PHYSICS 426 NOTES: PLASMA ASTROPHYSICS

Original version (2008): Jean Eilek
Physics Department, New Mexico Tech
Socorro, NM 87801, U.S.A.
JEILEK@AOC.NRAO.EDU

minor revisions (2009-2013): Lisa Young

These notes continue what we started in Physics 425. Our goal is still to explore the "physics of astrophysics". This term we'll develop the radiation-physics tools we'll need, and focus more on galaxies, their interstellar medium, and high-energy astrophysics.

Contents

1 Galaxies, normal and otherwise

To start, let's recall the large-scale structure of galaxies. We are going to focus on bright galaxies, spirals and ellipticals. A lot is known about these objects; in addition to being pretty, they are bright and extended, thus easy to study. We should remember, however, that most galaxies in the universe are neither S nor E. By counting galaxies we can determine the *luminosity function*, that is the number of galaxies at luminosity L (per volume usually). This is called the *Schechter function*, and has the approximate form,

$$\Phi(L) = \Phi_o \left(\frac{L}{L_*}\right)^{-\alpha} e^{-L/L_*} \qquad (1.1)$$

Here, Φ_o is a normalizing constant (which depends on the local environment: for instance Φ_o is much larger for a rich cluster than for the field). The slope $\alpha \sim 1.25$; and $L_* \sim 7 \times 10^{10} L_\odot$ is comparable to the characteristic luminosity of bright galaxies (values quoted by Elmegreen, for the V band). We know that the galaxy luminosity connects directly to the galaxy's mass: $M/L \sim 30 - 100$ (in solar units; depending somewhat on the galaxy type, and with some scatter). Thus the Schecter function also measures the mass distribution of galaxies.

Important point: this is a composite luminosity function, determined by adding all galaxy types. Most S and E galaxies sit within a factor of a few of L_*. But the LF has been measured over a range ~ 10 magnitudes, that is a factor $\sim 10^4$ in luminosity, and it keeps rising to smaller luminosities. It follows that *most of the galaxies in the universe are neither spirals nor ellipticals*. Most galaxies are either *Irregulars* (small, patchy structure, rotation supported) or *dwarf Ellipticals* (small, featureless, probably not rotation supported). Check Figure 23.34 of Carroll & Ostlie for a recent break-down of the total LF by galaxy type. In this course we will focus on the big galaxies, which are better understood; but don't forget the little guys.

1.1 Normal galaxies: spirals

Our galaxy is a medium-size spiral, and we can use it to study a "typical" spiral. As with all spirals, its most notable feature is its disk, which contains both stars and gas. The surface mass density of the stellar disk is exponential, $\Sigma(R) \propto e^{-R/H}$, with $H \sim$ few kpc (that number is typical of other nearby spirals). Thus the density of stars dies after several H distances. The gas (HI and molecular gas), however, extends much further; in big spirals gas can be traced to 40-50 kpc, and it has been traced to at least ~ 20 kpc in our galaxy. The surface density of the gas falls, in some galaxies exponentially and in others more irregularly; its radial scale length is generally larger than that of the stars.

What is the structure of the gravitating matter?

The disk is supported vertically by the random motions ("heat") of the stars and gas. Think about individual stars: each one moves up and down, through the galactic plane, in approximate harmonic motion. The vertical extent of its motion depends on the local gravity. Or, think about the stars as a "gas",[1] with random motions at speed σ. We can describe the ensemble of stars in a given volume by a "pressure" ($p \leftrightarrow nm\sigma^2 = \rho\sigma^2$, for number density n and stellar mass m). We then expect the vertical disk structure to be described by hydrostatic equilibrium. This last can be written as a vector equation, or alternatively we can isolate its z ("vertical") component:

$$\nabla p = \rho \mathbf{g}\,; \qquad \frac{dp}{dz} = \rho g_z \qquad (1.2)$$

The typical scale height of the disk, locally, is $H \sim 200 - 300$ pc (varying somewhat for different types of stars, or different phases of the ISM). The surface mass density in the disk can be found in two ways. One, we can measure the local density of stars and gas directly, to get a density of *luminous mass*. Two, we can use the vertical support condition (1.2) to find the local gravity, and from this (remembering Poisson's law for gravity, $\nabla^2\Phi_G = -\nabla \cdot \mathbf{g} = 4\pi G\rho$), we can find the density of *gravitating mass*. This latter approach gives $\Sigma \sim 75 M_\odot/\mathrm{pc}^2$, while the former (counting stars) gives a value that is only about half of this. Thus, about half of the gravitating mass in the local disk is *dark*: it does not emit any radiation that we have been able to detect.

In addition, the galaxy has more dark matter on larger scales. Consider our galaxy: the disk is supported "horizontally" by its rotation; our local rotation speed ~ 220 km/s, which gives a rotation period at the sun's distance (8.5 kpc) of ~ 230 Myr. The stellar orbits are nearly circular, and the galaxy is close to axisymmetric (or truly, we are assuming its mass distribution is spherically symmetric!), so we can use the simplest form of Kepler's law:

$$v_{rot}^2 = \frac{GM(r)}{r} \qquad (1.3)$$

[1] If you don't like this idea, look at §1.4 for some discussion.

where $M(r)$ is the mass within r. This provides a simple way to estimate the *gravitating* mass of the galaxy.

Turning to nearby external spiral galaxies, we can use the rotation curve to find the gravitating mass. Rotation curves in big spirals can be traced out to several tens of kpc using HI, and their rotation velocity v_{rot} stays constant out that far. Thus, the gravitating mass $M(< r) \propto r$: the mass of the galaxy keeps rising as we go to larger distances. This is *not* what happens to the total luminous mass (in stars plus gas), however: the integral of an exponential converges to a finite value. Thus, the ratio of total mass to luminous mass increases as we go to larger scales. By the outer $\sim 40 - 50$ kpc in big systems, the ratio of (gravitating mass)/(luminous mass) $\sim O(10)$; there is something like ten times more mass in the system than we can see.

A note on spiral arms...they are of course the most striking, defining features of spiral galaxies. They are not, however, fundametal to the galaxy's structure. Rather they are *waves* or *perturbations* in the self-gravitating disk of the galaxy. Some authors like to work with *linear* density waves – *i.e.* small-amplitude waves with a spiral shape. Other authors like to work with *global* perturbations of the galactic disk – which have a generally spiral shape but need not be linear. Still other authors consider local perturbations – in which a local overdensity, say, is enhanced (due to its own self-gravity) and sheared into a spiral fragment (due to the differential rotation of the disk). Probably each approach describes some fraction of spiral galaxies.

And a note on the dark matter ... we have no direct observation of the spatial distribution of the dark matter. It seems likely, however, that the dark haloes of spiral galaxies are spheroidal – *i.e.* supported by their internal, random motions ("heat"), just as elliptical galaxies are (as we discuss in the next section). This idea comes from numerical simulations of structure formation, as well as the fact that the dark matter (by assumption!) does not "cool", so can't dissipate its internal energy – thus it probably has not been able to flatten into a disk.

1.2 Normal galaxies: ellipticals

Elliptical galaxies are quite different. They are smooth and featureless structures, showing a core (of roughly constant density) and an outer envelope (or declining density). Their surface brightness follows the heuristic De Vaucoulers law: $\Sigma(r) = \Sigma_o e^{-(r/r_o)^{1/4}}$ (where r_o is a constant, a length scale). In a three-dimensional system such as this, the surface brightness is a projection of the underlying 3D spatial density:

$$\Sigma(R) = 2\int_R^\infty n(r)\frac{rdr}{\sqrt{r^2 - R^2}} \qquad (1.4)$$

The stellar density is decently well fit by $n(r) = n_o/[r(r+a)^2]$, or by similar (analytic) forms which have a characteristic "core" radius, $a \sim 1 - 2$ kpc. [2]

We might think they are very simple...but that's not the case. One hint comes from rotation: elliptical galaxies are not rotation supported. In a large elliptical the rotation speed is much smaller than the dispersion: $v_{rot} \lesssim 0.1\sigma$ typically. (For smaller galaxies, v_{rot} is a larger fraction of σ). In addition, E galaxies are truly *three-dimensional*. To see this, consider shape of the surface brightness isophotes. They are, of course, elliptical (and close to circular in some E galaxies). But, the direction of their major axis *rotates* going out from the galactic center. This *isophote twist* is a clear sign of a trixial system. Both of these facts tell us that the stellar orbits must be complex, randomly oriented, and not necessarily closed (that is a star can wander through much of the volume of the galaxy over it's lifetime).

Despite the complexity of the orbits, we can find a simple model of the structure of the galaxy. Assume spherical symmetry to simplify, and following the arguments above write down a "hydrostatic balance" equation for the stars:

$$\frac{d(\rho\sigma^2)}{dr} = \rho\frac{GM(r)}{r^2}\,; \qquad \frac{dM}{dr} = 4\pi\rho r^2 \qquad (1.5)$$

Now do some algebra, and combine these into one second-order ODE for $\rho(r)$. The resulting equation has two solutions. One is analytic, $\rho \propto 1/r^2$. This solution diverges in the center (that's not good) and its mass, $M(r) \propto \int \rho r^2 dr$, diverges at infinity (that's also not good). The other must be found numerically, but turns out to have a finite central density, and a characteristic *core radius*; at large radii it approaches the first, analytic, solution. And: you have no doubt recognized this solution, from last term. It is the *self-gravitating isothermal sphere*, an important solution in several different astrophysical applications.

[2] This is the "NFW" (Navarro, Frenk & White) profile; it turns up commonly in numerical, N-body simulations of collapsing, self-gravitating galaxies; and is a reasonable match to the profiles of real galaxies.

The divergence at infinity cannot be fixed analytically in this approach. The most attractive solution that I know of, comes from considering the internal velocity distribution of the stars; those at the highest velocities will exceed the escape velocity of the galaxy. King took this into account in numerical models of the isothermal sphere, and found solutions which are nicely well-behaved at infinity. He also gives an *approximate* expression for the density of the resulting system:

$$\rho_{King}(r) = \frac{\rho_o a^3}{(r^2 + a^2)^{3/2}} \tag{1.6}$$

What about the mass of an elliptical? Is dark matter important? Due to the complexity and variety of the shapes of allowed stellar orbits, we can't easily use Kepler-type arguments, as we could for spirals. We can of course use approximate arguments, such as the virial theorem, to estimate M_{grav} from the stellar velocity dispersion: $GM_{grav} \simeq \sigma^2 r$. Quantitatively, however, to get accuracies at the tens of percent level is not as easy as it is for spirals. Several approaches can be used:

• One, the strongest result in my opinion, uses the small (but non-trivial) minority of the population which have flattened, extended HI disks. These disks rotate, in simple Keplerian motion, which allows us to find the gravitating mass directly. The result: M/L for big E's is similar to that for big S's, ($\sim 10 - 30$ typically), and also tends to increase with radius.

• Another approach uses the hot, X-ray loud gas found in every big elliptical. This gas sits in approximate hydrostatic equilibrium in the potential well of the galaxy; from its spatial distribution we can estimate the gravitating mass. I defer details here until the next chapter; the results for M/L are generally consistent with those from flattened HI disks.

• Careful interpretation of the stellar kinematics involves making a model of the gravitational potential, doing numerical integrations to find out what the allowed stellar orbits look like, and verifying that if one populated those orbits with stars the resulting object would look like a real galaxy and would reproduce the potential assumed at the beginning. At present it is not very common to be able to do this in the far outer regions of elliptical galaxies. In the inner, bright portions of the galaxy the mass is dominated by stars rather than by dark matter.

1.3 Query: why are they different?

Although this isn't a course in galactic structure, it seems appropriate to ask why there are two characteristic types of big galaxies – one flat and supported by rotation, the other round and supported by random motions. Two possibilities have been discussed for ages:

• We might think that galaxies form in isolation: by gravitational collapse of their dark matter, and subsequent dissipational collapse of the baryonic gas (normal ISM!) within the dark halo. (An important point here is that normal, baryonic material can radiate away its interal energy; so it can cool, and collapse in a gravitational field. Dark matter, being non-baryonic, cannot). If this is the case, then the E/S difference might be due to how early in the process stars formed. Early star formation might lead to an E; later star formation, after the ISM has collapsed into a disk, could lead to an S.

• However there's another possibility: galaxies might influence each other during their formation process. Two facts are germane here. First, we know that E's are commonly found in regions of high galaxy density – rich clusters of galaxies. Second, simulations show that if two spiral galaxies collide, closely enough for their stars to remain bound, the energy of the collision will heat the stars and "puff up" the galaxy – *i.e.* making something that looks like an elliptical. So .... it may well be that E's are more common in clusters, because conditions (at least early on) were right for protogalaxy-protogalaxy collisions.

Which of these is right? I suspect the answer lies somewhere inbetween – this is still an active area of current research.

1.4 Interlude: stellar dynamics

How can we justify talking of a stellar "pressure"? Here's a quick overview of the argument, following Binney & Tremaine. Consider a distribution function of stellar velocities, $f(\mathbf{x}, \mathbf{v})$ (the number of stars "at $\mathbf{x}$ and $\mathbf{v}$"). If stars are conserved, and there are no collisions (in which the position and/or the velocity of the star changes instantaneously, by a finite amount), then the evolution of f is goverened by the motion of individual stars through $(\mathbf{x}, \mathbf{v})$ phase space:

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla_{\mathbf{x}} f + \mathbf{a} \cdot \nabla_{\mathbf{v}} f = 0 \tag{1.7}$$

(Think: does this make sense? We're just counting stars). Here, $\nabla_{\mathbf{x}}$ is the usual spatial gradient, $\nabla_{\mathbf{v}}$ is the gradient in phase space with respect to the velocity coordinates, $\mathbf{v}$ is the velocity and $\mathbf{a} \equiv d\mathbf{v}/dt$ is the acceleration. Here, we specify $\mathbf{a} = \mathbf{g} = -\nabla\Phi_g$, *i.e.* the gravitational acceleration. Now, we can derive a couple of useful results.

First, integrate (1.7) over all velocities. We get (specify to Cartesian coordinates to make the algebra more transparent):

$$\int \frac{\partial f}{\partial t} d^3\mathbf{v} + \int v_i \frac{\partial f}{\partial x_i} d^3\mathbf{v} + g_i \int \frac{\partial f}{\partial v_i} d^3\mathbf{v} = 0 \quad (1.8)$$

But now, the last term goes to zero (why? The i component of the integrand becomes $(\partial f/\partial v_i)dv_i$, a perfect differential; it integrates to zero if f has "reasonable" boundary conditions). Also, in the second term we can take the d/dx_i outside the integral (remember that $\mathbf{x}$ and $\mathbf{v}$ are independent coordinates). We can define the stellar density and mean velocity by

$$n = \int f d^3\mathbf{v} \; ; \qquad \langle \mathbf{v} \rangle = \frac{1}{n} \int f \mathbf{v} d^3\mathbf{v} \qquad (1.9)$$

Finally, (1.8) becomes

$$\frac{\partial n}{\partial t} + \frac{\partial}{\partial x_i} \left(n \langle v_i \rangle \right) = \frac{\partial n}{\partial t} + \nabla \cdot \left(n \langle \mathbf{v} \rangle \right) = 0 \quad (1.10)$$

(here and after I'm using ∇ to mean the usual $\nabla_{\mathbf{x}}$). But this is simple: it is just the continuity equation for stars.

Now, multiply (1.7) by $\mathbf{v}$ before integrating over velocity space. The algebra is longer, but the approach (and mathematical tricks) are the same. One version of the result is

$$n \frac{\partial \langle v_j \rangle}{\partial t} - \langle v_j \rangle \frac{\partial}{\partial x_i} \left(n \langle v_i \rangle \right) + \frac{\partial}{\partial x_i} \left(n \langle v_i v_j \rangle \right) = n g_j \quad (1.11)$$

Now, velocity $\mathbf{v}$ is partly due to streaming (all stars share the same mean velocity) and partly due to random motions about this mean. Thus, the mean value $\langle v_i v_j \rangle$ can be split into a part that describes the streaming motion ($\langle v_i \rangle \langle v_j \rangle$) and a part that describes the internal velocity dispersion,

$$\sigma_{ij}^2 = \langle v_i v_j \rangle - \langle v_i \rangle \langle v_j \rangle \qquad (1.12)$$

Using this, and letting $\mathbf{v}$ now represent the *mean* streaming velocity, we can write (1.11) as

$$n \frac{\partial \langle v_j \rangle}{\partial t} + n \langle v_i \rangle \frac{\partial \langle v_j \rangle}{\partial x_i} = n g_i - \frac{\partial}{\partial x_i} \left(n \sigma_{ij}^2 \right) \quad (1.13)$$

or – if we note that $nm\sigma_{ij}^2$ is equivalent to a pressure, p – this becomes

$$n \frac{\partial \mathbf{v}}{\partial t} + n\mathbf{v} \cdot \nabla \mathbf{v} = n\mathbf{g} - \frac{1}{m} \nabla p \qquad (1.14)$$

This is the general dynamical equation for collisionless stellar systems. In a steady state, with no bulk flows, this reduces to the usual hydrostatic equilibrium: $\nabla p = \rho \mathbf{g}$ (being slightly cavalier about the pressure, which is OK for our purposes here).

1.5 The not-so-normal ones: active galaxies

About one per cent of the bright galaxy population is "active". That is, they contain small, highly energetic, *non-stellar* events in their nuclei. (The fraction is less common in smaller galaxies; and more common in bright galaxies at high redshift). Some active galaxies are found optically, others are found in radio. We will return to this topic later in the course. For now, here, I just note the classes and general properties, with an eye to their historical discovery.

Seyfert galaxies are spirals with very bright nuclei. These nuclei are most easily detected by their strong, optical emission lines; they also have nonthermal continuum emission. They also have small ($\lesssim$ kpc), faint radio sources in their cores. The most important point here is that this phenomenon cannot be explained simply by stars: some non-stellar event is taking place in these nuclei.

Radio galaxies are ellipticals which have double-lobed radio sources, fed by jets emanating from galactic core. They also have compact, non-stellar nuclei, but on the weak side compared to bright Seyferts.

Quasars were originally called "quasi-stellar objects". They are bright, compact (inferred from from variability), have very strong emission lines (this is how they were first found), and a strong nonthermal continuum. Some (maybe 10%) are "radio-strong", with a radio-loud core and extended double-lobed radio structure. Some of these ("blazars"; maybe 10% again?) are violently variable, showing a large $\Delta L/L$ on short time scales. These generally also contain radio jets with superluminal motion. Initially no galaxy could be seen around these very bright, small sources – hence the name QSO. With dedication and better technology we now can image the underlying galaxies. This area is still under discussion, but it looks as if quasars show the same host-galaxy split as to Seyferts and AGN:

radio-loud from E's and radio-weak from S's. We now have found quasars out to $z \sim 5$, and probably higher by the time you read these notes.

What all of these objects have in common is an *Active Galactic Nucleus (AGN)*. An AGN is characterized by a high luminosity (which can be comparable to the luminosity of the entire host galaxy); a small intrinsic size (inferred from variability: $\lesssim$ light-day or light-month); a non-stellar spectrum (that is, from diffuse gas, often including nonthermal particles and B field); and a preferred axis (as shown by radio jets – showing us net angular momentum of the core?). The general model is accretion of matter onto massive black hole in the nucleus of the galaxy.

1.6 The beast in the core

A striking recent result is that every galaxy appears to have a massive dark object in its core. At this point we have no definite proof that these massive things are black holes; that would require resolving the event horizon and finding some definitive signature, for instance emission lines red/blue shifted due to the orbital speed of gas in the last stable orbit. What we can do, however, is use gravity to detect a massive dark object (MDO): that is, a total gravitating mass much larger than what can be accounted for by stars in the region.

1.6.1 Techniques: normal galaxies

There are two important detection techniques here, for galaxies without strong AGN. In what follows I use the term "bulges" to mean either elliptical galaxies, or the bulges of early-type spirals.

•**gas disks** Some bulges contain gas disks in (apparent) Kepler rotation around a central MDO. A small number of these, such as M87, have been resolved fairly close to the MDO. A few other galaxies have inner gas disks with maser spots which are easy to resolve with the VLBA and which can also give the gas velocities. A larger number of galaxies have gas disks which can be resolved with HST (albeit not so close to the MDO). Emission line velocities from these disks, again assuming Keplerian rotation, can be used to find the central mass.

•**stellar cusps** Most E's and spiral bulges, however, do not have nice gas disks in their cores. For these, we need to use the fact that a central point mass affects the distribution of nearby stars. A central star cluster with no MDO is well described by a self-gravitating isothermal sphere – which you remember from earlier in the course. The stellar density is (approximately) constant in the inner region of the cluster. On the other hand, if a large point mass, M_{BH}, sits at the center, it will cause the stars to form a *density cusp*, of characteristic scale $r_{cusp} \sim GM_{bh}/\sigma^2$ (if σ is the random stellar velocity).

• **Results** The results of this work is striking: *essentially every bulge yet observed (carefully enough) contains a massive black hole*. Furthermore, the mass of the central object correlates tightly with σ, the velocity dispersion of the nearby stars. This is called the $M_{bh} - \sigma$ relation, usually quoted as

$$M_{bh} \sim 1 \times 10^8 \left(\frac{\sigma}{200 \text{ km/s}} \right)^x M_\odot.$$

The exact value of the exponent is still being argued about: Kormendy (2001) gives $x = 3.65$, while Merritt & Ferrarese (2001) give $x = 4.80$. I'd take $x \sim 4$ as a reasonable guess for now. About 40 galaxies had been studied as of 2001 – about 2/3 of them by stellar cusps, 1/3 by gas disks – with BH masses extending from a few million to a few billion solar masses. [3]

1.6.2 Techniques: extend to AGN

The techniques just listed don't work for most AGN, due to the very bright nonstellar nucleus which swamps the non-AGN signal. Two other techniques are being developed. I personally don't yet find them as convincing as gas disks and stellar cusps, but they have their adherents and the techniques are becoming more reliable as time passes. They are:

• **Reverberation mapping**. Go back to our cartoon of the emission line clouds close to the AGN. The line widths tell us the velocity of the gas, which we assume is gravitationally bound to the BH. To get its distance, we look at variability. The gas emitting the broad lines is photoionized by the central engine; when the ionizing flux varies, so will the ionization level in the line-emitting gas, *but after a delay due to the light travel time*. Monitoring of both the (ionizing) continuum and

[3] Comment from the author: *this is very striking*, and was totally unexpected. Just about everyone in the field thought that only AGN would have a massive BH in the core. A minor complication was that quasars are much more common at high redshift – as discussed below – so people were aware that there must be some ex-quasars nearby. But few people suspected, before a few years ago, that MDO's would be so very common.

the emission lines can give us the distance of the line-emitting gas from the BH. Combining this with the linewidth gives us the mass of the BH. This has been done for only a handful of AGN so far; the results (as quoted in Merritt & Farrarese 2001) fit nicely on the $M_{bh} - \sigma$ curve determined for non-active bulges.

• **X-ray line profiles**. This attractive idea is still being pursued observationally. The Kα line of iron has now been seen in Xrays in several objects. The data suggest it is very broad ($\lesssim c/3$ linewidth), and has "an asymmetric red wing consistent with gravitational redshift". This is still a very new technique, and needs careful modelling of the accretion flow in order to do anything quantitative. Such work may be coming.

1.6.3 The galactic center: stellar orbits

The center of our galaxy is a special case, because it's so close (and easier to observe), and of course because it's of great personal interest to us. Different techniques can be used here than in external galaxies, because we can see fainter objects and resolve smaller scales. On the other hand, the GC is heavily obscured as seen from here, so we can't do anything optically. Radio, IR and high frequencies (X- and γ rays) are our tools.

As seen in the radio the region is quite a mess – a complex distribution of thermal gas (HII regions), cold molecular gas, and nonthermal emission (SNR and diffuse). Most of this is extraneous to our focus here, namely the existence and size of an MDO in the GC: we need to search for a compact object and/or ordered gas motions. Both things exist ...

• **Streaming gas** can be detected in radio continuum, and its velocity measured with radio recombination lines. Its structure has been called a "mini-spiral", although it is not as ordered as that name would suggest. Its physical extent $\sim 1 - 22$ pc, and its ordered rotation speed $\sim 100 - 200$ km/s. If this gas is in ordered, Keplerian motion it points to a gravitating mass of a few $\times 10^6 M_\odot$. This was the first strong sign, but the uncertainty of the gas orbits and the lack of strong constraints on nuclear stars (could the mass be just a dense star cluster?) kept the skeptical (like your author) from accepting this as a detection of a BH.

• **Sgr A*** has long been known to be an unresolved radio source at what seems to be the dynamical center of the galaxy. This argument is made as follows. The data verify that Sgr A* is at the dynamical center of the streaming gas ring/spiral, and also at the center of the nuclear star cluster (described next). A more indirect argument is it's lack of random motion. VLB monitoring over 16 years showed that it has only the parallax consistent with our motion around the GC; its own space velocity can be no more than 15 km/s. This is so much lower than other random motions that we can infer it's a massive object moving only very slowly. It is unresolved, even at VLB scales[4], making its physical size smaller than ~ 1 AU. It is a compact, variable synchrotron source: a small version of what we find in other AGN. We don't have the resolution at X- or γ-rays to separate its high-frequency emission from the general mess in the GC region; but an old report of an $e^+ - e^-$ annihiliation line, at 0.5 Mev, from the direction of the GC was tantalizing (and unfortunately has never been repeated).

Thus: there is suggestive evidence of a massive, compact thing in the GC. The recent result that tied this down and convinced the skeptics is IR imaging of the central star cluster. Individual stars can be distinguished easily within the central $\sim$ pc-sized star cluster; and 10 years of monitoring (as reported by Schödel etal, in a 23-author paper) allow measurement of the stars' proper motions. This is a great piece of work: the orbits of individual stars were followed long enough to track them through both peri-center and apo-center passage, which allows a good determination of the orbital parameters *and the mass of the central object*. This data fits well with other recent estimates of the mass of the MDO, such as from velocity dispersions; modelling seems to demonstrate robustly that the gravitating mass must be a point mass rather than a dense, but extended, star cluster. The bottom line: the core of our galaxy contains a MDO, almost certainly a BH, of $M_{bh} \sim 3 \times 10^6 M_\odot$.

References

Much of the general discussion is just "off the top of my head". Some useful general references on galaxies are

• Binney & Merrifield, *Galactic Astronomy*

• Elmegreen, *Galaxies & Galactic Structure*

[4] That really means its angular size is small enough to be affected by interstellar scattering, due to the radio waves propagating through the turbulent ISM

• Sparke & Gallagher, *Galaxies in the Universe*

For more detail than you want on stellar dynamics and the structure of galaxies, a very good book is

• Binney & Tremaine, *Galactic Dynamics*

I'll put up AGN references later in the course. For now, chapter 26 of Carroll & Ostlie is a nice introduction.

2 The Interstellar Medium

In the previous chapter we reviewed the gravitational structure of bright (S and E) galaxies. But the stars and dark matter are not all of the story. Each type of galaxy contains gas as well as stars: this is the interstellar medium (ISM). To set the stage, paraphrase from Elmgreen, who's talking specifically about our galaxy[1]: "the ISM is like the ocean of a galaxy, a fluid confined by gravity to a thin layer, and serving as a reservoir for all of the material in stars and planets that will ever form, evolve, and disperse." The physics of this ISM, and its connection to stellar birth and stellar death, will be one of our main applications in this course.

2.1 The diffuse ISM in our galaxy

What I call the *diffuse ISM* is the ISM that is truly "interstellar" – sitting in the potential well of the overall galactic disk, and not immediately involved with stars or star formation regions.

2.1.1 How we oberve the ISM

Our understanding of the physical state of the ISM has been driven by the data: recent work in radio astronomy and high-energy astronomy has dramatically changed the field. What are our current ways of looking at the ISM?

• **Optical:** Gas with temperatures in the range roughly 10^3 to 10^4 K emits both continuum and spectral line optical radiation, and cooler gas can be detected by the absorption lines it produces when there is a continuum optical source behind. We see interstellar absorption lines, stellar reddening, diffuse Hα and other emission lines, and dark dust clouds. In addition "stellar ISM" regions such as HII regions, supernova remnants (SNR), and planetary nebulae emit optically. Galactic dust tends to prevent us seeing through the plane of the disk beyond a kpc or so, aside from special lines of sight.

• **Radio:** the major discovery here was the HI 21 cm line, which we see in absorption and emission. This tracks atomic hydrogen, which is (of necessity) fairly cool ($T < 8000$ K). In addition, the diffuse ISM is a source of synchrotron radiation, which we see in the radio. This latter comes from relativistic electrons (the cosmic ray population) interacting with the galactic magnetic field.

• **Infrared:** another strong radiation source – one of the strongest cooling mechanisms – is IR radiation from dust grains. They absorb starlight and reradiate it at temperatures $\sim 10 - 100$ K. Dust exists in many places throughout the ISM, at several temperatures ("near IR", "far IR", *etc.*)

• **Millimeter:** most of the common molecules have rotational transitions in the mm region (with the notable exception of H_2); mm-wave astronomy is now a powerful tool for studying star formation regions and molecular clouds.

• **Ultraviolet:** this again typically samples only the local region, $\lesssim 1/2$ kpc, due to obscuration. At this band we see hot gas, $10^5 - 10^6$ K, from the so-called "local bubble"; in absorption, there are important transitions of atomic and molecular hydrogen. With the advent of UIT, the UV is also becoming a nice tool for studying the ISM in external galaxies.

• **X- and γ-rays:** the local bubble also emits soft X-rays. In addition, the galactic plane is a source of harder X-rays and γ-rays; some of these come from hot gas (such as in SNR), and the hardest component comes from nuclear reactions between the cosmic rays and the ISM.

• **Plasma propagation:** radio signals are affected by propagation through an ionized plasma. In addition, the ionized plasma can emit thermal bremsstrahlung. We can use the following three means to measure electron densities, path lengths and magnetic field:

$$\begin{aligned} \text{dispersion measure}: &\qquad \mathrm{DM} \propto \int n_e dl \\ \text{emission measure}: &\qquad \mathrm{EM} \propto \int n_e^2 dl \\ \text{rotation measure}: &\qquad \mathrm{RM} \propto \int n_e \mathbf{B} \cdot d\mathbf{l} \end{aligned} \tag{2.1}$$

The first refers to *plasma dispersion*, the fact that the phase speed of an EM wave depends on its frequency. The second measures the emissivity due to bremsstrahlung radiation. The third refers to *Faraday rotation*, the rotation of the plane of polarization due to passage through an ionized, magnetized plasma.

2.1.2 A multi-phase equilibrium?

From such observations, we find that the diffuse ISM is very complex. It comes in several phases, with com-

[1] in Burton, Elmegreen, & Genzel, eds., *The Galactic Interstellar Medium*, SAAS-FEE Advanced Course 21, 1991.

plex spatial structure. Different authors classify the phases differently, and papers on this topic can contain a flurry of acronyms (CNM, WNM, WIM, MOWIM, RWIM, HIM, etc, etc, etc). I summarize as follows.

• **Neutral gas** refers to atomic hydrogen, HI. (Molecular H_2 and other species are found in self-gravitating clouds, thus are discussed in the next section; they are not really part of the diffuse ISM). We now know that the spatial structure of neutral HI is quite complicated. First, there is "cold" HI (seen in absorption) and "warm" HI, still neutral, seen in emission. Heiles estimates the warm HI has $T \gtrsim 1000$ K, $n \sim 0.3$ cm^{-3}

The cold phase comes in sheets, filaments, and bubbles. Early observations – which had only spectral resolution (and so picked up gas at different velocities), not spatial, talked about "interstellar clouds". This terminology is still around, but our cartoon is no longer a raisins-in-a-pudding picture. Typical temperatures are $\sim 10 - 75$ K and typical densities are $\sim 20 - 2500$ cm^{-3}. There seems to be a wide range of "cloud" sizes (*i.e.*, path lengths through clumps or sheets), up to big "clouds" $\gtrsim 100\ M_\odot$.

• **Warm ionized gas** refers to partly to mostly ionized hydrogen. This used to be the "intercloud medium"; then for awhile it was thought to be located on interfaces between the cold HI clouds and the really hot, coronal gas. Now, observations of external galaxies show that there is also a diffuse, extended warm component, occupying filaments, clouds, bubbles and chimneys. We see it mainly in diffuse Hα and other optical lines; earlier work also found it locally in absorption lines.

Heiles argues that these are two separate types of warm, ionized gas. One is the diffuse component, throughout the galactic plane, maintained by photoionization by starlight. Heiles estimates $T \sim 8000$ K and $n \lesssim 0.1$ cm^{-3} for this gas (sometimes called the *Reynolds layer*, after the person who first studied it in depth). The second type is, indeed, the warm interfaces between cold and hot gas. Heiles gives $T \sim 8000$ K and $n \sim 0.1 - 0.4$ cm^{-3}.

• **Hot "coronal" gas** refers to the phase at $T \sim 10^5 - 10^6$K, $n \sim 10^{-3} - 10^{-2}$cm^{-3}. We detect this gas by its X-ray emission, and also some UV lines. It is associated with, but not confined to, the interiors of supernova remnants and "superbubbles" (large, multi-SN complexes).

If we look at all phases, we see that each has the product nT on the order of a few thousand (cm^{-3}K). Thus, within the accuracy of the diverse data, each has the same typical pressure: this product converts to an energy density ~ 1 eV/cm^{-3}. We are seeing three phases in approximate pressure balance with each other.

2.1.3 Other components of the ISM

There are three other significant components of the diffuse ISM.

• **Dust grains** tie up about 1% of the mass of the ISM, including most of the heavy elements. They are often found with, or comprised of, complex organic molecules. Polycyclic aromatic hydrocarbons (PAH's, the stuff of soot) are thought to be one major group. Dust absorbs and scatters optical starlight very effectively (this is why we can't see vary far optically) and reradiates it in the infrared, providing a major part of the bolometric luminosity of the galaxy.

• **Cosmic rays** are relativistic particles, electrons and ions, tied to the galaxy by its magnetic field. We detect them directly at the earth though their distribution is modified by passage through the solar wind and the earth's atmosphere. Their detected energies range from $\sim 10^{10}$ eV to $\sim 10^{21}$ eV. We also detect them indirectly by their diffuse synchrotron emission (in the galactic magnetic field) and by their high-energy photon emission (arising from nuclear reactions with the thermal ISM). Their energy density locally is comparable to that of the ISM gas, ~ 1 eV/cm^{-3}, and may increase by a factor of a few going inwards towards the center of the galaxy.

• **the Magnetic field** of the galaxy is detected through Faraday rotation, synchrotron emission, optical polarization of starlight (due to alignment of IS grains across $\mathbf{B}$) and Zeeman splitting. It is fairly ordered, with field lines tending to lie in the galactic plane, and somewhat along the spiral arms. Typical field strength is usually quoted as $\sim 3\mu$G, with strong spatial variation, and higher fields locally in some regions. Its energy density locally is also – you guessed it – $\gtrsim$ 1 eV/cm^{-3}. As with cosmic rays, there is some suggestion that the energy density in the field rises going towards the galactic center.

2.2 "Star stuff" in our galaxy

In addition to the diffuse medium, many well-known examples of so-called interstellar matter come from ISM closely related to stars. In particular, any of the

prettiest pictures come from nebulae associated with young stars or old, dying stars. They include:

• **HII regions** are regions of ionized hydrogen surrounding, and ionized by, hot young stars. They are strong optical and radio sources.

• **Planetary nebulae** are the outer layers of older stars, ejected by instabilities in the star's structure.

• **Stellar winds** are produced by nearly all stars at some level; hot, young stars have by far the strongest winds. While not as spectacular as the first two, winds are strong sources of mass and energy supply for the ISM.

• **Stellar jets** are produced both by young stars, in the formation process, and by compact stellar remnants (*e.g.*, neutron stars and their surrounding accretion disks). They are striking radio and optical sources when we can catch them; a few have apparent superluminal expansion velocities.

• **Supernovae remnants** result when massive stars eject something like half their mass, or more, in a violent explosion. We see SNR as strong radio, optical and X-ray sources; in addition they are thought to be strong sources of cosmic rays. They are also important contributors to the energy and mass budget of the ISM.

These nebulae all share the property that they are overpressured relative to the diffuse ISM: they arise from and are tied to stars. In addition, molecular gas is found in overpressured, self gravitating clumps called

• **Molecular Clouds**. These are detected in "trace" heavy-element molecules such as CO, which have strong millimeter line transitions; we infer that H_2 is also present, in larger quantites. These clouds can be very cold, $T \sim 30 - 100$K, and very big, up to $\sim 10^5 - 10^6 M_\odot$. They have line widths much greater than the Doppler width one would expect from their temperatures: thus they are either collapsing (under their own self-gravity) or supported by turbulent, disordered internal motions. They are the stellar nurseries of the galaxy, the sites of ongoing star formation.

2.3 Galactic ecology

Taking the galaxy as a whole, we must keep in mind that the ISM is not static. It is continually being consumed in new star formation (at a rate $\sim 3 - 10 M_\odot$/yr over the galaxy) and continually being replenished by stellar ejecta (everything from winds to supernovae). It loses energy by radiation, over all wavelengths, and gains energy as it is replenished by the stellar ejecta. Thus, we can think of stars as simply long-lived phases of the ISM; they are formed from it and they return to it.

However, there is a trend: not all stellar material is returned to the ISM. Low-mass stars become white dwarfs, and quietly cool forever; high-mass stars recycle much of their mass, but most are believed to leave compact cores, neutron stars or black holes. Thus, the overall trend of the system is towards cold, dense ex-stars. This will take awhile, however; at present we find quite a mixture of stellar and diffuse matter in the galaxy.

2.4 The ISM in Ellipticals

Our understanding of ellipticals (the stars as well as the ISM) has changed dramatically over the past couple of decades. Originally it was thought that these galaxies have no ISM. The older stellar population characteristic of E's, and the lack of strong star formation, also seemed to argue against any interstellar matter. Then people started looking...and it now appears that these galaxies also have a complex, multi-phase ISM, with total mass comparable to that in spirals. Unlike spirals, however, the ISM in E's tends to be hotter, that is with smaller amounts in cool, neutral or "warm" phases, and most of the gas being in the hot phase. This is still a new and evolving field, being limited both by detector sensitivity and by amount of effort (not as many people look at these faint, distant things). In these notes I summarize the current state of things.

2.4.1 Everything that isn't the hot phase

Some – not all – E's have now been shown to have neutral hydrogen, molecular gas, dust grains, and a "warm" ISM (the latter detected in optical emission lines, placing it at $\sim 10^4$K). Some have been shown to have onging star formation in their cores. Because a smaller fraction of the total ISM is "cool", its signal is faint; a non-detection does not necessarily mean there is no cool gas in a particular galaxy. On the other hand, there seems to be an ongoing argument as to whether galaxies detected strongly in, say, HI or CO are "typical" ("real ellipticals don't eat quiche," to quote Rupen[2]). Some authors argue that only unusual, or disturbed, E's contain detectable amounts of cool gas...thereby defining a "pure E" as one without such

[2] M. Rupen, private communication, a few years back.

gas. Given sensitivity limitations, I tend to think this view may eventually be proved wrong. But I agree that we do not, yet, understand the multi-phase cool gas in the ISM of an elliptical.

2.4.2 The hot phase – the x-ray loud gas

Our understanding of the ISM in E's started to change in 1985, when the Einstein X-ray observatory discovered that normal E's are strong X-ray sources. (This has since been explored in detail by ROSAT, and now CHANDRA is adding to the picture). "Hot" means at temperatures to emit X-ray bremsstrahlung: $T \sim 10^7$K. (We can estimate the temperature roughly from the fact that the gas is an X-ray source, and more specifically from X-ray emission lines). The spatial distribution of the gas is consistent with it sitting in hydrostatic equilibrium in the potential well of the galaxy. The amounts are large: Sarazin gives $10^9 - 10^{11} M_\odot$ in gas, or comparing to the optical (blue) light of the galaxy, $M_{gas}/M_\odot \simeq 0.2(L_B/L_{sun})$. This is most of the ISM; the fractions estimated in all of the cool components are much smaller.

Why is so much of the ISM hot in an elliptical? Think about the ecology of this ISM. As in a spiral, the gas is ejected from stars, by the usual mass-loss processes (stellar winds and SNe). Unlike a spiral, however, the stars have quite high random motions (200-300 km/s is typical for gravitational support). It follows that collisions between the ejected gas, and either gas ejected by nearby stars or the local ambient ISM will thermalize the kinetic energy of the injected gas. (This is a direct application of shock physics — which we'll see later in the course). This means the gas is effectively injected with at least $T \sim m\sigma^2/k \sim 7 \times 10^7 K(\sigma/300\ \mathrm{km\,s^{-1}})^2$ (any larger injection velocities, due to winds or SN flows, will only raise this number). We can also note that the radiative cooling from this hot gas will be less effective than from the cooler, denser gas of a spiral galaxy. The details of why this is so must wait until later in the course; it has to do with lower density (the same amount of gas spread over a larger volume) and strongly ionized gas at these hot temperatures (so emission lines, which are strong coolants, aren't as important).

References

Much of the general discussion is just "off the top of my head". Useful recent references come from meeting proceedings:

• Arnaboldi, Da Costa, & Saha, eds, 1996, *The Second Stromlo Symposium: the Nature of Elliptical Galaxies*. Several articles; especially that by Sarazin on the X-ray loud gas in E's.

• Duric, 1999, in *New Perspectives on the ISM*, eds. Taylor, Landecker & Jones; p. 161 – on cosmic rays and galactic B field.

• Sadler, 2002, in Da Costa & Jerjen, eds., *The Dynamics, Structure & History of Galaxies*, p. 215.

• Woodward, Bicay & Shull, eds, 2001, *Tetons 4: Galactic Structure, Stars, & the ISM*; especially the articles by Heiles & Dickey.

3 Some Radiation Basics

In this chapter I'll store some basic tools we need for working with radiation astrophysically. This material comes directly from Rybicki & Lightman ("RL"), where you can find a more complete discussion of it all.

Our apporoach will be ray optics. You remember that radiation can be approached as EM waves, as discrete photons, or in the 'ray optics' limit – we're thinking in terms of ray optics here. Some of this material will look pretty dry to you ... as you go along, look for these important quantities, which will be useful tools for us later.

Intensity, I_ν (equation 3.1)

Flux, F_ν (equation 3.2)

Energy density, u_ν (equation 3.4)

Intensity in thermal equilibrium (TE), B_ν (equation 3.17)

Emission coefficient (per solid angle), j_ν (equation 3.23)

Emissivity, $\epsilon_\nu \rightarrow 4\pi j_\nu$ (equation 3.24)

Absorption coefficient, κ_ν (per solid angle; equation 3.26)

Source function, $S_\nu = j_\nu/\kappa_\nu$ (equation 3.29)

Opacity, τ_ν (equation 3.27)

OK, here goes.

3.1 Radiation: some important definitions

We begin with some important definitions.

• **Intensity**. Consider a little (differential) area $d\mathbf{A} = dA\ \hat{\mathbf{n}}$, with a radiation beam passing through it. At a particular frequency ν, the energy passing through $d\mathbf{A}$ in a particular direction (θ, ϕ) (measured relative to $\hat{\mathbf{n}}$), per frequency range $d\nu$, per time dt, and per solid angle $d\Omega$, is given by

$$dE = I_\nu\ dA\ dt\ d\nu\ d\Omega. \tag{3.1}$$

This relation implicitly defines our basic quantity, the **intensity**: I_ν[1] Most of our other quantities are defined in terms of I_ν. Intensity is also called specific intensity, brightness or surface brightness. In cgs, its units look like erg cm^{-2} Hz^{-1} s^{-1} str^{-1}.

• **Flux** is the net energy passing through dA, in all directions:

$$F_\nu = \int I_\nu \cos\theta\ d(\cos\theta)\ d\phi \tag{3.2}$$

(with units erg cm^{-2} s^{-1} Hz^{-1} or W m^{-2} Hz^{-1}).

> • **Heads up here:** Flux and intensity are similar but not the same; you'll need to be able to work with both. There is some detailed discussion in Appendix I to this chapter. Basically, intensity is what is shown in an image of a source and flux is what you get if you integrate everything in the image.

• One can also define a **mean intensity**, averaged over all solid angles:

$$J_\nu = \frac{1}{4\pi}\int I_\nu\ d(\cos\theta)\ d\phi \tag{3.3}$$

• The **energy density** is clearly related to the intensity by a factor of lightspeed. The most useful definition is in terms of the angular mean.

$$u_\nu = \frac{4\pi}{c}J_\nu = \frac{1}{c}\int I_\nu\ d(\cos\theta)\ d\phi \tag{3.4}$$

The units are erg cm^{-3} Hz^{-1} (of course).

• **Frequency integrated.** You should also note that each of the quantities above can be integrated over frequency:

$$F = \int F_\nu d\nu\ ; \quad I = \int I_\nu d\nu\ ; \quad u = \int u_\nu d\nu \tag{3.5}$$

and so on.

3.2 Thermal equilibrium: an ideal gas

You've probably seen this elsewhere; I'll just review the basics here. Consider a small system – say, one atom in a gas – which is in *thermal contact* with a large system. That means that energy exchange is allowed, most likely through collisions with the rest of the gas. Let the big system – the so-called "reservoir"– have a temperature T. The fundamental result of classical

[1] NOTATION ALERT: I_ν is traditional notation in this field, and means "I is a function of ν". So, $I(\nu) \leftrightarrow I_\nu$, and ditto for j_ν, κ_ν, etc.

thermodynamics is that the probability of finding the small (test) system in a state of energy E is

$$\mathcal{P}(E) \propto e^{-E/k_BT} \tag{3.6}$$

This is the *Boltzmann factor*; the proportionality constant is used to normalize the probability, in a specific system (as, below).

Now, let's apply this to an ideal, monatomic gas. Let the test system be a single atom in the gas, and let the rest of the gas be the reservoir, at T. Each atom has a mass, m, and a random velocity, $\mathbf{v}$; the energy associated with this velocity is $E(\mathbf{v}) = \frac{1}{2}mv^2$. (This could of course describe a plasma as well as a neutral, atomic gas). The probability that a particle has energy E is then

$$\mathcal{P}(\mathbf{v}) \propto e^{-mv^2/2k_BT} \tag{3.7}$$

We want to use this to derive the distribution of particle velocities. In addition to the Boltzmann factor, we need to know the number of ways in which a particle of velocity $\mathbf{v}$ can have $E(\mathbf{v})$. If the gas is isotropic – if there are no restrictions on the possible orientation of the velocity vector – then the factor which weights the Boltzmann factor is just the number of possible directions the velocity vector can point. That is, $\mathcal{P}(v) \propto 4\pi v^2 e^{-mv^2/2k_BT}$. Now,if we normalize the answer to the total number density, so that

$$n = \int_{\mathbf{v}} \mathcal{P}(\mathbf{v}) d^3\mathbf{v} = \int_0^\infty f(v) dv \tag{3.8}$$

defines the *distribution function*, $f(v)$, we end up with

$$f(v) = 4\pi n \left(\frac{m}{2\pi k_BT}\right)^{3/2} v^2 e^{-mv^2/2k_BT} \tag{3.9}$$

which is (one way to write) the *Maxwell-Boltzmann distribution* of particle velocities.

Taking means (or moments) of this distribution, we find the mean particle velocity,

$$\langle v \rangle = \int_0^\infty v f(v) dv = \left(\frac{8k_BT}{\pi m}\right)^{1/2} \tag{3.10}$$

and

$$\langle E \rangle = \int_0^\infty \frac{1}{2}mv^2 f(v) dv = \frac{3}{2}k_BT \tag{3.11}$$

We can extend this type of analysis to find the pressure of the gas. For a simple, MB distribution, this recovers the usual ideal gas law:

$$p = nk_BT \tag{3.12}$$

This analysis can also be used to get the pressure of a relativistic ideal gas, or of a photon gas. I've put the details in Appendix II to this chapter.

3.3 Thermal equilibrium: radiation

Here, we consider a photon gas which is in thermal contact with something at temperature T. That "something" could be the walls of a closed box (the proverbial black body), or a dense gas cloud (which could be a dark interstellar cloud, or a star). As with a particle gas, thermal contact means the photons exchange energy with the "something" through collisions; here, this could mean particle-photon scattering, such as Thompson scattering of photons on electrons; or it could mean that the matter absorbs and re-emits photons. Now, in particle-particle collisions the particle number is conserved (in standard, elastic collisions, anyway). In photon-matter "collisions", on the other hand, the photon number is not conserved; it is quite possible for an atom to absorb one photon and re-emit several, still while conserving energy (for instance, the atom could absorb to the $n = 10$ level and re-emit $10 \to 8$, $8 \to 5$ and $5 \to 1$ photons as it decayed). Thus, in dealing with the particle gas, we normalized the probability function by assuming a certain total number density of particles. For radiation in TE, the number density of photons is *predicted* if we know T.

The analog of the Boltzmann factor for a photon gas is the Planck distribution: the probability of finding a photon at energy $E = h\nu$ is

$$\mathcal{P}(E) \propto \frac{1}{e^{E/k_BT} - 1} \tag{3.13}$$

which, of course, resembles the classical (non-quantum mechanical) Boltzmann factor for energies $E >> k_BT$. We find the density of states allowed at energy E just as we did for particles, by looking at the number of ways photons of wavevector $\mathbf{k} = \mathbf{p}/\hbar = E/\hbar c = 2\pi\nu/c = 2\pi/\lambda$ can be put into a volume. But, recalling standing waves, we remember that a one-dimensional box of length ℓ can contain standing waves of wavenumber $k = 2\pi q/\ell$ if q is any integer. So, the number of photon states in three dimensions in $(\mathbf{k}, \mathbf{k} + d\mathbf{k})$ is

$$d^3\mathbf{q} = \frac{\ell^3}{8\pi^3} 4\pi k^2 dk \tag{3.14}$$

From this, we find the density of states per volume by dividing by ℓ^3 and adding the usual factor 2 for the two

spin (polarization) states, and express things in terms of ν rather than k:

$$\text{density of states} = 8\pi\frac{\nu^2}{c^3}d\nu \qquad (3.15)$$

This is the factor which weights $\mathcal{P}(E)$ (from equation 3.6) to find the density of photons at energy $E = h\nu$. (Note that this is 4π times larger than RL's expression for ρ_s; they are working in density of states per solid angle.) However, it is common to multiply this density by $h\nu$ to find the energy density of radiation in TE at T:

$$u_{rad}(\nu, T)d\nu = \frac{8\pi\nu^2}{c^3}\frac{h\nu}{e^{h\nu/k_BT} - 1}d\nu \qquad (3.16)$$

(In dealing with black body radiation, watch out for units; different authors do different things. u_{rad} in equation (3.16) has units energy/volume-Hz; some authors divide by 4π to get energy/steradian-volume-Hz.)

Another way to express this result is in terms of the energy in radiation crossing a unit area per Hz per time (and usually per steradian). As we saw above, this quantity is the *intensity*, and is related to the energy density in radiation by $I(\nu) = cu_{rad}(\nu)/4\pi$. For (and only for) the specific case of Black Body radiation, the intensity is denoted $B(\nu, T)$ or $B_\nu(T)$, rather than I_ν. For black body radiation, we therefore have

$$B(\nu, T) = B_\nu(T) = \frac{2h\nu^3}{c^2}\frac{1}{e^{h\nu/k_BT} - 1} \qquad (3.17)$$

with units, energy/area-time-steradian-Hz. If you are working with wavelengths rather than frequencies, the analogous version is

$$B(\lambda, T) = B_\lambda(T) = \frac{2hc^2}{\lambda^5}\frac{1}{e^{hc/\lambda k_BT} - 1}$$

and it will give you an intensity in units of energy/area-time-steradian-wavelength.

Equations (3.16) and (3.17) are the basic result describing radiation in TE. However, several extensions and approximations are standard. First, we can integrate them over frequency to find the total energy (per volume, or per area per time):

$$\begin{aligned} u_{rad}(T) &= \int_0^\infty u_{rad}(\nu, T)d\nu \\ &= \frac{4\pi}{c}\int B_\nu(T)d\nu = aT^4 \end{aligned} \qquad (3.18)$$

where the constant $a = 8\pi^5 k_B^4/15c^3h^3 = 7.56 \times 10^{-15}$ erg/cm^{3}deg^4. In addition, we have

$$B(T) = \int_0^\infty B_\nu(T)d\nu = \frac{ac}{4\pi}T^4 \qquad (3.19)$$

and, finally, the emergent flux obeys $F = \pi B(T)$, so that

$$F(T) = \int F_\nu d\nu = \pi\int B_\nu d\nu = \sigma_{SB}T^4 \qquad (3.20)$$

where $\sigma_{SB} = ac/4 = 5.67 \times 10^{-5}$erg/cm^{2}deg^{4}s^{-1}.

Finally, a couple of limits of $B_\nu(T)$ are worth noting. For high frequencies with $h\nu \gg k_BT$, we have the Wien limit:

$$B_\nu(T) \simeq \frac{2h\nu^3}{c^2}e^{-h\nu/k_BT} \qquad (3.21)$$

which recovers the exponential form. For low frequencies $h\nu \ll k_BT$, we have the Rayleigh-Jeans limit:

$$B_\nu(T) \simeq \frac{2\nu^2}{c^2}k_BT = \frac{2}{\lambda^2}k_BT \qquad (3.22)$$

which is very frequently used as an approximation to the black body function in the radio part of the spectrum.

3.4 Radiative transfer

Start with a beam of radiation, described as usual by intensity I_ν. Consider such a beam, from some background source, hitting a slab of material. It will be absorbed by the material as it passes, and the material may well also emit radiation into the beam.

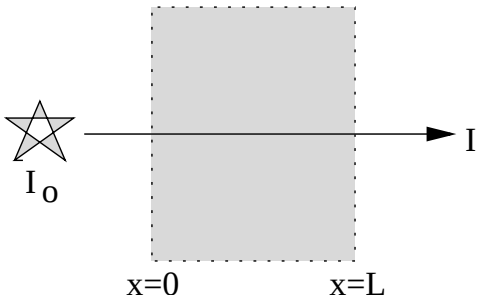

Figure 3.1 Geometry of radiation transfer through a slab of material, possibly with a background source I_o.

3.4.1 More definitions

We need more definitions now. In addition to the intensity, I_ν, we need to think about the plasma through which the beam passes.

• The plasma has an **emission coefficient** j_ν, defined in terms of the contribution to a radiation beam (I_ν as it propagates a distance dx:

$$dI_\nu = j_\nu dx \tag{3.23}$$

and the units of j_ν are erg cm^{-3} s^{-1} Hz^{-1} str^{-1}. This is the fundamental radiated power, at frequency ν, from the matter; its details depend on the local physics. Note also that the mean intenstiy J_ν that we saw earlier is not the same as the emission coefficient j_ν.

• For some problems it's more useful to work with the **volume emissivity**,

$$\epsilon_\nu = 4\pi j_\nu \tag{3.24}$$

(This assumes isotropic emission). We'll run into this later.

• We also need the **absorption coefficient** κ_ν,[2] which describes the fractional absorption or scattering of a radiation beam, per unit length dx:

$$dI_\nu = -\kappa_\nu I_\nu dx \tag{3.25}$$

This has units cm^{-1}. It can also be written microscopically (to reveal the physics), in terms of the number density of absorbers n and their absorption cross section σ_ν: $\kappa_\nu = n\sigma_\nu$. Question for the reader: How, then, is the absorption coefficient related to the mean free path of a photon?

3.4.2 Transfer analysis

With these definitions, the basic transfer equation can be written down,

$$\frac{dI_\nu}{dx} = j_\nu - \kappa_\nu I_\nu \tag{3.26}$$

Before solving this, we introduce an important and useful quantity, the *optical depth*:

$$\tau_\nu = \int_o^\ell \kappa_\nu dx \tag{3.27}$$

where the integral is taken from back to front through the slab of matter. From the discussion of κ_ν, above, we see that $\tau_\nu = \ell/\lambda$ measures the number of absorption mean free paths through the source. We would expect, then, that a system with $\tau_\nu \ll 1$ would have little effect on any source behind it (that is, it would be nearly transparent), and a system with $\tau_\nu \gg 1$ would be nearly opaque, absorbing most of the light. We can rewrite (3.26) with τ as the independent variable:

$$\frac{dI_\nu}{d\tau_\nu} = S_\nu - I_\nu \tag{3.28}$$

where we have defined the *source function*,

$$S_\nu = \frac{j_\nu}{\kappa_\nu} \tag{3.29}$$

Now, solve (3.28). If we put a source of intensity I_o behind the slab, the formal solution (remember integrating factors?) is

$$I_\nu(\tau_\nu) = I_o e^{-\tau_\nu} + \int_0^{\tau_\nu} S_\nu(\tau') e^{-(\tau_\nu - \tau')} d\tau' \tag{3.30}$$

Look at this: the first term is simply the attenuation of the background source by the slab. The variable τ' is a distance through the slab, but it's measured in dimensionless optical depth units. The second term describes the emission of radiation from within the slab, at position τ', and the attenuation of this radiation by the smaller optical depth, $\tau_\nu - \tau'$, between the emission point and the front of the slab.

3.4.3 Optically thick and thin limits

In the important case of a homogeneous source, with $I_o = 0$, (3.30) simplifies to

$$I_\nu(\tau_\nu) = S_\nu \left(1 - e^{-\tau_\nu}\right) \tag{3.31}$$

describing emission only from the cloud/slab itself. This has two important limits:

• *Optically thin*, $\tau \ll 1$: we see

$$I_\nu \simeq S_\nu \tau_\nu = j_\nu \ell \tag{3.32}$$

This limit just integrates the emissivity through the cloud, without modifying it by internal absorption.

• *Optically thick*, $\tau \gg 1$:

$$I_\nu \simeq S_\nu \tag{3.33}$$

In this limit, the emergent intensity is just equal to the source function. NOTE this may be quite different from the internal emissivity; transfer through the source has modified the spectrum.

From the optically thick limit we can make a couple of important connections.

[2] NOTATION ALERT: about half the literature uses α_ν for the absorption coefficient; the other half uses κ_ν, as I do here.

• Consider the case when radiation is in TE with the local plasma. This means $I_\nu \to B_\nu$ (the Planck function, (3.17 or its limits); and also $\tau_\nu \gg 1$ (because you need lots of collisions, *it i.e.* optically thick, to gain TE). Thus, from (3.31), we expect $S_\nu \simeq B_\nu$ (the source function approaches the Planck function), and from this we derive an important relation:

$$j_\nu \simeq B_\nu \kappa_\nu \tag{3.34}$$

This is called **Kirchoff's law**. It says: *if we can assume thermal equilibrium* (as we will do in chapters 3 and 4), the emissivity and absorption are related through the Planck function.

• Astronomers working at radio frequencies commonly quote the intensity, I_ν, in terms of the **Brightness Temperature**, T_B, defined by

$$I_\nu = 2\frac{\nu^2}{c^2} k_B T_B \tag{3.35}$$

But from comparing (3.22) and (3.35), we find that $T_B \to T$ as the source becomes optically thick: the brightness temperature approaches the physical temperature. (Question for you: if the source is optically thin, how are T and T_B related? How does your answer depend on the conditions in the source?)

3.5 Appendix I: some examples with intensity

Working with intensity, flux, etc, can be confusing (at least to your author!); so I'm putting some specific examples here – mostly directly from RL.

3.5.1 Isotropic radiation field

If the radiation field is isotropic life is simple: $J_\nu = I_\nu$ (is that obvious?) and $F_\nu = 0$ (there's no net energy flow; there's as much going "out" as going "in"). Also, by inspection of (3.4), we see that $u_\nu = 4\pi I_\nu / c$.

3.5.2 Intensity is constant along a ray

We should note one important fact: in the absence of absorption or emission, the intensity I_ν is constant along any ray. RL present one (rather formal) derivation of this, illustrated in Fig 3.2, which I summarize here. The key points are that intensity is defined per solid angle, and that energy is conserved. Think about the set of rays passing through both dA_1 and dA_2. The energy in that set of rays is

$$dE = I_1\, dA_1\, dt\, d\Omega_1\, d\nu = I_2\, dA_2\, dt\, d\Omega_2\, d\nu \tag{3.36}$$

But, thanks to the inverse square law, $d\Omega_1 = dA_2/R^2$, and $d\Omega_2 = dA_1/R^2$. Thus, because the same dE passes through both little areas, we must have $I_1 = I_2$. Q.E.D.

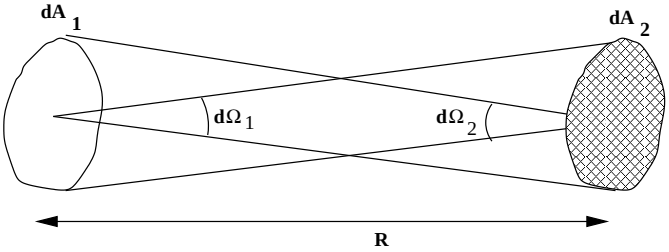

Figure 3.2 One way to establish the constancy of I along a ray. Two little areas, dA_1 and dA_2, are separated by R. dA_1 subtends a solid angle $d\Omega_2$, as seen by 2; and vice versa for dA_2 as seen by 1. Following RL Fig 1.5.

Wait .. does this seem unphysical? What about the inverse-square law that we know applies to radiated power? The key is that intensity is not the same as flux – and flux satisfies the $1/R^2$ law. This is contained in (3.36), because the solid angle $d\Omega = dA/R^2$; so that the energy *per area* passing through any dA falls off $\propto 1/R^2$.

Or, if this argument isn't very transparent, you might prefer the next example.

3.5.3 Flux from a sphere

Here's a nice example, to verify that radiative flux does indeed obey the inverse square law. Put yourself at distance D from a sphere of uniform brightness B; that means all rays leaving the surface have the same intensity, $I = I_o$, independent of direction. (The geometry is shown in Figure 3.3).

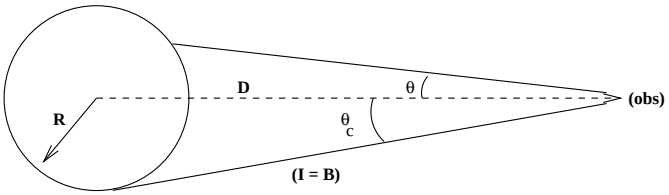

Figure 3.3 The geometry used for calculating the flux from a uniformly bright sphere. The observer is at a distance D away; the sphere radius is R; the radius subtends an angle θ_c as seen by the observer; the intensity is assumed uniform, $I = I_o$, along every ray that leaves the surface. Following RL Fig 1.6.

The intensity you see, then, is $I = I_o$ from rays which intersect the sphere; and $I = 0$ from other angles. The

flux you observe is thus

$$F = \int_0^{2\pi} d\phi \int_0^{\theta_c} I(\theta,\phi)\cos\theta d\cos\theta \qquad (3.37)$$

But this is easy to integrate:[3]

$$F = \pi I_o \left(1 - \cos^2\theta_c\right) = \pi I_o \sin^2\theta_c \qquad (3.38)$$

This nicely recovers the inverse square law as long as the distances involved are not cosmological:

$$F = \pi I_o \frac{R^2}{D^2} \qquad (3.39)$$

for a *uniform* source of circular cross-section. Also, note that at $R = D$ (when you're right at the surface of the star), the flux is

$$F(D = R) = \pi I_o \qquad (3.40)$$

We can also invert this solution. Say we observe flux F at earth.[4] The intensity at the source is, clearly,

$$I_\nu = \frac{F_\nu}{\pi}\frac{D^2}{R^2} \qquad (3.41)$$

But also, remember $\theta_c = \sin^{-1}(R/D) \simeq R/D$ (this last for a distance source); so the solid angle subtended by a distant source is $\Omega_c = \pi\theta_c^2$. Thus, we can go from the flux (at earth) to the intensity (at the source) by

$$I_\nu = \frac{F_\nu}{\Omega_c} \qquad (3.42)$$

3.6 Appendix II: more on pressure

3.6.1 Ideal gases: the pressure integral

We can also use the MB distribution to find the pressure of an ideal gas. (We will derive this is a fairly formal way, to use later on). (For this subsection, we use p for the single-particle momentum, and P for the pressure). The pressure is, of course, the force exerted per unit area from collisions of the gas particles. Consider some "test surface" within the gas, and we can find this force. One particle, with momentum $\mathbf{p}$, approaches this surface at angle θ. When it recoils from the surface, it transfers momentum $\Delta p = 2p\cos\theta$ to the surface. Now, the rate of particles approaching at this $\mathbf{p}$ and this θ is

$$j(p,\theta) = v(p)\cos\theta\frac{1}{2}f(p)\sin\theta$$

(that is, the normal velocity times the number of particles "at θ"; and noting that only half of the particles "at θ" are approaching the surface). The pressure, then, is this rate times the Δp per collision, integrated over all angles and all velocities:

$$P = \int_0^{\pi/2} d\theta \int_0^\infty dp 2p\cos\theta j(p,\theta) \qquad (3.43)$$

If we now assume the gas is isotropic, we can do the θ integral right away, and we end up with

$$P = \frac{1}{3}\int_0^\infty pv(p)f(p)dp \qquad (3.44)$$

If we put in the MB distribution, from (3.9), and assume a subrelativistic gas, so that $v(p) = p/2m$, we find $P \propto \int_0^\infty p^2 e^{-p^2/2mk_BT}dp$, and end up with

$$P = nk_BT \qquad (3.45)$$

as we expect.

Another interesting limit is that of a relativistic ideal gas, in which the single particle energy $E \gg mc^2$. In this limit, we have $p \simeq E/c$ and $v \simeq c$, so that

$$P \simeq \frac{1}{3}\int Ef(E)dE \qquad (3.46)$$

(where we have used $f(E)dE = f(p)dp$). But the integral is just the energy density of the gas, u; so we have

$$P = \frac{1}{3}u \qquad (3.47)$$

which is a general result for an internally relativistic gas (whether of particles or of photons).

3.6.2 Radiation pressure

Although not limited to radiation in TE, this is as good a place as any to put the basic facts about radiation pressure. For a general radiation field, we can use the general equation (3.43), with the photon flux written as

$$j(\nu,\theta) = \frac{c}{2h\nu}\cos\theta\, u_{rad}(\nu,\theta)\sin\theta \qquad (3.48)$$

[3] Just to be more difficult ... some authors (*e.g. Mihalas* "absorb the π (in 3.39) into the definition of flux", and call it "astrophysical flux". I think the moral is, be careful when you go from author to author – JAE.

[4] NOTATION ALERT: the flux at earth is often called S or S_ν, at least in radio astronomy.

To evaulate P_{rad}, we would have to know the angular distribution of u_{rad} (for instance, the basic radiation-pressure applications, such as an astronaut with a flashlight, or a spaceship with a light sail near the sun, assume a highly directed u_{rad}).

One simple limit is the case of an isotropic radiation field (which is a good description of a radiation field in which the photons scatter, such as the interior of a star – that is, a photon field which is probably also in TE). Here, we can start with equation (3.44), and we can certainly use the relativistic limit; thus, we get

$$P_{rad} = \frac{1}{3} u_{rad} \tag{3.49}$$

in general, for an isotropic radiation field.

The other simple limit is the case of a unidirectional radiation field – for instance radiation from the sun as seen at the distance of the earth. We could more properly talk about *radiation force* here: dimensionally that's the (radiation power)×(surface area of the absorber)/c. One application of this is the luminosity at which the radiation pressure (or force) from some object of mass M balances its gravity. This is the *Eddington luminosity*. If we're talking about ionized gas for which Thomson scattering (cross section σ_T) dominates, the Eddington luminosity is $L_{edd} = 4\pi c G M m_p / \sigma_T$ (which you remember from last term).

References

The discussion about ideal gas laws, pressure integrals, etc, can be found in any basic statistical mechanics book – two good ones are

• Reif, *Statistical and Thermal Physics*; Kittel, *Thermal Physics*.

The material on radiation comes straight from one of the fundamental references in the field,

• Rybicki & Lightman, *Radiation Processes in Astrophysics*

but also

• Mihalas, *Stellar Atmospheres*, has a good discussion of the intensity basics.

Key points

• Basic definitions, I_ν; F_ν; u_ν, and what they mean;

• Radiation in TE: Black body physics

• Radiative transfer: j_ν, κ_ν, and τ_ν: what they are, what they mean.

• Radiative transfer: solutions to $I_\nu(\tau_\nu)$, optically thick and thin limits.

• Brightness temperature and the approach to TE (as τ_ν gets big).

4 Bremsstrahlung radiation

Bremsstrahlung arises when a free electron is accelerated in the field of an ion – hence the English name "free-free," representing the transition between two unbound electronic states. The German name "braking radiation" refers to the acceleration of the electron.

4.1 Some basic tools

Before we start bremsstrahlung *per se*, we need to introduce two important general tools used for general analysis of astrophysical radiation.

4.1.1 Power; Larmor formula

You probably remember that if you shake an electron, it will radiate E&M waves. We want to connect the total power in the E&M radiation to "how hard the electron was shaken".

The formal result is that a charge e, which feels an acceleration $\mathbf{a}(t)$, radiates a power (erg s^{-1}) given by

$$\text{cgs} : \qquad P(t) = \frac{2}{3}\frac{e^2}{c^3}|\mathbf{a}(t)|^2 \tag{4.1}$$

Note this is cgs.[1]

To derive this, you need to work out the $\mathbf{E}$ and $\mathbf{B}$ fields which are produced by the accelerated charge; then fold them together into the Poynting flux, $\mathbf{S} = c\mathbf{E}\times\mathbf{B}/4\pi$.[2] That gives us the energy per unit area per unit time carried away from the particle by the fields, *i.e.* the radiated power. Griffiths has a good derivation in chapter 11; Rybicki & Lightman have a more terse derivation in chapter 3. You should note that this formula holds for non-relativistic motion; we'll extend it to the relativistic case, later.

4.1.2 Spectrum: Fourier analysis

The other basic tool we need is the **spectrum** of the radiation. When our electron is shaken, it emits a pulse of radiation, which has a finite duration. This pulse is a wave packet – a superposition of E&M waves of various frequencies. The pulse width in time, Δt, is related to the range of frequencies in the packet, $\delta\nu$, by the usual uncertainty principle: $\Delta t\Delta\nu \sim O(1)$ (as in Figure 4.1). The amplitude of frequency component ν gives what we'll identify as the radiation spectrum.

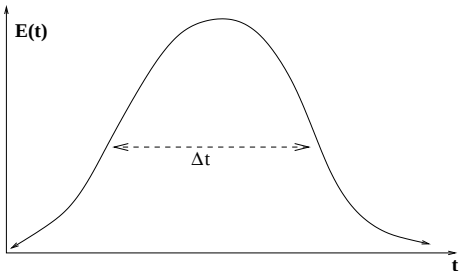

Figure 4.1 Remembering the wave content of a wave packet. The time duration of this packet $\sim \Delta t$ (note this is estimated "by eye" here); it contains frequencies $\lesssim 1/\Delta t$.

To get there formally, we use the Fourier transform ("FT"). I take this from chapter 2 of Rybicki & Lightman. Think about our pulse of radiation; let the electric field in the pulse have some time behavior, $\mathbf{E}(t)$. We can, of course, consider the Fourier transform of this, and its inverse (I'll drop vectors to simplify the notation):

$$\begin{aligned}\hat{E}(\omega) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} E(t)e^{i\omega t}dt \\ \Leftrightarrow \quad E(t) &= \int_{-\infty}^{\infty}\hat{E}(\omega)e^{-i\omega t}d\omega\end{aligned} \tag{4.2}$$

Two FT facts will be useful. First, because $E(t)$ is real, we know that

$$\hat{E}(-\omega) = \hat{E}^*(\omega)\ ;\ \ |\hat{E}(-\omega)|^2 = |\hat{E}(\omega)|^2 \tag{4.3}$$

(that is, the negative frequencies contain no new information). Second, a general result from Fourier transforms (Parseval's theorem) tells us that

$$\begin{aligned}\int_{-\infty}^{\infty} E(t)^2 dt &= 2\pi\int_{-\infty}^{\infty}|\hat{E}(\omega)|^2 d\omega \\ &= 4\pi\int_0^{\infty}|\hat{E}(\omega)|^2 d\omega\end{aligned} \tag{4.4}$$

(I've used 4.3 in the last step).

Now: the total energy per unit area emitted in the radiation pulse is

$$\frac{dW}{dA} = \frac{c}{4\pi}\int_{-\infty}^{\infty} E(t)^2 dt \tag{4.5}$$

(to see this, start with the Poynting flux, and remember that $E = B$ for an EM wave in cgs). Now by (4.4), we can rewrite this as

$$\frac{dW}{dA} = c\int_0^{\infty}|\hat{E}(\omega)|^2 d\omega \tag{4.6}$$

[1] In SI, the formula is

$$\text{SI} : \qquad P(t) = \frac{\mu_o}{6\pi}\frac{e^2}{c}|\mathbf{a}(t)|^2$$

[2] or, $\mathbf{S} = \mathbf{E}\times\mathbf{B}/\mu_o$ in SI.

So: we're just about there. What we do, essentially, is think of the integral in (4.6) as an integral over the frequency spectrum of the radiation pulse: that is, $c|\hat{E}(\omega)|^2$ measures the energy in the pulse "at ω".[3]

Thus: looking back to (4.1), we can connect this formalism to the Larmor result. Compare (4.1) to (4.6): they both described the total energy within the pulse. So, think about the Fourier transform of some component (x, y or z) of the acceleration:

$$a_i(t) = \int_{-\infty}^{\infty} \hat{a}_i(\omega) e^{-i\omega t} d\omega \tag{4.7}$$

We can thus identify

$$P(\omega) = \frac{8\pi}{3}\frac{e^2}{c^3}|\hat{a}(\omega)|^2 \tag{4.8}$$

with what we want, namely, the *spectrum* of the radiation. The 4π difference between the frequency-space definition here and the real-time definition (4.1) arises in the Fourier transform (4.4).

4.2 Bremsstrahlung I: single particle

We can now apply this to radiation from an electron-ion collision, using the usual geometry (which you remember from chapter 3 of our Phys 425 notes).

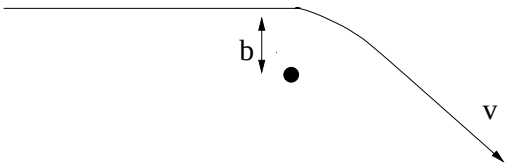

Figure 4.2 The usual geometry, as an electron is deflected by an ion. The impact parameter is b; let the electron move in the x direction to start, with velocity v.

The two components of acceleration are

$$\begin{aligned} a_x &= \frac{e^2 vt}{m(b^2+v^2t^2)^{3/2}} \\ a_z &= \frac{e^2 b}{m(b^2+v^2t^2)^{3/2}} \end{aligned} \tag{4.9}$$

where we have used the impact parameter, b, have set the origin of time at the time of closest approach, and have assumed the particle suffers only a small deflection, so that the **v** does not change by much. The radiated spectrum will depend on the FT of this acceleration. Without doing this out algebraically, we can predict the answer, using what we know about Fourier transforms. In particular, we note that

(i) The z-component of the acceleration will be the dominant factor over the course of the encounter because it does not go to zero at closest approach.

(ii) Since $a(t)$ is large only when the two particles are close together – just as we argued in the Coulomb collision discussion – $P(t)$ will be significant only for times $\lesssim 2b/v$.

(iii) Therefore, we expect the spectrum will be dominated by the FT of a_z, and that the FT will have power at frequencies $\omega \lesssim v/2b$.

Doing the actual work, Longair gives the result for both parallel and perpendicular acceleration:

$$\begin{aligned} \hat{a}_x(\omega) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^2 vt}{m_e}\frac{e^{i\omega t}}{(b^2+v^2t^2)^{3/2}}dt \\ &= \frac{1}{2\pi}\frac{e^2}{m_e bv}\frac{2\omega b}{v} iK_0\left(\frac{\omega b}{v}\right) \end{aligned} \tag{4.10}$$

and

$$\begin{aligned} \hat{a}_z(\omega) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^2 b}{m_e}\frac{e^{i\omega t}}{(b^2+v^2t^2)^{3/2}}dt \\ &= \frac{1}{2\pi}\frac{e^2}{m_e bv}\frac{2\omega b}{v} K_1\left(\frac{\omega b}{v}\right) \end{aligned} \tag{4.11}$$

where $K_o(\omega b/v)$ and $K_1(\omega b/v)$ are modified Bessel functions. We can find analytic forms for K_1 and K_o in the limits of large and small arguments:

$$\begin{aligned} K_0(x) &\to -\ln x \qquad x \ll 1 \\ &\to \sqrt{\frac{\pi}{2x}}e^{-x} \qquad x \gg 1 \end{aligned} \tag{4.12}$$

and

$$\begin{aligned} K_1(x) &\to \frac{1}{x} \qquad x \ll 1 \\ &\to \sqrt{\frac{\pi}{2x}}e^{-x} \qquad x \gg 1 \end{aligned} \tag{4.13}$$

From this, we can add both acceleration terms (squared) to get the radiated spectrum:

$$P(\omega) = \frac{8\pi e^2}{3c^3}\left[|\hat{a}_x(\omega)|^2 + |\hat{a}_z(\omega)|^2\right] \tag{4.14}$$

[3] There is an important technical detail here. The expression (energy/area) in (4.5) or (4.6) is *not* per unit time, rather it's integrated over the pulse. If we tried to do this argument "per dt" and "per $d\omega$", we'd violate the uncertainty relation between ω and t in the wave packet. However, if the pulses repeat frequency, one can formally take limits and get the same result ... *cf.* RL for details here.

Using the analytic expressions for the modified Bessel functions, we find the limiting forms for the spectrum radiated in a single-particle encounter:

$$\begin{aligned} P(\omega) &\simeq \frac{8}{3\pi}\frac{e^6}{m_e^2c^3v^2b^2} && \omega \ll \frac{v}{2b} \\ P(\omega) &\simeq \frac{8}{3\pi}\frac{e^6}{m_e^2c^3v^2b^2}e^{-\omega b/v} && \omega \gg \frac{v}{2b} \end{aligned} \tag{4.15}$$

You should note that the exponential cutoff in the second equation, here, means there is effectively no radiation above $\omega \sim v/b$ — which is consistent with what we expect from the duration of the wave packet.

The low frequency limit of (4.15) is worth commenting on. $P(\omega) \to$ constant as $\omega \to 0$, and is well behaved. However, the *photon* emissivity, $P(\omega)/\hbar\omega$ diverges as $1/\omega$ as $\omega \to 0$. This is a well-known problem in both classical and quantum electrodynamics, known as the "infrared divergence." This is essentially an artifact of our derivations, rather than a problem with the physics (for instance, see Jauch and Rohrlich, *The Theory of Photons and Electrons*). Here, we will simply note that the *energy* lost is finite, and that self-absorption in any finite system (see below) will keep us from ever seeing this divergence anyway. We will, therefore, carry on happily.

4.3 Bremsstrahlung II: from a plasma

Now, we want to extend this to consider a particle in a plasma, and to take all of its collisions into account. We did this last term, when we derived Coulomb scattering – we had to take all impact parameters into account. We'll do essentially the same thing here.

First, consider the range of impact parameters that one particle encounters. Since the number of ions that one electron, at velocity v, sees per second at impact parameter b is $2\pi n_i v b db$, we can find the total radiated spectrum from that electron,

$$\begin{aligned} P(\omega, v) &= \int_{b_{\rm min}}^{b_{max}} P(\omega, v, b) 2\pi n_i v b \, db \\ &= \int_{b_{min}}^{b_{max}} \frac{8}{3\pi}\frac{e^6}{m_e^2 b^2 v^2 c^3} n_i v 2\pi b \, db \\ &= \frac{16}{3}\frac{e^6 n_i}{m_e^2 v c^3}\ln\left(\frac{b_{max}}{b_{min}}\right) \end{aligned} \tag{4.16}$$

Again, the range of impact parameters must be chosen with some physics in mind. And again, luckily, our choice only affects the answer logarithmically. Typical choices are $b_{min} \sim e^2/m_e v^2$, and $b_{max} \sim v/\omega$ or $\sim \hbar/2m_e v$.

Next, we use this to find the total energy loss rate for one particle. We do this by integrating $P(\omega, v)$ over all frequencies. Since $P(\omega, v)$ is only a weak function of ω (it appears only in $\ln(b_{max}/b_{min})$), up to the frequency cutoff $\omega_{max} \simeq m_e v^2/\hbar$ (which is the highest photon frequency we can expect, from simple energy conservation), we have

$$\begin{aligned} P(v) &= \int_{\omega_{min}}^{\omega_{max}} P(\omega, v) d\omega \\ &= \frac{16}{3}\frac{e^6}{c^3 m_e \hbar}\ln\left(\frac{b_{max}}{b_{min}}\right) n_i v \end{aligned} \tag{4.17}$$

This has the functional form $P(E) \propto E^{1/2}$.

Returning to the single particle spectrum (4.16), we can now integrate over all particles in the plasma to get the total emissivity from that plasma. We need to know the distribution of electron speeds, and we will assume a Maxwell-Boltzmann distribution. We also switch from ω to $\nu = \omega/2\pi$, to connect with observations; and we derive $j_{ff}(\nu)$, the emissivity per steradian, to connect with the radiative transfer applications, above. (That means simply a 4π factor, since the single particle emission is essentially isotropic). Thus,

$$j_{ff}(\nu) = \frac{1}{4\pi}\int_0^\infty P(\omega, v) f(v) dv \tag{4.18}$$

with $f(v)$ assumed to be the Maxwellian of (3.9), normalized to n_e. This gives us

$$\begin{aligned} j_{ff}(\nu) = \frac{8}{3}\left(\frac{2\pi}{3}\right)^{1/2}&\frac{e^6}{m_e^{3/2}c^3} \\ &\times \frac{n_e n_i}{(k_B T)^{1/2}} g_{ff}(\nu, T) e^{-h\nu/k_B T} \end{aligned} \tag{4.19}$$

Numerically, with everything in cgs units, this is

$$\begin{aligned} j_{ff}(\nu) =& 5.44 \times 10^{-39} g_{ff}(\nu, T)\frac{n_e n_i}{T^{1/2}} e^{-h\nu/k_B T} \\ &\text{erg s}^{-1}\text{cm}^{-3}\text{Hz}^{-1}\text{str}^{-1} \end{aligned} \tag{4.20}$$

In this expression, we have implicitly defined the Gaunt factor, $g_{ff}(\nu, T)$. It arises from the velocity dependence of $\ln(b_{max}/b_{min})$, inside the integral in (4.18): the essential part of the integral is

$$\begin{aligned} \int \frac{1}{v} f(v) \ln\left(\frac{b_{max}(v)}{b_{min}(v)}\right) \\ \to \langle\frac{1}{v}\rangle\langle\ln\left(\frac{b_{max}(v)}{b_{min}(v)}\right)\rangle \end{aligned} \tag{4.21}$$

We see that the $\langle 1/v \rangle$ becomes the $(k_B T)^{-1/2}$ term; the mean of the logarithmic factor becomes $g_{ff}(\nu, t)$, the Gaunt factor. Note that both b_{max} and b_{min} might be functions of ν and of v. As with Coulomb scattering, different expressions, corresponding to different choices of b_{max} and/or b_{min}, are used in different situations. A couple of common cases are, first, in the radio range, with $h\nu \ll k_B T$:

$$\begin{aligned} g_{ff}(\nu, T) &\simeq \frac{\sqrt{3}}{\pi} \ln \left(\frac{2}{\pi} \frac{(k_B T)^{3/2}}{e^2 m_e^{1/2} \nu} \right) \\ &\simeq 10 \left(1.0 + 0.1 \log \frac{T^{3/2}}{\nu} \right) \end{aligned} \tag{4.22}$$

Second, in the X-ray range, with $h\nu \lesssim k_B T$, people use

$$g_{ff}(\nu, T) \simeq \frac{\sqrt{3}}{\pi} \ln \left(\frac{k_B T}{h\nu} \right) \tag{4.23}$$

Finally, the total emissivity of the plasma can be found, by integrating $j_{ff}(\nu)$ over all frequencies and all solid angles. This is

$$\begin{aligned} \varepsilon_{ff} &= 4\pi \int_0^\infty j_{ff}(\nu) d\nu \\ &= \left(\frac{2\pi k_B}{3 m_e} \right)^{1/2} \frac{32\pi e^6}{3 h m_e c^3} n_e n_i \langle g_{ff} \rangle T^{1/2} \\ &\simeq 1.4 \times 10^{-27} n_e n_i \langle g_{ff} \rangle T^{1/2} \quad \text{erg cm}^{-3}\text{s}^{-1} \end{aligned} \tag{4.24}$$

where $\langle g_{ff} \rangle$ is the mean Gaunt factor, averaged over frequency.

We are also interested in free-free absorption. This is the inverse of the emission process; a free electron absorbs a photon (the ion must be there, as well, to conserve momentum and energy at the same time). For absorption by a Maxwellian plasma, for which we have just derived the emissivity $j_{ff}(\nu)$, we can get the absorption coefficient, $\kappa_{ff}(\nu)$, immediately from Kirchoff's law (3.34):

$$\begin{aligned} \kappa_{ff}(\nu) &= \frac{4}{3\pi} \left(\frac{2\pi}{3} \right)^{1/2} \frac{e^6 n_e n_i g_{ff}(\nu, T)}{m_e^{3/2} c (k_B T)^{3/2} \nu^2} \\ &\simeq 0.018 g_{ff}(\nu, T) \frac{n_e n_i}{T^{3/2} \nu^2} \quad \text{cm}^{-1} \end{aligned} \tag{4.25}$$

This expression is valid in the Rayleigh-Jeans limit (3.22); and the second expression is all in cgs units). You should note that this expression also contains the Gaunt factor. The most common use of free-free absorption is in the radio range (given the $1/\nu^2$ form), and a commonly used form of the absorption, which includes a particular expression for the Gaunt factor, is

$$\kappa_{ff}(\nu) \simeq 0.08235 \frac{n_e n_i}{T^{1.35} \nu_{GHz}^{2.1}} \quad \text{pc}^{-1} \tag{4.26}$$

Note the oddball units: ν is in GHz; κ is in inverse pc(!); but n_e, n_i and T are still in cgs.[4] Avrett, *Frontiers of Astrophysics*, says this is good for $0.1 < \nu < 50$ GHz, and for $6000 < T < 18,000$ K (which describes HII regions and much of the warm, ionized ISM, for instance).

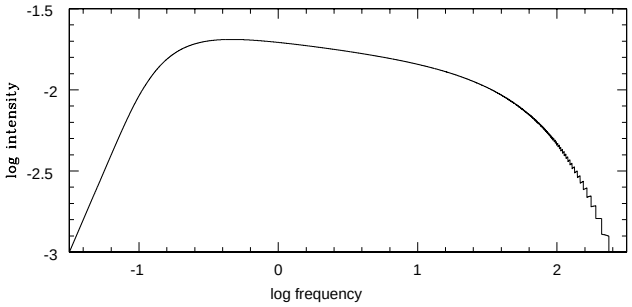

Figure 4.3 Illustrating the full bremsstrahlung spectrum we expect from a source which has a finite size, and thus a finite optical depth. Note, the jagged appearance of the high-frequency exponential is the fault of my simple plotting program, not the physics.

Finally: let's connect this to bremsstrahlung emission from a finite object. Think, for instance, about a galactic HII region (such as the Orion nebula). Typical temperatures are $T \simeq 10^4$ K, and typical densities might be $n \sim 100$ cm^{-3}; the size might be on the order of a pc. We can estimate $\tau_{ff}(\nu) = \kappa_{ff}(\nu) \ell$, where ℓ is again the line-of-sight thickness of the object, and we find $\tau_{ff}(\nu) \simeq 1$ for frequencies in the low radio range (a fraction of a GHz, say). Below this frequency the source will be optically thick, and the emergent intensity will obey $I_\nu = B_\nu(T) \propto \nu^2$, from the Rayleigh-Jeans limit of the Planck function. At higher frequencies, the source is optically thin, and the emergent intensity has the same frequency dependence as the fundamental emissivity: $I_\nu \propto j_{ff}(\nu, T)$. Thus, the intensity will be approximately constant at higher frequencies. At very high frequencies, $h\nu \sim k_B T$ (that is, tens of eV $\rightarrow 10^{16}$ Hz or so), the exponential cutoff will appear. Figure 4.3 sketches this behavior.

[4] That's astronomers for you ..

References

I'm mostly following a fairly nice discussion given in

• Longair (*High Energy Astrophysics, Vol II*);

and pulling some of the discussion on foundations from

• Rybicki & Lightman, *Radiative Processes in Astrophysics*

Key points (for chapter 4)

• Total power (Larmor formula);

• Wave content $\leftrightarrow$ radiation spectrum; "seat-of-the-pants" FT analysis.

• Bremsstrahlung: basic physical picture, single particle

• Bremsstrahlung: from a plasma; intrinsic spectrum and power

• Bremsstrahlung: emissivity, absorption coefficients, total spectrum

5 Thermal state of the ISM

What determines the energy balance of diffuse astrophysical gases? In general, systems which have had time to reach a thermal balance will have a temperature determined by the balance of heating and cooling rates. For most phases of the ISM (although probably not including the hot, coronal gas), the cooling is due to radiation, from an optically thin gas. The specific radiation mechanisms which cool a cloud or nebula will depend on the composition, the internal excitation states and the temperature of the cloud. This part of the problem is fairly well understood by now, at least for the low-density conditions which describe the ISM. The heating mechanism must also be specified; in general, the ISM can be heated by the local radiation field, and by the local cosmic ray population. Since the rate of energy transfer from either of these mechanisms can depend on the local ionized fraction, we must also consider ionization balance. In this chapter, we will first look at some general features of cooling and heating in this context. We will then look at two well-understood examples, one being the thermal state of an HII region, and the other being the multiphase ISM.

> *NOTE to the student*: This chapter contains quite a few details and a lot of unfriendly-looking equations. That's because we need these details in order to understand the various processes whose balance determines the thermal state of the ISM. In addition to the Key Points at the end of the chapter, I'm adding "Look ahead" comments to each of the major sections .

5.1 Heating and cooling: general considerations

The energetics of astrophysical fluids or plasmas ca be more complicated than lab fluids, because *dire* heating and cooling can be important. By "direct" mean that a small volume element, deep inside a clou can exchange energy directly with the outside wor via photons or other particles.

> *LOOKING AHEAD*: The biggest result here is the general cooling curve, $\Lambda(T)$, shown in Figure 5.1, and described in the text. The other important bit is the list of processes which contribute to heating the ISM.

5.1.1 Cooling

For energy loss, the most important astrophysical case is radiation. If the plasma or cloud is optically thin, radiation generated inside the volume element escapes the cloud without further interaction. The net energy loss (per volume per time, frequency-integrated) is often called

$$\Lambda(n, T) = n^2 \mathcal{L}(T) \tag{5.1}$$

and has dimensions, erg cm^{-3} s^{-1}. Note the separation into n^2 and a function $\mathcal{L}(T)$ of T only. It reflects the fact that radiative cooling processes are all two-body processes (electron-ion scattering, electron-atom collisions, etc). The total cooling rate from a thermal astrophysical plasma is due to the sum of all possible spectral lines emitted and also continuous emission due to electron-ion collisions. You've already seen the latter; it's *bremsstrahlung*. The former can be quite complicated – one has to know the excitation state of each level of each atom (which depends on the temperature of the gas and the microphysics of excitation and de-exciation processes) – and must be calculated numerically.

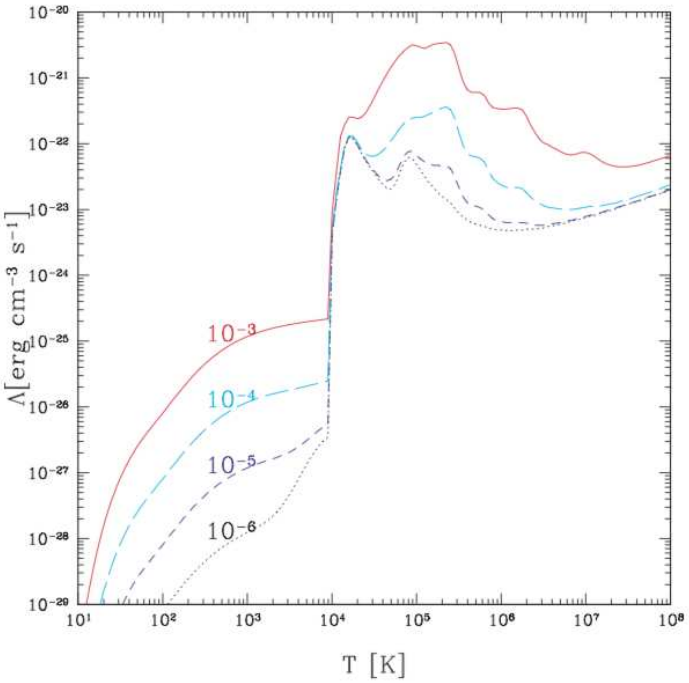

Figure 5.1 Cooling rates as a function of temperature, for diffuse astrophysical plasma at a density of 1 cm^{-3}. Different curves indicate differing heavy metal abundances (relative to hydrogen) by number. Cooling rates for other densities can be obtained by scaling as n^2. From Maio et al 2007, MNRAS 379, 963; see also an older version in Spitzer Figure 6.2.

Figure 5.1 shows the result, from a calculation (originally due to Raymond, Cox & Smith 1977), which adds up all of the different cooling mechanisms, to get what's called the standard *cooling curve*. The main coolants, in the various temperature regions, are:

• $T < 10^4$ K: collisional excitation of fine structure levels of various heavy elements; lines mostly in the IR.

• $T \sim 10^4$ K: excitation of hydrogen lines; the great abundance of hydrogen makes this the dominant coolant in the temperature range where H is significantly excited at or above the first excited state.

• $T \sim 10^4 - 10^7$ K: excitation of optical/UV lines of heavy elements, often forbidden lines.

• $T \gtrsim 10^7$ K: no species are left un-ionized, so bremsstrahlung is the only coolant.

At temperatures $T \gtrsim 10^9$ K, other reactions – such as pair production – become important; very few astrophysical plasmas are this hot, and if they are, other physics is probably important as well.

5.1.2 Heating

Direct heating of astrophysical plasmas isn't so simple, because it depends on the environment of the cloud, not just its internal state (density and temperature). We know, in general, the heating sources for the diffuse ISM. Stars are the fundamental energy sources; they lose energy, and supply power to the ISM, by starlight and also through dynamic input from such events as supernovae and stellar winds. What we need to quantify here, however, is the microphysics: how does this energy couple to some small test volume deep within an interstellar cloud? This area is still under discussion. Here I lean on the discussion from McKee (1995), which focuses on the diffuse ISM; but add photoionization heating which is important for HII regions.

• **Cosmic Rays**. These high-energy particles penetrate the cold ISM, and can transfer energy both through ionization of the neutral component and through Coulomb interactions with the ionized component. This is the heating mechanism first suggested by Field, Goldsmith, &Habing (1969), in their original model of the multiphase ISM. Since then other authors have argued that cosmic rays can't supply enough power, and that other possibilities are more important.

• **X-rays**. These have also been suggested, as they can also penetrate the ISM. They heat by ionization. However, most authors seem to agree that their heating rate is also too low to be useful.

• **Magnetic dissipation**. This is attractive but harder to quantify. The idea is that the large-scale galactic field will dissipate locally, either by local reconnection events (similar to solar flares), or by dissipation of MHD waves. The latter in particular could be a heating mechanism for the cool ISM, as the waves could propagate fairly far from their sources before dissipating. I have not seen this treated quantitatively, however, in any ISM modelling.

• **Photoelectric heating**. This seems to be the current favorite. Starlight is abundant in the galaxy (it also has a density ~ 1 eV/cm^3); can it couple to the ISM? Most starlight is too cool to ionize the ISM. It can, however, be absorbed by interstellar grains which then eject electrons, in a standard photoelectric effect. This can be quantitatively effective if very small grains and also large organic molecules (polycyclic aromatic hydrocarbons, PAH's) are included in the models.

• **Photoionization heating**. This will be treated in more detail immediately below. It is unlikely to be important in the diffuse ISM, because there are not very many "loose photons" with $h\nu > 13.6$ eV; but it is the dominant heating mechanism close to hot, young stars (and probably close to active galactic nuclei).

For our purposes here, the total heating rate – whatever it may be – will be called

$$\Gamma = n\mathcal{H} \tag{5.2}$$

and has units erg cm^{-3} s^{-1}. Once again, we've extracted the density dependence as most of the heating mechanisms described here depend on the first power of the density (*i.e.*, on how many atoms are around to be heated).

5.2 HII regions

Let's start with a spherical chicken. That is, we will start with the ionization and thermal structure of a simple, one-star HII[1] region. Consider one hot star (O or B type, with a significant part of its luminosity above the ionization edge of hydrogen, 13.6 eV), sitting in the ISM. The UV photons from this star will ionize and heat the local hydrogen, producing an HII region

[1]Notation: HII (called "H-two") refers to ionized hydrogen; HI ("H-one") is neutral hydrogen. And just to be confusing, H_2 (also called "H-two") is molecular hydrogen.

around the star. Within an HII region, the central star (or stars) will dominate the energy and ionization balances, which simplifies the problem (compared to the more general case of the diffuse ISM). To start, then, we will therefore consider the equilibrium of a purely photonionized nebula.

> *LOOKING AHEAD*: The important features in here are (i) the definitions of ionization and recombination rates (say equations 5.4 and 5.5); their use in ionization balance (equations 5.7, 5.8); the Stromgren sphere (5.17); the general discussion of heating and cooling (§5.2.2); and the final result, 8000-10,000 K "always".

5.2.1 Ionization structure

Let the central star have a spectrum $L(\nu)$ (in erg/s-Hz). At a distance r from the star, the mean intensity is

$$J(\nu, r) = \frac{L(\nu)}{4\pi r^2} e^{-\tau(\nu,r)}. \tag{5.3}$$

Look back to chapter 3: $J(\nu, r) = \frac{1}{4\pi}\int I(\nu, r) d\Omega$ is the mean intensity averaged over solid angle. The expression in (5.3) describes the intensity from the central star, attenuated by some optical depth $\tau(\nu, r)$ to be evaluated below. Photons above $h\nu_1 = 13.6\text{eV}$ will ionize hydrogen; the ionization cross section is $\sigma_{ion}(\nu) \simeq 6.6 \times 10^{-18}(\nu/\nu_1)^{-3}$ cm^{-2}. We can therefore write down the ionization rate per hydrogen atom,

$$\varphi_{uv}(r) = \int_{\nu_1}^{\infty} 4\pi \frac{J(\nu, r)}{h\nu} \sigma_{ion}(\nu)\, d\nu \tag{5.4}$$

Note that φ_{uv} has units s^{-1}; the ionization rate per volume is $n_{HI}\varphi_{uv}$, units cm^{-3} s^{-1}.

We also need the recombination rate to level j, from the continuum; call it $n_e\alpha_j(T)$ recombinations per second per atom. From the recombination cross section, again assuming a Maxwellian distribution of free electron velocities, $\alpha_j(T)$ is

$$\begin{aligned} n_e\alpha_j(T) &= \int v\sigma_{rec,j} f(v) dv \\ &\simeq \frac{2.1 \times 10^{-11} g_j n_e}{[1 + j^2 k_B T/13.6\text{eV})]\, j^2 T^{1.2}} \end{aligned} \tag{5.5}$$

where g_j is the statistical weight of that level. This can be summed over all levels to get the total recombination rate per atom,

$$\alpha(T) = \sum_j \alpha_j(T) \simeq 2.1 \times 10^{-11} \frac{\Phi(T)}{T^{1/2}} \text{cm}^3\text{s}^{-1} \tag{5.6}$$

where $\Phi(T)$ is yet another order-unity factor, which varies only slowly with temperature in the range $10^2 - 10^5$K.

In equilibrium, the ionization rate will be balanced by the recombination rate, $n_{HII} n_e \alpha(T)$ (recombinations per volume per second). The balance is then given by

$$n_{HI}\varphi_{uv} = n_{HII} n_e \alpha(T) \tag{5.7}$$

Now, the total hydrogen density is $n = n_{HI} + n_{HII}$; if the nebula is mostly hydrogen, $n_e \simeq n_{HII}$. Thus, we can define $x = n_{HII}/n$ as the ionized fraction (so that $n(1-x)$ is the neutral fraction), and write (5.7) as

$$\frac{1-x}{x^2} = \frac{n\alpha(T)}{\varphi_{uv}} \tag{5.8}$$

We note, φ_{uv} is a function of position in this case, from (5.4).

Consider the solutions to (5.8) for x. We know $0 \le x \le 1$, by definition. Now, close to the star, the right hand side of (5.8) will be $\ll 1$. We therefore know $x \simeq 1$, and the quadratic solution of (5.8) will be

$$1 - x \simeq \frac{n\alpha(T)}{\varphi_{uv}} \ll 1 \tag{5.9}$$

Far from the star, the right hand side will be $\gg 1$ (as φ_{uv} drops), and the solution will be

$$x \simeq \left(\frac{\varphi_{uv}}{n\alpha(T)}\right)^{1/2} \ll 1 \tag{5.10}$$

In addition, the region (of r) over which $n\alpha(T)/\varphi_{uv}$ changes from large to small turns out to be small (you can show that this region has width $\Delta r \sim$ the mean free path of a photon in the neutral gas). Thus, we expect the ionization structure to go from fully ionized, to hardly ionized, over a very short distance.

NOW, We want to study the structure of the nebula in a bit more detail, and to find the location of the ionized/neutral transition. To start, we need the optical depth, from the star out to radius r:

$$\begin{aligned} \tau(\nu, r) &= \int_0^r n(1-x)\sigma_{ion}(\nu) dr \\ &\simeq n(1-x)\sigma_{ion}(\nu) r \end{aligned} \tag{5.11}$$

where the last step assumes a uniform density and that $(1-x) \simeq 0 \simeq$ constant (which is true within the ionized region). Now, consider the local radiation density, $u(\nu, r) = J(\nu, r)/c$. Using (5.3), we can write

$$\begin{aligned} &\frac{d}{dr}\left(4\pi r^2 c u(\nu, r)\right) \\ &\quad = \frac{d}{dr}\left(L(\nu) e^{-\tau(\nu,r)}\right) \\ &\quad\quad = -L(\nu) e^{-\tau(\nu,r)} \frac{d\tau(\nu, r)}{dr} \end{aligned} \tag{5.12}$$

And, using (5.11), this becomes

$$\begin{aligned} &\frac{d}{dr}\left(4\pi r^2 c u(\nu, r)\right) \\ &\quad = -L(\nu) e^{-\tau(\nu,r)} n(1-x)\sigma_{ion}(\nu) \; . \end{aligned} \tag{5.13}$$

Now, we can (a) use (5.7) to eliminate $n(1-x)$ on the right hand side; (b) multiply both sides by $d\nu/n\nu$, and (c) integrate over ν:

$$\begin{aligned} &\frac{d}{dr}\left[4\pi r^2 c \int_{\nu_1}^{\infty} \frac{u(\nu, r)}{n\nu} d\nu\right] \\ &\quad = -4\pi r^2 c \frac{\alpha(T) n^2 x^2}{\varphi_{uv}} \int_{\nu_1}^{\infty} \frac{u(\nu, r)\sigma_{ion}(\nu)}{h\nu} d\nu \end{aligned} \tag{5.14}$$

At this point, we can identify terms. The quantity in brackets on the left of the equation is the number of UV photons per second crossing the surface at r, $S_{uv}(r)$. The integral on the right side is just $\varphi_{uv}(r)/c$. Thus, the equation simplifies quite nicely to

$$\frac{dS_{uv}(r)}{dr} = -4\pi r^2 \alpha(T) n^2 x^2 \tag{5.15}$$

For a constant density region, in which $x \simeq 1$, we can easily solve this equation:

$$S_{uv}(r) = S_{uv}(0) - \frac{4\pi}{3} r^3 \alpha(T) n^2 x^2 \tag{5.16}$$

where $S_{uv}(0)$ is the photon flux at the star.

From this, we can define the *Stromgren radius* as the distance at which the photon flux goes to zero:

$$S_{uv}(0) = \frac{4\pi}{3} R_s^3 \alpha(T) n^2 x^2 \tag{5.17}$$

This limit defines the Stromgren sphere. We note that – due to the assumption of local ionization balance – the expression for R_s only depends on the *total* number of UV photons, not on their energy, or how many central stars there are, or the ionization cross section, or any number of physical parameters one might have thought interesting. R_s is determined only by $S_{uv}(0)/n^2$. (Remember that $x \simeq 1$ inside R_s). The physical picture is, simply, that the size of the HII region – R_s – is set by the volume inside of which the number of recombinations per second exactly balanced the number of ionizing photons put out by the star per second.

5.2.2 Energy balance and temperature

Photoionization also provides the heating for the nebula. A photon of energy $\nu > \nu_1$ ionizes an atom, and the leftover energy $h(\nu - \nu_1)$ goes to kinetic energy of the free electron. This electron can then share the energy with the ions, through Coulomb collisions, so that this becomes a general heating mechanism for the gas.

The heating rate per HI atom can be written

$$\begin{aligned} &\int_{\nu_1}^{\infty} 4\pi \frac{J(\nu)}{h\nu} \sigma_{ion}(\nu) h(\nu - \nu_1) d\nu \\ &\quad\quad \simeq \varphi_{uv} \langle h(\nu - \nu_1)\rangle \end{aligned} \tag{5.18}$$

where the second expression defines the mean energy transfer per ionization (with the mean taken over the input photon spectrum). Note, we have suppressed the r dependence in $J(\nu)$, to ease the notation. Thus, the heating rate per volume is

$$\Gamma_{uv} = n_{HI} \varphi_{uv} \langle h(\nu - \nu_1)\rangle \tag{5.19}$$

We can again assume ionization balance, and this can be written in terms of the local density and temperature, as

$$\Gamma_{uv} = n^2 x^2 \alpha(T) \langle h(\nu - \nu_1)\rangle \tag{5.20}$$

To determine the equilibrium temperature, we want to balance Γ_{uv} against all of the important radiative cooling mechanisms (assuming the HII region is optically thin, which is a good assumption for visible ones – the ones in the pretty pictures). It turns out that there are two important types of coolants – recombination lines from hydrogen (and helium, which is a small correction), and collisionally excited lines from heavy elements. To do the calculation, we have to to compute the cooling rates numerically for these lines (or find someone who has done it already!).

The hydrogen recombination cooling rate is

$$\Lambda_{rec} = n_{HII} \sum_j \int \sigma_{rec,j}(v) v f(v) \frac{1}{2} m_e v^2 dv \tag{5.21}$$

Now, the integral-sum on the RHS is essentially the net recombination rate, (*cf.*equation 5.5), times the mean energy released per recombination:

$$\Lambda_{rec} \simeq n^2 x^2 \alpha(T) k_B T \quad . \tag{5.22}$$

The integral is over the electron distribution function, and the sum is over all relevant energy levels.

Can hydrogen cooling alone account for the temperature of an HII region? We observe temperatures $\lesssim 10^4$ K. Now, if the heating is by hydrogen ionization, and the cooling is by hydrogen recombination, our thermal balance would be

$$\Gamma_{uv}(T) = \Lambda_{H,rec}(T) \tag{5.23}$$

Comparing (5.20) and (5.22), we see that this solves to

$$k_B T \simeq \langle h(\nu - \nu_1) \rangle \tag{5.24}$$

(everything about the density and ionization state drops out, note – due to the ionization rate balance). But this is too hot; recall $h\nu_1 = 13.6$ eV. Thus, simple estimates don't work here – we need to include heavy element cooling, which turns out to be stronger, and results in the lower temperatures we see.

For the heavy element cooling rates, we can write schematically (for element X)

$$\Lambda_X = \sum_{j,k} n_{X,j} A_{X,jk} h\nu_{jk} \tag{5.25}$$

where the sum is over all "upper" and "lower" states, and the level populations $n_{X,j}$ must be determined by the methods discussed previously. An important fact to know is that the levels are collisionally excited (but radiatively de-excited, of course). Thus, the net cooling rate $\Lambda_x \propto n_e n_X \propto n^2$, since it is a collisional process.

Thermal balance in the nebula can then be written as

$$\Gamma_{uv}(T) = \Lambda_{rec}(T) + \sum_X \Lambda_X(T) \quad . \tag{5.26}$$

Solutions of this equation determine the temperature of the nebula. From (5.20), and the discussion just above, we see that both sides $\propto n^2$, so that the equilibrium temperature does not depend on the density. The temperature can be found numerically.

This is the big result: the temperature does not depend on the density; only on the details of the ionizing spectrum and coolants. Thus, the temperature of any (photoionized) HII region is $\sim 8000 - 10,000$ K. This result is independent of the density (nearly) and of the number of stars ionizing the region; it has some dependence on the ionizing spectrum and the abundance of heavy elements, primarily oxygen. Through the level populations it does have a weak dependence on the density.

5.3 The diffuse ISM: multiphase equilibrium

For the general diffuse ISM, things aren't quite so simple. However, Field etal (1969) developed a very nice semi-analytic formulation; here I simplify the algebra[2] by extracting only the dominant terms. We begin by determining what the most important cooling and heating mechanisms are. We start with cooling, which is the simpler of the two, as it only depends on microphyics.

> *LOOKING AHEAD*: The big results here are (i) the general idea of setting up a "heating = cooling" balance for the ISM; and (ii) the result, that the cold and warm phases which coexist in the ISM are thought to be two stable solutions of a "heating = cooling" balance.

5.3.1 Cooling: what dominates here?

The ISM cools by radiation (it's optically thin to many/most photons). The radiation is generated, typically, by collisional excitation of some atom or other, by a free electron. The atom de-excites by radiation, leading to a net energy loss from the system. If we want to find which of all possible line transitions are important for cooling, we can think of several criteria.

- The species must be fairly abundant;
- its excitation energy must be comparable to, or less than, the typical electron thermal energy;
- it must have a large cross section for such excitation;
- it must have a high probability to de-excite before another collision occurs;
- and it must generate photons to which the ISM is optically thin.

Given this general argument, we find that two of the most important coolants for the temperature range that describes the cold and warm hydrogen (which is a few $\times 10^2$ to a few $\times 10^3$ K) are particular lines from hydrogen and from carbon. We can approximate the cool-

[2] honestly, folks!

ing rate of these by saying that the line radiatively de-excites as soon as it is excited; thus, the cooling rate is just proportional to the excitation rate. But the collisional excitation rate is, typically,

$$q_{X;ij} \simeq q_{Xo} T^{-1/2} e^{-h\nu_{ij}/k_B T} \tag{5.27}$$

for excitation from level i to level j of species X. The term q_{Xo} is a numerical factor containing atomic constants.[3] For our two main coolants, the cooling rate can be written,

$$\Lambda_H \simeq n_e n_{HI} \frac{q_H}{T^{1/2}} e^{-h\nu_H/k_B T} h\nu_H \tag{5.28}$$

and

$$\Lambda_C \simeq n_e n_{HI} \frac{n_C}{n_H} \frac{q_C}{T^{1/2}} e^{-h\nu_C/k_B T} h\nu_C \tag{5.29}$$

Here, we are assuming that $x \ll 1$, so that $n_{HI} \simeq n$. We note that the line frequencies are given by $h\nu_C/k_B \simeq 92$ K, and $h\nu_H/k_B \simeq 10^5$ K. Thus, at low temperatures C cooling dominates, and at higher temperatures H cooling dominates.

5.3.2 Heating: by cosmic rays?

As an example of how this can be quantified, consider cosmic ray heating. This may not be the dominant heating; however it is easy to write down analytically (as Field etal did), and therefore provides a useful example of the analysis. In addition, cosmic ray heating is proportional to the local density, just as X-ray heating and photoelectric heating are (and probably as magnetic heating would be, too). Thus, the analysis I present here could be extended to PAH heating by changing the constants; the results for the multiphase ISM should be fairly robust.

In cosmic ray heating, the neutral fraction will be heated by ionization. If $\sigma_{ion}(E)$ is the cross section for ionization by a cosmic ray of energy E, $f_{cr}(E)$ is the density of cosmic rays at energy E, and ΔE is the excess energy the electron comes away with in this ionization, the ionization rate per volume can be written,

$$n_{HI}\varphi_{cr} = n(1-x) \int f_{cr}(E) v(E) \sigma_{ion}(E) dE \tag{5.30}$$

[3]The exponential, and the inverse $T^{1/2}$, arise from an integral of the collision rate, $1/n\sigma(v)v$, integrated over the electron velocity distribution; see if you can justify this for yourself.

From this, the heating rate due to cosmic ray ionization can be written,

$$\begin{aligned}\Gamma_{cr,ion} &= n(1-x) \int f_{cr}(E) v(E) \sigma_{ion}(E) \Delta E dE \\ &= n(1-x)\varphi_{cr}\langle \Delta E \rangle\end{aligned} \tag{5.31}$$

in direct analogy to (5.20). Numerically, this turns out to be

$$\Gamma_{cr,ion} \simeq 5 \times 10^{-12} \varphi_{cr} n(1-x) \ \text{erg cm}^{-3}\ \text{s}^{-1}$$

Current estimates for φ_{cr}, averaged over the disk, suggest $\varphi_{cr} \simeq 10^{-17}$ s^{-1}.

The ionized fraction of the gas will be heated by Coulomb collisions with the cosmic rays. A single cosmic ray, once it is relativistic, loses energy as a rate $P_C(E) \simeq 6 \times 10^{-19} nx$ erg/s to the ionized component (this number comes from the Coulomb collision rate for relativistic particles). Thus,

$$\Gamma_{cr,C} = \int f_{cr}(E) P_C(E) dE \tag{5.32}$$

To simplify the algebra below, we note that $\Gamma_{cr,C}$ can be related to φ_{cr}, since both involve an integral over $f_{cr}(E)$, the cosmic ray distribution function. We will write this relation as $\Gamma_{cr,C} = \mathcal{A}\varphi_{cr} nx$, where $\mathcal{A}$ contains a mean of $P_C(E)/\sigma_{ion}(E)$, and other (numerical) factors. Finally, we can also write down the ionization state of the gas, due to the cosmic rays:

$$\varphi_{cr} n(1-x) = n^2 x^2 \alpha(T) \tag{5.33}$$

which has limiting solutions for x, just as in (5.7).

5.3.3 Thermal balance and multiphase equilibrium

We will follow the method first presented by Field et al (1969), and will approximate the behavior of the heating and cooling functions to allow us to find an analytic solution for the equilibrium temperatures. This analysis addresses the two cooler phases of the ISM, the cold and warm HI distributions. (The hot, coronal gas is probably heated dynamically, by supernova shocks and/or stellar winds, and also cooled dynamically by its expansion out of the galactic plane; it will not be included in this analysis).

Now, the general thermal balance equation, for heating by cosmic rays and cooling by these two spectral lines,

from C and H, is

$$\varphi_{cr} n(1-x)\langle \Delta E\rangle + \varphi_{cr}\mathcal{A}nx = n^2 x(1-x) \times \left[\frac{q_H}{T^{1/2}} e^{-h\nu_H/k_BT} h\nu_H + \frac{n_C}{n_H}\frac{q_C}{T^{1/2}} e^{-h\nu_C/k_BT} h\nu_C\right] \tag{5.34}$$

This equation gives the implicit (n, T) solution, for the density-temperature relation of the equilibrium gas. We show this solution, below; but it seems worth getting some insight into its behavior first. Consider only the low-temperature regime, where we can ignore H cooling, and also ignore $\Gamma_{cr,C}$ (since $x \ll 1$). The balance is then, approximately,

$$n\varphi_{cr}(1-x)\langle \Delta E\rangle \simeq n^2 x(1-x)\frac{n_C}{n_H}\frac{q_C}{T^{1/2}} e^{-h\nu_C/k_BT} h\nu_C \tag{5.35}$$

But from this, using the $x \ll 1$ limit of (5.33), we find that

$$n \propto T\alpha(T) e^{2h\nu_C/k_BT} \propto T^{1/2} e^{2h\nu_C/k_BT} \tag{5.36}$$

This gives us an approximate analytic version of the (n, T) relationship for $T \sim 10^2 - 10^3$ K. We can find a similar expression for the high-T regions, where H cooling dominates. The net solution looks like the cartoon in Figure 5.2. This solution can also be turned into a (p, T) solution, using $p \propto nT$, also in Figure 5.2.

But this is, then, (almost) the end of the analysis. That is, since we know the pressure of the system – as we do, observationally ($nT \sim$ few $\times 10^3$ cm^{-3} K), we can find the (p, T) solutions which are allowed by this thermal balance. Figure 5.2 illustrates the solutions; in principle anywhere on the (n, T) or (p, T) loci is a solution. We restrict the possibilities by specifying the pressure – equating it to the ISM pressure (which is known). This gives us three possible solutions (as in Figure 5.2). But will all three exist in the ISM?

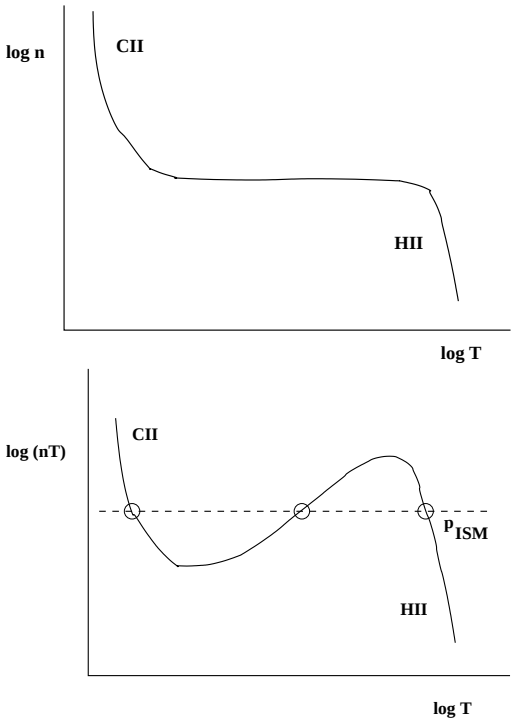

Figure 5.2 Top; sketch of the (n, T) solution from (5.34). The lower temperatures are controlled by CII cooling; the higher temperatures by HII cooling. Any (n, T) pair on this curve is a possible solution; but not all will exist in the ISM. Bottom: the same solution, but now plotting "pressure" nT against T. The dotted line indicates the ISM pressure, which picks out (n, T) solutions that might exist in the ISM. The small circles show three possible solutions; however only the two outer ones are thermally stable. When we put in real numbers, these stable solutions correspond to the *cold* and *warm* phases of the ISM.

5.3.4 Is the thermal balance solution stable?

The final step of the argument is to argue, as Field etal did, that the outer two solutions – $T \sim 10^2$ K and $T \sim$ few $\times 10^3$ K – are the two stable HI phases; and that the middle T solution is unstable. The argument goes as follows. Consider one particular equilibrium state, with density and temperature specified, as well as with external pressure fixed. For instance, look the middle T solution in Figure 5.2. Consider a slight compression of this state, in which the temperature drops a little bit (this is required from the (n, T) solution in this region). But the pressure must also drop in the perturbation (from the (p, T) curve); so the larger external pressure will compress the perturbation still further – and the density will drop still more, and it will cool faster . . . and so on. That is, this system is *thermally unstable*; a slight compression will collapse and cool, and a slight rarefaction will expand and reach higher temperatures. The timescale for this runaway $\sim t_{cool} \sim k_BT/\Lambda_{net}(T)$, as always; for the ISM, this is a short time, so the instability will develop rapidly.

Conversely, for the outer two solutions, the slope of the (p, T) curve is reversed, so that – say, a small compression will reach higher pressures, and expand back to its original state (or vice-versa for a small rarefaction). Thus, the two outer states should be stable, and

are believed to represent the two cool phases of the ISM.

More formally, Field etal (1969) worked with the net cooling function, $\mathcal{C} = \Lambda - \Gamma = n^2\mathcal{L} - n\mathcal{H}$. Again, the cooling term is $\Lambda = n^2\mathcal{L}$ as in Figure 5.1. The heating term is written as $\Gamma = n\mathcal{H}$ to highlight the linear dependence on density n. They showed that the system is unstable if

$$\frac{\partial \mathcal{C}}{\partial T}\Big|_p = \frac{\partial \mathcal{C}}{\partial T}\Big|_n - \frac{n}{T}\frac{\partial \mathcal{C}}{\partial T}\Big|_T < 0 \qquad (5.37)$$

Now, using $\mathcal{C} = 0$ (which means we start in thermal balance, before we perturb the system) in (5.37) turns this condition into

$$\frac{\partial \mathcal{C}}{\partial T}\Big|_p = n^2\frac{d\mathcal{L}}{dT} - n^2\frac{\mathcal{L}}{T} < 0 \qquad (5.38)$$

as the condition for instability. This can be rewritten as

$$\frac{d\ln \mathcal{L}}{d\ln T} < 1 \qquad (5.39)$$

for the instability condition; the system is stable otherwise. This makes it clear that the *slope* of the cooling curve in Figure 5.1 controls the thermal stability of the system. The middle phase of Figure 5.2 lies at a T where $\mathcal{L}$ is a very slow function of T; so this phase is unstable. The outer two phases of Figures 5.2 lie at T where $\mathcal{L}$ is a steep function of T; thus these phases are thermally stable.

References

Most of the formal material here can be found in

• Spitzer (*Physics of the Interstellar Medium*;

but the thermal stability details come from the original Field, Goldsmith & Habing paper (1969 ApJ), and I've also pulled more recent arguments (on heating, cooling mechcanisms) from the current literature.

Key points

• The ISM cooling curve (and what's inside it);

• Photonionization heating "always" gives $\sim 10^4$ K;

• Stromgren spheres – what they are, what size they are;

• General ISM thermal balance and why we have the two cooler phases;

• Why does ISM thermal balance *not* explain the hot (coronal) phase?

6 Dynamics of the ISM: energetics & shocks

Last term we worked with the mass and momentum conservation laws of fluid dynamics, and applied them to various problems. We now return to fluid dynamics, and consider the physics of shocks in fluid flow. But we can't do that until we look at the third important conservation law, energy conservation in fluids.

6.1 Fluids: energetics

We must consider two forms of energy: the kinetic energy density of bulk flows, $\rho v^2/2$, and internal energy density. The latter is the energy contained in random (thermal) motions of the particles. We will work with the internal energy per unit mass, $e = \frac{1}{\gamma-1}\frac{p}{\rho}$. For a subrelativistic, monatomic gas, for instance, $e = \frac{3}{2}\frac{k_B T}{m}$. The net energy in our volume V is $\int_V \rho(e + \frac{1}{2}v^2)dV$. The net rate of change of this energy from intrinsic changes and from flows is

$$\int_V \frac{\partial}{\partial t}\left[\rho\left(e+\frac{1}{2}v^2\right)\right]dV + \int_V \nabla\cdot\left[\rho\mathbf{v}\left(e+\frac{1}{2}v^2\right)\right]dV, \tag{6.1}$$

where we have used the divergence theorem to convert a surface integral to a volume integral as in Chapter 4 of the 425 notes. This net energy change must be accounted for by (a) work done by an external acceleration $\mathbf{f}$, which is often taken to be gravity; (b) work done by the external pressure; (c) direct energy gains or losses, most commonly direct heating (by photons or cosmic rays, say), or radiative losses, as in chapter 5. These three energy-change factors are

$$\int_V \rho\mathbf{f}\cdot\mathbf{v}dV - \int_A p\hat{\mathbf{n}}\cdot\mathbf{v}dA + \int_V(\Gamma-\Lambda)dV \tag{6.2}$$

Again we can use the divergence theorem to convert the pressure work term to a volume integral, and we can derive one version of the differential energy conservation law:

$$\frac{\partial}{\partial t}\left[\rho\left(e+\frac{1}{2}v^2\right)\right] + \nabla\cdot\left[\rho\mathbf{v}\left(e+\frac{1}{2}v^2\right)\right] = \rho\mathbf{f}\cdot\mathbf{v} - \nabla\cdot(p\mathbf{v}) + \Gamma - \Lambda \tag{6.3}$$

The forms we derived for the mass and momentum conservation equations are pretty standard. However, there does not seem to be one standard form for the energy conservation equation; rather, one uses the form that works best in a given application. Therefore, at the expense of a little algebra, we will look at several alternate forms of (6.3).

First, with the help of the continuity equation[1] we can separate out the $\partial\rho/\partial t$ and $\nabla\cdot(\rho\mathbf{v})$ terms in (6.3), we find

$$\rho\frac{\partial}{\partial t}\left(e+\frac{1}{2}v^2\right) + \rho\mathbf{v}\cdot\nabla\left(e+\frac{1}{2}v^2\right) = \rho\mathbf{f}\cdot\mathbf{v} - \nabla\cdot(p\mathbf{v}) + \Gamma - \Lambda \tag{6.4}$$

which is one alternate form that we will use again. We can isolate the rate of change of e, from (6.4), by subtracting $\mathbf{v}\cdot$(the momentum conservation equation), giving

$$\rho\frac{\partial e}{\partial t} + \rho\mathbf{v}\cdot\nabla e = -p\nabla\cdot\mathbf{v} + \Gamma - \Lambda \tag{6.5}$$

In this expression, we can see that the rate of change of the internal energy depends explicitly on compression work ("pdV" work), and on the net heating and cooling rates.

Yet another common form of the energy equation is found by defining the *convective*, *total* or *Lagrangian* derivative,

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla \tag{6.6}$$

With this, we can use the continuity equation to write $\nabla\cdot\mathbf{v}$ in terms of the density derivatives, and use $e = \frac{1}{\gamma-1}\frac{p}{\rho}$ to write

$$\frac{\rho}{\gamma-1}\frac{D}{Dt}\left(\frac{p}{\rho}\right) - \frac{p}{\rho}\frac{D\rho}{Dt} = \Gamma - \Lambda \tag{6.7}$$

or, if we collect the p and ρ derivatives separately, we get

$$\rho^\gamma\frac{D}{Dt}\left(\frac{p}{\rho^\gamma}\right) = (\gamma-1)(\Gamma-\Lambda) \tag{6.8}$$

which is the last of our alternate forms of the energy equation.

This last form allows us to consider a couple of important limits. The first is the *adiabatic limit*. If $\Gamma-\Lambda = 0$, so that there is no net gain or loss of energy to the system, (6.8) shows that

$$\frac{p}{\rho^\gamma} = \text{constant} \tag{6.9}$$

[1] You saw this last term, in chapter 4 of the P425 notes. One form is

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

which is the usual adiabatic law (the consequence of there being no gain or loss of heat from a system). The second limit is the *isothermal limit*. A good many calculations assume T = constant, which simplifies things enormously. From (6.5), we see that

$$p\nabla \cdot \mathbf{v} = \Gamma - \Lambda \tag{6.10}$$

is the condition that must be satisfied if T (or e) is constant.

6.2 Supersonic flow and shock fronts

In P425, we found a characteristic signal speed in a gas, namely the sound speed, c_s. This is a critical finding: because this is the speed at which a perturbation propagates, "information" about changes in the flow can only propagate at c_s. Figure 6.1 illustrates this, in a moving flow.

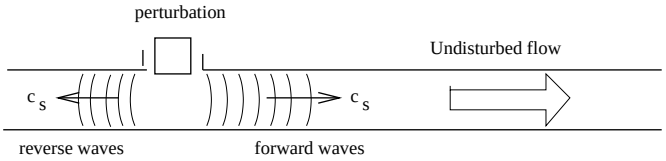

Figure 6.1 Illustrating signal propagation in a flow. A perturbation "whacks" the pipe at one spot; the information that this has happened travels upstream and downstream at c_s, *relative to the flow*. Following Thompson figure 8.6.

Consider, then, gas moving at a speed greater than the sound speed; if follows that information cannot propagate upstream. This means the gas generally cannot adjust smoothly to changes in the ambient or boundary conditions, but rather must adjust instantaneously - creating a discontinuity in the flow. Such a discontinuity is a *shock*. Examples are bow shocks around supersonic aircraft, or around the planets (since the solar wind is supersonic); or standing shocks, such as where supersonic flow runs into a zero-velocity surface (for instance, at the end of a radio jet, where it runs into the ambient plasma).

We treat a shock as an infinitely thin discontinuity in a flow. The true width of the shock is determined by collision processes within the fluid, and by assumption these operate on scales much smaller than those described by the fluid equations. The intent is to derive "jump conditions" – to use the basic conservation laws to derive relations between the fundamental variables (ρ, p, T, v) upstream and downstream of the shock. Let "1" describe upstream, and "2" describe

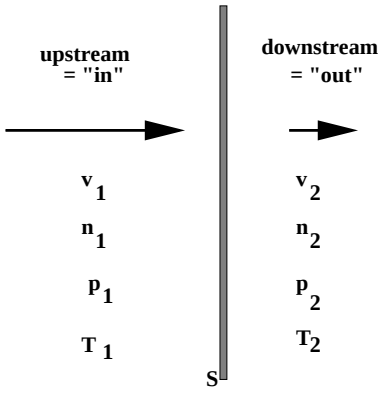

Figure 6.2 A close-up look at a shock, S, in a frame in which the shock is at rest. The incoming fluid is labelled with subscript "1"; outgoing with subscript "2". The incoming fluid must be supersonic, $v_1 > c_{s1}$. The outgoing fluid is slower ($v_2 < v_1$), denser, $n_2 > n_1$), and probably hotter ($T_2 > T_1$, for an adiabatic shock); the incoming kinetic energy is converted to heat when the shock decelerates the flow.

the downsteam flow, as seen in a frame moving with the shock (as in Figure 6.2). Let the *Mach number* be $\mathcal{M} = v/c_s$ (generally defined for upstream flow). Consider steady, one-dimensional flow, with no external forces, and with no net external heating or cooling. Referring back to the footnote on the prevous page, the continuity equation for the fluid in steady state becomes $\nabla \cdot (\rho \mathbf{v}) = 0$. If we integrate this over a small, Gaussian surface enclosing some part of the shock plane, we get

$$\rho_1 v_1 = \rho_2 v_2 \tag{6.11}$$

This is, of course, simply mass conservation: the flux in (per area) must equal the flux out. The force equation for the fluid is

$$\nabla \cdot (\rho \mathbf{v}\mathbf{v}) + \nabla p = 0 \tag{6.12}$$

(again, go back to your P425 notes, where this form was presented implicitly, as equation (4.3), but not elaborated on). Integrating this across the shock face, we get

$$\rho_1 v_1^2 + p_1 = \rho_2 v_2^2 + p_2 \tag{6.13}$$

These two equations are general. The energy equation is generally applied in either the adiabatic or isothermal limits.

6.2.1 Adiabatic shocks

The form (6.3) of the energy equation, with some algebra (always!) and applied to our conditions here, gives

us

$$\frac{\gamma}{\gamma-1}p_1v_1+\frac{1}{2}\rho_1v_1^3=\frac{\gamma}{\gamma-1}p_2v_2+\frac{1}{2}\rho_2v_2^3$$

Factoring out ρv from each side, and using (6.13), this can be written

$$\frac{\gamma}{\gamma-1}\frac{p_1}{\rho_1}+\frac{1}{2}v_1^2=\frac{\gamma}{\gamma-1}\frac{p_2}{\rho_2}+\frac{1}{2}v_2^2 \qquad (6.14)$$

Now, this can be combined with (6.11) and (6.13), to express three post-shock quantities, ρ_2, v_2 and p_2, in terms of their pre-shock counterparts. This solution is:

$$\begin{aligned}\frac{\rho_1}{\rho_2}&=\frac{\gamma-1}{\gamma+1}+\frac{1}{\mathcal{M}^2}\frac{2}{\gamma+1}\\ \frac{p_2}{p_1}&=\frac{2\gamma\mathcal{M}^2-(\gamma-1)}{\gamma+1}\\ \frac{v_2}{v_1}&=\frac{\gamma-1}{\gamma+1}+\frac{1}{\mathcal{M}^2}\frac{2}{\gamma+1}\end{aligned} \qquad (6.15)$$

When $\mathcal{M}\gg 1$, these equations simplify, to

$$\begin{aligned}\frac{\rho_1}{\rho_2}&=\frac{\gamma-1}{\gamma+1}\\ \frac{p_2}{p_1}&=\frac{2\gamma\mathcal{M}^2}{\gamma+1}\\ \frac{v_2}{v_1}&=\frac{\gamma-1}{\gamma+1}\end{aligned} \qquad (6.16)$$

In particular, if $\gamma=5/3$, we find $\rho_2/\rho_1=v_1/v_2=4$ (this is often quoted as the strong shock limit). And, in this limit, the temperature jump is $T_2/T_1=5\mathcal{M}^2/16$, giving $k_BT_2=3mv_1^2/32$; the upstream kinetic energy is converted to internal energy in an adiabatic shock.

6.2.2 Isothermal shocks

The energy equation is this case is simple: $T_2=T_1$ by assumption. The possibility of an isothermal shock depends on the cooling times. We would expect a general shock to have a structure as in Figure 6.3. The gas passing through the shock is initially heated, by adiabatic compression, and suffers a moderate density jump. This hotter gas (assuming an optically thin situation), can then cool by radiation, and while cooling the gas travels a distance $\sim v_2t_{cool}$. This "cooling distance" thus measures the effective width of the transition to an isothermal shock – assuming that some heating/cooling balance, as we described for the ISM, maintains the upstream and far-downstream temperature at T_1.

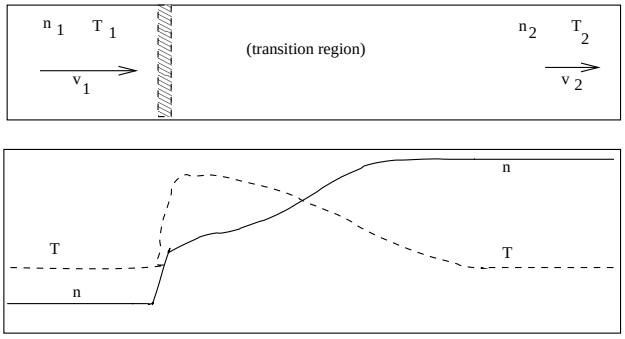

Figure 6.3 Schematic diagram of a radiating shock. In the upper diagram, the fluid is coming in from the left. At x_1 it enters the nonradiative shock. This is followed, to the right, by the transition region, where the temperature drops as the gas cools by radiation. The changes of density and temperature are shown schematically in the lower figure. Following Spitzer figure 10.1.

Combining $T_2=T_1$ with (6.11) and (6.13), to express isothermal jump conditions,

$$\begin{aligned}\frac{\rho_2}{\rho_1}&=\mathcal{M}^2\\ \frac{v_2}{v_1}&=\frac{1}{\mathcal{M}^2}\\ \frac{p_2}{p_1}&=\mathcal{M}^2\end{aligned} \qquad (6.17)$$

Thus, the compression factor, and deceleration factor, can be much higher for an isothermal shock than for an adiabatic one.

6.2.3 Magnetized shocks

The previous analysis ignored the magnetic field. This is probably too naive, as we know the ISM is magnetized. We can understand the effect of a shock on the field by considering flux freezing. Refer to the left part of Figure 6.4. In this case, the field is parallel to the shock face. In this geometry, the field is tied to the gas by flux freezing; thus the field behind the shock will be increased, in the same amount as the gas is compressed. By comparison, consider the right part of the figure, in which the shock propagates along the field lines. In this case, the density jump will not affect the magnetic field (why? can you see how this is consistent with flux freezing?).

6.2.4 Oblique shocks

Finally, what if the shock face is not perpendicular to the flow? The answer is qualitatively simple, and can be readily understood by thinking about an oblique in-

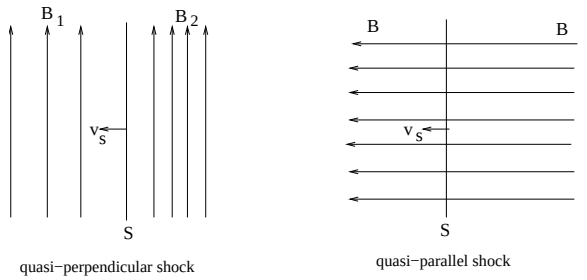

Figure 6.4 Schematic picture of magnetized shock, in two important limits The shock speed relative to the medium is v_s. Left, quasi-perpendicular shock (note $v_s \perp B$): flux freezing increases the post-shock field, by a factor $B_2/B_1 = \rho_2/\rho_1$. Right, quasi-parallel shock; the gas flows along the field lines without perturbing the field.

coming velocity (**w** in the Figure 6.5, which illustrates the geometry) in terms of its components parallel and perpendicular to the shock face. The component perpendicular to the shock face is decelerated, just as in the normal-shock results above. The component parallel to the shock face, however, is not affected (query to the reader: why not?). Thus, the net velocity bends *toward* the shock face.

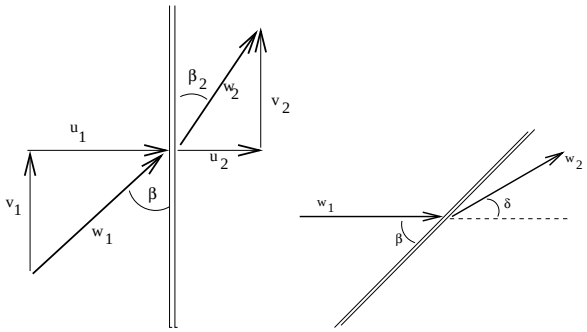

Figure 6.5 Oblique shocks. The velocity bends towards the shock face, as shown in the right figure; this can be understood by considering the effect of the shock on the velocity components, as in the left figure.

Where do we expect to deal with oblique shocks? In just about any 2D situation ... a good example is the terrestrial bow shock, where the supersonic solar wind encounters the earth. The flow coming in bends toward the shock normal, and thus is deflected around the earth. The jump conditions (extensions of 6.15 or 6.17) can be quite complicated algebraically; we won't deal with them in this course.

Key points

- Energy conservation: adiabatic and isothermal limits
- Shock fronts: jump conditions across the shock
- Shock fronts: adiabatic and isothermal limits
- Effects of B field or oblique angle: qualitative

7 Stellar Winds & Supernovae Remnants

Now let's look at some supersonic flow situations which involve shocks. You remember that we worked with smooth transonic flows last term – for instance the solar wind. These are rare; it's very easy to form shocks when supersonic flows are decelerated or bent by their surroundings. In this chapter we'll work with 2-1/2 examples: stellar wind shocks, and two types of supernova remnants.

7.1 Stellar winds and the surrounding ISM

We worked with smooth, transonic stellar wind flow last term ... and found a solution in which the outer regions of the wind flow are supersonic. But this can't carry on forever. At some point the wind flow must run into the ambient ISM. We expect the deceration to lead to an outer shock in the wind; what are the details? By way of review, I'll repeat the basics of the inner-wind solution here.

7.1.1 The basic solution

• Mass conservation in a steady, spherical flow is $\rho v r^2 =$ constant; or,

$$\frac{1}{\rho}\frac{d\rho}{dr} + \frac{1}{v}\frac{dv}{dr} + \frac{2}{r} = 0 \tag{7.1}$$

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

$$\rho v \frac{dv}{dr} + \frac{dp}{dr} = -\rho \frac{GM}{r^2} \tag{7.2}$$

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

$$\left(v - \frac{c_s^2}{v}\right)\frac{dv}{dr} = \frac{2c_s^2}{r} - \frac{GM}{r^2} \tag{7.3}$$

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 (7.3), as follows.

• First, 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 (7.3) must go to zero at the same point. This defines the condition that must be met at the sonic point:

$$v^2 = c_s^2 \qquad \text{at} \qquad r = r_s = \frac{GM}{2c_s^2} \tag{7.4}$$

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.

• Further, 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, (7.3) can be solved in the limit $r \gg r_s$:

$$v^2(r) \simeq 4c_s^2 \ln r + \text{constant} \tag{7.5}$$

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.

7.1.2 The outer shock

The pressure in this supersonic wind is dropping with radius (since $\rho \propto 1/vr^2$, with v being only slowly varying; thus $p \propto \rho T \propto 1/r^2$, approximately, in an isothermal wind). Therefore, when the wind pressure is close to the ISM pressure, the wind must slow down. At this outer boundary, we expect some sort of shock transition, since the wind is supersonic. Past this shock, the hot, shocked wind-gas will expand into the ISM (at about its own sound speed, to start); as long as this expansion is supersonic relative to the ISM, the expanding hot gas will drive a "snowplowed" shell of ISM, and a second shock, out into the ISM.

A cartoon of this region, at some point in time, would be that in Figure 7.1. Let region "a" be the wind; S_1 be the inner shock; region "b" be the wind-gas which has been through the shock; C be the contact surface between the wind and the ISM; region "c" be the shocked ISM; and S_2 be the outer shock (moving into the ISM). We expect S_1 to be an adiabatic shock (since the wind is probably hot and low density, and thus will have a long cooling time); region "b" will contain hot, shocked wind, with $T_b \sim \frac{3}{16}\frac{m v_{wind}^2}{k_B} \sim$ several $\times 10^7$ K

(noting that $m = \frac{1}{2}m_p$ is the mean mass per particle if region "b" is fully ionized). The outer shock will probably be isothermal, since the ISM is denser and cooler than the wind. Thus, the shocked ISM will be in a thin shell, containing all of the original ISM that lay between S_2 and the star.

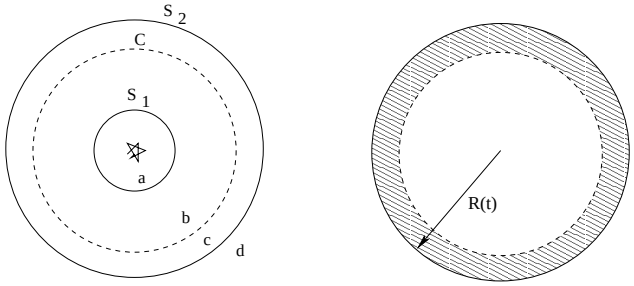

Figure 7.1 Cartoon of the structure of a stellar wind, and its interaction with the ISM. Left: the shock structure within the wind. Right: The outer shell of dense, snowplowed ISM. From Dyson & Williams figures 7.3 and 7.4.

To start, consider the position of this shell, $R(t)$, as a function of time. We start with a force equation. The mass is the ISM shell is the mass that was originally within $R(t)$: $\frac{4\pi}{3}\rho_o R^3(t)$. The force acting on the shell is just the pressure behind it, which drives it outward:

$$\frac{d}{dt}\left(\frac{4\pi}{3}R^3\rho_o\frac{dR}{dt}\right) = 4\pi R^2 p_b \qquad (7.6)$$

if p_b is the pressure, in the outer part of region "b", which acts on the shell. Next, we need an equation for $p_b(t)$. The energy within region "b" – which we will take to be nearly all of the volume within $R(t)$ – is $\simeq 2\pi p_b R^3(t)$; the rate of change of this energy is given by the difference between the input (from the wind) and the "pdV" work done by the expansion:

$$\frac{d}{dt}\left(2\pi R^3 p_b\right) = \dot{E}_{wind} - p_b\frac{d}{dt}\left(\frac{4\pi}{3}R^3\right) \qquad (7.7)$$

if $\dot{E}_{wind}$ is the energy input rate from the wind, assumed constant. These two equations can be combined, for instance by eliminating p_b, to get

$$aR^4\frac{d^3R}{dt^3} + bR^3\frac{dR}{dt}\frac{d^2R}{dt^2} + cR^2\left(\frac{dR}{dt}\right)^3 = \frac{3}{2\pi}\frac{\dot{E}_{wind}}{\rho_o} \qquad (7.8)$$

where a, b, c are numerical constants. Noting that each term on the left hand side has the same dependence on R and t ($\sim R^5t^{-3}$), we can try a power law solution, $R(t) = At^\alpha$. A small bit of algebra tells us that $\alpha = 3/5$ is the allowed solution, and we can also find an expression for A. Thus, the solution is

$$R(t) = 0.76\left(\frac{\dot{E}_{wind}}{\rho_o}\right)^{1/5} t^{3/5} \qquad (7.9)$$

and

$$v(t) = \frac{dR}{dt} = 0.46\left(\frac{\dot{E}_{wind}}{\rho_o}\right)^{1/5} t^{-2/5} \qquad (7.10)$$

Thus, the shell decelerates with time, as it ought to if it is picking up more and more ISM. One can then use these, and (7.6), to work out the pressure acting on the outer shell:

$$p_b(t) = \frac{7}{25}A^2\rho_o t^{-4/5} \qquad (7.11)$$

so that the outer pressure drops with time. Finally, one can also work out the kinetic energy of the shell, which must give the energy input to the general ISM from the wind. This turns out to be

$$\frac{2\pi}{3}R^3\rho_o\left(\frac{dR}{dt}\right)^2 \simeq 0.2\dot{E}_{wind}t \qquad (7.12)$$

so that about 20% of the wind energy goes to the ISM. (The rest goes to heating the bubble, and to "pdV" work).

7.1.3 What about the inner shock?

The location of the inner shock, S_1, is determined by a combination of the jump conditions, applied at S_1, and the pressure at the outside of the region, p_b, as follows.

• The jump conditions, at S_1, are $\rho_{sb} = 4\rho_{sa}$, if "sb" and "sa" subscripts refer to postshock (region "b") and preshock (region "a"), respectively. Also, $v_{sb} = v_{sa}/4$ – here we assume $\mathcal{M} \gg 1$, the strong shock limit, and also an adiabatic shock. At the shock, momentum conservation[1] tells us that $\rho v^2 + p$ is conserved; thus,

$$p_{sb} = \frac{3}{4}\rho_{sa}v_{sa}^2 + p_{sa} \simeq \frac{3}{4}\rho_{sa}v_{sa}^2$$

where we have used the fact that $v_{sa} \gg c_s$ in the wind region.

• In region "b", where we can ignore gravity (compared to the internal energy, $\frac{3}{2}p$), the momentum equation is

$$\rho v\frac{dv}{dr} + \frac{dp}{dr} \simeq 0$$

[1] Check back to chapter 4 from last term, P425; equation 4.4.

and, since $v \ll c_s$ in most of region "b", $\rho \simeq$ constant, and we have $\frac{1}{2}\rho v^2 + p \simeq$ constant. Now – we know that v drops from $v_{sb} = v_{sa}/4$, at S_1, to $v_{s2} \propto t^{-2/5} \ll c_{s,sb}$ at S_2 (from the shock conditions). Thus, we connect the conditions just past S_1 to the conditions at S_2 (remembering that we called the pressure in region "b" at S_2, p_b; ugly notation, I agree!),

$$p_b \simeq p_{sb} + \frac{1}{2}\rho_{sb} v_{sb}^2 = \frac{7}{8}\rho_{sa} v_{sa}^2$$

Thus, p_b, at S_2, is set by the dynamic pressure at S_1. If p_b is also set by external conditions – say the ISM pressure – then the location of S_1 must be where $\rho_{sa} v_{sa}^2$ satisfies the above condition.

• But we can find this location. We note, again, that the dynamic pressure in the wind region, $\rho v^2 \propto \dot{M} v / r^2$ drops approximately as $1/r^2$ (since v is slowly varying). Thus, the behavior of the dynamic pressure within the wind region is fixed by the basic wind solution. Thus, the shock S_1 must form where ρv^2, as determined by the wind solution, matches $\sim \frac{8}{7} p_b$.

7.2 Supernova remnants

First, set the stage: a star explodes. You probably recall the basic picture: stars meet a violent death about 4-5 times per century in a galaxy the size of ours. There are two possible types of supernovae. **Type I supernova** (more correctly Type Ia) arise from the explosion of a white dwarf in a close binary system, presumably initiated by sudden mass transfer from the companion which leads to a thermonuclear reaction in the dwarf star.[2] **Type II supernovae** come from evolved, massive stars in which nuclear burning has run out of fuel. The stellar core collapses inwards and bounces, producing the explosion.

For the purposes of these notes, both types of SNe result in very similar remnants. The dynamics of the ejected material will be sightly different at very early times, due to the differing local environments; but after a short time, both can be treated similarly. That is the approach we will take here. Think of an instantaneous release of energy ($E \sim 10^{51}$ ergs) from a point source, in ambient gas of density ρ_o. The energy released will heat the gas near the explosion to very high temperature and pressure, driving an expansion. This expansion will be very supersonic, setting up a spherical shock wave moving into the surroundings and sweeping up gas as it goes.

This picture is similar to our previous model of an expanding stellar wind bubble; but different in some respects. First, obviously, is the δ-function nature of the explosion. We expect that to change details, for instance the power-law evolution of the radius of the shock. In addition, we have to consider the radiative cooling rate at the outer shell. With stellar winds, we argued that the outer shock is radiative (isothermal); this is justified (after the fact) by the high densities and slow expansion speeds of the wind system. For SNR, however, the situation is different. They can occur in lower density regions (for instance an HII region?), and have much higher explosion speeds (thus much hotter post-shock temperatures). We must consider two phases, then: (a) an early **energy-conserving phase**, during which radiative losses are unimportant, and (b) a later **snowplow phase**, in which the shell becomes dense and cool, and the remnant evolves by momentum conservation.

7.2.1 Early: energy conserving (Sedov) phase.

The remnant in this stage can be thought of as a large hot bubble, filled with ambient gas that has been through the outer, strong shock. Again, let the radius of the outer shock be $R(t)$. In this case, both the internal and kinetic energies per mass are given by

$$e = \frac{1}{2}v^2 = \frac{9}{32}\dot{R}^2 \tag{7.13}$$

This is a direct post-shock result, and holds in a fixed reference frame – one in which the shock is advancing into a medium at rest. We will also assume that gradients within the bubble are small so that (7.13) holds throughout. (This latter is an OK, but not wonderful, assumption – *cf.* Figure 7.2.) The total energy is then

$$E_{tot} = \frac{4\pi}{3} R^3 \rho_o \left(e + \frac{1}{2}v^2 \right) = \frac{3\pi}{4}\rho_o R^3 \dot{R}^2 \tag{7.14}$$

But now, we require $E_{tot} = E_{SN}$, the input energy from the SN. We thus have an equation of motion for the shock:

$$R^3 \dot{R}^2 = \frac{4}{3\pi}\frac{E_{SN}}{\rho_o} \tag{7.15}$$

This solves to

$$R(t) = \left(\frac{25}{3\pi}\right)^{1/5} \left(\frac{E_{SN}}{\rho_o}\right)^{1/5} t^{2/5} \tag{7.16}$$

[2] Current work splits this hair. The distinction between Type I and Type II SNe was originally based on the presence, or absence, of strong hydrogen lines in the explosion spectra. Originally, all stars without H lines were thought to be explosions of white dwarfs; now I gather Type Ib and Ic are identified with core collapse.

and

$$V(t) = \dot{R}(t) = \frac{2}{5}\left(\frac{25}{3\pi}\right)^{1/5}\left(\frac{E_{SN}}{\rho_o}\right)^{1/5} t^{-3/5} \tag{7.17}$$

Compare these to (7.9, 7.10) for a stellar wind: note the different functional dependence on time.

To go further, the basic fluid equations can be used to determine the structure of the interior of the hot bubble. It is the well-known Sedov-Taylor solution. I do not reproduce it here, but show the results (which must be done numerically) in Figure 7.2.

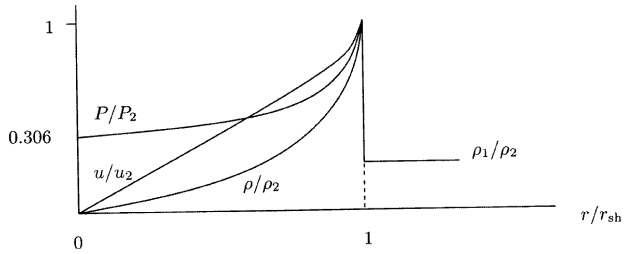

Figure 7.2 Solution for the interior structure of the hot bubble in the energy-conserving phase of the SNR. The radius is scaled to $r_{sh} = R(t)$ in our notation; this solution preserves its functional shape as the remnant expands. From Shu Fig. 17.3.

7.2.2 Late: momentum conserving (snowplow) phase.

At some point the outer shell will cool; as its pressure drops it will be compressed by the hot, expanding inner gas. This is reminiscent of the outer shell in the wind case, but with differences. Here, the dense shell contains only a small fraction of the ISM which was initially inside $R(t)$. In addition, the interior hot bubble is not receiving ongoing energy input, so the pressure on the outwards shell is dropping quickly with time.

The basic analysis of this late-time remnant, then, is based on the simple assumption of momentum conservation. Say the remnant enters this phase at some time t_o, when it has radius R_o and velocity $\dot{R}_o$. Conservation of momentum has,

$$\frac{4\pi}{3}R^3\rho_o\dot{R} = \text{constant} = \frac{4\pi}{3}R_o^3\rho_o\dot{R}_o \tag{7.18}$$

and this integrates to

$$R(t) = R_o\left[1 + 4\frac{\dot{R}_o}{R_o}(t-t_o)\right]^{1/4} \propto t^{1/4} \tag{7.19}$$

and

$$V(t) = \dot{R}(t) = \dot{R}_o\left[1 + 4\frac{\dot{R}_o}{R_o}(t-t_o)\right]^{-3/4} \propto t^{-3/4} \tag{7.20}$$

where the last proportionality statements are valid for $t \gg R_o/\dot{R}_o$. Thus, comparing this to (7.16,7.17) shows that in the momentum-conserving phase, the remnant expands more slowly . . . as one would expect, right?

Dyson & Williams present some numbers. The Sedov phase has $R(t) \simeq 3.6 \times 10^{-4} t^{2/5}$ pc, and $V(t) \simeq 4.4 \times 10^9 t^{-3/5}$ km/s (if t is in seconds). Strong cooling typically takes over for $\dot{R}_o \sim 250$ km/s, giving $R_o \sim 24$ pc and an age $\sim 30 \times 10^4$ years. At that point, about $1400 M_\odot$ of ISM has been swept up – much larger than the mass initially ejected (something like $4M_\odot$). Thus, we are looking almost entirely at the dynamics of the ISM which was dramatically heated in the star's explosion.

7.3 Plerions, a.k.a. pulsar wind nebulae

The preceding section discussed the "classical" picture of supernova remnants, in which energy is injected *at one instant* into the ambient ISM. Many galactic SNR are well described by this model. But not all: in some cases the exploding star leaves behind an active pulsar as its remnant. We now know that many (most? all?) pulsars drive a relativistic wind out from the star. The wind acts as a source of mass and energy, filling the interior of the SNR with hot (or relativistic) plasma. These "filled" SNR used to be called *plerions* (mostly in the SNR community); these days they are being called *pulsar wind nebulae* (in the ISM/X-ray community). As far as I know the two terms refer to the same type of object.

We have direct evidence of pulsar winds. For older pulsars (those not currently within SNR), we see structures in the nearby ISM which are clearly *bow shocks* associated with the star's high-speed motion through the ISM. From the standoff distance of the bow shock we can estimate the wind energy, and compare it to standard models of the pulsar. For young pulsars (those still within their SNR), recent CHANDRA images directly reveal the outflow from the pulsars (at this point there is a good handful of such images; the Crab and Vela pulsars are the most famous). These outflow are complex: they show *jets*, which presumably come out along the star's rotation axis (this is the only symme-

try axis in the system), and *equatorial winds*, which probably arise from the combined effects of the star's strong magnetic field and its rapid rotation.

When a relativistic wind hits the local ISM (which may be the ejecta from the SN), we expect strong shocks to form. Such shocks may be effective at accelerating particles in the local plasma to relativistic energies (we will discuss this in detail later in the course). Thus, the material which has been through the shock may contain a large fraction of relativistic particles – which could, for instance, maintain the nonthermal emission from the surrounding nebula (Crab or Vela). The dyanmics of such a system – assuming it's close to spherically symmetric – are of course very similar to the dynamics of a stellar wind hitting the ISM.

Key points

- Solar-wind solutions (from last term);
- The outer wind shock and getting its location from dimensional analysis;
- Supernova remnants, Sedov and snowplow phases;
- Plerions, what they are, how they work qualitatively.

8 Relativistic particles in astrophysics

Based on our knowledge of cosmic rays and, less directly, of the relativistic electrons which radiate in synchrotron sources, our general picture is of particles which are highly relativistic – that is, with $\gamma = E/mc^2 \gg 1$ – but which are also tied to the background, thermal plasma in which they find themselves. The distribution function of these particles is found to be a power law, rather than a Maxwellian; thus, these particles cannot have had time to thermalize (in the two-body sense). In this section, we will consider the dynamics of these particles and how they are tied to the background plasma). Later on we'll address how they may be accelerated.

8.1 Recap: basics for relativistic particles

First, basic special relativity. The total total energy of a relativistic particle is given by $E^2 = p^2c^2 + m^2c^4$; we also have the definition $E = \gamma mc^2$, where $\gamma^2 = 1/(1-\beta^2)$, and $\beta = v/c$. In the limit $E \gg mc^2$, we also have $E \simeq pc$ (which is exactly true for a photon, of course; as the particle gets more relativistic, its rest mass becomes less and less important). Note that the Lorentz factor γ is often used to represent the energy; the mc^2 factor is implicitly carried along if you need "real" units.

Power-law distribution functions (motivated by cosmic rays) are often used:

$$f(E) = f_o E^{-s} , \quad E_1 \le E \le E_2 \qquad (8.1)$$

or

$$n(\gamma) = n_o \gamma^{-s} , \quad \gamma_1 \le \gamma \le \gamma_2 \qquad (8.2)$$

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. That is, the total number of particles will be $N = \int f(E)dE = \int n(\gamma)d\gamma$. It follows, then, that $f(E)$ and $n(\gamma)$ have different units – watch out, this difference can bite you.[1]

[1]To be specific: the two DF's are equivalent if

$$f(E)dE = n(\gamma)d\gamma ; \quad f(E) = n(\gamma)d\gamma/dE$$

(so that we have the same number of particles "at $E = \gamma mc^2$"). For the specific power law case, above, this gives

$$f_o E^{-s} dE = n_o \gamma^{-s} d\gamma ; \quad f_o = n_o (mc^2)^{s-1}.$$

Remaining question to the reader: how is n_o related to the total number of particles N?

8.2 Quick overview of the observations

We have seen that astrophysical plasmas contain two species. One species is the thermal interstellar gas which we have been considering. In this context, its important properties are that it seems to be well described by a thermal equilibrium distribution function (a Maxwell Boltzmann velocity distribution), and that the energy per particle is subrelativistic. (Recall temperatures range from $O(10)$K to $O(10^6)$K).

In addition, many astrophysical plasmas – including the galactic ISM – contain a significant population of highly relativistic particles which are *not* in a thermal distribution. We saw last term that we have direct and indirect evidence of these particles; I'll review the arguments here.

Baryons. Here we have direct evidence – these are the cosmic rays. They are mostly protons, but there is a heavy element component, with approximately solar abundances (so they come from processed material). They are very isotropic in arrival direction, probably at all energies (although arrival directions for the very highest energy particles remain uncertain).

• The baryon 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.

Leptons. Here we have some direct evidence – the lepton component in the cosmic ray spectrum can be separated from the baryon component. The cosmic ray lepton distribution falls much more steeply than the baryon distribution, above energies ~ 1 GeV, so that its total contribution to the CR energy density is only $\sim 1\%$ that of the baryons.

In addition, there is also abundant *indirect* evidence fore highly relativistic electrons[2] throughout the universe. Synchrotron radiation – which we will study in detail in the next chapter – is common in many different settings. Because synchrotron radiation comes from highly relativistic particles (with $\gamma = E/mc^2 \gg 1$), we know immediately that synchrotron sources have a relativistic lepton component. Examples of this

[2]Well, really leptons; we'll discuss later whether we can distinguish electrons from positrons by their radiation signatures.

are the galactic disk; supernova remnants, both standard and "filled"; radio jets from compact stars in X-ray binaries; X- and γ-ray emission from pulsars; radio jets from active nuclei, and the radio lobes they create; diffuse synchrotron emission from the plasma in clusters of galaxies (including a few synchrotron-bright shocks created when two clusters collide); and quasars of course.

Looking ahead to the next chapter, we can note some important characteristics of synchrotron radiation. It requires magnetic fields and highly relativistic electrons. The spectrum from a single particle, with energy $E = \gamma mc^2$, peaks at a photon frequency

$$\nu_{sy} \sim \frac{3}{4\pi}\gamma^2 \frac{eB}{mc} \tag{8.3}$$

Thus, for a uniform B field, one particle energy γ maps directly to one (observed) photon frequency, ν. This single, radiating particle has an energy loss rate

$$\frac{dE}{dt} \simeq \frac{4}{3} c\sigma_T \gamma^2 \frac{B^2}{8\pi} \tag{8.4}$$

where $\sigma_T = 6.65 \times 10^{-25}\text{cm}^2$ is the Thompson cross section. Synchrotron radiation commonly has a power law spectrum, which tells us (assuming the B field is simple) that the underlying electron distribution is also power law (just as it is in galactic cosmic rays). Typical values for the photon spectrum are $j(\nu) \propto \nu^{-\alpha}$, with $\alpha \sim 0.5 - 1.0$; the associated electron distribution is $n(\gamma) \propto \gamma^{-s}$, with $s \sim 2.0 - 3.0$.

From(8.4), we see that higher energy electrons lose energy faster that the lower energy ones, because $dE/dt \propto E^2$; from this one can show that an initial electron power law spectrum will develop a break at the energy where the particle's radiative lifetime equals the age of the plasma. As the plasma gets older, this break will move to lower energies (and thus to lower photon frequencies).

8.3 Cosmic rays in the galactic setting

One of the big questions is, how are cosmic rays accelerated to such high energies, and how are they maintained there (why don't they eventually thermalize with the ISM)? Before we go there, let's recap a few points about cosmic rays from last fall (P425 notes).

Sources are thought to be two-fold. *Supernova remnant*s have long been thought to be the main source of CR; I personally suspect that *pulsars* are probably also an important source. Still another possibility is that the highest energy CR may have an *extragalactic origin*.

Propagation and Trapping. Once generated, CR do not just fly freely through space. Becuase they are charged, they are connected to the ISM by their gyromotion, and by scattering on turbulent Alfven waves in the ISM. 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. From nuclear abundances we learn that galactic CR typically spend most of their 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.

Losses. The leptons, being of smaller mass, are susceptible to radiative losses (synchrotron 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.

8.4 Particle acceleration, overview

Next, we consider the question of particle acceleration: what is the origin of cosmic rays (which we observe directly at earth), and of the relativistic electrons in such sources as supernova remnants and radio jets (which we observe indirectly *via* their synchrotron radiation)? How are these charged particles accelerated to relativistic energies, and how are they maintained in a non-Maxwellian distribution? The short answer is, this must be done by electric fields, $\mathcal{E} \neq 0$. Magnetic fields do no work on the particles, and gravity is a conservative force (so that any energy a particle gains by going into a gravitational potential well must be lost again when it leaves the well). However, is it not easy to maintain large-scale electric fields in space, where the abundant free charges in astrophysical plasmas will want to short out any static field. Two types of particle acceleration models exist — which I tend to call "first stage" and "second stage" mechanisms.

8.5 Particle acceleration, first stage mechanisms

One possibility is that an ordered, large-scale $\mathcal{E}$ field is maintained in the region of some massive object, like a star (thinking of solar flares) or a pulsar or an accretion disk, where dynamic processes can maintain a strong

dynamo; these are called first stage mechanisms. In this situation there is no minimum energy threshold for particle acceleration; any charged particle dropped in a region with $\mathcal{E} \neq 0$ will be energized. There may, however, be an upper limit to the energy that can be reached – due to the physical size of the system (and the consequent limit on the overall potential drop, $\int \mathcal{E} \cdot d\mathbf{s}$). There are a couple basic types of models here: reconnection sites and unipolar dynamos.

8.5.1 Magnetic reconnection

You saw this last term; I'll just store a recap here. Consider a region in a plasma in which the direction of the magnetic field reverses over a small spatial scale – for instance, the neutral sheet in the earth's magnetotail, or the base of a flux tube footed in the solar photosphere. The magnetic field in this region will find a lower-energy state by "reconnecting", that is by changing the magnetic field configuration. Such a region can be the site of rapid conversion of stored magnetic energy to kinetic and internal energy of the plasma, Reconnection is believed to be important in solar flares, and has been observed in the earth's magnetotail.

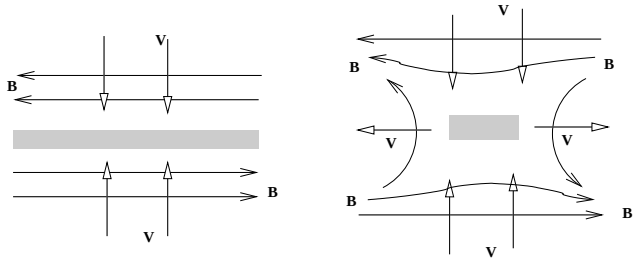

Figure 8.1 The geometry of magnetic reconnection. Left, a simple system before reconnection; the field lines have opposing directions and meet in a central current sheet (called a neutral sheet), shaded in this figure. Right, with reconnection ongoing; some field lines have changed their topology, and plasma is driven out of the neutral sheet across the reconnected field lines. Think: which way do the current and $\mathbf{E}$ field point in the neutral sheet?

The reason reconnection is important for particle acceleration, is that the neutral sheet (where the magnetic field goes through zero) contains an ordered electric field, of strength

$$\mathcal{E} = -\frac{c}{4\pi\sigma}\nabla \times \mathbf{B} - \frac{1}{c}\mathbf{v} \times \mathbf{B} \qquad (8.5)$$

(This is just Ohm's law, combined with Maxwell's equations; check the section earlier on flux freezing). Here, σ is the electrical conductivity (not a cross section). In the cartoons above, the $\mathcal{E}$ field will be in the plane of the neutral sheet and perpendicular to the paper. The maximum particle energy that can be reached, then, is set by the potential drop across the neutral sheet, $\Delta V = \int \mathcal{E} \cdot d\mathbf{l}$.

8.5.2 Unipolar dynamos

This is a jargon-word for massive, rotating, compact objects – pulsars and accretion disks – which can support strong fields. (Refer back to Fig 9.3 of your P425 notes.) If a magnetic field is tied to the rotating matter or plasma (say by a high conductivity, so that the field is approximately flux-frozen), the rotation will cause a field $\mathcal{E} = \frac{1}{c}\mathbf{v}_{rot} \times \mathbf{B}$ to be seen by a local observer.

The numbers that are in principle reached by rotating compact objects are impressive. For pulsars, flux-freezing considerations estimate $B \simeq 10^{12}$ G if the field was originally the stellar field. The implied electric field can accelerate a particle to energies of 10^{16} eV if it operates over a distance $\approx R_{ns} \approx 10$ km. For an accretion disk around a black hole, the relevant length scale is $r_g = GM_{bh}/c^2$, the magnetic field can be 10^3 G, and the inner parts of the accretion disk rotate at nearly lightspeed. In this case the potential drop can reach 10^{19} V.

On the other hand, it is not obvious that these high fields can be maintained – in fact it's very unlikely. The picture we have presented of a rotating, magnetized object is a vacuum argument. But the atmospheres of such objects are likely to have some amount of free charge, which may well short out a large part of these maximum predicted potential drops. For one example, think about a relativistic particle being accelerated along the magnetic field line in a pulsar by these strong $\mathcal{E}$ fields. As it gets accelerated, it radiates – and the γ-ray photons it radiates can pair produce in the pulsar's strong magnetic field. The newly created electron-positron pairs shield most of the rotation-induced $\mathcal{E}$ field ... leading to a net particle energy only $\sim 10^{-6}$ or so of the maximum predicted by the simple order-of-magnitude estimates.

These two families of models, reconnection and unipolar dynamos, represent most of the models of particle acceleration by large-scale electric fields. What are the important characteristics of these models, as far as their relevance to particle acceleration? First, all charges which see $\mathcal{E}$ are accelerated. Second, the maximum energy that can be reached is given by the potential drop; but real-world effects, mostly related to

local plasmas (shorting out part of the potential drop, or providing turbulence which can stop the speeding particles before they reach $E \sim e\Delta V$). Third, due to the effects just listed, much of the available energy goes to heating the plasma rather than accelerating a small fraction of the charged particles to high energies. The final values of the output particle energies depend on these details, of course. Observations of solar and magnetospheric reconnection suggest that some fraction of the charged particles are, indeed, accelerated to well above thermal energies; but not to relativistic energies. (The output energy may well be larger in pulsars and galactic-nucleus accretion disks; but we can only work from inference there). Thus, these mechanisms are often thought of as a "first stage" in the acceleration process; stochastic methods may take these "seed" particles and boost them to much higher energies.

8.6 Particle acceleration, second stage mechanisms

The other class of acceleration mechanisms invokes stochastic electric fields, such as are found in plasma turbulence. If the particles couple efficiently to the turbulent $\mathcal{E}$ fields, they can gain energy from the turbulence. There are again a couple of versions of this: true turbulence (disordered plasma motions, for instance somewhere in the dynamic ISM), or shock acceleration (using shock physics to drive the turbulence).

The fundamental idea of stochastic particle acceration was invented by Fermi in 1949; his physical picture was rather naive, but described the basic stochastic mechanism quite clearly. More recent work has improved the physical description of the scattering mechanism, while retaining the basic idea.

We have already seen the most likely scattering mechanism: resonant interaction with Alfven waves. In this situation there is a minimum particle energy which can interact with the waves, as we saw last term. Therefore, these models can be thought of as second stage mechanisms, in that they take "seed" particles created by first-stage mechanisms and accelerate them to very high energies. The maximum energy which can be reached here is also finite, and again tied to the size of the system (which determines the maximum Alfven wavelength which can exist, and also determines the rate at which particles can leak out of the acceleration region). But these upper limits can be higher, in some astrophysical settings, than those provided by first-stage mechanisms.

8.6.1 Fermi acceleration

This is the original picture. Think back to last term, when we talked about "magnetic mirrors". That is; for a particle moving in a region of spatially changing magnetic field, the magnetic moment, $\mu = p_\perp^2/2\gamma mB$ is conserved, as long as the time during which the particle sees the field to change is long compared to the particle's gyroperiod. But since $p^2 = p_\perp^2 + p_\parallel^2$ is fixed in the absence of external forces, $p_\perp^2 \leq p^2$. Thus, there is a maximum value of B which is allowed; the particle is kept out of high-field regions. This effect is called magnetic mirroring.

Now, to apply this to cosmic rays, envision a field of randomly moving mirrors (we can think of them as interstellar clouds, or as clumps of high magnetic field) on which the relativistic particles or cosmic rays scatter elastically. We want the effect of many collisions with these mirrors, on a particle distribution. Let the mirrors have random velocity v_m, and average spacing L. If a particle is moving faster than the mirrors, it can undergo either overtaking or head-on collisions. In a single collision, the particle (mass m, velocity v) suffers velocity and kinetic energy changes,

$$\begin{aligned} &\text{head}-\text{on}: \\ &\quad \Delta v \simeq 2v_m; \;\; \Delta E \simeq 2mv_m(v+v_m) \\ &\text{overtaking}: \\ &\quad \Delta v \simeq -2v_m; \;\; \Delta E \simeq 2mv_m(v-v_m) \end{aligned} \tag{8.6}$$

But, the rate at which the particle suffers headon collisions is $\simeq (v+v_m)/L$; its rate of overtaking collisions is $\simeq (v-v_m)/L$. Thus, the net rate of change of the particle's energy is

$$\begin{aligned} \frac{dE}{dt} \simeq &\frac{(v+v_m)}{L} 2mv_m(v+v_m) \\ &- \frac{(v-v_m)}{L} 2mv_m(v-v_m) \end{aligned} \tag{8.7}$$

and this collects to the final form,

$$\frac{dE}{dt} \simeq 16 \frac{E}{t_{coll}} \left(\frac{v_m}{v}\right)^2 \tag{8.8}$$

where we have written $t_{coll} = L/v$. While I wouldn't take the factor 16 too seriously, the fundamental form is: $dE/dt \propto E/t_{coll}$, which is the standard result for Fermi acceleration.

A couple of features are worth noting. One, this is a test particle approach; the energy loss of the mirrors is not taken into account. Two, the fractional energy gain per collision is small, since $v_m \ll v \simeq c$; thus, the acceleration time $t_{acc} \sim E/(dE/dt) \sim (v/v_m)^2 t_{coll}$.

What particle spectrum is predicted by this simple model? One way to find this is as follows. From the kinematics above, we found that the average energy gain per collision is $\Delta E \sim E(v_m/c)^2$ if the particles are relativistic. Thus, after p collisions, a particle which started at E_o will have energy E_p after p bounces:

$$E_p \simeq E_o \left(1 + \frac{v_m^2}{c^2}\right)^p \tag{8.9}$$

which can be written,

$$\ln \frac{E_p}{E_o} = p \ln \left(1 + \frac{v_m^2}{c^2}\right) \simeq p \frac{v_m^2}{c^2}. \tag{8.10}$$

Now, let the particles have some chance, η, of escaping from the acceleration region (something must end the acceleration process, after all!). If the escape time $\sim \tau$, then $\eta \sim t_{coll}/\tau$. But, the number of particles at E_p, $E_p f(E_p)$, is just the starting number, N_o, times the probability of the particle staying in the system for p bounces:

$$E_p f(E_p) = N_o P(p) = N_o (1-\eta)^p \tag{8.11}$$

which we can write as

$$\begin{aligned} \ln[E_p f(E_p)] &= \text{constant} + p \ln(1-\eta) \\ &\simeq \text{const} - p\eta \end{aligned} \tag{8.12}$$

Combining this with (8.10), eliminating p and dropping the p subscript, we find the predicted spectrum:

$$f(E) \propto E^{-(1+\eta c^2/v_m^2)} \tag{8.13}$$

Thus, this model – Fermi acceleration plus a constant probability of escape from the system – results in a power law spectrum, as is observed in cosmic rays and elsewhere.. However, as this model stands, the exponent of the power paw, $s = 1 + \eta c^2/v_m^2$, is a sensitive function of the mirror parameters and the escape time. This is not very appealing, since the observed values of s fall in the range $2 \lesssim s \lesssim 3$ almost everywhere (cosmic rays, supernova remnants, radio galaxies, etc.).

8.6.2 Plasma turbulence: Alfven waves

Modern work has replaced Fermi's picture of moving magnetic mirrors with small-scale structures which can scatter particles: most likely these are Alfven waves.

I'm sure you remember these waves, from last term. 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. Their dispersion relation is $\omega = k v_A$, with $v_A = B/\sqrt{4\pi\rho}$. Charged particles interact resonantly with Alfven waves. A particle moving along $\mathbf{B}$ at some velocity v sees a Doppler shifted frequency $\omega' = \omega(1 - v/v_A) = \omega - kv$. 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. But 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

$$\omega - kv = \pm\Omega(\gamma) \tag{8.14}$$

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

$$r_L(\gamma) = \gamma \frac{mc^2}{eB} \simeq \lambda_{res}(\gamma) \tag{8.15}$$

Numerically, note that particles with $\gamma \sim 10^3$ in a field $B \sim 1$ μG have a gyroradius – and thus a resonant Alfven wavelength – on the order of an AU.

If the Alfven waves are turbulent (say, just arising from a turbulent background plasma, such as the ISM), the picture above carries over directly; v_m becomes v_A, and the concept of a scattering length L must be replaced with the wave energy density and some measure of the scattering cross section, similar to that used above. This process is slow in many applications; it has $dE/dt \propto v_A^2/c^2$ (and thus is classically a "second order Fermi process"). In addition, because it relies on the wave-particle resonance, a particle at energy γ only sees waves at $\lambda_{res}(\gamma)$ – which can be a small fraction of the total energy in the turbulence.

8.6.3 Turbulent shock acceleration

A faster version of turbulent acceleration ties the turbulence to shock fronts. A shock must have $v_2 < v_1$ (the post-shock velocity must be less that the pre-shock

velocity; right?). Thus, in the frame of the shock, the flows on each side are converging. If the pre-shock and post-shock plasma also carry Alfven turbulence, a particle trapped in the shock region and bouncing back and forth will undergo Fermi acceleration, but will suffer only head-on collisions. With this picture, we can specify the parameters in the above Fermi analysis. We need two facts:

• The escape probability is the ratio of downstream to upstream fluxes: $\eta \simeq 4n_{rel}v_2/n_{rel}v = 4v_2/c$.

• The energy gain of a particle per scattering turns out to be

$$E' = E\frac{[1 + v_{e1}(v_1 - v_2)\cos\theta_{e1}/c^2]}{[1 + v_{e2}(v_1 - v_2)\cos\theta_{e2}/c^2]} \qquad (8.16)$$

if v_{e1}, v_{e2}, θ_{e1} and θ_{e2} are the velocities and angles of the particle (electron, say), before and after it is scattered. After p bounces, we have

$$\ln\frac{E_p}{E_o} \simeq \frac{4}{3}p\frac{(v_1 - v_2)}{c} \qquad (8.17)$$

With these, we can collect the algebra as above and find the spectrum predicted by this model:

$$f(E) \propto E^{-s}; \quad s = \frac{v_1 + 2v_2}{v_1 - v_2} \qquad (8.18)$$

Now, for a strong shock, $v_1 = 4v_2$, which predicts $s = 2$; higher s values will be produced if the velocity jump in the shock is < 4.

8.6.4 Energy limits for Alfven acceleration

What are the limits on particle energies that can be scattered and accelerated by Alfven waves? The particle-wave resonant condition (8.15) makes this easy to answer. For the lower limit, we noted last term that Alfven waves can only exist for frequencies $\omega < \Omega_p = eB/m_pc$, the proton gyrofrequency. This translates to a lower limit on the particles energies that can "see" the waves (given in chapter 8 of P425 notes).

The highest particle energy which can possibly be accelerated by Alfven waves is determined by the maximum wavelength that can exist in the system (which can't be any larger, clearly, than the physical size of the system). In practice, however, particle acceleration is limited by losses occuring at the same time as the acceleration. If the losses are due to synchrotron radiation, higher energy particles lose their energy more rapidly (equation 8.4); this leads to a second upper limit on the energy range of accelerated particles.

Key points

• Cosmic rays: in the galactic setting

• Relativistic electrons: indirectly "seen"

• First stage particle acceleration: where, what energies?

• Fermi acceleration, "classical"

• Alfven wave acceleration, shock and/or turbulent

9 Synchrotron radiation

A single particle, undergoing gyromotion in a magnetic field, is of course accelerated; and thus it will radiate. When the particle is subrelativistic, this process is called cyclotron radiation; when the particle is relativistic, the process is called synchrotron radiation. We have already studied bremsstrahlung, a fundamental continuous radiation mechanism which is important for thermal astrophysical plasmas. Synchrotron radiation is the other common and important continuous emission process. It is imortant for magnetized astrophysical plasmas which contain a significant component of relativistic electrons (or, yes, positrons) – commonly called a "nonthermal" plasma.

9.1 Total power

We start with the total power radiated by the particle, over all frequencies. Let the particle have acceleration $\mathbf{a}$, with components $a_\parallel$ along its direction of motion ($\mathbf{v}$), and $a_\perp$ across the direction of motion (NOT the direction of the magnetic field, here). In the particle's rest frame, the total power radiated as a function of time is

$$P(t) = \frac{2}{3}\frac{e^2}{c^3}|\mathbf{a}(t)|^2 \tag{9.1}$$

In the observer's frame, this becomes

$$P(t) = \frac{2}{3}\frac{e^2}{c^3}\gamma^4\left(a_\perp^2 + \gamma^2 a_\parallel^2\right) \tag{9.2}$$

Now, for gyromotion, we have $a_\parallel = 0$ and $a_\perp = eBc\beta_\perp/\gamma mc$. Let the particle have pitch angle θ – so that $\beta_\perp = \beta\sin\theta$. Then, the net power seen by the observer, per particle at energy γ and pitch angle θ, is

$$P_{sy} = \frac{2}{3}\frac{e^4B^2}{m^2c^3}\gamma^2\beta^2\sin^2\theta \tag{9.3}$$

You should note that this expression assumes $\gamma \gg 1$. From an ensemble of particles, with an isotropic distribution of pitch angles, we note that $\langle\sin^2\theta\rangle = 2/3$; thus the average single-particle power is

$$\langle P_{sy}\rangle = \frac{4}{9}\frac{e^4B^2}{m^2c^3}\gamma^2\beta^2 \tag{9.4}$$

Finally, we note that (9.4) can be rewritten in terms of the magnetic energy density, $u_B = B^2/8\pi$, and the Thompson scattering cross section, $\sigma_T = (8\pi/3)(e^2/m_ec^2)^2$, as

$$\langle P_{sy}\rangle = \frac{4}{3}c\sigma_T\gamma^2\beta^2u_B\ . \tag{9.5}$$

9.2 Single particle spectrum

From the discussion of bremsstrahlung, recall: the power as a function of time, $P(t)$, reflects the time dependence of the acceleration, $a(t)$; the distribution of this radiation over frequency – the spectrum – reflects the power (Fourier) spectrum of $a^2(t)$. In the synchrotron case, the *observed* radiation varies with time due to relativistic beaming effects, in addition to the fundamental gyromotion. The power spectrum of the observed $P(t)$ turns out to contain power at frequencies much higher than $\Omega/2\pi$; this determines the observed synchrotron emission spectrum.[1]

• **First, guesstimate the answer.** The important fact here is that the radiation from a relativistic particle is strongly beamed in the direction of the particle's motion. A nonrelativistic particle radiates in a dipole pattern, with the dipole axis aligned with the acceleration $\mathbf{a}$. However, for a relativistic particle, this dipole is squeezed into a narrow forward cone, aligned with the motion ($\mathbf{v}$) and with opening angle $\simeq 1/\gamma$.

So, think about a relativistic particle in gyromotion. You – as an observer – only see the particle's radiation once per orbit, when the particle's velocity is within $1/\gamma$ of your line of sight. Say that the beam stays within your line of sight for a time interval, Δt^{obs}. You will then see a very narrow pulse, of duration, Δt^{obs}, once every gyroperiod. From our experience with Fourier transforms, we know that most of the radiated power must appear at a frequency $\sim 1/\Delta t^{obs}$, rather than just at Ω (as you might naively guess if you didn't think about the beaming effects).

We can estimate Δt^{obs} from geometry. Let the part of the particle's orbit for which you see the radiation have length $\Delta s = a\Delta\phi$, where a is the radius of curvature of the path. We can find a from the equation of motion:

$$\frac{\Delta v}{\Delta t} = v^2\frac{\Delta\phi}{\Delta s} = \frac{e}{\gamma mc}vB\sin\theta \tag{9.6}$$

so that $a = \Delta s/\Delta\phi = \gamma v/\Omega_o\sin\theta$. Since $\Delta\phi = 2/\gamma$, we have $\Delta s = 2a/\gamma = 2v/\Omega_o\sin\theta$. Now, you see a pulse of radiation; its emitted duration is $\Delta t^{em} = \Delta s/v$; and the duration you observed is shortened by the light-travel time, so that $\Delta t^{obs} = \Delta t^{em} - \Delta s/c$. Putting all of this together, we find

$$\Delta t^{obs} = \frac{2}{\Omega_o\sin\theta}(1-\beta) \simeq \frac{1}{\gamma^2\Omega_o\sin\theta} \tag{9.7}$$

[1] As usual, $\Omega = eB/\gamma mc = \Omega_o/\gamma$ is the relativistic gyrofrequency.

where in the last step we have used $1/\gamma^2 \simeq 2(1-\beta)$, which is valid when $\gamma \gg 1$.

Thus – we find that the observed pulse has a duration $\simeq 2\pi/\gamma^2\Omega_o$ (converting to Hz); we expect this to be the highest frequency at which there is significant radiated power.

• **Now, do the math.** When the calculation is done more formally, the characteristic frequency for synchrotron radiation turns out to be

$$\nu_c = \frac{3}{4\pi}\gamma^2 \frac{eB}{mc}\sin\theta \tag{9.8}$$

When the power spectrum is evaluated, we find that the radiation can be separated into components which are linearly polarized along and across the magnetic field, *as seen projected on the sky*. These are called $P_\parallel$ and $P_\perp$. Considering the strong forward beaming effect, we would expect the component polarized at right angles to the field to dominate (why? Is this clear to you?) The form of the spectrum comes out to be

$$\begin{aligned} P_\perp(\nu,E) &= \frac{\sqrt{3}}{2}\frac{e^3 B\sin\theta}{mc^2}\left[F\left(\frac{\nu}{\nu_c}\right)+G\left(\frac{\nu}{\nu_c}\right)\right] \\ P_\parallel(\nu,E) &= \frac{\sqrt{3}}{2}\frac{e^3 B\sin\theta}{mc^2}\left[F\left(\frac{\nu}{\nu_c}\right)-G\left(\frac{\nu}{\nu_c}\right)\right] \end{aligned} \tag{9.9}$$

where

$$G(x) = xK_{2/3}(x)\ ;\ \ F(x) = x\int_x^\infty K_{5/3}(x')dx'$$

and the K's are Bessel functions. Since the $F(x)$ and $G(x)$ functions behave similarly, this verifies that the emissivity polarized transverse to the (projected) magnetic field is much stronger than the emissivity polarized parallel to the field.

Clearly, the total power

$$\begin{aligned} P(\nu,E) &= P_\perp(\nu,E) + P_\parallel(\nu,E) \\ &= \sqrt{3}\,\frac{e^3 B\sin\theta}{mc^2}F\left(\frac{\nu}{\nu_c}\right) \end{aligned} \tag{9.10}$$

Both $F(x)$ and $G(x)$ peak at $x \simeq 1$. The asymptotic behavior of $F(x)$ is

$$\begin{aligned} F(x) &\to \frac{4\pi}{\sqrt{3}\Gamma(1/3)}\left(\frac{x}{2}\right)^{1/3} \qquad x \ll 1; \\ F(x) &\to \left(\frac{\pi x}{2}\right)^{1/2} e^{-x} \qquad x \gg 1 \end{aligned} \tag{9.11}$$

The function $G(x)$ behaves similarly. Remembering that $x = \nu/\nu_c$, this also verifies our guess: the radiation peaks at ν_c and falls off exponentially above this. Further, we note that the particle energy $E = \gamma mc^2$ enters the emissivity expressions (9.9) only through the scaling of the argument, ν/ν_c.

9.3 Spectrum from a distribution of particle energies

The next step is to find the spectrum radiated by a distribution of electron energies, $f(E)$. Consider the total emission, the sum of both polarized components, with total power $P(\nu,E) = P_\parallel(\nu,E) + P_\perp(\nu,E)$. We have for the total emissivity per volume,

$$j_{sy}(\nu) = \frac{1}{4\pi}\int P(\nu,E)f(E)dEd\Omega \tag{9.12}$$

We will assume $f(E)$ is isotropic, and integrate over solid angle, in what follows.

Motivated by the observed cosmic ray spectrum, and the photon spectrum seen in many polarized radio sources, we choose the usual power-law particle spectrum,

$$f(E) = f_o E^{-s} \tag{9.13}$$

for the energy range $E_{min} < E < E_{max}$, where $E_{max} \gg E_{min}$ is usually assumed. Note that there must be an E_{min}; (9.13) diverges at low energies. There may well be an E_{max} also – we'll discuss that later.

We put this DF into equation (9.12), and write the single-particle power as $P(\nu,E) = P_o B F(\nu/c_o B E^2)$. From the form of $P(\nu,E)$ and the energy range chosen, note that we expect strong emission in the frequency range

$$\nu_{min} < \nu < \nu_{max} \tag{9.14}$$

where $\nu_{min} = \nu_c(E_{min})$ and $\nu_{max} = \nu_c(E_{max})$. Putting this $f(E)$ into (9.13), we get

$$j_{sy}(\nu) = P_o B f_o \int_{E_{min}}^{E_{max}} E^{-s}F(\nu/c_o B E^2)dE \tag{9.15}$$

Now, by changing the variable of integration from E to $x = \nu/c_o B E^2$, we end up with

$$\begin{aligned} j_{sy}(\nu) =& P_o B^{(s+1)/2} f_o \nu^{-(s-1)/2} \\ &\times \int_{x_{min}}^{x_{max}} x^{(s-3)/2}F(x)dx \end{aligned} \tag{9.16}$$

Now, the integral in (9.16) can be evaluated numerically if $x_{min} \to 0$ and $x_{max} \to \infty$ (which is a reasonable limit for the frequency range in (9.14); Pacholczyk, for instance, has numerical values. Thus, the emissivity in (9.16) can be expressed numerically as

$$\begin{aligned}\epsilon_{sy}(\nu) &= 4\pi j_{sy}(\nu) \\ &= 1.18 \times 10^{-22} a(s) f_o B^{(s+1)/2} \left(\frac{\nu}{2c_1}\right)^{-\alpha}\end{aligned} \tag{9.17}$$

where $c_1 = 6.3 \times 10^{18}$ Hz, $\alpha = (s-1)/2$ is called the *spectral index*, $a(s)$ is an order-unity function (noting the dependence of the x-integral in (9.16) on s), and everything in (9.17) is in cgs units.

Equation (9.17) is good only for the frequency range (9.14). Outside this range, the emissivity will be dominated by that of particles at E_{min} (for $\nu < \nu_{min}$), or of particles at E_{max} (for $\nu > \nu_{max}$). Thus, for $\nu < \nu_{min}$, we expect

$$\text{low } \nu\text{'s}: \quad j_{sy}(\nu) \propto \nu^{1/3} \; ; \tag{9.18}$$

and for $\nu > \nu_{max}$, we expect

$$\text{high } \nu\text{'s}: \quad j_{sy}(\nu) \propto \nu^{1/2} e^{-\nu/\nu_{max}} \; . \tag{9.19}$$

These limits may be repeated in the total spectrum from a plasma – we'll discuss this below.

9.4 Polarization

We noted that, since $P_\perp > P_\parallel$, the radiation from a single particle is linearly polarized. The fractional linear polarization is ususally defined as

$$\pi(\nu) = \frac{P_\perp - P_\parallel}{P_\perp + P_\parallel} \tag{9.20}$$

For a single particle energy, $\pi(\nu) = G(\nu/\nu_c)/F(\nu/\nu_c)$. For a power-law distribution of particle energies, both $P_\perp$ and $P_\parallel$ in (9.20) must be integrated over particle energy; the result is

$$\pi = \frac{\int G(x) E^{-s} dE}{\int F(x) E^{-s} dE} = \frac{s+1}{s+7/3} \tag{9.21}$$

where we have taken $x = \nu/\nu_c$ as above. In evaluating the integrals, we have used the fact that the integrals $\int_0^\infty x^p F(x) dx$ and $\int_0^\infty x^p G(x) dx$ can be expressed in closed form in terms of gamma functions. Thus, a source with $s \simeq 2-3$ will have $\sim 70\%$ polarization.

9.5 Synchrotron self-absorption

This is of course the inverse process – in which a free electron in a magnetic field can absorb a photon. We will treat this by relating the absorption probability to the emissivity, taking stimulated emission into account, using a powerful statistical method developed by Einstein, called Einstein coefficients.[2] In order to do this, we will represent the free electron energy state as a discrete state in a continuum (and after all, even the free electron phase space is quantized, remember), so that the absorption is a transition between a lower state, with energy $E - h\nu$, and momentum $\mathbf{p}(E - h\nu)$, and an upper state with energy E and momentum $\mathbf{p}(E)$.

Referring to the Appendix, we see that the absorption coefficient, at frequency ν, can be written in terms of the populations of all pairs of upper and lower electronic states which are separated by $h\nu$ of energy:

$$\kappa_\nu = \frac{h\nu}{4\pi} \sum_E \left[N(E - h\nu) B_{12} - N(E) B_{21} \right] \tag{9.22}$$

where $N(E) \simeq f(E) dE$ is something like, "the number of electrons in the Eth state", if $f(E)$ is the electron distribution function. Now, we note several facts:

- $B_{12} = B_{21}$ for free electrons, which have the same degeneracy factors $g_1 = g_2$ for the upper and lower states;
- $A_{21} = (2h\nu^3/c^2) B_{21}$;
- $P_{sy}(\nu, E) = h\nu A_{21}$ relates the single particle emissivity to the A_{21} coefficient;

Now, we switch back from describing discrete electron states to a continuous picture:

$$\sum_E n(E) \to \int f(E) dE \to \int f(\mathbf{p}) d^3\mathbf{p}$$

We can thus rewrite (9.22) as

$$\begin{aligned}\kappa_\nu = &\frac{c^2}{8\pi h\nu^3} \int \{ f[\mathbf{p}(E - h\nu) - f[\mathbf{p}(E)]\} \\ &\times P[\nu, E(\mathbf{p})] d^3\mathbf{p}\end{aligned} \tag{9.23}$$

Now, we can use the fact that the photon energy should be small compared to the electron energy, $h\nu \ll E$,

[2] We haven't seen these formally; in case you haven't seen them in another class, I'm putting the important discussion in the Appendix to this chapter.

and expand the difference inside the braces in the integrand:

$$f[\mathbf{p}(E-h\nu)] - f[\mathbf{p}(E)] \simeq \frac{h\nu}{c}\frac{df(p)}{dp} \qquad (9.24)$$

Finally, since we are still assuming an isotropic particle distribution, we can go back to energy space by noting that $d^3\mathbf{p} = 4\pi p^2 dp = 4\pi E^2 c^{-3} dE$ (where the latter assumes the particles are highly relativistic). This gives us a general expression for the synchrotron absorption from a distribution of electrons,

$$\kappa_\nu = -\frac{c^2}{8\pi\nu^2}\int P(\nu,E)E^2\frac{d}{dE}\left[\frac{f(E)}{E^2}\right]dE \quad (9.25)$$

Equation (9.25) is still general, except for the assumption of an isotropic particle distribution. Now, if we specify $f(E)$ to be the usual power law, and use the same variable transform as in (9.16), we find the synchrotron self-absorption from a power-law electron distribution:

$$\kappa_{sy}(\nu) = (s+2)\frac{c^2}{8\pi}P_o f_o c_o^{(s+8)/2} \times B^{(s+2)/2}\nu^{-(s+4)/2}\int x^{-(s+1)/2}F(x)dx \qquad (9.26)$$

The x-integral, again, can be evaluated as a function of s; note, s also appears in the exponents of other parameters in (9.27). If we pick $s = 2.5$ as typical, we can evaluate the constants in $\kappa_{sy}(\nu)$ (using cgs units, still!):

$$\kappa_{sy}(\nu) \simeq 8\times10^{-40} f_o B^{(s+2)/2}\left(\frac{\nu}{c_1}\right)^{-(s+4)/2} \qquad (9.27)$$

9.6 Total synchrotron spectrum

Finally, we want to consider overall shape of the photon spectrum seen from a synchrotron source. When we asked this question for a thermal source (such as bremmstrahlung), we only had to deal with the radiative transfer. Here, we must also deal with the underlying electron spectrum – as we cannot assume it's thermal.

First, consider the effect of self-absorption on the emergent spectrum from a synchrotron source. We recall the solution to the transfer equation:

$$I_\nu = S_\nu\left(1 - e^{-\tau_\nu}\right) \qquad (9.28)$$

where $S_\nu = j_\nu/\kappa_\nu$ and $\tau_\nu = \int \kappa_\nu dx$, integrated through the source. This has the limiting solutions, $I_\nu \to j_\nu x$ for $\tau_\nu \ll 1$ (corresponding to high frequencies for the synchrotron case), and $I_\nu \to S_\nu$ (corresponding to low frequencies). Further, we note that $S_\nu \propto \nu^{5/2}$ from (9.17) and (9.27); since we have assumed a non-Maxwellian electron distribution, we should expect our source function not to be the low-frequency limit of the black body spectrum.

The total spectrum from a source which has $\tau_\nu = 1$ at some observed frequency will thus have a low-frequency range,

$$I_\nu \propto \nu^{5/2} \qquad (9.29)$$

and a high frequency range,

$$I_\nu \propto \nu^{-(s-1)/2} \qquad (9.30)$$

(this is the optically thin range).

This is not the full answer, however. We pointed out earlier that the electron distribution must have a low-energy cutoff, E_{min} (with critical frequency $\nu_{min} = \nu_c(E_{min})$; and that it may also have a high-energy cutoff, E_{max} (with critical frequency $\nu_{max} = \nu_c(E_{max})$. If we consider these limits, but ignore transfer, then we expect three frequency ranges. Thus, a purely optically thin spectrum will have a low-frequency range,

$$I_\nu \propto \nu^{1/3} \qquad (9.31)$$

(why? compare the single-particle spectrum, 9.10 and 9.11). The source will have a mid-frequency range,

$$I_\nu \propto \nu^{-(s-1)/2} \qquad (9.32)$$

This source will also have a turnover at high frequencies, due to a cutoff in the electron energy distribution (say at γ_{max}). The most likely spectral form is

$$I_\nu \propto e^{-\nu/\nu_{max}} \qquad (9.33)$$

A variant of this can be obtained if the high-energy electron distribution does not cut off abruptly, but more slowly, as is the case in some models of electron aging.

Finally: what might be the cause of low-energy and high-energy cutoffs? The high-energy cutoff is commonly assumed to be due to what's called "spectral aging". That is: from (9.4) or (9.5), recall that the single particle power goes as $P \propto E^2$. Thus, if we start with a power law electron distribution (as in 9.13), the highest energy particles will lose energy the fastest. This

leads to a truncation of the electron spectrum, at energies whose synchrotron lifetime equals the age of the source. The low-energy cutoff is harder to specify. It is likely to be due to the fundamental particle acceleration mechanism. We discussed this in Chapter 8; it may be that the low-γ limit of the resonance between Alfven waves and accelerated particles produces this E_{min}.

References

This comes mostly from my own notes. More details can be found in

- Pacholczyk, *Radio Astrophysics*
- Rybicki & Lightmann, *Radiative Processes in Astrophysics*

Key points

- Single particle synchrotron power *and* spectrum;
- Synchrotron emission from a power-law particle DF;
- Synchrotron self-absorption;
- Total synchrotron spectrum: high and low ν cutoffs.

Appendix: Einstein coefficients

I'm taking this directly from Rybicki & Lightman. Remember Kirchoff's law, $j_\nu = \kappa_\nu B_\nu$ (which we saw back in radiative transfer). This relates emission to absorption for a thermal emitter; clearly there must be some relation between emission and absorption at the microscopic level (described by quantum mechanics). To find it, consider transitions between two discrete energy levels of an atom, the first with energy E and statistical weight g_1, the second with energy $E + h\nu_o$ and statistical weight g_2. There are three possible radiative transitions between them:

- **Spontaneous emission** occurs when the upper level spontaneously emits a photon; this occurs whether or not the atom sits in an external radiation field. We define A_{21} as the probability per unit time (sec^{-1}) of this happening (in principle A_{21} could be calculated from quantum physics if we knew all the details of the atom).
- **Absorption** occurs in the presence of photons of energy $h\nu_o$. We define another coefficient, B_{12}, such that $B_{12}J$ is the probability per time for absorption. An important detail here: the spectral line associated with the transition has some finite width, $\delta\nu$, about ν_o (due to energy level uncertainty, doppler and collisional broadening, etc). If $\phi(\nu)$ is the shape of this spectral line, we define $J = \int J_\nu \phi(\nu) d\nu$ as the weighted-mean intensity over the line. Note that, if J_ν varies slowly over the line, $\phi(\nu)$ acts like a delta function.
- **Stimulated emission** also occurs if there are photons of energy $h\nu_o$ around. This wasn't expected classically. Einstein found that it was needed in order to derive Planck's law; we now know it's why masers mase and lasers lase. We define a third coefficient, B_{21}, so that JB_{21} is the probability per time of a stimulated emission event.

The game, then, is to use thermodynamic equilibrium results to find relations between A_{21}, B_{21} and B_{12}. We know three such results:

(i). In TE, the number of radiative transitions into state 1 must equal the number of transitions out of state 1. If n_1 and n_2 are the number of atoms in the two states, we know $n_1 B_{12} J = n_2 A_{21} + n_2 B_{21} J$.

(ii). In TE the ratio of level populations is given by $n_2/n_1 = (g_2/g_1) e^{-h\nu_o/KT}$.

(iii). In TE the photon field is given by the Planck function: $J = B_\nu$ (for $\nu = \nu_o$).

These three facts are enough: we can solve the equations to show that the Einstein coefficients must satisfy

$$g_1 B_{12} = g_2 B_{21} \; ; \quad A_{21} = \frac{2h\nu^3}{c^2} B_{21} \tag{9.34}$$

That's the main result. We can – in principle – use quantum physics to determine A_{21} for any given transition; then (9.34) tells us what the B's must be.

The Einstein coefficients can also be used to determine the emission and absorption coefficients. Going back to the definitions (chapter 3), we can build

$$j_\nu = \frac{h\nu_o}{4\pi} n_2 A_{21} \phi(\nu) \tag{9.35}$$

and

$$\kappa_\nu = \frac{h\nu_o}{4\pi} \left(n_1 B_{12} - n_2 B_{21}\right) \phi(\nu). \tag{9.36}$$

10 Pair plasmas in Astrophysics

Electron-positron pairs can also be important in the physics of high energy plasmas and their interaction with radiation. To set the stage: recall that the electron rest mass energy is 511 keV (or a temperature of 6×10^9 K).

NOTATION: in this chapter, $\epsilon = h\nu/m_e c^2$ is the *normalized* photon energy (relative to the electron rest mass). The normalized lepton energy is $\gamma = E/m_e c^2$, as usual. Watch out: γ is also the "reaction notation" for a photon (as in eqn. 10.1).

10.1 Pair annihilation

Free positrons will, of course, annihilate on electrons (or any other form of "regular" matter). The most common decay is

$$e^+ + e^- \rightarrow \gamma + \gamma \tag{10.1}$$

The annihilation cross section, measured in the center of momentum (CM) frame, is

$$\sigma_{e^+e^-} = \frac{\pi r_o^2}{\gamma + 1} \left[\frac{\gamma^2 + 4\gamma + 1}{\gamma^2 - 1} \ln\left(\gamma + \sqrt{\gamma^2 - 1}\right) - \frac{\gamma + 3}{\sqrt{\gamma^2 - 1}} \right] \tag{10.2}$$

Here, γ is the lepton energy (in the CM frame), and r_o is the classical electron radius, defined as $r_o = e^2/m_e c^2$ (in cgs). The Thomson cross section can be written $\sigma_T = 8\pi r_o^2/3$.

The cross section has two useful limits. In the low-energy case, $\beta \ll 1$, the cross section becomes

$$\sigma_{e^+e^-} \simeq \frac{1}{\beta} \pi r_o^2 \tag{10.3}$$

This shows that the annihilation probability is very high for electrons nearly at rest. The decay produces two photons very close to the rest energy of the leptons: $h\nu \sim m_e c^2$, or $\epsilon = h\nu/m_e c^2 \simeq 1$. This is the *annihilation line*. In the high-energy case, $\gamma \gg 1$, the cross section becomes

$$\sigma_{e^+e^-} \simeq \frac{\pi r_o^2}{\gamma} (\ln 2\gamma - 1) \tag{10.4}$$

The decay at high energies still produces two photons, but the photons have a much broader energy spread. The annihilation line becomes a broad annihilation spectrum.

It is worth remembering that the electrons and positrons will undergo all of the usual plasma interactions and radiation, during the time that they exist (before they annihilate). They can support plasma waves, emit synchrotron radiation, and all of the other processes we have studied.

10.2 Pair creation

Electron-positron pairs can be created in a wide variety of nuclear and electromagnetic interactions. I summarize the three mechanisms that seem most important astrophysically.

10.2.1 Two-photon pair production

Classical physics says that EM radiation is linear; superposition is allowed, without affecting either incoming signal. This is not always true, however. We learn from quantum physics that two photons can interact to produce pairs, or other particles, if their energy is high enough. The process of interest here is

$$\gamma + \gamma \rightarrow e^+ + e^- \tag{10.5}$$

It can occur if the total incoming photon energy, in the rest frame, is at least as large as two electron rest masses. In the lab frame, this translates to $\epsilon_1 \epsilon_2 \geq 1$ (I'm keeping photon energies normalized to $m_e c^2$, to shorten the notation. If you don't accept this limit, do the Lorentz transform for yourself, to check!). The kinematics are simple in the CM frame: the two incoming photons must each have the same energy, and likewise for the two created leptons. In the CM frame, let β be the lepton velocity, and $\gamma = (1 - \beta^2)^{-1/2}$ be its energy. Energetics require $\beta^2 = 1 - 1/\epsilon_1\epsilon_2$. The cross section for this process is

$$\sigma_{\gamma\gamma} = \frac{\pi r_o^2}{2\gamma^2} \left[\beta(\beta^2 - 2) + (3 - \beta^4) \ln\left(\frac{1+\beta}{1-\beta}\right) \right] \tag{10.6}$$

This process can have an important impact on a luminous high-energy photon source (such as a gamma-ray burster, or a very hot accretion disk). Consider the optical depth of a gamma ray, at energy ϵ. It can react with any photon that has energy above $1/\epsilon$. If L is the luminosity of the source above this energy cutoff, then the optical depth of the γ ray photon[1] is

$$\tau_{\gamma\gamma} = \frac{L\sigma_{\gamma\gamma}}{4\pi R \epsilon m_e c^3} \tag{10.7}$$

[1] Can you derive this?

This shows that the source can be opaque to its own radiation, if it has high luminosity and small size. Such a source is called "compact".

It is also possible for pairs to be produced in photon-proton or photon-electron interactions: $\gamma + p \to p + e^+ + e^-$, and $\gamma + e \to e + e^+ + e^-$. These reactions have smaller cross sections than $\gamma\gamma$ pair production (smaller by about the fine structure constant), and seem to be less important astrophysically.

10.2.2 Pair production in pion decay

Another source of pair production is from pion decay. Pi and mu mesons are unstable particles; once created, they decay rapidly. The common decay chains are

$$\begin{aligned} \pi^\pm &\to \mu^\pm + \nu_\mu/\bar{\nu}_\mu \\ \pi^o &\to \gamma + \gamma \\ \mu^\pm &\to e^\pm + \nu_e/\bar{\nu}_e + \bar{\nu}_\mu/\nu_\mu \end{aligned} \tag{10.8}$$

Thus, neutral pions decay simply to γ-ray photons. Charged pions, however, decay to muons which in turn decay to leptons.

So: how are pions made astrophysically? We have seen one mechanism, the interation of cosmic-ray protons with the microwave background:

$$\gamma + p \to p + \pi + \cdots \tag{10.9}$$

This predicts that there should be some free pions, decaying to muons and leptons, everywhere in space.

In addition, particle-particle reactons can make pions. The most common (there are many more I'm not listing. . .) are proton-proton collisions:

$$\begin{aligned} p + p &\to p + n + \pi^+ \\ &\to p + p + \pi^o \\ &\to d + \pi^+ \end{aligned} \tag{10.10}$$

The velocity-weighted cross section for π production in a thermal pp reaction is

$$\langle \sigma_{pp} v \rangle \simeq 4 \times 10^{-16} \, (\ln T_{12}) \, cm^3 s^{-1} \tag{10.11}$$

if T_{12} is the proton temperature in units of 10^{12} K. This form is good for temperatures above ~ 100 MeV; at lower temperatures the cross section drops rapidly. One application of this process is to hot accretion disks. Some models predict that the inner part of the disks will be hot enough for pion production to take place.

10.3 Magnetic pair production

Relativistic kinematics tells us that a free photon, in vacuum, cannot create a massive particle (or particles); the process cannot conserve energy and momentum. Single-photon pair production is possible, however, in the presence of a strong magnetic field. This process requires high photon energies (in order to create the lepton rest masses), and also high magnetic fields. The field must be close to the so-called critical field, given by

$$\hbar \frac{e B_{crit}}{m_e c} = m_e c^2 \; ; \;\; B_{crit} \simeq 4.4 \times 10^{13} \text{ G} \tag{10.12}$$

Energetics require that the photon satisfy

$$\epsilon \sin\theta \geq 2 \tag{10.13}$$

if θ is the angle between the photon's wavevector and the magnetic field. The probability of this process occuring is given in terms of an attenuation coefficient,

$$\kappa(\chi) \simeq 1.5 \frac{\alpha}{\lambda_c} \frac{B}{B_{crit}} e^{-4/3\chi} \; ; \;\; \chi = \frac{\epsilon}{2} \frac{B}{B_{crit}} \sin\theta \tag{10.14}$$

Here, $\alpha = e^2/\hbar c$ is the fine structure constant, and $\lambda_c = h/m_e c$ is the Compton wavelength of the electron. This quantity κ is essentially an absorption coefficient: the "opacity" the photon sees is $\tau = \int \kappa(\chi) dx$.

Magnetic pair production is thought to be important in pulsars. You recall that these are strongly magnetized neutron stars. The high electric fields induced by the rapid rotation and strong B field pull charges off the star's surface and accelerate them to very high energies. These charges then emit γ rays, either by curvature radiation (related to synchrotron emission) or by inverse Compton scattering of thermal X-rays emitted by the star. These γ rays can then pair produce. The pairs themselves emit more γ rays, probably through synchrotron radiation. These secondary γ rays then make more pairs, which emit more photons . . . and a *pair cascade* develops. The resultant dense pair atmosphere is thought to be the source of the coherent pulsar radiation.

A very similar process may occur close to a massive black hole in the center of an active galaxy. While these models have not been as extensively developed as pulsar models, the basic ingredients are the same. The black hole, or the accretion disk feeding it, are almost certainly magnetized. Rotation will induce strong

electric fields, which can accelerate charges to high energies. In addition, the accretion flow is probably a strong γ ray source itself. Thus, both two-photon and single-photon pair cascades are possible in this setting. Such a pair plasma may be the initial content of the directed, relativistic radio jets created by the nuclear black holes.

Key points

- Pair annihiliation: what it is, the magnitude of the cross section.
- "Particle" pair production, two types, what it is.
- Magnetic pair production, what it is, its "probability".

11 (Inverse) Compton scattering

We have one more radiation mechanism to cover, which is particularly relevant to compact objects and relativistic (synchrotron) plasmas.

NOTATION: in this chapter $\epsilon = h\nu$ is the photon energy (in physical units, such as erg) ... sorry for the switch, folks, it's the standard notation.

11.1 Basic Tools

We need to use several different things here.

11.1.1 One event seen in the ERF

To start, go back to simple Compton scattering, as you saw it in modern physics. Work in a frame where the electron is at rest,[1] and hit it with an incoming photon, of energy ϵ'. The photon scatters through an angle θ_1', and to an energy ϵ_1', with

$$\epsilon_1' = \frac{\epsilon'}{1 + (\epsilon'/mc^2)(1 - \cos\theta_1')} \tag{11.1}$$

Clearly, the electron gains momentum and energy after the scattering, and the photon loses energy, $\epsilon_1' < \epsilon'$.

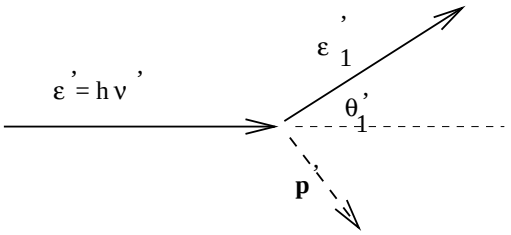

Figure 11.1 The geometry of compton scattering, in the electron rest frame (ERF). The photon comes in with $\epsilon' = h\nu'$, and leaves at angle θ_1' with ϵ_1'; the electron gains momentum $\mathbf{p}$.

11.1.2 Cross sections

In order to get the Compton-scattered spectrum, we'll need to know the probability that the electron scatters into angle θ_1' in the ERF, then average over all angles. The probability comes from quantum electrodynamics, and is called the Klein-Nishina formula:

$$\frac{d\sigma}{d\Omega} = \frac{r_o^2}{2}\left(\frac{\epsilon_1'}{\epsilon'}\right)^2\left(\frac{\epsilon'}{\epsilon_1'} + \frac{\epsilon_1'}{\epsilon'} - \sin^2\theta_1'\right) \tag{11.2}$$

where $r_o = e^2/m_e c^2$ is the classical electron radius, and $d\Omega$ is the differential solid angle for the scattering.

[1] This observing frame is called the Electron Rest Frame, ERF.

If we want the total scattering cross section, still in the ERF, we integrate (11.2) over $d\Omega$:

$$\sigma = \int \frac{d\sigma}{d\Omega} d\Omega$$

The general result is a long expression, and the limits are more useful. For $x = \epsilon/mc^2$:

$$\begin{aligned} x \ll 1: &\quad \sigma \simeq \sigma_T\ ; \\ x \gg 1: &\quad \sigma \simeq \frac{3}{8}\sigma_T \frac{1}{x}\left(\ln 2x + \frac{1}{2}\right) \end{aligned} \tag{11.3}$$

(recalling that $\sigma_T = 8\pi r_o^2/3$). Thus, for low photon energies, the scattering cross section is just σ_T; for high photon energies, it's more complicated. Note, finally, that these relations hold in the ERF. If we're looking at a scattering event in the lab frame, we know that the photon energy is $\epsilon' \simeq \gamma\epsilon$; thus the transition energy in (11.3), namely $x = 1$, becomes $\gamma\epsilon = mc^2$.

11.1.3 Remember your relativity

Here's a review if you need it.

• **Lorentz transforms.** Because our next step will be transforming between the ERF and the lab frame, we'll need to remember some Lorentz transforms. Figure 11.2 has the geometry. You remember that x position (along the motions) and time transform as

$$x' = \gamma(x - \beta ct)\ ;\ \ t' = \gamma(t - \beta x/c) \tag{11.4}$$

and, of course, the inverse transforms are just

$$x = \gamma(x' + \beta ct')\ ;\ \ t = \gamma(t' + \beta x'/c) \tag{11.5}$$

(Think about which way the "frame" is moving, and you can tell what to do about the $+$ and $-$ signs). You also probably remember that the coordinates y and z transverse to the motion don't change.

But now: several other important physical quantities transform just the same as position $\mathbf{x}$ and time do – these are called *4-vectors*. For a particle (massive or massless), momentum $\mathbf{p}$ and energy E are a 4-vector; so are wavenumber $\mathbf{k}$ and frequency ω if we're working with a wave. We therefore have (keeping track of units)

$$p_x' = \gamma(p_x - \beta E/c)\ ;\ \ E' = \gamma(E - \beta p_x c) \tag{11.6}$$

and

$$k_x' = \gamma(k_x - \beta\omega/c)\ ;\ \ \omega' = \gamma(\omega - \beta k_x c) \tag{11.7}$$

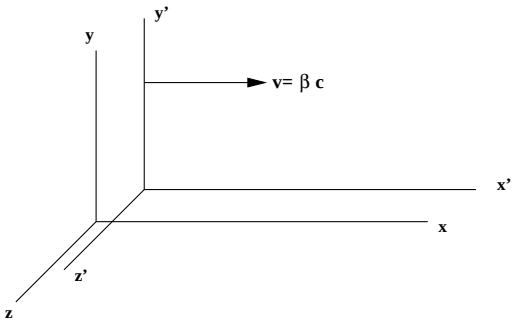

Figure 11.2 The geometry of our Lorentz transforms. The primed frame is moving at $v = \beta c$ along the x-axis; the unprimed frame is the "lab".

(and their respective inverses). Note that angle transforms can be found from the components of $\mathbf{p}$ or $\mathbf{k}$, as needed. If we apply these results to photons, which have $E = pc$ and $\omega = ck$, they collapse to one transform,

$$\omega' = \omega\gamma(1 - \beta\cos\theta) \tag{11.8}$$

where $\cos\theta = k_x/k$ selects out the $\mathbf{k}$ component along the direction of motion. We'll use this in a few minutes.

• **Invariants.** Some important quantities can be shown to be invariant – that is, to have the same value measured in either the lab frame or the moving frame. I'm not going to prove these here – you can look at Rybicki & Lightmann if you're curious. We'll need two important facts:

• We're interested in the power radiated (say by an accelerated particle). This is one invariant:

$$\frac{dE}{dt} = \frac{dE'}{dt'} \tag{11.9}$$

• We'll want to work with the photon spectrum; let $n(\epsilon)d\epsilon$ be the number of photons "at ϵ". It turns out that the ratio

$$\frac{n(\epsilon)d\epsilon}{\epsilon} = \frac{n'(\epsilon')d\epsilon'}{\epsilon'} \tag{11.10}$$

is also an invariant.

11.2 Scattering as seen in the lab

What then is "inverse" compton scattering? If the electron is moving, and has more (kinetic) energy than the photon, the photon will tend to gain energy in the collision (at the expense of the electron). The only difference between this and the simple version in (11.1) is the frame from which one views the collision ... but the moving-electron case is traditionally called Inverse Compton Scattering (ICS).

Thus, think of a situation where the photon has energy $\epsilon = h\nu$ and the electron has energy $\gamma m_e c^2$ before the scattering. We can analyze this by doing a Lorentz transform to a frame moving with the (γ, β) of the electron – the electron rest frame (ERF), and let it be the primed frame – in that frame (11.1) applies. In the ERF, we know the electron has $\beta' = 0$ (by definition!), and from (11.8), the photon has $\epsilon' = \gamma\epsilon(1 - \beta\cos\theta)$, where θ is the angle between the photon and electron momenta in the lab. Let the scattering take place, then transform back to the lab; the new photon energy in the lab is

$$\epsilon_1 = \epsilon_1'\gamma(1 + \beta\cos\theta_1') \tag{11.11}$$

(remember that ϵ', ϵ_1' are related by 11.1). Thus, for a given electron and photon energy before the scattering, ϵ' depends only on the scattering angle in the ERF.

11.2.1 Single particle radiation

The simplest application of this is to find the power emitted by an electron exposed to some photon field. By far the easiest way is to work in the ERF – where the scattering is simple – and transform to and from the lab as needed. We'll use basic Lorentz transforms, and the invariants (described above, in 11.9 and 11.10).

With this approach, it's not hard to find the total power scattered by our electron sitting in the radiation field $n(\epsilon)$. In the ERF, that power is

$$\frac{dE'}{dt'} = c\sigma_T \int n(\epsilon')\epsilon_1'(\epsilon', \theta_1')d\epsilon' \tag{11.12}$$

where we're assuming ϵ_1' from (11.1). The integrals here are over the photon spectrum. Because of the invariants, it's easy to write this in the lab frame:

$$\begin{aligned}\frac{dE}{dt} &= c\sigma_T \int (\epsilon')^2 \frac{n'd\epsilon'}{\epsilon'} \\ &= c\sigma_T\gamma^2 \int (1 - \beta\cos\theta)^2 \epsilon n(\epsilon)d\epsilon\end{aligned} \tag{11.13}$$

where we've used the basic Doppler shift, $\epsilon' = \gamma\epsilon(1 - \beta\cos\theta)$ in the last step.

Now: let the photon field be isotropic. The angle average of $(1 - \beta\cos\theta)^2$ is just $1 + \beta^2/3$; so the Compton-scattered power, angle-averaged and assuming an isotropic photon field, is

$$\frac{dE}{dt} = c\sigma_T\gamma^2 \left(1 + \frac{1}{3}\beta^2\right) u_{rad} \tag{11.14}$$

because $u_{rad} = \int \epsilon n(\epsilon) d\epsilon$ is the radiation density that the electron sees. Finally, then, we want the energy lost by the electron – its *Inverse Compton* power. That's just the scattered power minus the incoming power:

$$P_{IC} = \frac{dE}{dt} - c\sigma_T u_{rad} = \frac{4}{3} c\sigma_T \gamma^2 \beta^2 u_{rad} \quad (11.15)$$

(Does this form look familiar? Compare the relativistic limit, $\beta \simeq 1$, to the expression for synchrotron power, equation 9.5).

11.2.2 Single particle spectrum

Next, we want the spectrum of the scattered radiation. Unlike our previous applications (synchrotron and bremsstrahlung), we're not talking about Fourier analysis here. Rather, we know that a single scattering event leads to a photon energy ϵ_1, given by (11.11). But ϵ_1 depends only on the angles θ, θ_1'. One of these is the input condition (the angle between the incoming photons and the electron's motion). For the other, we know the probability of scattering into that angle – from $d\sigma/d\Omega$, (11.2). Thus, to get the photon spectrum – the probability of scattering giving an output ϵ' – we just can carry out the angle average of (11.11), using (11.2).

Because the power radiated depends on the photon field as well as the electron energy, we need to assume something about the radiation field. If the radiation is isotropic and monoenergetic, at photon energy $\epsilon_o = h\nu_o$, we can call its intensity $I(\epsilon) = F_o \delta(\epsilon - \epsilon_o)$. The angle averaging then gives (after much algebra),

$$P_{ic}(\epsilon_1) = 3\sigma_T F_o F_{ic}(x) \quad (11.16)$$

where we've defined

$$x = \frac{\epsilon_1}{4\gamma^2 \epsilon_o} \quad (11.17)$$

and the kernel function is

$$F_{ic}(x) = x\left(2x \ln x + x + 1 - 2x^2\right) \quad (11.18)$$

Compare the single-particle synchrotron spectrum, from (9.9) and (9.11): it is a narrow function which peaks at $\nu \simeq \nu_c$. Similarly, the function F_{ic} is a narrow function which peaks at $\nu \simeq 2\gamma^2 \nu_o$. Thus we have a "characteristic" frequency for inverse Compton scattering, just as we had for synchrotron.

11.3 Composite spectra

In our previous derivations, we integrated the single-particle spectrum over the distribution of particle energies, to find the volume emissivity (we did this before, for bremsstrahlung and synchrotron). It's more complicated than the previous derivations, because (i) we have to assume an input photon spectrum, and (ii) we really have to worry about whether a photon scatters once, or many times. Instead of doing the general problem, I'll just look at two simple limits.

11.3.1 Nonrelativistic electrons

In this case, each scattering leads to only a very small change in the photon energy: $\epsilon_1' \simeq \epsilon_1 \left[1 - (\epsilon'/mc^2)(1 - \cos\theta')\right]$. If we use (11.2), and do angle averaging, we find that the average energy gain per photon, scattering on electrons at temperature T, is

$$\frac{\delta\epsilon}{\epsilon} \simeq \left(\frac{4kT - \epsilon}{mc^2}\right) \quad (11.19)$$

Thus, one scattering can be significant to the radiation source (for instance the microwave background, scattering as it passes through a cluster of galaxies); but it makes little difference to the photon spectrum, as long as $kT \ll mc^2$.

11.3.2 Relativistic electrons, single scattering

The ICS kernel in (11.16, 11.18) peaks at $\epsilon_1 \simeq 2\gamma^2 \epsilon_o$ – that comes from the mean scattered photon energy. (It's easy to show that $4\gamma^2\epsilon_o$ is the maximum possible scattered energy; the mean will be somewhat less than this). In most astrophysical applications, the ICS opacity through a source is low, so the photons only scatter once (if at all). Thus IC scattering on electrons at energy γ boosts low-frequency radiation by $\sim \gamma^2$. The low frequency photons might be synchrotron photons from the electrons themselves (in which case this is called "synchrotron self-compton" radiation, SSC); or they might be microwave background photons.

11.3.3 Scattering from power-law electrons

Put a distribution of relativisitic electrons, $n(\gamma)$ again, in a photon field with spectrum $F(\nu')$.[2] If the photons are monoenergetic at ν_o, we have $F(\nu') = F_o \delta(\nu' -$

[2] *Notation alert*: in this section I'm letting ν' be the *incoming* photon energy – so that ν can refer to the scattered photon energy, which is our desired final result.

ν_o) (from above). Looking back to (11.16, 11.18), it will be useful to rewrite the ICS power from a single scattering by a relativistic electron as

$$P(\nu;\nu_o,\gamma) = \frac{3}{4}\frac{\sigma_T}{\gamma^2}F_o\frac{\nu}{\nu_o}f_{ic}(x) \tag{11.20}$$

where $f_{ic}(x) = 2x\ln x + x + 1 - 2x^2$. If we replace the monoenergetic photon field by a spectrum $F(\nu')$, we can find the emergent (singly) scattered ICS spectrum by integrating over the photon spectrum and the energy distribution:

$$4\pi j_{ic}(\nu) = 3\pi\sigma_T \int\int n(\gamma)\frac{1}{\gamma^2}\frac{\nu}{\nu'}F(\nu')f_{ic}(x)d\gamma d\nu' \tag{11.21}$$

where $x = \nu/(4\gamma^2\nu')$, as before. Because this is similar to the integrand we encountered in our synchrotron analysis, we can use a similar variable transform. Let x and ν be the integration variables now; for our power-law electron distribution take

$$n(\gamma) = n_o\gamma^{-s}$$

as usual. With this, the integral in (11.21) becomes

$$\begin{aligned} j_{ic}(\nu) =& \frac{3}{4}2^s\sigma_T n_o\nu^{-(s-1)/2}\int F(\nu')(\nu')^{(s-1)/2}d\nu' \\ &\times \int x^{(s-1)/2}f_{ic}(x)dx \end{aligned} \tag{11.22}$$

This looks horrible, yes; but we're almost there. The last integral, over x, is just a number, because $f_{ic}(x)$ is a simple function. In addition, the first integral, over ν', depends only on the photon spectrum. If we specify $F(\nu')$ (for instance the black body spectrum of the microwave background; or maybe the synchrotron spectrum of the electrons themselves), this first integral can be worked out. Thus, we have the important result: the ICS spectrum from a power-law electron distribution, for single scattering, obeys

$$j_{ic}(\nu) \propto n_o\nu^{-(s-1)/2} \tag{11.23}$$

That is: if the electrons are a power law, the ICS spectrum is also a power law, with spectral index $(s-1)/2$. This is, of course, the same spectral index as that for synchrotron emission from the same electrons – only we must remember that the ICS photons come out at much higher energies.

References

Once again, this is mostly from my own notes; but you can find more details in

- Rybicki & Lightmann.

Key points

- Compton scattering: what it is, physically, and what "inverse" is.
- Single particle ICS, power *and* spectrum.
- ICS from thermal (nonrelativistic) electrons.
- ICS from relativistic, power-law electrons.

12 Pulsars: overview and some physics

Carroll & Ostlie present the basic picture in some detail; another good reference is Longair's *High Energy Astrophysics*. In these notes I'll be brief about the basics, and emphasize the physics inside the basic model, as well as newer work on high-energy emission and pulsar winds.

12.1 The basic picture

What can cause a star's brightness to pulse as quickly as 100 times a second? That was the immediate question when pulsars were discovered. Stars with longer-period variability are "beating" – that is undergoing body-mode oscillations. But we know a fair bit about normal modes in stars, and none can have so high a frequency. So we must consider rotation; perhaps a hot spot on the star rotates past our line of sight? This was more promising ... but the most compact star then known, a white dwarf, is too big. Remember the rotation rate is limited by lightspeed at the star's surface, as well as by the stability (gravitational binding energy) of the star. Only neutron stars – which were predicted by theory but not yet proved to exist – could explain such a short period.

12.1.1 The cartoon

Thus the basic picture was born: an isolated pulsar is a rapidly rotating neutron star with a small, radio-loud "hot spot". But why is there a "hot spot"? Why does the star's rotation slow down? We think the star is strongly magnetized. We expect this to follow if the NS is made in the core collapse of a supernova; flux freezing will lead to a very strong magnetc field in the NS. But this can, in principle, answer both questions. Two things cause the star to slow down. A rotating magnetic field emits magnetic dipole radiation; by energy conservation this must lead to spindown. If the star sits in vacuum this is the only energy loss. If it sits in the ISM, however, there will also be some torque between the star and the ISM, probably mediated by the magnetic field and/or the plasma outflow from the star (as discussed below).

If the star is strongly magnetized *and if the magnetic dipole is set at an angle to the rotation axis*, we also have a ready cartoon for why it pulses, as follows. Close to the star the magnetic field will be dipolar and will rotate with the star, as will any plasma which is tied to these dipolar field lines. However, at the *light cylinder* radius, that is $r_{LC} = c/\Omega$ if Ω is the rotation rate, the plasma cannot corotate with the star. Magnetic field lines which start close to the magnetic pole will not be able to turn around ("close") before they reach r_{LC}; rather they must connect to the B field of the local ISM. Plasma can therefore flow out along these *open field lines*. If the B field is dipolar, field lines within angle $\sim (R_*/r_{LC})^{1/2}$ of the magnetic axis will be open. If this outflowing plasma can – somehow – make intense radio emission, which is strongly beamed forward along the open field lines, then we will see strong radio pulses only for the fraction of the rotation period when the beamed radiation intersects our line of sight.

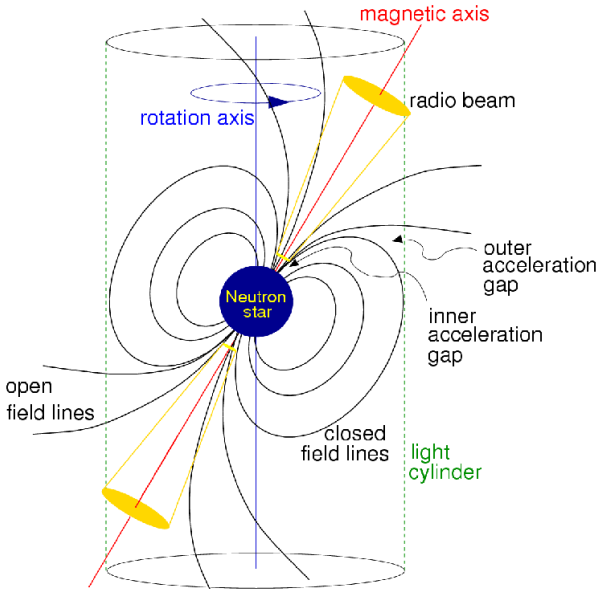

Figure 2.1.1 Cartoon of the standard picture of a pulsar. From Ransom & Condon, NRAO.

12.1.2 And some details

• **Typical numbers for single pulsars.** The star's radius $R_* \sim 10$ km (from NS models). The light cylinder is typically at $\sim 1000R_*$ (but this of course depends on the rotation rate Ω). The open field line region is a few degrees across at the star's surface. From the duty cycle of the radio pulses, and the assumption that the radio-loud plasma fills the open field line region, we infer the radio emission comes from $\sim 3 - 30R_*$ above the surface.

• **Energetics**. An isolated pulsar (one not in a binary system) is living off its rotational energy, In terms of its rotation rate Ω and the associated period $P = 2\pi/\Omega$,

this is

$$E_{rot} = \frac{1}{2} I\Omega^2 = \frac{2\pi^2 I}{P^2} \; ; \qquad \dot{E}_{rot} = I\Omega\dot{\Omega} = -4\pi^2 I \frac{\dot{P}}{P^3} \tag{12.1}$$

As you remember (yes?) from basic mechanics, I is the moment of inertia; if the NS were a homogeneous sphere, we would have $I = 2MR^2/5$. Pulsar people generaly take $M \gtrsim 1M_\odot$, and $R \sim 10$ km. These are probably pretty good guesses, although details of the as-yet-unknown equation of state in the star's core can matter here. The 2/5 factor in I is much less certain, due to our ignorance of the internal state of the star and the importance of general relativity in the star's structure. Many authors choose $I \simeq 10^{45}$ (cgs) as a "typical" value, and scale to I_{45}. Collecting these estimates, we find a large range for the power source, $\dot{E}_{rot} \sim 10^{32} - 10^{38}$ erg/s, for old to young pulsars.

• **How strong is the B field?** Remember we have no direct measure; and that very few NS sit in vacuum. Everyone in the field ignores this latter, and assumes that magnetic dipole radiation dominates the spindown. To remind you: magnetic dipole radiation comes from the time-dependence of the star's magnetic moment, $\mathbf{m}$. If this time dependence comes from the star's rotation, we get for a star rotating at Ω and a magnetic axis oriented at α relative to the rotation axis,

$$P_{mag\ dip} = \frac{2|\ddot{\mathbf{m}}|^2}{3c^3} = \frac{2}{3}\frac{(\Omega^2 m)^2}{c^3}\sin^2\alpha \tag{12.2}$$

Now, the magnetic moment is connected to the magnetic field at the star's surface and magnetic pole by $m = B_* R_*^3/2$. But P and $\dot{P}$ can be carefully measured; so, by equating $\dot{E}_{rot}$ (from 12.1) to $P_{mag\ dip}$ (12.2), we can "derive" the B field: $B_* \simeq 3.2 \times 10^{19}(P\dot{P})^{1/2}$. Putting in typical P's and $\dot{P}$'s for single pulsars, we find $B_* \sim 10^{11} - 10^{13}$ G is the expected range. Finally, note that things get interesting when the field is close to the quantum field, defined by $\hbar e B_{cr}/m_e c \sim m_e c^2$: $B_{cr} \sim 4.4 \times 10^{13}$G.

12.2 Spin a magnetic field

If the basic picture – a rapidly rotating, strongly magnetized neutron star – is correct, then some striking physics follows. If you spin a B field, you generate an E field. Think about $\partial\mathbf{B}/\partial t$, let $\hat{\mathbf{z}}$ be the rotation axis,[1] and remember that the rotation velocity is $\mathbf{v}_{rot} = \Omega \times \mathbf{r}$. The E field generated, measured in the inertial ("lab") frame, is

$$\mathbf{E}_{co} = -\mathbf{v}_{rot} \times \mathbf{B}/c = -(\Omega \times \mathbf{r}) \times \mathbf{B}/c \tag{12.3}$$

(I'm calling this the "corotation field", for reasons explained below). The important issue for the local physics is whether this E field is felt, at full strength, by the plasma around the star, and how that plasma responds.

To think about the impact of this on the NS, consider two scenarios. Both are still current in the field, and both lead to strong $\mathbf{E}$ fields, and high particle energies, in the open field line region.

12.2.1 Star in vacuum

Inside the star, we assume the matter and field corotate. The free charge must arrange itself into regions of positive and negative charge, so that the resulting E field just balances the $\mathbf{v} \times \mathbf{B}$ force. Thus, there must exist a *physical* $\mathbf{E}$ field, equal to that in (12.3), so that the net force measured in the corotating frame is zero. But this physical field has a non-zero divergence, so must be supported by a local charge density:

$$\rho_{co} = \frac{1}{4\pi}\nabla \cdot \mathbf{E}_{co} \simeq -\frac{\Omega \cdot \mathbf{B}}{2\pi c} \tag{12.4}$$

(the $\simeq$ means I'm dropping terms $\sim v_{rot}/c$; thus this is valid only well inside the light cylinder). Thus: the matter inside the star must have a net charge density, which is quadropolar: one sign towards the two poles, the opposite sign towards the equator.

Outside of the star, the $\mathbf{E}$ field will be a vacuum solution (by assumption, there are no charges outside), determined by the charge distribution of the star. But we know that from basic E&M, we can just solve Laplace's equation ($\nabla^2\Phi = 0$, if $\mathbf{E} = -\nabla\Phi$) in the vacuum region, *matching the potential and tangential field at the star's surface*. Those solutions can be worked out, but I won't bore you with the details. The interesting part of the solution is strength of the field, $E \sim B\Omega R_*/c$, and the fact that the $\mathbf{E}$ field in the polar cap region has a strong component parallel to $\mathbf{B}$ (call it $\mathbf{E}_\parallel$). From this, we can infer that any free charges in the region will be accelerated to high energies – at least in the open field line region. It may be that these strong fields can even pull charges from the surface of the star (which is mostly but not totally neutrons) ... so that the external vacuum may not last for long.

[1] I can't do bold Greek; so please pretend Ω is a vector, $\Omega\hat{\mathbf{z}}$.

12.2.2 Filled magnetosphere

Consider the other alternative, that the region above the star's surface is filled with plasma. We know the *charge* density needed if this plasma is to corotate with the star; it's just (12.4). Most authors assume that the plasma within the closed field line region has just this charge density, so that the free charge shields the rotation-induced E field, giving the net $\mathbf{E} = 0$ (measured in the corotating frame). With no net force, the plasma in the region will be static, again relative to the corotating frame. In fact, consider the consequences if the local plasma differed from ρ_{co}: there would be a finite unshielded E field, and the free charges in the plasma would move so as to cancel the field. Such a situation would not be stable, nor likely to last for very long. Thus corotating plasma is likely on closed field lines.

The situation in the open field line region can be more interesting, however. Charges leaving the star's surface must start at low velocity, and go through an acceleration region in which they reach their final energy. If the flow is steady, the charge density must vary inversely with the particle speed. In addition, particles can leave the star along the open field lines, presumably at high speeds. If the particle outflow carries a net charge, a *current* is driven out from the star; as with any electrical system, current density is sensitive to the global circuit (return path, driving voltage and net resistance). From these arguments we suspect that the rotation-induced $\mathbf{E}$ is *not* fully shielded in at least part of the open field line region. Any unshielded $\mathbf{E}_{||}$ will accelerate particles and drive the outflow/current.[2]

12.3 Radio emission and the pair cascade

Pulsars were discovered by their very intense, pulsed radio emission. Many, many papers have been generated about observations and models of this emission. However, it turns out to be only the small tail of the dog: the total radio power is a very small fraction of the spin-down power, $\dot{E}_{rot}$. In addition, we don't know what causes the radio emission. From the observed very high brightness temperatures ($T_B \sim 10^{35}$ might be typical), we know the emission cannot be due to any of the incoherent processes which apply elsewhere in astrophysics. No plasma can be physically so hot (why?? what would happen to the particles??) so T_B cannot be a physical temperature (as it would if it were thermal emission); nor can this be synchrotron radiation (remember the $T_B \sim 10^{12}$K limit for synchrotron, which we saw earlier). Thus we must be seeing a *collective* or *coherent* emission mechanism; and these are far from understood.

Despite lack of a solid model, a complex scenario has evolved to describe where and how the radio emission is likely to occur. To start, let's stay within the open field line region, where we have just argued that charges are accelerated to very high Lorentz factors. There are two ways in which these particles can emit very energetic photons.

• One way is *curvature radiation*. The charges must follow the magnetic field lines (their gyroradii are tiny; in fact their gyromotion is quantized). The field line curvature makes the particles emit curvature radiation. To understand this, go back to the arguments synchrotron radiation; we can apply them here to curvature emission. The characteristic photon frequency is $\sim 3\gamma^3 c/4\pi\rho_c$, if ρ_c is the radius of curvature of the field lines. The radiated power, integrated over frequency, is $\sim e^2 c\gamma^4/\rho_c^2$ (erg/s per particle). For $\rho_c \sim 10$ km, and the particle energies above, the curvature emission photons come out in the γ-ray region, in particular above the electron rest mass energy.

• Another alternative is *inverse Compton scattering*. The pulsar itself is warm (we know they are thermal X-ray sources). The primary charges can Compton scatter the thermal photons to higher energy, again making γ ray photons above $m_e c^2$. [NOTE: in this situation we cannot simply argue the scattered frequency $\sim \gamma^2 \nu_o$, because of the special geometry, with the photons and charges travelling nearly in the same direction. Doing the calculation carefully does verify that some of the scattered photons are hard enough to be interesting, however.]

Either one of these mechanisms will generate photons which are energetically capable of one-photon pair production, *via* $\gamma + B \rightarrow e^+ + e^- + B$. Recall that this mechanism has a low-B threshold, $h\nu B \gtrsim 0.1 m_e c^2 B_{crit}$, which is very likely satisfied in the high B fields near the pulsar polar cap. Once created, the leptons probably have enough energy to make more energetic photons (through synchrotron ra-

[2] The real question is what the net, unshielded, potential drop is, and thus particle energies are reached. Typical models suggest Lorentz factors $\gamma \sim 10^6 - 10^7$ are reached; but I wouldn't suggest overmuch confidence in this, the situation is complicated and still not well understood.

diation, most likely, also possibly further curvature and IC emission). Thus a pair cascade occurs ... and is thought to continue until most of the primary beam energy is converted to a dense pair plasma. Models suggest typical Lorentz factors of the pairs $\sim 10^2 - 10^3$.

Now what about the radio emission? As noted above, we need some collective process. The story becomes less clear at this point, as collective plasma emission is not well understood. A general guess is that the pair plasma is a necessary part of the picture. For instance, *plasma turbulence* may be involved; the charges in strong (large-amplitude),turbulent plasma waves can show collective behavior and emit intense radio pulses. If there is a residual $\mathbf{E}_{\|}$ in the pair-cascade region, it will generate relative streaming of the electrons and positrons; such streaming is known to lead to plasma turbulence.

12.4 High altitudes and currents

The ideas in the discussion above have been around for quite awhile, about as long as pulsars have been known. In recent years, new telescopes (X and γ ray) have expanded our picture of these stars and their interaction with their immediate environment.

12.4.1 High energy emission

Pulsars are now commonly detected in X-rays and γ-rays. To date, about 30 pulsars have been detected in X-rays and similar numbers in γ-rays (out to 100 MeV). These tend to be the young ones, which will have the largest $\dot{E}_{rot}$, and thus (possibly) have the strongest high-energy emission; thus we might guess that most or all pulsars would show high-energy emission if we had instruments sensitive enough to see them.

The high-energy emission is pulsed, but less narrowly so than is the radio emission. Thus, either the high-energy emission comes from a different location in the star, or (if contiguous with the radio emission) it is less strongly beamed. Many authors suggest this radiation comes from higher altitudes than the radio emission, possibly even from the light cylinder region.

For the stars well-measured up to now, we know that the radio emission is a very small part of the total power; most of the luminosity comes out at high energies, above an MeV. While uncertainties about distance and beaming factor make absolute power estimates difficult, we think that the bolometric power may be comparable to the spindown power. That means that pulsars are quite efficient at converting rotation energy to hard photons; and that the radio emission – the observation which first detected these stars – is but the small tail on a much larger dog.

What type of radiation are we seeing? Normal, incoherent synchrotron (from highly relativistic particles in the star's strong B field) seems to work well for the Xrays. Leptons in the pair cascade are created with finite pitch angles, and lose their energy to synchrotron radiation, much of which comes out as Xrays. The γ-rays are thought to be the leftovers of the pair cascade process, hard photons which escape the star without being turned into pairs.

12.4.2 The pulsar circuit?

Here's an important piece of unanswered physics in pulsar models: what does the current do? That is ... the basic low-altitude model says that free charges are accelerated away from the star's surface. This is what starts the cascade that leads to the radio emission. But this is a current; the rotating star acts as a unipolar dynamo (as we discussed earlier). But the star can't build up a net charge (why?) – so the current must close somehow. How this happens has been the topic of discussion ever since these stars were discovered.

Your author does not claim to know the answer ... but here's a speculation. We know that charges flow most easily along magnetic field lines – and undergo very little electrical resistance along the way. But because the polar current starts out along the open field lines, which must connect to the ISM, we also know the current must move across field lines, in order to connect to other field lines which return to the star. This is likely to happen in the outer magnetosphere (rather like the cross-field charge flow which leads to earth's aurora). Further, cross-field motion is likely to be more resistive; this dissipated energy may come out as observable radiation. Might this be connected to the high-energy pulsed emission, or even to a high-altitude component of the radio emission?

12.5 Winds and nebulae

Now, let's move out to larger scales, past the light cylinder. The standard pulsar model, above, predicts a strong Poynting flux radiated out from the star, and also an outflow of magnetized, relativistic particles. One might expect this energy outflow to couple to nor-

mal matter (to mass-load, somehow, when it reaches the ISM); will it drive a wind out from the star? In the inner regions, at least, this will be a relativistic, strongly magnetized outflow, with significant angular momentum – so we should not expect it to be spherical. Rather, it should come out perpendicular to the rotation axis, even if it is initially driven by the plasma outflow from the star's magnetic poles.[3]

12.5.1 Pulsar winds

This idea has been around for quite awhile; thanks to recent technology (mostly X-ray satellites) we are now seeing evidence of these winds. The data are striking. For older pulsars (those not currently within SNR), we sometimes see structures in the nearby ISM which are clearly *bow shocks* associated with the star's high-speed motion through the ISM. We infer that an unseen wind from the pulsar creates the pressure balance that leads to the observed bow shock. From the standoff distance of the bow shock we can estimate the wind energy, and compare it to standard models of the pulsar.

For young pulsars (those still within their SNR), recent CHANDRA images directly reveal the outflow from the pulsars (the Crab and Vela pulsars are the best examples here). These outflows are complex: they show *jets*, which presumably come out along the star's rotation axis (this is the only symmetry axis in the system), and *equatorial winds*, which probably arise from the combined effects of the star's strong magnetic field and its rapid rotation.

12.5.2 Pulsar wind nebulae

The wind coming out from the pulsar is (we think) still highly relativistic, with charged particles moving almost exactly along the magnetic field lines. Thus it should be nearly invisible (how would it radiate), except for its dynamical effects (such as the bow shocks). But what happens when it encounters the local ISM? If that ISM is dense enough, the wind will shock down, and the particles will be "thermalized" (they will gain a significant component of energy transverse to the local magnetic field). Thus, the shocked wind will become visible, as a synchrotron source. This makes what is called a *pulsar wind nebula*, PWN (we've already mentioned these in Chapter 7).

[3] Picture holding a rigid water hose at some angle, while you spin around your vertical axis. Which way will the water go?

To date there are a several good examples of PWNe; most of them are inside supernova remnants (which provides the high ambient density/pressure that creates the thermalizing shock in the wind). The Crab Nebula is the most striking example of this phenomenon: the shocked pulsar wind fills the nebula, and pushes a shell of cooler matter outwards. This shell is made up of the original SN ejecta and ISM which has been accumulated along the way; it's what we see as the optical and radio nebula.

12.6 Magnetars and Anomalous pulsars

Here's another new area. Based on some recent discoveries, of strong X/γ-ray flares, and/or strongly pulsed X-ray emission, the picture described above is being pushed in two ways. Two characteristics define these unusual objects.

- They have extremely strong magnetic fields, $B \sim 10^{14}$G. This is well above the quantum critical field, $B_{crit} \sim 4\times10^{13}$G – and in the regime where all kinds of interesting physics happens (photon splitting, vacuum birefringence, ...). Such a strong field may be too large to be due to simple flux freezing in the SN collapse of the parent star; various authors discuss dynamos taking place during the SN collapse.
- They are *not* living on their rotational energy: their strong flares have $L \gg \dot{E}_{rot}$. Where, then, does their energy come from? We think it's magnetic – that the very strong B field inside the star can occasionally break through the crust, in a cross between a "starquake" and a very strong "stellar flare". Once this flux tube emerges into the magnetosphere, reconnection will go, releasing the magnetic energy as X-ray and γ-ray radiation.

Because of the way these objects have been found, they are often called SGRs (Soft Gamma Ray Repeaters) or AXPs (Anomalous X-ray Pulsars), and appear to be different objects; but the growing consensus is that these are both observational variants of the magnetar phenomenon. Although none of these objects were initially found in radio, one or two have now been detected to have "normal" radio pulses as well.

Your author doesn't know just how magnetars fit into the general pulsar picture, they are too new – but their

existence points out that the range of extreme conditions which can exist on or around neutron stars seems to be more diverse than we've understood up to now.

Key points

- The basic picture, a pulsar as a rotating magnetized neutron star;
- Our cartoon of the pulsar magnetosphere: what is the plasma doing?
- The pair cascade, how it fills the magnetosphere;
- How the pair-filled magnetosphere might make radio and high-energy emission;
- Pulsar winds and nebulae.
- Magnetars.

13 Radio jets and radio galaxies

The idea of astrophysical jets first attracted interest as a theoretical prediction. Double-lobed radio galaxies (RGs) were detected in the first radio surveys (in the 1960's?). The data available then showed two radio-loud lobes (synchrotron sources, containing magnetized plasma and relativistic particles) on either side of a central, elliptical galaxy. Energetics and lifetime arguments soon found that the lobes were short-lived.[1] Either we were seeing them at a very special time in the life of the parent galaxy, or they were being resupplied with energy by an undetected pipeline: a radio jet. These jets were detected when the instruments improved – this was one of the first successes of the VLA. Models of active galactic nuclei – involving accretion onto a black hole – had to be amended to include the production of highly collimated, relativistic plasma jets, *if* the AGN happened to live in an elliptical galaxy.

We've learned a lot since then. We now know that jets are common on both small and large scales. On large scales, the massive black holes in galactic nuclei can produce jets while they are "active". Radio galaxies and (radio-loud) quasars, which live in elliptical galaxies, are the most dramatic examples. In these objects, the galactic nucleus contains a bright, compact radio core and a pc-scale, synchrotron-bright jet. The radio core is thought to be optically thick synchrotron emission from the base of the jet. On larger scales, the jets extend to at least several kpc, often farther; they connect to larger lobes or tails which arise from the interaction of the radio jet with the local extragalactic plasma.

On smaller scales, we now know that many star-sized galactic sources have relativistic jets. The most well-studied (and well-imaged) are jets from accretion flows in X-ray binary systems; SS433 is the prototype, and a few dozen are know known (*microquasars*). In addition, a few pulsars/PWNe systems are found to have jet outflows: the Crab and Vela pulsars are examples here. Finally, it is now thought that Gamma-Ray Bursters (GRBs) involve highly relativistic jets, and that core-collapse supernovae produce short-lived jets during the explosion. Also on stellar scales, but moving to less relativistic systems, we now know that the collimated outflows are produced during part of the collapse of a protostellar cloud to the final star; these are called (*protostellar jets*).

While all of these jets are interesting, in these notes I'll focus on what I'm most familiar with, namely, relativistic jets from AGN. My goal is to present an overview of the observations, some basic physics, and the current "cartoons" as to how RGs work.

13.1 Jets: the observational constraints

To start, what are the important properties of jets which theory must account for? In these notes, the discussion is strongly skewed toward extragalactic jets (radio galaxies and quasars), and relativistic galactic jets (microquasars), which is my personal interest and area of experience. Much of the basic dynamics apply to all jets, but some of the details and constraints are more relevant to relativistic jets.

Some critical facts and problems are:

• **Internal energy.** Radio jets from AGN and microquasars are dominantly synchrotron sources – thus we know they are magnetized and internally relativistic. Protostellar jets and some galactic jets are dominated by cooler, thermal radiation: emission lines from ionized gas, and even molecular emission.

• **Collimation.** These jets usually have very small opening angles – no more than a few degrees – and often retain their direction and collimation for a distance which is orders of magnitude larger than the scale on which they were initially produced.

• **Knots and bright spots.** Jets are rarely smooth when seen in high-quality images. Bright features are common. We don't know just why these features appear, but a mixture of shocks and strong-amplitude waves in the flow are likely. Note also that the radiation mechanisms are strongly nonlinear amplifiers: a weak density or B field enhancement, for example, can cause a strong enhancement of the emissivity.

• **Speed.** There is indirect evidence that many jets in RGs are supersonic (relative to themselves): structures are seen at their outer ends, where they run into the ambient medium, that can be identified as shock fronts within the jet. In addition, on pc scales, some jets (or at least waves in the jets) are moving at relativistic speeds; this is inferred from the apparent proper motion of knots in the jets, which can exceed lightspeed.

[1] Equipartition calculations give you the minimum energy in a radio source; divide this by its radio power, and you get a "lifetime" – which is short compared to any plausible estimate of the source age. Thus, you need energy resupply.

There's no compelling evidence for relativistic motion in kpc-scale jets; somehow the high-γ pc-scale flow has slowed down by the time it reaches kpc scales.

• **Plasma content**. Protostellar jets appear to be normal ISM; an ion-electron mixture. Compact-object and extragalactic jets are observed in synchrotron radiation, and therefore contain relativistic electrons and magnetic fields; we do not know if the electron charge is balanced by ions or by positrons.

• **Energization.** In at least some extragalactic jets there is a clear need for *local* re-acceleration of the relativistic particles. The synchrotron lifetime of the highest energy particles is significantly less that the shortest possible travel time down the jet (jet length / lightspeed). Somehow, local fluid/plasma processes must transfer energy from the bulk flow to the individual particles; a mixture of shocks and turbulence is probably the answer.

13.2 Some useful relativity

Two relativistic effects are important here.

13.2.1 Superluminal motion

The bright knots in many jets are observed to have apparent proper motion above lightspeed. As you've seen before (for instance in Caroll & Ostlie), this is a simple consequence of light-travel times. If the jet is oriented at angle θ to the line of sight, and has physical speed βc, its apparent speed – as seen by the observer – is

$$\beta_{app} = \frac{\beta \sin\theta}{1 - \beta\cos\theta} \qquad (13.1)$$

It's easy to show that this can lead to $\beta_{app} > 1$ for $\theta \ll 1$. Typical observed values are $\beta_{app} \sim$ a few; you should be able to work out what (γ, θ) values are needed for this.

13.2.2 Doppler beaming

A source of radiation is moving relativistically, at some γ, at angle θ to your line of sight. One effect, which you remember, is a Doppler shift. If the source emits photons at ν', you observe photons at ν, where the observed and emitted frequencies are related by $\nu' = \gamma\nu(1 - \beta\cos\theta)$ (note, the angle θ is measured in the observer's frame; and $\theta \to 0$ means motion towards the observer). This can be written,

$$\nu' = \frac{\nu}{\mathcal{D}}\,; \qquad \mathcal{D}(\theta) = \frac{1}{\gamma(1 - \beta\cos\theta)} \qquad (13.2)$$

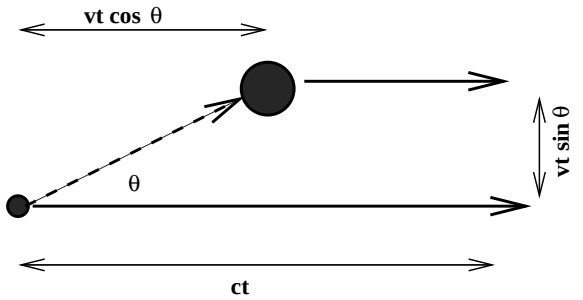

Figure 13.1 Illustrating how a feature in a jet can appear to be superluminal. The core source (small circle at left) emits a "blob" (large circle) at time $t = 0$, at angle θ to the line of sight; it also emits an EM signal directly towards us. After time t, the blob has moved a distance vt, and emits a second signal. The difference in arrival times between the first and second signals is $ct - vt\cos\theta$; the apparent separation of the core and blob is $vt\sin\theta$. Dividing the latter by the former gives us the apparent velocity of the blob, as in (13.1).

where $\mathcal{D}$ is called the *Doppler factor*. A second important fact is how to connect the spectral intensity you observe, I_ν, to that emitted by the source, $I'_{\nu'}$. It turns out that the two quantities are related by

$$\frac{I_\nu}{\nu^3} = \frac{I_{\nu'}}{(\nu')^3} \qquad (13.3)$$

(*cf.* Rybicki & Lightmann for the derivation of this).

As an example, let our emitting soure have a power-law synchrotron spectrum, say $S'_\nu \propto \nu^{-\alpha}$ in the rest frame. When you work out the details, an unresolved blob is Doppler boosted by $S_\nu(\theta) = S'_\nu\mathcal{D}^{3+\alpha}$; you get 3 $\mathcal{D}$'s from the frequency ratio in (13.3), and another "α of a $\mathcal{D}$" from the spectrum. Alternatively, a piece of a resolved jet is boosted by $S_\nu(\theta) = S'_\nu\mathcal{D}^{2+\alpha}$; time/space contractions applied to the piece you're resolving account for the change. Either way, because of the strong dependence of $\mathcal{D}$ on θ, this result means that an observer sees the radiation forward-beamed, into an angle $\sim 1/\gamma$. This makes a relativistic jet appear much brighter if you see it end-on.

13.3 Some useful physics

In this section I store a smattering of physical arguments we can make about jets.

13.3.1 Collimation

How to jets stay so well collimated? Why do they not expand? If the jet were propagating in vacuum, it would expand at its internal sound speed. But we ob-

serve jets which remain very collimated over very large distances, with an opening angle only a small fraction of a radian. This would require the flows to be very cold internally, which is inconsistent with other evidence that many of the jets are internally hot. Thus we believe the jet is confined. The two possibilities are

• **Confinement by external pressure.** We know jets propagate through external plasma – the ISM for microquasars, the extragalactic medium for jets from AGN. The external plasma may provide enough pressure to confine the jet.

• **Self-confiment** by magnetic fields. We've already seen that a current generates an azimuthal magnetic field, which can confine the plasma carrying the current. If this applies to jets, the question is then, where and how does the current return to the source?

Which of these two operates can, in principle, be learned from observations; and a given jet may change from being self-confined (close to its origins, say) to being pressure confined (farther out).

13.3.2 Jet transport

Some simple models of jet/RG evolution are based on the rate at which mass, momentum, and energy flow down the jet. Looking back to last term (when we did mass and momentum conservation .. right?), and earlier this term (for energy conservation) we can use the basic fluid equations, and/or simple common sense, to write down the rates. Let the jet have radius r_j, speed v_j, density ρ_j, and enthalpy $h_j = e_j + p_j/\rho_j$ (this is a useful way to collect terms involving the internal energy and pressure of the jet fluid). If everything is subrelativistic, and the magnetic field can be ignored for the moment, we have

$$\text{mass flux}: \quad \dot{M} = \pi r_j^2 \rho_j v_j \tag{13.4}$$

$$\text{momentum flux}: \quad \dot{P} = \pi r_j^2 \rho_j v_j^2 \left(1 + \frac{h_j}{c^2}\right) \tag{13.5}$$

$$\text{energy flux}: \quad \dot{E} = \pi r_j^2 \rho_j v_j \left(h_j + \frac{1}{2} v_j^2\right) \tag{13.6}$$

If t he flow is relativistic, but still ignoring B, these become (remember $v = \beta c$ and $\gamma = (1-\beta^2)^{-1/2}$:

$$\text{mass flux}: \quad \dot{M} = \pi r_j^2 \rho_j \gamma_j \beta_j c \tag{13.7}$$

$$\text{momentum flux}: \quad \dot{P} = \pi r_j^2 \gamma_j^2 \beta_j^2 \rho_j \left(c^2 + h_j\right) \tag{13.8}$$

$$\text{energy flux}: \quad \dot{E} = \pi r_j^2 \gamma_j^2 \beta_j c \rho_j \left(c^2 + h_j\right) \tag{13.9}$$

If the flow is strongly magnetized, we need to include the field energy in the bookkeeping; you remember that electromagnetic energy is transported in a Poynting flux. The details of this are complex and more than we need here; I'll return to this general idea below.

13.4 Larger Scales: the Radio Galaxy

If a jet propagated into vacuum, we might imagine that it would remain unchanged, carrying on forever. But jets don't live in vacuum; rather they propagate into the surrounding plasma – which can be relatively dense intracluster medium (ICM: if the parent galaxy lives in a cluster), or the tenuous intergalactic medium (IGM; if the parent galaxy lives in the "field"). So: when the jet interacts with the surrounding medium[2] the nature of this interaction shapes the Radio Galaxies (RGs) that we see. Observed RGs tend to fall into two morphological types[3], suggesting two different physical situations.

13.4.1 Classical Double radio galaxies (FR II's)

Radio galaxies classified as FR II's – cartooned in Figure 13.2 – are identified by their symmetric, two-sided "lobes" which have bright "hot spots" at their outer edges. Often a narrow, well-collimated jet can be seen propagating through the lobe and almost to the hot spot. These tend to be the brightest ones – so that, even though they are rare by number (only a few per cent of the population), they have received the most attention.

Here's the basic scenario for this type of source. Put a massive black hole down in the center of a galaxy, and turn it on. The jet propagates out, into the ICM; as time goes on, the head of the jet will reach farther and farther from the AGN. We can use simple conservation laws to estimate how this souce evolves with time. The mass and energy flowing down the jet carry momentum; this momentum flux allows the end of the jet to make its way into the external medium (with density ρ_x). Think about a simple ram pressure balance: if the end of the jet advances at $v_D = dD/dt$, it sees a head-on ram pressure $\rho_x v_D^2$. Balancing this against the momentum flux in the jet tells us what v_D must be.

[2] I'll refer to the ambient medium as the "ICM" for short ... meaning ICM/IGM.

[3] Jargon: two authors, Fanaroff & Riley, first pointed out this duality – so RGs today are still usually called "FR I's" (FR Type I's) or "FR II's" (FR Type II's).

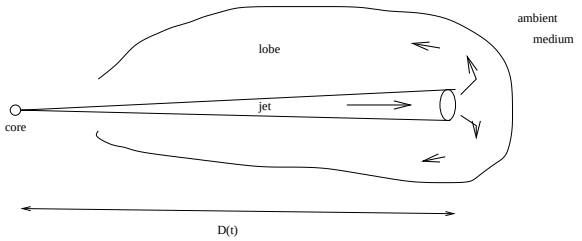

Figure 13.2 Cartoon of (half of) an FRII radio galaxy. Let a jet have constant opening angle, propagating into an external medium (at density ρ_x, say). Because the length of the jet does not grow as fast as the plasma speed within the jet, the plasma must slow down (possibly shock down), and move "sideways", creating a larger "lobe" which surrounds the jet. Because the shock compresses the jet plasma and B field, and accelerates relativistic particles, we expect it to be a bright synchrotron source – *i.e.*the "hot spot" seen in many such sources.

If the jet is less dense than the external medium, then v_D will be less than the jet speed; if the jet is also supersonic we expect a shock to form at this transition point.

Two things should happen at this shock. First, the jet plasma will be compressed and heated when it shocks down. The higher plasma density, and higher B field, will make the post-shock plasma a stronger synchrotron source; if any particle acceleration takes place at this shock, that will further enhance the synchrotron power. Thus, the post-shock plasma should be a localized "hot spot" – matching nicely with what we observe in classical double RGs. Second, the shocked plasma must go somewhere – we expect the high post-shock pressure to "push it off to the side". As the RG grows, this shocked jet plasma will expand to fill a "lobe" or "cocoon" which surrounds the jet. This also matches what we see – in fact the lobes are the brightest, defining, parts of this type of RG.

Recent observational note: if the jet is strong enough, we expect its advance speed, v_D, to still be supersonic relative to the ICM. If this holds, the advancing jet will drive a bow shock in the ICM – such shocks have now been seen in a few cases.

13.4.2 Tailed radio galaxies (FR I's)

Radio galaxies classified as FR I's – as cartooned in Figure 13.3 – are characterized by long, diffuse-looking "tails". These tails begin close to the AGN (sometimes you can see a well-collimated inner jet that suddenly changes to a broader tail; other times you can't), and carry on – broadening gently – for up to 100s of kpc. Unlike FR II's, which are brightest at their outer hot spots, FR I tails tend to be brightest close to the galaxy; the tails gradually become fainter going away from the core. The physical end of the flow – where it runs into the ICM – may or may not be visible.

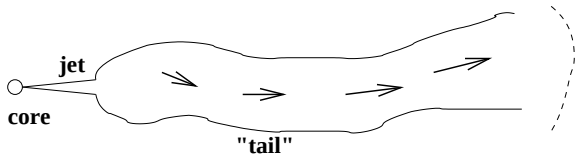

Figure 13.3 Cartoon of (half of) an FRI radio galaxy. When the jet initially leaves the AGN, it is well collimated (probably relativistic and supersonic as well, at least in some cases). But it soon destabilizes – sometimes very suddenly, as in this cartoon. The plasma flow carries on, away from the AGN, but in a more disorganized way.

FR I's are much more common than FR II's; most RG's we know are FR I's. But FR II's have been better studied, partly because they are the bright ones (so were the first ones to be well-observed), and partly because they seem to involve simpler physics.

So, what is the physical picture for FR I's? The cartoon above, for FR II's, relies on the jet remaining stable, collimated, and (internally) supersonic.[4]

The situation will be different if the jet becomes unstable, uncollimated, or subsonic. In this case, the jet flow not retain the "jet/lobe" morphology characteristic of FRII's. Instead, we expect the jet flow to be much more sensitive to local conditions in the ICM. The flow can be affected by local ICM "weather" (flows, turbulence in the ICM); if the parent galaxy is moving rapidly through the local ICM, we can see strongly bent radio tails. If the flow is slow enough, and less dense than its surroundings (which is almost always the case), it will also be affected by buoyancy.

Clearly a wide range of radio-tail morphologies is possible here, depending on local conditions (ICM weather), and also the details of the jet flow. However, the range of possible flows should have one thing in common: they are probably brighter synchrotron sources closer to the AGN (whereas FRII's are brighter closer to their hot spots). We expect this because the radio tail usually expands as it propagates – leading

[4]These conditions are related – it turns out that subsonic, or subalfvenic, jets can easily be destabilized as they pass through a surrounding plasma. (Think of a firehose flapping side to side if the pressure gets high enough ...) Supersonic/superalfvenic jets, on the other hand, tend to be relatively stable in the same situation.

to lower plasma density and B field – thus lower synchrotron power. Synchrotron aging may also matter – the ends of the tails can simply fade away as the relativistic electrons lose their energy.

Simple dynamical models (as we have for FR II's) have yet to be developed for FR I's. While we expect the basic momentum-conservation picture, from above, to hold here, the added dynamical effects of the interaction with the ICM make it harder to find simple models of FR I evolution.

13.5 Unresolved issues

While much of the above picture has remained stable for quite awhile, some issues and questions are getting new attention.

13.5.1 What is the life cycle of a RG?

Two questions come to mind here.

First, what is the duty cycle of a radio-loud AGN? If $\sim 10\%$ of bright ellipticals host a radio galaxy, does that mean that the black hole in every such galaxy is active only 10% of the time? Do AGN, or RG's, alternate between "on" and "off" phases?

Second, where are the old radio galaxies? The simple models described above predict that RG's should keep growing, and remain relatively bright synchrotron sources (*i.e.*, detectable), as long as the jet stays "on". If the jet turns off, the same models predict that the RG's synchrotron luminosity should slowly fade, as the relativistic electrons lose their energy. Because the synchrotron lifetime tends to be long for typical RG conditions, this fading should be slow – we should observe a fair number of "old" RG's (jetless, probably steep radio spectrum).

But we don't see what these models predict. We almost never see RGs that could be called "old"; they all have currently-active jets. In addition, dynamical models (such as you'll see in the homework) generally estimate RG ages only $\sim$ 100 Myr – much younger than the age of the parent galaxy. So: why don't we see either older, jet-on sources, or old, jet-off sources?

13.5.2 How does a jet affect its environment?

This is a very active current area. Clearly the environment affects the AGN; accretion from the *local* environment is what makes the massive black hole "active", and creates the jet. We know the jet transports mass, momentum and energy out from the AGN: what effect does this transport have on the environment?

If the jet/RG is well-coupled to its environment – and that's a big "if" – then the jet/RG can have a strong effect. It carries enough energy to heat the local ICM significantly. Two applications here:

•**Cool cores in clusters of galaxies**. In most clusters (as we've seen in the homework), the plasma density is low enough that the radiative cooling time (from bremsstrahlung) is longer than the Hubble time. Thus, most of the plasma in most clusters hasn't cooled down by much since the cluster first formed. In some clusters, however, the central plasma is dense enough that radiative cooling does matter; these "cool cores" must have an ongoing heating mechanism. All of these cool cores are observed to be centered on a currently-active AGN and RG; do these RG heat the cluster core enough to offset radiative losses?

•**Feedback in galaxy formation**. We now know that a massive black hole sits at the heart of every galaxy. We suspect that black hole formed, and was "active", at about the same time as the galaxy originally formed (we'll discuss this later, in chapter 15). So: did the energy released by that actively accreting black hole play an important feedback role in the galaxy formation process?

Both of the above questions are getting a lot of attention these days. In my opinion, the big, unanswered physics question is, "how effectively does the jet/RG couple to its environment?" That is: how much of its energy does the radio jet/RG actually transfer to the local ICM? The answer depends on the details of how the RG grows and evolves. Does it simply do "pdV" work as it pushes the ICM/ISM out of its way? Does it drive significant shocks or turbulence in the ICM/ISM (which then dissipate and heat the ICM/ISM)? Does the relativistic plasma mix effectively with the ICM – thus releasing the relativistic plasma directly into the ICM/ISM? None of these questions have been answered yet.

13.6 How are jets made?

Finally, a few words about how this all starts.

It seems very likely that every accretion flow involves a jet outflow. And, nearly all jets that we know about are tied to accretion flows (the one exception might be jets from single pulsars; we don't yet know much about them). However, we don't yet have a clear and agreed-

upon picture of how jets get formed. Models out there can be grouped into two broad categories, fluid-based and MHD-based. I think the MHD models are most likely to prove correct; but will include a brief discussion of wind models as well.

13.6.1 Wind (fluid-based) models

These models are probably not the right answer, and pretty much disregarded in recent work. However they were the first type of model proposed; and the internal physics may well still be useful in the more modern MHD models, below. To think about wind models, remember the solar wind: it accelerates from a very slow start, to supersonic speeds at large distances. It does this smoothly, by passing through a "gravitational nozzle" at the sonic point. The wind is driven by its internal energy (and thus, ultimately, by whatever heats it). Linear, one-dimensional flows can also undergo a smooth transition, if the area of the confining channel has a minimum at the sonic point. The first models of astrophysical jets used this analogy – arguing that the flow is accelerated by its internal energy, and that some combination of channel geometry (provided perhaps by the walls of a fat accretion disk?) and gravity produce a high-speed, somewhat collimated outflow.

13.6.2 MHD models

In this discussion I'm partly following several recent papers[5] and addressing jet formation from magnetized accretion flows around a black hole.

What might the magnetic field of an accretion flow look like? Think about a field which threads the disk plasma, but is also "tied" to the distant ISM, which is rotating much more slowly than the accretion flow. The accreting gas will draw the field with it, imparting a ϕ component to the field. One might expect dissipative processes to balance the field growth induced by the accretion, leading to a field shaped something like an expanding helix. Plasma ejected from the disk (for instance in a wind) will be both channeled and accelerated by such a field. Magnetic pressure gradients up the rotation axis will accelerate plasma "up and out" of the system; magnetic tension (from the helical field lines) will exert a "hoop stress" and confine the plasma as it goes.

[5] e.g. Meier, D. L. 2005, *Astrophysics & Space Science*, 300, 55; Meier, D. L. & Nakamura, M. 2006, ASP Conference Series, 350, 195

There are also a couple of variants worth mentioning.

• **Poynting flux models.** These describe the limit in which the plasma density close to the source is small compared to the field energy. As we've seen, a rapidly rotating $\mathbf{B}$ field generates an $\mathbf{E}$ field (just as in pulsars). There will very likely be a component of the Poynting flux, $\mathbf{S} \propto \mathbf{E} \times \mathbf{B}$, along the rotation axis, which will be a strong source of power lost from the accretion system. In order to produce what we observe as radio jets, these Poynting-flux jets must "mass-load"; they must either pick up stray charges from the ambient medium, or generate their own plasma through magnetized pair production. Such a process may well account for the origin of relativistic jets close to a black hole; they would then become visible as they gained mass.

• **Penrose jets.** Another class of models extracts rotation energy from the black hole to power the jets. (I'm inventing the name; the authors of two similar models are Blandford & Znajek, and Punsly & Coroniti). Think back to rotating black holes. You recall that *frame dragging* is important near the event horizion – this is the azimuthal motion induced by the hole's rotation. Think now about a piece of magnetized accreting matter, with field lines again tied to some slowly-rotating distant point. As the matter gets close to and crosses the event horizon, frame dragging will speed it up, thus generating helical (Alfven) waves which move out along the field lines. This will again generate an outwards Poynting flux (which can mass-load and become a visible jet); and the reaction force ends up making the plasma counter-rotating as it passes through the event horizon. Thus some of the black hole's rotation is lost, to supply the power carried out by the yet.

13.6.3 Duty cycles?

Accretion flows in galactic X-ray binaries provide a possibly interesting clue. Think back to our discussion of accretion disks, from last term: they are mostly thermal (Black body) sources, possibly with an optically thin corona (could be a bremsstrahlung source). We found that accretion disks around small (star-sized) compact objects becomes hot enough to radiate in the X-ray band. It turns out that X-ray emission from galactic binaries has two states. It can be "low/hard", meaning lower X-ray power and a harder (nonthermal?) X-ray spectrum. Or, it can be "high/soft", meaning higher X-ray power and a softer (thermal) X-ray spectrum. It turns out that steady radio jets are present

in the low/hard state, but *not* in the high/soft state. The difference between the two states is thought to be the accretion rate – a lower $\dot{M}$ leads to lower X-ray power (which is constent with the simple accretion-disk models we saw last term). Thus: perhaps a lower $\dot{M}$ somehow changes the accretion mode in galactic microquasars, and allows a jet to form?

It is not clear whether this scenario also applies to AGN. There are arguments on both sides, and the observations don't (yet?) support the existence of two such states in AGN. If AGN do turn out to work this way, we may have a physical origin for AGN duty cycles – but to my mind, this issue isn't settled yet.

Key points

- Jet phenomenology: what we know from observations.
- Important relativity: superluminal motion and beaming.
- Basic jet "fluid physics": collimation, propagation.
- Radio galaxy types & cartoons: FR I's, IIs
- How does the jet affect its surroundings?
- Current ideas of jet origins – mostly MHD.

14 Quasars and Active Galactic Nuclei

AGN astronomy started in the 1960's. Early radio surveys had found bright, compact radio object with no clear optical identification. Most people thought they were simply radio-loud stars, but some thought they were extragalctic ... in 1963 a spectrum was obtained of the radio source 3C273. The likely optical ID was a 13th magnitude blue object, apparently stellar (not extended), with faint linear emission nebulosity. This turned out to have very unusual spectra for a star, rich with emission lines. These were thus called "quasi-stellar objects" (QSOs), or quasars for short.

Identifying the emission lines was hard, until someone realized the object was at the very high (at the time) redshift, $z = 0.16$. After that, people started looking seriously for these objects. The bright emission lines and blue continuum were easy to find. By now we know of several thousand...number counts find about 100 quasars per square degree of sky, with $z \lesssim 2$ and blue magnitude $m_B < 22$ (with more, of course, at fainter magnitudes and higher z).

14.1 Basic properties: observations

Although we now know (or think we know) that all the varieties of AGN are driven by accretion onto a massive black hole in the core of the galaxy, that was not at all obvious when the field started. Thus, the literature is dominated by the unusual observational properties of bright AGN.

14.1.1 Spectral lines

Strong optical emission lines are characteristic of (almost) all AGN. They are more or less the same lines that we see in galactic nebulae – planetary nebulae and HII regions. Emission lines from AGN seem to obey very similar physics – and thus arise in similar conditions. They are probably ionized by the observed nuclear continuum source (the so-called "central engine"; AGN are strong in the UV/soft Xray region), although shock ionization is also likely in some cases.[1] The striking feature of the emission lines is their width. Permitted lines (such as Hα, the $n = 3-2$ transition of H) have typical linewidths $\Delta v = c\Delta\nu/\nu \sim 10^4$ km/s. The forbidden lines (from heavy elements; OII, OIII, NII, etc.) can be narrower, more like $\Delta v \sim 10^3$ km/s. These widths are much too high to be due to thermal broadening;[2] these widths must be due to bulk motion of the emitting gas clouds – either random or orbital motion.

The difference between broad and narrow emission lines seems to suggest that the two line types are formed in different regions – perhaps the permitted lines come from denser gas, closer to the central engine while the forbidden lines come from less dense gas, a bit further from the engine. Not all AGN have this separation, however; some Seyferts have mostly narrow-line emission, while some quasars have mostly broad lines.

14.1.2 Continuum emission

AGN radiate in every frequency in which we've looked: from radio, through infrared, to optical and UV, thence to X-rays and γ-rays. The underlying spectrum is broad-band, with 2 or 3 separate features. Remember your radiation: telescopes generally measure intensity, f_ν say, which is power/Hz. The energy in the spectrum is better determined by $\int f_\nu d\nu \sim \nu f_\nu$; this is often what's plotted. The total emission νf_ν is typically dominated by two (or maybe 3) peaks: one IR/optical (which itself may be two components), and a second peak in the hard X-ray to γ-ray region. It seems likely that the bulk of the energy comes out in the optical/UV region, say $\nu \sim 10^{14} - 10^{16}$ Hz. (This dominant "bump" in the νf_ν plot is called the Big Blue Bump, I kid you not.) However, some radio-loud sources have been observed out to hard γ-rays; their broadband spectra are dominated by the hardest frequencies, $\nu \sim 10^{20} - 10^{25}$ Hz (compare: what frequency corresponds to the electron rest mass?)

[1] Detailed analysis of the emission line strengths can tell us the likely cause of ionization, as well as the local density and temperature of the emission line gas. One piece of terminology is necessary. Remember that quantum mechanics gives us the probability that an atom in an excited (upper) state can make a spontaneous transition to a lower state, emitting a photon in the process. *Permitted lines* have high transition probabilities; what are called *forbidden lines* have lower transition probabilities – but are not forbidden in the QM sense. Rather, their radiative transition rates are low enough that in dense conditions they will be de-excited by collisions first. Thus, there is a *quenching density* associated with each particular transition.

[2] how hot would the gas be? how could ions such as OII or OIII ever exist at such temperatures? how low would the hydrogen recomination rate be, and how would you ever see hydrogen permitted lines?

14.2 The AGN zoo

We now know that quasars are unusual nuclei in otherwise normal galaxies. Other types of nuclear activity are also possible; for historical reasons (who found what first, in what observing band), a wide variety of names and acronyms exist in this field. All of these objects, as well other subclasses and acronyms I'm not bothering to put down, are grouped together as Active Galactic Nuclei (AGN).

14.2.1 The radio-quiet ones

Although quasars were originally detected based on their radio emission, radio-loud quasars turn out to be rare, something like 10% of the quasar population. Radio-quiet quasars (usually "QSO's") are domiated by their bright core, with strong optical emission lines as well as strong continuous emission (optical/UV, also X-rays).

Seyfert galaxies and **Markarian** galaxies[3] are spiral galaxies with nuclei which are quasar-like in their broadband and line emission properties, but not so bright as the quasars. Seyferts are never radio-loud; they have weak radio cores and can have small, weak, poorly collimated radio jets which do not propagate out of the galaxy (due to the denser ISM? due to different conditions in the core?) Weaker versions of Seyferts, also in spiral galaxies, are called **LINERS** (Low Ionization Nuclear Emission Region) – with emission lines but less nonthermal continuum than the Seyferts.

14.2.2 The radio-loud ones

As we saw in the previous chapter, The term "radio galaxy" refers to an elliptical galaxy with a radio jet and double-lobed (or tailed) radio structure on supergalactic scales (linear extent, side-to-side, can range from $\sim$ 100 kpc to a few Mpc). The term "radio-loud quasar" (QSR) refers to a quasar with the same sort of radio jet-lobe structure. The distinction is mainly a question of how strong the nuclear activity is (as seen by us, anyway): how strong is the nonthermal continum and/or the emission lines?

[3]both named for the person who originally cataloged them – based on unusually strong nuclear emission lines or UV continuum.

14.2.3 Blazars and friends

These are a subset of the radio-loud ones (about 10%) with unusually bright, active, variable radio-to-optical cores. People in this field speak in acronyms:

OVVs (optically violent variables). These are the highly variable, clearly relativistic ones. They show a flat radio spectrum (suggestive of a compact, synchrotron self-absorbed core and a pc-scale radio jet with a few bright knows). They are highly variable, timescales from days to years.[4] These are the ones that show clear superluminal motion.

BLLs, as described, are named for BL Lacertae (an object previously thought to be variable star locally). The emission lines are quite faint compared to the cointinuum. As with OVVs, the BLL radio emission is strongly polarized, flat spectrum, & highly variable ($\delta t \gtrsim$ days).

Both of these are often grouped together as *blazars*. The general picture – motivated by the observations of superluminal motion – is that we are looking "down the pipe" of the jet, close to its axis. In addition to explaining the superluminal motion, such a geometry will enhance the variability (by relativistic effects and by the Doppler-beaming effects on a jet with some slight inhomogeneities.

14.2.4 Parent galaxies

What are the parent galaxies? We noted above that radio galaxies are found in elliptical galaxies, while Seyferts (and related radio-weak AGN) are found in spirals. What about quasars? There has been a sense that radio-loud quasars live in E's, while radio-quiet live in S's. This view seems to be changing, however, at least for the brightest of the population. Quoting from Dunlop etal (2003): "Virtually all powerful AGN live in normal, massive ellipticals: the parents agree with local bright E's in morphology, luminosity, scale-length, consistency with fundamental plane, axial-ratio distribution, and colors (evolved stellar population, age 10-13 Gyr)." "The inevitable conclusions is that these galaxies, or at least most of their stars, must have formed at high redshift." "Spheroidal hosts become more prevalent with increasing nuclear luminosity; for bright enough nuclei, the hosts of both radio-loud and

[4]There is an argument currently going on about variability on much shorter δt's, less than a day; these are the Intra-Day Variables. The variability is probably due to interstellar scintillation, although some people want it to be intrinsic.

radio-quiet AGN are massive ellipticals."

14.3 The usual model: a massive BH

We know the answer of course (or think we do): all of this comes from accretion onto a massive black hole. Let's see how (or if) this model can explain the variety of observations.

14.3.1 Zoom in: the central kpc and within

The first clues may have come from the emisison lines. In most objects we need photoionization; with detailed modelling one can pinpoint the photon *density*, which means the distance from the central engine. Applying this to the broad and narrow lines shows that the NLR must be at $\sim$ kpc from the engine, while the BLR is closer, more like pc to tens of pc away. Both broad and narrow line emission must be "patchy", for instance coming from dense clouds.[5] It may not be correct, however, to picture the clouds as uniformly distributed over the entire central kpc. We are learning, from HST imaging, that the emission-line gas in the central kpc is often confined to emission "cones" – what you would expect, for instance, if the ionizing radiation escaped from the central engine only in a moderately narrow cone.

14.3.2 Zoom in further: the central pc and within

This is, of course, the region of the central engine; the region we probe indirectly through variabilty, and directly through milliarcsecond radio imaging (as with the VLBA). The generic picture is, as we've discussed, accretion onto a massive black hole, probably with a jet being driven out the rotation axis. Figure 14.1 illustrates the general thinking. The accretion disk is mostly thermal, that is resembling the models you saw in P425: spatially thin and optically thick. It emits as a black body, and is thought to be the origin of the "big blue bump". Nonthermal emission – high energy, X- and γ-rays – comes from an optically thin region somewhere close to the black hole and the inner region of the disk; very high gas temperatures, possibly relativistic plasmas, and electron-positron physics may be going on here. A two-sided jet is driven out; if this is in an elliptical the jet propagates to extragalactic scales. The gas emitting the broad emission lines is sketched as discrete clumps, moving at high speed (here guessed to be random). The intercloud region is probably a hot wind, driven out from the accretion disk by a combination of its own temperature and magnetic/centrifugal effects.

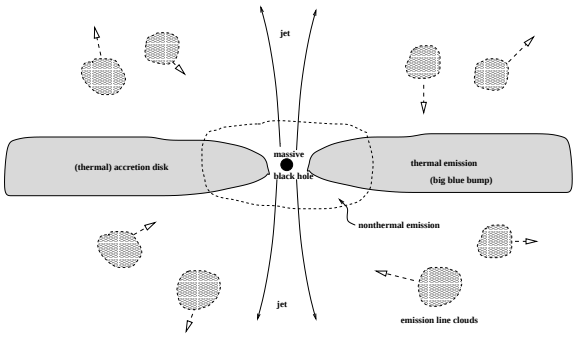

Figure 14.1 Generic cartoon of the central pc of the AGN. An accretion disk feeds a massive black hole. Details are in the text. The scale of this cartoon is a few pc (remember the event horizon of the black hole is only $\sim 1 - 10$ AU.

14.3.3 Why radio-loud vs. radio-quiet?

We still don't have much of an answer to this question, or even a standard "toy model". Radio galaxies, and maybe all of the brightest quasars, live in ellipticals; Seyfert nuclei, and maybe most of the radio-quiet quasars, live in spirals. Why? I wish I knew.

14.3.4 Why are only some galaxies "active"?

This one is even harder to answer. We now know that every galaxy contains a massive black hole at its core (check back to chapter 1 for discussion). But this black hole is "active" in only a few per cent of galaxies. Why? Once again, I wish I knew.

14.4 Unification Models

Can these various disparate categories and acronyms be explained by one simple, unified picture? Maybe, maybe not .. this issue has strong adherents, because simple pictures are pleasing and attractive. But one can push too much for simplicity, at the expense of ignoring some of the physics ... some of the community belong to "the unification church", while others (including your author) remain agnostic.

There seem to be two important motivations for this view, as follows.

[5] Why? One, we can see the engine, so the line-emitting gas doesn't completely cover it. Two, we know the density of the line emitting gas, from quenching and line-ratio arguments; comparing this to the total volume at the clouds' distance also requires the volume be incompletely filled.

14.4.1 Relativistic beaming

We've already met this several times. If the jet is driven out at relativistic speeds, with Lorentz factor γ – and we know it is, from the observation of superluminal motion – then a viewing angle within an angle $\sim 1/\gamma$ of the jet axis is special. From this vantage, *only*, we will see $v_{app} > c$, strong forward beaming, enhanced variability, and etc – that is we will see a blazar.

14.4.2 Obscuration and tori

We noted above that the optical line emission is anisotropic in some nearby AGN, being located in an ionization cone. Such a cone might arise from a fat torus (think of a donut) surrounding the central engine, allowing hard photons to escape only in a fairly broad cone along the axis of the torus. In addition, emission lines from Seyferts (which are nearby, and bright, enough to do this measurement) appear different in polarized than total light. Some Seyferts that show only a narrow line region (it's called a Seyfert 2, if you care) turn out to have broad line wings when seen in polarized light. Remember that Thompson scattering polarizes the light. Thus, if the light from the central engine, which sits at the heart of the donut, is scattered (by some diffuse ionized gas, above or below the system), an observer can detect the broad lines – which come from very close to the black hole – at angles at which the direct emission is obscured by the torus.

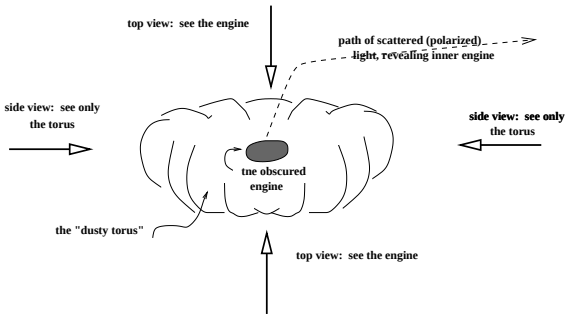

Figure 14.2 Extremely generic cartoon of the unified model. The central engine may be surrounded by a fat, opaque torus (maybe containing dust). Whether or not you see the central engine depends on your viewing angle. The scale of this cartoon is a few hundred pc.

14.5 AGN demographics

To summarize: we've walked through the general picture of an AGN: a massive black hole sits in the core of a galaxy. Galactic material (ISM) accretes onto the black hole. Side effects of that accretion are the generation of a strong radiation over all frequency bands (by a mix of thermal and nonthermal processes), and the ejection of collimated plasma jets (which may or may not propagate out of the galactic core).

The remaining questions, then, are demographic. How common are AGN, and how were they distributed over the history of the universe? Why does every bulge contain a black hole, and which formed first (the bulge or the BH)? Does every nuclear black hole make an AGN? How did the black hole come to exist in the galactic core (is it the chicken or the egg?) How have these AGN evolved over the course of the universe? To answer these questions, we need to go to high redshift - $z \sim 2 - 10$, say, to see what happened at early times.

These questions are far from being answered, but we do have some important clues. In case you're not familiar with high-redshift thinking, I've put some basic cosmography in an Appendix to this chapter.

14.5.1 Was there a "quasar era?"

Quasars – like other AGN – are rare in today's universe; most galaxies do not contain a currently active central engine (even if, as we now know, they probably contain a massive BH). This is not the case at early epochs, however. Two facts become clear from quasar surveys.

• At a given epoch – today, for instance – there are more faint quasars than bright ones. The number per volume at an absolute (optical, blue) magnitude M_B can be written very approximately as $\Phi(M_B) \sim \Phi_o M_B^{-0.7}$.

• The absolute number of quasars – either at a given magnitude, or integrated over all luminosities – was greater at early epochs. (This is the same thing as saying that the constant above, Φ_o, is a function of z.) The space density of quasars rises out to $z \sim 2$, has a broad peak in the range $z \sim 2-3$, then decays again at higher z. This epoch, which saw a lot of quasar activity, is called the "quasar era".

What has changed, with quasars, since the quasar era? What is this evolution of the number density saying? Up to a few years ago, there were (at least) two alternative possibilities. One is that the fraction of quasars, per galaxy population, has stayed the same with time, but the luminosity of a given quasar was much brighter at $z \sim 2$ than it is now. Because all quasar surveys must be flux limited, we would be detecting more of

them early on. This is called *luminosity evolution*. An alternative theory is that the mean quasar luminosity has not evolved with time, but that there were simply more of them back then ... so that the increase in measured numbers of quasars back to $z \sim 2$ is what it appears to be. The real answer is probably somewhere inbetween these two extremes...but we must remember that almost every nearby galaxy (at least) harbors a black hole. If we define this as a dark quasar, then luminosity evolution must be the preferred model. However, we are learning more about galaxy formation and evolution, and the question may be more complex.

14.5.2 What about galaxy formation?

The quasar counts, described above, have been known for some time. A currently very active research area tries to understand when (at what z) galaxies formed. This is much harder than counting quasars (which are easy to pick out, they are bright and small and blue). How can we pick out galaxies in the act of formation? Think about a protogalaxy: picture a self-gravitating clump of stuff collapsing due to its own gravity, and/or being enhanced by the accumulation of smaller-sized clumps (mergers). The baryonic matter will lose energy (by radiation), fall through the dark matter, and must eventually form stars. We know ellipticals have little active star formation today; their stars must have formed in one great rush, at early times. Spirals today do have ongoing star formation, but they probably also had a strong starburst phase when they first formed.

So, the general idea is to look for unique signs pointing to early bursts of star formation. We know a lot about local star formation regions. Optically, they are bright in Hα, and also in the UV (due to all those hot young stars). Moreover, we know they are dust-enshrouded, and thus very bright in the IR or sub-mm band. Thus: we can try to find distant objects that are bright in the UV, or in Hα, or – the currently most promising – in the IR/sub-mm.

The tentative result of these surveys (the work is still ongoing, and the arguments as large as the error bars at high z) is very interesting. Given the variety of observational probes, which mix apples and oranges, people generally turn their data into "the rate of star formation as a function of z".[6] This is called a "Madau plot", after the person who first did one.

Difficulties in the analysis aside, everyone in this field agrees on the low-z result. The number of strongly-starbursting galaxies increases with redshift, out to $z \sim 2$ – which is just the quasar era. The uncertainties come at higher redshift: it is not clear if the star formation rate declines again, at higher z (like the quasars), or continues more or less steady (out to, say, $z \sim 5$, the current limit of this sort of work).

14.6 Ending with questions

All of this raises some intriguing questions, which seem as good a way as any to end these notes.

- Did the bulk of galaxy formation take place at the same time as quasars were most active?
- Is a strong starburst a part of the mass-accretion process which kept the AGN powerful at that epoch?
- Are the merger and dissipation events that made the bulge or E galaxy the same events that made the quasar shine?
- Are the majority of nuclear BH dark today because they aren't being fed? Are galaxy mergers necessary to transport the gas (i.e. BH food) close to the nucleus?

And ... that's it, folks! It's been fun – have a good summer!

Key points

- The phenomenology: important observational trends..
- The usual model: what's there on sub-pc scales?
- "Practical cosmology": what does "high z" mean (ages, timescales, etc)?
- The QSO epoch and how it might connect to galaxy formation.

---

[6] This can be a long daisy chain. For instance: measure an IR luminosity, say; from that determine how many bright, massive stars are needed to power the dust; divide by the main sequence lifetime of the massive stars; and you have "derived" a "star formation rate". Or ... measure the radio luminosity; use the fact that the radio and IR powers are well correlated for nearby spiral galaxies, to estimate the IR luminosity; and return to top..

14.7 Appendix: a little practical cosmology

Before we can talk about BH and AGN in the early universe we need to review the language used. The context is cosmology: we work with the *scale factor*, $R(t)$. This is a quantity which describes ("scales") the distance between two fixed points in an expanding universe – say two nearby galaxies. The evolution of $R(t)$ reflects the fight between gravity (an attractive force), initial conditions (how much expansion did the universe start with?), and the effect of any vacuum energy density (the cosmological constant; "dark energy"). For our purposes here, let's assume that $R(t)$ is a known function (found from the solution of Einstein's field equations). Everything we can measure about a distant object – size, luminosity – depend on how much the universe has expanded between the time it emitted a light signal (t_{em}) and the time we receive that signal (today; usually called t_o).

14.7.1 Just what is the redshift?

The redshift is defined is $1 + z = \delta\lambda/\lambda$ (the shift in an emission wavelength, say of a spectral line, relative to its rest value). The redshift of a cosmological object is an important quantity. It tells us not just the recession speed (which isn't very interesting), but – more importantly – the distance and age of the object.

There are at least 3 physical causes of a redshift.

• You remember the simple Dopper shift: for low speeds, $1 + z = v/c$; and for relativistic speeds, the Lorentz transform in simple geometry gives $1 + z = \gamma(1 + \beta)$.

• There is also a gravitational redshift, which occurs when light emitted in a potential well (say at the surface of a star) climbs out to the observer. For a Schwarzchild geometry this becomes $1 + z = 1/(1 - r_s/r)^{1/2}$.

• Finally, there is a cosmological redshift: the frequency of a signal decreases due to the expansion of the universe while that signal propagates to the observer. This is the one we want here: it becomes $1 + z = R(t_o)/R(t_{em})$.

Because the cosmological redshift is intimately tied to the expansion of the universe, it becomes a handy way to describe the distance and age of an object. Two things are worth noting here.

14.7.2 The Hubble diagram

Think about a source at some distance D. In Euclidean space its light spreads out over a surface of area $4\pi D^2$ before it gets to the observer, so that one sees a flux $\propto L/D^2$. But, this changes if space is not Euclidean. The key fact is that the area of the D-sphere is not $4\pi D^2$. The details of how the area changes depend on the cosmological parameters (curvature and cosmological constant), but the general trend can be sketched, as in Figure 15.1. NOTE that for $z \ll 1$, curvature effects don't matter; $d_l \propto z$, just as in Euclidean space. The original Hubble's law ($v = H_o z$) applies in this limit. Also note that the so-called Hubble constant, H_o, is the slope of this curve *as* $z \to 0$ (that is, it isn't a constant, but a variable).

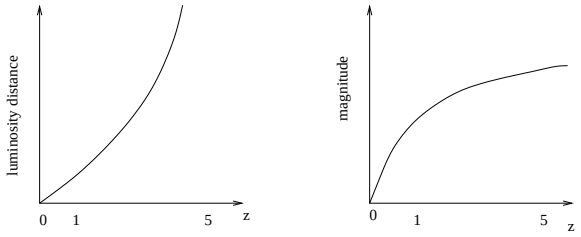

Figure 14.3 How the curvature of space affects the luminosity distance, d_l (defined so that the observed luminosity of a source at redshift z is $L/4\pi d_l(z)^2$). Note that for small redshifts, $d_l \propto z$, and the classical Hubble's law is recovered. The right hand diagram is often called the Hubble diagram; its high-z extension can, in principle, be used to measure cosmological parameters.

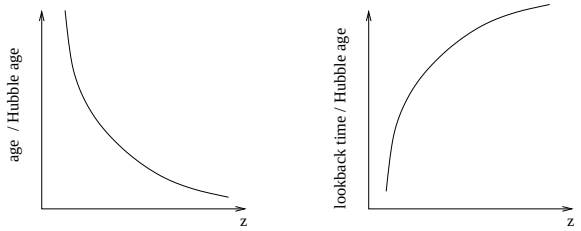

Figure 14.4 How the age of the universe, and of an object we observe, depend on redshift. Left: the age of the universe (counted from $t = 0$, the big bang) when an object had redshift z. Right, the lookback time – the difference between the age of the universe at z and the age of the universe now. This tells us how long ago an object existed, if we see it at z today. The "Hubble age" is defined as $1/H_o$.

14.7.3 The lookback time

How long ago did an object exist, if we see it now at redshift z? If space were Euclidean, that would be simple: (age at emission) = (today's age) - (distance

/ lightspeed); and the lookback time would be proportional to the redshift. But as with the Hubble diagram, space curvature makes this more interesting for $z \gtrsim 1$. Numbers: $z = 2$ gives a lookback time $\sim 0.5/H_o$; $z = 5$ gives a time $\sim 0.6/H_o$, and for higher z's the lookback time doesn't change by a lot.