbf{E} = 0$. If we assume electrostatic perturbations such that $\mathbf{E}_1 = -\nabla\phi_0$, and consider waves of the form $\phi_1 = \phi_1 \exp i(\mathbf{k}\cdot\mathbf{r} - \omega t)$, Poisson's equation can be written

$$k_x^2\epsilon_{xx} + 2k_x k_z \epsilon_{xz} + k_z^2\epsilon_{zz} = 0$$

Note that we have chosen a coordinate system that has $\mathbf{k}$ lying in the x-z plane, so that $k_y = 0$. We next substitute for $\epsilon_{xx}, \epsilon_{xz}$, and ϵ_{zz} from the dielectric tensor expression and express $Z'(\zeta_n)$ in terms of $Z(\zeta_n)$ with the identity

$$Z'(\zeta_n) = -2[1 + \zeta Z(\zeta_n)]$$

via integration by parts. The equation becomes

$$
\begin{aligned}
k_x^2 + k_z^2 + &\sum_s \frac{\omega_p^2}{\omega^2} e^{-b} \zeta_0 \sum_{n=-\infty}^{\infty} I_n(b) \\
&\times [k_x^2 \frac{n^2}{b} Z - 2(\frac{2}{b})^{1/2} n k_x k_z (1 + \zeta_n Z) - 2k_z^2 \zeta_n (1 + \zeta_n Z)] = 0
\end{aligned}
$$

The expression in the square brackets can be simplified in a few algebraic steps to $2k_z^2[\zeta_{-n} + \zeta_0^2 Z(\zeta_n)]$ by using the definitions $b = k_x^2 v_{th}^2 / 2\omega_c^2$ and $\zeta_n = (\omega + n\omega_c)/k_z v_{th}$. Further noticing that $2k_z^2 \omega_p^2 \zeta_0 / \omega^2 = 2\omega_p^2 / v_{th}^2 \equiv k_D^2$ for each species, we can write the equation as

$$
k_x^2 + k_z^2 + \sum_s k_D^2 e^{-b} \sum_{n=-\infty}^{\infty} I_n(b) [\zeta_{-n}/\zeta_0 + \zeta_0 Z(\zeta_n)] = 0
$$

The term $\zeta_{-n}/\zeta_0 = 1 - n\omega_c/\omega$. Since $I_n(b) = I_{-n}(b)$, the term $I_n(b) n\omega_c/\omega$ sums to zero when n goes from $-\infty$ to ∞; hence, ζ_{-n}/ζ_0 can be replaced by 1. Defining $k^2 = k_x^2 + k_z^2$, we obtain the general dispersion relation for Bernstein waves:

$$
1 + \sum_s \frac{k_D^2}{k_\perp^2} e^{-b} \sum_{n=-\infty}^{\infty} I_n(b) [1 + \zeta_0 Z(\zeta_n)] = 0
$$

(A) *Electron Bernstein Waves.* Let us first consider high-frequency waves in which the ions do not move. These waves are not sensitive to small deviations from perpendicular propagation, and we may set $k_z = 0$, so that $\zeta_n \to \infty$. There is, therefore, no cyclotron damping; the gaps in the spectrum that we shall find are not caused by such damping. For large ζ_n, we may replace $Z(\zeta_n)$ by $-1/\zeta_n$. (???) The $n = 0$ term in teh second sum of the above equation then cancels out, and we can divide the sum into two sums, as follows:

$$
k_\perp^2 + \sum_s k_D^2 e^{-b} \Big[\sum_{n=1}^{\infty} I_n(b)(1 - \zeta_0/\zeta_n) + \sum_{n=1}^{\infty} I_{-n}(b)(1 - \zeta_0/\zeta_{-n}) \Big] = 0
$$

or

$$
k_\perp^2 + \sum_s k_D^2 e^{-b} \sum_{n=1}^{\infty} I_n(b) \Big[2 - \frac{\omega}{\omega + n\omega_c} - \frac{\omega}{\omega - n\omega_c} \Big] = 0
$$

The bracket collapses to a single term upon combining over a common denominator:

$$
1 = \sum_s \frac{k_D^2}{k_\perp^2} e^{-b} \sum_{n=1}^{\infty} I_n(b) \frac{2n^2 \omega_c^2}{\omega^2 - n^2 \omega_c^2}
$$

Using the definitions of k_D and b, one obtains the well-known (NOT TO ME!!!) $k_z = 0$ dispersion relation

$$1 = \sum_s \frac{\omega_p^2}{\omega_c^2} \frac{2}{b} e^{-b} \sum_{n=1}^{\infty} \frac{I_n(b)}{(\omega/n\omega_c)^2 - 1}$$

We now specialize to the case of electron oscillations. Dropping the sum over species, we obtain

$$\frac{k_\perp^2}{k_D^2} = 2\omega_c^2 \sum_{n=1}^{\infty} \frac{e^{-b} I_n(b)}{\omega^2 - n\omega_c^2} n^2 \equiv \alpha(\omega, b)$$

... (ADD fig.7-33!!!)

To obtain the fluid limit, we repalce $I_n(b)$ by its small-b value $(b/2)^n/n!$. Only the $n = 1$ term remains in the limit $b \to 0$, and we obtain

$$1 = \frac{\omega_p^2}{\omega_c^2} \frac{2}{b} \frac{b}{2} \left(\frac{\omega^2}{\omega_c^2} - 1 \right)^{-1} = \frac{\omega_p^2}{\omega^2 - \omega_c^2}$$

or $\omega^2 = \omega_p^2 + \omega_c^2 = \omega_h^2$, which is the upper hybrid oscillation. As $k_\perp \to 0$, this frequency must be one of the roots. If ω_h falls between two high harmonics of ω_c, the shape of the $\omega - k$ curves changes near $\omega = \omega_h$ to allow this to occur. ...

(B) *Ion Bernstein Waves.* In the case of waves at ion cyclotron harmonics, one has to distinguish between *pure* ion Bernstein waves, for which $k_z = 0$, and *neutralized* ion Bernstein waves, for which k_z has a small but finite value. The difference, as we have seen earlier for lower hybrid oscillations, is that finite k_z allows electrons to flow along $\mathbf{B}_0$ to cancel charge separations. Though the $k_z = 0$ case has already been treated in ...

9 Stability

This chapter discusses plasma equilibrium and stability, or more precisely, instability. We are interested in when, where, and how instabilities arise in plasmas.

Magnetic fields are ubiquitous in cosmological objects. General questions exist about why are they of the form that they are:

- Why does this particular form remain?
- How does it get to this form?

Anything that reorganizes magnetic field is a transport phenomenon therefore instabilities are of great interest. From hydrodynamic instabilities to plasma instabilities, we need to realize the role of magnetic field and kinetic physics in the instability process:

- Existing hydro-instability may be affected by the presence of magnetic field
- Instability may be driven by presence of magnetic fields
- Instability that does not exist in hydrodynamics may arise in the kinetic description.

There exists a massive catalogue of plasma instabilities.

Type	Description
Beam acoustic instability	
Bump-on-tail instability	
Buneman instability	
Cherenkov instability	
Chute instability	
Coalescence instability	
Collapse instability	
Counter-streaming instability	
Cyclotron instabilities	Alfvén, electron, electrostatic, ion, magnetoacoustic
Disruptive instability	
Double mission instability	
Drift instability	
Electrothermal instability	
Fan instability	
Filamentation instability	
Firehose instability	Section 10.1

Type	Description
Free electron maser instability	
Gyrotron instability	
Helical instability	Helix
Interchange instability	Rayleigh-Taylor, flute, ballooning, kink, sausage
Ion beam instability	
Lower hybrid drift instability	Section 9.9
Magnetic drift instability	
Modulation instability	
Non-Abelian instability	Chromo-Weibel
Pair instability	
Parker instability	Magnetic buoyancy
Peratt instability	
Pinch instability	
Tearing mode instability	Section 9.7.5
Two stream instability	Kelvin-Helmholtz, Section 9.5
Weak beam instability	
Weibel instability	
Z-pinch instability	Bennett pinck

The basic methodology of examining instabilities

- Take the equations
- Linearize about an equilibrium solution
- Add some perturbations and see what happens
  - Look for *normal mode* solutions $e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)}$ and find a relationship between the growth rate $\Im(\omega)$ and the wavenumbers of the disturbance $\mathbf{k}$ and the parameters of the problem (i.e. dispersion relation).
  - Use the *MHD energy principle*: calculate δW, the change in potential energy associated wit the disturbance, and look for disturbances with $\delta W < 0$ (the existence of any one means unstable; need $\delta W > 0 \forall$ types of disturbances).

9.1 Analogy of the Energy Principle

First, let us have an intuitive inspection from the energy principle. In classical mechanics, we learn that motion acts to lower the energy in the system. For the simplest Rayleigh-Taylor instability between two fluids with different density, we can easily decide if there is an instability by checking the total potential energy change. In a current-carrying plasma, similar motions can be developed. The energy can be expressed as

$$W = \frac{1}{2}LI^2$$

where L is the inductance and I is the current. The magnetic flux can be written as

$$\Phi = LI$$

so the energy can also be written as

$$W = \frac{\Phi^2}{2L}$$

The relation between magnetic flux and current is similar to charge and voltage, $Q = CV$, where C is the electric capacity. In a system where magnetic flux Φ is conserved (e.g. ideal MHD), plasma lowers its potential by increasing its inductance L. Check (Bellan 2008) P315.

1. current loop, hoop force, increase area;
2. sausage mode;
3. current wire into helix.

9.2 Implication of Single Particle Motion on Instabilities

This section provides some qualitative understanding of instabilities in plasma. Thinking of instabilities from the prospective of single particle motion provides us the physical intuition of the causes and development of instabilities from a very basic level.

9.2.1 Equilibrium stability of a plasma from drifts

A key problem in nuclear fusion is to confine plasma. In the first stage, we need current loops around a torus tube (ADD FIGURE!) to provide a toroidal magnetic field $\mathbf{B}_t$ along the torus. Let us take a look at a torus cut. (ADD FIGURE!) ∇B_0 is pointing towards the inside. $\mathbf{B}_0$ points outside the plane, so the gradient-curvature drift will lead electrons upward and ions downward, which in turn creates a E field pointing upwards. $\mathbf{E}\times\mathbf{B}$ drift then will lead both electrons and ions towards the outer boundary. Eventually we will lose the plasma. One way to fix this is to add $\mathbf{J}_t$, an internal plasma current in the toroidal direction to generate a poloidal $\mathbf{B}_p$, so that the magnetic field is stronger on the boundary than that at the center. In this way the total magnetic field becomes a helix.

Consider the poloidal magnetic field generated by $\mathbf{J}_t$. One step further, our question is: will the plasma tube be stable under infinitesimal perturbations to the ideal cylinder configuration? The answer is no. Imagine a small perturbation shown as in Figure 9.1 b. ∇B points from weak B to strong B regions. On the convex side, the gradient-curvature drift will lead ions to the left and electrons to the right, which in turn generates an electric field pointing left to right. Thereafter the $\mathbf{E} \times \mathbf{B}$ drift will further drag the plasma to the convex region, and the whole system can never return to equilibrium. This the current-carrying plasma instability is called *kink instability*. The situation described here is sometimes referred to as *linear kink instability*. The kink mode can carry currents. Another similar mode is the *sausage instability* as shown in Figure 9.1. Also note that a sufficiently strong B_z (not poloidal/toroidal component) can stabilize these instabilities.

Another famous instability is the Rayleigh-Taylor instability. In fluid dynamics, Rayleigh-Taylor instability happens due to gravity. Here in plasma physics, the role of gravity force is replaced by the electromagnetic force. Imagine a situation where plasma are located at $z > 0$ region, below which is a vacuum space. There is a B field pointing outside the plane while ∇B is pointing upwards ($B_{\text{up}} > B_{\text{down}}$). Due to gradient drift on the boundary, ions will move to the left while electrons will move to the right, where a E field pointing left to right is created. Thus the $\mathbf{E} \times \mathbf{B}$ drift will lead plasma from upper region to lower region, and eventually breaks the interface. (Actually I have some questions for this figure: it seems to me that it is impossible to decide which part of the interface is changing first?)

9.2.2 Stability of magnetic mirror in the scope of single particle motion

We can deduce the stability of magnetic mirror by assuming a initial small perturbation along the boundary. In the center cross-section cut, first there is a centrifugal force pointing outward, which will cause electrons drifting to one way and ions drifting to the other way. The charge separation will generate an electric field. The $\mathbf{E} \times \mathbf{B}$ drift will then pull the plasma further out if there's a ripple, which will lead to instability. Several names describe the same thing: flute instability, R-T instability, interchangeable instability, gravitational instability and so on. This instability propagates at Alfvén speed.

In general, we can define two configuration categories: the unstable situation, where $\mathbf{B}$ has a "bad" unfavorable curvature, and the stable situation, where $\mathbf{B}$ has a "good" favorable curvature. This depends on whether or not the magnetic pressure in the vacuum is stronger than that inside the plasma. In a basic magnetic mirror, the plasma on the center boundaries are unstable, while those around the coil curvature are stable.

A famous Russian scientist Ioffe introduced conducting bars around the mirror to create an *absolute minimum B-geometry*, where in any point away from the center the B field is stronger. This indeed supresses the R-T instability, but the whole system is, unfortunately, still unstable due to micro-instabilities caused by lost cone distribution. The inverted population (more high speed particles than low speed ones) will lead to instabilities. Later scientists came up with the idea of adding cold gas to modify the distribution, but the cold gas injection procedure eventually kills the mirror configuration.

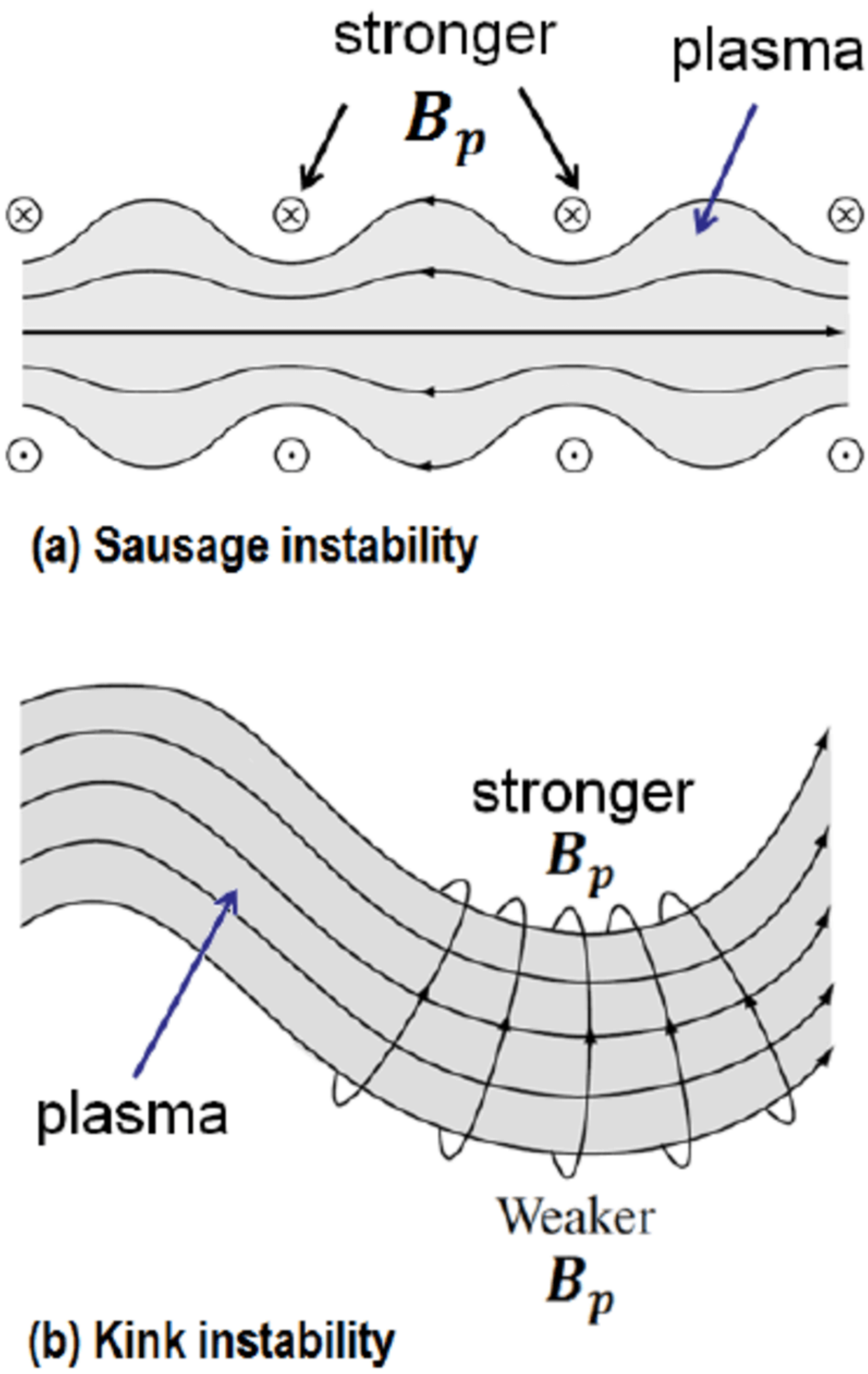

Figure 9.1: Sausage and kink modes.

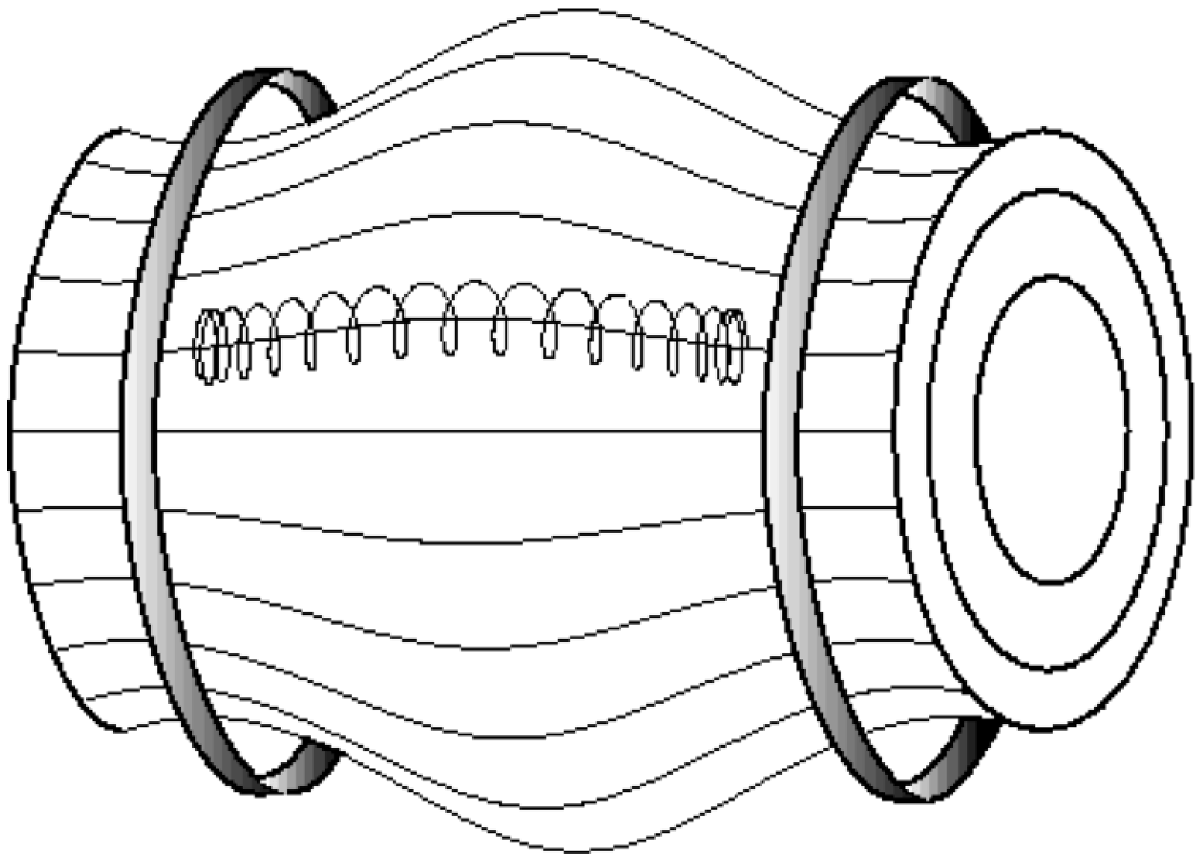

Figure 9.2: Basic magnetic mirror configuration.

9.2.3 Stability in the Tokamak

In a classical Tokamak geometry, the poloidal and toroidal magnetic field together created a spiral around the torus.

$$\mathbf{B} = \mathbf{B}_T + \mathbf{B}_P$$
$$B = \sqrt{{B_T}^2 + {B_P}^2} \approx B_T$$

The field strength goes down as R increases, which implies that the inner semi-tube is in good curvature and the outer semi-tube is in bad curvature. By plotting mechanical potential=μB along the B-line as a function of θ, we can see that there are bumps and valleys. Particles with low $v_\parallel$ are trapped, and those with large $v_\parallel$ are transit particles. Tokamak is an *average minimum B geometry*, because particles spend longer time on hills (stable region) and less time in valleys (unstable region). This geometry is not as robust as Ioffe bar magnetic mirror, since some particles with small $v_\parallel$ are always trapped in bad curvature region.

See more in (F. F. Chen 2016), Third Edition, Chapter Application of Plasmas.

9.3 Two-Stream Instability

The general procedure of obtaining the dispersion relation for electrostatic wave is:

1. Get the linear equations from governing equations.
2. Get the relation between linearized velocity and perturbed electric field.
3. Get the relation between linearized density and perturbed electric field.
4. Get the current response to the perturbed electric field.
5. Get the dielectric tensor from Maxwell's equation. Let $|\boldsymbol{\epsilon}| = 0$. we finally obtain the dispersion relation. For an isotropic case, the dielectric tensor shrinks to a scalar, so we simply have $\epsilon = 0$.

Assume the simple equilibrium state in 1D (static and "cold" ions, "cold" electrons): $m_i = \infty, E_0 = B_0 = 0, p_i = p_e = 0, n_i = n_0, T_e = T_i = 0$. Whenever we say "cold" for plasma, it does not mean that the plasma is at absolute zero degree. This only means that we are considering a situation where the kinetic pressure plays no roles in the dispersion relation. This is also a *non-magnetized plasma* because $\mathbf{B_0} = 0$. The variables including perturbations are

$$\begin{aligned}
n_e &= n_0 + n_1 \quad n_0 \gg n_1 \\
v_e &= v_0 + v_1 \quad v_0 \gg v_1 \\
E &= \cancel{E_0} + E_1 \\
n_1(x,t) &= \tilde{n}_1 e^{-i\omega t + ikx}
\end{aligned}$$

The electron continuity and the momentum equations read

$$\frac{\partial}{\partial t}(n_0+n_1)+\frac{\partial}{\partial x}[(n_0+n_1)(v_0+v_1)]=0$$
$$\frac{\partial}{\partial t}(v_0+v_1)+(v_0+v_1)\frac{\partial(v_0+v_1)}{\partial x}=-\frac{e}{m_e}[(E_0+E_1)+(\mathbf{v}_0+\mathbf{v}_1)\times(\mathbf{B}_0+\mathbf{B}_1)_x]$$

Again, for electrostatic waves, $\mathbf{B}_1=0$. Neglecting high order terms, we get the linearized equations

$$\frac{\partial n_1}{\partial t}+\frac{\partial}{\partial x}(n_0v_1)+\frac{\partial}{\partial x}(n_1v_0)=0$$
$$\frac{\partial v_1}{\partial t}+v_0\frac{\partial v_1}{\partial x}=-\frac{e}{m_e}E_1$$

Assume wave-like perturbations $e^{ikx-i\omega t}$ as in the Vlasov theory, from the linearized momentum equation we have

$$(-i\omega+ikv_0)v_1=-\frac{e}{m_e}E_1$$
$$\Rightarrow v_1=\frac{\frac{e}{m_e}E_1}{i(\omega-kv_0)}.$$

Substituting into the linear continuity equation, we get

$$-i\omega n_1+ikn_0v_1+ikn_1v_0=0$$
$$\Rightarrow n_1=\frac{kn_0v_1}{\omega-kv_0}=\frac{kn_0}{\omega-kv_0}\frac{\frac{e}{m_e}E_1}{i(\omega-kv_0)}.$$

This is the density perturbation in response to electric perturbation E_1 in 2-fluid theory.

Then we can use the Poisson's equation to generate the dielectric function

$$\nabla\cdot(\epsilon_0\mathbf{E}_1)+en_1\equiv\nabla\cdot(\epsilon\mathbf{E}_1)=0$$
$$ik\epsilon_0\tilde{E}_1+e\tilde{n}_1=ik\epsilon\tilde{E}_1$$
$$\Rightarrow\epsilon=\epsilon_0\Big[1-\frac{{\omega_{pe}}^2}{(\omega-kv_0)^2}\Big]$$

which is the same as the result given by Vlasov theory (Y.Y. Problem Set #6.2). The advantage of using 2-fluid method is that we do not need to consider velocity space, which simplifies the derivation.

If we have two streams of electron described by $g(v)$ as

$$g(v)=\frac{1}{2}[\delta(v-v_1)+\delta(v-v_2)]$$

with the oscillation frequency ω_{p1}, ω_{p2} and number density $n_{p1,1}, n_{p2,1}$ respectively. In consideration of linear superposition property, we expect the dielectric function to be

$$\frac{\epsilon}{\epsilon_0} = 1 - \frac{{\omega_{p1}}^2}{(\omega - kv_1)^2} - \frac{{\omega_{p2}}^2}{(\omega - kv_2)^2}$$

If $g(v)$ is a continuous distribution in general, $g(v) = \sum_j g_j(v)$, then

$$\begin{aligned}
\frac{\epsilon}{\epsilon_0} &= 1 - \int_{-\infty}^{\infty} \sum_j \frac{{\omega_{p,j}}^2 g_j(v)\mathrm{d}v}{(\omega - kv)^2} \\
&= 1 - \frac{{\omega_{pe}}^2}{k^2} \int_{-\infty}^{\infty} \frac{g(v)\mathrm{d}v}{(v - \omega/k)^2}.
\end{aligned}$$

Note that $\delta'(x) = x^{-1}\delta(x)$. Here we reconstruct the result of Vlasov theory from 2-fluid theory. The equivalence of the two approaches is explored more thoroughly in later section Fluid Descriptions of Kinetic Modes (ADD LINK!).

Then what happens if ion motion is included? We still have "cold" ions at rest in equilibrium but now with ion perturbations in density. The Poisson's equation should include ion density perturbation

$$\begin{aligned}
\nabla \cdot (\epsilon_0 \mathbf{E}_1) &= \sum_{j=i,e} q_j n_{1j} \\
\Rightarrow \frac{\epsilon}{\epsilon_0} &= 1 - \frac{{\omega_{pe}}^2}{k^2} \int_{-\infty}^{\infty} \mathrm{d}v \frac{\partial g_e(v)/\partial v}{v - \omega/k} - \frac{{\omega_{pi}}^2}{k^2} \int_{-\infty}^{\infty} \mathrm{d}v \frac{\partial g_i(v)/\partial v}{v - \omega/k}
\end{aligned}$$

In the simplest equilibrium case, $g_e(v) = \delta(v - v_0),\ g_i(v) = \delta(v)$

$$\frac{\epsilon}{\epsilon_0} = 1 - \frac{{\omega_{pe}}^2}{(\omega - kv_0)^2} - \frac{{\omega_{pi}}^2}{\omega^2}$$

Let $\epsilon = 0$, we get the dispersion relation $\omega = \omega(k)$. An example dispersion relation and dielectric function property are shown in Figure 9.3 and Figure 9.4, respectively. Note that if you have a real wave number k, you will get a pair of conjugate ω, one of which that lies between 0 and kv_0 is an unstable mode. This will exhibit 2-stream instability as shown by the velocity space distribution in Figure 9.5.

$$\begin{aligned}
\epsilon(\omega, k)/\epsilon_0 = 1 - \frac{{\omega_{pe}}^2}{(\omega - kv_0)^2} - \frac{{\omega_{pi}}^2}{\omega^2} &= 0 \\
1 = \frac{{\omega_{pe}}^2}{(\omega - kv_0)^2} + \frac{{\omega_{pi}}^2}{\omega^2}
\end{aligned}$$

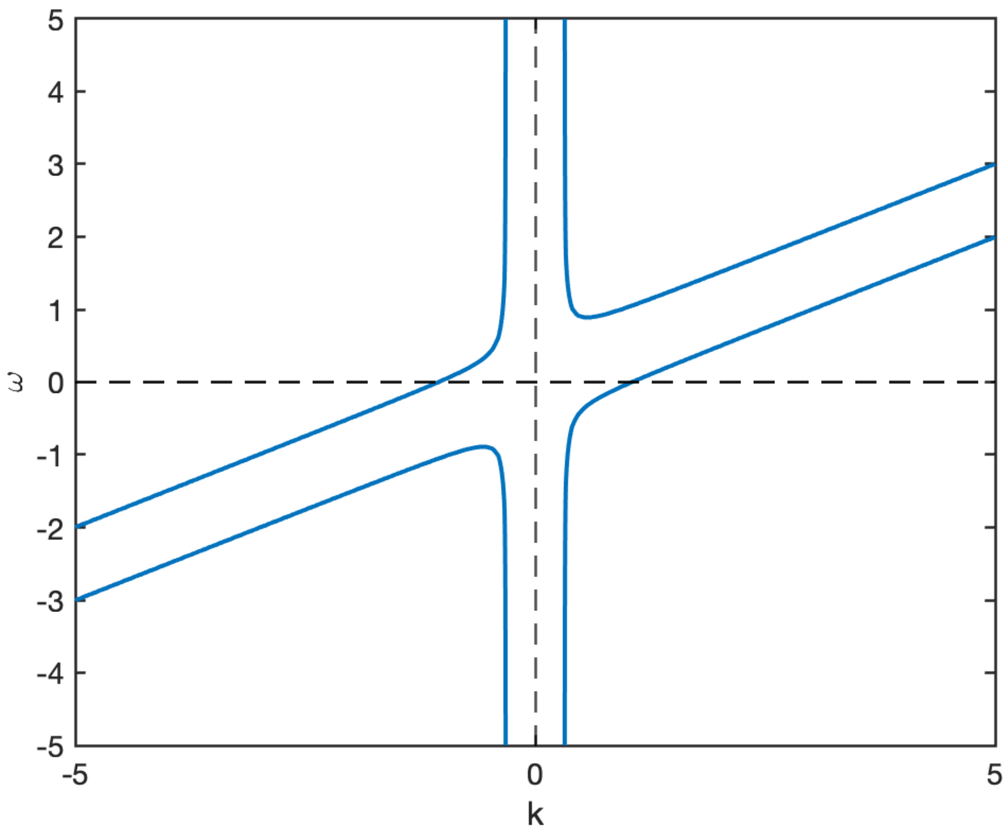

Figure 9.3: 2-stream dispersion relation.

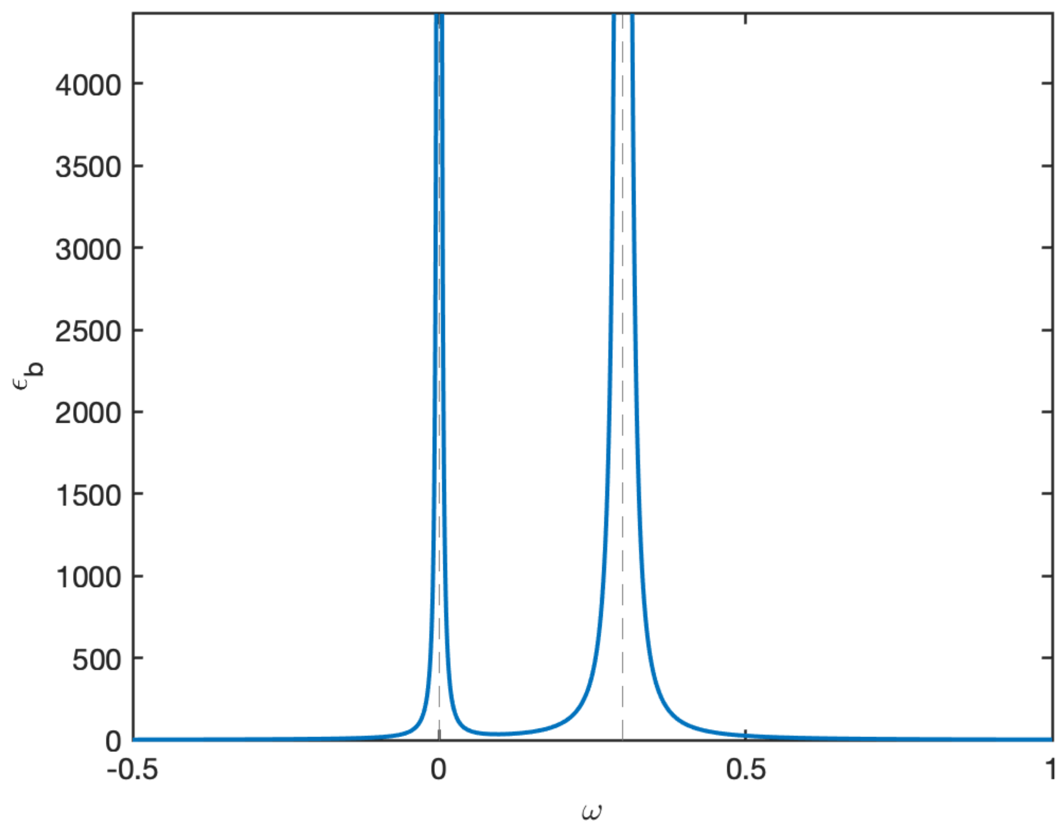

Figure 9.4: 2-stream dielectric function.

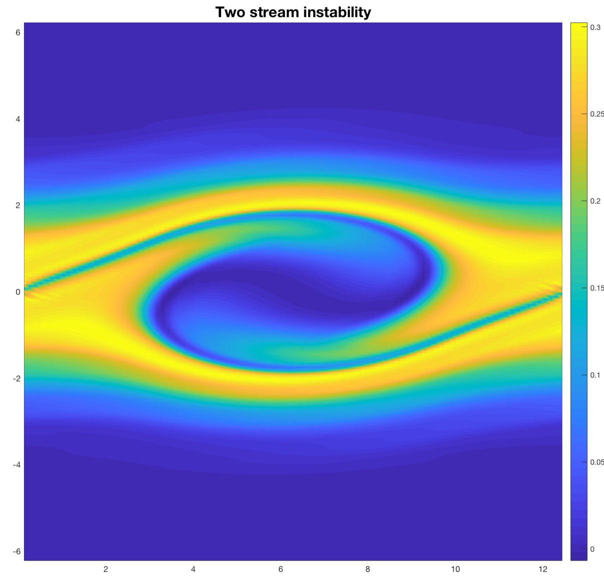

Figure 9.5: 2-stream velocity space distribution.

Let $\omega/\omega_{pe} = z$, $\frac{kv_0}{\omega_{pe}} = \lambda$, such that z and λ are dimensionless numbers. Let the right-hand side be $f(z)$, then

$$f(z) = \frac{1}{(z-\lambda)^2} + \frac{{\omega_{pi}}^2/{\omega_{pe}}^2}{z^2} = \frac{z^2 + (z-\lambda)^2\omega_{pi}^2/{\omega_{pe}}^2}{(z-\lambda)^2 z^2}$$

We can plot f to find the threshold of k when the instability will happen. See Y.Y's Problem Set 6.3 for details.

9.4 Rayleigh-Taylor Instability

The Rayleigh-Taylor instability is probably the most important MHD instability. It is also called *gravitational instability* , *flute instability* or more generally, *interchange instability*. In ordinary hydrodynamics, a Rayleigh-Taylor instability arises when one attempts to support a heavy fluid on top of a light fluid: the interface becomes "rippled", allowing the heavy fluid to fall through the light fluid. In plasmas, a Rayleigh-Taylor instability can occur when a dense plasma is supported against gravity by the pressure of a magnetic field.

The situation would not be of much interest or relevance in its own right, since actual gravitational forces are rarely of much importance in plasmas. However, in curved magnetic fields, the centrifugal force on the plasma due to particle motion along the curved field-lines acts like a "gravitational" force (Section 4.1.3). For this reason, the analysis of the Rayleigh-Taylor instability provides useful insight as to the stability properties of plasmas in curved magnetic fields. Rayleigh-Taylor-like instabilities driven by actual field curvature are the most virulent type of MHD instability in non-uniform plasmas.

Here we use a 2-fluid model and a so-called "diffuse boundary" model (F. F. Chen 2016) to describe it mathematically. Recall the structure of magnetic mirror: we have curved magnetic field lines and high density plasma at the center. From the discussion in Section 9.2.2, we know that the central part of magnetic mirror is unstable for Rayleigh-Taylor instability because of centrifugal force. Let us simplify the scenario and study the problem in Cartesian coordinates. The centrifugal force is irrelevant to particle charge and proportional to particle mass, so both ions and electrons have the same acceleration due to it. Let us replace the centrifugal force with gravity $\mathbf{g}$. In Figure 9.6, there are high density plasma on top and low density plasma on the bottom, with a distribution $\partial n_o/\partial x < 0$.

9.4.1 2-Fluid Diffuse Boundary Model

This section is similar to Section 6.7 in (F. F. Chen 2016).

The continuity and momentum equations are:

$$
\begin{aligned}
\frac{\partial n_j}{\partial t} + \nabla\cdot(n_j\mathbf{v_j}) = 0 \\
\frac{\partial \mathbf{v_j}}{\partial t} + \mathbf{v_j}\cdot\nabla\mathbf{v_j} = \frac{q_j}{m_j}(\mathbf{E}+\mathbf{v_j}\times\mathbf{B}) - \frac{\nabla P_j}{n_j m_j} + \mathbf{g}
\end{aligned}
$$

where $j = e^-, i^+$ for electrons and ions.

Assume an one-dimensional case

$$
\begin{aligned}
T_e = T_i = 0 \Rightarrow P_e = n_e k_B T_e = 0,\ P_i = n_i k_B T_i = 0 \\
n_0 = n_0(x), \frac{\partial n_0}{\partial x} < 0 \text{ (nonuniform plasma)} \\
\mathbf{g} = g\hat{x}, g = \text{const.} > 0 \\
\mathbf{B}_0 = B_0\hat{z},\ B_0 = \text{const.},\ \mathbf{E}_0 = 0 \text{ (no gradient/curvature drift)}
\end{aligned}
$$

Note that there is no diamagnetic current if $P_e = P_i = 0$ (no electric field so no current along $\mathbf{B}_0$?):

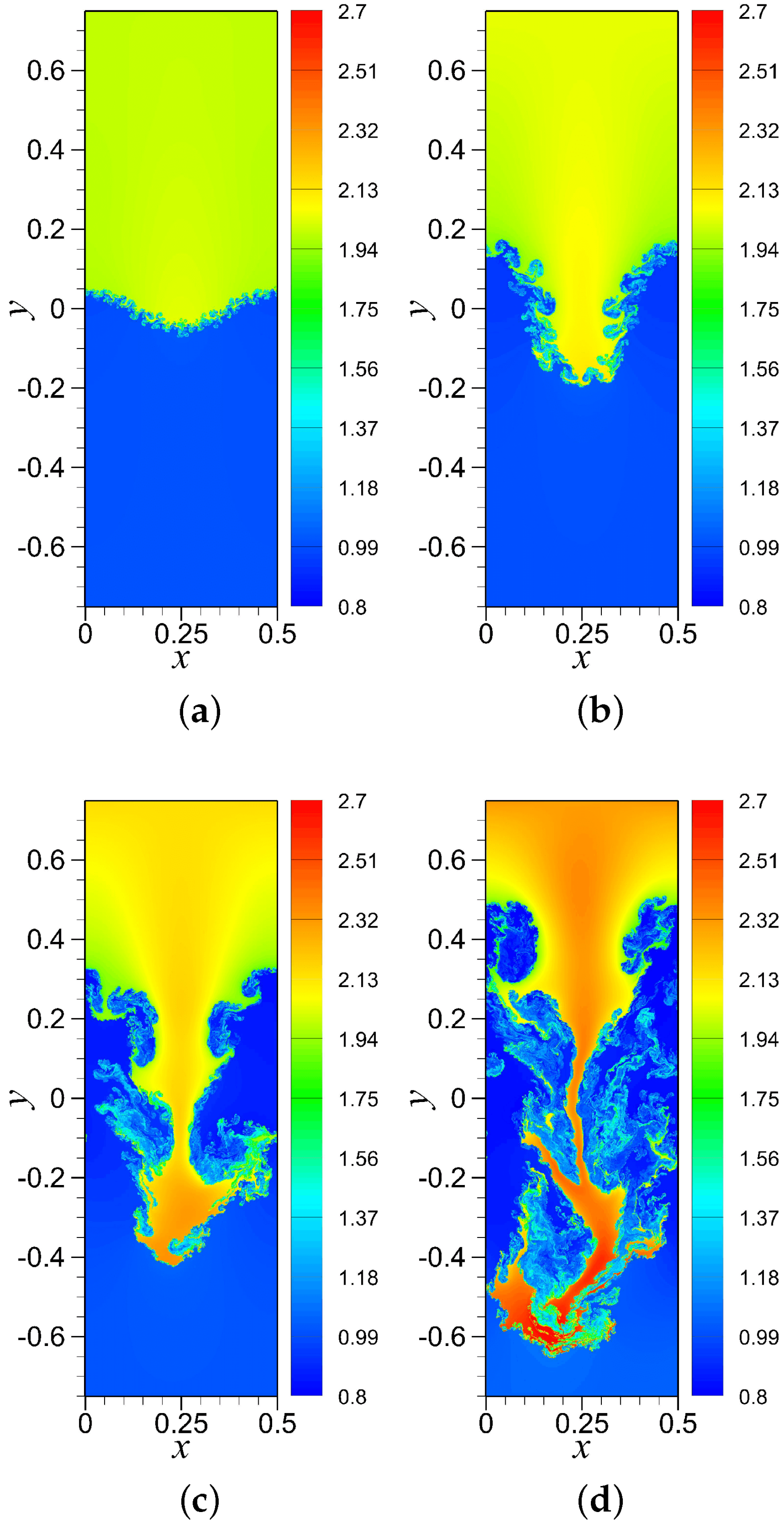

Figure 9.6: Simultaion of Rayleigh-Taylor instability (Rahman and San 2019).

$$
\begin{aligned}
&\mathbf{J}_{de} \times \mathbf{B}_0 = -\nabla P_e = 0 \\
&\mathbf{J}_{di} \times \mathbf{B}_0 = -\nabla P_i = 0 \\
\Rightarrow &\mathbf{J}_{de} = \mathbf{J}_{di} = 0
\end{aligned}
$$

Instability arises when an equilibrium state is violated. What is the force that balances the gravity? It turns out to be the Lorentz force $\mathbf{v} \times \mathbf{B}$ term: the separation of electrons and ions creates currents, and currents lead to force.

ADD FIGURE!

In equilibrium, $\frac{\partial}{\partial t} = 0$, $\frac{\partial}{\partial y}[n_{0j}(x)v_{0j}] = 0$, $v_{0j} = \text{const.}$,

$$
\begin{aligned}
&\frac{\partial n_{oj}}{\partial t} + \nabla \cdot (n_{oj}\mathbf{v}_{oj}) = 0 \\
&\frac{q_j}{m_j}\mathbf{v}_j \times \mathbf{B}_0 + \mathbf{g} = 0 \\
\Rightarrow &\begin{cases} \mathbf{v}_i = \frac{gm_i}{q_i B_0}(-\hat{y}) = -\frac{g}{\Omega_i}\hat{y} = -\hat{y}v_{0i} \\ \mathbf{v}_e = \frac{gm_e}{q_e B_0}(\hat{y}) = \frac{g}{\Omega_e}\hat{y} = \hat{y}v_{0e} \approx 0 (v_{0e} \ll v_{0i}) \end{cases}
\end{aligned}
$$

where $\Omega_i,\ \Omega_e$ are the gyro-frequency for ions and electrons respectively.

Now, we introduce an electrostatic perturbation on this equilibrium state ($\mathbf{B}_1(\mathbf{x}, t) = 0$, $\nabla \times \mathbf{E}_1 = -\frac{\partial \mathbf{B}_1}{\partial t} = i\omega \mathbf{B}_1 = 0$, $\mathbf{E}_1$ can be written as a gradient of a scalar potential)

$$
\mathbf{E}_1(\mathbf{x}, t) = -\nabla \Phi_1 = -\nabla [\phi_1(x) e^{ik_y y - i\omega t}]
$$

In addition, we adopt the so-called "local approximation", i.e. we assume $\partial \phi_1/\partial x = 0$, $\frac{\partial}{\partial x}[E_1, \mathbf{v}_1, n_1] = 0$. This is a very drastic assumption that greatly simplifies the problem but cannot be justified. This assumption is commonly used in many textbooks, both explicitly and implicitly (e.g. (Bellan 2008) used this to treat universal instability. Remember in solving the Vlasov equations, we integrate along the unperturbed orbits, which also requires this assumption.)

In this case,

$$
\mathbf{E}_1 = 0\hat{x} + E_{1y}\hat{y} = \hat{y}\widetilde{E}_{1y} e^{ik_y y - i\omega t}
$$

where $\widetilde{E}_{1y} = -ik_y \phi_1$ is a constant.

Linearization:

$$
\begin{aligned}
&\frac{\partial}{\partial t}(n_0+n_1)+\nabla\cdot[(n_0+n_1)(\mathbf{v}_0+\mathbf{v}_1)]=0\\
&\frac{\partial}{\partial t}(\mathbf{v}_0+\mathbf{v}_1)+(\mathbf{v}_0+\mathbf{v}_1)\cdot\nabla(\mathbf{v}_0+\mathbf{v}_1)=\frac{q}{m}[\mathbf{E}_0+\mathbf{E}_1+(\mathbf{v}_0+\mathbf{v}_1)\times(\mathbf{B}_0+\mathbf{B}_1)]+\mathbf{g}
\end{aligned}
$$

Getting rid of the equilibrium and high-order terms, we have (Notice that $\mathbf{g}$ does not even appear here! In MHD, it does, in a very explicit way.)

$$
\begin{aligned}
i(k_y v_{0y}-\omega)n_1&=-n_0 i k_y v_{1y}-v_{1x}\frac{\partial n_0}{\partial x}\\
\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{v}_1=i(k_y v_{0y}-\omega)\mathbf{v}_1&=\frac{q}{m}(\mathbf{E}_1+\mathbf{v}_1\times\mathbf{B}_0)
\end{aligned}
$$

Now, from the linearized momentum equation, we can get the x and y components of perturbed velocity; intuitively, you can guess the expression:

$$
\begin{aligned}
&v_{1,ix}=\frac{E_{1y}}{B_0},\ v_{1,ex}=\frac{E_{1y}}{B_0}\\
&v_{1,iy}=\frac{1}{\Omega_i}\frac{\mathrm{d}}{\mathrm{d}t}\Big(\frac{E_{1y}}{B_0}\Big)=\frac{i(k_y v_{0i}-\omega)}{\Omega_i}\Big(\frac{E_{1y}}{B_0}\Big),\ v_{1,ey}=\frac{1}{\Omega_e}\frac{\mathrm{d}}{\mathrm{d}t}\Big(\frac{E_{1y}}{B_0}\Big)\approx 0
\end{aligned}
$$

where in the x direction, it is the $\mathbf{E}\times\mathbf{B}$ drift, and in the y direction, it is the polarization drift.

From the linearized continuity equation

$$
\begin{aligned}
i(k_y v_{0,yi}-\omega)n_{1i}&=-n_0 i k_y v_{1,yi}-v_{1,xi}\frac{\partial n_0}{\partial x}\\
-i\omega n_{1e}&=-v_{1,xe}\frac{\partial n_0}{\partial x}
\end{aligned}
$$

Then we can get $n_{1e}=n_{1e}(E_{1y}),\ n_{1i}=n_{1i}(E_{1y})$. Setting $n_{1e}=n_{1i}$, we have the dispersion relation

$$
\omega(\omega-k_y v_{0i})=g\frac{1}{n_0}\frac{\partial n_0}{\partial x}
$$

When $k_y\rightarrow 0$,

$$
\omega^2=g\frac{1}{n_0}\frac{\partial n_0}{\partial x}<0\Rightarrow\ \text{instability!}
$$

Let's think about this 2-fluid approach for a while. Apparently, we cannot treat a sharp boundary, namely $\frac{\partial n_0}{\partial x} = \delta(x)$, with exactly the same equations. However, it's quite a surprise that MHD approach can easily do that, as we will see in the next section.

9.4.2 Single fluid MHD method

In equilibrium, $\mathbf{g} = \hat{x}g,\ \mathbf{B}_0 = \hat{z}B_0(x),\ \mathbf{U}_0 = 0,\ \mathbf{E}_0 = 0,\ \rho_0(x),\ p_0(x)$.

$$\begin{aligned}\frac{\partial \rho_0}{\partial t} + \nabla\cdot(\rho_0\mathbf{U}_0) = 0\\ 0 = \frac{1}{\mu_0}\Big[-\frac{1}{2}\frac{\partial}{\partial x}({B_0}^2)\Big] - \frac{\partial}{\partial x}p_0(x) + \rho_0(x)g\end{aligned}$$

Note that there's a difference between cases where different pressure is dominant. For example, in z-pinch the magnetic pressure is dominant, while in a laser pulse, the thermal pressure is usually dominant.

Assume perturbations of the form

$$\begin{aligned}p_1(\mathbf{x},t) &= p_1(x)e^{ik_y y - i\omega t}\\ \rho_1(\mathbf{x},t) &= \rho_1(x)e^{ik_y y - i\omega t}\\ \mathbf{U}_1 &= \frac{\partial \boldsymbol{\xi}_1}{\partial t} = -i\omega\boldsymbol{\xi}_1\end{aligned}$$

where $\boldsymbol{\xi}_1$ is the displacement.

We can calculate each linear term:

$$\begin{aligned}\big[(\mathbf{B}_0+\mathbf{B}_1)\cdot\nabla\big](\mathbf{B}_0+\mathbf{B}_1) \approx (\mathbf{B}_0\cdot\nabla)\mathbf{B}_1 + (\mathbf{B}_1\cdot\nabla)\mathbf{B}_0 = (B_0(x)\frac{\partial}{\partial z})\mathbf{B}_1\\ B^2 = (\mathbf{B}_0+\mathbf{B}_1)\cdot(\mathbf{B}_1+\mathbf{B}_1) \approx 2\mathbf{B}_0\cdot\mathbf{B}_1\\ \mathbf{U} = \mathbf{U}_0 + \mathbf{U}_1 = \mathbf{U}_1\end{aligned}$$

The tension term has no x or y component, so we can just ignore it. Then the linearized momentum equation can be written as

$$\rho_0\frac{\partial \mathbf{u}_1}{\partial t} = \rho_0\frac{\partial^2\boldsymbol{\xi}_1}{\partial t^2} = -\nabla\Big(\frac{\mathbf{B}_0\cdot\mathbf{B}_1}{\mu_0}\Big) - \nabla p_1 + \rho_1\mathbf{g}$$

which can be separated into two scalar equations

$$
\begin{aligned}
-\rho_0\omega^2\xi_{1x} &= -\frac{\partial}{\partial x}\Big(\frac{\mathbf{B}_1\cdot\mathbf{B}_1}{\mu_0}+p_1\Big)+\rho_1 g \\
-\rho_0\omega^2\xi_{1y} &= -ik_y\Big(\frac{\mathbf{B}_1\cdot\mathbf{B}_1}{\mu_0}+p_1\Big)
\end{aligned}
$$

Assume incompressibility

$$
\nabla\cdot\mathbf{u}=0\Rightarrow\nabla\cdot\mathbf{u}_1=0,\ \nabla\cdot\boldsymbol{\xi}_1=\frac{\partial\xi_{1x}}{\partial x}+ik_y\xi_{1y}=0
$$

The linearized continuity equation (Section 7.11) yields

$$
\rho_1=-\nabla\cdot(\rho_0\boldsymbol{\xi}_1)=\xi_{1x}\frac{\partial}{\partial x}\rho_0
$$

Combining the last four equations, we have

$$
-\rho_0\omega^2\xi_{1x}=-\frac{\partial}{\partial x}\Big[\rho_0\omega^2\frac{1}{{k_y}^2}\frac{\partial\xi_{1x}}{\partial x}\Big]-g\xi_{1x}\frac{\partial\rho_0}{\partial x} \tag{9.1}
$$

This is the governing equation for the Rayleigh-Taylor instability, which is the same as Eq.(10.15) in (Bellan 2008). Note that here we have no assumption on the x-dependence; if we simply use the local approximation as before, this immediately gives you the identical result.

To treat the sharp boundary problem, we assume

$$
\rho_0=\begin{cases}\text{const.} & \text{if } x<0\\ 0 & \text{if } x>0\end{cases}
$$

Then for $x<0$,

$$
\begin{aligned}
&\frac{\partial^2\xi_{1x}}{\partial x^2}-{k_y}^2\xi_{1x}=0\\
\Rightarrow\ &\xi_{1x}=Ae^{k_y x}+Be^{-k_y x}
\end{aligned}
$$

and for $x>0$,

$$
\begin{aligned}
&\frac{\partial^2\xi_{1x}}{\partial x^2}-{k_y}^2\xi_{1x}=0\\
\Rightarrow\ &\xi_{1x}=Ce^{k_y x}+De^{-k_y x}
\end{aligned}
$$

The coefficient B and C must be zero because of infinite field requirement. Due to continuity at $x = 0$, we set

$$A = D = \xi_0$$

The density profile obeys

$$\frac{\partial \rho_0}{\partial x} = -\rho_0 \delta(x)$$

Integrating the governing Equation 9.1 from $x = 0^-$ to $x = 0^+$ yields

$$\begin{aligned} &-\frac{\rho_0 \omega^2}{{k_y}^2} \frac{\xi_{1x}}{\partial x}\bigg|_{x=0^-}^{x=0^+} - g\xi_{1x}(-\rho_0) = 0 \\ \Rightarrow &\omega^2 = k_y g \end{aligned}$$

Therefore the growth rate is $\gamma = \Im(\omega) = \sqrt{k_y g}$. You may realize that $\mathbf{k} \cdot \mathbf{B}_0 = 0$ here, so this magnetic stablizing term vanishes in the dispersion relation.

9.4.3 2-fluid sharp boundary model

Now let's go back and see if we can treat the sharp boundary problem with 2-fluid model. This is actually not easy: it is first solved by S.Chandraserkhar in the view of particle orbit theory. I believe there is a more 'modern' way of doing exactly the same thing, but here I just list the original derivation.

We consider a plasma at uniform temperature lying above a horizontal plane in a uniform gravitational field directed vertically downwards. There is a horizontal magnetic field in x direction uniform in each half volumn with a jump in field strength produced by a uniform horizontal current sheet at the boundary plane $z = 0$. The gravitational force is balanced by a pressure gradient in the plasma and by the jump in the magnetic pressure at $z = 0$.

We now suppose the boundary of the plasam at $z = 0$ to be rippled by a sinusoidal disturbance as shown in fig-RT_perturb. We may write for the displacement of the interface (ADD FIGURE!)

$$\Delta z = a \sin ky \tag{9.2}$$

where a, the amplitude of the disturbance, is considered small and $k(= 2\pi/\lambda)$ is the wave number of the disturbance in the y-direction. The drift resulting from gravity is given by

$$\mathbf{V}_g = \frac{m}{q} \frac{\mathbf{g} \times \mathbf{B}}{B^2}$$

Since the magnetic field is in the x-direction, the electrons will drift to the right and the positive ions will drift to the left. The gravity drift, therefore, causes a charge separation in the plasma and the resulting boundary has the form shown in fig-RT_displacement. The surface charge $\delta\sigma$ due to the separation ($\delta\Delta z$) of ions and electrons is given by

$$\begin{aligned}\delta\sigma &= Ne\delta\Delta z \\ &= Ne\frac{\partial\Delta z}{\partial y}\delta y \\ &= e\frac{\partial\Delta z}{\partial y}V_g\delta t\end{aligned}$$

Therefore, the time rate of change of the surface charge density is given by

$$\begin{aligned}\frac{\partial\sigma}{\partial t} &= NeV_g\frac{\partial}{\partial y}\Delta z \\ &= -Ne\frac{Mg}{eB}ak\cos ky \\ &= -\frac{NMeg}{B}ak\cos ky\end{aligned} \tag{9.3}$$

where in writing these expressions, we have neglected the electron drift, as being small in the ratio m/M compared to the ion drift. The electric field resulting from the surface charge can be computed in a straight-forward manner. In the region away from the boundary, we must have

$$\nabla\cdot(\epsilon\mathbf{E}) = 0 \tag{9.4}$$

where $\epsilon = \epsilon_0(1 + \frac{{\omega_{pi}}^2}{{\Omega_i}^2}) = ...$ is the dielectric constant of the plasma. At the interface the electric field is determined by

$$\nabla\cdot(\epsilon\mathbf{E}) = \frac{1}{\mu_0}\frac{\sigma}{dz}$$

where σ is the surface charge density and dz is the infinitesimal thickness of the layer. We now integrate this equation over an element of volumn dSdz. The right-hand side gives the charge within the column element (σdS). Making use of Gauss's theorem to transform the left-hand side, we obtain

$$\epsilon E_z dS = \frac{1}{\mu_0}\sigma dS = \frac{1}{\mu_0}\sigma_0\cos ky dS$$

Thus the electric field at the interface is given by

$$\epsilon E_z = \frac{\sigma_0}{\mu_0}\cos ky \tag{9.5}$$

The electric field which satisfies Equation 9.4 within the plasma and the boundary condition Equation 9.5 at $z = 0$ has the components

$$E_y = \frac{\sigma_0}{\mu_0 \epsilon} \sin ky \, e^{-kz}$$
$$E_z = \frac{\sigma_0}{\mu_0 \epsilon} \cos ky \, e^{-kz}$$

with $E_x = 0$. These electric fields give rise to the drifts which can be computed from the equation

$$\mathbf{V} = \frac{\mathbf{E} \times \mathbf{B}}{B^2}$$

Remembering that $\mathbf{B}$ is in the x-direction, we obtain

$$V_y = \frac{E_z}{B}, \quad V_z = -\frac{E_y}{B}$$

From the foregoing equations we obtain

$$V_y = \frac{\sigma_0}{\mu_0 B} \cos ky \, e^{-kz}$$
$$V_z = -\frac{\sigma_0}{\mu_0 B} \sin ky \, e^{-kz}$$

It is clear from the solutions that the velocity field is divergence free and, therefore, does not cause any change in the density of the plasma except at the boundary. We have

$$\frac{\partial}{\partial t} \Delta z(z = 0) = V_z(z = 0) = -\frac{\sigma}{\mu_0 B} \sin ky \tag{9.6}$$

From Equation 9.2 and Equation 9.6, we obtain the equation of motion for the amplitude a:

$$\frac{\mathrm{d}a}{\mathrm{d}t} = -\frac{\sigma_0}{\mu_0 B}$$

Equation 9.3 and Equation 9.5 yield

$$\frac{\mathrm{d}\sigma_0}{\mathrm{d}t} = -\frac{NMg}{\mu_0 B} ak$$

From the above two equations, we obtain

$$\frac{\mathrm{d}^2 a}{\mathrm{d}t^2} = \frac{1}{\mu_0 \epsilon B} \frac{NMg}{B} ka$$
$$\approx gka, \quad (\text{for } \epsilon \gg 1)$$

($g =$ ... from equilibrium). The solution of this equation is given by

$$a(t) = a_0 e^{\pm\sqrt{gk}t}$$

It is interesting to note that the rate of growth of the instability is exactly the same as in the Rayleigh-Taylor instability of a fluid supported against gravity by a second fluid which is weightless. The charge separation is able to overcome exactly the restraining influence of the magnetic field. This exact compensation occurs only in the limit of $\epsilon \gg 1$.

The same result can also be obtained using the rigorous formulation of the Boltzmann transport equation. However, in more complicated cases, the first order orbit theory gives results which agree with those obtained from the Boltzmann equation only in some special cases.

The essential mechanism which gives rise to the instability is the charge separation resulting from the gravity drift — drift arising from a force which does not depend upon the sign of the charge. If we consider a plasma configuration in a torus, the particles experience the centrifugal force $m{v_\parallel}^2/R$ and the gradient B force $m{v_\perp}^2/R$ which are both independent of the sign of the charge. Therefore, we should expect instabilities in a plasma confined to a torus.

9.5 Kelvin-Helmholtz instability

Kelvin-Helmholtz (KH) instability happens due to a velocity shear and produces surface waves at the shear layer that eventually roll up into nonlinear vortices. Typical examples are:

- plane crash
- flapping of flags
- diocotron instability in an electron sheet
- water wave, nonlinear phase
- laser ablation of metal
- low-latitude flanks of the magnetopause during periods of northward interplanetary magnetic field
- accretion discs
- pulsar winds
- jets

It presents a key mechanism for mediating mass, momentum and energy transport at boundary layers with a velocity shear via processes such as viscous interaction, turbulence, mode conversion and magnetic reconnection.

To understand why shear flow can lead to instability, we'll first introduce the Bernoulli theorem in fluid mechanics. From the MHD momentum equation, take $\partial/\partial t = 0,\ \mathbf{J} = 0,\ \mathbf{g} = 0$, we obtain

$$\rho\mathbf{v}\cdot\nabla\mathbf{v} = -\nabla p$$

Using the natural coordinates, let $\hat{t}$ be the unit tangent vector on a streamline, $\hat{n}$ be the unit normal vector pointing from concave to convex side, and ds be the infinitisimal distance along streamline, we have

$$\begin{aligned}\rho\mathbf{v}\cdot\nabla\mathbf{v} &= (\rho v\frac{\partial}{\partial s})\mathbf{v} = \rho v_s\frac{\partial}{\partial s}(\hat{t}v) = \rho v\Big(\frac{\partial v}{\partial s}\hat{t} + \frac{\partial \hat{t}}{\partial s}v\Big)\\ &= -\frac{\partial p}{\partial s}\hat{t} - \nabla_\perp p\\ \Rightarrow \hat{t}:\ pv\frac{\partial v}{\partial s} + \frac{\partial p}{\partial s} &= \frac{\partial}{\partial s}\Big(\frac{1}{2}\rho v^2 + p\Big) = 0\end{aligned}$$

Therefore, $\frac{1}{2}\rho v^2 + p =$ const. along a streamline for an incompressible flow. This is the classical Bernoulli equation.

Now, consider two flow layer with velocity shear at the plane interface in fig-two_layer (ADD FIGURE!). Imaging there's a ripple on the layer interface pointing upward at Q. If we examine the cross section at P and Q for the lower layer respectively, we find

$$\text{flux at P} = \rho vA|_P = \text{flux at Q} = \rho vA|_Q$$

Since density along field lines are constant (incompressible) and area $A_P < A_Q$, we have $v_P > v_Q$. From the Bernoulli equation, $p_P < p_Q$. Similar for the upper layer, we get the pressure at P is larger than that at Q. Therefore, the total pressure is pointing away from interface, which let the ripple grow.

Now we are ready to do more careful derivations. The governing equations are

$$\begin{aligned}\nabla\cdot\mathbf{v} &= 0\\ \rho\Big(\frac{\partial\mathbf{v}}{\partial t} + \mathbf{v}\cdot\nabla\mathbf{v}\Big) &= -\nabla p\end{aligned}$$

In equilibrium, suppose there is a velocity shear in the x direction, and the interface lies along the y direction,

$$\begin{aligned}\rho_0 &= \text{const.}\\ p_0 &= \text{const.}\\ \mathbf{v}_0 &= \hat{y}v_{0y}(x)\end{aligned}$$

Assume linear perturbations of the form

$$\begin{aligned}v_{1x}(\mathbf{x},t) &= v_{1x}(x)e^{ik_y y - i\omega t}\\ p_1(\mathbf{x},t) &= p_1(x)e^{ik_y y - i\omega t}\end{aligned}$$

so the linearized momentum equation is

$$\rho_0\big(\frac{\partial\mathbf{v}_1}{\partial t} + \mathbf{v}_0\cdot\nabla\mathbf{v}_1 + \mathbf{v}_1\cdot\nabla\mathbf{v}_0\big) = -\nabla p_1$$

where

$$\mathbf{v}_0 \cdot \nabla \mathbf{v}_1 = ik_y v_{0y} \mathbf{v}_1$$
$$\mathbf{v}_1 \cdot \nabla \mathbf{v}_0 = v_{1x} \frac{\partial}{\partial x} v_{0y}(x) \hat{y}$$

Then the x and y components of the linearized momentum equation give

$$-i\omega\rho_0 v_{1x} + \rho_0 ik_y v_{0y} v_{1x} = -\frac{\partial p_1}{\partial x}$$
$$-i\omega\rho_0 v_{1y} + \rho_0 ik_y v_{0y} v_{1y} + \rho_0 v_{1x} \frac{\partial}{\partial x} v_{0y}(x) = -ik_y p_1$$

Together with the linearized incompressibility condition

$$\frac{\partial v_{1x}}{\partial x} + ik_y v_{1y} = 0$$

by eliminating p_1 and v_{1y} we get

$$\frac{\partial^2 v_{1x}}{\partial x^2} - \Big[{k_y}^2 - \frac{k_y {v_{0y}}''(x)}{\omega - k_y v_{0y}(x)}\Big] v_{1x} = 0 \tag{9.7}$$

Now we are half way from obtaining the dispersion relation. For simplicity, let us assume three layer regions with Region I on the left ($x < 0$), Region II in the middle ($x \in (0, \tau)$), and Region III on the right ($x > \tau$). Let the shear layer II thickness be τ. Set the velocity on the two sides $v_1 = 0,\ v_2 = v_1 + \Delta v \equiv v$, and $k_y = k$. Then

$$v_{0y}''(x) = \Big(\frac{v}{\tau}\Big)\big[\delta(x) - \delta(x-\tau)\big]$$

except at $x = 0$ and $x = \tau$. Equation 9.7 can be simplified to

$$\frac{\mathrm{d}^2 v_{1x}}{\mathrm{d}x^2} - k^2 v_{1x} = 0$$

In region I ($x < 0$),

$$v_{1x} = \xi_0 e^{kx}, \quad x < 0$$
$$\frac{\partial v_{1x}}{\partial x}\Big|_{x=0_-} = k\xi_0$$

In region III ($x > \tau$),

$$v_{1x} = \xi_\tau e^{-k(x-\tau)}, \quad x > \tau$$
$$\frac{\partial v_{1x}}{\partial x}\Big|_{x=\tau^+} = -k\xi_\tau$$

In region II ($x \in (0, \tau)$), we are looking for a solution which is a superposition of the solutions from both sides and is continuous at the boundaries

$$
\begin{aligned}
v_{1x} &= \xi_\tau \frac{\sinh kx}{\sinh k\tau} + \xi_0 \frac{\sinh k(x-\tau)}{\sinh -k\tau}, \quad 0 < x < \tau \\
\frac{\partial v_{1x}}{\partial x}\Big|_{x=\tau^-} &= k\xi_\tau \frac{\cosh k\tau}{\sinh k\tau} - k\xi_0 \frac{1}{\sinh k\tau} \\
\frac{\partial v_{1x}}{\partial x}\Big|_{x=0^+} &= k\xi_\tau \frac{1}{\sinh k\tau} - k\xi_0 \frac{\cosh k\tau}{\sinh k\tau}
\end{aligned}
$$

The continuity at $x = 0$ requires $V_{0y} = 0, V_{1x} = \xi_0$. Integrating the governing Equation 9.7 from $x = 0^-$ to $x = 0^+$ yields

$$
-\xi_0 k \frac{\cosh k\tau}{\sinh k\tau} + \xi_\tau k \frac{1}{\sinh k\tau} - k\xi_0 + \frac{kv}{\omega\tau}\xi_0 = 0
$$

Integrating the governing Equation 9.7 from $x = \tau^-$ to $x = \tau^+$ yields

$$
-\xi_\tau k + \xi_0 k \frac{1}{\sinh k\tau} - \xi_\tau k \frac{\cosh k\tau}{\sinh k\tau} - \frac{kv}{\omega\tau}\xi_\tau = 0
$$

Combining the last two equations, we obtain

$$
1 = \left[\sinh k\tau + \cosh k\tau + \frac{kv}{\omega - kv}\frac{\sinh k\tau}{k\tau}\right]\left[\cosh k\tau + \sinh k\tau - \frac{kv}{\omega}\frac{\sinh k\tau}{k\tau}\right]
$$

which is the dispersion relation for the KH instability.

With the identity $\sinh k\tau + \cosh k\tau = e^{k\tau}$, the dispersion relation can be simplified to

$$
\begin{aligned}
1 &= \left[e^{k\tau} + \frac{kv}{\omega - kv}\frac{\sinh k\tau}{k\tau}\right]\left[e^{k\tau} - \frac{kv}{\omega}\frac{\sinh k\tau}{k\tau}\right] \\
1 &= e^{2k\tau} + e^{k\tau}\frac{(k\tau)^2}{\omega(\omega - kv)}\frac{\sinh k\tau}{k\tau} - \frac{(k\tau)^2}{\omega(\omega - k\tau)}\left(\frac{\sinh k\tau}{k\tau}\right)^2 \\
1 &= e^{2k\tau} + \frac{(k\tau)^2}{\omega(\omega - kv)}\frac{\sinh k\tau}{k\tau}\left[e^{k\tau} - \frac{\sinh k\tau}{k\tau}\right]
\end{aligned}
$$

Multiplying both sides by $\omega(\omega - kv)$, we get

$$
\omega(\omega - k\tau)(1 - e^{2k\tau}) = (kv)^2 \frac{\sinh k\tau}{k\tau}\left[e^{k\tau} - \frac{\sinh k\tau}{k\tau}\right]
$$

Assuming $k\tau \ll 1$ (long wavelength approximation), $e^{k\tau} \approx 1 + k\tau$, we obtain

$$
\omega(\omega - kv) + \frac{(kv)^2}{2} \approx 0
$$

the solution of which is

$$\omega = \frac{1}{2}kv(1 \pm i), \quad k\tau \ll 1$$

In general, the growth rate of KH is

$$\omega_i = \frac{1}{2}|k_y \Delta v|, \quad k_y\tau \ll 1$$

9.5.1 Diocotron instability on electron sheet

HAVEN'T CHECKED!

A diocotron instability is a plasma instability created by two sheets of charge slipping past each other. Energy is dissipated in the form of two surface waves propagating in opposite directions, with one flowing over the other. This instability is the plasma analog of the Kelvin-Helmholtz instability in fluid mechanics.

For the simplest case, we have a uniform electron sheet and a parallel constant magnetic field in the plane of the sheet, as illustrated in fig-electron_sheet. Due to space charge of electron sheet, there is electric field pointing towards the sheet in the upper and lower region. Consider a small perturbation (sinusoidal ripple) on an electron sheet. The coulomb force expels electrons outward, so the electrons will drift, according to the right-hand rule, to the left. The deficit of electrons is equivalent to some positive charge distribution, and thus created an electric field. The $\mathbf{E}\times\mathbf{B}$ drift is pointing outward, so the perturbation is growing. Note that even though this problem looks innocent, but it is actually not easy. People, even the giants in plasma physics, made a lot of mistakes in the derivation!

If there is also a magnetic field inside the sheet, the $\mathbf{E}\times\mathbf{B}$ drift will form a velocity gradient within the sheet, and lead to K-H instability. Denote σ_0 as the surface charge density and ρ_0 as charge density, we have

$$\sigma_0 = \rho_0\tau = en_0\tau,$$

and the velocity shear across the sheet

$$\Delta v = \frac{E_2}{B_0} + \frac{E_1}{B_0} = -\frac{1}{B_0}\frac{\sigma_0}{\epsilon_0} = -\frac{en_0\tau}{B_0\epsilon_0}.$$

Then from the dispersion relation of K-H mode, we have the growth rate

$$\omega_i = \frac{1}{2}k_y|\Delta v| = \frac{1}{2}k_y\left|\frac{en_0\tau}{B_0\epsilon_0}\right| = \frac{1}{2}k_y\tau\frac{{\omega_{pe}}^2}{|\Omega_e|},$$

which is valid as long as $k_y\tau \ll 1$,i.e., long wave length limit.

(Bellan 2008) P537.

FIGURE NEEDED from H.W.3.4 Consider the diocotron instabity on a MELBA-like annular electron beam which propagates inside a metallic drift tube. Let $V = $ beam voltage, $I =$ beam current, $a =$ beam radius, $\tau =$ annular beam thickness ($\tau \ll 1$)m $L =$ length of drift tube, $T = $ beam's pulselength, $B =$ axial magnetic field. Note that the combined self-electric and self-magnetic field of the beam produces a slow rotational $\mathbf{E} \times \mathbf{B}$ drift in the $\theta-$direction. This azimuthal drift velocity, $v_{0\theta}$, is much less than the axial velocity of the beam, but it is sheared.

In equilibrium,

$$\begin{aligned}
0 &= q(\mathbf{v} \times \mathbf{B} + \mathbf{E}) \\
\mathbf{v} &= v_{0\theta}\hat{\theta} + v_{0z}\hat{z} \\
\mathbf{B} &= B_{0\theta}\hat{\theta} + B_{0z}\hat{z} \\
\Rightarrow & v_\theta - v_z B_\theta + E_r = 0
\end{aligned}$$

From Ampère's law,

$$\begin{aligned}
B_{0\theta} \cdot 2\pi(a+\tau) \approx B_\theta \cdot 2\pi a &= \mu_0 2\pi a \tau J_z \\
\Rightarrow J_z &= \frac{1}{\mu_0 \tau} B_{0\theta} = -en_0 V_{0z}
\end{aligned}$$

From Gauss's law,

$$\begin{aligned}
\nabla \cdot \mathbf{E} &= \frac{\rho}{\epsilon_0} = \frac{-en_0}{\epsilon_0} \\
E_r \cdot 2\pi a \Delta &= \frac{-en_0}{\epsilon_0} \cdot 2\pi a \tau \Delta \\
\Rightarrow E_r &= \frac{1}{\epsilon_0} \tau(-e) n_0
\end{aligned}$$

Substituing E_r and J_z into the radial force balance equation, we obtain

$$\begin{aligned}
V_\theta\Big|_{r=a+\tau} &= \frac{1}{B_z}[V_{0z}B_\theta - E_r]\Big|_{r=a+\tau} \\
&= \frac{1}{\gamma^2}\frac{E_r}{B_z}
\end{aligned}$$

where $\gamma = (1-\beta)^{-1/2} = 1 + V/(511 \text{ keV})$, and $\beta = V_{0z}/c$.

Let $\nu = \frac{I}{\beta I_A}$ be the Budker parameter, $I_A = 4\pi\epsilon_0 mc^2/e = 17\,\text{kA}$ be the Alfvén-Lawson current, and $\Omega = \frac{eB_{0z}}{m_e}$ be the nonrelativistic cyclotron frequency associated with the axial B field, we have

$$V_\theta\big|_{r=a+\tau} = \frac{2c^2\nu}{\Omega a\gamma^2}$$

At $r = a$, $V_\theta = 0$ because there's no E field. Therefore the velocity shear in $\hat{\theta}$ is

$$\Delta V_\theta = V_\theta\big|_{r=a+\tau} - V_\theta\big|_{r=a} = \frac{2c^2\nu}{\Omega a\gamma^2}$$

Then from the analysis of K-H instability, the temporal growth rate ω_i is given by

$$\begin{aligned}\omega_i &= \frac{1}{2}|k_\theta \Delta V_{0\theta}| \\ &= \frac{1}{2}\frac{m}{a}\frac{2c^2\nu}{\Omega a\gamma^2}\end{aligned}$$

For long wavelength limit, let $m = 1$.

For MELBA-like beam with the following parameters, $V = 700\text{keV},\ I = 1\text{kA},\ a = 5\text{cm},\ \tau = 0.5\text{cm},\ T = 500\text{ns},\ L = 1\text{m},\ B = 2\text{kG}$,

$$\omega_i = 1.18 \times 10^7\text{s}^{-1}$$

The total number of e-folds of the instability growth during the pulse time T is

$$\omega_i T \approx 5.9$$

Even though this is large, K-H instability will not stay at the initial position and grow in time; instead it will be transported. It is more meaningful to estimate the spatial growth by evaluating the total number of e-folds experienced by a signal of some frequency as it propagates along the machine length L:

$$k_i L = \frac{\omega_i}{V_{0z}} L = \omega_i \frac{L}{\beta c} \approx 0.04$$

Therefore we don't need to worry too much about this instability.

9.6 MHD Stability

$$\mathbf{J} \times \mathbf{B} = \nabla p$$

$$\mathbf{J}_\perp = \frac{\mathbf{B} \times \nabla p}{B^2}$$

The current is often called the *diamagnetic current.* It arises from the plasma pressure gradient.

Using Ampère's law we can write the magnetic force in the form

$$\mathbf{J} \times \mathbf{B} = -\nabla\Big(\frac{B^2}{2\mu_0}\Big) + \frac{1}{\mu_0}(\mathbf{B} \cdot \nabla)\mathbf{B} \tag{9.8}$$

which separates into the magnetic pressure term and the magnetic tension term.

If we use an anisotropic description of the thermal pressure term, Equation 9.8 can be written as

$$\mathbf{J} \times \mathbf{B} = -\nabla_\perp \left(p_\perp + \frac{B^2}{2\mu_0}\right) + \left(1 + \frac{p_\perp - p_\parallel}{B^2/\mu_0}\right) \mathbf{B} \cdot \nabla\mathbf{B} \tag{9.9}$$

9.6.1 Harris Current Sheet

An example of a MHD equilibrium configuration is the Harris current sheet, in which the variations in the magnetic field and plasma pressure over the current sheet balance each other In a 1D Harris current sheet the magnetic field (assumed here to be in the z-direction) is given by

$$\mathbf{B} = B_0 \tanh\left(\frac{z}{\lambda}\right)\hat{y}$$

The pressure is given by

$$p = p_0 \cosh^{-2}\frac{z}{\lambda}$$

where $p_0 = B_0^2/(2\mu_0)$. The current density is then

$$J_y(z) = \frac{B_0}{\mu_0\lambda}\mathrm{sech}^2\Big(\frac{z}{\lambda}\Big)$$

Harris current sheet can be taken as the first approximation of the Earth's magnetotail that can stay stable for long time periods.

9.6.2 -Pinch and Z-Pinch

-pinch and Z-Pinch are both 1D equilibrium configurations expressed in cylindrical coordinates. In a -pinch cylindrical coils drive an elecric current and the magnetic field is axial, while in a Z-pinch the electric current is axial and the magnetic field is toroidal.

9.6.3 Force-Free Field

If $\beta \ll 1$ in MHD equilibrium, the pressure gradient is negligible and thus

$$\mathbf{J} \times \mathbf{B} = 0 \tag{9.10}$$

Such configurations are called *force-free fields* because the magnetic force on the plasma is zero. According to Equation 9.8 in a force-free field the magnetic pressure gradient $\nabla(B^2/2\mu_0)$ is balanced by the magnetic tension force $\mu_0^{-1}(\mathbf{B} \cdot \nabla)\mathbf{B}$. In reality the force-free equilibrium is often a very good approximation of the momentum equation. It is also evident from Equation 9.10 that in a force-free field the electric current flows along the magnetic field. Such currents are commonly called *field-aligned currents* (FAC).

Using Ampère's law we can write Equation 9.10 as

$$(\nabla \times \mathbf{B}) \times \mathbf{B} = 0$$

From this we see that the innocent-looking equation $\mathbf{J} \times \mathbf{B} = 0$ is in fact nonlinear and thus difficult to solve.

The field-alignment of the electric crrent can be expressed as

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} = \alpha(\mathbf{r})\mathbf{B}$$

where α is a function of position. Taking divergence of this we get

$$\mathbf{B} \cdot \nabla\alpha = 0$$

i.e. α is constant along the magnetic field.

In the case α is *a constant in all directions*, the equation

$$\nabla \times \mathbf{B} = \alpha \mathbf{B} \tag{9.11}$$

is linear. Taking a curl Equation 9.11 we get the *Helmholtz equation*:

$$\nabla^2 \mathbf{B} + \alpha^2 \mathbf{B} = 0$$

Solution to the Helmoltz with helical equation in cylindrical symmetry was presented by Lundquist in 1950 in terms of Bessel functions J_0 and J_1:

$$\begin{aligned}
B_R &= 0 \\
B_A &= B_0 J_0\Big(\frac{\alpha_0 r}{r_0}\Big) \\
B_T &= \pm B_0 J_1\Big(\frac{\alpha_0 r}{r_0}\Big)
\end{aligned}$$

where B_r, B_A, and B_T are radial, axial and tangential magnetic field components, respectively. The solution is a magnetic *flux rope* where magnetic field lines form helixes whose pitch angle increases from the axis (**?@fig-flux_rope**). r is the radial distance from the flux rope axis, r_0 is the radius of the flux rope and B_0 is the maximum magnetic field magnitude at the center of the flux rope $r = 0$.

A special case of a force-free magnetic field is the current-free configuration $\nabla \times \mathbf{B} = 0$. Now the magnetic field can be expressed as the gradient of a scalar potential $\mathbf{B} = \nabla \Psi$, and since $\nabla \cdot \mathbf{B} = 0$ it can be found via the *Laplace equation*

$$\nabla^2 \Psi = 0$$

with appropriate boundary conditions and using the methods of potential theory.

For example, the Sun's magnetic field structure is often modeled by the so-called Potential Field Source Surface (PFSS) model (**?@fig-PFSS**). The magnetic field is computed from the Laplace equation using spherical coordinates from the photosphere to the "source surface", nominally chosen to be at 2.5 Solar radii. At the source surface the Sun's magnetic field is assumed to be purely radial. The inner boundary conditions are obtained from solar magnetograms. Thus, PFSS assumes that there is no electric current in the corona.

9.7 MHD Modes

A simple but representative dispersion relation writes

$$\omega^2 = (\mathbf{k}\cdot\mathbf{V}_A)^2 - \mathbf{k}\cdot\mathbf{g}, \quad \text{where } \mathbf{V}_A = \frac{\mathbf{B}_0}{B_0}\cdot V_A$$

If we treat plasma as a single magnetized fluid,

$$\rho\Big(\frac{\partial\mathbf{u}}{\partial t} + \underbrace{\mathbf{u}\cdot\nabla\mathbf{u}}_{\text{K-H inst.}}\Big) = \underbrace{-\nabla p}_{\text{ballooning inst.}} + \underbrace{\mathbf{j}\times\mathbf{B}}_{\text{kink, sausage inst.}} + \underbrace{\rho\mathbf{g}}_{\text{R-T inst.}}$$

Qualitatively, we can identify the source for each kind of instability in plasma. We will discuss them separately and in a set of combination below.

9.7.1 Kink Mode

A kink instability, is a *current-driven* plasma instability characterized by transverse displacements of a plasma column's cross-section from its center of mass without any change in the characteristics of the plasma. It typically develops in a thin plasma column carrying a strong axial current which exceeds the Kruskal–Shafranov limit and is sometimes known as the Kruskal–Shafranov (kink) instability.

The kink instability was first widely explored in fusion power machines with Z-pinch configurations in the 1950s. It is one of the common magnetohydrodynamic instability modes which can develop in a pinch plasma and is sometimes referred to as the $m = 1$ mode.

If a "kink" begins to develop in a column the magnetic forces on the inside of the kink become larger than those on the outside, which leads to growth of the perturbation. As it develops at fixed areas in the plasma, kinks belong to the class of "absolute plasma instabilities", as opposed to convective processes.

```
KeyNotes.plot_kink()
```

The kink instability is the most dangerous instability in Tokamak. We have discussed this kind of micro-instability from the view of single particle motion in Section 9.2; here, we will explore this a little bit further.

String model

First, image a current-carrying plasma column, shown in the x-z plane in fig-kink_column. The metallic wire carries current under tension T, and μ =mass/length is the mass per length. From the basic mechanics, $C_s = \sqrt{T/\mu}$ is the acoustic velocity in the system. Let the background field $\mathbf{B}_0 = \hat{z}B_0$ and the displacement $\boldsymbol{\xi} = \boldsymbol{\xi}(\mathbf{x}, t)$. We can show that, if the current $\mathbf{I}$ is sufficiently strong, there will be kink instability.

ADD PLASMA KINK COLUMN FIGURE!

Assume the displacement in x-y plane has the form

$$\boldsymbol{\xi} = (\xi_x, \xi_y)e^{ik_z z - i\omega t}$$

The force law gives (i.e. the basic string model in mechanics textbooks)

$$\mu\frac{\partial^2\boldsymbol{\xi}}{\partial t^2} = T\frac{\partial^2\boldsymbol{\xi}}{\partial z^2} + \text{force per unit length}$$

Here, the external force *per length* is the Lorentz force (which is why we say the R-T instability is current-driven),

$$\begin{aligned}\mathbf{I}\times\mathbf{B} &= \Big(\hat{x}I\frac{\partial\xi_x}{\partial z} + \hat{y}I\frac{\partial\xi_y}{\partial z} + \hat{z}0\Big)\times B_0\hat{z} \\ &= IB_0\Big(\hat{x}\frac{\partial\xi_y}{\partial z} - \hat{y}\frac{\partial\xi_x}{\partial z}\Big)\end{aligned}$$

In scalar forms, the force law gives

$$\begin{aligned}\hat{x}:\ -\mu\omega^2\xi_x &= T(-{k_z}^2)\xi_x + IB_0\frac{\partial\xi_y}{\partial z} \\ \hat{y}:\ -\mu\omega^2\xi_y &= T(-{k_z}^2)\xi_y - IB_0\frac{\partial\xi_x}{\partial z}\end{aligned}$$

Combining these two equations, we can easily get the dispersion relation

$$\omega^2 = {k_z}^2{C_s}^2 \pm \frac{IB_0}{\mu_0\mu}k_z$$

The dispersion relation is a representation of the force-law. The first term on the right-hand side is a stabilizing term due to tension; the second term with a minus sign is a destabilizing term due to Lorentz force. Note that the expression is very similar to R-T instability. (Which one?)

We can immediately estimate the scenario in a Tokamak. Take the radius of the column cut as a, wave number $k_z \sim 1/R$ (i.e. wave length is on the order of tokamak radius), ${C_s}^2 = {V_A}^2 = {B_{0z}}^2/(\mu_0\rho_0)$ (i.e. tension in plasma give rises to Alfvén wave), then the current is

$$I = J_z(\pi a^2) = \frac{B_\theta 2\pi a}{\mu_0} \sim \frac{B_\theta a}{\mu_0}$$

and the mass per unit length is

$$\mu = \rho_0(\pi a^2) \sim \rho_0 a^2$$

The criterion for stability then becomes

$$\begin{aligned}
&{k_z}^2{C_s}^2 > \frac{IB_0}{\mu_0}k_z \Rightarrow \frac{1}{R}\frac{{B_0 z}^2}{\mu_0\rho_0} > \frac{B_\theta a}{\mu_0}\frac{B_{0z}}{\rho_0 a^2} \\
\Rightarrow &\frac{a}{R}\frac{B_{0z}}{B_{0\theta}} > 1
\end{aligned}$$

which is called the *Kruskal-Shafranov stability criterion*. Usually we define

$$q \equiv \frac{a}{R}\frac{B_{0z}}{B_{0\theta}} = \frac{a}{R}\frac{B_t}{B_p}$$

as the safety factor. A real value for q is about 2 to 3.

Ideal MHD Approach

Now we use a more standard way to treat the kink mode. (Section 10.9 (Bellan 2008)) Assume we have a plasma column with radius a. Inside the column, we assume infinite conductivity, $\sigma = \infty$; outside the column, we assume vacuum so that we can only have current flow on surface $r = a$. Thus, besides the universal background magnetic field in the z direction, we also have an azimuthal field due to surface current. (You will see later that the decay in θ actually drives the kink instability.)

In equilibrium,

$$\begin{aligned}
r < a : \mathbf{B}_0 &= \hat{z}B_0 = \text{const.} \\
&p_0 = \text{const.},\ \mathbf{v}_0 = \text{const.},\ \mathbf{J}_0 = 0,\ \rho_0 = \text{const.} \\
r > a : \mathbf{B}_0 &= B_0\hat{z} + B_{0\theta}\frac{a}{r}\hat{\theta} \\
&p_0 = 0,\ \rho_0 = 0
\end{aligned}$$

The force equation in equilibrium

$$\nabla p_0 = \mathbf{J} \times \mathbf{B}_0$$

is satisfied automatically both for $r > a$ and $r < a$.

Let us introduce a small perturbation

$$\boldsymbol{\xi}_{1r}(\mathbf{x}, t) = \tilde{\xi}_{1r}(r)e^{ikz+im\theta-i\omega t}$$

such that at $r = a$,

$$\boldsymbol{\xi}_{1r}(\mathbf{x}, t)\Big|_{r=a} = \tilde{\xi}_{1a}(r)e^{ikz+im\theta-i\omega t}$$

Before running into linearized equations, we can first take a look at different wave modes. That is, what will the perturbation looks like at a fixed time t with different m? For simplicity, let us assume $t = 0$. (You can always make a time shift.) The actual displacement is the real part of $\boldsymbol{\xi}$,

$$\xi_{1r} = \xi_{1a}\cos(k_z z + m\theta)$$

For $m = 0$,

$$\xi_{1r} = \xi_{1a}\cos(k_z z)$$

which is the sausage mode.

For $m = 1$,

$$\xi_{1r} = \xi_{1a}\cos(k_z z + \theta)$$

If we draw the displacement down for $k_z z = 0, \frac{\pi}{2}, \pi, \frac{3}{2}\pi$, you can see one rotation in a 2π period, which indicates a shape of helix. This is often called the kink mode.

For higher m,

$$\xi_{1r} = \xi_{1a}\cos(k_z z + m\theta)$$

which looks like m intertwine helixes in one axial wavelength.

Now let's return to the perturbed equations. Here we will assume incompressibility as the equation of state,

$$\nabla \cdot \mathbf{v} = 0$$

The procedure to get the dispersion relation goes as follows:

1. Express the perturbed magnetic field as a function of displacement inside and outside the surface.
2. Relate the two regions by total force balance.

(I) $r < a$

$$\mathbf{v}_1 = \frac{\partial \boldsymbol{\xi}_1}{\partial t} \Rightarrow \nabla \cdot \boldsymbol{\xi}_\mathbf{1} = 0$$

The linearized continuity equation gives

$$\rho_1 = -\nabla \cdot (\rho_0 \boldsymbol{\xi}_1) = -\boldsymbol{\xi}_1 \cdot \nabla \rho_0 - \rho_0 \nabla \cdot \boldsymbol{\xi}_1 = 0$$

The linearized force law gives

$$\rho_0 \frac{\partial^2 \boldsymbol{\xi}_1}{\partial t^2} = -\nabla p_1 + \frac{(\nabla \times \mathbf{B}_1) \times \mathbf{B}_0}{\mu_0} + \cancel{\mathbf{J}_0} \times \mathbf{B}_1$$

And the Ohm's law gives

$$\begin{aligned} -\frac{\partial \mathbf{B}_1}{\partial t} &= \nabla \times \mathbf{E}_1 = \nabla \times (-\mathbf{v}_1 \times \mathbf{B}_0) \\ \mathbf{B}_1 &= \nabla \times (\boldsymbol{\xi}_1 \times \mathbf{B}_0) = B_0 \nabla \times (\boldsymbol{\xi}_1 \times \hat{z}) = ik_z B_0 \boldsymbol{\xi}_1 \end{aligned}$$

The last equivalence is obtained from the expansion of the second term into four terms and cancellation of zero terms.

In cylindrical coordinates,

$$(\nabla \times \mathbf{B}_1) \times \mathbf{B}_0 = ik{B_0}^2 (\nabla \times \boldsymbol{\xi}_1) \times \hat{z}$$

and

$$\nabla \times \boldsymbol{\xi}_1 = \frac{1}{r} \begin{bmatrix} \hat{r} & r\hat{\theta} & \hat{z} \\ \partial_r & \partial_\theta & \partial z \\ \xi_{1\theta} & r\xi_{1\theta} & \xi_{1z} \end{bmatrix}$$

so the linearized force law gives

$$
\begin{aligned}
-\omega^2\xi_{1r} &= -\frac{1}{\rho_0}\frac{\partial p_1}{\partial r} + ikv_A{}^2[ik\xi_{1r} - \frac{\partial \xi_{1z}}{\partial r}] \\
-\omega^2\xi_{1\theta} &= -\frac{im}{\rho_0 r}p_1 + kv_A{}^2[-\frac{im}{r}\xi_{1z} + ik\xi_{1\theta}] \\
-\omega^2\xi_{1z} &= -ik_z p_0/\rho_0
\end{aligned}
$$

Substituting the expression of ξ_{1z} into the other two equations, we can get

$$\boldsymbol{\xi}_1 = \frac{1}{\omega^2}\nabla(\frac{p_1}{\rho_0})$$

From the incompressibility condition, we have

$$\nabla\cdot\boldsymbol{\xi}_1 = 0 \Rightarrow \nabla^2 p_1 = 0$$

or in cylindrical coordinates,

$$\frac{1}{r}\frac{\partial}{\partial r}(r\frac{\partial p_1}{\partial r}) - \frac{m^2 p_1}{r^2} - k^2 p_1 = 0$$

Assume long wavelength limit and $m = 1$ (kink mode),

$$kr < ka \ll 1$$

we have

$$\frac{1}{r}\frac{\partial}{\partial r}(r\frac{\partial p_1}{\partial r}) - \frac{p_1}{r^2} = 0$$

the solution of which from Legendre polynomials (I need to check) is

$$p_1 = Ar + \frac{B}{r} = Ar$$

because p_1 is finite at $r = 0$.

So we have

$$
\begin{aligned}
\boldsymbol{\xi}_1 &= \frac{1}{\omega^2}\nabla(\frac{p_1}{\rho_0}) = \hat{r}\frac{A}{\rho_0\omega}e^{-i\omega t + i\theta + ikz} \\
\Rightarrow\ \xi_{1ra} &= \frac{A}{\rho_0\omega^2}
\end{aligned}
$$

Then the perturbed kinetic pressure on the surface is

$$p_1(r = a^-) = Aa = \rho_0 \omega^2 \xi_{1ra} a,$$

and the perturbed magnetic field is

$$\mathbf{B}_1 = ikB_0 \boldsymbol{\xi}_1 \Rightarrow B_{1z}(r = a^-) = ikB_0 \xi_{1z} = -k^2 B_0 \xi_{1ra}$$

(II) $r > a$

$$\nabla \cdot \mathbf{B}_1 = 0, \ \nabla \times \mathbf{B}_1 = 0 \Rightarrow \mathbf{B}_1 = \nabla \Psi_1, \ \nabla^2 \Psi_1 = 0$$

The solution of Laplace equation in cylindrical coordinates is

$$\Psi = (\frac{C}{r} + \cancel{D}r)e^{-i\omega t + i\theta + ikz}$$

where $D = 0$ because $\Psi < \infty$ when $r \to \infty$.

The perturbed magnetic field is then

$$\mathbf{B}_1 = \mathbf{B}_{1e} \nabla \Psi = C[-\frac{\hat{r}}{r^2} + i\frac{\hat{\theta}}{r^2} + ik\frac{\hat{z}}{r}]e^{-i\omega t + i\theta + ikz}$$

and at $r = a$,

$$\mathbf{B}_{1e}(r = a) = B_{1ra}[\hat{r} - i\hat{\theta} - ika\hat{z}]e^{i\theta + ikz}.$$

Now, we want to relate ξ_{1ra} and B_{1ra} by the "frozen-in" law. To first-order approximation, let $\hat{n}$ be the direction normal to the perturbed boundary, we have

$$(\hat{n} \cdot \mathbf{B})_1 = 0$$

The equation for the perturbed boundary (Eq.(10.146) of (Bellan 2008)) gives

$$r - \xi_r - a = 0$$

so

$$
\begin{aligned}
\hat{n} &= \frac{\nabla(r-\xi_r-a)}{|\nabla(r-\xi_r-a)|} \\
&= \frac{\hat{r}-\frac{i}{r}\xi_{1ra}\hat{\theta}-ik\xi_{1ra}\hat{z}}{|\hat{r}-\frac{i}{r}\xi_{1ra}\hat{\theta}-ik\xi_{1ra}\hat{z}|} \\
&= \hat{r}-\frac{i}{r}\xi_{1ra}\hat{\theta}-ik\xi_{1ra}\hat{z}
\end{aligned}
$$

where the last equivalence holds because ξ_{1ra}/r and $k\xi_{1ra}$ are both second-order in magnitude.

Therefore we get

$$
(\hat{n}\cdot\mathbf{B})_1 = (\hat{r}-\frac{i}{r}\xi_{1ra}\hat{\theta}-ik\xi_{1ra}\hat{z})\cdot(B_{1r}\hat{r}+B_{0\theta}\hat{\theta}+B_{0z}\hat{z}) = 0
$$

$$
B_{1r} = B_{0\theta}\frac{i\xi_{1ra}}{a}+ik_z\xi_{1ra}B_{0z} = B_{1ra} = \frac{i\xi_{1ra}}{a}[B_{0\theta}+k_z aB_{0z}] = i\xi_{1ra}(\mathbf{k}\cdot\mathbf{B})
$$

where $\mathbf{k} = k_\theta\hat{\theta}+k_z\hat{z} = \frac{m}{r}\hat{\theta}+k_z\hat{z} = \frac{m}{a}\hat{\theta}+k_z\hat{z}$.

Finally, $p+B^2/2\mu_0$ is continuous across perturbed boundary,

$$
\begin{aligned}
\Big(p+\frac{B^2}{2\mu_0}\Big)_{1,interior} &= p_{1i}+\frac{2\mathbf{B}_{oi}\cdot\mathbf{B}_{1i}}{2\mu_0} = \rho_0\omega^2 a\xi_{1ra}-\frac{1}{\mu_0}k^2{B_0}^2 a\xi_{1ra} = a\rho_0\xi_{1ra}(\omega^2-k^2{v_A}^2) \\
\Big(p+\frac{B^2}{2\mu_0}\Big)_{1,exterior} &= 0+\frac{1}{2\mu_0}[{B_{0e}}^2+2\mathbf{B}_{oe}\cdot\mathbf{B}_1 e]_{1,r=a+\xi_r} \\
&= \frac{1}{2\mu_0}\Big[\cancel{{B_{0ze}}^2}+{B_{0\theta a}}^2(\frac{a}{a+\xi_{1r}})^2+2(B_{0\theta}B_{1e\theta}+B_{0z}B_{1ez})\Big]_1 \\
&= \frac{1}{2\mu_0}\Big[-\frac{2\xi_{1ra}{B_{0\theta a}}^2}{a}+2\big[B_{0\theta}(-iB_{1ra})+B_{0z}(-ikaB_{1ra})\big]\Big] \\
&= \frac{1}{2\mu_0}\Big[-\frac{2\xi_{1ra}{B_{0\theta a}}^2}{a}+\frac{2\xi_{1ra}}{a}\big[B_{0\theta}+kaB_{0z}\big]^2\Big]
\end{aligned}
$$

where in one of the equivalence $\frac{\xi_{1ra}}{a}\to 0$,

$$
\Big(\frac{1}{1+\frac{\xi_{1r}}{a}}\Big)^2 \approx -2\frac{\xi_{1ra}}{a}
$$

$$\left(p + \frac{B^2}{2\mu_0}\right)_{1,interior} = \left(p + \frac{B^2}{2\mu_0}\right)_{1,exterior}$$

$$\omega^2 = k^2 {v_A}^2 + \frac{1}{a\mu_0 a^2 \rho_0}[k^2 a^2 {B_{0z}}^2 + 2kaB_{0\theta}B_{0z}]$$

$$\omega^2 = \frac{1}{a\mu_0 a^2 \rho_0}[2k^2 a^2 {B_{0z}}^2 + 2kaB_{0\theta}B_{0z}] = \frac{2k^2 {B_{0z}}^2}{\mu_0 \rho_0}\left[1 + \frac{B_{0\theta}}{kaB_{0z}}\right]$$

For stability, $1 > \left|\frac{B_{0\theta}}{kaB_{0z}}\right| \Rightarrow |ka| > \frac{B_{0\theta}}{B_{0z}}$. Take $|k| = R^{-1}$, where R is the major radius, the stability condition becomes

$$q \equiv \frac{a}{R}\frac{B_{0z}}{B_{0\theta}} > 1$$

and the q the called the safety factor. This is again the Kruskal-Shafranov limit for $m = 1$ kink mode. For sausage mode $m = 0$, the same approach as above can get

$$B_{0\theta} < \sqrt{2}B_{0z}$$

for stability.

Note:

1. This 2-region model can be generalized to 3-region model, which is more realistic compared to experiments. In the liner inertial fusion experiment, there is a mixture of R-T, sausage, kink and many high order modes.
2. In general, the dispersion relation can be written as

$$\omega^2 = (\mathbf{k} \cdot \mathbf{v_A})^2 - \text{destablizing term}$$

where the destablizing term can be gravity $\mathbf{k} \cdot \mathbf{g}$, current I/I_{crit}, tearing mode due to resistivity, etc.

3. If we do the same analysis in Cartesian coordinates (x,y,z), there will be no sausage or kink mode! See HW3.

3-Region Problem

In 572 HW3.1, we have shown that a Cartesian liner is always stable for kink and sausage mode while a cylindrical liner isn't. However, if you also consider the gravity (or equivalently, centrifugal forc, etc.) then you can have magnetic Rayleigh-Taylor (MRT) instability.

Here we show the derivation of dispersion relation in a 3-region problem in Cartesian liner. This method can be found in Chandrasekhar's book on instability, P429. Consider Fig.1 shown in Prof.Lau's paper "Anisotropy and feedthrough in magneto-Rayleigh-Taylor instability". The interface of the regions are $x = \alpha,\ x = \beta$. The governing equation including gravity is

$$\frac{\mathrm{d}}{\mathrm{d}x}\Big\{\rho_0[\omega^2 - (\mathbf{k}\cdot\mathbf{v}_A)^2]\frac{\partial \xi_{1x}}{\partial x}\Big\} - k^2[\rho_0(\omega^2 - (\mathbf{k}\cdot\mathbf{v}_A)^2) + g\frac{\partial \rho_0}{\partial x}]\xi_{1x} = 0$$

If $\rho_0 = \text{const.} \Rightarrow g\frac{\partial \rho_0}{\partial x} = 0$. Within region I, II, and III, the governing equation can be simplified to

$$\frac{\mathrm{d}^2\xi_{1x}}{\mathrm{d}x^2} - k^2\xi_{1x} = 0$$

the general solution of which reads

$$\xi_{1x} = Ae^{-kx} + Be^{kx} = C\sinh k(x-\alpha) + D\sinh k(x-\beta)$$

where A, B, C, D are coefficients. Cancelling out the unphysical solutions in each region, we have

$$\begin{aligned}
\text{I: } \xi_{1x} &= \xi_\beta e^{k(x-\beta)} \\
\text{II: } \xi_{1x} &= \xi_\alpha \frac{\sinh k(x-\beta)}{\sinh k(\alpha-\beta)} + \xi_\beta \frac{\sinh k(x-\alpha)}{\sinh k(\beta-\alpha)} \\
\text{III: } \xi_{1x} &= \xi_\alpha e^{-k(x-\alpha)}
\end{aligned}$$

Note that there's a jump for $\rho_0(x)$ at $x = \alpha, \beta$. Now we can integrate the governing equation across the boundaries,

$$\begin{aligned}
&\int_{x=\alpha^-}^{x=\alpha^+} \mathrm{d}x \Big[\frac{\mathrm{d}}{\mathrm{d}x}\Big\{\rho_0[\omega^2 - (\mathbf{k}\cdot\mathbf{v}_A)^2]\frac{\partial \xi_{1x}}{\partial x}\Big\} - k^2[\rho_0(\omega^2 - (\mathbf{k}\cdot\mathbf{v}_A)^2) + g\frac{\partial \rho_0}{\partial x}]\xi_{1x}\Big] = 0 \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \Rightarrow F_1(\xi_\alpha, \xi_\beta) = 0 \\
&\int_{x=\beta^-}^{x=\beta^+} \mathrm{d}x \Big[\frac{\mathrm{d}}{\mathrm{d}x}\Big\{\rho_0[\omega^2 - (\mathbf{k}\cdot\mathbf{v}_A)^2]\frac{\partial \xi_{1x}}{\partial x}\Big\} - k^2[\rho_0(\omega^2 - (\mathbf{k}\cdot\mathbf{v}_A)^2) + g\frac{\partial \rho_0}{\partial x}]\xi_{1x}\Big] = 0 \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \Rightarrow F_2(\xi_\alpha, \xi_\beta) = 0
\end{aligned}$$

Finally, from F_1 and F_2, we can get the dispersion relation

$$\omega^4 - R\omega^2 + S = 0$$

where R and S are functions of $B_{01}, B_{02}, B_{03}, \rho_{01}, \rho_{02}, \rho_{03}, g, \Delta, k_y$ and k_z.

Next, we can examine the temporal evolution of sinusoidal ripples at interfaces in the form

$$\xi_{\alpha,\beta}(t) = A_1 e^{i\omega_1 t} + A_2 e^{-i\omega_3 t} + A_3 e^{i\omega_3 t} + A_4 e^{-i\omega_3 t}$$

where ω_1 and ω_3 are two eigen mode from the dispersion relation. See the details in the paper. The details of dispersion relation is given in M.Weis, et. al., Phys. Plasmas 21, 122708 (2014), and the coulping of MRT, sausage and kink mode in a cyclindrical liner is given by M.Weis, et. al., Phys. Plasma 22, 032706 (2015).

Note: for RT/MRT,

$$\xi_{RT}(t) \sim \xi_0 e^{\gamma t} < \xi_0 e^{\sqrt{gk}t}$$

the growth rate

$$\gamma_{\max} = \sqrt{gk}$$

If we use $s = \frac{1}{2}gt^2 \Rightarrow \sqrt{2s} = \sqrt{|g|}t$,

$$\xi_{RT}(t) < \xi_0 e^{\sqrt{2ks}}$$

which only depends on wavenumber and distance.

9.7.2 Sausage Mode

The $m = 0$ mode is known as the sausage instability. There is no B_z so we have purely toroidal field. The initial equilibrium is established with radially inward Lorentz force and outward pressure gradient. But it is unstable to interchange due to curvature. When a perturbation that causes the rings to shrink, magnetic field in the plasma increases so that the $\mathbf{J}\times\mathbf{B}$ Lorentz force increases. There is nothing to counter this radially inward force increase, which leads to instability.

The dispersion relation is

$$\omega^2 = -2\frac{p_0}{\rho_0}\frac{k}{R_c^2}$$

Adding B_z to the interior plasma stabilizes sausage instability: the magnetic pressure caused by B_z pushes back to oppose squeezing. The pressure balance at the interface gives

$$p_0 + \frac{B_{0z}^2}{2\mu_0} = \frac{B_\phi^2}{2\mu_0}$$

With the modified dispersion relation

$$\omega^2 = -2\frac{p_0}{\rho_0 a^2} + \frac{B_{0z}^2}{\mu_0 \rho_0 a^2}$$

we have the condition for stability

$$B_{0z}^2 > \frac{1}{2} B_\phi^2$$

THIS IS PROBABLY DUPLICATE WITH THE PREVIOUS SECTION.

9.7.3 Kink Mode

However, even when the sausage mode is suppressed, the configuration is still unstable to the kink mode. This $m = 1$ mode retains circular cross-section of the tube and the perturbation is a kink of the tube into a helix. Without B_z, the system is unstable for all k; with B_z, it is unstable for wavelengths long enough such that the pitch of the perturbation follows the pitch of the helix, i.e. the crests/troughs of the perturbations follow the fieldlines of tube

$$B_\phi / R + k B_Z \geq 0$$

In terms of a twist $\Phi = 2LB_\phi / RB_z$, this criterion is equivalent to

$$k \geq -\frac{\Phi}{2L}$$

The perturbation Lorentz force $\mathbf{j}_1 \times \mathbf{B}_0$ is zero. (???)

9.7.4 Ballooning Mode

The ballooning instability is a type of internal pressure-driven plasma instability usually seen in tokamak fusion power reactors or in space plasmas (Hameiri, Laurence, and Mond 1991). It is important in fusion research as it determines a set of criteria for the maximum achievable plasma beta. The name refers to the shape and action of the instability, which acts like the elongations formed in a long balloon when it is squeezed. In literature, the structure of these elongations are commonly referred to as "fingers".

The narrow fingers of plasma produced by the instability are capable of accelerating and pushing aside the surrounding magnetic field in order to cause a sudden, explosive release of energy. Thus, the instability is also known as the explosive instability.

The interchange instability can be derived from the equations of the ballooning instability as a special case in which the ballooning mode does not perturb the equilibrium magnetic field.(Hameiri, Laurence, and Mond 1991) This special limit is known as the Mercier criterion.

```
KeyNotes.plot_balloon()
```

9.7.5 Tearing Mode

```
KeyNotes.plot_tearing()
```

Tearing mode is closely related to magnetic reconnection (Section 9.8). The reconnection process is very important because it is one of the main way of burst energy transformation. It is known that in collisionless systems current sheets are unstable against tearing instability, a process where the current tends to collapse into filaments. The tearing instability produces magnetic islands that then interact and merge together giving rise to a nonlinear instability phase, where the reconnection process is enhanced. See (Bellan 2008) P413 for more.

Linear Tearing Mode

Consider the interface between two plasmas containing magnetic fields of different orientations. The simplest imaginable field configuration is that illustrated in **?@fig-tearing_config_simple**. Here, the field varies only in the x-direction, and points only in the y-direction. The field is directed in the $-y$-direction for $x < 0$, and in the $+y-$direction for $x > 0$. The interface is situated at $x = 0$. The sudden reversal of the field direction across the interface gives rise to a z-directional current sheet at $x = 0$.

With the neglect of plasma resistivity, the field configuration shown in **?@fig-tearing_config_simple** represents a stable equilibrium state, assuming, of course, that we have normal pressure balance across the interface. But, does the field configuration remain stable when we take resistivity into account? If not, we expect an instability to develop which relaxes the configuration to one possessing lower magnetic energy. As we shall see, this type of relaxation process inevitably entails the breaking and reconnection of magnetic field lines, and is, therefore, termed *magnetic reconnection*. The magnetic energy releases during the reconnection process eventually appears as plasma thermal energy. Thus, magnetic reconnection also involves plasma heating.

In the following, we shall outline the standard method for determining the linear stability of the type of magnetic field configuration shown in **?@fig-tearing__config__simple**, taking into account the effect of plasma resistivity. We are particularly interested in plasma instabilities which are stable in the absence of resistivity, and only grow when the resistivity is non-zero. Such instabilities are conventionally termed *tearing modes*. Since magnetic reconnection is, in fact, a nonlinear process, we shall then proceed to investigate the nonlinear development of tearing modes.

The equilibrium magnetic field is written

$$\mathbf{B}_0 = B_{0y}(x)\hat{y}$$

where $B_{0y}(-x) = -B_{0y}(x)$. There is assumed to be non-equilibrium plasma flow. The linearized equations of resistive-MHD, assuming incompressible flow, take the form

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}_0) + \frac{\eta}{\mu_0}\nabla^2 \mathbf{B} \tag{9.12}$$

$$\rho_0 \frac{\partial \mathbf{v}}{\partial t} = -\nabla p + \frac{(\nabla \times \mathbf{B}) \times \mathbf{B}_0}{\mu_0} + \frac{(\nabla \times \mathbf{B}_0) \times \mathbf{B}}{\mu_0} \tag{9.13}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{9.14}$$

$$\nabla \cdot \mathbf{v} = 0 \tag{9.15}$$

Here, ρ_0 is the equilibrium plasma density, $\mathbf{B}$ is the perturbed magnetic field, $\mathbf{v}$ the perturbed plasma velocity, and p the perturbed plasma pressure. The assumption of incompressible plasma flow is valid provided that the plasma velocity associated with the instability remains significantly smaller than both the Alfvén velocity and the sonic velocity.

Suppose that all perturbed quantities vary like

$$A(x, y, z, t) = A(x)e^{iky+\gamma t} \tag{9.16}$$

where γ is the instability growth rate. The x-component of Equation 9.12 and the z-component of the curl of Equation 9.13 reduce to ???

$$\gamma B_x = ikB_{0y}v_x + \frac{\eta}{\mu_0}\Big(\frac{\mathrm{d}^2}{\mathrm{d}x^2} - k^2\Big)B_x \tag{9.17}$$

$$\gamma\rho_0\Big(\frac{\mathrm{d}^2}{\mathrm{d}x^2} - k^2\Big)v_x = \frac{ikB_{0y}}{\mu_0}\Big(\frac{\mathrm{d}^2}{\mathrm{d}x^2} - k^2 - \frac{B_{0y}''}{B_{0y}}\Big)B_x \tag{9.18}$$

respectively, where use has been made of Equation 9.14 and Equation 9.15. Here, $'$ denotes $\mathrm{d}/\mathrm{d}x$.

It is convenient to normalize Equation 9.17 and Equation 9.18 using a typical magnetic field-strength, B_0, and a typical scale-length, a. Let us define the *Alfvén time-scale*

$$\tau_A = \frac{a}{V_A}$$

where $V_A = B_0/\sqrt{\mu_0\rho_0}$ is the Alfvén velocity, and the resistive diffusion time-scale

$$\tau_R = \frac{\mu_0 a^2}{\eta}$$

The ratio of these two time-scale is the *Lundquist number*:

$$S = \frac{\tau_R}{\tau_A}$$

Let $\psi = B_x/B_0$, $\phi = i\,k\,V_y/\gamma$, $\bar{x} = x/a$, $F = B_{0y}/B_0$, $F' \equiv dF/\mathrm{d}\bar{x}$, $\bar{\gamma} = \gamma\,\tau_A$, and $\bar{k} = k\,a$. It follows that

$$\bar{\gamma}\,(\psi - F\,\phi) = S^{-1}\left(\frac{\mathrm{d}^2}{\mathrm{d}\bar{x}^2} - \bar{k}^2\right)\psi \tag{9.19}$$

$$\bar{\gamma}^2\left(\frac{\mathrm{d}^2}{\mathrm{d}\bar{x}^2} - \bar{k}^2\right)\phi = -\bar{k}^2\,F\left(\frac{\mathrm{d}^2}{\mathrm{d}\bar{x}^2} - \bar{k}^2 - \frac{F''}{F}\right)\psi \tag{9.20}$$

The term on the right-hand side of Equation 9.19 represents plasma resistivty, whilst the term on the left-hand side of Equation 9.20 represents plasma inertia.

It is assumed that *the tearing instability grows on a hybrid time-scale which is much less than τ_R but much greater than τ_A*. It follows that

$$
\begin{aligned}
\tau_A \ll \gamma^{-1} &\ll \tau_R \\
\gamma\tau_A \ll 1 &\ll \gamma\tau_R \\
\bar{\gamma} \ll 1 &\ll S\,\bar{\gamma}
\end{aligned}
$$

Thus, throughout most of the plasma we can neglect the right-hand side of Equation 9.19 and the left-hand side of Equation 9.20, which is equivalent to the neglect of plasma resistivity and inertia. In this case, the two equations reduce to

$$\phi = \frac{\psi}{F} \tag{9.21}$$

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}\bar{x}^2} - \bar{k}^2\psi - \frac{F''}{F}\psi = 0 \tag{9.22}$$

Equation 9.21 is simply the flux freezing constraint, which requires the plasma to move with the magnetic field. Equation 9.22 is the linearized, static force balance criterion $\nabla \times (\mathbf{j} \times \mathbf{B}) = 0$. These two equations are known collectively as the equations of ideal-MHD, and are valid throughout virtually the whole plasma. However, it is clear that these equations *break down* in the immediate vicinity of the interface, where $F = 0$ (i.e. where the magnetic field reverses direction). Witness, for instance, the fact that the normalized "radial" velocity, ϕ, becomes infinite as $F \to 0$, according to Equation 9.21.

The ideal-MHD equations break down close to the interface because the neglect of plasma resistivity and inertia becomes untenable as $F \to 0$. Thus, there is a thin layer, in the immediate vicinity of the interface, $\bar{x} = 0$, where the behaviour of the plasma is governed by the full MHD equations, Equation 9.19 and Equation 9.20. We can simplify these equations, making use of the fact that $\bar{x} \ll 1$ and $\mathrm{d}/\mathrm{d}\bar{x} \gg 1$ in a thin layer, to obtain the following layer equations:

$$\bar{\gamma}\,(\psi - \bar{x}\,\phi) = S^{-1}\frac{\mathrm{d}^2\psi}{\mathrm{d}\bar{x}^2} \tag{9.23}$$

$$\bar{\gamma}^2\,\frac{\mathrm{d}^2\phi}{\mathrm{d}\bar{x}^2} = -\bar{x}\,\frac{\mathrm{d}^2\psi}{\mathrm{d}\bar{x}^2} \tag{9.24}$$

Note that we have redefined the variables ϕ, $\bar{\gamma}$, and S, such that $\phi \to F'(0)\,\phi$, $\bar{\gamma} \to \gamma\,\tau_H$, and $S \to \tau_R/\tau_H$. Here,

$$\tau_H = \frac{\tau_A}{k\,a\,F'(0)}$$

is the *hydromagnetic time-scale.*

The tearing mode stability problem reduces to solving the non-ideal-MHD layer equations, Equation 9.23 and Equation 9.24, in the immediate vicinity of the interface, $\bar{x} = 0$, solving the ideal-MHD equations, @Equation 9.21 and Equation 9.22, everywhere else in the plasma, matching the two solutions at the edge of the layer, and applying physical boundary conditions as $|\bar{x}| \to \infty$. This method of solution was first described in a classic paper (Furth, Killeen, and Rosenbluth 1963). The steps are listed in the farside note. The procedure is similar to the 3 layer solution of the K-H instability. After some maths, tearing mode dispersion relation is given as

$$\gamma = \left[\frac{\Gamma(1/4)}{2\pi\,\Gamma(3/4)}\right]^{4/5} \frac{(\Delta')^{4/5}}{\tau_H^{2/5}\,\tau_R^{3/5}}$$

where

$$\begin{aligned}
\Delta &= 2\pi\,\frac{\Gamma(3/4)}{\Gamma(1/4)}\,S^{1/3}\,Q^{5/4} \\
Q &= \gamma\,\tau_H^{2/3}\,\tau_R^{1/3}
\end{aligned}$$

The tearing mode is unstable whenever $\Delta' > 0$, and grows on the hybrid time-scale $\tau_H^{2/5}\,\tau_R^{3/5}$.

9.8 Magnetic Reconnection

There are several key elements for understanding the physics of reconnection at a deeper level (Ji et al. 2022). First, the frozen-in properties in an electron-ion plasma are associated with the electron fluid due to its light weight. This is expressed in terms of the time changing rate of magnetic flux (Equation 4.24) through an arbitrary area, $\mathbf{S}$ (enclosed by loop l), convecting with the electron flow as $\mathrm{d}\Phi/\mathrm{d}t = \oint(\mathbf{E} + \mathbf{u}_e \times \mathbf{B})\cdot\mathrm{d}l = 0$. Thus, the frozen-in condition is regulated by the electron momentum equation (i.e. the generalized Ohm's law)

$$\underbrace{\mathbf{E} + \mathbf{V}_e \times \mathbf{B}}_{\text{Ideal}} = \underbrace{\mathbf{R}_\text{col}}_{\text{Collisional}} \underbrace{-\frac{\nabla\cdot\mathbf{P}_e}{en} - \frac{m_e}{e}\frac{\mathrm{d}\mathbf{u}_e}{\mathrm{d}t}}_{\text{Kinetic}} \tag{9.25}$$

where $\mathbf{R}_\text{col}$ is the collisional force per electron charge. In a fully ionized collisional plasma, $\mathbf{R}_\text{col} \approx \eta\mathbf{j}$ where η is the resistivity due to Coulomb collisions. If the terms on the right-hand side of Equation 9.25 are negligible, then $\mathrm{d}\Phi/\mathrm{d}t \approx 0$ and the magnetic flux is "frozen-in" to the electron flow. For the generic reconnection layer illustrated in the middle panel of **?@fig-reconnection_configurations**, deviations from ideal evolution occur within the "diffusion region" (blue), where either finite resistivity or kinetic effects (electron inertia and pressure tensor) are important. Within the diffusion region, field lines converging from opposite sides of the layer can change connectivity.

The next key element for understanding the physics of magnetic reconnection is to grasp the remarkable global consequences of changing field line connectivity within a localized region. In particular, the newly reconnected field lines have a large curvature $\mathbf{B}\cdot\nabla\mathbf{B}$ which produces a tension force — closely analogous to a stretched rubber band. Allowing for the possibility of pressure anisotropy, the MHD momentum equation perpendicular to magnetic field is

$$\rho\frac{\mathrm{d}\mathbf{u}_\perp}{\mathrm{d}t} = -\nabla_\perp\left(p_\perp + \frac{B^2}{2\mu_0}\right) + \left(1+\frac{p_\perp - p_\parallel}{B^2/\mu_0}\right)\mathbf{B}\cdot\nabla\mathbf{B} - \mathbf{F}_{\mathrm{col}\perp} \tag{9.26}$$

where $\mathbf{F}_{\mathrm{col}}$ is due to collisions between ions (viscous force) and between ions and neutral particles (frictional force). For newly reconnected field lines, the magnetic tension in Equation 9.26 drives an outflow jet approaching the Alfvén speed $V_{\mathrm{out}} \approx V_A \equiv B_0/\sqrt{\mu_0\rho}$ (B_0 is the reconnecting field component). The resulting deficit in magnetic pressure pushes new field lines into the diffusion region with a maximum inflow velocity $V_{\mathrm{in}} \sim (0.01 \rightarrow 0.1)V_A$. As this process continues, the larger stressed region is relaxed, leading to a reconfiguration of the global magnetic field on fast Alfvénic time scales $\sim L/V_A$. Assuming the diffusion regions remain small in comparison to the global scales, most plasma enters the flow jet across the magnetic separatrices as illustrated in **?@fig-reconnection_configurations**. This inflow is a consequence of changing the field line connectivity within the diffusion region, which causes the entire extent of the reconnected field lines to join the outflow. In this limit, the majority of the energy release is associated with the relaxation of field-line tension within outflow jets over long distances. Since the spatial extent of these jets is limited only by the macroscopic configuration, they are one of the most prominent signatures of reconnection in both in-situ and remote-sensing observations. Our primary focus is on situations where the available magnetic energy to drive reconnection is comparable or larger than the initial plasma thermal energy, corresponding to $\beta \equiv p/(B_0^2/2\mu_0) \lesssim 1$. In these regimes, global relaxation of field-line tension is the ultimate "engine" for reconnection and is essentially "ideal", thus operating in a similar manner for most applications, but with a few important exceptions. Plasma heating preferentially along the magnetic field ($p_\parallel \gg p_\perp$) can weaken the magnetic tension force in Equation 9.26, whereas in partially ionized regimes the jet formation is more complicated due to interactions with neutrals. In addition, non-ideal kinetic physics may persist along magnetic separatrices (see red lines in **?@fig-reconnection_configurations**) to larger distances, while in very large systems shocks may form along these boundaries and play a role in the energy conversion.

In contrast to the ideal physics driving the jet, the non-ideal terms within the diffusion region are intimately dependent upon the plasma conditions and spatial scales, and thus, a variety of different regimes are possible, as illustrated in the various panels of **?@fig-reconnection_configurations**. Since the outflow is always energetically limited to V_A in a quasi-steady state, mass conservation implies that the geometry of the diffusion region determines the dimensionless reconnection rate, $R \equiv u_{\mathrm{in}}/V_A \approx \Delta/L$ where Δ is the diffusion region thickness. The current understanding of the diffusion region physics has evolved over more than 60 years in three main stages, yet a full understanding remains elusive for large space and astrophysical problems.

9.8.1 Reconnection Rate

Reconnection rate quantifies the amount of magnetic flux transportation through the X-line. There are two kinds of reconnection rate:

1. The dimensionless reconnection rate, defined as the ratio between the inflow magnetic flux and the outflow magnetic flux:

$$R \equiv \frac{B_{\mathrm{in}} u_{\mathrm{in}}}{B_{\mathrm{out}} u_{\mathrm{out}}} \approx \frac{u_{\mathrm{in}}}{u_{\mathrm{out}}} \tag{9.27}$$

The approximation is usually valid when we take the field values far away from the reconnection site. Observationally, this ratio is on the order of $\mathcal{O}(0.1)$. There has been some argument based on geometry to explain this universal ratio in recent years.

2. The dimensional reconnection rate, defined as the difference between the inflow magnetic flux and the outflow magnetic flux:

$$R \equiv B_{\mathrm{in}} u_{\mathrm{in}} - B_{\mathrm{out}} u_{\mathrm{out}} \tag{9.28}$$

When we think of a 2D X-line configuration in the X-Z plane, this is equivalent to the out-of-plane electric field E_y.

9.8.2 Hall Field

Let X be the inflow direction, Z the outflow direction, and Y the out-of-plane direction. During magnetic reconnection,

- a quadrupolar out-of-plane or *Hall* magnetic field B_y is formed by the field aligned currents $j_\parallel$ that flow in the vicinity of the magnetic separatrices when ion and electrons decouple on the length scales less than the ion inertial length (Sonnerup 1979). The quadruple Hall magnetic field was analyzed as a tearing mode eigenmode of the current layer (Terasawa 1983). The Hall effect introduces a coupling between the tearing mode perturbations (2D configuration in-plane $\mathbf{B}$ components) and shear Alfvén mode perturbations (out-of-plane $\mathbf{B}$ component).
- In the context of collisionless plasmas, ion-electron collisions are absent so that ions and electrons behave differently at micro-scale. On the ion scale, ions start to become demagnetized whereas electrons are still coupled to the magnetic field. The difference in the ion and electron motion produces *charge separation*, *Hall electric fields* and the *associated electric current.* (On the even smaller electron-scale, electrons are demagnetized

and correspond to the "breaking" of the field line.) The spatial variation of Hall electric fields is mostly in X, normal to and pointing towards the current layer. Relevance of Hall effects in reconnection is considered important at the ion inertial scale, providing a source for non-ideal reconnection electric field. (In observations, the perpendicular scale of the Hall fields and current layer can be as small as electron scales as well?) The parallel scale of the Hall fields are much larger. Equivalently, Hall fields are associated with a wave vector $\mathbf{k}$ so that $k_x \gg k_z$ and $1/k_x$ on the ion-scale.

The Hall fields are continuously produced quasi-steady structures accompanied with a timescale much longer than the ion gyroperiod. Their associated kinetic and magnetic energies propagate away from the reconnection layers carried by Alfvén waves. Numerical simulations suggest that it is primarily kinetic Alfvén waves (KAW, Section 7.7) which are excited in the vicinity of reconnection regions (Rogers+, 2001; Shay+, 2011), but further downstream in the exhaust it is primarily shear Alfvén waves (Gurram, Egedal, and Daughton 2021).

In observations, signatures of Hall magnetic fields are used as a working definition of collisionless magnetic reconnection in the Earth's and planetary magnetosphere.

(Dai and Wang 2023) emphasized that the physics underlying the Hall effect is similar to that in KAW:

- Both can be described by the two-fluid equations.
- The temporal and spatial scales of Hall fields are exactly the same as those of KAW.
- The polarity of the Hall electric field (in X) and the Hall magnetic field (in Y) is also consistent with that of KAW.
- The ratio E_x/B_y of Hall fields is on the order of V_A, similar to KAW.
- The current system of Hall fields are in the XZ-plane.

As the perpendicular wavelength of Alfvén wave becomes comparable to the ion gyroradius, the ions cannot follow the $\mathbf{E}\times\mathbf{B}$ drift. This is because the wave electric field now is non-uniform in the trajectory of ion gyromotion. As a consequence, the ion transverse motion is modified by the finite-Larmor-radius (FLR) effect. Considering the kinetic correction, the ion actually follows $\mathbf{E} = \mathbf{V}_i \times \mathbf{B} + \frac{1}{n_i q_i}\nabla\cdot\mathbf{P}_i$ in the transverse motion. The FLR effect is manifested in the ion pressure gradient term: the FLR effect is equivalent to a finite temperature for the ions on the fluid level. At the wavelength of ion-scale wavelength, electrons cannot follow ions in the transverse motion. The difference in the transverse motion of ions and electrons produces a charge separation and coupling to the electrostatic mode, and a quasi-electrostatic electric field is formed in the transverse direction. To remain quasi-neutral, electrons quickly move in the direction parallel to the magnetic field. In the parallel direction, these electrons undergo a force balance between a small parallel electric field and the electron pressure gradient in the case of KAW. The parallel motion of electrons created a field-aligned current (FAC) that is a distinct nature of Alfvén mode. The FAC induces a wave magnetic field perpendicular to the DC (background) magnetic field. A new physical scale length r_{iL}, the ion gyroradius based on the electron temperature, for the current layer and outflow layer in fast reconnection. Through the diffusion region near the X-line, intense KAW turbulence is observed, suggesting that magnetic reconnection is a *source* of KAW. Thus KAW has been invoked as ingredients for fast reconnection.

In the direction parallel to the magnetic field ($\hat{z}$), B_y reverses but E_x remains its sign. This feature is related to the *superposition* of two KAW mode waves that propagate in the opposite direction and outward from the reconnection region. In the perpendicular direction, the Hall fields appears to exist only for one wavelength along $\hat{x}$. This is related to the eigenmode (standing mode) structure of KAW in the current layer.

In the literature, the Hall electric field is usually considered to arise from the $\mathbf{j} \times \mathbf{B}/n_e q$ term in the context of the generalized Ohm's law Equation 5.17. The Hall term starts to become important at the ion inertial scale, contributing to the Hall electric field. Interestingly, the $\mathbf{j} \times \mathbf{B}$ also represents a major force term in the one-fluid MHD momentum equation. The $\mathbf{j} \times \mathbf{B}$ force could be important on the large MHD-scale and the argument for the origin of Hall electric field becomes ambiguous; from the MHD momentum equation, $\mathbf{j} \times \mathbf{B}$ term is equivalent to the ion inertial effect, ion pressure effect and the electron pressure effect:

$$\rho \frac{d\mathbf{u}}{dt} + \nabla \cdot (\overleftrightarrow{P}_i + \overleftrightarrow{P}_e) = \mathbf{J} \times \mathbf{B} \tag{9.29}$$

which is derived from the sum of the two-fluid momentum equations multiplied by the mass of each species. The one-fluid MHD momentum equation Equation 9.29 together with the general Ohm's law is nearly equivalent (with some extent of reduction?) to the two-fluid momentum for two species. The Hall electric field is equivalently described by either set of equations.

From the ion momentum equation

$$\mathbf{E} + \mathbf{u}_i \times \mathbf{B} = \frac{1}{n_i q_i} \nabla \cdot \overleftrightarrow{P}_i + \frac{m_i}{q_i} \frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t} \tag{9.30}$$

The perpendicular component of the non-ideal electric field $\mathbf{E} + \mathbf{u}_i \times \mathbf{B}$ is the Hall electric field at the ion kinetic scale; as seen on the RHS, two sources of Hall electric fields are the ion inertial term and the ion pressure gradient. From the KAW perspective, the Hall electric field is supported by the ion pressure gradient. In the regime of the temporal and spatial scale of KAW, the ion inertial term is small and on the order of $\mathcal{O}(\omega^2/\Omega_i^2)$.

If the ion inertial term is neglected, Equation 9.30 becomes

$$\mathbf{E} + \mathbf{u}_i \times \mathbf{B} = \frac{1}{n_i q_i} \nabla \cdot \overleftrightarrow{P}_i$$

The $\mathbf{u}_i \times \mathbf{B}$ term is kept since the ion flow may not be neglected near the X-line.

9.8.3 Kinetic Signatures of Diffusion Region

(The following part is from [Kinetic signatures of the region surrounding the X-line in asymmetric (magnetopause) reconnection] and [Fluid and kinetics signatures of reconnection at the dawn tail magnetopause-Wind observations].)

Electron diffusion region (EDR) is defined to be the electron-scale region surrounding the X-line in which magnetic connectivity is ultimately broken. Note that this definition is fundamentally non-local in nature.

1. enhanced dissipation [e.g., Zenitani+, 2012]
2. non-gyrotropic particle behavior [e.g., Scudder+, 2008; Aunai+, 2013; Swisdak 2016]
3. electron distribution functions [e.g., Chen+, 2008, Ng+, 2011]
4. In the two inflow regions, we expect a reduction in total pressure, magnetic field B_L and the plasma density n_i relative to ambient upstream conditions.

Some signatures also exist downstream of the diffusion region and along the separatrices. In practice people use a complementary approach for the identification.

Specifically for EDR, it should exhibit a number of properties:

1. the violation of the electron frozen-in condition
2. non-gyrotropic electron distributions [e.g., Scudder+, 2008; Aunai+, 2013; Swisdak 2016]
3. enhanced dissipation [Zenitani+, 2011]

A simple and practical indicator of EDR is the presence of a sunward pointing E_N at the midplane (called the "shoulder") as this signature coincides with the region of enhanced dissipation, non-gyrotropic electrons at midplane, and counterstreaming electron beams due to electron meandering orbits around the X-line. This EN signature is straightforward to measure experimentally because it is the largest component of the electric field at the magnetopause.

The crescent shape distribution is not as localized as the E_N shoulder especially for ion.

At the dayside magnetopause, reconnection between magnetosheath and geomagnetic field lines is expected to produce

1. a finite magnetic field component normal to the magnetopause, B_N;
2. Alfvénic plasma flow acceleration associated with a rotational discontinuity at the magnetopause;
3. ion distributions on reconnected field lines consisting of a mixture of magnetosheath and magnetosphere populations where the transmitted magnetosheath population has a "D-shaped" distribution, with a low-energy cutoff at the deHoffmann-Teller velocity [Cowley, 1982];
4. reflected ions in the layers adjacent to the magnetopause;
5. opposite streaming along reconnected field lines of outgoing magnetospheric electrons and incoming magnetosheath electrons, resulting in large parallel electron heat flux;

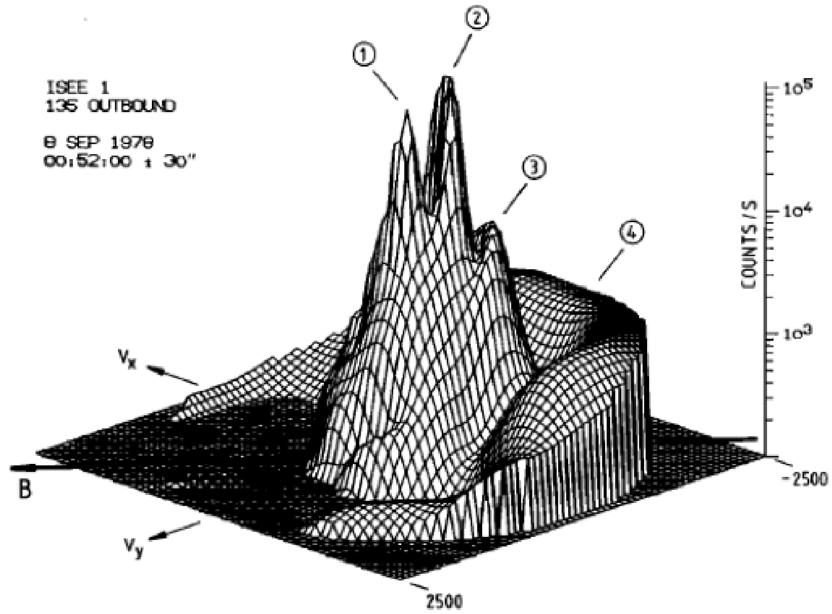

Fig. 12. **Relief plot of two-dimensional count rate distributions in the GSE (v_x, v_y) plane for velocities up to 2500 km s^{-1}, measured just after the final magnetopause crossing on September 8, 1978. The plot is based on a one-minute accumulation of count rates in 16 velocity and azimuth channels. The three peaks near the center represent (1) the main magnetosheath plasma, (2) ions which have been reflected at the magnetopause, and (3) a small contribution of He^+ ions, respectively. The crescent-shaped distribution at high energies (4) is interpreted as ions of magnetospheric (ring current) origin leaking out along reconnected field lines. They are seen to move preferentially antiparallel to the projected magnetic field B, as expected for a crossing north of the X line (Figure 1*b*). The abrupt outer termination of the distribution of energetic ions reflects the high-energy cutoff of the instrument at 40 keV.**

Figure 9.7: Ion crescent shape distribution at Earth's magnetopause revealed by ISEE satellite at 1981.

6. an offset between the ion and electron edges at the inner boundary of the low-latitude boundary layer (LLBL) due to a time-of-flight effect resulting from the fact that entering magnetosheath electrons have much higher parallel speeds than ions while their transverse motions are the same.

9.8.4 D-shape Distribution

The kinetic description of reconnection can also be quantitatively verified. An important kinetic signature is the "D-shaped" ion distribution. Magnetosheath particles can either be reflected at or cross the magnetopause. In the deHoffmann-Teller frame of reference in which the electric field vanishes, only ions traveling toward the magnetopause will cross the magnetopause. Thus, when viewed in the spacecraft frame, only magnetosheath ions with parallel velocity greater than the deHoffmann-Teller velocity can be seen earthward of the magnetopause, resulting in a "D-shaped" distribution. [Cowley, 1982]

The existence of a finite BN at the MP requires field lines on both sides of the MP to move together. In this scenario there must be a reference frame (the HT frame) which slides along the MP with the field-line velocity. In this frame, the convective electric field vanishes ($\mathbf{E}_c = -\mathbf{v} \times \mathbf{B} = 0$), i.e. the flows are field aligned on the two sides of the MP.

In practice, the HT frame for a set of plasma and field measurements across the MP can be found as the reference frame in which the mean square of the convective electric field, $D = \langle |(\mathbf{v} - \mathbf{V}_{HT}) \times \mathbf{B}|^2 \rangle$, is as small as possible. The angle bracket $< ... >$ denotes an average of an enclosed quantity over a set of measurements. The velocity $\mathbf{v}$ for which $D(\mathbf{v})$ is a minimum is the deHoffmann-Teller velocity, $\mathbf{V}_{HT}$. The ratio D/D_0, where $D_0 = \langle |(\mathbf{v} \times \mathbf{B}^| 2 \rangle$, is used as a measure of the quality of the HT frame. For a good HT frame D/D_0 should be small ($\ll 1$).

In another paper [Characteristics of the flank magnetopause: Cluster observations], $\mathbf{V}_{HT}$ is defined by minimizing $\mathbf{E}' = \mathbf{E} + \mathbf{V}_{HT} \times \mathbf{B}$, and the correlation between $\mathbf{E}_c = -\mathbf{v} \times \mathbf{B}$ and $\mathbf{E}_{HT} = -\mathbf{V}_{HT} \times \mathbf{B}$ describes how well the frame is determined.

9.8.5 Classification of Discontinuity Types

In a fluid description, the magnetopause can be described either as a tangential discontinuity (TD) or as a rotational discontinuity (RD). A TD implies a complete separation of two plasma regimes (in this case the magnetosheath on one side and the magnetosphere on the other side). The boundary as a whole may move, but there is no transport of plasma across the discontinuity, and there is no magnetic field along the boundary normal. An RD-like magnetopause, on the other hand, implies transport across the boundary and a normal magnetic field, and indicates the presence of reconnection. In the vicinity of the X-line, the plasma flow is Alfvénic, i.e., the Walén relation is satisfied:

$$\mathbf{v} - \mathbf{V}_{HT} = \pm \mathbf{V}_A$$

where $\mathbf{v}$ is the plasma velocity, $\mathbf{V}_{HT}$ is the deHoffmann-Teller frame velocity, and $\mathbf{V}_A$ is the local Alfvén velocity.

In some other literature, a more complete form is written as

$$\mathbf{v} - \mathbf{V}_{HT} = \pm(1-\alpha)^{1/2}\mathbf{V}_A$$

which takes the pressure anisotropy factor $\alpha = (p_\parallel - p_\perp)\mu_0/B^2$ into account.

This is often used to classify the discontinuity type of the magnetopause. For an RD, the flow across the boundary is proportional to the normal magnetic field, i.e., $v_n \propto B_n$. A positive (negative) slope of the regression means that normal magnetic field and flow have the same (opposite) signs. At the magnetopause, we can assume that the flow is inward, i.e., from the magnetosheath into the magnetosphere.

9.8.6 Reconnection Efficiency

This is very confusing. From [Kivelson+ 1997]:

> There is an interesting link between flow speeds over the polar cap and reconnection efficiency at the nose of its magnetopause. This follows from the fact that the voltage drop across the magnetosphere is a fraction of the voltage drop across the same distance in the corotating flow upstream. Thus, the convection electric field within the magnetosphere is a fraction of the corotation electric field determined by the efficiency of reconnection. The model allows us to estimate the convective flow speeds of plasma over the polar cap if the reconnection efficiency is known, and conversely allows us to estimate the reconnection efficiency if the flow speed over the polar cap is known.

Let me think in this way: the convective electric field in the upstream, $\mathbf{E} = -\mathbf{v}\times\mathbf{B}$, describes how many magnetic field lines are moving into the reconnection site at a given point. The integral of electric field along the X-line describes how many magnetic field lines are getting reconnected into the magnetosphere. This is not the same as electric potential drop, where in the simplest cases requires a static EM field. Because of the difficulty of finding the separatrices, we use an indirect approach assuming equi-potential field lines to calculate the electric field along the magnetosphere boundary curve away from the reconnection site.

Be careful for the word "cross polar cap potential" here. This may be a misleading term, since how can one define an electric potential in a EM field?

9.8.7 Dissipation Mechanisms

Perhaps the most import problem associated with reconnection is the understanding of the mechanisms by which the magnetic field can dissipate its energy, and subsequently produce particle heating and acceleration [Huba 1979]. Since many field reversed plasmas are essentially collisionless, instabilities are likely to play an important role in the dissipation process.

A macro-instability (MHD type), like the tearing mode can dissipate magnetic energy even in a collisionless plasma via electron inertia, electron or ion Landau damping. However, the collisionless tearing mode saturates at a very small amplitude. Micro-instabilities can also dissipate magnetic energy by producing an anomalous resistivity which can either dissipate the magnetic energy directly or enhance the growth of the tearing mode.

In any case, the kinetic physics corresponding to the fast rate of collisionless reconnection is a long-standing question. *Non-gyrotropy particle distributions* and the associated *off-diagonal term of the pressure tensor* have been considered to provide the reconnection electric field at electron-scale and ion-scale.

The *lower-hybrid-drift instability* (LHDI, Section 9.9) is often considered to be an important microinstability for dissipation near the diffusion region.

9.8.8 Sweet-Parker Solution

From Yi-Hsin Liu's presentation of space weather.

mass conservation:

$$\nabla\cdot(n\mathbf{V})=0\Rightarrow\ V_{in}L\simeq V_{out}\delta$$

momentum conservation:

$$\frac{\mathbf{B}\cdot\nabla\mathbf{B}}{4\pi}=nm_i\mathbf{V}\cdot\nabla\mathbf{V}\Rightarrow V_{out}\simeq\frac{B}{\sqrt{4\pi nm_i}}=V_A$$

normalized reconnection rate

$$R\equiv\frac{V_{in}}{V_A}\sim\frac{\delta}{L}$$

However, this model has a small δ/L, the rate is too small to explain the time-scales in solar flare. To explain the flares, it requires $R\sim 0.1$.

9.8.9 Petschek Solution

Reconnection in this model is much larger because $R \sim \delta/L$ goes up. However, this is not a self-consistent solution.

In PIC simulations, the diffusion region is localized like the Petschek solution. The reason we use PIC instead of MHD is that PIC captures the key physics that breaks the frozen-in condition in nature. In the GEM challenge study, PIC, hybrid, Hall-MHD, MHD with a localized artificial resistivity all give similar $R \sim 0.1$ fast reconnection rate in disparate systems. In terms of the pressure descriptions from the moment equations, it has been found that 6-moment cannot trigger fast reconnection while 10-moment can, which essentially indicates that the off-diagonal terms in the pressure tensor is critical in getting the right physics.

One possible explanation for this 0.1 rate comes from geometrical consideration. In the small δ/L limit, $R \sim \delta/L \sim 0$. In the large δ/L limit, $\delta/L \to 1$, $R \to 0$. It turns out that there should be an optimized $R_{\max}$ in between. In the large δ/L limit,

$$\frac{\mathbf{B}\cdot\nabla\mathbf{B}}{4\pi} \simeq \frac{\nabla(B^2)}{8\pi} + nm_i\mathbf{V}\cdot\nabla\mathbf{V}$$

At the inflow region, the large angle decrease the reconecting magnetic field. At the outlfow region, it decrease the outflow speed. Both of them cause R to decrease. Constraints imposed at the inflow and outflow region (upper) bound the rate! In [Liu+, PRL 2017], an analytical expression of R is given by

$$R \equiv \frac{V_{in}}{V_A} \simeq \frac{\delta}{L}\cdot \underbrace{\left[\frac{1-(\delta/L)^2}{1+(\delta/L)^2}\right]}_{\text{reduction of reconnection B}} \cdot \underbrace{\sqrt{1-\left(\frac{\delta}{L}\right)^2}}_{\text{reduction of } V_{out}}$$

Reconnection tends to proceed near the most efficient state with $R \sim \mathcal{O}(0.1)$. Nicely, the rate is insensitive to δ/L near this state (BECAUSE THE SLOPE IS SMALL???).

Amitava Bhattacharjee argued that scaling is a controversial subject:

1. How does it scale with ion/electron skin depth, resistivity, plasma beta, and system size?
2. Is the reconnection rate insensitive to the details of the electron layer (current sheet layer), and controlled by ions?

- The whistler waves generate an out-of-plane quadrupolar magnetic field.
- The ratio of the horizontal electron outflow to the horizontal magnetic perturbation scales as k for the dispersive whistler (or kinetic Alfvén) wave. (WHICH WAVE IS IT? I have been confused by the relation between whistler waves ann KAWs. Both contains a region with super-Alfvénic phase speed for large k and especially with frequency between Ω_i and Ω_e; whister wave involves electrons whereas KAW can only rely on ions.)

From the GEM Challenge perspective:

- Reconnection is insensitive to the mechanism that breaks field lines (electron inertia or resistivity). The length of the reconnection layer $\Delta_i \sim 10d_i$. Reconnection rate is a "universal constant", $V_in \approx 0.1V_A$.
- In the presence of Hall currents, whistler waves mediate reconnection. The characteristic outflow speed is the whistler phase speed (based on the upstream magnetic field).
- The inflow velocity $v_{\text{in}} = \Omega_e \delta_e^2/\Delta_i$ where $\Delta_i \sim k^{-1} \ll L$ (system size). This rate is independent of m_e.

The GEM perspective is not universally accepted. An alternate point of view provides evidence that:

- Reconnection is not a universal constant, and depends on system parameters (such as ion/electron skin depth, plasma beta, boundary conditions).
- Reconnection rate is not independent of the system size, and in fact, often decreases as the system size increases.

He provided three examples:

1. Forced reconnection without guide field
2. Undriven reconnection with guide field
3. Undriven reconnection with open boundaries

9.8.10 3D Nature of Reconnection

How about the freedom coming from the extra dimension?

Distinct 3D features, including

- flux ropes
- kink instability
- turbulence

3D diffusion region can be fundamentally different. For example, you may find bifurcation of electron diffusion region in 3D reconnection simulations. Clue: bifurcated layer is located in between these intertwined flux ropes, and tearing modes give rise to these flux ropes! 2D only allows the parallel tearing mode, i.e., no bifurcation. 3D allows a spectrum of oblique tearing modes, unlike 2D.

9.9 Lower Hybrid Drift Instability

The *lower-hybrid-drift instability* (LHDI) is a microinstability which has thought to be an anomalous transport mechanism in both laboratory and space plasmas. The local linear theory of this instability is well understood. The mode is driven by the diamagnetic current produced by pressure inhomogeneities and is characterized at maximum growth by

$$\begin{aligned} \omega \sim kV_d &\lesssim \omega_{lh} \\ \gamma &\lesssim \omega \\ k &\sim \omega_{lh}/v_i \\ \mathbf{k}\cdot\mathbf{B} &= 0 \end{aligned}$$

where V_d is the diamagnetic drift velocity, ω_{lh} is the lower hybrid frequency, and v_i is the ion thermal velocity.

$$\omega_{lh} = [(\Omega_i\Omega_e)^{-1} + \omega_{pi}^{-2}]^{-1/2}$$

where Ω_i is the ion cyclotron frequency, Ω_e is the electron cyclotron frequency, and ω_{pi} is the ion plasma frequency. It is so called "hybrid" because it is a mixture of two frequencies. It is unusual in that the ion and electron masses play an equally important role.

The drift velocity is

$$V_d = \frac{\nabla p \times \mathbf{B}}{neB^2}$$

Physically, the instability is reactive (fluid-like) in the strong drift velocity regime ($V_d > v_i$) and dissipative (kinetic) in the weak drift velocity regime ($V_d < v_i$).

9.10 Kinetic Mode

9.11 Pressure Anisotropy Instabilities

When we shift to the MHD with anisotropic pressure tensor

$$P_{ij} = p_\perp\delta_{ij} + (p_\parallel - p_\perp)B_iB_j/B^2$$

where $p_\perp$ and $p_\parallel$ are the pressures perpendicular and parallel w.r.t. the magnetic field, respectively. For the strong magnetic field approximation, the two pressures are related to the plasma density and the magnetic field strength by two adiabatic equations,

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{p_\parallel B^2}{\rho^3}\right) &= 0 \\
\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{p_\perp}{\rho B}\right) &= 0
\end{aligned}
$$

This is also known as the *double adiabatic theory*, which is also what many people remember to be the key conclusion from the CGL theory ((Chew, Goldberger, and Low 1956), Equation 5.32 and Equation 5.34). Here I want to emphasize the meaning of *adiabatic* again: this assumes zero heat flux. If the system is not adiabatic, the conservation of these two quantities related to the parallel and perpendicular pressure is no longer valid, and additional terms may come into play such as the stochastic heating.

Now imagine an increasing magnetic field B. For an initially Maxwellian distributed plasma like a circle in the phase space, it will get stretched in the perpendicular direction and become an oval; conversely, if the magnetic field is decreasing, it will get stretched in the parallel direction.

9.11.1 Firehose Instability

See more in Section 10.1.

9.11.2 Mirror Instability

Southwood & Kivelson had a nice paper explaining the physics of mirror instability. See also Section 10.2 for a more thorough mathematical description of the instability.

The mirror instability is prevalent in planetary and cometary magnetosheaths and other high β environment. It is recognized as one of the two magnetohydrodynamic instabilities that occur in the presence of extreme velocity space (pitch angle) anisotropy in a uniform plasma, the other instability being the firehose. The mirror mode has gained increasing interest following its identification in spacecraft data from the magnetosheath and solar wind. Although the instability was originally derived from magnetohydrodynamic fluid theory, later work showed that there were significant differences between the fluid theory and a more rigorous kinetic approach, as we will see in Section 10.2.

The instability occurs when β, the ratio of plasma to magnetic pressure. The anisotropy $(p_\perp - p_\parallel)/p_\parallel$ required for instability $\propto \beta^{-1}$. In planetary magnetosheaths the source of anisotropy is likely to be the planetary bow shock and in cometary environments the ion pickup process is a natural source of anisotropy. Both types of environment tend to have relatively large values of β.

Any nonlinear saturation mechanism of the mirror instability is likely to leave the plasma spatially structured, as is also strongly suggested by the many observations. In practice, in almost any experimental detection of a plasma instability, the wave fields and the plasma population will have evolved to some quasi-steady condition that represents a nonlinear saturated state of the instability. In the spatially structured magnetic field associated with the minor instability in both its linear and nonlinear phase, different parts of the ion distribution will resonate with an ion cyclotron wave as the wave propagates along the inhomogeneous field. This effect is likely to inhibit the growth of the ion cyclotron mode, which theoretically also exists in the same frequency band.

The mirror instability is referred to as a “fluid” instability, alluding to the fact that the phase space (pitch angle) anisotropy of the bulk of the hot plasma distribution serves as the source of energy. The instability grows because of a subtle coupling between a group of particles with small velocity parallel to the field and the rest of the population. The mirror instability has zero parallel phase velocity in the plasma frame of reference. It follows, by analogy with other uses of the term that we can call particles with near zero parallel velocity *resonant*.

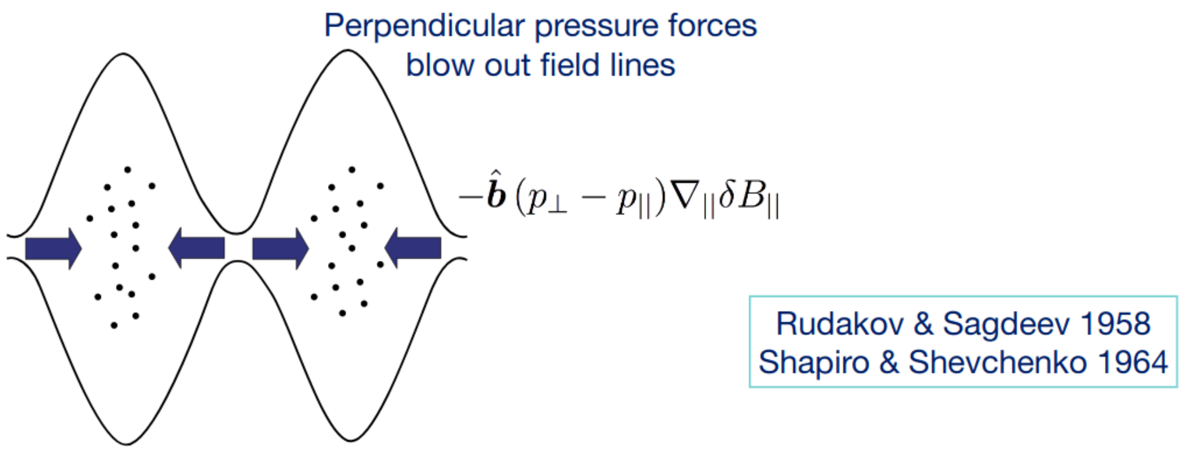

Figure 9.8: Illustration of mirror instability.

(Song+ 1994) presented a set of identification criteria for distinguishing among various forms of high β MHD waves, with mirror modes having the properties:

- Compressional fluctuations:
$$(\delta\mathbf{B}^2 - \delta\mathbf{B}_{\|}^2)/\delta\mathbf{B}_{\|}^2 < 1$$
- Anti-correlated δPi and δP_B:
$$\delta P_i/\delta P_B < 0$$
- Linear magnetic field polarization

- Stationary in plasma rest frame (i.e. zero phase speed):

$$\left(\delta\mathbf{v}^2/\mathbf{v}_0^2\right)/\left(\delta\mathbf{B}^2/\mathbf{B}_0^2\right) \approx 0$$

The last criterion is referred to as the Doppler ratio and is used to distinguish between slow mode waves (value ≥ 1) and mirror modes (value near zero).

Note that mirror mode is not really a MHD mode, but from observation in the Earth's magnetosheath we often see mirror modes on large MHD scales. However, in the solar wind, mirror mode structures seem to be much smaller than those observed in the magnetosheath. From Equation 10.17, electron physics may be involved on the electron scales.

It is still an unknown mystery from MMS observation that downstream of an interplanetary shock, very few mirror modes were observed. This is surprising, because usually the shock is a source of pressure anisotropy. Another explained fact is that typically mirror modes are not observed during fast solar wind (as of 2022).

9.12 Cyclotron Instability

I still don't see the dispersion relations for these types of instabilities...

9.12.1 Ion Cyclotron Instability

There is another instability (besides mirror instability) that occurs at frequencies below the ion gyrofrequency in the presence of ion pitch angle anisotropy, the *electromagnetic ion cyclotron instability* (EMIC). This is a resonant instability in which the energy for the instability is fed from a subset of the particle population that are in gyroresonance with the unstable wave.

From Earth's magnetosheath observations, mirror modes dominates even though the linear dispersion relation predicts a smaller growth rate compared with EMIC waves. Gary argued in the 1990s that this is because of a small portion of heavier species (e.g. helium) that modifies the growth rate. Local hybrid simulations (e.g. the example from hybrid VPIC in ISSS14) supports this. However, researchers have not reached a consensus on whether ion cyclotron instability suppress mirror instability or vice versa.

10 Kinetic MHD

This is taken from the hand-written lecture notes from Prof. Alexander Schekochihin. Theoretical physicists love CGS units, but I tend to use SI units here. In later part of the note, there may be mixed units, so be careful.

We start by showing how for magnetized, weakly collisional plasmas ($\nu_{\text{colli}} \ll \Omega_s$, $r_L \ll \lambda_{\text{mfp}}$, where λ_{mfp} is the mean free path), low-frequency ($\omega \ll \Omega_s$), long-wavelength ($kr_L \ll 1$) dynamics can be decided by a set of equations that look almost like the familiar MHD. We will see later on that the ways in which they are not MHD will profoundly affact the dynamics — indeed we do not fully understand the full implication of this in high-β plasmas. This is consequently one of the frontier topics in theoretical plasma astrophysics.

Let us start from first principles. Any plasma that is going to be of interest to us is described by the Vlasov-Maxwell-Landau system of equations:

$$\frac{\partial f_s}{\partial t} + \mathbf{v} \cdot \nabla f_s + \frac{q_s}{m_s}\Big[\mathbf{E} + \mathbf{v} \times \mathbf{B}\Big]\frac{\partial f_s}{\partial \mathbf{v}} = C(f_s) \tag{10.1}$$

The Maxwell's equation can be simplified based on our assumptions.

$$\nabla \cdot \mathbf{E} = \epsilon_0 \sum_s q_s n_s,\, n_s = \int \mathrm{d}\mathbf{v} f_s$$

$\nabla \cdot \mathbf{E}$ is small when $k^2 \lambda_{De} \ll 1$, and this simply gives the quasineutrality condition.

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

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

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j} + \cancel{\epsilon_0 \mu_0 \frac{\partial \mathbf{E}}{\partial t}},\, \mathbf{j} = \sum_s n_s q_s \mathbf{u}_s, \mathbf{u}_s = \frac{1}{n_s}\int \mathrm{d}\mathbf{v}\mathbf{v} f_s$$

The displacement current can be neglected since $\omega \ll kc$ for low frequency waves and non-relativistic motions.

Intuitively, we tend to think of the plasma as a fluid (or a multi-fluid of several species) with some density n_s, velocity $\mathbf{u}_s$ and perhaps pressure, temperature, etc. This is rooted in our experience with collisional gases ($\nu \gg \omega$), which are in local Maxwellian equilibrium:

$$f_s = \frac{n_s}{(\pi v_{\text{th},s}^2)^{3/2}} e^{-\frac{(\mathbf{v}-\mathbf{v}_s)^2}{v_{\text{th},s}^2}},\ v_{\text{th},s} = \sqrt{\frac{2k_B T_s}{m_s}}$$

where $n_s, \mathbf{u}_s$ and T_s are governed by fluid equations.

With this desire to think of plasmas as fluid, let us break the motion of the particles into two parts:

$$\mathbf{v} = \mathbf{u}_s(t, \mathbf{r}) + \mathbf{w}$$

where $\mathbf{u}_s$ represents the mean velocity of species s (fluid-like description) and $\mathbf{w}$ represents the "peculiar" velocity or internal motion (kinetic description). This amounts to a transformation of variables

$$(t, \mathbf{r}, \mathbf{v}) \to (t, \mathbf{r}, \mathbf{w}), \quad \mathbf{w} = \mathbf{v} - \mathbf{u}_s(t, \mathbf{r})$$

under which the derivatives in the new basis shall be written as

$$\begin{aligned}
\frac{\partial}{\partial t} &\to \frac{\partial}{\partial t} - \frac{\partial \mathbf{u}_s}{\partial t} \cdot \frac{\partial}{\partial \mathbf{w}} \\
\nabla &\to \nabla - (\nabla \mathbf{u}_s) \cdot \frac{\partial}{\partial \mathbf{w}} \\
\frac{\partial}{\partial \mathbf{v}} &\to \frac{\partial}{\partial \mathbf{w}}
\end{aligned} \tag{10.2}$$

This can be derived from the chain rule with

$$\begin{aligned}
t' &= t \\
\mathbf{r}' &= \mathbf{r} \\
\mathbf{w} &= \mathbf{v} - \mathbf{u}_s(t, \mathbf{r})
\end{aligned}$$

Note that the three variables are independent, for any quantity f, we have

$$\begin{aligned}
\frac{\partial f}{\partial t} &= \frac{\partial f}{\partial t'} \cdot \frac{\partial t'}{\partial t} + \frac{\partial f}{\partial \mathbf{r}'} \cdot \frac{\partial \mathbf{r}'}{\partial t} + \frac{\partial f}{\partial \mathbf{w}} \cdot \frac{\partial \mathbf{w}}{\partial t} \\
&= \frac{\partial f}{\partial t} - \frac{\partial \mathbf{u}_s}{\partial t} \cdot \frac{\mathbf{f}}{\partial \mathbf{w}}
\end{aligned}$$

Similarly we can derive the other two relations in Equation 10.2. Then the Boltzmann equation becomes

$$\Big(\frac{\partial}{\partial t} + \mathbf{u}_s \cdot \nabla\Big) f_s + (\mathbf{w} \cdot \nabla) f_s + \Big(\frac{q_s}{m_s}\mathbf{w} \times \mathbf{B} + \mathbf{a}_s - \mathbf{w} \cdot \nabla \mathbf{u}_s\Big) \cdot \frac{\partial f_s}{\partial \mathbf{w}} = C(f_s) \tag{10.3}$$

where

$$\mathbf{a}_s = \frac{q_s}{m_s}\Big(\mathbf{E} + \mathbf{u}_s \times \mathbf{B}\Big) - \frac{\mathrm{d}\mathbf{u}_s}{\mathrm{d}t}$$

and now we always have $\int \mathrm{d}\mathbf{w}\mathbf{w} f_s = 0$ by definition. The strategy now is to take moments of Equation 10.3. The zeroth-order moment ($\int \mathrm{d}\mathbf{w}$) gives

$$\begin{aligned}
\int \frac{df_s}{\mathrm{d}t}\mathrm{d}\mathbf{w} + \int (\mathbf{w} \cdot \nabla) f_s \mathrm{d}\mathbf{w} + \int \Big[...\Big] \cdot \frac{\partial f_s}{\partial \mathbf{w}}\mathrm{d}\mathbf{w} &= 0 \\
\frac{\mathrm{d}}{\mathrm{d}t}\int f_s \mathrm{d}\mathbf{w} + \cancel{\nabla \cdot \int \mathbf{w} f_s \mathrm{d}\mathbf{w}} - \int (\mathbf{w} \cdot \nabla \mathbf{u}_s) \cdot \frac{\partial f_s}{\partial \mathbf{w}}\mathrm{d}\mathbf{w} &= 0 \\
\frac{dn_s}{\mathrm{d}t} + \int f_s \frac{\partial}{\partial \mathbf{w}}(\mathbf{w} \cdot \nabla \mathbf{u}_s)\mathrm{d}\mathbf{w} &= 0 \\
\frac{dn_s}{\mathrm{d}t} + \int f_s \frac{\partial}{\partial w_j}(w_i \frac{\partial}{\partial x_i} u_{sj}) dw_j &= 0 \\
\frac{dn_s}{\mathrm{d}t} + \int f_s \delta_{ij} \frac{\partial}{\partial x_i} u_{sj} dw_j + \int f_s w_i \cancel{\frac{\partial^2 u_{sj}}{\partial x_i \partial w_j}} dw_j &= 0 \\
\frac{dn_s}{\mathrm{d}t} + \int f_s \frac{\partial}{\partial x_i} u_{si} dw_i &= 0 \\
\frac{dn_s}{\mathrm{d}t} + \nabla \cdot \mathbf{u}_s \int f_s \mathrm{d}\mathbf{w} &= 0 \\
\frac{dn_s}{\mathrm{d}t} + (\nabla \cdot \mathbf{u}_s) n_s &= 0
\end{aligned}$$

or

$$\frac{\partial n_s}{\partial t} + \nabla \cdot (n_s \mathbf{u}_s) = 0 \tag{10.4}$$

The first-order moment ($\int \mathrm{d}\mathbf{w} m_s \mathbf{w}$) gives

$$\nabla \cdot \int \mathrm{d}\mathbf{w} m_s \mathbf{w}\mathbf{w} f_s - m_s n_s \mathbf{a}_s = \int \mathrm{d}\mathbf{w} m_s \mathbf{w} C(f_s) \equiv \mathbf{R}_s$$

where $\int \mathrm{d}\mathbf{w} m_s \mathbf{w}\mathbf{w} f_s = \mathbf{P}_s$ is the pressure tensor and $\mathbf{R}_s$ is the collisional friction. Unpacking $\mathbf{a}_s$, we have the momentum equation for each species s

$$m_s n_s \frac{\mathrm{d}\mathbf{u}_s}{\mathrm{d}t} = -\nabla \cdot \mathbf{P}_s + q_s n_s (\mathbf{E} + \mathbf{u}_s \times \mathbf{B}) + \mathbf{R}_s \tag{10.5}$$

Summing over all the species,

$$\begin{aligned}
\sum_s m_s n_s \frac{\mathrm{d}\mathbf{u}_s}{\mathrm{d}t} = -\nabla \cdot \sum_s \mathbf{P}_s + \cancel{\sum_s q_s n_s \mathbf{E}} + \sum_s q_s n_s \mathbf{u}_s \times \mathbf{B} + \cancel{\sum_s \mathbf{R}_s} \\
\rho \frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t} = -\nabla \cdot \mathbf{P} + \mathbf{j} \times \mathbf{B} \\
\rho \frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t} = -\nabla \cdot \mathbf{P} + \mu_0^{-1} (\nabla \times \mathbf{B}) \times \mathbf{B} \\
\rho \frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t} = -\nabla \cdot \left[\mathbf{P} + \frac{B^2}{2\mu_0}\mathbf{I} - \mathbf{B}\mathbf{B} \right]
\end{aligned}$$

It is useful to emphasize that $\mathrm{d}/\mathrm{d}t = \partial/\partial t + \mathbf{u} \cdot \nabla$. Later we will see the notation of D/Dt, which is used to remind us of the fact that $\mathbf{w}$ is involved.

We also need an equation for the magnetic field. It is Faraday's law:

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

From Equation 10.5,

$$\mathbf{E} = -\mathbf{u}_s \times \mathbf{B} + \frac{\nabla \cdot \mathbf{P}_s}{q_s n_s} - \frac{\mathbf{R}_s}{q_s n_s} + \frac{m_s}{q_s} \frac{\mathrm{d}\mathbf{u}_s}{\mathrm{d}t}$$

Based on the following arguments:

- $\nabla \cdot \mathbf{P}_s / (q_s n_s)$ is small since $k r_s / M_A \ll 1$(long wave + incompressible plasma???),
- $\mathbf{R}_s / (q_s n_s)$ is small since $\nu_s / \Omega_s \ll 1$,
- $(m_s / q_s) \mathrm{d}\mathbf{u}_s / \mathrm{d}t$ is small since $\omega / \Omega_s \ll 1$

we have the simplest Ohm's law and in turn $\mathbf{u}_s = \mathbf{E} \times \mathbf{B} / B^2 = \mathbf{u}_\perp$, the perpendicular component of the velocity is the same for all species. Then we get the induction equation from Faraday's law:

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{u} \times \mathbf{B}) \tag{10.6}$$

or

$$\frac{\mathrm{d}\mathbf{B}}{\mathrm{d}t} = \mathbf{B}\cdot\nabla\mathbf{u} - \mathbf{B}\nabla\cdot\mathbf{u} \tag{10.7}$$

The three equations we have so far are very similar to MHD, except for the pressure tensor. Obviously, all the kinetic magic is hidden in $\mathbf{P}$.

Going back to Equation 10.3, it is key to notice that

$$\frac{q_s}{m_s}\mathbf{w}\times\mathbf{B}\cdot\frac{\partial f_s}{\partial \mathbf{w}} = -\Omega_s\Big(\frac{\partial f_s}{\partial}\Big)_{w_\perp,w_\parallel}$$

where θ is the gyroangle in the perpendicular plane. This can be proved by changing to cylindrical coordinates

$$\mathbf{w} = (w_\perp\cos\theta, w_\perp\sin\theta, w_\parallel)$$

with changing of variables:

$$\begin{aligned}
\frac{\partial f_s}{\partial\theta} &= \frac{\partial f_s}{\partial w_{\perp 1}}\frac{\partial w_{\perp 1}}{\partial\theta} + \frac{\partial f_s}{\partial w_{\perp 1}}\frac{\partial w_{\perp 1}}{\partial\theta} + \frac{\partial f_s}{\partial w_\parallel}\cancel{\frac{\partial w_\parallel}{\partial\theta}} \\
&= \frac{\partial f_s}{\partial w_{\perp 1}}\frac{\partial w_\perp\cos\theta}{\partial\theta} + \frac{\partial f_s}{\partial w_{\perp 2}}\frac{\partial w_\perp\sin\theta}{\partial\theta} \\
&= -\frac{\partial f_s}{\partial w_{\perp 1}}w_\perp\sin\theta + \frac{\partial f_s}{\partial w_{\perp 2}}w_\perp\cos\theta \\
&= -w_{\perp 2}\frac{\partial f_s}{\partial w_{\perp 1}} + w_{\perp 1}\frac{\partial f_s}{\partial w_{\perp 2}} \\
&= \mathbf{w}\times\hat{b}\cdot\frac{\partial f_s}{\partial\mathbf{w}}
\end{aligned} \tag{10.8}$$

This is why we say the third term in Equation 10.1 represents a rotation in the velocity space, or more exactly, in the perpendicular velocity plane.

From Equation 10.3, if we apply the lowest order of approximation,

$$\Omega_s\Big(\frac{\partial f_s}{\partial}\Big)_{w_\perp,w_\parallel} = \underbrace{\frac{\mathrm{d}f_s}{\mathrm{d}t}}_{\substack{\omega/\Omega_s\ll 1\\ kr_s u_s/v_{th,s}\ll 1}} + \underbrace{\mathbf{w}\cdot\nabla f_s}_{kr_s\ll 1} + (\underbrace{\mathbf{a}_s}_{kr_s\ll 1} - \underbrace{\mathbf{w}\cdot\nabla\mathbf{u}_s}_{kr_s M_A\ll 1})\cdot\frac{\partial f_s}{\partial\mathbf{w}} - \underbrace{C(f_s)}_{\nu_s\ll\Omega_s} = 0 \tag{10.9}$$

which essentially tells us that $f_s = f_s(w_\perp, w_\parallel, \theta) = f_s(w_\perp, w_\parallel)$ is gyrotropic. Let us use <> to denote averaging over a gyroperiod:

$$\langle A \rangle = \int_0^{2\pi} A \mathrm{d}\theta$$

We can use gyrotropy to simplify the pressure tensor:

$$\begin{aligned}
\mathbf{P}_s &= \int \mathrm{d}\mathbf{w} m_s \langle \mathbf{w}\mathbf{w} \rangle f_s(\mathbf{r}, w_\perp, w_\parallel, t) \\
&= \int \mathrm{d}\mathbf{w} m_s [\frac{w_\perp^2}{2}(\mathbf{I} - \hat{b}\hat{b}) + w_\parallel^2 \hat{b}\hat{b}] f_s(\mathbf{r}, w_\perp, w_\parallel, t) \\
&= (\mathbf{I} - \hat{b}\hat{b}) \int \mathrm{d}\mathbf{w} \frac{m_s w_\perp^2}{2} f_s + \hat{b}\hat{b} \int \mathrm{d}\mathbf{w} m_s w_\parallel^2 f_s \\
&= \begin{pmatrix} p_{\perp s} & 0 & 0 \\ 0 & p_{\perp s} & 0 \\ 0 & 0 & p_{\parallel s} \end{pmatrix}
\end{aligned}$$

where

$$\begin{aligned}
p_\perp &= \int \mathrm{d}\mathbf{w} \frac{m_s w_\perp^2}{2} f_s \\
p_\parallel &= \int \mathrm{d}\mathbf{w} m_s w_\parallel^2 f_s
\end{aligned}$$

Equation 10.5 becomes

$$\rho \frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t} = -\nabla\Big(\underbrace{p_\perp + \frac{B^2}{2\mu_0}}_{\text{total scalar pressure}} \Big) + \nabla \cdot \Big[\hat{b}\hat{b}(\underbrace{p_\perp - p_\parallel}_{\text{pressure anisotropy stress}} + \underbrace{\frac{B^2}{\mu_0}}_{\text{Maxwell stress}}) \Big] \tag{10.10}$$

The pressure anisotropy stress is the key new feature compared to usual MHD. It should be important provided $p_\perp - p_\parallel \gtrsim B^2/\mu_0$, or $(p_\perp - p_\parallel)/p \gtrsim 2/\beta$. Therefore this is more likely to matter in high-β plasmas.

To summarize what we have gotten so far: to work out motions and magnetic fields in a plasma, solve Equation 10.10 for $\mathbf{u}$ and Equation 10.6 for $\mathbf{B}$, where

$$\begin{aligned}
\rho &= \sum_s m_s \int \mathrm{d}\mathbf{w} f_s \\
p_\perp &= \sum_s \int \mathrm{d}\mathbf{w} \frac{m_s w_\perp^2}{2} f_s \\
p_\parallel &= \sum_s \int \mathrm{d}\mathbf{w} m_s w_\parallel^2 f_s
\end{aligned}$$

We still need the kinetic equation to calculate f_s — this kinetic equation will need to be somewhat reduced to solve for the lowest-order, gyrotropic $f_s(w_\perp, w_\parallel)$. In pursuit of instant justification, we can postpone doing this and first derive some results that do not need the f_s equation (i.e. the Firehose instability) as in Section 10.1. For mirror modes, let us continue from the kinetic Equation 10.9 for higher orders. We have already known that the lowest order approximation gives gyrotropic distributions.

To the first order,

$$\Omega_s\Big(\frac{\partial f_s^1}{\partial\theta}\Big)_{w_\perp,w_\parallel} = \frac{\mathrm{d}f_s^0}{\mathrm{d}t} + \mathbf{w}\cdot\nabla f_s^0 + (\mathbf{a}_s - \mathbf{w}\cdot\nabla\mathbf{u}_s)\cdot\frac{\partial f_s^0}{\partial\mathbf{w}} - C(f_s^0)$$

The left-hand side can be eliminated by integrating over θ, so we have

$$\Big\langle\frac{\mathrm{d}f_s}{\mathrm{d}t} + \mathbf{w}\cdot\nabla f_s + (\mathbf{a}_s - \mathbf{w}\cdot\nabla\mathbf{u}_s)\cdot\frac{\partial f_s}{\partial\mathbf{w}} - C(f_s)\Big\rangle = 0$$

where $f_s = f_s(w_\perp, w_\parallel)$. To do this averaging, we tranform variables from $(t,\mathbf{r},\mathbf{w}) \rightarrow (t,\mathbf{r},w_\perp,w_\parallel,\theta)$. With

$$\begin{aligned} w_\parallel &= \mathbf{w}\cdot\hat{b}(t,\mathbf{r}) \\ w_\perp &= |\mathbf{w} - w_\parallel\hat{b}| \end{aligned}$$

and some algebras (??? Check online notes.), we have

$$\frac{Df_s}{Dt} + \frac{1}{B}\frac{DB}{Dt}\frac{w_\perp}{2}\frac{\partial f_s}{\partial w_\perp} + \Big(\frac{q_s}{m_s}E_\parallel - \frac{D\mathbf{u}_s}{Dt}\cdot\hat{b} - \frac{w_\perp^2}{2}\frac{\nabla_\parallel B}{B}\Big)\frac{\partial f_s}{\partial w_\parallel} = C(f_s) \tag{10.11}$$

where

$$D/Dt = \mathrm{d}/\mathrm{d}t + w_\parallel\hat{b}\cdot\nabla = \partial/\partial t + \mathbf{u}_s\cdot\nabla + w_\parallel\hat{b}\cdot\nabla$$

This is not terribly transparent and it is perhaps better to write this equation in different, "more physical" variables. Let

$$f_s(w_\perp, w_\parallel) = F_s(\mu,\epsilon)$$

where $\mu = m_s w_\perp^2/2B$ is the magnetic moment of a gyrating particle and $\epsilon = m_s w^2/2 = m_s(w_\perp^2 + w_\parallel^2)/2$. Since μ is conserved when $\omega \ll \Omega_s$, F_s satisfies (???)

$$\frac{DF_s}{Dt} + \Big[m_s w_\parallel\Big(\frac{q_s}{m_s}E_\parallel - \frac{D\mathbf{u}_s}{Dt}\cdot\hat{b}\Big) + \mu\frac{\mathrm{d}B}{\mathrm{d}t}\Big]\frac{\partial F_s}{\partial\epsilon} = C(F_s) \tag{10.12}$$

- The first term is the convective derivative in the guiding center coordinates.
- The second term is the acceleration by parallel electric field, where $E_\parallel$ is determined by imposing $\sum_s q_s n_s = 0$.
- The third term takes account of the fact that ϵ does not include the bulk velocity.
- The fourth term is the *betatron acceleration* due to μ conservation:

$$
\begin{aligned}
\epsilon &= \mu B + \frac{m_s w_\parallel^2}{2} \\
\dot{\epsilon} &= \mu \dot{B}\, (w_\parallel \text{ constant???})
\end{aligned}
$$

Betatron acceleration refers to situations in which the magnetic field strength increases slowly in time (compared with a gyroperiod), so that μ remains constant, but the particle kinetic energy is not constant due to the presence of electric fields (associated with the time-varying magnetic field). Then, the perpendicular energy is increased due to constancy of μ. As we will see soon in Section 10.2, this is the key for explaining mirror modes.

10.1 Firehose Instability: Linear Theory

Suppose we have some "macroscopic" solution of our (yet to be fully derived) equilibrium. We allow low-frequency, short-wavelength perturbations ($\omega \ll u/l, kl \gg 1$) of this solution, and seek solutions in the form $\mathbf{X} + \delta\mathbf{X}$ with infinitesimal perturbations $\propto e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}$. Note that the velocity $\mathbf{u}$ is treated as a perturbation term (background velocity is simply a drift).

From Equation 10.6

$$
\begin{aligned}
-\omega\delta\mathbf{B} &= \mathbf{B}\cdot\mathbf{k}\delta\mathbf{u} - \mathbf{B}\mathbf{k}\cdot\delta\mathbf{u} \\
&= B(k_\parallel\delta\mathbf{u}_\perp - \hat{b}\mathbf{k}_\perp\cdot\delta\mathbf{u}_\perp)
\end{aligned}
\tag{10.13}
$$

Inserting Equation 3.8 into Equation 10.10, we have

$$
\begin{aligned}
-\omega\rho\delta\mathbf{u} &= -\mathbf{k}\Big(\delta p_\perp + \frac{B\delta B}{\mu_0}\Big) + \mathbf{k}\cdot\Big[\Big(\delta\hat{b}\hat{b} + \hat{b}\delta\hat{b}\Big)\Big(p_\perp - p_\parallel + \frac{B^2}{\mu_0}\Big) + \hat{b}\hat{b}\Big(\delta p_\perp - \delta p_\parallel + \frac{2B\delta B}{\mu_0}\Big)\Big] \\
&= -\mathbf{k}_\perp\Big(\delta p_\perp + \frac{B\delta B}{\mu_0}\Big) - \hat{b}k_\parallel\Big[\delta p_\parallel + (p_\perp - p_\parallel)\frac{\delta B}{B}\Big] + \delta\hat{b}k_\parallel\Big(p_\perp - p_\parallel + \frac{B^2}{\mu_0}\Big)
\end{aligned}
\tag{10.14}
$$

$\delta\hat{b}$ has two parts: the Alfvénic part $\delta\mathbf{B}_\perp/B$ and the compressional part $\delta\mathbf{B}_\parallel/B$. From Equation 10.13, the Alfvénic perturbation of $\hat{b}$ can be written as

$$\delta\hat{b} = \frac{\delta \mathbf{B}_\perp}{B} = -\frac{k_\parallel}{\omega}\delta \mathbf{u}_\perp$$

Isolate the Alfvénic response in Equation 10.14 by cross-producting with $\mathbf{k}_\perp$:

$$\begin{aligned}
-\omega\rho\mathbf{k}_\perp \times \delta\mathbf{u}_\perp &= k_\parallel\Big(p_\perp - p_\parallel + \frac{B^2}{\mu_0}\Big)\mathbf{k}_\perp \times \delta\hat{b} \\
\omega\rho\mathbf{k}_\perp \times \delta\mathbf{u}_\perp &= k_\parallel\Big(p_\perp - p_\parallel + \frac{B^2}{\mu_0}\Big)\mathbf{k}_\perp \times \frac{k_\parallel}{\omega}\delta\mathbf{u}_\perp \\
\omega^2 = k_\parallel^2\Big(\frac{B^2}{\mu_0\rho} + \frac{p_\perp - p_\parallel}{\rho}\Big) &= k_\parallel^2 v_{th\parallel}^2\Big(\frac{p_\perp - p_\parallel}{p_\parallel} + \frac{2}{\beta_\parallel}\Big)
\end{aligned}$$

Let $A = (p_\perp - p_\parallel)/p_\parallel$. The system will be unstable if $A < -2/\beta_\parallel$, which leads to a growth rate

$$\gamma = k_\parallel v_{th\parallel}\sqrt{\left|A + \frac{2}{\beta_\parallel}\right|}$$

Thus, negative A ($p_\parallel > p_\perp$) locally weakens tention, i.e. slows down Alfvén waves, and makes it energetically easier to bend the field lines. For $A < -2/\beta_\parallel$, the elasticity of field lines is lost and we have the firehose instability.

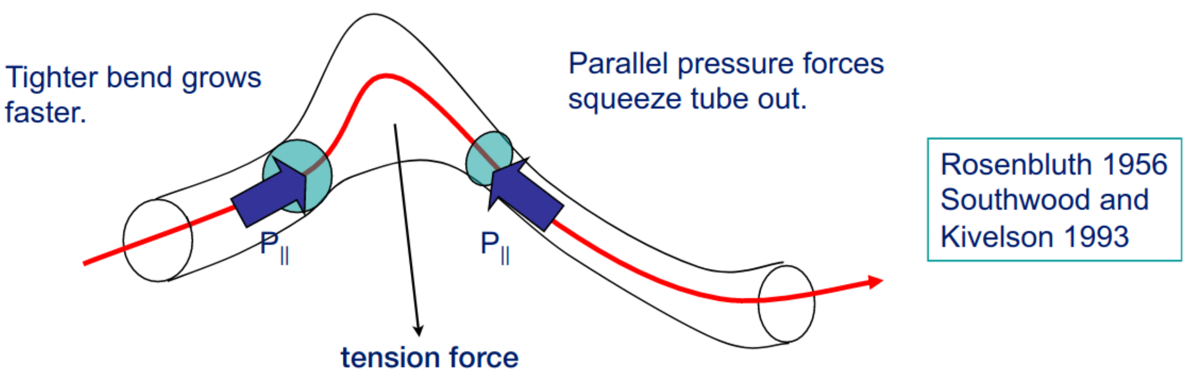

Figure 10.1: Along a flux tube, $p_\parallel$ is the destabilizing force, the curvature force is the stabilizing force.

Key points:

- Nothing surprising that $p_\parallel > p_\perp$ leads to an instability: it is a non-equilibrium situation, so a source of free energy.

- $\gamma \propto k_\parallel$ leads to UV catastrophe: within KMHD ($\omega \ll \Omega_i, kr_i \ll 1$), the wavenumber of peak *gamma* is not captured. Including finite larmor radius gives (Oxford MNRAS 405, 291? ARE THE EXPRESSIONS CORRECT?)

$$
\begin{aligned}
\gamma_{\text{peak}} &\sim \left| A + \frac{2}{\beta_\parallel} \right| \Omega_i \\
k_{\parallel\text{peak}} r_i &\sim \left| A + \frac{2}{\beta_\parallel} \right|^{1/2}
\end{aligned}
$$

so the instability is very fast ($\gamma \propto \Omega_i$ very large with strong B field) at microscale. Any high-β macroscopic solution with $p_\parallel > p_\perp$ will blow up instantly. What happens next is decided by the nonlinear saturation of the firehose. It was a transformative moment when Justin Kasper in 2002 discovered that the firehose stability boundary constrains most observed solar wind states, followed by Hellinger in 2006. Bale in 2009 showed that there is an increased fluctuation level at the boundary.

ADD FIGURE ABOUT THE FIREHOSE STABILITY REGIME FIGURE!

10.2 Mirror Instability: Linear Theory

Let us go back to Equation 10.13 and Equation 10.14 and get apart from Alfvénic what other perturbations there are and when they are stable. We have already looked at the Alfvénic perturbation $\delta\hat{b} = \delta\mathbf{B}_\perp / B$. Now consider

$$
\frac{\delta B}{B} = \frac{\delta B_\parallel}{B}
$$

From Equation 10.13, we have the perpendicular compression increases B:

$$
\omega \frac{\delta B}{B} = \mathbf{k}_\perp \cdot \delta\mathbf{u}_\perp
$$

Take $\mathbf{k}_\perp\cdot$ Equation 10.14:

$$
\omega\rho\mathbf{k}_\perp \cdot \delta\mathbf{u}_\perp = \rho\omega^2 \frac{\delta B}{B} = k_\perp^2 \Big(\delta p_\perp + \frac{B\delta B}{\mu_0} \Big) + k_\parallel^2 \Big(p_\perp - p_\parallel + \frac{B^2}{\mu_0} \Big) \frac{\delta B}{B} \tag{10.15}
$$

Note the $p_\perp$ term here: we need kinetic theory to calculate this! Fortunately we have Equation 10.12 ready for calculating

$$\delta p_\perp = \int \mathrm{d}\mathbf{w} \frac{m_s w_\perp^2}{2} \delta f_s(w_\perp, w_\|)$$

$\delta f_s(w_\perp, w_\|)$ can be obtained by calculating $F_s(\mu, \epsilon)$ and transforming back to $w_\perp, w_\|$.

Here is a cute subtlety: our macroscopic equilibrium, around which we are expanding the distribution is

$$F_{0s}(\mu, \epsilon) = f_{0s}(w_\perp, w_\|) = f_{0s}\Big(\sqrt{\frac{2B_0\mu}{m_s}}, \sqrt{\frac{2(\epsilon - \mu B_0)}{m_s}}\Big)$$

which contains B_0 the unperturbed magnetic field. μ in F_0 contains $B_0 + \delta B$, and this has to be taken into account when transforming to $w_\perp, w_\|$. Now when we perturb everything:

$$\begin{aligned}
F_s(\mu, \epsilon) &= F_{0s}(\mu, \epsilon) + \delta F_s \\
&= f_{0s}(w_\perp, w_\|) + \delta f_s \\
&= f_{0s}\Big(\sqrt{\frac{2\mu(B_0 + \delta B)}{m_s}}, \sqrt{\frac{2[\epsilon - \mu(B_0 + \delta B)]}{m_s}}\Big) + \delta f_s \\
&= f_{0s}\Big(\sqrt{\frac{2\mu B_0}{m_s}}\sqrt{1 + \frac{\delta B}{B_0}}, \sqrt{\frac{2(\epsilon - \mu B_0)}{m_s}}\sqrt{1 - \frac{m_s \mu \delta B}{(\epsilon - \mu B_0)}}\Big) + \delta f_s \\
&\approx f_{0s}\Big(\sqrt{\frac{2\mu B_0}{m_s}}, \sqrt{\frac{2(\epsilon - \mu B_0)}{m_s}}\Big) + \frac{2\mu}{m_s}\delta B\Big(\frac{\partial f_{0s}}{\partial w_\perp^2} - \frac{\partial f_{0s}}{\partial w_\|^2}\Big) + \delta f_s
\end{aligned}$$

Thus

$$\delta f_s = \delta F_s - w_\perp^2 \frac{\delta B}{B}\Big(\frac{\partial f_{0s}}{\partial w_\perp^2} - \frac{\partial f_{0s}}{\partial w_\|^2}\Big)$$

If f_{0s} is a bi-Maxwellian,

$$f_{0s} = \frac{n_s}{\pi^{3/2} v_{\mathrm{th}\perp s}^2 v_{\mathrm{th}\| s}} \exp\Big(- \frac{w_\perp^2}{v_{\mathrm{th}\perp s}^2} - \frac{w_\|^2}{v_{\mathrm{th}\| s}^2}\Big)$$

then this can be further written as

$$\delta f_s = \delta F_s + w_\perp^2 \frac{\delta B}{B}\Big(\frac{1}{v_{\mathrm{th}\perp s}^2} - \frac{1}{v_{\mathrm{th}\| s}^2}\Big) f_{0s} = \delta F_s + w_\perp^2 \frac{\delta B}{B}\frac{m_s n_s}{2}\Big(\frac{1}{p_{\perp s}} - \frac{1}{p_{\| s}}\Big) f_{0s}$$

We can eliminate the partial derivatives via integration by parts:

$$\begin{aligned}
\int \mathrm{d}\mathbf{w} \frac{\partial f_{0s}}{\partial w_\parallel^2} &= \int_0^{2\pi} \mathrm{d}\theta \int dw_\perp \int \frac{1}{2w_\parallel} \frac{\partial f_{0s}}{\partial w_\parallel} dw_\parallel \\
&= \int_0^{2\pi} \mathrm{d}\theta \int dw_\perp \int \frac{1}{2w_\parallel} \mathrm{d} f_{0s} \\
&= \int_0^{2\pi} \mathrm{d}\theta \int dw_\perp \Big[\cancel{\frac{f_{0s}}{2w_\parallel}\Big|_{-\infty}^{\infty}} - \int f_{0s} \mathrm{d} \frac{1}{2w_\parallel} \Big] \\
&= \int_0^{2\pi} \mathrm{d}\theta \int dw_\perp \int \frac{1}{2w_\parallel^2} f_{0s} dw_\parallel \\
&= \int \mathrm{d}\mathbf{w} \frac{1}{2w_\parallel^2} f_{0s} \\
\int \mathrm{d}\mathbf{w} w_\perp w_\perp^4 \frac{\partial f_{0s}}{\partial w_\perp^2} &= 2\pi \int dw_\parallel \Big[\cancel{\frac{1}{2} w_\perp^4 f_{0s}\Big|_{-\infty}^{+\infty}} - \frac{1}{2} \int f_{0s} dw_\perp^4 \Big] \\
&= -2\pi \int dw_\parallel 2w_\perp^2 f_{0s} w_\perp dw_\perp \\
&= -2 \int \mathrm{d}\mathbf{w} w_\perp^2 f_{0s}
\end{aligned}$$

This then gives us

$$\begin{aligned}
\delta p_{\perp s} &= \int \mathrm{d}\mathbf{w} \frac{m_s w_\perp^2}{2} \delta f_s \\
&= \int \mathrm{d}\mathbf{w} \frac{m_s w_\perp^2}{2} \delta F_s - \int \mathrm{d}\mathbf{w} \frac{m_s w_\perp^4}{2} \Big(\frac{\partial f_{0s}}{\partial w_\perp^2} - \frac{\partial f_{0s}}{\partial w_\parallel^2} \Big) \frac{\delta B}{B} \\
&= \int \mathrm{d}\mathbf{w} \frac{m_s w_\perp^2}{2} \delta F_s - \int_0^{2\pi} \mathrm{d}\theta \int dw_\perp w_\perp \int \mathrm{d}w_\parallel \frac{m_s w_\perp^4}{2} \Big(\frac{\partial f_{0s}}{\partial w_\perp^2} - \frac{\partial f_{0s}}{\partial w_\parallel^2} \Big) \frac{\delta B}{B} \\
&= \int \mathrm{d}\mathbf{w} \frac{m_s w_\perp^2}{2} \delta F_s + 2 \int \mathrm{d}\mathbf{w} \frac{m_s w_\perp^2}{2} \delta f_{0s} \frac{\delta B}{B} + \int \mathrm{d}\mathbf{w} \frac{2(\frac{1}{2} m_s w_\perp^2)^2}{m_s w_\parallel^2} f_{0s} \frac{\delta B}{B} \\
&= \int \mathrm{d}\mathbf{w} \frac{m_s w_\perp^2}{2} \delta F_s + \frac{\delta B}{B} \Big(2p_{\perp s} - \frac{2p_{\perp s}^2}{p_{\parallel s}} \alpha_s)
\end{aligned} \tag{10.16}$$

where α_s is some coefficients of order 1 if f_{0s} is not bi-Maxwellian.

δF_s can be obtained by ignoring collisions and linearizing and Fourier-transforming Equation 10.12 ($\mathbf{u}_s = 0$):

$$
\begin{aligned}
-i(\omega-k_\parallel w_\parallel)\delta F_s &= -\Big[m_s w_\parallel\Big(\frac{q_s}{m_s}E_\parallel - i(\omega-k_\parallel w_\parallel)\delta u_{\parallel s}\Big) - i\omega\mu\delta B\Big]\frac{\partial F_{0s}}{\partial\epsilon} \\
\delta F_s &= -i\frac{w_\parallel q_s E_\parallel}{\omega-k_\parallel w_\parallel}\frac{\partial F_{0s}}{\partial\epsilon} - \delta u_{\parallel s} m_s w_\parallel\frac{\partial F_{0s}}{\partial\epsilon} - \frac{\omega}{\omega-k_\parallel w_\parallel}\mu\delta B\frac{\partial F_{0s}}{\partial\epsilon}
\end{aligned}
$$

The first term can be ignored if $\beta \gg 1$ (??? See the complete calculation in another note!); otherwise $E_\parallel$ can be got by imposing $\sum_s q_s n_s = 0$. The second term can be shown to be equivalent to $\delta u_{\parallel s}\partial f_{0s}/\partial w_\parallel$:

$$
\begin{aligned}
\frac{\partial f_{0s}}{\partial w_\parallel} &= \frac{\partial f_{0s}}{\partial\epsilon}\frac{\partial\epsilon}{\partial w_\parallel} + \frac{\partial f_{0s}}{\partial\mu}\cancel{\frac{\partial\mu}{\partial w_\parallel}} \\
&= \frac{\partial F_{0s}}{\partial\epsilon} m_s w_\parallel
\end{aligned}
$$

so this will not contribute to $\delta p_\perp$ because it integrates to 0.

The third term can be written as

$$
\frac{\omega}{\omega-k_\parallel w_\parallel}\mu\delta B\frac{\partial F_{0s}}{\partial\epsilon} = \frac{\omega}{\omega-k_\parallel w_\parallel}\frac{m_s w_\perp^2}{2}\frac{\delta B}{B}\frac{1}{w_\parallel}\frac{\partial f_{0s}}{\partial w_\parallel} = \frac{\omega}{\omega-k_\parallel w_\parallel}m_s w_\perp^2\frac{\delta B}{B}\frac{\partial f_{0s}}{\partial w_\parallel^2}
$$

Thus, the "relevant" part of δF_s is

$$
\delta F_s = -\frac{\omega}{\omega-k_\parallel w_\parallel}m_s w_\perp^2\frac{\delta B}{B}\frac{\partial f_{0s}}{\partial w_\parallel^2}
$$

and its contribution to $\delta p_{\perp s}$ is

$$
\int \mathrm{d}\mathbf{w}\frac{m_s w_\perp^2}{2}\delta F_s = \frac{\delta B}{B}\frac{\omega}{|k_\parallel|}\int\frac{dw_\parallel}{w_\parallel-\frac{\omega}{|k_\parallel|}}\Big[\frac{\partial}{\partial w_\parallel^2}\int \mathrm{d}\mathbf{w}_\perp\frac{m_s^2 w_\perp^4}{2}f_{0s}\Big]
$$

Here we have $|k_\parallel|$ because if $k_\parallel < 0$, we can change the variable $w_\parallel \to -w_\parallel$. This involves the Landau integral, which can be evaluated with the residual theorem Equation 3.3 when integrate in the complex plane mostly along the real axis and the large semicircle in the upper half plane except for a small semicircle just below the pole (ADD FIGURE!):

$$
\frac{1}{w_\parallel-\frac{\omega}{|k_\parallel|}} = P\frac{1}{w_\parallel-\frac{\omega}{|k_\parallel|}} + i\pi\delta\Big(w_\parallel-\frac{\omega}{|k_\parallel|}\Big)
$$

so

$$\int \mathrm{d}\mathbf{w} \frac{m_s w_\perp^2}{2} \delta F_s = \frac{\delta B}{B} \Big[\cancel{\frac{\omega}{|k_\parallel|} P \int \frac{dw_\parallel}{w_\parallel - \frac{\omega}{|k_\parallel|}}} [...] + i\pi \frac{\omega}{|k_\parallel|} [...]_{w_\parallel} = \omega/|k_\parallel| \Big]$$

The first term is small when we assume $\omega \ll k_\parallel v_{\mathrm{th}s\parallel}$; the second term must be kept because it is the lowest-order imaginary part which will lead to instability.

For a bi-Maxwellian,

$$\Big[\frac{\partial}{\partial w_\parallel^2} \int \mathrm{d}\mathbf{w}_\perp \frac{m_s^2 w_\perp^4}{2} f_{0s} \Big]_{w_\parallel = \omega/|k_\parallel|} = -\frac{2p_{\perp s}^2}{p_{\parallel s}} \frac{e^{-\frac{\omega^2}{k_\parallel^2 v_{\mathrm{th}\parallel s}^2}}}{\sqrt{\pi} v_{\mathrm{th}\parallel s}}$$

The exponential term is nearly 1. If it is not a bi-Maxwellian, then we need to multiply by a coefficient $\alpha_s \sim 1$.

Equation 10.16 becomes

$$\delta p_{\perp s} = \frac{\delta B}{B} \Big[2p_{\perp s} - \frac{2p_{\perp s}^2}{p_{\parallel s}} \Big(\alpha_s + i\sqrt{\pi} \frac{\omega}{|k_\parallel| v_{\mathrm{th}\parallel s}} \sigma_s \Big) \Big]$$

This goes into Equation 10.15:

$$\rho\omega^2 = k_\perp^2 \frac{B^2}{\mu_0} \Big[\sum_s (1 - \frac{p_{\perp s}}{p_{\parallel s}} \alpha_s) \beta_{\perp s} - i \sum_s \sigma_s \frac{p_{\perp s}}{p_{\parallel s}} \beta_{\perp s} \sqrt{\pi} \frac{\omega}{|k_\parallel| v_{\mathrm{th}\parallel s}} + 1 \Big] + k_\parallel^2 \frac{B^2}{\mu_0} \Big[\sum_s \frac{\beta_{\perp s}}{2} (1 - \frac{p_{\parallel s}}{p_{\perp s}}) + 1 \Big]$$

The left-hand side can be neglected because $\omega \ll k_\parallel v_{\mathrm{th}\parallel s}$. The electron thermal velocity $v_{\mathrm{th}\parallel e}$ in the denominator can be neglected because $v_{\mathrm{th}\parallel e} \gg v_{\mathrm{th}\parallel i}$. The growth rate γ is the imaginary part of ω. Reorganize the last equation:

$$\sigma_i \frac{p_{\perp i}}{p_{\parallel i}} \beta_{\perp i} \sqrt{\pi} \frac{\gamma}{|k_\parallel| v_{\mathrm{th}\parallel i}} = \sum_s \Big(\frac{p_\perp s}{p_\parallel s} \alpha_s - 1 \Big) \beta_{\perp s} - 1 - \frac{k_\parallel^2}{k_\perp^2} \Big[\sum_s \frac{\beta_{\perp s}}{2} (1 - \frac{p_{\parallel s}}{p_{\perp s}}) + 1 \Big] \tag{10.17}$$

where $\Lambda \equiv \frac{k_\parallel^2}{k_\perp^2} \sum_s \Big(\frac{p_\perp s}{p_\parallel s} \alpha_s - 1 \Big) \beta_{\perp s} - 1$ triggers instability if this is positive:

$$\sum_s \Big(\frac{p_\perp s}{p_\parallel s} \alpha_s - 1 \Big) \beta_{\perp s} > 1$$

Examining where this comes from, we see that this amounts to $\delta p_\perp$ modifying the magnetic pressure force and turning it from positive to negative:

$$\delta p_\perp + \frac{B\delta B}{\mu_0} = \frac{B\delta B}{\mu_0}\Big[\underbrace{1}_{\text{B pressure}} - \underbrace{\sum_s\Big(\frac{p_{\perp s}}{p_{\parallel s}}\alpha_s - 1\Big)\beta_{\perp s}}_{\substack{\text{non-resonant}\\ \text{particle pressure}}} + \underbrace{\cdots}_{\substack{\text{resonant particle}\\ \text{pressure}}}\Big]$$

Thus, fundamentally, pressure anisotropy makes it easier to compress or rarefy magnetic field — and things become unstable when the sign of the pressure flips and it becomes energetically profitable to create compressions and rarefications. (ADD FIGURE!) The dispersion relation Equation 10.17 is basically a statement of pressure balance between the magnetic pressure, the non-resonant particle pressure $\delta p_\perp$ and the resonant particle pressure $\propto \gamma$, which came from the *betatron acceleration* $\mu \mathrm{d}B/\mathrm{d}t$ in Equation 10.12.

The betatron acceleration term refers to what happens in the stable case. When magnetic pressure opposes formation of δB perturbations (say, troughs), to compensate it, we must have $\gamma < 0$ and energy goes from δB to resonant particles, which are accelerated by the mirror force. The corresponding decaying of δB is the well-known *Barnes damping* (landau damping of "mirror field", Barnes 1966, also known as transit-time damping from Stix's book. See more discussion on the physics in Southwood & Kivelson 1993.)

To finish the job, note that, from Equation 10.17 (ADD FIGURE!) for a given $k_\perp$

$$\frac{\partial\gamma}{\partial k_\parallel}\Big|_{k_\perp} \propto \Lambda - \frac{k_\parallel^2}{k_\perp^2}\Big[\sum_s \frac{\beta_{\perp s}}{2}\Big(1 - \frac{p_\parallel s}{p_\perp s}\Big) + 1\Big]$$

The maximum growth rate is reached when the right-hand side goes to 0, which is equivalent to $\frac{2}{3}\Lambda$, so the maximum growth rate

$$\gamma_{\max} = \frac{|k_\parallel| v_{\mathrm{th}\parallel i}}{\sqrt{\pi}}\frac{2}{3}\Lambda\frac{p_{\parallel i}}{p_{\perp i}}\frac{1}{\sigma_i\beta_{\perp i}}$$

We have assumed $\gamma \ll k_\parallel v_{\mathrm{th}\parallel s}$, which is indeed true if

$$\Lambda\frac{1}{\beta_{\perp i}} = (\sum_s A_s\beta_{\perp s} - 1)\frac{1}{\beta_{\perp i}} \ll 1$$

so our approximations are consistent.

If we are close to marginal instablity,

$$\frac{k_\parallel}{k_\perp} \sim \sqrt{\Lambda} \ll 1$$

so mirror modes are highly oblique near the threshold.

Another important point is that again we encounter the UV catastrophe since $\gamma \propto k_\parallel$. The mirror mode is a fast, microscale instability whose peak growth rate is outside KMHD regime. Including finite larmor radius gives (Hellinger 2007 PoP 14, 082105?)

$$\gamma_{\rm peak} \sim \Big(A - \frac{1}{\beta}\Big)^2 \beta\Omega_i, \quad k_{\rm peak} r_i \sim \Big(A - \frac{1}{\beta}\Big)\beta$$

Thus, any high-β macroscopic solution of KMHD with $p_\perp > p_\parallel$ will blow up, just like the case for $p_\parallel > p_\perp$, and again what happens next depends on how mirror instability saturates. Note that A_e is ignored since $A_e \ll A_i$ (?). The mirror instability condition is

$$\begin{aligned} \frac{p_{\perp i}}{p_{\parallel i}} - 1 > \frac{1}{\beta_{\perp i}} &= \frac{1}{\beta_{\parallel i}} \frac{p_{\parallel i}}{p_{\perp i}} \\ \frac{p_{\perp i}}{p_{\parallel i}}\Big(\frac{p_{\perp i}}{p_{\parallel i}} - 1\Big) &> \frac{1}{\beta_{\parallel i}} \end{aligned}$$

Figure 10.2 shows observation from Wind spacecraft. The solar wind indeed seems to stay within these boundaries. (ADD REFS!)

10.3 Origin of Pressure Anisotropy

So far we have seen that the bottom line is that any macroscopic, high-β KMHD solution that has $p_\perp \neq p_\parallel$ (more precisely, $|p_\perp - p_\parallel|/p \gtrsim 1/\beta$) will be voilently unstable to either firehose or mirror — both of which are fast and micro-scale modes giving rise to fluctuations outside the KMHD regime (and, by the way, also outside gyrokinetics — too close to cyclotron frequency, $k_\parallel/k_\perp$ not small enough, $\delta\mathbf{B}/B$ also not small enough). How worried should this make us about the applicability of KMHD to high-β plasmas that are not collisional enough to be fully fluid (i.e. $\nu \ll \Omega_s$)?

The answer is, *very worried*! A key property of low-frequency, weakly collisional dynamics is that the magnetic moment $\mu = m_s w_\perp^2/2B$ is conserved by particles. The mean μ of particles of species s is

$$< \mu >_w = \frac{1}{n_s} \int \mathrm{d}\mathbf{w} \mu f_s = \frac{p_{\perp s}}{n_s B} = \text{const.}$$

For the purpose of a qualitative discussion, let us pretend for a moment that $n_s = $ const (incompressible plasmas, $\beta \gg 1$). Then the above conservation relation says that, locally in a fluid element ($\mathbf{w}$ is peculari velocity), every time you change $\mathbf{B}$, you must change $p_{\perp s}$ proportionally (but not $p_{\parallel s}$). Thus we expect (?)

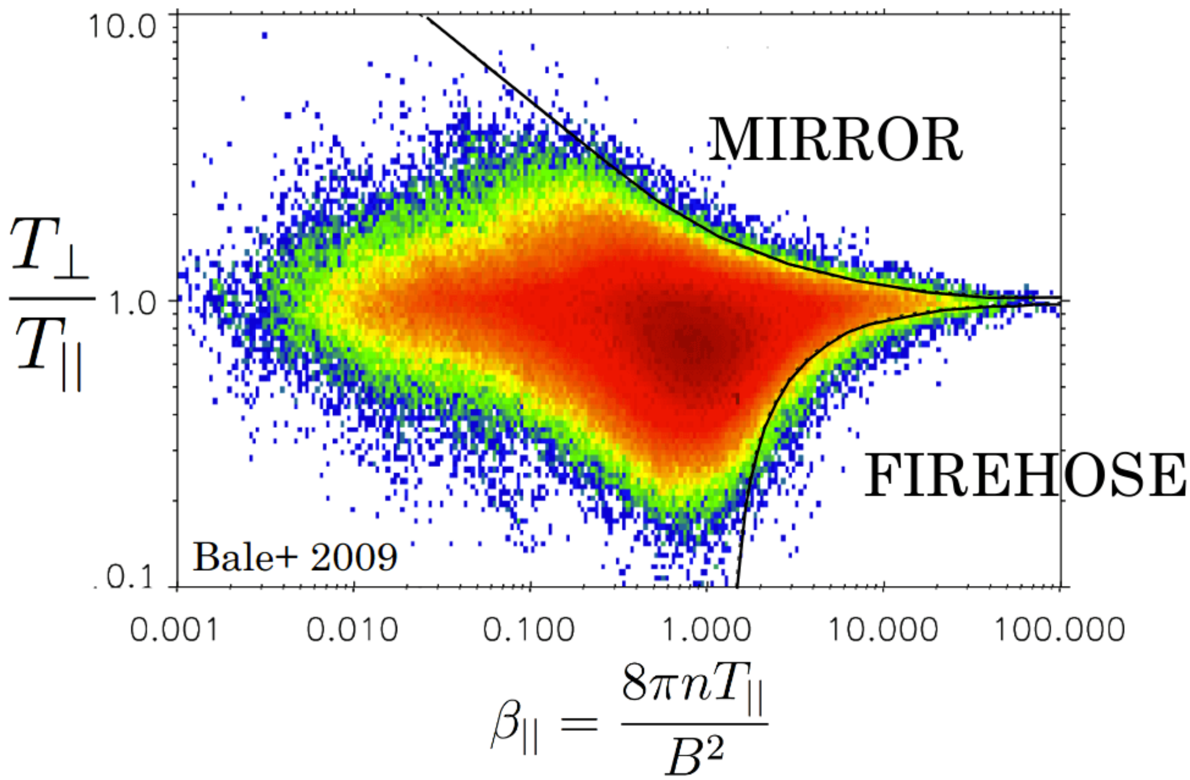

Figure 10.2: Collective solar wind observation data ($\sim 1e6$) from Wind spacecraft. The lines represent the instability thresholds for mirror and firehose instability, respectively.

$$\frac{1}{p_{\perp s}}\frac{\mathrm{d}p_{\perp s}}{\mathrm{d}t} \sim \underbrace{\frac{1}{B}\frac{\mathrm{d}B}{\mathrm{d}t}}_{\mu \text{ conservation}} - \underbrace{\nu_s \frac{p_{\perp s} - p_{\parallel s}}{p_{\perp s}}}_{\substack{\text{relaxation of pressure} \\ \text{anisotropy by collisions}}} \tag{10.18}$$

It is useful to remind ourselves that $\mathrm{d}/\mathrm{d}t$ is in the $\mathbf{u}_s$ frame. Balancing the two effects on the right-hand side,

$$\Delta_s = \frac{p_{\perp s} - p_{\parallel s}}{p_{\perp s}} \sim \frac{1}{\nu_s}\frac{1}{B}\frac{\mathrm{d}B}{\mathrm{d}t} \tag{10.19}$$

This expression is valid only if $\Delta_s \ll 1$, i.e. $\nu_s \gg \frac{1}{B}\frac{\mathrm{d}B}{\mathrm{d}t}$, otherwise Δ_s will grow with time as B is changed. Thus

- B increases locally $\to \Delta_s > 0 \to$ mirror
- B decreases locally $\to \Delta_s < 0 \to$ firehose

As nealy any large-scale dynamics involves local changes in $\mathbf{B}$, this means that nearly any macroscopic solution of KMHD in the high-β regime will be unstable. A very good example is the dynamo problem: when magnetic field is randomly stretched by turbulence, leading (in MHD) to exponential growth of magnetic energy (and, eventually, to saturated fields we observe), locally one find structures of this sort: (ADD FIGURE!)

Generally speaking, in order to understand long-time evolution, we need some sort of mean-field theory for the large-scale effect of the microscale instabilities on the dynamics. Presumably, this is to keep preserve anisotropy, marginal out (?) the instabilities (as indeed appears to be confirmed by the solar wind measurements — see Bale + 2009).

There are two ways in which this can happen — firehose and mirror fluctuations might scatter particles, leading to higher effective collisionality and ? control the pressure anisotropy

$$-\frac{2}{\beta} \lesssim \frac{p_\perp - p_\parallel}{p} \lesssim \frac{1}{\beta}$$

They might inhibit changes of B, which is another way of keeping Δ under control. Which of these matters for dynamics, see the speculative overview of the possible consequences of either mechanisim in MNRAS 440, 3226 (2014).

- nonlinear firehose: Rosin+ MNRAS 413 2011
- nonlinear mirror: Rincon+ MNRAS 447, 2015
- PIC simulations: Kunz+ PRL, 2014

10.4 Remarks

10.4.1 Remark I

If we assume incompressibility, the magnetic induction Equation 10.7 becomes

$$\frac{\mathrm{d}\mathbf{B}}{\mathrm{d}t} = \mathbf{B} \cdot \nabla \mathbf{u}$$
$$\frac{1}{B}\frac{\mathrm{d}B}{\mathrm{d}t} = \hat{b}\hat{b} : \nabla \mathbf{u}$$

Then, from Equation 10.19,

$$p_{\perp s} - p_{\parallel s} \sim \frac{p_s}{\nu_c}\hat{b}\hat{b} : \nabla \mathbf{u}$$

where p_s/ν_c is the parallel dynamical viscosity. Putting this back into Equation 10.10, we get the lowest order ??? MHD equation. So, from the large-scale point of view, pressure anisotropy is viscous stress — but the resulting equations are ill-posed (blow up via instabilities with $\gamma \propto k_\parallel$).

10.4.2 Remark II

More rigorously, Equation 10.18 can be obtained via "CGL equations", i.e. the evolution equations of $p_{\perp s}$ and $p_{\parallel s}$. Namely $\int \mathrm{d}\mathbf{w}\frac{m_s w_\perp^2}{2}$ Equation 10.11:

$$p_{\perp s}\frac{\mathrm{d}}{\mathrm{d}t}\ln\frac{p_{\perp s}}{n_s B} = -\nabla \cdot (q_{\perp s}\hat{b}) - q_{\perp s}\nabla \cdot \hat{b} - \nu_s(p_{\perp s} - p_{\parallel s}) \tag{10.20}$$

$\int \mathrm{d}\mathbf{w} m_s w_\parallel^2$ Equation 10.11:

$$p_{\parallel s}\frac{\mathrm{d}}{\mathrm{d}t}\ln\frac{p_{\parallel s}B^2}{n_s^3} = -\nabla \cdot (q_{\parallel s}\hat{b}) + 2q_{\perp s}\nabla \cdot \hat{b} - 2\nu_s(p_{\parallel s} - p_{\perp s}) \tag{10.21}$$

The left-hand side is the conservation of $J = \oint dl w_\parallel$, i.e. "bounce invariant". The new feature here is heat fluxes:

$$q_{\perp s} = \int \mathrm{d}\mathbf{w}\frac{m_s w_\perp^2}{2} w_\parallel f_s$$
$$q_{\parallel s} = \int \mathrm{d}\mathbf{w} m_s w_\parallel^3 f_s$$

They are here because particles can flow in and out of a fluid element and thus affect the conservation (or otherwise) of $< \mu >_w$ and $< J^2 >_w$ within it.

Finally, from Equation 10.20 and Equation 10.21,

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}(p_{\perp s} - p_{\parallel s}) = (p_{\perp s} + 2p_{\parallel s})\frac{1}{B}\frac{\mathrm{d}B}{\mathrm{d}t} + (p_{\perp s} - 3p_{\parallel s})\frac{1}{n_s}\frac{dn_s}{\mathrm{d}t} \\
- \nabla \cdot [(q_{\perp s} - q_{\parallel s})\hat{b}] - 3q_{\perp s}\nabla \cdot \hat{b} - 3\nu_s(p_{\perp s} - p_{\parallel s})
\end{aligned}
$$

???

$$
\begin{aligned}
\Delta_s &= \frac{p_{\perp s} - p_{\parallel s}}{p_s} \\
&\approx \frac{1}{\nu_s}\left\{\frac{1}{B}\frac{\mathrm{d}B}{\mathrm{d}t} - \frac{2}{3}\frac{1}{n_s}\frac{dn_s}{\mathrm{d}t} - \frac{\nabla \cdot [((q_{\perp s} - q_{\parallel s})\hat{b})] + 3q_{\perp s}\nabla \cdot \hat{b}}{3p_s}\right\}
\end{aligned}
$$

Typically, what prominent here are electron heat fluxes. So heat fluxes can also lead to anisotropies and so macroscopic solutions of KMHD involving temperature gradients will also go unstable at microscales!

So here we are, we cannot change B at large scales, we cannot compress/rarefy the plasma and we cannot have temperature gradients without having to deal everything exploding and needing new equations. Enjoy!

11 Gyrokinetics

This is taken from the slide given by Frank Jenko from Max-Planck-Institut für Plasmaphysik.

Gyrokinetics is a theory for describing plasmas at a much finer level. If kinetic effects (e.g. finite Larmor radius, Landau damping, magnetic trapping etc.) play a role, MHD is not applicable, and one has to use a kinetic description.

Once again we start from the Vlasov-Maxwell equation. Removing the fast gyromotion under the assumption $\omega \ll \Omega$ leads to a dramatic speed-up. Thus we can think the basic element of charged rings as quasiparticles: it is described by gyrocenter coordinates and can keep the kinetic effects.

11.1 A Brief Historical Review

The word "gyrokinetics" appeared in the literature in the late 1960s, first proposed by Rutherford, Frieman, Taylor and Hastie. The goal is to provide an adequate formalism for the linear kinetic drift waves study in general magnetic configurations, including finite Larmor radius effects. The first nonlinear set of equations for the perturbed distribution function δf was given by Friemann and Liu Chen in 1982, who also introduced the gyrokinetic ordering. Gyrokinetic theory is based on Hamiltonian methods, which means that from a Lagrangian description, we remove the gyro-angle dependency by the change of coordinate systems to describe the equation of motion.

11.2 Coordinate Transformation

We perform the following coordinate transformation from the charged particle's phase space $(\mathbf{x}, \mathbf{v})$ to the corresponding guiding-center phase space $(\mathbf{X}, \mathbf{V})$, where:

$$\mathbf{X} = \mathbf{X}_\perp + X_\parallel \mathbf{b}_0, \quad \mathbf{X}_\perp = \mathbf{x}_\perp + \mathbf{r}_L, \quad \mathbf{r}_L = \mathbf{v} \times \mathbf{b}_0 / \Omega_c$$

$$\mathbf{V} = [\epsilon = v^2/2, \mu = v_\perp^2/2B_0, \sigma = \mathrm{sgn}(v_\parallel)]$$

Here, $\mathbf{b}_0 = \mathbf{B}_0/B_0$, $\mathbf{r}_L$ is the gyroradius vector, $v_\parallel = \mathbf{v}\cdot\mathbf{b}_0$, μ is the magnetic moment adiabatic invariant and, assuming there is no equilibrium electrostatic potential, ϵ is an equilibrium constant of motion.

In the guiding-center phase space, charged particle dynamics is naturally separated into the fast cyclotron motion and the slow guiding-center motion. One can then apply the gyrokinetic orderings and systematically average out the fast cyclotron motion (i.e., the gyrophase averaging) and obtain the asymptotically dominant (in terms of the smallness parameter ϵ) perturbed distribution function response. This perturbed distribution function in the guiding-center phase space can then be *inversely transformed* back to the charged particle phase space and applied toward the field equations (i.e., Maxwell's equations) for a self-consistent kinetic description.

- Obtain Vlasov equation in the guiding center coordinates.
- Obtain Maxwell's equations in the guiding center coordinates.

Why do we need extra steps before using Maxwell's equations? It is because in Maxwell's equations the particle information $(\rho, \mathbf{j})$ are described not in the guiding center coordinates. The distribution we obtain from Vlasov equation must be transformed back to ordinary coordinates and then we can do the moment integral.

In the guiding center coordinates, density and current density can be expressed as

$$n = \bar{N}_0 + \bar{\nabla}\cdot\left(\frac{c\bar{N}_0}{B\Omega}\bar{\nabla}_\perp\delta\phi\right) + \frac{\bar{N}_0\delta B_\parallel}{B} + \int \bar{F}_1 d\mathbf{p} \tag{11.1}$$

$$\begin{aligned}\mathbf{j} = &-\frac{\bar{N}_0 q^2}{cm}\delta A_\parallel \hat{b} + \int \frac{q\bar{\mathbf{P}}_\parallel}{m}\hat{b}\mathrm{d}\bar{\mathbf{P}} + \frac{cq\bar{N}_0}{B}\hat{b}\times\bar{\nabla}\delta\phi + \frac{3c^2\bar{N}_0\bar{T}}{2B^2\Omega}\hat{b}\times\bar{\nabla}\bar{\nabla}_\perp^2\delta\phi \\ &+ \frac{2c\bar{N}_0\bar{T}}{B^2}\hat{b}\times\bar{\nabla}\delta B_\parallel + \frac{\hat{b}}{B}\times\bar{\nabla}(c\bar{N}_0\bar{T}) + \hat{b}\times\bar{\nabla}\int c\bar{\mu}\bar{F}_1\mathrm{d}\bar{\mathbf{P}}\end{aligned} \tag{11.2}$$

Note that

- In Vlasov equation in the guiding center coordinates, the $\mathbf{E}\times\mathbf{B}$ drift, gradient drift and curvature drift all appears but the polarization drift is missing.
- In Equation 11.1, only the polarization term correction appears. From Hamilton's mechanics, all term that has a explicit time dependency will not contribute here since it will break the energy conservation of the system.

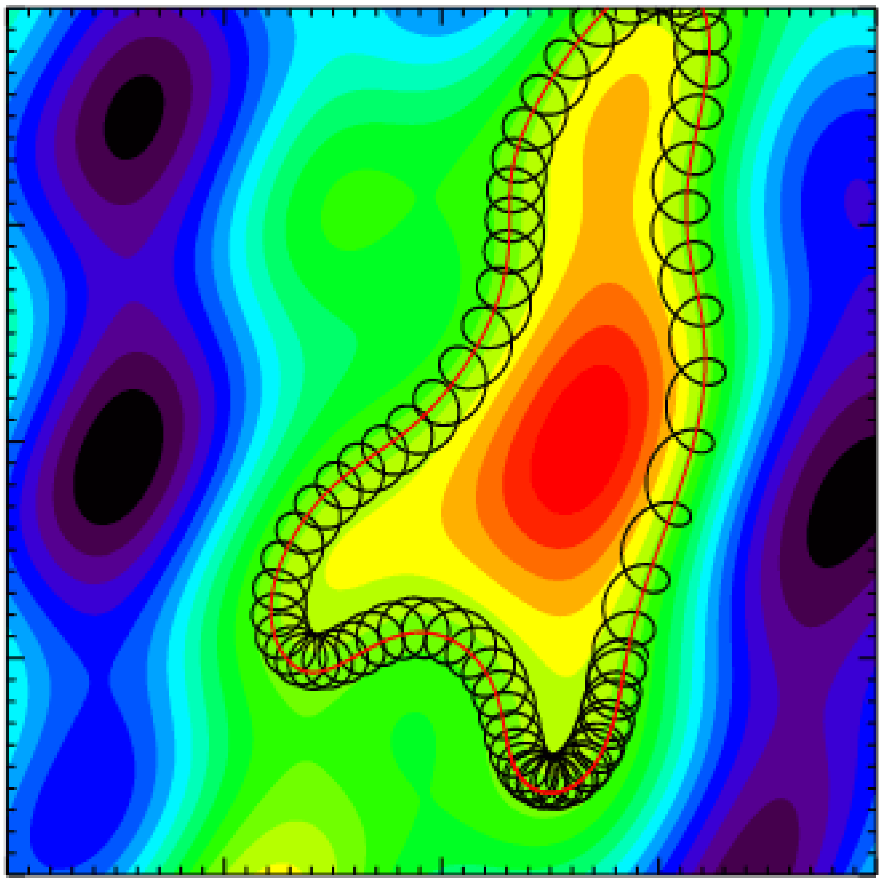

Figure 11.1: Charged particle orbit in a magnetic field pointing into the plane with electrostatic potential fluctuations.

11.3 The Gyrokinetic Ordering

11.3.1 From Kinetics to Gyrokinetics

Figure 11.1 shows the basic idea of gyrokinetic approximation. There is a strong magnetic field pointing into the plane. The electrostatic potential fluctuations are shown by the colored contours. The particle orbit is composed of two parts: a fast gyromotion and a slow $\mathbf{E} \times \mathbf{B}$ drift. In gyrokinetics we simply remove the fast gyromotion and introduce charged rings as quasiparticles, i.e. go from particle to gyrocenter coordinates.

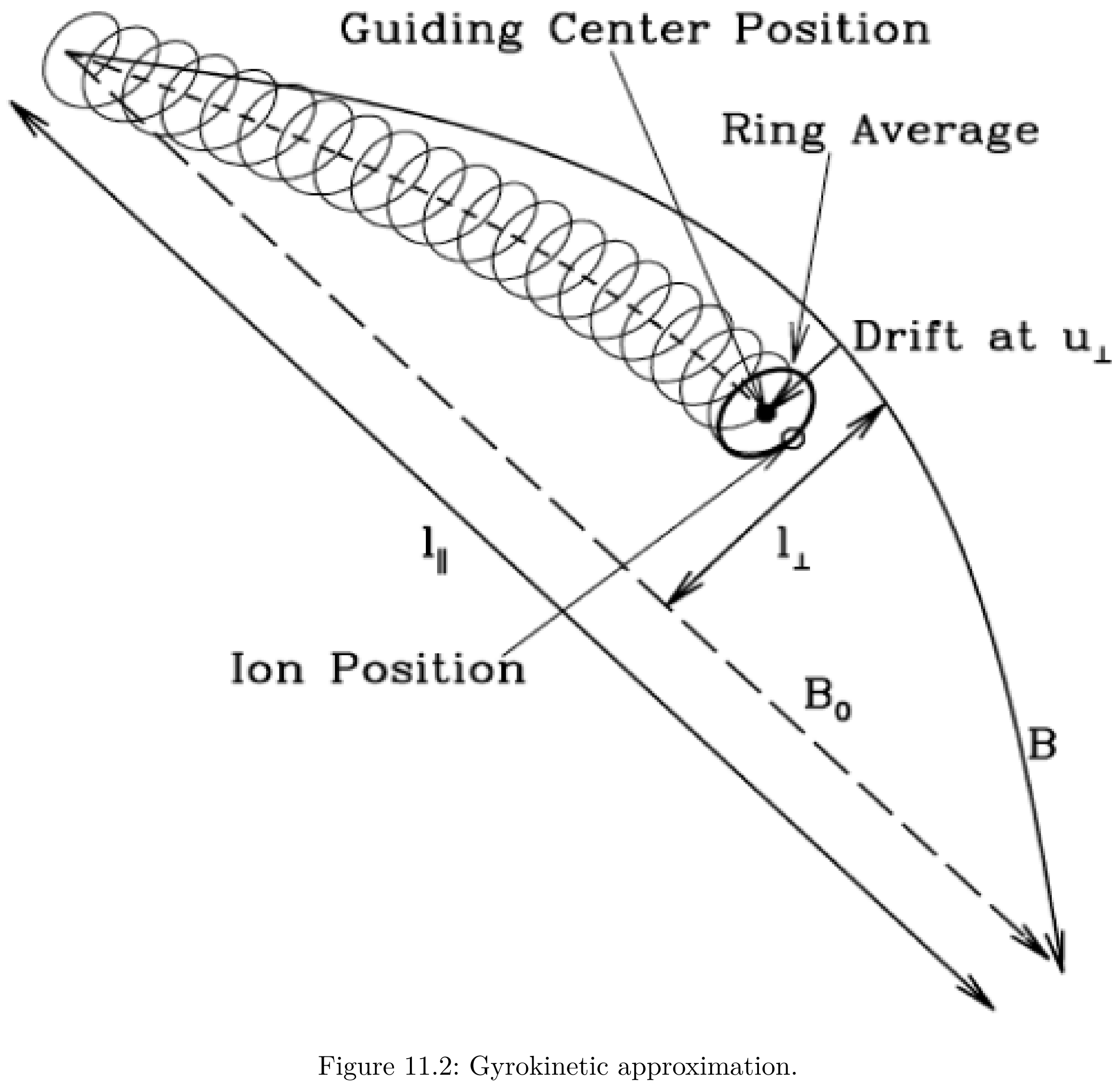

Figure 11.2: Gyrokinetic approximation.

- Slow time variation as compared to the gyromotion time scale:

$$\omega/\Omega_i \ll 1$$

- Spatial equilibirum scale much larger than the Larmor radius:

$$r_L/L \ll 1$$

- Strong anisotropy, i.e. only perpendicular components of the fluctuating quantities can be large:

$$k_\parallel/k_\perp \ll 1$$

- Small amplitude perturbations, i.e. energy of perturbation much smaller than the thermal energy:

$$e\phi/k_B T_e \ll 1$$

There exists a natural smallness parameter, $\epsilon = r_L/L$, which we are going to use in the ordering. In magnetically confined plasmas, typically we have $\epsilon \lesssim \mathcal{O}(10^{-2}) \ll 1$.

$$\frac{\omega}{\Omega} \sim \frac{r_L}{L} \sim \frac{k_\parallel}{k_\perp} \sim \frac{\delta F}{F_0} \sim \frac{\delta T}{T_0} \sim \frac{\delta n}{n_0} \sim \frac{|\delta \mathbf{B}|}{|\mathbf{B}_0|} \sim \frac{q\delta\phi}{T} \sim \epsilon$$

Usually low-frequency ($|\omega/\Omega_i| \sim \epsilon$) but short-wavelength ($k_\perp r_L \sim 1$) fluctuations are of interest in gyrokinetics. To include Landau resonance (???),

$$k_\parallel v_\parallel \sim \omega, \quad \text{or} \quad |k_\perp r_i| \sim 1$$

Noting, furthermore, for $|k_\perp r_i| \sim 1$ and the ratio of plasma ion pressure to the background magnetic field energy density $\beta_i \lesssim 1$:

$$\beta = \frac{P_{0i}}{P_{0B}} = \frac{m_i n_i v_{\perp i}^2/2}{B_0^2/2\mu_0} = \frac{v_{\perp i}^2}{v_A^2}$$

$$\left|\frac{\omega}{k_\perp v_A}\right| \sim \left|\frac{\omega}{\Omega_i}\right| \left|\frac{1}{k_\perp r_{iL}}\right| \beta_i^{1/2} \lesssim \epsilon$$

i.e. fast waves are systematically suppressed in the gyrokinetic orderings.

11.4 A Lagriangian Approach

If the Lagriangian of a dynamical system is known, e.g. for charged particle motion in non-canonical coordinates $(\mathbf{x}, \mathbf{v})$:

$$
\begin{aligned}
L &= \left(\frac{e}{c}\mathbf{A}(\mathbf{x}, t) + m\mathbf{v}\right) \cdot \dot{\mathbf{x}} - H(\mathbf{x}, \mathbf{v}) \\
H &= \frac{m}{2}v^2 + e\phi(\mathbf{x}, t)
\end{aligned}
$$

with $\mathbf{B} = \nabla \times \mathbf{A}$ and $\mathbf{E} = -\nabla\phi - \partial_t\mathbf{A}/c$. The equation of motion are given by the Lagrange equations (I almost forget everything from theoretical mechanics…):

$$
\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = 0, \quad i = 1, ..., 6
$$

For charged particles,

$$
\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{\mathbf{v}}} - \frac{\partial L}{\partial \mathbf{v}} = 0?
$$

$$
\begin{aligned}
\dot{\mathbf{x}} &= \mathbf{v} \\
\dot{\mathbf{v}} &= \frac{e}{m}(\mathbf{E} + \mathbf{v} \times \mathbf{B})
\end{aligned}
$$

In gyrokinetics we add low-frequency, anisotropic, small-amplitude fluctuations:

$$
\frac{\omega}{\Omega_i} \sim \frac{k_\parallel}{k_\perp} \sim \frac{e\phi}{T_e} \sim \epsilon
$$

We need a transition from particle coordinates $(\mathbf{x}, \mathbf{v})$ to guiding center coordinates $(\mathbf{R}, v_\parallel, \mu, \varphi)$. The easy way is to construct a new Lagrangian using Lie transforms (???)

$$
\Gamma = \left(mv_\parallel\hat{b}_0 + \frac{e}{c}\bar{\mathbf{A}}_{1\parallel}\hat{b}_0 + \frac{e}{c}\mathbf{A}_0\right) \cdot \mathrm{d}\mathbf{X} + \frac{mc}{e}\mu\mathrm{d}\theta - \left(\frac{m}{2}v_\parallel^2 + \mu B_0 + \mu\bar{B}_{1\parallel} + e\bar{\phi}_1\right)\mathrm{d}t
$$

where μ is the magnetic moment, θ is the gyroangle, and the overbar denotes a gyroaveraging operation. This gives the Euler-Lagrange equations

$$
\begin{aligned}
\dot{\mathbf{X}} &= v_\parallel \hat{b} + \frac{B}{B_\parallel^*}\Big[\frac{v_\perp}{B}\bar{\mathbf{B}}_{1\perp} + \frac{c}{B^2}\bar{\mathbf{E}}_1 \times \mathbf{B} + \frac{\mu}{m\Omega}\hat{b} \times \nabla(B + \bar{B}_{1\parallel}) + \frac{v_\parallel^2}{\Omega}(\nabla \times \hat{b})_\perp\Big] \\
\dot{v}_\parallel &= \frac{\dot{\mathbf{X}}}{mv_\parallel} \cdot (e\bar{\mathbf{E}}_1 - \mu\nabla(B + \bar{B}_{1\parallel})) \\
\dot{\mu} &= 0
\end{aligned}
\tag{11.3}
$$

Equation 11.3 contains all the drifts we have seen in Chapter 4.

Applying the gyrokinetic approximation, the effective gyroaveraged potential over one gyroperiod can be written using Fourier transform (???):

$$
\begin{aligned}
\phi^{\text{eff}}(\mathbf{x}, r_L) &= \frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}\theta \phi(\mathbf{x} + \mathbf{r}_L) \\
&= \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty} \mathrm{d}\mathbf{k} e^{i\mathbf{k}\cdot\mathbf{x}}\phi(\mathbf{k})J_0(kr_L)
\end{aligned}
$$

where J_0 is the zeroth order Bessel function.

11.4.1 Linear Gyrokinetics

We shall limit our considerations to that of a simple uniform plasma with an isotropic Maxwellian equilibrium distribution function. Assuming, furthermore, β (ratio between the plasma and magnetic pressures) $\ll 1$, such that there is negligible magnetic compression, the particle velocity distribution is then given by:

$$
f(\mathbf{x}, \mathbf{v}, t) = f_M(\epsilon) + \delta f(\mathbf{x}, \mathbf{v}, t)
$$

where $f_M(\epsilon) = n_0/(\pi^{3/2}v_t^3)\exp(-\epsilon/v_t^2)$ is the Maxwellian distribution function, v_t is the thermal speed (HOW TO UNDERSTAND THIS???):

$$
\delta f = \frac{q}{T} f_M(\epsilon)\delta\phi + e^{-\mathbf{r}_L\cdot\nabla}\delta g
$$

$T = mv_t^2/2$, δg satisfies the following linear gyrokinetic equation:

$$
\begin{aligned}
\Big(\frac{\partial}{\partial t} + v_\parallel \mathbf{b}_0 \cdot \nabla\Big)\delta g &= \frac{q}{T} f_M(\epsilon)\frac{\partial}{\partial t}\left\langle \delta L_g \right\rangle_\alpha \\
\delta L_g &= e^{-\mathbf{r}_L\cdot\nabla}\delta L \\
\delta L &= \delta\phi - v_\parallel \delta A_\parallel / c
\end{aligned}
\tag{11.4}
$$

and $< ... >_\alpha$ denotes averaging over the gyrophase angle, α. Here, the field variables are the scalar and vector potentials, $\delta\phi$ and $\delta\mathbf{A}$, with $\delta A_\parallel = \delta\mathbf{A}\cdot\mathbf{b}_0$ and the $\nabla\cdot\delta\mathbf{A} = 0$ Coulomb gauge. The operator $e^{-\mathbf{r}_L\cdot\nabla}$, meanwhile, represents the transformation between the particle and guiding center positions.

The corresponding field equations are the Poisson's equation and the parallel Ampère's law, $\nabla^2\delta A_\parallel = -4\pi\delta J_\parallel/c$. In the low-frequency and $|k\lambda_D|^2 \ll 1$ limit with λ_D being the Debye length, Poisson's equation can be approximated as the quasi-neutrality condition; $\sum_j n_{0j}q_j < \delta f_j >_v \simeq 0$. Here, $< ... >_v = \int \mathrm{d}\mathbf{v}(...)$ is the velocity-space integral, and subscript j runs over the particle species. Meanwhile, substituting the parallel Ampère's law into the $\nabla\cdot\delta\mathbf{J}\simeq 0$ quasi-neutrality condition as derived by Equation 11.4 yields a generalized linear gyrokinetic vorticity equation, which is often convenient to use in studying shear/kinetic Alfvén wave dynamics.

Linear kinetic Alfvén wave properties

(WARNING: SUPER HARD TO FOLLOW!) For plane wave $\omega, \mathbf{k}$ perturbations, Equation 11.4 gives:

$$\delta g_{\mathbf{k}} = -\frac{q}{T} f_M J_0(k_\perp \mathbf{r}_L)\frac{\omega}{k_\parallel v_\parallel - \omega}\Big(\delta\phi - \frac{v_\parallel}{c}\delta A_\parallel\Big)_{\mathbf{k}}$$

J_0 is the Bessel function and $J_0(k_\perp\mathbf{r}_L)$ corresponds to the gyro-averaging of the coordinate transformation, that is:

$$< \exp(-\mathbf{r}_L\cdot\nabla) >_\alpha = J_0(k_\perp r_L)$$

In SAW/KAW analyses, it is sometimes convenient to introduce an *effective induced parallel potential* defined by $\mathbf{b}_0\cdot\nabla\delta\Psi = -\partial_t\delta A_\parallel/c$ or:

$$\delta\Psi_{\mathbf{k}} = \omega\delta A_{\parallel\mathbf{k}}/(ck_\parallel)$$

$\delta\Psi$, thus, gives rise to the induced parallel electric field; that is, the net parallel electric field is given by:

$$\delta E_\parallel = -\mathbf{b}_0\cdot\nabla(\delta\phi - \delta\Psi) \quad \text{or}$$
$$\delta E_{\parallel\mathbf{k}} = -ik_\parallel(\delta\phi - \delta\Psi)_{\mathbf{k}}$$

The quasi-neutrality condition then straightforwardly yields: (Chen and Hasegawa 1991)

$$\sum_j \Big(\frac{n_0 q^2}{T_0}\Big)_j \{\delta\phi_{\mathbf{k}} + \Gamma_{0kj}[\xi_{kj}Z_{kj}\delta\phi_{\mathbf{k}} - (1 + \xi_{kj}Z_{kj}\delta\Psi_{\mathbf{k}})]\} = 0 \tag{11.5}$$

Here, $\xi_{kj} = \omega/|k_\parallel|v_{tj}, Z_{kj} = Z(\xi_{kj})$ with Z being the well-known plasma dispersion function (Equation 8.33), and $\Gamma_{0kj} = I_0(b_{kj})\exp(-b_{kj})$ with I_0 the modified Bessel function and $b_{kj} = k_\perp^2 r_{Lj}/2 = k_\perp^2(T_j/m_j)/\Omega_j^2$. The linear gyrokinetic vorticity equation, meanwhile, is given by: (Chen and Hasegawa 1991)

$$i\frac{c^2}{4\pi\omega}k_\parallel^2 k_\perp^2 \delta\Psi_{\mathbf{k}} - i\sum_j \left(\frac{n_0 q^2}{T_0}\right)_j (1-\Gamma_{0kj})\omega\delta\phi_{\mathbf{k}} = 0 \tag{11.6}$$

Nothing that, for KAW, $|k_\perp r_{Li}| \sim \mathcal{O}(1)$ and $|k_\perp r_{Le}| \ll 1$ and, thus, $\Gamma_{0ke} \simeq 1$, Equation 11.5 and Equation 11.6 then become

$$\begin{aligned} \epsilon_{s\mathbf{k}}\delta\phi_{\mathbf{k}} &= [1 + \xi_{ke}Z_{ke} + \tau(1 + \Gamma_k \xi_{ki} Z_k i)]\delta\phi_{\mathbf{k}} \\ &= [1 + \xi_{ke}Z_{ke} + \tau\Gamma_k(1 + \xi_{ki}Z_{ki})]\delta\Psi_{\mathbf{k}} \end{aligned} \tag{11.7}$$

and

$$\omega^2\delta\phi_{\mathbf{k}} = k_\parallel^2 v_A^2 \frac{b_k}{1-\Gamma_k}\delta\Psi_{\mathbf{k}} \tag{11.8}$$

Here, $\tau = T_{0e}/T_{0i}, b_k = b_{ki}, \Gamma_k = \Gamma_{0ki}$, and $\epsilon_{s\mathbf{k}}$ is the dielectric constant for the slow-sound (ion-acoustic) wave (SSW).

It is also instructive, as done in some literatures, to define the effective parallel potential, $\delta\phi_{\parallel\mathbf{k}} = \delta\phi_{\mathbf{k}} - \delta\Psi_{\mathbf{k}}$, and rewrite Equation 11.7 and Equation 11.8 as

$$\epsilon_{s\mathbf{k}}\delta\phi_{\parallel\mathbf{k}} = -\tau(1-\Gamma_k)\delta\Psi_{\mathbf{k}} \tag{11.9}$$

and

$$\left[\omega^2 - k_\parallel^2 v_A^2 \frac{b_k}{1-\Gamma_k}\right]\delta\Psi_{\mathbf{k}} = -\omega^2\delta\phi_{\parallel\mathbf{k}} \tag{11.10}$$

Equation 11.9 and Equation 11.10 demonstrate the coupling between SAW and SSW via the finite $|k_\perp r_{Ls}|$ term. In the $|k_\perp r_{Li}| \sim \mathcal{O}(1)$ short-wavelength limit, SAW evolves into KAW due to both the finite $|k_\perp r_{Li}|$ and $|k_\perp r_{Ls}|$ effects. (???) More specifically, the coupled KAW-SSW dispersion relation becomes

$$\omega_{\mathbf{k}}^2\left[1 - \frac{\tau(1-\Gamma_k)}{\epsilon_{s\mathbf{k}}}\right] = k_\parallel^2 v_A^2 \frac{b_k}{1-\Gamma_k} \tag{11.11}$$

Let us concentrate on the KAW branch and, to further simplify the analysis, assume $1 \gg \beta_i \sim \beta_e \gg m_e/m_i$. With $|\omega| \sim |k_\parallel v_A|$, we then have $|\xi_{ki}| = |\omega/k_\parallel v_{ti}| \sim \beta_i^{-1/2} \gg 1 \gg |\xi_{ke}| \sim (m_e/m_i\beta_e)^{1/2}$, and, keeping only the lowest order $\mathbf{O}(1)$ terms:

$$\epsilon_{s\mathbf{k}} \simeq 1 + \tau(1 - \Gamma_k) \equiv \sigma_k$$

From Equation 11.11, we then have

$$\omega_{\mathbf{k}}^2 \simeq k_\parallel^2 v_A^2 \frac{\sigma_k b_k}{1 - \Gamma_k} \tag{11.12}$$

As to wave polarizations, which are useful for wave identification in observations, we can readily derive:

$$\left|\frac{c\delta\mathbf{E}_\perp}{\delta\mathbf{B}_\perp}\right| = v_A \Big[\frac{b_k}{\sigma_k(1 - \Gamma_k)}\Big]^{1/2} \tag{11.13}$$

and

$$\left|\frac{c\delta E_\parallel}{\delta\mathbf{B}_\perp}\right| = v_A \left|\frac{k_\parallel}{k_\perp}\tau\right| \Big[\frac{b_k(1 - \Gamma_k)}{\sigma_k}\Big]^{1/2} \tag{11.14}$$

ADD PLOTS for Equation 11.12, Equation 11.13, and Equation 11.14!!!

Equation 11.14 show that, for a fixed $|k_\parallel/k_\perp|$, $|\delta E_\parallel/\delta\mathbf{B}_\perp|$ increases with b_k. Since wave-particel energy and momentum exchanges are proportional to $|\delta E_\parallel|$, short-wavelength KAW are, thus, expected to play crucial roles in the heating, acceleration, and transport of charged particles.

In addition to having a significant $\delta E_\parallel$, another important property of KAW, in contrast to SAW, is that KAW has a finite perpendicular (to $\mathbf{B}_0$) group velocity, $\mathbf{v}_{g\perp}$. Assuming $|k_\perp r_{Li}|^2 \ll 1$, we have, letting $\omega_A^2 \equiv k_\parallel^2 v_A^2$:

$$\omega_{\mathbf{k}}^2 \simeq \omega_A^2(1 + k_\perp^2 \hat{r}^2) \tag{11.15}$$

where

$$\hat{r}^2 = (3/4 + \tau) r_{Li}^2 \tag{11.16}$$

Thus

$$\mathbf{v}_{g\perp} = \frac{\partial \omega_{\mathbf{k}}}{\partial \mathbf{k}_\perp} \simeq \frac{\omega_A^2}{\omega_{\mathbf{k}}} \hat{r}^2 \mathbf{k}_\perp \tag{11.17}$$

Linear mode conversion of KAW

Equation 11.15 has a significant implication in non-uniform plasmas. Consider, again, a slab plasma with a non-uniform $\omega_A^2(x)$ and $k_\perp^2 = k_x^2(x)$ being the WKB wavenumber in the non-uniformity x-direction. Equation 11.15 then indicates that KAW is propagating ($k_x^2 > 0$) in the $\omega_{\mathbf{k}}^2 > \omega_A^2(x)$ region, and it is cutoff ($k_x^2 < 0$) in the $\omega_{\mathbf{k}}^2 < \omega_A^2(x)$ region. That $\mathbf{v}_{g\perp}$ is finite also suggests that, in constrast to SAW, an initial smooth perturbation will not only evolve into short wavelengths, but also propagate toward the lower $\omega_A^2(x)$ region. These features are illustrated in **?@fig-gk__KAW__evolution**(b); where the spatial-temporal evolution of KAW is solved explicitly according to the following wave equation: (DO IT MYSELF!!!)

$$\left[\hat{r}^2\frac{\partial^2}{\partial x^2} - 1 - \frac{1}{\omega_A^2(x)}\frac{\partial^2}{\partial t^2}\right]\delta B_y(x,t) = 0 \tag{11.18}$$

Equation 11.18 can be readily derived from Equation 11.15 by letting $\omega_{\mathbf{k}} = i\partial/\partial t$ and $k_\perp = -i\partial/\partial x$. The spatial profile of $\omega_A^2(x)/\omega^2 = 1/(1+x^2/L^2)$ is shown in **?@fig-gk__KAW__evolution**(a), with L indicating the profile length-scale, so that the KAW wave-packet frequency is assumed to be consistent with the SAW frequency at $x = 0$. **?@fig-gk__KAW__evolution**(b) shows the propagation of the KAW wave-packet in the direction of radial non-uniformity, consistent with Equation 11.17.

That there exists a finite perpendicular group velocity also implies, in the steady state, the removal of "singular" resonance and linear mode conversion process (Hasegawa and Chen 1976, see also Chapter 12). More specifically, the corresponding wave equation is given by:

$$\left[\hat{r}^2\frac{\partial^2}{\partial x^2} + \left(\frac{\omega_0^2}{\omega_A^2(x)} - 1\right)\right]\delta\hat{B}_y(x) = \delta\hat{B}_{y0} \tag{11.19}$$

Here, ω_0 is the external driving frequency. In the ideal SAW ($\hat{r}\to 0$) limit, there is the resonance singularity at x_0, where $\omega_0^2 = \omega_A^2(x_0)$. Noting that, near $x = x_0$, $\omega_A^2(x) \simeq \omega_0^2 + (\omega_A^2)'(x_0)(x-x_0) \equiv \omega_0^2 - (\omega_0^2/L_A)(x-x_0)$, Equation 11.19 can be approximated as an inhomogeneous Airy equation and solved analytically. Equation 11.19 can then be solved, with appropriate boundary conditions, by connecting the solutions valid away from the $x = x_0$ resonance layer via the analytic solution of the inhomogeneous Airy equation valid near $x = x_0$ (Hasegawa and Chen 1975, 1976). The solutions away from the singular layer are given by:

$$\delta\hat{B}_y(x) = \begin{cases} \frac{\delta\hat{B}_{y0}}{\epsilon_A(x)} & \text{for } \omega_0^2 < \omega_A^2(x) \\ \frac{\delta\hat{B}_{y0}}{\epsilon_A(x)} - \frac{\sqrt{\pi}\delta\hat{B}_{y0}}{(\hat{r}/L_A)^{1/2}}\left(\frac{\hat{r}^2}{\epsilon_A(x)}\right)^{1/4}\exp\left[i\int_{x_0}^{x}\left(\frac{\epsilon_A(x')}{\hat{r}^2}\right)^{1/2}\mathrm{d}x' + i\frac{\pi}{4}\right] & \text{for } \omega_0^2 > \omega_A^2(x) \end{cases} \tag{11.20}$$

where

$$\epsilon_A(x) = \frac{\omega_0^2}{\omega_A^2(x)} - 1$$

The corresponding numerical solutions are plotted in **?@fig-gk__KAW__MHD**. (I DON'T UNDERSTAND!!!)

Both the analytical results and mode conversion process exhibit two important features. One is, instead of being singular, the amplitude at $x = x_0$ (where $\omega_A(x_0) = \omega_0$) is amplified by the Airy swelling factor $(L_A/\hat{r})^{2/3}$. Here, we recall L_A is the scale length of ω_A and $\hat{r}$, from Equation 11.16, is of $\mathcal{O}(r_{Li})$, and, hence, $|L_A/\hat{r}| \gg 1$. The other is the singularity at $x = x_0$ is being replaced by the Airy scale length $\Delta_0 = (\hat{r}^2 L_A)^{1/3}$. Recalling, from Equation 7.31, $|k_x| \simeq |\omega_A'|t \simeq (\omega/L_A)t$, there then exists a KAW formation time scale given by $(\omega_0/L_A)t_0 \simeq 1/\Delta_0$; i.e., $\omega_0 t_0 \simeq (L_A/\hat{r})^{2/3}$. Taking, for an example, a typical laboratory plasma, $L_A/\hat{r} \simeq \mathcal{O}(10^3)$, we have $\omega_0 t_0 \simeq \mathcal{O}(10^2)$, suggesting that it is reasonable to anticipate, in the presence of SAW continuous spectrum, the appearance of KAW in such plasmas.

The main mode identification method for KAWs is based on the measurement of the wave polarization, $|c\delta\mathbf{E}_\perp/v_A\delta\mathbf{B}_\perp|$. Observations can be found from *Van Allen Probes* in the Earth's inner magnetosphere and the *Cluster* satellites in the solar wind.

11.4.2 Nonlinear Gyrokinetics

The idea of deriving the gyrokinetic equations is very similar to the derivation of 5/10-moment equations. Extra care is needed because of the coordinate transformation to the guiding center coordinates. For species α,

$$\frac{\partial f_\alpha}{\partial t} + \mathbf{v}\cdot\frac{\partial f_\alpha}{\partial \mathbf{x}} + \frac{q_\alpha}{m_\alpha}\Big(\mathbf{E} + \frac{\mathbf{v}\times\mathbf{B}}{c}\Big)\cdot\frac{\partial f_\alpha}{\partial \mathbf{v}} = \sum_{\alpha'} C_{\alpha\alpha'}(f_\alpha, f_\alpha')$$

We expand f in different orders:

$$f = f_0 + \delta f = f_0 + f_1 + f_2 + ...$$

and treat the velocity as a first order quantity. For simplicity, we now ignore the species subscript α. Separating the equilibrium and perturbation terms, we have

$$
\begin{aligned}
&\frac{\partial f_0}{\partial t} + \frac{\partial \delta f}{\partial t} \\
&+ \mathbf{v}_\perp \cdot \frac{\partial \delta f}{\partial \mathbf{x}} + \mathbf{v}_\parallel \cdot \frac{\partial \delta f}{\partial \mathbf{x}} \\
&+ \frac{q}{m}\Big(\delta\mathbf{E} + \frac{\mathbf{v}\times\mathbf{B}_0}{c} + \frac{\mathbf{v}\times\delta\mathbf{B}_0}{c}\Big)\cdot\frac{\partial f_0}{\partial \mathbf{v}} + \frac{q}{m}\Big(\delta\mathbf{E} + \frac{\mathbf{v}\times\mathbf{B}_0}{c} + \frac{\mathbf{v}\times\delta\mathbf{B}}{c}\Big)\cdot\frac{\partial \delta f}{\partial \mathbf{v}} = \\
&C(f_0,f_0) + C(f_0,\delta f) + C(\delta f, f_0) + C(\delta f,\delta f)
\end{aligned}
$$

Based on the gyrokinetic ordering,

$$
\begin{aligned}
\frac{\partial f_0}{\partial t} &\sim \omega_{\text{heat}} f_0 \sim \epsilon^3 \Omega f_0 \\
\frac{\partial \delta f}{\partial t} &\sim \omega \delta f \sim \epsilon^2 \Omega f_0 \\
\mathbf{v}_\perp \cdot \frac{\partial \delta f}{\partial \mathbf{x}} &\sim v_{\text{th}} k_\perp \delta f \sim \epsilon \Omega f_0 \\
\mathbf{v}_\parallel \cdot \frac{\partial \delta f}{\partial \mathbf{x}} &\sim v_{\text{th}} k_\parallel \delta f \sim \epsilon^2 \Omega f_0 \\
\frac{q}{m}\nabla_\perp \delta\phi \cdot \frac{\partial f_0}{\partial \mathbf{v}} &\sim \frac{q}{m} k_\perp \delta\phi \frac{f_0}{v_{\text{th}}} \sim \epsilon \Omega f_0 \\
\frac{q}{m}\nabla_\parallel \delta\phi \cdot \frac{\partial f_0}{\partial \mathbf{v}} &\sim \frac{q}{m} k_\parallel \delta\phi \frac{f_0}{v_{\text{th}}} \sim \epsilon^2 \Omega f_0 \\
\frac{q}{mc}\frac{\partial \delta\mathbf{A}}{\partial t} \cdot \frac{\partial f_0}{\partial \mathbf{v}} &\sim \frac{q}{m} \omega \delta A \frac{f_0}{v_{\text{th}}} \sim \epsilon^2 \Omega f_0 \\
\frac{q}{mc}(\mathbf{v}\times\mathbf{B}_0) \cdot \frac{\partial f_0}{\partial \mathbf{v}} &\sim \frac{q}{mc} v_{\text{th}} B_0 \frac{f_0}{v_{\text{th}}} \sim \Omega f_0 \\
\frac{q}{mc}(\mathbf{v}\times\delta\mathbf{B}) \cdot \frac{\partial f_0}{\partial \mathbf{v}} &\sim \frac{q}{mc} v_{\text{th}} \delta B \frac{f_0}{v_{\text{th}}} \sim \epsilon \Omega f_0 \\
\frac{q}{m}\nabla_\perp \delta\phi \cdot \frac{\partial \delta f}{\partial \mathbf{v}} &\sim \frac{q}{m} k_\perp \delta\phi \frac{\delta f}{v_{\text{th}}} \sim \epsilon^2 \Omega f_0 \\
\frac{q}{m}\nabla_\parallel \delta\phi \cdot \frac{\partial \delta f}{\partial \mathbf{v}} &\sim \frac{q}{m} k_\parallel \delta\phi \frac{\delta f}{v_{\text{th}}} \sim \epsilon^3 \Omega f_0 \\
\frac{q}{mc}\frac{\partial \delta\mathbf{A}}{\partial \mathbf{v}} &\sim \frac{q}{m} \omega \delta A \frac{\delta f}{v_{\text{th}}} \sim \epsilon^3 \Omega f_0 \\
\frac{q}{mc}(\mathbf{v}\times\mathbf{B}_0) \cdot \frac{\partial \delta f}{\partial \mathbf{v}} &\sim \frac{q}{mc} v_{\text{th}} B_0 \frac{\delta f}{v_{\text{th}}} \sim \epsilon \Omega f_0 \\
\frac{q}{mc}(\mathbf{v}\times\delta\mathbf{B}) \cdot \frac{\partial \delta f}{\partial \mathbf{v}} &\sim \frac{q}{mc} v_{\text{th}} \delta B \frac{\delta f}{v_{\text{th}}} \sim \epsilon^2 \Omega f_0
\end{aligned}
$$

ω_{heat} is ???

The zeroth order equation is

$$\frac{q}{mc}(\mathbf{v}\times\mathbf{B}_0)\cdot\frac{\partial f_0}{\partial\mathbf{v}} = \Omega(\mathbf{v}_\perp\times\hat{b}_0)\cdot\frac{\partial f_0}{\partial\mathbf{v}} = 0$$

If we tranform to a cylindrical coordinates in velocity space (Equation 10.8), this can be written as

$$(\mathbf{v}_\perp\times\hat{b}_0)\cdot\frac{\partial f_0}{\partial\mathbf{v}} = -\frac{\partial f_0}{\partial\theta} = 0$$

This means that the equilibrium distribution function f_0 does not depend on the gyrophase, i.e. is gyrotropic.

The first order equation is

$$\mathbf{v}_\perp\cdot\frac{\partial\delta f_1}{\partial\mathbf{x}} + \frac{q}{m}\Big(-\nabla_\perp\delta\phi + \frac{\mathbf{v}\times\delta\mathbf{B}}{c}\Big)\cdot\frac{\partial f_0}{\partial\mathbf{v}} + \frac{q}{mc}(\mathbf{v}\times\mathbf{B}_0)\cdot\frac{\partial\delta f_1}{\partial\mathbf{v}} = C(f_0,f_0)$$

Muliplying with $(1+\ln f_0)$ and integrating over phase space, we get

$$\int\ln(f_0)C(f_0,f_0)\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{v} = 0$$

From the proof of Boltzmann's H-theorem, we conclude:

$$\begin{aligned} f_0 = f_{0,M}(v) &= \frac{n_0}{\pi^{3/2}v_{\text{th}}^3}e^{-v^2/v_{\text{th}}^2} \\ C(f_{0,M},f_{0,M}) &= 0 \\ \frac{\partial f_{0,M}}{\partial\mathbf{v}} = -2f_{0,M}\frac{\mathbf{v}}{v_{\text{th}}^2} &= -f_{0,M}m\frac{\mathbf{v}}{T} \end{aligned}$$

The equilibrium distribution is Maxwellian. The first order equation thus becomes (???)

$$\mathbf{v}_\perp\cdot\frac{\partial\delta f_1}{\partial\mathbf{x}} - \Omega\frac{\partial\delta f_1}{\partial\theta} = -f_{0,M}\mathbf{v}\cdot\nabla_\perp\Big(\frac{q\delta\phi}{T}\Big)$$

The solution to this is a combination of a homogeneous solution and a particular solution. For the homogeneous part,

$$\mathbf{v}_\perp\cdot\frac{\partial h}{\partial\mathbf{x}} - \Omega\frac{\partial h}{\partial\theta}\bigg|_{\mathbf{x}} = 0,\ \mathbf{x} = \mathbf{X}+\mathbf{r}_L, \quad \Omega\frac{\partial h}{\partial\theta}\bigg|_{\mathbf{X}} = 0$$

For the particular part, the solution is

$$-q\delta\phi(\mathbf{x},t)f_0(\mathbf{v})/T$$

Thus the complete solution is

$$\delta f_1(\mathbf{x},\mathbf{v},t) = h(\mathbf{X},v_\parallel,v_\perp,t) - q\delta\phi(\mathbf{x},t)f_0(\mathbf{v})/T$$

We still need an equation for h for closure. Therefore we have to go to the second order equation

$$\begin{aligned}\frac{\partial \delta f_1}{\partial t} + v_\parallel \hat{b}_0 \cdot \nabla \delta f_1 + \frac{q}{m}\Big(-\nabla_\perp \delta\phi + \frac{\mathbf{v}\times\delta\mathbf{B}}{c}\Big)\cdot\frac{\partial \delta f_1}{\partial \mathbf{v}} \\ -\frac{q}{m}(\nabla_\parallel \delta\phi)\frac{\partial f_0}{\partial v_\parallel} - \frac{q}{mc}\frac{\partial \delta\mathbf{A}}{\partial t}\cdot\frac{\partial f_0}{\partial \mathbf{v}} \\ +\mathbf{v}_\perp\cdot\nabla\delta f_2 + \frac{q}{mc}(\mathbf{v}\times\mathbf{B}_0)\cdot\frac{\partial \delta f_2}{\partial \mathbf{v}} \\ = C(\delta f_1, f_0) + C(f_0,\delta f_1)\end{aligned}$$

From particle coordinates to guiding center coordinates

$$\begin{aligned}\frac{\partial h}{\partial t} + \frac{\mathrm{d}\mathbf{X}}{\mathrm{d}t}\cdot\frac{\partial h}{\partial \mathbf{X}} + \frac{q}{m}\Big(-\nabla_\perp\delta\phi + \frac{\mathbf{v}\times\delta\mathbf{B}}{c}\Big)\cdot\frac{\partial h}{\partial \mathbf{v}} = \\ C(h,f_0) + C(f_0,h) + \frac{q}{T}f_0\frac{\partial \chi}{\partial t} + \Omega\frac{\partial \delta f_2}{\partial \theta}\bigg|_{\mathbf{X}}\end{aligned}$$

where $\xi \equiv \delta\phi - \mathbf{v}\cdot\delta\mathbf{A}/c$.

For any quantity Q, the gyrophase averaging at fixed guiding center position is

$$\langle Q(\mathbf{x},\mathbf{v},t)\rangle_{\mathbf{X}} \equiv \frac{1}{2\pi}\int_0^{2\pi} Q\Big(\mathbf{X} - \frac{\mathbf{v}\times\hat{b}_0}{\Omega},\mathbf{v},t\Big)\mathrm{d}\theta$$

We have

$$\begin{aligned}\langle h(\mathbf{x},\mathbf{v},t)\rangle_{\mathbf{X}} &= h \\ \langle \mathbf{v}_\perp\cdot\nabla Q\rangle_{\mathbf{X}} = -\Omega\left\langle \frac{\partial Q}{\partial \theta}\bigg|_{\mathbf{X}}\right\rangle_{\mathbf{X}} &= 0\end{aligned}$$

The equation for the gyrocenter distribution function h is then

$$\frac{\partial h}{\partial t} + v_\parallel \hat{b}_0 \cdot \frac{\partial h}{\partial \mathbf{X}} + \frac{c}{B_0}\Big(\hat{b}_0 \times \frac{\partial \langle \chi \rangle_{\mathbf{X}}}{\partial \mathbf{X}}\Big) \cdot \frac{\partial h}{\partial \mathbf{X}} = \langle C(h) \rangle_{\mathbf{X}} + \frac{q}{T} f_0 \frac{\partial < \chi >_{\mathbf{X}}}{\partial t}$$

The quasi-neutrality $\sum_\alpha q_\alpha \delta n_\alpha = 0$ now becomes

$$\sum_\alpha \Big[-\frac{q_\alpha^2 n_{0\alpha}}{T_{0\alpha}} \delta\phi(\mathbf{x}, t) + q_\alpha \int h_\alpha \Big(\mathbf{x} + \frac{\mathbf{v} \times \hat{b}_0}{\Omega_\alpha}, \mathbf{v}, t\Big) \mathrm{d}\mathbf{v}\Big] = 0$$

$$\sum_\alpha \Big(-\frac{q_\alpha^2 n_{0\alpha}}{T_{0\alpha}} \delta\phi(\mathbf{x}, t) + q_\alpha \int < h_\alpha > \mathrm{d}\mathbf{v}\Big) = 0$$

Last but not least we need the modified Ampère's law:

$$\nabla \times \delta \mathbf{B}(\mathbf{x}, t) = \frac{4\pi}{c} \sum_\alpha q_\alpha$$

The distribution function now depends on 6 variables, $f = f(\mathbf{X}, v_\parallel, \mu; t)$. The Vlasov equation can be written as

$$\frac{\partial f}{\partial t} + \dot{\mathbf{X}} \cdot \frac{\partial f}{\partial \mathbf{X}} + \dot{v}_\parallel \frac{\partial f}{\partial v_\parallel} = 0$$

where $\mathbf{X}$ is the gyrocenter position, $v_\parallel$ is the parallel velocity, and μ is the magnetic moment. The appropriate field equations are now

$$\begin{aligned} \frac{n_1}{n_0} &= \frac{\bar{n}_1}{n_0} - (1 - |I_0^2|)\frac{e\phi_1}{T} + |eI_0 I_1| \frac{B_{1\parallel}}{B} \\ \nabla_\perp^2 A_{1\parallel} &= -\frac{4\pi}{c} \sum \bar{J}_{1\parallel} \\ \frac{B_{1\parallel}}{B} &= -\sum \epsilon_\beta \Big(\frac{\bar{p}_{1\perp}}{n_0 T} + |x I_0 I_1| \frac{e\phi_1}{T} + |x^2 I_1^2| \frac{B_{1\parallel}}{B}\Big) \end{aligned}$$

These equations remove the irrelevant space-time scales and become nonlinear 5D equations instead 6D.

11.5 Applications

Gyrokinetics has emerged as the standard approach to plasma turbulence.

Some issues under investigation where gyrokinetics may shed light on:

- Role of microturbulence in space physics?
- transport across boundary layers
- fast magnetic reconnection
- coronal and solar wind heating
- How does MHD turbulence dissipate energy?
- How do fast particles interact with turbulence?

11.6 Limitations

What gyrokinetics cannot resolve:

- cyclotron resonance
- gyro-related phenomena

At least the orginal form of gyrokinetic cannot include EMIC waves and whistler waves.

12 Field Line Resonance

The resonant mode coupling is one of major physical processes in the space plasmas including the magnetic reconnection and collisionless shock wave. The field line resonance describes *the resonant interaction between compressional fast waves and shear Alfvén waves in a non-uniform plasma* such as the Earth magnetosphere. The concept of this resonant mode coupling was first outlined by Tsutomu Tamao in 1961 and later discussed in his seminal paper on hydromagnetic coupling resonances (Tamao, 1965). Experimental work by John Samson+ (1971) made resonant mode coupling the current paradigm of planetary magnetospheric ULF wave research. Thinking of magnetosphere as a microwave oven: instead of heating food with microwaves, we heat Earth with EM waves.

12.1 Historial Review

Before 1950s, no people thought about the idea of standing waves in space simply because a vacuum had been assumed. Right after the space age began, Dungey proposed that standing Alfvén waves could be excited on geomagnetic field lines.

- Ground observations have shown discrete frequencies for EM wave power. In the magnetosphere the Alfvén speed is typically $\sim 1000\,\text{km/s}$, while typical periods of geomagnetic pulsations are 10-600 s. Thus typical wavelengths are $10^4 - 10^6\,\text{km}$, or $1 - 100\,\text{R}_E$, comparable with the size of the magnetosphere itself. So uniform plasma theory is clearly inadequate.
- In the 1950s people realized that MHD waves of poloidal and toroidal modes can be coupled, and if we thought the modes as standing waves, we might explain the discrete frequencies. However, good agreements were not found because there were poor estimates of the magnetospheric plasma density.
- In the 1960s, [Sugiura 1961] showed that waves are observed simultaneously at both ends of the same field line, which indicated that the waves were guided. [Nagata+ 1963] showed pulsations at conjugate points could be matched cycle for cycle. The phase comparison could indicate the possible harmonics of standing waves. [Patel 1965] reported discovery of both transverse and compressional waves in space from Explorer 12 magnetometer data that are correlated with ground measurements, confirming that they were the same. [Cummings+ 1969] showed long-lived and frequent waves with nice numerical comparison of their periods from the poloidal and toroidal mode equations.

- In the 1970s a simpler box model was proposed to explain field line resonance. The idea of standing waves could not explain why only some field lines are preferentially excited (with the foot point of the field line near magnetic latitude 70^o). In the box model, discrete frequencies are treated as sources that reflect the properties of the magnetopause thickness with a preferred KHI growth rate.
- In the 1980s the cavity model was proposed to link the discrete frequencies to the eigenmodes if we treated the magnetosphere as a cavity.
- In the 1985s the waveguide model was proposed based on the cavity idea to loose the constrain in the azimuthal direction. This could explain many ground observation of discrete frequency, large amplitude Pc5 waves but satellite measurements often saw small amplitude Pc3/4 waves. Unlike field line resonances, which are described by a single eigennumber (k_z), cavity modes have three eigennumbers, corresponding to the number of wavelengths along a field line (k_z), azimuthally around the Earth (k_y), and radially between the magnetopause (or bow shock) and the inner reflection point (n). It is argued in the 1990s that if the lowest order mode has a frequency of, say, 2 mHz, then above perhaps 10 mHz the higher order modes will be so numerous and close in frequency that they could probably not be resolved in the data given their inherent width and the frequency resolution of a typical spectrum. Rather they would appear as a continuum. As a continuum, they can not be responsible for exciting discrete frequencies in this higher frequency band (corresponding to Pc3/4). Maybe the waveguide model is only relevant to Pc5 pulsations?

Now, the big question is: where do the quantized numbers of the observed ULF waves come from [Kivelson & Southwood, 1986]?Researchers borrowed ideas from ionospheric radio propagation, laser fusion and plasma physics and came up with the names *cavity* and *waveguide*. For the box model that we will discuss in the next section, if the z boundaries are perfectly reflecting, wave fields must have standing structure in the z direction, and allowed parallel wave numbers are quantized ($k_z = m\pi/c$, as in Equation 12.19). If the boundaries are weakly absorptive, the parallel wave numbers are complex, but the real parts are still quantized as above. If we impose periodic boundary conditions in y, the wavenumber in y k_y needs to be quantized as well.

Another question: if these discretized frequencies correspond to standing waves propagating along the field lines, at the footprints (i.e. ground measurements) the amplitudes shall be the smallest? Why are we still able to observe that on the ground?

12.2 Theory

The theory starts from linearized cold MHD Equation 7.24 and Equation 7.25.

12.2.1 Axisymmetric Spherical Coordinates

First, let's treat the problem in spherical coordinates (r, θ, ϕ), as done by (Radoski and Carovillano 1966).

The background density is taken to be only a function of the radial distance, $\rho = \rho(r)$. In the analysis below the perturbed variables, $\mathbf{B}_1$, $\mathbf{E}$, $\mathbf{j}$ and $\mathbf{u}$ are treated as axisymmetric, i.e. in spherical coordinates any perturbed function has the form

$$= \delta_r(r,\theta)\hat{r} + \delta_\theta(r,\theta)\hat{\theta} + \delta_\phi(r,\theta)\hat{\phi}$$

Under this symmetry, the variables can be separated into two independent sets referred to as the *toroidal* and *poloidal* variables:

- Toroidal variables

$B_{1\phi}, E_r, E_\theta, u_\phi, j_r, j_\theta$; $\nabla\cdot\mathbf{u} = \frac{1}{r\sin\theta}\frac{\partial u_\phi}{\partial\phi} \equiv 0$ (incompressible)

- Poloidal variables

$B_{1r}, B_{1\theta}, E_\phi, u_r, u_\theta, j_\phi$; $\nabla\cdot\mathbf{E} = \frac{1}{r\sin\theta}\frac{\partial E_\phi}{\partial\phi} \equiv 0$

The name of the set were originally based on the magnetic field perturbation directions. One insight into these two sets of variables can be gained from the Poynting vector $\mathbf{S}$. The simplest MHD Ohm's law constrains the electric field $\mathbf{E}$ to be perpendicular to the background magnetic field $\mathbf{B}_0$. For the toroidal mode, $\mathbf{E}$ lies in a meridional plane and $B_{1\phi}$ is out of plane, so $\mathbf{S}$ is directed along the field lines. For the poloidal mode, $\mathbf{S}$ is in a meridional plane and parallel to the wave vector $\mathbf{k}$, because

$$\mathbf{B}_1 = -i\mathbf{k}\times\mathbf{E} \Rightarrow \mathbf{S} \parallel \mathbf{E}\times\mathbf{B}_1 \parallel \mathbf{E}\times(\mathbf{k}\times\mathbf{E}) = E^2\mathbf{k} - \cancel{(\mathbf{E}\cdot\mathbf{k})}\mathbf{E} \parallel \mathbf{k}$$

for a transverse EM wave ($\mathbf{E}\perp\mathbf{k}$). Thus the toroidal mode is an Alfvén mode with dispersion relation $\omega = V_A k\cos\theta$, $\mathbf{S}\parallel\mathbf{B}_0$, while poloidal mode is a fast mode with dispersion relation $\omega = V_A k$, $\mathbf{S}\parallel\mathbf{k}$.

To clarify my early misunderstanding: *a toroidal mode is not propagating in the toroidal direction*! Similarly, *a poloidal mode does not propagate in the poloidal direction.* In fact, it is more common to have a fast poloidal mode propagating in the azimuthal direction, then coupling to the Alfvén toroidal mode along the ambient magnetic field direction, then being observed on the ground.

For the axisymmetric poloidal electric field $E_\phi \sim e^{i\omega t}$, Equation 7.24 becomes (I HAVEN'T DERIVED THIS!)

$$B_0^2[\nabla^2 - (r^2 \sin^2 \theta)^{-1}]E_\phi = -\mu_0 \rho \omega^2 E_\phi$$

This is the decoupled poloidal wave equation. For a dipole field, $B_0^2 = \frac{M^2(1+3\mu^2)}{r^6}$, where M is the magnetic moment and $\mu = \cos\theta$ is the colatitude. A solution for E_ϕ is sought after a separation of the unknowns of the form

$$E_\phi = \sum_{l=1}^{\infty} f_l(r) P_l^m(\mu)$$

where P_l^m is the associated Legendre function with integer indices l and m. Why this form you may ask? It's a mathematical hypothesis based on experience and talents. With this representation the poloidal wave equation becomes

$$\sum_{l=1}^{\infty} [(1+3\mu^2)O_l + 2K^2] f_l P_l^m = 0 \tag{12.1}$$

where $K^2 = \frac{\mu_0 \rho}{2} \left(\frac{\omega^2 r^2}{M^2}\right)^2$ (maybe the coefficient is wrong) and O_l is the spherical Bessel operator

$$O_l = \frac{1}{r}\frac{\partial^2}{\partial r^2} r - \frac{l(l+1)}{r^2}$$

(This is almost out of my control now. If I want to fully understand this, I need to go back to math equations for physics!!!) The special property of Legendre polynomials allows us to write the solution to the wave equation as

$$\sum_{s=0,\pm 2} a_{l+s} O_{l+s} f_{l+s} + K^2 f_l = 0 \tag{12.2}$$

where the constants coefficients are

$$\begin{aligned}
a_{l+2} &= \frac{3(l+2)(l+3)}{2(2l+3)(2l+5)} \\
a_l &= \frac{5l(l+1)-6}{(2l-1)(2l+3)} \\
a_{l-2} &= \frac{3(l-2)(l-1)}{2(2l-3)(2l-1)}
\end{aligned}$$

From Equation 12.2, we know that radial amplitudes of opposite parity do not couple. Since only $l \geq 1$ occurs in Equation 12.1, there are two fundamental unknown radial amplitudes: one for odd l and one for even l.

We assume perfect reflection at the boundaries so that the transverse components of E, i.e. E_θ and E_ϕ, are zeros. The walls are assumed to be rigid so that $u_r = 0$. The normal component of $\mathbf{B}_1$ also vanishes from the governing equations.

The toroidal wave equation seems to be too complicated for me following Radoski's derivations…

From the boundary condition, the length l of the field line between two reflection point must be a multiple of half the wavelength λ, implying

$$n\lambda = 2l, \quad n = 1, 2, 3, ...$$

From the dispersion relation, with the average Alfvén speed $< V_A >$, one finds:

$$\omega_n =< V_A > k =< V_A > \frac{2\pi}{\lambda} = \frac{n\pi < V_A >}{l}$$

12.2.2 Cylindrical Coordinates

In cylindrical coordinates (r, ϕ, z), assuming perturbations of the form $e^{i(m\phi - \omega t)}$, we can separate the linearized equations ((Hughes 1994) I HAVEN'T DERIVED THIS!),

$$\Big[\omega^2\mu_0\rho - \frac{1}{r}(\mathbf{B}_0 \cdot \nabla) r^2 (\mathbf{B}_0 \cdot \nabla)\Big]\frac{u_\phi}{r} = \omega m \frac{\mathbf{B}_0 \cdot \mathbf{B}_1}{r} \tag{12.3}$$

$$\Big[\omega^2\mu_0\rho - rB^2(\mathbf{B}_0 \cdot \nabla)\frac{1}{r^2B^2}(\mathbf{B}_0 \cdot \nabla)\Big] rE_\phi = i\omega B^2 (\mathbf{B}_0 \times \nabla)_\phi \Big(\frac{\mathbf{B}_0 \cdot \mathbf{B}_1}{B^2}\Big) \tag{12.4}$$

$$i\omega \mathbf{B}_0 \cdot \mathbf{B}_1 = \frac{1}{r}(\mathbf{B} \times \nabla)_\phi (rE_\phi) - imB^2 \frac{u_\phi}{r}$$

The first two equations are from the momentum equation. $\mathbf{B}_0 \cdot \nabla$ is the derivative along the direction of $\mathbf{B}$. These are wave equations coupled by the terms on the RHS which depend on $\mathbf{B}_0 \cdot \mathbf{B}_1$, the compressional part of the magnetic perturbation. The third equation is from Faraday's law and Ohm's law. It shows how $\mathbf{B}_0 \cdot \mathbf{B}_1$, E_ϕ and u_ϕ are related and closes the set. Since the transverse dispersion relation depends only on $k_\parallel$, it is tempting to think of the LHS of Equation 12.3 and Equation 12.4 as representing pseudo-transverse mode oscillations, and the RHS as representing the coupling due to the fast mode which has a compressional component. When $m \neq 0$, the phase variation of the toroidal mode in the ϕ direction leads

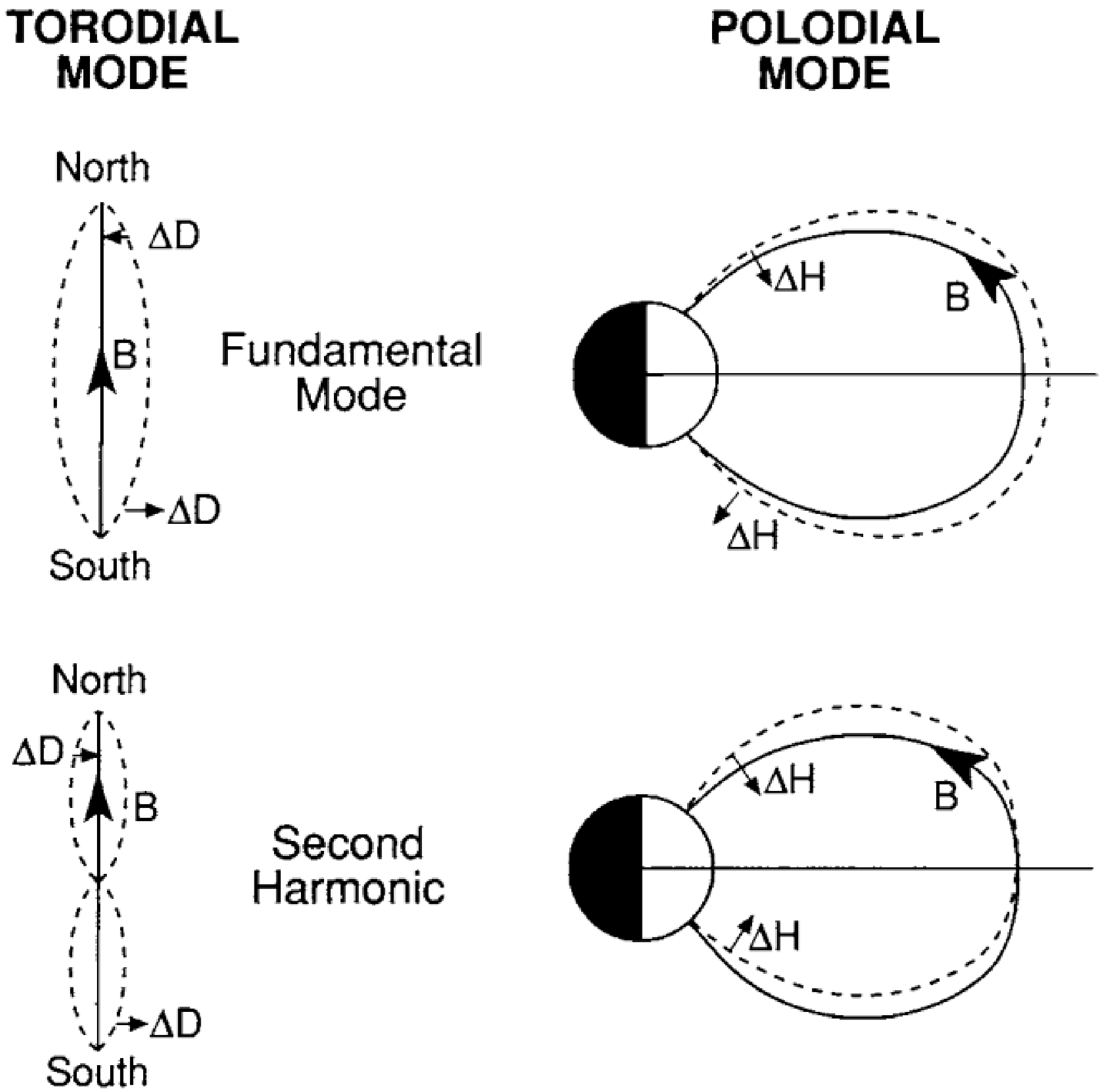

Figure 12.1: Cartoons showing the oscillation of a field line in the two lowest frequency toroidal modes (left) and poloidal modes (right). On the left the field line is drawn stretched out from north to south, and we look towards the Earth as the field line is displaced east and west. On the right, the dipolar field line is displaced within its meridian plane so all magnetic perturbations are radial. Note that for the toroidal mode, the magnetic perturbations (AD) have the opposite sense north and south in the fundamental and the same sense in the second harmonic. The opposite is true of the AH perturbation in the poloidal mode.

to a compressional perturbation in the magnetic field. The polarizations of the toroidal and poloidal oscillations are no longer orthogonal, and thus there is coupling.

If $m = 0$ the RHS of Equation 12.3 vanishes. The LHS then describes a mode in which the electric field is purely radial and the magnetic and velocity perturbations are azimuthal. Magnetic L shells decouple and each shell oscillates azimuthally independently of each other. This is the Alfvénic *toroidal* mode. If $m \rightarrow \infty$, for the RHS to remain finite, $\mathbf{B}_0 \cdot \mathbf{B}_1 \rightarrow 0$ so the RHS of Equation 12.4 vanishes. Equation 12.4 then describes a mode in which $\mathbf{E}$ is azimuthal and $\mathbf{u}$ and $\mathbf{b}$ are contained in a meridian plane. This is the compressional *poloidal* mode.

12.2.3 Cartesian Box Model

Earliest theories had cold ideal MHD equations Equation 7.24 or Equation 7.25 expressed in the spherical or cylindrical coordinates, which are not easy to solve. As we will see, the essence of FLR can be obtained in the Cartesian coordinates. We simplify the actual dayside magnetosphere, which like a compressed dipole, to something we can solve analytically. Imagine a field line with both footpoints connecting to the conducting ionosphere, we can map this curved field line into a straight line extending along z. In the x-direction, the outer boundary is the magnetopause, and the inner boundary is the reflection point (plasmapause or ionosphere). The y-direction represents the azimuthal direction, therefore at the front of the magnetosphere it is more aligned with the y-direction in GSE/GSM and at the flanks/sides it is more aligned with the x-direction. If the length in y is finite ($b < \infty$), we call it a *cavity*; if the length in y is infinite ($b \rightarrow \infty$), we call it a *waveguide*. The names cavity and waveguide are inherited from electrodynamics. This is shown in the schematic Figure 12.2 and Figure 12.3.

Along the z-direction, we impose the ionospheric boundary for closed field lines $u_x = u_y = B_{1z} = 0$. Because of the ideal MHD assumption, the electric field along the field lines $E_z = 0$. In the y-direction, we impose either a periodic boundary condition for the cavity model that leads to a quantized wavenumber k_y, or a infinite boundary. In the x direction, the boundaries are provided by large V_A gradients at both the outer boundary x_m (magnetopause) and the inner boundary x_t (plasmapause). We impose a reflective boundary $u_x = 0$ at $x = x_t, x_m$. *The inhomogeity in x means that we can only consider Fourier components in the y and z directions.* Therefore, we seek wave modes of the form $\propto \exp[i(k_y y + k_z z - \omega t)]$ with the ansatz

$$
\begin{aligned}
\mathbf{k} &= (0, k_y, k_z) \\
\mathbf{E} &= (E_x(x), E_y(x), 0)e^{i(k_y y + k_z z - \omega t)} \\
\mathbf{B}_1 &= (B_{1x}, B_{1y}, B_{1z})e^{i(k_y y + k_z z - \omega t)} \\
\mathbf{B}_0 &= B_z \hat{z} = B\hat{z} \\
\rho &= \rho(x) \\
\mathbf{u} &= \mathbf{u}_0 + \mathbf{u}_1 = \mathbf{u}_1
\end{aligned}
$$

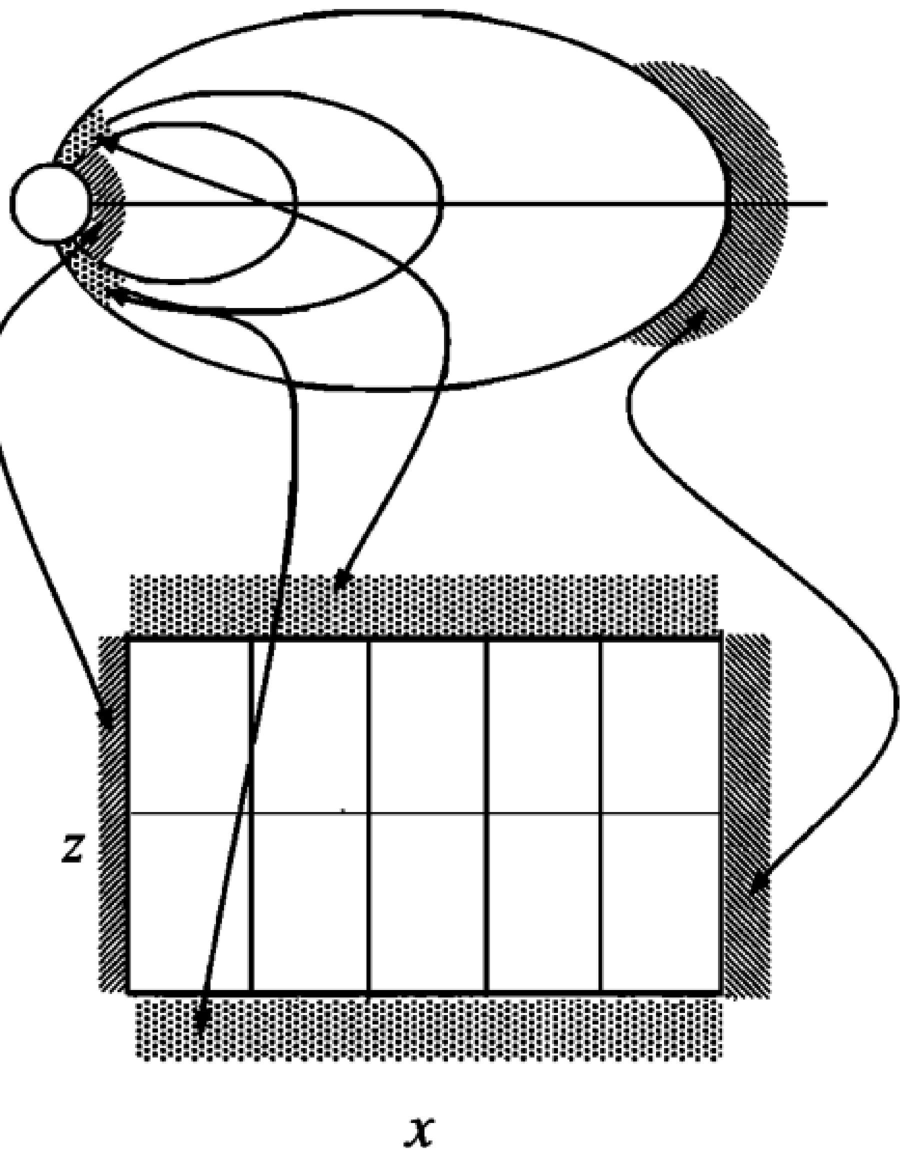

Fig. 2. Schematic of the mapping of the quasidipolar magnetosphere onto the cuboidal box used for the analysis of this paper. Boundaries corresponding to the magnetopause, the northern hemisphere and southern hemisphere ionospheres, and a low latitude near-equatorial boundary surface are illustrated.

Figure 12.2: From one of Southwood & Kivelson's paper.

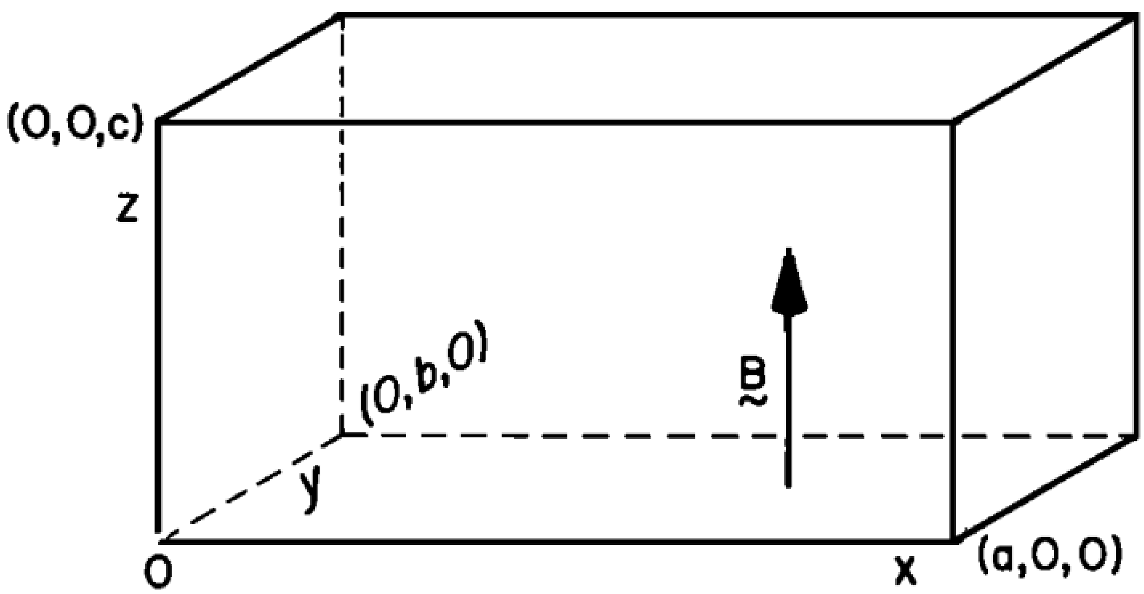

Fig. 3. **Schematic representation of the magnetosphere as a rectangular box containing a uniform magnetic field parallel to the z axis. The ionospheric boundaries are planes at $z = 0$ and $z = c$. The magnetopause is a plane at $x=a$ and a reflecting boundary is assumed at $x=0$. The y-direction represents the azimuthal direction in the magnetosphere. Plasma density is assumed to be either constant (the uniform model) or to vary with x (the nonuniform model).**

Figure 12.3: From one of Southwood & Kivelson's paper.

The linearized momentum Equation 7.23 gives

$$
\begin{aligned}
\rho\Big(\frac{\partial \mathbf{u}}{\partial t} + \cancel{\mathbf{u}\cdot\nabla\mathbf{u}}\Big) &= \mathbf{j}\times\mathbf{B} \\
\rho\frac{\partial \mathbf{u}}{\partial t} &= \frac{1}{\mu_0}(\nabla\times\mathbf{B})\times\mathbf{B}
\end{aligned}
$$

Separating the two perpendicular x and y directions and applying plane wave decomposition,

$$
\begin{aligned}
-i\omega\rho\mu_0 u_x &= (\nabla\times\mathbf{B}_1)_y B_z - (\nabla\times\mathbf{B}_1)_z \cancel{B_y} \\
&= \Big(\frac{\partial B_{1x}}{\partial z} - \frac{\partial B_{1z}}{\partial x}\Big)B \\
&= -\frac{\partial}{\partial x}\Big(BB_{1z}\Big) + ik_z BB_{1x}
\end{aligned}
$$

$$
\begin{aligned}
-i\omega\rho\mu_0 u_y &= (\nabla\times\mathbf{B}_1)_z \cancel{B_x} - (\nabla\times\mathbf{B}_1)_x B_z \\
&= \Big(\frac{\partial B_{1y}}{\partial z} - \frac{\partial B_{1z}}{\partial y}\Big)B \\
&= k_z BB_{1y} - k_y BB_{1z}
\end{aligned}
$$

From the simplest form of the generalized Ohm's law $\mathbf{E} = -\mathbf{u}_1\times\mathbf{B}_0$,

$$
\begin{aligned}
E_x &= -u_y B \\
E_y &= -u_x B
\end{aligned}
$$

Inserting into the linearized momentum equation, we have

$$
\begin{aligned}
-i\omega\rho\mu_0\frac{E_y}{B} &= -\frac{\partial}{\partial x}\Big(BB_{1z}\Big) + ik_z BB_{1x} \\
\omega\rho\mu_0\frac{E_x}{B} &= -k_y BB_{1z} + k_z BB_{1y}
\end{aligned} \tag{12.5}
$$

From linearized Faraday's law $\nabla\times\mathbf{E} = -\dot{\mathbf{B}}_1$,

$$
\begin{aligned}
-\frac{\partial E_y}{\partial z} &= -\dot{B}_{1x} \\
\frac{\partial E_x}{\partial z} &= -\dot{B}_{1y} \\
\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} &= -\dot{B}_{1z}
\end{aligned}
$$

$$
\begin{aligned}
B_{1x} &= \frac{1}{\omega}(-k_z E_y) \\
B_{1y} &= \frac{1}{\omega}(k_z E_x) \\
B_{1z} &= \frac{1}{\omega}\Big(-i\frac{\partial E_y}{\partial x} - k_y E_x\Big)
\end{aligned}
\tag{12.6}
$$

Substituting Equation 12.6 into Equation 12.5 gives

$$
\begin{aligned}
-i\omega\rho\mu_0\frac{E_y}{B} = -\frac{\partial}{\partial x}\Big[B\frac{1}{\omega}\Big(-i\frac{\partial E_y}{\partial x} - k_y E_x\Big)\Big] + ik_z B\frac{-k_z}{\omega}E_y \\
-i(\omega^2\rho\mu_0/B^2 - k_z^2)E_y = k_y\frac{\partial E_x}{\partial x} - i\frac{\partial^2 E_y}{\partial x^2}
\end{aligned}
$$

$$
\begin{aligned}
-i\omega\rho\mu_0\frac{E_x}{B} = k_z B\frac{k_z}{\omega}E_x - k_y B\frac{1}{\omega}\Big(-i\frac{\partial E_y}{\partial x} - k_y E_x\Big) \\
(\omega^2\rho\mu_0/B^2 - k_y^2 - k_z^2)E_x + ik_y\frac{\partial E_y}{\partial x} = 0
\end{aligned}
$$

Let $R^2 = \mu_0\rho(x)\omega^2/B^2 = \omega^2/V_A(x)^2$, we have

$$
\begin{aligned}
-i(R^2 - k_z^2)E_y = k_y\frac{\partial E_x}{\partial x} - i\frac{\partial^2 E_y}{\partial x^2} \\
(R^2 - k_y^2 - k_z^2)E_x + ik_y\frac{\partial E_y}{\partial x} = 0
\end{aligned}
\tag{12.7}
$$

Eliminating E_x from Equation 12.7 gives

$$
\frac{\partial^2}{\partial x^2}E_y - k_y^2\frac{\partial}{\partial x}R^2(x)\frac{1}{(R^2 - k_z^2)(R^2 - k_z^2 - k_y^2)}\frac{\partial E_y}{\partial x} + (R^2 - k_z^2 - k_y^2)E_y = 0 \tag{12.8}
$$

There is a very important observation about polarization from Equation 12.7:

$$
\frac{E_x}{E_y} = -\frac{ik_y}{R^2 - k_y^2 - k_z^2}\frac{\partial E_y}{\partial x}\frac{1}{E_y} \tag{12.9}
$$

Our familiar fast MHD wave is *no longer linearly polarized in a nonuniform plasma*! A key thing on the RHS is the dependence on the sign of k_y as well as $\partial E_y/\partial x$. For fast waves propagating in opposite directions (i.e. eastward and westward in the azimuthal direction), the sense of polarization (RH, LH) will change as k_y changes sign across local noon. Right at the resonance point where $R^2 - k_z^2 - k_y^2 = 0$, we get linearly polarized Alfvén wave. $\partial E_y/\partial x$

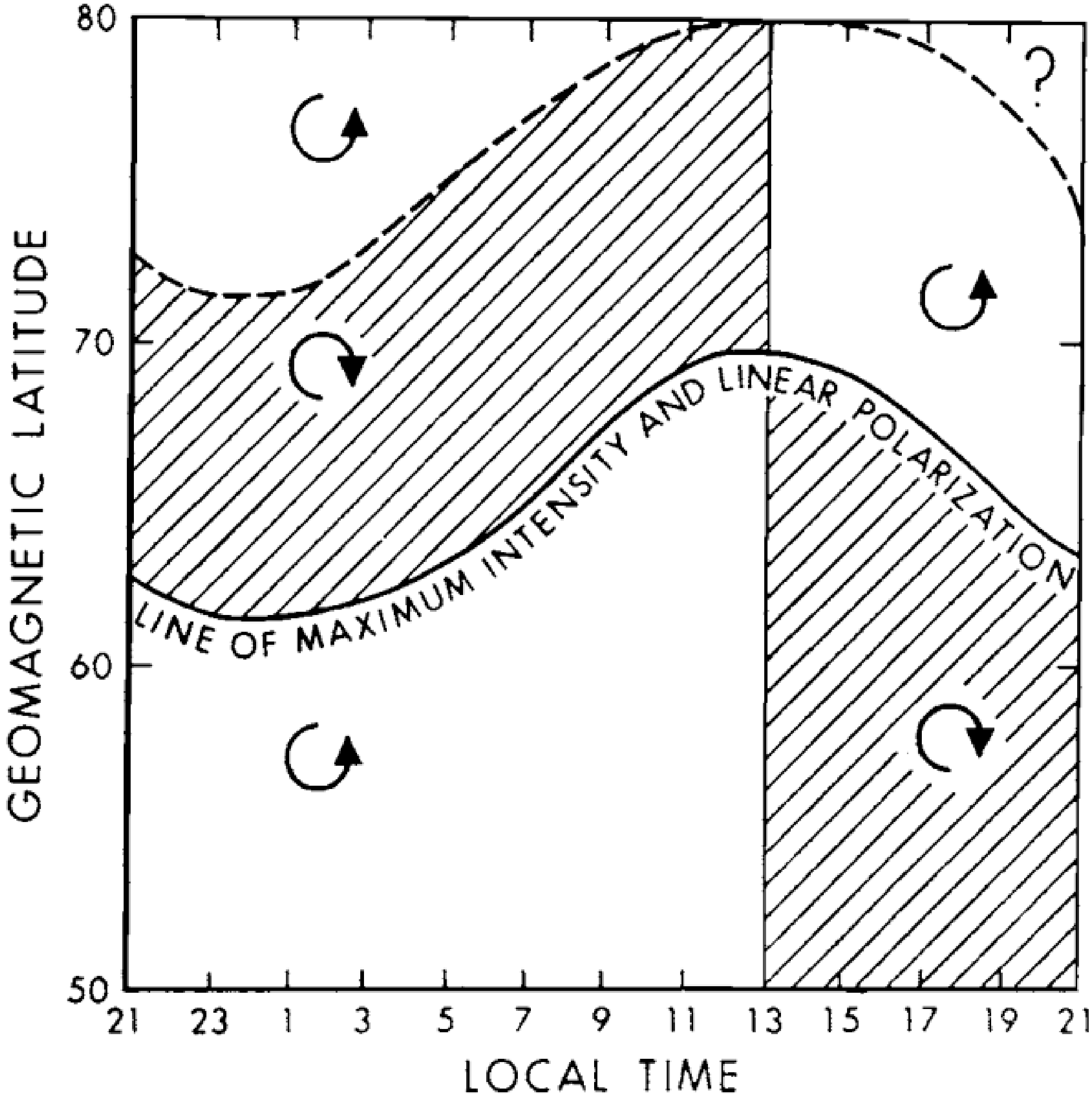

Figure 12.4: The variation of amplitude and the sense of polarization of Pc5's seen at high latitudes with latitude and magnetic local time. (After [Samson +, 1974].)

changes sign across the resonance point, thus we also have another change in polarization detected on the ground for different L-shells (or magnetic latitudes).

Equation 12.8 can also be written as a second order differential equation of B_{1z}, the compressional component of the magnetic field (see Equation 3.6), as (Kivelson and Southwood 1985) did in proposing the idea of cavity modes:

$$\frac{\partial^2 B_{1z}}{\partial x^2} - \frac{\frac{\partial R^2}{\partial x}}{R^2 - k_z^2}\frac{\partial B_{1z}}{\partial x} + (R^2 - k_y^2 - k_z^2)B_{1z} = 0$$

or equivalently,

$$\frac{\partial^2 B_{1z}}{\partial x^2} - \frac{\omega^2 \partial V_A^{-2}/\partial x}{\omega^2/V_A^2 - k_z^2}\frac{\partial B_{1z}}{\partial x} + \Big(\frac{\omega^2}{V_A^2} - k_y^2 - k_z^2\Big)B_{1z} = 0 \tag{12.10}$$

Let us first get some intuitions about Equation 12.8 (following (Glassmeier et al. 1999), but note that there are sign errors and wrong equations in the original paper!). This equation exhibits strong singularities found in the denominator of its second term, much as first described by Tamao (1965). The following solutions are possible. If $R^2 - k_z^2 - k_y^2 > 0$, from which $R^2 - k_z^2 > 0$ follows, no singularities appear. Assuming $k_y \approx 0$, the above equation reduces to

$$\frac{\partial^2}{\partial x^2}E_y + (R^2 - k_z^2)E_y = 0$$

For a linear density profile, i.e. $R^2 = \alpha_0^2 + \alpha^2 x$, with the definition of the turning point x_t via $R^2 = k_z^2$, and the transformation $s = \alpha^{2/3}(x - x_t) > 0$,

$$\begin{aligned} \frac{\partial E_y}{\partial x} &= \frac{\partial E_y}{\partial(\alpha^{-2/3}s + x_t)} = \alpha^{2/3}\frac{\partial E_y}{\partial s} \\ \frac{\partial^2 E_y}{\partial x^2} &= \alpha^{4/3}\frac{\partial^2 E_y}{\partial s^2} \end{aligned}$$

$$\frac{\partial^2 E_y}{\partial s^2} + sE_y = 0$$

This is the Airy or Stokes equation (Section 3.10.1) with the two principal solutions displayed in **?@fig-airy_flr**. The solution $Bi(s)$ is unphysical as it implies unlimited growth of E_y behind the turning point at $s = 0$. Thus $E_y(s) = Ai(s)$ is the required solution. The turning point actually is the point of total reflection of the wave field. Its appearance can be understood on the following grounds. The "effective" local wavenumber in x is given by

$$k_x^2(x,\omega) = \frac{\omega^2}{V_A^2(x)} - k_y^2 - k_z^2 \tag{12.11}$$

If V_A^2 increases with x, that is with s in the Airy function plot, k_x^2 has to decrease as ω, k_y and k_z stay constant. Eventually k_x^2 may become negative, which implies an imaginary wave number k_x. At this turning point the wave will be reflected.

```
KeyNotes.plot_airy_minus_x()
```

Next if $R^2 - k_z^2 - k_y^2 < 0$, $R^2 - k_z^2 = 0$ may occur. Assuming again a linear density profile, defining a resonance point x_r via $R^2 = k_z^2$, and $s = x - x_r$, the electric field perturbation transforms close to the resonance point $x - x_r = 0$ into

$$\frac{\partial^2 E_y}{\partial s^2} - \frac{1}{s}\frac{s_t}{s - s_t}\frac{\partial E_y}{\partial s} + \alpha^2(s - s_t)E_y = 0 \tag{12.12}$$

where $s_t = k_y^2/\alpha^2$.

At the resonance point its solution exhibits a clear singularity with unlimited growth of E_y. In front of the turning point $s > s_t$ the solution is similar to an Airy function, while behind it singular behaviour is observed at the resonance point $s = 0$. I do not show the solution BECAUSE I DON'T KNOW HOW TO WRITE IT DOWN! But let us now follow (Kivelson and Southwood 1985) and check the equivalent perturbation Equation 12.10 for B_{1z}.

When we also assume linear mass density variation with x in Equation 12.10 such that

$$\begin{aligned} R^2 - k_z^2 &= \alpha^2(x - x_r) \\ R^2 - k_z^2 - k_y^2 &= \alpha^2(x - x_t) \end{aligned}$$

where $x_r = k_z^2/\alpha^2$, $x_t = x_r + k_y^2/\alpha^2$, we have

$$\frac{\partial^2 B_{1z}}{\partial x^2} - \frac{1}{x - x_r}\frac{\partial B_{1z}}{\partial x} + \alpha^2(x - x_t)B_{1z} = 0 \tag{12.13}$$

The solution of Equation 12.13 is discussed by (Kivelson and Southwood 1986). The singular point where $x = x_r$ and $R^2 = k_z^2$ corresponds to the *resonance point* where the wave frequency matches the Alfvén mode frequency. The point where $x = x_t$ and $R^2 = k_y^2 + k_z^2$ is the *turning point* of the equation. When $x > x_t$, any wave described by the equation oscillates with x; when $x < x_t$ independent solutions grow or evanesce with x on each side of the resonance singularity x_r.

Series solutions valid in the neighborhood of $x = x_r$ and $x = x_t$ have been derived by a variety of authors. The Frobenius method of solution of partial differential equations can be used to find solutions valid near x_r, where the analytic solution is of the form

$$R_t(x) = k_y^2(x - x_r)^2 + \mathcal{O}[(x - x_r)^4] \tag{12.14}$$

and the singular solution is of the form

$$s_t(x) = 1 + \frac{1}{2}R_t(x)\ln(x_r - x) + \mathcal{O}[(x - x_r)^2] \tag{12.15}$$

For $k_y^2 = \alpha^{4/3}$, the two solutions for B_{1z} is shown in Figure 12.5. On the far right side of the plot the functions are close to spatial quadrature and have zeros whose spacing decreases with increasing x. In this regime, as we will mention later, the WKB approximation is good and the amplitude and phase are proportional to $s^{1/4}$ and $s^{3/2}$, respectively. The amplitudes have been plotted just such that an incoming wave from large x would be represented by $R_t(x)+iS_t(x)$ if a time variation $e^{-i\omega t}$ is specified. The corresponding mode propagating toward large x would be $R_t(x) - iS_t(x)$. Thus $R_t(x), S_t(x)$ represent standing wave forms. However, the presence of the resonance (at the far left of the plots) preludes there being perfect reflection or perfect standing wave solutions.

```
@sco KeyNotes.plot_b1z_ode()
```

The following physical interpretation is tempting. The MHD wave propagating into the magnetosphere is a fast mode wave generated by, e.g. plasma instabilities at the magnetopause. Eventually the wave reaches the turning point where reflection occurs. If conditions are favourable, that is if a resonance point occurs, part of the fast mode wave energy can tunnel into the resonance, where coupling from the fast mode disturbance into an Alfvén mode perturbation takes place. The resonance point *always* lies beyond the turning point, but energy tunnels to the resonance point and the reflection is found to be less than perfect. This scenario is schematically shown in Figure 12.6, where the reflection or turning point is assumed to coincide with the magnetopause. The point of maximum coupling or "resonant mode coupling" is given by $\omega = k_\parallel V_A(x)$, where ω is the fast mode wave frequency, $k_\parallel$ is the field-aligned component of the fast mode wave vector and $V_A(x)$ is the local Alfvén velocity. This is also the Alfvén dispersion relation.

Equation 12.10 has a long history that appeared first in the context of a radio wave obliquely incident on a region of increasing electron density, all the way back to 1951, as mentioned in (Kivelson and Southwood 1986). It also appears in calculating the absorption of laser energy by a plasma and heating a plasma by radio wave injection to excite internal Alfvén waves. Energy is absorbed at the location in the plasma corresponding to the singular point of the governing differential Equation 12.10 or its equivalent. Without the introduction of additional effects

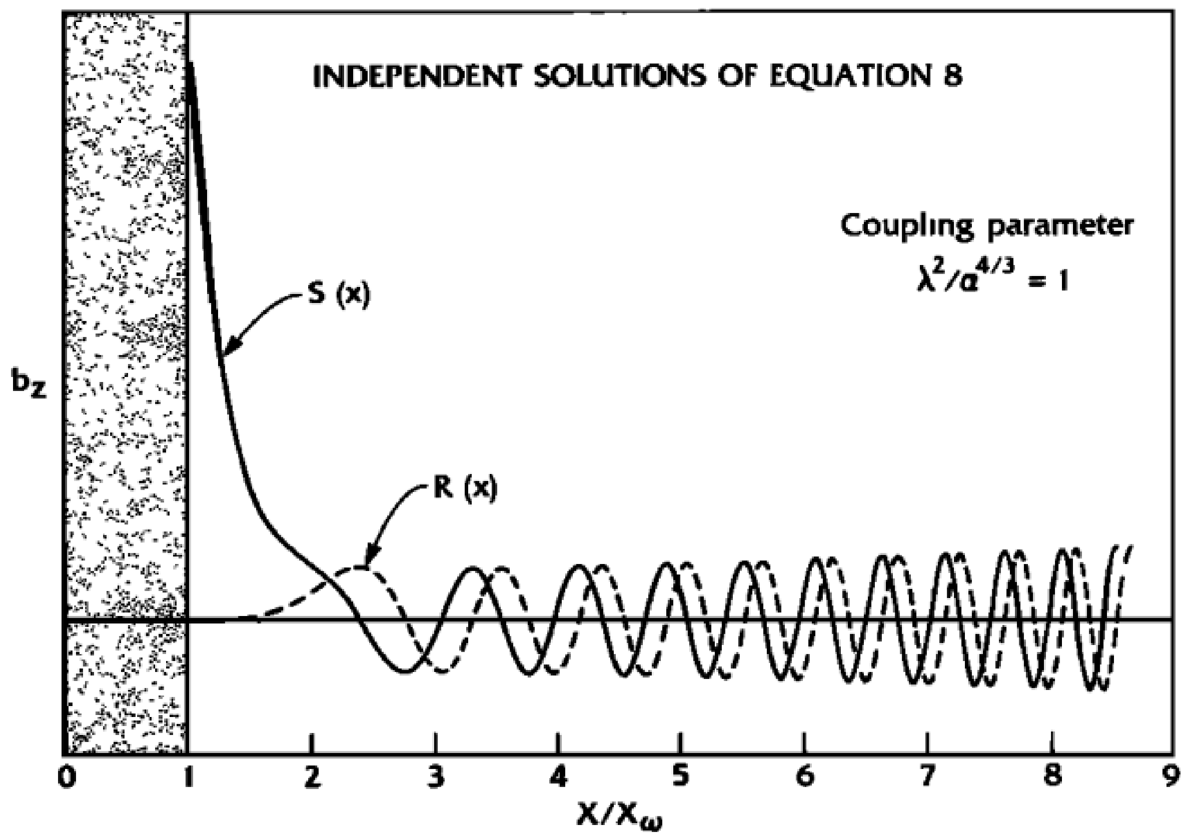

Figure 12.5: Two independent solutions of Equation 12.13 for $x > x_r$. The particular pair illustrated are solutions for the case $\lambda^2/\alpha^{4/3} = 1$ for which $x_t/x_r = 2$. The dashed curve is the solution that is analytic at the singular point (cf. Equation 12.14). The solution plotted as a solid curve is finite at the singularity with infinite derivatives as evident from Equation 12.15. Amplitudes have been selected to match at large x where the oscillating solutions are in spatial quadrature.

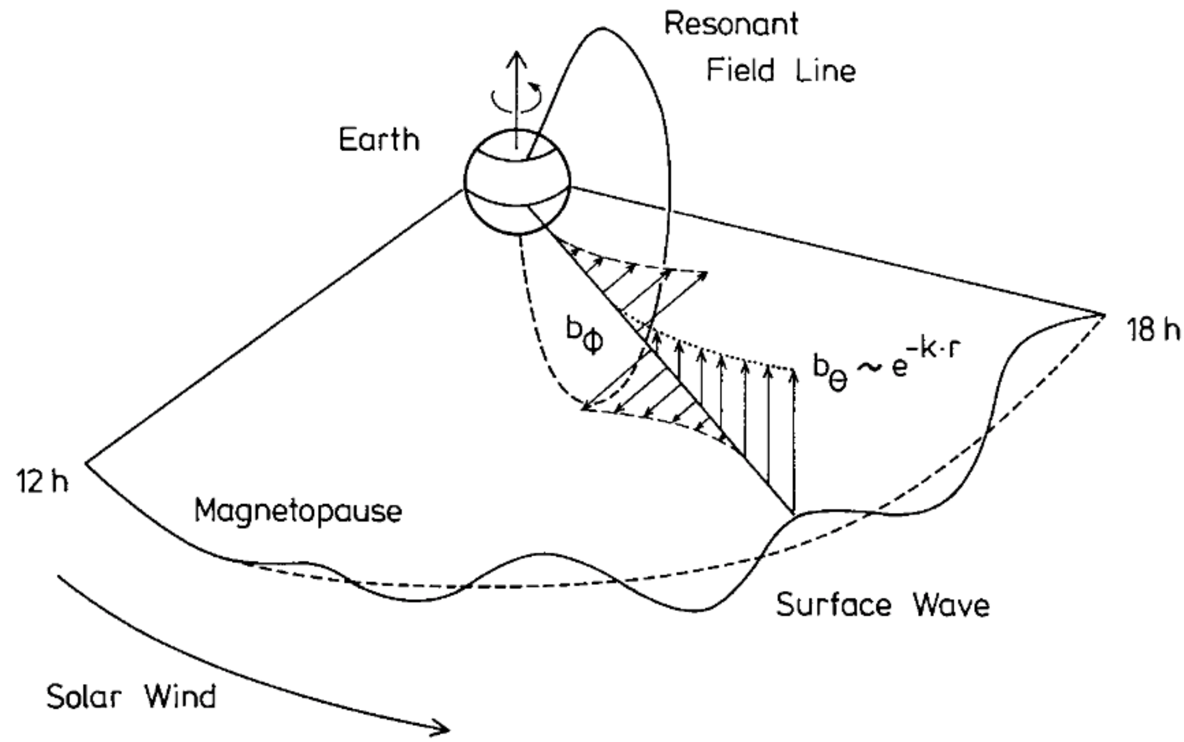

Figure 12.6: Schematic view of the field line resonance, driven by unstable surface wave.

such as dissipation or dispersion near resonance, the singularity in the differential equation implies that the amplitude of the Alfvén wave grows secularly.

Some studies (e.g. (Kivelson and Southwood 1986), (Wright 1994)) prefer to use plasma displacement $\boldsymbol{\xi}$ instead of $\mathbf{u}$, where $\mathbf{u} = \dot{\boldsymbol{\xi}}$, and perturbed magnetic field component B_{1z} to describe the model:

$$
\begin{aligned}
\frac{1}{V_A^2}\frac{\partial^2 \xi_x}{\partial t^2} - \frac{\partial \xi_x}{\partial z^2} &= -\frac{1}{B_0}\frac{\partial B_{1z}}{\partial x} \\
\frac{1}{V_A^2}\frac{\partial^2 \xi_y}{\partial t^2} - \frac{\partial \xi_y}{\partial z^2} &= -\frac{1}{B_0}\frac{\partial B_{1z}}{\partial y} \\
B_{1z} &= -B_0\Big(\frac{\partial \xi_x}{\partial x} + \frac{\partial \xi_y}{\partial y}\Big)
\end{aligned}
\tag{12.16}
$$

This equation set is equivalent to Equation 7.25. The first two equations are the x and y component of the momentum equation, and the last equation is the magnetic field advection equation (Faraday's law + Ohm's law). Recall that B_{1z} is the compressional perturbation to the ambient magnetic field. Then on the RHS of the second equation, it is the azimuthal gradient in the fast mode $\partial B_{1z}/\partial y$ that drives the response in the azimuthal place displacement ξ_y. It is like the LHS is a simple harmonic oscillator and the RHS is the forcing term. Recall that we are seeking for solutions of the form $e^{i(k_y y + k_z z - \omega t)}$. From here we can see that:

- if $k_y = 0$ for the fast mode, then there are no azimuthal gradients to drive FLRs and hence you get no wave coupling;
- if $k_y \to \infty$, then $\xi_y \to 0$ and ξ_x describes a decoupled poloidal Alfvén wave.
- if $k_y \neq 0$ or ∞, the wave modes are coupled together and energy initially in the fast mode may mode convert to Alfvén wave energy on localized field lines.

Uniform density and background magnetic field

Let $\rho = \text{const.}$, $V_A = \text{const.}$. From Equation 12.16,

$$
\begin{aligned}
\ddot{B_{1z}} &= -B_0\Big(\frac{\partial \ddot{\xi_x}}{\partial x} + \frac{\partial \ddot{\xi_y}}{\partial y}\Big) = V_A^2\Big(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\Big)B_{1z} \\
&= V_A^2 \nabla^2 B_{1z}
\end{aligned}
$$

Consider normal modes of waveguides of the form $e^{\mathbf{k}\cdot\mathbf{r} - i\omega t}$, where $\mathbf{k} = (k_x, k_y, k_z)$. The dispersion relation for the fast mode is then

$$
\omega^2 = V_A^2(k_x^2 + k_y^2 + k_z^2) \tag{12.17}
$$

In an infinite medium this relation gives nondispersive propagation. Since we are interested primarily in propagation along the waveguide (i.e., in the y direction) we shall define v_p and v_g to be the phase and group velocities along $\hat{y}$. Employing Equation 12.17 gives

$$v_p = \frac{\omega}{k_y} = \frac{k}{k_y} V_A$$
$$v_g = \frac{\partial \omega}{\partial k_y} = \frac{k_y}{k} V_A$$

which yields the familiar waveguide relation

$$v_p v_g = V_A^2 \tag{12.18}$$

The shear Alfvén wave has the same velocity parallel to the field line (because it is cold). However, in a waveguide the boundary conditions in the x and z directions restrict the choice of wavenumbers and introduce dispersion. Suppose that the boundaries in x are perfectly reflecting (e.g. $\xi_x = \frac{\partial B_{1z}}{\partial y} = 0$) as are those in z ($\xi_x = \xi_y = B_{1z} = 0$) which represent the ionospheric boundary for closed field lines. Then

$$\begin{aligned} k_x &= \pm n\pi/a, \quad n = 1,2,3... \\ k_z &= \pm m\pi/c, \quad m = 1,2,3... \end{aligned} \tag{12.19}$$

where a is the box length in x and c is the box length in z (i.e. length of the field lines).

Given values of k_x and k_z, we may use Equation 12.17 to find k_y as a function of ω,

$$k_y^2 = \frac{\omega^2}{V_A^2} - k_x^2 - k_y^2$$

If k_y is real then the mode may propagate along the guide. However, if we believe that the boundary at $x = x_m = a$ is not a reflector but is driven, it would be appropriate to impose a wavenumber k_y along the outer boundary and solve for k_x given k_z. This is the case when the magnetopause is driven by the Kelvin-Helmholtz instability (i.e. k_y given by the surface waves), which was proposed in the 1970s to be the driver of field line resonances. In this situation k_x is found to be imaginary, and the mode is evanescent (exponentially decaying) in x.

Nonuniform density and magnetic field

The simplest assumptions in this case would be $\rho = \rho(x)$, $\mathbf{B}_0 = B_0\hat{z}$, and that the Alfvén speed V_A is monotonically decreasing with x, $\partial V_A/\partial x < 0$. Then again we get Equation 12.10. Solving this equation for given boundary conditions in x yields a set of orthogonal eigenfunctions $B_{1z}(x)$ and eigenfrequencies.

First of all, we shall notice that a singularity/resonance occurs at x_r where

$$\omega^2 = k_z^2 V_A^2(x_r) \tag{12.20}$$

Across x_r there will be a 180^o phase shift for the perturbed terms (Equation 12.9). Energy is been absorbed at the resonance point from fast mode to Alfvén mode. Dissipation is required, otherwise the amplitude of the Alfvén wave grows secularly. Ionospheric dissipation near the resonant field line is likely to be one important process limiting the growth of the resonance amplitude.

Secondly, there is a turning point at x_t defined via $k_x^2 = 0$ (fixed k_y, k_z, ω):

$$\omega^2 = (k_y^2 + k_z^2) V_A^2(x_t) \tag{12.21}$$

On the right of x_t, the positive exponent represents a wave propagating in the positive x direction towards the magnetopause, and the negative exponent represents one propagating away from the magnetopause. This is like an imperfect standing wave solution. Across x_t on the left, the oscillating solutions convert to a decaying solution and an exponentially growing solution (Section 3.10.1).

If the Alfvén speed V_A is monotonically decreasing with x, from Equation 12.20 and Equation 12.21 we always have $0 < x_r < x_t < x_m$. In the low Alfvén speed region $[x_t, x_m]$ the fast mode may propagate, while in the high Alfvén speed region $[0, x_t]$ the fast mode is evanescent. Thus the resonant singularity x_r is in the evanescent tail of the fast mode.

The first-order derivative term in Equation 12.10 dominate near $x = x_r$, but is small in the propagating interval $[x_t, x_m]$. This is particularly true in the WKB limit (Section 3.10), where we consider waves of short wavelength in x; i.e. second-order derivative term is much larger than the first-order derivative term. In the lowest order WKB analysis we may neglect the first-order derivatives and solve

$$\frac{\partial^2 B_{1z}}{\partial x^2} + \Big(\frac{\omega^2}{V_A^2} - k_y^2 - k_z^2\Big) B_{1z} = 0 \tag{12.22}$$

and the WKB solution is given in the form

$$B_{1z}(x) = A k_x^{-1/2} e^{i\int k_x dx} + B k_x^{-1/2} e^{-i\int k_x dx}$$

where A and B are constants, and k_x is the "effective" wavenumber defined in Equation 12.11, an explicit function of x.

The Bohr-Sommerfeld (or phase integral, Section 3.10.4) condition which the quasi-standing wave must satisfy between x_t and x_m is

$$\int_{x_t}^{x_m} k_x(x, \omega_n) dx = (n + \alpha)\pi \qquad n = 1, 2, 3... \tag{12.23}$$

where $x_m = a$ is the location of the outer boundary in x and the phase factor α is determined by the boundary conditions in x. For a perfect reflector at a ($\xi_x = 0$) and evanescent decay at small x, we find $\alpha = -1/4$. If there is a resonance in the waveguide, it will modify α; however, since the resonance is in the evanescent tail there will be no leading order change in α.

Equation 12.23 is an integral relation for the eigenfrequencies of the fast mode: we specify k_y, k_z, and n (the mode number in x), then find the n-th eigenfrequency ω_n as that frequency for which the criterion is met. Once we compute ω_n, we can then compute the locations of the excited field line x_r from Equation 12.20. These are the most observable discrete modes.

Driven FLR vs Cavity Modes vs Waveguide Modes

The surface wave driven FLR seems reasonable. However, the biggest problem for this hypothesis to be valid is that the mapped phase speeds to the magnetopause were too high compared to in-situ measurements. In addition, in spite of numerous magnetopause crossings by ISEE 1 and 2, regular oscillations of the magnetopause corresponding to the Kelvin-Helmholtz waves were not seen. These two facts indicated that we should seek for a new explanation for the driver.

In the 1980s, cavity modes and waveguide modes were proposed to remove the constraint on driven FLR. The analogy of a magnetosphere to a cavity came before the analogy to a waveguide. In a cavity model the azimuthal direction ϕ or the y direction in the box geometry is finite, and the wavenumber k_y is determined by the given scale size b. In the axisymmetric cylindrical coordinates (r, ϕ, z), where ϕ is the azimuthal angle and z is the coordinate along the field lines, the field has the form $\sim e^{im\phi}$, $m = 0, 1, 2.$[1] This is a weak point of the theory because the magnetosphere is far from axisymmetric and there are no other obvious boundaries to define the cavity. If the system is similar to a waveguide, then k_y can have a continuum of values, and the waveguide allows propagation over a wide band of frequencies.

If instead we use a waveguide model, the azimuthal direction (or y-direction) will have no imposed boundary condition. Theoretically a continuum of wave modes k_y can excite FLR, and the discrete resonance frequencies are determined by the Bohr-Sommerfeld condition eq-cavity_eigen_wkb where harmonics of toroidal and poloidal are given by k_z. The discrete frequency spectrum of the fast waveguide modes is suitable for driving a series of Alfvén resonance; however, it is not obvious that the continuous frequency spectrum of the modes will be able to drive resonances at discrete frequencies. A key observation to make waveguide mode more valid is that *only small k_y contributes to the resonance*. ((Walker et al. 1992),[Wright 1992], (Wright 1994)). If k_y is not fixed (but k_z is fixed), then Equation 12.23 is the dispersion relation for the waveguide expressing ω as a function of k_y. The Alfvén speed is a strong function of the L shell, typically $\sim L^{-3}$. Thus, if the turning point x_t is reasonably deep within

[1] $E(\phi = 0) = E(\phi = 2\pi)$

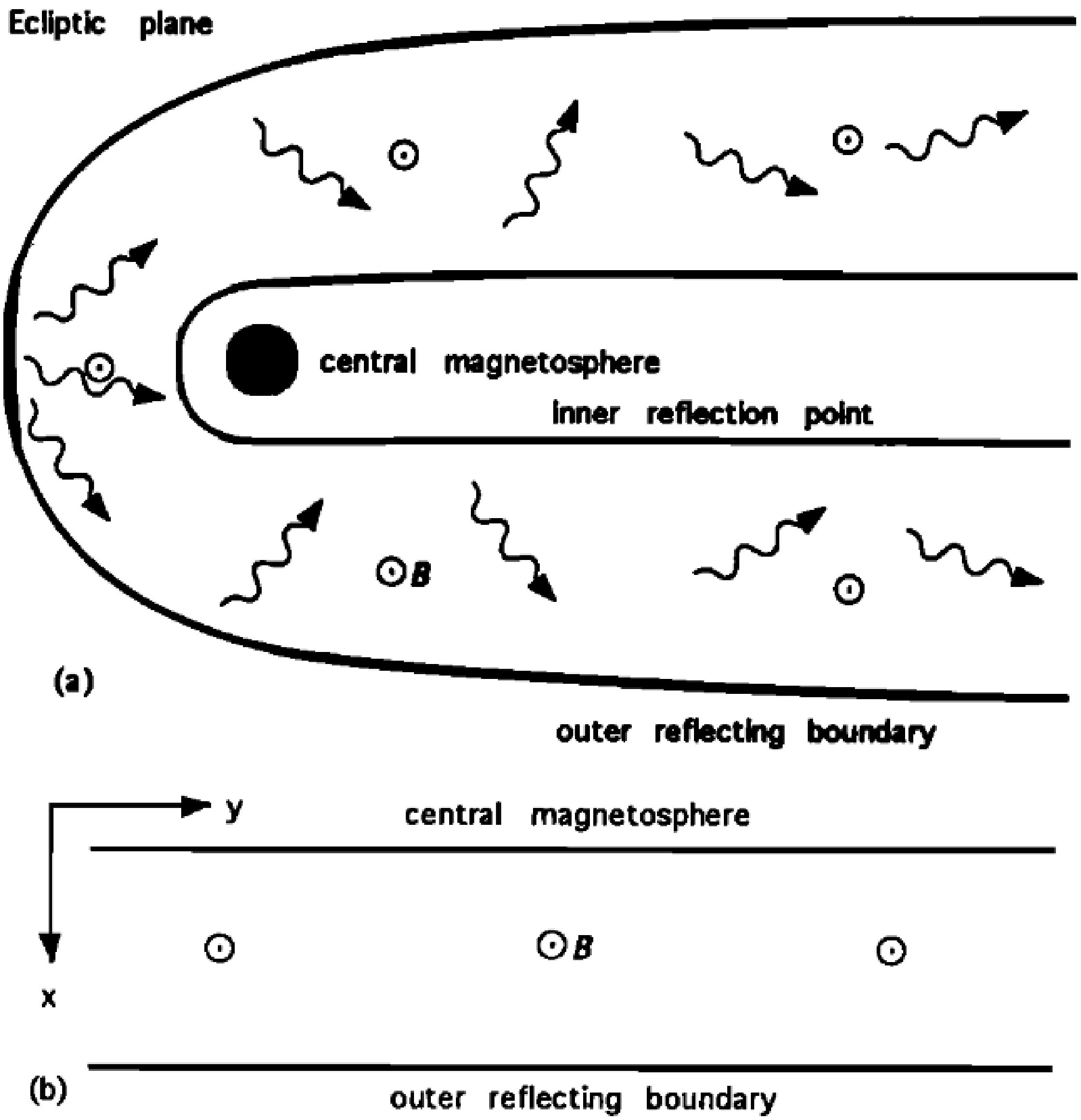

Fig. 1. (*a*) **The magnetospheric waveguide viewed in the ecliptic plane. Fast mode waves (depicted as wiggly arrows) enter the waveguide and subsequently suffer reflection from an outer boundary (perhaps the magnetopause or the bow shock) and an inner reflection point (the plasmapause or a "turning point").** (*b*) **Half of the waveguide (from noon, around the dusk flank and into the magnetotail) is approximated as an infinitely long waveguide containing a uniform magnetic field and plasma whose density depends on** x **and** y**.**

Figure 12.7: From Wright 1992.

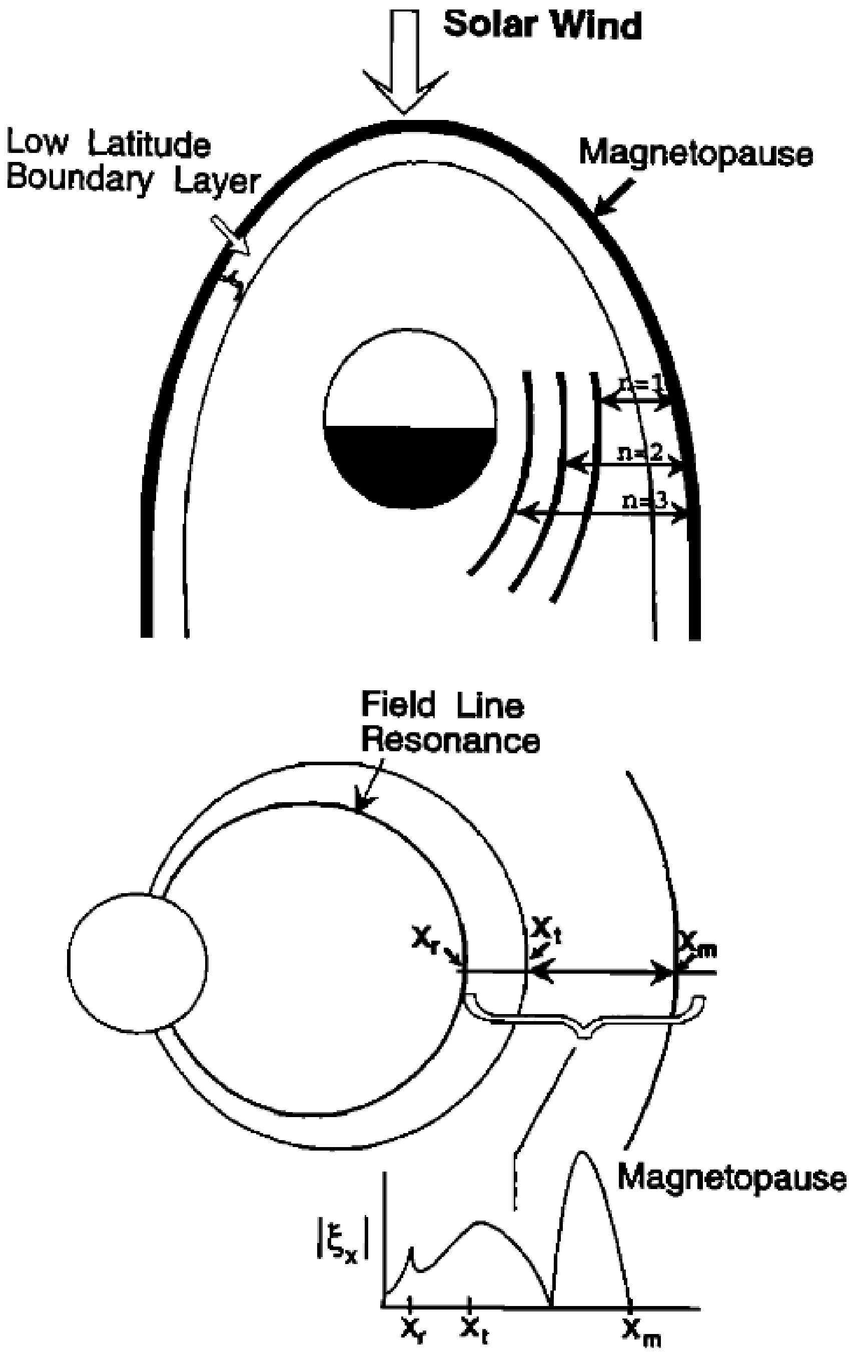

Fig. 3. The waveguide model we have used for our calculations. Top: The turning points in the equatorial plane. The first three eigenmodes are shown. Bottom: A cut in the local dawn meridian. A schematic of the radial displacement for n=2 is given in the insert at the bottom.

Figure 12.8: From (Samson et al. 1992).

the magnetosphere, the range over which $\omega^2/V_A^2 \gg k_y^2 + k_z^2$ contributes most significantly to the integral Equation 12.23. The result is that over a wide range of wavelengths the frequency is very insensitive to k_y. Only when k_y is quite large this is not true, and then the turning point is near the magnetopause, and the wave does not penetrate very deeply into the magnetosphere. Each waveguide mode has a phase velocity ω/k_y and a group velocity $\mathrm{d}\omega/dk_y$. Since Equation 12.23 defines ω as a function of k_y, we can substitute Equation 12.11 into Equation 12.23 and differentiate with respect to k_y (assuming α is a constant):

$$\frac{\partial}{\partial k_y}\int_{x_t}^{x_m}\Big(\frac{\omega^2}{V_A(x)^2} - k_y^2 - k_z^2\Big)^{1/2} dx = 0$$

$$\int_{x_t}^{x_m}\frac{1}{2}k_x^{-1}\Big(\frac{1}{V_A^2}\frac{\partial\omega^2}{\partial k_y} - 2k_y\Big)dx = 0$$

$$\frac{\omega}{k_y}\frac{\partial\omega}{\partial k_y}\int_{x_t}^{x_m} k_x^{-1}V_A^{-2}dx = \int_{x_t}^{x_m} k_x^{-1}dx$$

$$v_p v_g \equiv \frac{\omega}{k_y}\frac{\mathrm{d}\omega}{dk_y} = \frac{\int_{x_t}^{x_m} k_x^{-1}dx}{\int_{x_t}^{x_m} V_A^{-2}k_x^{-1}dx}$$

The RHS is the reciprocal of the weighted mean of V_A^{-2} with weighting function k_x^{-1}. Thus we can write

$$v_p v_g = \left\langle V_A^{-2}\right\rangle^{-1} \tag{12.24}$$

In a uniform waveguide filled with plasma for which the Alfvén speed was V_A the well-known waveguide relation Equation 12.18 would be recovered. In this more complicated case where the Alfvén speed is not constant throughout the waveguide the RHS of Equation 12.24 is an appropriately weight mean of V_A^{-2}. In general the phase velocity v_p is much larger than the mean Alfvén velocity in the guide, and the group velocity v_g is much smaller. This is also consistent with the weak dependence of ω on k_y such that v_g is small. The consequence is that even if the ends of the guide are open, the time taken for energy to be propagated along it is long compared with the period of the oscillations.

Therefore, the waveguide model proposes the following explanation. Disturbances in the solar wind move along the magnetopause, producing a disturbance on the boundary which is propagated tailward with speed v_w. Generally $v_w \gg V_A$. Such a disturbance is likely to excite waveguide modes with phase velocities equal to v_w. Because the frequency of the modes is very insensitive to k_y, the same frequencies (determined by eigenmodes in x and z) are excited no matter what the velocity. The disturbance moving with the wave can feed energy continuously into it. This then requires

$$k_y = \frac{\omega}{v_w}$$

i.e., *the azimuthal wavenumber is determined by the velocity of the source rather than by a boundary condition on y*. Thus we have a discrete value of wavenumber and a discrete fast mode frequency spectrum, without relying on the resonance modes from the cavity model.

Such a disturbance could arise from more than one type of source. If it arose from an impulse in the solar wind, then one would expect the waveguide response to be a ringing at the natural frequencies of its modes with growth on the time scale required to establish the mode and decay on the time scale for leakage to the resonance and loss in the ionosphere or elsewhere. Alternatively the waveguide modes could be excited by the Kelvin-Helmholtz instability. In this case it could be expected that the waveguide modes would all be excited continuously and would be driven so long as the solar wind velocity and the characteristics of the magnetopause were maintained.

12.3 Energy Consideration

???

The energy that is carried into the magnetosphere across the background field by the non-guided fast mode is accumulated in the plane of resonant mode coupling (i.e. the y–z plane through x_r in Figure 12.3) in the form of the guided Alfvén wave. It is this localized accumulation of energy due to resonant mode coupling between a non-guided mode and a guided mode that constitutes a field line resonance. This wave energy accumulation can be described by

$$\frac{\partial w}{\partial t} + \nabla \cdot \mathbf{S} = -h$$

where w is the wave energy density, $\mathbf{S}$ is the Poynting flux, and h a dissipation term, describing energy loss due to e.g. ionospheric Joule heating. As the background parameters only vary in the radial direction, this equation reduces to

$$\frac{\partial w}{\partial t} = -\frac{\mathrm{d}}{\mathrm{d}r} S_x - h$$

Integration across the width of the coupling region in a radial direction leads to the following rate equation:

$$\frac{\partial W}{\partial t} = c_e S_{ng} - S_h^{\mathrm{off}} - H$$

where W is the energy per area that is being accumulated in the coupling, c_e is a coupling efficiency, S_{ng} the incoming Poynting flux of the non-guided mode, and S_g^{off} the "off-angle" component of the Poynting flux of that mode to which the non-guided mode couples. Including this term S_g^{off} allows us to consider energy losses due to coupling to not strictly guided modes. A finite off-angle component of the coupled wave mode would render the energy accumulation less efficient or may even inhibit the build-up of a resonance. Off-angle components may arise if the transverse scale of the coupled wave become small enough for finite ion gyroradius or finite electron inertia becoming important. In this case the coupled mode is a *kinetic Alfvén mode*. The parameter c_e denotes the coupling efficiency, that is the fraction of energy of the non-guided mode that is converted into the guided mode. Finally, H gives the dissipative losses, integrated in the x direction.

It is now instructive to evaluate the rate Equation for the ideal MHD regime. As the Alfvén mode is a strictly guided mode its Poynting flux is directed exactly parallel to the background magnetic field. In other words, $S_h^{\text{off}} = 0$, and Equation reads

$$\frac{\partial W}{\partial t} = c_e S_{\text{fast}}$$

where $S_{\text{fast}} = S_{ng}$ is the Poynting flux of the non-guided MHD mode. Thus, the energy density in the coupling region is continuously increasing as there is no outward transport of energy to balance the incoming Poynting flux of the fast mode. However, ionospheric Joule heating provides a significant dissipation mechanism with H limiting the energy density.

Resonant mode coupling is only a *necessary* condition for field line resonances to occur. A *sufficient* condition is critical coupling to a strictly guided mode and absence of dissipation to a degree that there is enough energy for resonance.

12.4 Non-MHD Effects

In higher frequency cases, the Alfvén resonance condition Equation 12.20 is modified as

$$\omega^2 = k_\parallel^2 V_A^2 (1 - \omega^2/\omega_{ci}^2)^2$$

because of the finite ion Larmor radius effects [Stix, 1992].

The Alfvén resonance singularity can be removed by including non-MHD effects such as electron inertia or ion Larmor radius corrections. (I don't quite follow it, but check the papers by Lin, Hasegawa, Chen, Johnson, and Cheng.) There are very few studies considering the electron inertia correction, because it is related to r_{Ls}/λ_e, where $\lambda_e = c/\omega_{pe}$ is the plasma skin depth or electron inertial length, and $r_{Ls} = \sqrt{T_e/m_i}/\omega_{ci}$. It is typically not important except at locations where $k_\parallel \to 0$ such as in a reconnection geometry.

Thus what's usually been considered to remove the singularity is the ion kinetic effects, where we end up coupling fast waves with kinetic Alfvén waves of the dispersion relation Equation 7.44.

12.5 3D FLR

13 Nonlinear Effects

In many experiments waves are no longer describable by the linear theory by the time they are observed. Consider, for instance, the case of drift waves. Because they are unstable, drift waves would, according to linear theory, increase their amplitude exponentially. This period of growth is not normally observed — since one usually does not know when to start looking — but instead one observes the waves only after they have grown to a large, steady amplitude. The fact that the waves are no longer growing means that the linear thoery is no longer valid, and some *nonlinear* effect is limiting the amplitude. Theoretical explanation of this elementary observation has proved to be a surprisingly difficult problem, since the observed amplitude at saturation is rather small.

A wave can undergo a number of changes when its amplitude gets large. It can change its shape — say, from a sine wave to a lopsided triangular waveform. This is the same as saying that Fourier components at other frequencies (or wave numbers) are generated. Ultimately, the wave can "break", like ocean waves on a beach, converting the wave energy into thermal energy of the particles. A large wave can trap particles in its potential troughs, thus changing the properties of the medium in which it propagates (e.g. nonlinear Landau damping). If a plasma is so strongly excited that a continuous spectrum of frequencies is present, it is in a state of *turbulence*. This state must be described statistically, as in the case of hydrodynamics. An important consequence of plasma turbulence is *anomalous resistivity*, in which electrons are slowed down by collisions with random electric field fluctuations, rather than with ions. This effect is used for ohmic heating of a plasma to temperature so high that ordinary resistivity is insufficient.

Nonlinear phenomena can be grouped into three broad categories:

1. *Basically nonlinearizable problems.*

- Diffusion in a fully ionized gas, for instance, is intrinsically a nonlinear problem because the diffusion coefficient varies with density.
- Problems of hydromagnetic equilibrium are nonlinear.
- Plasma sheath.

2. *Wave-particle interactions.*

- Particle trapping can lead to nonlinear damping (Section 8.6).

- A classic example is the quasilinear effect, in which the equilibrium of the plasma is changed by the waves. Consider the case of a plasma with an electron beam (Fig.8-1). Since the distribution function has a region where df_0/dv is positive, the system has inverse Landau damping, and plasma oscillations with v_ϕ in the positive-slope region are unstable (Equation 8.32). The resonant electrons are the first to be affected by wave-particle interactions, and their distribution function will be changed by the wave electric field. The waves are stabilized when $f_e(v)$ is flattened by the waves, as shown by the dashed line in Fig.8-1 **?@fig-unstable_f**, so tha the new equilibrium distribution no longer has a positive slope.
- Plasma wave echoes (Section 13.5).

3. *Wave-wave interactions.*

- Waves can interact with each other even in the fluid description, in which individual particle effects are neglected. A single wave can decay by first generating harmonics of its fundamental frequency. These harmonics can then interact with each other and with the primary wave to form other waves at the beat frequencies. The beat waves in turn can grow so large that they can interact and form many more beat frequencies, until the spectrum becomes continuous. It is interesting to discuss the direction of energy flow in a turbulent spectrum. In fluid dynamics, long-wavelength modes decay into short-wavelength modes, because the large eddies contain more energy and can decay only by splitting into small eddies, which are each less energetic. The smallest eddies then convert their kinetic motion into heat by viscous damping. In a plasma, usually the opposite occurs. Short wave-length modes tend to coalesce into long-wavelength modes, which are less energetic. This is because the electric field energy $E^2/8\pi$ is of order $k^2\phi^2/8\pi$, so that if $e\phi$ is fixed (usually by k_BT_e), the small-k, long-λ modes have less energy. As a consequence, energy will be transferred to small k by instabilities at large k, and some mechanism must be found to dissipate the energy. No such problem exists at large k, where Landau damping can occur. For motions along $\mathbf{B}_0$, nonlinear "modulational" instabilities could cause the energy at small k to be coupled to ions and to heat them. For motions perpendicular to $\mathbf{B}_0$, the largest eddies will have wavelengths of the order of the plasma radius and could cause plasma loss to the walls by convection.

Although problems still remain to be solved in the linear theory of waves and instabilities, the mainstream of plasma research has turned to the much less well understood area of nonlinear phenomena. The examples in the following sections will give an idea of some of the effects that have been studied in theory and in experiment.

13.1 The Necessity for Sheaths

In all practical plasma devices, the plasma is contained in a vacuum chamber of finite size. What happens to the plasma at the wall? For simplicity, let us confine our attention to a one-dimensional model with no magnetic field **?@fig-sheath**. Suppose there is no appreciable electric field inside the plasma; we can then let the potential ϕ be zero there. When ions and electrons hit the wall, they recombine and are lost. Since electrons have much higher thermal velocities than ions, they are lost faster and leave the plasma with a net positive charge. The plasma must then have potential positive with respect to the wall; i.e., the wall potential ϕ_w is negative. This potential cannot be distributed over the entire plasma, since Debye shielding will confine the potential variation to a layer of the order of several Debye lengths in thickness. This layer, which must exist on all cold walls with which the plasma is in contact, is called a sheath. The function of a sheath is to form a potential barrier so that the more mobile species, usually electrons, is confined. electrostatically. The height of the barrier adjusts itself so that the flux of electrons that wave enough energy to go over the barrier to the wall is just equal to the flux of ions reaching the wall.

13.1.1 The Planar Sheath Equation

Recall that the Debye length is derived from linearing Poisson's equation. To examine the exact behavior of $\phi(x)$ in the sheath, we must treat the nonlinear problem: we shall find that there is not always a solution. **?@fig-sheath_potential** shows the situation near one of the walls. At the plane $x=0$, ions are imagined to enter the sheath region from the main plasma with a drift velocity u_0. This drift is needed to accout for the loss of ions to the wall from the region in which they were created by ionization. For simplicity, we assume $T_i=0$, so that all ions have the velocity u_0 at $x=0$. We consider the steady state problem in a collisionless sheath region. The potential ϕ is assumed to decrease monotonically with x. Actually, ϕ could have spatial oscillations, and then there would be trapped particles in the steady state. This does not happend in practice because dissipative processes tend to destroy any such highly organized state.

If $u(x)$ is the ion velocity, conservation of energy requires

$$
\begin{aligned}
\frac{1}{2}mu^2 &= \frac{1}{2}mu_0^2 - e\phi(x) \\
u &= \Big(u_0^2 - \frac{2e\phi}{m_i}\Big)^{1/2}
\end{aligned}
\tag{13.1}
$$

The ion equation of continuity then gives the ion density n_i in terms of the density n_0 in the main plasma:

$$
\begin{aligned}
n_0 u_0 &= n_i(x) u(x) \\
n_i(x) &= n_0\Big(1 - \frac{2e\phi}{m_i u_0^2}\Big)^{1/2}
\end{aligned}
\tag{13.2}
$$

In steady state, the electrons will follow the Boltzmann relation closely,

$$n_e(x) = n_0 \exp(e\phi/k_BT_e)$$

Poisson's equation is then

$$\epsilon_0 \frac{\mathrm{d}^2\phi}{\mathrm{d}x^2} = e(n_e - n_i) = en_0\Big[\exp\Big(\frac{e\phi}{k_BT_e}\Big) - \Big(1 - \frac{2e\phi}{m_iu_0^2}\Big)^{1/2}\Big] \tag{13.3}$$

The structure of this equation can be seen more clearly if we simplify it with the following changes in notation:

$$\chi \equiv -\frac{e\phi}{k_BT_e} \quad \xi \equiv \frac{x}{\lambda_D} = x\Big(\frac{n_0e^2}{\epsilon_0k_BT_e}\Big)^{1/2} \quad \mu \equiv \frac{u_0}{(k_BT_e/m_i)^{1/2}}$$

Then it can be written as

$$\chi'' = \Big(1 + \frac{2\chi}{\mu^2}\Big)^{-1/2} - e^{-\chi} \tag{13.4}$$

where the prime denotes $\mathrm{d}/\mathrm{d}\xi$. This is the nonlinear equation of a plane sheath, and it has an acceptable solution only if μ is large enough. The reason for the symbol μ will become apparent in the following section on shock waves.

13.1.2 The Bohm Sheath Criterion

Equation 13.4 can be integrated once by multiplying both sides by χ':

$$\int_0^\xi \chi'\chi''\mathrm{d}\xi_1 = \int_0^\xi \Big(1 + \frac{2\chi}{\mu^2}\Big)^{-1/2}\chi'\mathrm{d}\xi_1 - \int_0^\xi e^{-\chi}\chi'\mathrm{d}\xi_1$$

where ξ_1 is a dummy variable. Since $\chi = 0$ at $\xi = 0$, the integration easily yield

$$\frac{1}{2}({\chi'}^2 - {\chi_0'}^2) = \mu^2\Big[(1 + \frac{2\chi}{\mu^2})^{1/2} - 1\Big] + e^{-\chi} - 1 \tag{13.5}$$

If $\mathbf{E} = 0$ in the plasma, we must set $\chi_0' = 0$ at $\xi = 0$. A second integration to find χ would have to be done numerically; but whatever the answer is, the right-hand side of Equation 13.5 must be positive for all χ. In particular, for $\chi \ll 1$, we can expand the right-hand terms in Taylor series:

$$
\begin{aligned}
\mu^2\Big[1+\frac{\chi}{\mu^2}-\frac{1}{2}\frac{\chi^2}{\mu^4}+...-1\Big]+1-\chi+\frac{1}{2}\chi^2+...-1>0 \\
\frac{1}{2}\chi^2\Big(-\frac{1}{\mu^2}+1\Big)>0 \\
\mu^2>1 \quad \text{or} \quad u_0>(k_BT_e/m_i)^{1/2}
\end{aligned}
\tag{13.6}
$$

This inequality is know as the *Bohm sheath criterion*. It says that ions must enter the sheath with a velocity greater than the acoustic velocity v_s. To give the ions this directed velocity u_0, there must be a finite electric field in the plasma. Our assumption that $\chi'=0$ at $\xi=0$ is therefore only an approximate one, made possible by the fact that the scale of the sheath region is usually much smaller than the scale of the main plasma region in which the ions are accelerated. The value of u_0 is somewhat arbitrary, depending on where we choose to put the boundary $x=0$ between the plasma and the sheath. Of course, the ion flux n_0u_0 is fixed by the ion production rate, so if u_0 is varies, the value of n_0 at $x=0$ will vary inversely with u_0. If the ions have finite temperature, the critical drift velocity u_0 will be somewhat lower.

The physical reason for the Bohm criterion is easily seen from a plot of the ion and electron densities vs. χ **?@fig-sheath_n_variation**. The electron density n_e falls exponentially with χ, according to the Boltzmann relation. The ion density also falls, since the ions are accelerated by the sheath potential. If the ions start with a large energy, $n_i(\chi)$ falls slowly, since the sheath field causes a relatively minor change in the ions' velocity. If the ions start with a small energy, $n_i(\chi)$ falls fast, and can go belwo the n_e curve. In that case, n_e-n_i is positive near $\chi=0$; and Equation 13.4 tells us that $\phi(x)$ must curve upward, in contradiction to the requirement that the sheath must repel electrons. In order for this not to happen, the slope of $n_i(\chi)$ at $\chi=0$ must be smaller (in absolute value) than that of $n_e(\chi)$; this condition is identical with the condition $\mu^2>1$.

13.1.3 The Child-Langmuir Law

Since $n_e(\chi)$ falls exponentially with χ, the electron density can be neglected in the region of large χ next to the wall (or any negative electrode). Poisson's equation is then approximately

$$
\chi''\approx\Big(1+\frac{2\chi}{\mu^2}\Big)^{-1/2}\approx\frac{\mu}{(2\chi)^{1/2}}
$$

Multiplying by χ' and integrating from $\xi_1=\xi_s$ to $\xi_1=\xi$, we have

$$
\frac{1}{2}({\chi'}^2-{\chi_s'}^2)=\sqrt{2}\mu(\chi^{1/2}-\chi_s^{1/2})
$$

where ξ_s is the place where we started neglecting n_e. We can redefine the zero of χ so that $\chi_s=1$ at $\xi=\xi_s$. We shall also neglect χ_s', since the slope of the potential curve can be expected to be much steeper in the $n_e = 0$ region than in the finite-n_e region. Then the above equation becomes

$$\begin{aligned}
{\chi'}^2 &= 2^{3/2}\mu\chi^{1/2} \\
\chi' &= 2^{3/4}\mu^{1/2}\chi^{1/4} \\
\mathrm{d}\chi/\chi^{1/4} &= 2^{3/4}\mu^{1/2}\mathrm{d}\xi
\end{aligned}$$

Integrating from $\xi = \xi_s$ to $\xi = \xi_s + \mathrm{d}/\lambda_D = \xi_{\mathrm{wall}}$, we have

$$\frac{4}{3}\chi_w^{3/4} = 2^{3/4}\mu^{1/2}\mathrm{d}/\lambda_D$$

or

$$\mu = \frac{4\sqrt{2}}{9}\frac{\chi_w^{3/2}}{\mathrm{d}^2}\lambda_D^2$$

Changing back to the variables u_0 and ϕ, and noting that the ion current into the wall is $J = en_0u_0$, we then find

$$J = \frac{4}{9}\left(\frac{2e}{m_i}\right)^{1/2}\frac{\epsilon_0|\phi_w|^{3/2}}{\mathrm{d}^2} \tag{13.7}$$

This is just the well-known Child-Langmuir law of space-charge-limited current in a plane diode.

The potential variation in a plasma-wall system can be divided into three parts. Nearest the wall is an electron-free region whose thickness d given by Equation 13.7. Here J is determined by the ion production rate, and ϕ_w is determined by the equility of electron and ion fluxes. Next comes a region in which n_e is appreciable; as shown in the linear theory from which the scale of Debye length is derived. Finally, there is a region with much larger scale length, the “presheath”, in which the ions are accelerated to the required velocity u_0 by a potential drop $|\phi| \geq \frac{1}{2}k_BT_E/e$. Depending on the experiment, the scale of the presheath may be set by the plasma radius, the collisional mean free path, or the ionization mechanism. The potential distribution, of course, varies smoothly; the division into three regions is made only for convenience and is made possible by the disparity in scale lengths. In the early days of gas discharges, sheaths could be observed as dark layers where no electrons were present to excite atoms to emission. Subsequently, the potential variation has been measured by the electrostatic deflection of a thin electron beam shot parallel to a wall.

13.1.4 Electrostatic Probes

The sheath criterion Equation 13.6 can be used to estimate the flux of ions to a negatively biased probe in a plasma. If the probe has a surface area A, and if the ions entering the sheath have a drift velocity $u_0 \geq (k_B T_e/m_i)^{1/2}$, then the ion current collected is

$$I = n_s e A (k_B T_e/m_i)^{1/2}$$

The electron current can be neglected if the probe is sufficiently negative (several times $k_B T_e$) relative to the plasma to repel all but the tail of the Maxwellian electron distribution. The density n_s is the plasma density at the edge of the sheath. Let us define the sheath edge to be the place where u_0 is exactly $(k_B T_e/m_i)^{1/2}$. To accelerate ions to this velocity requires a presheath potential $|\phi| \geq \frac{1}{2} k_B T_e/e$, so that the sheath edge has a potential

$$\phi_s \simeq -\frac{1}{2} k_B T_e/e$$

relative to the body of the plasma. If the electrons are Maxwellian, this determines n_s:

$$n_s = n_0 e^{e\phi_s/k_B T_e} = n_0 e^{-1/2} \approx 0.61 n_0$$

For our purposes it is accurate enough to replace 0.61 with a round number like 1/2; thus, the "saturation ion current" to a negative probe is approximately

$$I_B \simeq \frac{1}{2} n_0 e A (k_B T_e/m_i)^{1/2}$$

I_B, sometimes called the "Bohm current", give the plasma density easily, once the temperature is known.

If the Debye length λ_D, and hence the sheath thickness, is very small compared to the probe dimensions, the area of the sheath edge is effectively the same as the area A of the probe surface, regardless of its shape. At low densities, however, λ_D can become large, so that some ions entering the sheath can orbit the probe and miss it. Calculations of orbits for various probe shapes were first made by I. Langmuir and L. Tonks — hence the name "**Langmuir probe**" ascribed to this method of measurement. Though tedious, these calculations can give accurate determinations of plasma density because an arbitrary definition of sheath edge does not have to be made. By varying the probe voltage, the Maxwellian electron distribution is sampled, and the current-voltage curve of a Langmuir probe can also yield the electron temperature. The electrostatic probe was the first plasma diagnostic and is still the simplest and the most localized measurement device. Unfortunately, material electrodes can be inserted only in low-density, cool plasma.

More detailed explanation of how to measure the plasma density using a Langmuir probe is given here.

13.2 Ion Acoustic Shock Waves

When a jet travels faster than sound, it creates a shock wave. This is a basically nonlinear phenomenon, since there is no period when the wave is small and growing. The jet is faster than the speed of waves in air, so the undisturbed medium cannot be "warned" by precursor signals before the large shock wave hits it. In hydrodynamic shock waves, collisions are dominant. Shock waves also exist in plasmas, even when there are no collisions. A magnetic shock, the "bow shock", is generated by the earth as it plows through the interplanetary plasma while dragging along a dipole magnetic field. We shall discuss a simpler example: a collisionless, one-dimensional shock wave which develops from a large-amplitude ion wave.

(Normal space plasma textbooks discuss this part starting from R-H relations. I don't like that.)

13.2.1 The Sagdeev Potential

?@fig-shock_potential shows the idealized potential profile of an ion acoustic shock wave. The reason for this shape will be given presently. The wave is traveling to the left with a velocity u_0. If we go to the frame moving with the wave, the function $\phi(x)$ will be constant in time, and we will see a stream of plasma impinging on the wave from the left with a velocity u_0. For simplicity, let T_i be zero, so that all the ions are incident with the same velocity u_0, and let hte electrons be Maxwellian. Since the shock moves much more slowly than the electron thermal speed, the shift in the center velocity of the Maxwellian can be neglected. The velocity of the ions in the shock wave is, from energy conservation,

$$u = \Big(u_0^2 - \frac{2e\phi}{m_i}\Big)^{1/2}$$

If n_0 is the density of the undisturbed plasma, the ion density in the shock is

$$n_i = \frac{n_0 u_0}{u} = n_0\Big(1 - \frac{2e\phi}{m_i u_0^2}\Big)^{-1/2}$$

The electron density is given by the Boltzmann relation. Poisson's equation then gives

$$\epsilon_0 \frac{\mathrm{d}^2\phi}{\mathrm{d}x^2} = e(n_e - n_i) = en_0\Big[\exp\Big(\frac{e\phi}{k_B T_e}\Big) - \Big(1 - \frac{2e\phi}{m_i u_0^2}\Big)^{1/2}\Big]$$

This is, of course, the same Equation 13.3 as we had for a sheath. A shock wave is no more than a sheath moving through a plasma. We now introduce the dimensionless variables

$$\chi \equiv \frac{e\phi}{k_B T_e} \quad \xi \equiv \frac{x}{\lambda_D} = x\Big(\frac{n_0 e^2}{\epsilon_0 k_B T_e}\Big)^{1/2} \quad \mu \equiv \frac{u_0}{(k_B T_e/m_i)^{1/2}}$$

Note that we have changed the sign in the definition of χ so as to keep χ positive in this problem. The quantity μ is called the *Mach number* of the shock. Poisson's equation can now be written

$$\chi'' = e^{\chi} - \Big(1 - \frac{2\chi}{\mu^2}\Big)^{-1/2} = -\frac{dV(\chi)}{\mathrm{d}\chi} \tag{13.8}$$

which differs from the sheath Equation 13.4 only because of the change in sign of χ.

The behavior of the solution of Equation 13.8 was made clear by R. Z. Sagdeev, who used an analogy to an oscillator in a potential well. The displacement x of an oscillator subjected to a force $-mdV(x)/\mathrm{d}x$ is given by

$$\ddot{x} = -dV/\mathrm{d}x$$

If the right-hand side of Equation 13.8 is defined as $-dV/\mathrm{d}x$, the equation is the same as that of an oscillator, with the potential χ playing the role of χ, and $\mathrm{d}/\mathrm{d}\xi$ replacing $\mathrm{d}/\mathrm{d}t$. The quasipotential $V(\chi)$ is sometimes called the Sagdeev potential. The function $V(x)$ can be found from Equation 13.8 by integration with the boundary condition $V(\chi) = 0$ at $\chi = 0$:

$$V(\chi) = 1 - e^{\chi} + \mu^2\Big[1 - \Big(1 - \frac{2\chi}{\mu^2}\Big)^{1/2}\Big] \tag{13.9}$$

For μ lying in a certain range, this function has the shape shown in **?@fig-shock_quasipotential**. If this were a real well, a particle entering from the left will go to the right-hand side of the well ($x > 0$), reflect, and return to $x = 0$, making a single transit. Similarly, a quasiparticle in our analogy will make a single excursion to positive χ and return to $\chi = 0$, as shown in **?@fig-shock_soliton**. Such a pulse is called a *soliton*: it is a potential and density disturbance propagating to the left in **?@fig-shock_soliton** with velocity u_0. (I DON'T FULLY UNDERSTAND!!!)

Now, if a particle suffers a loss of energy while in the well, it will never return to $x = 0$ but will oscillate (in time) about some positive value of x. Similarly, a little dissipation will make the potential of a shock wave oscillate (in space) about some positive value of ϕ. This is exactly the behavoir depicted in **?@fig-shock_potential**. Actually, dissipation is not needed for this: reflection of ions from the shock front has the same effect. To understand this, imagine that the ions have a small thermal spread in energy and that the height $e\phi$ of the wave front is just large enough to reflect some of the ions back to the left, while the rest go over the potential hill to the right. The reflected ions cause an increase in ion density in the upstream region to the left of the shock front (**?@fig-shock_potential**). This means that the quantity

$$\chi' = \frac{1}{n_0}\int_0^{\xi}(n_e - n_i)\mathrm{d}\xi_1$$

is decreased. Since χ' is the analog of $\mathrm{d}x/\mathrm{d}t$ in the oscillator problem, our virtual oscillator has lost velocity and is trapped in the potential well of **?@fig-shock_quasipotential**.

13.2.2 The Critical Mach Numbers

Solutions of either the soliton type or the wave-train type exist only for a range of μ. A lower limit for μ is given by the condition that $V(\chi)$ be a potential well, rather than a hill. Expanding Equation 13.9 for $\chi \ll 1$ yields

$$\mu^2 > 1 \quad u_0 > (k_B T_e/m_i)^{1/2}$$

This is exactly the same, both physically and mathematically, as the Bohm criterion for the existence of a sheath (Equation 13.6).

An upper limit to μ is imposed by the condition that the function $V(\chi)$ of **?@fig-shock_quasipotential** must cross the χ axis for $\chi > 0$; otherwise, the virtual particle will not be reflected, and the potential will rise indefinitely. From Equation 13.9, we require

$$e^{\chi} - 1 < \mu^2\left[1 - \left(1 - \frac{2\chi}{\mu^2}\right)^{1/2}\right]$$

for some $\chi > 0$. If the lower critical Mach number is surpassed ($\mu > 1$), the left-and side, representing the integral of the electron density from zero to χ, is initially larger than the right-hand side, representing the integral of the ion density. As χ increases, the right-hand side can catch up with the left-hand side if μ^2 is not too large. However, because of the square root, the largest value χ can have is $\mu^2/2$. This is because $e\phi$ cannot exceed $\frac{1}{2}m_i u_0^2$; otherwise, ions would be excluded from the plasma in the downstream region. Inserting the largest value of χ into the above equation, we have

$$\exp(\mu^2/2) - 1 < \mu \quad \text{or} \quad \mu < 1.6$$

This is the upper critical Mach number. Shock waves in a cold-ion plasma therefore exist only for $1 < \mu < 1.6$.

As in the case of sheaths, the physical situation is best explained by a diagram of n_i and n_e vs. χ (**?@fig-shock_n**). This diagram differs from **?@fig-sheath_n_variation** because of the change of sign of ϕ. Since the ions are now decelerated rather than accelerated, n_i will appraoch infinity at $\chi = \mu^2/2$. The lowe critical Mach number ensures that the n_i curve lies below the n_e curve at small χ, so that the potential $\phi(x)$ starts off with the right sign for its curvature. When the curve n_{i1} crosses the n_e curve, the soliton (**?@fig-shock_soliton**) has an inflection point. Finally, when χ is large enough that the areas under the n_i and n_e curves are equal, the soliton reaches a peak, and the n_{i1} and n_e curves are retraced as χ goes back to zero. The equality of the areas ensures that the net charge in the soliton is zero; therefore, there is no electric field outside. If μ is larger than 1.6, we have the curve n_{i2}, in which the area under the curve is too small even when χ has reached its maximum value of $\mu^2/2$.

13.2.3 Wave Steepening

If one propagates an ion wave in a cold-ion plasma, it will have the phase velocity given by

$$v_p = \Big(\frac{k_B T_e + \gamma_i k_B T_i}{m_i}\Big)^{1/2} = \Big(\frac{k_B T_e}{m_i}\Big)^{1/2}$$

corresponding to $\mu = 1$. How, then, can one create shocks with $\mu > 1$? One must remember that the above phase velocity was a linear result valid only at small amplitudes. As the amplitude is increased, an ion wave speeds up and also changes from a sine wave to a sawtooth shape with a steep leading edge (**?@fig-wave_steepening**). The reason is that the wave electric field has accelerated the ions. In **?@fig-wave_steepening**, ions at the peak of the potential distribution have a large velocity in the direction of v_ϕ than those at the trough, since they have just experienced a period of acceleration as the wave passed by. In linear theory, this difference in velocity is taken into account, but not the displacement resulting from it. In nonlinear theory, it is easy to see that the ions at the peak are shifted to the right, while those at the trough are shifted to the left, thus steepening the wave shape. Since the density perturbation is in pahse with the potential, more ions are accelerated to the right than to the left, and the wave causes a net mass flow in the direction of propagation. This causes the wave velocity to exceed the acoustic speed in the undisturbed plasma, so that μ is larger than unity.

13.2.4 Double Layers

A phenomenon related to sheaths and ion acoustic shocks is that of the double layer. This is a localized potential jump, believed to occur naturally in the ionosphere, which neither propagates nor is attached to a boundary. The name comes from the successive layers of net positive and net negative charge that are necessary to create a step in $\phi(x)$. Such a step can remain stationary in space only if there is a plasma flow that Doppler shifts a shock front down to zero velocity in the lab frame, or if the distribution functions of the transmitted and reflected electrons and ions on each side of the discontinuity are specially tailored so as to make this possible. In natural situations double layers are likely to arise where there are gradients in the magnetic field $\mathbf{B}$, not where $\mathbf{B}$ is zero or uniform, as in laboratory simulations. In that case, the $\mu\nabla B$ force can play a large role in localizing a double layer away from all boundaries. Indeed, the thermal barrier in tandem mirror reactors is an example of a double layer with strong magnetic trapping.

13.3 The Ponderomotive Force

Light waves exert a radiation pressure which is usually very weak and hard to detect. Even the esoteric example of comet tails, formed by the pressure of sunlight, is tainted by the added effect of particles streaming from the sun. When high-powered microwaves or laser beams are used to heat or confine plasmas, however, the radiation pressure can reach several hundred thousand atmospheres! When applied to a plasma, this force is coupled to the particles in a somewhat subtle way and is called the *ponderomotive force*. Many nonlinear phenomena have a simple explanation in terms of the ponderomotive force.

The easiest way to derive this nonlinear force is to consider the motion of an electron in the oscillating $\mathbf{E}$ and $\mathbf{B}$ fields of a wave. We neglect dc $\mathbf{E}_0$ and $\mathbf{B}_0$ fields. The electron equation of motion is

$$m\dot{\mathbf{v}} = -e[\mathbf{E}(\mathbf{r}) + \mathbf{v} \times \mathbf{B}(\mathbf{r})]$$

This equation is exact if $\mathbf{E}$ and $\mathbf{B}$ are evaluated at the instantaneous position of the electron. The nonlinearity comes partly from the $\mathbf{v} \times \mathbf{B}$, which is second order because both $\mathbf{v}$ and $\mathbf{B}$ vanish in the equilibrium, so that the term is no larger than $\mathbf{v}_1 \times \mathbf{B}_1$, where $\mathbf{v}_1$ and $\mathbf{B}_1$ are the linear-theory values. The other part of the nonlinearity, as we shall see, comes from evaluating $\mathbf{E}$ at the actual position of the particle rather than its initial position. Assume a wave electric field of the form

$$\mathbf{E} = \mathbf{E}_s(\mathbf{r}) \cos \omega t$$

where $\mathbf{E}_s(\mathbf{r})$ contains the spatial dependence. In first order, we may neglect the $\mathbf{v} \times \mathbf{B}$ term in the equation of motion and evaluate $\mathbf{E}$ at the initial position $\mathbf{r}_0$. We have

$$\begin{aligned} m\dot{\mathbf{v}}_1 &= -e\mathbf{E}(\mathbf{r}_0) \\ \mathbf{v}_1 &= -(e/m\omega)\mathbf{E}_s \sin \omega t = \dot{\mathbf{r}}_1 \\ \delta\mathbf{r}_1 &= (e/m\omega^2)\mathbf{E}_s \cos \omega t \end{aligned}$$

It is important to note that *in a nonlinear calculation, we cannot write* $e^{i\omega t}$ *and takes its real part later.* Instead, we write its real part explicitly as $\cos\omega t$. This is because products of oscillating factors appera in nonlinear theory, and the operations of multiplying and taking the real part do not commute.

Going to second order, we expand $\mathbf{E}(\mathbf{r})$ about $\mathbf{r}_0$:

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}(\mathbf{r}_0) + (\delta\mathbf{r}_0 \cdot \nabla)\mathbf{E}|_{r=r_0} + ...$$

We must now add the term $\mathbf{v}_1 \times \mathbf{B}_1$, where $\mathbf{B}_1$ is given by Maxwell's equation:

$$\begin{aligned} \nabla \times \mathbf{E} &= -\dot{\mathbf{B}} \\ \mathbf{B}_1 &= -(1/\omega)\nabla \times \mathbf{E}_s|_{r=r_0} \sin\omega t \end{aligned}$$

The second-order part of the equation of motion is then

$$m\dot{\mathbf{v}}_2 = -e[(\delta\mathbf{r}_1 \cdot \nabla)\mathbf{E} + \mathbf{v}_1 \times \mathbf{B}_1]$$

Inserting the expressions of $\mathbf{v}_1, \delta\mathbf{r}_1$ and $\mathbf{B}_1$ into the above and averaging over time, we have

$$m\left\langle \frac{\mathrm{d}\mathbf{v}_2}{\mathrm{d}t} \right\rangle = -\frac{e^2}{m\omega^2}\frac{1}{2}[(\mathbf{E}_s \cdot \nabla)\mathbf{E}_s + \mathbf{E}_s \times (\nabla \times \mathbf{E}_s)] = \mathbf{f}_{NL} \tag{13.10}$$

Here we have used $\left\langle \sin^2\omega t \right\rangle = \left\langle \cos^2\omega t \right\rangle = \frac{1}{2}$. The double cross product can be written as the sum of two terms, one of which cancels the $(\mathbf{E}_s \cdot \nabla)\mathbf{E}_s$ term.

What remains is

$$\mathbf{f}_{NL} = -\frac{1}{4}\frac{e^2}{m\omega^2}\nabla E_s^2$$

This is the effective nonlinear force on a single electron. The force per m^3 is $\mathbf{f}_{NL}$ times the electron density n_0, which can be written in terms of ω_p^2. Since $E_s^2 = 2\left\langle E^2 \right\rangle$, we finally have for the ponderomotive force the formula

$$\mathbf{F}_{NL} = -\frac{\omega_p^2}{\omega^2}\nabla\frac{\left\langle \epsilon_0 E^2 \right\rangle}{2} \tag{13.11}$$

If the wave is electromagnetic, the second term in Equation 13.10 is dominant, and the physical mechanism for $\mathbf{F}_{NL}$ is as follows. Electrons oscillate in the direction of $\mathbf{E}$, but the wave magnetic field distorts their orbits. That is, the Lorentz force $-e\mathbf{v} \times \mathbf{B}$ pushes the electrons in the direction of $\mathbf{k}$ (since $\mathbf{v}$ is in the direction of $\mathbf{E}$, and $\mathbf{E} \times \mathbf{B}$ is in the direction of $\mathbf{k}$).

The phases of $\mathbf{v}$ and $\mathbf{B}$ are such that the motion does not average to zero over an oscillation, but there is a secular drift along $\mathbf{k}$. If the wave has uniform amplitude, no force is needed to maintain this drift; but if the wave amplitude varies, the electrons will pile up in regions of small amplitude, and a force is needed to overcome the space charge. This is why the effective force $\mathbf{F}_{NL}$ is proportional to the *gradient* of $\langle E^2 \rangle$. Since the drift for each electron is the same, $\mathbf{F}_{NL}$ is proportional to the density — hence the factor ω_p^2/ω^2.

If the wave is electrostatic, the first term in Equation 13.10 is dominant. Then the physical mechanism is simply that an electron oscillating along $\mathbf{k} \parallel \mathbf{E}$ moves farther in the half-cycle when it is moving from a strong-field region to a weak-field region than vice versa, so there is a net drift.

Although $\mathbf{F}_{NL}$ acts mainly on the electrons, the force is ultimately transmitted to the ions, since it is a low-frequency or dc effect. When electrons are bunched by $\mathbf{F}_{NL}$, a charge-separation field $\mathbf{E}_{cs}$ is created. The total force felt by the electrons is

$$\mathbf{F}_e = -e\mathbf{E}_{cs} + \mathbf{F}_{NL}$$

Since the ponderomotive force on the ions is smaller by $\Omega_p^2/\omega_p^2 = m_e/m_i$, the force on the ion fluid is approximately

$$\mathbf{F}_i = e\mathbf{E}_{cs}$$

Summing the last two equations, we find that the force on the plasma is $\mathbf{F}_{NL}$.

A direct effect of $\mathbf{F}_{NL}$ is the self-focusing of laser light in a plasma. In **?@fig-laser_self-focusing** we see that a laser beam of finite diameter causes a radially directed ponderomotive force in a plasma. This force moves plasma out of the beam, so that ω_p is lower and hte dielectric constant ϵ is higher inside the beam than outside. The plasma then acts as a convex lens, focusing the beam to a smaller diameter.

13.4 Parametric Instabilities

The most thoroughly investigated nonlinear wave-wave interactions are the "parametric instabilities", so called because of an analogy with parametric amplifiers, well-known devices in electrical engineering. A reason for the relatively advanced state of understanding of this subject is that the theory is basically a linear one, but linear about anoscillating equilibirum.

13.4.1 Coupled Oscillators

Consider the mechanical model of **?@fig-parametric_amplifier**, in which two oscillators M1 and M2 are coupled to a bar resting on a pivot. The pivot P is made to slide back and forth at a frequency ω_0, while the natural frequencies of the oscillators are ω_1 and ω_2. It is clear that, in the absence of friction, the pivot encounters no resistance as long as M1 and M2 are not moving. Furthermore, if P is not moving and M2 is put into motion, M1 will move; but as long as 2 is not the natural frequency of M1, the amplitude will be small. Suppose now that both P and M2 are set into motion. The displacement of M1 is proportional to the product of the displacement of M2 and the length of the lever arm and, hence, will vary in time as

$$\cos\omega_2 t\cos\omega_0 t = \frac{1}{2}\cos[(\omega_2+\omega_0)t] + \frac{1}{2}\cos[(\omega_2-\omega_0)t]$$

If ω_1 is equal to either $\omega_2+\omega_0$ or $\omega_2-\omega_0$, M1 will be resonantly excited and will grow to large amplitude. Once M1 starts oscillating, M2 will also gain energy, because one of the beat frequencies of ω_1 with ω_0 is just ω_2. Thus, once either oscillator is started, each will be excited by the other, and the system is unstable. The energy, of course, comes from the "pump" P, which encounters resistance once the rod is slanted. If the pump is strong enough, its oscillation amplitude is unaffected by M1 and M2; the instability can then be treated by a linear theory. In a plasma, the oscillators P, M1, and M2 may be different types of waves.

13.4.2 Frequency Matching

The equation of motion for a simple harmonic oscillator x_1 is

$$\ddot{x}_1 + \omega_1^2 x_1 = 0 \tag{13.12}$$

where ω_1 is its resonant frequency. If it is driven by a time-dependent force which is proportional to the product of the amplitude E_0 of the driver, or pump, and the amplitude x_2 of a second oscillator, the equation of motion becomes

$$\ddot{x}_1 + \omega_1^2 x_1 = c_1 x_2 E_0 \tag{13.13}$$

where c_1 is a constant indicating the strength of the coupling. A similar equation holds for x_2:

$$\ddot{x}_2 + \omega_1^2 x_2 = c_2 x_1 E_0 \tag{13.14}$$

Let $x_1 = \bar{x}_1\cos\omega t$, $x_2 = \bar{x}_2\cos\omega' t$, and $E_0 = \bar{E}_0\cos\omega_0 t$. Equation 13.14 becomes

$$
\begin{aligned}
(\omega_2^2 - {\omega'}^2)\bar{x}_2 \cos\omega' t &= c_2 \bar{E}_0 \bar{x}_1 \cos\omega_0 t \cos\omega t \\
&= c_2 \bar{E}_0 \bar{x}_1 \frac{1}{2}\{\cos[(\omega_0 + \omega)t] + \cos[(\omega_0 - \omega)t]\}
\end{aligned}
$$

The driving terms on the right can excite oscillators x_2 with frequencies

$$
\omega' = \omega_0 \pm \omega \tag{13.15}
$$

In the absence of nonlinear interactions, x_2 can only have the frequency ω_2, so we must have $\omega' = \omega_2$. However, the driving terms can cause a frequency shift so that ω' is only approximately equal to ω_2. Furthermore, ω' can be complex, since there is damping (which has been neglected so far for simplicity), or there can be growth (if there is an instability). In either case, x_2 is an oscillator with finite Q and can respond to a range of frequencies about ω_2. If ω is small, one can see from that both choices for ω' may lie within the bandwidth of x_2, and one must allow for the existence of two oscillators, $x_2(\omega_0 + \omega)$ and $x_2(\omega_0 - \omega)$.

Now let $x_1 = \bar{x}_1 \cos\omega'' t$ and $x_2 = \bar{x}_2 \cos[(\omega_0 \pm \omega)t]$ and insert into Equation 13.13:

$$
\begin{aligned}
&(\omega_1^2 - {\omega''}^2)\bar{x}_1 \cos\omega'' t \\
&= c_1 \bar{E}_0 \bar{x}_2 \frac{1}{2}\big(\cos\{[\omega_0 + (\omega_0 \pm \omega)]t\} + \cos\{[\omega_0 - (\omega_0 \pm \omega)]t\}\big) \\
&= c_1 \bar{E}_0 \bar{x}_2 \frac{1}{2}\big(\cos[(2\omega_0 \pm \omega)t] + \cos\omega t\big)
\end{aligned} \tag{13.16}
$$

The driving terms can excite not only the original oscillation $x_1(\omega)$, but also new frequencies $\omega'' = 2\omega_0 \pm \omega$. We shall consider the case $|\omega_0| \gg |\omega_1|$, so that $2\omega_0 \pm \omega$ lies outside the range of frequencies to which x_1 can respond, and $x_1(2\omega_0 \pm \omega)$ can be neglected. We therefore have three oscillators, $x_1(\omega)$, $x_2(\omega_0 - \omega)$, and $x_2(\omega_0 + \omega)$, which are coupled by Equation 13.13 and Equation 13.14:

$$
\begin{aligned}
(\omega_1^2 - \omega^2)x_1(\omega) - c_1 E_0(\omega_0)[x_2(\omega_0 - \omega) + x_2(\omega_0 + \omega)] &= 0 \\
[\omega_2^2 - (\omega_0 - \omega)^2]x_2(\omega_0 - \omega) - c_2 E_0(\omega_0)x_1(\omega) &= 0 \\
[\omega_2^2 - (\omega_0 + \omega)^2]x_2(\omega_0 + \omega) - c_2 E_0(\omega_0)x_1(\omega) &= 0
\end{aligned} \tag{13.17}
$$

The dispersion relation is given by setting the determinant of the coefficients equal to zero:

$$
\begin{vmatrix}
\omega^2 - \omega_1^2 & c_1 E_0 & c_1 E_0 \\
c_2 E_0 & (\omega_0 - \omega)^2 - \omega_2^2 & 0 \\
c_2 E_0 & 0 & (\omega_0 + \omega)^2 - \omega_2^2
\end{vmatrix} = 0 \tag{13.18}
$$

A solution with $\Im(\omega) > 0$ would indicate an instability.

For small frequency shifts and small damping or growth rates, we can set ω and ω' approximately equal to the undisturbed frequencies ω_1 and ω_2. Equation 13.15 then gives a frequency matching condition:

$$\omega_0 \approx \omega_2 \pm \omega_1 \tag{13.19}$$

When the oscillators are waves in a plasma, ωt must be replaced by $\omega t - \mathbf{k} \cdot \mathbf{r}$. There is then also a wavelength matching condition

$$\mathbf{k}_0 \approx \mathbf{k}_2 \pm \mathbf{k}_1 \tag{13.20}$$

describing spatial beats; that is, the periodicity of points of constructive and destructive interference in space. The two conditions Equation 13.19 and Equation 13.20 are easily understood by analogy with quantum mechanics. Multiplying the former by Planck's constant $\hbar$, we have

$$\hbar\omega_0 = \hbar\omega_2 \pm \hbar\omega_1$$

E_0 and x_2 may, for instance, be electromagnetic waves, so that $\hbar\omega_0$ and $\hbar\omega_2$ are the photon energies. The oscillator x_1 may be a Langmuir wave, or plasmon, with energy $\hbar\omega_1$. Equation 13.17 simply states the conservation of energy. Similarly, Equation 13.16 states the conservation of momentum $\hbar\mathbf{k}$.

For plasma waves, the simultaneous satisfaction of Equation 13.15 and Equation 13.16 in one-dimensional problems is possible only for certain combinations of waves. The required relationships are best seen on an $\omega - k$ diagram **?@fig-parallelogram_construction**. (A) shows the dispersion curves of an electron plasma wave ω_2 (Bohm-Gross wave) and an ion acoustic wave ω_1. A large-amplitude electron wave $(\omega_0, \mathbf{k}_0)$ can decay into a backward moving electron wave $(\omega_2, \mathbf{k}_2)$ and an ion wave $(\omega_1, \mathbf{k}_1)$. The parallelogram construction ensures that $\omega_0 = \omega_1 + \omega_2$ and $\mathbf{k}_0 = \mathbf{k}_1 + \mathbf{k}_2$. The positions of $(\omega_0, \mathbf{k}_0)$ and $(\omega_2, \mathbf{k}_2)$ on the electron curve must be adjusted so that the difference vector lies on the ion curve. Note that an electron wave cannot decay into two other electron waves, because there is no way to make the difference vector lie on the electron curve.

B shows the parallelogram construction for the "parametric decay" instability. Here, $(\omega_0, \mathbf{k}_0)$ is an incident electromagnetic wave of large phase velocity $(\omega_0/k_0 \approx c)$. It excites an electron wave and an ion wave moving in opposite directions. Since $|\mathbf{k}_0|$ is small, we have $|\mathbf{k}_1| \approx -|\mathbf{k}_2|$ and $\omega_0 = \omega_1 + \omega_2$ for this instability.

C shows the $\omega - k$ diagram for the "parametric backscattering" instability, in which a light wave excites an ion wave and another light wave moving in the opposite direction. This can also happen when the ion wave is replaced by a plasma wave. By analogy with similar phenomena in solid state physics, these processes are called, respectively, "stimulated Brillouin scattering" and "stimulated Raman scattering."

D represents the two-plasmon decay instability of an electromagnetic wave. Note that the two decay waves are both electron plasma waves, so that frequency matching can occur only if $\omega_0 \simeq 2\omega_p$. Expressed in terms of density, this condition is equivalent to $n \simeq n_c/4$, when n_c is the critical density associated with ω_0 ($n_C = m\epsilon_0\omega_0^2/e^2$). This instability can therefore be expected to occur only near the "quarter-critical" layer of an inhomogeneous plasma.

13.4.3 Instability Threshold

Parametric instabilities will occur at any amplitude if there is no damping, but in practice even a small amount of either collisional or Landau damping will prevent the instability unless the pump wave is rather strong. To calculate the threshold, one must introduce the damping rates Γ_1 and Γ_2 of the oscillators x_1 and x_2. Equation 13.12 then becomes

$$\ddot{x_1} + \omega_1^2 x_1 + 2\Gamma_1 \dot{x_1} = 0$$

For instance, if x_1 is the displacement of a spring damped by friction, the last term represents a force proportional to the velocity. If x_1 is the electron density in a plasma wave damped by electron–neutral collisions, Γ_1 is $v_c/2$ (???) Examination of Equation 13.13, Equation 13.14, and Equation 13.18 will show that it is all right to use exponential notation and let $\mathrm{d}/\mathrm{d}t \to i\omega$ for x_1 and x_2, as long as we keep E_0 real and allow $\bar{x}_1$ and $\bar{x}_2$ to be complex. Equation 13.13 and Equation 13.14 become

$$\begin{aligned} (\omega_1^2 - \omega^2 - 2i\omega\Gamma_1)x_1(\omega) &= c_1 x_2 E_0 \\ [\omega_2^2 - (\omega - \omega_0)^2 2i(\omega - \omega_0)\Gamma_2]x_2(\omega - \omega_0) &= c_2 x_1 E_0 \end{aligned} \tag{13.21}$$

We further restrict ourselves to the simple case of two waves — that is, when $\omega \simeq \omega_1$ and $\omega_0 - \omega \simeq \omega_2$ but $\omega_0 + \omega$ is far enough from ω_2 to be nonresonant — in which case the third row and column of Equation 13.18 can be ignored. If we now express x_1, x_2, and E_0 in terms of their peak values, as in Equation 13.16, a factor of 1/2 appears on the right-hand sides of Equation 13.21. Discarding the nonresonant terms and eliminating x_1 and x_2 from Equation 13.21, we obtain

$$(\omega^2 - \omega_1^2 + 2i\omega\Gamma_1)[(\omega_0 - \omega)^2 - \omega_2^2 - 2i(\omega_0 - \omega)\Gamma_2] = \frac{1}{4}c_1 c_2 \bar{E}_0^2$$

At threshold, we may set $\Im(\omega) = 0$. The lowest threshold will occur at exact frequency matching; i.e., $\omega = \omega_1, \omega_0 - \omega = \omega_2$. Then the last equation gives

$$c_1 c_2 (\bar{E}_0^2)_{\text{threshold}} = 16\omega_1\omega_2\Gamma_1\Gamma_2$$

The threshold goes to zero with the damping of *either* wave.

13.4.4 Physical Mechanism

The parametric excitation of waves can be understood very simply in terms of the ponderomotive force (Section 13.3). As an illustration, consider the case of an electromagnetic wave $(\omega_0, \mathbf{k}_0)$ driving an electron plasma wave $(\omega_2, \mathbf{k}_2)$ and a low-frequency ion wave $(\omega_1, \mathbf{k}_1)$ (fig-parallelogram_construction B). Since ω_1 is small, ω_0 must be close to ω_p. However, the behavior is quite different for $\omega_0 < \omega_p$ than for $\omega_0 > \omega_p$. The former case gives rise to the "oscillating two-stream" instability (which will be treated in detail), and the latter to the "parametric decay" instability.

Suppose there is a density perturbation in the plasma of the form $n_1 \cos k_1 x$; this perturbation can occur spontaneously as one component of the thermal noise. Let the pump wave have an electric field $E_0 \cos \omega_0 t$ in the x direction, as shown in **?@fig-oscillating_2stream**. In the absence of a dc field $\mathbf{B}_0$, the pump wave follows the relation $\omega_0^2 = \omega_p^2 + c^2 k_0^2$, so that $k_0 \approx 0$ for $\omega_0 \approx \omega_p$. We may therefore regard E_0 as spatially uniform. If ω_0 is less than ω_p (HOW IS THIS POSSIBLE???), which is the resonant frequency of the cold electron fluid, the electrons will move in the direction opposite to E_0, while the ions do not move on the time scale of ω_0. The density ripple then causes a charge separation, as shown in **?@fig-oscillating_2stream**. The electrostatic charges create a field E_1, which oscillates at the frequency ω_0. The ponderomotive force due to the total field is given by Equation 13.11:

$$\mathbf{F}_{NL} = -\frac{\omega_p^2}{\omega^2} \nabla \frac{\langle (E_0 + E_1)^2 \rangle}{2} \epsilon_0$$

Since E0 is uniform and much larger than E_1, only the cross term is important:

$$\mathbf{F}_{NL} = -\frac{\omega_p^2}{\omega^2} \frac{\partial}{\partial x} \frac{\langle 2E_0 E_1 \rangle}{2} \epsilon_0$$

This force does not average to zero, since E1 changes sign with E0. As seen in fig-oscillating_2stream, F_{NL} is zero at the peaks and troughs of n_1 but is large where ∇n_1 is large. This spatial distribution causes F_{NL} to push electrons from regions of low density to regions of high density. The resulting de electric field drags the ions along also, and the density perturbation grows. The threshold value of F_{NL} is the value just sufficient to overcome the pressure $\nabla n_{i1}(k_B T_i + k_B T_e)$, which tends to smooth the density. The density ripple does not propagate, so that $\Re(\omega_1) = 0$. This is called the *oscillating two-stream instability* because the sloshing electrons have a time-averaged distribution function which is double-peaked, as in the two-stream instability (Sect. ???).

13.4.5 The Oscillating Two-Stream Instability

I do not understand parametric instability so far. Too hard.

13.4.6 The Parametric Decay Instability

13.5 Plasma Echoes

Echoes in a plasma are the excitement of new waves due to nonlinear interaction. The excitement may happen at an arbitrarily large time, which is the main source of difficulties in understanding Landau damping. Since Landau damping *does not involve collisions or dissipation*, it is a *reversible* process. That this is true is vividly demonstrated by the remarkable phenomenon of plasma echoes.

13.6 Nonlinear Landau Damping

When the amplitude of an electron or ion wave excited, say, by a grid is followed in space, it is often found that the decay is not exponential, as predicted by linear theory, if the amplitude is large. Instead, one typically finds that the amplitude decays, grows again, and then oscillates before settling down to a steady value. Although other effects may also be operative, these oscillations in amplitude are exactly what would be expected from the nonlinear effect of particle trapping discussed in Section 8.6.

13.7 Equations of Nonlinear Plasma Physics

13.7.1 The Korteweg-de Vries Equation

$$\frac{\partial U}{\partial \tau} + U\frac{\partial U}{\partial \xi} + \frac{1}{2}\frac{\partial^2 U}{\partial \xi^3} = 0$$

where U is amplitude, and τ and ξ are timelike and spacelike variables, respectively.

13.7.2 The Nonlinear Schrödinger Equation

$$i\frac{\partial\psi}{\partial t}+p\frac{\partial^2\psi}{\partial x^2}+q|\psi|^2\psi=0 \tag{13.22}$$

where ψ is the wave amplitude, $i=\sqrt(-1)$, and p and q are coefficients whose physical significance will be explained shortly. Equation 13.22 differs from the usual Schrödinger equation

$$i\hbar\frac{\partial\psi}{\partial t}+\frac{\hbar^2}{2m}\frac{\partial^2\psi}{\partial x^2}-V(x,t)\psi=0$$

in that the potential $V(x,t)$ depends on ψ itself, making the last term nonlinear.

14 Turbulence

Turbulence is perhaps the most beautiful unsolved problem of classical physics. History of turbulence in interplanetary space is, perhaps, even more interesting since its knowledge proceeds together with the human conquest of space. The behavior of a flow which rebels against the deterministic rules of classical dynamics is called *turbulent*. Practically turbulence appears everywhere when the velocity of the flow is high enough. At a first sight turbulence looks strongly irregular and chaotic, both in space and time: it seems to be impossible to predict any future state of the motion. However, it is interesting to recognize the fact that, when we take a picture of a turbulent flow at a given time, we see the presence of a lot of *different turbulent structures of all sizes* which are actively present during the motion. We recognize that the gross features of the flow are reproducible but details are not predictable. We have to use a *statistical approach* to turbulence, just as it is done to describe stochastic processes, even if the problem is born within the strange dynamics of a deterministic system.

Small fluctuations in plasmas lead to turbulence, and turbulent eddies can very effectively transport mass, momentum and heat from the hot core across confining magnetic field lines out to the cooler plasma edge. Predicting this phenomenon of turbulent-transport is essential for solar wind study and the understanding and development of fusion reactors.

Turbulence became an experimental science since Osborne Reynolds who, at the end of 19th century, observed and investigated experimentally the transition from laminar to turbulent flow. He noticed that the flow inside a pipe becomes turbulent every time a single parameter, a combination of the mass density ρ, viscosity coefficient η, a characteristic velocity U, and length L, would increase. This parameter $\mathrm{Re} = UL\rho/\eta$ is now called the *Reynolds number*. When Re increases beyond a certain threshold of the order of $\mathrm{Re} \simeq 4000$, the flow becomes turbulent.

Predictability in turbulence can be recast at a statistical level. In other words, when we look at two different samples of turbulence, even collected within the same medium, we can see that details look very different. What is actually common is a generic stochastic behavior. This means that the global statistical behavior does not change going from one sample to the other. *Fully developed turbulent flows are extremely sensitive to small perturbations but have statistical properties that are insensitive to perturbations.* Fluctuations of a certain stochastic variable ψ are defined as the difference from the average value $\delta\psi = \psi - \langle\psi\rangle$, where brackets mean some averaging process. There are, at least, three different kinds of averaging procedures that may be used to obtain statistically-averaged properties of turbulence:

1. Space averaging over flows that are statistically homogeneous over scales larger than those of fluctuations.
2. Ensemble averaging where average is taken over an ensemble of turbulent flows prepared under nearly identical external conditions. Each member of the ensemble is called a *realization*.
3. Time averaging, which is useful only if the turbulence is statistically stationary over time scales much larger than the time scale of fluctuations. In practice, because of the convenience offered by locating a probe at a fixed point in space and integrating in time, experimental results are usually obtained as time averages. The ergodic theorem (Halmos, 1956) assures that time averages coincide with ensemble averages under some standard conditions.

A different property of turbulence is that *all dynamically interesting scales are excited*, that is, *energy is spread over all scales* and a kind of self-similarity is observed.

Since fully developed turbulence involves a hierarchy of scales, a large number of interacting degrees of freedom are involved. Then, there should be an asymptotic statistical state of turbulence that is independent on the details of the flow. Hopefully, this asymptotic state depends, perhaps in a critical way, only on simple statistical properties like energy spectra, as much as in statistical mechanics equilibrium where the statistical state is determined by the energy spectrum (Huang, 1987). Of course, we cannot expect that the statistical state would determine the details of individual realizations, because realizations need not to be given the same weight in different ensembles with the same low-order statistical properties.

It should be emphasized that there are no firm mathematical arguments for the existence of an asymptotic statistical state. Reproducible statistical results are obtained from observations, that is, it is suggested experimentally and from physical plausibility. Apart from physical plausibility, it is embarrassing that such an important feature of fully developed turbulence, as the existence of a statistical stability, should remain unsolved. However, such is the complex nature of turbulence.

14.1 MHD

Nunov highly recommended a review (Schekochihin 2022). Another nice reference for the solar wind turbulence is presented (Bruno and Carbone 2013), where I cite many materials in this note.

14.1.1 Navier–Stokes equation and Reynolds number

14.1.2 Coupling between charged fluid and magnetic field

Elsässer variables are used to extract the Alfvénic component from MHD. They are defined as

$$\mathbf{z}^{+} = \mathbf{v} + \mathbf{b}$$
$$\mathbf{z}^{-} = \mathbf{v} - \mathbf{b}$$

where $\mathbf{b} = \mathbf{B}/\sqrt{\mu_0 \rho}$.

The perturbations are written as

$$\delta\mathbf{z}^{+} = \delta\mathbf{v} + \delta\mathbf{b}$$
$$\delta\mathbf{z}^{-} = \delta\mathbf{v} - \delta\mathbf{b}$$

$\mathbf{z}^{\pm}$ corresponds to anti-parallel/parallel propagating modes:

- Parallel wave: $\delta\mathbf{v} = -\delta\mathbf{b} \Rightarrow \delta\mathbf{z}^{+} = 0, \delta\mathbf{z}^{-} = 2\delta\mathbf{v}$
- Anti-parallel wave: $\delta\mathbf{v} = \delta\mathbf{b} \Rightarrow \delta\mathbf{z}^{+} = 2\delta\mathbf{v}, \delta\mathbf{z}^{-} = 0$

The incompressible MHD wave equation in fluctuating Elsässer form is:

$$\frac{\partial \delta\mathbf{z}^{\pm}}{\partial t} \mp v_A \nabla_{\parallel} \delta\mathbf{z}^{\pm} + \delta\mathbf{z}^{\mp} \cdot \nabla \delta\mathbf{z}^{\pm} = -\nabla p \tag{14.1}$$

$$\nabla \cdot \delta\mathbf{z}^{\pm} = 0 \tag{14.2}$$

The zero divergence means that there are no forcing or dissipation terms.

Equation 14.1 unveils the interesting phenonmena in Alfvénic turbulence study. The second term on the left-hand side is a linear term (v_A constant?) that represents propagation of waves parallel to the mean field. The third term represent the non-linear interaction of counter-propagating waves, during which energy is transferred to smaller scales. This is exactly turbulence.

14.1.3 Critical balance

If we match the Alfvén (linear) timescale, $\tau_A \sim l_\parallel / v_A$, with the nonlinear timescale, $\tau_{nl} \sim l_\perp / \delta z$, we have the so-called *critical balance*

$$\tau_A = \tau_{nl}$$

Assume a constant energy cascade rate $\epsilon \sim \delta z^2 / \tau_{nl}$, and an energy injection rate at scale L $\epsilon_L \sim v_A^2 / \tau_{nl} = v_A^3 / L$. When we match the two rates and … (I have some memory from Nunov's lecture at Nordita), we have $k_\parallel \propto k_\perp^{2/3}$. This means that anisotropy grows with decreasing scale. The power spectra are given as

$$\begin{aligned} E(k_\perp) &\propto k_\perp^{-5/3} \\ E(k_\parallel) &\propto k_\parallel^{-2} \end{aligned}$$

The critical balance scalings were seen in observations when analysed with local mean field techniques [Chen 2016]. However, why this is the case in nature is still under debate.

14.1.4 Residual Energy

Residual energy is the difference in energy between magnetic and velocity fluctuations. The normalized residual energy is

$$\sigma_r = \frac{|\delta \mathbf{v}|^2 - |\delta \mathbf{b}|^2}{|\delta \mathbf{v}|^2 + |\delta \mathbf{b}|^2} \tag{14.3}$$

σ_r is zero for an Alfvén wave but is generally negative ($|\delta \mathbf{b}|^2 > |\delta \mathbf{v}|^2$) at MHD scales for solar wind observations, meaning that they are mostly likely turbulence but not waves.

14.1.5 Normalized Cross Helicity

Cross helicity is the difference in energy between $\mathbf{z}^+$ and $\mathbf{z}^-$ fluctuations (Section 15.3). The normalized cross helicity is

$$\sigma_c = \frac{|\delta \mathbf{z}^+|^2 - |\delta \mathbf{z}^-|^2}{|\delta \mathbf{z}^+|^2 - |\delta \mathbf{z}^-|^2} = \frac{2\delta \mathbf{v} \cdot \delta \mathbf{b}}{|\delta \mathbf{v}|^2 + |\delta \mathbf{b}|^2} \tag{14.4}$$

- $|\sigma_c| = 1$: unidirectional Alfvén waves (no turbulence)
- $|\sigma_c| \lesssim 1$: imbalanced turbulence
- $|\sigma_c| = 0$: balanced turbulence (or non-Alfvénic fluctuations)

The solar wind is typically imbalanced towards the anti-sunward direction. From Equation 14.3 and Equation 14.4 we can easily see that

$$\sqrt{\sigma_r^2 + \sigma_c^2} \leq 1$$

The observed distribution is shown in Figure 14.1. Note, however, σ_r depends on how you evaluate δb. In the MHD case $\mathbf{b} = \mathbf{B}/\sqrt{\mu_0 \rho}$ is used, but in the kinetic case there will be an extra coefficient. The results look pretty different!

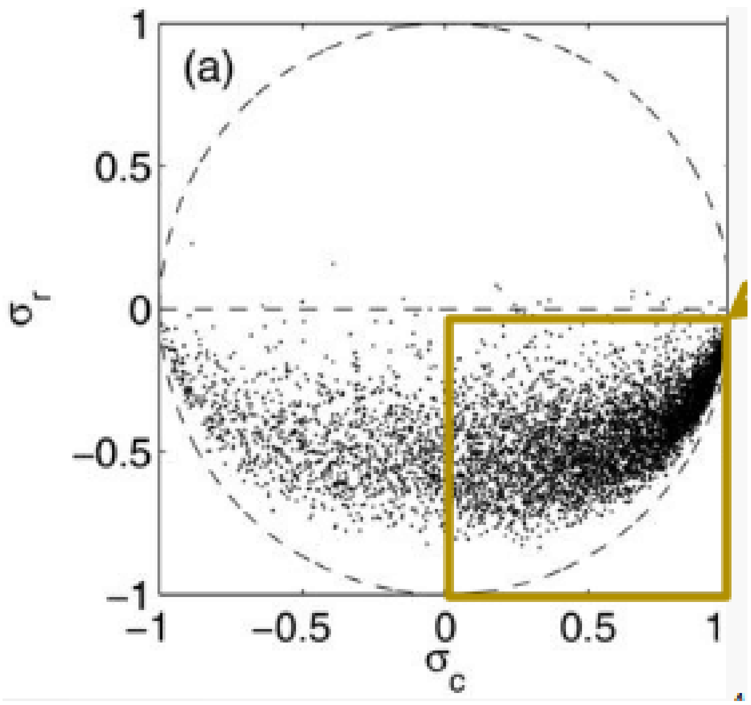

Figure 14.1: Solar wind normalized residual energy and cross helicity statistics from Chen+ 2013.

14.2 Solar Wind Turbulence

14.2.1 Taylor's Hypothesis

Solar wind is a supersonic flow ($v_A/v_{\mathrm{SW}} \ll 1$), with advection timescale much shorter than any dynamical timescales in the plasma. This means that for spacecraft observations, the time series represents an instantaneous spatial cut through the solar wind plasma. We can thus relate spacecraft frequency, f_{sc}, to wavenumber in the plasma frame, k, in a simple way via

$$k = \frac{2\pi f_{\rm sc}}{v_{\rm SW}}$$

This is often not appropriate in the magnetosheath, and modified Taylor's hypothesis is required close to the Sun.

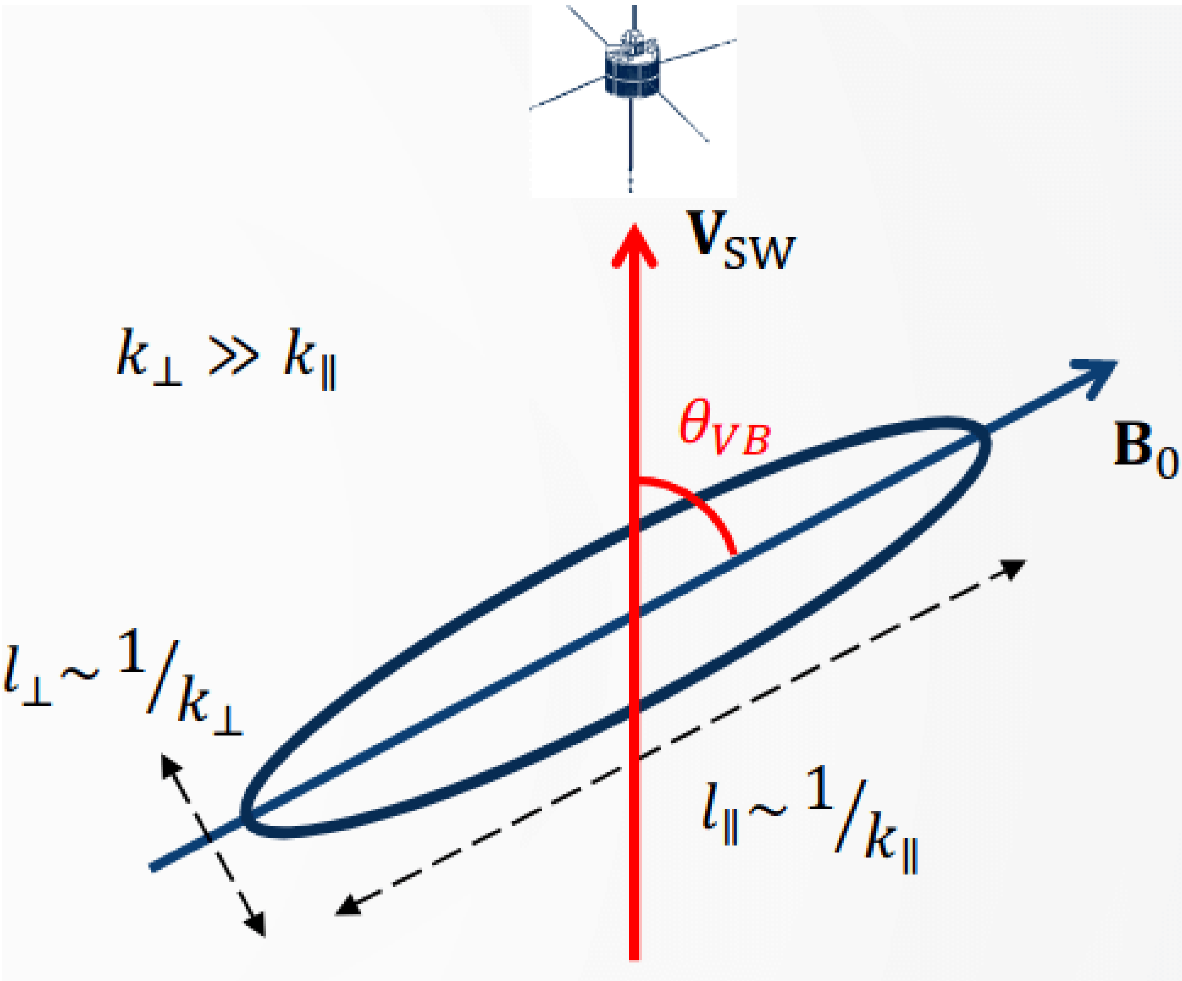

Figure 14.2: Wave scale from typical L1 satellite solar wind observation.

The wavenumber k determined from Taylor's hypothesis is really the flow-aligned component of $\mathbf{k}$. For a given angle θ_{vb}, one cannot distinguish $k_\perp$ and $k_\parallel$; these can be possibly measured for small ($\sim 0^o$ for $k_\parallel$) and large ($\sim 90^o$ for $k_\perp$) at different times. In the solar wind, $P(k_\perp) \gg P(k_\parallel)$, where P is the power of perturbation. For a Parker spiral-like magnetic field at 1 AU, the angle between $\mathbf{B}_0$ and $\mathbf{v}_{\rm sw}$ is rarely small ($\sim 45^o$), the power spectra are typically dominated by the contribution from the $k_\perp$ fluctuations (Figure 14.2).

14.2.2 Solar Wind Power Spectrum

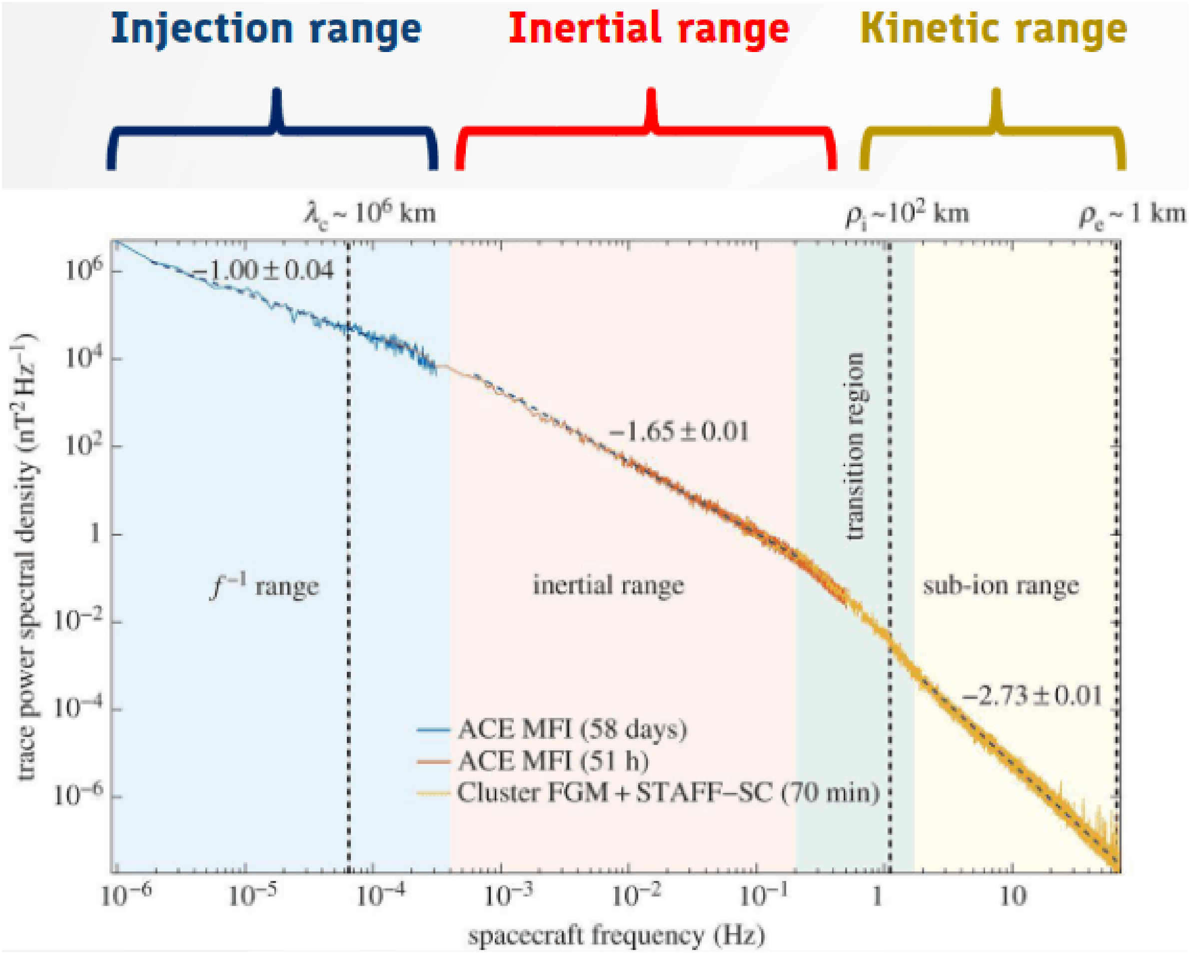

Figure 14.3: Solar wind power spectrum from Kiyoni+, 2015.

When we plot the solar wind power spectrum ($\delta|\mathbf{B}|$), it is usually representative of $k_\perp$. There are three distinct power-law ranges in the spectrum from spacecraft solar wind observations (Figure 14.3):

- Injection range:
  - Large amplitude, low frequency Alfvén waves originating from the Sun.
  - The f^{-1} spectrum is known as the "pink noise".
  - Non-interacting
  - Inject energy to the MHD cascade at higher frequencies
- Inertial range:

  - Mostly incompressible Alfvénic turbulence.
  - $\sim f^{-5/3}$ spectrum (Kolmogorov type)
  - a cascade of energy to smaller scales

- Kinetic range:
  - Scales at which particles are heated
  - Typically $f^{-2.8}$ spectrum
  - Possibly KAW (Section 7.7) or whistler mode (Section 7.5) turbulence

When we move to smaller scales, anisotropy increases. This is demonstrated in Figure 14.4 as a result of $k_\perp \gg k_\parallel$, which is also why gyrokinetics is the theory for turbulence.

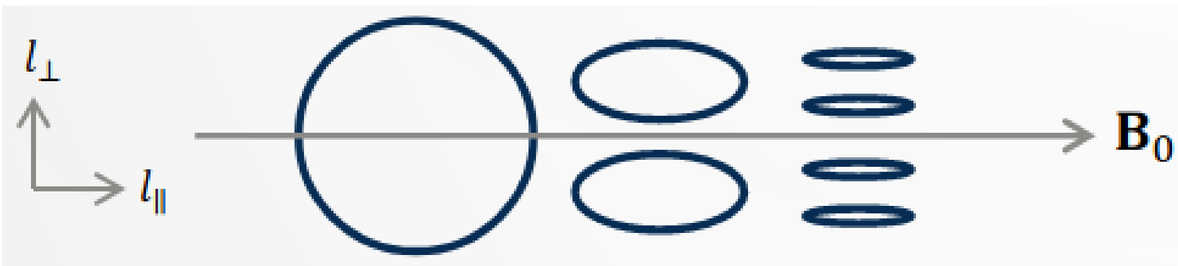

Figure 14.4: Demonstration of parallel and perpendicular length scales of turbulence.

Fluctuation modes in the inertial range consists of 90% incompressible (Alfvén) modes and 10% compressible slow/mirror modes. Alfvénic turbulence is thought to be passive to the compressive modes: compressive modes scatter off Alfvén modes without affecting them significantly (i.e. being decoupled from each other). The compressive modes are expected to damp $\sim k_\parallel v_A$. Since $k_\perp \gg k_\parallel$, the damp is not significant for compressive modes, so they tend to be more anisotropic than the Alfvénic turbulence.

15 Geometry

This chapter introduces the topology-related concepts in plasma physics, including ropes, knots, boundaries, and null points. Usually, observers and modelers have different views of topology because of the tools at hand: observers have probes which give measurements as a function of both time and space, while modelers have full spatial-temporal information under the given resolution. It always amazes me how observers can deduce the general picture of plasma structures with such limited data. Incorporating observation experience into physics as well as diagnosing numerical simulations with physics are our main goals in studing geometry.

Generally speaking, there are two ways of tackling the geometry problems: physics-based methods and statistical methods (classical and machine learning). We may find harmony when combining these two families and reach optimal results.

15.1 Helicity

In fluid dynamics, helicity is, under appropriate conditions, an invariant of the Euler equations of fluid flow, having a topological interpretation as a measure of *linkage* and *knottedness* of vortex lines in the flow (Moffatt 1969).

Let $\mathbf{u}(x,t)$ be the velocity field and $\nabla\times\mathbf{u}$ the corresponding vorticity field. Under the following three conditions, the vortex lines are transported with (or "frozen in") the flow:

1. the fluid is inviscid;
2. either the flow is incompressible ($\nabla\cdot\mathbf{u}=0$), or it is compressible with a barotropic relation $p=p(\rho)$ between pressure p and density ρ;
3. any body forces acting on the fluid are conservative. Under these conditions, any closed surface S on which $n\cdot(\nabla\times\mathbf{u})=0$ is, like vorticity, transported with the flow.

Let V be the volume inside such a surface. Then the helicity in V is defined by

$$H=\int_V \mathbf{u}\cdot(\nabla\times\mathbf{u})\,dV \tag{15.1}$$

For a localised vorticity distribution in an unbounded fluid, V can be taken to be the whole space, and H is then the total helicity of the flow. H is invariant precisely because the vortex lines are frozen in the flow and their linkage and/or knottedness is therefore conserved, as recognized by Lord Kelvin (1868). Helicity is a pseudo-scalar quantity: it changes sign under change from a right-handed to a left-handed frame of reference; it can be considered as a measure of the handedness (or chirality) of the flow. Helicity is one of the four known integral invariants of the Euler equations; the other three are energy, momentum and angular momentum.

The invariance of helicity provides an essential cornerstone of the subject topological fluid dynamics and MHD, which is concerned with global properties of flows and their topological characteristics.

15.2 Magnetic Helicity

The helicity of a smooth vector field defined on a domain in 3D space is the standard measure of the extent to which the field lines wrap and coil around one another. As to magnetic helicity, this "vector field" is magnetic field. It is a generalization of the topological concept of linking number to the differential quantities required to describe the magnetic field. As with many quantities in electromagnetism, magnetic helicity (which describes magnetic field lines) is closely related to fluid mechanical helicity (which describes fluid flow lines).

If magnetic field lines follow the strands of a twisted rope, this configuration would have nonzero magnetic helicity; left-handed ropes would have negative values and right-handed ropes would have positive values.

Formally,

$$H = \int \mathbf{A} \cdot \mathbf{B} \mathrm{d}\mathbf{r}^3 \tag{15.2}$$

where

- H is the helicity of the entire magnetic field
- $\mathbf{B}$ is the magnetic field strength
- $\mathbf{A}$ is the vector potential of $\mathbf{B}$ and $\mathbf{B} = \nabla \times \mathbf{A}$

Magnetic helicity has units of Wb^2 in SI units and Mx^2 in Gaussian Units. Note that $\mathbf{A} \cdot \mathbf{B}$ should not be considered as "helicity density" because of gauge freedom.

It is a conserved quantity in electromagnetic fields, even when magnetic reconnection dissipates energy. The concept is useful in solar dynamics and in dynamo theory. Helicity is approximately conserved during magnetic reconnection and topology changes. Helicity can be injected into a system such as the solar corona. When too much builds up, it ends up being expelled through coronal mass ejections. The simple proof in ideal MHD can be found on Wikipedia.

Magnetic helicity is a gauge-dependent quantity, because $\mathbf{A}$ can be redefined by adding a gradient to it (gauge transformation):

$$\mathbf{A}' = \mathbf{A} + \nabla\phi$$

However, for perfectly conducting boundaries or periodic systems without a net magnetic flux, the magnetic helicity is gauge invariant. A gauge-invariant relative helicity has been defined for volumes with non-zero magnetic flux on their boundary surfaces. If the magnetic field is turbulent and weakly inhomogeneous a magnetic helicity density and its associated flux can be defined in terms of the density of field line linkages.

The topological properties of a magnetic field are interpreted in terms of magnetic helicity. The total helicity of a collection of flux tubes arises from the linking of flux tubes with one another (mutual helicity) and the internal magnetic structure of each flux tube (self-helicity). Reconnection changes the topology and magnetic connectivity of flux tubes. This can also be viewed as a redistribution of self- and mutual helicities. If total magnetic helicity is approximately conserved, it is possible to put quantitative limits upon the changes in self- and mutual helicities. This can be interpreted as the change in magnetic flux tube linkage (due to reconnection) and amount of twist present in the reconnected flux tubes. [Wright & Berger, 1989]

Simple examples:

1. A single untwisted closed flux loop has $H = 0$.
2. A single flux rope with a magnetic flux of ϕ that twists around itself T times has a helicity of $H = T\phi^2$.
3. Two interlinked untwisted flux loops with fluxes ϕ_1 and ϕ_2 have $H = \pm 2\phi_1\phi_2$ where the sign depends on the sense of the linkedness.

There are generalizations to allow for gauge-invariant definitions of helicity. [Berger & Field (1984)] defined the relative magnetic helicity to be

$$H = \int_V \mathbf{A}\cdot\mathbf{B} - \mathbf{A}_0\cdot\mathbf{B}_0 dV \tag{15.3}$$

where $\mathbf{B}_0 = \nabla\times\mathbf{A}_0$ is the potential field inside V with the same field outside of V (see also Finn & Antonsen 1985).

In toroidal laboratory experiments, it is natural to consider the volume contained within conducting wall boundaries that are coincident with closed flux surfaces (i.e., the magnetic field along the wall is parallel to the boundary).

The time evolution of magnetic helicity is given by

$$
\frac{\mathrm{d}H}{\mathrm{d}t} = -2c\int_V \mathbf{E}\cdot\mathbf{B}dV + 2c\int_S \mathbf{A}_p \times \mathbf{E}\cdot\mathrm{d}\mathbf{S}
$$

where we choose $\nabla\times\mathbf{A}_p = 0$ and $\mathbf{A}_p\cdot\mathrm{d}\mathbf{S} = 0$ on S. The first term represents helicity dissipation when $E_\parallel \neq 0$, which is always zero in ideal MHD. The second term represents helicity fluxes in and out of the system, for example, flux emergence from the solar photosphere corresponds to helicity injection in the corona.

15.3 Cross Helicity

The cross helicity measures the imbalance between interacting waves, which is important in MHD turbulence (Section 14.1.5). It is given by

$$
H_C = \int_V \mathbf{v}\cdot\mathbf{B}dV \tag{15.4}
$$

In ideal MHD, the rate of change of H_C is

$$
\frac{\mathrm{d}H_C}{\mathrm{d}t} = -\oint_S \mathrm{d}\mathbf{S}\cdot\Big[\Big(\frac{1}{2}v^2 + \frac{\gamma}{\gamma-1}\frac{p}{\rho}\Big)\mathbf{B} - \mathbf{v}\times(\mathbf{v}\times\mathbf{B})\Big]
$$

This vanishes when $\mathrm{d}\mathbf{S}\cdot\mathbf{B} = d\mathbf{S}\cdot\mathbf{V} = 0$ along the boundary S or when the boundary conditions are periodic. Cross helicity is an ideal MHD invariant when this integral vanishes.

There are discretized forms of cross helicity from the observation point of view. Check it out if you want to know more.

15.4 Flux Rope Identification

Using turbulence parameters to find flux ropes (Zhao et al. 2020):

$$
\begin{aligned}
\sigma_r &< -0.5 \\
|\sigma_c| &< 0.4 \\
|\sigma_m| &> 0.7
\end{aligned}
$$

where σ_m is the *normalized reduced magnetic helicity*, which is a measure of $\mathbf{B}$ rotation. Strictly speaking, the magnetic helicity depends on the spatial properties of the magnetic field topology, and thus cannot be directly evaluated from single spacecraft measurements. However, Matthaeus+ (1982) described a reduced form of magnetic helicity that can be estimated with measurements from a single spacecraft based on the magnetic power spectrum. The normalized reduced magnetic helicity can be estimated by

$$\sigma_m(\nu,t) = \frac{2\Im[W_T^*(\nu,t)\cdot W_N(\nu,t)]}{|W_R(\nu,t)|^2 + |W_T(\nu,t)|^2 + |W_N(\nu,t)|^2} \tag{15.5}$$

where ν is the frequency associated with the Wavelet function and the sampling period of the measured magnetic field in the radial tangential normal (RTN) coordinate system. The spectra $W_R(\nu,t)$, $W_T(\nu,t)$, and $W_N(\nu,t)$ are the wavelet transforms of time series of $\mathbf{B}_{1R}$, $\mathbf{B}_{1T}$, and $\mathbf{B}_{1N}$, respectively; and $W_T^*(\nu,t)$ is the conjugate of $W_T(\nu,t)$. From the resulting spectrogram of the magnetic helicity, σ_m, one can determine both the magnitude and the handedness (chirality) of underlying fluctuations at a specific scale. A positive value of σ_m corresponds to right-handed chirality and a negative value to left-handed chirality.

15.5 Reconnection Identification

Here we present analytical fields for an X-point configuration and an O-point configuration, respectively.

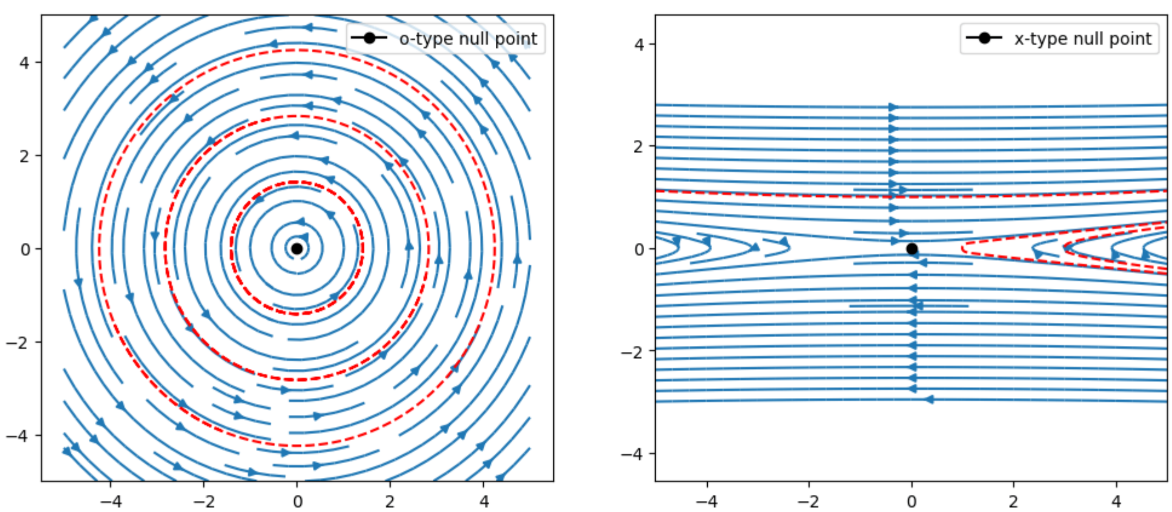

Figure 15.1: Example of (a) O-point and (b) X-point.

15.5.1 2D

Identification of 2D reconnection sites is easy. Usually we start with a Harris current sheet topology. We can define a flux function, and the null points are simple saddle points and extremas of the flux function.

15.5.2 3D

Identification of 3D reconnection sites is not easy.

Implement Lapenta's method.

The four-field junction (FFJ) method is proposed by Finnish researchers (???).

BBF

15.6 Magnetospause Identification

From pressure balance argument, let $\beta^* = (p_{\text{th}} + p_{\text{dyn}})/p_B$, we can have a simple criterion:

$$\beta^* \simeq 1$$

16 Solar Wind

The *solar wind* is a high-speed particle stream continuously blown out from the Sun into interplanetary space. It extends far beyond the orbit of the Earth, and terminates in a shock front, called the heliopause, where it interfaces with the weakly ionized interstellar medium. The heliopause is predicted to lie between 110 and 160 AU (1 Astronomical Unit is $1.5 \times 10^{11}\,\mathrm{m}$) from the center of the Sun. Voyager 1 & 2, which were launched in 1977, have passed through the heliopause, and are still functional!

In the vicinity of the Earth, (i.e., at about 1 AU from the Sun) the solar wind velocity typically ranges between 300 and 1400 $\mathrm{km\,s^{-1}}$. The average value is approximately $500\,\mathrm{km\,s^{-1}}$, which corresponds to about a 4 day time of flight from the Sun. Note that the solar wind is both *super-sonic* and *super-Alfvénic*.

The solar wind is predominately composed of protons and electrons.

Amazingly enough, the solar wind was predicted theoretically by Eugine Parker in 1958, a number of years before its existence was confirmed using satellite data. Parker's prediction of a super-sonic outflow of gas from the Sun is a fascinating scientific detective story, as well as a wonderful application of plasma physics.

The solar wind originates from the solar corona. The solar corona is a hot, tenuous plasma surrounding the Sun, with characteristic temperatures and particle densities of about $10^6\,\mathrm{K}$ and $10^{14}\,\mathrm{m^{-3}}$, respectively. Note that the corona is *far hotter* than the solar atmosphere, or photosphere. In fact, the temperature of the photosphere is only about 6000 K. It is thought that the corona is heated by Alfvén waves emanating from the photosphere together with a turbulent cascading process. The solar corona is most easily observed during a total solar eclipse, when it is visible as a white filamentary region immediately surrounding the Sun.

Let us start, following Chapman, by attempting to construct a model for a static solar corona. The equation of hydrostatic equilibrium for the corona takes the form

$$\frac{\mathrm{d}p}{\mathrm{d}r} = -\rho\,\frac{G\,M_\odot}{r^2} \tag{16.1}$$

where $G = 6.67 \times 10^{-11}\,\mathrm{m^3\,s^{-2}\,kg^{-1}}$ is the gravitational constant, and $M_\odot = 2 \times 10^{30}\,\mathrm{kg}$ is the solar mass. The plasma density is written

$$\rho \simeq n\,m_p \tag{16.2}$$

where n is the number density of protons. If both protons and electrons are assumed to possess a common temperature, $T(r)$, then the coronal pressure is given by

$$p = 2nk_B T \tag{16.3}$$

The thermal conductivity of the corona is dominated by the electron thermal conductivity, and takes the form (???)

$$\kappa = \kappa_0 \, T^{5/2}$$

where κ_0 is a relatively weak function of density and temperature. For typical coronal conditions this conductivity is extremely high: i.e., it is about twenty times the thermal conductivity of copper at room temperature. The coronal heat flux density is written

$$\mathbf{q} = -\kappa \nabla T$$

For a static corona, in the absence of energy sources or sinks, we require

$$\nabla \cdot \mathbf{q} = 0$$

Assuming spherical symmetry, this expression reduces to

$$\frac{1}{r^2}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2 \kappa_0 \, k_B T^{5/2}\,\frac{dT}{\mathrm{d}r}\right) = 0$$

Adopting the sensible boundary condition that the coronal temperature must tend to zero at large distances from the Sun, we obtain

$$T(r) = T(a)\left(\frac{a}{r}\right)^{2/7} \tag{16.4}$$

The reference level $r = a$ is conveniently taken to be the base of the corona, where $a \sim 7 \times 10^5$ km, $n \sim 2 \times 10^{14}\,\mathrm{m}^{-3}$, and $T \sim 2 \times 10^6$K.

Equation 16.1, Equation 16.2, Equation 16.3, and Equation 16.4 can be combined and integrated to give

$$p(r) = p(a)\exp\left\{\frac{7}{5}\,\frac{G\,M_\odot\,m_p}{2k_B T(a)\,a}\left[\left(\frac{a}{r}\right)^{5/7} - 1\right]\right\}$$

Note that as $r \to \infty$ the coronal pressure tends towards a finite constant value:

$$p(\infty) = p(a)\,\exp\left\{-\frac{7}{5}\,\frac{G\,M_\odot\,m_p}{2k_B T(a)\,a}\right\}$$

There is, of course, nothing at large distances from the Sun which could contain such a pressure (the pressure of the interstellar medium is negligibly small). Thus, we conclude, with Parker, that the static coronal model is unphysical.

Since we have just demonstrated that a static model of the solar corona is unsatisfactory, let us now attempt to construct a dynamic model in which material flows outward from the Sun.

16.1 Parker Solar Wind Model

By symmetry, we expect a purely radial coronal outflow. The radial momentum conservation equation for the corona takes the form

$$\rho\,u\,\frac{du}{\mathrm{d}r} = -\frac{\mathrm{d}p}{\mathrm{d}r} - \rho\,\frac{G\,M_\odot}{r^2} \tag{16.5}$$

where u is the radial expansion speed. The continuity equation reduces to

$$\frac{1}{r^2}\frac{\mathrm{d}(r^2\,\rho\,u)}{\mathrm{d}r} = 0 \tag{16.6}$$

In order to obtain a closed set of equations, we now need to adopt an equation of state for the corona, relating the pressure, p, and the density, ρ. For the sake of simplicity, we adopt the simplest conceivable equation of state, which corresponds to an *isothermal* corona. Thus, we have

$$p = \frac{2k_B\rho T}{m_p} \tag{16.7}$$

where T is a constant. Note that more realistic equations of state complicate the analysis, but do not significantly modify any of the physics results.

Equation 16.6 can be integrated to give

$$r^2\,\rho\,u = I \tag{16.8}$$

where I is a constant. The above expression simply states that the mass flux per unit solid angle, which takes the value I, is independent of the radius, r. Equation 16.5, Equation 16.7, and Equation 16.8 can be combined together to give

$$\frac{1}{u}\frac{du}{\mathrm{d}r}\left(u^2 - \frac{2k_BT}{m_p}\right) = \frac{4k_BT}{m_p\,r} - \frac{G\,M_\odot}{r^2} \tag{16.9}$$

Let us restrict our attention to coronal temperatures which satisfy

$$T < T_c \equiv \frac{G\,M_\odot\,m_p}{4k_B\,a}$$

where a is the radius of the base of the corona. T_c is the defined temperature where the right-hand side of Equation 16.9 is zero at $r = a$. For typical coronal parameters (see previous section), $T_c \simeq 5.8 \times 10^6$K, which is certainly greater than the temperature of the corona at $r = a$. For $T < T_c$, the right-hand side of Equation 16.9 is negative for $a < r < r_c$, where (???)

$$\frac{r_c}{a} = \frac{T_c}{T} \tag{16.10}$$

and positive for $r_c < r < \infty$. The right-hand side of Equation 16.9 is zero at $r = r_c$, implying that the left-hand side is also zero at this radius, which is usually termed the "critical radius". There are two ways in which the left-hand side of Equation 16.9 can be zero at the critical radius. Either

$$u^2(r_c) = u_c^2 \equiv \frac{2k_BT}{m_p} \tag{16.11}$$

or

$$\frac{du(r_c)}{\mathrm{d}r} = 0 \tag{16.12}$$

Note that u_c is the coronal sound speed.

As is easily demonstrated, if Equation 16.11 is satisfied then $du/\mathrm{d}r$ has the same sign for all r, and $u(r)$ is either a monotonically increasing, or a monotonically decreasing, function of r. On the other hand, if Equation 16.12 is satisfied then $u^2 - u_c^2$ has the same sign for all r, and $u(r)$ has an extremum close to $r = r_c$. The flow is either super-sonic for all r, or sub-sonic for all r. These possibilities lead to the existence of four classes of solutions to Equation 16.9, with the following properties:

1. $u(r)$ is sub-sonic throughout the domain $a < r < \infty$. $u(r)$ increases with r, attains a maximum value around $r = r_c$, and then decreases with r.
2. a unique solution for which $u(r)$ increases monotonically with r, and $u(r_c) = u_c$.
3. a unique solution for which $u(r)$ decreases monotonically with r, and $u(r_c) = u_c$.

4. $u(r)$ is super-sonic throughout the domain $a < r < \infty$. $u(r)$ decreases with r, attains a minimum value around $r = r_c$, and then increases with r.

These four classes of solutions are illustrated in **?@fig-parker_sol**.

Each of the classes of solutions described above fits a different set of boundary conditions at $r = a$ and $r \to \infty$. The physical acceptability of these solutions depends on these boundary conditions. For example, both Class 3 and Class 4 solutions can be ruled out as plausible models for the solar corona since they predict super-sonic flow at the base of the corona, which is not observed, and is also not consistent with a static solar photosphere. Class 1 and Class 2 solutions remain acceptable models for the solar corona on the basis of their properties around $r = a$, since they both predict sub-sonic flow in this region. However, the Class 1 and Class 2 solutions behave quite differently as $r \to \infty$, and the physical acceptability of these two classes hinges on this difference.

Equation 16.9 can be rearranged to give

$$\frac{du^2}{\mathrm{d}r}\left(1 - \frac{{u_c}^2}{u^2}\right) = \frac{4\,{u_c}^2}{r}\left(1 - \frac{r_c}{r}\right)$$

where Equation 16.10 and the definition of T_c have been used. The above expression can be integrated to give

$$\left(\frac{u}{u_c}\right)^2 - \ln\left(\frac{u}{u_c}\right)^2 = 4\,\ln r + 4\,\frac{r_c}{r} + C \tag{16.13}$$

where C is a constant of integration.

Let us consider the behaviour of Class 1 solutions in the limit $r \to \infty$. It is clear from **?@fig-parker_sol** that, for Class 1 solutions, u/u_c is less than unity and monotonically decreasing as $r \to \infty$. In the large-r limit, Equation 16.13 reduces to

$$\ln\frac{u}{u_c} \simeq -2\,\ln r$$

so that

$$u \propto \frac{1}{r^2}$$

It follows from Equation 16.8 that the coronal density, ρ, approaches a finite, constant value, ρ_∞, as $r \to \infty$. Thus, the Class 1 solutions yield a finite pressure,

$$p_\infty = \frac{2k_B\,\rho_\infty\,T}{m_p}$$

at large r, which cannot be matched to the much smaller pressure of the interstellar medium. Clearly, Class 1 solutions are unphysical.

Let us consider the behaviour of the Class 2 solution in the limit $r \to \infty$. It is clear from **?@fig-parker__sol** that, for the Class 2 solution, u/u_c is greater than unity and monotonically increasing as $r \to \infty$. In the large-r limit, Equation 16.13 reduces to

$$\left(\frac{u}{u_c}\right)^2 \simeq 4\,\ln r$$

so that

$$u \simeq 2\,u_c\,(\ln r)^{1/2}$$

It follows from Equation 16.8 that $\rho \to 0$ and $r \to \infty$. Thus, the Class 2 solution yields $p \to 0$ at large r, and can, therefore, be matched to the low pressure interstellar medium.

We conclude that the only solution to Equation 16.9 which is consistent with physical boundary conditions at $r = a$ and $r \to \infty$ is the Class 2 solution. This solution predicts that the solar corona expands radially outward at relatively modest, sub-sonic velocities close to the Sun, and gradually accelerates to super-sonic velocities as it moves further away from the Sun. Parker termed this continuous, super-sonic expansion of the corona the *solar wind*.

Note that while the velocity prediction accords well with satellite observations, the Parker model's prediction for the density of the solar wind at the Earth is significantly too high compared to satellite observations. Consequently, there have been many further developments of this model. In particular, the unrealistic assumption that the solar wind plasma is isothermal has been relaxed, and two-fluid effects have been incorporated into the analysis.

16.2 Parker Spiral

The *interplanetary magnetic field* (IMF), also called *Parker spiral*, is the component of the solar magnetic field that is dragged out from the solar corona by the solar wind flow to fill the Solar System. Depending on the polarity of the photospheric footpoint, the heliospheric magnetic field spirals inward or outward; the magnetic field follows the same shape of spiral in the northern and southern parts of the heliosphere, but with opposite field direction. These two magnetic domains are separated by a current sheet (an electric current that is confined to a curved plane). This heliospheric current sheet has a shape similar to a twirled ballerina skirt (Figure 16.1), and changes in shape through the solar cycle as the Sun's magnetic field reverses about every 11 years.

```
KeyNotes.plot_IMF_3D()
```

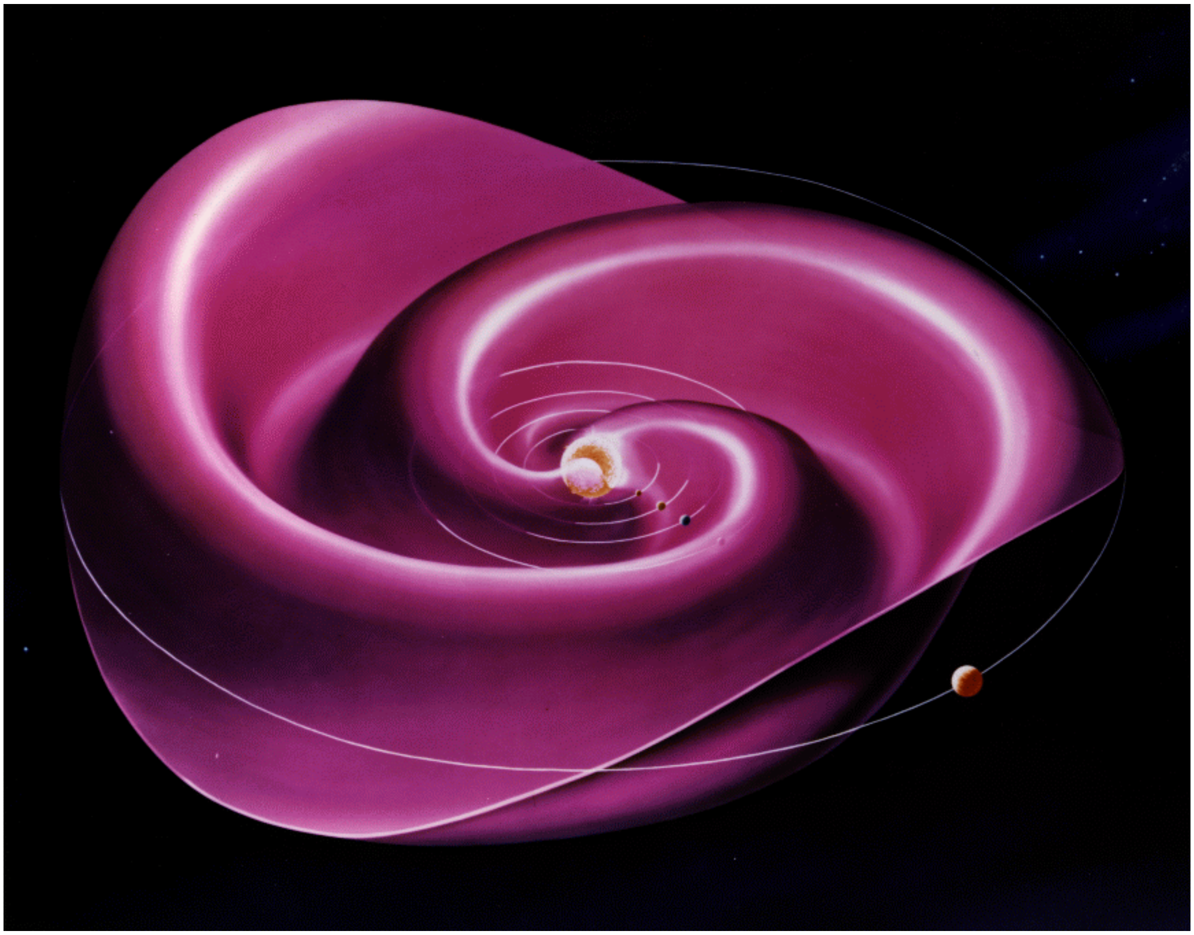

Figure 16.1: The heliospheric current sheet is a three-dimensional form of a Parker spiral that results from the influence of the Sun's rotating magnetic field on the plasma in the interplanetary medium.

16.3 Anisotropy in the Solar Wind

Observationally, Pioneer 6 showed that the ion temperature anisotropy in the solar wind at 1AU generally has $T_\parallel > T_\perp$, together with other two interesting discoveries:

1. high fluctuations of flow velocity outside the solar ecliptic plane;

2. anisotropic ion thermal distribution ($T_\parallel/T_\perp \sim [2,5]$);
3. presence of a 3rd species, helium, from charge-to-mass ratio analysis of the angular and energy distributions.

It may possibly be explained by the conservation of the 1st adiabatic invariant Scarf, Wolfe, and Silva (1967).

The magnetic moment $\mu = mv_\perp^2/2B$ is conserved as the collisionless solar wind flows outward from the sun. Near the solar equator the mean field magnitude declines with

$$B_r(r) \simeq B_r(r_0)\Big(\frac{r_0}{r}\Big)^2$$

and

$$B_\phi(r) \simeq \frac{\Omega_r}{u(r)} B_r(r)$$

from the Parker spiral solar wind model and $\Omega_r = 2.94\times 10^{-6}$ rad/s being the angular frequency of the rotation of the sun.

The adiabatic equation in the perpendicular direction indicates that the perpendicular thermal energy $\langle mv_\perp^2/2\rangle = k_B T_\perp$ declines with B. Assuming in the rest frame the distribution function is a bi-Maxwellian of the form

$$f(v) = \Big(\frac{m}{2\pi}\Big)^{3/2} \frac{1}{k_B T_\perp (k_B T_\parallel)^{1/2}} \exp\Big(-\frac{mv_\perp^2}{2k_B T_\perp} - \frac{mv_\parallel^2}{2k_B T_\parallel}\Big)$$

The conservation of the total thermal energy

$$W = \int \mathrm{d}^3 v \frac{mv^2}{2} f(v)$$

yields

$$W = k_B T_\perp + k_B T_\parallel/2 = \text{const.}$$

These allows us to evaluate the variations in $T_\perp$ and $T_\parallel$ originating from isotropic distribution on the surface of the sun. Starting from $T_\perp(0.3AU) = T_\parallel(0.3AU) \simeq 1.3\times 10^5$K, the predicted anisotropy $T_\parallel/T_\perp$ at Earth can go beyond 20! Therefore, in fact, the reasonable question to ask is why the actual solar wind anisotropy factor is so small. In simulations, we always apply isotropic distribution in the upstream solar wind condition, which is primarily due to the fact that we are mostly using fluid models (i.e. MHD) for global magnetosphere-solar wind interactions. For kinetic models, *we need more realistic distribution setups*.

Well, we know now when pressure anisotropy develops, two kinds of plasma instabilities can be triggered: firehose when $T_\perp < T_\parallel$ and mirror when $T_\perp > T_\parallel$. Further studies require kinetic theory to describe their behaviors.

On the other hand, the opposite case, $T_\perp > T_\parallel$, is also observed and believed to be related to local ion heating by macroscale compressions (e.g. high/low speed streams interaction) or plasma instabilities (Bame et al. 1975).

The mirror instability criterion as an additional relation to determine the pressure anisotropy downstream of the shock from the book Plasma instabilities and nonlinear effects by Hasegawa 1975,

$$1 + \sum_{\text{species}} \beta_\perp \left(1 - \frac{\beta_\perp}{\beta_\parallel}\right) < 0$$

16.4 Switchbacks

When Parker Solar Probe (PSP) sent back the first observations from its voyage to the Sun, scientists found signs of a wild ocean of currents and waves quite unlike the near-Earth space much closer to our planet. This ocean was spiked with what became known as *switchbacks*: rapid flips in the Sun's magnetic field that reversed direction like a zig-zagging mountain road.

Our current understanding is that switchbacks may be generated from either magnetic reconnection, turbulence, or plasma velocity shears. There are models for each, but they all requires more testing from future PSP data.

See more in Switchbacks Science: Explaining Parker Solar Probe's Magnetic Puzzle.

16.5 MHD Description

Because of the presence of a strong magnetic field carried by the wind, low-frequency fluctuations in the solar wind are usually described with MHD. However, due to some peculiar characteristics, the solar wind turbulence contains some features hardly classified within a general theoretical framework. (Tu and Marsch 1995) presents a thorough review of the turbulent phenomena in the solar wind from observations in the ecliptic. In the 1990s, with the launch of the Ulysses spacecraft, investigations have been extended to the high-latitude regions of the heliosphere.

Ruelle and Takens (1971) who showed that a strange attractor in the phase space of the system is the best model for the birth of turbulence. This gave a strong impulse to the investigation of the phenomenology of turbulence from the point of view of dynamical systems. Turbulence in the solar wind has been used as a big wind tunnel to investigate scaling laws of turbulent fluctuations, multifractals models, etc. Therefore, the solar wind can be seen as a very big laboratory where fully developed turbulence can be investigated not only per se, rather as far as basic theoretical aspects are concerned.

17 Shock

Consider a subsonic disturbance moving through a conventional neutral fluid. As is well-known, sound waves propagating ahead of the disturbance give advance warning of its arrival, and, thereby, allow the response of the fluid to be both smooth and adiabatic. Now, consider a supersonic disturbance. In this case, sound waves are unable to propagate ahead of the disturbance, and so there is no advance warning of its arrival, and, consequently, the fluid response is *sharp* and *non-adiabatic*. This type of response is generally known as a *shock*.

In space, shock is a dissipative structure in which the kinetic and magnetic energy of a directed plasma flow is partly transferred to heating of the plasma. The dissipation does not take place, however, by means of particle collisions for a shock in space. Collisionless shocks can be divided into *super-* and *sub-critical*: the *critical Mach number* M_A^* is defined by equating the normal component of the downstream flow velocity in the shock frame to the sound speed. (In aerodynamics, the critical Mach number M^* of an aircraft is the *lowest* Mach number at which the airflow over some point of the aircraft reaches the speed of sound, but does not exceed it.) Above the critical Mach number, resistivity alone is unable to provide all the dissipation needed for the required *Rankine-Hugoniot* shock jump. Supercritical shocks are important because they usually produce much greater ion heating than subcritical shocks. In contrast to sub-critical shocks, resistivity in super-critical shocks cannot provide all the necessary dissipation for a shock transition according to the *Rankine–Hugoniot* relationships. Thus, other processes like wave-particle interactions provide the dissipation required for supercritical shock formation. This is the reason why they are able to accelerate particles to high energies. Note that most of the time we have supercritical bow shock at Earth, i.e. the sound speed in the magnetosheath is larger than the bulk speed.

Let us investigate shocks first in MHD fluids. Since information in such fluids is carried via three different waves – namely, *fast* or *compressional-Alfvén waves*, *intermediate* or *shear-Alfvén waves*, and *slow* or *magnetosonic waves* – we might expect MHD fluids to support three different types of shock, corresponding to disturbances traveling faster than each of the aforementioned waves.

In general, a shock propagating through an MHD fluid produces a significant difference in plasma properties on either side of the shock front. The thickness of the front is determined by a balance between convective and dissipative effects. However, dissipative effects in high temperature plasmas are only comparable to convective effects when the spatial gradients in plasma variables become extremely large. Hence, MHD shocks in such plasmas tend to be *extremely narrow*, and are well-approximated as *discontinuous* changes in plasma parameters.

The MHD equations, and Maxwell's equations, can be integrated across a shock to give a set of jump conditions which relate plasma properties on each side of the shock front. If the shock is sufficiently narrow then these relations become independent of its detailed structure. Let us derive the jump conditions for a narrow, planar, steady-state, MHD shock.

17.1 MHD Theory

The conserved form of MHD equations can be written as:

$$\begin{aligned}
\nabla \cdot \mathbf{B} &= 0 \\
\frac{\partial \mathbf{B}}{\partial t} - \nabla \times (\mathbf{V} \times \mathbf{B}) &= 0 \\
\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho\, \mathbf{V}) &= 0 \\
\frac{\partial (\rho\, \mathbf{V})}{\partial t} + \nabla \cdot \mathbf{T} &= 0 \\
\frac{\partial U}{\partial t} + \nabla \cdot \mathbf{u} &= 0
\end{aligned} \tag{17.1}$$

where

$$\mathbf{T} = \rho\, \mathbf{V}\, \mathbf{V} + \left(p + \frac{B^2}{2\mu_0}\right) \mathbf{I} - \frac{\mathbf{B}\mathbf{B}}{\mu_0}$$

is the total (i.e., including electromagnetic, as well as plasma, contributions) stress tensor, $\mathbf{I}$ the identity tensor,

$$U = \frac{1}{2}\rho V^2 + \frac{p}{\gamma - 1} + \frac{B^2}{2\mu_0}$$

the total energy density, and

$$\mathbf{u} = \left(\frac{1}{2}\rho V^2 + \frac{\gamma}{\gamma - 1} p\right) \mathbf{V} + \frac{\mathbf{B} \times (\mathbf{V} \times \mathbf{B})}{\mu_0}$$

the total energy flux density.

Let us move into the rest frame of the shock. For a 1D shock, suppose that the shock front coincides with the y-z plane. Furthermore, let the regions of the plasma upstream and downstream of the shock, which are termed regions 1 and 2, respectively, be *spatially uniform* and *time-static*, i.e. $\partial/\partial t = \partial/\partial x = \partial/\partial y = 0$. Moreover, $\partial/\partial x = 0$, except in the immediate vicinity of the shock. Finally, let the velocity and magnetic fields upstream and downstream of the shock all lie in the x-y plane. The situation under discussion is illustrated in the figure below.

Here, ρ_1, p_1, $\mathbf{V}_1$, and $\mathbf{B}_1$ are the downstream mass density, pressure, velocity, and magnetic field, respectively, whereas ρ_2, p_2, $\mathbf{V}_2$, and $\mathbf{B}_2$ are the corresponding upstream quantities.

The basic RH relations are listed in MHD shocks. In the immediate vicinity of the planar shock, Equation 17.1 reduce to

$$\begin{aligned}
\frac{\mathrm{d}B_x}{\mathrm{d}x} &= 0\\
\frac{\mathrm{d}}{\mathrm{d}x}(V_x\,B_y - V_y\,B_x) &= 0\\
\frac{\mathrm{d}(\rho\,V_x)}{\mathrm{d}x} &= 0\\
\frac{\mathrm{d}T_{xx}}{\mathrm{d}x} &= 0\\
\frac{\mathrm{d}T_{xy}}{\mathrm{d}x} &= 0\\
\frac{\mathrm{d}u_x}{\mathrm{d}x} &= 0
\end{aligned}$$

Integration across the shock yields the desired jump conditions:

$$\begin{aligned}
\lfloor B_x\rceil &= 0\\
\lfloor V_x\,B_y - V_y\,B_x\rceil &= 0\\
\lfloor \rho\,V_x\rceil &= 0\\
\lfloor \rho\,V_x^2 + p + B_y^2/2\mu_0\rceil &= 0\\
\lfloor \rho\,V_x\,V_y - B_x\,B_y/\mu_0\rceil &= 0\\
\left\lfloor \frac{1}{2}\,\rho\,V^2\,V_x + \frac{\gamma}{\gamma-1}\,p\,V_x + \frac{B_y\,(V_x\,B_y - V_y\,B_x)}{\mu_0}\right\rceil &= 0
\end{aligned} \tag{17.2}$$

where $\lfloor A\rceil = A_2 - A_1$ is the difference across the shock. These relations are often called the Rankine-Hugoniot relations for MHD. There are 6 scalar equations and 12 scalar variables all together. Assuming that all of the upstream plasma parameters are known, there are 6 unknown parameters in the problem–namely, $B_{x\,2}$, $B_{y\,2}$, $V_{x\,2}$, $V_{y\,2}$, ρ_2, and p_2. These 6 unknowns are fully determined by the six jump conditions. If we loose the planar assumption, then we typically write the x-component as the normal component (v_n, B_n) and the combined $y-$ and $z-components$ as the tangential component (v_t, B_t). Luckily this is still deterministic. However, the general case is very complicated!

The categories of the solution of Equation 17.2 are shown in the following table. The $\pm$ signs denote the changes of the downstream compared with the upstream (+ means increase, $-$ means decrease). By definition shocks are transition layers across which there is a transport of particles, whereas discontinuities are transition layers across which there is no particle transport.

Type	Particle Transport	ρ	$\mathbf{v}$	p	$\mathbf{B}$	T
Contact	No	$\pm$	continuous	continuous	continuous	$\pm$
Tangential	No	$\pm$	continuous	$\pm$	$B_n = 0$	$\pm$
Slow	Yes	+	-	+	B_t -	+
Intermediate	Yes	continuous	$\pm$	$\pm$	rotation?	$\pm$
Fast	Yes	+	-	+	B_t +	+

1. The Earth's magnetopause is generally a tangential discontinuity. When there is no flux rope been generated, the magnetopause can be treated as the surface of pressure balance between magnetic pressure, ram pressure and thermal pressure. However, when reconnection triggers flux rope generation, it may become a rotational discontinuity (TO BE CONFIRMED!).
2. A special case is called *rotational discontinuity*. All are isentropic. Their existence are still a matter of debate.
3. The Earth's bow shock is a fast mode shock.

17.1.1 Coplanarity

Knowing that for the shock $v_n \neq 0$ and $\lfloor v_n \rceil \neq 0$, we can eliminate $\lfloor v_t \rceil$ from the RH relations listed on Wiki and obtain

$$\lfloor v_n \mathbf{B}_t \rceil = \frac{B_n^2}{\mu_0 (something)} \lfloor \mathbf{B}_t \rceil$$

Hence the cross product of the left with the right hand side must vanish:

$$\begin{aligned} \lfloor \mathbf{B}_t \rceil \times \lfloor v_n \mathbf{B}_t \rceil = 0 \\ (v_{n1} - v_{n2})(\mathbf{B}_{t1} \times \mathbf{B}_{t2}) = 0 \end{aligned} \tag{17.3}$$

The resulting coplanarity theorem implies that the magnetic field across the shock has a 2-D geometry: upstream and dowstream tangential fields are parallel to each other and coplanar with the shock normal $\hat{n}$.

17.1.2 Parallel Shock

The first special case is the so-called parallel shock in which both the upstream and downstream plasma flows are parallel to the magnetic field, as well as perpendicular to the shock front. In other words,

$$\begin{aligned} \mathbf{V}_1 &= (V_1,\, 0,\, 0), \quad \mathbf{V}_2 = (V_2,\, 0,\, 0) \\ \mathbf{B}_1 &= (B_1,\, 0,\, 0), \quad \mathbf{B}_2 = (B_2,\, 0,\, 0) \end{aligned} \tag{17.4}$$

Substitution into Equation 17.2 yields

$$\begin{aligned} \frac{B_2}{B_1} &= 1 \\ \frac{\rho_2}{\rho_1} &= r \\ \frac{v_2}{v_1} &= r^{-1} \\ \frac{p_2}{p_1} &= R \end{aligned} \tag{17.5}$$

with

$$\begin{aligned} r &= \frac{(\gamma+1)M_1^2}{2+(\gamma-1)M_1^2} \\ R &= 1+\gamma M_1^2(1-r^{-1}) = \frac{(\gamma+1)r-(\gamma-1)}{(\gamma+1)-(\gamma-1)r} \end{aligned} \tag{17.6}$$

Here, $M_1 = v_1/v_{s1}$, where $v_{s1} = \sqrt{(\gamma p_1/\rho_1)}$ is the upstream sound speed. Thus, the upstream flow is supersonic if $M_1 > 1$, and subsonic if $M_1 < 1$. Incidentally, as is clear from the above expressions, a parallel shock is unaffected by the presence of a magnetic field. In fact, this type of shock is identical to that which occurs in neutral fluids, and is, therefore, usually called a *hydrodynamic shock*.

It is easily seen from Equation 17.4 that there is no shock (i.e., no jump in plasma parameters across the shock front) when the upstream flow is exactly sonic: i.e., when $M_1 = 1$. In other words, $r = R = 1$ when $M_1 = 1$. However, if $M_1 \neq 1$ then the upstream and downstream plasma parameters become different (i.e., $r \neq 1$, $R \neq 1$) and a true shock develops. In fact, it is easily demonstrated that

$$
\begin{aligned}
\frac{\gamma-1}{\gamma+1} \leq r \leq \frac{\gamma+1}{\gamma-1}\\
0 \leq R \leq \infty\\
\frac{\gamma-1}{2\,\gamma} \leq M_1^2 \leq \infty
\end{aligned} \tag{17.7}
$$

Note that the upper and lower limits in the above inequalities are all attained simultaneously.

The previous discussion seems to imply that a parallel shock can be either compressive (i.e., $r > 1$) or expansive (i.e., $r < 1$). Is there a preferential direction across the shock? In other words, can we tell the upstream and the downstream? Yes, with the additional physics principle of the *second law of thermodynamics*. This law states that the *entropy* of a closed system can spontaneously increase, but can never spontaneously decrease. Now, in general, the entropy per particle is different on either side of a hydrodynamic shock front. Accordingly, the second law of thermodynamics mandates that the downstream entropy must *exceed* the upstream entropy, so as to ensure that the shock generates a net increase, rather than a net decrease, in the overall entropy of the system, as the plasma flows through it.

The (suitably normalized) entropy per particle of an ideal plasma takes the form

$$
S = \ln\left(\frac{p}{\rho^\gamma}\right)
$$

Hence, the difference between the upstream and downstream entropies is

$$
\lfloor S\rceil = \ln R - \gamma\,\ln r
$$

Now, using Equation 17.6,

$$
r\frac{\mathrm{d}\lfloor S\rceil}{dr} = \frac{r}{R}\frac{dR}{dr} - \gamma = \frac{\gamma(\gamma^2-1)(r-1)^2}{[(\gamma+1)r-(\gamma-1)][(\gamma+1)-(\gamma-1)r]}
$$

Furthermore, it is easily seen from Equation 17.7 that $\mathrm{d}\lfloor S\rceil/dr \geq 0$ in all situations of physical interest. However, $\lfloor S\rceil = 0$ when $r = 1$, since, in this case, there is no discontinuity in plasma parameters across the shock front. We conclude that $\lfloor S\rceil < 0$ for $r < 1$, and $\lfloor S\rceil > 0$ for $r > 1$. It follows that the second law of thermodynamics requires hydrodynamic shocks to be *compressive*: i.e., $r \equiv \rho_2/\rho_1 > 1$. In other words, the plasma density must always *increase* when a shock front is crossed in the direction of the relative plasma flow. It turns out that this is a general rule which applies to all three types of MHD shock. In the shock rest frame, the shock is associated with an irreversible (since the entropy suddenly increases) transition from supersonic to subsonic flow.

The upstream Mach number, M_1, is a good measure of shock strength: i.e., if $M_1 = 1$ then there is no shock, if $M_1 - 1 \ll 1$ then the shock is weak, and if $M_1 \gg 1$ then the shock is strong. We can define an analogous downstream Mach number, $M_2 = V_2/(\gamma\, p_2/\rho_2)^{1/2}$. It is easily demonstrated from the jump conditions that if $M_1 > 1$ then $M_2 < 1$. In other words, in the shock rest frame, the shock is associated with an irreversible (since the entropy suddenly increases) transition from supersonic to subsonic flow. Note that $r \equiv \rho_2/\rho_1 \to (\gamma+1)/(\gamma-1)$, whereas $R \equiv p_2/p_1 \to \infty$, in the limit $M_1 \to \infty$. In other words, as the shock strength increases, the compression ratio, r, asymptotes to a finite value, whereas the pressure ratio, P, increases without limit. For a conventional plasma with $\gamma = 5/3$, the limiting value of the compression ratio is 4: i.e., the downstream density can never be more than four times the upstream density. We conclude that, in the strong shock limit, $M_1 \gg 1$, the large jump in the plasma pressure across the shock front must be predominately a consequence of a large jump in the plasma *temperature*, rather than the plasma *density*. In fact, the definitions of r and R imply that

$$\frac{T_2}{T_1} \equiv \frac{R}{r} \to \frac{2\gamma(\gamma-1)M_1^2}{(\gamma+1)^2} \gg 1$$

as $M_1 \to \infty$. Thus, a strong parallel, or hydrodynamic, shock is associated with intense plasma heating.

As we have seen, the condition for the existence of a hydrodynamic shock is $M_1 > 1$, or $V_1 > V_{S\,1}$. In other words, in the shock frame, the upstream plasma velocity, V_1, must be supersonic. However, by Galilean invariance, V_1 can also be interpreted as the propagation velocity of the shock through an initially stationary plasma. It follows that, in a stationary plasma, a parallel, or hydrodynamic, shock propagates along the magnetic field with a supersonic velocity.

17.1.3 Perpendicular Shock

The second special case is the so-called *perpendicular shock* in which both the upstream and downstream plasma flows are perpendicular to the magnetic field, as well as the shock front. In other words,

$$\begin{aligned} \mathbf{V}_1 &= (V_1,\, 0,\, 0), \quad \mathbf{V}_2 = (V_2,\, 0,\, 0) \\ \mathbf{B}_1 &= (0,\, B_1,\, 0), \quad \mathbf{B}_2 = (0,\, B_2,\, 0) \end{aligned} \tag{17.8}$$

Substitution into Equation 17.2 yields

$$
\begin{aligned}
\frac{B_2}{B_1} &= r \\
\frac{\rho_2}{\rho_1} &= r \\
\frac{v_2}{v_1} &= r^{-1} \\
\frac{p_2}{p_1} &= R
\end{aligned}
\tag{17.9}
$$

where

$$R = 1 + \gamma\, M_1^2\,(1 - r^{-1}) + \beta_1^{-1}\,(1 - r^2) \tag{17.10}$$

and r is a real positive root of the quadratic

$$F(r) = 2\,(2-\gamma)\,r^2 + \gamma\,[2\,(1+\beta_1) + (\gamma-1)\beta_1 M_a^2]r - \gamma\,(\gamma+1)\,\beta_1\,M_1^2 = 0 \tag{17.11}$$

Here, $\beta_a = 2\mu_0 p_1/B_1^2$.

Now, if r_1 and r_2 are the two roots of Equation 17.11 then

$$r_1 r_2 = -\frac{\gamma(\gamma+1)\beta_1 M_1^2}{2(2-\gamma)}$$

Assuming that $\gamma < 2$, we conclude that one of the roots is negative, and, hence, that Equation 17.11 only possesses one physical solution: i.e., there is only one type of MHD shock which is consistent with Equation 17.8. Now, it is easily demonstrated that $F(0) < 0$ and $F(\gamma+1/\gamma-1) > 0$. Hence, the physical root lies between $r = 0$ and $r = (\gamma+1)/(\gamma-1)$.

Using similar analysis to that employed in the previous subsection, it is easily demonstrated that the second law of thermodynamics requires a perpendicular shock to be compressive: i.e., $r > 1$. It follows that a physical solution is only obtained when $F(1) < 0$, which reduces to

$$M_1^2 > 1 + \frac{2}{\gamma\,\beta_1}$$

This condition can also be written

$$\mathbf{v}_1^2 > \mathbf{v}_{s1}^2 + \mathbf{v}_{A1}^2$$

where $v_{A1} = B_1/\sqrt{(\mu_0 \rho_1)}$ is the upstream Alfvén speed. $v_{+1} = (v_{S\,1}^2 + v_{A\,1}^2)^{1/2}$ can be recognized as the velocity of a fast wave propagating perpendicular to the magnetic field (Section 7.6.4). Thus, the condition for the existence of a perpendicular shock is that the relative upstream plasma velocity must be *greater* than the upstream fast wave velocity. Incidentally, it is easily demonstrated that if this is the case then the downstream plasma velocity is less than the downstream fast wave velocity. We can also deduce that, in a stationary plasma, a perpendicular shock propagates across the magnetic field with a velocity which exceeds the fast wave velocity.

In the strong shock limit, $M_1 \gg 1$, Equation 17.10 and Equation 17.11 become identical to Equation 17.6. Hence, a strong perpendicular shock is very similar to a strong hydrodynamic shock (except that the former shock propagates perpendicular, whereas the latter shock propagates parallel, to the magnetic field). In particular, just like a hydrodynamic shock, a perpendicular shock cannot compress the density by more than a factor $(\gamma+1)/(\gamma-1)$. However, according to Equation 17.9, a perpendicular shock compresses the magnetic field by the same factor that it compresses the plasma density. It follows that there is also an upper limit to the factor by which a perpendicular shock can compress the magnetic field.

17.1.4 Oblique Shock

Let us now consider the general case in which the plasma velocities and the magnetic fields on each side of the shock are neither parallel nor perpendicular to the shock front. It is convenient to transform into the so-called *de Hoffmann-Teller frame* in which $|\mathbf{v}_1 \times \mathbf{B}_1| = 0$, or

$$v_{x1} B_{y1} - v_{y1} B_{x1} = 0 \tag{17.12}$$

In other words, it is convenient to transform to a frame which moves at the local $\mathbf{E}\times\mathbf{B}$ velocity of the plasma.[1] It immediately follows from the jump condition of the magnetic convection equation that

$$v_{x2} B_{y2} - v_{y2} B_{x2} = 0 \tag{17.13}$$

or $|\mathbf{v}_2 \times \mathbf{B}_2| = 0$. Thus, in the de Hoffmann-Teller frame, the upstream plasma flow is parallel to the upstream magnetic field, and the downstream plasma flow is also parallel to the downstream magnetic field. Furthermore, the magnetic contribution to the jump condition Equation 17.2 (last eq.) becomes identically zero, which is a considerable simplification.

Equation 17.12 and Equation 17.13 can be combined with the general jump conditions Equation 17.2 to give

[1] for a parallel shock, the HT frame can be just itself; for a perpendicular shock, the HT frame can be chosen by shifting the velocity by $\mathbf{v}_1$ s.t. $\mathbf{v}_1 = 0$.

$$
\begin{aligned}
\frac{\rho_2}{\rho_1} &= r \\
\frac{B_{x\,2}}{B_{x\,1}} &= 1 \\
\frac{B_{y\,2}}{B_{y\,1}} &= r\left(\frac{v_1^2-\cos^2\theta_1\,v_{A\,1}^2}{v_1^2-r\,\cos^2\theta_1\,v_{A\,1}^2}\right) \\
\frac{v_{x\,2}}{v_{x\,1}} &= \frac{1}{r} \\
\frac{v_{y\,2}}{v_{y\,1}} &= \frac{v_1^2-\cos^2\theta_1\,v_{A\,1}^2}{v_1^2-r\,\cos^2\theta_1\,v_{A\,1}^2} \\
\frac{p_2}{p_1} &= 1+\frac{\gamma v_1^2(r-1)}{v_{s1}^2 r}\left[1-\frac{r v_{A1}^2[(r+1)v_1^2-2r v_{A1}^2\cos^2\theta_1]}{2(v_1^2-r v_{A1}^2\cos^2\theta_1)^2}\right]
\end{aligned}
\tag{17.14}
$$

where $v_1 = v_{x1} = v_1 \cos\theta_1$ is the component of the upstream velocity normal to the shock front, and $theta_1$ is the angle subtended between the upstream plasma flow and the shock front normal. Finally, given the compression ratio, r, the square of the normal upstream velocity, v_1^2, is a real root of a cubic equation known as the *shock adiabatic*:

$$
\begin{aligned}
0 =&(v_1^2-r\cos^2\theta_1 v_{A1}^2)^2\{[(\gamma+1)-(\gamma-1)r]v_1^2-2r v_{s1}^2\} \\
&r\sin^2\theta_1 v_1^2 v_{A1}^2\{[\gamma+(2-\gamma)r]v_1^2-[(\gamma+1)-(\gamma-1)r]r\cos^2\theta_1 v_{A1}^2\}
\end{aligned}
\tag{17.15}
$$

As before, the second law of thermodynamics mandates that $r > 1$.

Weak shock limit

Let us first consider the weak shock limit $r \to 1$. In this case, it is easily seen that the three roots of the shock adiabatic reduce to the slow, intermediate (or Shear-Alfvén), and fast waves, respectively, propagating in the normal direction to the shock front:

$$
\begin{aligned}
v_1^2 &= v_{-\,1}^2 \equiv \frac{v_{A\,1}^2+v_{S\,1}^2-[(v_{A\,1}+v_{S\,1})^2-4\,\cos^2\theta_1\,v_{S\,1}^2\,v_{A\,1}^2]^{1/2}}{2} \\
v_1^2 &= \cos^2\theta_1\,v_{A\,1}^2 \\
v_1^2 &= v_{+\,1}^2 \equiv \frac{v_{A\,1}^2+v_{S\,1}^2+[(v_{A\,1}+v_{S\,1})^2-4\,\cos^2\theta_1\,v_{S\,1}^2\,v_{A\,1}^2]^{1/2}}{2}
\end{aligned}
$$

We conclude that slow, intermediate, and fast MHD shocks degenerate into the associated MHD waves in the limit of small shock amplitude. Conversely, we can think of the various MHD shocks as *nonlinear* versions of the associated MHD waves. It is easily demonstrated that

$$
v_{+1} > \cos\theta_1 v_{A1} > v_{-1}
$$

In other words, a fast wave travels faster than an intermediate wave, which travels faster than a slow wave. It is reasonable to suppose that the same is true of the associated MHD shocks, at least at relatively low shock strength. It follows from Equation 17.14 that $B_{y2} > B_{y1}$ for a fast shock, whereas $B_{y\,2} < B_{y\,1}$ for a slow shock. For the case of an intermediate shock, we can show, after a little algebra, that $B_{y\,2} \rightarrow -B_{y\,1}$ in the limit $r \rightarrow 1$. We can conclude that (in the de Hoffmann-Teller frame) fast shocks refract the magnetic field and plasma flow (recall that they are parallel in our adopted frame of the reference) *away* from the normal to the shock front, whereas slow shocks refract these quantities *toward* the normal. Moreover, the tangential magnetic field and plasma flow generally *reverse* across an intermediate shock front. This is illustrated in **?@fig-mhd__shock__rest__frame**.

When r is slightly larger than unity it is easily demonstrated that the conditions for the existence of a slow, intermediate, and fast shock are $v_1 > V_{-\,1}$, $v_1 > \cos\theta_1\, V_{A\,1}$, and $v_1 > V_{+\,1}$, respectively.

Strong shock limit

Let us now consider the strong shock limit, $v_1^2 \gg 1$. In this case, the shock adiabatic yields $r \rightarrow r_m = (\gamma+1)/(\gamma-1)$, and

$$v_1^2 \simeq \frac{r_m}{\gamma-1}\,\frac{2v_{S1}\sin^2\theta_1\,[\gamma+(2-\gamma)\,r_m]\,v_{A1}^2}{r_m - r}$$

There are no other real roots. The above root is clearly a type of fast shock. The fact that there is only one real root suggests that there exists a critical shock strength above which the slow and intermediate shock solutions cease to exist. (In fact, they merge and annihilate one another.) In other words, there is a limit to the strength of a slow or an intermediate shock. On the other hand, there is no limit to the strength of a fast shock. Note, however, that the plasma density and tangential magnetic field cannot be compressed by more than a factor $(\gamma+1)/(\gamma-1)$ by any type of MHD shock.

$\theta_1 = 0$

Consider the special case $\theta_1 = 0$ in which both the plasma flow and the magnetic field are normal to the shock front. In this case, the three roots of the shock adiabatic are

$$\begin{aligned} v_1^2 &= \frac{2r\,v_{S1}^2}{(\gamma+1)-(\gamma-1)\,r} \\ v_1^2 &= r\,v_{A1}^2 \\ v_1^2 &= r\,v_{A1}^2 \end{aligned}$$

We recognize the first of these roots as the hydrodynamic shock discussed in Section 17.1.2. This shock is classified as a slow shock when $V_{S\,1} < v_{A\,1}$, and as a fast shock when $V_{S\,1} > v_{A\,1}$. The other two roots are identical, and correspond to shocks which propagate at the velocity

$v_1 = \sqrt{r}\, v_{A\,1}$ and "switch-on" the tangential components of the plasma flow and the magnetic field: it can be seen from Equation 17.14 that $v_{y\,1} = B_{y\,1} = 0$ whilst $v_{y\,2} \neq 0$ and $B_{y\,2} \neq 0$ for these types of shock.

There we have "switch-on" and "switch-off" shocks which refer to the generation and elimination of tangential components of the plasma flow and the magnetic field. Incidentally, it is also possible to have a "switch-off" shock which eliminates the tangential components of the plasma flow and the magnetic field. According to Equation 17.14, such a shock propagates at the velocity $v_1 = \cos\theta_1\, v_{A\,1}$. Switch-on and switch-off shocks are illustrated in **?@fig-shock_switch_on_off**.

$\theta_1 = \pi/2$

Consider another special case $\theta_1 = \pi/2$. As is easily demonstrated, the three roots of the shock adiabatic are

$$
\begin{aligned}
v_1^2 &= r\left(\frac{2v_{S1}^2 + [\gamma + (2-\gamma)\, r]\, v_{A1}^2}{(\gamma+1) - (\gamma-1)\, r}\right) \\
v_1^2 &= 0 \\
v_1^2 &= 0
\end{aligned}
$$

The first of these roots is clearly a fast shock, and is identical to the perpendicular shock discussed in Section 17.1.3, except that there is no plasma flow across the shock front in this case. (IS IT BECAUSE OF THE HT FRAME?) The fact that the two other roots are zero indicates that, like the corresponding MHD waves, slow and intermediate MHD shocks do not propagate perpendicular to the magnetic field.

MHD shocks have been observed in a large variety of situations. For instance, shocks are known to be formed by supernova explosions, by strong stellar winds, by solar flares, and by the solar wind upstream of planetary magnetospheres.

17.2 Double Adiabatic Theory

The classical approach by Chew, Goldberger, and Low (Chew, Goldberger, and Low 1956) utilizes the MHD framework by assuming isotropic distributions parallel and perpendicular to the magnetic field, which results in scalar pressures on the two sides of the shock. This is now known as the CGL theory.

When we shift to the MHD with anisotropic pressure tensor

$$
P_{ij} = p_\perp \delta_{ij} + (p_\parallel - p_\perp) B_i B_j / B^2
$$

where $p_\perp$ and $p_\parallel$ are the pressures perpendicular and parallel w.r.t. the magnetic field, respectively. For the strong magnetic field approximation, the two pressures are related to the plasma density and the magnetic field strength by two adiabatic equations,

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\Big(\frac{p_\parallel B^2}{\rho^3}\Big) &= 0 \\
\frac{\mathrm{d}}{\mathrm{d}t}\Big(\frac{p_\perp}{\rho B}\Big) &= 0
\end{aligned}
$$

This is also known as the double adiabatic theory, which is also what many people remember to be the key conclusion from the CGL theory. (These are constants at a fixed location in time: it is *not* correct to apply these across the shock!) Here I want to emphasize the meaning of *adiabatic* again: this assumes zero heat flux. If the system is not adiabatic, the conservation of these two quantities related to the parallel and perpendicular pressure is no longer valid, and additional terms may come into play such as the stochastic heating.

The general jump conditions for discontinuities in a collisionless anisotropic magnetoplasma in the CGL approximation were derived by (Abraham-Shrauner 1967).

The general jump conditions for an anisotropic plasma are given by (Hudson 1970)

$$
\begin{aligned}
\lfloor \rho v_n \rceil &= 0 \\
\lfloor v_n \mathbf{B}_t - \mathbf{v}_t B_n \rceil &= 0 \\
\lfloor p_\perp + (p_\parallel - p_\perp)\frac{B_n^2}{B^2} + \frac{B_t^2}{8\pi} + \rho v_n^2 \rceil &= 0 \\
\lfloor \frac{B_n \mathbf{B}_t}{4\pi}\Big(\frac{4\pi(p_\parallel - p_\perp)}{B^2} - 1\Big) + \rho v_n \mathbf{v}_t \rceil &= 0 \\
\lfloor \rho v_n \Big(\frac{\epsilon}{\rho} + \frac{v^2}{2} + \frac{p_\perp}{\rho} + \frac{B_t^2}{4\pi\rho}\Big) + \frac{B_n^2 v_n}{B^2}(p_\parallel - p_\perp) & \\
-\frac{\mathbf{B}_t \cdot \mathbf{v}_t B_n}{4\pi}\Big(1 - \frac{4\pi(p_\parallel - p_\perp)}{B^2}\Big) \rceil &= 0 \\
\lfloor B_n \rceil &= 0
\end{aligned}
$$

where ρ is the mass density, v and B are the velocity and magnetic field strength. Subscripts t and n indicate tangential and normal components with respect to the discontinuity. (THIS IS IN CGS UNITS!) Quantities $p_\perp$ and $p_\parallel$ are the elements of the plasma pressure tensor perpendicular and parallel with respect to the magnetic field. Quantity ϵ is the internal energy, $\epsilon = p_\perp + p_\parallel/2$, and $\lfloor Q \rceil = Q_2 - Q_1$, where subscripts 1 and 2 signify the quantity upstream and downstream of the discontinuity. These equations refer to the conservation of physical quantities, i.e. the mass flux, the tangential component of the electric field, the normal and tangential components of the momentum flux, the energy flux, and, finally, the normal component of the magnetic field. To solve the jump equations for anisotropic plasma conditions upstream and downstream of the shock, one has to use an additional equation, since the set of equations is underdetermined. One common choice is the magnetic field/density jump ratio.

The following derivations follow (Erkaev, Vogl, and Biernat 2000). Let us introduce two dimensionless parameters, A_s and A_m, which are determined for upstream conditions as

$$
\begin{aligned}
A_s &= \frac{p_{\perp 1}}{\rho_1 v_1^2} \\
A_m &= \frac{1}{M_A^2}
\end{aligned}
$$

where M_A is the upstream Alfvén Mach number. For common solar wind conditions, both of these parameters are quite small (~ 0.01).

For shocks, the tangential components of the electric and magnetic fields are *coplanar* (Equation 17.3). Thus, the components of the magnetic field upstream of the shock are given as $B_{n1} = B_1 \cos\theta_1$ and $B_{t1} = B_1 \sin\theta_1$, where θ_1 is the angle between the magnetic field vector and the vector $\hat{n}$ normal to the discontinuity. Similarly, the components of the bulk velocity upstream of the shock are chosen as $v_{n1} = v_1 \cos\alpha$ and $v_{t1} = B_1 \sin\alpha$, where the angle between the bulk velocity and the normal component of the velocity. Furthermore, a parameter λ is used to denote the pressure anisotropy

$$
\lambda = p_\perp / p_\parallel
$$

and another parameter r is used to denote the ratio of density

$$
r \equiv \frac{\rho_2}{\rho_1} = \frac{v_{n1}}{v_{n2}}
$$

17.2.1 Perpendicular Shock

For a perpendicular shock, $B_n = 0$, we have the conservation relations reduce to

$$
\begin{aligned}
\lfloor \rho v_n \rceil &= 0 \\
\lfloor v_n \mathbf{B}_t \rceil &= 0 \\
\lfloor p_\perp + \frac{B_t^2}{8\pi} + \rho v_n^2 \rceil &= 0 \\
\lfloor \rho v_n \mathbf{v}_t \rceil &= 0 \\
\lfloor \rho v_n (\frac{\epsilon}{\rho} + \frac{v^2}{2} + \frac{p_\perp}{\rho} + \frac{B_t^2}{4\pi\rho}) \rceil &= 0
\end{aligned}
$$

The quantities downstream of the discontinuity are

$$
\begin{aligned}
B_{t2} &= rB_{t1} \\
v_{t2} &= v_{t1} \\
p_{\perp 2} &= p_{\perp 1} + \frac{B_{t1}^2}{8\pi}(1-r^2) + \rho_1 v_{n1}^2 (1-\frac{1}{r})
\end{aligned}
$$

Substituting these into the energy equation leads to

$$
\begin{aligned}
&2\lambda_1(3\lambda_2+1)\xi^3 - \lambda_1(4\lambda_2+1)(2A_S+A_M+2)\xi^2 \\
&+\lambda_2[2\lambda_1(4A_S+1+2A_M)+2A_S]\xi + A_M\lambda_1 = 0
\end{aligned}
$$

where $\xi = 1/r$.

Now we can do some simple estimations. Assume we have isotropic upstream solar wind with $n = 2$amu/cc, $\mathbf{v} = [600, 0, 0]$km/s, $\mathbf{B} = [0, 0, -5]$nT in GSM coordinates, and $T = 5 \times 10^5$K. We want to estimate the downstream anisotropy given a density/tangential magnetic field jump of 3.

```
KeyNotes.shock_estimation()
```

Another thing to note is that, if you set the jump ratio to 4 (maximum value when $\gamma = 5/3$) in the above calculations, the downstream anisotropy will become 0.6. This indicates that under this set of upstream conditions, the jump ratio shall never be close to 4 if the anisotropy $T_\perp/T_\parallel > 1$!

17.2.2 Parallel Shock

Parallel shocks are more special in that the magnetic field strength remains unchanged so the equations effectively describe pure gasdynamic solutions. (Kuznetsov and Osin 2018) presents a simplified solution in a 1D parallel shock case with parallel and perpendicular thermal energy heat fluxes $S_\parallel$ and $S_\perp$ included. Note again the original CGL theory assumes 0 heat fluxes.

17.3 Location of Shock

In the observation comparison paper (Slavin and Holzer 1981) for quasi-perpendicular shocks, they concluded that the variations in shock stand-off distance and shape are ordered by the sonic Mach number M_s and not other Mach numbers involve magnetic field. In other words, they think the bow shock is a gasdynamic structure.

However, even in neutral fluid theory, the determination of shock location as well as shape is still an ongoing research. Imagine the simplest scenario where there is a static ball in the air with infinite mass. Assuming purely homogenous air with known density, velocity and pressure in the upstream, can you tell me the exact location of shock stand-off distance with pen and paper?

On top of that, the introduction of EM field complicates the story. Especially in the case of a parallel shock, the plasmas get “shocked” both upstream and downstream, and the stand-off distance of the shock may not be a single point theoretically. In some sense, normal magnetic field to the boundary “thickens” the shock front.

17.4 Earth Bow Shock

Using data from the AMPTE/IRM spacecraft, (Hill et al. 1995) have shown that the double adiabatic equations do not hold in the magnetosheath. Moreover, the thermal behaviour of the magnetosheath is studied by (Phan et al. 1996) using WIND spacecraft data. They report that most parts of the magnetosheath are marginally mirror unstable: electron observations showed $T_{e\perp}/T_{e\parallel} \sim 1.3$ in the magnetosheath.

18 Magnetosphere

The big picture when considering the interaction of the solar wind and the magnetosphere is as follows:

1. Is there something that can penetrate from the solar wind into the magnetosphere?
2. Is there something that can be triggered from the interaction?
3. Is a physical process internally or externally driven?

18.1 ULF Waves

Ultra-low frequency (ULF) waves refer to waves within frequency range [0.001, 10] Hz. The name does not tell us anything about their physical origin, but simply observational fact. At Earth's magnetosphere, this frequency range overlaps largely with the MHD waves. This is the reason why early pioneers in space physics relied on MHD theory for large spatial and temporal scales to explain the physics behind these waves, albeit some deviations and deficiencies which require more refined models such as the Vlasov description. ULF waves permeate the near-Earth plasma environment and play an important role in its dynamics, for example in transferring energy from the solar wind to the magnetosphere or accelerating electrons in the Earth's radiation belts.

ULF waves were originally called micropulsations or magnetic pulsations since they were first observed by ground magnetometers. ULF pulsations are classified into two types: *pulsations continuous (Pc)* and *pulsations irregular (Pi)* with several subclasses (Pc1–5 and Pi1–2) according to their frequencies and durations. The division is based on their physical and morphological properties, and the boundaries are not strict.

Notation	Period Range [s]	Property
Pc1	0.2 - 5	EMIC
Pc2	5 - 10	EMIC, Mirror
Pc3	10 - 45	Foreshock, FLR, Mirror
Pc4	45 - 150	FLR
Pc5	150 - 600	SW, FLR
Pi1	1 - 40	

Notation	Period Range [s]	Property
Pi2	40 - 150	

With respect to polarization, field line resonant ULF waves can be categorized into three modes: *compressional* ($\Delta B_\parallel$, ΔE_ϕ), *poloidal* (ΔB_r, ΔE_ϕ), and *toroidal* (ΔB_ϕ, ΔE_r). Here, B_r (E_r), $B_\parallel$, and B_ϕ (E_ϕ) are the radial, parallel (or compressional), and azimuthal components in the local magnetic field system, respectively. Referring to the basic MHD theory, the compressional modes are fast modes, whereas the poloidal and toriodal modes are Alfvén modes. The perturbed EM fields are related by $\mathbf{B}_1 = \frac{\mathbf{k}}{\omega}\times\mathbf{E}_1$. Think of a closed field line near the equatorial plane inside the magnetosphere: if the wave vector $\mathbf{k}$ is along the field line, i.e. $\mathbf{k}=(0,0,k_z)$, then there will be two cases for the EM field: poloidal where $\mathbf{E}_1$ in $\hat{\phi}$, $\mathbf{B}_1$ in $\hat{r}$ and toroidal where $\mathbf{E}_1$ in $\hat{r}$, $\mathbf{B}_1$ in $\hat{\phi}$. If the wave vector $\mathbf{k}$ is perpendicular to the field line, i.e. $\mathbf{k}=(k_x,0,0)$, since there is no $E_\parallel$ in MHD, we only have one option $\mathbf{E}_1$ in $\hat{\phi}$ and $\mathbf{B}_1$ in $\hat{z}$. A phase shift is allowed, and actually in real observations (e.g. THEMIS) it is rare that you can find B and E changing in-phase.

18.1.1 Pc1 & Pc2

- Usually observed in the noon-afternoon MLT sector, easily detectable when following sudden impulses (SI) produced by sudden changes in the pressure of the solar wind plasma.
- A sudden compression of the magnetosphere by increased solar wind pressure causes maximum distortion of the quiet magnetospheric plasma near noon at high latitudes. It is on the fieldlines which thread this disturbed plasma that one is most likely to witness ULF emissions.
- Conversely, as suggested by Hirasawa (1981) sudden rarefactions of the magnetosphere would be expected to quench ULF wave growth by reducing the anisotropy and β of the plasma. (INTERESTING ONE, SHOULD CHECK AT SOMETIME!)
- Delay of 1-3 mins between the occurance of SI and the onset of the ULF emission (ground-based magnetometers)[^growth_rate]
- Drive the trapped proton radiation, greatly enhanced eV energy range protons along the B field, and energization of keV range protons caused by betatron acceleration (Arnoldy et al. 2005).

At Earth, Electromagnetic ion cyclotron (EMIC) waves are typically observed in Pc1 and Pc2 range. In the outer radiation belt, the frequency typically ranges between 0.1 to 5 Hz. The characteristic fine structure appearance of “pearl” Pc1 waves was attributed to dispersive field-aligned wave packet propagation in the LH ion mode on successive bounces between hemispheres. However, this still lacks observation support. Spacecraft measurements have shown that EMIC wave propagation is almost exclusively away from the equator at latitudes greater than about 11^o, with minimal reflection at the ionosphere.

The preferential region of occurrence of EMIC waves is known to be the afternoon magnetic local time (MLT) sector from 12:00 to 18:00 MLT in the region near the plasmapause and the plasmaspheric plume.

Mirror Instability & Ion Cyclotron Instability

Already, early observations in the 1970s have shown that the magnetosheath is populated by intense magnetic field fluctuations at time sclaes from 1 s to 10 s of seconds. Later research based primarily on data from ISEE and AMPTE satellites has shown that the mirror mode waves (Section 10.2) and kinetic Alfvén ion cyclotron (AIC) waves (i.e. EMIC waves) constitute a large majority of magnetosheath ULF waves:

1. AIC/EMIC are found predominantly near the bow shock and in the plasma depletion layer[1] (Song, Russell, and Gary 1994), (Hubert et al. 1998).
2. Mirror mode waves dominate in the central and downstream magnetosheath but can occur immediately downstream of quasi-perpendicular shocks too. (Hubert et al. 1989)
3. ULF waves are generally stronger in the dayside magnetosheath.
4. More frequency EMIC wave occurrence during quasi-parallel shocks.

The ion cyclotron instability responsible for the generation of AIC waves often grows under the same conditions as the mirror instability and in the linear approximation should dominate in lower β plasmas. The mirror instability, on the other hand, should dominate in high ion β plasmas (Lacombe and Belmont 1995). Since the initial confirmation of the existence of mirror modes in the Earth's magnetosheath, they have been observed throughout the heliosphere. A long-standing puzzle in space plasmas is the fact that *mirror modes are often the dominant coherent magnetic structures even for low β plasmas*.

People tried to find an answer to this puzzle. A bunch of studies in late 1980s and early 1990s (e.g. [Gary+]) argued that the presence of He^{++} tends to increase the EMIC threshold while the mirror mode growth is less affected by the presence of He^{++} ions. Yoshiharu Omura and his student Shoji presented another possibility in 2009 with hybrid PIC simulations that even though EMIC modes have higher linear growth rate, they saturates an an earlier stage than the mirror modes, especially in higher dimensions (by comparing 2D and 3D results), so that mirror mode waves can gain more free energy from temperature anisotropy.

18.1.2 Pc3 & Pc4

As already noted above, in the beginning when people proposed the ULF wave Pc divisions, many underlying physics are still unclear. The boundaries are chosen based on the observation data back then and does not necessarily contain any physical meaning.

[1] The plasma depletion layer (PDL) is a layer on the sunward side of the magnetopause with lower plasma density and higher magnetic field compared to the corresponding upstream magnetosheath value.

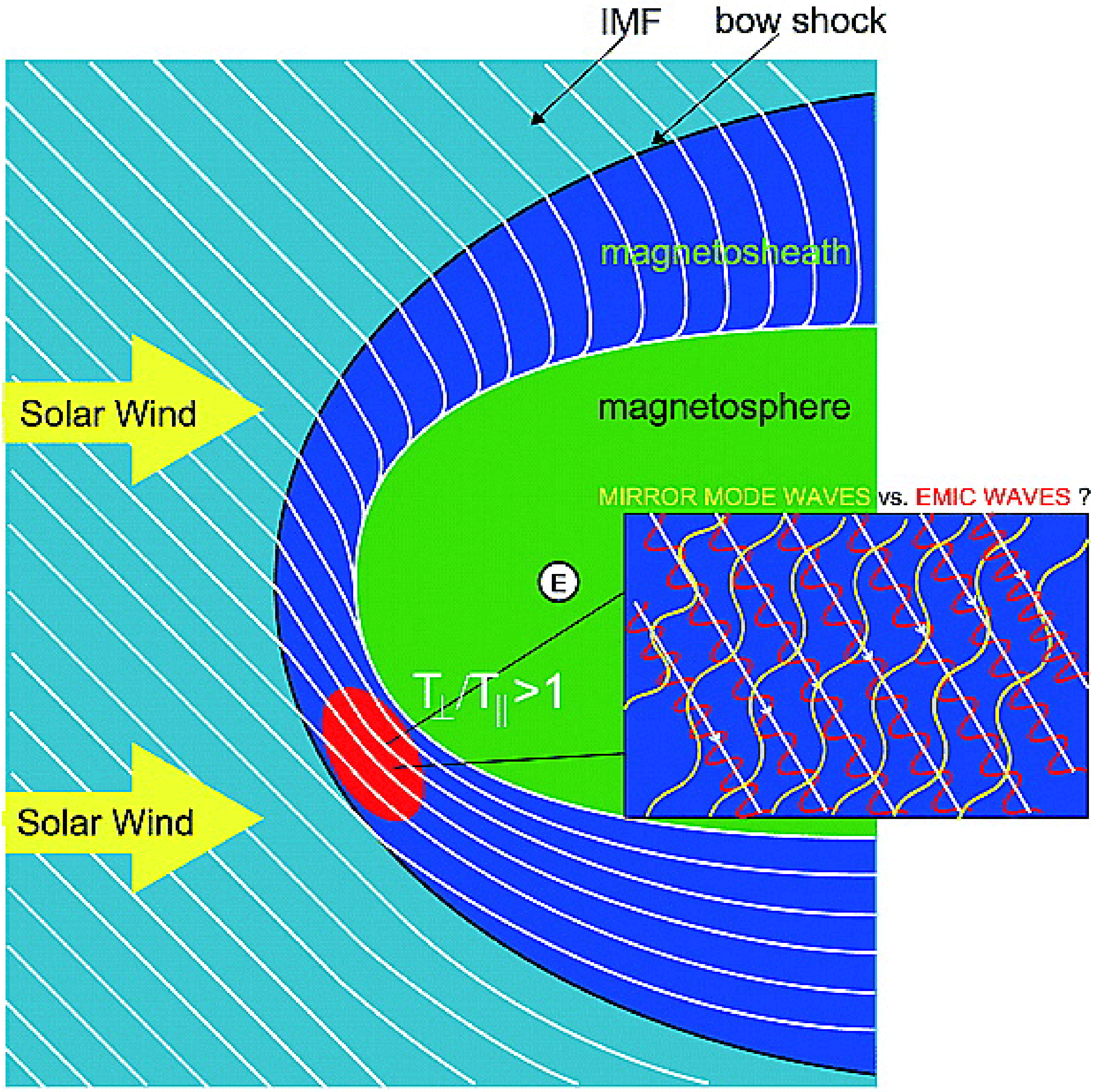

Figure 18.1: Mirror mode waves vs. EMIC waves in the magnetosphere.

ULF waves in the Pc3 range, with periods between 10-45 s, are a common feature of the dayside magnetosphere, where they are frequently observed both by spacecraft and ground-based observatories. They are thought to originate from the ion foreshock, extending upstream of the Earth's quasi-parallel bow shock (the angle between IMF and shock normal $\theta_{Bn} \leq 45^o$). There, ULF waves in the Pc3 frequency range are produced by *ion beam instabilities*, due to the interaction of shock-reflected suprathermal ions with the incoming solar wind.

For the foreshock-related Pc3/4 waves, we have the following picture. After foreshock waves are generated, they propagate through the magnetosheath (with very few observations) and reach the magnetopause. They enter the dayside magnetopause and travel antisunward into the magnetosphere as compressional Pc3 fluctuations, transporting the wave energy towards the nightside. In the inner magnetosphere, they may couple to Alfvénic field line resonances (FLRs), where their frequency matches the eigenmodes of the Earth's magnetic field lines. Pc3 FLRs was observed at low latitudes and Pc4 at midlatitudes [Yumoto+, 1985]. The amplitude of the compressional mode decays when moving further into the magnetosphere, yet they can sometimes be observed all the way to the midnight sector. Compressional Pc3 wave power associated with transmitted foreshock waves is confined near the equator. Statistical study also shows that equatorial Pc3 wave power is stronger in the prenoon or noon sector (under various geomagnetic activity levels), consistent with the foreshock extending upstream of the dawn flank bow shock for a Parker-spiral IMF orientation. However, contrary to Pc5 pulsations (150-600 s), Pc3 wave activity does *not* show a clear correlation with the level of geomagnetic disturbances.

Also, note that not all Pc3 waves are related to foreshock waves, thus we may have different survey results about the distribution of Pc3 waves. This hints the fact that we are far from understanding the whole physical mechanism of wave generation.

There are several mechanisms by which Pc3-4 ULF waves may propagate to high latitudes:

- Harmonics of fundamental mode Pc5 resonances. Such harmonics would be expected to exhibit the same form of amplitude and phase properties that characterize FLRs and should occur at the same time as the fundamental.
- Cavity modes (Chapter 12).
- Fast mode waves propagate without mode conversion through the magnetosphere directly to the ionosphere. Such waves are subject to refraction and diffraction on their passage through the magnetosphere and may be directed to high latitudes via Fermat's Principle (???).
- Transistor model that invokes beams of precipitating *electrons* [Engebretson+ 1991] (???). The transistor model requires no wave mode coupling or wave propagation across field lines, rather the modulated precipitation of electrons in response to pressure fluctuations in the magnetosheath. The latter are attributed to the upstream ion-cyclotron resonance mechanism. The modulated electron beams convey wave information from the outer magnetosphere region containing the parent population of trapped electrons, to the near-cusp ionosphere. The resultant periodic precipitation would modulate the ionospheric

conductivity and hence ionospheric currents equatorward of the cusp. Overhead field lines could then be excited by these modulated currents equatorward of the cusp, with the same frequency as the modulated electrons. Engebretson likened this behavior to that of a transistor, where a small base current modulates a larger flow from collector to emitter. These ULF waves are characterized by noise-like appearance and low coherence lengths.

18.1.3 Pc5

We learned from ground, ionospheric, and space observations about the existence of *only one or at least a few* resonant field line vibrations (i.e. eigenoscillations) in the Pc5 range in the magnetosphere. As the Alfvén velocity is varying in the radial direction and as most sources of magnetospheric hydromagnetic waves are broadband sources, the resonance condition can be matched at an infinite number of geomagnetic field lines. Thus, every field line should be in resonance for a broad enough energy source. Therefore the observational fact of "magic" frequencies requires a magnetospheric frequency selection rule. (Kivelson and Southwood 1985) suggested that perturbations due to a broadband source at, e.g., the magnetopause first couple to the discrete eigenoscillations of global compressional eigenmodes. These narrow band compressional modes then couple to Alfvénic perturbations due to the field line resonance mechanism. The global modes thus select the frequency components participating in the resonant coupling. An alternative way of selection has been proposed by [Fujita+, 1996], who demonstrated that the K-H instability at the surface of a non-uniform magnetospheric plasma introduces dispersive properties of the unstable waves, which then gives rise to a narrower source spectrum. Thus, the field line resonance concept as outlined in Chapter 12 is able to explain the major features of observed resonant ULF pulsations in the terrestrial magnetosphere.

18.2 Bow Shock

Bow shock is the shock (Chapter 17) resulted from the interaction of supersonic, super-Alfvénic solar wind with the magnetosphere of an astrophysical object.

18.2.1 Foreshock

The word "foreshock" is borrowed from geophysics for an earthquake that occurs before a larger seismic event (the "mainshock") and is related to it in both time and space. In space physics, *foreshock* is the region upstream from the bow shock that is magnetically connected to the bow shock and contains both solar wind plasma and also charged particles backstreaming from the bow shock. It is typically associated with quasi-parallel shocks, as it is much easier for charged particles to move freely along the normal direction of the shock if it is aligned with **B**. If their speed relative to the ambient plasma is fast enough, plasma shocks without Coulomb collisions cannot dissipate all the incoming plasmas; some of them have to "return". The counter-streaming between the foreshock plasma and the incoming plasma is very unstable and thus excites many types of waves.

The dominant wave mode in the ion foreshock is produced by the *ion-ion beam right-hand instability* and has a typical period around 30 s in Earth's foreshock. Their exact period however varies significantly depending on the solar wind conditions, in particular the IMF strength, between 10 and 80 s. One easy way to think about this is by remembering the gyrofrequency $\omega_c = qB/m$: the beam instabilities are associated with the cyclotron resonance, and the gyrofrequency is directly related to B. Their wavelength is of the order of $1\,\mathrm{R_E}$ and they are left-hand polarized in the spacecraft frame. Their intrinsic polarization is however right-handed, indicative of a fast mode. The polarization reversal is due to the waves attempting to propagate sunward, while they are effectively carried earthward by the faster solar wind flow. The fast mode is also consistent with the observed positively correlated magnetic field strength and plasma density fluctuations.

An interesting question arises about the transmission of these fast waves through the bow shock into the downstream. A list of possible mechanisms are summarized in **?@fig-foreshock_wave_transmission**. (Turc et al. 2022) argued that the earthward magnetosonic disturbances can then propagate to and disturb the magnetopause, whereas shock reformation and mode conversion play only minor roles.

As usual in plasma physics, accompanying the ion foreshock we have electron foreshock at a much smaller scale.

The formation of ion/electron foreshocks requires counter-streaming of ions and electrons, respectively. Since there is only one velocity for each species in the electron/ion two-fluid equations, it is not possible to trigger this instability in a multi-moment model. Theoretically we may have an extremely simplified case where the ion counter-streaming is represented by two ion fluids, but there are very few studies on this topic. The minimum requirement may be gyrokinetics.

18.3 Magnetosheath

18.3.1 Jets

Magnetosheath jets are regions of high dynamic pressure. Think of jets as raindrops in the magnetosheath falling on the magnetopause. They are widely associated with the quasi-parallel magnetosheath, suggesting that their origin is tied to the interactions between the foreshock and the bow shock.

The formation of magnetosheath jets may be related to both external and internal drivings.

- External
    - Solar wind pressure pulses
    - Rotational discontinuities
- Internal
    - Local ripples from foreshock turbulence could lead to the refraction and funneling of plasma.
    - Short large-amplitude magnetic structures (SLAMS), which are steepened foreshock fluctuations in a short time-scale but large spatial scale.

18.4 Magnetosphere

18.4.1 Dungey Cycle

Dungey (1961) was the first to propose a cycle of magnetospheric convection driven by magnetic reconnection at the dayside magnetosphere. Magnetic field connected on the dayside of Earth is transported by the solar wind to the night side where it forms a long tail behind the Earth. This transfer of flux to the nightside forces the magnetosphere to undergo systematic changes in configuration that eventually lead to nightside magnetic reconnection, which returns flux to the dayside along the flanks of the magnetosphere via the different response modes.

18.4.2 Vasyliunas Cycle

The Jovian equivalent to the Dungey cycle of the Earth's magnetosphere, where the centrifugal force plays a critical role in affecting the plasma convection.

18.4.3 Low-Latitude Boundary Layer

The Earth's *low-latitude boundary layer* (LLBL) is the region where magnetosheath and magnetospheric plasmas are mixed along the magnetospheric side of the low-latitude magnetopause. It has been suggested that three processes — magnetic reconnection at high-latitudes, the Kelvin-Helmholtz instability, and Kinetic Alfvén waves — primarily contribute to the formation, all of which have been observed by in-situ satellites inside the LLBL. Recent studies further suggested that the couplings of these processes effectively enhance the formation rate.

19 Ionosphere

Ionosphere is like a transition region from neutral gas to plasma. Therefore, collisions as well as kinetic effects may become important. Section 6.8 presents some basic physics related to the collisional effect in the ionosphere.

19.1 Current Systems

Electric fields associated with the viscous interaction and reconnection-driven plasma flow are transmitted along field lines to the ionosphere where they drive currents through the ionosphere. The high latitude FAC are known as *region-1 currents*, while the lower latitude currents are called *region-2 currents*. The ionospheric current connecting these two FAC systems flows parallel to the projected electric field and is called the *Pedersen current*. Since the Pedersen current flows through a resistive ionosphere in the direction of the electric field it causes Ohmic heating $\mathbf{j}\cdot\mathbf{E} > 0$. The two shells of FAC form a solenoid so that the magnetic perturbations they create are confined to the region between them and cannot be detected on the ground. Because of the interaction of the motion of ionized gases with the neutral atmosphere, the ions and electrons undergo different drift motion. Ions, due to higher collision rates with the neutral atmosphere, have a component of motion perpendicular to the $\mathbf{E}\times\mathbf{B}$ drift direction, while the electrons generally follow $\mathbf{E}\times\mathbf{B}$. The non-dissipative ($\mathbf{j}\cdot\mathbf{E} = 0$) Hall current flows at right angles to both the electric and magnetic fields. The magnetic effects of the Hall current driven by the region-1 and 2 currents can be observed on the ground and are known as the DP-2 current system.

In the context of ionospheric energy dissipation, the different current systems can be classified into two types based on how energy is processed.

1. If the ionospheric motion and dissipation is coupled directly to the solar wind, the system is said to be "directly-driven". The directly-driven process manifests itself as the DP-2 (two cell pattern) ionospheric current system.
2. If magnetic energy is first stored in the tail lobes and then released some time later, driving additional convection and field-aligned and ionospheric currents, it is called "unloading", and is associated with the DP-1 (substorm current wedge) current system.

Both processes cause precipitation of charged particles that also deposit energy in the atmosphere.

The characteristic feature of the driven DP-2 current system is the existence of the eastward and westward electrojets flowing toward midnight along the auroral oval. Rough measures of the strength of these currents are the *auroral upper* (AU) and *auroral lower* (AL) indices. These are respectively the largest positive northward magnetic perturbation (H) measured on the ground under the eastward electrojet by any magnetic observatory in the afternoon to dusk sector, and the largest negative southward perturbation measured in the late evening to morning sector. Both AU and AL begin to grow in intensity soon after the IMF turns southward and dayside reconnection begins. The characteristic feature of the unloading DP-1 current system is the sudden development of an additional westward current that flows across the bright region of the expanding auroral bulge. This is the ionospheric segment of the substorm current wedge. The onset of this current is recorded in the AL index as a sudden decrease, corresponding to an increase in intensity of the westward current.

19.2 M-I Coupling

Ionospheric properties, principally conductivity, provide boundary conditions for magnetospheric convection, and the ionosphere is often treated as a passive part of the system. Especially during substorms, however, the boundary conditions change in a time-dependent and spatially localized fashion, allowing ionospheric feedback that can alter the magnetospheric dynamics. The coupling from the ionospheric perspective differs primarily in that the ionospheric conductance is anisotropic due to the influence of the neutral atmosphere, involving Hall as well as Pedersen conductivity. These conductivities are altered both by the connecting currents and the precipitating electrons associated with upward field-aligned currents, which increase Pedersen conductivity and field line tying. An important role is also played by field-aligned electric fields, set up locally, primarily in upward field-aligned current regions, which are the cause of auroral intensifications and, specifically auroral arcs.

For magnetosphere simulations, the simplest inner boundary is treated as a conducting sphere ($\mathbf{B}_{\mathrm{normal}} = 0$, $\mathbf{E} = \mathbf{v} = 0$). However, this is often not realistic enough to reveal the nature. In magnetosphere simulations, the location of tail main reconnection site will be closer to the Earth by simply applying a conducting boundary. The next level extension is to employ a magnetospheric-ionospheric electrostatic coupling model. This means that we seek nonzero $\mathbf{E}$ and $\mathbf{v}$ at the inner boundary. The inner boundary, where the MHD quantities are connected to the ionosphere, is taken to be a shell of radius r_{in} (e.g. $r_{\mathrm{in}} \sim 3\,\mathrm{R_E}$). The ionosphere locates at $r_{\mathrm{ion}} \sim 1000\,\mathrm{km} \sim 0.15\,\mathrm{R_E}$. Ideally r_{in} shall be as close to r_{ion} as possible, but typically it is restricted by computational limitations, such as extraneously high Alfvén speeds and very large $\mathbf{B}$ field gradients closer to the Earth. Inside this shell we do not solve the governing equations (MHD/PIC/Vlasov), but assume a static dipole field. The important physical processes within the shell are the flow of *field-aligned currents* (FACs) and the closure of these currents in the ionosphere. At each time step,

1. The magnetospheric FACs are mapped along the field lines from the inner boundary to the ionosphere assuming $j_\parallel/B = \text{const.}$, which are the input to the ionospheric potential equation (Raeder, Walker, and Ashour-Abdalla 1995)

$$\nabla \cdot (\mathbf{\Sigma} \cdot \nabla \Phi) = -j_\parallel \sin I \tag{19.1}$$

where $\mathbf{\Sigma}$ is the conductance tensor, Φ is the electric potential, $j_\parallel$ is the mapped FAC density with the downward considered positive and corrected for flux tube convergence, and I is the inclination of the dipole field at the ionosphere, $\sin I = \cos(\frac{\pi}{2} - I) = \hat{b} \cdot \hat{r}$.

There is another form derived by [Wolf 1983]:

$$\nabla_\perp \cdot \begin{pmatrix} \sigma_P/\cos^2\delta & -\sigma_H \cos\delta \\ \sigma_H/\cos\delta & \sigma_P \end{pmatrix} \cdot \nabla_\perp \Phi = j_\parallel \cos\delta \tag{19.2}$$

where δ is the magnetic field dip angle:

$$\cos\delta = -2\frac{\cos\theta}{\sqrt{1+3\cos^2\theta}}$$

for the northern hemisphere, where θ is the polar angle (magnetic colatitude). I DON'T KNOW THE RELATION BETWEEN THESE TWO!

2. Equation 19.1 is solved on the surface of a sphere $r = r_{\text{ion}}$. Commonly there are two types of boundary conditions: (1) $\Phi = 0$ at the equator (Raeder, Walker, and Ashour-Abdalla 1995), or (2) constant potential at or near the low-latitude boundary (e.g. LFM, BATSRUS). From here, one can either choose a static analytic model of Hall and Pederson conductance that accounts for multiple physics, or simply adopt a uniform Pederson conductance, or the height-integrated conductivity, $\Sigma_p = 5\,\text{Siemens}$, while the Hall conductance Σ_H is assumed to be zero. The latter one is simplified to solve

$$\nabla^2 \Phi = -j_\parallel \sin I / \Sigma_p$$

A more realistic conductance requires considering EUV and diffuse auroral contributions (???) as well. The solar EUV contribution to $\mathbf{\Sigma}$ is considered constant in time, but naturally it varies with the solar zenith angle. For example, the empirical formulas by [Moen and Brekke 1993] can be used. The solar EUV radiation is approximated by the 10.7 cm radio flux (commonly known as *F10.7*), a widely used proxy solar UV activity, whose standard values is taken to be $100 \times 10^{-22}\,\text{W/m}^2$.

The total conductance can be then expressed as

$$\Sigma_{P,H} = \sqrt{(\Sigma_{P,H}^{e-})^2 + (\Sigma_{P,H}^{\mathrm{UV}})^2}$$

This is because $\sigma_{P,H} \propto n_e$, which in a stationary state is proportional to the square root of the production rate, and it is the production rates that can be summed linearly. (???)

Does this look well? Not yet. We know that while the high-latitude ionospheric convection is driven by the solar wind and magnetosphere interaction, at lower latitudes atmospheric neutral winds start to dominate. The next level approximation needs to take this into account. Because there is a gap between r_{ion} and r_{in}, the ionospheric footprint of their grid has a low-latitude boundary somewhere in the midlatitudes, e.g., 45° when $r_{\mathrm{in}} \sim 2\,\mathrm{R_E}$ (Figure 19.1). Global magnetospheric models, unless they are fully coupled to models of the inner magnetosphere and the ionosphere, lack details of the ionospheric convection at latitudes equatorward of their low-latitude ionospheric boundary. To some extent, such details can be translated to the global model via the low-latitude boundary condition used to solve Equation 19.2. The easy way is to set the ionospheric potential to zero everywhere on the boundary. This corresponds to no flow across the boundary in the ionosphere or the inner boundary of the magnetosphere simulation in the equatorial plane. The choice of this boundary condition is usually justified by the argument that it helps to shield the inner magnetosphere from the cross-tail electric field.

(Merkin and Lyon 2010) tested three different boundary conditions for the potential equation:

- STANDARD: Dirichlet, the potential at the low-latitude boundary is set to zero.
- NEUMANN: the electric field component normal to the low-latitude boundary was set to zero. This condition requires all ionospheric plasma to move normal to the boundary.
- LOWERBC: Dirichlet, but the location of the low-latitude boundary was moved to 2° above the equator, thus allowing the plasma to move across the magnetosphere inner boundary. (???) Equation 19.3 is singular at $\theta = \pi/2$, which is why the calculation has to stop just short of the equator. (???)

3. Sparse linear algebra, GMRES together with an imcomplete LU preconditioner (default for many modern solvers) are usually applied to solve the potential equation. This is generally an easy equation to solve mathematically.

4. Once the potential equation is solved the ionospheric potential is mapped back to the r_{in} shell and used as a boundary condition for the magnetospheric flow by taking $\mathbf{v} = (-\nabla\Phi) \times \mathbf{B}/B^2$.

19.2.1 Caveats

- The mapping assumes conservation, which is not perfect. In practice $r_{\mathrm{in}} \sim 4\,\mathrm{R_E}$ is a minimum requirement for reasonable FACs.

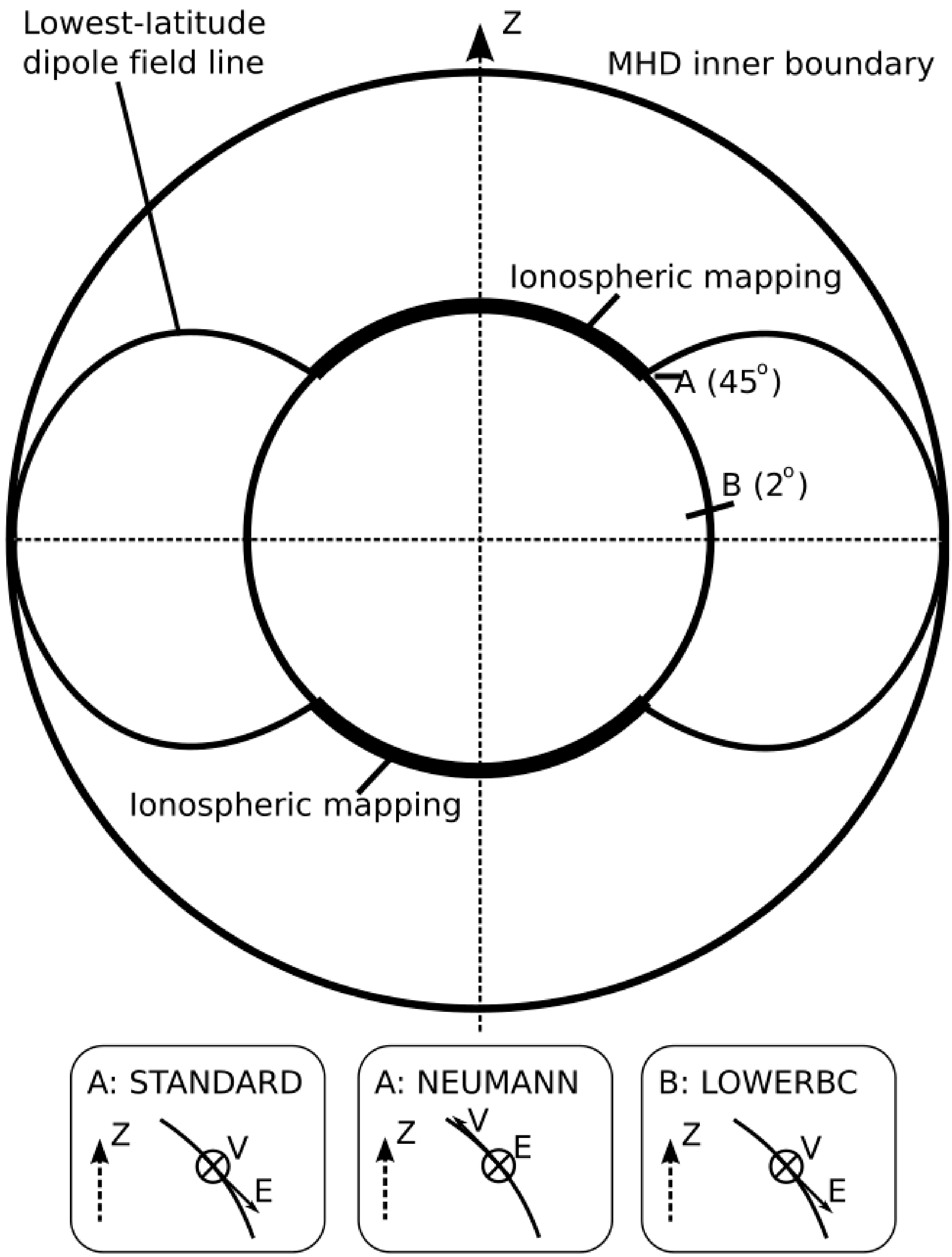

Figure 19.1: A schematic depiction of the inner boundary of the magnetosphere simulation and its ionospheric mapping. The location of the low-latitude boundary of the ionospheric grid is determined by the radius of the inner MHD boundary mapped along the dipole field from the equator. Point A (at 45°) denotes a typical location of the low-latitude boundary, where the STANDARD and NEUMANN boundary conditions are applied, while point B is where the LOWERBC boundary condition is applied. The three inset plots at the bottom depict schematically the configuration of velocity and electric field vectors with respect to the surface of the ionosphere in a meridional plane. The inset plot titles identify the type of the boundary condition and the point where it is applied. Adopted from (Merkin

- Most numerical codes couples a Cartesian grid to a spherical ionosphere grid, while some couples a spherical grid to a spherical grid. For magnetosphere simulations we need a relatively simple but super fast electric potential solver, therefore structured mesh is often adopted. If a spherical grid is used, care must be taken near the pole since it is a singular from the grid but not physics. Equation 19.2 under spherical coordinates is written as

$$
\begin{aligned}
&\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\Big[\sin\theta\frac{\Sigma_P}{\cos^2\delta}\frac{\partial\Phi}{\partial\theta}-\frac{\Sigma_H}{\cos\delta}\frac{\partial\Phi}{\partial\phi}\Big]\\
&\quad+\frac{\partial}{\partial\phi}\Big[\frac{\Sigma_P}{\sin^2\theta}\frac{\partial\Phi}{\partial\phi}+\frac{\Sigma_H}{\sin\theta\cos\theta}\frac{\partial\Phi}{\partial\theta}\Big]=j_\parallel R^2\cos\delta
\end{aligned}
\tag{19.3}
$$

- Be careful about distinguishing $E_\parallel$ and $j_\parallel$. $E_\parallel = 0$ from advection and Hall terms, but $j_\parallel = \nabla\times\mathbf{B}\cdot\hat{b}/\mu_0$ can be nonzero at the MHD inner boundary.
- How important it is to use a more realistic conductance model? (Merkin and Lyon 2010) shows that different BCs may give $> 10\%$ CPCP values, but I have no clue about the effect of a more complicated conductance.

19.3 Ionosphere Modeling

- What equation set is the model solving? How is the model solving them? What is being neglected?
- What species are included? How is the chemistry solved?
- Parameterizations in things such as heating, cooling, viscosity, conduction, chemistry, diffusion, collision, absorption, ionization cross sections, reaction rates…
- How are upper and lower boundaries treated? How is the pole or the open/closed field-line boundary treated?
- How are electrodynamics considered? How is the aurora specified? How is the magnetospheric electric field imposed? Is ion precipitation considered?

For an ionosphere, a magnetic field-aligned grid is often used. This is complicated by the non-orthogonal nature of the magnetic field coordinate system. Assumptions are made.

How to solve each direction (coupled or independently)? For the ionosphere, along the field-line is treated differently than across the field-lines.

As the equations of motion are solved for, the source terms must be added. How to treat these with respect to solving in the different directions?

Build a simple 1D ionosphere model is easy (CAN I MAKE ONE MYSELF?).

Steady-state - assume $\partial \mathbf{X}/\partial t = 0$. Strangely, this is applied in situations in which the value can change on a time scale much faster than the time-step. For example, the ion velocity is often assumed to be steady-state. Ion chemistry is sometimes assumed to be steady-state.

19.3.1 Chemistry

$$\frac{\partial N}{\partial t} = S - L$$

where N is the number density of the species, S represents the sources and L represents the losses. Losses can almost always be expressed as:

$$L = RMN$$

where R is the reaction rate, M is the density of the species it is reacting with.

For a steady-state,

$$\begin{aligned} S - L &= 0 \\ N &= \frac{S}{RM} \end{aligned}$$

This is quite stable, but can be very wrong in regions of slowly changing ion densities, such as in the F-region. This is perfect for the E-region, though. It is quite easy to implement in a simple environment, but can be much more complicated as non-linear terms are include (recombination, in which M can depend on N).

An explicit time step chemistry is trivial to implement in almost all situations, but it is also the least stable, since the loss terms can become larger than the source terms and the density can quickly be driven to negative values. Subcycling can help with this, but not always.

An implicit time step chemistry is relatively stable and easier to implement than steady-state. For example in GITM, there is a blend between sub-cycling and a simplified implicit scheme that switches depending on the size of the loss term compared to the density.

Now we need to also look at the source terms. The ionization rates can be obtained from Q_{EUV} and substituting $\sigma_s^i\lambda$ instead of $\sigma_s^a\lambda$ (not in τ). [Schunk and Nagy] Chapter 8 list a bunch of chemical equations.

After that, we write down all the sources and losses, decide a time-stepping scheme, and run the model. However, if we don't have any ion advection, the F-region will just build and build!

19.3.2 Ion Advection

In the simplest form, we assume an advection speed and let the densities advect upward and downward.

20 Radiation Belt

A Van Allen radiation belt is a zone of energetic charged particles, most of which originate from the solar wind, that are captured by and held around a planet by that planet's magnetosphere. Earth possesses an inner belt and an outer belt.

1. the inner belt

- MeV protons, 100 keV electrons
- 0.2 - 2 R_{E} (L 1 - 3)

2. the outer belt

- 0.1 - 10 MeV electrons
- 3 - 10 R_{E} (outer boundary is the magnetopause)

Our understanding of the physics mechanisms until 1990s:

1. Low energy electrons get injected into the magnetosphere from the solar wind.
2. Electrons transport towards the planet by reconnections, substorms and associated electric fields.
3. Electrons drift around the planet.
4. Magnetic fluctuations cause inward/outward diffusions.
5. Energy is gained by conservation of the 1st adiabatic invariant.
6. Loss by collisions with the atmosphere, reaching the magnetoapuse, and outward radial diffusion.

During magnetic storms the magnetopause is compressed, so more electrons are lost to the magnetopause. If the mirror point is deep inside the atmosphere, charged particles will precipitate into the atmosphere and lost due to collisions. There are still many mysteries both due to lack of observations and theories. For example, the classical theory cannot explain the electron variation (intermittent injection, rapid loss) timescales!

In 1998, two new theories came out:

1. Enhanced radial diffusion

- Solar wind flows past the magnetosphere, drives K-H instability, and in turn triggers ULF waves.
- The ULF waves trigger FLR (Chapter 12).

- The outcome is a faster radial diffusion process?

2. Wave acceleration

- Substorms, convection, plasma instabilities.
- Waves accelerate electrons to MeV.

The wave acceleration, or more specifically, wave-particle interaction, has been widely explored both theoretically and compared with observation data at Earth, Jupiter, and Saturn.

20.1 Waves in the Radiation Belt

- Lower hybrid waves: ???
- Whistler waves (Section 7.5)
- Chorus waves
  - 1-5 kHz
  - highly nonlinear
  - not generated by lightning, but by natural plasma instabilities (same for hiss).
- Hiss waves
  - broadband structureless signal in the plasmasphere and resembles audible hiss

Before 1990s, waves in the radiation belts are known for transferring energy from charged particles, which act as a loss process to the particles. Richard Thorne, together with his collegues, proposed that they could also the sources of energetic particles ([Horne & Thorne], 1998, 2003, 2005a,b). If that is the case, *the wave-particle interactions must break the adiabatic invariants.*

The theory of wave-particle interaction starts with hot plasma kinetic theory (Section 8.10.1). When the denominator goes to 0, the resonance condition is fulfilled. From observations, we learned that the wave frequencies are smaller than the gyro-frequencies. How can we have resonance then? The answer is, Doppler shift needs to be taken into account.

20.1.1 Sources of Waves

resonance can generate current –> this current can generate higher frequency radiations?

Whistler waves can be explained by the linear theory. However, chorus and hiss waves are highly nonlinear and thus cannot be explained by a linear theory. We conjecture that they are caused by natural plasma instabilities, but we still have little idea what exactly these instabilities are.

As a rough physical picture, during the cyclotron resonance:

1. Waves diffuse source electrons into loss cone –> electron loss and wave growth.
2. Waves diffuse trapped electrons –> energy diffusion leads to electron acceleration.

20.2 Modeling

3D Fokker-Planck diffusion model (e.g. (Glauert, Horne, and Meredith 2014)) has been built to model the radiation belt electrons. In the Earth's radiation belts, the evolution of the phase-averaged phase-space density $f(p,r,t)$ can be described by a diffusion equation (see also Equation 8.19):

$$\frac{\partial f}{\partial t} = \sum_{i,j} \frac{\partial}{\partial J_i}\left[D_{ij}\frac{\partial f}{\partial J_j}\right] \tag{20.1}$$

Here D_{ij} are diffusion coefficients and J_i are the action integrals, $J_1 = 2\pi m_e \mu/|q|$, $J_2 = J$, and $J_3 = q|\phi|$, where μ, J, and ϕ are the adiabatic invariants of charged particle motion, m_e is the electron mass, and q the charge. The adiabatic invariants are awkward variables to visualize and relate to data so many authors transform to other coordinates. One choice is to use pitch angle, energy, and $L^* = 2\pi M/(|\phi| R_\mathrm{E})$ (Equation 4.25), where $M = 8.22\times 10^{22}\,\mathrm{A\,m^2}$ is the magnetic moment of the Earth's dipole field and R_E is the Earth's radius, as the three independent variables.

Assuming a dipole field, changing coordinates to equatorial pitch angle α, kinetic energy E, and L^*, and neglecting some cross derivatives, Equation 8.19 can be written as

$$\begin{aligned}\frac{\partial f}{\partial t} =& \frac{1}{g(\alpha)}\frac{\partial}{\partial \alpha}\bigg|_{E,L} g(\alpha)\left(D_{\alpha\alpha}\frac{\partial f}{\partial \alpha}\bigg|_{E,L} + D_{\alpha E}\frac{\partial f}{\partial E}\bigg|_{\alpha,L}\right) \\ &+ \frac{1}{A(E)}\frac{\partial}{\partial E}\bigg|_{\alpha,L} A(E)\left(D_{EE}\frac{\partial f}{\partial E}\bigg|_{\alpha,L} + D_{E\alpha}\frac{\partial f}{\partial \alpha}\bigg|_{E,L}\right) \\ &+ L^2\frac{\partial}{\partial L}\bigg|_{\mu,J}\left(\frac{D_{LL}}{L^2}\frac{\partial f}{\partial L}\bigg|_{\mu,J}\right)\end{aligned} \tag{20.2}$$

where

$$g(\alpha) = T(\alpha) \sin 2\alpha$$
$$A(E) = (E + E_0)(E(E + 2E_0))^{1/2}$$

and E_0 is the electron rest mass energy. $T(\alpha)$ is related to the bounce period and in a dipole field can be approximated by

$$T(\alpha) = (1.3802 - 0.3198(\sin \alpha + \sin^{1/2} \alpha))$$

$D_{\alpha\alpha}$, $D_{\alpha E} = D_{E\alpha}$, D_{EE}, and D_{LL} are the *drift-* and *bounce-averaged* pitch angle, mixed pitch angle energy, energy and radial diffusion coefficients, respectively. When we need to clarify the fact that they are drift and bounce averaged, they will be explicitly written as $\langle D_{\alpha\alpha} \rangle^d, \langle D_{\alpha E} \rangle^d, \langle D_{EE} \rangle^d$, and $\langle D_{LL} \rangle^d$, respectively.

The bounce-averaged pitch angle and energy diffusion rates are defined by

$$\langle D_{\alpha\alpha} \rangle \equiv \left\langle \frac{(\Delta\alpha)^2}{2\Delta t} \right\rangle$$
$$\langle D_{\alpha E} \rangle \equiv \left\langle \frac{(\Delta\alpha \Delta E)^2}{2\Delta t} \right\rangle$$
$$\langle D_{EE} \rangle \equiv \left\langle \frac{(\Delta E)^2}{2\Delta t} \right\rangle$$

with units of $\mathrm{s}^{-1}, \mathrm{Js}^{-1}$, and $\mathrm{J}^2\mathrm{s}^{-1}$, respectively.

A loss term of the form f/τ_L can be added to represent losses to the atmosphere due to collisions. Here, τ_L, the loss timescale, is equal to 1/4 of the bounce time in the loss cone and infinite elsewhere. As a first-order approximation, we can neglect the cross terms and simplify Equation 20.2 as

$$\begin{aligned} \frac{\partial f}{\partial t} =& \frac{1}{g(\alpha)} \frac{\partial}{\partial \alpha}\bigg|_{E,L} g(\alpha) \left(D_{\alpha\alpha} \frac{\partial f}{\partial \alpha}\bigg|_{E,L} \right) \\ &+ \frac{1}{A(E)} \frac{\partial}{\partial E}\bigg|_{\alpha,L} A(E) \left(D_{EE} \frac{\partial f}{\partial E}\bigg|_{\alpha,L} \right) \\ &+ L^2 \frac{\partial}{\partial L}\bigg|_{\mu,J} \left(\frac{D_{LL}}{L^2} \frac{\partial f}{\partial L}\bigg|_{\mu,J} \right) \\ &- \frac{f}{\tau_L} \end{aligned} \tag{20.3}$$

The first three terms on the right-hand side represent pitch angle, energy, and radial diffusion, respectively. The final term accounts for losses to the atmosphere. Pitch angle diffusion has contributions from wave-particle interactions and Coulomb collisions with the atmosphere, though the latter are only significant inside the loss cone. Energy diffusion is due to wave-particle interactions and radial diffusion to interactions with large-scale fluctuations in the Earth's magnetic and electric fields. The diffusion coefficients are based on statistical wave models derived from data. Depending on the choice of the diffusion coefficients, the Fokker-Planck equation can be linear, quasi-linear, or nonlinear.

Let us make our lives even simpler by considering only the radial diffusion and loss term. The quasi-linear assumptions are:

- Broadband wave spectrum (with random phase), typically assumed as a Gaussian centered around a single dominant mode
- Low amplitude fluctuations (no nonlinear interactions)
- "Resonance limit" (linear growth rate of instabilities goes to 0)
- ULF wave-particle interactions lumped into the diffusion coefficient

Even when quasi-linear theory works fine, there are multiple sources of uncertainties in the radial diffusion coefficient:

- Background magnetic field model
- Bounce + drift average eliminates MLT dependence
- Azimuthal wave structure (i.e. power distribution over mode numbers)

Given the limited observation data, we want to estimate the unknown parameters D_{LL}, τ_L. This is called an *inverse problem.* In 2000, Brautigam and Albert proposed a dependence of the radial diffusion coefficent on K_p and L:

$$D_{LL} \approx 10^{-0.506 K_p(t) - 9.325} L^1 0$$

The standard statistical approach is to apply a Bayesian parameter estimation, where we are looking for a distribution of parameters and correcting results based on new observations. However, it is difficult to apply in high dimensions, which typically requires the Markov-chain Monte Carlo (MCMC) approach. For example, our assumption is that the parameters can be written in the following forms:

$$\begin{aligned} D_{LL}(L,t) &= 10^(a K_p(t) + b) L^c \\ \tau_L(L,t) &= \begin{cases} (a_0 + a_1 L + a_2 L^2)/K_p(t) & \text{for } L \leq L_{pp} \\ a_3/K_p(t) & \text{for } L > L_{pp} \end{cases} \end{aligned}$$

where L_{pp} is the plasmapause location, which can be estimated with some approximations. We introduce a bunch of random variables, and the posterior distributions of these variables after MCMC give us the best fits.

A new trend is to use physics-informed machine learning to learn the coefficients. First we may need to constrain the forms of the coefficients to avoid an ill-posed inverse problem. Let the physical loss term be described by a drift coefficient C:

$$\frac{\partial f}{\partial t} = L^2 \frac{\partial}{\partial L}\bigg|_{\mu,J} \left(\frac{D_{LL}}{L^2} \frac{\partial f}{\partial L}\bigg|_{\mu,J} \right) - \frac{\partial C f}{\partial L}$$

The job of *Physical-Informed Neural Network* (PINN) is to deduce the optimal D_{LL} and C based on statistics by embedding the expected form of the equation into the loss term in the neural network. As this is statistics, we can build and test any kind of relations between variables without physical interpretation, which has been critizised a lot. For example, one may ask how you can let the phase space distribution f depend only on t and L? The information is incomplete from a physics point of view, but datawise it may just "work", in the sense that neural network is essentially a *universal high-dimensional function approximator*.

20.2.1 Coupling physical processes at different time scales

If we only look at low frequency regime, the high frequency wave-particle interactions behave as a diffusion process to the plasma. Thus the coupling idea is not to simulate everything on the same time scale, but simulate the radiation belt effect by macroscopic diffusion coefficients.

21 Aurora

The aurora is a visible and fascinating consequence of the complex magnetospheric processes that are driven by the interaction between the solar wind and the geomagnetic field. The aurora can appear as a diffuse plae band crossing the sky from east to west, or as rapidly moving bright and colorful curtains and rays covering a large fraction of the sky. We will not attempt to describe all the various forms that auroras may appear in, but only distinguish between *diffuse* and *discrete* auroras. The particles that cause these two main types are precipitated into the atmosphere by rather different physical processes. Magnetospheric substorms and the mechanisms leading to the formation of discrete auroras are still at the frontier of magnetospheric research. This chapter is mostly based on Chapter 10 in the lecture notes of Prof. Kjell Rönnmark.

21.1 Auroral Light Emission

There are numerous ancient legends and beliefs about the aurora from various parts of the world. Many of the mythological ideas, as well as more scientific theories from the eighteenth century, explained the aurora as sunlight reflected, refracted, or scattered by various divine or natural processes. Observations reveal that it consists of discrete spectral lines, which proves that auroral light is emitted by a gas, and rules out reflected or scattered sunlight. The strongest line the auroral spectrum is found to be 557.7 nm.

Auroras are caused by charged particles with energy in the range from 10 eV to 30 keV. These particles collide in the upper atomosphere with atoms and molecules that are left in an excited state after the collision. About 90% of the aurora is caused by electrons, and the rest by protons. At altitudes about 100 km where most of the auroral light is emitted, the atmosphere consists mainly of nitrogen and oxygen as shown in Fig. ???. In these gases the quantum mechanically allowed transitions that produce strong spectral lins are all outside the visible part of the spectrum — otherwise air would not be transparent and colourless. The visible auroral emissions are due to transitions from forbidden metastable states, with low transition probabilities and long lifetimes. If the gas pressure is too high, these metastable states will relax to the ground state through collisions long before they decay by emitting light. However, at altitudes above 100 km collisions are sufficiently rare to allow the decay of an excited state of atomic oxygen. This state has an excitation energy of 4.17 eV and a lifetime of 0.8 s. It decays in a two step process. The green line is emitted in the first step, which takes the atom to a state at 1.96 eV. This intermediate state has a very long lifetime, 110 s, and the

the red line emission at 630.0 nm that takes it to the ground state mainly occurs at altitudes above 200 km. This red line can also be produced by direct excitation to the state at 1.96 eV by collisions with low energy electrons. Other strong lines in the auroral spectrum are emitted by nitrogen. Molecular N_2^+ ions are created in an excited state by collisions, and they decay to their ground state by emitting a bluish-voilet line at 427.8 nm. Sometimes the lowest part of strong auroras are colored red by emissions near 600 nm from neutral N_2 molecules.

Figure 21.1 shows the typical altitude distribution of the auroral emissions. The cross section for excitation of the different lines depends on the energy of the incoming auroral particles. Usually, the precipitating particles have higher energy at night than during the day. Low energy electrons give up their energy at higher altitudes, and produces more red emissions. Another difference between day and night is that resonant scattering by sunlight enhances the N_2^+ line 427.8 nm relative to the green line.

21.2 Diffuse Aurora

Some diffuse aurora is nearly always found in the auroral zone. On a clear and dark night it is often seen as a diffuse band, which may appear gray if the intensity is below the color threshold of the eye. Even if no aurora can be seen from the ground, we know from satellite observations of precipitating particles and auroral spectral lines in the UV-range that the diffuse aurora forms a rather continuous band aroud the auroral zone.

Magnetosheath particles that enter through the magnetopause cusps are a source of diffuse aurora on the dayside. In the cusps there are field lines that connect the ionosphere to the magnetosheath, and magnetosheath particles with sufficiently small pitch angles will precipitate along these field lines. Around local noon, there will be a continuous flux of low energy ($\leq 100\,\mathrm{eV}$) particles, which at altitudes above 200 km cause a diffuse band of emissions at 630 nm.

On the nightside, and far into the evening and morning, diffuse auroras are caused by particles of plasmasheet origin. The energy spectrum of these particles extends from below 100 eV to above 20 keV. The continuous precipitation of electrons from the plasmasheet presents a problem, which has not been completely solved yet. Most of the plasmasheet is on closed field lines, and the loss cone in a stationary magnetosphere should be empty (because particles will be lost). To explain the diffuse aurora, we need some mechanisms that can scatter the particles into the loss cone.

Plasma waves in the equatorial magnetosphere can cause strong pitch-angle scattering if they are resonant with the particles. Resonant in this context means that the parallel velocity of the particle and the phase velocity of the wave are related so that the particle feels an electric field that oscillates at the gyrofrequency. Significant changes in the velocities of the particles occur since they systematically gain or lose energy while they are resonant. At least for electrons with energy higher than a few keV, the required pitch-angle scattering can be provided by a

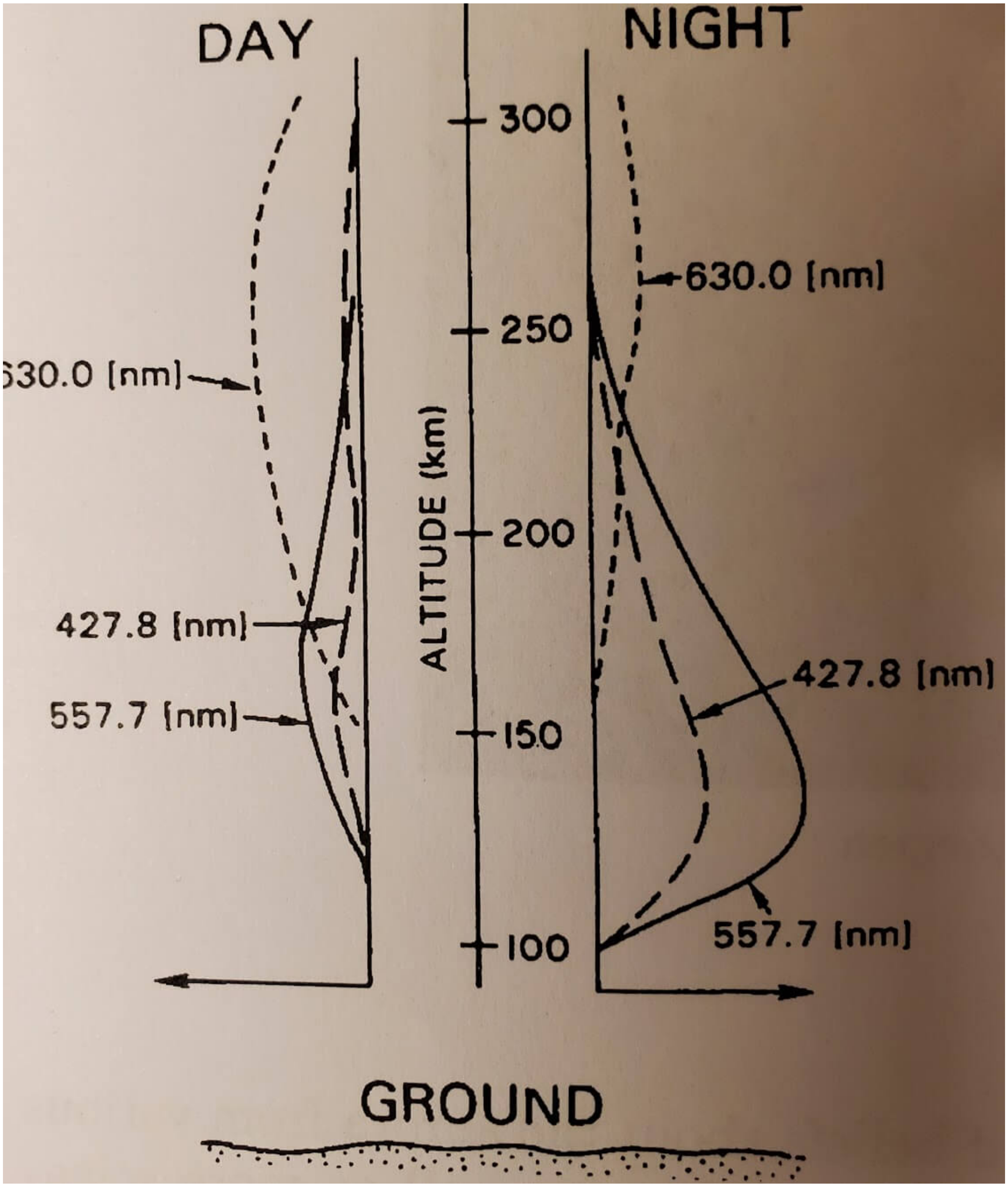

Figure 21.1: The altitude distribution of auroral emissions at day and night.

type of waves in the whistler mode, known as *magnetospheric hiss.* However, the phase velocity of whistler mode waves may be too high to give resonance with lower energy electrons. Low energy electrons can be scattered efficiently by waves in the upper hybrid mode. Upper hybrid waves are electrostatic waves, but it is not clear that these waves are sufficiently common to explain the diffuse precipitation of low energy electrons.

21.3 Auroral Waves and Ion Heating

Plasma waves play several roles on the auroral stage. As we have seen, waves near the equator provide the pitch angle scattering that causes diffuse aurora. Closer to Earth there are several types of plasma waves that are generated by the intense flux of particles precipitating into the aurora.

Discrete auroras are strong sources of whistler modes that appear rather different from whistlers generated by lightning. The auroral emissions are characterized by a broadband spectrum covering frequencies from a few kHz to hundreds of kHz, and they are continously generated with only slow amplitude variations. If the signal is amplified and connected to a loudspeaker a hissing noise is heard, and these waves are called *auroral hiss.* The auroral hiss propagates from the upper ionosphere to the equatorial magnetosphere where it contributes to the magnetospheric hiss. Auroral hiss also propagates to the ground, and at auroral latitudes it will cause noise in radio receivers tuned to low frequencies.

An interesting wave emission connected to discrete auroral arcs is called Auroral Kilometric Radiation (AKR). This is an intense radio wave radiation with peak intensity at frequencies just below 300 kHz, corresponding to a kilometric wavelength. However, the frequency can vary from 50 kHz to 1 MHz. The AKR is generated by a mechanism involving subtle relativistic effects ina low density plasma at a frequency very close to the local electron gyrofrequency. The highest frequencies are generated cloeet to Earth, where the magnetic field is strongest.The maximum power transmitted as AKR has been estimated to 1 GW, which makes auroras the strongest sources of radio waves from Earth. EM radiation at these frequencies can propagate freely out of the magnetosphere where the density is low, but it will be reflected from the ionosphere at the level where the wave frequency equals ω_{pe}. Hence, AKR signals cannot be received on the ground, and although these emissions are very strong they were not discovered until 1965. In fact, the corresponding auroral emission from Jupiter, which is known as Jovian decametric radiation and has higher frequency, was observed by radio astronomers ten years earlier.

Electrostatic plasma oscillations and other electrostatic waves are also generated above auroral arcs. Electrostatic waves with low frequencies, in the vicinity of the ion gyrofrequencies, have important effects on the magnetospheric dynamics. These waves are generated by the auroral electrons, but they are damped by the ions. The wave energy absorbed by the ions goes mainly into their perpendicular velocities. The magnetic moment then increases, and the ions are pushed out of the ionosphere by the magnetic mirror force $\mu\nabla B$. This process can cause a flow of at least $10^{25}\,O^+$ per second from the auroral ionosphere during times of auroral activity. This flow gives the magnetosphere a significant content of ionospheric particles, and reduces the plasma density in the upper ionosphere above auroral arcs. In the auroral cavity dug out by ion heating the electron density is often about $3\times10^5\,\mathrm{m}^{-3}$ down to altitudes below $1\,\mathrm{R_E}$, at least when the ionosphere is in darkness. This density reduction is essential for the generation of AKR, and as we shall see also for the generation of discrete auroras.

21.4 Substorms and Discrete Aurora

Magnetospheric substorms arise due to an imbalance between the dayside and nightside reconnection rates. As a simple example, this can arise during a sudden rotation of the IMF from northward to southward. A southward IMF leads to an increased reconnection rate at the magnetopause. More plasma and more magnetic flux is then transported to the magnetotail in $\mathcal{O}(10^2)\,\mathrm{s}$, and the pressure in the tail lobes increases. The plasmasheet is compressed, which on the ground can be seen as an equatorward motion and brightening of a pre-existing diffuse arc. The increased precipitation from the squeezed plasmasheet will also intensify ion heating, which leads to lower densities in the upper ionosphere above the auroral arc. As the density of the cross tail current increases and the neutral sheet becomes thinner, an unstable situation builds up. At this stage, reconnection starts in some part of the tail (**?@fig-substorm**). This marks the onset of a magnetospheric substorm, and the process of storing energy in the magnetotail described above is often called the *substorm growth phase*.

The original usage of "substorm" comes from Akasofu and Chapman (1961), and was used to describe the short-term magnetic variations during the main phase of a magnetic storm. The current definition did not develop until a decade later, after it became clear that substorms and storms were distinctly different geomagnetic phenomena. The collection of phenomena that includes auroral breakup and expansion, the substorm current wedge, near-Earth dipolarization, and Pi2 pulsations, became collectively known as the magnetospheric substorm.

The concept of the *substorm current wedge* (SCW) has played an important role in understanding the coupling of the magnetotail to the ionosphere during substorms. It provides a simple explanation for the magnetic perturbations observed at mid and low latitudes during substorms, and is useful in understanding the magnetic variations seen in the auroral zone. In its simplest form, a model of the current wedge consists of a single loop with line currents into and out of the ionosphere on dipole field lines connected by a westward ionospheric line current and by an eastward magnetospheric line current (**?@fig-substorm**).

When the magnetic field lines in the tail start to reconnect, part of the plasmasheet will be ejected from the tail and flow Earthward with a speed that often exceeds 400 km/s. The braking of this flow at the inner edge of the plasmasheet requires a substantial $\mathbf{j}\times\mathbf{B}$ force, and hence a substantial dusk-to-dawn current, which leads to dipolarization of the inner portion of the magnetic field. At the edges of the injection region, this current is diverted into field-aligned currents that drive the auroral electrojet, a strong westward current in the ionosphere. Note that the auroral electrojet via Ohm's law implies a westward electric field and an equatorward flow of the ionospheric plasma. Flows from the reconnection region and dipolarization of the magnetic field are associated with field-aligned currents coupling to the ionosphere, whose net effect is then the SCW. This sequence of changes, from energy storage through explosive release, is called a *magnetospheric substorm*. The reconnection process is temporally and spatially varying, which structures the flows in scale sizes of the order of a few $\mathrm{R_E}$ and time scales of a few minutes.

Field-aligned auroral currents in the ionosphere are observed to have current densities of 10 A/m^2. Considering only field-aligned currents, the current density must decrease higher up where the flux tube widens. At altitudes around $1\,\mathrm{R_E}$ the current density is still about 1 A/m^2. In the auroral cavity where the electron density n_e is about $3 \times 10^5\,\mathrm{m^{-3}}$, we can estimate the velocity the electrons must have to carry this current as $v_\parallel = j_\parallel / e n_e \sim 2 \times 10^7$ m/s, which corresponds to a kinetic energy slightly higher than 1 keV. When the fast flow injected from the tail by reconnection is stopped, the field-aligned current is carried by electrons that must be accelerated to keV energies. These electrons will cause aurora, and the sudden buildup of this current leads to a breakup of the quiet auroral arc that existed during the growth phase of the substorm.

The auroral breakup occurs in the region of upward field-aligned current, at the western end of the electrojet. The upward current is carried by downgoing electrons that have been accelerated by $E_\parallel$ at altitudes around $1\,\mathrm{R_E}$. When these electrons start to precipitate, a very bright and dynamic auroral display begins. A suitably located observer may see a large part of the sky filled with rapidly moving auroral forms. As illustrated in the classical drawings of an auroral substorm shown by Akasofu 1968 (viewing from the north pole), the aurora spreads poleward, and after a few minutes more stable discrete auroral arcs start to form.

The inertia of the plasma flowing in from the tail with high speed will carry it into a region where the ambient, mainly magnetic, pressure is higher than the pressure in the injected plasma. As the inflow continues the pressure increases, and the injected plasma will start to expand towards the evening and morning side of the magnetosphere. This happens somewhat further from Earth where the ambient magnetic pressure is lower, on field lines that reach the ionosphere poleward of the initial breakup. The azimuthal plasma flows associated with this expansion cause the discrete arcs that form poleward of the breakup.

The simple substorm current wedge model described here is only a crude approximation to the currents that actually exist in space. It is generally believed that the upward current is localized in the premidnight sector while the downward current is more broadly distributed along the auroral oval post-midnight.[1] Upward currents are carried by downward moving electrons, while the downward current is a combination of upward flowing electrons and precipitating ions. The actual currents probably do not flow on the same L-shells. It has also been suggested that the current wedge includes currents closing in meridian planes. In this more complex model there is a second current wedge of opposite sense flowing on a lower L-shell with a different

[1] Up/down direction is determined w.r.t. the ionosphere.

current strength. The effect of this loop on the ground is to reduce the apparent strength of the higher latitude current wedge.

The Dungey cycle is the source of the two-cell (DP-2) ionospheric convection pattern. Intervals of steady magnetospheric convection, where the dayside and nightside reconnection rates are roughly balanced, approach the idealized state originally envisioned by Dungey (1961). Yet reconnection is not a steady process, and even during intervals of SMC, when the solar wind driver is relatively constant, plasmasheet convection can be intermittent and bursty. Detailed examination of the responses reveals that a pair of field-aligned currents in the form of a current wedge forms during each of the phenomenological stages. Thus, the SCW system is a key phenomenon across all magnetospheric dynamics related to the Dungey cycle, present from weak to extreme activity conditions, and is the link between magnetospheric dynamics and the ionosphere.

21.4.1 Knight's relation

By studying the adiabatic motion of electrons in a parallel electric field, Knight [1973] derived a formula for the auroral current-voltage (C-V) relation. Analyzing data from sounding rockets, Lyons et al. [1979] found that the energy flux of precipitating electrons was proportional to the square of the potential drop. Soon after, it was noticed that for the parameters of interest in the upward current region, Knight's relation could be approximated by [Fridman and Lemaire, 1980] (I feel like this is NOT the original Knight's relation??? Kosta mentioned the linear relation between FAC and B?)

$$j_z \approx -K\Delta\phi \tag{21.1}$$

where j_z is the field-aligned current density below the potential drop and $\Delta\phi$ is the potential difference between the ionosphere and the equatorial magnetosphere. The constant K, known as the Lyons-Evans-Lundin constant, is given by

$$K = \frac{n_G e^2}{\sqrt{2\pi m T_G}}$$

where N_G and T_G are the density and temperature of electrons in the equatorial generator region. Notice that since N_G and T_G may vary in space and time, the term "constant" in this context means "independent of ϕ". The linear C-V relation Equation 21.1 has since been adopted as a part of theoretical and numerical models. However, there are observations both support and against this relation.

Using a fluid description of the plasma, Rönnmark [1999] showed that within a purely electrostatic quasineutral model the potential drop should be proportional to the square of the upward field-aligned current

$$\Delta\phi = \frac{m_e}{2e^3}\frac{j_z^3}{n^2} \tag{21.2}$$

where n is the density in the acceleration region above the ionosphere.

21.4.2 Current circuit

The current circuit connected with a discrete auroral arc is shown in Figure 21.2. Comparing with **?@fig-substorm** we see that the flows have been rotated from radial to azimuthal. The length scales are also different, since the auroral electrojet is hundreds of kilometers long and the ionospheric current in Figure 21.2 which runs across an auroral arc is at most a few km. Still the physics is very similar. When pointing out some details, we will here use the geometry shown in Figure 21.2.

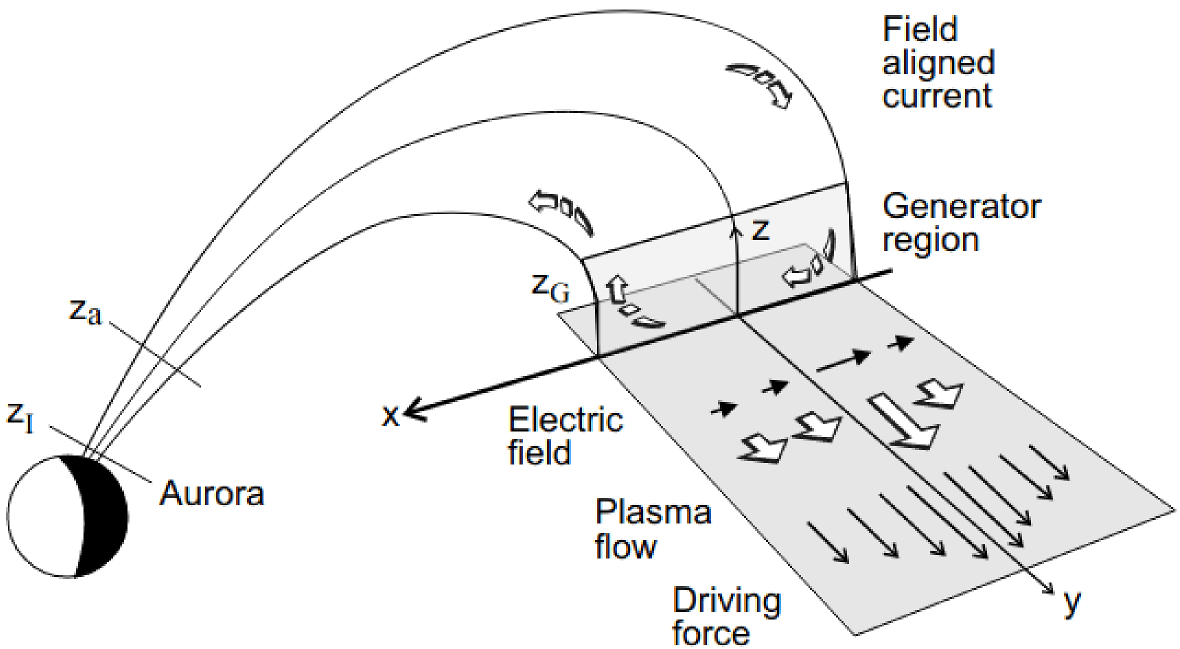

Figure 21.2: Geometry of the auroral current circuit and the generator region in the equatorial magnetosphere.

Let us assume that the auroral flux tube, extending from the ionosphere to the equatorial plane, can be separated into three parts.

1. At low altitudes we have the collision dominated ionosphere, where field-aligned currents are connected to horizontal currents.

2. The magnetospheric plasma above the ionosphere is collisionless. In a stationary state, and in the absence of collisions, the magnetic moment μ and the total energy H are conserved along the phase-space trajectory of a particle. These assumptions imply that there are no currents perpendicular to the magnetic field lines (???), and the field-aligned current in a flux tube is conserved in the second, main, part of the flux tube.
3. The third part is the equatorial generator region. Perpendicular currents are in the generator region driven by kinetic and dynamic pressure gradients, and the divergence of these perpendicular currents is diverted to field-aligned currents. In the real magnetosphere the boundaries between these parts may be rather diffuse, and the generator region may extend far from the equatorial plane.

For simplicity we assume that the main part of an auroral flux tube is separated from the ionosphere and the generator region by well-defined boundaries. Because of quasineutrality, the density of the light and mobile electrons is determined by the ion density. The ion density will remain approximately constant during transitions between different stationary states. Such transitions, for example increases of the field-aligned current, are accomplished by shear Alfvén waves propagating up and down the field lines. The time variation of $E_\perp$ associated with these Alfvén waves will cause an *ion polarization current* $j_\perp$ given by

$$k_\perp = \frac{n_i m_i}{B^2} \frac{\partial E_\perp}{\partial t}$$

Combining this with the ion continuity equation

$$e\frac{\partial n_i}{\partial t} \approx -\partial_\perp j_\perp = \frac{n_i m_i}{B^2} \frac{\partial \partial_\perp E_\perp}{\partial t}$$

we can integrate to find the density change Δn_i during the growth of $E_\perp$. Let $L_\perp$ be the thickness of the current sheet that $j_\perp$ connects to. Choosing $E_\perp = 0.1\,\mathrm{V/m}$ and $B = 10\,\mu\mathrm{T}$ as typical values for the auroral acceleration region we find

$$\left|\frac{\Delta n_i}{n_i}\right| \sim \frac{m_i}{eB^2}|\partial_\perp E_\perp| \sim \frac{10\,\mathrm{m}}{L_\perp}$$

Clearly, this process can increase the plasma density significantly only within extremely thin current sheets. On the other hand, if the current sheet has a thickness of a few hundred meters or more, the density will remain almost constant. Hence, it seems reasonable to consider the plasma density in the main part of the flux tube as fixed when the current and voltage vary.

In a state of steady field-aligned current the contribution to the current by ions of mass m_i is about a factor $\sqrt{m_e/m_i}$ smaller than the contribution by electrons of mass m_e. If we as a first approximation consider this ratio as fixed, the ion and electron currents are separately conserved. In the main part of the flux tube, where the current is purely field-aligned, the plasma density will then be unaffected by the presence of a steady current. However, altitude variations in plasma properties such as ion composition and electron and ion temperatures may cause variations in the ratio between ion and electron current, and this will cause slow decreases or increases of the plasma density.

Pressure forces in the equatorial plane try to establish a strong velocity shear $\partial_x u_y$, which implies a strong $\partial_x E_x$ in Figure 21.2 ($\mathbf{E} = -\mathbf{u}\times\mathbf{B}$). Recalling eq-ionosphere_potential_derivation we find that the gradient of the ionospheric electric field is determined by the field-aligned current as

$$\frac{\partial E_x}{\partial x} = \frac{j_z}{\Sigma_P}$$

As long as there is no potential drop along the field lines, the magnetospheric and ionospheric electric fields are simply related, and the mapping of a strong velocity shera in the equatorial plane to the ionosphere demands a strong field-aligned current. However, a strong current through the auroral cavity means that the electrons must be accelerated by a potential drop (e.g. Equation 21.2).

The appearance of this potential drop on field lines that carry currents up from the ionosphere means that the ionospheric and magnetospheric electric fields become decoupled. The frozen-in condition does not hold in the acceleration region, and this breakdown of ideal MHD also means that the equatorial plasma is free to flow without dragging the ionosphere along. The potential drop $\Delta\phi$ limits the current density, and hence the braking $\mathbf{j}\times\mathbf{B}$ force. Notice that it is the low altitude acceleration in this potential drop, which in the energy spectrum of the precipitating electrons is observed as a peak at the energy $e\Delta\phi$, that distinguishes the discrete from the diffuse aurora.

The field lines carrying current up from the ionosphere connect to regions with a strong inward gradient of the driving force, which is the outer part of the volume with enhanced pressure. When the brakes in this region are released, the flow will build up a narrower region of high pressure at larger y. ??? Only the outer part of this smaller volume will continue to expand, and this process quickly creates a narrow flow channel in the y-direction. The outer edge of this channel maps to a discrete auroral arc in the ionosphere.

In the model described above, the flows are constrained to the y-direction. In the real magnetosphere, the flows can be deviated in the radial x-direction to form curls and vortices. As illustrated by Akasofu, the geometry of real auroral arcs is very complicated. The patterns are similar to the turbulence seen when water flows from a river into a lake.

I still have many questions regarding the substorms and SCW. Reading the review by (Kepko et al. 2015) makes things worse: my impression is that there has not yet existed a model for explaining the whole current system. No wonder the sawtooth study with MHD-EPIC ended up in a strange way.

21.5 Omega Band Auroras

Auroral luminosity undulations of the poleward boundaries of diffuse auroras were first described by Akasofu and Kimball (1964) and were named "omega bands" due to the similarity of their shapes to inverted (poleward-opening) versions of the Greek letter Ω. Omega bands are generally observed in the post-midnight and morning sectors during the recovery phases of magnetospheric substorms. They typically have sizes of 400–1000 km and usually drift eastward at speeds of 0.3–2 km/s.

21.6 Pulsation Auroras

21.7 Summary

- Enhanced dayside reconnection
- thin tail current sheet
- reconnection-accelerated electrons
- carry current into the auroral cavity

22 Boundary Conditions

Boundary condition is an important topic both in theory and simulation. This chapter aims at providing an overview of the boundary conditions we apply for electromagnetism and plasma simulations.

22.1 Conducting Boundary

The general boundary conditions on the EM field at an interface between medium 1 and medium 2 are

$$
\begin{aligned}
\hat{n} \cdot (\mathbf{D}_1 - \mathbf{D}_2) &= \tau \\
\hat{n} \times (\mathbf{E}_1 - \mathbf{E}_2) &= \mathbf{0} \\
\hat{n} \cdot (\mathbf{B}_1 - \mathbf{B}_2) &= 0 \\
\hat{n} \times (\mathbf{H}_1 - \mathbf{H}_2) &= \mathbf{K}
\end{aligned}
\tag{22.1}
$$

where τ is used for the interfacial surface change density (to avoid confusion with the conductivity), and $\mathbf{K}$ is the surface current density. Here, $\hat{n}$ is a unit vector normal to the interface, directed from medium 2 to medium 1. We have learned from EM that at normal incidence, the amplitude of an electromagnetic wave falls off very rapidly with distance inside the surface of a good conductor. In the limit of perfect conductivity (i.e., $\sigma \to \infty$), the wave does not penetrate into the conductor at all, in which case the internal tangential electric and magnetic fields vanish. This implies, from Equation 22.1, that the tangential component of $\mathbf{E}$ vanishes just outside the surface of a good conductor, whereas the tangential component of $\mathbf{H}$ may remain finite. Let us examine the behavior of the normal field components.

Let medium 1 be a conductor, of conductivity σ and dielectric constant ϵ_1, for which $\sigma/\epsilon_1\,\epsilon_0\,\omega \gg 1$, and let medium 2 be a perfect insulator of dielectric constant ϵ_2. The change density that forms at the interface between the two media is related to the currents flowing inside the conductor. In fact, the conservation of charge requires that

$$
\hat{n} \cdot \mathbf{j} = \frac{\partial \tau}{\partial t} = -i\,\omega\,\tau
$$

However, $\hat{n} \cdot \mathbf{j} = \hat{n} \cdot \sigma\,\mathbf{E}_1$, so it follows from Equation 22.1 that

$$
\begin{aligned}
\hat{n}\cdot(\epsilon_0\epsilon_1\mathbf{E}_1-\epsilon_0\epsilon_1\mathbf{E}_2) &= \frac{i\sigma}{\omega}\hat{n}\cdot\mathbf{E}_1\\
\left(1+\frac{i\,\omega\,\epsilon_0\,\epsilon_1}{\sigma}\right)\hat{n}\cdot\mathbf{E}_1 &= \frac{i\,\omega\,\epsilon_0\,\epsilon_2}{\sigma}\,\hat{n}\cdot\mathbf{E}_2
\end{aligned}
$$

Thus, it is clear that the normal component of $\mathbf{E}$ within the conductor also becomes vanishingly small as the conductivity approaches infinity.

If $\mathbf{E}$ vanishes inside a perfect conductor then the curl of $\mathbf{E}$ also vanishes, and the time rate of change of $\mathbf{B}$ is correspondingly zero. This implies that there are no oscillatory fields whatever inside such a conductor, and that the fields just outside satisfy

$$
\begin{aligned}
\hat{n}\cdot\mathbf{D} &= -\tau\\
\hat{n}\times\mathbf{E} &= \mathbf{0}\\
\hat{n}\cdot\mathbf{B} &= 0\\
\hat{n}\times\mathbf{H} &= -\mathbf{K}
\end{aligned}
$$

Here, $\hat{n}$ is a unit normal at the surface of the conductor pointing into the conductor. Thus, the electric field is normal, and the magnetic field tangential, at the surface of a perfect conductor. For good conductors, these boundary conditions yield excellent representations of the geometrical configurations of the external fields, but they lead to the neglect of some important features of real fields, such as *losses in cavities* and *signal attenuation in waveguides*. Here is an estimation related to the skin depth.

22.2 Perfectly Matched Layer

In EM field solvers, often we need boundless free-space simulation to prohibit reflecting waves. Back in 1993, a techinque called *perfectly matched layer* (PML) for the *absorption* of EM waves was proposed to handle this problem, so that we don't necessarily need to set boundaries sufficiently far enough from the scatterer when solving interaction problems. With the new medium the theoretical reflection factor of a plane wave striking a vacuum-layer interface is null at any frequency and at any incidence angle, contrary to the previously designed medium with which such a factor is null at normal incidence only. So, the layer surrounding the computational domain can theoretically absorb without reflection any kind of wave travelling towards boundaries, and it can be regarded as a perfectly matched layer. The new medium as the PML medium and the new technique of free-space simulation as the PML technique.

22.2.1 Derivation

The theoretical derivation starts in discussing the transverse electric wave propagation. One key concept that confused me was the magnetic conductivity denoted as σ^*. Maybe this is just a jargon: permeability is a magnetic analogy to conductivity in electric circuits. Reluctance in a magnetic circuit is inversely proportional to permeability just as electric resistance is inversely proportional to conductivity. The relationships between length and cross-sectional area are also the same. Consequently calling permeability "magnetic conductivity" is a fine way to reinforce the analogy and understand magnetic circuits using an electronic analogy.

I guess I need to review my EM courses to fully understand the derivations.

23 Particle-in-Cell

23.1 Phase Space Sampling

Let $\hat{f}(\mathbf{z})$ be a general probability density function with normalization:

$$\int_{\Omega} \hat{f}(\mathbf{z}) dV = 1$$

Now sample the phase space with N markers according to $\hat{f}(\mathbf{z})$ then the *marker distribution*, g is:

$$g(\mathbf{z}) = N\hat{f}(\mathbf{z})$$

and the phase space volume element occupied by a marker is

$$dV_i = \frac{1}{g(\mathbf{z}_i)}$$

where $\mathbf{z}_i$ is the phase space coordinate of the marker and dV_i the volume element occupied by it.

Define the weight of a marker as the number of physical particles in the corresponding phase space volume occupied by the marker:

$$w_i = f(\mathbf{z}_i) dV_i = \frac{f(\mathbf{z}_i)}{g(\mathbf{z}_i)}$$

This marker distribution can represent the total f or the δf. (???)

We can now take moments of the distribution function in the phase space (such as density), and the Monte Carlo approximation to the integration is as follows:

$$M(A) = \int_{\Omega} A(\mathbf{z}) f(\mathbf{z}) dV = \sum_{i=1}^{N} A(\mathbf{z}_i) w_i + \frac{1}{\sqrt{\mathcal{N}}}$$

where N is the number of markers in the phase space region Ω. This Monte Carlo approximation makes the particle simulation feasible for a number of problems, even in 3D, at the cost of introducing an error on the order of $N^{-1/2}$.

23.2 Computational Cost

The key feature that makes PIC more feasible than the brute-force nbody method is the reduction of computational cost. Given N macro-particles in the system, a brute force method has a complexity of $\mathcal{O}(N^2)$, and a PIC method has $\mathcal{O}(N \log N)$. A rough idea how this is possible can be obtained by ignoring the pair interactions and consider collective effects especially for those particles at a large distance.

A commonly used trick is to introduce a reduced ion-to-electron mass ratio. In a 3D simulation, the cost scales with $(m_i/m_e)^3$. However, note that in low mass ratio runs (e.g. $m_i/m_e = 25$), the thermal speed of the electrons may be comparable to the Alfvén speed which causes higher wave decay through Landau damping on the electrons.

23.3 f Method

Consider the growth rate of a single unstable wave in a 1D-1V Vlasov-Poisson system of plasma. The energy conservation leads to

$$\frac{\mathrm{d}}{dt}\Big(\frac{\epsilon_0}{2}E(x)^2\Big) + \int qvE(x)f(x,v)dv = 0$$

For a single mode $E(x,t) = A(t)\sin(kx - \omega t)$, the growth rate is

$$\gamma = \frac{1}{A}\frac{\mathrm{d}A}{\mathrm{d}t} = -\int qvEfdv = -\sum_j q_j w_j E(x_j)$$

where w_j is the particle weight. All kinetic information lies in the marker particle position distribution.

The idea of $\delta - f$ method is to decompose the total distribution function into two parts,

$$f = f_0 + \delta f, \quad |\delta f| \ll f_0$$

and let the particle weight be proportional to the perturbation part,

$$w \propto \delta f, \quad w_j = \frac{\delta f(x_j, v_j)}{f(x_j, v_j)}$$

If the contribution from f_0 is known, for example

$$\int qvEf_0 dv = \int qvf_0(v) A \sin(kx - \omega t) dv = 0$$

then the growth rate can be written as

$$\gamma = -\int qvE\delta f dv = -\int qvE\frac{\delta f}{f} f dv = \sum_j q_j v_j w_j E(x_j)$$

Since now the particle weight is proportional to $\delta f/f$, the system noise would be significantly reduced by $(\delta f/f)^2$; in other words, for the same noise level the number of macro-particles required would be reduced by $(\delta f/f)^2$, which is typically $\mathcal{O}(10^6)$. Of course this requires the prior assumption $|\delta f| \ll f$ to be valid.

24 Hybrid Methods

The hybrid model is valid for low-frequency physics with $\omega \sim \Omega_i$ and $kr_{Li} \sim 1$ (wavelength $\lambda \sim 6r_{Li}$), where ω is the wave frequency, k is the wave number, Ω_i is the ion gyrofrequency, and r_{Li} is the ion Larmor radius. For this range of wave frequency and wavelength, the ion kinetic physics in the near-Earth instabilities are resolved with grid sizes $\sim r_{Li}$ or ion inertial length d_i. The finite ion gyroradius effects are resolved with particle time steps Δt much smaller than the gyroperiod.

A typical time step in a global hybrid magnetosphere model is $0.05\,\Omega_0^{-1}$, where Ω_0 is the upstream (solar wind) ion gyrofrequency. For IMF $B_0 \sim 10\,\mathrm{nT}$, $\Omega_0 \sim 1\,\mathrm{rad/s}$, which corresponds to $f_0 \sim 1/2\pi\,\mathrm{s}^{-1}$. Based on in-situ observations, the typical ion inertial length in the tail is $\sim 0.2\,\mathrm{R_E}$, but smaller near the dayside magnetopause. Therefore usually discrete grid size $\Delta x = 0.1\,d_i$ is barely enough to resolve tail ion kinetic dynamics but not dayside kinetic structures.

24.1 Classical Hybrid Model

Define the electric charge density ρ and current density $\mathbf{J}$ as

$$\rho = \sum_s q_s n_s - e n_e$$

$$\mathbf{J} = \sum_s q_s n_s \mathbf{V}_s - e n_e \mathbf{V}_e$$

where $q_s, n_s, \mathbf{V}_s$ are the charge, number density and bulk velocity of ion species s calculated by taking moments of the distribution function

$$n_s = \int \mathrm{d}^3 v f_s(\mathbf{r}_s, \mathbf{v}_s, t)$$

$$\mathbf{u}_s = \frac{1}{n_s}\int \mathrm{d}^3 v \mathbf{v}_s f_s(\mathbf{r}_s, \mathbf{v}_s, t)$$

or in the corresponding discrete forms where the distribution function is represented as a group of macro-particles with a specific shape function. In this way it behaves more like a particle cloud.

The crucial assumption in the hybrid model is the *quasi-neutrality*, that is, the electrons move fast enough to cancel any charge-density fluctuations and $\rho = 0$ is always satisfied. By assuming quasi-neutrality, we can

- avoid solving the conservation equations for electrons
- avoid solving the Maxwell's equations entirely and instead use the generalized Ohm's law instead

The electron density thus can be written by using ion densities $n_e \approx n_i \equiv \sum_s q_s n_s / e$. In addition, the electron bulk velocity may also be eliminated using Ampère's law

$$\mathbf{J} = \mu_0 \nabla \times \mathbf{B}$$

and the relation

$$\mathbf{v}_e = \mathbf{v}_i - \mathbf{J}/n_e e \tag{24.1}$$

The basic equations used in the conventional PIC hybrid model first has a particle pusher for individual ions

$$\frac{\mathrm{d}\mathbf{x}_j}{\mathrm{d}t} = \mathbf{v}_j$$

$$\frac{\mathrm{d}\mathbf{v}_j}{\mathrm{d}t} = \frac{q_j}{m_j}(\mathbf{E} + \mathbf{v}_j \times \mathbf{B})$$

where the subscript j and e indicate the indices for individual ions and the electron fluid and other notations are standard. The lowercase velocities are velocities for each macro-particle.

Alternatively, if we rely on a Vlasov system, we directly solve for the distribution function $f(\mathbf{r}_s, \mathbf{v}_s, t)$ from the Vlasov equation

$$\frac{\partial f_s}{\partial t} + \mathbf{v}_s \frac{\partial f_s}{\partial \mathbf{r}_s} + \mathbf{a}_s \cdot \frac{\partial f_s}{\partial \mathbf{v}_s} = 0$$

where

$$\mathbf{a}_s = \frac{q_s}{m_s}(\mathbf{E} + \mathbf{v}_s \times \mathbf{B})$$

The generalized Ohm's law is used to determine the time evolution of the electric field, derived from the electron momentum equation assuming $m_e \to 0$,

$$\mathbf{E} = -\mathbf{V}_i \times \mathbf{B} + \frac{1}{n_i e}(\nabla \times \mathbf{B}) \times \mathbf{B} - \frac{1}{n_i e}\nabla \cdot \overleftrightarrow{P}_e + \eta \mathbf{J} \tag{24.2}$$

where the current in the Hall term has already been replaced by the curvature of B. The last term can either represent collision/physical resistivity, or artificial resistivity/numerical diffusion.

The magnetic fields evolve according to Faraday's law

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

Finally, by determining the electron pressure tensor by using an appropriate equation of state, the evolution of the system can be followed in time. For example, let $\overleftrightarrow{P}_e = P_e \overleftrightarrow{I}$ where P_e is the isotropic scalar electron pressure. In the simplest form

$$P_e = n_e k_B T_e$$

where $n_e \approx n_i$ and $T_e = T_i$. Note however in a plasma electron pressure is usually higher than ion temperature, so this is a very crude assumption. Another commonly used assumption is an adiabatic process

$$P_e = n_e^\gamma k_B T_e = n_0 (n/n_0)^\gamma k_B T_{e0}$$

where $\gamma = 5/3$ is the adiabatic index for a monatomic ideal gas.

24.1.1 Pros and Cons

Strengths:

- No approximations to ion physics.
- Valid for $\omega/\Omega_i \sim kr_i \sim 1$.
- No issues for high-β regimes.
- Simple implementation of particle push that can be readily optimized.
- Removes stiffest electron scales.

Limitations:

- Need to resolve ion gyrofrequency.
- Stiff EMHD whistler waves $\Delta t_{CFL} \sim \Delta x^2$.
- No electron Landau damping.
- Explicit time-stepping schemes can be complex.
- No existing method conserves momentum or energy.

24.2 Finite Electron Inertia

The conventional hybrid simulation model dealing with kinetic ions and a massless charge-neutralizing electron fluid is known to be susceptible to numerical instability due to divergence of the whistler-mode wave dispersion, as well as division-by-density operation in regions of low density. The Alfvén wave at short wavelength comparable to ion inertia length has dispersion due to the decoupling between ion and electron dynamics. There thus appears the whistler mode whose frequency diverges as $\omega \propto k^2$. This means that the maximum phase velocity in the system increases rapidly without bound, implying numerical difficulty. The division-by-density issue originates from Equation 24.1 and appears in the Hall and electron pressure gradient terms. Consequently, a pure vacuum region is not allowed to exist in the simulation domain unless some ad hoc technique is used.

The finite electron inertia correction is proposed to solve the whistler-mode wave dispersion issue. The conventional way to include a finite electron inertia correction into the hybrid model is to introduce the following so-called generalized electromagnetic field $\widehat{\mathbf{E}}, \widehat{\mathbf{B}}$, defined as

$$\widehat{\mathbf{E}} = \mathbf{E} - \frac{\partial}{\partial t}\left(\frac{c}{\omega_{pe}^2}\nabla \times \mathbf{B}\right)$$

$$\widehat{\mathbf{B}} = \mathbf{B} + \nabla \times \left(\frac{c^2}{\omega_{pe}^2}\nabla \times \mathbf{B}\right)$$

in which the terms proportional to $\nabla \times \mathbf{B}$ represent electron inertia correction.

From the equation of motion for the electron fluid, it may be shown that

$$\widehat{\mathbf{E}} = -\mathbf{V}_i \times \mathbf{B} + \frac{1}{n_i e}(\nabla \times \mathbf{B}) \times \mathbf{B} - \frac{1}{n_i e}\nabla \cdot \overleftrightarrow{P}_e - \frac{m_e}{e}(\mathbf{V}_e \cdot \nabla)\mathbf{V}_e$$

which is similar to the generalized Ohm's law but now with the last term which also represents the correction. $\mathbf{V}_e$ is obtain from Equation 24.1.

Given the generalized electric field $\widehat{\mathbf{E}}$, one can advance the generalized magnetic field $\widehat{\mathbf{B}}$ by using Faraday's law, which can be easily checked to satisfy

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

Further simplifications are commonly adopted; for example, the electric field correction term and electron-scale spatial variation of density are often ignored. In this case, the magnetic field may be recovered by solving the equation

$$\widehat{\mathbf{B}} = \left(1 - \frac{c^2}{\omega_{pe}^2}\nabla^2\right)\mathbf{B}$$

and $\widehat{\mathbf{E}} = \mathbf{E}$ is assumed. The nice feature with this approach is that the correction can be implemented as a post process to the each integration step of a standard procedure.

24.3 Low Density Treatment

Apparently new methods are still being proposed because the inclusion of electron inertia term along cannot solve all the issues. (Amano, Higashimori, and Shirakawa 2014) suggests another way to solve for the electric field

$$(\omega_{pe}^2 - c^2\nabla^2)\mathbf{E} = \frac{e}{m_e}(\mathbf{J}_e \times \mathbf{B} - \nabla \cdot \overleftrightarrow{P}_e) + (\mathbf{V}_e \cdot \nabla)\mathbf{J}_e + \eta\mathbf{J}$$

which can be reduced to the Laplace equation in near-vacuum region, presenting no numerical difficulty.

Besides, the electron velocity is redefined

$$\mathbf{V}_e = \frac{\mathbf{J}_e}{\max(\rho_e, \rho_{e,min})}$$

where the minimum density $\rho_{e,min}$ is an artificially set value.

In a hybrid system, the maximum phase velocity is the electron Alfvén speed, which goes to infinity when $m_e \approx 0$. However, when doing calculations we only have ion Alfvén speed

$$v_{p,max} \simeq \frac{1}{2}\frac{B}{\sqrt{\mu_0 n_e m_e}} = \frac{1}{2}V_{A,i}\sqrt{\frac{m_i}{m_e}}$$

To keep the maximum phase velocity always below the CFL condition, one may use a modified electron mass ratio m_e' defined as

$$\frac{m_e'}{m_i} = \max\Big(\frac{m_e}{m_i}, V_A^2(\frac{\Delta t}{2\alpha\Delta x})^2\Big)$$

instead of the physical electron mass m_e. Here V_A is the Alfvén speed calculated from the local density and magnetic field, and α is the maximum allowed Courant number (≤ 0.5).

24.4 Comparison with Hall MHD

The zeroth and first moments of the ion Vlasov equation are

$$\frac{\partial n}{\partial t} + \nabla \cdot (n\mathbf{u}_i) = 0$$

$$\frac{\partial mn\mathbf{u}_i}{\partial t} + \nabla \cdot \Big[mn\mathbf{u}_i\mathbf{u}_i - \frac{\mathbf{BB}}{\mu_0} + \frac{B^2}{2\mu_0}\mathbf{I} + \overleftrightarrow{P}\Big] = 0$$

The difference between Hall MHD and hybrid model is the treatment of the pressure tensor term. For Hall MHD with constant T_{i0}/T_{e0},

$$\overleftrightarrow{P} = p_e(1 + T_{i0}/T_{e0})\mathbf{I}$$

For hybrid models,

$$\overleftrightarrow{P} = p_e\mathbf{I} + \int m_i f_i \mathbf{w}\mathbf{w}\mathrm{d}\mathbf{w}$$

Thus Hall-MHD is a "cold-ion" model in the sense that it does not include ion finite Larmor radius (FLR) or other kinetic effects from warm distribution functions.

24.5 Normalization

There are five basic quantities in the hybrid model (length, mass, time, current density, and number density???) and three physical constants (μ_0, q, m). If we add temperature, then correspondingly k_B would appear. Usually even though the particle mass is a constant, we treat it as a parameter to represent a proton system or electron system or other particle system. Thus we need three $(5-2)$ independent reference quantities for the normalized units in a hybrid model. For instance, We can take a magnetic field scale B_0, number density scale n_0, and mass scale m_0. The general variable transformation from the original units to normalized units is

$$\chi = \chi_0 \tilde{\chi}$$

where $\tilde{\chi}$ denote the variable in the normalized units and the scale χ_0 shall be in the original units (e.g. SI).

Typically we use the inverse of gyrofrequency for the time scale

$$t_0 = \Omega_{c0}^{-1} = \frac{m_0}{eB_0}$$

and Alfvén speed for the velocity scale

$$v_0 = v_A = \frac{B_0}{\sqrt{\mu_0 n_0 m_0}}$$

Then the length scale is taken to be the ion skin-depth, or inertial length

$$L_0 = d_i = \frac{c}{\Omega_{c0}} = \sqrt{\frac{m_0}{\mu_0 e^2 n_0}}$$

A common trick we can use to speed up the simulation is to artificially increase the ion mass such that the length scale is increased $\propto \sqrt{m_0}$. For example, in many global hybrid Earth magnetosphere models, d_i in the upstream solar wind is artificially increased to $0.1\,\mathrm{R}_E$ (by increasing the ion mass), which is 6.8 times the realistic $d_i = 0.015\,\mathrm{R}_E$ for $n_{sw} = 6\,\mathrm{amu/cc}$. If our grid resolution is $0.05\,\mathrm{R}_E \approx 300\,\mathrm{km}$ (which is common as of 2020s), we will have about 10 points per ion-scale wave, which is enough to resolve the ion-scale kinetics.

However, a consequence of this scaling due to computational limitation must be emphasized. The reference Alfvén speed

$$V_{A0} = d_{i0}\Omega_{i0}$$

will also be larger than reality since we artificialy increase d_{i0} but not Ω_{i0}.

Note the difference between gyrofrequency and frequency, which differs by a factor of 2π:

$$\omega = 2\pi f$$

I once made a mistake in dealing with a code that uses SI units. You may wonder how come the ion inertial length is defined by speed of light divided by the plasma ion frequency in the unit of [rad/s], and time scale in the unit of Ω_i^{-1} which is [s/rad]. In practice, we do not include 2π in neither of them!

The pressure scale can be equivalently derived from the magnetic pressure or dynamic pressure

$$P_0 = \frac{B_0^2}{\mu_0} = \rho_0 v_0^2$$

Note the drop of the factor of 2 here: it is then kept in the dimensionless equations, e.g. $p_B' = {B'}^2/2$.

The temperature scale is then

$$T_0 = \frac{p_0}{n_0 k_B} = \frac{B_0^2}{2\mu_0 k_B n_0}$$

Note the factor of 2 appeared in the pressure and temperature scales: this is to make the derivations consistent. The unit conversions are summarized in Table 24.1.

24.6 Numerical Stability

Nonlinear numerical simulations typically need some dissipation for stability. This is achieved either via:

1. Explicit terms in equations ("physical dissipation").
2. Upwinding of advective terms (implicit dissipation via discretization)

Table 24.1: hybrid model unit conversion.

(a) Example basic quantities in the hybrid model

Basic variable	Notation	Definition	Value
Magnetic field	B_0	$B_0 = B_{\text{ref}}$	$1 \times 10^{-8}\,\text{T}$
Number density	n_0	$n_0 = n_{\text{ref}}$	$1 \times 10^{6}\,\text{m}^{-3}$
Mass	m_0	$m_0 = m_i$	$1.67 \times 10^{-27}\,\text{kg}$

(b) Example derived quantities in the hybrid model

Derived variable	Notation	Definition	Value
Length	l_0	$l_0 = \sqrt{m_0/(\mu_0 e^2 n_0)}$	$2.28 \times 10^{5}\,\text{m}$
Velocity	v_0	$v_0 = B_0/\sqrt{\mu_0 m_i n_0}$	$2.18 \times 10^{5}\,\text{m} \cdot \text{s}^{-1}$
Time	t_0	$t_0 = m_i/(e * B_0)$	$1.04\,\text{s}$
Mass density	ρ_0	$\rho_0 = n_0/m_i$	$1.67 \times 10^{-21}\,\text{m}^{-3}$
Pressure	p_0	$p_0 = \rho_0 v_0^2$	$7.96 \times 10^{-11}\,\text{N} \cdot \text{m}^{-2}$
Temperature	T_0	$T_0 = p_0 * m_i/(k_B\,\rho_0)$	$5.76 \times 10^{6}\,\text{K}$

Hybrid models usually follow 1) by adding dissipation in Ohm's law:

$$\mathbf{E} = \underbrace{\mathbf{E}^*}_{\text{frictionless }\mathbf{E}} + \underbrace{\eta\mathbf{j}}_{\text{resistivity}} - \underbrace{\eta_H \nabla^2 \mathbf{j}}_{\text{hyper-resistivity}}$$

The reason that we need hyper-resistivity is because the Hall term is badly behaved (stiff?). Slide 33 Stainer??? The hyper-resistivity term has a similar form of an electron collisional viscosity $\eta_H \nabla^2 \mathbf{u}_e$. But the coefficient is too large for space, sometimes argued as "anomalous viscosity".

Extra care shall be taken for conservation when including the frictional terms. For momentum conservation, we should use $\mathbf{E}^*$ for the macro-particle pusher or Vlasov solver, and only use $\mathbf{E}$ for updating $\mathbf{B}$. For energy conservation, it requires a separate electron pressure equation with frictional heating $H_e = \eta j^2 + \eta_H \nabla\mathbf{j} : \nabla\mathbf{j}$ and heat flux $\mathbf{q}_e$:

$$(\gamma - 1)^{-1}\Big[\frac{\partial p_e}{\partial t} + \nabla\cdot(\mathbf{u}_e p_e)\Big] + p_e \nabla\cdot\mathbf{u}_e = H_e - \nabla\cdot\mathbf{q}_e$$

24.6.1 Finite Grid Instability

Imagine a scenario where cold ion beams move through uniform spatial mesh. It was shown by (Rambo 1995) that non-conservative (explicit) schemes are unstable for Ti /Te « 1 regardless of spatial resolution. The precise threshold in T_i/T_e and beam velocity depends on shape-function for macroparticles. This instability causes unstable (exponential) heating of ions until some saturation value and also violates momentum conservation. Implict momentum and energy conserving schemes are stable w.r.t. these instabilities.

24.7 Magnetosphere Inner Boundary Conditions

Inner boundary condition is often the most tricky part in magnetopshere modeling. In Angeo3D [Lin+ 2014], the inner magnetosphere ($r < 6\,\mathrm{R_E}$) is assumed to be dominated by a cold, incompressible ion fluid, which coexists with particle ions. The number density of the cold ion fluid is assumed to be

$$n_f = \frac{n_{\mathrm{eq}}}{r^3}[1 - \tanh(r - 6.5)]$$

where r is in the unit of $\mathrm{R_E}$, and $n_{\mathrm{eq}} = 1000\,\mathrm{cm}^{-3}$.

The inclusion of the cold ion fluid in the inner magnetosphere simplifies the conditions for the fluid-dominant low-altitude, inner boundary. Ion particles are set to be reflected at the magnetospheric inner boundary (e.g. $r = 3.5\,\mathrm{R_E}$). This simple reflection of the ion parallel velocity $\mathbf{v}_{i\parallel}$ means that loss cone effects are omitted. For a particle distribution with an isotropic pitch angle distribution in a dipole field, the particles in a full loss cone are only 0.3% of the total, which is reasonably neglected. $\mathbf{B}$ is assumed to maintain the dipole field values at the inner boundary. The ionospheric conditions (1000 km altitude) are incorporated into the hybrid code, as in global MHD models (Raeder, Walker, and Ashour-Abdalla 1995). See Section 19.2.

24.8 Tests

24.8.1 Proton Cyclotron Anisotropy Instability

This is an electromagnetic and multi-ion verification test. We have an 1D-3V electromagnetic instability driven by $p_{i\perp}/p_{i\parallel} > 1$. Maximum growth happens at $\mathbf{k}\times\mathbf{B} = 0$ with a finite real frequency. For the initial perturbation, we choose $k_x\Delta x = 0.065$. We set an 1D simulation with 64 cells and 10K particles/cell, $\Delta t\Omega_{ci} = 0.01$, dissipation=0.

PCAI results for nominal parameters:

- Transverse velocity and magnetic field components grow from noise (left-handed Alfvén waves).
- Agree with linear theory for these parameters ($\gamma/\Omega_{ci} = 0.162$).
- Pressure anisotropy decreases until saturation.

When we add a 20% density fraction of a minor species of alpha particles (He^+), the growth rate is modified.

The growth rates across a range of β and anisotropy can be computed and compared with a linear solve, e.g. HYDROS.

24.8.2 Landau Damped Ion Acoustic Wave

The fundamental electrostatic mode in the hybrid-PIC model is the ion acoustic wave. This is also driven by pressure perturbations. In fluid models (e.g. Hall MHD), this wave is undamped. However, in the hybrid-PIC, Landau resonance damps the wave and locally flattens ion VDF, which is analogous to electron Landau damping of Langmuir oscillations.

The dispersion relation is

$$\frac{dZ(\zeta)}{\mathrm{d}\zeta} = 2\frac{T_i}{T_e}, \quad \zeta \equiv \frac{\omega - i\gamma}{kv_{\mathrm{th},i}}$$

The nominal simulation parameters are:

$$T_i = 1/3,\ \gamma = 5/3,\ c_s = \sqrt{\gamma T_e/m_i} = 1,\ k_x = \pi/8,\ \delta n = 2 \times 10^{-2}$$

Results for nominal parameters

- Damping rate: $\gamma = -0.0932$.
- Initial perturbation damps to noise floor. Noise can be reduced by:
  1. Use more particles/cell (noise $\sim 1/\sqrt{(}N_p)$).
  2. Binomial smoothing/higher order shape functions.
  3. Using low-discrepancy quasi-Random numbers to seed particles (noise $\sim 1/N_p$).
  4. Most efficient: Delta-F (See Section 23.3).

24.8.3 Magnetic Reconnection Island Coalescence

Magnetic islands are 2D versions of flux-ropes. Here we set a self-driven reconnection problem of coupling of ideal island motion to micro-scale reconnection physics. Ion kinetic effects are crucial in reconnection studies.

Unstable Fadeev island equilibrium: the magnetic field $\mathbf{B}$ is given by $\nabla \times \mathbf{A}$, where in this setup we only need

$$A_y = -\lambda B_0 \ln[\cosh(z/\lambda) + \epsilon \cos(z/\lambda)]$$

and the density is given by

$$n = n_0(1 - \epsilon^2)/[\cosh(z/\lambda) + \epsilon \cos(z/\lambda)]^2 + n_b$$

The pressure balance gives

$$\beta = \frac{2\mu_0 n_0 k_B (T_{i0} + T_{e0})}{B_0^2} = 1$$

The nominal simulation parameters are:

$$\lambda = 5d_i,\ \epsilon = 0.4,\ n_b = 0.2n_0,\ T_i/T_e = 1,\ \eta = 10^{-3},\ \eta_H = 5 \times 10^{-3},\ \gamma = 1$$

and for the numerical parameters, we have a 2D space with 256×128 cells, 50 particles/cell, $\Delta t \Omega_{ci} = 0.005$.

24.8.4 Collisionless Shock

This is a 2D magnetospheric shock problem, with a $M_A = 11.4$ shock injected from the right (open) boundary and a reflecting left boundary to drive the collisionless shock.

25 Tests

Verifying the validity of models is critical in performing consistent research. This chapter goes through the numerical tests we can do in plasma physics, and is heavily influenced by the tests in the Athena MHD code. Detail can be found in (Stone et al. 2008).

25.0.1 Unit Conversion

Most test parameters are listed in dimensionless units. If SI units are required, we need unit conversions. A rule of thumb is to check the number of constants in the system: making one constant equal to 1 requires to shrink the number of basic variables by 1. In ideal MHD Equation 5.38, there are 4 basic quantities (length, mass, time, and current density) and 1 physical constant μ_0. If we include temperature, then there will be 5 basic quantities and 2 physical constants μ_0, R. This indicates that we need 3 reference quantities in order to make the 2 physical constants equal to 1 in the normalized units. For example, in the Earth's magnetosphere, we select a reference length, e.g. $l_0 = 1\,\mathrm{R_E}$, a reference mass density, e.g. $\rho_0 = 1.67\times 10^{-17}\,\mathrm{kg\cdot m^{-3}}$, and a reference magnetic field, e.g. $B_0 = 3.12\times 10^{-5}\,\mathrm{T}$ (Earth's magnetic field strength at the equator). All the rest conversion factors can be derived from these together with the physical constants μ_0, R, m_i expressed SI units. (R is needed for temperature and m_i is needed for number density.) Inserting the initially chosen values, we get a full set of conversion factors from variables in normalized units $n', B', u', p', T', \mathcal{E}'$ to SI units $n, B, u, p, T, \mathcal{E}$:

$$U_{\mathrm{SI}} = U' * U_0$$

where each conversion factor is summarized in Table 25.1. Note that in the definition of number density, the denominator is m_i instead of m_0 and $R = 2k_B/m_i$ only appears in temperature. This is because mass does not appear in ideal MHD: for a given ρ, the same results are obtained for a heavy species with small number density and a light species with large number density, e.g. $\rho = 2m_i n_i = m_i 2n_i$. In principle we can choose the reference mass arbitrarily, but here in order to make the values in SI units "look good" we choose m_i. If you need to compare with a hybrid model, an easy way to make self-consistent conversion is to make mass $m_0 = m_i$ as a substitute for any of the basic variables. In this case, the dimensionless $\rho = n$. Also note that in Table 25.1 the MHD pressure and temperature are the sum of electron's and ion's pressure and temperature, which is the reason a factor of "2" pops out in the definition of temperature. If we are comparing to a hybrid model which assumes massless electrons, a proper modification would be $p_0 = \rho_0 v_0^2/2$ and $T_0 = p_0 * m_i/(k_B\,\rho_0)$.

Table 25.1: MHD unit conversion.

(a) Example basic quantities in MHD

Basic variable	Notation	Definition	Value
Length	l_0	$l_0 = \mathrm{R_E}$	$6.371 \times 10^6\,\mathrm{m}$
Magnetic field	B_0	$B_0 = B_\mathrm{E}$	$3.12 \times 10^{-5}\,\mathrm{T}$
Mass density	ρ_0	$\rho_0 = \rho_\mathrm{E}$	$1.67 \times 10^{-17}\,\mathrm{kg \cdot m^{-3}}$

(b) Example derived quantities in MHD

Derived variable	Notation	Definition	Value
Mass	m_0	$m_0 = \rho_0 l_0^3$	$4.32 \times 10^3\,\mathrm{kg}$
Velocity	v_0	$v_0 = B_0/\sqrt{\mu_0\rho_0}$	$6.81 \times 10^6\,\mathrm{m \cdot s^{-1}}$
Time	t_0	$t_0 = l_0/v_0$	$0.94\,\mathrm{s}$
Number density	n_0	$n_0 = \rho_0/m_i$	$10^{10}\,\mathrm{m^{-3}}$
Pressure	p_0	$p_0 = \rho_0 v_0^2$	$7.75 \times 10^{-4}\,\mathrm{N \cdot m^{-2}}$
Temperature	T_0	$T_0 = p_0 * m_i/(2k_B\,\rho_0)$	$2.81 \times 10^9\,\mathrm{K}$
Energy density	$\mathcal{E}_0$	$\mathcal{E}_0 = p_0$	$7.75 \times 10^{-4}\,\mathrm{J \cdot m^{-3}}$

25.0.1.1 Example: $\rho, \rho\mathbf{u}, \mathcal{E}, \mathbf{B}$ to n, u, T

When dealing with velocity distribution functions, one easy way is to write $f(\mathbf{v}) = f(\mathbf{v}; n, \mathbf{u}, T)$. For instance, a 3D Maxwellian is given as

$$f(\mathbf{v}) = n\left(\frac{m}{2\pi k_B T}\right)^{3/2}\exp\left(-\frac{m(\mathbf{v}-\mathbf{u})^2}{2k_B T}\right)$$

From dimensionless conserved variables $(\rho, \rho\mathbf{u}, \mathcal{E}, \mathbf{B})$ to SI units $(n, \mathbf{u}, T, \mathbf{B})$, we have

$$\begin{aligned}
n_\mathrm{SI} &= n_0\, n = \rho_\mathrm{SI}/m_i = \frac{\rho_0}{m_i}\rho \\
u_\mathrm{SI} &= u_0\left(\frac{\rho u}{\rho}\right) \\
T_\mathrm{SI} &= T_0\frac{(\gamma-1)}{\rho}\left[\mathcal{E} - \frac{1}{2}\rho u^2 - \frac{1}{2}B^2\right] \\
B_\mathrm{SI} &= B_0\, B
\end{aligned}$$

25.1 Errors

1. Diffusion error: the rate at which the amplitude of the wave decreases.
2. Dispersion error: the difference between the speed at which the wave propagates in the numerical versus the analytic solution.

The error norm usually contains contributions from both. However, one could plot the error as a function of position or phase to track each individually.

25.2 Wave Tests

25.2.1 Linear wave

Linearization of Equation 5.38 ends up in a system of equations $\mathbf{W}_t = \mathbf{A}(\mathbf{W})\mathbf{W}_x$, where $\mathbf{W}$ is composed of either the primitive variables or conserved variables in 1D, and the subscripts t and x are corresponding partial derivatives. The eigenvalues and eigenvectors of ideal MHD can be used to set the characteristic wave perturbations. See the details in Appendix A and B from (Stone et al. 2008).

The density, velocity, magnetic field, and total energy are all set to constant values initially. These values can be chosen so that the wave speeds are all well separated, and so that (in MHD) the wavevector is at an arbitrary angle to $\mathbf{B}$. The precise values chosen for the tests described here are given in the appropriate results section below.

The wave is added to as a perturbation to these constant values of the form $\delta\mathbf{U} = A\mathbf{R}\sin(2\pi x)$. Here $\mathbf{U}$ is the vector of conserved variables, A is an amplitude, and $\mathbf{R}$ is the right-eigenvector corresponding to the desired wave family. The components of $\mathbf{R}$ for each test are listed in the appropriate results section below. For all the tests shown here, $A = 10^{-6}$.

The length of the computational domain is set to be one wavelength. Periodic boundary conditions are used. After the wave has propagated one wavelength, we measure the error in the numerical solution by computing the norm of the vector resulting from summing the absolute value of errors in each variable over the grid,

$$
\begin{aligned}
\epsilon = ||\Delta U|| &= \sqrt{\sum_k (\Delta U_k)^2} \\
&= \sqrt{\sum_k (\sum_{i=1}^{N_x} |U^n_{k,i} - U^0_{k,i}|/N_x)^2}
\end{aligned}
$$

Here, $U_{k,i}$ is the numerical solution for the k-th component of the vector of conserved quantities at grid point i and time level n, $U^0_{k,i}$ is the initial numerical solution, and N_x is the number of grid points. Note the initial solution $U^0_{k,i}$ is just the analytic solution which has been discretized to the grid. Since the discretization of the initial condition also introduces error, to measure the error in the integration algorithm it is important to compute errors relative to the initial condition $U^0_{k,i}$ rather than the analytic solution. In 2D (3D), summation over $j(k)$ is required as well.

To compare to the results given here, it is important to

1. Use the same amplitude for the wave
2. Compute the errors in exactly the same way
3. Compute the errors at exactly the same time

This is an excellent quantitative test of the accuracy and convergence rate of a numerical algorithm. The only drawback is that it involves only linear amplitude oscillations. Thus, this test is not characteristic of the kind of problem codes are written to solve in the first place (after all, the dynamics of linear waves can be treated analytically). A code which does well on this test may still be very poor at shock-capturing. Still, it is nice to know an algorithm reduces to the correct answer in the linear regime. Moreover, since virtually all schemes are first-order for discontinuities such as shocks, smooth problems like this are the only way to measure the actual convergence rate of higher-order schemes.

This test has proved very useful at detecting coding bugs:

- The errors for left- and right-going waves of the same family should be identical.
- If the errors do not converge, something is wrong somewhere.
- It is necessary to use double precision and very small wave amplitudes to eliminate round-off error and nonlinear effects.

25.2.1.1 1D Adiabatic Hydrodynamics

Given Equation 5.38 (with $\mathbf{B} = 0$),

$$\begin{aligned} \rho &= 1 \\ \mathbf{u} &= \mathbf{0} \\ p &= 1/\gamma \end{aligned}$$

with $\gamma = 5/3$. The conserved variables are

$$\mathbf{U} = [\rho, \rho u_x, \rho u_y, \rho u_z, E]$$

and the right-eigenvector for a left-going wave is

$$\mathbf{R} = [1.0, -1.0, 0.0, 0.0, 1.5]$$

When switching from $\mathcal{E}$ to p, we can see from the background and $\mathbf{R}$ that

$$\delta p = (\gamma - 1)\left(\delta\mathcal{E} + \rho\, u\delta u + \frac{1}{2}u^2\delta\rho\right) = 1.0$$

The parameters are summarized in Table 25.2.

Table 25.2: 1D adiabatic hydrodynamic linear wave parameters

Variable	Background	Perturbation **R**
ρ	1	1.0
ρu_x	0	-1.0
ρu_y	0	0.0
ρu_z	0	0.0
p	$\frac{1}{\gamma}$	1.0
$\mathcal{E}$	$\frac{1}{\gamma(\gamma-1)}$	1.5

The absolute error in propagating a sound wave to the left one wavelength as a function of the number of grid points N_x

25.2.1.2 1D Adiabatic MHD

Given Equation 5.38,

$$\begin{aligned}\rho &= 1\\ \mathbf{u} &= \mathbf{0}\\ p &= 1/\gamma\\ \mathbf{B} &= [1, \sqrt{2}, 0.5]\end{aligned}$$

with $\gamma = 5/3$. Thus, the characteristic speeds along x is

$$\begin{aligned}v_s &= \sqrt{\frac{\gamma p}{\rho}} = 1.0\\ v_A &= \sqrt{\frac{b_x^2 + b_y^2 + b_z^2}{\rho}} = \frac{\sqrt{13}}{2} \approx 1.8\\ v_{xA} &= \sqrt{\frac{b_x^2}{\rho}} = 1.0\\ v_{x,\text{fast}} &= \sqrt{\frac{1}{2}\left[v_s^2 + v_A^2 + \sqrt{(v_s^2 + v_A^2)^2 - 4v_s^2 v_{xA}^2}\right]} = 2.0\\ v_{x,\text{slow}} &= \sqrt{\frac{1}{2}\left[v_s^2 + v_A^2 - \sqrt{(v_s^2 + v_A^2)^2 - 4v_s^2 v_{xA}^2}\right]} = 0.5\end{aligned}$$

If the conserved variables are ordered as

$$\mathbf{U} = [\rho, \rho u_x, \rho u_y, \rho u_z, E, B_y, B_z]$$

then the right eigenvectors for left going waves are as follows:

- For a fast magnetosonic wave:

```
U = [
4.472135954999580e-01
-8.944271909999160e-01
4.216370213557840e-01
1.490711984999860e-01
2.012457825664615e+00
8.432740427115680e-01
2.981423969999720e-01 ]
```

- For a Alfvén wave:

```
U = [
0.0
0.0
-3.333333333333333e-01
9.428090415820634e-01
0.0
-3.333333333333333e-01
9.428090415820634e-01 ]
```

- For a slow magnetosonic wave:

```
U = [
8.944271909999159e-01
-4.472135954999579e-01
-8.432740427115680e-01
-2.981423969999720e-01
6.708136850795449e-01
-4.216370213557841e-01
-1.490711984999860e-01 ]
```

The parameters are summarized in Table 25.3.

In some cases, we need to convert from conserved variables to other set of variables, e.g. $(n, \mathbf{u}, p, \mathbf{B})$. The number density perturbation is

$$\delta n = \frac{\delta \rho}{m} = \frac{\rho_0}{m} A \sin 2\pi x \tag{25.1}$$

Table 25.3: 1D adiabatic MHD linear wave parameters

(a) fast magnetosonic wave

Variable	Background	Perturbation **R**
ρ	1.0	4.472135954999580e-1
ρu_x	0	-8.944271909999160e-1
ρu_y	0	4.216370213557840e-1
ρu_z	0	1.490711984999860e-1
$\mathcal{E}$	$\frac{1}{\gamma(\gamma-1)} + \frac{13}{8}$	2.012457825664615
B_x	1.0	0.0
B_y	$\sqrt{2}$	8.432740427115680e-1
B_z	0.5	2.981423969999720e-1

(b) Alfvén wave

Variable	Background	Perturbation **R**
ρ	1.0	0.0
ρu_x	0	0.0
ρu_y	0	-3.333333333333333e-1
ρu_z	0	9.428090415820634e-1
$\mathcal{E}$	$\frac{1}{\gamma(\gamma-1)}$	0.0
B_x	1.0	0.0
B_y	$\sqrt{2}$	-3.333333333333333e-1
B_z	0.5	9.428090415820634e-1

(c) slow magnetosonic wave

Variable	Background	Perturbation **R**
ρ	1.0	8.944271909999159e-1
ρu_x	0	-4.472135954999579e-1
ρu_y	0	-8.432740427115680e-1
ρu_z	0	-2.981423969999720e-1
$\mathcal{E}$	$\frac{1}{\gamma(\gamma-1)}$	6.708136850795449e-1
B_x	1.0	0.0
B_y	$\sqrt{2}$	-4.216370213557841e-1
B_z	0.5	-1.490711984999860e-1

Assuming background $u_0 = 0$,

$$\delta u = \frac{1}{\rho}\left[\delta(\rho u) - (\delta\rho)u\right] = \frac{1}{\rho_0 + \delta\rho}\delta(\rho u) \tag{25.2}$$

Given the perturbation of energy density

$$\begin{aligned}\mathcal{E} &= \frac{p}{\gamma - 1} + \frac{\rho u^2}{2} + \frac{B^2}{2\mu_0} \\ p &= (\gamma - 1)\left[\mathcal{E} - \frac{\rho u^2}{2} - \frac{B^2}{2\mu_0}\right]\end{aligned}$$

the pressure perturbation is then

$$\delta p = (\gamma - 1)\left[\delta\mathcal{E} - \frac{1}{2}(\rho_0 + \delta\rho)\delta u^2 - \frac{1}{2}\left(2B_0\delta B + \delta B^2\right)\right] \tag{25.3}$$

If the input is T instead of p, we have

$$\delta T = \frac{p_0 + \delta p}{\rho_0 + \delta\rho} - \frac{p_0}{\rho_0} \tag{25.4}$$

Note that while in the linear perturbation theory we can ignore the high order (>2) terms, we do not need to do so here numerically.

25.2.1.3 2D Adiabatic MHD

In this case, we use the same values as for the 1D adiabatic MHD test, but we use a 2D grid of size $0 \le x \le 2$ and $0 \le y \le 1$. We use twice as many grid points in the x-direction at every resolution (e.g. our highest resolution is 512×256), thus the grid is rectangular, but each cell is square. The wave propagates along the diagonal of the grid, at an angle $\theta = \tan^{-1}(0.5) \approx 26.6$ degrees with respect to the x-axis. Since the wave does not propagate along the diagonals of the grid cells, we guarantee the x- and y-fluxes are different; that is the problem is truly multi-dimensional.

25.2.2 3D Adiabatic MHD

Now we use a 3D grid of size $0 \le x \le 3$ and $0 \le y \le 1.5$, and $0 \le z \le 1.5$. The grid is of size $2N \times N \times N$. The wave propogates along the grid diagonal, again guaranteeing a truly multidimensional test. The background state is identical to the 1D test values.

25.2.3 Anisotropic MHD

This test is taken from (Daldorff et al. 2014). The parallel and perpendicular pressures vary differently in the collisionless plasma even if the unperturbed plasma has isotropic pressure. The isotropic MHD solutions are not valid in models like anisotropic MHD, hybrid, or PIC.

The following equations are an exact solution of the anisotropic MHD equations (Equation 5.38 but replace the energy equation with two pressure equations as in Equation 5.14)

$$\begin{aligned}
\rho &= \rho_0 \left[1 + \delta \sin(kx - \omega t)\right] \\
u_x &= V_f \delta \sin(kx - \omega t) \\
u_y &= 0 \\
u_z &= 0 \\
p &= p_0 \left[1 + \gamma\delta \sin(kx - \omega t)\right] \\
p_\parallel &= p_0 \left[1 + \delta \sin(kx - \omega t)\right] \\
B_x &= 0 \\
B_y &= B_0 \left(1 + \delta \sin(kx - \omega t)\right) \\
B_z &= 0
\end{aligned}$$

where $p = \frac{p_\parallel + 2p_\perp}{3}$ and

$$V_f = \frac{\omega}{k} = \sqrt{\frac{B_0^2 + 2p_0}{\rho_0}}$$

is the propagation speed of the fast wave moving perpendicular relative to the magnetic field direction. This corresponds to a case where $\theta = 90^o$ in Section 7.6.4. Note that p and $p_\parallel$ have different wave amplitudes, i.e. this solution is different from the isotropic (collisional) fast wave. Also note the $2p_0$ (actually $2p_\perp$) term instead of the usual γp term in the phase speed.

Initially, we set $\rho_0 = 1, p_0 = 4.5 \times 10^{-4}$ and $B_0 = 0.04$ which results in $V_f = 0.05$. The wavelength is set to $\lambda = 32$ so that $k = 2\pi/\lambda$. The above solution also satisfies the Vlasov-Maxwell equations solved by a kinetic solver if the ion and electron gyro-radii are small relative to the wavelength λ and the propagation speed $V_{\rm ph}$ is much less than the speed of light c.

25.2.4 Circularly polarized Alfvén wave

$$
\begin{aligned}
\rho &= 1.0 \\
p &= 0.1 \\
v_\perp &= 0.1\sin(2\pi x_\parallel) \\
v_z &= 0.1\cos(2\pi x_\parallel) \\
B_\parallel &= 0 \\
B_\perp &= 0.1\sin(2\pi x_\parallel) \\
B_z &= 0.1\cos(2\pi x_\parallel)
\end{aligned}
$$

with $\gamma = 5/3$ and $x_\parallel = (x\cos\alpha + y\sin\alpha)$ where α is the angle at which the wave propagates with respect to the grid. Here $v_\perp$ and $B_\perp$ are the components of velocity and magnetic field (in the x-y plane) perpendicular to the wavevector. They are related to the components stored on the grid B_x and B_y via

$$
\begin{aligned}
B_\perp &= -B_x\sin\alpha + B_y\cos\alpha \\
B_\parallel &= B_x\cos\alpha + B_y\sin\alpha
\end{aligned}
$$

By inverting the system, we have

$$
\begin{aligned}
B_x &= B_\parallel\cos\alpha - B_\perp\sin\alpha \\
B_y &= B_\parallel\sin\alpha + B_\perp\cos\alpha
\end{aligned}
$$

For 2D the computational domain is of size $L_x = 2L_y$, with $N_x = 2N_y$. Thus, the grid is rectangular, but each cell is square. The wave propagates along the diagonal of the grid, at an angle $\alpha = \tan^{-1}(0.5) \approx 26.6$ degrees with respect to the x-axis. Since the wave does not propagate along the diagonals of the grid cells, we guarantee the x- and y-fluxes are different; that is the problem is truly multi-dimensional.

The wave is an exact nonlinear solution to the MHD equations, allowing one to test the algorithm in the nonlinear regime. Although nonlinear amplitude Alfvén waves are subject to a parametric instability which causes them to decay into magnetosonic waves, the instability should not be present for the parameters defined here. Since the problem is smooth, it can be used for convergence testing. Running the test with smaller pressure (higher) and/or larger amplitudes is a good test of how robust is the algorithm.

See more in reference and (Tóth 2000).

25.2.5 Whistler wave

This test is taken from (Daldorff et al. 2014), and involves the propagation of whistler waves. The whistler waves have variations in the transverse components of the magnetic field and the velocity. The longitudinal components, the density and the pressure are not perturbed. The following solution of the Hall MHD equations describes a right-hand polarized whistler wave propagating in the $+x$-direction:

$$
\begin{aligned}
u_y &= -\delta u \cos(kx - \omega t) \\
u_z &= +\delta u \sin(kx - \omega t) \\
B_y &= +\delta B \cos(kx - \omega t) \\
B_z &= -\delta B \cos(kx - \omega t)
\end{aligned}
$$

where the wave speed is

$$
v_{\mathrm{ph}} = \frac{\omega}{k} = W + \sqrt{\frac{B_x^2}{\rho} + W^2}, \quad W = \frac{m}{e}\frac{kB_x}{2\rho}
$$

Note that the wave speed depends on the wave number k. The magnetic and velocity perturbations are related as

$$
\frac{\delta u}{\delta B} = \frac{B_x}{v_{\mathrm{ph}}\rho}
$$

We set $\rho = 1, u_x = 0, B_x = 0.2$ and $p = 5.12 \times 10^{-4}$, and wavelength $\lambda = 32$. This results in $v_{\mathrm{ph}} = 0.22059$ that is about 10% faster than the Alfvén speed $V_A = 0.2$, so the Hall term is small but not negligible at this wavelength. The amplitude of the transverse magnetic field perturbation is set to $\delta B = 0.01$ which is 5% of the background field.

These parameters are chosen to satisfy

- $c \gg v_{\mathrm{ph}}$
- $m_e \ll m_i$
- $\beta \ll 1$
- $\omega \ll \Omega_e$

such that the kinetic dispersion relation reduces to the Hall MHD dispersion relation for the right-hand polarized whistler waves.

25.2.6 Light wave

25.2.7 Firehose instability

25.2.8 Mirror mode instability

25.3 Shock Tests

25.3.1 Brio-Wu shock tube

This test is an MHD shocktube, where the right and left states are initalized to different values. The initial left/right values are

$$
\begin{aligned}
\rho &= 1.0,\, 0.125 \\
u_x &= 0.0,\, 0.0 \\
u_y &= 0.0,\, 0.0 \\
u_z &= 1.0,\, -1.0 \\
B_x &= 0.75,\, 0.75 \\
B_y &= 0.0,\, 0.0 \\
B_z &= 0.0,\, 0.0 \\
p &= 1.0,\, 0.1
\end{aligned}
$$

and $\gamma = 2$. The hydrodynamic portion of the initial conditions are the same as for the Sod shock tube problem.

This is a standard test for MHD codes for checking whether the code can accurately represent the shocks, rarefactions, contact discontinuities, and the compound structures of MHD.

25.3.2 Ryu and Jones Test 2A

This test is an MHD shocktube, where the right and left states are initialized to different values. It involves a three-dimensional field and velocity structure and rotation of the plane of the magnetic field. The initial left/right values are

$$\begin{aligned}
\rho &= 1.08,\ 1.0 \\
u_x &= 1.2,\ 0.0 \\
u_y &= 0.01,\ 0.0 \\
u_z &= 0.5,\ 0.0 \\
B_x &= 3.6,\ 2.0 \\
B_y &= 2.0,\ 4.0 \\
B_z &= 2.0,\ 2.0 \\
p &= 0.95,\ 1.0
\end{aligned}$$

This test contains fast shocks, slow shocks, and rotational discontinuities which propagate to each side of the contact discontinuity. The ability of the scheme to capture all 7 waves in MHD can be checked with this single test.

25.3.3 Spherical blast waves

We used a rectangular domain, $-0.5 \leq x \leq 0.5; -0.75 \leq y \leq 0.75$. The boundary conditions are periodic everywhere. This non-square domain and periodic boundary conditions produces complex shock-shock and shock-CD interactions at late times.

The initial conditions are

$$\begin{aligned}
&\rho = 1.0 \\
&\mathbf{u} = [0.0, 0.0, 0.0] \\
&p = 0.1
\end{aligned}$$

with $\gamma = 5/3$. Initial velocities are zero everywhere. Within the region $r < 0.1$, $p = 10.0$ (that is, 100 times the ambient pressure). For the MHD problem, the initial magnetic field is uniform everywhere with $B_x = B_y = 1/\sqrt{2}$.

Although this test is not very quantitative, it makes great movies!

At early times, it is important that the out-going blast wave is spherical and shows no grid alignment effects. At late times, the interaction of the blast wave with the CD at the edge of the evacuated bubble in the center produces filaments of dense gas by the Richtmyer-Meshkov instability. It is important these fingers are sharp and not diffused away. Moreover, for the hydrodynamical problem, the pattern of the fingers should be exactly symmetric top-to-bottom and left-to-right. For the MHD problem, the Richtmyer-Meshkov instability is suppressed, and no fingers are evident.

25.4 Instability Tests

25.4.1 Kelvin-Helmholtz instability

We use a square domain, $-0.5 \leq x \leq 0.5; -0.5 \leq y \leq 0.5$. The boundary conditions are periodic everywhere. For $|y| > 0.25$, we set $\rho = 1$ and $u_x = -0.5$, for $|y| \leq 0.25$, $\rho = 2$ and $u_x = 0.5$. The pressure is 2.5 everywhere, and $\gamma = 1.4$, giving a Mach number of about 0.377 in the $\rho = 2$ gas, and about 0.267 in the $\rho = 1$ gas. The interface between the two oppositely directed streams is a discontinuity, that is a "slip surface". We use different densities in the two fluids to make visualization of the interface easier.

For the MHD problem, the initial magnetic field is uniform everywhere with $B_x = 0.5$.

To seed the instability, we add random numbers to both the x- and y-components of the velocity with peak-to-peak amplitude of 0.01.

At early times, one can check that the growth rate of the transverse component of the velocity agrees with the prediction from linear theory. This requires initializing a single-mode perturbation rather than a spectrum of perturbations as we have done here.

At late times, once the instability has gone fully nonlinear, it is difficult to make quantitative comparisons. However, the sharpness of the boundary between the two streams is an indication of the numerical diffusion of the scheme. For example, if the HLLE Riemann solver is used, diffusion at the interface is significant enough to suppress the instability.

25.4.2 Rayleigh-Taylor instability

For the single-mode test, we use a rectangular domain, $-0.25 \leq x \leq 0.25; -0.75 \leq y \leq 0.75$. The boundary conditions are periodic at $|x| = 0.25$, and reflecting walls at $|y| = 0.75$. For $y > 0$ the density is 2.0, while for $y \leq 0$ it is 1.0. A constant gravitational acceleration $g = 0.1$ must be added to the equations of motion. The pressure is given by the condition of hydrostatic equilibrium, that is $p = p_0 - 0.1\rho y$, where $p_0 = 2.5$, and $\gamma = 1.4$. This gives a sound speed of 3.5 in the low density medium at the interface.

The structures which appear in the nonlinear regime are very sensitive to the nature of the perturbations used to seed the instability. To avoid gridding errors associated with perturbing the interface, we instead perturb the velocities. For the single-mode perturbation, we set $u_y = 0.01[1 + \cos(4\pi x)][1 + \cos(3\pi y)]/4$.

For the multimode perturbation, we use a domain of size $-0.25 \leq x \leq 0.25; -0.375 \leq y \leq 0.375$, and set $u_y = A[1 + \cos(8\pi y/3)]/2$, where A is a random number at each zone with a peak-to-peak amplitude of 0.01.

The way in which source terms are included in the algorithm can have a strong effect on the outcome of this test. For example, for Godunov schemes, if the source term is added using operator splitting, grid noise generated by the lack of an exact numerical equilibrium can perturb the interface and seed structure. If the source terms are included directly in the reconstruction and integration steps, it is able to hold hydrostatic equilibrium automatically.

At early times, one can check that the growth rate of the vertical component of the velocity agrees with the prediction from linear theory.

At late times, once the instability has gone fully nonlinear, it is difficult to make quantitative comparisons. However, the sharpness of the boundary between the two fluids is an indication of the numerical diffusion of the scheme. Also, the amount of fine scale struture induced by secondary KH instabilities is sensitive to the way the interface is perturbed, and how sharp the algorithm preserves the contact discontinuity. It is not always clear that sharper is better, however. For example the “contact steepener” in the PPM algorithm can introduce “stair stepping” in contact discontinuities in multidimensions, which in turn can cause KH rolls to be seeded by grid noise.

25.5 Turbulence Tests

25.5.1 Orszag-Tang vortex

We use a square domain, $0 \leq x \leq 1; 0 \leq y \leq 1$. The boundary conditions are periodic everywhere. The density ρ is $25/(36\pi)$ and the pressure is $5/(12\pi)$ everywhere, and $\gamma = 5/3$. Note that this choice gives $u_s^2 = \gamma p/\rho = 1$. The initial velocities are periodic with $u_x = -\sin(2\pi y)$ and $u_y = \sin(2\pi x)$. The magnetic field is initialized using a periodic vector potential defined at zone corners; $A_z = B_0(\cos(4\pi x)/(4\pi) + \cos(2\pi y)/(2\pi))$, with $B_0 = 1.0$. Face-centered magnetic fields are computed using $\mathbf{B} = \nabla \times \mathbf{A}$ to guarantee $\nabla \cdot \mathbf{B} = 0$ initially. This gives $B_x = -B_0 \sin(2\pi y)$ and $B_y = B_0 \sin(4\pi x)$.

The Orszag-Tang vertex is a well-known model problem for testing the transition to supersonic 2D MHD turbulence. Thus, the problem tests how robust the code is at handling the formation of MHD shocks, and shock-shock interactions. The problem can also provide some quanititative estimates of how significant magnetic monopoles affect the numerical solutions, testing the $\nabla \cdot \mathbf{B} = 0$ condition. Finally, the problem is a very common test of numerical MHD codes in two dimensions, and has been used in many previous studies. As such, it provides a basis for consistent comparison of codes.

25.6 Reconnection Tests

25.6.1 GEM challenge

This 2D setup is based on (Birn et al. 2001). The computation is carried out in a rectangular domain $-L_x/2 \leq x \leq L_x/2$ and $-L_z/2 \leq z \leq L_z/2$. The system is taken to be periodic in the x direction with ideal conducting boundaries at $z = \pm L_z/2$. Thus the boundary conditions on the magnetic fields at the z boundaries are $B_z = \partial B_x/\partial z = \partial B_y/\partial z = 0$ with corresponding conditions on the electric fields and particle or fluid quantities. Open boundary conditions are used for all quantities.

The equilibrium chosen for the reconnection challenge problem is a Harris equilibrium with a floor in the density outside of the current layer. The magnetic field is given by

$$B_x(z) = B_0 \tanh(z/\lambda)$$

where λ is the current sheet scale size, and the density by

$$n(z) = n_0 \mathrm{sech}^2(z/\lambda) + n_\infty$$

The electron and ion temperatures, T_e and T_i, are taken to be uniform in the initial state. We assume the plasma $\beta = (p_i + p_e)/p_B = 1$ initially.

The normalization of the space and time scales of the system is chosen to be the ion inertial length $d_i = c/\omega_{pi}$ and the ion cyclotron frequency Ω_i^{-1}, where $\omega_{pi}^2 = n_0 e^2/\epsilon_0 m_i$ is evaluated with the density n_0 and the ion gyrofrequency $\Omega_i = eB_0/m_i$ is evaluated at the peak magnetic field. The velocities are then normalized to the Alfvén speed v_A. In the normalized units, $B_0 = 1$ and $n_0 = 1$. Specific parameters for the simulations are $L_x = 25.6, L_z = 12.8, \lambda = 0.5, n_\infty/n_0 = 0.2$, and $T_e/T_i = 0.2$. $m_i/m_e = 25$ is assumed if required.

The initial magnetic island is specified through the perturbation in the magnetic flux,

$$\psi(x, z) = \psi_0 \cos(2\pi x/L_x) \cos(\pi z/L_z)$$

where the magnetic perturbation is given by $\mathbf{B} = \hat{y} \times \nabla\psi$, or more specifically,

$$B_{1x} = -\psi_0 \Big(\frac{\pi}{L_z}\Big) \cos(2\pi x/L_x) \sin(\pi z/L_z)$$
$$B_{1z} = \psi_0 \Big(\frac{2\pi}{L_x}\Big) \sin(2\pi x/L_x) \cos(\pi z/L_z)$$

In normalized units $\psi_0 = 0.1$, which produces an initial island width which is comparable to the initial width of the current layer. The rationale for such a large initial perturbation is to put the system in the nonlinear regime of magnetic reconnection from the beginning of the simulation.

The initial setup is shown in Figure 25.1.

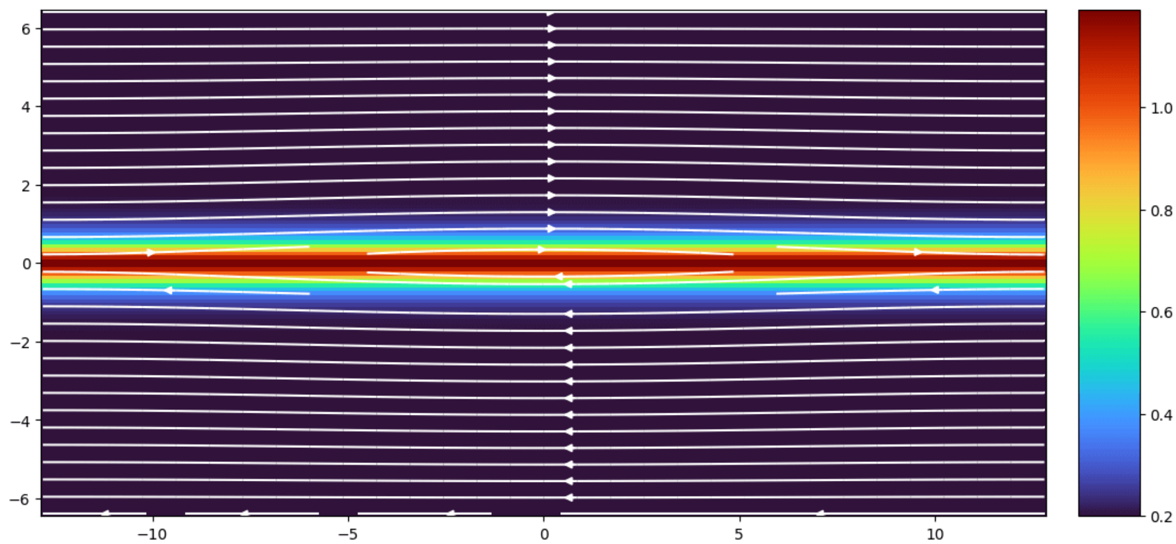

Figure 25.1: Initial density and magnetic field in the GEM current sheet setup.

25.6.1.1 Why do we need to resolve ion inertial length

Physically, electrons and ions separate at the scale of ion inertial length. Numerically, Hall term is important only when cell size is small enough to resolve the ion initial length. The reason is as follows. The Ohm's law is

$$\mathbf{E} = -(\mathbf{U} + \mathbf{U}_H) \times \mathbf{H}$$

Assume the typical flow velocity is Alfvén velocity: $\mathbf{U} = \mathbf{V}_A$. The Hall velocity is estimated as:

$$\mathbf{U}_H = -\frac{\mathbf{J}}{ne} = -\frac{\nabla \times \mathbf{B}}{\mu_0 ne} \sim -\frac{|dB|}{\mu_0 ne \Delta x} \sim -\frac{|B|}{\mu_0 ne \Delta x}$$

The approximation $dB \approx B$ is valid for relatively coarse grid size; for fine discrete cell sizes, this approximation does not hold, so magnetic field cannot cancel out in the following estimation. Let us now assume we can make this assumption. Then the ratio of Hall velocity and Alfvén velocity is:

$$\frac{|\mathbf{U}_H|}{|\mathbf{V}_A|} = \frac{c/\omega_{pi}}{\mu_0 \Delta x}$$

Ion inertial length is also important for PIC: if particle's velocity is assumed to be Alfvén velocity, then ion inertial length is the same as ion gyroradius.

25.6.2 Current sheet

The grid is a square with $-0.5 \leq x \leq 0.5$ and $-0.5 \leq y \leq 0.5$. The density and pressure are uniform everywhere, with $\rho = 1$ and $p = \beta/2$ where β is an input parameter. For $|x| > 0.25$ we set $B_y/\sqrt{4\pi} = 1$, otherwise $B_y/\sqrt{4\pi} = -1$. The velocities are $u_x = A \sin(2\pi y)$ (where A is an amplitude) and $u_y = 0$. The "standard" test uses $\beta = 0.1$ and $A = 0.1$, although part of the point of this test is to see how small (large) a value of $\beta(A)$ is required to break the code.

Although we do not know the analytic solution for this problem, it may be an excellent test of the robustness of the algorithm. For ideal MHD, initially the solution should be nonlinearly (???) polarized Alfvén waves propagating along the field in the y-direction (which quickly generate magnetosonic waves since the magnetic pressure does not remain constant). However, because of the two current sheets in the problem (at $x = \pm 0.25$), reconnection inevitably occurs. Because $\beta < 1$, this reconnection drives strong over-pressurized regions that launch magnetosonic waves transverse to the field. Moreover, as reconnection changes the topology of the field lines, magnetic islands will form, grow, and merge. The point of the test is to make sure the algorithm can follow this evolution for as long as possible without crashing. Keeping $\nabla \cdot \mathbf{B} = 0$ as the field toplogy undergoes complex changes could be important.

25.7 Divergence-free Field test

25.7.1 Magnetic field loop

Appendix

This is the appendix.

```
example_dataframe()
```

26 References

Abraham-Shrauner, Barbara. 1967. "Shock Jump Conditions for an Anisotropic Plasma." *Journal of Plasma Physics* 1 (3): 379–81. https://doi.org/10.1017/S0022377800003366.

Amano, Takanobu, Katsuaki Higashimori, and Keisuke Shirakawa. 2014. "A Robust Method for Handling Low Density Regions in Hybrid Simulations for Collisionless Plasmas." *Journal of Computational Physics* 275: 197–212. https://doi.org/10.1016/j.jcp.2014.06.048.

Arnoldy, Roger L, MJ Engebretson, RE Denton, JL Posch, Marc R Lessard, NC Maynard, DM Ober, et al. 2005. "Pc 1 Waves and Associated Unstable Distributions of Magnetospheric Protons Observed During a Solar Wind Pressure Pulse." *Journal of Geophysical Research: Space Physics* 110 (A7). https://doi.org/10.1029/2005JA011041.

Bame, SJ, JR Asbridge, WC Feldman, S Peter Gary, and Michael D Montgomery. 1975. "Evidence for Local Ion Heating in Solar Wind High Speed Streams." *Geophysical Research Letters* 2 (9): 373–75. https://doi.org/10.1029/GL002i009p00373.

Bellan, Paul M. 2008. *Fundamentals of Plasma Physics*. Cambridge University Press.

Bezanson, Jeff, Alan Edelman, Stefan Karpinski, and Viral B Shah. 2017. "Julia: A Fresh Approach to Numerical Computing." *SIAM Review* 59 (1): 65–98.

Birn, J, JF Drake, MA Shay, BN Rogers, RE Denton, M Hesse, M Kuznetsova, et al. 2001. "Geospace Environmental Modeling (GEM) Magnetic Reconnection Challenge." *Journal of Geophysical Research: Space Physics* 106 (A3): 3715–19. https://doi.org/10.1029/1999JA900449.

Bruno, Roberto, and Vincenzo Carbone. 2013. "The Solar Wind as a Turbulence Laboratory." *Living Reviews in Solar Physics* 10 (1): 1–208. https://doi.org/10.12942/lrsp-2013-2.

Budden, Kenneth George. 1961. *Radio Waves in the Ionosphere. Radio Waves in the Ionosphere.*

Chen, Francis F. 2016. *Introduction to Plasma Physics and Controlled Fusion*. Vol. 1. Springer.

Chen, Liu, Fulvio Zonca, and Yu Lin. 2021. "Physics of Kinetic Alfvén Waves: A Gyrokinetic Theory Approach." *Reviews of Modern Plasma Physics* 5 (1): 1–37. https://doi.org/10.1007/s41614-020-00049-3.

Chew, GF, ML Goldberger, and FE Low. 1956. "The Boltzmann Equation and the One-Fluid Hydromagnetic Equations in the Absence of Particle Collisions." *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 236 (1204): 112–18. https://doi.org/10.1098/rspa.1956.0116.

Dai, Lei, and Chi Wang. 2023. "Kinetic Alfven Wave Eigenmode in Magnetosphere Magnetic Reconnection." *Reviews of Modern Plasma Physics* 7 (1): 1–38. https://doi.org/10.1007/s41614-022-00107-y.

Daldorff, Lars KS, Gábor Tóth, Tamas I Gombosi, Giovanni Lapenta, Jorge Amaya, Stefano Markidis, and Jeremiah U Brackbill. 2014. "Two-Way Coupling of a Global Hall Magnetohydrodynamics Model with a Local Implicit Particle-in-Cell Model." *Journal of Computational Physics* 268: 236–54. https://doi.org/10.1016/j.jcp.2014.03.009.

Erkaev, NV, DF Vogl, and HK Biernat. 2000. "Solution for Jump Conditions at Fast Shocks in an Anisotropic Magnetized Plasma." *Journal of Plasma Physics* 64 (5): 561–78. https://doi.org/10.1017/S002237780000893X.

Furth, Harold P, John Killeen, and Marshall N Rosenbluth. 1963. "Finite-Resistivity Instabilities of a Sheet Pinch." *The Physics of Fluids* 6 (4): 459–84. https://doi.org/10.1063/1.1706761.

Glassmeier, Karl-Heinz, Carsten Othmer, Rüdiger Cramm, Martin Stellmacher, and Mark Engebretson. 1999. "Magnetospheric Field Line Resonances: A Comparative Planetology Approach." *Surveys in Geophysics* 20 (1): 61–109. https://doi.org/10.1023/A:1006659717963.

Glauert, Sarah A, Richard B Horne, and Nigel P Meredith. 2014. "Three-Dimensional Electron Radiation Belt Simulations Using the BAS Radiation Belt Model with New Diffusion Models for Chorus, Plasmaspheric Hiss, and Lightning-Generated Whistlers." *Journal of Geophysical Research: Space Physics* 119 (1): 268–89. https://doi.org/10.1002/2013JA019281.

Gurram, Harsha, Jan Egedal, and W Daughton. 2021. "Shear Alfvén Waves Driven by Magnetic Reconnection as an Energy Source for the Aurora Borealis." *Geophysical Research Letters* 48 (14): e2021GL094201. https://doi.org/10.1029/2021GL094201.

Hameiri, E, P Laurence, and M Mond. 1991. "The Ballooning Instability in Space Plasmas." *Journal of Geophysical Research: Space Physics* 96 (A2): 1513–26.

Hill, P, G Paschmann, RA Treumann, W Baumjohann, N Sckopke, and H Lühr. 1995. "Plasma and Magnetic Field Behavior Across the Magnetosheath Near Local Noon." *Journal of Geophysical Research: Space Physics* 100 (A6): 9575–83. https://doi.org/10.1029/94JA03194.

Hubert, D, C Lacombe, CC Harvey, M Moncuquet, CT Russell, and MF Thomsen. 1998. "Nature, Properties, and Origin of Low-Frequency Waves from an Oblique Shock to the Inner Magnetosheath." *Journal of Geophysical Research: Space Physics* 103 (A11): 26783–98. https://doi.org/10.1029/GL016i002p00159.

Hubert, D, C Perche, CC Harvey, C Lacombe, and CT Russell. 1989. "Observation of Mirror Waves Downstream of a Quasi-Perpendicular Shock." *Geophysical Research Letters* 16 (2): 159–62. https://doi.org/10.1029/GL016i002p00159.

Hudson, PD. 1970. "Discontinuities in an Anisotropic Plasma and Their Identification in the Solar Wind." *Planetary and Space Science* 18 (11): 1611–22. https://doi.org/10.1016/0032-0633(70)90036-X.

Hughes, W Jeffrey. 1994. "Magnetospheric ULF Waves: A Tutorial with a Historical Perspective." *Solar Wind Sources of Magnetospheric Ultra-Low-Frequency Waves* 81: 1–11. https://doi.org/10.1029/GM081p0001.

Ji, Hantao, William Daughton, Jonathan Jara-Almonte, Ari Le, Adam Stanier, and Jongsoo Yoo. 2022. "Magnetic Reconnection in the Era of Exascale Computing and Multiscale

Experiments." *Nature Reviews Physics* 4 (4): 263–82. https://doi.org/10.1038/s42254-021-00419-x.

Kepko, Larry, RL McPherron, O Amm, S Apatenkov, Wolfgang Baumjohann, Joachim Birn, M Lester, R Nakamura, Tuija I Pulkkinen, and V Sergeev. 2015. "Substorm Current Wedge Revisited." *Space Science Reviews* 190 (1): 1–46. https://doi.org/10.1007/s11214-014-0124-9.

Kilpua, Emilia, and Hannu Koskinen. 2017. *Introduction to Plasma Physics*. Limes ry.

Kivelson, Margaret G, and David J Southwood. 1985. "Resonant ULF Waves: A New Interpretation." *Geophysical Research Letters* 12 (1): 49–52. https://doi.org/10.1029/GL012i001p00049.

———. 1986. "Coupling of Global Magnetospheric MHD Eigenmodes to Field Line Resonances." *Journal of Geophysical Research: Space Physics* 91 (A4): 4345–51. https://doi.org/10.1029/JA091iA04p04345.

Kuznetsov, VD, and AI Osin. 2018. "On the Parallel Shock Waves in Collisionless Plasma with Heat Fluxes." *Physics Letters A* 382 (31): 2052–54. https://doi.org/10.1016/j.physleta.2018.05.029.

Lacombe, Catherine, and Gérard Belmont. 1995. "Waves in the Earth's Magnetosheath: Observations and Interpretations." *Advances in Space Research* 15 (8-9): 329–40. https://doi.org/10.1016/0273-1177(94)00113-F.

Livadiotis, G, and DJ McComas. 2013. "Understanding Kappa Distributions: A Toolbox for Space Science and Astrophysics." *Space Science Reviews* 175 (1): 183–214. https://doi.org/10.1007/s11214-013-9982-9.

Merkin, VG, and JG Lyon. 2010. "Effects of the Low-Latitude Ionospheric Boundary Condition on the Global Magnetosphere." *Journal of Geophysical Research: Space Physics* 115 (A10). https://doi.org/10.1029/2010JA015461.

Phan, TD, DE Larson, RP Lin, JP McFadden, KA Anderson, CW Carlson, RE Ergun, et al. 1996. "The Subsolar Magnetosheath and Magnetopause for High Solar Wind Ram Pressure: WIND Observations." *Geophysical Research Letters* 23 (10): 1279–82. https://doi.org/10.1029/96GL00845.

Radoski, Henry R, and Robert L Carovillano. 1966. "Axisymmetric Plasmasphere Resonances: Toroidal Mode." *The Physics of Fluids* 9 (2): 285–91. https://doi.org/10.1063/1.1761671.

Raeder, J, RJ Walker, and M Ashour-Abdalla. 1995. "The Structure of the Distant Geomagnetic Tail During Long Periods of Northward IMF." *Geophysical Research Letters* 22 (4): 349–52. https://doi.org/10.1029/94GL03380.

Rahman, Sk Mashfiqur, and Omer San. 2019. "A Relaxation Filtering Approach for Two-Dimensional Rayleigh–Taylor Instability-Induced Flows." *Fluids* 4 (2): 78.

Rambo, PW. 1995. "Finite-Grid Instability in Quasineutral Hybrid Simulations." *Journal of Computational Physics* 118 (1): 152–58. https://doi.org/10.1006/jcph.1995.1086.

Samson, JC, BG Harrold, JM Ruohoniemi, RA Greenwald, and ADM Walker. 1992. "Field Line Resonances Associated with MHD Waveguides in the Magnetosphere." *Geophysical Research Letters* 19 (5): 441–44. https://doi.org/10.1029/92GL00116.

Scarf, FL, JH Wolfe, and RW Silva. 1967. "A Plasma Instability Associated with Thermal

Anisotropies in the Solar Wind." *Journal of Geophysical Research* 72 (3): 993–1005. https://doi.org/10.1029/JZ072i003p00993.

Schekochihin, Alexander A. 2022. "MHD Turbulence: A Biased Review." *Journal of Plasma Physics* 88 (5): 155880501. https://doi.org/10.1017/S0022377822000721.

Slavin, James A, and Robert E Holzer. 1981. "Solar Wind Flow about the Terrestrial Planets 1. Modeling Bow Shock Position and Shape." *Journal of Geophysical Research: Space Physics* 86 (A13): 11401–18. https://doi.org/10.1029/JA086iA13p11401.

Song, P, CT Russell, and SP Gary. 1994. "Identification of Low-Frequency Fluctuations in the Terrestrial Magnetosheath." *Journal of Geophysical Research: Space Physics* 99 (A4): 6011–25. https://doi.org/10.1029/93JA03300.

Sonnerup, B. U. Ö. 1979. "Space Plasma Physics: The Study of Solar-System Plasmas."

Stone, James M, Thomas A Gardiner, Peter Teuben, John F Hawley, and Jacob B Simon. 2008. "Athena: A New Code for Astrophysical MHD." *The Astrophysical Journal Supplement Series* 178 (1): 137. https://doi.org/10.1086/588755.

Swanson, Donald Gary. 2012. *Plasma Waves*. Elsevier.

Tóth, Gábor. 2000. "The · b= 0 Constraint in Shock-Capturing Magnetohydrodynamics Codes." *Journal of Computational Physics* 161 (2): 605–52. https://doi.org/https://doi.org/10.1006/jcph.2000.6519.

Turc, L, OW Roberts, D Verscharen, AP Dimmock, P Kajdič, M Palmroth, Y Pfau-Kempf, et al. 2022. "Transmission of Foreshock Waves Through Earth's Bow Shock." *Nature Physics*, 1–9. https://doi.org/10.1038/s41567-022-01837-z.

Walker, ADM, Jo M Ruohoniemi, KB Baker, RA Greenwald, and JC Samson. 1992. "Spatial and Temporal Behavior of ULF Pulsations Observed by the Goose Bay HF Radar." *Journal of Geophysical Research: Space Physics* 97 (A8): 12187–202. https://doi.org/10.1029/92JA00329.

Wang, Liang, Ammar H Hakim, Jonathan Ng, Chuanfei Dong, and Kai Germaschewski. 2020. "Exact and Locally Implicit Source Term Solvers for Multifluid-Maxwell Systems." *Journal of Computational Physics* 415: 109510. https://doi.org/10.1016/j.jcp.2020.109510.

Wright, Andrew N. 1994. "Dispersion and Wave Coupling in Inhomogeneous MHD Waveguides." *Journal of Geophysical Research: Space Physics* 99 (A1): 159–67. https://doi.org/10.1029/93JA02206.

Zhao, L-L, GP Zank, L Adhikari, Q Hu, JC Kasper, SD Bale, KE Korreck, et al. 2020. "Identification of Magnetic Flux Ropes from Parker Solar Probe Observations During the First Encounter." *The Astrophysical Journal Supplement Series* 246 (2): 26. https://doi.org/10.3847/1538-4365/ab4ff1.