PHYSICS 425 NOTES: PLASMA ASTROPHYSICS

Original version (2007): Jean Eilek
Physics Department, New Mexico Tech
Socorro, NM 87801, U.S.A.
JEILEK@AOC.NRAO.EDU

Minor revisions (2008, 2009+): Lisa Young

The goal of this course is to explore the "physics of astrophysics". What physics governs the behavior of the astrophysical objects we observe? How can we interpret our observations, in light of the relevant physics, to understand what's going on inside a particular star, nebula or galaxy? To reach this goal we need to bring together a diverse range of physics – some of which you will have seen in other courses, some of which will be new. These ideas cover a broad range of material, not all of which is in a single textbook. Thus, we've got the course notes for our text.

You should note *units and dimensions*. These notes are in cgs, as is most of the astrophysical literature. That makes very little difference for "rocks" (analyses that involve mass, length, time); but it makes a big difference for electrodynamics. The $\mathbf{E}$ and $\mathbf{B}$ fields, as well as the fundamental charge, have different dimensions in cgs than in SI; and the coupling constants in Maxwell's equations are different.

Contents

1 Astrophysical plasmas we'll meet

We will be using several different astrophysical systems as examples of plasma astrophysics. While you've probably run into all of these before, it may be worth collecting a brief description of them in one place. For some objects I'll put in "typical" parameters (size, density, temperature, etc); for others, such numbers are hard to pin down, & we'll introduce them as needed later on.

The terrestrial ionosphere and magnetosphere Start at the surface of the earth: the atmosphere you're breathing, as you read this, is very close to charge neutral. The density of electrons is only $\sim 10 - 100$ cm^{-3} (how does this compare to the total number density of the low-altitude atmosphere?). But this changes when you get to altitudes $\gtrsim 100$ km. The ionized fraction, and total ionized number density, grows suddenly (due to what? what ionizes the upper atmosphere?), until at a few hundred km you reach the highly ionized *ionosphere*. The number density here $\sim 10^6$ cm^{-3} (but remember this changes rapidly with height, due to the exponential density of the atmosphere); the temperature $\sim 10^3$ K. This is our nearest-by example of an astrophysical plasma; we see its signature in the propagation, bending (refraction) or non-transmission or radio waves.

Continuing upwards, the ionospheric density drops off more rapidly than does the earth's magnetic field (which is a dipole field, ~ 1 G at the earth's surface). Thus you reach the *magnetosphere*, a region where the plasma density is small and the plasma is dominated dynamically by MHD and plasma effects. On the sunward side of the planet, the magnetopause extends to several earth radii, and is bounded by the bow shock where the solar wind runs into the earth's B field. On the "downstream" (antisunward) side of the planet, the magnetotail extends to several tens of R_E.

The solar corona and solar wind Now, start at the sun's surface and move outward ... the *photosphere* is the visible "surface" of the sun, with $T \simeq 5800$ K. It's about 500 km thick, mostly neutral. We know a good bit about magnetic fields here, because we can observe sunspots & related phenomena. The mean field $B \sim 1$ G; sunspot fields are ~ 1 kG. The *chromosphere* is a region 2000-3000 km thick, just above the photosphere. Going upwards, the temperature first drops to ~ 4000K, then rises to $\sim 10^4$ K; the density drops rapidly, from $\gtrsim 10^{15}$ cm^{-3} at the base to $\sim 10^9$ cm^{-3} at the top. Above this is the *corona*, in which the temperature rises abruptly to $\gtrsim 10^6$ K, and the density continues to drop (but more gently). Magnetic fields are "outlined" by the beautiful filaments and prominences which extend from the chromosphere up into the corona. These structures have typical thicknesses ~ 6000 km, lengths $\sim 100,000$ km, and extend to $\sim 50,000$ km above the surface. The plasma in a quiescent prominence is about 300 times colder and denser than the surrounding coronal gas; $B \sim 10$ G is usually quoted as typical.

The corona is also the source of the *solar wind*. The outflow starts in coronal holes, large open regions (visible in X-ray images) that are associated with "open" B field lines and high-speed solar wind streams. Solar wind numbers are of course a function of radius. At the earth, the wind density ~ 10 cm^{-3}, $T \sim 10^5$ K, $B \sim 10^{-4}$ G, and the wind is very supersonic, with $v \sim 500$ km/s. Elsewhere in the wind, the velocity changes only slowly, $v^2 \propto \ln r$, and we think the temperature doesn't change by much; the behavior of the other parameters (density, field, pressure) is set by conservation laws.

Stars You know stars are held together by gravity, and supported against collapse by their internal pressure. They are the classic example of *hydrostatic equilibrium*, which we'll work with later in the course. In addition, stars are almost totally ionized (how do you know? What's the temperature, interior and surface, of your favorite mass of star?), and thus are plasmas. We also know that stars are magnetized; we measure the sun's surface B field directly, and that of other stars indirectly (through X-ray and radio observations of stellar flares, for instance). That means that *magnetohydrodynamic* (MHD) effects – the forces exerted by the B fields and the currents which support them – must be considered. We won't talk a lot about "normal" stars in this course, but you should remember that MHD effects are important to many aspects of their formation and evolution.

HII regions Put a hot young star down in a region of neutral ISM. If the star is hot enough – if it produces a substantial amount of UV photons (with $h\nu > 13.6$

eV) – it will photoionize the nearby ISM, making an HII region. These vary a lot across the galaxy in their density and size. The spectrum ranges from young, ultra-compact ones, with diameter < 0.03 pc and density $> 10^4$ cm^{-3} (which are hidden behind thick dust clouds, so that you can only see them with radio or IR); to the older, big, bright, famous ones, which have blown off their dust shrouds and are beautiful optical sources (such as the Orion nebula; diameter ~ 1 pc, very inhomogeneous, but maybe it has a typical density $\sim 10^3$ cm^{-3}). The temperature of a photoionized region is regulated by the microphysics: $T \lesssim 10^4$K always (see Physics 426 for the proof).

Supernova remnants The physical picture is easy to describe: a star goes bang, and ejects a rapidly moving shell of matter. This shell moves out into the local interstellar medium (or the wind ejected by the pre-SN star), pushing the ambient matter ahead of it and decelerating as it goes. We'll talk about SNR in more detail next term; for now, typical sizes are a few pc (with older ones being bigger, of course). Typical densities and temperatures are harder ... the outer shell is defined by a shock, and the conditions therein are complex. Inside of the shell, the shocked gas is hot, $\sim 10^7$ K – we see it in X-rays.

Another type of SNR, about 10% of the population, is a *filled remnant*, also called a *plerion* or *pulsar wind nebula*. These are the remnants with active pulsars inside; the relativistic-plasma wind from the pulsar fills the SNR. The Crab nebula is a well-known example of this.

Our galaxy & the interstellar medium Our galaxy is a typical big spiral. It is rotation supported in the plane (stars + gas in circular orbits) and supported by "heat" (*i.e.*, random motions) transverse to the plane. The stellar disk extends ~ 15 kpc from the center, with the sun at 8.5 kpc out; the gas disk in a typical spiral extends much further, out to $\sim 30 - 50$ kpc. The disk thickness depends a bit on which species you measure (stars, hot gas, cool gas ..); typically it's $\sim 1/2$ kpc thick.

The diffuse interstellar medium (ISM) is multiphase: there is a cold, neutral component; a "warm", mostly ionized component; and a hot ("coronal gas") component. As everywhere in diffuse astrophysical plasmas, the chemical composition is almost all hydrogen; all heavier elements contribute no more than a few per cent. Each of these phases has a range of temperatures and densities. For "typical" numbers, let's say the cold, neutral HI is at $n \sim 1$ cm^{-3} and $T \sim 100$ K; the warm, partly ionized HII is at $n \sim 0.2$ cm^{-3} and $T \sim 6000$ K; and the coronal gas is at $n \sim 10^{-2}$cm^{-3} and $T \sim 10^6$ K. Note that each of these phases are in approximate pressure balance: $p = nk_BT$ is about the same for each. Measured as an energy density each phase is at ~ 1 eV/cm^3. The ISM is of course magnetized, and also contains an energetically important relativistic plasma (the cosmic rays); each of these is also at ~ 1 eV/cm^3.

The galaxy also has a hot, extended halo. We detect it mostly through its synchrotron emission (from relativistic particles undergoing gyromotion in the local magnetic field). Based on other galaxies, our halo probably extends a few kpc above and below the plane; information on its composition (thermal gas, B field, energy density, etc) is harder to come by.

Cosmic rays These are worth their own paragraph. Most of the ISM is "thermal" – that is it has a well-defined temperature (subrelativistic: $kT \ll mc^2$) and a Maxwellian distribution of the particle velocities. However it also contains a component of highly energetic ($E = \gamma mc^2$; the Lorentz factor $\gamma \gg 1$) charged particles. These particles are accelerated "somewhere" in or out of the galaxy and remain tied to B field lines as they move through the ISM. They are "nonthermal": their energy/velocity distribution is not a Maxwellian, rather a power law (which means they haven't had time to thermalize *via* collisions with the ISM. Their lowest energy $\sim 10^{11}$ eV (at least the lowest that we detect); their highest energy $\sim 10^{20} - 10^{21}$eV.

Elliptical galaxies These are the big round ones. Sizes: they typically have an inner core, radius $\sim 1-2$ kpc, and an extended, power-law outer halo (in which the stellar density $\propto 1/r^x$, where $x \sim 2 - 3$. They are supported by the random motions of their stars; they show little or no organized rotation. It used to be thought that they had no ISM, because you couldn't see it in optical pictures, and because they show little ongoing star formation; we now know that's wrong. Their ISM is mostly hot, $T \sim 10^7$ K (we see it with X-ray telescopes) – so it's less inclined to star formation than the cooler, denser ISM in a spiral galaxy. Its density distribution is roughly similar to that of the stars; the density ~ 0.1 cm^{-3} in the core, and falls off roughly as a power law outside the core.

Ellipticals do, however, contain a smaller amount of cooler ISM, which can be detected with radio and millimeter-wave telescopes. It stands to reason – by analogy with spiral galaxies and clusters of galaxies – that the ISM in an elliptical should also be magnetized and contain a relativistic plasma component. There has been very little observational work on this question, so I can't offer any numbers for the field or the cosmic rays here.

Radio jets and radio galaxies Let matter accrete onto a compact object (which could be the core of a protostar, or a neutron star, or a black hole). Some part of the accreted matter and energy is driven away, into an outflow – which in some (many) cases can be highly collimated. The outflowing plasma is very often (probably always) magnetized, and it often contains a relativistic particle component which we see in radio frequencies through its synchrotron radiation. These are thus called *radio jets*. Their size varies from AU (protostellar jets) to a few pc (jets from galactic accretion sources) to 100's of kpc (jets from supermassive black holes in active galaxies). Their internal state (density, temperature, composition) also varies a lot between these different objects. What they have in common is that they all involve energetic, relativistic plasma which is controlled dynamically by MHD effects.

Clusters of galaxies These are the largest self-gravitating structures in the universe. Their structure is sort of like a big elliptical galaxy, with an inner core, radius $\sim 300 - 500$ kpc, and an outer halo where density (of gas or galaxies) decays roughly as a power law. The outer halo can be traced to a radius of more than a Mpc in big clusters. The cluster has a plasma atmosphere – the *intracluster medium*, ICM. It is composed partly of primordial material which accumulated as the cluster formed, and partly of processed material that has been through star formation in the galaxies and then ejected back into the ICM. Typical temperatures, $\sim 10^8$ K; typical densities $\sim 10^{-3}$ cm^{-3} in the core (and again decaying outwards).

We are beginning to learn that the ICM is magnetized, and that it also contains a relativistic ("cosmic ray") component. Numbers here are not yet well determined. A "typical" field seems to be $B \lesssim \mu$G, but B can reach tens of μG in high field regions. The mean cosmic ray energy density might be $\sim 1 - 10\%$ of the thermal plasma pressure.

What's left? Here's a question for the student: what astrophysical objects have I *not* mentioned? How many objects can you think of that are neither plasmas nor affected somehow by MHD affects?

Key points

After each chapter I'll try to highlight the most important issues in that chapter. Here, it's just the objects themselves: you should be familiar with the range of astrophysical plasmas we'll encounter, including typical sizes, densities, *etc* for each one.

2 Some basic plasma tools

In these notes I'm storing the ideas and important expressions for some basic plasma tools.

2.1 Distribution functions

This is an important tool for understanding the microphysics of a plasma: what is the *distribution function* (DF) of plasma particles (free charges) with momentum, or energy? In terms of momentum, this is defined so that $f(\mathbf{p})d\mathbf{p}$ is "the number of particles at $\mathbf{p}$; the total number of particles (usually per volume) is

$$n = \int f(\mathbf{p})d\mathbf{p} \tag{2.1}$$

If the particle distribution in momentum space is isotropic, the $d\mathbf{p}$ can be expanded out to give

$$\text{isotropic}: \quad n = \int f(p) 4\pi p^2 dp \tag{2.2}$$

You've probably seen the *Maxwell-Boltzmann* distribution, for a thermal, subrelativistic plasma:

$$f(p) = Ae^{-p^2/2mkT} \tag{2.3}$$

where the constant A is written in terms of the total number density n by plugging (2.3) into (2.2). We can also take suitable moments of the DF to get other useful things. For instance, the mean kinetic energy per particle, averaged over the MB distribution, is

$$\begin{aligned} \langle KE \rangle &= \frac{1}{n} \int A \frac{p^2}{2m} e^{-p^2/2mkT} 4\pi p^2 dp \\ &= \frac{1}{2} m v_{th}^2 = \frac{3}{2} kT \end{aligned} \tag{2.4}$$

Note, the $1/n$ term is part of the definition of the mean energy (do you understand why?). This expression (2.4) also defines the *thermal speed*, $v_{th} = \sqrt{3kT/m}$ (this holds for monatomic particles; more complex molecules have more degrees of freedom & a different numerical factor).

We can also work with relativsitic plasmas.

• You recall that the total energy of a relativistic particle is given by

$$E^2 = p^2c^2 + m^2c^4 \tag{2.5}$$

we also have the definition

$$E = \gamma mc^2 \tag{2.6}$$

where

$$\beta = v/c \quad \text{and} \quad \gamma^2 = 1/(1-\beta^2) \tag{2.7}$$

In the limit $E \gg mc^2$, we also have $E \simeq pc$; thus the integrals in (2.1) or (2.2) can be written in terms of particle energy E (or just the Lorentz factor γ). An alternative DF, motivated by observations (for instance of cosmic rays) is often used with relativistic plasmas: the *power law* distribution,

$$f(E) = f_o E^{-s}\ , \quad E_1 \le E \le E_2 \tag{2.8}$$

or

$$n(\gamma) = n_o \gamma^{-s}\ , \quad \gamma_1 \le \gamma \le \gamma_2 \tag{2.9}$$

The exponent s depends on the system; the scaling constant f_o or n_o connects to the total number (or number density) of particles.

• NOTE the way these DF's are normalized: we want the total number of particles to be, say,

$$N = \int f(E)dE = \int n(\gamma)d\gamma \tag{2.10}$$

This means that $f(E)dE = n(\gamma)d\gamma$ .. so that $f(E)$ and $n(\gamma)$ have different units.[1]

2.2 Collective effects

Two important pieces of physics appear when we think about a "clump" of plasma.

2.2.1 Plasma waves

These will be treated more formally later on, but we can see the basics with a simple cartoon. Start with a layer of charge-neutral plasma, with number density $n_+ = n_- = n$. Now displace the electrons, relative to the (heavier) positive charges, by some distance ξ in the x-direction, as in the figure. The excess charge surface density is $ne\xi$; this generates an E field, $\mathbf{E} = 4\pi ne\xi\hat{\mathbf{x}}$.[2] Each charge layer feels a net force equal to its charge times the E field. Thus, we can write an equation of motion for the electrons,

$$m_e n\xi \frac{d^2\xi}{dt^2} = (ne\xi)(4\pi ne\xi) \tag{2.11}$$

[1] To the student: what are those units?? Work out the dimensions – they will look funny, but that's the way it is.

[2] To the student: why?? think about Gauss's law and capacitors.

But this is clearly an equation for simple harmonic motion: $\xi(t) = \xi_o e^{-i\omega_p t}$, where ξ_o is the amplitude of the displacement, and

$$\omega_p^2 = 4\pi n e^2 / m_e \tag{2.12}$$

is the (square of the) *electron plasma frequency*. This is a fundamental mode of oscillation of the plasma; it's very easy to excite such waves, and in fact we expect any plasma to have some level, albeit low, of plasma wave turbulence.

We'll see later that ω_p is a cutoff frequency for EM wave propagation: only waves with $\omega > \omega_p$ can propagate in an unmagnetized plasma. We'll also see later that $\nu_p = \omega_p / 2\pi$ is the frequency at which some plasmas emit coherent radiation (for instance in solar flares, or pulsar radio emission).

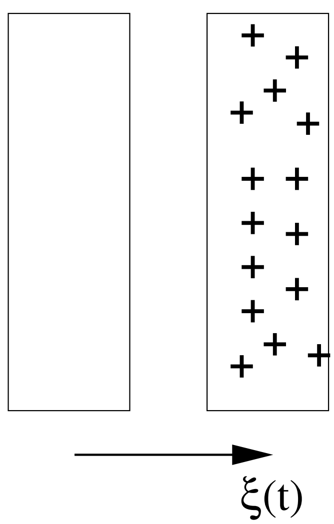

Figure 2.1 A simple cartoon illustrating plasma waves. The positive charges are displaced by $\xi(t)$ from the negatives; the attractive force turns this system into a simple harmonic oscillator with frequency ω_p.

2.2.2 Debye shielding

An important feature of plasmas is that their charges are very mobile; they can easily shield out any external E field we try to apply. Say we put a postive charge Q down somewhere in the plasma. Plasma particles of the opposite sign will scoot over to Q and form a charge cloud of the opposite sign around Q, thus neutralizing its effect on the rest of the system. If the plasma were cold (if thermal motions didn't matter), the charge cloud would be very thin around Q, and the shielding outside would be perfect. On the other hand, if the temperature is finite, particles on the edge of the cloud, where the E field is weak, have enough energy to escape from the potential well. Thus the nearby shielding is only partial – there is a region of finite size within which Q causes a finite E field. The size of this region is the *Debye length*; it is given by

$$\lambda_D^2 = kT / 4\pi n e^2 \tag{2.13}$$

This is one of the important length scales in plasma physics.

One immediate use for λ_D is to turn it into $N_D = (4\pi/3) n \lambda_D^3$ – which measures the number of particles in a "Debye sphere". If $N_D \gg 1$, then Debye shielding is indeed a valid statistical concept, and we can treat the plasma as macroscopically charge neutral (that's what we'll do in this course). On the other hand, at high temperatures and/or low densities, it may be that $N_D \lesssim 1$, and we need to worry about single-particle motion as well as macroscopic effects.

2.3 Single particle motions

We also need to understand how individual particles move within the plasma.

2.3.1 Gyromotion

You have seen this before (right?). Just recall the basic analysis: the equation of motion for a particle with charge q, in a $\mathbf{B}$ field, is (in cgs!)

$$\frac{d\mathbf{p}}{dt} = q\frac{\mathbf{v}}{c} \times \mathbf{B} \tag{2.14}$$

For a subrelativistic particle, with $\mathbf{p} = m\mathbf{v}$, and putting the $\hat{\mathbf{z}}$ axis along $\mathbf{B}$, the solution to (2.14) describes gyromotion:

$$\begin{aligned} v_{x,y} &= v_\perp e^{i\Omega t} e^{i\phi} \\ v_z &= v_{zo} \end{aligned} \tag{2.15}$$

where $v_\perp$ is the (constant) amplitude of the motion across $\mathbf{B}$; v_{zo} is the (constant) velocity along $\mathbf{B}$; ϕ is the phase of the x or y motions; and the *gyrofrequency* is

$$\Omega = \frac{qB}{mc} \tag{2.16}$$

From this we can also get the *gyroradius* (also called Larmor radius),

$$r_L = \frac{m v_\perp c}{qB} \tag{2.17}$$

Or, in words: the general motion of a charged particle in a $\mathbf{B}$ field can be described as *gyromotion* about a *guiding center*, plus motion of the guiding center through space. In this simple case, the guiding center just moves along $\mathbf{B}$; but that will change in the next section.

• How does this change for a relativistic particle? In the homework, you will show that the gyrofrequency and gyroradius depend on the particle's energy, as

$$\Omega = \frac{qB}{\gamma mc}\,; \quad r_L = \frac{\gamma m v_\perp c}{qB} \tag{2.18}$$

2.3.2 Particle drifts, external forces

Now let's expand the example above, by adding an $\mathbf{E}$ field. The equation of motion is now

$$\frac{d\mathbf{p}}{dt} = q\frac{\mathbf{v}}{c} \times \mathbf{B} + q\mathbf{E}. \tag{2.19}$$

If $\mathbf{E}$ has components $(0, E_y, E_z)$ (Fig. 2.2) the solution for a subrelativistic particle is

$$\begin{aligned} v_x &= v_\perp e^{i\Omega t} - \frac{E_y}{B} \\ v_y &= i v_\perp e^{i\Omega t} \\ v_z &= v_{zo} + \frac{qE_z}{m} t \end{aligned} \tag{2.20}$$

Thus, we see (i) simple acceleration along $\mathbf{B}$, if $\mathbf{E}$ has a component in that direction; and (ii) sideways drift, *across* $\mathbf{B}$, at a rate

$$\mathbf{v}_E = c\frac{\mathbf{E} \times \mathbf{B}}{B^2} \tag{2.21}$$

This is called "$\mathbf{E} \times \mathbf{B}$ drift". Comments:

• What causes the drift? Think about energetics: the particle alternately gains and loses energy, every half of its orbit. Thus, its Larmor radius gets alternately larger, then smaller. This leads to a net drift, as illustrated in the figure.

• We can also look at this particular drift in terms of Lorentz transforms. You may remember that $\mathbf{E}$ and $\mathbf{B}$ fields are not "relativistically pristine"; changing reference frames can turn $\mathbf{E}$ into $\mathbf{B}$, and vice versa. This connects directly to $\mathbf{E} \times \mathbf{B}$ drift. What direction do the particles go? Note, here, both positive and negative charges drift in the same direction. You can understand this from the cartoon; also note, $\mathbf{v}_E$ is independent of charge.

• We can generalize this. The key ingredient in what we just did, was the presence of a non-magnetic force

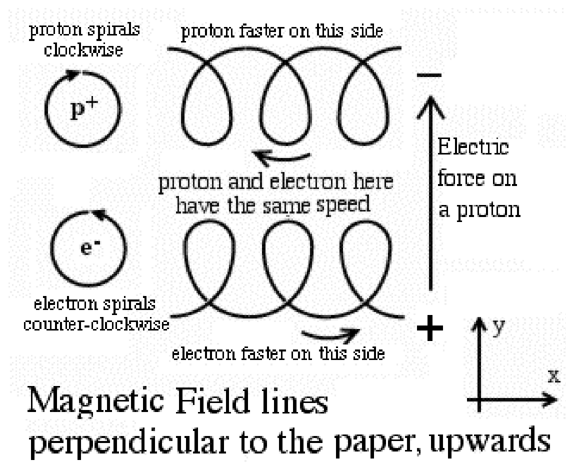

Figure 2.2 Illustrating $\mathbf{E} \times \mathbf{B}$ drift.

(call it $\mathbf{F}$) in the equation of motion. Repeating the above analysis for this F, we find a generalized drift velocity,

$$\mathbf{v}_F = c\frac{\mathbf{F} \times \mathbf{B}}{eB^2} \tag{2.22}$$

$\mathbf{F}$ can be anything relevant: gravity is the most common application.

2.3.3 Particle drifts, non-uniform B field

We also find drifts when a charge moves in a non-uniform $\mathbf{B}$ field. I'm not going to derive these formally (the algebra gets really tedious); rather we can do it by cartoon. One case is illustrated in Figure 2.3 – let $\mathbf{B}$ vary in space, and let $\nabla\mathbf{B}$ have a component perpendicular to $\mathbf{B}$. Once again, the Larmor radius of the particle will change during its gyro-orbit; and once again, this will cause a net drift across $\mathbf{B}$. One simple way to find the drift speed is to remember that a magnetic moment, $\boldsymbol{\mu}$, feels a force $\mathbf{F} = -\boldsymbol{\mu}\nabla\mathbf{B}$ when it sits in a nonuniform $\mathbf{B}$ field. You remember (of course....) that the gyromotion creates a magnetic moment (defined for a nonrelativistic particle),[3]

$$\mu = \frac{m v_\perp^2}{2B} \tag{2.23}$$

With this definition, the $\boldsymbol{\mu}\nabla\mathbf{B}$ force gives us the rate of "$\nabla\mathbf{B}$ drift":

$$\mathbf{v}_{\nabla\mathbf{B}} = \frac{v_\perp^2}{2\Omega}\frac{\mathbf{B} \times \nabla\mathbf{B}}{B^2} \tag{2.24}$$

[3] why does this definition make sense? To derive this, you'll need to know that the definition of magnetic moment, in cgs, is $\mu = Ia/c$, for a current I going in a circle of area a.

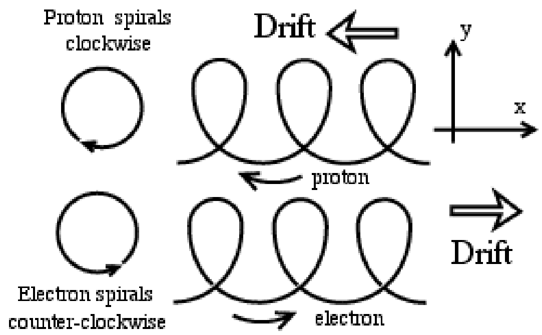

Figure 2.3 Illustrating ∇B drift. In this cartoon, $\mathbf{B}$ is out of the paper again; and it is stronger at the top of the drawing ($\nabla\mathbf{B} \parallel \hat{\mathbf{y}}$).

Another effect comes from a particle moving along a curved $\mathbf{B}$ line. Let $\mathbf{R}$ be radius of curvature of the field line, and define $\boldsymbol{\kappa} = \hat{\mathbf{R}}/R$.[4] The curved path creates a centrifugal force, and again causes a sideways *curvature drift*, at a rate

$$\mathbf{v}_{curv} = -\frac{v_\parallel^2}{R\Omega}\frac{\boldsymbol{\kappa}\times\mathbf{B}}{B} = -\frac{v_\parallel^2}{\Omega}\frac{\mathbf{R}\times\mathbf{B}}{R^2 B} \tag{2.25}$$

2.4 Adiabatic invariants

Let's continue with our particle in a nonuniform $\mathbf{B}$ field. It turns out that several useful constants of the motion can be found. We'll just look at one, the magnetic moment. Here's the result:

- If $\mathbf{B}$ is constant, or varies only slowly (compared to the gyroperiod), then μ is a constant of the motion.

Here's the outline of the proof. We're interested in nonuniform B fields. First, remember that if $\mathbf{B}$ changes with time (this is as seen by the particle in its gyro-orbit), it generates an EMF:

$$\oint \mathbf{E}\cdot d\mathbf{l} = -\frac{1}{c}\int \frac{\partial \mathbf{B}}{\partial t}\cdot d\mathbf{S}$$

so the rate of work done (current times EMF, right?) is

$$\frac{d}{dt}\left(\frac{1}{2}mv_\perp^2\right) = \frac{q\Omega}{2\pi}\frac{\pi r_L^2}{c}\frac{dB}{dt} \to \mu\frac{dB}{dt} \tag{2.26}$$

But also, from (2.23), the definition of μ:

$$\frac{d}{dt}(\mu B) = \frac{d}{dt}\left(\frac{1}{2}mv_\perp^2\right) \tag{2.27}$$

[4] Both $\mathbf{R}$ and $\boldsymbol{\kappa}$ are defined pointing *inward* relative to the arc of the circle (*i.e.* the curved field line).

Now, compare (2.26) and (2.27): clearly these two are consistent only if $d\mu/dt = 0$; thus we've proved that μ is constant.

- I did this for a nonrelativistic particle. It can be generalized to the relativistic case: if we define

$$\mu_{rel} = \frac{\gamma m v_\perp^2}{2B} = \frac{p_\perp v_\perp}{2B} \tag{2.28}$$

it can be shown that μ_{rel} is also a constant of the motion. So we can apply this analysis to cosmic rays – that's good.

2.5 Applications

Two applications are particularly interesing.

2.5.1 magnetic mirrors

This is a straightforward consequence of the invariance of μ. Think about a particle moving into a region of higher B (*i.e.*, converging field lines, in the usual cartoon). Because μ is constant, $v_\perp^2 \propto B$ must increase. But, the particle's energy is constant – so the gain in $v_\perp$ must come at the expense of $v_\parallel$. This clearly has a limit: when all of the particle's initial energy has been turned into $v_\perp^2$, there is no more $v_\parallel$, and the particle can't go any further along that field line. This point is called a *magnetic mirror*.

Mirrors are important for particle trapping .... if you want to keep a plasma confined magnetically, you might create it in a region of low B, which is bounded (at least along the field lines) by a region of high B. That's a "magnetic bottle". If each end of the confining field has a high-B region, particles are in principle trapped forever ... they can move back and forth along a field line (while undergoing gyromotion), but they can never escape the region (unless you add more physics ... as we'll talk about).

2.5.2 Particle acceleration

Magnetic mirror geometries can be used to accelerate the charged particles trapped therein.

For one method, think about a closed magnetic bottle, and now contrive to have the high-B regions approach each other. The trapped particles will gain a little bit of energy each time they collide with the mirror point (*e.g.*, think about a particle bouncing off a moving brick wall – go back to basic physics). This is called *Fermi acceleration*; it was the first mechanism proposed to accelerate cosmic rays.

For another method, let the magnetic bottle's geometry stay fixed (no moving end mirrors), but now let the B field go up and down with time, in some cyclic fashion. In the $dB/dt > 0$ phase, particles will gain perpendicular energy. If nothing else happens to the particles, their perpendicular energy will go up and down with the field (not very interesting). But what if the particles collide with each other before the field starts its downwards cycle? A collision between two particles will, statistically, redistribute energy between $v_\perp$ and $v_\|$. The parallel part of the velocity will not decrease when B goes back down – so the particle will have a net gain of energy on each cycle. This is *betatron acceleration* or *magnetic pumping*.

2.5.3 Earth's radiation belts

The region above the (mostly neutral) atmosphere, out to about 10 Earth radii, is more properly called the "inner magnetosphere" – but I'm using the older name, here, to differentiate from the full magnetosphere, which we'll visit later. The motion of particles in this region is mainly governed by single particle effects – gyromotion, $\mathbf{E} \times \mathbf{B}$ drifts, $\nabla \mathbf{B}$ drifts, and adiabatic invariants, such as we've seen in this chapter.

Some authors talk about three particle populations in this region. (1) The *cool, thermal plasma*, at particle energies below 100 eV, is mostly the magnetospheric extension of the ionosphere (*i.e.* terrestrial in origin). (2) The *ring current plasma*, particle energies from 100 eV up to several hundred keV, is "injected" into the radiation belts from magnetic storms in the Earth's magnetotail. Their drift motions result in a net current around the Earth, hence the name. (3) *Trapped radiation belt* or *van Allen belt* particles are even more energetic, at and above 1 MeV per particle. These are the ones whose radiation was detected in the 1950's (and thought to be a sign of nefarious, warlike activities on the part of other terrestrial nations); we now know these particles come from the solar wind, *via* the magnetotail.

Key points

• Relativistic particles: if you don't remember $p = \gamma\beta mc$, $E = \gamma mc^2$, etc., you should go back and review this material.

• Distribution functions: how they're defined; thermal *vs.* non-thermal.

• Plasma waves: what they are, what is ω_p?

• Debye shielding: what it is, what is λ_D?

• gyromotion: Ω, r_L, in cgs; for subrelativistic and relativstic.

• particle drifts: $\mathbf{E} \times \mathbf{B}$, $\nabla \mathbf{B}$, curvature drift.

• adiabatic invariants: μ is constant; magnetic mirrors.

3 Collisions in Plasmas

We want to understand the behavior of a collection of charges: that is, a plasma. We start with what a "collision" means for a set of charges. To begin, we recall the basics of hard-sphere collisions. If a "gas" of billiard balls, say, has a number density n and each particle has a random velocity v and a radius a, we define the collision cross section,

$$\sigma_c = \pi a^2 \tag{3.1}$$

From this we find the mean free path (the average distance between collisions),

$$\lambda \simeq \frac{1}{n\sigma_c} \tag{3.2}$$

and the mean time between collisions,

$$\tau_{coll} \simeq \frac{1}{n\sigma_c v}\ . \tag{3.3}$$

This last can be inverted to describe the collision rate per particle, $\tau_{coll}^{-1} \simeq n\sigma_c v$.

For hard spheres this analysis is straightforward, of course; they will not interact unless there is a direct "hit", and the geometrical cross section is the relevant one to describe energy exchange. Neutral atoms and molecules behave similarly, in that they need a very close hit; their cross sections can be calcalated from basic physics. Typical atomic cross sections $\sim 10^{-14}\ \mathrm{cm}^2$. Another "hard-sphere" type of calculation would describe direct hits between stars. Two stars must pass within a couple of stellar radii of each other for either of them to be strongly disturbed by the encounter; the cross section could be estimated from eq. (1), with $a \sim 2 - 3 \times R_*$.

3.1 The Spitzer collision cross section

There is, however, another type of encounter which is important in astrophysics: a long-range encounter between two objects which feel a $1/r^2$ force. This will describe collisions between charges in a plasma (an ionized gas), and will also describe distant collisions between stars (or any gravitating bodies). In addition it is the physical mechanism underlying bremsstrahlung radiation. I'm following the discussion in Longair, *High Energy Astrophysics, Vol. I*, chapter 2.

3.1.1 the basics

Start with a single encounter, in which particle A (an electron, say) scatters on particle B (a proton, say; with $m_p \gg m_e$, we can assume the proton stays at rest. Let the incoming particle have velocity v and mass m_e, and let in come in at impact parameter b.

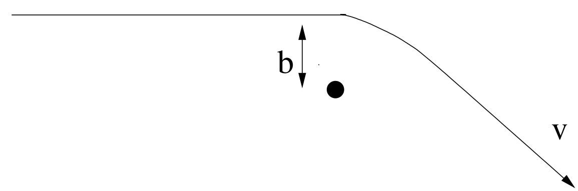

Figure 3.1 A 'soft collision' at impact parameter b.

We can solve this problem exactly, from classical mechanics, and find the deflection angle, θ, and the resultant velocity and momentum changes, $\Delta\mathbf{p} = m\Delta\mathbf{v}$. Here, we will approximate this analysis, following Longair.

The net impulse on the electron will be $\Delta\mathbf{p} = \int \mathbf{F}(t)dt$, integrated over the collision. Now, the force is strong only when the two particles are close. Since they are close for a period of time $\Delta t \simeq 2b/v$, we can approximate $F \simeq e^2/b^2$ and $\Delta p \simeq 2Fb/v$. (Since we know the net deflection is perpendicular to the initial direction of motion, we can also drop the vector notation). This gives us the net energy gain per collision,

$$\Delta E = \frac{(\Delta p)^2}{2m_e} \simeq \frac{2e^4}{m_e b^2 v^2}$$

We want to extend this analysis, to find the net rate of energy exchange with the plasma. But the collision rate of our electron, with particles at impact parameter b is, (collisions/second) $= 2\pi nbv\ db$, we find the net energy exchange rate by integrating over all allowed b:

$$\frac{dE}{dt} = \int_{b_{min}}^{b_{max}} \frac{2e^4}{m_e b^2 v^2} 2\pi bnv db = \frac{4\pi e^4 n}{m_e v} \ln\left(\frac{b_{max}}{b_{min}}\right) \tag{3.4}$$

Now, we want to express this in terms of a cross section:

$$\frac{dE}{dt} = \frac{E}{\tau_{coll}} = nv\sigma_c E \tag{3.5}$$

This defines the Coulomb cross section, σ_c:

$$\sigma_c = 8\pi\left(\frac{e^2}{m_e v^2}\right)^2 \ln\Lambda \tag{3.6}$$

if $\ln\Lambda = \ln\left(b_{max}/b_{min}\right)$ is defined as the Coulomb logarithm.

The Coulomb logarithm depends on the largest and smallest impact parameters that are important (clearly, we cannnot integrate from $b_{min} = 0$ to $b_{max} = \infty$, since the integral in (3.4) would diverge). b_{min} is usually taken to be the distance corresponding to maximum energy transfer,

$$b_{min} \sim e^2/m_e v^2$$

b_{max} is less straightforward. Longair describes Coulomb scattering for an energetic particle hitting an electron bound in an atom, and for this case he likes $b_{max} \simeq v/\nu_o$, if $\nu_o = h^3/(2\pi)^2 m_e e^4$ is the electron's orbital frequency. (He argues that collisions slower than $1/\nu_o$ will violate the free-electron model assumed in the derivation). For unbound electrons, a more common choice is the Debye shielding length (the scale over which an extra charge causes charge separation in a plasma):

$$b_{max} \simeq \lambda_D = \left(k_B T/4\pi n e^2\right)^{1/2}$$

(k_B is the Boltzmann constant, and T is the temperature). Thus, the best choice of $\ln\Lambda$ clearly depends on the exact situation one is considering. Luckily, for our purposes, this is only a logarithmic uncertainty, and will not be critical for most of our calculations. The choices above, with typical astrophysical parameters, give $\ln\Lambda \simeq 10 - 20$, in almost any diffuse-matter setting.

Numerically, for a thermal plasma with $\frac{1}{2} m_e v^2 = k_B T$, the Coulomb cross section becomes,

$$\sigma_c \simeq 7 \times 10^{-13} \frac{\ln\Lambda}{T_4^2} \text{ cm}^2 \qquad (3.7)$$

where $T_4 = T/10^4$K; so that

$$\tau_{coll} \simeq 4 \times 10^4 \frac{T_4^{3/2}}{n \ln\Lambda} \text{ sec} \qquad (3.8)$$

and

$$\lambda \simeq 1 \times 10^{12} \frac{T_4^2}{n \ln\Lambda} \text{ cm} . \qquad (3.9)$$

3.1.2 mnemonics and extensions

A useful short way to remember the Coulomb cross section is as follows. Similarly to the b_{min} estimate above, we can define an effective "radius", a_{eff}, by equating potential and kinetic energies:

$$\frac{e^2}{a_{eff}} = \frac{1}{2} m_e v^2 \qquad (3.10)$$

and then, estimating $\sigma_c = 2\pi a_{eff}^2 \ln\Lambda$. This recovers the form of equaiton (3.6), and resembles the hard-sphere cross section, (3.1), "with a factor of $\ln\Lambda$ tacked on". The factor of 2 is retained in this estimate of σ_c, to match (3.6). In extending this to other examples, as we will do just below, the exact numerical factor that scales $\pi a_{eff}^2 \ln\Lambda$ cannot be recovered by this method of guessing; one would have to do a more formal analysis to get the correct order-unity numerical factor for each cross section.

Two other inverse-square-law cross sections can be immediately written down from this guesstimation. First, extend the Coulomb cross section to a relativistic plasma. If the particle energy is $\gamma m_e c^2$, where $\gamma = (1-\beta^2)^{-1/2}$ and $\beta = v/c$, we get

$$\sigma_c \simeq 2\pi \left(\frac{e^2}{\gamma m_e c^2}\right)^2 \ln\Lambda \qquad (3.11)$$

so that $\sigma_c \propto 1/E^2$ for this limit as well.

The other extension is to the cross section for energy exchange in a gravitating system (such as a star cluster). Here, we estimate a_{eff} from

$$\frac{G m_*^2}{a_{eff}} \simeq \frac{1}{2} m_* v^2$$

(we have assumed all of the stars have the same mass, m_*), and the gravitational cross section is

$$\sigma_{grav} \simeq 2\pi \left(\frac{2 G m_*}{v^2}\right)^2 \ln\Lambda \qquad (3.12)$$

3.2 Anomalous effects

In the preceding section, we worked through a specific, well-formulated, very concrete example of particle collisions. That is, a free charge feels the electric field of an adjacent charge, or of all the nearby charges in the plasma, and changes its momentum and energy accordingly. This is attractive because we can write it down explictly. Unfortunately, it isn't always the whole answer. We know of several astrophysical situations where the predictions (for timescale, for instance) of Spitzer collisions disagree strongly with the observations.

• Plasmas in which the Spitzer collision time (or mean free path) is much longer than the characteristic age (or size) of the system are called *collisionless*. This is maybe an unfortunate terminology, because...

• ... "collisionless" doesn't really mean "no particle ever changes its energy or momentum due to collective effects". Rather, collisionless plasmas often "act collisional" (for instance they support shocks, which require dissipation). For a while this was not understood; now we believe that *plasma turbulence* is involved. That means that a random background of plasma waves (such as those discussed in §2.2.1) exists. A charged paraticle will scatter on the randomly fluctuating $\mathbf{E}$ fields associated with the turbulence; this will have the same effect as physical collisions, but (often) much faster than actual Spitzer/Coulomb collisions. Any such effect associated with plasma turbulence is called *anomalous* (resistivity, conductivity, *etc*. The collective effect of the turbulence is very hard to calculate from first principles; in these notes, we'll just assume that τ_{coll} refers either to Coulomb collisions, or anomalous effects, as needed.

3.3 Apply this: conductivity.

As an example of this, consider the conductivity in an ionized plasma. You recall that, in the simple case, the conductivity σ relates the current density to the local $\mathbf{E}$ field: $\mathbf{j} = \sigma\mathbf{E}$.

> *Notation alert*. Yes, I know, the Greek letter σ is doing double duty here. We attempt to keep them straight by reserving σ_c for a Coulomb cross section and σ for the conductivity.

3.3.1 isotropic conductivity

Let's start simply, ignoring the effects of any $\mathbf{B}$ field. We can then find σ simply, in terms of the mean time between collisions, τ_{coll} (or the collision frequency, $\nu_{coll} = 1/\tau_{coll}$), as follows. Consider a free electron, in a plasma, subjected to an external electric field E. The net force on the particle can be estimated,

$$F_{net} \simeq eE - \frac{\Delta p}{\Delta t} \tag{3.13}$$

where $\Delta p/\Delta t$ is the mean rate of momentum change per collision. But if the charges have a net drift velocity v_D, we can estimate $\Delta p/\Delta t \sim m_e v_D/\tau_{coll}$; then, in a steady state we have $F_{net} \simeq 0$, and the drift velocity must be $v_D = eE\tau_{coll}/m_e$. Next, we can use this in the (static) Ohm's law, to relate the conductivity to the drift velocity:

$$j = n_e e v_D = \sigma E \tag{3.14}$$

where the second equality defines σ. Collecting everything, we end up with

$$\sigma = \frac{n_e e^2}{m_e}\frac{1}{\nu_{coll}} = \frac{\omega_p^2}{4\pi}\tau_{coll} \tag{3.15}$$

(In the last expression, ω_p is the electron plasma frequency, which we've already seen). For our purposes here, the important fact is that $\sigma \propto \tau_{coll}$. Thus, in an ionized, low density plasma, collisions are infrequent, and the conductivity is very high.

3.3.2 anisotropic conductivity

Now, include a $\mathbf{B}$ field; in a general situation in which $\mathbf{E}$ and $\mathbf{B}$ exist. We expect single particle motion to have components along $\mathbf{B}$ (regulated by collisions, just as in the nonmagnetized case); across $\mathbf{B}$ (driven totally by collisions); and also $\mathbf{E} \times \mathbf{B}$ drift, in the third direction.

To start here, let's work out the single particle motion. We have

$$e\left(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{B}\right) - m\mathbf{v}\nu_{coll} = 0 \tag{3.16}$$

Put $\mathbf{B}$ along $\hat{\mathbf{z}}$, and write this in components:

$$\begin{aligned} eE_x + \frac{e}{c}v_y B - m\nu_{coll}v_x &= 0 \\ eE_y - + \frac{e}{c}v_x B - m\nu_{coll}v_y &= 0 \\ eE_z - m\nu_{coll}v_z &= 0 \end{aligned} \tag{3.17}$$

Now do some algebra, and solve for each component of $\mathbf{v}$:

$$\begin{aligned} v_x\left(1 + \frac{\Omega^2}{\nu_{coll}^2}\right) &= \frac{e}{m\nu_{coll}}E_x + \frac{\Omega^2}{\nu_{coll}^2}\frac{cE_y}{B} \\ v_y\left(1 + \frac{\Omega^2}{\nu_{coll}^2}\right) &= \frac{e}{m\nu_{coll}}Ey - \frac{\Omega^2}{\nu_{coll}^2}\frac{cE_x}{B} \\ v_z &= \frac{e}{m\nu_{coll}}E_z \end{aligned} \tag{3.18}$$

Or, this can be written in terms of the vectors $\mathbf{v}_\parallel, \mathbf{v}_\perp$, and $\mathbf{E}_\parallel, \mathbf{E}_\perp$ (relative to $\mathbf{B}$:

$$\begin{aligned} \mathbf{v}_\perp\left(1 + \Omega^2\tau_{coll}^2\right) &= \frac{e\tau_{coll}}{m}\mathbf{E}_\perp - \Omega^2\tau_{coll}^2 c\frac{\mathbf{E}\times\mathbf{B}}{B^2} \\ \mathbf{v}_\parallel &= \frac{e\tau_{coll}}{m}\mathbf{E}_\parallel \end{aligned} \tag{3.19}$$

Thus, single particle motion is a mix of direct flow along $\mathbf{E}_\parallel$, $\mathbf{E} \times \mathbf{B}$ drift, and a new effect, cross-field flow due to the collisions.

These three effects each contribute to a current. Remembering that in general, $\mathbf{j} = ne\mathbf{v}$ (for each charge species), the net current can be written,

$$\mathbf{j} = \sigma_o \mathbf{E}_{||} + \sigma_\perp \mathbf{E}_\perp + \sigma_H \hat{\mathbf{b}} \times \mathbf{E} \tag{3.20}$$

where $\hat{\mathbf{b}}$ is a unit vector along $\mathbf{B}$. We have *three* separate conductivities:

$$\begin{aligned} \sigma_o &= \frac{ne^2}{m\nu_{coll}}\ ; \quad \sigma_\perp = \sigma_o \frac{\nu_{coll}^2}{\nu_{coll}^2 + \Omega^2} \\ \sigma_H &= \sigma_o \frac{\nu_{coll}\Omega}{\nu_{coll}^2 + \Omega^2} \end{aligned} \tag{3.21}$$

where $\Omega = eB/mc$ is the gyrofrequency, as usual. These three terms are called the collisional, Pederson and Hall conductivities. Comparing the two cross-field terms, we see that the Hall current dominates if Ω is large (so that gyromotion is much faster than collisions); or that the Pederson current wins if collisions dominate.

3.4 Apply this: diffusion

Another example is diffusion of one species into another (for instance, think about diffusion of cosmic rays into a thermal, subrelativistic part of the ISM).

3.4.1 isotropic diffusivity

We can pull the same trick as above, to estimate the diffusion rate for the isotropic case. Here, forget about any applied $\mathbf{E}$ field, but put a number of particles in a density gradient, ∇n. Keep the temperature constant, so that the density gradient connects to a pressure gradient, $\nabla p = kT\nabla n$. If the particles have number density n, and undergo collisions just as they did in the previous section, the net force per unit volume is

$$F_{net} \simeq -\nabla p - \frac{mnv_{diff}}{\tau_{coll}} \tag{3.22}$$

Note I've relabeled the drift velocity here, trying to clarify the notation.

> Why does equation (3.22) hold?? This can be proved in two ways. One way is to use macroscopic conservation laws – you know that pressure is a force per area, so a pressure gradient exerts a net force on a unit volume of stuff. We'll do this formally in the next chapter. Alternatively, one can connect the micro to the macro by starting with a conservation law in phase space (called the Boltzmann equation) for all of the particles, multiply by $\mathbf{v}$ twice, and integrate over $d\mathbf{v}$ ... to get to the same point. We won't go through this second approach, but it does make clear the microscopic, statistical effect of a density gradient.

So, in this case the net force goes away for a diffusion-drift velocity, which we express in terms of the *flux* (particles/cm^2-s), as

$$nv_{diff} \simeq -\frac{kT}{m}\tau_{coll}\nabla n \tag{3.23}$$

We have, thus, a diffusion velocity that depends on the local density gradient – also note the sign (particles move towards lower density regions). This result is usually incorporated into the continuity equation:[1]

$$\frac{\partial n}{\partial t} = -\nabla \cdot (n\mathbf{v}) = \nabla \cdot (D\nabla n) \tag{3.24}$$

where

$$D = \frac{kT}{m}\tau_{coll} \tag{3.25}$$

is the *diffusion coefficient*. Note, D has dimensions of cm^2/s, and is often written $D \simeq v_{char}^2\tau_{coll} \sim v_{char}\lambda$, where v_{char} is the "characteristic" speed (for instance thermal speed) of the particle distribution, and $\lambda = v_{char}\tau_{coll}$ is the mean free path of a particle. Thus, simple diffusion can be thought of as a random walk with step length λ.

3.4.2 anisotropic diffusivity

How does the presence of a $\mathbf{B}$ field change these results? About as you'd expect – particles have a much harder time diffusing across a field than along it. Refer back to the discussion around equations (3.16) through (3.18). We proceed similarly here, ignoring $\mathbf{E}$ but including a pressure gradient. Motion along $\mathbf{B}$ isn't changed; the equations of motion across $\mathbf{B}$ become

$$\begin{aligned} mnv_x\nu_{coll} &= -kT\frac{\partial n}{\partial x} + en\frac{v_y}{c}B \\ mnv_y\nu_{coll} &= -kT\frac{\partial n}{\partial y} - en\frac{v_x}{c}B \end{aligned} \tag{3.26}$$

[1] You've probably seen this before, most likely in E&M (remember conservation of charge). If not, trust me for a little bit, we'll derive in in the next chapter.

and after more algebra, we get to[2]

$$
\begin{aligned}
v_y\left(1+\frac{\Omega^2}{\nu_{coll}^2}\right) &\simeq -\frac{D}{n}\frac{\partial n}{\partial y} \\
v_x\left(1+\frac{\Omega^2}{\nu_{coll}^2}\right) &\simeq -\frac{D}{n}\frac{\partial n}{\partial x}
\end{aligned}
\tag{3.27}
$$

(with the parallel motion unchanged). Similarly to the conductivity study, we thus have parallel and perpendicular diffusion coefficients:

$$
D_{\|} = D = \frac{kT}{m}\tau_{coll}\ ; \quad D_{\perp} = \frac{D\nu_{coll}^2}{\nu_{coll}^2+\Omega^2} \tag{3.28}
$$

Thus – as with electrical conductivity – collisions slow down diffusion across $\mathbf{B}$. In the limit $\Omega \gg \nu_{coll}$, the perpendicular diffusion coefficient becomes $D_\perp \simeq kT\nu_{coll}/m\Omega^2 = r_L^2\nu_{coll}$. Thus, cross-field diffusion can be thought of as a random walk with step length r_l (instead of λ).

Key points

• Collisions: mean free path, collision time, *etc.* – if you haven't used these before, make sure you understand them.

• "Spitzer" or "Coulomb" collisions: what they are, how to *estimate* the cross section.

• Anomalous effects: what they are, what we think they are due to.

• Electrical conductivity: how to build isotropic σ from basics; qualitative effects of $\mathbf{B}$.

• Diffusion coefficients: how to build isotropic D from basics; qualitative effects of $\mathbf{B}$.

[2] Doing the full analysis finds a third term, analogous to the $\mathbf{E}\times\mathbf{B}$ term in the anisotropic conductivity. It leads to a "density gradient drift", a.k.a. "diamagnetic drift", $\mathbf{v}_D \propto \nabla p \times \mathbf{B}$. It's usually slow compared to our diffusion, so I'm omitting it from these notes.

4 Basic fluid dynamics

Thus far in the course we've used a microscopic approach, emphasizing the effects of individual charges and single particle trajectories in the behavior of a plasma. Now, we move to a macroscopic approach. We want to describe the collective dynamical behavior of the system in terms of a few macroscopic variables – density, pressure, temperature, velocity, and (later on) currents and fields. This approach will be valid on scales large enough that we can ignore the discrete nature of the fluid (the fact that the gas, or plasma, or fluid, is composed of point-like particles). Thus, we must be working on scales large compared to the interparticle distance. In addition, some of our results will depend on collisions between the particles being important – this allows a shock to form, for instance. In these applications, we have the additional condition that the *effective collision length* must be small compared to the system size. This effective length can be (i) the Coulomb or hard-sphere mean free path; or (ii) the mean free path of a particle to "collisions" with microturbulence in the plasma (small-scale, small amplitude waves involving electric and/or magnetic field fluctuations); or (iii) the gyroradius of a charged particle, if the plasma contains a tangled magnetic field (in which case the particle motion across the field is severely restricted, and the effective "collision length" is somewhere between the gyroradius and the characteristic "tangling length" of the field).

Comments to the Reader.

In this chapter and the next I present the fundamental laws of astrophysical fluid mechanics: conservation of mass, momentum and (nearly) magnetic flux.[1] The fundamental relations are expressed as partial differential equations, and I have chosen (for the sake of having a thorough reference) to present them semi-formally. The down side of this is that they may look slightly intimidating to a student who does not often solve PDE's for fun and relaxation. I therefore offer a summary table – giving the most useful forms of the basic conservation laws, and some other critical results. These are the forms with which said student should be particularly familiar. There are also some simple, and important, applications coming. These important formal results can be found at:

Key (math) points

Mass conservation	eq. (4.2)
Momentum conservation	eq. (4.4)
Induction	eq. (5.11)

Key applications

Hydrostatic equilibrium	eq. (4.7)
Sound waves and sound speed	eq. (4.13)
Bernoulli's fact	eq. (4.18,19)
Flux freezing	eq. (5.14)
Alfven waves	eq. (5.9)

You might also note that the formal equations, while (of course!) necessary, are not the heart of the material at this level. Rather, the focus in class and in the homework will be on physical insight and simple applications of these basic laws.

4.1 Fluids: basics

Hydrodynamics starts with three basic equations, describing mass, energy and momentum conservation in the fluid.[2] We will consider the first two in this chapter. To extend to magnetohydrodynamics (MHD), we need a fourth basic equation, connecting the field to the fluid, and vice versa. That will come in chapter 5.

4.1.1 mass conservation

Consider an arbtrary volume of fluid, V, bounded by a closed surface, A; let the surface have an outward normal, $\hat{\mathbf{n}}$. The mass within this volume is $\int_V \rho dV$, if ρ is the mass density. The net rate of change of this mass is

$$\frac{d}{dt}\int_V \rho dV \quad ;$$

if there are no sources or sinks of matter, this quantity must equal zero. Now, there are two ways this integral can change with time. (i) there can be intrinsic variation of ρ, $\partial\rho/\partial t \neq 0$; or (ii) there can be flow into or

[1] We'll defer energy conservation to next term,after we've developed more tools.

[2] Terminology warning: hydrodynamics normally talks about "fluids"; but "gases" obey the same macroscopic laws, as do "plasmas".

out of the volume, at a rate $\rho\mathbf{v}\cdot\hat{\mathbf{n}}$ per surface area. The sum of (i) and (ii) must balance out to zero:

$$\frac{d}{dt}\int_V \rho dV = \int_V \frac{\partial\rho}{\partial t}dV + \int_A \rho\mathbf{v}\cdot\hat{\mathbf{n}}dA = 0 \quad (4.1)$$

But the surface integral can be written as $\int_A \rho\mathbf{v}\cdot\hat{\mathbf{n}}dA = \int_V \nabla\cdot(\rho\mathbf{v})dV$. Since V is arbitrary, we can set the integrand to zero, and we get the differential form of this basic equation:

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\mathbf{v}) = 0 \quad (4.2)$$

This is, of course, the continuity equation, applied to mass conservation.

4.1.2 momentum conservation

Consider again our surface A, enclosing volume V. The momentum within this surface is $\int_V \rho\mathbf{v}dV$. The net rate of change of this quantity again must reflect intrinsic ($\partial/\partial t \neq 0$) variation and advection (flow across the surface). Thus, we write the net rate of change of momentum as

$$\begin{aligned}\int_V \frac{\partial}{\partial t}(\rho\mathbf{v})dV &+ \int_A (\rho\mathbf{v})\mathbf{v}\cdot\hat{\mathbf{n}}dA \\ &= \int_V \frac{\partial}{\partial t}(\rho\mathbf{v})dV + \int_V \nabla\cdot(\rho\mathbf{v}\mathbf{v})dV\end{aligned} \quad (4.3)$$

In the second expression, we have used Gauss's law for tensors (noting that $\rho\mathbf{v}\mathbf{v}$ is a second-rank tensor).[3]

Now, the net rate of change of momentum in the volume must be equal to the net force exerted on the volume. We consider external forces which act throughout the volume ("body" forces, such as gravity, electomagnetism, buoyancy, radiation pressure if the fluid is optically thin; we let $\mathbf{f}$ be the net force per mass), and also the force exerted on the surface by the fluid outside V. The net force on the volume V is, then,

$$\int_V \rho\mathbf{f}dV - \int_A p\hat{\mathbf{n}}dA = \int_V \rho\mathbf{f}dV - \int_V \nabla p dV$$

where we have again used vector identities in the last step. If we take its differential form, expand the derivatives in the LHS and use (4.2) to simplify, we get

$$\rho\frac{\partial\mathbf{v}}{\partial t} + \rho(\mathbf{v}\cdot\nabla)\mathbf{v} = -\nabla p + \rho\mathbf{f} \quad (4.4)$$

This is our basic force equation (also known sometimes as Euler's equation, or the Navier-Stokes equation).[4]

4.1.3 Lagrangian derivative

Look at the LHS of (4.2) or (4.4): both terms describe the "intrinsic" ways in which the mass, or momentum, in the elemental volume can change. It can be useful to collect them as

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla \quad (4.5)$$

which is called the "Lagrangian derivative". It describes the rate of change of whatever (mass, in 4.2; $\mathbf{v}$, in 4.4; etc) along the *trajectory* of the particle/fluid element. With this, the continuity equation becomes

$$\frac{D\rho}{Dt} + \rho\nabla\cdot\mathbf{v} = 0 \quad (4.6)$$

and the momentum/Euler equation can be written similarly (simplifying to Cartesian geometry);

$$\rho\frac{D\mathbf{v}}{Dt} = -\nabla p + \rho\mathbf{f} \quad (4.7)$$

4.2 Apply: hydrostatic equilibrium

After all that math, a couple of simple examples are in order. The first is a familiar one: consider Euler's equation, for force balance, in a fluid at rest (so that $\mathbf{v} = 0$). The most common application of this is in a gravitational field, $\mathbf{f} = \mathbf{g}$. This gives us just the condition for hydrostatic equilibrium:

$$\nabla p = \rho\mathbf{g} \quad (4.8)$$

[3] *Shriek!* you're probably saying... what the heck does the notation $\mathbf{ab}$ mean? In Cartesian coordinates, it's a 3x3 matrix, where the ijth component is constructed from the ith component of $\mathbf{a}$ and the jth component of $\mathbf{b}$:

$$[\mathbf{ab}]_{ij} = a_i b_j$$

I've added a page on "vector identities" to the class web page which has a bit more detail (in fact more than you'll need in this course).

[4] You should note that one important additional force term has not been included: the force between two adjacent fluid elements, due to the friction of their relative motion. This is the viscous term, and involves second derivatives of $\mathbf{v}$ in the space coordinates. We won't use this term in our examples, but it is often included in terrestrial applications.

4.2.1 planar atmosphere

For instance, if $\mathbf{g}$ is uniform, we can use the ideal gas law,

$$p = \rho \frac{k_B}{m} T \tag{4.9}$$

to write (4.8) as a DE in ρ. If $\mathbf{g} = -g\hat{\mathbf{z}}$, so that gravity is in the z direction, and if the gas is isothermal,[5] we can show this leads to the usual solution for the exponential atmosphere:

$$\rho(z) = \rho_o e^{-z/H} \tag{4.10}$$

where $H = k_B T/gm$. This is familiar as a description of the earth's atmosphere; it also applies to the ISM in the disk of the galaxy. Around our location in the plane, $\mathbf{g}$ is nearly vertical, so this describes the thickness of the ISM disk.

4.2.2 stellar equilibrium

Change the geometry to spherical: think about a star. The HSEq equation (4.8) becomes

$$\frac{dp}{dr} = -\rho \frac{GM(r)}{r^2} \tag{4.11}$$

where $M(r)$, the mass inside radius r, is of course

$$M(r) = 4\pi \int_0^r r^2 \rho(r) dr \tag{4.12}$$

So far so good. However, solving (4.11) is far from simple, because we can't assume the interior of the star is isothermal – so we have to introduce some further physics. That physics gets complicated: we must account for the energy sources within the star (from nuclear fusion), and how that energy is transported out from where it's generated (mostly by radiation, but also by convection). The details of the transport, combined with overall energy balance, determine the temperature structure inside the star ... folding that back into (4.11) eventually gives us the star's density structure.

Doing all that is very complicated, requires numerical solutions, and is far too much to go into in this course. However one simple scaling argument is useful. Look back at (4.11). We can approximate the LHS by

$$\frac{dp}{dr} \sim \frac{\Delta p}{\Delta r} \sim \frac{p_o}{R} \tag{4.13}$$

[5] BIG simplification here!!

if R is the star's radius and p_o is the gas pressure near the star's core. The RHS, evaluated at the star's surface, is M/R. Thus, remembering $\rho = nm$ (if m is the mean mass per particle), (4.11) can be approximated as

$$\frac{nkT}{R} \sim \frac{\rho GM}{R^2} \; ; \quad \frac{kT}{m} \sim \frac{GM}{R} \tag{4.14}$$

By this point, we're interpreting n, ρ, T as "typical" values in the core of the star. Cute result: the second expression is just energy balance. When the star is in hydrostatic equilibrium, its internal energy (per mass) is approximately equal to its potential energy (per mass).

That's a nice result; now let's consider what happens when HSEq is not satisfied.

4.2.3 star formation: gravitational instability

A star is a gravitationally bound system. Thus, the most fundamental idea is that a piece of the ISM can form a star when it is gravitationally unstable – as (Sir James) Jeans first pointed out. You have probably already seen an informal approach to this problem: if the (magnitude of the) gravitational potential energy exceeds the internal energy, the gas cloud (protostar) will collapse. For a cloud of radius R and fixed mass M, one way to express this is for a single particle of mass m:

$$\frac{GMm}{R} \gtrsim k_B T \; : \quad \text{or} \quad \frac{GM^2}{R} \gtrsim M \frac{k_B}{m} T \tag{4.15}$$

(if m is the mean mass per particle). This is clearly an *upper limit* on the size of a gravitationally unstable cloud. Alternatively, if we consider a piece of the ISM, we might want to hold the density fixed: the same criteria now becomes

$$\frac{4\pi}{3} G R^2 \rho \gtrsim \frac{k_B T}{m} \; ; \; R \gtrsim R_J = \left(\frac{3}{4\pi} \frac{k_B T}{G m \rho} \right)^{1/2} \tag{4.16}$$

and

$$M \gtrsim M_J = \left(\frac{k_B T}{mG} \right)^{3/2} \left(\frac{3}{4\pi\rho} \right)^{1/2} \tag{4.17}$$

which is now a *lower limit* for the radius of an unstable region. This latter is usually identified with the original Jeans analysis: the length and mass scales in (4.16 and 4.17) are called the *Jeans' length* and *Jeans' mass*.[6]

[6] Comment from the author...please note where the apostrophe goes!

Free-fall time. What happens if the protostar is gravitationally unstable? How long does it take to collapse? If life is simple, this time is close to the *free-fall time*. Consider a piece of star at radius r, which suddenly loses gravitational support. Through "potential = kinetic" energy, its collapse speed will be $v_{ff}^2 \sim 4\pi G\rho R^2$; so the time it takes to fall a distance R is

$$t_{ff} \sim 1/\sqrt{G\rho} \tag{4.18}$$

You should also note that two important effects are not included in this simple approach: *rotation* and *magnetic fields*. Both will fight against gravitational collapse. We'll return to this in chapter 5.

4.3 Apply: Sound waves

This is an important concept: if you "hit" a fluid, how fast does the information (that the fluid has been hit) travel? Here's a physical approach, similar to the way we derived plasma waves.

Let some perturbation $(\delta\rho, \delta p, \delta T)$ be moving at some c_s. Ahead of the wave the fluid has $v = 0$; behind the wave the fluid has δv, in the same direction as the wave motion. Mass conservation at the wave front, in a frame moving with the wave front, gives

$$\rho c_s = (\rho + \delta\rho)(c_s - \delta v)$$

and to lowest order small, this gives

$$\delta v \simeq c_s \frac{\delta\rho}{\rho} \tag{4.19}$$

Thus, $\delta v > 0$ if $\delta\rho > 0$; the passage of a compression wave leaves behind a fluid moving in the direction of the wave. Now, apply momentum balance: the net force on some control volume, from the pressure difference, equals the rate of change of momentum in that volume. That is,

$$p - (p + \delta p) = \rho c_s \left[(c_s - \delta v) - c_s\right] \tag{4.20}$$

so that, again to first order small, we have

$$\delta p \simeq \rho c_s \delta v \tag{4.21}$$

Combining (4.19) and (4.21), we get a condition on the wave speed (to allow mass and momentum balancs):

$$c_s^2 = \frac{\delta p}{\delta\rho} \tag{4.22}$$

This is a big result: the fundamental signal speed in an unmagnetized fluid. Referring back to the ideal gas law, we also see that $c_s^2 \simeq kT/m$ (the "$\simeq$" describes the uncertainty in how T varies when ρ does).

4.4 Apply: the Bernoulli effect

You may have seen this before. It is basically an expression of energy conservation, for a moving fluid element. But you remember that one can derive energy conservation from momentum conservation.[7] Here, I go through the rather formal derivation for the hydrodynamic equivalent.

Start with Euler's equation, in the form (4.4). But now, note two useful facts. The first is that if the fluid is *barotropic* – that is if $p = p(\rho)$ only (as in an adiabatic gas), we have

$$\frac{1}{\rho}\nabla p = \nabla \int \frac{dp}{\rho} \tag{4.23}$$

(this can be verified using the chain rule; take $p = F(\rho)$, F being some function, and go from there). Thus, this term is a perfect differential. The second useful fact is that

$$\mathbf{v}\cdot\nabla\mathbf{v} = -\mathbf{v}\times\boldsymbol{\omega} + \nabla\left(\frac{1}{2}v^2\right) \tag{4.24}$$

(this is easiest to verify by expanding out in Cartesian coordinates). Thus, this term is also a perfect differential. The first term on the right hand side is written in terms of $\boldsymbol{\omega} = \nabla\times\mathbf{v}$, the local *vorticity* (which is useful in advanced applications). Specify the force to gravity, which can be expressed in terms of a potential: $\mathbf{g} = \nabla\Phi_g$. If we then consider steady flow, we can rewrite (4.4) as

$$\nabla\left[\frac{1}{2}v^2 + \int\frac{dp}{\rho} + \Phi_g\right] = \mathbf{v}\times\boldsymbol{\omega} \tag{4.25}$$

But now: the right hand side of (4.25) is normal to both the local flow field (that is normal to streamlines) and to the local vorticity $\boldsymbol{\omega}$. Thus, we have one form of *Bernoulli's relation*: in inviscid, steady flow, the term in brackets has zero gradient in the direction of the local velocity field. We therefore have one version of Bernoulli's law:

$$\frac{1}{2}v^2 + \int\frac{dp}{\rho} + \Phi_g = \text{constant along streamline} \tag{4.26}$$

[7] Don't believe me? Start with "$F = ma$", let F come from the gradient of some potential, and integrate once; you'll get (kinetic energy) + (potential energy) = constant. Try it!

Further, in an adiabatic gas, $p \propto \rho^\gamma$ if γ is the adiatabic index (the ratio of specific heats). The second term simplifies, so that Bernoulli's relation for an inviscid adiabatic gas is

$$\frac{1}{2}v^2 + \frac{\gamma}{\gamma - 1}\frac{p}{\rho} + \Phi_g = \text{constant along streamline} \tag{4.27}$$

Alternatively, in an incompressible fluid, ρ is constant, and the second term in (4.27) becomes simply p/ρ. Thus, for an incompressible fluid, Bernoulli's relation is

$$\frac{1}{2}v^2 + \frac{p}{\rho} + \Phi_g = \text{constant along streamline} \tag{4.28}$$

4.4.1 example: free expansion

Now, we can apply this to the case of a piece of fluid expanding into a low-pressure environment. (For instance, this might describe a newly created HII region, which has suddenly been heated to $T \simeq 10^4$K, and has an internal pressure $\gg$ that of its surroundings). If we are describing a cloud in the ISM, we can probably ignore gravity. First, we note that (4.27) can be written,

$$\frac{c_s^2}{\gamma - 1} + \frac{1}{2}v^2 = \text{constant} \tag{4.29}$$

Now, consider a spherical gas cloud, with finite density and pressure at its center, which is expanding into vacuum. This is clearly a time-dependent problem; but for times between its initial "violent" expansion, and its eventual final dispersion, we might get away with a nearly steady-state description. (That is, for these intermediate times, the rate at which the spatial dependence of the flow field changes is small compared to the rate at which an individual piece of fluid moves from the center to the outer edge). We can then relate conditions at the center of the cloud to conditions at the edge, by applying (4.29) at these two regions. Now, at the center, $v \simeq 0$ (since this is a spherical expansion, with a center at rest); and the cloud must have some finite central temperature, so that the central sound speed is c_{so}. Thus, the constant in (4.29) is $c_{so}^2/(\gamma - 1)$. Now, as $r \to \infty$, $p \to 0$, by assumption, so that $c_s \to 0$. (There can be conditions in which this last does not follow, if the external environment is very low-density but has a high internal energy per particle. However, such conditions probably do not hold in the ISM.) Thus, the constant from (4.29), evaluated at $r \to \infty$, must be $\frac{1}{2}v_\infty^2$. Finally, we can equate the constant evaluted at the center of the cloud, to the constant evaluated far away from the cloud, and infer

$$v_\infty^2 = \frac{2}{\gamma - 1}c_{so}^2 \tag{4.30}$$

Thus, a cloud having finite pressure, in a near-zero-pressure environment, will expand at its internal sound speed. (This makes sense; in the absence of driving forces, the best the cloud can do is to use its internal energy to drive the expansion; and c_{so} is a measure of this internal energy).

Key (physical) points

- mass conservation: how does the basic idea relate to the ways we use it mathematically?
- momentum conservation: ditto?
- hydrostatic equilibrium and gravitational (in)stability
- sound waves and the sound speed
- the Bernoulli effect, and how it applies to free expansion

5 Basic MHD

In this chapter we carry on with our approach of the previous chapter, but now include what's special to plasmas: their ability to carry currents, and to support and react to magnetic fields.

5.1 The Lorentz force

The effect of the $\mathbf{B}$ field on the force equation, (2.4), is straightforward. We simply add the Lorentz force to the momentum equation:

$$\rho \frac{D\mathbf{v}}{Dt} = -\nabla p + \frac{\mathbf{j}}{c} \times \mathbf{B} + \mathbf{F} \tag{5.1}$$

(where $\mathbf{F}$ is any other external force such as gravity, and I have ignored viscosity here). Now: expand out the Lorentz force as

$$\begin{aligned} \frac{\mathbf{j}}{c} \times \mathbf{B} &= \frac{1}{4\pi}(\nabla \times \mathbf{B}) \times \mathbf{B} \\ &= -\frac{1}{8\pi}\nabla B^2 + \frac{1}{4\pi}(\mathbf{B}\cdot\nabla)\mathbf{B} \end{aligned} \tag{5.2}$$

This is an important breakdown of the Lorentz force; it demonstrates that the field exerts a **magnetic tension** and a **magnetic pressure** on the fluid. The first term in (5.2) represents the gradient of a scalar pressure, $p_B = B^2/8\pi$. It appears in the momentum equation parallel to the fluid pressure....you can think of trying to compress a magnetic field, by pushing at right angles to the field lines, with the field resisting the compression ("fighting back"). The second term is non-zero only if the field varies parallel to itself. A simple illustration is a curved field line. The curvature means there is a current flowing along the field line; the $\mathbf{j}\times\mathbf{B}$ force points inwards (relative to the curvature). Thus, curved field lines "want to straighten out"...Some authors combine both effects by describing magnetic field lines as "elastic bands within the fluid", which resist being stretched: either pushed together, or pulled transverse to their length.

5.2 Apply: plasma confinement

I don't know of any stars that are held together by magnetic fields[1] (think: can you come up with a spherically symmetric magnetic field that can confine a plasma? I bet not..) However, magnetic confinement is quite possible in other geometries. In fact, plasma confinement is the fundamental problem for laboratory plasmas, and may well be relevant to some astrophysical applications as well.

The issue is, can the plasma pressure can just balance the Lorentz forces from the fields? Most commonly, flows are ignored, as are resistivity and gravity. The general condition for equilibrium is, then,

$$\frac{\mathbf{j}}{c} \times \mathbf{B} = \nabla p \tag{5.3}$$

This is subject, of course, to the constraints

$$\nabla\cdot\mathbf{B} = 0\,; \qquad \nabla\cdot\mathbf{j} = 0\,; \qquad \mathbf{j} = \frac{c}{4\pi}\nabla\times\mathbf{B} \tag{5.4}$$

(The second relation holds in steady state, right?). A system which satisfies (5.3) also obeys

$$\mathbf{j}\cdot\nabla p = 0\,; \quad \mathbf{B}\cdot\nabla p = 0 \tag{5.5}$$

That is, constant-pressure surfaces are also "magnetic surfaces" and "current surfaces": $\mathbf{B}$ and $\mathbf{j}$ lines lie in constant-p surfaces.

Now... (5.3), with its auxilliaries (5.4) and (5.5), is "all" that is needed for laboratory confinement. We just have to solve it (and then test for stability). In these notes I confine myself to infinitely long plasmas in cylindrical geometry, which involve the simplest math, and may well be relevant to astrophysical jets. The basic equation, (5.3), becomes

$$\frac{dp}{dr} + \frac{d}{dr}\left(\frac{B_\phi^2 + B_z^2}{8\pi}\right) + \frac{B_\phi^2}{4\pi r} = 0 \tag{5.6}$$

Perhaps the most interesting application of this, for us, is the possibility that a current-carrying plasma can confine itself.

Example: linear pinch or z pinch Consider a purely azimuthal field: $\mathbf{B} = (0, B_\phi, 0)$, so that magnetic tension confines the plasma. The plasma can contribute to its own confinement by carrying just the right net current. The basic relation is

$$\frac{d}{dr}\left(p + \frac{B_\phi^2}{8\pi}\right) = -\frac{B_\phi^2}{4\pi r} \tag{5.7}$$

One possible equilibrium (illustrated in the figure) is $B_\phi(r) = B_o r/(1 + r^2/a^2)$ (exercise for the student: what are the corresponding pressure and current density profiles?) This type of pinch can be self-confining. Note that the current within radius r is

[1] although there does exist strong evidence that some molecular clouds, which are the sites of stellar birth in our galaxy, are held *up* by magnetic pressure ... that's for a later discussion.

$I(r) = \int_0^r 2\pi j_z r dr$, and from Maxwell the B field is $B_\phi(r) = 2I(r)/rc$. Using these and the pressure balance condition (5.7),

$$\int_0^a p r dr = \frac{1}{4\pi c^2} I_a^2 \tag{5.8}$$

where I_a is the current in the entire pinch (out to radius a), and we've assumed $p(a) = 0$. (To the student: can you derive this?). Thus, the plasma can self-confine if it carries the right current. This type of pinch is attractive, in that particles don't escape out the ends, and the current is carried by the plasma itself. However, this configuration turns out to be seriously unstable, thus is also of little practical interest in the lab.[2]

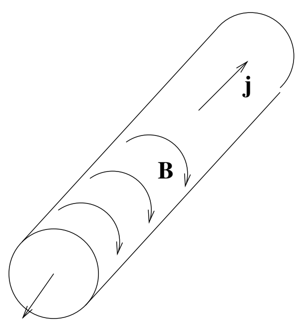

Figure 5.1a The geometry of a linear pinch: the field is azimuthal and the current is axial.

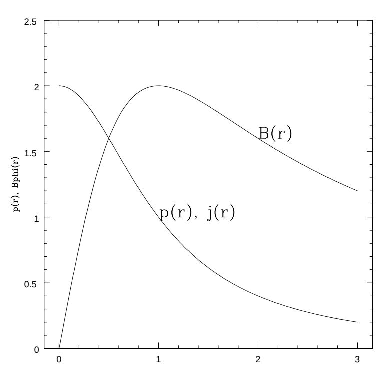

Figure 5.1b Qualitative solutions for magnetic field, pressure and current within a linear pinch.

Example: theta pinch We can switch the geometry above, and consider a "plasma solenoid". That is, run a current azimuthally around a cylinder of plasma – you will of course generate a B field parallel to the axis of the cylinder (as in Figure 5.2). This is not particularly interesting for astrophysics,[3] but it's common in laboratory plasmas (which live in tin cans). I'm including it here because I want to point to the figure later, when I discuss flux freezing.

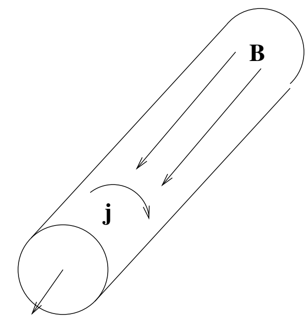

Figure 5.2 The geometry of a theta pinch: an azimuthal current supports an axial magnetic field.

5.3 Apply: Alfven waves

We saw above that any small disturbance in a non-magnetized gas will create sound waves. In a magnetized gas, one of the analogous waves (and probably the most important in astrophysical applications) is an Alfven wave. These waves also carry information, and (if the plasma is cold) are the signal-carrying waves.

What is an Alfven wave? These are waves in which the magnetic field dominates. It exerts the restoring force; fluctuations in the plasma density and pressure are either exactly zero, or unimportant. More specifically, an Alfven wave is a transverse wave, which is not compressive, and which propagates (in the simplest case) along the magnetic field. Thus, they can be thought of as propagating wiggles in the field lines, as in the figure.

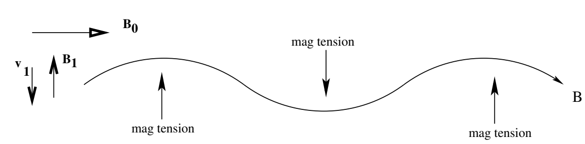

Figure 5.3 Schematic of simple Alfven waves; the perturbed B_1 and v_1 terms are perpendicular to the background field B_o. Following Cravens Figure 4.16.

Deriving the full details of the wave is long, but we can guesstimate the likely wave speed. Recall that waves in

[2]It might, however, be a useful model for radio jets – giving the plasma an axial velocity doesn't change the confinement physics. One must, however, think about how and where the circuit closes: how does the current return to the "origin" (the compact object that creates the radio jet)?

[3]Why? Can this geometry be self-confining, or does it need some external pressure to hold it together?

an elastic wire propagate due to the tension; as the field lines in a plasma exert a tension $B_o^2/4\pi$, one might expect a wave speed

$$v_A = \frac{B_o}{(4\pi\rho_o)^{1/2}} \tag{5.9}$$

This is the *Alfven speed*, and it is, indeed, a useful scaling speed for waves in a magnetized plasma. We can also note directly that $\nabla \cdot \mathbf{B} = 0 \Rightarrow \mathbf{k} \cdot \mathbf{B}_1 = 0$; so that the magnetic field perturbation must be normal to the wavevector.

5.4 The induction equation

Now we turn to the magnetic field in the fluid (which must be ionized, and thus a plasma, in order to interact with the field, right??)

Consider an arbitrary surface, S, within a fluid, bounded by some curve C. The magnetic flux within this surface is $\Phi_B = \int_S \mathbf{B} \cdot \hat{\mathbf{n}} dA$. We want to find an expression for $d\Phi_B/dt$. To get this, we start with Maxwell's equations; in particular the $\nabla\times\mathbf{B}$ and $\nabla\times\mathbf{E}$ ones. Also, we need Ohm's law for a moving fluid:

$$\mathbf{j} = \sigma\left[\mathbf{E} + \frac{1}{c}\mathbf{v} \times \mathbf{B}\right] \tag{5.10}$$

where σ is the conductivity of the fluid or plasma. Now, if we take the curl of (5.10), and also note that

$$\nabla \times (\nabla \times \mathbf{B}) = -\nabla^2\mathbf{B} = \frac{4\pi}{c}\nabla \times \mathbf{j}$$

(since $\nabla \cdot \mathbf{B} = 0$), we find

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) + \eta\nabla^2\mathbf{B} \tag{5.11}$$

where we've defined the *magnetic diffusivity*, $\eta = c^2/4\pi\sigma$. This describes the behavior of the magnetic field in a moving fluid with a specified conductivity. The first term describes induction due to the motion of the fluid, while the second (noting the second derivative) acts as a "diffusion" term, allowing field lines to "leak out"of high-field areas, for instance.

What happened to displacement current? Those of you who are fans of E&M will have noticed that there is no $\partial\mathbf{E}/\partial t$ term in the derivation of (5.11). This is a standard approximation in MHD; the reasoning goes as follows.

(1) Our fluids are very good conductors (why's that? Remember how ineffective Coulomb collisions are at dissipating currents), so we don't expect any free-charge $\mathbf{E}$ to stay around.

(2) Therefore the only $\mathbf{E}$ fields we expect are induced ones, which are $O(vB/c)$; thus $\partial\mathbf{E}/\partial t \sim O[(v/l)(vB/c)] \rightarrow O[(cB/l)(v^2/c^2)]$. But $v \ll c$ for subrelativistic flows (almost always our limit here), so the displacement current is of order (small)^2, and we can ignore it.

5.4.1 Ideal limit: flux freezing

This is an important application; we'll use it a lot.

(a) Derivation. From the definition of Φ_B, we have

$$\frac{d\Phi_B}{dt} = \int_S \frac{\partial B}{\partial t} \cdot \hat{\mathbf{n}} dA + \oint_C \mathbf{B} \cdot (\mathbf{v} \times d\mathbf{l}) \tag{5.12}$$

where the second term, a line integral around the boundary of the surface, accounts for changes in the enclosed flux due to the motion of the surface. This line integral can be made a surface integral, and we find

$$\frac{d\Phi_B}{dt} = \int_S \left[\frac{\partial B}{\partial t} - \nabla \times (\mathbf{v} \times \mathbf{B})\right] \cdot \hat{\mathbf{n}} dA \tag{5.13}$$

From this, and using (5.11), we find our desired result:

$$\frac{d\Phi_B}{dt} = \int_S \eta\nabla^2\mathbf{B} \cdot \hat{\mathbf{n}} dA \tag{5.14}$$

Thus, the rate of change of the magnetic flux depends on the inverse of the conductivity. In particular, astrophysical fluids are often highly conductive, so that $\sigma \rightarrow \infty$ and $\eta \rightarrow 0$. In that limit, we have $\Phi_B \simeq$ constant: the magnetic flux through some loop which is "tied to the plasma" is a constant.

(b) Do magnetic field fines exist? The concept of a magnetic line of force is an abstraction. In general no identity can be attached to these lines (they cannot be labelled in a varying field), nor can we speak of "motion" of field lines. In a perfect conductor, however, the concept of field lines becomes meaningful, due to flux freezing – and turns out to be a very useful way to envision what's going on.

Consider a material line in the fluid (say a chain of labelled droplets, or particles painted pink), defined by intersecting material surfaaces. Choose these surfaces everywhere tangential to $\mathbf{B}$ at $t = 0$. The flux through both surfaces is therefore zero to start, and their intersection defines a field line at that point. Flux freezing guarantees that these surfaces continue to satisfy $\Phi_B = 0$ at any later time. Thus, their intersection continues to define a field line, in fact the same field line – it has become identifiable; labelling the material (painting it pink) has labelled the field line, and the local fluid velocity $\mathbf{v}(\mathbf{x}, t)$ is also the velocity of that section of the field line. *The field line is attached to – "frozen into" – the fluid.*

(c) Flux freezing in practice. If we think of field lines as real entities, that move with the plasma, we can easily predict the effects of Lens' law on B fields in a moving plasma. Note, the easiest way to understand the physics, is to evaluate Φ_B over some imaginary surface *that is parallel to the field lines*.

For one example, think about a plasma cylinder with the B field along the axis, as in Figure 5.2. This might, for instance, be an astrophysical jet (never mind, right now, how the B field is maintained). The useful surface is just a cross-section through the cylinder. Now let the cylinder radius, R, increase – maybe the jet expands as it its source. Flux freezing means the product $B\pi R^2$ is constant; thus $B \propto 1/R^2$.

For another example, think about a plasma cylinder (or astrophysical jet) with azimuthal B – as in Figure 5.1a. The useful surface here might be a square, oriented with one edge along the jet axis, and the other edge along the outer surface of the jet. Now let the cylinder expand, R increase, but without any compression or stretching parallel to the axis. Exercise for the student: how does B vary with R now?

We'll see quite a few other examples as we go along – star formation, solar wind, the earth's magnetotail, neutron star formation in a supernova, accretion flows onto a black hole – all lean heavily on flux freezing.

5.4.2 resistive limit: flux annihilation

In a fluid with finite conductivity, flux freezing no longer holds. We can explore this by going to the other limiting case, when σ is small so that η becomes large. This is *diffusive limit*. If we simply ignore the advection term, equation (5.11) becomes

$$\frac{\partial \mathbf{B}}{\partial t} = \eta \nabla^2 \mathbf{B} \tag{5.15}$$

This describes the effect of Ohmic dissipation on the magnetic field; note that it is a standard diffusion equation.

> **Do magnetic field lines diffuse?** We know how solutions to (5.15) behave: an initial field will decay on a timescale $\sim L^2/\eta$. Some authors discuss this in terms of field line "diffusion" or "slippage" out of the fluid. Remember that the density of field lines is related to the strength of the field; so a lower density of field lines, with time, should correspond to field lines "diffusing" out of the field. In particular, when η is finite, field lines are no longer tied to parcels of the plasma; some authors talk of field lines "moving through" the plasma in dissipative regions.

5.5 Protostellar collapse, revisited

OK, now let's return to our collapsing molecular cloud (MC). Assume that the MC somehow fragments into star-sized pieces. Next, consider the collapse of one such piece. The simplest physics will be when the initial cloud/protostar is nonrotating and unmagnetized. Let's pretend for now that this is the case.

A self-gravitating isothermal sphere is a useful model for this idealized collapse of the protostar. Because this is a simple HSEq solution, we might expect the collapsing cloud to want to find such a structure, as long as it can cool (and stay isothermal). The density structure of an isothermal sphere must obey

$$\frac{k_B T}{m}\frac{d\rho}{dr} = -\rho \nabla \Phi_g \tag{5.16}$$

where $M(r) = \int_0^r 4\pi r^2 \rho(r) dr$ and Φ_g is the gravitational potential (related to the gravitational field by $\mathbf{g} = -\nabla\Phi_g$). We must solve this in conjunction with Poisson's equation for gravity:

$$\nabla^2 \Phi_g = 4\pi G \rho \tag{5.17}$$

(Look familiar? Think about the analogous equation for the electric potential). When this is done, it turns out that one simple solution for the density is

$$\rho(r) \propto \frac{1}{r^2}$$

everywhere. Now, this clearly has two problems: $\rho(r) \to \infty$ as $r \to 0$, and $M(r) \to \infty$ as $r \to \infty$. Thus, real and finite clouds cannot satisfy this everywhere. This is not the only possible solution, however. A more satisfying physical solution has a core: the density is nearly constant (at ρ_o, say) for $r < r_o$, where $r_o^2 \propto T/\rho_o$. In addition, a satisfying physical solution must be truncated at large radii: $\int \rho(r) r^2 dr$ must converge.

Now, clouds collapsing under their own self-gravity can be modelled numerically. Such solutions of the collapse do find that the collapsing cloud moves through a series of nearly isothermal solutions – modified by a central density plateau, and by an outer edge (naturally; any simulation must be finite). Such a cloud will collapse "from the inside out" – since the higher central density will result in a shorter free-fall time (*cf.* 4.13). We would expect, then, that the core of the cloud would at some point become opaque to its own radiation, so that it can no longer cool; further collapse will heat the core. When the temperature reaches $\sim 10^7$K, nuclear burning will start and a star is born.

In the real world, however, we cannot neglect either the angular momentum or the magnetic field of the protostar.

Angular momentum If the collapsing cloud conserves its angular momentum, its angular velocity must increase as

$$\Omega(r) \propto \frac{1}{r^2}$$

Thus, a cloud which starts with only a slow rotation (for instance the differential galactic rotation,

$$\Delta\Omega_{cloud} \simeq \frac{d\Omega}{dR}\Delta R \sim \frac{\Omega(R)}{R} r_{cloud}$$

if R is the galactic radial coordinate and r_{cloud} is the cloud radius), will quickly speed up as it collapses. Without any loss of angular momentum, this spin-up will quickly provide centrifugal force which can balance the self-gravity: we would expect a large, rotating disk rather than a star. Clearly the initial angular momentum must be lost somehow in the collapse process. (Also, we note that the current angular momentum of the sun, for instance, $\ll \Delta\Omega_{cloud}(r_{cloud}/R_\odot)^2$; in agreement with this.)

Magnetic fields and flux freezing A simple picture also predicts that the contracting cloud will conserve magnetic flux. Referring back to (5.14), we recall that the magnetic flux, Φ_B, is nearly constant. This meanss the mean magnetic field in the cloud will increase, as

$$B \propto \frac{1}{r^2}$$

as the cloud colapses. As with angular momentum, the enhanced field will stop the collapse long before a star is reached; we can verify, again, that for the sun $B_\odot \ll B_{ISM}(r_{cloud}/R_\odot)^2$.

How are these problems resolved? We do not know the answer here, in detail. It is likely that a couple of processes are important. The first is the possibility of the conductivity being smaller than the discussion above suggests – which will happen if the cloud is mostly neutral. This will shorten t_{coll} and reduce σ – which will allow the flux inside the cloud to decrease, and the collapse to continue. One can picture the charged particles in the gas, to which the field is tied, "slipping past" the neutral gas as it collapses. This is called *ambipolar diffusion*. In order to estimate the time for the flux to change, t_{flux}, one must know the ion-neutral cross section; specific calculations suggest $t_{flux} \sim 10 - 100 \times t_{ff}$ for MC conditions.

This will reduce the magnetic flux within the cloud, and allow slow collapse. However, it does nothing for the angular momentum problem. The resolution of this is probably brought about by the torque exerted on the collapsing cloud by the magnetic field which threads the cloud. The magnetic field lines are very likely to connect to the ISM outside of the cloud (rather than to be contained wholly within the cloud); flux freezing in the ISM will tend to tie the "ends" of the field lines down, and they will thus exert a torque on the rotating, collapsing cloud. This will tend to slow the cloud down, and to transfer its angular momentum to the surroundings.

Key points

• Magnetic pressure and tension (when the field coexists with a plasma!)

• Self-confinement of a current-carrying plasma

• Flux freezing (when resistivity isn't important) and how it's applied.

• Magnetic "diffusion" or dissipation; when resistivity is important

• How do B fields affect protostellar collapse?

6 One-dimensional flows

In this chapter we'll continue exploring simple (well, fairly simple) examples of steady fluid/plasma flows.

6.1 The sound speed is important

In chapter 4 we introduced the sound speed: $c_s^2 = \partial p/\partial \rho$. There are important differences between subsonic and supersonic flows. Subsonic flows can be thought of as quasi-hydrostatic. That is, the flow field is strongly influenced by pressure gradients which are determined by conditions a long distance away (such as at boundaries). Supersonic flows, however, are quasi-ballistic. Pressure gradients have only a limited range of influence, and conditions far away have little or no effect on a solution locally.

The reason the sound speed is critical to the dynamics of a fluid or plasma flow, is that it is the speed at which information can propagate. We can illustrate this with 1D and 3D cartoons.

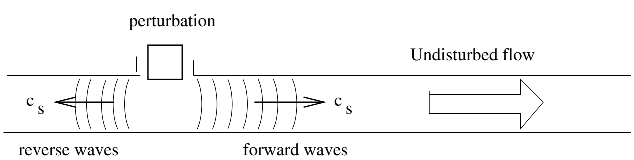

Figure 6.1 Physical illustration of simple waves. The information that the flow has been "whacked" at some point, propagates by simple sound waves, moving at speed c_s relative to the fluid in the pipe. Following Thompson figure 8.6.

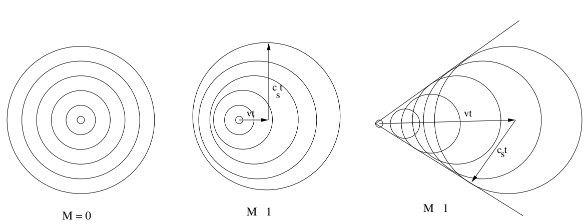

Figure 6.2 Mach's construction for the propagation of a disturbance. Consider a point source of sound (Thompson suggests a bumblebee) in a moving medium. If the source and flow are stationary, the sound propagates spherically from the source. If the source/flow are moving subsonically. the motion only distorts the spherical wavefronts. If, however, the motion is supersonic, all disturbances are confined to a *Mach cone*; an observer located outside of this cone does not receive any information about the bee.

We can also explore this by checking the magnitude of the terms in the (steady state) force equation:

$$\begin{array}{ccccc} \rho\,(\mathbf{v}\cdot\nabla)\,\mathbf{v} & = & -\,\nabla p & + & \rho\mathbf{f} \\ (\rho v^2/L) & & (p/L) & & (\rho f) \\ (v^2/c_s^2) & & (1) & & (U_f/c_s^2) \end{array} \tag{6.1}$$

In the first line I've written the basic equation, in terms of some body force $\mathbf{f}$; in the second line I've estimated the magnitude of each term, for some scale length L; in the third line I've compared the three terms, using $p \sim c_s^2\rho$ and defining a "potential energy" or "work" associated with $\mathbf{f}$, $U_f \sim fL$. Thus: we see that the pressure gradient dominates the inertial ($\mathbf{v}\cdot\nabla\mathbf{v}$) terms for subsonic flow, and vice versa for supersonic flow.

While subsonic flows can easily be smooth and continuous (with internal structure driven by a pressure gradient), it turns out that supersonic flows can (and usually do) contain discontinuous jumps in the flow properties (shocks).

6.2 Outflow: 1D channel flow

In general, a flow can't adjust smoothly from subsonic to supersonic; one or more shocks are generated by the transition. However, there are some special cases in which a smooth transition is possible.

Let's start with flow in a channel; and let the channel have cross section A. If A varies only slowly along the flow, then we can treat this is a 1D problem. If the flow is steady, mass conservation requires $\rho v A =$ constant. (Why?) Differentiating this, we find

$$\frac{1}{\rho}\frac{d\rho}{dx} + \frac{1}{v}\frac{dv}{dx} + \frac{1}{A}\frac{dA}{dx} = 0 \tag{6.2}$$

For the same flow, momentum conservation gives

$$\rho v\frac{dv}{dx} + \frac{dp}{dx} = 0. \tag{6.3}$$

(Why? What terms have we retained or ignored from 4.4?) We need a third equation relating p and ρ: we use $c_s^2 = \partial p/\partial\rho$. Combining these results gives the basic equation for channel flow:

$$\left(\frac{v^2}{c_s^2} - 1\right)\frac{1}{v}\frac{dv}{dx} = \frac{1}{A}\frac{dA}{dx} \tag{6.4}$$

This is an interesting result. Consider the differences between subsonic ($v^2 < c_s^2$) and supersonic ($v^2 > c_s^2$) flow.

- Subsonic: a converging channel ($dA/dx < 0$) accelerates the flow (leads to $dv/dx > 0$), and a diverging channel ($dA/dx > 0$ decelerates it ($\Rightarrow dv/dx < 0$).

- Supersonic: a converging channel ($dA/dx < 0$) *decelerates* the flow (leads to $dv/dx < 0$), and a diverging channel ($dA/dx > 0$ *accelerates* it ($\Rightarrow dv/dx > 0$).

Consider, then, a flow which starts subsonic in a converging channel. It will accelerate as the channel narrows. If things are set up *just right*, the flow will reach $v = c_s$ just at the narrowest point of the channel. If the channel broadens again, the flow can accelerate smoothly to supersonic speeds. Referring to (6.4), we see that "just right" means the flow must reach $v = c_s$ exactly at the narrowest point of the channel. If this is not the case, the flow can do one of several things. It can (i) change from acceleration to deceleration (or vice versa); (ii) it may not be able to remain steady; or (iii) it may set up internal shocks to enable it to adjust to the local conditions in the channel.

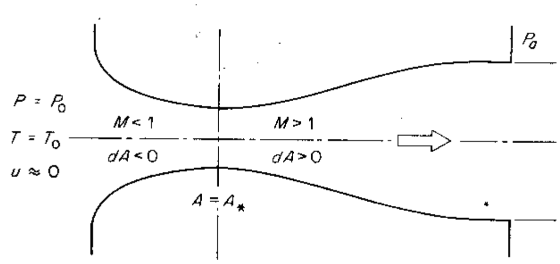

Figure 6.3 Transonic flow in a convergent-divergent nozzle. If the throat occurs at just the right place, relative to the flow, then a smooth transition from subsonic to supersonic is possible. If, however, the flow is not exactly at $v = c_s$ when it reaches the throat, it cannot remain steady: shocks and/or time-unsteady flow happen. From Thompson figure 6.3.

6.3 Outflow: stellar winds

Spherical stellar winds – taking the solar wind as a well-studied example – are an interesting extension of channel flow. We will break the problem into two parts: first, demonstrate that the extended atmosphere of the sun can't be static; and second, the nature of the solar wind outflow. We'll return to the interaction of the flow with a cooler, finite-density ISM at a later point.

6.3.1 Why must there be a solar wind?

We know, from simple observations, that the sun has a hot atmosphere (the chromosphere, and the more extended corona). Are static solutions possible for the solar atmosphere? These would be solutions of

$$\frac{dp}{dr} = -\rho \frac{GM_\odot}{r^2} \tag{6.5}$$

If the temperature structure of the atmosphere $T(r)$, is known, then (6.4) can be integrated easily. We can consider two simple cases:

• First, think about an isothermal atmosphere. In this case, solutions of (6.5) predict a *finite* pressure at infinity:

$$p_\infty = p_\odot \exp\left(-\frac{GM_\odot m_p}{k_B T_\odot R_\odot}\right) \tag{6.6}$$

where $T_\odot \simeq 1.5 \times 10^6$K and $p_\odot \simeq 0.3$N/m^2 are the temperature and pressure at the base of the corona. Evaluating this limit, one finds that $p_\infty \gg p_{ISM} \simeq 10^{-8}$N/m^2. Thus: a static atmosphere would have a pressure at infinity that greatly exceeds the surrounding pressure of the ISM – and so it can't be static. The solution we've just derived would have a tendency to expand outward.

• Can we devise another $T(r)$ profile to alleviate this? Not easily ... we might expect if the temperature falls rapidly enough outwards, one might come up with an acceptible p_∞. However, the temperature gradient can't be steeper than that allowed by thermal conduction (heat flow from the hot solar surface). This turns out to be $T(r) \propto 1/r^{2/7}$; which still has an overly large p_∞.

Thus, we need to look at a *dynamic* solution of the momentum equation – one which gives an outflow or "wind".

6.3.2 The basic wind solution

Next, how does the wind behave (ignoring magnetic fields for the time being)? The basic solution is due to Parker. Consider a steady, spherical outflow. Mass conservation in this case is $\rho v r^2$ =constant; or,

$$\frac{1}{\rho}\frac{d\rho}{dr} + \frac{1}{v}\frac{dv}{dr} + \frac{2}{r} = 0 \tag{6.7}$$

while the momentum equation becomes in this case (noting that gravity from the central star is important),

$$\rho v \frac{dv}{dr} + \frac{dp}{dr} = -\rho \frac{GM}{r^2} \tag{6.8}$$

Writing $dp/dr = c_s^2 d\rho/dr$, these two equations combine to give the basic wind equation,

$$\left(v - \frac{c_s^2}{v}\right)\frac{dv}{dr} = \frac{2c_s^2}{r} - \frac{GM}{r^2} \tag{6.9}$$

This does not have analytic solutions over the whole range of r. However, we can learn quite a bit about the nature of the solutions simply by inspection of (6.9), as follows.

• The left hand side contains a zero, at $v^2 = c_s^2$. If we want to consider well-behaved flows, that is to say those in which the derivative dv/dr does not blow up, then the right hand side of (6.9) must go to zero at the same point. This defines the condition that must be met at the sonic point:

$$v^2 = c_s^2 \qquad \text{at} \qquad r = r_s = \frac{GM}{2c_s^2} \tag{6.10}$$

Whether or not a particular flow satisfies this condition depends on the starting conditions, such as with what velocity and temperature it left the stellar surface, and also what the boundary conditions at large distances are. If it does not start in such a way to satisfy this condition, it either stays subsonic (corresponding to finite pressure at infinity), or cannot establish a steady flow.

• The solution beyond the sonic point depends on the temperature structure of the wind. The only solutions with $dv/dr > 0$ for $r > r_s$ are those for which $c_s^2(r)$ drops off more slowly than $1/r$; it is only these for which the right-hand side stays positive. In the case of an isothermal wind, with c_s^2 = constant, (6.9) can be solved in the limit $r \gg r_s$:

$$v^2(r) \simeq 4c_s^2 \ln r + \text{constant} \tag{6.11}$$

Thus, the wind will be supersonic, by a factor of a few, as $r \to \infty$. The question of how the solar wind manages to stay nearly isothermal is not solved; it is probably due to energy transport by some sort of waves (MHD or plasma waves, for instance) which are generated in the photosphere and damped somewhere far out in the wind.

• Inside the sonic point, the gravity term will dominate the right hand side of (6.9). Thus, solutions with $dv/dr > 0$, and $v^2 < c_s^2$, will obey

$$\frac{c_s^2}{v}\frac{dv}{dr} \simeq -\frac{GM}{r^2}$$

This equation looks as if gravity is driving the wind out! This unlikely-looking result comes from the fact that the flow is nearly subsonic in this region; therefore, the dp/dr term in (6.8) – which actually drives the wind out – is nearly equal to the gravity term.

6.3.3 What about MHD effects?

How can we justify ignoring the magnetic field in this analysis? To explore this, go back to the basic momentum equation (5.1), write it for steady flow, and estimate the magnitude of each term:

$$\begin{array}{cccc} \rho\mathbf{v}\cdot\nabla\mathbf{v} = -\nabla p & + \dfrac{\mathbf{j}}{c}\times\mathbf{B} & - & \rho\dfrac{GM_\odot}{r^2} \\ (\rho v^2/L) \quad (p/L) & (p_B/L) & & (\rho GM/r^2) \\ (v^2/v_A^2) \quad (c_s^2/v_A^2) & (1) & & (U_{grav}/v_A^2) \end{array} \tag{6.12}$$

In the first line of (6.12) I've written the real equation. In the second line I've estimated the magnitude of each term (for some scale length L, and gravitational potential energy U_{grav}), and in the third line I've compared the relative magnitudes of each term to the Lorentz force term. Thus: we can ignore MHD effects when the Alfven speed is low – when the B field is low, or when the density is high.

Looking at numbers we know for the solar wind, this limit holds (i) very close to the sun, where the density is high; and (ii) past about 10 solar radii, where $v \sim v_A$ and $U_{grav} \ll v_A^2$. Between these limits, in the range $\sim 2 - 10R_\odot$, we find that the $\mathbf{j}\times\mathbf{B}$ term dominates. It is in this region that the B field controls the geometry of the flow. This is the region in which the field is changing between "emerging flux ropes", fully connected to the sun's surface, and open field lines, connected to the solar wind; the plasma is constrained to flow along open field lines, and thus the configuration of these open field lines determines the "area function" that channels the plasma flow. Finally, past $\sim 10R_\odot$, the flow is strongly superalfvenic as well as supersonic; the plasma is capable of "pushing the field around". We can use flux freezing here, and think of the field lines as being stretched out by the outflowing, supersonic plasma.

The outer, intertia-dominated regions of the solar wind include close to earth, where space probes provide us with good, detailed information about the nearby structure of the wind. We know that the equatorial plane of the solar wind contains four *sectors*, two with outwards-pointing magnetic field, and two with inwards-pointing field. We also know that the global field direction reverses direction above and below the equatorial plane (as you would expect when the basically dipolar magnetic field of the sun is stretched out by the solar wind). Thus, the equatorial plane must contain a *current sheet* (think: which way must the current flow?). It turns out that this current sheet is also the explanation of the sectors.

6.3.4 What about shocks?

Just a quick note here. The solar wind flow is highly supersonic, and superalfvenic, by the time it reaches earth. But here and there it must slow down: when it is forced to go around a planet, and when it runs into the local ISM. We expect a shock transition at each of these sites. We find a *bow shock* where the solar wind encounters the earth's magnetosphere – and similar structures around the other planets. We also find an *outer shock* – called the *heliopause* – where the solar wind's pressure has dropped to approximately that of the surrounding ISM. We'll talk more about these, later on in the course.

6.4 Inflow: Spherical Accretion

Accretion flows are common in many areas of astrophysics. Galactic binary systems involving accretion are common. Cataclysmic variables, novae, and Type I supernovae involve accretion (usually non-steady) onto white dwarfs. X-ray binaries (which come in several different flavors) involve accretion (usually thought to be steady) onto neutron stars or black holes. In addition, star formation must proceed through accretion (as the outer regions of the protostar accrete onto the inner regions). Millisecond pulsars ("recycled pulsars") are often found in binary systems, and are believed to have been spun-up by accretion of mass and angular momentum from a compansion. Finally, accretion onto a massive nuclear black hole is the most likely explanation for active galactic nuclei (quasars, radio galaxies, Seyfert galaxies, and all related phenomena).

Most astrophysical accretion flows involve angular momentum, and thus we need to worry about accretion disks. We'll do that later; for now, let's consider simple spherical accretion.

6.4.1 Basic ideas

To start, we first address simple considerations of energetics and temperature. Following this, we will look at models of spherical and then disk accretion.

Energetics. The basic consideration is that gravitational potential energy is released, at a rate $\dot{E}_g$; and can be turned into radiation with some (as yet unspecified) efficiency ε:

$$\dot{E}_g \sim \frac{GM\dot{M}}{r} \qquad \Rightarrow \qquad L = \varepsilon \frac{GM\dot{M}}{r} \tag{6.13}$$

This simple formulation is the foundation for a wide range of models (and pure wild speculation) wherever accretion is happening in astrophysics (neutron stars, star-sized black holes, supermassive black holes ..); we'll come back to it again and again.

Temperatures. What can we say, simply, about the temperature of the inflowing material? One simple limit is when the material is optically thick. In this case, the luminosity from (6.13) is re-radiated as a black body, for which we know (from thermodynamics) that the luminosity per surface area is $\sigma_{SB}T^4$, wher eσ_{SB} is the Stefan-Boltzmann constant. Equating the luminosity in (6.13) to that lost by black body radiation determines the temperature the gas will reach. This calculation finds $T \sim 10^7$K for accretion onto a solar-sized black hole (and thus we have galactic X-ray binaries); and $T \sim 10^4 - 10^5$K for accretion onto a massive $M \sim 10^9 M_\odot$ black hole (as in a galactic nucleus).

6.4.2 Spherical (Bondi) accretion

To continue, consider spherical accretion specifically. This would describe, for instance, the rate at which a compact object would accumulate matter from the general ISM; or it could describe the growth of a protostar inside a dense cloud (until angular momentum and magnetic fields become important).

We look only at a simple case of spherical accretion, that of smooth, adiabatic inflow. This simple case will be describable by the basic wind equation, (6.9), with $dv/dr < 0$ solutions. The rate of inflow, then, must be determined by conditions "at infinity", that is, in the local ISM. Equation (6.9) has smooth, transonic, steady solutions which start at low velocity at large distances, pass smoothly through the sonic point (6.10) and become supersonic at small radii. These solutions must of course shock down somewhere close to the stellar surface (unless we have a black hole at the origin) – but if the sonic point is well outside the surface, we can assume nearly-steady flow as described by the wind equation, over most of space.

Here, we will consider adiabatic accretion; assuming the inflowing gas is tenuous and hot, and does not have time to cool as it is compresses. Rather than consider full solutions of the force equation (6.9 in this application), we will use a simpler approach to study the nature of these solutions. In particular, we want to estimate the mass inflow rate, $\dot{M} = 4\pi r^2 \rho(r) v(r)$, evaluated at some r where we know ρ and v.

For adiabatic flow, write $p = K\rho^\gamma$ and $c_s^2 = \gamma p/\rho$. From this,

$$\frac{1}{\rho}\frac{dp}{dr} = \frac{\gamma}{\gamma-1}\frac{d}{dr}\left(\frac{p}{\rho}\right)$$

Thus, the basic momentum equation,

$$v\frac{dv}{dr} + \frac{1}{\rho}\frac{dp}{dr} = -\frac{GM}{r^2}$$

tells us that

$$\frac{1}{2}v^2 + \frac{1}{\gamma-1}c_s^2 - \frac{GM}{r} = \text{constant} \qquad (6.14)$$

We can now evaluate this constant

- at $r \to \infty$: $v = 0$ by assumption, so the constant has the value $c_{s,\infty}^2/(\gamma-1)$.
- at $r = r_s$, the sonic point: $v = c_s$, and $GM/r = 2c_s^2$; so the constant is $\frac{1}{2}c_s^2(r_s) + \frac{1}{\gamma-1}c_s^2(r_s) + 2c_s^2(r_s)$.

Equating these two, we can relate the sound speed (that is, the internal energy) at r_s to that of the ISM:

$$c_s^2(r_s) = \left(\frac{2}{5-3\gamma}\right)c_{s,\infty}^2 \qquad (6.15)$$

We can also evaluate the density at r_s, from $c_s^2 = \gamma p/\rho = K\gamma\rho^{\gamma-1}$:

$$\rho(r_s) = \rho_\infty \left(\frac{c_s(r_s)}{c_{s,\infty}}\right)^{2/(\gamma-1)} \qquad (6.16)$$

Thus, we can evalute $\dot{M}$ at r_s, and get our desired result, in terms of ISM quantities:

$$\dot{M} = 4\pi r_s^2 \rho_\infty c_{s,\infty}\left(\frac{c_s(r_s)}{c_{s,\infty}}\right)^{(\gamma+1)/(\gamma-1)} \qquad (6.17)$$

where we have used (6.10 and 6.15) to express r_s in terms of the stellar mass, M, and conditions at ∞. With more algebra we can show

$$\dot{M} = \pi(GM)^2\frac{\rho_\infty}{c_{s,\infty}^3}\left(\frac{c_s(r_s)}{c_{s,\infty}}\right)^{(5-3\gamma)} \qquad (6.18)$$

Thus, the important dependence is, $\dot{M} \propto M^2\rho_\infty/c_{s,\infty}^3$; the $c_s(r_s)/c_{s,\infty}$ term is just an order-unity numerical constant (from 6.15).[1]

Key points

- Why the sound speed is important; why subsonic and supersonic flows can behave differently.
- Stellar winds: the sonic point; behavior inside and outside r_s.
- What happens with $\mathbf{B}$ fields in the solar wind.
- Spherical accretion: basic energetics
- Spherical accretion: Bondi model, how $\dot{M}$ relates to conditions at ∞.

[1] You might be worried that this constant blows up when $\gamma = 5/3$, which is our favorite value for γ. Not so; with l'Hopital's rule you can show that

$$\lim_{\gamma\to 5/3}\left(\frac{2}{5-3\gamma}\right)^{(5-3\gamma)/2(\gamma-1)} \to 1$$

Try it for yourself .. it's a nice little exercise.

7 Wave propagation in plasmas

In this chapter we revisit waves in plasmas, more formally. This can be quite a mathematical topic. I'll store some of the math details in these notes, but try to highlight the physics (especially in class). If you want more detail, or background, good references are Clarke & Carswell, *Principles of Astrophysical Fluid Dynamics*; Chen, *An Introduction to Plasma Physics and Controlled Fusion*; and Choudhuri, *The Physics of Fluids and Plasmas*.

7.1 Plasma Oscillations

First, let's revisit the plasma waves, which we've seen once already. We want a more formal derivation: how do the particle motions and charge separation connect to each other?

7.1.1 Cold plasma

Start with a cold plasma – that is, ignore thermal effects; the only particle motions are those in response to the electric field. The basic equation of motion for particles of charge q is[1]

$$mn\frac{d\mathbf{v}}{dt} = mn\left[\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v}\cdot\nabla\mathbf{v}\right] = qn\mathbf{E} \tag{7.1}$$

But the charge distribution determines $\mathbf{E}$; thus we need to add two more equations,

$$\frac{\partial n}{\partial t} + \nabla\cdot(n\mathbf{v}) = 0\,; \qquad \nabla\cdot\mathbf{E} = 4\pi\rho \tag{7.2}$$

(where ρ is the *net* charge density). For simple plasma waves, we will assume the ions don't move (because they are so heavy), and just look at the electrons.

Now, we want to look for small-amplitude, wavelike disturbances. Thus, we first *linearize* - assume all dependent variables can be broken down into (unperturbed) + (perturbed) parts, with the perturbations being small. That is:

$$n \to n_o + n_1\,; \quad \mathbf{E} \to \mathbf{E}_o + \mathbf{E}_1\,; \quad \mathbf{v} \to \mathbf{v}_o + \mathbf{v}_1 \tag{7.3}$$

We put these into (7.1) and (7.2), and sort terms by their "order in small": the zero-th order terms (n_o, etc) represent any unperturbed equilibrium state, & should cancel out; the second-order-small terms (n_1^2, etc) are small & can be dropped; so we keep just the first-order small terms. That gives us,

$$\begin{gathered}\frac{\partial n_1}{\partial t} + n_o\nabla\cdot\mathbf{v}_1 = 0\,; \quad \nabla\cdot\mathbf{E}_1 = -4\pi e n_1\,; \\ m\frac{\partial \mathbf{v}_1}{\partial t} = -e\mathbf{E}_1\end{gathered} \tag{7.4}$$

These are the equations we want to solve. We make two simplifying choices here. One, we assume[2] each perturbation $f_1(\mathbf{r},t) \to \tilde{f}_1 e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}$, and two, we choose the geometry: pick $\mathbf{E}_1 \parallel \mathbf{v}_1 \parallel \mathbf{k}$. Our three equations (7.4) become

$$\begin{gathered}ik\tilde{E}_1 = -4\pi e\tilde{n}_1\,; \quad -i\omega\tilde{n}_1 = -n_o ik\tilde{v}_1\,; \\ -i\omega m\tilde{v}_1 = -e\tilde{E}_1\end{gathered} \tag{7.5}$$

(remember that $\tilde{n}_1$, etc, are the amplitudes of the perturbations). But now: (7.5) is a linear, homogeneous system in (n_1, E_1, v_1); it has non-trivial solutions only if the determinant of the coefficients is zero. This translates to an important condition on the frequency:

$$\omega^2 = \omega_p^2 = \frac{4\pi n_o e^2}{m} \tag{7.6}$$

Thus, we've recovered the *plasma frequency* – the frequency at which charge-separated perturbations oscillate. NOTE that these are *not* propagating waves, because ω is independent of the wavenumber k (thus the group velocity $v_g = d\omega/dk = 0$).

7.1.2 Warm plasma waves

Now, extend this to include the effects of internal energy in the plasma. That is, add a ∇p term to (7.1):

$$mn\frac{d\mathbf{v}}{dt} = mn\left[\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v}\cdot\nabla\mathbf{v}\right] = qn\mathbf{E} - \nabla p \tag{7.7}$$

Here, $p = nkT$ as usual. We'll assume an adiabatic perturbation, so that $\nabla p = (\gamma p/n)\nabla n$; and for one-dimensional motion, with one degree of freedom, we have $\gamma = 3$. Carry through the same analysis as above, with the extra pressure term; and you find

$$\omega^2 = \omega_p^2 + k^2\frac{3kT}{m} = \omega_p^2 + \frac{3}{2}k^2 v_{th}^2 \tag{7.8}$$

[1] We can think of the LHS as the "total" time rate of change, as seen by the particle moving with velocity $\mathbf{v}$; you can also compare the LHS of the force equation for a fluid, eqn (4.4).

[2] Why can we do this? Think about Fourier analysis – any perturbation can be expressed as the sum of its Fourier components. We can get away with this because equations (7.4) are *linear* in the dependent variables.

where the (1D) thermal speed is $v_{th}^2 = 2kT/m$. These are now propagating waves; they have a nonzero group velocity, $v_g = d\omega/dk$.

7.1.3 Damping: collisional

Plasma waves are subject to two types of damping. One – which you'd expect – is *collisional*. If the electrons collide with other charges as they respond to the wave motion, the wave energy will of course go to heating. The equation of motion for the electrons gains a collisional term:

$$\begin{aligned} mn\frac{d\mathbf{v}}{dt} &= mn\left[\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v}\cdot\nabla\mathbf{v}\right] \\ &= qn\mathbf{E} - \nu_{coll} nm\mathbf{v} \end{aligned} \tag{7.9}$$

where ν_{coll} is the collision rate of an individual charge. (This can be due to Coulomb collisions in a fully ionized plasma, or electron-neutral collisions in a partly ionized one.) It's easy to show that the effect of this is to add an imaginary component to the wave frequency, $\omega = \omega_R + i\omega_I$; and thus our waves are exponentially damped: $E_1 \propto e^{i\omega t} \propto e^{-\omega_I t}$ (the sign of ω_I determines whether the wave grows or is damped; physically for simple collisions it must be damped, right?)

7.1.4 Damping: collisionless

Another, less intuitive, process is *collisionless* damping, also called Landau damping. The derivations of this are usually quite mathematical — but we really need a physical understanding more, so that's all I'll do in these notes. Think about a warm-plasma wave, which travels at a phase speed $v_{ph} = \omega/k$; if a charged particle has v_x exactly equal to v_{ph}, it will be in "resonance" with the wave – it will see exactly the same phase of the wave as it moves along. Now, consider another particle moving at a velocity v_x which is very close to v_{ph}. This particle will see almost a constant phase of the wave; in fact, it will be "captured" by the wave, and will "ride along" with a crest or trough of the wave. If the particle is slightly slower – if $v_x \lesssim v_{ph}\omega/k$ – it will be accelerated up to v_{ph} – draining a wee bit of the wave energy in the process. Conversely, if $v_x \gtrsim v_{ph}$, the particle will be decelerated as it is "captured", thus giving a bit of energy up to the wave.

Think, then, about the net effect of this particle-wave interaction in a plasma. Two cases are possible (refer to class notes for a cartoon).

- If the distribution function $f(v_x)$ of the plasma is "normal", say a Maxwellian as you'd expect from thermal equilibrium, then it has negative slope: $df/dv_x < 0$ at $v_x \simeq v_{ph}$. Thus, there will be slightly more particles at $v_x < v_{ph}$ then at $v_x > v_{ph}$. That means that more particles will gain energy from the wave than will lose energy to the wave: the wave is *damped*, without needing any particle-particle collisions. This is *Landau damping*.
- What if the distribution function has a positive slope – $df/dv_x > 0$ at $v_x = v_{ph}$? Well, the opposite will occur: the wave will gain energy at the expense of the plasma. But how do you get such a distribution function? Think about two plasmas impinging on one another – or trying to send a "beam" or "stream" of charges through a thermalized background plasma. This situation is unstable: the energy of the relative motion between the plasma streams is fed into waves of the appropriate phase velocity. This is called the *two-stream instability*. Because of this instability, one plasma trying to penetrate another cannot do so – instead a background of turbulent plasma waves is generated.

Isn't this last result curious? You might naively think that two low-density plasmas could interpenetrate each other without any problem, as long as Coulomb collisions are unimportant. But that's not the case: the two-stream instability, which also arises from long-range Coulomb interactions, tries to keep the plasmas separate.

7.2 EM wave propagation: B = 0

In the last section we considered intrinsic oscillations in the plasma. Now we change the setup: send in an EM wave (and let's stick to cold plasma). This means we'll have two important changes in our analysis – (i) we need to include $\mathbf{B}$ effects (at least the $\mathbf{B}$ field of the wave), and (ii) now we take $\mathbf{E}_1, \mathbf{B}_1 \perp \mathbf{k}$ (EM waves are transverse, after all). We want to know how the wave propagates and how it affects the plasma; and – for astrophysics – how to interpret observations in terms of what's going on in the plasma.

7.2.1 Basic: the dispersion relation

Start here with the full Maxwell set:

$$\begin{aligned} \nabla\cdot\mathbf{E} = 4\pi\rho\,; &\quad \nabla\times\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} \\ \nabla\cdot\mathbf{B} = 0\,; &\quad \nabla\times\mathbf{B} = \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi}{c}\mathbf{j} \end{aligned} \tag{7.10}$$

We are still imposing plane waves, $e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}$; putting these into Maxwell, we get

$$\begin{aligned} i\mathbf{k}\cdot\mathbf{E} = 4\pi\rho\,; \quad i\mathbf{k}\times\mathbf{E} = i\frac{\omega}{c}\mathbf{B} \\ i\mathbf{k}\cdot\mathbf{B} = 0\,; \quad i\mathbf{k}\times\mathbf{B} = -i\frac{\omega}{c}\mathbf{E} + \frac{4\pi}{c}\mathbf{j} \end{aligned} \tag{7.11}$$

Now work on the source terms; we know $\mathbf{j} = -ne\mathbf{v}$ (again, only the electons move). We still have the simple equation of motion, (7.1), and its linearized/wave mode form, $im\omega\mathbf{v} = e\mathbf{E}$. Thus, we can write[3]

$$\mathbf{j} = \sigma\mathbf{E}\,; \qquad \sigma = \frac{ine^2}{m\omega} \tag{7.12}$$

We need to close the system with conservation of charge:

$$\frac{\partial\rho}{\partial t} + \nabla\cdot\mathbf{j} = 0 \quad \Rightarrow \quad i\omega\rho = i\mathbf{k}\cdot\mathbf{j} \tag{7.13}$$

Now, put these back in (7.11) and collect terms:

$$\begin{aligned} i\mathbf{k}\cdot(\epsilon\mathbf{E}) = 0\,; \qquad i\mathbf{k}\times\mathbf{E} = i\frac{\omega}{c}\mathbf{B} \\ i\mathbf{k}\cdot\mathbf{B} = 0\,; \quad i\mathbf{k}\times\mathbf{B} = -i\frac{\omega}{c}(\epsilon\mathbf{E}) \end{aligned} \tag{7.14}$$

Here, we've identified the *dielectric factor*

$$\epsilon = 1 - \frac{4\pi\sigma}{i\omega} = 1 - \frac{\omega_p^2}{\omega^2} \tag{7.15}$$

(Comment: many books call this the dielectric constant, in analogy with EM propagation in dielectric materials – but here ϵ is a function of frequency, so it's hardly constant). By analogy with what you've seen of EM waves elsewhere, once we have the equations in the form (7.14), we know the wave solution is $\omega^2\epsilon = c^2k^2$. Thus, our dispersion relation for the wave propagating in the plasma is

$$\omega^2 = k^2c^2 + \omega_p^2 \tag{7.16}$$

and we want to look at simple applications of this.

7.2.2 Applications and extensions

Several important phenomena deserve mention here.

• **Plasma cutoff**. It's clear from (7.16) that real k's are allowed only if $\omega > \omega_p$: *waves below the plasma frequency do not propagate*. Why is this? Mathematically, think about k being imaginary: the wave form $e^{ikx} \to e^{-|k|x}$, that is the wave damps exponentially. Physically, think about how the plasma responds to an incoming EM wave of frequency ω. For $\omega \ll \omega_p$, the plasma charges can easily "move up and down" with the wave – that is they can easily absorb the wave energy. But for $\omega \gg \omega_p$, the charges can't move fast enough (think about driving an oscillator well above its natural frequency) – so the wave happily propagates through the plasma.

• **Plasma dispersion**. Because $\omega(k)$ in (7.16) is dispersive, waves at different frequencies move at different phase speeds. This is important when you're looking at an astrophysical object – such as a pulsar - which emits very short pulses. The arrival time of such a pulse at earth, from a distance D away, depends on the frequency. Remembering that $v_g = d\omega/dk$, the arrival time of a pulse is (switching to $\nu = \omega/2\pi$, to connect to observations)

$$t_p(\nu) = \int_0^D \frac{ds}{v_g} \simeq \int_0^D \left(1 + \frac{\nu_p^2}{\nu^2}\right)^{1/2} \frac{ds}{c} \tag{7.17}$$

(You should note that I've assumed $\nu \gg \nu_p = \omega_p/2\pi$ here.) The frequency-dependent term in (7.17) is usually scaled as

$$\int_0^D \frac{ne^2}{\pi mc}\frac{1}{\nu^2}ds \simeq 4.15\times 10^{15}\frac{(\mathrm{DM})}{\nu^2}\ \mathrm{sec} \tag{7.18}$$

where the *dispersion measure*, $\mathrm{DM} = \int_0^D n ds$, is measured in cm^{-3}pc, and the frequency ν is in Hz.

• **collisional damping**. As with plasma waves, EM waves also can damp out in a plasma. Equation (7.9) is still the starting point. Working this through for our EM wave, we find the dispersion relation is

$$\omega^2 = \frac{\omega_p^2\omega}{\omega + i\nu_{coll}} + c^2k^2 \tag{7.19}$$

This looks rough; it's a cubic equation in ω. However things simplify if the damping rate is small compared to the wave frequency (or its real part). The solutions to (7.19) must be complex. To approach them, we can either hold ω real and find complex k, or vice versa. The first approach corresponds to driving a plasma with an incoming wave of fixed ω; the imaginary part of k corresponds to the spatial damping (wave absorption by the plasma). Alternatively, we can hold k fixed and

[3] What does an imaginary conductivity, σ, mean? Remember that $e^{-x} = \cos x + i\sin x$; what does an imaginary σ tell us about the phase relation between $\mathbf{j}$ and $\mathbf{E}$?

look for the imaginary part of ω – some authors prefer this approach. In either case: (7.19) is a cubic, and potentially intimidating; but solutions simplify if the damping is weak, $\nu_{coll} \ll \omega$ (or $\nu_{coll} \ll \Re(\omega)$ if you prefer).

7.3 EM wave propagation: finite B

Well, this has been so much fun, let's do it again. Now add a magnetic field $B = B\hat{\mathbf{z}}$ to the background plasma, and consider waves propagating along the field: $\mathbf{k} \parallel \mathbf{B}$ (this is the only simple geometry!).

7.3.1 Basic: the dispersion relation

We'll only do the outline here, you have the details above. The equation of motion for the electrons now is

$$m\frac{d\mathbf{v}}{dt} = -e\mathbf{E} - \frac{e}{c}\mathbf{v} \times \mathbf{B} \tag{7.20}$$

(caution: here $\mathbf{E}$ is the wave field, but $\mathbf{B}$ is the background field. I'm assuming the wave B field is small & ignoring it). Now: for the incoming EM wave we choose circular polarization:

$$\mathbf{E} = Ee^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}\left(\hat{\mathbf{x}} \pm i\hat{\mathbf{y}}\right) \tag{7.21}$$

where the $\pm$ signs pick out RH or LH circular polarization. (Why? think about the response of the plasma charges – the electrons have a preferred sense of gyromotion – so we might expect RH and LH circularly polarized waves to have different phase speeds – as the electrons might help or hinder them.) Carry out the same type of analysis – the electron response is

$$\mathbf{v} = \frac{-ie}{m(\omega \pm \Omega)}\mathbf{E} \tag{7.22}$$

where $\Omega = eB/mc$ is the gyrofrequency. The dielectric factor becomes

$$\epsilon_{R,L} = 1 - \frac{\omega_p^2}{\omega(\omega \pm \Omega)} \tag{7.23}$$

Thus, RH and LH waves do propagate at different phase speeds in the medium. The important application of this is to the angle of linear polarization of an incoming wave. We also get the dispersion relation, directly, from this, from the definition of ϵ:

$$\omega^2 = c^2k^2 + \frac{\omega_p^2}{(1 \pm \Omega/\omega)} \tag{7.24}$$

7.3.2 Applications and extensions

Once again we have some important applications.

• **"Plasma" cutoffs** also occur here, they're just a bit more complicated. We again want the frequency, from (7.23), at which $k = 0$ – that's the transition between real (propagating) and imaginary (damped) waves. The answer depends on the sign choice in (7.23) (that is, on whether the waves are RH or LH polarized). We get the critical frequencies for transmission:

$$\omega_{R,L} = \frac{1}{2}\left[\pm\Omega + \left(\Omega^2 + 4\omega_p^2\right)^{1/2}\right] \tag{7.25}$$

Thus, LH waves only propagate for $\omega > \omega_L$. However, RH waves propagate above ω_R and also below Ω (you check the algebra: there are two domains of $k^2 > 0$ for this polarization).

• **Faraday rotation** is one of the most important astrophysical methods for measuring magnetic fields. Remember that the phase of a wave, which has travelled a distance D, is $\phi = \int_0^D kds$. Thus the phase difference between R,L waves is

$$\Delta\phi = \int_0^D (k_R - k_L)ds\ ; \quad k_{R,L} = \frac{\omega}{c}\sqrt{\epsilon_{R,L}} \tag{7.26}$$

Now, because a linearly polarized wave can be written as the sum of RH and LH circularly polarized waves, the angle χ by which the polarization is rotated is 1/2 of the phase difference: $\chi = (\Delta\phi)/2$.

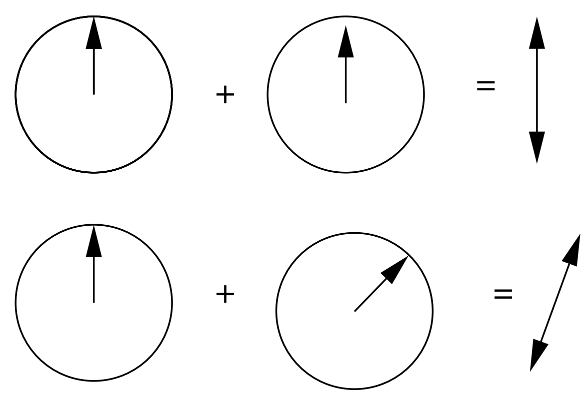

Figure 7.1 Decomposition of linear polarization into components of right and left circular polarization. Top, RC and LC in phase; bottom, phase shift $\Delta\phi$ between RC and LC (as due to Faraday rotation), rotates the plane of polarzation by $\chi = \Delta\phi/2$. Following Rybicki & Lightman figure 8.1.

But now, going back to (7.21) and assuming that our observed wave frequency is well above the plasma and cyclotron frequencies of the plasma, we can expand k as

$$k_{R,L} \simeq \frac{\omega}{c}\left[1 - \frac{\omega_p^2}{2\omega^2}\left(1 \mp \frac{\Omega}{\omega}\right)\right] \tag{7.27}$$

Putting this back in (7.22) and doing more algebra, we get

$$\chi = \frac{\lambda^2}{2\pi}\frac{e^3}{m^2c^4}\int_0^D n\mathbf{B}\cdot d\mathbf{s} \tag{7.28}$$

(Note that I've gone to the more general $\mathbf{B}\cdot d\mathbf{s}$ in the integrand; for a general direction of wave propagation relative to $\mathbf{B}$, it turns out to be the component of $\mathbf{B}$ along the line of sight that matters.) This result, (7.27), is generally scaled as $\chi = \lambda^2(\mathrm{RM})$, where the *rotation measure* is

$$\mathrm{RM} = \frac{1}{2\pi}\frac{e^3}{m^2c^4}\int_0^D n\mathbf{B}\cdot d\mathbf{s} \tag{7.29}$$

A convenient numerical scaling for rotation measure is

$$(\mathrm{RM}) \simeq 810\int_0^D n\mathbf{B}\cdot d\mathbf{s}$$

for n in cm^{-3}, B in μG, s, D in kpc, and RM in rad/m^2.

Key (physical) points

• Wave analyses (e.g. plasma waves): the mathematical attack (start with basic equations; linearize; assume $e^{i(kx-\omega t)}$ single-frequency solutions; do the math and see what you get.

• Plasma waves: when do they propagate, how fast, how do they damp?

• Collisional dissipation (of various types of waves)

• EM waves in plasma: plasma dispersion effects

• EM waves in plasma: $\mathbf{B} \neq 0$ effects, Faraday rotation

8 MHD: more applications

In this chapter and the next, we'll work through some important fluid-type applications in which MHD effects are important. Our focus here will be MHD in the galactic setting. We'll revisit Alfven waves, and look at how they interact with galactic cosmic rays. We'll also look at magnetic effects on buoyant instabilities, and again apply it to the galactic setting.

8.1 MHD waves, again

First, let's revisit Alfven waves, which we've seen once already. We again want a more formal derivation: how do the particle motions and magnetic tension (the restoring force here) connect to each other?

8.1.1 Alfven waves: gory details

You remember these waves, from chapter 5. They are transverse waves, which are not compressive, and which propagate (in the simplest case) along the magnetic field. Thus, they can be thought of as propagating wiggles in the field lines, as in Figure 8.1.

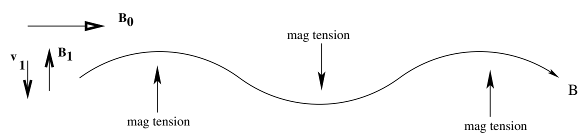

Figure 8.1 Schematic of (shear) Alfven waves, propagating along a background field $\mathbf{B}_o$. The perturbed $\mathbf{B}_1$ and $\mathbf{v}_1$ terms are perpendicular to $\mathbf{B}_o$. Following Cravens Figure 4.16.

In chapter 5 we used simple, restoring-force arguments to guess their dispersion relation as $\omega = kv_A$, with $v_A = B/\sqrt{4\pi\rho}$. Now let's derive that, more formally, to understand how the waves work. To do that, we'll use the same mathematical techniques as in the last chapter. Consider a uniform fluid at rest: zero velocity, uniform B and mass density ρ. Consider the behavior of a small perturbation – $\mathbf{B} = \mathbf{B}_o + \mathbf{B}_1$. Pick $\mathbf{B}_o$ along $\hat{\mathbf{z}}$, and the wave field $\mathbf{B}_1 = B_1\hat{\mathbf{y}}$ along $\hat{\mathbf{y}}$. Because variable $\mathbf{B}$ leads to $\mathbf{E}$ (from Maxwell, right?), we also have the wave electric field $\mathbf{E}_1 \parallel \hat{\mathbf{x}}$.

• **Starting points & linearization.** We need to follow both $\mathbf{E}_1$ and $\mathbf{B}_1$, so we need both Maxwell equations:

$$\nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}\ ;\ \ \nabla \times \mathbf{B} = \frac{4\pi}{c}\mathbf{j} + \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} \quad (8.1)$$

We linearize as usual, picking a perturbation $\propto e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)}$. Maxwell then becomes

$$i\mathbf{k}\times\mathbf{E}_1 = \frac{i}{c}\omega\mathbf{B}_1\ ; \quad i\mathbf{k}\times\mathbf{B}_1 = \frac{4\pi}{c}\mathbf{j}_1 - \frac{i}{c}\omega\mathbf{E}_1 \quad (8.2)$$

The current density $\mathbf{j} = ne(\mathbf{v}_i - \mathbf{v}_e)$, so we need the equation of motion, for charge q, as usual:

$$\begin{aligned} m\frac{d\mathbf{v}}{dt} &= q\mathbf{E} + \frac{q}{c}\mathbf{v}\times\mathbf{B}\ ; \quad \Rightarrow \\ -\, im\omega\mathbf{v} &= q\mathbf{E}_1 + \frac{q}{c}\mathbf{v}\times\mathbf{B}_o \end{aligned} \quad (8.3)$$

In the last, we've linearized *and* implicitly assumed $\mathbf{v}$ is "small" (so we've dropped $\mathbf{v}\times\mathbf{B}_1$).

• **Particle motion.** We need to follow both electrons and ions. For ions, we keep all the terms in (8.3), and write out both components:

$$\begin{aligned} -i\omega m_i v_x &= eE_1 + e\frac{v_y}{c}B_o\ ; \\ -i\omega m_i v_y &= -e\frac{v_x}{c}B_o \end{aligned} \quad (8.4)$$

These solve to give the ion motion:

$$\begin{aligned} v_{ix} &= \frac{ieE_1}{m_i\omega}\left(1 - \frac{\Omega_i^2}{\omega^2}\right)^{-1}\ ; \\ v_{iy} &= \frac{eE_1}{m_i\omega}\frac{\Omega_i}{\omega}\left(1 - \frac{\Omega_i^2}{\omega^2}\right)^{-1} \end{aligned} \quad (8.5)$$

For the electrons, we can simplify by assuming $\Omega_e \gg \omega$; this gives us

$$v_{ex} = 0\ ; \quad v_{ey} = -\frac{E_1}{B_o}c \quad (8.6)$$

• **Cut to the chase.** These results are all the necessary bits. We put (8.6) and (8.5) into the definition for $\mathbf{j}$, then put this into (8.2) and combine the two equations there into one (by eliminating $\mathbf{B}_1$, say). After a page or so of algebra, we get the dispersion relation for these Alfven waves:

$$\omega^2\left(1 + \frac{4\pi n m_i c^2}{B^2}\right) = c^2k^2 \quad (8.7)$$

With some manipulation, this can be written as

$$\omega^2 = \frac{v_A^2k^2}{\left(1 + v_A^2/c^2\right)} \quad (8.8)$$

where $v_A^2 = B^2/4\pi\rho$, as we saw before (the mass density is $\rho = nm_i$, because the electron mass is so small). Thus, when $v_A \ll c$ (as is usually the case), we have $\omega \simeq kv_A$; or when $v_A \gg c$ (the low-density, high-field limit), we get $\omega \rightarrow ck$.

8.1.2 Magnetosonic waves

Another type of wave is also of interest. Think of a compressive wave, something like a sound wave, but propagating *across* the magnetic field. The magnetic pressure will modify the restoring force; we might expect the wave speed to be a mixture of compressive effects (through c_s) and magnetic effects (through v_A). These are called *magnetosonic waves*.

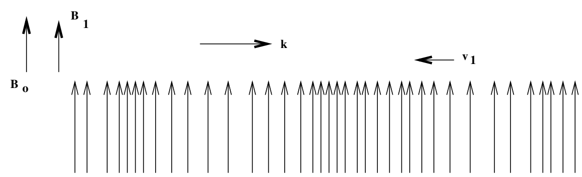

Figure 8.2 Schematic of magnetosonic waves, illustrating a compressive wave propagating at right angles to the background B_o. Following Cravens figure 4.17.

The full analysis of the wave is messy, and not worth doing here; but *for the simple case of propagation across* $\mathbf{B}$ one finds a wave speed which is what we might well guessed anyway:

$$v_{MS}^2 = c_s^2 + v_A^2 \tag{8.9}$$

Because these waves are compressive, they tend to damp out more easily than Alfven waves (think of frictional effects in the dense regions). Thus they may be of less interest in general astrophysical situations, than Alfven waves (which turn out to be only very weakly damped).

8.2 The cosmic ray-Alfven wave connection

Alfven waves are particularly interesting because they have resonant interactions with relativistic particles, such as galactic cosmic rays (CR).

8.2.1 Cosmic rays: a quick overview of the observations.

We have already noted that many astrophysical plasmas – including the ISM – contain a significant population of highly relativistic particles which are *not* in a thermal distribution. We have direct and indirect evidence of these particles, which I review very briefly here.

Baryons. Here we have direct evidence – these are the cosmic rays. We can directly measure their flux (including details of composition, energy, charge, and isotropy) at earth; and from there work backwards to model their distribution above the earth's atmosphere. (Connecting further back, to their composition outside the heliopause, is harder). Significant facts about their distribution include

• Their energy distribution is a power law, $N(E) \propto E^{-s}$, with a break at $E \sim 10^{15}$eV (the "knee"), and another at $E \sim 10^{19}$eV (the "ankle"). The exponent $s \sim 2.7$ below the ankle, and higher above. Comparison of the gyroradius to the scale of the galaxy suggests that the highest energy CR, above the ankle, are extragalactic, while the lower energy ones are galactic in origin.

• Their composition is mostly protons, but there is a heavy element component, with approximately solar abundances (so they come from processed material).

• Cosmic rays are very isotropic in arrival direction, probably at all energies.[1] In terms of their origin, this requires one of three things. (i) The CR are very local, from something like the Oort cloud; this seems unlikely. (ii) If the CR are galactic in origin, they must be isotropized in propagating from their sources (which would lie in the plane of the galaxy) to us. (iii) The CR could be cosmological in origin; this idea has its own difficulties which arise in interactions between the CR and photons of the microwave background.

Leptons. Here we have some direct evidence – the lepton component in the cosmic ray spectrum can be distinguished from the baryon component. The cosmic ray lepton distribution falls much more steeply with energy than that of the baryons, which is probably due to the stronger radiative losses the leptons suffer. We'll return to this topic next term.

8.2.2 Cosmic rays in the galactic setting

We have two big questions: how are cosmic rays accelerated to such high energies, and how do they interact with the ISM? We'll defer the acceleration question to next term, but can think here about how they interact with the ISM.

[1] NOTE here: this has been an ongoing discussion. Early work suggested that this was the case only at low energies; above the ankle, it had been believed that the highest enery CR were anisotropic, with a tendency to come from the Virgo cluster (remember that's the cluster that hosts one of the nearest AGN, namely M87). Since then this has come under debate. Some newer work however shows that even the highest energy CR show no strong evidence of anisotropy, but other authors still support an anisotropy.

Sources are thought to be two-fold. *Supernova remnants* have long been thought to be the main source of CR. This is because of their importance in the overall ISM energy budget; because we know they eject heavy elements and the isotopic composition of CR is roughly that of processed stellar material; and because we believe that their outer shocks should be good at accelerating individual particles in the ejecta, up to the relativistic energies typical of CR. In addition, it occurs to me that *pulsars* are probably also an important source. We will see that most models predict pulsars should accelerate single charges to very high energies ($\gamma \sim 10^7$ for instance) close to the star's surface. Some of these particles must go to drive the pulsar wind, but others may well escape directly into the ISM to be a secondary CR source. Still another possibility is that the highest energy CR may have an *extragalactic origin*. This is suggested (i) because of the possibile anisotropy, and (ii) because the highest energy ones are unlikely to be "trapped" by the ISM. How these very fast particles are accelerated is one of the major unsolved puzzles – one possibility is that they are generated around the massive black holes in nearby active galactic nuclei.

Propagation. Once generated, CR do not just fly freely through space. Because they are charged, they are connected to the ISM by their gyromotion, and by scattering on turbulent Alfven waves in the ISM (more details on this, below). Thus the CR distribution we observe at earth may well have been seriously changed, relative to their "birth" distribution, by propagation and scattering through the ISM on their way to us.

Losses. The leptons, being of smaller mass, are susceptible to radiative losses (synchrotron radiation in the galactic magnetic field, inverse Compton scattering on whatever radiation is around) as well as Coulomb losses (scattering on the plasma component of the ISM). This also modifies the electron energy distribution, compared to the source, and of course reduces the net energy in the electron component of the CR. The baryons, on the other hand, don't radiate much – but there is an interesting argument on their confinement lifetime in the ISM, as follows.

Lifetimes. To investigate the confinement of CR baryons, we must look at the evidence on radioactive isotopes and on spallation rates (that is nuclear interactions with ISM protons). The upshot is, that the time an average CR has hung around the galaxy (determined by radioactive decay) is about ten times longer than thet time it has spent propagating through the ISM (determined by spallation rates). From this we learn that an average CR spends most of its life in the galaxy, *not* in the disk, but rather in the more extended halo; and that its lifetime to escape from that halo ~ 20 Myr.

8.2.3 Alfven waves and wave-particle resonance

We have already seen that cosmic rays are "tied to magnetic field lines", due to their gyromotion. There is another important effect: Alfven waves have a strong resonant interaction with the cosmic rays. This interaction can (i) scatter and isotropize the particles; (ii) slow down an initially anisotropic particle beam from $v \sim c$ to $v \sim v_A$ (this is called the "Alfven speed limit" in the trade); and (iii) possibly accelerate the particles.

Particle-Wave Resonance. Charged particles interact resonantly with Alfven waves. A particle moving along $\mathbf{B}$ at some velocity v sees a Doppler shifted frequency $\omega' = \omega - kv = \omega(1 - v/v_A)$. Now, the particle will interact with the fluctuating E field of the wave; if the particle "stays in phase" with this fluctuating wave, it will undergo a strong interaction. This happens if the Doppler shifted wave frequency is close to the particle's natural frequency, its gyrofrequency. That is, the interaction is strong when

$$\omega - kv = \pm\Omega \tag{8.10}$$

For relativistic particles, with $v \gg v_A$, this condition solves to an approximate equality between the particles gyroradius and the wave's wavelength:

$$r_L(\gamma) = \gamma\frac{mc^2}{eB} \simeq \lambda_{res}(\gamma) \tag{8.11}$$

Numerically, note that particles with $\gamma mc^2 \sim 1$ GeV, in a field $B \sim 1\mu$G, have a gyroradius – and thus a resonant Alfven wavelength – on the order of an AU.

Wave-particle scattering. Assume, now, that we have a field of Alfven waves, including waves at the right wavelength λ to resonate with particles at some momentum p, and some gyroradius $r_g(p)$. Let the waves *at this wavelength* have amplitude B_1, and energy density $U_{res} = B_1^2/8\pi$. Their energy will be small compared to the mean field energy, $U_{Bo} = B_o^2/8\pi$. Thus the deviation of the field lines in the wave, from the mean field direction, is $\delta\phi \sim B_1/B_o$. We can describe the effect of the waves on the particles by noting that, when a resonant particle stays in phase with a wave over one wave period, its pitch angle will have changed by $\sim \delta\phi$ compared to its initial pitch angle. However, this is a random process; after N encounters with Alfven waves, the particle's net change in pitch angle is $\phi = \sqrt{N}\delta\phi$. Thus, the waves act as scattering centers. We can define a scattering mean free path as the distance over which a particle's pitch angle changes by $\phi \sim 1$ rad:

$$\lambda_{mfp} = N\lambda \simeq \left(\frac{B_o}{B_1}\right)^2 r_g = \frac{U_{Bo}}{U_{res}} r_g \qquad (8.12)$$

Thus: if we know the energy density of *resonant* waves, we can find the mean free path, and thus the collision time, for the particles to scatter on the waves. This is the mechanism by which relativistic particles are tied to the galactic system; and it is probably important in the acceleration of the particles to these high energies.

Wave growth. When Alfven waves exist at the right $\lambda_{res}(\gamma)$, then, particles with energies satisfying (8.11) will be strongly tied to the background plasma. We should note that there are a couple of ways in which Alfven waves can be generated. First, as mentioned above, any disturbance to the background plasma will generate Alfven and sound waves. Wave sources could include stellar winds, cloud motions, novae and supernovae, stellar random velocities, etc. In addition, it turns out that the resonant particles themselves can generate the waves. It turns out that particles with (i) an anisotropic velocity distribution, such as in a directed beam, and (ii) streaming speed $v \gg v_A$ will generate Alfven waves,[2] at wavelengths given by (8.11). These self-generated waves will then scatter and isotropize the particles, reducing their streaming speed to $\sim v_A$.

Energy limits for wave-particle interaction. What are the limits on particle energies that can be scattered and accelerated by Alfven waves? The upper limit is given by the maximum wavelength that can exist in the system; this can't be larger, clearly, than the scale size of the system. For the lower limit, it turns out that Alfven waves can only exist for frequencies $\omega < \Omega_p = eB/m_p c$, the *subrelativistic* proton gyrofrequency. (At higher frequencies, the proton response complicates the wave behavior, and the wave changes nature). Thus, there is a maximum wavenumber $k_{max} = \Omega_p/v_A$, and a minimum resonant wavelength, $\lambda_{min} = 2\pi/k_{max}$, which can exist for Alfven waves. From (8.10), we see that there is a minimum particle energy which can "see" Alfven waves. When one works out the details of the algebra, it turns out for protons,

$$\gamma_{min,p} = 1 + \frac{2v_A^2}{c^2\mu^2} \qquad (8.13)$$

where μ is the cosine of the particle's pitch angle. For electrons, the equation is best solved in limiting cases. In a high-density plasma, with $v_A^2 \ll \left(m_e^2/m_p^2\right) c^2\mu^2$, the minimum electron energy that satisfies (8.7) is

$$\gamma_{min,e} = 1 + \frac{v_A^2}{c^2\mu^2}\frac{m_p}{m_e} \qquad (8.14)$$

In the opposite limit, for a low-density plasma, the minimum energy for resonance becomes

$$\gamma_{min,e} = \frac{m_p}{m_e}\frac{v_A}{c\mu} \qquad (8.15)$$

Thus, the acceleration of low energy (say, thermal) particles by any mechanism that depends on resonant Alfven waves is quite different for electrons and protons. A significant number of thermal protons can resonate with Alfven waves, and can (in principle) be accelerated by them. But thermal electrons have less chance of acceleration; in a low density plasma, especially, they must already have $\gamma \gg 1$ to resonate with the waves.

8.3 Magnetic Buoyancy

Magnetic fields can affect buoyant behavior in various ways. Several situations are possible, involving simple buoyant bubbles, or buoyant flux tubes. Applications range from solar prominances, to the gaseous halo of our galaxy, to the structure of some radio galaxies. You've already worked with the buoyant rise speed (in previous homework). Here we need to think about whether or not an equilibrium (such as a hydrostatic atmosphere, or the galactic disk) is buoyantly unstable.

8.3.1 Convective stability (unmagnetized)

To set the stage, we first need to look at the problem without magnetic effects. Following Shore[3], let's think about a duck.

[2] Think back to chapter 7 and the two-stream (beam) instability, in which a plasma "beam" trying to penetrate another plasma generates turbulent plasma waves. The detailed physics is different here and there, but the general picture is the same – a beam/plasma system and a resonant interaction.

[3] *An Introduction to Astrophysical Hydrodynamics* (Academic Press) 1992, ch. 9.

Picture a duck sitting calmly on a pond. . . If we say that the bird is buoyant, we mean that if we depress him a bit by pushing from above, he will bob back to the surface and, ignoring his agitation, bounce up and down for awhile. This is called *neutral buoyancy*. If, on the other hand, our duck is not well preened and therefore is not waterproof, and we push down on him, he may sink. Now, think of a blob which is hotter than its surroundings. It will begin to rise, since we already know that its density will be lower than that of the surrounding medium, and it will thus be buoyant. If it remains underdense, it will continue to rise – we call this an *instability*. It will continue to rise until it reaches a level at which it is neutrally buoyant again. On the other hand, if the blob is pushed down, and if it remains overdense, it will sink until it reaches a point at which the density again allows for stable balance.

Now, move on to a more relevant example, namely an unmagnetized, hydrostatic atmosphere. Think about a blob ("parcel") in this atmosphere: assume it starts at some vertical position z, with density and pressure in balance with its surroundings (as in Figure 3). Thus: it starts at $\rho_{in} = \rho_{out} = \rho_1$ and $T_{in} = T_{out} = T_1$. Now, raise it some distance dz, and assume it evolves *adiabatically*. Thus, it reaches a new density and temperature,

$$\rho^*_{in} = \rho_1 + \left(\frac{d\rho}{dz}\right)_{ad} \;;\; T^*_{in} = T_1 + \left(\frac{dT}{dz}\right)_{ad} \qquad (8.16)$$

The surroundings, however, are not necessarily adiabatic: they have some other dT/dz and $d\rho/dz$ values (specified by the situation – for instance the heating/cooling balance for the outer layers of a star).

But now: the blob will be unstable if, when it has risen this δz, it is at a *lower* density than the surrounding atmosphere – in that case it will keep rising. Thus, our condition for instability is just a condition on the external density and temperature gradients. Therefore, the atmosphere is buoyantly unstable if $(d\rho/dz)_{ad} < (d\rho/dz)_{atm}$. Because we usually consider situations with $d\rho/dz < 0$, for instance a hydrostatic atmosphere, the condition for *instabilty* is often written in terms of absolute values:

$$\left|\frac{d\rho}{dz}\right|_{ad} > \left|\frac{d\rho}{dz}\right|_{atm} \quad : \quad \left|\frac{dT}{dz}\right|_{ad} < \left|\frac{dT}{dz}\right|_{atm} \qquad (8.17)$$

Thus: if the outside (atmospheric) temperature changes too rapidly with altitude, the atmosphere is convectively unstable. An underdense blob will continue to rise, and an overdense blob will sink.

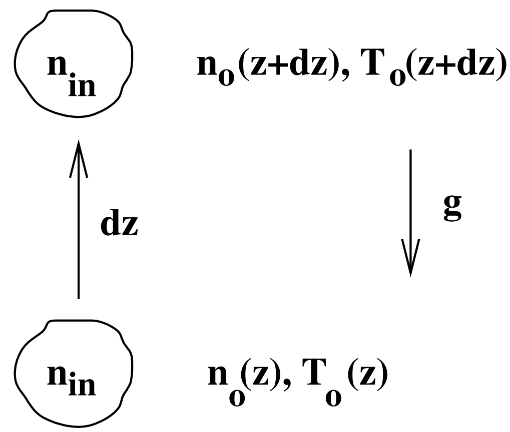

Figure 8.3 Setting the stage for the buoyant (in)stability. A blob starts at height z, in a local gravitational field $\mathbf{g}$. The blob is initially at the same pressure and density as its surroundings: $n_{in} = n_o, T_{in} = T_o$. It rises slowly, staying in pressure balance with outside; it expands adiabatically as it rises, going to some n^*_{in}. The question is, how does its new density compare to the density outside?

8.3.2 Bouyant instability, magnetized

How will a B field change things? To think about it, let's replace our blob with a more realistic geometry, such as a flux tube rising from the surface of the sun (as in Figure 4). We will find another condition is necessary for instability (in addition to the structure of the surrounding atmosphere). In this geometry, instability requires overcoming magnetic tension (holding the ends of the flux rope down) as well as simple convective instability. To see this, consider an isolated flux tube, such as might give rise to a sunspot. Let the external gas have a density scale height $H = k_B T/mg$ (refer back to earlier work for hydrostatic equilibrium with no magnetic field). If the magnetic field is confined in the flux tube, (no B field outside), and the intial state is in pressure balance, we can again write internal-external pressure balance as $p_i + B^2/8\pi = p_e$. Assuming the gas inside is at the same temperature, it must be at lower density than the outside. Thus leads to a buoyancy force, as you found before:

$$F_{buoy} = g\,(\rho_e - \rho_i) = g\Delta\rho = \frac{B^2}{8\pi H} \qquad (8.18)$$

Say, now, that the flux tube is bent upwards, locally, with a radius of curvature R. If R is short, magnetic tension will pull the tube back towards its initial position, giving a stable system. If, however, R is long, the

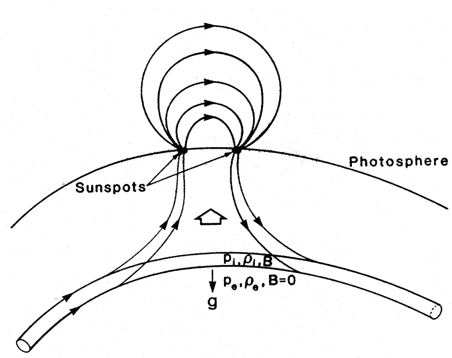

Figure 8.4 The geometry of a sub-surface flux tube before it erupts from the sun due to buoyancy, and its possible post-eruption state; from Tajima &Shibata figure 3.17.

buoyancy force will overcome the tension, leading to instability. Comparing these two forces, we find instability occurs if

$$g\Delta\rho > \frac{B^2}{4\pi R} \tag{8.19}$$

and thus the flux rope is unstable if $R > 2H$.

8.3.3 Parker instability

Now, let's apply this to a different geometry, a 1D planar system (such as the ISM in the galaxy). In this context magnetic buoyancy is referred to as the *Parker instability*. Let the magnetic field be horizontal, and fully mixed with the gas. Picking $\hat{\mathbf{z}}$ as the vertical direction, that means we take $\mathbf{B} = B_y(z)\hat{\mathbf{y}}$; and describe the gas by density $\rho(z)$, pressure and sound speed $p(z) = c_s^2\rho(z)$, and gravitational field $\mathbf{g}$.

We need a model for the unperturbed state. The simplest assumption we can make is that the ratio of gas to magnetic pressure is constant at all altitudes: $B^2/8\pi p = \alpha_o$ = constant. Magnetostatic balance, then, can be written

$$(1+\alpha_o)\frac{dp}{dz} = -\frac{p}{c_s^2}g \tag{8.20}$$

This again gives us a simple exponential atmosphere. The difference here is that the scale height involves the magnetic as well as thermal pressures:

$$\frac{p}{p_o} = \frac{\rho}{\rho_o} = \frac{B^2}{B_o^2} = e^{-z/\Lambda} \tag{8.21}$$

where $\Lambda = \left(c_s^2 + v_A^2/2\right)/g$ is the new scale height. (Check: yes this is larger than the unmagnetized H – which makes sense, due to the extra pressure suport of the magnetic field).

Now, consider a small perturbation: let a horizontal layer (or flux tube) be raised some small Δz. The plasma in this layer can flow freely along the field lines – thus it will "slide down to the bottom", leaving the top part of the perturbation at a lower density than its surroundings. If the perturbation has horizontal scale λ, simple geometry gives us its radius of curvature: $r \simeq \lambda^2/8\Delta z$. To check the stability of this perturbation, we again compare the buoyant forces to the restoring force, magnetic tension. Let "o" refer to the altitude of the undisturbed sheet. The density in the gas at the top of the perturbation does not know about the B field, because the gas can slide down along the field lines. So if the perturbation is taken slowly, to reach a new quasi-hydrostsatic balance "inside", the density inside is

$$\rho_i(\Delta z) \simeq \rho_o e^{-\Delta Z/H} \simeq \rho_o\left(1 - \frac{\Delta z}{H}\right) \tag{8.22}$$

The external density does know about magnetic pressure, however:

$$\rho_e(\Delta z) \simeq \rho_o e^{-\Delta Z/\Lambda} \simeq \rho_o\left(1 - \frac{\Delta z}{\Lambda}\right) \tag{8.23}$$

Thus, the difference between the two scale heights drives the instability in this case. If we again compare buoyancy to the restoring force of magnetic tension, we find the instability condition for this case:

$$\lambda^2 > \frac{16\Lambda^2\alpha_o}{(1+\alpha_o)^2} \tag{8.24}$$

where α_o is still the ratio of gas to magnetic pressure.

Key points

- Alfven waves, magnetosonic waves; what they are, how they work (reprise);
- Cosmic rays – basic picture, in the galactic setting;
- Resonant interaction between CR and Alfven waves;
- Buoyant instability, non-magnetized;
- Magnetic buoyancy, Parker instability.

9 Magnetic Topology: Dynamos and Reconnection

Magnetic flux freezing is a very good approximation in most astrophysical environments (and is a very handy tool). When flux freezing holds, the *topology* of the magnetic field lines remains constant.[1] We know, however, that flux freezing can be violated on small scales, where resitivity becomes important (such as a thin current sheet). It follows that the topology of the field is no longer invariant – the field lines can "break and reconnect" (well, sort of) in resistive regions.

There are two particularly important applications of resistive MHD: magnetic reconnection and dynamo theory.

9.1 Magnetic Reconnection

In these notes I'm emphasizing the details of simple, 2D reconnection. For context, note that reconnection provides one method of heating a magnetized plasma.

I just argued that magnetic topology must be preserved in the absence of resistivity. But is the converse obvious? To illustrate how resistivity can "break" and "reconnect" field lines, think about the geometry in Figure 9.1. We know resistivity is important in regions of high current density - such as the central region (around OP). If we set up this geometry and waited awhile, the central magnetic field would decay as the current layer supporting them is dissipated. This would deplete the magnetic pressure in this region. The plasma above and below this region would be pushed inwards by its own pressure, bringing a in a fresh supply of field (and plasma).

Now, look at this in more detail (I'm following Choudhuri's discussion here). The field lines ABCD and A'B'C'D' move inwards, with velocity v_{in}. Eventually the BC and B'C' parts of the field lines decay away. The AB part of the field line is moved to EO, and the A'B' part of that field line is moved to E'O. Thus, these "fragments" of two original field lines now make up one new field line, EOE'. And similarly, the parts CD and C'D' eventually make up a new field line, FPF'. Thus, "cutting and pasting" of field lines (otherwise known as reconnection) takes place in the central region. And, of course, there must be plasma flow away from the region – sideways in this cartoon – to conserve mass.

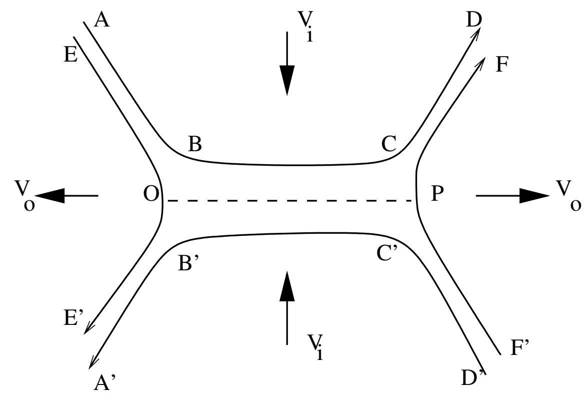

Figure 9.1 Illustrating a sinple reconneciton geometry; see text for discussion. Following Choudhuri figure 15.2.

We can be more quantitative about this geometry, and find simple scaling laws to describe this situation. The model I'm describing here is *Sweet-Parker reconnection* — it's illustrated in Figure 9.2.

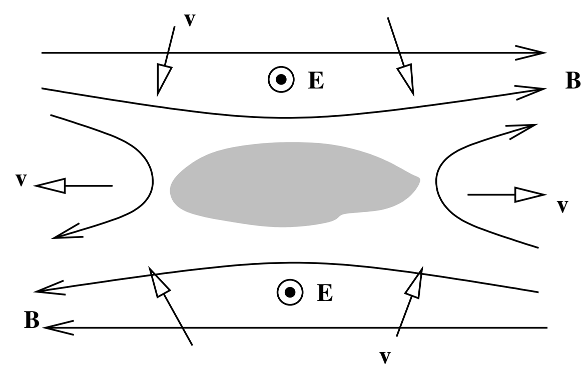

Figure 9.2 Geometry of Sweet-Parker reconnection. The current sheet (grey shaded area) has thickness l and lateral extent L The input velocity is v_{in} and the output velocity is v_{out}. Quantities far away from the current sheet are labelled with subscript o. Following Cowley figure 5.5.

First, assume the flow is incompressible – that it stays at constant density (which turns out to be a good approximation if v_{in} and v_{out} are both subsonic. Mass conservation then requires

$$v_{in} L = v_{out} l \tag{9.1}$$

Next, consider force balances. In the vertical direction, we note that $B \to 0$ at the center of the current sheet, and that $p \to p_{max}$ there (its maximum value). Pressure balance in this direction therefore requires

$$\frac{B_o^2}{8\pi} \simeq p_{max} - p_o \tag{9.2}$$

[1] Why is this? You can think about the field lines as being tied to the fluid; they move and stretch as the fluid moves, but cannot cross each other (how could they do that without "untying" themselved from the flow?)

Along and within the current sheet (call that the $\hat{\mathbf{y}}$ direction), there is no $\mathbf{v} \times \mathbf{B}$ force, so only the pressure gradient accelerates the flow. We have then,

$$\rho v_y \frac{\partial v_y}{\partial y} \simeq -\frac{\partial p}{\partial y}\,; \quad \rho \frac{v_{out}^2}{2} \simeq p_{max} - p_o \qquad (9.3)$$

Thus, the outflow speed must be

$$v_{out}^2 \simeq v_A^2 = \frac{B_o^2}{4\pi\rho} \qquad (9.4)$$

Now, what about the inflow speed v_{in}? Clearly, by mass conservation, it must be $v_{out} \simeq v_{in} L/l$; but what sets l? We can try two different arguments here.

1. First, go back to the induction equation (5.11). In a steady state, it is $\nabla \times (\mathbf{v} \times \mathbf{B}) + \eta \nabla^2 \mathbf{B} = 0$, which gives (by dimensional/scaling analysis)n

$$\frac{v_{in} B}{l} \simeq \frac{\eta B}{l^2}\,; \quad v_{in} \simeq \frac{\eta}{l} \qquad (9.5)$$

This tells us that we can keep the diffusion region steady if the rate at which flux is brought in is equal to the rate at which it is annihilated.

2. Alternatively, note that there must be an $\mathbf{E}$ field "out of the page", as shown in figure 9.2, to maintain the current sheet. From Maxwell, we get

$$\mathbf{j} = \sigma \mathbf{E}\,; \ \nabla \times \mathbf{B} = \frac{4\pi}{c}\mathbf{j} \ \Leftrightarrow \ \frac{B}{l} \simeq \frac{4\pi}{c}\sigma E \qquad (9.6)$$

where σ is the electrical conductivity. But now, the incoming plasma charges must $\mathbf{E} \times \mathbf{B}$ drift[2] into the region: thus

$$v_{in} = c\frac{E}{B} \simeq \frac{c^2}{4\pi\sigma}\frac{1}{l} = \frac{\eta}{l} \qquad (9.7)$$

(in that last step I've used the definition of the magnetic resistivitiy, $\eta = c^2/4\pi\sigma$, from chapter 5). And look: this agrees with (9.5)!

OK: now, combine our answer (9.5 or 9.7) with (9.1), and we get our result, the inflow velocity and thickness of the dissipation layer:

$$v_{in}^2 \simeq \frac{v_A \eta}{L}\,; \qquad l^2 \simeq \frac{\eta L}{v_A} \qquad (9.8)$$

This gives us, finally, the *spontaneous reconnection rate*; the rate of slow inflow that allows things to go steadily. This is indeed slow – in many situations (for instance solar flares), the plasma conductivity is high (given by the Coulomb collision value), so that η is low; and the reconnection timescale ($\simeq L/v_{in}$) is much to long to explain the observations. People have, therfore, spent a lot of time trying to invent faster versions of this model. Some are as follows ..

• **Petschek reconnection** has gotten a lot of attention in the literature. The idea was that internal structures in the flow (shocks toward the edges of the current sheet) would affect the velocity field and narrow the width of the sheet, to $L \sim l$. This would of course make the inflow rate nearly independent of η. The last I heard is that lab experiements (*cf.*review by Kulsrud, 1998) were not confirming this model.

• **Compressible flow**. Another suggestion is that the plasma in the dissipation region (DR) is compressed. If this is the case, then (9.1) is replaced by $v_{in}\rho_o L \simeq v_{out}\rho_{DR} v_{out}$. This will increase v_{in} relative to v_{out} – its a plausible idea, but I'm not aware of any lab tests yet.

• **Anomalous diffusion**. This is my personal favorite, and is a good example of how plasma microphysics can be important in macroscopic situations. That is: is the usual value of η, based on Coulomb collisions in the plasma, the right value? The answer is almost surely that it is not. If the current density is high enough, we have a "two-stream"-like situation in the current sheet, and can expect plasma turbulence to be generated.

But can this be quantified? One commonly used estimate is as follows. Recall the microscopic origin of plasma conductivity, from chapter 3:

$$\sigma = \frac{j}{E} =\simeq \frac{ne^2}{m}\tau_{coll}$$

But now, what is τ_{coll} when we're working with plasma turbulence? A common guesstimate, which is thought to be an upper limit to the resistivity (lower limit to τ_{coll}) is to set $\tau_{coll} = 2\pi/\omega_p$, if $\omega_p = \left(4\pi n e^2/m\right)^{1/2}$ is the usual plasma frequency. This gives an estimate of the *anomalous resistivity*.

9.2 Reconnection: other approaches

Reconnection is a very active field these days; the simple 2D models we've just seen are probably way too simple. I'll just store a few comments here, thinking about what interests me most.

[2] Check back in Chapter 2: $\mathbf{v}_E = c(\mathbf{E} \times \mathbf{B})/B^2$, from (2.17).

9.2.1 Non-steady reconnection

Here is my personal impression: who says steady-state reconnection is relevant to any natural situation? That is: the arguments above show that steady reconnection models must be forced, sometimes rather severely, to connect with what we think is occuring in nature. But examples of non-steady reconnection events are easy to find:

• Reconnection in solar flares. Flares are seriously transient events; the large amount of energy released is believed to be due to very fast annihilation of magnetic field, in a reconnection event. Flares have motivated much of the work on steady-state models; however as an outsider I suspect time-dependent, patchy, localized reconnection must be taking place.

• Reconnection at the magnetopause, where the solar wind hits the earth's magnetic field. Spacecraft observations suggest this is very patchy, localized and time-dependent. Picture, for instance, magnetic flux tubes being carried along in the solar wind; and let one of them impact the magnetopause, which also has its field bunched into ropes. This process can allow solar wind plasma, and field, to penetrate into the magnetosphere.

9.2.2 Driven reconnection

It's worth remembering that the 2D models above are all "spontaneous": put two misaligned $\mathbf{B}$ fields together and wait to see how quickly they reconnect. But nature does not always work this way. One can envision a situation in which the two anti-parallel magnetic structures are *driven together*, by large-scale flows in the system (one example of this is MHD turbulence, in which different parts of the plasma move in random directions, at a speed set by the energetics of the turbulence). In this case one (this one at least) expects that the inflow speed (v_{in}) in our notation above) will be set by the large-scale flow (say the turbulent speed). How can the reconnection site adjust to this? Some authors (*e.g.* Parker) suggest that the internal structure of the reconnection layer – its thickness, density or resistivity (set by microscale turbulence therein) – will adjust as necessary.

9.2.3 Three-dimensional reconnection

Another observation: only rarely can a reconnection event be well described by a two-dimensional analysis. My last example in fact assumed this – because the intersection of two magnetic flux ropes is clearly a three-dimensional process. Going to 3D is challenging, and work is only starting here (helped significantly by numerical simulations).

9.3 MHD Dynamos

Where do magnetic fields come from? In the lab, the answer is easy: "currents". In magnetic solids, the currents are those of well-ordered electrons spins in ferromagnetism. More typically, currents in the lab — and their consequent $\mathbf{B}$ fields — come from obvious things like batteries and wires. The issue is then, what drives the currents? My dictionary defines a dynamo as

> "a device for converting mechanical energy into electrical energy, usually by expending the mechanical energy in producing a periodic motion of a conductor and a surrounding magnetic field".

A simple lab version of this is called the *unipolar dynamo*, in Figure 9.3. This involves a conducting disk, threaded by a $\mathbf{B}$ field, which rotates about its axis. This induces a radial $\mathbf{E}$ field, $\mathbf{v} \times \mathbf{B}/c$, and thus a potential drop between the axis and the edge of the disk. If you hook up wires in the right way you'll have a current – and this current will create its own magnetic field.

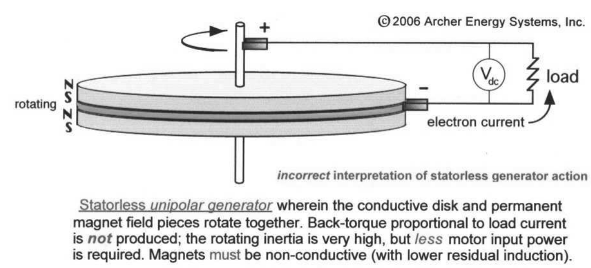

Figure 9.3 A simple unipolar dynamo (in a less than simple figure from www.stardrivedevice.com, the best figure I could find). The conducting disk moves through an (externally supported) $\mathbf{B}$ field as it rotates about its axis. The resultant EMF supports a potential drop between the axis and edge of the disk — which can drive a current.

What about astrophysical magnetic fields? To be specific, what is the origin of the earth's field, or the sun's field? It's easy to think of what doesn't work. One, even solid planets like the earth can't be ferromagnetic (because the core temperature is well above the Curie temperature at which permanent magnetism disappears); and clearly stars and galaxies can't be ferromagnetic at all. Two, we can't assume the fields are primordial — were somehow created when the sun/earth/galaxy formed — because we know the resistivity of the plasmas in question, and thus we know how long it would take a primordial current to dissipate. Such calculations predict that primordial fields would long ago have died away; but we know that stars, planets, and galaxies are still magnetized.[3]

Thus, we still must ask, "What supports astrophysical $\mathbf{B}$ fields?" The answer is still currents, but what drives astrophysical currents? We can't expect a device such as in Figure 9.3 exists inside a planet, or star, or whatnot ... so we need to find a way to drive *fluid motions* which can maintain the $\mathbf{B}$ fields we observe. This question gets us into what's called *dynamo theory*. To approach this, go back (yet again) to the induction equation,

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) + \eta \nabla^2 \mathbf{B} \tag{9.9}$$

We know the second term describes resistive decay; if we're lucky the first term can be a *growth* term. The first term describes magnetic induction, when $\mathbf{v} \times \mathbf{B}$ creates a local $\mathbf{E}$ field, and thus an EMF, which can drive a current. If the geometry is right, this current can make the initial (seed) magnetic field grow — giving us a *dynamo*. But the devil is in the details – how can the right flow field be created and maintained naturally?

9.3.1 Cowling's theorem

We can start by seeing what *won't* work. That is, most astrophysical models assume simple, symmetric geometries; but these can't support a dynamo.

To be specific, we need to prove Cowling's theorem: *it is not possible to maintain a steady dynamo in an axisymmetric system*. To do this, I follow Cowling's original (1934) argument, as presented by Choudhuri. Start by assuming we do have an axisymmetric dynamo: one with $\partial/\partial t = \partial/\partial\phi = 0$. Consider a plane through the symmetry axis: the projections of the field lines on this plane must be closed curves (think of a simple magnetic dipole). There will be at least one neutral point in this plane (a point where the closed field lines center) – and j_ϕ must be non-zero here, while $\mathbf{B}$ has only a ϕ component at this point. Take a line integral of Ohm's

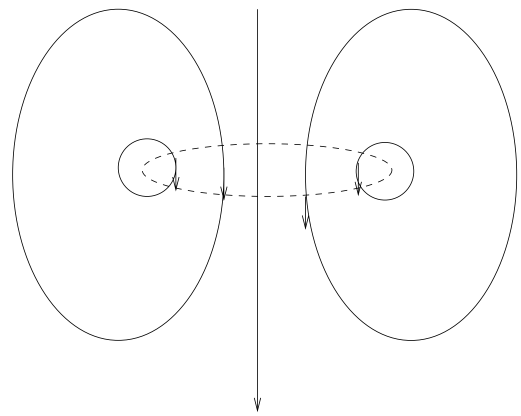

Figure 9.4 Illustration of geometry for Cowling's theorem. Following Choudhuri Figure 16.3.

law (e.g. equation 5.10) along a closed loop through these neutral points, enclosing the symmetry axis:

$$\frac{1}{\sigma} \oint j_\phi dl = \oint \mathbf{E} \cdot d\mathbf{l} + \oint \mathbf{v} \times \mathbf{B} \cdot d\mathbf{l} \tag{9.10}$$

But now: the second term vanishes, because $\mathbf{B} \parallel d\mathbf{l}$ if this loop goes entirely through neutral points. The first term vanishes, because

$$\oint \mathbf{E} \cdot d\mathbf{l} = \int \nabla \times \mathbf{E} \cdot d\mathbf{S} = -\int \frac{\partial \mathbf{B}}{\partial t} \cdot d\mathbf{S}$$

and this last is zero by our steady-state assumption. However, the LHS of (9.10) is non-zero, as j_ϕ is finite. Thus, we have a contradiction; and Cowling's theorem is proved.

It follows, then, that we must relax the assumption of axisymmetry; and yet we want to maintain the large-scale axisymmetry which we know describes objects like the sun, the earth, or the galaxy. The answer is to introduce small-scale asymmetries — best done, astrophysically, with small-scale disordered fluid motion, such as convection or turbulence.

9.3.2 Parker's solar dynamo

The classic dynamo model is due to Parker (1955), and is meant to describe the solar magnetic field. It is best presented qualitatively — refer to Figure 9.5 for the cartoon.

Say the solar field starts mainly dipolar (this is roughly consistent with observations of the global field, just above the solar surface). The sun does not rotate as a solid body; near the surface, the equator rotates faster than the polar regions. This will stretch our dipolar

[3] In addition, we know that the sun's field reverses pretty regularly, every 11 years or so; and the earth's field reverses less regularly, every $10^4 - 10^5$ years. This clearly requires some internal, self-governing mechanism.

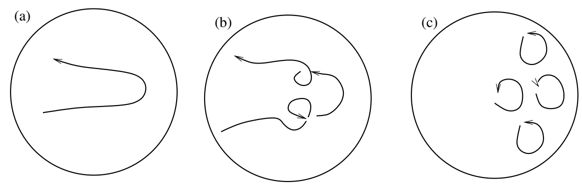

Figure 9.5 Parker's model of the solar dynamo, at the cartoon level. (a) Differential rotation (the sun's equator rotates faster than the poles) stretches initially dipolar field in the toroidal direction. (b) Coriolis forces acting on surface convective cells generates local poloidal fields. (c) The opposite sense of the Coriolis force in the north and south hemispheres, combined with the opposite sense of the initially toroidal field, results in a strong net poloidal field, rather than a randomly directed set of field loops. (These loops are shown projected in the meridional plane.) Following Choudhuri, Figures 16.4.

field, generating toroidal components. Thus, it is no problem to generate toroidal field if the body has differential rotation. But this cannot be all of the story. Such a stretched toroidal field will have many local field-line reversals, and if nothing else happens it will simply decay away due to resistive dissipation.

However, the upper layers of the sun are convectively unstable. In this region, plasma blobs rise and fall.[4] Now, these vertically moving blobs are subject to a Coriolis force, due to the sun's overall rotation. The blobs therefore rotate as they rise; they act like little cyclones, and formally we say that their their motion has a net *helicity* (that means the small-scale motions do not have mirror symmetry: for instance a flow with $\mathbf{v} \cdot (\nabla \times \mathbf{v}) \neq 0$ is helical). Look at (b) of Figure 9.5: this cyclonic motion twists the magnetic field back into poloidal loops. Remember that both the direction of B_ϕ and of the Coriolis rotation are opposite in the north and south hemispheres: this means the direction of the poloidal field component generated is the *same* in the two hemispheres. We therefore have a fully working dynamo: poloidal fields are generated by the helical convective (turbulent) motions, while toroidal fields are generated by differential rotation. The whole system must be stabilized by dissipation – that is resistivity will keep each field component from getting too large.

[4] In chapter 8 we talked about (in)stability to buoyancy – an unstable atmosphere will develop strong convection.

9.3.3 Scale separation and turbulent dynamos

We argued "by cartoon" that helical, convective motions on the sun (or the earth) can maintain the large-scale $\mathbf{B}$ field. That is, we're arguing that *small-scale* turbulent motions can add up to a net *large-scale* dynamo.

To get a sense of how this works, and what's needed to make it work, we need to be a bit formal. Split the velocity and magnetic fields into mean and fluctuating parts:

$$\mathbf{B} = \mathbf{B} + \mathbf{b}\,; \qquad \mathbf{v} = \mathbf{V} + \mathbf{v} \tag{9.11}$$

where we're assuming that $\mathbf{b}$ and $\mathbf{v}$ have zero mean,, and also that they are small-scale – that they vary over much smaller spatial scales than $\mathbf{V}$ and $\mathbf{B}$ do. What effect do they have on the induction equation (5.11, also 9.9)? Let's split it into large-scale (mean) and small-scale (fluctuating) parts. For the mean field, we get

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{V} \times \mathbf{B}) - \nabla \times \boldsymbol{\varepsilon} + \eta \nabla^2 \mathbf{B} \tag{9.12}$$

where the important new term is

$$\boldsymbol{\varepsilon} = -\langle \mathbf{v} \times \mathbf{b} \rangle \tag{9.13}$$

This describes the net EMF due to the fluctuating $\mathbf{v}$ and $\mathbf{b}$.

But now, we must ask whether $\boldsymbol{\varepsilon}$ has any interesting large-scale effect. If $\mathbf{v}$ and $\mathbf{b}$ are rapidly varying, have zero mean, and uncorrelated, we'd expect the mean of their product to be zero. It turns out ("can be shown") that things are interesting (non-zero) if the turbulence satisfies two conditions: (i) it must be helical, satisfying $\mathbf{v} \cdot \nabla \times \mathbf{b} \neq 0$; and (ii) it must be resistive; $\eta \neq 0$. If both of these conditions are met, the turbulent EMF, $\boldsymbol{\varepsilon}$, will be non-zero *on large scales*. In particular, it may be the case that $\langle \mathbf{v} \times \mathbf{b} \rangle = \alpha \mathbf{B}$ (*i.e.*, that $\boldsymbol{\varepsilon}$ has a component along $\mathbf{B}$). If this is so, then we have an effective dynamo term:

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\alpha \mathbf{B}) + (\text{other stuff}) \tag{9.14}$$

If this works – if $\boldsymbol{\varepsilon} = \alpha \mathbf{B}$ — then we can see two useful astrophysical consequences. One is balancing ohmic losses, as in the sun or the earth — and (in principle) accounting for the occasional field reversals in each body. The second is "growing" the $\mathbf{B}$ field in the first place. To see this, note that (9.14) allows solutions $\mathbf{B} \propto e^{\alpha t}$, if α is constant in time. That's a growing $\mathbf{B}$ field, with growth time $\sim L/\alpha$ (some large-scale length scale L). We might expect that a small seed field would grow exponentially until some other physics (dissipation? back reaction on the driving fluid?) comes into play.

9.3.4 Astrophysical dynamos in the lab

Finally, a few words about trying to do this in the lab. Everything above is still pure theory — it would be good to verify directly that an $\alpha\omega$ dynamo (rotation plus turbulence, as in the sun), or an α^2 dynamo (pure turbulence) can really make a large-scale ordered $\mathbf{B}$ field. Several groups are working on this, including NMT's very own Stirling Colgate. The experiments use liquid metal — usually liquid sodium — in some sort of rotating system (the ω in an $\alpha\omega$ dynamo), and try various ways to induce turbulence (the α) in the flow. The last I heard, no one had sucessfully made their dynamo work — but I think the field's progressing. Check the Feb 2006 issue of *Physics Today* if you'd like more details.

Key points

• Reconnection – simple 2D model; "what is a reconnection rate?"

• Reconnection – extensions of the 2D model (anomalous effects, non-steady, driven, etc).

• Dynamos – what they are; what they need (helicity). Parker's model for the sun.

• Dynamos — why turbulence matters; $\alpha\omega$ and α^2.

10 Accretion in astrophysics I: star formation

We argued in chapter 6 that accretion flows are common. They are found in many places – from Young Stellar Objects (YSO's), to compact objects (black holes, neutron stars) in galactic binary systems, to massive black holes in Active Galactic Nuclei (AGN). They are potentially important to astrophysics in all settings. In this chapter we'll look at star formation and YSO's, with an eye to the role of accretion in the process. In the next two chapters we'll return to "classical" accretion, as it applies to compact objects and standard accretion disks.

10.1 Star formation, recall the basics

The ISM is the source of new stars. We think these new stars form in some gravitational collapse process . . . and that star formation should be an ongoing process, with each subsequent generation of stars having a somewhat richer heavy element content, due to nuclear processing by previous generations of stars.

Think back to our general discussion of gravitational collapse and star formation (chapters 4, 5). We noted that a piece of the ISM will be gravitationally unstable if "gravity wins"; that is, if its gravitational potential energy exceeds its internal energy. Written in terms of initial density and temperature, this becomes a condition on the size, or mass, of the perturbation. If the initial perturbation has $R > R_J$ (the *Jeans length*), or $M > M_J$, (the *Jeans mass*), where

$$\frac{4\pi}{3} G R_J^2 \rho \simeq \frac{k_B T}{m} \; ; \; M_J \simeq \left(\frac{k_B T}{mG}\right)^{3/2} \left(\frac{3}{4\pi\rho}\right)^{1/2} \tag{10.1}$$

then it is gravitationally unstable.

We also noted, in chapter 5, that two classic problems – conservation of magnetic flux and of angular momentum – make the collapse process more complicated than simple "gravity wins". We suggested that these two problems are probably interrelated – for instance torques exerted by the B field may contribute to slowing down the rotation and transferring angular momentum to the surroundings. Another possibility is that the collapsing cloud is mostly neutral; because B field lines are only tied to the ionized fraction of the cloud, it may be that the neutral gas can "slip through" the ionized fraction. This process goes by the fancy name, "ambipolar diffusion". Both ideas are probably part of the truth. The picture still needs to be expanded, however, and put in the context of star formation research, which is one of the most active areas in current astrophysics. That's where we're going in this chapter.

10.2 Molecular Clouds as Precursors

We know that young stars are gregarious. They do not form singly; rather, they form in clusters, inside dense, molecular clouds (MC's) in the ISM. We therefore need to consider the physical state of the MC's, and how stars — as well as planets and possibly life — form within them.

10.2.1 Observational constraints

Most of what we know about MCs comes directly from observations.

• **Physical conditions**. MC's are found in a range of sizes, from $\sim 10^2\ M_\odot$ to $\sim 10^6\ M_\odot$, with an approximate $\propto M_{mc}^{-3/2}$ number distribution. Interestingly, there seems to be a correlation between cloud mass, cloud distribution in the galaxy, and cloud temperature. There seem to be two populations of MC's: the small clouds (SMC's), which have internal temperatures $\sim$ 10K, and the "giant" clouds (GMC's), which have internal temperatures $\gtrsim$ 20 K. The SMC temperature is consistent with heating by the background cosmic ray population in the galaxy; they do not seem to have any internal energy sources (such as hot stars). This correlates with the absence of young, hot (O and B) stars in these clouds. The SMC population is distributed throughout the galaxy, with no particular preference for spiral arms. The GMC's, on the other hand, require internal energy sources, and we do observe young O and B stars forming in these clouds. Their higher temperature is thought to be due to heating by these massive protostars. GMC's are not distributed uniformly in the galaxy, but are found in spiral arms (they provide the young, bright stars by which we see pretty spiral arms in external galaxies).

• **Lifetimes**. We know that MC's are self-gravitating: their pressure generally exceeds the typical ISM pressure, so they cannot be in pressure balance with the ISM. We also know their mass exceeds the Jeans' mass for their temperatures and densitites. We would expect them to be in a state of gravitational collapse. If this were the case, we can use the free-fall time (check back to equation 4.18) to guesstimate how how quickly they should collapse. We find $t_{ff} \sim 10^7$ years for typical mean MC densities, $n \sim 100\text{cm}^{-3}$. Combining this with the total mass in MC's in our galaxy, $\sim 5 \times 10^9 M_\odot$, we can estimate the current star formation rate that this simple picture (free-fall collapse of all MC's) would predict: we get a rate $\sim 500\ M_\odot$/yr. This prediction is much larger than the observed rate, $\sim$ a few $M_\odot$/yr. Thus, only about one percent of the clouds we know about can be collapsing; the rest must be supported by gravity somehow, and have a lifetime ~ 100 times longer than this prediction.

• **Turbulence**. This support against gravity almost certainly is provided by internal random motions in the clouds, which are generically called "turbulence". We can measure internal Δv's, from linewidths; we find $\Delta v \gg v_{th} \simeq (k_B T/m)^{1/2}$. Thus, the internal motions are highly supersonic. Further, the linewidth does not come from rotation (which could be detected), nor does is come from free-fall collapse (from the argument above). Thus, it must be from random internal motions – the clouds must contain subclumps, or waves, which move through the "cloud" at supersonic velocities, $v_{ran} \sim \Delta v$. These random velocities do appear able to provide the virial support: the correlation $\Delta v^2 \propto nR^2 \propto M_{mc}/R$ is observed, consistent with virial balance; the numbers also work out, to have $v_{ran}^2 \simeq GM_{mc}/R$.

MC's are also magnetized. We know this from Zeeman splitting of spectral lines, also from polarization of starlight, which comes from dust grain alignment with the magnetic field (the grains are dielectrics). The fields can, in principle, also help support the cloud against collapse. In practice, the fields and turbulence are probably intimately mixed (gas flows stretch and twist field lines; the field fights back and limits turbulent velocities). We also know that turbulent (and thus magnetic) decay times are very likely short compared to the MC lifetimes. Thus "turbulence" or "B fields" aren't the full answer; something inside the cloud must drive the turbulence. Current thinking has the driver being newly formed YSOs within the cloud – their winds and jets may dump enough energy back into the MC to stabilize it.

10.2.2 How do they fragment?

We know that the Jeams mass in a MC is much larger than the mass of a single star (even a big one). We also know that stars form in large numbers within MCs. It follows that the gravitational collapse process within the MC must involve fragmentation into a large number of star-sized structures. We don't know how this happens, but can find a couple of hints in the data.

•**The Jeans mass depends on temperature and density**, as $M_J \propto T^{3/2}/\rho^{1/2}$. Thus, as a large cloud collapses, and its density increases, M_J will drop *if* the cloud stays cool. This should allow smaller and smaller subclumps to become unstable, and fragment out of the larger-scale collapse (recalling $t_{ff} \propto 1/\rho^{1/2}$, so that denser subclumps fragment faster). This process should continue as long as the collapsing cloud can stay cool. People tend to argue that the cloud will stay cool as long as it stays transparent, and can radiate effectively. Once it goes opaque (noting that opacity $\propto \rho R \propto M/R^2$ for a cloud of fixed mass[1]), it may have trouble radiating away the energy generated by the gravitational collapse. If this happens, T will increase, and M_J will increase; this will provide a smallest fragment size. This process is called "opacity-limited fragmentation".

• **The Initial Mass Function** (IMF), which describes the mass distribution of newly formed stars, can be determined from the mass distribution of stars that are currently around. If a star has a visible lifetime $\tau(m)$ (which is a function of the mass m of the star), and if stars are created in the galaxy at a rate $S(m,t)$, the current distribution of stars $N(m,t)$ is given by

$$\frac{dN(m,t)}{dt} = S(m,t) - \frac{N(m,t)}{\tau(m)} \tag{10.2}$$

If $S(m,t)$ is not varying in time, this has the simple solution

$$N(m,t) = S(m)\left[1 - e^{-t/\tau(m)}\right] \tag{10.3}$$

which allows us to relate $N(m,t)$ (which can, in principle, be observed) to $S(m)$, the *initial mass function* (which we want to know).

In general $S(m)$ is expected to be a decreasing function of m, giving many low-mass stars and few high-mass stars. It is determined by the complex interplay between fragmentation, coalescence, accretion, and outflow, so at present it is quite poorly understood but is held to be one of the holy grails of astrophysics. From the observational side it seems that at high masses ($m \gtrsim M_\odot$), $S(m)$ can be approximated by a power law. The form $S(m) \propto m^{-2.35}$ is called the *Salpeter initial mass function*, and has

[1] We'll see what this means next term

been around for a long time. Below $m \sim M_\odot$, it gets harder to determine $N(m,t)$ accurately. It is clear that $S(m)$ flattens out; whether or not if turns down below $m \sim 0.1 - 0.3\ M_\odot$ is less clear. In addition, several people have suggested that the locally measured IMF being the superposition of two IMF's, resulting from two types or modes of star formation. The suggestion is that low-mass stars (compared to $\sim 1 M_\odot$) form in one region – probably the SMC's – and that high-mass stars (above $\sim 1 M_\odot$) form elsewhere – probably the GMC's – and that different physics governs the two modes. This is called *bimodal star formation*.

The range of allowed stellar masses is, on the other hand, fairly well determined by the physics of the stars themselves. At the high-mass end, stars with $M \gtrsim 60 M_\odot$ are unstable; their own luminosity generates too much radiation pressure for the star to be able to attain hydrostatic equilibrium from its own gravity. At the low-mass end, $0.1\ M_\odot$ is about the lowest mass object which can sustain hydrogen fusion. Objects with masses lower than that certainly do exist but are not called stars. If they can fuse deuterium they're called brown dwarfs, and if they are too low mass for that they're planets.

10.3 Young Stellar Objects: how do they evolve?

Somehow or other, a star-sized fragment of a MC separates out and collapses to form the star. Current thinking — based on an impressive amount of new data, as well as theory — identifies four stages of this process[2] (as illustrated in Figure 10.1). A caveat: what follows is thought to describe the formation process for *low-mass* stars only.

(1) Initial Contraction and core collapse. When the protostar first detaches from its MC environment and starts to collapse, it's probably still nearly spherical, and collapsing only slowly. Refer back to chapter 4, where we discussed self-gravitating isothermal spheres. These are characterized by a core of size $r_o \propto (T/\rho_o)^{1/2}$; they are probably a good approximation to the initial state of the protostar. As the cloud slowly gets denser (and/or cools), the core gets smaller; so we expect an inside-out collapse (the core collapses most rapidly, the outer layers follow later). Much of the MC magnetic field must be lost in this stage, possibly by ambipolar diffusion. Note the protostellar core is well hidden at this stage, being shrouded by its dusty outer envelope.

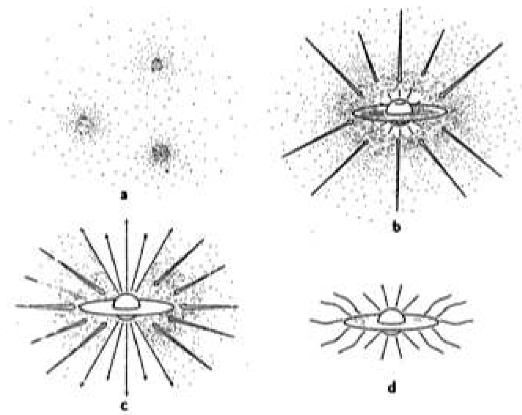

Figure 10.1 Illustrating the four likely stages of low-mass star formation; from Shu etal, *Ann.Rev.Ast.Ap*, 1987. See text for details. Note, this cartoon needs to be updated, to include outflows, which we now know are an important part of the process.

(2) Disk formation and outflows. As the inside-out collapse proceeds, the inner parts settle into what will become the core of the protostar. Material continues to accrete from the surroundings, but now — due to angular momentum — it settles into a disk around the core. That was expected; but here's where the surprise came. People did not expect gravitational collapse also to involve outflow (it is the wrong direction), but it does. Observations show that just about every protostar has *bipolar outflows*, which can be quite broad, not well collimated. They are usually seen in molecular lines. Some YSOs also produce well-collimated *jets*, usually seen in the optical. The optically bright "nebulae" called *Herbig-Haro objects*, which used to seem only a curiosity, are now known to be associated with shocks in these YSO jets.

In addition, some YSO's are found to have a *disk* structure, oriented perpendicular to the jet or bipolar outflow. These are presumably the accretion disks — or at least the rotation-supported outer regions of the "reservoir" from which material accretes to form the star. Some authors estimate that $\sim 1/3$ of the infalling matter goes out again in the outflow/jet, and $\sim 2/3$ makes it onto the protostar. Because there is still a lot of (dusty) reservoir material, YSOs in this stage are mostly shrouded objects, detectable only in radio/molecular lines; one author talks about "deeply

[2] Jargon warning: many authors break YSOs up into Class 0, Class I, Class II, Class III objects. These Classes correlate decently with the four phases I list here, but are more observationally based — so I won't describe them in detail here.

buried protostars with their infalling envelopes and associated bipolar outflows".

(4) Post-outflow, pre-Main Sequence star. Eventually, the outflow slows, as does the general infall of matter. At some point the YSO becomes visible in optical and IR, as the surrounding dust/gas is cleared away. When nuclear burning starts, these young stars[3] appear on the HR diagram. They still have disks and outflows; T Tauri stars are examples here (several of them have detectable disks). Finally, the outflows reduce to normal stellar-wind level, and the disk either is all accreted or (mostly) dissipates. Remnants of these circumstellar disks, presumably, become planetary systems.

Caveat 1: magnetic activity. It seems that YSOs, at least in the PMS/T Tauri stage, have unusually high levels of magnetic activity (think of solar flares, eruptive prominences, but at a much stronger level). They can show strong X-ray emission (which requires plasma much hotter than the conventional picture, above, would predict), and also nonthermal radio emission (which requires relativistic electrons and a magnetic field). This interesting area is just starting to be explored ... stay tuned.

Caveat 2: what about high mass stars? High-mass star formation seems still to be less well understood; the problem is that the strong radiation pressure after nuclear burning starts disrupts the inflow/accretion process ... and people do not seem sure of what happens at that point. Once again, stay tuned.

Key points

• Basic gravitational (in)stability criteria;

• Isothermal spheres: what they are, what characterizes the solution;

• Molecular clouds: their general nature, how they relate to the YSO's which they contain;

• YSO's: what we observe, and what we think their evolutionary stages are.

[3] alternatively, pre-main-sequence, "PMS", star

11 Accretion II: compact objects

Now, let's get back to accretion onto compact objects – dense stellar remnants and black holes.

11.1 Basic ideas

Before we talk about specific objects, let's start with an overview of important physical ideas.

11.1.1 Energetics ("Accretion Power")

Recall what we discussed briefly in chapter 6. The basic idea is that accreting matter liberates its gravitational energy, at a rate $\dot{E}_g \sim GM\dot{M}/r$, for some mass accretion rate $\dot{M}$ and some "conversion radius" r. Physically, this energy must be converted to internal energy of the accreting stuff (*via* friction, for instance) and the hot matter then radiates with some (as yet unspecified) efficiency ε:

$$L = \varepsilon \frac{GM\dot{M}}{r} \tag{11.1}$$

This applies to general accretion flows, be they onto YSO's or neutron stars.

Another approach is sometimes used for accretion onto black holes. Let's think about the conversion radius r. The smallest it's likely to be is some factor times the "gravitational radius" around a black hole, $r_g = GM/c^2$.[1] If we scale r in (11.1) to r_g, we have

$$L = \varepsilon \frac{GM\dot{M}}{r_g}\frac{r_g}{r} = \varepsilon \frac{r_g}{r}\dot{M}c^2 \tag{11.2}$$

Thus, for conversion close to r_g, ε measures the output luminosity as a fraction of the infalling rest mass energy; this scaling is often used in black hole accretion models.

What are typical numbers? A typical galactic X-ray binary might have $L \sim 10^{37}$ erg/s; this requires $\varepsilon\dot{M} \sim 3.5 \times 10^{-7} M_\odot$/yr, for conversion close to r_g. Or, a bright quasar might have $L \sim 10^{46}$erg/s; this requires $\varepsilon\dot{M} \sim 0.35 M_\odot$/yr.

11.1.2 Eddington luminosity

This is an important reference point: at what luminosity can the radiation pressure from a central source of luminosity L can offset gravity from a central mass M, and stop the flow? For most accretion problems, the infalling matter is fully ionized hydrogen. That means we must consider the gravitational force on a proton, and note that the radiation pressure is communicated *via* Thomson (electron-photon) scattering (which has a cross section $\sigma_T = 6.65 \times 10^{-25}\mathrm{cm}^2$).[2] With this we can derive the Eddington luminosity:

$$\frac{L_{edd}}{4\pi r^2}\frac{\sigma_T}{c} = \frac{GMm_p}{r^2} \;\Rightarrow\; L_{edd} = \frac{4\pi GMm_p c}{\sigma_T} \tag{11.3}$$

and numerically, this is $L_{edd} \simeq 1.3 \times 10^{38}(M/M_\odot)$ erg/s.

A related quantity is the *Eddington mass flux*. This is the mass flow that produces L_{edd} in a given system. Defining

$$L_{edd} = \varepsilon \frac{GM\dot{M}_{edd}}{r}$$

gives us

$$\dot{M}_{edd} = \frac{4\pi}{\varepsilon}\frac{r}{r_g}\frac{GMm_p}{c\sigma_T} \tag{11.4}$$

11.1.3 Thermal state

This is critical to interpreting observations of accretion systems. While the full story here can be quite complex,[3] one useful simple estimate is possible. If the infalling matter *and radiation it generates* are in something close to thermal equilibrium, we know from basic thermodynamics that it radiates as a black body. That means the luminosity coming from its surface has an intensity (energy per time per surface area) $\sigma_{SB}T^4$, where $\sigma_{SB} = 5.67 \times 10^{-5}$ (cgs) is the Stefan-Boltzmann constant. We also know that the typical photon energy $h\nu \sim kT$. Equating the luminosity in (11.1) to that lost by black body radiation determines the lowest temperature the gas will reach. You will find that $T \sim 10^7$ K for accretion onto a solar-sized black hole (and thus we have galactic X-ray binaries); and $T \sim 10^4 - 10^5$ K (and thus we have UV and optical sources) for accretion onto a massive $M \sim 10^9\ M_\odot$ black hole (as in a galactic nucleus).

We must remember, however, that the situation in an accretion flow is rarely that simple. One complication

[1] Think about the event horizon of a black hole, or the smallest stable circular orbit – both of these are $\propto r_g$. We'll return to this later in this chapter.

[2] Other physical states for the accreting matter call for other values of σ and m.

[3] We would need to know how rapidly the plasma is heated by friction, turbulence, magnetic dissipation, etc; and how effectively it can lose energy by radiation – which depends on all sorts of details about the plasma density, temperature, etc..

is that the radiation emitted by the infalling matter may well not be in thermal equilibrium with the matter; that will be the case if the flow is transparent to the radiation, meaning that the mean free path for a photon to be absorbed or scattered is large compared to the size of the system. Just how the radiation comes out will be a major topic for next term. For now, I'll just note that the inflowing plasma will be *hotter*, and transparent (parts of) accretion flows can reach temperatures $\sim 10^8 - 10^{10}$ K. Another complication is that some of the accretion energy does not simply to go heating, but (somehow) goes to accelerating a few particles to relativistic energies. We know this because we know that many accretion sources are "nonthermal", with significant emission by these relativistic particles due to their gyromotion in the local magnetic field.

11.1.4 The transtion to disk accretion

We've worked so far with spherical accretion – because it's simple, and because it's probably a useful limiting case (for accretion from large distances in a quasi-symmetric system). But think about conservation of angular momentum. If the quasi-spherical inflow has any net angular momentum at all, its angular velocity will increase as the matter flows inward; eventually we can no longer assume spherical inflow, and rotation will dominate *perpendicular to the rotation axis*. However, parallel to the rotation axis the collapse can continue – and in most systems will be helped along by radiative cooling (as the collapsing matter loses its internal energy and pressure support against gravity). Thus we expect an inflow that is initially spherical – say at large distances from its core – to change to disk-like accretion closer to the core.

From there, the matter must lose angular momentum in order to keep accreting. This is thought to come from viscosity in quasi-steady flows (as we'll see in chapter 12), or possibly from instabilities in unsteady flows (*i.e.*, the flow might fragment into blobs, some of which move inward).

11.1.5 Size Matters

The size of the accretion region is critical to both the energetics and the thermal state of the accreting matter. That means it's critical to both the luminosity and the spectrum of the accretion flow. For quick reference, the interesting sizes are:

• The *YSO radius* is probably comparable to the radius the YSO will have when it reaches the main sequence. If the YSO is strongly magnetized, however, magnetic pressure may stop the accretion flow at a significantly larger radius.

• The *neutron star radius* is about 10 km. Most of the action for accretion onto a neutron star probably occurs close to this radius.

• *Gravitational radius* of a star-sized black hole is $r_g \simeq 1.5(M/M_\odot)$ km. Most of the action for accretion onto a black hole probably occurs from a few to a few tens of r_g.

• *Gravitational radius* of a galaxy-sized black hole is $r_g \simeq 1.5 \times 10^{14}(M/10^9 M_\odot)$ cm $\simeq$ 10 AU for a $10^9 M_\odot$ black hole — the radius of Saturn's orbit. So most of the action for a massive black hole in a galactic nucleus occurs well below 1 pc.

11.1.6 Jets and outflows

Finally, we should note that just about every accretion **inflow** involves an **outflow**. This comes from the data: **jets** (and sometimes less collimated outflows as well) are found in connection with just about every accretion disk, everywhere. This has long been known to be true for massive black holes in AGN. We now know that many star-sized binary accretion systems within the galaxy drive out well-collimated jets. We also now know (as in chapter 10) that jets are commonly found associated with YSO's, presumably in their disk-accretion phase.

With these general ideas in mind, let's look at some of the objects and astrophysical settings for "classical accretion".

11.2 The Setting: Compact Stellar Remnants

Our focus in this chapter is accretion onto compact objects – neutron stars and black holes. While this is not a course in stellar structure, one of our applications has been the ISM. We remember that the state of being a "star" ties up that piece of the ISM for quite while. Let's bypass that and jump from YSO's to the very end of the star's life.

11.2.1 From main sequence stars to remnants

We can recall the likely events at the end of a star's main sequence life, after it has exhausated its nuclear fuel and nuclear burning has stopped. Its evolution depends on its mass. One scaling mass is the Chandrasekhar mass, $M_{ch} \simeq 1.4 M_{\odot}$; this is the maximum mass of a star which can be supported by electron degeneracy pressure. Current thinking as to the end of a star's life is as follows – though note this picture is vastly oversimplified:

• $M \lesssim M_{ch}$: the star quietly settles to a state supported by electron degeneracy pressure, that is a white dwarf.

• $M_{ch} \lesssim M \lesssim 8M_{\odot}$: the star develops an electron degenerate core, at $M \sim M_{ch}$, which becomes a white dwarf. The rest of the star's mass is ejected, in some form such as a strong stellar wind or a planetary nebula.

• $M \gtrsim 8M_{\odot}$: the star meets a violent end. When its fuel is exhausted, the core collapses suddenly. This collapse drives the core through the electron degeneracy density, into a more dense and more compact remnant. The outer layers of the star "bounce", and are driven outwards at high speed. This is a Type II supernova.[4] Detailed models of SN currently predict that the remnant is a neutron star if the original star's mass is $\lesssim 30 M_{\odot}$, and is a black hole for larger original masses.

We should note that these arguments are based in the best current stellar evolution models – but that they are still somewhat uncertain (the number "8" really means "several to 10"). In addition, there is still a discrepancy between the supernova rate in the galaxy and the required pulsar birth rate – so we don't understand everything yet. What we can say with certainty, is that the normal process of stellar evolution ties up some of the total (baryonic) mass of the galaxy into remnants – white dwarfs, neutron stars and black holes – which just sit there, providing gravity but not interacting with the ISM or galactic evolution any further.

11.2.2 The result: (star-sized) compact objects

Our interest here is the connection between these stellar remnants and accretion processes in astrophysics. To that end, I'll group them by how we observe them, and/or their role in high-energy astrophysics.

• **White dwarf stars** are the low-mass end of the compact-object set. We are not going to say much about them...they are not as important in high-energy astrophysics.

• **Isolated neutron stars** can appear as pulsars. These small (radius $\sim$ 10 km), rapidly rotating stars have strong, narrowly beamed "hot spot" sources of coherent radio emission; when the beams rotate into our line of sight we see a "pulse". A few are also pulsed X-ray and γ-ray sources, as the radio emission region seems also to emit beamed X-ray and γ-ray photons. These stars have very high magnetic fields ($\sim 10^{12}$ G), and a dense, corotating charged magnetosphere (probably composed of an electron-positron "pair" plasma). They emit coherent radio radiation – generated by some (as yet uncertain) plasma process in which bunches of charges oscillate together. Finally, it's thought that they drive *relativistic winds*, which carry mass and energy away from the star; these winds may be what feeds the filled supernova remnants (called *plerions*).

• **Neutron stars in binary systems** have many options. Some of these are pulsars, particularly the rapidly rotating millisecond pulsars which are thought to have been spun-up by accretion of matter from their companion. In addition, binary-system neutron stars are often strong X-ray sources – not beamed but more isotropic. The energy source here is very likely quasi-steady accretion, through an accretion disk, from the companion star. In the accretion process, gravitational energy is turned into heat (possibly also relativistic, nonthermal particles), and from there goes to radiation. Jets are also common – but not universal – in such systems (such objects as SS433, or the "microquasars" being found recently). As with protostellar jets, the existence of jets here reinforces the jet-accretion disk connection.

• **Black holes in binary systems** can also be accretion-powered X-ray and γ-ray sources. Just as with neutron star binaries, the radiation in BH binaries comes from the the accretion process; the fact that the "action" is occuring at somewhat smaller radii should lead to some differences in the details. The most important point about these systems is probably that they exist. That is, the parameters of the binary orbit allow us to determine the mass of the compact, accreting companion. There are a handful of systems for which the compact object's mass is comfortably above $3M_{\odot}$ (the theoretical "Chandrasekhar"-type upper limit on the mass of a neutron star) — and for these systems, we can argue strongly that they contain a star-sized black hole.

[4] You recall the two types of supernovae. Type I are believed to come from accretion of matter onto a white dwarf, which drives a thermonuclear explosion. They are very uniform in their spectra and their light curves, with potential use as standard candles. Type II are much more varied in their properties, and are thought to come from stellar collapse as described above.

• Finally, an historical comment. **Gamma-ray bursters** used to be included in this list. Until a few years ago, people argued as to whether they are galactic or extragalactic (there was no clear measure of their distance). Mose people favored a galactic location; at such a distance, the energy of their bursts suggested some explosive accretion event involving a neutron star. Recently, however, it has been shown conclusively that they are extragalactic (from the high redshift of spectral lines in the associated optical sources; finding such sources was a major step forward). This pushes their energies up by a lot – if they were isotropic emitters, the energy released would approach the regime where models might involve an explosive release of the entire binding energy of a neutron star (as in a NS-NS collision?); larger than the canonical 10^{51} erg known to be released in a supernova. But again, another important observation clarified this: a brand new supernova was found at the exact spot where a GRB had just gone off. How can we reconcile the energetics? It can all work if the GRB is *beamed* – if part of the supernova process is the creation of a short-lived relativistic jet of material. Relativistic effects make the γ-radiation emitted by this jet appear much brighter when viewed close to the jet's direction of motion – that's called *forward beaming*. And yet the game isn't over: there are two types of GRBs (short pulse and long pulse); only one type seems to be consistent with SN explosions; the origin of the other type is still being discussed.

11.3 The Setting: Active Galactic Nuclei

Black holes in galactic nuclei aren't really "stellar", but they are related to star-sized compact objects in the ways they shine (and the ways in which they are modelled). Active galactic nuclei – Seyfert galaxies, radio galaxies, quasars – are believed to be powered by accretion onto supermassive black holes ("SMBH"; $\sim 10^8 - 10^9 M_\odot$). In a very few objects spectral lines or masers can be localized close to the central mass, with clear signs of ordered (disk-like) rotation. The inferred velocities are used to determine the central mass. It's worth pointing out that this isn't directly a detection of a black hole; it's a detection of a small, "massive dark object" (MDO, the term in some of the literature). Whether MDO or BH, models of AGN generally assume an accretion disk flow, with all the associated physics that can occur in star-sized accretion systems. Some AGN – those in radio galaxies – are also, of course, associated with strong, collimated, relativistic jets. In addition, we now have strong (if indirect) evidence that *every* galaxy contains a MDO in its heart – with mass related to the mass of the "bulge" part of the galaxy. But in most galaxies the MDO is not "active"; we detect it only by its gravitational effects.[5] More on this next term.

11.4 Black Holes (a quick visit)

This isn't a course in general relativity, so these notes are not the place for a full exposition of general relativity. I will assume that you have seen the basics, such as in Carroll & Ostlie, and will focus on those aspects of black holes which are relevant to their role in high-energy astrophysics. We will only have a once-over-lightly visit, emphasizing the aspects of black holes that are important for their accretion-related astrophysics.

In terms of accretion physics – or the black hole's impact on its surroundings – we only need to understand a few critical radii. We need to know the size of the *event horizon* – that's the surface inside of which nothing (not a rock, a photon, you or me) can escape. We also need to know a little bit about *stable orbits*, as follows.

11.4.1 Stable orbits

To set the stage, think about circular orbits around a mass M in Newtonian gravity. They're easy: gravity provides the centripetal force. Thus, $GM/r^2 = v^2/r = L^2/r^3$ connects r to v (or to the angular momentum per mass, $L = rv$), uniquely. At any radius r, the orbit is described by $L^2 = GMr$. In addition, you know this orbit is *stable*: think about perturbing a planet in orbit around the sun. It will just oscillate radially around its initial radius – that's an epicycle. In other words, any circular orbit in Newtonian gravity is stable. (This also holds for closed elliptical orbits, the math just gets longer). You also know about open (hyperbolic) orbits: if I start at infinity and throw a rock at the sun, and it has any angular momentum at all, it will pass by the sun and escape back to infinity. It won't be captured unless I drop it directly at the sun (that means it has zero angular momentum).

However this changes in General Relativity. At large radii we can indeed find a stable circular orbit; and as $r \to \infty$, the solution approaches the Newtonian one.

[5] Does this mean it's not accreting matter? If so, why not?

But for small radii, *we can no longer find a stable circular orbit*. If you try to put a rock in orbit close to the black hole, it will either fall in or fly away. You won't be able to find an equilibrium. The *innermost stable radius* is critical for disk accretion – it's the smallest radius for which quasi-stable accretion disks can exist. This innermost radius is also relevant to *capture orbits*: if a rock comes too close to the black hole, it will be captured (it's trajectory will pass within the event horizon), even if it has finite angular momentum. The capture radius is usually comparable to the innermost stable radius although it has to be derived by more complicated methods.

Now, we carry on to describe (not derive!) black hole solutions of Einstein's field equations.

11.4.2 Schwarzchild black holes

Formally, a "black hole" is the name we give to a particular vacuum solution of Einstein's field equations. The fundamental quantity in classical GR is the metric: the distance between two space-time points. The GR field equations are a set of second-order PDE's which describe the metric and its connection to the sources.[6] Solutions to these equations must assume some type of symmetry, and will involve one or more constants of integration.

Two important vacuum solutions have been worked out.[7] The simplest is an isotropic, static solution, which has one constant of integration (we call it M). This is the **Schwarzchild metric**. It is the most familiar: it describes the space around a non-rotating object of mass M.

A Schwarzchild black hole has one critical surface, the *event horizon*, at $r_s = 2r_g = 2GM/c^2$. This is the surface of no escape; no trajectory (particle or photon) that starts within r_s can reach the outside world. It's the surface of infinite time dilation: periodic signals starting at r_s are dilated to infinite period. It's the surface of stationarity: outside of r_s, you can sit still (think of firing your rocket motors "downward", to hold yourself in a fixed position relative to the star), but inside r_s, this is not possible. (Mathematically, it's easy to show that $dt > 0$ – advancing time – requires $ds < 0$ – motion in space – inside of r_s). However, you should note that r_s is not in any way a physical barrier for something moving inwards; except for tidal forces, you don't notice anything dramatic as you cross the event horizon. The unusual effects are related to how you or your signals connect with the outside world.

What about astrophysical applications that do not involve a mass crossing the event horizon? This is where we need to analyze orbits in this geometry. When this is carried out, we find there is a *minimum* radius for which stable circular orbits are possible. It is:

$$r_{ms} = 6r_g = 3r_s \tag{11.5}$$

No stable orbits are possible at smaller radii. This is, thus, another important scale for a Schwarzchild black hole; it is often taken as the inner edge of an accretion disk around such a BH.

11.4.3 Kerr black holes

The next simplest is an axisymmetric, static solution, which has two constants of integration (M and J). This is the **Kerr metric**. It describes the space around an object of mass M and angular momentum J. Notation: it's common to work with the parameter $a = J/Mc$, the normalized angular momentum per mass. Well-behaved solutions exist only for $a < M$. Because of this mathematical limit,[8] the astrophysical speculation is that more rapidly rotating systems can never become black holes.

A Kerr BH has two important scaling radii. It has an *event horizon*,

$$r_o = \frac{r_s}{2} + \left[\left(\frac{r_s}{2}\right)^2 - a^2\right]^{1/2} \tag{11.6}$$

which is the surface of no escape, just as in the Schwarzchild case. Note this is a spherical surface,

[6] Want a familiar example? Think of Maxwell's equations: they are PDE's with the field terms $(\mathbf{E}, \mathbf{B})$ "on the left", and the source terms $(\rho, \mathbf{j})$ "on the right". The GR field equations are analogous – the terms on the left involve the metric; the terms on the right involve the sources, namely, the distribution of mass-energy. A vacuum solution is analogous to a point charge in empty space – no external sources.

[7] A third solution is the **Reissner-Nordstrom metric**; it is not a vacuum solution, but rather allows a radial electric field to exist throughout space. This solution adds a third constant, Q, corresponding to the object's charge. There is little evidence that any astrophysical body carries significant charge, and good arguments against that being the case (think: if you charged up a star, how long would it stay charged, given all those free charges around in the ISM?) Thus, the R-N solution is rarely invoked; the first two are the common ones.

[8] which is in units with $G = c = 1$, which relativists love; in more normal units, the limit is $J/Mc < GM/c^2$.

also as in the Schwarzchild case. The *surface of stationarity* is distinct in the Kerr metric:

$$r_+ = \frac{r_s}{2} + \left[\left(\frac{r_s}{2}\right)^2 - a^2\cos^2\theta\right]^{1/2} \tag{11.7}$$

You can escape from within r_+, but you can't sit still. The direction of increasing time, inside r_+, is also the direction of increasing ϕ (the angular coordinate). Note that this surface is not spherical; it bulges out at the equator.

An interesting related effect is *frame dragging*. A particle with no angular momentum at infinity will still want to move in the direction of increasing ϕ. The rotation speed of such a particle, seen at r as measured by a distant observer, is in general

$$\frac{d\phi}{dt} = \frac{rr_s a}{(r^2+a^2) - a^2\Delta} \tag{11.8}$$

where $\Delta = r^2 - 2Mr + a^2$. A more useful form of this can be found by going to the equatorial plane ($\theta = \pi/2$), and taking $r \gg a, M$. Putting all back in physical units, and converting to an angular speed, we have

$$\omega_{LT} \simeq \frac{2GJ}{r^3c^2} \tag{11.9}$$

Orbital mechanics in the Kerr metric are complex. For circular orbits, one again finds that there is a minimum stable radius. Its limits tell most of the story. When $a \to 0$, $r_{ms} \to 6r_g$ (which is the Schwarzchild limit); when $a \to M$ (the maximum possible value), $r_{ms} \to r_g$. These limits are for prograde orbits; retrograde orbits can't get so close in.

Key points

• Accretion energetics;

• Eddington luminosity;

• Thermal state, black body radiation;

• Compact objects, star sized: what they are, how they relate to accretion flows;

• SMBH: where we find them, how they relate to accretion flows;

• MDOs: why can't we assume they're all SMBH?

12 Accretion III: Disk models

We've already seen spherical (Bondi) accretion and talked about the general energetics of accretion flow. We now need a more concrete picture of disk accretion, which we expect to occur when the accreting gas has significant angular momentum.

The disk geometry may occur in at least two ways. The common picture is that of matter being dumped onto one member of a binary star system from its companion; the mass in this case clearly has significant orbital angular momentum and is confined more or less to a plane. Another possible case would be that of gas which is initially accreting spherically, with some angular momentum, and which can also lose energy radiatively. Radiative dissipation will reduce the internal energy and thus lead to a thin disk; but whatever dissipates the angular momentum of the gas may well be less efficient, so that the accreting gas remains supported by rotation in one plane, while cooling and flattening in the other direction.

Consider a binary star accretion disk, to be concrete. Mass will move from one star to the other when one of the stars (the companion) expands to fill its Roche lobe. The mass coming through the Lagrange point will have significant angular momentum, and will probably be initially in a non-circular orbit. However, it can lose energy quickly, by radiation, and will settle itself into a circular orbit (the lowest energy orbit for a given angular momentum). This will be a Keplerian orbit, with $v_\phi = (GM/r)^{1/2}$ (orbital velocity), and $l = (GMr)^{1/2}$ (specific angular momentum). This matter can only spread in radius if some of it loses angular momentum; this will happen, slowly, due to viscosity (as described below). Thus, if the mass flux stays fairly steady, the system will develop a steady accretion disk.

12.1 Models of thin (alpha) disks

One type of accretion disk model – the oldest, the first one developed – can be treated analytically. To start, we assume the gas in the disk moves in a circular orbit, at the local Kepler velocity, $v_\phi = r\Omega(r)$ where the angular speed is $\Omega(r) = (GM/r^3)^{1/2}$. The gas slowly drifts inward, at a radial/inflow velocity, $v_r \ll v_\phi$, and a mass accretion rate,

$$\dot{M} = 2\pi r v_r \int \rho dz \qquad (12.1)$$

We will limit ourselves to physically thin disks; this means we can reduce the problem to one dimension. The vertical thickness of the disk is determined by hydrostatic equilibrium. This condition is, again, $\nabla p = \rho \mathbf{g}$, or $\partial p/\partial z = \rho g_z$. In the thin-disk limit, we write $\partial p/\partial z \simeq p/H$ if H is the disk scale height. For a disk which is dominated by the gravity of its central mass (M), rather than by its own self gravity, we have $g_z \simeq GMH/r^3$. Thus, the vertical support condition is

$$\frac{H}{r} \simeq \left(\frac{p}{\rho}\frac{r}{GM}\right)^{1/2} \simeq \frac{c_s}{v_\phi} \qquad (12.2)$$

so that a cool disk is a thin disk. The surface density is, then,

$$\Sigma = \int \rho dz \simeq \rho H \qquad (12.3)$$

Past this point, the analysis gets furry. There are many different models of accretion disks and accretion flows. The literature is a bit daunting, with a plethora of differing assumptions (and consequently differing results). However,

don't panic

The basic physics of steady accretion disks can be understood by looking at the simplest of the models that are out there, namely steady-state, spatially thin disks. These models are governed by a few simple principles: mass and momentum conservation and the effects of viscosity. However, even in the simplest model which these notes address, understanding how these basics affect the structure of the disk requires some algebra. I'm trying to lay out the argument in detail in these notes, following *Accretion Power in Astrophysics*, by Frank, King & Raine. Some of the important results in these notes are:

Mass conservation	equation (12.5), (12.6)
Viscous torques	equation (12.8), (12.9)
Angular momentum	equation (12.11), (12.12)
Inflow velocity	equation (12.21)
Alpha (α)	equation (12.22)
Accretion luminosity	equation (12.25)

So fasten your seatbelt ..

12.1.1 Mass conservation

Let's start simply. Consider a ring of disk material, lying between r and $r + dr$. It has total mass, $M(r, r+$

$dr) = 2\pi r\Sigma dr$. Now, this mass changes due to flows into and out of the ring; that is,

$$\frac{\partial}{\partial t}(2\pi r\Sigma dr)$$
$$= v_r(r)2\pi r\Sigma(R) - v_r(r+dr)2\pi(r+dr)\Sigma(r+dr)$$

And thus, we have

$$\frac{\partial}{\partial t}(2\pi r\Sigma dr) \simeq -2\pi(dr)\frac{\partial}{\partial r}(rv_r\Sigma) \qquad (12.4)$$

Note we have assumed there are no local sources or sinks of mass, just the flows into and out of the ring. From this, as $dr \to 0$, we get the mass conservation equation:

$$r\frac{\partial\Sigma}{\partial t} + \frac{\partial}{\partial r}(rv_r\Sigma) = 0 \qquad (12.5)$$

(Compare (4.2), our original form for mass conservation; can you see the connection to 12.5?). And: in a steady-state system, we have the expected expression for mass conservation:

$$\dot{M} = 2\pi r\Sigma v_r = \text{constant} \qquad (12.6)$$

Thus, as expected, $\dot{M}$ is constant with radius in a steady-state flow.

12.1.2 Viscosity and torque

What happens about the angular momentum? How does gas with finite angular momentum ever manage to move inwards? To answer this, consider a parcel of gas at r. It has specific angular momentum $l \propto r^{1/2}$, and this must be lost if the parcel is to move inwards. This is accomplished by the friction between rings of the differentially rotating disk. The friction is transmitted by viscosity.

> *Big fudge coming:* Formally viscosity is a microscopic process, transporting momentum "sideways" by particle collisions. We know how to treat this for a plasma, using Coulomb collisions as always. However, astrophysically particle-based viscosity is often very small (due to ionized plasmas being such good conductors), and turbulent viscosity will dominate. If the turbulence has mean velocity v_{turb} and mean scale λ, the coefficient of viscosity $\nu \sim \frac{1}{3}\lambda v_{turb}$. This is conceptually just fine, but very hard to write down analytically – so just about all accretion disk work makes a standard assumption (read "fudge"), as discussed below.

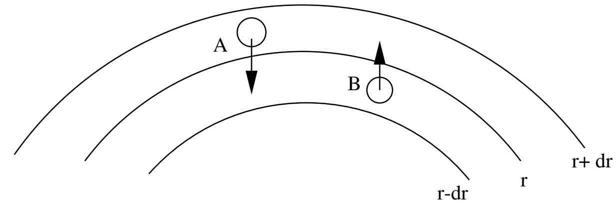

Figure 12.1 Viscous angular momentum transport in a shearing medium. Parcels A, B can be thought of either as single particles (for microscopic viscous transport) or as turbulent "eddies" (for turbulent viscosity). Following Figure 15 from Frank, King & Raine; connect notation to the text as $\lambda \leftrightarrow dr$

We now illustrate the effect of viscosity on angular momentum transport by considering a ring of matter at r, and two adjacent rings, at $r - \lambda$ and $r + \lambda$. The rings have $\Omega(r)$, $\Omega(r+\lambda) \simeq \Omega(r) + \frac{d\Omega}{dr}\lambda$, and $\Omega(r-\lambda) \simeq \Omega(r) - \frac{d\Omega}{dr}\lambda$. The scale λ can be any local differential, but it is most useful to connect it to the mean free path of whatever accounts for the viscosity.

To first order, viscous transport between the rings (which may be carried by single-particle collisions, or by more efficient turbulent motions) exchanges equal amounts of mass between the layers. We can write this mass flux as $2\pi\rho v_{turb}rH$ if v_{turb} is some characteristic turbulent or microscopic turbulent transport velocity. Note the rate of mass flux inwards must be the same as outwards, in a steady state disk. Now, consider an observer in corotation with the fluid on the surface $r =$ constant. The fluid at $r - \lambda/2$ will appear to move with velocity $v_\phi(r-\lambda/2) = (r-\lambda/2)\Omega(r-\lambda/2) - \Omega(r)r$. Thus the average angular momentum flux per unit lengh carried through r =constant in the outward direction is

$$\rho v_{turb}H\left(r - \frac{\lambda}{2}\right)\left[\left(r-\frac{\lambda}{2}\right)\Omega\left(r-\frac{\lambda}{2}\right) - r\Omega\right]$$

A similar expression, changing the sign of λ, gives the mean inward momentum flux per length. The difference between in and out gives the net outward momentum flux (which is also the torque per length). If we multiply this by $2\pi r$ we get the net torque exterted by the outer ring on the inner (and = - the torque of the inner on the outer, right?):

$$G(r) \simeq 2\pi r\lambda v_{turb}\Sigma r^2\frac{d\Omega}{dr} \qquad (12.7)$$

But now, we can collect λv_{turb} in the viscosity coefficient, ν, to write this as

$$G = 2\pi r^3\nu\Sigma\frac{d\Omega}{dr} \qquad (12.8)$$

where we've defined $\nu = \lambda v_{turb}$.

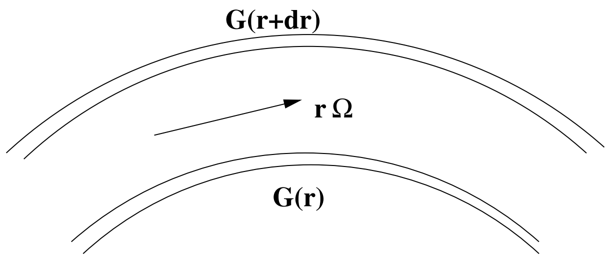

Figure 12.2 The net shear force on an annulus at r is the difference between the shear $G(r)$ on the outer and inner edges. This net shear force acts on the ϕ-component of angular momentum.

Finally, then, the net torque on a ring at r is the difference between the torque exerted by the inner and outer rings:

$$G(r+dr) - G(r) \simeq \frac{dG}{dr}dr \tag{12.9}$$

Thus, a Keplerian system in which $d\Omega/dr < 0$ has a net *outwards* flux of angular momentum. It is this which allows the inwards flow of matter, as a packet of gas slowly loses its l.

12.1.3 Angular momentum equation

Next: we need a conservation law for specific angular momentum. We'll use the same approach as we used for mass conservation – consider a ring of matter at r. It has total angular momentum $2\pi r(dr)\Sigma r^2\Omega$. This changes due to matter flowing in and out of the ring, and also due to the net torque on the ring (exerted because we've assumed each adjacent ring is in Keplerian motion, and because the matter is viscous). So we can repeat the same bookkeeping as above ... giving

$$\frac{\partial}{\partial t}\left(2\pi r(dr)\Sigma r^2\Omega\right) \simeq \\ -2\pi(dr)\frac{\partial}{\partial r}\left(r\Sigma v_r r^2\Omega\right) + \frac{\partial G}{\partial r}(dr)$$

Taking the limit $dr \to 0$, this becomes exact, and we find the conservation law for angular momentum:

$$r\frac{\partial}{\partial t}\left(\Sigma r^2\Omega\right) + \frac{\partial}{\partial r}\left(r\Sigma v_r r^2\Omega\right) = \frac{1}{2\pi}\frac{\partial G}{\partial r} \tag{12.10}$$

Now – unlike the mass equation, the steady solution here leads to a "flow of angular momentum" which is a function of radius. In a steady state, (12.10) becomes

$$2\pi\frac{d}{dr}\left(r^3\Sigma v_r\Omega\right) = \frac{d\dot{J}}{dr} = \frac{dG}{dr} \tag{12.11}$$

where I've defined $\dot{J} = 2\pi r^2\Sigma v_r\Omega$ as the rate of angular momentum flow (outwards); note it is a function of r.

Now, integrate (12.11) over r:

$$\nu\Sigma\frac{d\Omega}{dr} = \Sigma v_r\Omega + \frac{C}{2\pi r^3} \tag{12.12}$$

which can also be written,

$$r^2\Sigma v_r\Omega = \frac{G}{2\pi} + C \tag{12.13}$$

Here, C is a constant of integration.

The conventional approach evaluates this at the inner boundary of the disk, at r_1 say. *If* we're talking about accretion onto a hard-surface star, r_1 is the stellar surface; the flow must approach solid-body rotation at the surface, so that $d\Omega/dr \to 0$ there. Alternatively, *if* we're talking about accretion onto a black hole, r_1 is taken as the minimum stable orbit. Inside of that the matter just "falls right across the event horizon", so that (it might be reasonable to assume) there is no torque on the last ring, $G \to 0$ there. In either case, this argument shows that the integration constant $C = -\dot{M}\left(GMr_1\right)^{1/2}$.

We get two important results from this. The first is a direct solution of (12.12):

$$\nu\Sigma = \frac{\dot{M}}{3\pi}\left[1 - \left(\frac{r_1}{r}\right)^{1/2}\right] \tag{12.14}$$

This expression for the product $(\nu\Sigma)$ will be useful below. The second is an expression for the rate of angular momentum flow, $\dot{J}$. Going back to the definition, in (12.11), using (12.12) and the value of C, we get

$$\dot{J}(r) = -r^{1/2}(GM)^{1/2}\dot{M} \tag{12.15}$$

This thus shows explicitly that J flows *outward* when M flows *inward*, and also that $\dot{J}$ is a function of radius.

12.1.4 Accretion rate and radial velocity

Now that we have an expression for the viscous force, we can look at its effect on flow within the disk. We start with the basic equations of mass conservation,

$$r\frac{\partial\Sigma}{\partial t} + \frac{\partial}{\partial r}\left(r\Sigma v_r\right) = 0 \tag{12.16}$$

and angular momentum conservation,

$$r\frac{\partial}{\partial t}\left(\Sigma r^2\Omega\right) + \frac{\partial}{\partial r}\left(r\Sigma v_r r^2\Omega\right) = \frac{1}{2\pi}\frac{\partial G}{\partial r} \tag{12.17}$$

Combining these, and noting $\partial\Omega/\partial t = 0$, we get

$$r\Sigma v_r \frac{\partial}{\partial r}\left(r^2\Omega\right) = \frac{1}{2\pi}\frac{\partial G}{\partial r} \tag{12.18}$$

(compare 12.11). We can also combine this with mass conservation to get

$$r\frac{\partial \Sigma}{\partial t} = -\frac{\partial}{\partial r}\left[\frac{1}{2\pi}\left[\frac{\partial(r^2\Omega)}{\partial r}\right]^{-1}\frac{\partial G}{\partial r}\right] \tag{12.19}$$

This result will give us the behavior of Σ, once Ω and G (which involves the viscosity ν) are specified. If we now assume Kepler orbits, (12.19) becomes

$$\frac{d\Sigma}{\partial t} = \frac{3}{r}\frac{\partial}{\partial r}\left[r^{1/2}\frac{\partial}{\partial r}\left(\nu\Sigma r^{1/2}\right)\right] \tag{12.20}$$

This last is the basic equation governing the time/space evolution of the density in a Keplerian disk. Solutions to this are not simple – they depend on the local viscosity (discussed below), and also on the local energetics (which we haven't even discussed – temperature and whatnot). When we do get a solution – and the literature is full of them – we can go back to (12.18) to find the radial/accretion velocity:

$$v_r = -\frac{3}{\Sigma r^{1/2}}\frac{\partial}{\partial r}\left(\nu\Sigma r^{1/2}\right) \tag{12.21}$$

Thus: the radial velocity (and thus $\dot{M}$, from equation 12.6) is determined by the local viscosity and the local surface density. We don't have enough information here to solve the system – that needs further arguments about internal energetics which I'm not going to go into. But we can see, directly from (12.21), that the inflow velocity $v_r \sim \nu/r$ (to order of magnitude).

12.1.5 What is ν?

The catch, of course, is that we do not know how to find the viscosity ν. This is the quantity which determines the rate at which matter can move inwards (the accretion rate); in addition, all of the details of the disk models (r-dependence of Σ, v_r, temperature, etc.) depend sensitively on ν. The common fudge in disk models is to parameterize the viscous stress in terms of the pressure, through a factor α = (viscous stress) / (pressure). Now, viscous forces come from adjacent rings of matter which are not moving at the same speed – so one slips against the other if $d\Omega/dr \neq 0$. Collecting all this, with the definition of α and the proper way to write the viscous stress, we get

$$\alpha p \simeq \nu\rho r\frac{d\Omega}{dr} \tag{12.22}$$

Remebering that ν has dimensions (turbulent velocity)$\times$(turbulent length scale), or (mean particle speed)$\times$(collision mean free path), you should see that (12.22) makes dimensional sense. This entire, grand hand-wave is collected in the term "α-disk". It's common to take $\alpha = 0.1$ in the literature.[1]

12.1.6 Energy dissipation and luminosity

One nice result is that we can find the radiated energy, and thus the efficiency ε, without knowing the details of α. To get this, we use the fact that shear stress dissipates energy. The local heating rate for this disk, integrated over the disk thickness – call it $D(r)$ (for dissipation) – is

$$D(r) = \frac{G}{4\pi r}\frac{d\Omega}{dr} = \nu\Sigma r^2\left(\frac{d\Omega}{dr}\right)^2 \tag{12.23}$$

(The first equality is the definition of heating by viscous shear stress; the second used the result 12.8 for G). From this, in a Kepler disk, we can find the local heating rate in terms of the basic quantities:

$$D(r) = \frac{3GM\dot{M}}{8\pi r^3}\left[1 - \left(\frac{r_1}{r}\right)^{1/2}\right] \tag{12.24}$$

Now: where does this energy go? Viscous dissipation heats the local gas. *If the inflow speed is slow* – an important assumption in this type of disk model – then the gas must be able to radiate this away locally. That means we can equate $D(r)$ to the local luminosity per area from the disk. Integrating this over r, we find an expression for the luminosity:

$$L = \int_{r_1}^{\infty} D(r)2\pi r dr = \frac{GM\dot{M}}{2r_1} \tag{12.25}$$

Thus: if r_1 is small – as for a black hole ($r_1 \simeq r_{ms}$, the minimum stable orbit); or even for a neutron star – then the efficiency (refer back to 11.1) in this type of model can be high: $\varepsilon = r_s/2r_1$ where r_s is the Schwarzchild radius.

[1] Some order-of-magnitude estimates are perfectly valid in astrophysics. But perhaps one should not put overmuch confidence in the details of complicated, ornate accretion disk solutions – houses of cards – based on assumptions such as this α.

12.2 Extensions of the model

The basic, thin-disk, α-disk model in the previous section has been expanded and reworked extensively in the literature, in order to try to match the models to the data. Some of the extensions are as follows.

12.2.1 Hot and/or thick disks

In the previous model we assumed the disk is cold (that is, $c_s \ll v_\phi$), and therefore thin. This allowed a one-dimensional treatment. This isn't the only possibility, of course – the disk may be warm or hot, and therefore thick. Two versions of this exist in the literature. One is a "thick disk", a disk which is hot throughout. Such disks are still being worked on – their dynamics can be complicated, and they are probably globally unstable (they break up into non-axisymmetric structures – big "lumps" in the disk). Another variant is a cool, thin disk with a hot atmosphere ("corona"), or possibly a hot outflowing wind. This hot component does not radiate as a black body; it can be a source of high-frequency radiation (for instance γ rays, $h\nu \lesssim m_e c^2$).

12.2.2 Accretion flows

The α-disk model, above, was really developed for accretion onto a hard-surface star. Quite different inner boundary conditions, and inner flows, can occur if the disk sits around a black hole. The innermost gas can just slide across the event horizon. In such flows, the radial velocity, v_r, can go through a sonic point (much like Bondi accretion) before it reaches the last stable orbit. If this happens, a parcel of gas doesn't have time to radiate "locally" – so that the assumptions going into (12.23, 12.24) don't hold. These solutions – called ADAFs (for Advection Dominated Accretion Flows) – are therefore *underluminous*, with $L \ll GM\dot{M}/r$. Thus, an accreting black hole can be faint (relative to its Eddington luminosity) for at least two reasons. It may have a low accretion rate ($\dot{M} \ll \dot{M}_{edd}$); or its accretion flow may have a high $\dot{M}$ but be ADAF-like.

12.2.3 MHD effects

Everything we've done in this chapter has been field-free, *i.e.* purely hydrodynamic. That's probably unrealistic, because the accreting plasma is almost certainly magnetized. Think about the B field in the accreting material being "tied to infinity", to start. Flux freezing means that the field lines will be dragged along with the plasma; the field will be amplified, and will develop a toroidal component (forming sort of a helix). Two things follow from this. The helical field can channel, and even accelerate, a wind-type outflow ("MHD winds"). In addition, the field lines tied to the rotating plasma create a $\partial\mathbf{B}/\partial t$, which creates a $\mathbf{E}$ field: $\mathbf{E} = \mathbf{v}_{rot} \times \mathbf{B}/c$. This E field may be able to accelerate particles (are accretion disks a source of cosmic rays?). In addition, this system creates a $\mathbf{E} \times \mathbf{B}$ Poynting flux, also directed out along the rotation axis; this may be how jets are made.

Key points

- Thin accretion disk: the basic picture;
- Viscosity: what it is, why its important, how it controls $\dot{M}$;
- Energetics: what is the disk's luminosity?