Department of Physics

ASTM-052: Extragalactic Astrophysics

by

Peter Clegg

Note 4. Active Galaxies

Table of Contents

Chapter 4 : Active Galaxies 1
1. Introduction 1
2. Properties of Active Galaxies 1
2.1 Overview 1
2.2 Seyfert Galaxies 2
2.3 Quasars and BL Lacertae Objects 2
2.4 Radio Galaxies 3
3. Models of Active Galactic Nuclei 4
3.1 The Source of Energy in AGN 4
3.1.1 Nuclear or Gravitational Power? 4
3.1.2 Accretion Power 4
3.1.3 The Eddington Limit 5
3.2 Characteristic Temperatures 6
3.3 Accretion Discs 7
3.3.1 Structure of the Disc 7
3.3.2 Conservation of Mass 7
3.3.3 Conservation of Angular Momentum 8
3.3.4 Conservation of Energy 8
3.3.5 Viscosity in the Disc 9
3.3.6 Spectrum Radiated by Disc 10
4. Models of Radio Sources 12
4.1 Synchrotron Radiation 12
4.1.1 The Frequency of Synchrotron Radiation 12
4.1.2 The Power emitted by a Relativistic electron 12
4.1.3 The Synchrotron Spectrum 12
4.1.4 Lifetime of Relativistic Electrons 13
4.2 Models of the Radio Lobes 13
4.2.1 The Energy in the Lobes 13
4.2.2 The Supply of Energy to the Lobes 15
4.2.3 Motion of the Jets and Lobes 16
4.2.3.1 Superluminal Velocity in Jets 16
4.2.3.2 Velocity of the Lobes 17
4.2.4 Confinement of the radio Lobes 18
4.2.4.1 Diameter-Separation Ratio 18
4.2.4.2 Inertial Confinement 19
4.2.4.3 Ram-Pressure Confinement 19
4.2.4.4 Thermal Confinement 20
4.2.4.5 Summary and Conclusions 20
4.3 The Particle Energy Spectrum 20
4.3.1 The Need for an Acceleration Mechanism 20
4.3.2 Stochastic Acceleration 20
4.3.2.1 General Scheme 20
4.3.2.2 Elastic Collisions between Particles 20
4.3.2.3 Growth of Energy 21
4.3.2.4 Loss-rate 22
4.3.2.5 Resultant Energy Distribution 22
4.3.3 Shock Acceleration 22
4.3.3.1 Energy-Gain 22
4.3.3.2 Loss-rate 23
4.3.3.3 Resultant Energy Distribution 23
Bibliography for Chapter 4 23

CHAPTER 4: ACTIVE GALAXIES

1. Introduction

The number of known types of galaxies showing unusual activity has grown over the years to the point at which the activity itself can hardly be considered unusual, even if its origin is not well understood. Moreover, similar sorts of activity, albeit on a smaller scale, are seen in objects usually considered to be "normal". Many astronomers are therefore coming to the view that we are seeing different mixtures, on different scales, of one or two basic phenomena.

In this section, I shall first identify the various types of active galaxy and discuss their properties. I shall then review the whole phenomenon.

2. Properties of Active Galaxies

2.1 Overview

Active galaxies are frequently described as showing violent or energetic activity. Perhaps a more useful definition is galaxies displaying phenomena that cannot be ascribed to normal stellar processes, although this begs the question of what constitutes a "normal" stellar process. It could be, for example, that the activity seen in some nuclei is the result of a violent burst of star-formation. In all cases, the activity seems to be connected with the nucleus of the galaxy, whence the term *active galactic nucleus* (AGN). In general, AGN show some combination of the following features:

1. They have bright, star-like nucleus. In a short-exposure image of the galaxy, only the point-like nucleus is apparent. On longer exposure, the nucleus saturates the detector – photographic plate or CCD – and the rest of the galaxy becomes apparent, although for *quasars*, this is only seen with great difficulty.
2. The nucleus radiates over a very wide range of wavelengths, from radio (in some cases) to X-rays. This is unlike stars, which radiate mostly at optical wavelengths, and normal galaxies, which are dominated by starlight and the infrared emission from cool interstellar dust.
3. The spectra of AGN are unlike those of normal galaxies, being very blue and having no absorption lines characteristic of starlight, but having strong emission lines.
4. The spectral energy distributions (SED), defined by

$$\mathrm{SED} \coloneqq \nu S(\nu), \tag{2.1}$$

of AGN are quite unlike those of normal galaxies. For most AGN, the SED is roughly constant but follows a power law from the hard X-ray region of the spectrum to the far-infrared (far-IR), the radiation being unpolarised. In *radio-loud* AGN, this continues out to the radio region. In *blazars*, the SED has a smooth broad hump and is strongly polarised.

5. The SEDs typically have two "bumps" on top of the power-law.

   The *blue bump* is an excess, rising from the optical to the ultraviolet (UV). There is some evidence that this bump is seen descending again in the soft X-ray region in which case it presumably peaks in the extreme ultraviolet (XUV), at $\nu \sim 10^{16}$ Hz; this corresponds to a temperature of ~ 300,000 K. The peak cannot be observed directly because the Galactic interstellar medium is opaque in this region of the spectrum.

   The X-ray bump occurs in the hard X-ray region with $S(\nu) \propto \nu^{-(\sim 0.7)}$ [so that $\nu S(\nu)$ is rising with frequency] and is still rising at $E > 20$ keV, corresponding to a temperature of ~ 10^8 K.
6. The strong emission lines are much too highly excited to be the result of ionisation from OB stars and are presumably ionised by the photons from the blue bump. The lines can be very broad, presumably the result of Doppler broadening with velocities ~ 10,000 km s^{-1}.
7. Many radio sources have double *radio lobes*, symmetrical "blobs" on either side of the central galaxy. These are very large structures, tens of kiloparsecs in size and separated from the central galaxy by ~ 50 kpc to 1 Mpc. The spectrum of radio emission from these lobes is a power law of the form $S(\nu) \propto \nu^{-(\sim 0.7)}$. There is often a bridge of radio emission between the lobes and the central galaxy.
8. Many radio galaxies have relativistic jets of material pointing at the lobes. These jets seem to be made up of blobs of material that appear to be moving away from the galaxy faster than light! Although we shall see that this is an optical illusion, it *does* mean that the material in the jets must be travelling at velocities close to that of light.
9. The brightness of AGN can vary very rapidly. The X-rays from NGC 4051, for example, change by a factor of 2 in 30 minutes! The variability tends to depend upon wavelength:

X-ray	Hours to days
Optical	Weeks to years
Far-infrared	Invariable

10. Since an object's output cannot change coherently faster than the time it takes light to cross it, this puts limits on the sizes of the emitting regions and suggests enormous powers are being generated in very small regions, sometimes smaller than the solar system (see below).
11. The total power of AGN is hard to estimate because of the great spread of wavelengths and also because of the variability. The weakest known AGN are in

nearby galaxies with luminosities $L \sim 10^{33}$ W, whilst the most powerful distant quasars have $L \sim 10^{40}$ W. You should compare this with our Galaxy, which has $L \sim 10^{10}$ L_{sun} or $\sim 10^{33}$ W. In the most luminous cases, the nucleus can outshine the whole of the rest of the galaxy!

Before trying explain these observations, let us look briefly at the various types of active galaxy; a fuller discussion is given in [1]

2.2 Seyfert Galaxies

Carl Seyfert first noticed a class of spiral galaxy with very bright nuclei in 1943. These were the first AGN to be recognised and represent the mildest form of activity. As well a dominating the light from the rest of the galaxy, the output from Seyfert nuclei can vary in less than a year. This sets an upper limit to the size of the region responsible for the emission.

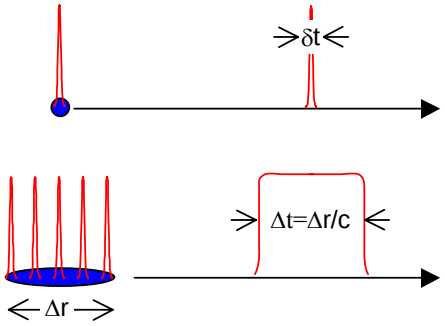

Figure 4-1. Limits on size of variable objects.

The upper part of Figure 4-1 shows schematically a *point* source whose luminosity increases and decreases on a time-scale δt. The lower part of the figures shows an extended source, each part of which brightens and fades on the same time-scale δt. The *overall* brightening and fading of the source, as seen by an observer, occurs over a longer time-scale Δt because the observed change in luminosity of the more distant parts of the source arrives later than that of the nearer parts. In general, we deduce that in an object whose luminosity varies on a time-scale Δt, size of the region responsible for the is emission is cannot be larger than Δr, where

$$\Delta r \lesssim c\Delta t \ . \qquad (2.2)$$

One light year, which is about a third of a parsec, is extremely small compared with sizes of galaxies, which are the tens of kiloparsecs in diameter.

As with other AGN, the spectra of Seyferts differ from those of "normal" galaxies. The latter have a continuum spectrum – corresponding to the blackbody curves of stars – containing the absorption lines of the stellar atmospheres. Seyfert spectra have *emission* lines sitting on top of their continua: this suggests the presence of hot gas. The continua themselves, are not the thermal spectra that are seen, for example, in HII regions.

Spectroscopy of Seyferts also distinguishes two types, Seyfert 1 and Seyfert 2. In Seyfert 1 spectra, the hydrogen Balmer lines (and other *permitted* lines) have velocities some thousands of km s^{-1} wide. It is clear that the thermal motions of individual atoms cannot produce these widths: the temperatures necessary for such thermal broadening would ionise the hydrogen and there would no Balmer lines to be broadened! If the broadening does not arise from the motion of the individual molecules, it must be caused by *bulk* motion of the gas: either the gas is rotating at high speed around the centre of the galaxy or it is being expelled at high speed. Remember that typical orbital velocities in normal spiral galaxies are hundreds of km s^{-1}. If the widths are caused by rotation, it implies that the Seyfert galaxies have masses of $\sim 10^9$ M_{sun} within less than a light-year of their centres. This represents a very high density. Could Seyfert 1 galaxies harbour a black hole in their centres?

Seyfert 1 spectra also contain *forbidden* lines which are much narrower (hundreds of km s^{-1}) than the hydrogen lines. Presumably these lines are formed further from the nucleus where the orbital velocities are lower. [The velocities are still higher than typical in normal galaxies.] Because they are *forbidden* lines, they must certainly come from regions of relatively low density.

In Seyfert 2 spectra, permitted and forbidden lines have the same widths, both being several hundred to a thousand km s^{-1}. It is likely that Seyfert 1s and 2s are not completely different types of object, the broad-line regions existing in Seyfert 2s but being obscured by dust. There is support for this in the fact that Seyfert 2s tend to be stronger infrared sources than Seyfert 1s, suggesting that the ultraviolet radiation from the broad-line region is absorbed by dust which then re-radiates it in the infrared. Alternatively, Seyfert 2s could be Seyfert 1s in which the central source of power has been "turned off" so that the gas in the broad-line region is no longer excited. This is a possible explanation for NGC 4151's having been seen to change from type 1 to type 2.

Seyfert galaxies make up about 1% of the population of spirals. An obvious interpretation is that 1% of spirals "choose" to be Seyferts all their lives. There is an alternative explanation, however. Because of the time-scales involved, we are unable to wait long enough to see galaxies evolve significantly. When we look at the sky, therefore, we see a snapshot of all galaxies at one moment in their evolution. It is possible that all galaxies go through a "Seyfert phase", spending 1% of their lives as Seyferts. Since there are signs of some level of activity in almost all galaxies, this is perhaps the more likely explanation.

2.3 Quasars and BL Lacertae Objects

The first quasar to be identified was 3C48. As its name implies, this is a radio source, being the 48th source in the Third Cambridge Catalogue of radio sources. The closest optical object to the radio position was a 16^m object whose image was indistinguishable from that of a star, hence the name quasi-stellar (radio) source or

quasar. I have put the word *radio* in brackets because I want to emphasise that, although the first quasars to be identified were radio sources, we now know that only about 1% or so of quasars are strong radio sources. Radio-quiet quasars are often QSOs; less often (radio-loud) quasars are called QSRs. The terminology is not universal or unambiguous.

Although 3C48 looked star-like on the phtographic plates, its spectrum was very different from that of a star, being very blue and having unidentified broad emission lines and a non-thermal continuum, very similar to that of Seyfert 1s. The emission lines were first identified by Maarten Schmidt of Palomar Observatory. Working on another radio quasar, 3C 273, he showed that the pattern of some lines could be explained as hydrogen Balmer lines with what was, for the time, a huge redshift of 0.158. Interpreting the redshift as a Doppler effect gave a velocity of recession nearly 16% of the velocity of light. If 3C273 were a local object, its enormous velocity of recession would be hard to explain. Nowadays, quasars with redshift greater than 5 are known and it becomes almost impossible to maintain that they are local objects. The redshift of quasars is therefore generally accepted to be cosmological in origin. This view is not universally held. Halton Arp, for example, argues that some quasars at least are physically associated with galaxies that have much lower redshifts. If this is the case, then some "new physics" is required to explain the redshifts of these quasars (cf. for example [2]). The alternative explanation of gravitational redshift possibility is easily ruled out (cf. page 317 ff. of [3]).

The problem with the cosmological interpretation of the redshift of quasars is that it implies enormous luminosities of up to about 10^{40} W. Moreover, like Seyfert galaxies, the output of quasars can vary, on time-scales of as little as days or even hours. A source that can generate up to 10^{13} solar luminosities within a volume comparable to that of the solar system is truly awesome. Nevertheless, the consensus among astronomers is that quasars *are* at cosmological distances and that they are the active nuclei of galaxies[1]. There are many arguments in support of this view. First, the spectrum of a quasar can look so like the spectrum of a Seyfert galaxy that it is tempting to believe we are seeing extreme examples of Seyfert activity. To reverse the argument, if bright Seyferts were moved out to the distances of low-redshifts quasars, they would look like faint quasars. Secondly, the spectra of high-redshift quasars exhibit a "forest" of absorption lines with a range of redshifts lower than that of the quasar. These lines presumably arise in material, such as the halos of faint galaxies lying between the quasar and us. Lastly, modern techniques show the image – and the spectrum – of the host galaxy around many quasars.

[1] This view is not universally held. Halton Arp, for example, argues that some quasars at least are physically associated with galaxies that have much lower redshifts. If this is the case, then some "new physics" is required to explain the redshifts of these quasars (cf. for example []).

BL Lacertae objects (or BL Lacs) are named after the prototype of these sources, the radio source BL Lacertae. The brightness of these strange point-like sources can vary by a factor of several on the time-scale of days. In this respect, therefore, they are like an extreme form of Seyfert nuclei or quasar. Most BL Lacs, however, shows no emission lines in their spectra, which are of typical non-thermal form. Their distances are therefore difficult to estimate. Some BL Lacs, including BL Lac itself, have a faint evidence of material around the point source with the spectral characteristics of a galaxy. Its redshift is 0.07 corresponding to a distance of 200 h^{-1} Mpc.

2.4 Radio Galaxies

The optical luminosities of galaxies range from about 10^5 to 10^{10} L_{sun}. As might be expected from their ranges of size and mass, ellipticals span this whole range whilst spirals are confined roughly to the upper two decades. Since the solar luminosity is 3.9 10^{26} W, the total optical power radiated by galaxies ranges from 10^{32} to 10^{37} W.

Compared with these values, most galaxies radiate very little power in the radio region of the spectrum, luminous spirals emitting up to about 10^{33} W, for example. There are, however, strong radio sources that, in some cases, emit more radio power than the brightest optical galaxies. Some of these radio sources are quasars but not all quasars are radio sources.

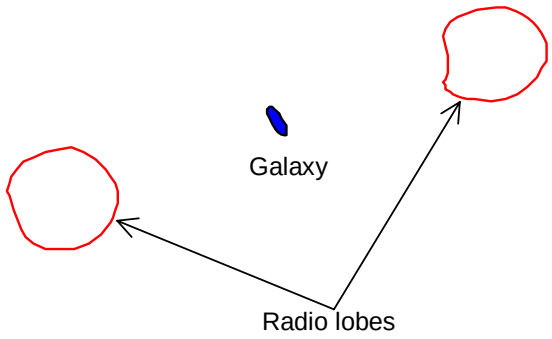

Figure 4-2. Cartoon of Cygnus A.

The strongest radio sources are typified by Cygnus A (3C405), one of the most luminous. It has a radio output of some 10^{38} W. Figure 4-2 is a schematic representation of a low-resolution map of the radio emission from Cygnus A. It shows the typical structure of a "classical" double radio source, with two lobes of radio material on either side of a peculiar galaxy and about 50 kpc away from it. The core of the galaxy itself is a radio emitter and I shall return to this shortly. Over a wide range of frequencies, the flux density $S(\nu)$ from the lobes has a power-law dependence on ν:

$$S(\nu) = S_o \times \left(\frac{\nu}{\nu_o}\right)^{\alpha} \qquad (2.3)$$

or

$$\log S(\nu) = \alpha \log \nu + \text{constant}, \qquad (2.4)$$

where S_o is the flux-density at the frequency ν_o and α is a constant called the spectral index[2]. Equation (2.4) shows that the spectrum is a straight line on a *log-log* plot of flux-density against frequency.

3. Models of Active Galactic Nuclei

3.1 The Source of Energy in AGN

3.1.1 Nuclear or Gravitational Power?

We have to explain luminosities of at least 10^{39} W (10^{13} L_{sun}), generated within regions only 10^{13} m or less across and lasting for some 10^8 years. In other words, we need to account for the generation of about 10^{54} J of energy. [A related, but smaller, problem is the explanation of the 10^{52} J or so found in the lobes of radio galaxies.] Relativity tells us that, in order to produce a luminosity L, we need to convert mass into energy at the rate $\dot{m}$ given by

$$\dot{m} \equiv \frac{dm}{dt} = -\frac{L}{c^2}; \qquad (3.1)$$

the minus sign occurs because the mass is decreasing with time. From equation (3.1) we find that we need to convert about a tenth of a solar mass a year to produce 10^{39} W. At first sight, this seems modest: a quasar would use up only 10^5 solar masses – perhaps 10^{-6} of the mass of a galaxy – in 10^8 years. What we have not taken into account, however, is the *efficiency* with which mass may be converted into energy. Let us define the efficiency η of an energy-generating process as the ratio of the energy E produced to the total rest-mass energy of the material generating the power:

$$\eta := \frac{E}{mc^2}. \qquad (3.2)$$

Taking this efficiency into account, equation (3.1) becomes

$$\dot{m} = -\frac{1}{\eta}\frac{L}{c^2}. \qquad (3.3)$$

Suppose we try to produce 10^{39} W of luminosity by as nuclear burning within stars[3]. The most efficient nuclear process is the conversion of hydrogen into helium, with $\eta = 0.7\%$. The rate of star-formation – that is the rate $\dot{m}$ at which interstellar material is converted into stars – is given by equation (3.3) as

$$-\dot{m} \approx \frac{10^{39}}{0.007 \times \left(3 \times 10^8\right)^2} \text{ kg s}^{-1} = 1.6 \times 10^{24} \text{ kg s}^{-1} \approx 25 \text{ M}_{sun} \text{ y}^{-1}. \qquad (3.4)$$

This is a very high rate: in 10^8 y, a galaxy would consume some 10^9 M_{sun}, or perhaps 0.1% of its mass. And, if we are to explain the variability of AGN, we need all this mass to be contained within a region not more than 10^{13} m across. The binding energy Ω of mass m contained within a region R in radius is given by

$$-\Omega \sim \frac{Gm^2}{R} \qquad (3.5)$$

and is about 10^{55} J for the above figures. This binding energy, which could in principle be released by gravitational contraction, is about an order of magnitude *more* than we are trying to produce by nuclear means! This suggests that the release of *gravitational* energy is the more likely source of power in AGN, although we need to look at the efficiency of this process as well.

3.1.2 Accretion Power

Let us see if we can find a more efficient process. Consider the decrease in gravitational potential energy $-\Delta\Omega$ of a body of mass m when it falls to within distance r of another body of mass M:

$$-\Delta\Omega = \frac{GMm}{r}. \qquad (3.6)$$

If the body is in free-fall, however, this decrease in potential energy will simply go into increasing the kinetic energy of the body. We have to find a way of converting the potential energy into radiation. The usual model of this process is to consider the "body" to be gaseous material in an *accretion disc* and which is slowly spiralling into the central mass[4]. In order to spiral, the material must be slowly losing angular momentum and this is envisaged as arising from viscosity in the disc – mutual friction between the material successive turns of the spiral[5]. This viscosity heats the disc, converting the kinetic energy of the orbiting material into internal energy of the molecules and atoms of the disc. Finally, this heat is radiated away by photons. Let us look at this process more quantitatively.

---

[2] Note that some authors define the spectral index by:

$$S(\nu) = S_o \times \left(\frac{\nu}{\nu_o}\right)^{-\alpha}.$$

This is because most of the early observations were of sources whose spectral index was, by this definition, positive. The slope of log $S(\nu)$ against log ν can have either sign, however, and it is becoming more usual – but not universal – to use the form I have chosen.

[3] Even if we were to do this, we should still have to explain how to convert ordinary stellar radiation into relativistic particles needed to explain the observed spectra – see later.

[4] This is almost certain to happen anyway because the in-falling material will have angular momentum with respect to the central object that it will have to get rid of.

[5] It has to be said that no satisfactory viscous mechanism has been found!

Since the material is in quasi-stationary orbit about the central mass, the kinetic energy $T(r)$ and potential energy $\Omega(r)$ of material at distance r from the centre must obey the virial theorem:

$$2T(r)+\Omega(r)=0\,. \tag{3.7}$$

The total energy of the system – that is the sum of the kinetic energy, the potential energy and any energy $R(r)$ that has been radiated by the time the material has reached r – must be conserved, so that

$$T(r)+\Omega(r)+R(r)=0\,, \tag{3.8}$$

where I have taken the potential energy of the material at infinite distance from the central mass to be zero. Eliminating T from equations (3.7) and (3.8), we get

$$\begin{aligned} R(r)&=-\frac{1}{2}\Omega(r)=\frac{1}{2}\frac{GMm}{r};\\ T(r)&=-\frac{1}{2}\Omega(r)=\frac{1}{2}\frac{GMm}{r}. \end{aligned} \tag{3.9}$$

Only half the potential energy is, therefore, available to be converted into radiation.

It appears at first sight as if we could extract infinite energy from this process by allowing r to go to zero. There is, however, a natural limit to how close one can approach to an object of mass M. General relativity tells us that, if a body becomes smaller that its *Schwarzschild radius* r_S, given by

$$r_S := \frac{2GM}{c^2}\,, \tag{3.10}$$

it will collapse to form a black hole from which nothing can escape. Putting in numbers, we get

$$r_S(\text{light hours})=2.76\times10^{-9}\left(\frac{M}{M_{\text{sun}}}\right). \tag{3.11}$$

Hence the Schwarzschild radius represents an absolute minimum to the value of r occurring in equation (3.9) and then only if the body of mass M is, indeed, a black hole. Because we get *maximum* efficiency from a black hole, let us shall assume that this is the case. General relativity also says that there is no stable orbit closer to a black hole than $3r_S$. Material closer than this plunges straight into the hole without having time to get rid of its remaining gravitational energy[6]. In practice, therefore, the maximum radiant energy $R_{\max}$ that I can extract from letting matter of mass m fall into a black hole is given by

$$R_{\max}=\frac{1}{2}\frac{GMm}{3r_s}=\frac{1}{12}mc^2 \tag{3.12}$$

with a corresponding maximum luminosity L given by

$$L_{\max}\equiv\frac{dE}{dt}=\frac{1}{12}\dot{m}c^2\,. \tag{3.13}$$

The minimum mass-inflow rate $\dot{m}$ needed to produce a luminosity L is therefore given by

$$\dot{m}=12\frac{L}{c^2} \tag{3.14}$$

Comparing equation (3.14) with equation (3.3), we see that the efficiency of gravitational accretion is 1/12 or about 8%, which is an order of magnitude more efficient than nuclear processes. This means that only about 2 solar masses a year are needed to fuel the observed luminosities. Actually, we can expect the efficiency to be even higher than this. I have dealt only with a non-rotating black hole. In practice, any black hole formed from the collapse of material to the centre of a galaxy is likely to be rotating because the material from which it formed would have had angular momentum about the centre of the galaxy. The maximum efficiency of extracting potential energy from a rotating black hole can be shown to be about 40%.

Note that the luminosity predicted by this model depends only on the mass-infall rate and is independent of the mass of the black hole itself. The observed variability, however, limits the size of the region and hence the Schwarzschild radius. Variability on the time scale of an hour demands a black hole of at least $10^9\,M_{\text{sun}}$.

3.1.3 The Eddington Limit

It is easy to show that there is a limiting "accretion-luminosity" for a body of a given mass. Out-flowing photons exert a force on the in-flowing matter and, if the flux of photons is large enough, this force will exceed the gravitational attraction of the central mass. The situation is illustrated in Figure 4-3.

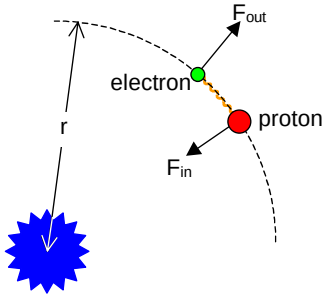

Figure 4-3. Origin of the Eddington Limit

The details are as follows. The general relation between the energy E and momentum $\tilde{p}$ of a relativistic particle of mass m is given by

$$E^2=\tilde{p}^2c^2+m^2c^4\,. \tag{3.15}$$

[6]This is a slight over-simplification but it serves our purpose.

Since photons are massless, each carries momentum $\tilde{p}$ given by

$$\tilde{p} = E/c \tag{3.16}$$

and the total momentum $\tilde{P}$ carried per unit time by all the photons emitted by a source of luminosity L is given by

$$\tilde{P} = \frac{L}{c}. \tag{3.17}$$

The pressure p_γ exerted by these photons is, by definition, the momentum flux, that is the rate flow of momentum per unit area of surface perpendicular to the flow of radiation. At a distance r from the source, therefore,

$$p_\gamma(r) = \left(\frac{L}{4\pi r^2}\right)/c. \tag{3.18}$$

This pressure acts upon the Thomson cross-sections σ of the (ionised) in-falling particles, mainly protons and electrons. For a particle of charge e and mass m,

$$\sigma(e,m) = \frac{2}{3}\left(\frac{e^2}{4\pi\varepsilon_o mc^2}\right)^2 \tag{3.19}$$

It is clear from equation (3.19) that, because of the ratio of their masses, the Thomson cross-section of electrons is some six orders of magnitude greater than that of protons. The main outward force $F_{out}(r)$ is therefore that exerted by the photons on the electrons and is given by

$$F_{out}(r) = p_\gamma(r)\times\sigma_T = \frac{L\sigma_T}{4\pi r^2 c}, \tag{3.20}$$

where σ_T is the Thomson cross-section of the electron. Because the electrons in the plasma are coupled to the protons *via* electromagnetic forces, the outward force given by equation (3.20) is transferred to the protons as well.

The inward gravitational force $F_{in}(r)$ on a particle of mass m is given as usual by

$$F_{in}(r) = \frac{GMm}{r^2}. \tag{3.21}$$

Equation (3.21) shows that the gravitational force on the proton is three orders of magnitude higher than that on the electron so that the main inward force is given by

$$F_{in}(r) = \frac{GMm_P}{r^2}, \tag{3.22}$$

where m_p is the proton's mass. Again, though, this force is transferred to the electrons electromagnetically.

For accretion to take place, we need inward force given by equation (3.22) to be greater than the outward force given by equation (3.20):

$$\frac{GMm_P}{r^2} \geq \frac{L\sigma_T}{4\pi r^2 c}. \tag{3.23}$$

This means that, if a source is to derive its luminosity from accretion, its luminosity L must be less than the *Eddington luminosity*, $L_{Eddington}$:

$$L \leq L_{Eddington} := 4\pi\frac{GMm_p c}{\sigma_T}. \tag{3.24}$$

Putting in numbers, we get

$$L_{Eddington}(\mathrm{W}) = 1.26\times10^{31}\left(\frac{M}{M_{sun}}\right). \tag{3.25}$$

Relation (3.24) can be re-arranged to give a lower limit on the mass of a source with a given observed velocity L:

$$M \geq M_{min} = \frac{1}{4\pi}\frac{L\sigma_T}{Gm_p c}. \tag{3.26}$$

Again putting in numbers, we get

$$M_{min}(\mathrm{M_{sun}}) = 7.9\times10^{-32}L(\mathrm{W}). \tag{3.27}$$

Equation (3.27) shows that we need a black hole of at least 10^8 solar masses to give an accretion luminosity of 10^{39} W.

3.2 Characteristic Temperatures

Let us make some rough estimates of the temperatures involved in accretion radiation. First, consider that coming from the inner edge of the accretion disc at radius r_{min}. If we assume that the radiation is blackbody with effective temperature T_{eff}, then the luminosity L is given by

$$L \sim r_{min}^2\sigma T_{eff}^4, \tag{3.28}$$

where σ is the Stefan-Boltzmann constant. If we assume that the disc is radiating at the Eddington limit and that r_{min} is the radius of the last stable orbit, then we have from equations (3.28), (3.24) and (3.10),

$$T_{eff} \sim \left[\frac{\pi}{9}\frac{m_p c^5}{GM\sigma\sigma_T}\right]^{1/4}. \tag{3.29}$$

Putting in numbers, we have

$$T_{eff} \sim 4\times10^7\left(\frac{M}{M_{sun}}\right)^{-1/4}. \tag{3.30}$$

For a 10^8 solar mass black hole, we find an effective temperature of about 10^5 K. Using the relation

$$\frac{hc}{\lambda} \sim kT\,, \tag{3.31}$$

we see that this radiation peaks at around 150 nm, or in the UV to soft X-ray region of the spectrum where the big blue bump lies. Moreover, from equation (3.10), we see that the size of this region is about a light-hour across, so we should not be surprised at variations on the time-scale of hours.

On the other hand, think of protons that fall straight from "infinity" to $r_{\min}$ and are then thermalised. The gain in ΔT in kinetic energy of each proton is given, from equation (3.9), by

$$\Delta T = \frac{GMm_{\rm p}}{3r_{\rm S}} = \frac{1}{6}m_{\rm p}c^2\,. \tag{3.32}$$

If this energy is now thermalised to a temperature $T_{\rm therm}$, we have

$$\Delta T = \frac{3}{2}kT_{\rm therm} \tag{3.33}$$

so that

$$T_{\rm therm} = \frac{1}{9}\frac{m_{\rm p}c^2}{k} \tag{3.34}$$

or, numerically

$$T_{\rm therm} = 1.2\times10^{12}\ {\rm K}\,.^{7} \tag{3.35}$$

From equation (3.31), we find that the energy of the photons corresponding to this temperature is about 100 MeV, many times the energy (1 MeV) required for the production of electron-positron pairs. From this, we may tentatively conclude that:

- the big blue bump arises from optically thick thermal radiation from the inner parts of the accretion disc;
- the X-rays originate from a region of electron-positron plasma.

3.3 Accretion Discs

3.3.1 Structure of the Disc

Figure 4-4 shows schematically an accretion disc of density ρ(r) encircling a black hole and though which mass is spiralling down at a rate $\dot{m}$ [8]. The upper part of the figure shows the plan view and the lower part a section. Consider a thin annulus of the disc of width δr at r. Define the *surface density* $\sigma(r)$ of the disc at r by

$$\sigma(r) := \int_{-h(r)/2}^{+h(r)/2} \rho(r')dr'\,, \tag{3.36}$$

where $h(r)$ is the thickness of the disc at r.

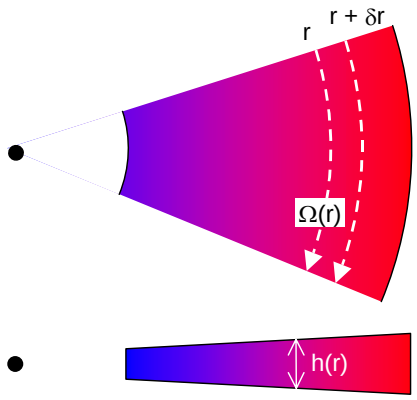

Figure 4-4. Accretion disc.

3.3.2 Conservation of Mass

Consider an annulus of width δr at r. The rate of build-up of mass $\partial m(r)/\partial t$ within this annulus is given by

$$\frac{\partial m(r)}{\partial t} \equiv \frac{\partial}{\partial t}\left[2\pi r\delta r\times\sigma(r)\right]. \tag{3.37}$$

This build-up must be provided by the net flow of material $\dot{m}_{\rm net}$ into the annulus:

$$\frac{\partial m(r)}{\partial t} = \dot{m}_{\rm net} = \dot{m}_{\rm in} - \dot{m}_{\rm out}\,, \tag{3.38}$$

where $\dot{m}_{\rm in}$ is the rate of flow of mass into the annulus from the outer regions and $\dot{m}_{\rm out}$ is rate of flow of mass from the annulus into the inner regions. Now

$$\dot{m}_{\rm out} \equiv \dot{m}(r) = -2\pi\, r\times\sigma(r)\times v_r(r)\,. \tag{3.39}$$

and

$$\begin{aligned}\dot{m}_{\rm in} &\equiv \dot{m}(r+\delta r)\\ &= -2\pi(r+\delta r)\sigma(r+\delta r)v_r(r+\delta r),\end{aligned} \tag{3.40}$$

where $v_r(r)$ is the radial component of velocity of the material in the disc (positive in the direction of increasing r). Hence,

$$\dot{m}_{\rm net} = -2\pi\delta r\times\frac{\partial}{\partial r}\left[r\sigma(r)v_r(r)\right] \tag{3.41}$$

to first order in δr. From equations (3.37), (3.38) and (3.41), we have

[7] Note that these results are independent of the mass of the black hole.

[8] If the disc is to be in a steady state, that is of there is to be no build up of matter anywhere in it, the mass infall rate must be independent of radius.

$$2\pi r\delta r\frac{\partial}{\partial t}\left[\sigma(r)\right]=-2\pi\delta r\times\frac{\partial}{\partial r}\left[r\sigma(r)v_r(r)\right] \qquad (3.42)$$

or

$$\frac{\partial\sigma(r)}{\partial t}+\frac{1}{r}\frac{\partial}{\partial r}\left[r\sigma(r)v_r(r)\right]=0\,. \qquad (3.43)$$

In the steady state, $\partial\sigma(r)/\partial t$ must be zero so that

$$r\sigma(r)v_r(r)=\text{constant}\,. \qquad (3.44)$$

Comparing equations (3.44) and (3.39), we see that

$$\dot{m}(r)=-2\pi\, r\times\sigma(r)\times v_r(r)=\text{constant}=\dot{m}\,, \qquad (3.45)$$

independent of r. That is, the radial mass flow is independent of radius, as we expected.

3.3.3 Conservation of Angular Momentum

Now consider the transfer of angular momentum across the annulus. The rate $\dot{L}_{\text{mass flow}}(r)$ at which clockwise angular momentum is flowing out to the inner regions of the disc is given by

$$\dot{L}_{\text{mass flow}}(r)=\dot{m}(r)\times r\times v_\phi(r), \qquad (3.46)$$

where $v_\phi(r)$ is the azimuthal velocity of the material at r. taken to be positive clockwise. Using similar arguments to those above, we find that the rate of build-up $\partial L(r)/\partial t$ of angular momentum in the annulus, resulting from the mass flow, is given by

$$\frac{\partial L_{\text{mass flow}}(r)}{\partial t}=\dot{m}\frac{\partial}{\partial r}\left[rv_\phi(r)\right]\delta r\,, \qquad (3.47)$$

where I have used the constancy of mass-flow, equation (3.45)

There is, however, another source of angular momentum – the action of any torque $\Gamma(r)$ acting in the disc: I shall derive an expression for $\Gamma(r)$ later. If we take positive $\Gamma(r)$ to be clockwise acting on material interior to r, then the net clockwise torque $\delta\Gamma(r)$ acting on the annulus is given by

$$\begin{aligned}\delta\Gamma(r)&=\Gamma(r+\delta r)-\Gamma(r)\\&\approx\frac{\partial\Gamma(r)}{\partial r}\delta r\end{aligned} \qquad (3.48)$$

so that, from equations (3.47) and (3.48)

$$\begin{aligned}\frac{\partial L(r)}{\partial t}&\equiv\frac{\partial L_{\text{mass flow}}(r)}{\partial t}+\frac{\partial\Gamma(r)}{\partial r}\\&=\dot{m}\frac{\partial}{\partial r}\left[rv_\phi(r)\right]+\frac{\partial\Gamma(r)}{\partial r}\end{aligned}\,. \qquad (3.49)$$

In the steady state, $\partial L(r)/\partial t$ must be zero so

$$\dot{m}\frac{\partial}{\partial r}\left[rv_\phi(r)\right]=-\frac{\partial\Gamma(r)}{\partial r}\,. \qquad (3.50)$$

If we integrate equation (3.50) between r_{lso}, the radius of the last stable orbit, and r, we get

$$\dot{m}\int_{r_{\text{lso}}}^{r}\frac{\partial}{\partial r'}\left[r'v_\phi(r')\right]dr'=-\int_{r_{\text{lso}}}^{r}\frac{\partial\Gamma(r')}{\partial r'}dr'\,, \qquad (3.51)$$

or

$$\begin{aligned}\dot{m}\left[rv_\phi(r)-r_{\text{lso}}v_\phi(r_{\text{lso}})\right]&=-\left[\Gamma(r)-\Gamma(r_{\text{lso}})\right]\\&=-\Gamma(r),\end{aligned} \qquad (3.52)$$

where I have assumed that the torque at the last stable orbit is zero, because there is nothing for it to act on.

3.3.4 Conservation of Energy

Finally, let us consider the conservation of energy. We see that the rate $\dot{T}(r)$ at which kinetic energy is flowing out of the annulus into the inner regions of the disc is given by

$$\dot{T}(r)=\frac{1}{2}\dot{m}\times v_\phi^2(r)\,. \qquad (3.53)$$

Proceeding along similar lines to the above, we find that the net rate of flow $\partial T(r)/\partial t$ of kinetic energy into the annulus is given by

$$\frac{\partial T(r)}{\partial t}=\frac{1}{2}\dot{m}\frac{\partial}{\partial r}\left[v_\phi^2(r)\right]\delta r\,. \qquad (3.54)$$

We must also take account of the flow of potential energy V. The rate $\dot{V}(r)$ at which potential energy is flowing out of the annulus to inner regions is given by

$$\begin{aligned}\dot{V}(r)&=-2\pi\, r\sigma(r)v_r(r)\times\left(-\frac{GM}{r}\right)\\&=-\frac{GM\dot{m}}{r},\end{aligned} \qquad (3.55)$$

where I have used equation (3.45) The net rate of increase $\partial V(r)/\partial t$ in the annulus is therefore given by

$$\frac{\partial V(r)}{\partial t}=-GM\dot{m}\frac{\partial}{\partial r}\left(\frac{1}{r}\right)\delta r\,. \qquad (3.56)$$

Finally, the differential torque does work on the annulus. This net rate of working $\partial W(r)/\partial t$ is given by

$$\frac{\partial W(r)}{\partial t} = \Gamma(r+\delta r)\Omega(r+\delta r) - \Gamma(r)\Omega(r)$$
$$\approx \frac{\partial}{\partial r}\left[\Gamma(r)\Omega(r)\right]\delta r = \frac{\partial}{\partial r}\left[\Gamma(r)\frac{v_\phi(r)}{r}\right]\delta r \quad (3.57)$$
$$= -\dot{m}\frac{\partial}{\partial r}\left\{\left[1-\frac{r_{\rm lso}}{r}\frac{v_\phi(r_{\rm lso})}{v_\phi(r)}\right]v_\phi^2(r)\right\}\delta r,$$

where $\Omega(r)$ is the angular velocity of the material at r and where I have used the momentum conservation equation (3.52) to substitute for $\Gamma(r)$. Hence, the overall rate $\partial E(r)/\partial t$ of build-up of total energy in the annulus is given by

$$\frac{\partial E(r)}{\partial t} = \frac{\partial T(r)}{\partial t} + \frac{\partial V(r)}{\partial t} + \frac{\partial W(r)}{\partial t}$$
$$= \dot{m}\frac{\partial}{\partial r}\left\{\begin{array}{l}\frac{1}{2}v_\phi^2(r) - \frac{GM}{r} \\ -\left[1-\frac{r_{\rm lso}}{r}\frac{v_\phi(r_{\rm lso})}{v_\phi(r)}\right]v_\phi^2(r)\end{array}\right\}\delta r \quad . \quad (3.58)$$
$$= -\dot{m}\frac{\partial}{\partial r}\left\{\begin{array}{l}\left[\frac{1}{2}-\frac{r_{\rm lso}}{r}\frac{v_\phi(r_{\rm lso})}{v_\phi(r)}\right]v_\phi^2(r) \\ +\frac{GM}{r}\end{array}\right\}\delta r$$

If the material is in almost circular Keplerian orbits about the black hole ($v_r << v_\phi$) , we have

$$\frac{v_\phi^2}{r} = \frac{GM}{r^2} \quad . \quad (3.59)$$

Substituting from equation (3.59) into equation (3.58), we get

$$\frac{\partial E(r)}{\partial t} = -\dot{m}\frac{\partial}{\partial r}\left\{\left[\frac{3}{2}-\left(\frac{r_{\rm lso}}{r}\right)^{1/2}\right]v_\phi^2(r)\right\}\delta r \quad . \quad (3.60)$$

Let the surface energy density of the disc be $U(r)$. Then the total surface energy $E(r)$ of the annulus is given by

$$E(r) = 2\pi r\,\delta r \times U(r), \quad (3.61)$$

From equations (3.60) and (3.61), we get

$$\frac{\partial U(r)}{\partial t} = \frac{1}{2\pi r\,\delta r}\frac{\partial E(r)}{\partial t}$$
$$= -\frac{GM\dot{m}}{2\pi r}\frac{\partial}{\partial r}\left\{\frac{1}{r}\left[\frac{3}{2}-\left(\frac{r_{\rm lso}}{r}\right)^{1/2}\right]\right\} \quad (3.62)$$
$$= \frac{3}{4\pi}\frac{GM\dot{m}}{r^3}\left[1-\left(\frac{r_{\rm lso}}{r}\right)^{1/2}\right].$$

If the disc is in thermal equilibrium, then this energy must be radiated away. The luminosity per unit area $l(r)$ of the disc must therefore be given by

$$l(r) = \frac{1}{2}\frac{\partial U(r)}{\partial t} = \frac{3}{8\pi}\frac{GM\dot{m}}{r^3}\left[1-\left(\frac{r_{\rm lso}}{r}\right)^{1/2}\right]. \quad (3.63)$$

where the extra factor of a half appears because energy is radiated by the disc has two sides[9]. Note that the luminosity tends to zero as we approach the last stable orbit.

3.3.5 Viscosity in the Disc

It is interesting to note that the above result is entirely independent of the *mechanism* responsible for the torque $\Gamma(r)$ acting in the disc. Such a torque must exist in order to conserve angular momentum, as shown by equation (3.52) but its nature is unimportant as far as the luminosity of the disc is concerned. Let us, though, investigate a possible mechanism.

Suppose the material in the disc has some chaotic motion with mean-squared velocity <v> and mean-free-path λ, superimposed on its circular and radial motions. Consider a blob of material A in Figure 4-5, moving from the initial position shown to a new position at smaller radius. Then the mass-transfer rate $\dot{m}_{\rm chaos}$ produced by such movement is given by

$$\dot{m}_{\rm chaos} = 2\pi r \times \sigma(r) \times \langle v \rangle \quad (3.64)$$

On average, such a blob will transfer angular momentum corresponding to material at $r + \lambda$ across the surface of at radius r. Similarly, blob B will transfer angular momentum corresponding to material at $r - \lambda$ across the same surface. This net transfer of angular momentum will exert a torque on the disc at r.

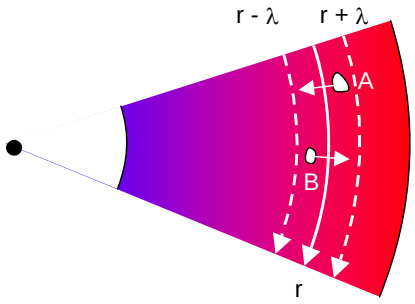

Figure 4-5. Origin of Viscosity.

Material at r has circular velocity $r\Omega(r)$. If the disc rotated as a rigid body, so that $\Omega(r)$ were constant, then the material at $(r + \lambda/2)$ would have circular velocity $(r + \lambda/2)\Omega(r)$. In fact, the material at $(r + \lambda/2)$, where A originates, has circular velocity $(r + \lambda/2)\Omega(r + \lambda/2)$. The

[9] I am assuming that the disc is optically thick.

relative velocity of material at $(r + \lambda/2)$ with respect to that at r is therefore given by

$$\begin{aligned}&\left(r+\frac{\lambda}{2}\right)\Omega\left(r+\frac{\lambda}{2}\right)-\left(r+\frac{\lambda}{2}\right)\Omega(r)\\&=\left(r+\frac{\lambda}{2}\right)\left[\Omega\left(r+\frac{\lambda}{2}\right)-\Omega(r)\right]\\&\approx\left(r+\frac{\lambda}{2}\right)\left[\Omega(r)+\frac{\lambda}{2}\Omega'(r)-\Omega(r)\right]\\&\cong\frac{1}{2}\lambda\, r\,\Omega'(r),\end{aligned}\tag{3.65}$$

where I have used a Taylor expansion and kept only first-order terms in λ. The rate of transfer of angular momentum $\dot{L}_{\rm in}(r)$ across r by material such as A is therefore given by

$$\dot{L}_{\rm in}(r)=\dot{m}_{\rm chaos}\times\left[r\times\frac{1}{2}\lambda\, r\,\Omega'(r)\right].\tag{3.66}$$

Similarly, the rate of transfer of angular momentum $\dot{L}_{\rm out}(r)$ across r by material such as B is given by

$$\dot{L}_{\rm out}(r)=\dot{m}_{\rm chaos}\times\left[r\times\frac{1}{2}(-\lambda)r\,\Omega'(r)\right].\tag{3.67}$$

The net rate of transfer of angular momentum inward $\dot{L}(r)$ at r is therefore given by

$$\begin{aligned}\dot{L}(r)&=\dot{L}_{\rm in}(r)-\dot{L}_{\rm out}(r)\\&=\lambda r^2\,\Omega'(r)\dot{m}_{\rm chaos}\\&=2\pi r^3\sigma(r)\lambda\langle v\rangle\Omega'(r),\end{aligned}\tag{3.68}$$

where I have used equation (3.64). But the net rate of transfer of angular momentum across a surface is equal to the torque exerted on that surface. The torque $\Gamma(r)$ exerted, in the direction of rotation, by the outer material on the inner material at r is therefore given by

$$\Gamma(r)=2\pi r^3\sigma(r)\mu\langle v\rangle\Omega'(r),\tag{3.69}$$

where

$$\mu(r):=\lambda\langle v\rangle\tag{3.70}$$

is the coefficient of viscosity. Note that, if $\Omega(r)$ is actually constant, independent of r, then $\Gamma(r)$ is zero: we only get a torque of there is shear motion in the disc.

3.3.6 Spectrum Radiated by Disc

Let us assume that the disc radiates as a black body so that

$$I(r)=\sigma T^4(r),\tag{3.71}$$

where $T(r)$ is the temperature of the disc at r and σ is the Stefan-Boltzmann constant. From equations (3.71) and (3.63), we have

$$\begin{aligned}T(r)&=\left(\frac{3}{8\pi}\frac{GM\dot{m}}{r^3\sigma}\right)^{1/4}\left[1-\left(\frac{r_{\rm lso}}{r}\right)^{1/2}\right]^{1/4}\\&=\left(\frac{3}{8\pi}\frac{GM\dot{m}}{r^3\sigma}\right)^{1/4}\left[1-\left(\frac{6GM}{c^2 r}\right)^{1/2}\right]^{1/4}.\end{aligned}\tag{3.72}$$

The spectrum $I(\nu,T)$ of black-body radiation at temperature T is given by

$$I(\nu,T)=\frac{2h\nu^3}{c^2}\frac{1}{e^{h\nu/kT}-1}\tag{3.73}$$

so that the spectral luminosity $l(\nu,r)$ per unit area of the accretion disc is given by[10]

$$l(\nu,r)=\frac{2\pi h\nu^3}{c^2}\frac{1}{e^{h\nu/kT(r)}-1},\tag{3.74}$$

where $T(r)$ is given by equation (3.72). The total spectral luminosity $L(\nu)$ of the disc is therefore given by

$$\begin{aligned}L(\nu)&=2\times\int_{r_{\rm lso}}^{r_{\rm max}}l(\nu,r)\times 2\pi\, r dr\\&4\pi\int_{r_{\rm lso}}^{r_{\rm max}}\left[\frac{2\pi h\nu^3}{c^2}\frac{1}{e^{h\nu/kT(r)}-1}\right]r dr,\\&=\frac{8\pi^2 h\nu^3}{c^2}\int_{r_{\rm lso}}^{r_{\rm max}}\frac{r dr}{e^{h\nu/kT(r)}-1}\end{aligned}\tag{3.75}$$

where $r_{\rm max}$ is the outer radius of the disc and the first factor of two recognises that the disc has two faces.

Although the integration of equation (4.26) has to be done numerically, we can get a good idea of what this spectrum looks like by making some approximations. At low frequencies, we have

$$\frac{1}{e^{h\nu/kT}-1}\approx\frac{kT}{h\nu};\qquad\frac{h\nu}{kT}<<1.\tag{3.76}$$

This is the *Rayleigh-Jeans* approximation. The lowest temperature in the disc occurs at $r_{\rm max}$. For frequencies such that

$$h\nu<<kT(r_{\rm max}),\tag{3.77}$$

therefore, we have

[10] Note that a black surface radiates into π steradians.

$$
\begin{aligned}
L(\nu) &\approx \frac{8\pi^2\nu^2}{c^2} k \int_{r_{\rm lso}}^{r_{\rm max}} T(r) r dr \\
&= \frac{8\pi^2\nu^2}{c^2} k \left(\frac{GM\dot{m}}{8\pi\sigma}\right)^{1/4} \\
&\times \int_{r_{\rm lso}}^{r_{\rm max}} r^{1/4} \left[1-\left(\frac{r_{\rm lso}}{r}\right)^{1/2}\right]^{1/4} dr \\
&\approx \frac{32\pi^2}{5}\left[\frac{GM\dot{m}k^4}{8\pi\sigma c^8} J(r_{\rm lso}, r_{\rm max})\right]^{1/4} \nu^2 \\
&\propto \nu^2,
\end{aligned}
\tag{3.78}
$$

where, $J(r_{\rm lso},r_{\rm max})$ is the integral in the third line of (3.78). At low frequencies, therefore, the spectrum is a power law.

At high frequencies, we have

$$
\frac{1}{e^{h\nu/kT}-1} \approx e^{-h\nu/kT}; \qquad \frac{h\nu}{kT} >> 1. \tag{3.79}
$$

Differentiation of equation (3.72) shows that the maximum temperature occurs at $r_{T\,\rm max}$ given by

$$
r_{T\,\rm max} = \left(\frac{4}{3}\right)^2 r_{\rm lso} = 1.8\, r_{\rm lso}. \tag{3.80}
$$

For frequencies such that

$$
h\nu >> kT(r_{T\,\rm max}), \tag{3.81}
$$

we have

$$
\begin{aligned}
L(\nu) &\approx \frac{8\pi^2 h\nu^3}{c^2} \int_{r_{\rm lso}}^{r_{\rm max}} e^{-h\nu/kT(r)} r dr \\
&\approx \frac{8\pi^2 h\nu^3}{c^2} e^{-h\nu/kT(r_{\rm max})} \int_{r_{\rm lso}}^{r_{\rm max}} r dr, \\
&\approx \frac{4\pi^2 h\nu^3}{c^2} r_{\rm max}^2 e^{-h\nu/kT(r_{\rm max})},
\end{aligned}
\tag{3.82}
$$

where, in the second approximation, I have replaced the exponential term with its maximum value and, in the third, have assumed that $r_{\rm max} >> r_{\rm lso}$. At high frequencies, the spectrum falls off exponentially with frequency.

What about in between? The blackbody function (3.73) peaks at frequency ν given by

$$
\nu \sim \frac{kT}{h} \tag{3.83}
$$

and the total luminosity per unit area is σT^4 equation (3.71). We can approximate $I(\nu,T)$ very crudely as

$$
I(\nu, r) \approx \sigma T(r)^4 \delta\left[\nu - \frac{kT(r)}{h}\right], \tag{3.84}
$$

$\delta(x)$ is the Dirac delta function. Substituting from equation (3.84) into the first line of equation (3.75), we get

$$
\begin{aligned}
L(\nu) &= 4\pi \int_{r_{\rm lso}}^{r_{\rm max}} \left\{\sigma T^4(r) \delta\left[\nu - \frac{kT(r)}{h}\right]\right\} r dr \\
&= 4\pi\sigma\left(\frac{GM\dot{m}}{8\pi\sigma}\right) \int_{r_{\rm lso}}^{r_{\rm max}} \left[T(r)^4 \delta\left(\nu - \frac{kT}{h}\right)\right] r dr
\end{aligned}
\tag{3.85}
$$

The integral could be evaluated by changing the variable of integration from r to T using equation (3.72). Unfortunately, the term

$$
\left[1-\left(\frac{r_{\rm lso}}{r}\right)^{1/2}\right]^{1/4}
$$

complicates this considerable. Since we are only looking for an approximate behaviour, I shall ignore this term. Then

$$
\begin{aligned}
L(\nu) &= 4\pi\sigma\left(\frac{GM\dot{m}}{8\pi\sigma}\right)^{5/3} \int_{T(r_{\rm max})}^{T(r_{\rm lso})} \left[T^{1/3} \delta\left(\nu - \frac{kT}{h}\right)\right] dT \\
&= \frac{16\pi}{3}\sigma\left(\frac{GM\dot{m}}{8\pi\sigma}\right)^{5/3}\left(\frac{h}{k}\right)^{1/3} \nu^{1/3} \\
&\propto \nu^{1/3}
\end{aligned}
\tag{3.86}
$$

so that $L(\nu)$ is proportional to $\nu^{1/3}$ in the intermediate region. The overall spectrum is sketched in Figure 4.6.

How does this fit with observation? It is not a bad representation of the "blue bump" in AGN spectra although ν^0 would be better than $\nu^{1/3}$ Nevertheless, especially in view of the approximations involved, it is not a bad attempt.

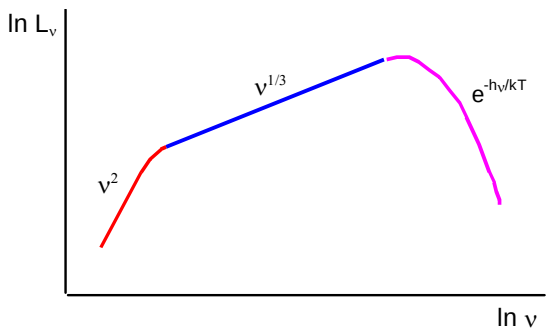

Figure 4.6. Overall spectrum of accretion disc.

4. Models of Radio Sources

4.1 Synchrotron Radiation

4.1.1 The Frequency of Synchrotron Radiation

A power-law spectrum of the form given in equation (2.3) is consistent with the *synchrotron radiation* emitted by a collection of charged particles, of various energies, moving with relativistic velocities in a magnetic field. Synchrotron emission is also consistent with the observed polarisation of the radiation. I shall give a simplified treatment of such radiation.

An electron moving perpendicularly to a magnetic field of induction B describes a circular orbit with the *cyclotron frequency* ν_c given by

$$\nu_c = \frac{1}{2\pi}\frac{eB}{m_e}, \tag{4.1}$$

where m_e is the rest-mass of the electron. Notice that ν_c is independent of the velocity (or energy) of the electron. Because circular motion involves acceleration, the electron will emit electromagnetic radiation with frequency ν_c. If the electron is moving relativistically with velocity u however, it emits radiation over a broad band of frequencies peaked at ν given by

$$\nu = \frac{\gamma^2}{2}\nu_c = \frac{1}{4\pi}\frac{eB}{m_e}\gamma^2, \tag{4.2}$$

where the Lorentz factor γ is given by

$$\gamma = \frac{1}{\sqrt{1-u^2/c^2}} = \frac{1}{\sqrt{1-\beta^2}}; \qquad \beta := \frac{u}{c}. \tag{4.3}$$

Using the fact that the energy E of the relativistic electron is given by

$$E = \gamma m_e c^2, \tag{4.4}$$

we get

$$\nu = \frac{1}{4\pi}\left(\frac{e}{m_e^3 c^4}\right)BE^2 = aBE^2, \tag{4.5}$$

where

$$a := \frac{1}{4\pi}\left(\frac{e}{m_e^3 c^4}\right). \tag{4.6}$$

Putting in numerical values, we get

$$\nu(\mathrm{Hz}) = 2.1\times 10^{36}\, B(\mathrm{T})E^2(\mathrm{J}). \tag{4.7}$$

4.1.2 The Power Emitted by a Relativistic Electron

The power P(E) radiated by an electron of energy E is given by

$$P(E) = \frac{2}{3}\frac{1}{4\pi\varepsilon_o}\left(\frac{e^4B^2}{m_e^2 c}\right)(\beta\gamma)^2 = \frac{1}{6\pi\varepsilon_o}\left(\frac{e^4}{m_e^4 c^5}\right)\times\beta^2B^2E^2 \tag{4.8}$$

where ε_o is the permittivity of free space and where the second line come from equation (3.39).

In any collection if electrons in space, there must be an equal number of protons to preserve charge neutrality, and these protons will also radiate in a magnetic field. Equation (4.8) shows, however, that the power radiated by a particle in inversely proportional to the fourth power of its mass. Even if the protons are moving relativistically with the same energy as the electrons, therefore, the power they radiate will be relatively negligible because they are some two thousand times more massive[11].

In all the cases we shall deal with, the electrons are ultra-relativistic so that

$$\beta \approx 1. \tag{4.9}$$

We can therefore re-write equation (3.63) as

$$P(E) \approx bB^2E^2, \tag{4.10}$$

where

$$b = \frac{1}{6\pi\varepsilon_o}\left(\frac{e^4}{m_e^4 c^5}\right). \tag{4.11}$$

Putting in numerical values, we get

$$P(\mathrm{W}) = 2.4\times 10^{12}\, B^2(\mathrm{T})E^2(\mathrm{J}). \tag{4.12}$$

4.1.3 The Synchrotron Spectrum

Suppose that the number $N(E)dE$ of relativistic electrons with energies between E and $E + dE$ is given by

$$N(E)dE = \begin{cases} N_o\left(\frac{E}{E_o}\right)^{-p}dE, & E_{\min} < E \le E_{\max} \\ 0 & \text{otherwise.} \end{cases} \tag{4.13}$$

Each electron of energy E radiates power $P(E)$ so that the luminosity $dL(E)$ of the electrons with energies in the range E to $E + dE$ is given by

[11] Equation (3.72) shows that they will also radiate at much lower frequencies.

$$dL(E) = N(E)dE \times P(E) = N_o E_o^p bB^2 E^{2-p} dE, \quad (4.14)$$

where I have used equation (3.73). We can use equation (3.71) to eliminate the energy E from this equation:

$$dL(\nu) = N_o E_o^p bB^2 \left(\frac{\nu}{aB}\right)^{(2-p)/2} d\left(\frac{\nu}{aB}\right)^{1/2} = \frac{1}{2} N_o E_o^p \left[\frac{b}{a^{(3-p)/2}}\right] B^{\left(\frac{1+p}{2}\right)} \nu^{\left(\frac{1-p}{2}\right)} d\nu. \quad (4.15)$$

This spectrum has the same form as that given in equation (2.3) provided that

$$\alpha = \frac{1-p}{2}. \quad (4.16)$$

Local cosmic rays have an energy spectrum of the form (4.13) with $p \sim 2.5$, which would give $\alpha \sim -0.75$. This is remarkably close to the commonly observed value of –0.7. We may therefore feel that we have "explained" in terms of "cosmic rays" in the radio sources. We have yet to explain, though, why p for cosmic rays should have this value! I shall return to this point later.

4.1.4 Lifetime of Relativistic Electrons

Because the electron is radiating power $P(E)$, it is losing energy at this rate so

$$-\frac{dE}{dt} = P(E). \quad (4.17)$$

Substituting for $P(E)$ from equation (3.58) we find

$$-\frac{dE}{dt} = bB^2 E^2. \quad (4.18)$$

Hence, an electron with energy E loses energy ΔE in a time Δt given by

$$\Delta t = \frac{\Delta E}{bB^2 E^2}. \quad (4.19)$$

Crudely, we can put ΔE equal to E and obtain a characteristic time $\tau(E)$ for an electron to lose its energy, where

$$\tau(E) = \frac{1}{bB^2 E}. \quad (4.20)$$

Notice that the more energetic the electron, the more rapidly it loses its energy. $\tau(E)$ can be considered as the lifetime for radiation from an electron of energy E.

[More formally, we can integrate equation (4.18):

$$\int_{E_i}^{E} \frac{dE'}{E'} = -bB^2 t, \quad (4.21)$$

where E_i is the initial velocity of the electron at $t = 0$. Carrying out the integration on the left-hand side, we find that

$$E = \frac{E_i}{1 + bB^2 E_i t}. \quad (4.22)$$

Even if the electron were to start with infinite energy, it would have energy E after a finite time $t(E)$ given by

$$t(E) = \frac{1}{bB^2 E}. \quad (4.23)]$$

We can use equation (3.71) to eliminate the energy E from (3.41) and obtain the characteristic lifetime $\tau(\nu)$ of an electron radiating at frequency ν:

$$\tau(\nu) = \left(\frac{a^{1/2}}{b}\right) B^{-3/2} \nu^{-1/2}. \quad (4.24)$$

Substituting numerical values, we get

$$\tau(\mathrm{s}) = 6 \times 10^5 B(\mathrm{T})^{-3/2} \nu(\mathrm{Hz})^{-1/2}. \quad (4.25)$$

4.2 Models of the Radio Lobes

4.2.1 The Energy in the Lobes

In this section, I shall assume that all electrons have the same energy E_o. That is, I shall replace the true energy spectrum given by equation (3.59) by the simplified spectrum

$$N(E)dE = N_o \delta(E - E_o)dE. \quad (4.26)$$

This will make treatment much easier without changing the overall conclusions. Suppose the source has total luminosity L. From equations (4.26) and (3.78), we have

$$L \equiv \int N(E)dE \times P(E) = N_o bB^2 \int E^2 \delta(E - E_o)dE = N_o bB^2 E_o^2 = N_o P(E_o). \quad (4.27)$$

The total energy $E_{\text{electrons}}$ contained in the electrons is given by

$$E_{\text{electrons}} = \int N(E)dE \times E = N_o \int E\delta(E - E_o)dE = N_o E_o. \quad (4.28)$$

From equations (4.8) and (4.13) we get

$$E_{\text{electrons}} = \frac{N_o E_o}{N_o P(E_o)} L = \frac{E_o}{P(E_o)} L = \left(\frac{a^{1/2}}{b}\right) L B^{-3/2} \nu_o^{-1/2}, \quad (4.29)$$

where I have used equation (4.5) to eliminate E_o. Putting numerical values into, we obtain

$$E_{\text{electrons}}(\text{J}) = 6\times10^{-5}\, L(\text{W}) B(\text{T})^{-3/2} \nu_o(\text{Hz})^{-1/2}. \quad (4.30)$$

Equation (3.41) takes account only of the energy in the electrons. If the material is to be electrically neutral, the electrons must be accompanied by an equal number of protons, which will also have energy. To allow for this, let us say that the energy E_{protons} in the protons is related to $E_{\text{electrons}}$ by

$$E_{\text{protons}} = K E_{\text{electrons}}. \quad (4.31)$$

The total energy $E_{\text{particles}}$ in particles is then given by

$$E_{\text{particles}} \equiv E_{\text{electrons}} + E_{\text{protons}} = (1+K) E_{\text{electrons}} = (1+K)\left(\frac{a^{1/2}}{b}\right) L B^{-3/2} \nu_o^{-1/2}. \quad (4.32)$$

We have no way of determining K directly. We may have a clue in the cosmic ray flux measured near the Earth in which the protons carry 100 times as mush energy as the electrons. In the absence of better information, therefore, let us take K to be 100. Then it is easy to show that the total particle energy in the radio lobes of Cygnus A, for example, is equivalent to the rest-mass of about 10^5 M_{sun}.

A magnetic field of induction B has an energy u_{field} associated with it, given by

$$u_{\text{field}} = \frac{B^2}{2\mu_o}, \quad (4.33)$$

where μ_o is the permeability of free space. The total energy E_{field} contained in the field is therefore given by

$$E_{\text{field}} = V \frac{B^2}{2\mu_o}, \quad (4.34)$$

where V is the volume of the emitting region. Numerically, we have

$$E_{\text{field}}(\text{J}) = TBD \times V(\text{kpc})^3 B(T)^2, \quad (4.35)$$

The total energy E_{total} in the lobes needed to give the observed luminosity is therefore given, from equations (3.72) and (4.34), by

$$E_{\text{total}} \equiv E_{\text{particles}} + E_{\text{field}} = (1+K)\left(\frac{a^{1/2}}{b}\right) L B^{-3/2} \nu_o^{-1/2} + V \frac{B^2}{2\mu_o}. \quad (4.36)$$

Unfortunately it is rarely possible to measure the magnetic field independently. How, therefore, are we to estimate the total energy E_{total}? The first term in equation (4.36) is a decreasing function of the magnetic induction B whereas the second term is an increasing function of B. There must therefore be a minimum in the total energy needed to give an observed luminosity L. This is shown in Figure 4.7, which plots the logarithm of the particle, field and total energies against the logarithm of the induction.

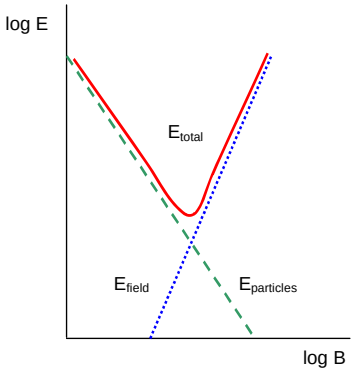

Figure 4.7. Variation of particle (dashed curve), field (dotted curve) and total (solid curve)energies with B.

It is usually assumed that the magnetic field has the value that minimises the total energy for no better reason than that, even *with* this assumption, the total energies required – of around 10^{54} J – are embarrassingly high!

To find the minimum, we differentiate E_{total} with respect to B and set the result equal to zero to get B_{min}. We find

$$B_{\text{min}} = \left\{\left(\frac{3}{2}\frac{a^{1/2}}{b}\mu_o\right)\left[(1+K)\left(\frac{L\nu_o}{V}\right)\right]\right\}^{2/7}. \quad (4.37)$$

The first term in round brackets depends solely on fundamental constants. The terms in square brackets are estimated or observed quantities.

Putting in numerical values, we get

$$B_{\text{min}}(\text{T}) = 0.23 \times \left[(1+K)\frac{L(\text{W})\nu_o(\text{Hz})}{V(\text{m}^3)}\right]^{2/7}. \quad (4.38)$$

It is easy to show, from equations (4.32) and (4.35) that the magnetic induction B_{min}, makes the energy in the particles nearly equal to that in the magnetic field:

$$\frac{E_{\text{particles}}}{E_{\text{field}}} = \frac{4}{3}. \tag{4.39}$$

Some people prefer to *start* with the assumption that the energy is equally divided between the particles and the field – the so-called equipartition of energy – rather than to assume that the energy is minimised. From equation (4.39), it is obviously immaterial which assumption is made.

From equations (3.52) and (3.75), we have for the minimum total energy E_{min},

$$\begin{aligned} E_{\text{min}} &= \left(1+\frac{4}{3}\right)E_{\text{field,min}} = \\ &= \frac{7}{6}\left(\frac{V}{\mu_{\text{o}}}\right)^{3/7}\left[\frac{3}{2}(1+K)\frac{a^{1/2}}{b}L\nu_{\text{o}}^{-1/2}\right]^{4/7}. \end{aligned} \tag{4.40}$$

4.2.2 The Supply of Energy to the Lobes

Perhaps the most natural assumption, given their disposition on either side of the galaxy, is that the radio lobes were ejected as entities from the galaxy and have been travelling independently since through intergalactic space. There is certainly nothing in the vicinity of the lobes themselves that could be responsible for them. The lobes are tens of kiloparsecs – some 50 kpc or 150 000 light years in the case of Cygnus A, for example – or more away from the galaxy. As we shall see later, they are moving out into the intergalactic medium at no more than a tenth of the velocity of light so that they would have taken $\sim 10^6$ years to get there. Yet it can be shown from equation (4.25) that the lifetime of an electron in a radio lobe is typically $\sim 10^5$ years. The electrons would therefore have lost all their energy in the time taken by the lobes to get to their present position.

In fact, the problem is far worse than appears at first sight! If the lobes had been ejected from the central galaxy, it is reasonable to assume that they had come from the nucleus of the galaxy – where the activity is seen – rather than from the benign elliptical galaxy surrounding the core. But the variability – on the time-scale of a year or less – of the activity in the nucleus shows that it must be less that a few parsecs in size. Let us take an upper limit of 10 parsecs. The radio lobes themselves, however are of the order of several *kilo*parsecs. Hence, the lobes must have undergone expansion by a factor of around 1000 in their passage from the nucleus of the galaxy. Let us denote the factor of expansion by f. The we can say that the lobe expands form an initial size d to a final size d' given by fd. Symbolically:

$$d \to d' = fd. \tag{4.41}$$

Now the lobes contain *plasma* – highly ionised matter – and the magnetic field B is therefore "frozen" into the matter[12]. As the lobes expand, therefore, the field lines get further apart and the field itself decreases. Quantitatively, the flux threading the lobe must be conserved. Since the size of the lobe is d, the flux Φ threading it is given by

$$\Phi \sim Bd^2 = B'd'^2 = \text{constant}, \tag{4.42}$$

where B' is the final value of the induction. From equations (4.41) and (4.42), therefore,

$$B' = f^{-2}B. \tag{4.43}$$

The total energy E'_{field} within the field after expansion is given by

$$\begin{aligned} E'_{\text{field}} &= V'\frac{B'^2}{2\mu_{\text{o}}} \sim d'^3\frac{B'^2}{2\mu_{\text{o}}} \\ &= f^3d^3\frac{f^{-4}B^2}{2\mu_{\text{o}}} = f^{-1}E_{\text{field}}, \end{aligned} \tag{4.44}$$

where V' is the expanded volume of the lobes and E_{field} is the total field energy before expansion.

What about the electrons? The radius a of the orbit of a highly-relativistic electron of energy E in the magnetic field B is given by

$$a = \frac{E}{eBc} \tag{4.45}$$

so that the flux Φ_a threading the electron's orbit is given by

$$\Phi_a \sim a^2B = \frac{1}{e^2c^2}\frac{E^2}{B}. \tag{4.46}$$

For changes that are slow compared with the time to complete one orbit – that is, for *adiabatic* changes – Φ_a must be constant so that

$$\frac{E'^2}{B'} = \frac{E^2}{B} \tag{4.47}$$

or, using equation (4.18),

$$E' = f^{-1}E. \tag{4.48}$$

If no new electrons are added to the plasma, the total number N of electrons contained in the lobes must be constant so that

$$\begin{aligned} E'_{\text{particles}} &= (1+K)NE' \\ &= (1+K)Nf^{-1}E = f^{-1}E_{\text{particles}}. \end{aligned} \tag{4.49}$$

[12]Strictly, the conductivity of the plasma would have to be infinite for the field to be completely tied to the matter, but the approximation is good.

The field and particle energies therefore scale in the same way with the expansion factor f, preserving the "equipartition" between the particle and field energies:

Using equation (3.84) for the power radiated by an individual electron, we have for the total power P'_{total} radiated by the lobes after expansion

$$\begin{aligned} P'_{\text{total}} &= NbB'^2E'^2 \\ &= Nbf^{-4}B^2f^{-2}E^2 = f^{-6}P_{\text{total}}, \end{aligned} \tag{4.50}$$

where P_{total} is the total power radiated *before* the expansion. Equation (4.50) says that the power radiated by the lobes decreases by a factor f^6 during the expansion. As we have seen, f is of order 10^3 so the power decreases by eighteen orders of magnitude during the expansion! If the ejection hypothesis were true, therefore, one would expect to find that sources with small separation between the lobes and the central galaxy were on average more powerful than the larger sources. On the contrary, it is the *larger* sources that are, on the whole, more powerful. We must therefore seek another explanation of the origin of the radio lobes.

The accepted explanation is that there is a beam or jet of particles – emanating from the central galaxy – which feeds the lobes with fresh relativistic electrons. As we have seen, there is direct evidence for such beams in radio sources.

4.2.3 Motion of the Jets and Lobes

4.2.3.1 Superluminal Velocity in Jets

Figure 4.8 shows knots or blobs of material in a jet leaving a central galaxy with velocity V. By taking radio measurements – separated by months or years – of these blobs, we can measure their *proper motion* μ, that is their angular velocity on the sky. If we can establish the distance r_o the galaxy, we can calculate the rate of change $\dot{p}$ of the projected distance p:

$$\dot{p} = r_o\mu\,. \tag{4.51}$$

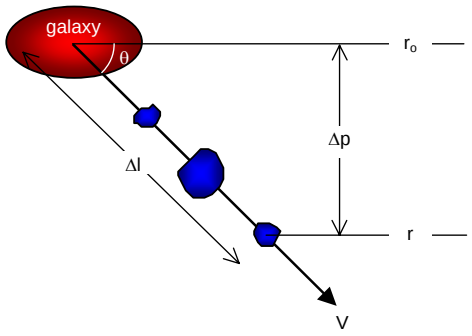

Figure 4.8. Illustration of superluminal motion.

As we have already seen, $\dot{p}$ is found to be greater than the velocity of light c in several sources. I shall now show that this an optical illusion and does not imply that the blobs are violating the tenets of relativity. Suppose the furthest blob was ejected from the galaxy at time t_e. The radiation emitted by the blob at this time is received at the Earth at time t_r given by

$$t_r = t_e + \frac{r_o}{c}, \tag{4.52}$$

where r_o is the distance of the galaxy from Earth[13]. At some time $t_e + \Delta t_e$, the blob is in the position shown in the diagram and is distant $r_o - \Delta r$ from the Earth. In this position, it emits radiation which is received on Earth at time $t_r + \Delta t_r$ where

$$(t_r + \Delta t_r) = (t_e + \Delta t_e) + \frac{r_o - \Delta r}{c}. \tag{4.53}$$

From equations (4.52) and (4.53), we have

$$\Delta t_r = \Delta t_e - \frac{\Delta r}{c}. \tag{4.54}$$

Note that the time between the reception of the two signals is less than the time between emission because the light has less far to travel: it is this difference which gives rise to the optical illusion.

In the time Δt_r, the blob is observed to have moved a distance Δp perpendicular to the line-of-sight where, from Figure 4.8,

$$\Delta p = \Delta l \times \sin\theta = V\Delta t_e \times \sin\theta\,, \tag{4.55}$$

where V is the velocity of the blob. From equations (4.54) and (4.55), we have

$$\dot{p} \equiv \frac{\Delta p}{\Delta t_r} = \frac{\Delta t_e}{\Delta t_r} \times V\sin\theta = \frac{V\sin\theta}{\left(1 - \frac{1}{c}\frac{\Delta r}{\Delta t_e}\right)}. \tag{4.56}$$

But it is clear from the figure that

$$\Delta r = \Delta l \times \cos\theta = V\Delta t_e \times \cos\theta \tag{4.57}$$

so that

$$\dot{p} = \frac{V\sin\theta}{\left(1 - \frac{V}{c}\cos\theta\right)} = \left(\frac{\beta\sin\theta}{1 - \beta\cos\theta}\right) \times c\,, \tag{4.58}$$

where

$$\beta := \frac{V}{c} \tag{4.59}$$

is the velocity of the blob expressed as a fraction of the velocity of light. Obviously, what we need if we are to

[13]For simplicity, I assume that the galaxy is at rest with respect to the Earth. This assumption does not affect the conclusion.

observe a blob apparently moving as fast as, or faster than, light is to have

$$\left(\frac{\beta\sin\theta}{1-\beta\cos\theta}\right)\geq 1\,. \tag{4.60}$$

We can re-arrange equation (4.60) to give

$$\beta\geq\frac{1}{\sin\theta+\cos\theta} \tag{4.61}$$

and it is easy to that the *minimum* value $\beta_{\min}$ of β needed for apparent superluminal velocity is $1/\sqrt{2}$:

$$V_{\min}\equiv c\beta_{\min}=\frac{c}{\sqrt{2}}=0.707c\,. \tag{4.62}$$

The fact that we *do* see such motion shows that the material in the jets feeding the radio lobes is moving relativistically[14]. Note that $(\sin\theta+\cos\theta)$ is symmetrical about $\theta=\pi/4$ so that the same value of β is needed for the angle $(\pi/4-\Delta\theta)$ as for $(\pi/4+\Delta\theta)$. The reason for this is that, what we lose in a longer light travel time ($\cos\theta$ decreasing) we gain in greater projected distance ($\sin\theta$ increasing), and *vice versa*.

Inequality (4.61) can be manipulated to give the range of values for θ which will given superluminal velocities for a given value of β:

$$\sin^{-1}\left[\frac{1}{2\beta}\left(1-\sqrt{2\beta^2-1}\right)\right]\leq\theta\leq\sin^{-1}\left[\frac{1}{2\beta}\left(1+\sqrt{2\beta^2-1}\right)\right] \tag{4.63}$$

4.2.3.2 Velocity of the Lobes

How can we get an estimate of the velocities with which the lobes are separating from the central galaxy? We can use the symmetry of the apparent images to get both upper and lower limits[15].

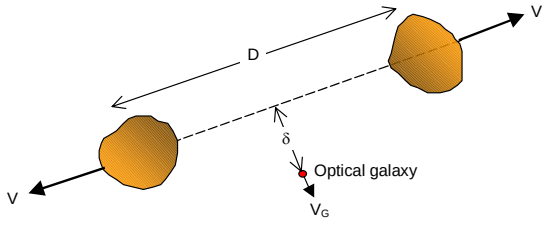

Figure 4.9. Lobe and galaxy velocities.

Figure 4.9 shows schematically a central galaxy between the two radio lobes that are separated by a distance D. Observation puts limits on the ratio of the distance δ the galaxy has moved away from the line joining the lobes:

$$\frac{\delta}{D}\lesssim\varepsilon\,, \tag{4.64}$$

where ε is to be determined from observation. Suppose that the lobes are separating from the galaxy with velocity V whilst the galaxy itself has a component of velocity V_G at right angles to the line joining the lobes. Then

$$D\sim Vt\,, \tag{4.65}$$

whilst

$$\delta\sim V_G t\,, \tag{4.66}$$

where t is the age of the lobes. From (4.64), (4.65) and (4.66), we have

$$V\gtrsim\frac{V_g}{\varepsilon}\,. \tag{4.67}$$

Because we cannot measure the velocity the galaxy directly, we have to apply (4.67) statistically. Typical random velocities of galaxies are around 500 km s^{-1} and typical values of δ are less than 0.01 kpc, giving a lower limit to values of V of some tens of thousands of kilometres a second or ~ 0.1 c.

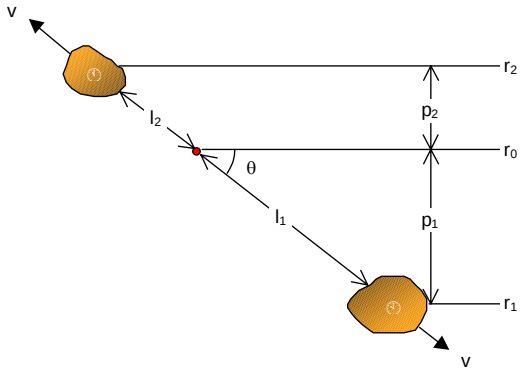

Figure 4.10. Asymmetry of radio sources

We can also get an upper limit on the velocity from the appearance of the source. Figure 4.10 shows lobes 1 and 2 of a radio source, separated from the central galaxy by distances l_1 and l_2 respectively. Their *projected* distances on the sky, at right angles to the observer's line of sight, are p_1 and p_2 respectively. Let the respective distances from earth of lobe 1, the central galaxy and lobe 2 be r_1, r_0 and r_2. I suppose that the separation of the two lobes from the central galaxy was zero at time t_o.

The argument is similar to that used in the discussion of superluminal velocity. Consider radiation received at the earth from all three components at time t_r. The radiation from the central galaxy was emitted at time t_e given by

[14]It is important to realise that we are here speaking of the *bulk* motion of the jets. The electrons within the jets may also moving with pseudo-random relativistic velocities.

[15]There are asymmetrical sources, such as "head-tail" sources, to which these arguments clearly do not apply.

$$t_e = t_r - \frac{r_0}{c}. \tag{4.68}$$

Similarly, the radiation from the two lobes was emitted at times t_I given by

$$t_i = t_r - \frac{r_i}{c}. \tag{4.69}$$

where i takes the values 1 or 2 for the two lobes respectively. From equations (4.68) and (4.69) we have

$$t_i = t_e + \frac{r_0 - r_i}{c}. \tag{4.70}$$

But, from the figure,

$$r_i = r_0 \mp l_i \cos\theta, \tag{4.71}$$

where the upper sign refers to lobe 1 and the lower to lobe 2. From equations (4.70) and (4.71), we have

$$\begin{aligned} t_i &= t_e + \frac{r_0 - (r_0 \mp l_i \cos\theta)}{c} \\ &= t_e \pm \frac{l_i \cos\theta}{c} \end{aligned} \tag{4.72}$$

Note that the radiation we receive at any time from lobe 1 is emitted later than that from lobe 2 received at the same time because lobe 1 is nearer to us than lobe 2. Since the lobes have been travelling with velocity V for a time $(t_i - t_0)$, we have for the distances l_i of the lobes from the central galaxy *at the time they emitted the radiation that is received on earth at time* t_r,

$$l_i = V \times (t_i - t_0) = V\left[(t_e - t_0) \pm \frac{l_i \cos\theta}{c}\right] \tag{4.73}$$

so that

$$l_i = \frac{V(t_e - t_0)}{\left(1 \mp \frac{V}{c}\cos\theta\right)} \tag{4.74}$$

and

$$\frac{l_1}{l_2} = \frac{1 + \frac{V}{c}\cos\theta}{1 - \frac{V}{c}\cos\theta}. \tag{4.75}$$

Finally, the *projected* separations p_i are given by

$$p_i = l_i \sin\theta \tag{4.76}$$

so that they are also related by the right hand side of equation (4.75):

$$\frac{p_1}{p_2} = \frac{1 + \frac{V}{c}\cos\theta}{1 - \frac{V}{c}\cos\theta}. \tag{4.77}$$

Note that p_1 is greater than p_2 because the light started later from lobe 1, which therefore had a longer time to travel. If observation shows that

$$\frac{p_1}{p_2} \le \eta, \tag{4.78}$$

then equation (4.77)shows that

$$V\cos\theta \le \frac{\eta - 1}{\eta + 1} c. \tag{4.79}$$

Unfortunately, the angle θ cannot be measured. We can apply inequality (4.79) *statistically*, though, assuming that the orientation of radio galaxies is random. The result is that the velocities of radio lobes do not exceed about a tenth of the velocity of light.

4.2.4 Confinement of the Radio Lobes

4.2.4.1 Diameter-Separation Ratio

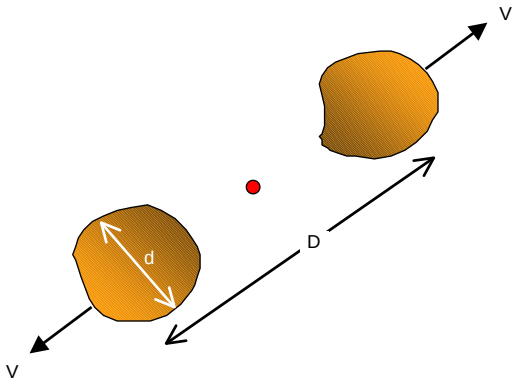

Figure 4.11. Diameter-separation ratio of radio lobes.

Figure 4.11 is a sketch of a typical pair of radio lobes. Consider the ratio χ of the separation D of the lobes to their diameter d:

$$\chi := \frac{D}{d} \tag{4.80}$$

Observed values of χ are around ten or more. Let us assume that the two lobes, constantly replenished with fresh electrons, have been moving way from the central galaxy with velocity V for a time t. Clearly,

$$\frac{D}{2} \le ct, \tag{4.81}$$

since otherwise they would have been travelling faster than light. Now, if there is nothing to stop them, the lobes themselves will expand and they will do so at the

velocity of sound u_S in the gas of which they are composed, giving

$$d \sim u_S t \,. \tag{4.82}$$

In general, the velocity of sound of sound in a fluid is given by (cf. Note 3)

$$u_S^2 = \left(\frac{\partial p}{\partial \rho}\right)_S \tag{4.83}$$

where p is the pressure and ρ the density. For a relativistic fluid, where the pressure is related to the energy-density u by

$$p = \frac{1}{3}u = \frac{1}{3}\rho c^2 , \tag{4.84}$$

we have

$$u_S = \frac{c}{\sqrt{3}} . \tag{4.85}$$

From relations (4.81), (4.82) and (4.85), we can deduce that

$$\frac{D}{d} \lesssim 2\sqrt{3} \tag{4.86}$$

in contradiction of the *observed* relation (4.119). We have therefore to abandon our hypothesis that the lobes are unconfined and seek some confinement mechanism.

4.2.4.2 Inertial Confinement

Let us first try an *inertial* mechanism for confining the lobes. Let us suppose that, in addition to the relativistic electrons, the lobes contain much more *non-relativistic* material of mass M. If we also assume that this material is ionised, it will be "glued" to the electrons by electromagnetic forces, thus providing them with additional inertial mass. In this case, the *pressure* of the lobes will be dominated by that of the relativistic electrons whilst their *mass* will be dominated by the non-relativistic material. The velocity of sound will therefore be given by [cf. Equation (4.83)]

$$u_S^2 \sim \frac{p_{\text{relativitic electrons}}}{\rho_{\text{non-relativistic matter}}} \sim \frac{u}{\left(M/d^3\right)} \sim \frac{U}{M} \tag{4.87},$$

where $U \sim (d^3 \times u)$ is the total energy in the lobes. Hence,

$$d \sim u_S t \sim \left(\frac{U}{M}\right)^{1/2} t \,. \tag{4.88}$$

On the other hand, if the lobes are separating from the nucleus at velocity V, we must have (cf. Figure 4.11)

$$D \sim 2Vt \,. \tag{4.89}$$

Using relations (4.88) and (4.89), we get

$$\left(\frac{D}{d}\right)^2 \sim \frac{4MV^2}{U} \sim 8\frac{T}{U} \tag{4.90}$$

where T is the total *bulk* kinetic energy of a lobe. From equations (4.90) and (4.80), we have

$$T \sim \frac{1}{8}\chi^2 U \,. \tag{4.91}$$

From the observed values of χ, it seems that we need at least an order of magnitude more energy T, in the form of bulk motion of the lobes, to confine the lobes in this way. Since U is already embarrassingly high, this does not seem an attractive solution!

4.2.4.3 Ram-Pressure Confinement

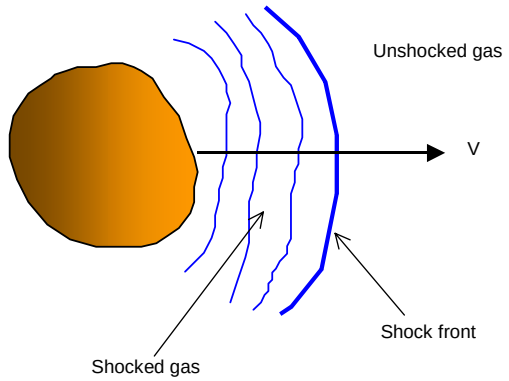

Figure 4.12.Ram-pressure confinement.

Another possibility is to attempt to confine the lobes by means of their supersonic motion through the extragalactic gas. Such supersonic motion will create a shock wave, as sketched in Figure 4.12. The material that is being pushed out of the way by the lobe will exert a pressure on it. Remembering that

$$\begin{aligned} &\text{force} = \text{rate of change of momentum;} \\ &\text{pressure} = \text{force per unit area,} \end{aligned} \tag{4.92}$$

we can see that the ram-pressure p_{ram} exerted on the lobe by the external gas is given by

$$P_{\text{ram}} \sim \frac{\text{mass invloved} \times \text{change in its velocity}}{\text{time taken to change velocity}} \times \frac{1}{\text{area}} \sim \frac{\left(\rho_{\text{external}} d^3\right) \times V}{\left(\frac{d}{V}\right)} \times \frac{1}{d^2} = \rho_{\text{external}} V^2 . \tag{4.93}$$

where ρ_{external} is the density of the extragalactic gas. The work W done by the lobes against this pressure is given by

$$W \sim \text{force exerted} \times \text{distance moved}$$
$$\sim \left(pd^2\right) \times D \sim \left[\left(\frac{U}{d^3}\right)d^2\right] \times D = \left(\frac{D}{d}\right) \times U \qquad (4.94)$$

where I have used equation (4.84). We see that, ram-pressure confinement again requires at least several times the internal energy of the lobes in order to do be effective.

4.2.4.4 Thermal Confinement

Finally, let us see what happens if the lobes are moving *subsonically* through the extragalactic medium and that they are confined by the thermal pressure p_{external} exerted on them by that medium. We must have

$$p_{\text{external}} = p_{\text{internal}} \ , \qquad (4.95)$$

where p_{internal} is the pressure within the lobes. Because this pressure is exerted by the electrons within the lobes, it is given by equation (4.84):

$$p_{\text{internal}} = \frac{1}{3} u_{\text{internal}} \ . \qquad (4.96)$$

where u_{internal} is the energy density in the lobes. On the other hand, the pressure in the extragalactic medium is thermal and is therefore given by

$$p_{\text{external}} = n_{\text{external}} kT_{\text{extarnal}} \ , \qquad (4.97)$$

where n_{external} and T_{external} are the particle density and temperature of the external medium respectively. We have, therefore,

$$n_{\text{external}} kT_{\text{external}} = \frac{1}{3} u_{\text{internal}} \ . \qquad (4.98)$$

Observations of X-ray emission from intra-cluster gas (cf. Note 5) show that the condition given by equation (4.98) is feasible, at least in clusters.

4.2.4.5 Summary and Conclusions

Inertial ram-pressure confinement models can be constructed in which the lobes move, with about a tenth of the velocity of light, through an external medium with density 10^{-25} to 10^{-24} kg m^{-3}. These densities are acceptable and consistent with X-ray results for the intra-cluster medium but they cannot be universal. The results from thermal confinement models of low-luminosity radio sources are also consistent with the X-ray observations of clusters but require too high a density for high-luminosity sources..

4.3 The Particle Energy Spectrum

4.3.1 The Need for an Acceleration Mechanism

Although we have tentatively identified the *source* of energy we see in AGN, we have not yet explained how the particles in radio sources are accelerated to relativistic energies. Nor have we any explanation for the observed power-law spectrum of particle energies given by equation (4.13):

$$N(E) = N_{\text{o}} \left(\frac{E}{E_{\text{o}}}\right)^{-p}$$

with p often equal to about 2.5. There is no final answer to these questions but we can get some insight into likely processes. All current ideas involve multiple interactions with macroscopic *scattering centres* such as turbulent eddies, plasma waves or shock-fronts.

4.3.2 Stochastic Acceleration[16]

4.3.2.1 General Scheme

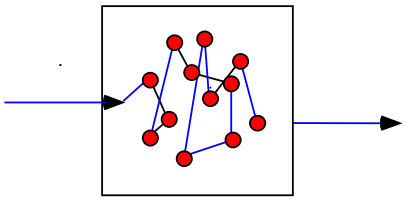

Figure 4.13. Scattering centres.

We imagine the interaction of fast-moving particles – mainly electrons and protons – with slower moving, but much more massive, scatterers that have much more energy than the particles. The actual mechanism by which the particle is scattered turns out to be immaterial but is presumably electromagnetic. Figure 4.13 represents a particle entering a region with a population of scatterers, undergoing multiple collisions and then escaping. The probability of the particle having a given energy on escape is determined by the balance between the rate of growth of its energy and its rate of escape.

4.3.2.2 Elastic Collisions between Particles

Suppose the particle of mass m and velocity v collides with the scatterer of mass M and velocity V. Let the velocities of the particle and scatterer after collisions be v' and V' respectively. In one dimension, conservation of momentum gives us that[17]

$$mv + MV = mv' + MV' \qquad (4.99)$$

whilst conservation of energy gives

$$\frac{1}{2}mv^2 + \frac{1}{2}MV^2 = \frac{1}{2}mv'^2 + \frac{1}{2}MV'^2 \ . \qquad (4.100)$$

Equations (4.99) and (4.100) can be solved to give

[16] This problem is really a relativistic one, but the treatment given here gives the right qualitative result.

[17] I use classical, rather than relativistic, mechanics but this does not affect the conclusions.

$$v' = -\left(\frac{M-m}{M+m}\right)v + \left(\frac{2M}{M+m}\right)V;$$
$$V' = +\left(\frac{2m}{M+m}\right)v + \left(\frac{M-m}{M+m}\right)V. \quad (4.101)$$

If $M >> m$, equations (4.101) reduce to

$$v' \approx -v + 2V;$$
$$V' \approx V. \quad (4.102)$$

In *overtaking* collisions, where v is in the same direction as V,

$$|v'| = |v| - 2|V|, \quad (4.103)$$

representing a *loss* in energy. In *head-on* collisions, on the other hand, where v is in the opposite direction to V,

$$|v'| = |v| + 2|V|, \quad (4.104)$$

so that the particle gains energy. If E and E' are the initial and final energies respectively of the particle, then

$$E = \frac{1}{2}mv^2;$$
$$E' = \frac{1}{2}mv'^2 = \frac{1}{2}m|v'|^2 = \frac{1}{2}m(v \pm 2V)^2. \quad (4.105)$$

If $v >> V$, then

$$E' \approx \frac{1}{2}m\left(v^2 \pm 4vV\right) \quad (4.106)$$

so that the change in energy ΔE is given by

$$\Delta E \equiv E' - E \approx \pm\frac{1}{2}m \times 4vV = \pm 4E\frac{V}{v}. \quad (4.107)$$

The change in the particle's energy is therefore proportional to the energy itself

This is solution to the one-dimensional problem. For two- or three-dimensional scattering, the final velocities are under-determined unless we specify the angle through which the particle is scattered. But, for *isotropic scattering*, the one-dimensional solution is a good approximation to the *average* change in velocity. Realistic scattering, will not necessarily be either isotropic or elastic. A more detailed analysis typically gives a result similar to (4.107) but, with the factor of four replaced by unity:

$$\Delta E \approx \pm E\frac{V}{v}.. \quad (4.108)$$

Assuming that the velocity of the particle is already comparable to c, we have

$$\Delta E \approx \pm E\frac{V}{c}.. \quad (4.109)$$

where the positive sign holds for head-on collisions and the negative for overtaking collisions.

4.3.2.3 Growth of Energy

At first sight, we seem to have achieved nothing: what we gain in head-on collisions, we lose in overtaking ones. But, head-on collisions are *slightly more frequent* because the number of collisions per unit time depends on *relative* velocity v_{rel} of particle and "target" (cf. Note 3), the rate being given by

$$R = n\sigma v_{\text{rel}}, \quad (4.110)$$

where n is the number-density of the scatterers and σ is their cross-section. The rate R_+ of head-on collisions is therefore given by

$$R_+ = \frac{1}{2}n\sigma(c+V), \quad (4.111)$$

where the factor of one half reflects the fact that the particles are equally likely to be going in either direction. Similarly, the rate R_- of overtaking collisions is given by

$$R_- = \frac{1}{2}n\sigma(c-V). \quad (4.112)$$

The total rate R is given by

$$R \equiv R_+ + R_- = n\sigma c \quad (4.113)$$

and the net rate R_{net} by

$$R \equiv R_+ - R_- = n\sigma V \quad (4.114)$$

The net rate of energy gain $\dot{E}$ is given by

$$\dot{E} = R_+\Delta E_+ + R_-\Delta E_-$$
$$= (R_+ - R_-)E\frac{V}{c} = n\sigma V \times E\frac{V}{c} \quad (4.115)$$
$$= n\sigma E\frac{V^2}{c} = n\sigma cE\beta^2_{\text{scatterer}},$$

where $\beta_{\text{scatterer}}$ is the velocity of the scatterer measured as a fraction of the velocity of light:

$$\beta_{\text{scatterer}} := \frac{V}{c}. \quad (4.116)$$

The average energy gain *per collision* ΔE_{net} is therefore given by

$$\Delta E_{\text{net}} \equiv \frac{\dot{E}}{R} = \left(n\sigma E \frac{V^2}{c} \right) \Big/ (n\sigma c) = E \times \beta^2_{\text{scatterer}} . \quad (4.117)$$

Suppose the average time between collisions is τ. Each collision gives ΔE_{net} so that the energy gained dE in time dt is given by

$$dE = \frac{dt}{\tau} \times \Delta E_{\text{net}} \quad (4.118)$$

or

$$\frac{dE}{dt} = \frac{\Delta E_{\text{net}}}{\tau} = \beta^2_{\text{scatterer}} \times \frac{E}{\tau} . \quad (4.119)$$

Equation (4.83) predicts *exponential* growth of energy with time:

$$E = E_{\text{o}} \exp\left[\beta^2_{\text{scatterer}} \frac{t}{\tau} \right] . \quad (4.120)$$

Inversely, the time $t(E)$ required to reach energy E is given by

$$t(E) = \frac{1}{\beta^2_{\text{scatterer}}} \ln\left(\frac{E}{E_{\text{o}}} \right) \times \tau . \quad (4.121)$$

4.3.2.4 Loss-rate

Let $N(E)dE$ be number density of particles with energies in range E to $E + dE$. Suppose that the probability P of an electron's escape from the scattering region is independent of time and energy. Then, in time dt, the decrease $dN_{\text{escape}}(E)$ in the density of particles through escape is given by

$$dN_{\text{escape}}(E) = N(E) P dt \quad (4.122)$$

so that the rate $\dot{N}_{\text{escaope}}(E)$ of escape of particles with energy E is given by

$$\dot{N}_{\text{escaope}}(E) \equiv \frac{dN_{\text{escape}}(E)}{dt} = \frac{N(E)}{T} , \quad (4.123)$$

where the *leakage time* T is given by

$$T := \frac{1}{P} . \quad (4.124)$$

4.3.2.5 Resultant Energy Distribution

The equation of continuity equation in one-dimensional energy-space is[18]

$$\frac{\partial N(E)}{\partial t} + \frac{\partial}{\partial E}\left[N(E) \frac{dE}{dt} \right] = -\dot{N}_{\text{escape}}(E) . \quad (4.125)$$

Let us consider the steady state in which the density $N(E)$ of electrons in energy-space is constant. Setting $\partial N(E)/\partial t$ to zero in equation (4.125) and substituting from equation (4.119) for dE/dt, we have

$$\frac{\partial}{\partial E}\left[N(E) \beta^2_{\text{scatterer}} \frac{E}{\tau} \right] = -\frac{N(E)}{T} , \quad (4.126)$$

which has solution

$$N(E) = N_{\text{o}} \left(\frac{E}{E_{\text{o}}} \right)^{-\gamma} , \quad (4.127)$$

where N_{o} and E_{o} are constants and

$$\gamma = 1 + \frac{1}{\beta^2_{\text{scatterer}}} \frac{\tau}{T} . \quad (4.128)$$

Hence this so-called *second-order Fermi acceleration* mechanism gives the desired power-law. Unfortunately, τ, T and $\beta_{\text{scatterer}}$ depend upon the physical processes going on in the source and does not predict the "universal" value of 2.5 for γ. Let us try something different.

4.3.3 SHOCK ACCELERATION

4.3.3.1 Energy-Gain

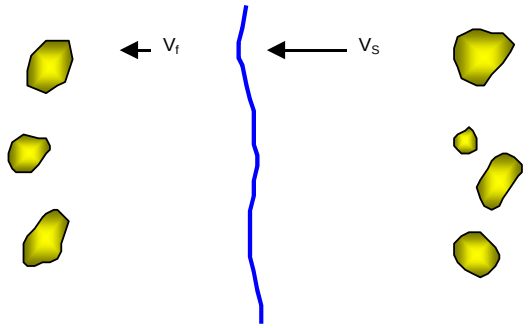

Figure 4.14. Shock acceleration.

An alternative mechanism is provided by the sort of shock fronts that we might expect to exist in the highly excited regions that exist around AGN. Figure 4.14 shows – in the rest frame of the shock front – scattering centres both upstream and downstream of the shock. These centres are more or less at rest in the material surrounding them and are therefore approaching each other at velocity ΔV, the difference in velocity between the speed relative to the shock of the upstream and downstream material:

$$\Delta V = V_S - V_f , \quad (4.129)$$

[18] This can be obtained by adapting the continuity equation of Note 3 to one dimension.

where V_S is the velocity of the shock advancing into the unshocked region and V_f is velocity of the material flowing away downstream of the shock front (cf. Note 3).

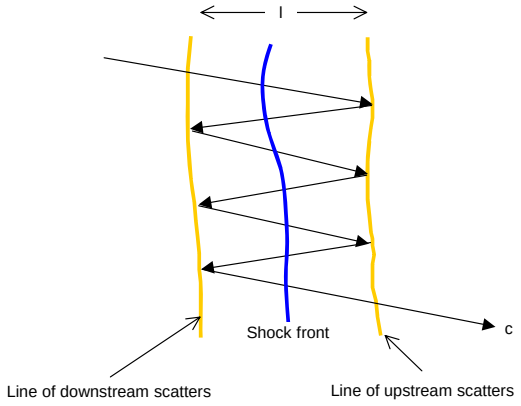

Figure 4.15. Multiple crossing of shock front.

Now consider electrons moving, at velocities nearly equal to c, crossing the shock front, bouncing off a scatterer, travelling back across the front, bouncing off another scatterer to re-cross the front again, etc, as shown schematically in Figure 4.15. The electrons meet the scatterers *head on* each time so that the gain in energy at each double crossing of the front is given by twice equation (4.108) with the *positive sign* and with V replaced by ΔV:

$$\Delta E = 2E\frac{\Delta V}{c}. \qquad (4.130)$$

Compare this first-order process, with the random collisions of section 4.3.2, in which the fractional change in energy was only *second-order* in $\Delta V/c$. In the present case, the equivalent of equation (4.119) is

$$\frac{dE}{dt} = \frac{\Delta E}{\tau} = \beta\frac{E}{\tau}, \qquad (4.131)$$

where τ is the mean time between scatterings, as before, and where β is now given by

$$\beta = 2\frac{\Delta V}{c}. \qquad (4.132)$$

4.3.3.2 Loss-rate

What about the time T required for the loss of electrons from the scattering region? Although the electrons are individually moving almost with velocity c, the *net* rate at which material crosses the front is only ΔV, the difference between the upstream and downstream velocities. If ℓ is the typical distance between scatterers on either side of the front, we must have

$$\frac{\ell}{T} \sim \Delta V. \qquad (4.133)$$

On the other hand, the electron crosses the front many times during this process at velocity c so that the mean time τ between double scatterings, is given by

$$\frac{2\ell}{\tau} \sim c. \qquad (4.134)$$

From equations (4.133), (4.132) and (4.134), therefore,

$$T \sim \frac{\ell}{\Delta V} \sim \frac{c\tau}{\beta c} = \frac{\tau}{\beta}. \qquad (4.135)$$

This is the key to getting a process-independent spectral index: the escape time is directly related to the mean time between scatterings.

4.3.3.3 Resultant Energy Distribution

With this new mechanism, the solution to equation (4.126) is still given by (4.127):

$$N(E) = N_o\left(\frac{E}{E_o}\right)^{-\gamma}$$

but γ is now given by

$$\gamma = 1 + \frac{1}{\beta}\frac{\tau}{T} = 1 + \frac{1}{\beta}\frac{\tau}{(\tau/\beta)} = 2, \qquad (4.136)$$

where I have used equation (4.135). We were trying to get a value of 2.5 for γ and have managed to get a value of 2. Not quite, therefore, but pretty close and at least totally independent of the detailed conditions in the source. It is likely, therefore that some modification of the process, including the effects of radiation losses, will prove to be the correct model.

Bibliography for Chapter 4

[1] Robson, I. *Active Galactic Nuclei*, Wiley 1996. ISBN 0-471-96050-0
[2] Field, G B, Arp, H and Bahcall, J N, *The Redshift Controversy*, W A Benjamin Inc, MA, 1973
[3] Shu, F H. *The Physical Universe*, University Science Books, CA, 1982. ISBN 0-19-855-706-X