Astrophysics 3 2001–2002
Physics of Stars and Nebulae

J.A. Peacock

Room R5, Royal Observatory; jap@roe.ac.uk

http://www.roe.ac.uk/japwww/teaching/stars_neb.html

Synopsis

Students are introduced to the physics of stars and their influence on their galactic environment. The basic observational properties of stars are reviewed including the HR diagram, followed by a discussion of the physical structure of stars on the Main Sequence to their final states as exotic objects: white dwarfs, neutron stars, and black holes. The interactions of stars with their surroundings are described. Topics covered are HII regions around young stars, stellar winds, and supernova remnants.

Recommended books

D.A. Ostlie & B.W. Carroll, *Modern Stellar Astrophysics*, Addison-Wesley Publishing Company, Inc. ISBN 0 201 59880 9.
J.E. Dyson & D.A. Williams, *The Physics of the Interstellar Medium, 2nd ed.*, Institute of Physics. ISBN 0 7503 0460 X.

Additional texts for consultation:

A.C. Phillips, *The Physics of Stars*
D.E. Osterbrock, *Astrophysics of Gaseous Nebulae*
L. Spitzer Jr, *Searching Between the Stars*
H. Scheffler & H. Elsasser, *Physics of the Galaxy and Interstellar Matter*
R.J. Tayler, *The Stars: Their Structure and Evolution*

Syllabus

1 Stellar classification & the HR diagram

2 The equations of stellar structure

3 The properties of Main Sequence and degenerate stars

4 An overview of the physical state of the Interstellar Medium (ISM)

5 The physics of HII regions

6 Fluids and shocks

7 The growth of stellar wind bubbles in the ISM

8 Supernova remnants

Tutorials

Tutorials will be held Mondays at 2.00 pm in JCMB (Rooms 5215 and 5326)

Schedule:

Tutorial 1: 22 October

Tutorial 2: 5 November

Tutorial 3: 19 November

Tutorial 4: 3 December

1 Preamble: fluxes, magnitudes and colours

The quantitative study of the radiation from stars is based on their **flux density** f_ν, which is the energy received per unit time, per unit area of detector, per unit frequency range (so the unit of f_ν is $\mathrm{W\,m^{-2}\,Hz^{-1}}$). It is of course possible to use wavelength instead of frequency as a measure of bandwidth, and there may well be practical reasons for doing so in terms of optical spectrographs, which produce a nearly linear wavelength scale. We then use f_λ, where

$$f_\lambda = \frac{d|\nu|}{d|\lambda|} f_\nu = \frac{c}{\lambda^2} f_\nu \qquad \text{units : W m}^{-3}. \tag{1}$$

Real instruments commonly deal with bandwidths that are large by comparison with features in the spectrum of the source under study. The total energy measured is then the integral under the source flux times some frequency-dependent effective filter response. This last quantity includes all the factors that modify the energy arriving at the top of the Earth's atmosphere. The main factor is the instrumental filter, but atmospheric absorption and frequency-dependent sensitivity of the detector also matter.

A notion of relative brightness can still be maintained by using the instrumental output, O, which is what leads to the notion of magnitude. The definition of magnitude is one of astronomy's unfortunate pieces of historical baggage, since it originates as a quantification of the ancient system where the brightest stars to the eye were 'first magnitude', and fainter stars were of larger magnitude. The eye is a logarithmic detector and the empirical constant of proportionality is 2.5, leading to

$$\boxed{m = -2.5\ \log_{10}\left(\frac{\mathrm{O_{object}}}{\mathrm{O_0}}\right).} \tag{2}$$

A magnitude system is then defined by the total effective filter and by the **zero point** (the object of zero magnitude that produces output $\mathrm{O_0}$). This is often (but not always) taken to be the A0 star Vega.

Once we have the apparent magnitude, this can be converted to the magnitude form of intrinsic luminosity, or **absolute magnitude** This is the apparent magnitude that would be observed if the source lay at a distance of 10 pc:

$$\boxed{M = m - 5\log_{10}\left(\frac{D}{10\ \mathrm{pc}}\right).} \tag{3}$$

(1 pc $= 3.0856 \times 10^{16}$ m).

Table 1 Parameters for common filter systems

Filter	$\lambda_{\rm eff}$/nm	$\Delta\lambda$/nm
U	360	65
B	440	100
V	540	80
R	680	95
I	900	230
J	1220	150
H	1630	170
K	2190	190
L	3450	280

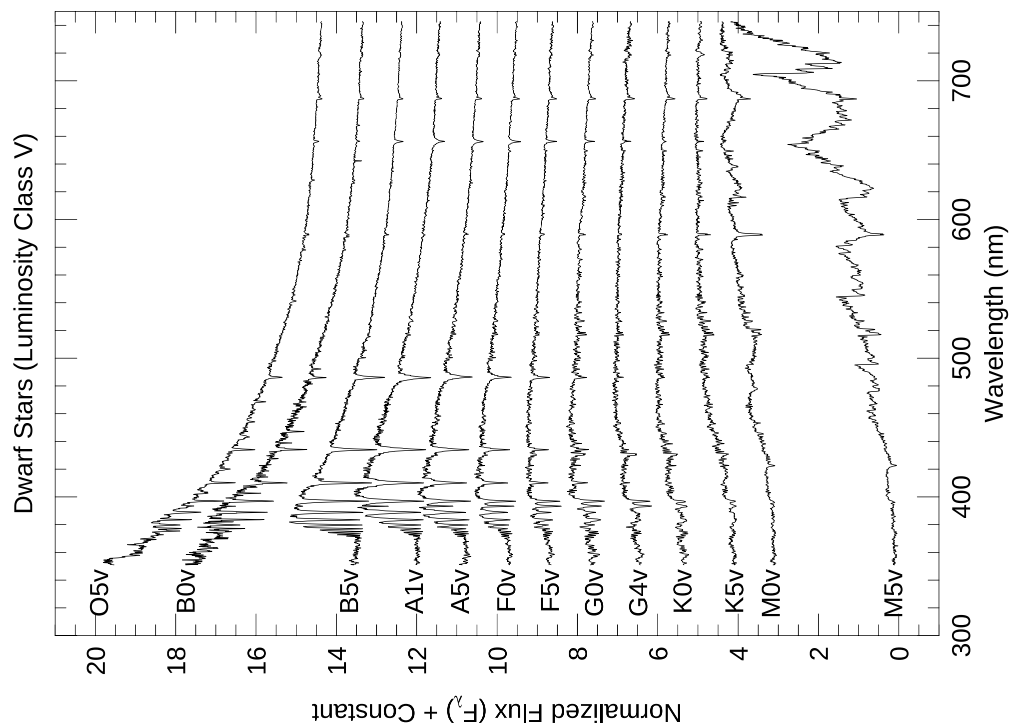

Figure 1. Illustrating how the spectra of main-sequence stars change systematically from types O5v through M5v ('early' to 'late') spectral types. Taken from `http://www-astronomy.mps.ohio-state.edu/~pogge/Ast162`

2 Stellar Classification

Stars may be classified on the basis of their spectra and colours. The spectra of stars are close to a black body in shape:

$$I_\nu = \frac{2h\nu^3}{c^2} \left(e^{h\nu/kT} - 1\right)^{-1} . \tag{4}$$

Here, h is one of two incarnations of **Planck's constant**: $h = 2\pi\hbar$, where $\hbar = 1.055 \times 10^{-34}$ Js. This 'Planck function' is the thermal specific intensity (flux per unit solid angle), and it peaks at wavelength $\lambda_{\rm peak}/\mu{\rm m} \simeq 2900\,{\rm K}/T$. The spectra of stars, however, show in addition large numbers of absorption lines. These arise largely in the thin surface layers comprising the atmospheres of the stars. These atomic absorption lines come in families, of which the most familiar is the spectrum of Hydrogen:

$$h\nu = E_m - E_n = 13.6(1/m^2 - 1/n^2)\ {\rm eV} \tag{5}$$

(where photon energy $h\nu$ is approximately 1.25 electron volts (eV) at $\lambda = 1\mu$m. The most prominent lines in the optical are the **Balmer series**, which have $n = 2$ as their lowest level.

Given their non-thermal spectrum, what exactly do we mean by the temperature of a star? The **effective temperature**, $T_{\rm eff}$, is defined such that a perfect black body of this temperature would radiate the same total luminosity as the real star:

$$\boxed{L = 4\pi R^2\, \sigma\, T_{\rm eff}^4, \quad \text{where} \quad \sigma = 5.67\times 10^{-8}\ \mathrm{W\,m^{-2}K^{-4}};} \tag{6}$$

R is the radius of the star, and σ is the **Stefan-Boltzmann constant**.

Some of the stronger lines in the hotter stars, with black-body temperatures $8000 < T_{\rm eff} <$ 40000 K, are due to incompletely ionized hydrogen, but absorption lines due to a wide range of chemical elements are found: Si, N, Ca, Fe, etc. The coolest stars, with $T_{\rm eff} < 4000$ K, even show absorption bands from molecules like TiO, CN, CH, and others.

Stars also show a range of colours across the visual spectrum, from blue to red. It was discovered already in the 1800s that the colours of stars were, in the main, determined by the star's surface temperature. This is to be expected for a black body spectrum, since f_ν peaks at a frequency ν that increases with T. Also, Pickering and Cannon in the 1880s developed a set of spectral classes (or 'Harvard types') based on the strongest absorption lines apparent in a star's spectrum. This is the classification scheme commonly used today. Unfortunately, their logical classification (ABCDE..., with 10 sub-classes 0...9) was based on the relative strength of hydrogen absorption lines, which do not vary monotonically with temperature – see below. It was Annie Jump Cannon who argued in the early 20th Century for a physical classification in order of temperature, but unfortunately the old letters were retained.

Table 2 Principal characteristics of spectral types

Spectral type	Spectral features
O	He II lines visible; H lines weak; High-ionization species (C III etc.); UV continuum
B	He I lines strong; He II lines absent; Lower-ionization species (C II etc.)
A	H lines strongest; Mg II & Si II strong
F	H lines weaker; Ca II stronger; lines from neutral and once-ionized metals
G	Solar-type; Ca II strong; neutral metals prominent; G band (CH) strong
K	Neutral metal lines dominate; molecular bands (CH, CN) developing
M	Strong molecular bands, particularly TiO; very red continuum

Of all these stars, the most familiar is the Sun, of type G2. For many years, it was the only star whose physical properties could be determined directly; these are listed in the following table.

Table 3 Properties of the Sun

Mass	Radius	Mean density	Luminosity	$T_{\rm eff}$
1.99×10^{30} kg	6.96×10^{8} m	$1410\,{\rm kg\,m^{-3}}$	3.83×10^{26} W	5770 K

2.1 The Main Sequence

One of the most important relationships in all of stellar astronomy is that between the absolute magnitude of a star and its spectral type. It was noticed in 1905 by Hertzsprung and independently in 1913 by Russell that stars fall mainly on a narrow locus in the $M - Sp$ plane (M = absolute magnitude; Sp = spectral type). This locus is called the **Main Sequence**. It was noticed by Hertzsprung that stars of the same spectral type, as determined by their absorption lines, were sometimes very dim stars (dwarfs), and sometimes very bright (giants) – i.e. some stars deviate from the Main Sequence. The plot of absolute stellar magnitudes against stellar type is credited to both and is known as the **Hertzsprung-Russell diagram** (or often just as the 'HR diagram').

2.2 MK Classification

Subsequently, in the 1940s and 1950s, Morgan and Keenan generalised the classification of stars by their absolute magnitudes into luminosity classes:

Ia-O	Supergiants
I	Supergiants
II	Bright Giants
III	Giants
IV	Subgiants
V	Main Sequence (dwarfs)
VI	Subdwarfs

In practice, this classification uses the shape of spectral lines to measure surface gravities of stars. The gravitational acceleration on the surface of a giant star is much lower than for a dwarf; the lower gravity means that gas pressures and densities are lower, so that spectral lines are narrower in these stars. Classifying a star according to both its luminosity class and its spectral type forms the basis of the MK Classification scheme of stars, central to stellar astronomy.

2.3 Colour-magnitude diagrams

Instead of plotting M against Sp, it is often convenient to plot M against colour, since both of these may be measured directly from stellar photometry with no need for doing spectroscopy. Stellar colours most commonly involve the UBV system developed by Johnson and Morgan in 1951. A colour such as $B - V$ is simply the difference of the magnitudes measured in these two wavebands, and has the following physical significance:

$$\text{larger colour} \quad \Rightarrow \quad \text{redder spectrum} \quad \Rightarrow \quad \text{lower temperature} \tag{7}$$

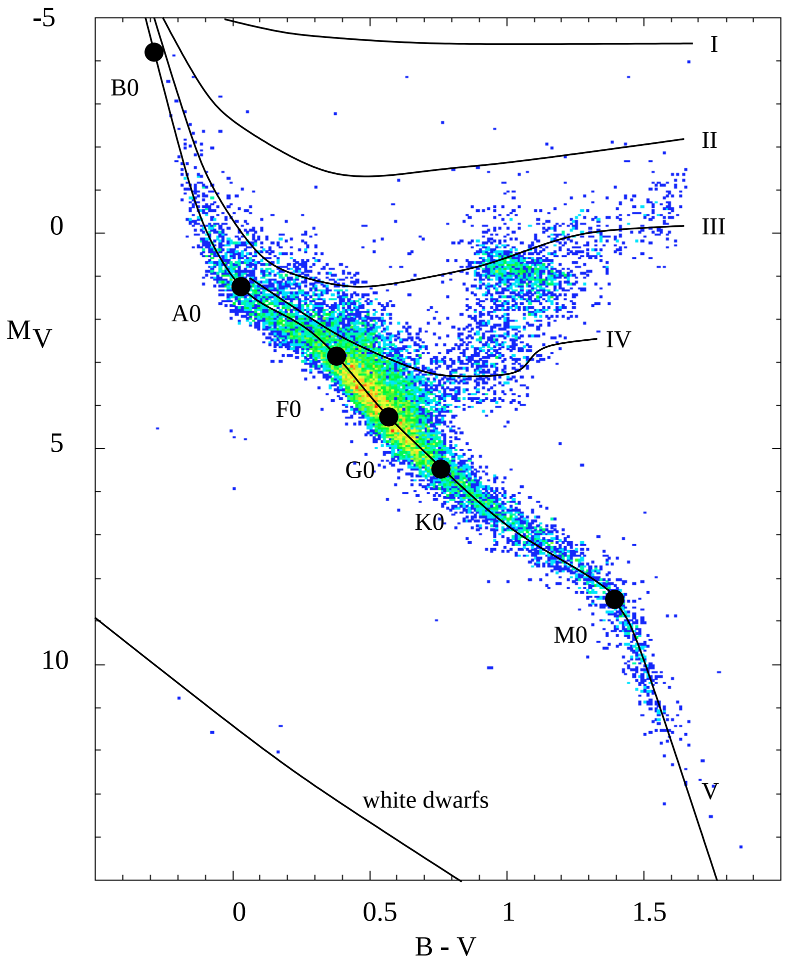

Figure 2. The colour-magnitude diagram (the observational version of the Hertzsprung–Russell diagram) for the Solar neighbourhood, showing 16631 stars from the Hipparcos catalogue with parallax distances known to better than 10%. The loci of the different luminosity classes are shown, as well as the locations of the different main-sequence spectral types.

Since colour is related to temperature, which is related to spectral type, the colour-magnitude diagram shows the same type of stellar loci as in the HR diagram.

The colour-magnitude diagram shown in Figure 2 fails to do justice to the precise physical correlations seen in systems of stars. The main sequence is blurred because the parallax distances are not perfectly accurate, and because the location of the main sequence depends on metallicity (more absorption lines in high-metal stars). Moreover, the Solar neighbourhood contains a mixture of stars of different ages, and age is a critical feature of the CMD. This is illustrated in Figure 3, which shows the CMD for a globular cluster. Here, all the stars are at the same distance, age and metallicity, so we can see how well defined the stellar loci are, illustrating not only the main sequence, but the giant branches. To anticipate our future conclusions, the form of this diagram is due to stars at the top left of the main sequence having used up a large amount of the 'fuel' in their cores, which causes them to move off the main sequence. They become giant stars with lower surface temperatures, but hugely greater sizes and hence greater luminosities.

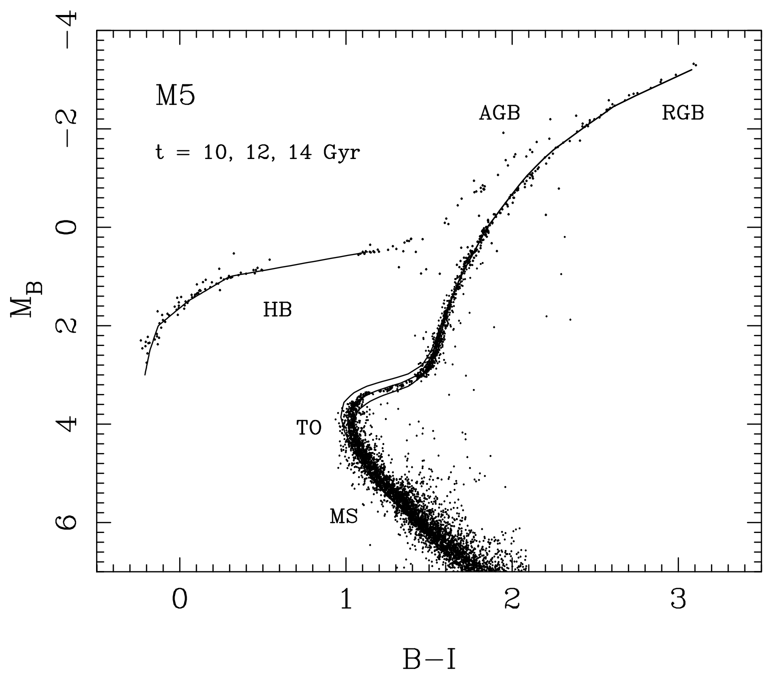

Figure 3. A colour–magnitude plot for stars in the globular cluster M5. This illustrates well the main features of stellar evolution: the main sequence (MS) and its turn-off point (TO), followed by the red giant branch (RGB), horizontal branch (HB) and asymptotic giant branch (AGB). The data are well-fitted by a theoretical isochrone of age 12 Gyr.

One very important systematic error in the CMD technique, however, is reddening due to dust. The cross-section for scattering photons decreases with increasing wavelength roughly linearly:

$$\sigma(\lambda) \propto \lambda^{-1}. \tag{8}$$

This means more blue light will be scattered by interstellar dust (dust distributed along with gas between a star and us). The result will be a systematic reddening of the starlight the greater the amount of dust, and a systematic mis-classification to 'later' stellar types. This may be corrected using a two-colour diagram.

3 Formation of spectral lines

The task of stellar astrophysics is to explain some of these systematics. Our conclusion will be that the sequence of different stellar types is driven mainly by a single parameter: the mass of the star. The observational properties of stars largely depend on their temperature. Later, we will show how this is related to temperature is related to mass. First, we will see why the spectral types change with temperature as they do.

The main thing to appreciate is that stars are hot enough that the material in their outer regions is partly ionized. How does the ionization depend on temperature? Consider a single hydrogen atom, for which we want to know the relative probabilities that the electron is bound

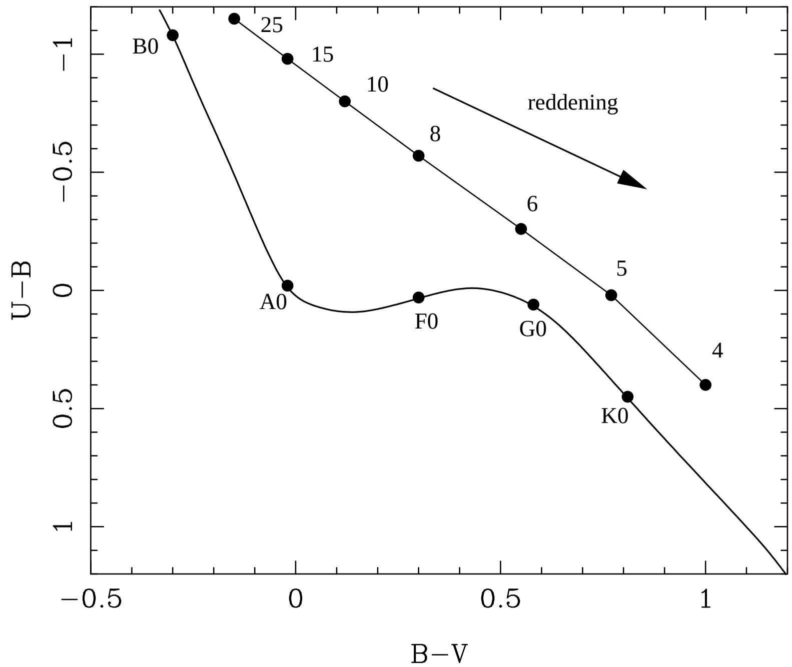

Figure 4. The two-colour plot for $U - B$ vs $B - V$, showing the expected locus of the main sequence, plus perfect black bodies (with temperatures in units of 1000 K)

or free. In equilibrium, the probability of a given state being occupied is proportional to the Boltzmann factor,

$$\boxed{P \propto \exp(-E/kT).} \tag{9}$$

Whether an electron is free or bound depends on the ratio of the **partition functions** (a fancy name for the sum of all the relevant Boltzmann factors):

$$\frac{P_{\text{free}}}{P_{\text{bound}}} = \frac{\sum_{\text{free}} g_i e^{-E_i/kT}}{\sum_{\text{bound}} g_i e^{-E_i/kT}} = \frac{Z_{\text{free}}}{Z_{\text{bound}}} \tag{10}$$

(where g is a degeneracy factor). In the case of this *single* object (the electron) the degeneracy factor $g = 2$, from spin, times the degeneracy factor for atomic orbitals. Now, a good approximation is that $Z_{\text{bound}} = g \exp(\chi/kT)$ where χ is the **ionization potential** of hydrogen. This will be valid at low temperatures $kT \lesssim \chi$, where the contribution of excited states to Z is negligible.

In order to deal with the free-electron states, we need to know how many such states fit into a box of volume V. The requisite **density of states** will be derived next term. Briefly, the box (of side L) should be imagined as filling all space, via periodic replication. This means that the wavevectors of the free-electron quantum states can only take certain values. If the electron wave function is $\psi \propto \exp(i\mathbf{k}\cdot\mathbf{x})$, where $\mathbf{k} = (k_x, k_y, k_z)$, then periodicity requires

$$k_x = n_x \frac{2\pi}{L} \quad \text{where} \quad n_x = 1, 2, \ldots \tag{11}$$

Using **de Broglie's relation**, the momentum of the electron relates to its wave vector via $\mathbf{p} = \hbar \mathbf{k}$. The allowed states in **momentum space** therefore lie on a lattice with spacing $L/(2\pi\hbar)$, and we can evaluate the free-electron partition function in a box of volume V by integrating over momentum:

$$Z_{\rm free} = g \frac{V}{(2\pi\hbar)^3} \int e^{-p^2/2mkT} \, d^3p = (2\pi mkT)^{3/2} \frac{gV}{(2\pi\hbar)^3}. \tag{12}$$

What value of V do we use? The volume 'available' to each electron is effectively $1/n_e$ – the reciprocal of the electron number density. Using this, and defining the fractional ionization x ($n_e = xn$, where n is the total **nucleon number density**, $n_{\rm H} + n_p$), we get the **Saha equation**

$$\boxed{\frac{x^2}{1-x} = \frac{(2\pi m_e kT)^{3/2}}{n(2\pi\hbar)^3} e^{-\chi/kT}.} \tag{13}$$

It is now clear why e.g. the Balmer lines show a non-monotonic dependence on temperature. These lines require the existence of neutral H atoms that are excited into the $n = 2$ state. These do not exist at very high temperatures, where the material is almost completely ionized. Conversely, at very low temperatures, the neutral H atoms exist almost entirely in the $n = 1$ ground state. The ratio of the abundance of atoms in these two states is

$$N_2/N_1 = (g_2/g_1) \exp(-E_{12}/kT) \tag{14}$$

and the overall density of $n = 2$ atoms is approximately

$$n_2 = (1-x)n \times \frac{N_2}{N_1 + N_2}. \tag{15}$$

This has a peak at some intermediate temperature.

We can put some typical numbers into this for the surface of the Sun. There, the nucleon density is about $n = 10^{19}\,\mathrm{m}^{-3}$ (or a mass density of only about $10^{-8}\,\mathrm{kg\,m}^{-3}$ – very much lower than the mean density). The important energies are $\chi = 13.6$ eV and $E_{12} = 10.2$ eV, and g_2/g_1 is 4. Therefore, for temperature T_5 in units of 10^5 K,

$$\begin{aligned} \frac{x^2}{1-x} &= (38750 T_5)^{3/2} \exp(-1.58/T_5) \\ N_2/N_1 &= 4 \exp(-1.18/T_5) \end{aligned} \tag{16}$$

The overall intensity of the Balmer lines peaks where this product is a maximum, which is at $T \simeq 13,500$ K. This is the surface temperature of an A star, which explains why the Balmer lines peak in this species.

In 1925, Celia Payne was the first to apply this sort of analysis to the atmospheres of stars, which allowed her to deduce the abundances of the different elements from their spectral lines. Remarkably, it was only as recently as this that it was first appreciated that Hydrogen was the dominant element in the universe.

4 Equations of Stellar Structure

A star is characterised by its mass M, radius R, luminosity L, and temperature T. We'll also see that its composition or **metallicity** Z also matters ('metals' are anything other than H & He: the Sun has about 2% of its mass in metals and 28% in He). Some of these quantities are related. For instance, since a star shines like a black body:

$$L = 4\pi R^2 \, \sigma \, T^4_{\text{surface}}, \quad \text{where} \quad \sigma = 5.67 \times 10^{-8} \text{ W m}^{-2}\text{K}^{-4}; \tag{17}$$

σ being the **Stefan-Boltzmann constant**. The mass and radius are related through the average density $\bar{\rho}$:

$$M = 4\pi R^3 \bar{\rho}/3. \tag{18}$$

More generally, if $\rho(r)$ denotes the local density of the star at a radius r from its centre, then the mass ΔM of the star in a shell of thickness Δr is $M = 4\pi r^2 \rho(r) \Delta r$, or

$$\boxed{\frac{dM(r)}{dr} = 4\pi r^2 \rho(r)} \tag{19}$$

This is the **first equation of stellar structure** (mass conservation, or the **equation of continuity**). $M(r)$ means the mass interior to radius r.

Now, the local density $\rho(r)$ and temperature $T(r)$ are also related. For a star on the main sequence, its hot interior behaves like an **ideal gas**:

$$P(r) = n(r)\, kT(r) \quad \Rightarrow \quad P(r) = [\rho(r)/\bar{m}]\, kT(r), \tag{20}$$

where k is **Boltzmann's Constant** ($= 1.38 \times 10^{-23}$ J K^{-1}), $n(r)$ is the number density of all particles (electrons plus nuclei, since the material in the Sun is ionized). The mass $\bar{m}$ is the **mean mass per particle**: $n(r) = \rho(r)/\bar{m}$. We shall generally assume $\bar{m} = m_p/2$: most of the Sun is Hydrogen plasma, and each electron-proton pair weighs basically just the proton mass – or $m_p/2$ on average.

The internal pressure of the star varies with radius, and this pressure gradient supports the star against its own gravity. Consider a parcel of gas of mass m inside the star, with area A and radial extent ΔR: The pressure force acting on the bottom is $F_{\text{bottom}} = P(r)A$; on the top, it is

$$F_{\text{top}} = P(r + \Delta r)A \simeq P(r)A + \frac{dP}{dr} \Delta r \, A. \tag{21}$$

The net pressure force acting is thus

$$F_{\text{bottom}} - F_{\text{top}} = P(r)A - \left[P(r)A + \frac{dP}{dr}\Delta r \, A \right] = -\frac{dP}{dr} \Delta r \, A. \tag{22}$$

In **hydrostatic equilibrium**, this net pressure force must balance gravity.

The gravitational acceleration at r produced by the star is

$$g = \frac{GM(r)}{r^2} \quad \text{(radially inwards)}, \tag{23}$$

and the gravitational force acting on the parcel of gas (of mass m) is $F = mg$. In order to balance the net pressure force, we must have

$$-\frac{dP}{dr}\Delta r\, A = m\,\frac{GM(r)}{r^2}, \tag{24}$$

and dividing by m gives the **equation of hydrostatic equilibrium**:

$$\boxed{-\frac{1}{\rho}\frac{dP}{dr} = \frac{GM(r)}{r^2}} \tag{25}$$

(using $\rho(r) = m/(\Delta r\, A)$, because $\Delta r\, A$ is the volume of the gas parcel.

This equation can tell us about conditions in the invisible stellar interior. As an example, we can estimate the pressure at the centre of the Sun, $P_{\rm C}$. As a rough approximation, we can assume a typical pressure gradient $dP/dR \simeq P_{\rm C}/R$ (i.e. the surface pressure is negligibly small). Hydrostatic equilibrium relates this typical gradient to a typical density and a typical acceleration. We take the former to be the mean density, $\bar{\rho}$, and the latter to be the surface acceleration, GM/R^2. This gives

$$P_{\rm C} \simeq GM\bar{\rho}/R = 3GM^2/4\pi R^4. \tag{26}$$

A more careful argument shows that this is an inequality. If we divide the equation of hydrostatic equilibrium by the equation of mass conservation, we get

$$-\frac{dP/dr}{dM/dr} = -\frac{dP}{dM} = \frac{GM(r)}{4\pi r^4} > \frac{GM(r)}{4\pi R^4} \tag{27}$$

Integrating dM from 0 to M then gives

$$P_{\rm C} > GM^2/8\pi R^4, \tag{28}$$

which is almost the previous result. This pressure is $4.5\times10^{13}\ \mathrm{N\,m^{-2}}$, or $10^{8.6}$ times atmospheric pressure.

What is the corresponding temperature? We assume the **perfect gas law**,

$$P = \rho kT/\bar{m}, \tag{29}$$

so the central temperature is

$$T_{\rm C} = \bar{m}P_{\rm C}/k\rho_{\rm C}. \tag{30}$$

If we assume that the central density is of order the mean density, this suggests a minimum temperature of

$$kT_{\rm C} \gtrsim \frac{G\bar{m}M}{R} \simeq 1.2\times10^7\ \mathrm{K}. \tag{31}$$

(taking $\bar{m} = m_p/2$, and discarding factors of order unity). In other words, the typical thermal energy is of order the gravitational binding energy. This approximate equality between kinetic and potential energies is very common in self-gravitating structures, and is known as the **virial theorem**. This temperature has been deduced using some rather dubious assumptions, but the final figure for the temperature is not so far wrong: the correct central temperature for the Sun is 1.6×10^7 K.

Now, we would like to get an equation for $L(r)$: the luminosity passing through each layer r in the star. [$L(R)$ is the luminosity emanating from the stellar surface, which is the star's luminosity]. To do this, we must consider the source of a star's energy.

5 Energy Generation in stars

The Sun loses energy at the rate $L_\odot = 3.8 \times 10^{26}$ W. How long would it take the Sun to use up all of its thermal energy? If each particle in the Sun carries a thermal energy $3kT/2$ on average (by **equipartition of energy**), then we have for the total thermal energy of these particles

$$U_{\text{thermal}} \simeq (M_\odot/\bar{m}) \times 3kT/2 \simeq (M_\odot/\bar{m}) \times \frac{GM_\odot \bar{m}}{R_\odot} = \frac{3GM_\odot^2}{2R_\odot} \simeq 6 \times 10^{41}\,\text{J}. \tag{32}$$

How long does it take the Sun to radiate away all this thermal energy?

$$t_{\text{thermal}} = \frac{U_{\text{thermal}}}{L_\odot} = \frac{6 \times 10^{41}\,\text{J}}{3.8 \times 10^{26}\,\text{W}} \simeq 1.6 \times 10^{15}\,\text{s} \simeq 50\,\text{Myr}. \tag{33}$$

This is much less than the age of the Earth, which is 4.56 Gyr. This was a fundamental problem recognised already by the end of the 19th century by Kelvin & Helmholtz (sometimes t_{thermal} is called the **Kelvin-Helmholtz time**). It tells us that the Sun must have some additional energy-generating mechanism. One of the great achievements of 20th Century physics was to work out what powers the stars. It just comes down to $E = mc^2$.

As soon as nuclear reactions were discovered in the early 20th Century, it was clear that there were a much greater potential source of energy than any chemical reaction, and were thus a plausible source of energy for a star. Consider **fusing** 4 protons (from the ionised hydrogen in the star) into a helium nucleus:

$$4p \to {}^4\text{He} + \text{Energy } Q_{\text{eff}} \tag{34}$$

(Note that ^{4}He consists of 2 protons and 2 neutrons, so 2 of the protons would need to convert into neutrons plus positrons in this process.) Now, the measured mass of ^{4}He is *less* than the total mass of the individual 4 protons (this is true even though the mass of a neutron is slightly more than the mass of a proton).

We know then that some mass Δm was lost, corresponding to an energy $Q_{\text{eff}} = \Delta mc^2$. Energy must be conserved in the fusion process, so this much energy is carried away by radiation: that's what Q_{eff} represents. Physically, there is a small but negative **binding energy** holding the 2 protons and 2 neutrons together in the ^{4}He nucleus. This binding energy represents the amount of energy needed to cause the nucleus to undergo **fission** into its component parts. Quantitatively, we find $\Delta m \simeq 0.007m({}^4\text{He})$, so that nearly 1% of the mass of the protons is converted into energy.

This is all very well, but fusion reactions will only happen if the protons can approach within the range of the strong nuclear interaction (about 10^{-15} m). The protons are electrically charged and repel each other, so it is not clear whether fusion will happen in practice.

5.1 Classical view

Start with a classical argument (which will turn out to be wrong). Two protons at large separation have a total energy that is roughly their kinetic energy, $E_{\text{tot}} \simeq E_K$, because their Coulomb energy decays with separation: $E_{\text{C}} = e^2/4\pi\epsilon_0 r$. To fuse, they must be brought together into the very small space of a nucleus: $r_{\text{N}} \simeq 10^{-15}$ m.

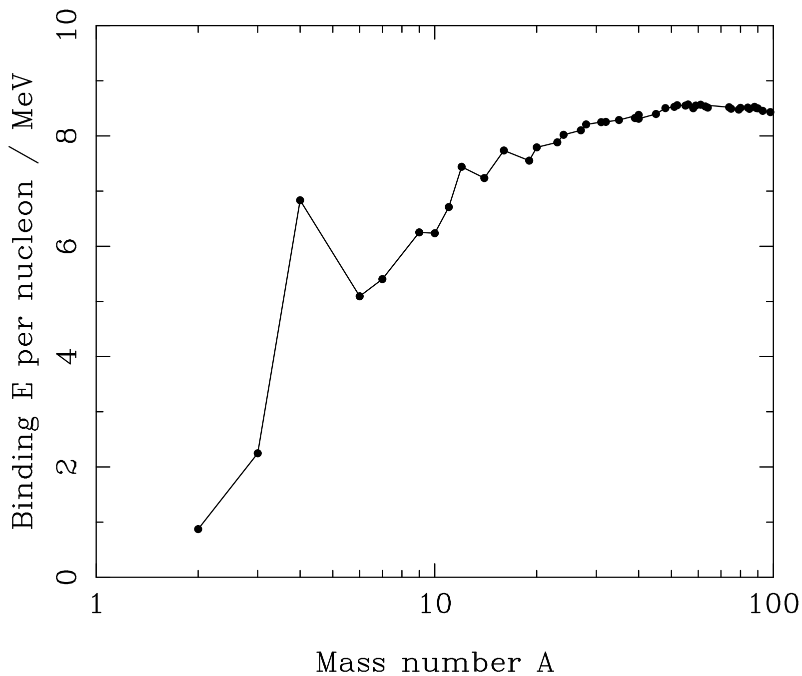

Figure 5. As we move to heavier nuclei, the inter-nuclear forces cause them to be more strongly bound. The best way of quantifying the relative strengths of these bonds is via the binding energy per nucleon (as measured by the mass of the nucleus compared to its component parts). In this plot (showing the most abundant isotopes of each element), two things stand out: (1) the relatively strong binding of ^{4}He; (2) the maximum at Fe ($Z = 26$, $A = 56$), which is the most stable element.

If they were initially given just enough kinetic energy that the Coulomb repulsion brings them exactly to rest as they reach a separation $r_{\rm N}$, then their final total energy will just be electrostatic potential energy. Conservation of energy then says

$$E_{\rm tot}^{\rm init} = E_{\rm tot}^{\rm final} \quad \Rightarrow \quad E_K^{\rm init} = E_{\rm C}^{\rm final} = \frac{e^2}{4\pi\epsilon_0 r_{\rm N}} \simeq 2\times 10^{-13}\,{\rm J}. \qquad (35)$$

If the protons have a temperature T, then the mean kinetic energy of two protons will be $E_K = 2 \times (3kT/2)$, and the required temperature will be

$$T = \frac{E_{\rm C}}{3k} \simeq 10^{10}\,{\rm K}. \qquad (36)$$

However, this is *much* hotter than the Sun's centre. This was a major stumbling block to advocates of nuclear fusion as the energy source in stars. Still, some weren't dissuaded, like Sir Arthur Eddington who said it must somehow work at the Sun's temperature, '...we tell him [the critic] to go and find a hotter place!'. The answer was recognised by George Gamow, and lies in the intrinsically non-classical phenomenon known as **quantum tunneling**.

5.2 Quantum tunneling

The Coulomb potential energy curve between two nucleons of charges $Z_{\rm A}e$ and $Z_{\rm B}e$ is

$$V(r) = \frac{Z_{\rm A} Z_{\rm B} e^2}{4\pi\epsilon_0 r}. \tag{37}$$

As we have seen, classical physics says that a particle of energy E is prevented by the **Coulomb barrier** from approaching within a distance $r_{\rm C}$, where

$$E = V(r_{\rm C}) = \frac{Z_{\rm A} Z_{\rm B} e^2}{4\pi\epsilon_0 r_{\rm C}}. \tag{38}$$

Quantum physics says a particle will nevertheless be able to tunnel through the barrier.

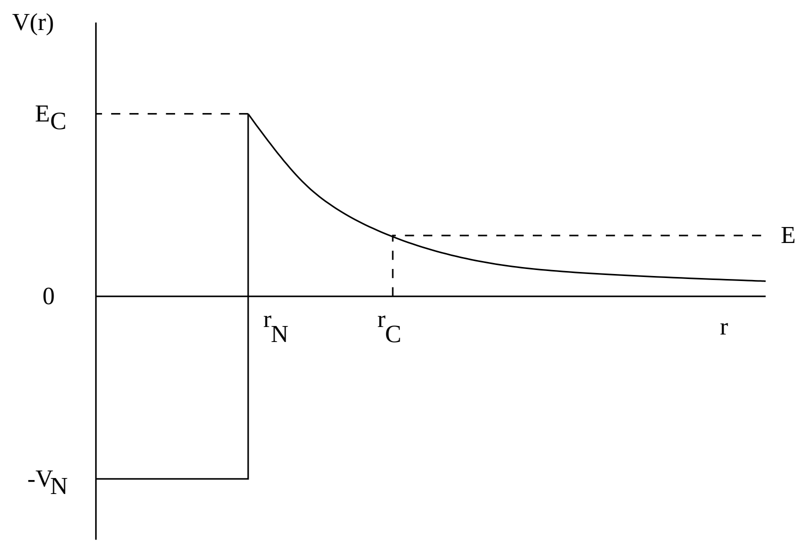

Figure 6. The potential experienced by a charged particle approaching a nucleus. The potential is attractive for $r < r_{\rm N}$, but repels at larger r. A particle of energy E at large r has zero energy at radius $r = r_{\rm C}$. If the energy is too small, then classical physics says the particle cannot reach the centre.

To give a proper quantum treatment, we need to give the approaching particle a **wavefunction**, $\psi(r)$, which obeys **Schrödinger's equation**:

$$\boxed{\frac{-\hbar^2}{2m_r}\nabla^2\psi + V\psi = E\psi,} \tag{39}$$

where the potential is

$$V(r) = \begin{cases} \frac{Z_{\rm A} Z_{\rm B} e^2}{4\pi\epsilon_0 r} & r > r_{\rm N} \\ -V_{\rm N} & r < r_{\rm N} \end{cases} \tag{40}$$

To be clear, we cannot really treat one particle as stationary and the other as moving. However, we will see in the quantum mechanics lecture course next term that this can be dealt with by treating the 2-particle system as a single particle having 'reduced mass' $m_r \equiv m_{\rm A} m_{\rm B}/(m_{\rm A}+m_{\rm B})$. Defining $\chi^2(r) = 2m_r(V-E)/\hbar^2$ converts Schrödinger's equation to

$$\nabla^2 \psi = \chi^2(r)\psi. \tag{41}$$

This equation would be easy to solve if χ was a constant: if $\chi^2 < 0$, we just get travelling plane waves, $\psi \propto \exp(i|\chi|x)$. However, if $\chi^2 > 0$, the wavefunction varies exponentially with position. We can therefore divide the region at $r > r_{\rm N}$ into two:

$$\boxed{\begin{array}{rlcl} \textbf{Classically allowed}: & V < E & \Rightarrow & \chi \text{ is imaginary and } \psi \text{ oscillates} \\ \textbf{Classically forbidden}: & V > E & \Rightarrow & \chi \text{ is real and } \psi \text{ varies exponentially} \end{array}} \tag{42}$$

We will not solve the problem of matching the wavefunctions between these different regions, nor will we take proper account of the spherical geometry. For now, it will be enough to note that the amplitude of the wavefunction decays exponentially in the forbidden region: $\psi \propto \exp(-\chi x)$ for a particle that penetrates a distance x into a forbidden zone in which χ has a constant value. When χ varies, we replace χx by $\int \chi(x)\, dx$ (again, the proper justification for this will come next term). Finally, we have to remember the critical fact that the probability density in quantum mechanics is proportional to $|\psi^2|$. Therefore, the relative probability of finding the particle at $r_{\rm C}$ and $r_{\rm N}$ is

$$\boxed{P_{\rm tunnel} \simeq \left| e^{-\int \chi(r)\, dr} \right|^2,} \tag{43}$$

where

$$\int_{r_{\rm C}}^{r_{\rm N}} \chi(r)\, dr = \sqrt{2m_r/\hbar^2} \int_{r_{\rm C}}^{r_{\rm N}} \left(\frac{Z_{\rm A} Z_{\rm B} e^2}{4\pi\epsilon_0 r} - E \right)^{1/2} dr. \tag{44}$$

This integral may be done by recalling $E = Z_{\rm A} Z_{\rm B} e^2/4\pi\epsilon_0 r_{\rm C}$, so that the main bracket in the integral is $Er_{\rm C}/r - E$. We can then substitute $r = r_{\rm C}\cos^2\theta$, so that $\theta = 0$ for $r = r_{\rm C}$ and $\theta \simeq \pi/2$ for $r = r_{\rm N}$ since $r_{\rm N} \ll r_{\rm C}$. We then get

$$\int_{r_{\rm C}}^{r_{\rm N}} \chi(r)\, dr = \sqrt{2m_r/\hbar^2}\, \frac{2Z_{\rm A} Z_{\rm B} e^2}{4\pi\epsilon_0 E^{1/2}} \int_0^{\pi/2} \sin^2\theta\, d\theta = \frac{1}{2}\left(E_G/E\right)^{1/2}, \tag{45}$$

(using $\sin^2\theta = (1-\cos 2\theta)/2$ to reduce the trig integral to $\pi/4$). The **Gamow energy**, E_G, is defined by

$$E_G = (\pi\alpha Z_{\rm A} Z_{\rm B})^2\; 2m_r c^2 \quad \text{where} \quad \alpha \equiv \frac{e^2}{4\pi\epsilon_0 \hbar c} \simeq \frac{1}{137}. \tag{46}$$

We then get for the tunneling probability

$$P_{\rm tunnel} \simeq \left| e^{-\int \chi(r)\, dr} \right|^2 = e^{-(E_G/E)^{1/2}}, \tag{47}$$

So particles with energies E_G and above should easily be able to tunnel through the Coulomb barrier and fuse.

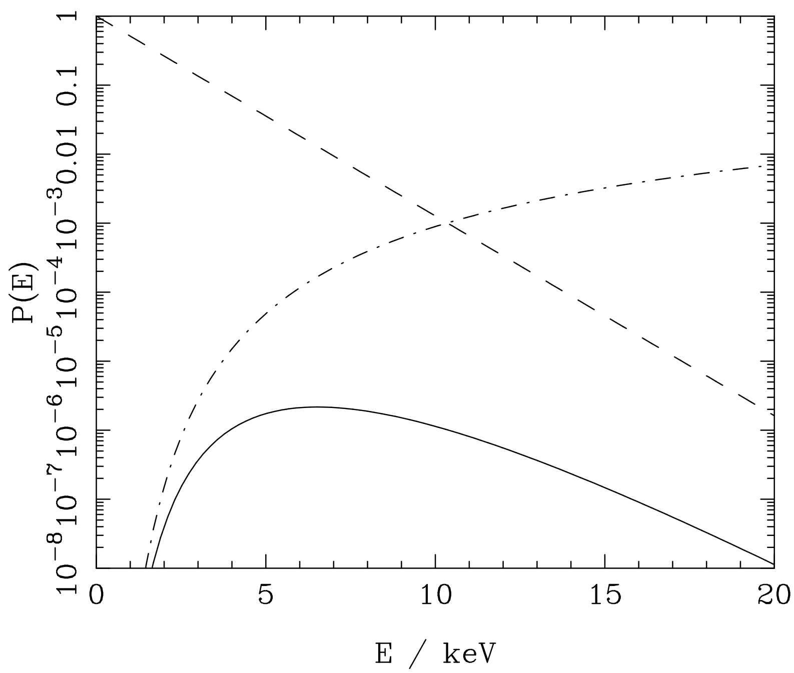

Figure 7. The probability that a proton in the Sun of energy E will be able to fuse is a product of the Boltzmann thermal probability distribution (dashed) times the tunneling probability (dot-dashed). The product (solid line) peaks in the Sun at about 6 keV, or 4 times the typical thermal energy.

But how many such particles are there? In thermal equilibrium, the particle energies are distributed according to the Boltzmann factor

$$P_{\text{Boltz}} \propto \exp(-E/kT), \tag{48}$$

where T is the temperature of the gas. So the total probability for particles to fuse – the fusion reaction rate R_{AB} – then depends on particle energies according to

$$\boxed{R_{\text{AB}} \propto P_{\text{tot}} = P_{\text{Boltz}} \times P_{\text{tunnel}} \simeq \exp\left[-E/kT - (E_G/E)^{1/2}\right].} \tag{49}$$

This rate peaks at $E = E_0 = (kT)^{2/3} E_G^{1/3}/2^{2/3}$. For the fusion of two protons, $E_G = 493$ keV, whereas in the Sun's core ($T = 1.7 \times 10^7$ K) $kT = 1.5$ keV, so that $E_0 \simeq 6.5$ keV $\simeq 4.2kT$. Notice that this automatically gives a strong temperature dependence of the reaction rate: fusion in the Sun involves protons that are on the rare exponential tail of the probability distribution, so that a small change in temperature causes a large change in the reaction rate.

6 Nuclear Reactions

Nuclear reactions fusing $4p \to {}^4\text{He}$ occur by two principal mechanisms: (1) The PP chain; (2) The CNO Cycle.

6.1 PP chain (for stars with $M \lesssim M_\odot$)

The dominant reaction branch gives deuterium initially (^{2}H):

$$\begin{aligned} p + p &\to d + e^+ + \nu_e \\ p + d &\to {}^3\mathrm{He} + \gamma \\ {}^3\mathrm{He} + {}^3\mathrm{He} &\to {}^4\mathrm{He} + 2p \end{aligned} \tag{50}$$

The sum total gives

$$4p \to {}^4\mathrm{He} + Q_{\mathrm{eff}} + 2e^+ + 2\nu_e, \tag{51}$$

where $Q_{\mathrm{eff}} = 26.2$ MeV includes the energy lost directly as photons and by the annihilation of $2e^+$ with ambient electrons in the plasma; it does not include the energy carried away as neutrinos (an additional 2%). For T close to 10^7 K, the **energy generation rate per unit mass**, ϵ, is given approximately by

$$\epsilon_{pp} \simeq 1.07 \times 10^{-8} \rho X^2 T_7^4 \ \mathrm{W\,kg^{-1}}, \tag{52}$$

where X is the mass fraction in H and $T_7 \equiv T/10^7$ K.

6.2 CNO Cycle ($M \gtrsim M_\odot$)

This is a more complex chain of nuclear reactions involving proton capture and decay:

$$\begin{aligned} p + {}^{12}\mathrm{C} &\to {}^{13}\mathrm{N} + \gamma \\ {}^{13}\mathrm{N} &\to {}^{13}\mathrm{C} + e^+ + \nu_e \\ p + {}^{13}\mathrm{C} &\to {}^{14}\mathrm{N} + \gamma \\ p + {}^{14}\mathrm{N} &\to {}^{15}\mathrm{O} + \gamma \\ {}^{15}\mathrm{O} &\to {}^{15}\mathrm{N} + e^+ + \nu_e \\ p + {}^{15}\mathrm{N} &\to {}^{12}\mathrm{C} + {}^4\mathrm{He} \end{aligned} \tag{53}$$

The original ^{12}C nucleus has gone back into circulation by the end, resulting in the cycle – it therefore acts as a catalyst for the nuclear generation of Helium. As for the pp chain, the net result is

$$4p \to {}^4\mathrm{He} + Q_{\mathrm{eff}}, \tag{54}$$

where now $Q_{\mathrm{eff}} = 23.8$ MeV (the neutrinos would increase this by 6%). For T close to 10^7 K, the energy generation rate is approximately given by

$$\epsilon_{\mathrm{CNO}} \simeq 6.54 \times 10^{-11} \rho X X_{\mathrm{CNO}} T_7^{20} \ \mathrm{W\,kg^{-1}}, \tag{55}$$

where X_{CNO} is the total mass fraction in C, N and O. The stronger temperature dependence of the CNO cycle means that it dominates over the pp chain in stars with masses slightly greater than the Sun's.

6.3 Equation of energy generation

How much energy will be generated in a shell of thickness Δr? Since ϵ gives the energy-generation rate per unit mass (in W kg^{-1}), the rate of energy generation from the shell per unit volume is $\epsilon\rho(r)$, where $\rho(r)$ is the density of the star in the shell at radius r. The volume of the shell is $4\pi r^2 \Delta r$, so the rate of energy generation in the shell that contributes to the total luminosity of the star (in W) is

$$L = (4\pi r^2 \Delta r)\epsilon\rho. \tag{56}$$

Taking the limit of infinitesimal Δr, we get the **equation of energy generation**:

$$\boxed{\frac{dL}{dr} = 4\pi r^2 \epsilon\rho.} \tag{57}$$

Here, ϵ is in general $\epsilon = \epsilon_{pp} + \epsilon_{\rm CNO}$ since both the pp-chain and the CNO cycle will contribute to the total energy generation rate at some level. Most of the energy is generated in the central core of the star where the temperature is highest.

7 Radiative diffusion

How does the energy generated in the central core of a star escape to the star's surface? The high density of atoms and ions in a star act as efficient scatterers of photons. A typical average distance a photon in the core of a star can travel before scattering – the **mean free path** – is 1 mm. The very short mean free path of a photon has two important consequences: (1) The radiation field takes on a blackbody spectral shape; (2) The radiation leaks out slowly by radiative diffusion.

Recall the diffusion equation for particles:

$$J = -D\, dn/dx, \tag{58}$$

where J is the **flux density** of particles (number of particles crossing a unit area per unit time), n is the number density of the particles, and D is the diffusion coefficient. Statistical mechanics gives

$$D = \ell v/3, \tag{59}$$

where ℓ is the mean free path for particles moving at velocity v. The same relation will hold for photons of number density n_ν where the subscript of 'ν' is added to allow for the spectrum of the radiation (photons of different frequencies ν have different number densities). We then have for the diffusion equation of photons of frequency ν

$$J_\nu = -D_\nu dn_\nu/dr \tag{60}$$

for diffusion in the radial direction outwards from the star's centre. Here

$$D_\nu = \ell_\nu c/3 \tag{61}$$

is the diffusion constant for the photons, all of which move at the speed of light c between scatterers, but have a mean free path ℓ_ν that in general will depend on ν.

The energy flux density is then given by

$$f_\nu = (h\nu)J_\nu = -D_\nu d(h\nu n_\nu)/dr = -D_\nu dU_\nu/dr, \qquad (62)$$

where each photon of frequency ν has energy $h\nu$, and the energy density of such photons is

$$U_\nu = h\nu n_\nu. \qquad (63)$$

To get the total energy flux passing through a shell at distance r, we need to integrate over all frequencies:

$$F = \int_0^\infty f_\nu \, d\nu = -\int_0^\infty D_\nu \, dU_\nu/dr \, d\nu \equiv -\bar{D} \, dU/dr \qquad (64)$$

where $U = \int U_\nu \, d\nu$ is the total energy density of radiation field, and $\bar{D}$ is a frequency-averaged diffusion coefficient.

By convention, we define an **opacity** κ by

$$\boxed{\kappa = \frac{c}{3\rho\bar{D}}.} \qquad (65)$$

The opacity has units of $\mathrm{m^2kg^{-1}}$. It describes the average scattering cross section of a photon on passing through 1 kg of material responsible for the scattering. So we can write

$$F = -\frac{c}{3\rho\kappa}\frac{dU}{dr}. \qquad (66)$$

Since the radiation is black body, $U = aT^4$, where T is the temperature of the stellar material. So we finally obtain the **equation of radiative diffusion**:

$$\boxed{F = \frac{L}{4\pi r^2} = -\frac{4ac}{3\rho\kappa} T^3 \frac{dT}{dr}.} \qquad (67)$$

At low temperatures, the gas is only partially ionized. The opacity is then dominated by bound-free absorption. At higher temperatures, when the ionization is nearly complete, the opacity is dominated by free-free absorption. The resulting frequency-averaged opacity depends on density and temperature according to (for bound-free transitions)

$$\kappa \propto \rho T^{-3.5} \qquad (68)$$

This is known as **Kramer's Law**.

Scattering by electrons is also always present. If σ_T denotes the Thomson cross section for scattering of photons by free electrons, this gives an opacity of

$$\kappa_\mathrm{es} = \frac{n_e\sigma_\mathrm{T}}{\rho} = (1+X)\frac{\sigma_\mathrm{T}}{2m_\mathrm{H}} \simeq 0.02(1+X)\ \mathrm{m^2kg^{-1}}, \qquad (69)$$

where n_e is the number density of electrons, and X is the mass fraction of hydrogen in the star.

Examples:

(a) For material of solar abundances, at $\rho = 10^4\ \mathrm{kg\,m^{-3}}$ and $T = 2\times10^6$ K, $\kappa \simeq 10\ \mathrm{m^2kg^{-1}}$. This gives a mean free path of $\ell = 1/\kappa\rho \simeq 0.1$ mm.

(b) At a higher temperature $T = 10^7$ K, the opacity decreases to $\kappa \simeq 0.1\ \mathrm{m^2kg^{-1}}$ and the photon mean free path increases to $\ell \simeq 1$ mm.

7.1 Summary of the equations of stellar structure

Define:

$\rho(r)$ = mass density at radius r
$T(r)$ = temperature at radius r
$P(r)$ = pressure at radius r
$M(r)$ = mass within radius r
$L(r)$ = luminosity escaping through a surface at r
$F(r)$ = flux of radiation escaping through a surface at r: $F(r) = L(r)/(4\pi r^2)$
$\epsilon(r)$ = nuclear energy generation rate per unit mass at r
$\kappa(r)$ = opacity of stellar material at r

$$\boxed{\begin{array}{l} \text{(1) Equation of Continuity}: \ \dfrac{dM(r)}{dr} = 4\pi r^2 \rho(r) \\[2ex] \text{(2) Equation of Hydrostatic Equilibrium}: \ \dfrac{1}{\rho}\dfrac{dP}{dr} = -\dfrac{GM(r)}{r^2} \\[2ex] \text{(3) Equation of Energy Generation}: \ \dfrac{dL}{dr} = 4\pi r^2 \epsilon\rho \\[2ex] \text{(4) Equation of Radiative Diffusion}: \ F = \dfrac{L}{4\pi r^2} = -\dfrac{4ac}{3\rho\kappa}\, T^3 \dfrac{dT}{dr} \end{array}} \qquad (70)$$

We would like to solve these subject to the perfect gas law $P(r) = [\rho(r)/\bar{m}]\, kT(r)$ and simple power-law scalings for the nuclear energy generation rate and the opacity:

$$\epsilon = \epsilon_0 \rho T^\alpha; \qquad \kappa = \kappa_0 \rho^\beta / T^\gamma. \qquad (71)$$

8 Stars on the main sequence

8.1 Simple approach

Before even thinking about how to solve the equations of stellar structure, it is important to emphasise that the key element in understanding the properties of stars is that the energy emerges via diffusion, and that this allows the systematic properties of the main sequence to be understood very simply.

Consider for simplicity a star of uniform density. Introducing a density profile only changes things by some dimensionless constants. In **virial equilibrium** the potential and kinetic energies are of the same order: $GM^2/R \sim nkTR^3$, so that $T \propto M/R$. Now, if photons suffer Thomson scattering inside the star with mean free path λ ($= (n\sigma_{\rm T})^{-1} \propto R^3/M$), the typical time for escape is that taken to diffuse a distance R, $t \sim R^2/c\lambda$ (because the distance diffused in a random walk of N scatterings is $\lambda\sqrt{N}$). The luminosity corresponds to radiating away the energy content $E \propto R^3 T^4$ over this time, so $L \propto RT^4 \lambda \propto R^4 T^4 / M$. Using the virial temperature above, this is just $L \propto M^3$. Putting in the dimensional constants gives

$$\boxed{L \simeq \frac{G^4 m_p^5}{\hbar^3 c^2 \sigma_{\rm T}}\, M^3.} \qquad (72)$$

In deducing this expression, we have made no assumptions at all about the internal nuclear reactions, which is quite remarkable.

Another way of writing L is to say that the surface radiates like a black body, $L \propto R^2 T_{\rm eff}^4$. If we assume that the virial scaling also applies to $T_{\rm eff}$, $T_{\rm eff} \propto M/R$, we additionally deduce $T_{\rm eff} \propto M^{1/2}$, $R \propto M^{1/2}$. These relations in fact apply reasonably well to stars several times more massive than the sun. They fail for the sun (i) because some energy is transported via convection; (ii) because cooler stars have their opacity dominated by ionic processes: free–free and bound–free scattering, where **Kramer's opacity** applies.

8.2 Solving for stellar structure

Contrast the simple approach with a more mathematical approach using the equations of stellar structure. The key idea here is one of **dimensionless variables**: the differential equations cannot be solved exactly, and must be put on a computer. This means that the quantities involved must be pure numbers, whereas M, R etc. have units. The solution is to make dimensionless versions of these variables by dividing by the natural unit supplied by the star itself. In some cases, this is the value at the surface, so we define

$$\tilde{r} \equiv r/R; \quad \tilde{M}(\tilde{r}) \equiv M(r)/M; \quad \tilde{L}(\tilde{r}) \equiv L(r)/L; \tag{73}$$

but the natural values of T, P and ρ are the central values:

$$\tilde{T} \equiv T/T_{\rm C}; \quad \tilde{P} \equiv P/P_{\rm C}; \quad \tilde{\rho} \equiv \rho/\rho_{\rm C}; \tag{74}$$

Rewriting the equations of stellar structure in these units creates dimensionless combinations of the reference values. For example, consider the equations of continuity and hydrostatic equilibrium:

$$\frac{dM(r)}{dr} = 4\pi r^2 \rho(r) \quad \Rightarrow \quad \frac{d\tilde{M}(\tilde{r})}{d\tilde{r}} = \left(\frac{R^3 \rho_{\rm C}}{M}\right) 4\pi \tilde{r}^2 \tilde{\rho}(\tilde{r}) \tag{75}$$

and

$$\frac{1}{\rho}\frac{dP}{dr} = -\frac{GM(r)}{r^2} \quad \Rightarrow \quad \frac{1}{\tilde{\rho}}\frac{d\tilde{P}}{d\tilde{r}} = -\left(\frac{GM\rho_{\rm C}}{RP_{\rm C}}\right)\frac{\tilde{M}(\tilde{r})}{\tilde{r}^2} \tag{76}$$

The terms in brackets must be some dimensionless constants, and this is very important: *we only need to solve the equations once*. Suppose we integrate the differential equations starting from the centre, at $\tilde{r} = 0$. Here, $\tilde{M} = \tilde{L} = 0$ and $\tilde{T} = \tilde{P} = \tilde{\rho} = 1$. At the edge, we want to have reached $\tilde{M} = \tilde{L} = 1$ and $\tilde{P} = \tilde{\rho} = 0$. This will only happen if we choose the right values for dimensionless combinations like $(R^3 \rho_{\rm C}/M)$.

Solving the equations of stellar structure to deduce the correct dimensionless numbers can be complicated, and we can't go into much detail. A simple start is to choose trial forms for the desired functions that satisfy the boundary conditions. For example, in the equation of continuity we have to satisfy

$$\frac{d\tilde{M}(\tilde{r})}{d\tilde{r}} = \left(\frac{R^3 \rho_{\rm C}}{M}\right) 4\pi \tilde{r}^2 \tilde{\rho}(\tilde{r}), \tag{77}$$

which requires $d\tilde{M}(\tilde{r})/d\tilde{r}$ to vanish at both $d\tilde{r} = 0$ and $d\tilde{r} = 1$. A simple form that achieves this is $d\tilde{M}(\tilde{r})/d\tilde{r} = A\tilde{r}^2(1-\tilde{r})$, and $\tilde{M}(1) = 1$ gives $A = 12$. Solving the equation then gives

$\tilde{\rho} = (1 - \tilde{r})$ and $(R^3 \rho_C / M) = 12/4\pi$. Proceeding similarly with the other equations gives initial guesses for the other dimensionless parameters. We can then try to integrate the equations numerically. At the first pass, the exact boundary conditions will not be satisfied, but we will have a more accurate approximation for $\tilde{\rho}(\tilde{r})$ etc. The parameter guesses can be improved, and then the whole process iterated.

The end result of such a calculation comes quite close to more sophisticated calculations of the interior conditions in the Sun, which are reproduced below in Table 4.

Table 4 The Solar interior

$r/R_\odot$	$M(r)/M_\odot$	$\rho/\bar{\rho}$	$T/T_{\rm eff}$	$L(r)/L_\odot$
0	0	115	2,740	0
0.02	0.001	104	2,700	0.01
0.09	0.057	68	2,360	0.36
0.22	0.399	20.4	1,520	0.97
0.32	0.656	6.9	1,110	1.00
0.52	0.908	0.75	650	1.00
0.71	0.977	0.13	390	1.00
0.91	0.999	1.38×10^{-2}	89	1.00
0.99	1.000	1.82×10^{-4}	0.76	1.00
0.999	1.000	9.15×10^{-7}	0.23	1.00
1.000	1.000	1.55×10^{-7}	0.10	1.00

8.3 Mass–luminosity relation

So, having found the correct value of a number like $(R^3 \rho_C / M)$, it must apply for any star, so we deduce how the size and density of a star will change if we alter its mass: $\rho_C \propto M/R^3$. Similarly, from hydrostatic equilibrium, we get $P_C \propto M \rho_C / R$, so that

$$P_C \propto M^2 / R^4, \tag{78}$$

as we had before.

The perfect gas law says $P_C \propto \rho_C T_C$, so the four equations of stellar structure give four constant combinations involving the five quantities M, L, R, ρ_C, T_C:

$$\boxed{\frac{\rho_C R^3}{M}; \quad \frac{\rho_C R^4 T_C}{M^2}; \quad \frac{\rho_C M T_C^\alpha}{L}; \quad \frac{\rho_C^\beta M L}{T_C^{4+\gamma} R^4}} \tag{79}$$

These can be manipulated to find out how any given quantity depends on mass. After a little (certainly non-examinable) effort, the luminosity-mass relation comes out as

$$\boxed{L \propto M^{\frac{(3+\alpha)(3+\gamma-\beta)+(2+\alpha)(3\beta-\gamma)}{(3+\alpha)-\gamma+3\beta}}.} \tag{80}$$

This doesn't look very simple, but notice that for Thompson opacity ($\beta = \gamma = 0$), this gives $L \propto M^3$, independent of α (as we had earlier). For large α, we get approximately $L \propto M^{3+2\beta}$, independent of the temperature dependence of the opacity (so $L \propto M^5$ in the large-α Kramers case).

These simple predictions for the scaling of properties of stars with their mass along the main sequence can be contrasted with the actual properties, as summarised below in Table 5. The overall behaviour is quite close to what we have predicted theoretically: between types M5 and O5, the effective power-law scaling of the stellar properties is $L \propto M^{3.2}$; $R \propto M^{0.67}$; $T_{\rm eff} \propto M^{0.44}$.

Table 5 Physical properties of main-sequence stars

Spectral type	$T_{\rm eff}/{\rm K}$	$M/M_\odot$	$R/R_\odot$	$\bar\rho/\bar\rho_\odot$	$\log_{10}(L/L_\odot)$	M_V
O5	38,000	60.0	12.0	0.035	5.90	−5.7
B0	30,000	17.5	7.4	0.043	4.72	−4.0
B5	16,400	5.9	3.9	0.099	2.92	−1.2
A0	10,800	2.9	2.4	0.21	1.73	+0.6
A5	8,620	2.0	1.7	0.41	1.15	+1.9
F0	7,240	1.6	1.5	0.47	0.81	+2.7
F5	6,540	1.3	1.3	0.59	0.51	+3.5
G0	5,920	1.05	1.1	0.79	0.18	+4.4
G5	5,610	0.92	0.92	1.18	−0.10	+5.1
K0	5,240	0.79	0.85	1.29	−0.38	+5.9
K5	4,410	0.67	0.72	1.79	−0.82	+7.4
M0	3,920	0.51	0.60	2.36	−1.11	+8.8
M5	3,120	0.21	0.27	10.70	−1.96	+12.3

This robust prediction of a steep power law dependence of luminosity on mass is the key theoretical fact underlying the existence of the main sequence. It also governs evolution off the main sequence. Stars can shine only for a time τ in which they convert a significant quantity of their mass into radiation. The total energy radiated is $L\tau$, and this must be of order the nuclear efficiency ($\sim 10^{-3}$) of Mc^2:

$$\boxed{\tau \propto M/L \propto M^{-(2+2\beta)}.} \tag{81}$$

In other words, although massive stars have more fuel, they use it up more quickly, and are the first to evolve away from the main sequence. This lifetime is about 9 Gyr for the Sun, which is therefore about half-way through its lifespan.

8.4 Convective instability

We now need to come clean about a complication that has been neglected so far: energy is not always transmitted in stars by diffusion of photons. This is because there is a large radial temperature gradient in stars, and we can immediately see a potential problem: hot fluid rises. In other words, there is a possibility of **convection** in which the central parts of stars 'boil' and send streamers of hot material outwards, thus transporting energy far more efficiently than diffusion.

Consider a small parcel of fluid in the star, of mass ΔM; this starts in equilibrium, so that the pressure forces acting on it balance gravity, according to the equation of hydrostatic equilibrium. Now suppose that this parcel is displaced upwards (to larger r). In order to stay in hydrostatic equilibrium at its new radius, $r + dr$ it would need to have the same density as the rest of the fluid at that radius: $\rho(r + dr)$. If it is denser, it will sink, but if it is lighter it will float to larger r and the motion will be unstable. Which of these happens depends on the the magnitude of dT/dr. Let the parcel move a distance Δr, so that the surrounding density is now

$$\rho = \rho_0 + (d\rho/dr)\Delta r, \tag{82}$$

where ρ_0 is the starting density of star and parcel. Note that $d\rho/dr$ is negative. If the change in density of the parcel is larger in magnitude than this, convection will happen:

$$|\Delta\rho_{\rm parcel}| > |(d\rho/dr)\Delta r| \quad \Rightarrow \quad \text{convection} \tag{83}$$

(we use the modulus to avoid worries about signs).

The change in density of the parcel is governed by the fact that the fluid inside suffers **adiabatic** change: there is no time to exchange heat with the surroundings. Therefore, the density and pressure are connected by the relation

$$P \propto \rho^\gamma \quad \Rightarrow \quad \Delta \ln \rho = \frac{1}{\gamma}\Delta \ln P, \tag{84}$$

where γ is the ratio of specific heats: $\gamma = 5/3$ for a monatomic gas. The gas is also perfect, so $P = nkT \propto \rho T$. This means that the general gradients in P, T and ρ are related:

$$\frac{d\ln\rho}{dr} = \frac{d\ln P}{dr} - \frac{d\ln T}{dr}. \tag{85}$$

Therefore, since we want $\Delta\ln\rho_{\rm parcel} < (d\ln\rho/dr)\Delta r$ (both changes being negative), the criterion for convection is

$$\frac{1}{\gamma}\Delta\ln P < \left(\frac{d\ln P}{dr} - \frac{d\ln T}{dr}\right)\Delta r, \tag{86}$$

and taking the limit of infinitesimal Δr gives

$$\boxed{\left|\frac{d\ln T}{dr}\right| > \left(1 - \frac{1}{\gamma}\right)\left|\frac{d\ln P}{dr}\right| \quad \Rightarrow \quad \text{convection.}} \tag{87}$$

This makes sense in terms of everyday experience: it shows that the temperature inversions (with hot air above cold) often experienced on clear winters days are stable.

The question is thus whether there could be one or more **convective zones** inside the star, and this depends on the detailed solutions of the structure for diffusive energy transport. In practice, convection is important in the cores of stars above a few $M_\odot$: this is because CNO burning is important in these hotter stars. CNO is so temperature sensitive that one can think of the energy generation as occurring almost in a central delta function, and the central temperature gradient becomes very high. Most stars also become convective in their outer parts: for the Sun, convection sets in at a radius of about $0.7R_\odot$. This complication of convection does not greatly alter the general picture of the scaling of stellar properties with mass, but it certainly makes detailed study of stars a much messier problem.

9 Stellar evolution and the giant branch

We deduced earlier that stars will have a finite lifetime, which is about 9 Gyr for the Sun, and scales very roughly as M^{-3}. After this time, a significant proportion of the initial nuclear fuel in the core will have been fused, and the star can no longer shine on the main sequence. In this course, we will not look at the subsequent evolution in much detail, since it is the subject of a module in Astrophysics 4. In outline, the main events are as seen in Figure 3:

(1) The giant branch. Once hydrogen is exhausted in the core, it continues to burn in a shell around the core. The helium-rich core contracts until it is supported by **electron degeneracy pressure**, which we will study below. This phase is associated with an expansion in size of the outer parts of the star by roughly a factor 10 to form a red giant. In the initial phases of the process, the turnoff stars are subgiants and become somewhat cooler at constant luminosity, before increasing size and luminosity greatly at roughly constant temperature as they ascend the giant branch proper. It would be nice if there existed a simple reason why giant stars increase in size so much, but one of the embarrassments of the subject is the lack of any such explanation. A common and deceptively simple argument is that it is all to do with the virial theorem, where the kinetic and potential energies must satisfy $2K + V = 0$ in equilibrium: since the contraction of the core causes V to become more negative, the balance can be restored if the outer parts expand to become less tightly bound. This argument works because the expansion is approximately isothermal, but it misses out the fact that the expansion is associated with a vast increase in nuclear energy output. Table 6 below gives characteristic sizes and luminosities for red giants. Note that these depend largely on the spectral type, and not on the mass: stars of a wide range of masses follow similar tracks up the HR diagram, becoming redder and more luminous as they grow. This whole process is rather fast: the subgiant phase lasts roughly 10% of the main-sequence lifetime, and the red-giant phase is roughly 5% of the main-sequence lifetime.

(2) The horizontal branch. As red giants burn more hydrogen, the mass and temperature of the helium core increases, until it too can fuse. Stars then follow their usual pattern of defying intuition by reducing luminosity somewhat but becoming smaller and hotter. They end up more or less at the top end of the original main sequence, which is reassuring, since we argued that this was largely independent of the details of nuclear energy generation for stars where the energy originated in the core.

(3) The AGB. History now repeats itself as helium in the core is exhausted. Shell burning returns, with both a helium-burning shell and an outer hydrogen-burning shell, and the star ascends the giant branch once more, now termed the **asymptotic giant branch**, although its location in the HR diagram is rather similar to the red giant branch.

(4) Late stages of evolution. Stars that come to the end of the AGB stage contain high-density cores enhanced in heavy elements. The normal behaviour is for the remaining outer layers to be lost, in a greatly exaggerated analogue of the Solar wind, leaving behind a **white dwarf**, which simply cools via radiation without generating further nuclear energy. These stars lie below the main sequence, being of very low luminosity for their temperature.

Table 6 Radii and luminosities of red giants

	G0	G5	K0	K5	M0	M5
$\log_{10}(R/R_\odot)$	0.8	1.0	1.2	1.4	1.6	1.9
$\log_{10}(L/L_\odot)$	1.5	1.7	1.9	2.3	2.6	3.0

10 Degeneracy pressure in stars

10.1 Equation of state

The uncertainty principle imposes a limit on the range of position that a particle with a given range of momentum may occupy. In 1D,

$$(\Delta x)(\Delta p_x) \gtrsim \hbar. \tag{88}$$

Since at most two electrons, for the two spin states, can occupy a given quantum state in space, there is a limit to how tightly packed electrons of a given range of momenta may be made. This results in **electron degeneracy pressure**. We may estimate the magnitude of the pressure for a given electron density ρ_e as follows.

Microscopically, pressure arises from flux of momentum. If the number density of electrons is n_e, then the flux of electrons in the x direction is just the number of electrons crossing unit area per unit time, or $n_e v_x$. The pressure is then approximately

$$P_e \simeq p_x \, n_e \, v_x, \tag{89}$$

where p_x is the momentum of the particles.

(1) non-relativistic case: if the particles are non-relativistic, then $v_x = p_x/m_e$, so

$$P_e \simeq n_e \, p_x^2/m_e = \rho_e \, p_x^2 m_e^{-2} \qquad (\rho_e = m_e n_e) \tag{90}$$

By the uncertainty principle, the minimum volume these particles may be squeezed into is

$$(\Delta x)^3 \simeq (\hbar/p_x)^3, \tag{91}$$

corresponding to a mass density $\rho_e \simeq m_e/(\Delta x)^3 = m_e(p_x/\hbar)^3$, so that $p_x \simeq \hbar(\rho_e/m_e)^{1/3}$, and we obtain

$$P_e \simeq \rho_e p_x^2/m_e^2 \simeq \hbar^2 \rho_e^{5/3} m_e^{-8/3}. \tag{92}$$

(2) relativistic case: if the particles are highly relativistic, then their flux will simply be $n_e c$. In this case

$$P_e \simeq p_x n_e c = p_x c \rho_e/m_e \tag{93}$$

As before, the uncertainty principle gives $p_x \simeq \hbar(\rho_e/m_e)^{1/3}$, so we have now

$$P_e \simeq \hbar c \rho_e^{4/3} m_e^{-4/3}. \tag{94}$$

These expressions relating p_e and e may be made exact by starting with Schrödinger's equation. We previously considered the idea of counting the number of states that can exist in a box of volume $V = L^3$, and we showed that this corresponded to a density of states in k space:

$$dN = g \, \frac{V}{(2\pi)^3} \, d^3k; \qquad (d^3k \equiv dk_x dk_y dk_z), \tag{95}$$

where g is a degeneracy factor for spin *etc.* This expression is nice because it is **extensive** (number of states $\propto V$) and hence the number density of particles in the box is independent of V:

$$\boxed{n = g\,\frac{1}{(2\pi\hbar)^3}\int_0^\infty f(p)\;d^3p,} \tag{96}$$

where we have converted from **k space** to **momentum space** using $p = \hbar k$ as before; $f(p)$ is the **occupation number** of the mode – i.e. the number of particles in the box with that particular wavefunction. For **bosons** (particles such as photons), the occupation number is unrestricted. Next term, it will be shown that the above formalism applies for black-body radiation treated as a gas of photons, for which the occupation number is

$$f_{\rm BB} = \left(\exp[\hbar\omega/kT] - 1\right)^{-1}. \tag{97}$$

For **fermions** (particles such as electrons, whose spin angular momentum is $\hbar/2$), things are completely different, and the **Pauli exclusion principle** says that

$$\boxed{f(p) \leq 1.} \tag{98}$$

This criterion immediately imposes a restriction on how dense an electron gas can be before it has to be treated in a manner very different from the classical one. Normally, the distribution of momenta would be treated as a **Maxwellian distribution**, in which each component of velocity has a Gaussian distribution with standard deviation σ:

$$d\mathrm{Prob} = \frac{1}{(2\pi)^{3/2}\sigma^3}\exp(-v^2/2\sigma^2)\;d^3v, \tag{99}$$

where v is the particle velocity. The dispersion in velocities, σ, is related to the temperature by equipartition of energy: $m\sigma^2/2 = kT/2$. We can convert this to a number of particles in a given range of momentum space, by multiplying by the total number density, n, and using $p = mv$:

$$dn = \frac{n}{(2\pi mkT)^{3/2}}\exp(-p^2/2mkT)\;d^3p. \tag{100}$$

Comparing with the general expression for density, we deduce that there is a **critical density**, at which the classical law would yield $f > 1$:

$$\boxed{n_{\rm crit} = \frac{g(mkT)^{3/2}}{(2\pi)^{3/2}\hbar^3}.} \tag{101}$$

A simpler way of expressing this is $\hbar/(mkT)^{1/2} \simeq n_{crit}^{-1/3}$: the momentum dispersion is $p^2 \sim mkT$, so this just says that the typical interparticle spacing cannot be smaller than the limit set by the uncertainty principle. If we take a fixed density, the gas will be in the classical regime for high T, but quantum effects become important as we go towards zero T. This is illustrated in Figure 8. In the limit of zero temperature, states are occupied only up to the **Fermi momentum**, $p_{\rm F}$:

$$\boxed{n = g\,\frac{1}{(2\pi\hbar)^3}\int_0^{p_{\rm F}} d^3p = g\,\frac{1}{(2\pi\hbar)^3}\,\frac{4\pi}{3}\,p_{\rm F}^3.} \tag{102}$$

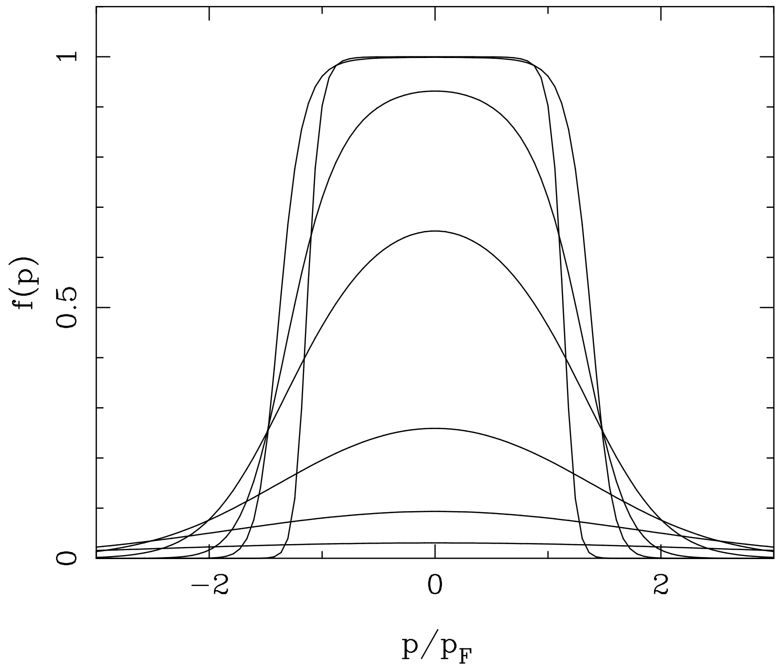

Figure 8. The occupation number for a gas of fermions as a function of their density relative to the critical density, ranging from $n/n_{\rm crit} = 0.03$ to $n/n_{\rm crit} = 30$. For the lowest densities (or highest temperatures), we have almost exactly the classical Maxwellian velocity distribution. For densities well above critical, the occupation number tends to a 'top-hat' distribution: unity for momenta less than the Fermi momentum, and zero otherwise.

The thermodynamic properties of the electron gas can be worked out by using simple properties of each state, plus symmetry. Each mode has a definite value of momentum, and has some energy that depends on the momentum, $\epsilon(p)$. It therefore has an effective mass of ϵ/c^2, and an effective velocity given by momentum divided by mass, or $v = pc^2/\epsilon$. The flux density of momentum in a given direction (x, say) is just $p_x\, n_e\, v_x$, and so we can deduce the pressure in the x direction, which must be equal to the pressure in any direction:

$$\boxed{P = g\,\frac{1}{(2\pi\hbar)^3}\int_0^\infty \frac{p^2c^2}{3\epsilon}\; f(p)\; 4\pi\; p^2\; dp.} \tag{103}$$

Here, we have used

$$\int p_x^2\; dp_x\, dp_z\, dp_z = \frac{1}{3}\int (p_x^2 + p_y^2 + p_z^2)\; dp_x\, dp_z\, dp_z = \frac{1}{3}\int p^2\; 4\pi p^2\; dp. \tag{104}$$

This can be regarded as obvious by isotropy, or one could use spherical polars in momentum space to get the same answer. Consider now the extreme non-relativistic and relativistic limits of the exact expressions for the pressure, in the zero-temperature case. Hereafter, we will explicitly adopt $g = 2$ for the two electron spin states.

(1) non-relativistic case: $\epsilon = m_e c^2$, or $v = p/m_e$, so

$$P_e = \frac{8\pi}{(2\pi\hbar)^3}\;\frac{1}{3}\int_0^{p_{\rm F}} (p^2/m_e)\; p^2\; dp = \frac{1}{15\pi^2\hbar^3 m_e} p_{\rm F}^5. \tag{105}$$

Since $n_e = (3\pi^2\hbar^3)^{-1} p_{\rm F}^3$ from above, we have, using $\rho_e = m_e n_e$,

$$p_{\rm F} = \left(\frac{3\pi^2\hbar^3\,\rho_e}{m_e}\right)^{1/3} \tag{106}$$

and so

$$P_e = \frac{1}{15\pi^2\hbar^3}\left(3\pi^2\hbar^3\right)^{5/3} \rho_e^{5/3} m_e^{-7/3} \equiv K_1 \rho_e^{5/3} \tag{107}$$

where

$$K_1 = \frac{\pi^2\hbar^2}{5m_e^{8/3}}\left(\frac{3}{\pi}\right)^{2/3}. \tag{108}$$

(2) relativistic case: $\epsilon = pc$, or $v = c$, so

$$\begin{aligned} P_e &= \frac{8\pi}{(2\pi\hbar)^3}\,\frac{1}{3}\int_0^{p_{\rm F}} pc\,p^2\,dp = \frac{c}{12\pi^2\hbar^3} p_{\rm F}^4 \\ &= \frac{c}{12\pi^2\hbar^3}\left(\frac{3\pi^2\hbar^3\rho_e}{m_e}\right)^{4/3} \equiv K_2 \rho_e^{4/3} \end{aligned} \tag{109}$$

where

$$K_2 = \frac{\pi\hbar c}{4m_e^{4/3}}\left(\frac{3}{\pi}\right)^{1/3}. \tag{110}$$

Note that for both cases, the electron degeneracy pressure P_e depends only on the electron density ρ_e. For partially degenerate gas, however, P_e will depend on both ρ_e and T. Note also that the degeneracy pressure depends on a negative power of m_e: the degeneracy pressure from the protons that are also present is therefore negligible, and this justifies us having considered only the electron component.

10.2 The lower limit to the main sequence

As a first application of degeneracy pressure, we can calculate how massive a star needs to be before it can shine. Stars start life as collapsing gas clouds, and will radiate energy ('cool') on the **Kelvin timescale**, which comes from assuming that the body must radiate away its binding energy:

$$L \sim \frac{\partial}{\partial t}\left(\frac{GM^2}{R}\right) \quad \Rightarrow \quad t_{\rm K}\left(\equiv \frac{R}{|\dot{R}|}\right) \sim \frac{GM^2}{RL}. \tag{111}$$

As we saw earlier, this gives about 3×10^7 years for the Sun, showing the need for a non-gravitational source of energy. The collapse will cause the star to heat up, as may be seen from the virial theorem: (potential energy) + $2\times$ (kinetic energy) = 0. In terms of the number of particles N and the mean temperature T this gives

$$2 \times \frac{3}{2} NkT = \frac{\alpha GM^2}{R} \quad \Rightarrow \quad T = \frac{\alpha GM^2}{3NkR} \propto \frac{M}{R} \tag{112}$$

(where α is a constant of order unity; $\alpha = 3/5$ for a uniform-density sphere). So, normally collapse will cause the temperature to rise until nuclear burning can begin. The only way this can be prevented from happening is if electron degeneracy sets in first and halts the collapse. As we have seen, the condition for the onset of degeneracy is just that the electron interparticle spacing becomes small enough for the uncertainty principle to matter:

$$n_e^{-1/3} p \lesssim \hbar \quad \Rightarrow \quad \text{degeneracy}, \tag{113}$$

where n_e is electron number density and p momentum. We can roughly estimate the condition for this to occur by assuming the body to be of uniform density and temperature. Equipartition says that $p^2/2m_e = kT/2$, so $p^2 = m_e kT$. We will combine this with the virial relation

$$kT = \frac{GM^2}{5NR} = \frac{GMm_p}{5R}, \tag{114}$$

where we have assumed that the star is mainly hydrogen, so that $N = M/m_p$. The condition that the star is *non*-degenerate is then

$$\left(\frac{m_p}{M}\frac{4\pi R^3}{3}\right)^{1/3} (m_e kT)^{1/2} \gtrsim \hbar. \tag{115}$$

This equation mixes T, M and R, but we can eliminate R by using the equation for T again. This gives

$$(m_e kT)^{1/2} \left(\frac{m_p}{M}\frac{4\pi}{3}\right)^{1/3} \frac{GMm_p}{5kT} \gtrsim \hbar, \tag{116}$$

or

$$kT \lesssim \left(\frac{4\pi}{3}\right) \frac{G^2 m_p^{8/3} m_e}{25} M^{4/3}. \tag{117}$$

For a given mass, we can therefore express in practical units an estimate of the maximum temperature that can be attained before degeneracy becomes important:

$$\boxed{T_{\max} \simeq 10^{8.5} \left(\frac{M}{M_\odot}\right)^{4/3} \text{ K.}} \tag{118}$$

Since fusion reaction rates drop exponentially for $T \lesssim 10^6$ K, a minimum mass for normal stars on the range $0.1 - 0.01\, M_\odot$ might be expected. The accepted figure for this limit when realistic calculations are performed is in fact $0.08 \pm 0.01 M_\odot$. Objects much less massive than this will generate energy only gravitationally, and will therefore cool to a state of virtual invisibility on the Kelvin timescale. Such objects constitute one of the holy grails of stellar astrophysics, and are known as **brown dwarfs**. The search for these objects has been somewhat controversial, since it is very hard to measure the masses of faint cool stars directly. However, there are now a number of good candidates for objects that sit on this interesting transition between star and planet.

11 White dwarfs and the Chandrasekhar limit

Degeneracy pressure is also important in more massive stars as they evolve off the main sequence. We have appealed to the effect to hold up the cores once hydrogen burning ceases, and the same effect supports the white dwarfs that are formed at the very end of the evolutionary sequence. The maximum mass for a white dwarf (the **Chandrasekhar mass**) is about 1.4 $M_\odot$, and this arises roughly as follows.

Consider the energy density for zero temperature in the limit that the electrons are all highly relativistic (*i.e.* the Fermi momentum is $p_{\rm F} \gg m_e c$, which will be true at high enough density, since the typical spacing for degenerate electrons is $\sim \hbar/p_{\rm F}$). The energy density can be calculated in exactly the same way as we obtained the number density and pressure: by integrating over momentum space, putting in the energy per mode, $\epsilon(p)$:

$$\boxed{U = g\,\frac{1}{(2\pi\hbar)^3}\int_0^\infty \epsilon(p)\ f(p)\ 4\pi\ p^2\,dp.} \tag{119}$$

In the zero-temperature limit, we set $f(p) = 1$ up to the Fermi momentum. In the relativistic limit, $E = pc$, so the energy density is

$$U_e = g\,\frac{1}{(2\pi\hbar)^3}\,4\pi c p_{\rm F}^4/4, \tag{120}$$

which can be expressed in terms of the number density as

$$U_e = \tfrac{3}{4}(3\pi^2)^{1/3}\,\hbar c\, n_e^{4/3}. \tag{121}$$

In the opposite limit of highly non-relativistic electrons, the energy density is similarly

$$U_e = \frac{3\hbar^2}{10 m_e}\,(3\pi^2)^{2/3}\,n_e^{5/3}. \tag{122}$$

Suppose the star is in the relativistic regime, so that the total kinetic energy of the electrons is $E_K \propto U_e V \propto n_e^{4/3} V \propto M^{4/3}/r$. The gravitational energy is proportional to $-M^2/r$, so that the total energy can be written as

$$E_{\rm tot} = (AM^{4/3} - BM^2)/r. \tag{123}$$

There thus exists a critical mass where the two terms in the bracket are equal. If the mass is smaller, then the total energy is positive and will be reduced by making the star expand until the electrons reach the mildly relativistic regime and the star can exist as a stable white dwarf. If the mass exceeds the critical value, the binding energy increases without limit as the star shrinks: gravitational collapse has become unstoppable.

To find the exact limiting mass, we need to find the coefficients A and B above, where the argument has implicitly assumed the star to be of constant density. In this approximation, the kinetic and potential energies are

$$E_K = \left(\frac{243\pi}{256}\right)^{1/3}\frac{\hbar c}{r}\left(\frac{M}{\mu m_p}\right)^{4/3}; \qquad E_V = -\frac{3GM^2}{5r}, \tag{124}$$

where the mass per electron is μm_p. This estimate of the critical mass is

$$\boxed{M_{\rm crit} = \frac{3.7}{\mu^2}\left(\frac{\hbar c}{G}\right)^{3/2} m_p^{-2} \simeq \frac{7}{\mu^2}\,M_\odot.} \tag{125}$$

For a star near the end of the evolutionary track, much of the initial fuel has been burned to iron, so that $\mu \simeq 2$. In fact, more exact calculations show that the critical mass in this case is about 1.4 $M_\odot$.

11.1 Sizes and densities of white dwarfs

White dwarfs are extremely compact objects, and become more so as their mass increases. To find their radius, we assume that the electrons are non-relativistic, so that their energy density is $\propto n_e^{5/3}$ so that the total kinetic energy is $E_K = CM^{5/3}R^{-2}$. As before, we write $E_V = -BM^2/R$, and the total energy is the sum of these terms. The equilibrium radius is where $dE/dr = 0$, so that we have

$$R = (2C/B)M^{-1/3}, \tag{126}$$

and the star contracts as its mass is increased. To put numerical values into this, the electron number density is

$$n_e = (M/\mu m_p) \,/\, (4\pi R^3/3), \tag{127}$$

and we have the energy density in terms of the number density

$$U_e = \frac{3\hbar^2}{10m_e}\,(3\pi^2)^{2/3}\, n_e^{5/3}, \tag{128}$$

which gives

$$C = \frac{3\hbar^2}{10m_e}\,(3\pi^2)^{2/3}\,(\mu m_p)^{-5/3}\,(4\pi/3)^{5/3}. \tag{129}$$

Combined with our previous $B = 3G/5$, we get

$$R = \frac{3}{2}\left(\frac{3\pi^2}{2}\right)^{1/3} \frac{\hbar^2}{Gm_e(\mu m_p)^{5/3}}\, M^{-1/3}. \tag{130}$$

Expressed in terms of the Chandrasekhar mass, this is

$$\boxed{R = 3\sqrt{\pi/5}\frac{\hbar}{\mu m_p m_e}\left(\frac{\hbar}{cG}\right)^{1/2}\left(\frac{M}{M_{\rm crit}}\right)^{-1/3} \quad\simeq\quad 5975\left(\frac{M}{M_{\rm crit}}\right)^{-1/3}\ \text{km},} \tag{131}$$

where the last figure assumes $\mu = 2$. In other words, a Chandrasekhar-mass white dwarf is about the size of the Earth. This is an extremely high density, which we can evaluate as

$$\rho = \frac{M}{4\pi R^3/3} = \frac{125\mu m_p m_e^3 c^3}{192\pi^2\hbar^3}\left(\frac{M}{M_{\rm crit}}\right)^2 \quad\simeq\quad 10^{9.6}\left(\frac{M}{M_{\rm crit}}\right)^2\ \text{kg m}^{-3}. \tag{132}$$

These expressions contain tedious numerical coefficients, which come from working out the constant-density model. Even if the expressions were simpler, there would be no point in memorising the exact coefficients, because the constant-density model is not exact. If we wanted to do better, we would need to solve for the internal structure of a white dwarf. This is not so hard, since we know that the material of the white dwarf obeys the **equation of state**

$$P = K\rho^\gamma, \tag{133}$$

where $\gamma = 5/3$ in the non-relativistic case. This is known as a polytropic equation of state, and stars described by it are called **polytropes**. We can insert this in the equations of hydrostatic equilibrium and mass conservation:

$$-\frac{1}{\rho}\frac{dP}{dr} = \frac{GM(r)}{r^2}; \qquad \frac{dM(r)}{dr} = 4\pi r^2 \rho(r), \tag{134}$$

giving

$$4\pi r^2 \rho(r) = -\frac{d}{dr}\left(\frac{r^2}{\rho G}\frac{dK\rho^\gamma}{dr}\right). \tag{135}$$

As before, we can place this in dimensionless form by defining $\tilde{r} \equiv r/R$ and $\tilde{\rho} \equiv \rho/\rho_{\rm c}$:

$$4\pi \tilde{r}^2 \tilde{\rho} = -\left[\frac{K\rho_{\rm c}^{\gamma-2}}{R^2}\right] \frac{d}{d\tilde{r}}\left(\frac{\tilde{r}^2}{\tilde{\rho}}\frac{d\tilde{\rho}^\gamma}{d\tilde{r}}\right). \tag{136}$$

This gives a dimensionless differential equation to be solved subject to boundary conditions $\tilde{\rho}(0) = 1$; $\tilde{\rho}(1) = 0$ and $d\tilde{\rho}/d\tilde{r} = 0$ at $\tilde{r} = 0$ (because the pressure gradient must vanish at the centre, by spherical symmetry). Numerical solution of this equation gives a central density that is 6 times the mean if $\gamma = 5/3$ and fixes the constant in the density–radius relation:

$$R \propto \rho_{\rm c}^{\gamma/2-1}. \tag{137}$$

Similarly, we can solve for relativistic polytropes with $\gamma = 4/3$. The key outcome there is that the mass is independent of the radius and density:

$$M \propto \rho_{\rm c} R^3, \quad \text{but} \quad R \propto \rho_{\rm c}^{-1/3}. \tag{138}$$

So, non-relativistic white dwarfs are stable over a range of masses, but as their mass increases, they shrink. By the uncertainty principle, this must force the electrons closer and closer to becoming relativistic, and the limit of this is a star in which the electrons are all relativistic, which has a unique mass. It appears that degeneracy pressure has no way of supporting more massive bodies.

11.2 Cooling and ages of white dwarfs

All these calculations have assumed that the white dwarf is of zero temperature. This is not true initially, as the white dwarf starts life as the core of an evolved star. Nevertheless, degeneracy pressure allows the white dwarf to be supported without fusion energy generation, so it rapidly cools, and the zero-temperature approximation becomes a good one.

As usual, the black-body approximation can be used to deduce an effective surface temperature in terms of luminosity:

$$L = 4\pi R^2\, \sigma\, T_{\rm eff}^4, \quad \text{where} \quad \sigma = 5.67 \times 10^{-8}\ \mathrm{W\,m^{-2}K^{-4}}; \tag{139}$$

Using the mass–radius relation derived above, we can obtain the temperature corresponding to a given luminosity:

$$T_{\rm eff} = 10^{4.8}(L/L_\odot)^{1/4}(M/M_{\rm crit})^{1/6}\ \mathrm{K}. \tag{140}$$

Because they are so small, luminous white dwarfs must be extremely hot; conversely, white dwarfs with effective temperatures closer to that of the sun are very low-luminosity objects, and extremely hard to detect.

Such cool white dwarfs are interesting, because they are natural clocks: the coolest ones are the oldest. One might think of working out the age–temperature relation just by using the above temperature to calculate the thermal energy of the star, and equating the luminosity to rate of change of thermal energy. This goes wrong because the internal temperature tends to be much higher than the surface. The interior is at a single temperature owing to thermal conduction by the degenerate electrons, but the surface layer is non-degenerate and serves as an effective insulator. The lifetimes of white dwarfs are therefore longer than the naive sum would indicate – but still finite. There are claims of a cutoff in numbers at about $L = 10^{-4.5} L_\odot$, which would correspond to an age of assembly for the disk of the Milky Way of about 9 Gyr, but this is still an area of active research.

12 Neutron stars and black holes

12.1 Neutron stars

For a white dwarf slightly more massive than the Chandrasekhar limit, it appears that gravitational collapse must follow, with a corresponding increase in density. As the density of free electrons and nuclei goes up, the reaction of **inverse beta decay** is favoured – i.e. an electron can combine with a proton in a nucleus to yield a neutron. Once the density reaches a critical level neutron rich nuclei (like ^{118}Kr) start to release free neutrons, in the phenomenon of **neutron drip**. This happens at a density of $\rho \simeq \rho_{\rm drip} = 4 \times 10^{14}\,{\rm kg\,m^{-3}}$. Being fermions, these neutrons can provide a degeneracy pressure just in the same way as the electrons. As more and more electrons combine with protons, their number density falls, and the number density of free neutrons rises. At a density of $\rho = 4 \times 10^{15}\,{\rm kg\,m^{-3}}$, half the pressure is provided by the neutrons. At $\rho \simeq 2.4 \times 10^{17}\,{\rm kg\,m^{-3}}$, the remaining nuclei touch, leaving essentially a giant ball of neutrons, with some electrons and protons mixed in. Effectively, the star has become a single nucleus, of colossal proportions. In short, a last-gasp strategy for evading total collapse of a white dwarf is to combine its electrons and protons into neutrons. The same stability analysis would then be applied with neutrons as the fermions but now with $\mu \simeq 1$. This apparently gives a limiting mass 4 times larger than for white dwarfs, but is the Newtonian analysis valid for these **neutron stars**?

An interesting way of expressing the size of these stars is in terms of their gravitational potential. Recall that stability of a degenerate star requires it to be on the relativistic borderline, *i.e.* $p_{\rm F} \sim m_e c$ for the Fermi momentum. For a degenerate configuration, the number density is thus $n_e \sim (m_e c/\hbar)^3$, *i.e.* a particle spacing of about one de Broglie wavelength. The number density is also $n_e \sim (M/\mu m_p)/R^3$, which gives $R \sim (M/\mu m_p)^{1/3}(\hbar/m_e c)$. At the critical mass, $M \sim (\hbar c/G)^{3/2}(\mu m_p)^{-2}$, this gives a gravitational potential at the star's surface of

$$\frac{\Phi}{c^2} \sim \frac{GM}{Rc^2} \sim GM^{2/3}(\mu m_p)^{1/3} m_e/\hbar c \sim \frac{m_e}{\mu m_p}. \tag{141}$$

Newtonian theory is therefore adequate for white dwarfs, but fails entirely for neutron stars where $m_e \to m_n$ and $\mu \to 1$. The best detailed calculations give a maximum neutron star mass of

$$M_{\rm max} \simeq 2 - 3\ M_\odot \tag{142}$$

and a radius

$$R_{\rm max} \simeq 10 - 15\ {\rm km}. \tag{143}$$

These figures are still uncertain, owing to the exotic nature of neutron star material.

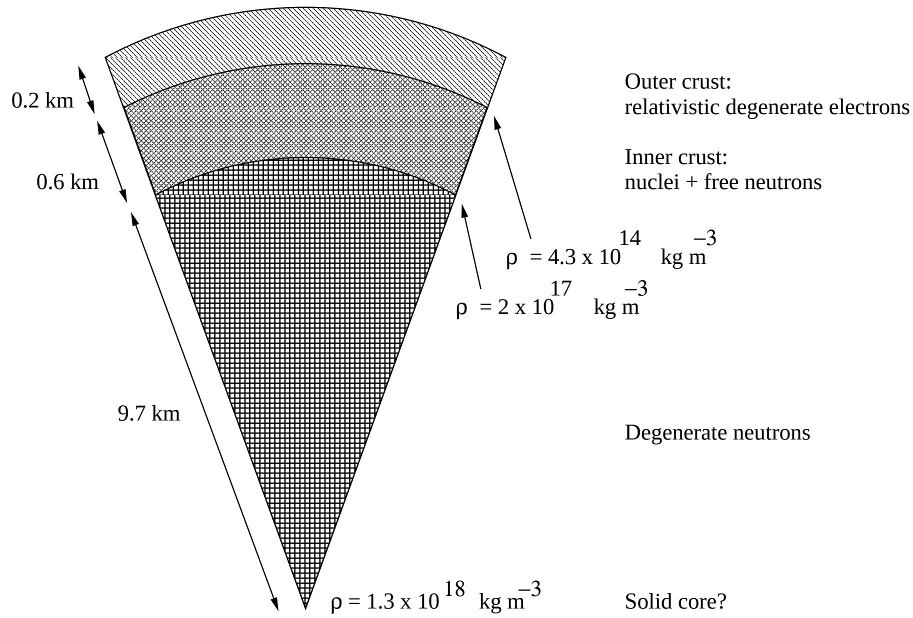

Figure 9. The density of a neutron star is highest at the centre and decreases outward. In the outer crust, the pressure is dominated by relativistic degenerate electrons. The majority of the volume, however, is dominated by a relativistic gas of degenerate neutrons.

12.2 Black holes

For a neutron star above the theoretical mass limit, there is no known equation of state of nuclear matter that can prevent collapse. What then will be the fate of a more massive star once it stops burning, and so loses its thermal pressure support to balance gravity? Usually the entire disruption of the star will result in a spectacular supernova. But the best models we have for supernovae predict that for progenitor (pre-supernova) stars that are sufficiently massive, a core more massive than $2\,M_\odot$ will remain. The only known answer is that such a star will undergo a catastrophic collapse that, according to General Relativity, is unstoppable, resulting in a singularity in space- time. We don't see the singularity, however. Instead we see a **black hole**.

This name reflects the fact that light cannot escape at all from a sufficiently strong gravitational field. By something of a coincidence, a Newtonian argument gives the right answer for this, in an argument due to Laplace. A particle emitted with velocity v at radius r has total energy

$$E_{\rm tot} = mv^2/2 - GMm/r, \tag{144}$$

which is negative if $r < 2GM/v^2$, so that the particle cannot escape to infinity. Letting $v \to c$ implies that there is an **event horizon** of radius $2GM/c^2$ from which light cannot escape. A proper calculation in general relativity shows that this conclusion is actually correct. Moreover, as physical systems approach this radius, gravitational time dilation slows the apparent ticking of clocks to zero, so that emission of photons effectively ceases: the central regions of a black hole cannot be detected directly. Instead, we rely on emission from hot material in orbit at radii of a few times the horizon radius. Here, the orbital speeds are a good fraction of c, and material being accreted can be heated to the point where it emits X-rays.

13 An overview of the interstellar medium

The ISM is a component of great significance in astronomy:

(1) Birthplace of stars
(2) Stellar graveyard – 'metal' enrichment (Nebulae)
(3) Dynamic arena of stars: HII regions, winds, supernova remnants
(4) Significant part of Galaxy: $M_{\rm gas} \simeq 0.05 M_{\rm Galaxy}$

Its major components are as follows:

(1) **Neutral Hydrogen** (H^0 or HI regions); revealed by 21cm radiation; contains $\gtrsim 90\%$ of total mass of ISM; typical $T \simeq 80$ K with $n_{\rm H} \simeq 3\times10^6$ m^{-3}; denser clouds colder: $T \simeq 30$ K; concentrated along spiral arms of Galaxy

(2) **Molecular Hydrogen** (H_2); H_2 forms in densest regions of HI clouds; $T \simeq 30$ K with $n({\rm H_{tot}})$ reaching $10^9 - 10^{12}$ m^{-3}.

(3) **HII Regions** (H^+); located around hot, blue stars (O & B); $n_e \simeq 10^8 - 10^{10}$ m^{-3} near O stars; $n_e \simeq 3\times10^5$m^{-3} near B stars, planetary nebulae; $T \simeq 8000$ K; $M_{\rm HII} \simeq 0.1 M_{\rm ISM}$; concentrated along Galaxy's spiral arms.

(4) **Dust** Small particles or grains (C & Si mostly): radii < 1 μm; $M_{\rm dust} \simeq M_{\rm ISM}/160$; Revealed by extinction: Apparent magnitudes of stars are typically reduced by 4 magnitudes in HI clouds (since the clouds contain dust).

(5) **Abundances** 70% H, 30% He by mass – similar to Sun; 1–2% in **metals**: C, N, O, Fe, Si, Na, Mg, etc., but with different relative ratios compared with the Sun. This is believed to be due to condensation of some of the metals into dust grains – some elements solidify more easily than others.

Table 7 Properties of typical interstellar gas clouds

Cloud Type	Bok Globule	IR/HII Cloud	Dark Cloud	Diffuse Cloud	Cloud Complex
A_v/mag	4	30	4	0.2	4
$N_{\rm H}$/cm^{-2}	8×10^{21}	6×10^{22}	8×10^{21}	4×10^{20}	8×10^{21}
$n_{\rm H}$/cm^{-3}	7×10^3	4×10^4	2×10^3	20	200
R/pc	0.3	0.4	1	5	10
T/K	10	50	10	80	10
$M/M_\odot$	30	400	300	400	3×10^4

13.1 Dust Extinction

Let f_λ denote the flux of radiation received from a source of luminosity L_λ at a distance D. Then

$$f_\lambda = \frac{L_\lambda}{4\pi D^2} \tag{145}$$

in free space. If there is intervening material, as there is between the stars, then the received flux will be reduced by the factor $\exp(-\tau_\lambda)$, where τ_λ is the **optical depth** of the intervening

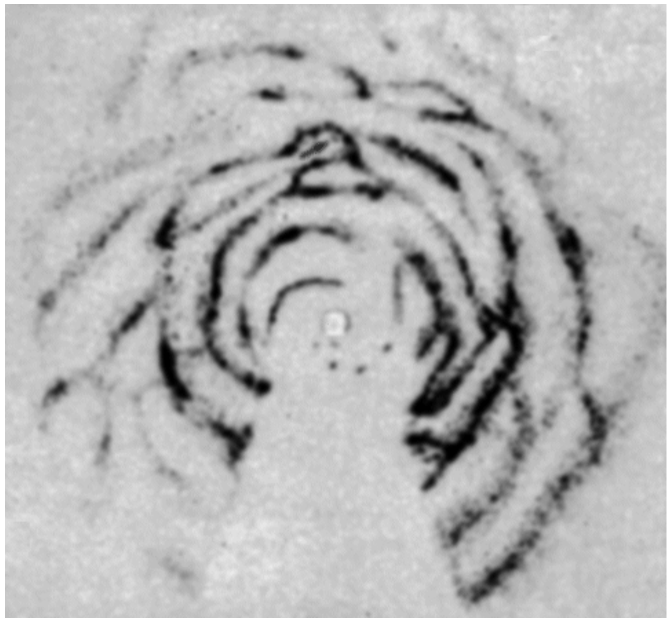

Figure 10. The distribution of neutral hydrogen in the Milky Way, as measured by the 21cm emission line. The location of the Sun (8 kpc from the centre) is shown by the arrow top centre. The concentration of neutral hydrogen towards spiral arms is very clear.

material. The subscript λ denotes that the amount of flux reduction – or 'extinction' – will in general depend on the wavelength of the light.

If the intervening material consists of particles with cross-section σ_λ and n is the number density of the particles, the optical depth over a path length ℓ will be

$$\tau_\lambda = n\sigma_\lambda\ell. \tag{146}$$

It is convenient to introduce the **column density** $N = n\ell$, which is just the number of atoms per unit area in a cylinder of length ℓ. Then

$$\tau_\lambda = N\sigma_\lambda \tag{147}$$

In the case of a source at distance D, the observed flux will then be suppressed exponentially:

$$\boxed{f_\lambda^{(\mathrm{ext})} = \frac{L_\lambda}{4\pi D^2}\, e^{-\tau_\lambda},} \tag{148}$$

where $\tau_\lambda = N\sigma_\lambda$ and $N = nD$. In astronomy, the amount of extinction is often expressed as a magnitude A_λ. The apparent magnitude at wavelength λ is related to the flux density by $m_\lambda = -2.5\log_{10} f_\lambda + \mathrm{constant}$. Thus, in the presence of extinction, we have

$$\begin{aligned} m_\lambda^{(\mathrm{ext})} &= -2.5\log_{10}\left[\frac{L_\lambda}{4\pi D^2}\, e^{-\tau_\lambda}\right] + \mathrm{constant} \\ &= -2.5\log_{10}\left(\frac{L_\lambda}{4\pi D^2}\right) - 2.5\log_{10}\left(e^{-\tau_\lambda}\right) + \mathrm{constant} \\ &= m_\lambda^{(\mathrm{no\ ext})} + 2.5\tau_\lambda \log_{10}(e) \end{aligned} \tag{149}$$

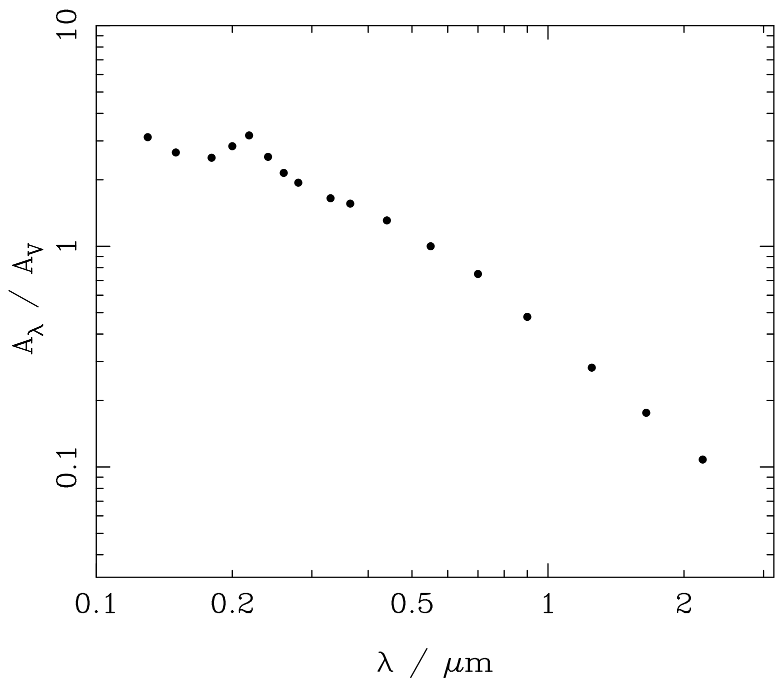

Figure 11. The extinction at wavelength λ relative to V, for dust with the composition characteristic of the Milky Way's ISM.

We call the magnitude increase due to extinction A_λ:

$$\boxed{A_\lambda = m_\lambda^{(\mathrm{ext})} - m_\lambda^{(\mathrm{no\ ext})} = 2.5\tau_\lambda \log_{10}(e).} \tag{150}$$

Observationally, we find in the visual band V ($\lambda \simeq 5480$Å) that

$$A_V \simeq 0.6 \, \frac{N_{\mathrm{HI}}}{10^{25}\ \mathrm{m}^{-2}}, \tag{151}$$

where $N_{\mathrm{HI}} = n_{\mathrm{HI}} D$ is the HI **column density**. We believe most of this extinction arises from dust grains mixed in with the gas.

In the B band ($\lambda \simeq 4400$Å), we find a higher amount of absorption. We define the **reddening** as

$$\boxed{E_{B-V} = A_B - A_V.} \tag{152}$$

In practice, $E_{B-V} \simeq A_V/3$. This constrains models for the dust grains, which are believed to be a mixture of carbon and silicate grains. The colour excess E_{B-V} may be measured from the displacement of stars in a colour-colour plot from the Main Sequence. (Recall the two-colour diagram in Figure 4.) Such studies allow the relative extinction at different wavelengths to be measured, as shown in Figure 10. A common approximation is to say that this scales as

$$A_\lambda \propto 1/\lambda, \tag{153}$$

But this ignores the UV 'bump' and the faster dropoff towards the infrared. These features show where the wavelength of light becomes comparable to detailed structure in the dust grains. If the grains were very much smaller than the wavelength, we would expect the $A_\lambda \propto 1/\lambda^4$ law of **Rayleigh scattering**, as seen in the Earth's atmosphere. The larger dust grains also differ in an important way: they genuinely **absorb** photons, rather than simply **scattering** them in a different direction. Either effect dims an object seen through a screen of dust, but we will see later in HII regions that the difference between absorption and scattering can be important.

14 HII Regions

14.1 Photoionization

Incoming photons with enough energy to lift an electron out of its Coulomb potential well into the continuum are said to photoionize the electron: this is the **photoelectric effect**. For H, the ionization potential is

$$I_{\rm H} = 13.6\,\mathrm{eV}. \tag{154}$$

Consider a star with a temperature $T_* > 30\,000$ K. The average energy of a photon from the star

$$\langle \epsilon_* \rangle \simeq kT_* = 4 \times 10^{-19}\ \mathrm{J} = 0.2 I_{\rm H} \tag{155}$$

is thus too small to ionise hydrogen. But recall the spectrum of a star is a black body, which has a long high energy tail. For an O star with $L \simeq 20 L_\odot$, typically $S_* \simeq 10^{49}\ \mathrm{s}^{-1}$ in ionizing photons. What will be the effect of these ionizing photons on the surrounding neutral gas (HI)?

In a time t, $N_\gamma = S_* t$ photons will be produced. These will be all used up in ionizing the surrounding gas. If $n_{\rm H}$ is the number density of hydrogen atoms (number of H atoms per unit volume), then

$$N = \frac{4\pi R^3}{3} n_{\rm H} \tag{156}$$

hydrogen atoms will be ionized where

$$N = N_\gamma = S_* t \quad \Rightarrow \quad R = \left(\frac{3}{4\pi} \frac{S_* t}{n_{\rm H}} \right)^{1/3}. \tag{157}$$

So as time passes, an ever-increasing ionized sphere will surround the star. This is called an HII region.

In the above, we assumed $n_{\rm H}$ was uniform. This was unnecessary. If the HII region has already grown to a radius $R_{\rm I}$, then in a short time Δt it will grow an additional amount ΔR given by

$$4\pi R_{\rm I}^2 \Delta R_{\rm I} n_{\rm H}(R) = S_* \Delta t, \tag{158}$$

and so we have

$$\frac{\Delta R_{\rm I}}{\Delta t} \rightarrow \frac{dR_{\rm I}}{dt} = \frac{S_*}{4\pi R_{\rm I}^2 n_{\rm H}(R_{\rm I})}, \tag{159}$$

for any $n_{\rm H}(R)$. This tells us how fast the HII region grows.

Examples:

(1) $S_* = 10^{49}\ \mathrm{s}^{-1}$ (typical O6 star) Typical surrounding density is $10^{10}\ \mathrm{m}^{-3}$ for a cloud where young O stars form. Then after $t = 10\,000$ years, $R_{\mathrm{I}} \simeq 4\times 10^{16}\ \mathrm{m} \simeq 1.4$ pc.

(2) Consider a somewhat more interesting case. The clouds surrounding young stars tend to be denser at their centres (where the stars form), and decrease in density outward. Suppose $n_{\mathrm{H}} = n_0 (R/R_0)^{-3/2}$; our growth equation becomes $4\pi R_{\mathrm{I}}^2 n_{\mathrm{H}} dR_{\mathrm{I}}/dt = S_*$, implying

$$\int_{R_{\mathrm{I}}=0}^{R_{\mathrm{I}}(t)} 4\pi R_{\mathrm{I}}^2 n_0 \left(\frac{R_{\mathrm{I}}}{R_0}\right)^{-3/2} dR_{\mathrm{I}} = \int S_*\, dt = S_* t \tag{160}$$

(if $S_* =$ constant). Now,

$$\begin{aligned}\int_{R_{\mathrm{I}}=0}^{R_{\mathrm{I}}(t)} 4\pi R_{\mathrm{I}}^2 n_0 \left(\frac{R_{\mathrm{I}}}{R_0}\right)^{-3/2} dR_{\mathrm{I}} &= 4\pi n_0 R_0^3 \int_{x=0}^{R_{\mathrm{I}}(t)/R_0} x^2 x^{-3/2} dx \qquad x \equiv \frac{R_{\mathrm{I}}}{R_0} \\ &= 4\pi n_0 R_0^3 \frac{2}{3}\left[x^{3/2}\right]_0^{R_{\mathrm{I}}(t)/R_0} = \frac{8\pi}{3} n_0 R_0^3 \left(\frac{R_{\mathrm{I}}}{R_0}\right)^{3/2}\end{aligned} \tag{161}$$

so

$$R_{\mathrm{I}}(t) = \left(\frac{3}{8\pi}\,\frac{S_* t}{n_0 R_0^3}\right)^{2/3} R_0 \tag{162}$$

We see now that instead of growing like $R_{\mathrm{I}}(t) \propto t^{1/3}$ as in the uniform density case, the HII region now grows faster, as $R_{\mathrm{I}}(t) \propto t^{2/3}$. This is because the gas density of the cloud decreasing with R – the cloud is thinning out – and so the star is able to ionize a greater volume of cloud for the same number of photons. The sensitivity to the density gradient can in part explain why an HII region is so filamentary on its appearance. We shall see, however, that this is too simple a view of an HII region, and only describes its very early stages.

14.2 Radiative Recombination

When we look at an HII region, most of what we see is light emission from hydrogen. This is dominated by the $n = 3 \to n = 2$ Balmer Hα line at $\lambda = 6563$Å: How did an electron get into the $n = 3$ state? Through **radiative recombination**, which is the capture by a positive ion (or proton in the case of H) of a free electron, with the consequent release of radiation. If the electron has initial energy E_e, the amount of energy released during the recombination to a level of energy E_n is

$$E_\gamma = E_e + |E_n|. \tag{163}$$

Once captured, the electron can make its way down by spontaneous transitions to levels of lower energy, resulting in a complex cascade of transitions. These inevitably lead to the production of Balmer ($n > 2 \to n = 2$) photons if the nebula is optically thick to Lyα photons (as it usually is).

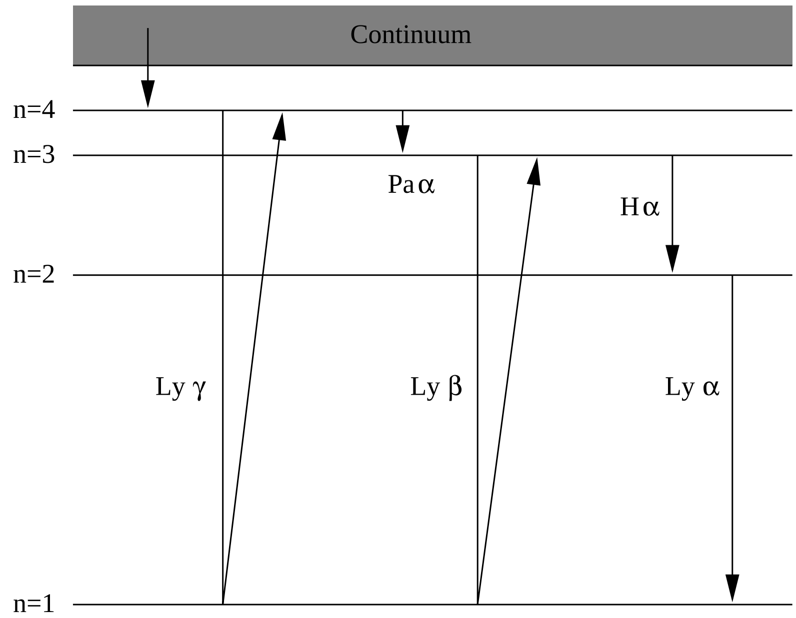

Figure 12. The recombination cascade in hydrogen. A continuum electron is absorbed, placing the atom in the $n = 4$ excited state in this example. Decay direct to the ground state produces a Lyγ photon, which is immediately reabsorbed by another atom. The path to the ground state instead produces a sequence of Balmer and lower-energy photons, plus Lyα.

Most of the hydrogen that's neutral is in the ground $n = 1$ state. That means that any Lyman photons will be immediately re-absorbed (by a nearby H atom in the $n = 1$ state). The result is that Lyα photons undergo **resonant scattering** in the nebula, only escaping by a very slow diffusive process. In a typical HII region, the odds are high that on its effectively long journey through the HII region, a Lyα photon will eventually hit a dust grain and be absorbed by it. As a consequence, very few Lyα photons can be seen from HII regions.

Another consequence of the recombinations is that hydrogen atoms that were once ionized recombine, producing a new neutral atom which will then be ionized again. As a consequence, there will always be some neutral hydrogen atoms in the HII region. But it also means the full amount of ionizing photons won't make it to the edge of the HII region (if some are lost on route). This slows down the advance of the ionization front, and can even stop it altogether. Let's see how this can happen. Suppose we consider a particular free electron. How long will it take to meet a proton and so recombine? Obviously the rate of collisions will be proportional to the density of protons: doubling the number of protons in a box of fixed volume will double the chance of the electron recombining with a proton during the same interval of time. Thus the rate of recombinations per electron is

$$\mathcal{R} = \alpha n_p, \tag{164}$$

where we call the proportionality constant α the **radiative recombination coefficient**. The time it will take the electron to recombine is then

$$\boxed{t_{\rm rec} = \frac{1}{\mathcal{R}} = \frac{1}{\alpha n_p},} \tag{165}$$

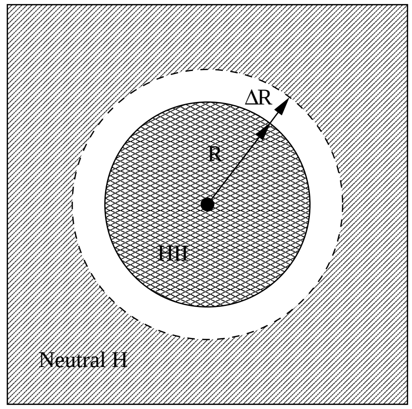

Figure 13. The geometry for the growth of an HII region. If the ionized region is of radius R, the time taken to ionize a further shell of thickness ΔR depends on the rate of production of ionizing photons, minus the rate at which they are required to balance recombinations within R.

where $t_{\rm rec}$ is called the **recombination time**. Typically $\alpha \simeq 3 \times 10^{-19}$ m^{3}s^{-1}, although it depends on temperature. Then for $n_p = 10^{10}$ m^{-3}, $t_{\rm rec} = 3 \times 10^8\,{\rm s} = 10$ yr.

Now, if there are n_e electrons per unit volume in the HII region, the total recombination rate is

$$\mathcal{R}_{\rm tot} = \frac{4\pi}{3}\, R_{\rm I}^3 n_e\, \alpha n_p, \tag{166}$$

where the first factor is the total number of electrons in HII region of radius $R_{\rm I}$. But this can't exceed the rate at which the star is producing ionizing photons S_*. Thus the HII region must stop growing when the rate of ionizing photons balances the total rate of recombinations:

$$S_* = \mathcal{R}_{\rm tot}, \tag{167}$$

or

$$\boxed{S_* = \frac{4\pi}{3} R_{\rm S}^3 \alpha n_e n_p,} \tag{168}$$

where $R_{\rm S}$ is called the **Strömgren radius**.

For a pure hydrogen nebula, $n_e = n_p$ by ionization balance, and this is $\simeq n_{\rm H}$, the original total H-atom density, since the gas is highly ionized. So

$$R_{\rm S} \equiv \left(\frac{3}{4\pi}\,\frac{S_*}{\alpha n_{\rm H}^2}\right)^{1/3}. \tag{169}$$

Example: Consider the O star again with $S_* = 10^{49}$ s^{-1} in a cloud of density $n_{\rm H} = 10^{10}$ m^{-3}. Then $R_{\rm S} \simeq 4.3 \times 10^{15}\,{\rm m} = 0.14$ pc. So indeed the Strömgren radius is not so large.

How long does it take for the ionization radius to reach $R_{\rm S}$? Consider an advancing shell as before: the number of hydrogen atoms in a shell of thickness ΔR is

$$n_{\rm H}(R_{\rm I}) \times 4\pi R_{\rm I}^2 \Delta R_{\rm I}. \tag{170}$$

The number of ionizing photons produced by the star in the time available to ionize the atoms is $S_*\Delta t$. But we know not all the photons will make it all the way to $R_{\rm I}$ because some will be absorbed on route by the newly recombined hydrogen atoms. How many?

In ionization equilibrium, every newly recombined atom is compensated by a new photoionization. So we only need count up the total number of recombinations inside the sphere:

$$\frac{4\pi}{3} R_{\rm I}^3 n_e\,(\alpha n_p)\Delta t. \tag{171}$$

The number of photons that reach $R_{\rm I}$ is thus reduced by this amount, needed to keep the parts at $R < R_{\rm I}$ ionized, and so:

$$4\pi R_{\rm I}^2 n_{\rm H}(R_{\rm I})\Delta R_{\rm I} = S_*\Delta t - \frac{4\pi}{3} R_{\rm I}^3 \alpha n_{\rm H}^2 \Delta t \tag{172}$$

(taking again $n_e = n_p \simeq n_{\rm H}$), or

$$\boxed{4\pi R_{\rm I}^2 n_{\rm H}(R_{\rm I})\frac{dR_{\rm I}}{dt} = S_* - \frac{4\pi}{3} R_{\rm I}^3 \alpha n_{\rm H}^2(R_{\rm I})} \tag{173}$$

This is a key result, which gives the rate of advance of the ionization front. Now we can see explicitly that $dR_{\rm I}/dt = 0$ when

$$R_{\rm I} = R_{\rm S} = \left(\frac{3}{4\pi}\,\frac{S_*}{\alpha n_{\rm H}^2}\right)^{1/3}, \tag{174}$$

so that the HII region stops growing.

We see we have two characteristic quantities, defined by the parameters $n_{\rm H}$ & S_* of the situation: the Strömgren radius $R_{\rm S}$ and the recombination time

$$t_{\rm rec} = \frac{1}{n_{\rm H}\alpha}. \tag{175}$$

Let us make the equation for $dR_{\rm I}/dt$ dimensionless by defining

$$\lambda \equiv R_{\rm I}/R_{\rm S}; \qquad \tau \equiv t/t{\rm rec}. \tag{176}$$

Multiplying through by $t_{\rm rec}/R_{\rm S}^3$ gives

$$4\pi\lambda^2 n_{\rm H}\,\frac{d\lambda}{d\tau} = \frac{S_* t_{\rm rec}}{R_{\rm S}^3} - \frac{4\pi}{3}\lambda^3 n_{\rm H}, \tag{177}$$

and dividing through by $4\pi n_{\rm H}/3$ gives

$$3\lambda^2\frac{d\lambda}{d\tau} = 1 - \lambda^3. \tag{178}$$

The left-hand side is just $d\lambda^3/d\tau$, so the solution is clearly

$$\lambda^3 = 1 - \exp(-\tau), \tag{179}$$

or

$$\boxed{R_{\rm I}(t) = R_{\rm S}\left[1 - \exp(-t/t_{\rm rec})\right]^{1/3}.} \tag{180}$$

14.3 Temperatures in nebulae

How do we know the temperature of the gas in an HII region? This is measured via **collisional excitation**, which is the excitation of an atom or ion from its ground state to a higher energy (excited) state caused by the collision with a free electron:

$$e^- + X(n=1) \quad \overset{\text{excitation}}{\longrightarrow} \quad X(n>1) \quad \overset{\text{decay}}{\longrightarrow} \quad X(n=1) + \gamma. \tag{181}$$

By looking at the intensity of lines produced by the spontaneous decay following the excitation, we can measure the temperature of the electrons. This works because the kinetic energies of electrons in thermal equilibrium follow a Boltzmann distribution:

$$f(E) \propto e^{-E/kT}. \tag{182}$$

If it takes an energy $E_{n'} - E_1$ to excite an atom from the $n = 1$ state to the n' state, the number of electrons with sufficient energy will be $\propto \exp\left(E_{n'} - E_1\right)/kT$. Thus, the larger $E_{n'} - E_1$ is, the dimmer the line $n' \to n = 1$ will be, since there will be fewer transitions to the higher n' levels. Specifically, consider two such levels: n' and n'' where $E_{n''} > E_{n'}$, Then the ratio of collisions to the n'' level is the n' level will scale like the corresponding electron densities with the required energies:

$$\frac{f\left(E_{n''} - E_1\right)}{f\left(E_{n'} - E_1\right)} = \frac{e^{-(E_{n''}-E_1)/kT}}{e^{-(E_{n'}-E_1)/kT}} = e^{-(E_{n''}-E_{n'})/kT}. \tag{183}$$

If $I_{n'',1}$ and $I_{n,1}$ represent the intensities of the lines that result, then

$$\boxed{\frac{I_{n',1}}{I_{n'',1}} \propto \frac{f\left(E_{n'} - E_1\right)}{f\left(E_{n''} - E_1\right)} = e^{\Delta E/kT},} \tag{184}$$

where $\Delta E = E_{n''} - E_{n'}$.

Example: OIII. The emission-line ratio is

$$\frac{I_{\lambda 4959} + I_{\lambda 5007}}{I_{\lambda 4363}} \propto e^{\Delta E/kT}. \tag{185}$$

Typical measured temperatures lie in the range $6000 < T < 12\,000$ K. What determines the temperature?

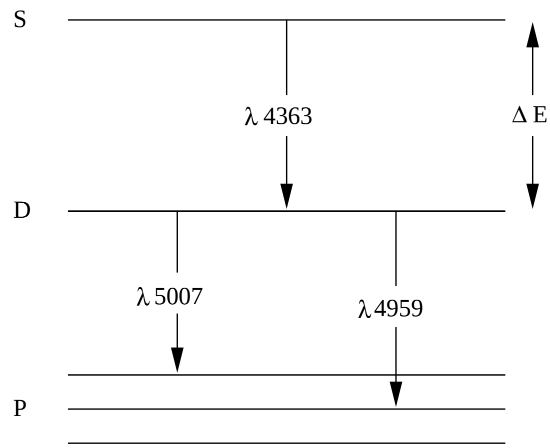

Figure 14. The energy level diagram for the lowest terms of OIII (twice ionized oxygen). The ground state consists of two electrons in the $2p$ state, but fine-structure interactions split this according to the total electron spin and (mainly) total angular momentum (denoted S,P,D etc.). The relative intensities of the two sets of lines shown (one a singlet, the other a doublet) is set by the relative occupancies of the upper levels, separated by ΔE.

Heating: Heat is added to the HII region by the photoionization process. The typical energy of a photo-ejected electron is kT_*, where T_* is the effective black-body temperature of the star. So the total rate of heat input by the star is:

$$G = S_* \times kT_*. \tag{186}$$

Cooling: The nebula loses energy due to radiative recombinations. Every time an electron recombines, an amount of energy kT_e (on average) is radiated away. So the total rate of energy loss is the total recombination rate times kT:

$$L_R = \frac{4\pi}{3} R_{\rm I}^3 \alpha n_{\rm H}^2 \times kT_e \tag{187}$$

In equilibrium, we have $L_R = G$ and $R_{\rm I} = R_{\rm S}$, where

$$\frac{4\pi}{3} R_{\rm S}^3\, \alpha n_{\rm H}^2 = S_* \quad \text{(Strömgren sphere)}. \tag{188}$$

So

$$L_R = \frac{4\pi}{3} R_{\rm S}^3\, \alpha n_{\rm H}^2\, kT_e = S_* kT_e = G = S_* kT_* \tag{189}$$

Thus, we find

$$\boxed{T_e = T_*.} \tag{190}$$

But $T_* \simeq 30\,000 - 50\,000$ K typically (O & B stars), which is much higher than measured. So what went wrong?

14.4 Forbidden-line cooling

The problem with the above argument is that it neglects the energy emitted from the nebula by processes other than hydrogen recombination. The general name for the total loss of energy by radiation is **cooling**. this is a bad name, since it implies diffusion of heat, but we are stuck with it. In discussing the use of species such as OIII as temperature diagnostics, we introduced the idea of collisional excitation, and such excitation can contribute to the cooling in a variety of ways:

(1) Neutral H: emits Lyα λ1216 plus Hα λ6563; Lyα re-radiated in IR by dust;
(2) OII (O^+): emits λ3726, λ3729; (λ2470);
(3) OIII (O^{++}): emits (λ4363), λ4959, λ5007 (as we saw earlier);
(4) NII (N^+): emits λ6548, λ6583

We might expect the collisional excitations of HI to dominate because it is so abundant, but the trouble is it turns out that the electrons are quite cool, and they lack the energy to excite HI (since the energy of a UV photon is required). In any case, most of the hydrogen is ionized, rather than neutral. It turns out that the main **coolant** for HII regions is the emission from OII (O^+) ions. The collisional excitation loss for O^+ is

$$L_{\mathrm{O}^+} = \frac{4\pi}{3} R_{\mathrm{I}}^3 \mathcal{L}_{\mathrm{O}^+}, \tag{191}$$

where the **emissivity** of O^+ is approximately

$$\mathcal{L}_{\mathrm{O}^+} = 1.1 \times 10^{-35} \left(\frac{n_{\mathrm{O}^+}}{n_{\mathrm{O}}^{\mathrm{tot}}} \right) \frac{n_{\mathrm{H}}^2}{T_4^{1/2}} e^{-3.9/T_4} \ \ \mathrm{J\,m^{-3}s^{-1}}. \tag{192}$$

Here, $T_4 \equiv T_e/10^4\,\mathrm{K}$, and we have used high ionization of hydrogen plus the known cosmic abundance of oxygen to replace the natural collisional term $n_e\, n_{\mathrm{O}^+}$ by one $\propto n_{\mathrm{H}}^2$.

So, balancing radiative losses with photoionization heating gives

$$L_{\mathrm{O}^+} = \frac{4\pi}{3} R_{\mathrm{S}}^3 \mathcal{L}_{\mathrm{O}^+} = G = S_* k T_* = \frac{4\pi}{3} R_{\mathrm{S}}^3\, \alpha n_{\mathrm{H}}^2\, k T_* \tag{193}$$

$$\boxed{\Rightarrow \quad \mathcal{L}_{\mathrm{O}^+} = \alpha n_{\mathrm{H}}^2 k T_*.} \tag{194}$$

The hydrogen recombination coefficient can be approximated by

$$\alpha = 2 \times 10^{-16}\, T_e^{-3/4} \ \ \mathrm{m^3 s^{-1}}, \tag{195}$$

so that, taking $n_{\mathrm{O}^+}/n_{\mathrm{O}}^{\mathrm{tot}} \simeq 1$,

$$\mathcal{L}_{\mathrm{O}^+} = 1.1 \times 10^{-35} \frac{n_{\mathrm{H}}^2}{T_4^{1/2}} e^{-3.9/T_4} = 2 \times 10^{-19} T_4^{-3/4} n_{\mathrm{H}}^2 k T_* \tag{196}$$

$$\boxed{\Rightarrow \quad T_4^{1/4} e^{-3.9/T_4} = 2.5 \times 10^{-7} T_*.} \tag{197}$$

Solving this equation, we find that the nebula can indeed maintain itself at a temperature well below that of the ionizing star, in agreement with observations:

T_*/K	T_e/K
20 000	7450
40 000	8500
60 000	9300

14.5 Forbidden lines and densities

Collisionally excited species are useful not only for measuring and understanding the temperatures in nebulae, but also for measuring electron number densities. This is because another name for a line such as the $\Lambda 3727$ doublet of OII is a **forbidden transition**. What this means is that, owing to the symmetry of the wavefunctions of the upper and lower states, the transition rate between the energy levels involved in the lines is very low. For example, the D level in OIII from which the $\lambda 4959$ & $\lambda 5007$ lines is produced has a lifetime of about 30 s (a **metastable state**). This is very much longer than the corresponding figure for **permitted** transitions, such as the hydrogen series: for example, the $n = 2$ level decays spontaneously to $n = 1$ in a lifetime of about 10^{-9} s, producing a Lyα photon.

This long lifetime of forbidden lines is useful, because the line is fragile: if the excited ion is disturbed before it can decay, the corresponding transition will not be produced. The natural way to achieve this is **collisional de-excitation**: the same collisions with electrons that excite ions can also depopulate the excited levels if they occur too frequently. The collision rate for a given ion is proportional to the electron number density, with a temperature-dependent coefficient $\Gamma(T)$. The production of transitions down from a level with lifetime τ therefore requires

$$\boxed{\Gamma(T)n_e < \tau^{-1}.} \tag{198}$$

Clearly, there is a **critical density** at which $n_e = (\Gamma\tau)^{-1}$; above this, the transition is quenched and the spectral line does not occur. Table 8 shows critical densities for some of the oxygen lines, computed at $T = 10^4$ K (not terrible temperature sensitive). We see that the intensity ratio for the $\lambda 3726$ & $\lambda 3729$ OII lines is an excellent way to measure the density, if is is around $10^{10}\,\mathrm{m}^{-3}$. We can also see why these transitions are unlikely to be observed on Earth: very low-density gas is needed. For some time, the OIII $\lambda 4959/\lambda 5007$ doublet was thought to result from a new element: **nebulium**. The correct explanation was only given by Ira Bowen in 1927.

Table 8 Critical densities for nebular forbidden lines

Ion	line	critical n_e/m^{-3}
OII	$\lambda 3726$	1.6×10^{10}
OII	$\lambda 3729$	3.1×10^{9}
OIII	$\lambda 4959$	7.0×10^{11}
OIII	$\lambda 5007$	7.0×10^{11}

15 Astrophysical fluid dynamics

So far, the IGM has been treated as static, but this is not realistic. At a minimum, we have seen that young stars can heat the IGM in their vicinity; such energy input would make the gas tend to expand. We therefore need to consider the processes that govern the time dependence of the density and velocity in a fluid.

15.1 The fluid approximation

Before getting on with this task, we should ask whether the normal concept of a fluid is valid in astrophysics. Ultimately, all fluids are a collection of independent particles. The smooth and continuous behaviour we associate with e.g. water arises because the mean free path of an individual particle is normally very short: it is the interactions between neighbouring particles that give the fluid the ability to show collective behaviour. Since the IGM is a plasma with a density that is very low by terrestrial standards, we need to worry about how big the mean free path really is. The mean free path is $\lambda = (n\sigma)^{-1}$, where n is the number density and σ the cross section. A rough estimate for the latter is to say that scattering in a plasma happens if the electrostatic potential energy exceeds the thermal energy:

$$\frac{e^2}{4\pi\epsilon_0 r} \gtrsim kT. \tag{199}$$

Taking $\sigma = \pi r^2$ and ignoring π etc., we get

$$\lambda \sim \frac{(\epsilon_0 kT)^2}{e^4 n} \simeq (T/30\mathrm{K})^2\,(n/10^{10}\,\mathrm{m}^{-3})^{-1}\ \mathrm{m}. \tag{200}$$

This is usually small enough, but not always. For material in the Solar wind near the Earth, λ exceeds the size of the Earth, so the plasma is effectively **collisionless**. Nevertheless, fluid-like behaviour is seen, and this is because magnetic field can act as extra 'glue' to make the plasma act collectively.

15.2 Equations of motion

The key concepts governing the behaviour of a fluid are (1) conservation of mass; (2) acceleration of fluid elements by pressure gradients. We will be content with studying these issues in one dimension – i.e. the velocity, density etc. will be taken to depend only on x, and be constant over the yz plane.

To deduce the equations of motion, consider a small box, of area A and thickness Δx. The amount of mass in this box is $M = \rho V = \rho\, A\, \Delta x$, where ρ is the mass density of the fluid. This mass can change through the difference between the rates at which mass flows in and out of the walls. The rate at which mass crosses unit area (the **flux density** of mass) is ρv, as is easily seen. In time t, the fluid moves a distance $v\Delta t$, and so a volume of fluid $A \times v(x)\Delta t$ crosses the wall of the box from the left. Similarly, a volume $A \times v(x+\Delta x)\Delta t$ leaves the right-hand wall of the box. The net change in mass in the box is given by the difference in these volumes, weighted respectively by $\rho(x)$ and $\rho(x+\Delta x)$. Since the volume of the box is fixed, this change in mass is $\Delta M = \Delta\rho V$:

$$\Delta M = \Delta\rho V = \Delta\rho A\Delta x = Av(x)\rho(x)\Delta t - Av(x+\Delta x)\rho(x+\Delta x)\Delta t. \tag{201}$$

We can consider small intervals so that $\Delta\rho/\Delta t$ becomes the derivative of ρ with respect to time, and $v(x+\Delta x)\rho(x+\Delta x) \simeq v(x)\rho(x) + \Delta x d(v\rho)/dx$. This gives a very simple equation, called the **equation of continuity**

$$\boxed{\frac{\partial\rho}{\partial t} = -\frac{\partial}{\partial x}(\rho v).} \tag{202}$$

The only subtlety here is that the derivatives involved must be *partial*: d/dx at constant t and vice-versa.

The other equation we will need is an equation of motion, or force law. Here, it is easiest to consider a box that moves with the fluid (i.e. we work in a frame of reference in which the fluid is instantaneously at rest). Alternatively, we are considering a little parcel of fluid whose walls are defined by particles, so that no material crosses them. Nevertheless, momentum crosses the walls, since the pressure in the fluid on either side of the walls acts on them. The net force on the $+x$ direction is just the pressure acting on the left, minus that on the right, times the area A:

$$F = AP(x) - AP(x+\Delta x). \tag{203}$$

This force must equal the mass of the fluid element times its acceleration: $\rho A\Delta x\dot{v}$. Therefore, we get a simple equation of motion, known as **Euler's equation**:

$$\boxed{\dot{v} = -\frac{1}{\rho}\,\partial P/\partial x.} \tag{204}$$

The only subtlety here is the meaning of $\dot{v}$. This is a time derivative as seen by an observer who moves with the fluid, and it is a mixture of time and spatial derivatives as seen in the lab:

$$\dot{v} \equiv \partial v/\partial t + v\,\partial v/\partial x \tag{205}$$

(because in a time Δt, the fluid moves a distance $v\Delta t$). The idea here is that the changes experienced by an observer moving with the fluid are inevitably a mixture of temporal and spatial changes. If I start to feel rain as I cycle towards my destination, it might be a good idea to cycle harder in the hope of arriving before the downpour really starts, but it could also be that it is raining near my destination, and I should stop and wait for the local shower to finish. With only local information on $d[\mathrm{rain}]/dt$ there is no way to tell which is the right strategy.

15.3 Sound waves

Now we have the fundamental partial differential equations that govern fluid flow, we can perform a very important analysis and ask what happens if we **perturb** the fluid by changing its density etc. by small amounts:

$$\rho \to \rho + \delta\rho. \tag{206}$$

We further assume that the unperturbed state is as simple as possible: a fluid of uniform density ρ_0 at rest ($v_0 = 0$). The equation of continuity is

$$\begin{aligned} \frac{\partial(\rho_0 + \delta\rho)}{\partial t} &= -\frac{\partial}{\partial x}\,(\rho_0 + \delta\rho)(v_0 + \delta v). \\ \Rightarrow \quad \frac{\partial \delta\rho}{\partial t} &= -\frac{\partial}{\partial x}\,(\rho_0 + \delta\rho)\delta v \\ \Rightarrow \quad \frac{\partial \delta\rho}{\partial t} &= -\rho_0 \frac{\partial}{\partial x}\,\delta v - \frac{\partial}{\partial x}\,\delta\rho\,\delta v \end{aligned} \tag{207}$$

We can make this even simpler by exploiting the fact that the perturbations are small: the last term contains the quadratic quantity $\delta\rho\,\delta v$, which must be negligible compared to $\rho_0 \partial/\partial x\,\delta v$. The equation of continuity is then

$$\frac{\partial \delta\rho}{\partial t} = -\rho_0 \frac{\partial}{\partial x}\,\delta v. \tag{208}$$

Similar reasoning makes Euler's equation simple: v_0 is zero, so the equation is

$$\partial\delta v/\partial t + \delta v\,\partial v/\partial x = -\frac{1}{\rho_0 + \delta\rho}\,\frac{\partial \delta P}{\partial x}. \tag{209}$$

Neglecting all second-order terms, this becomes

$$-\rho_0\,\partial\delta v/\partial t = \frac{\partial \delta P}{\partial x} \tag{210}$$

We can eliminate δv from these equations by taking the time derivative of the first and the space derivative of the second to yield $-\rho_0 \partial^2 \delta v/\partial t \partial x$ in both cases. This gives

$$\frac{\partial^2 \delta\rho}{\partial t^2} = \frac{\partial^2 \delta P}{\partial x^2}. \tag{211}$$

Finally, we need a relation between perturbations in pressure and in density. If we define a symbol

$$c_s^2 \equiv dP/d\rho, \tag{212}$$

then the equation becomes

$$\boxed{\frac{\partial^2 \delta\rho}{\partial t^2} - c_s^2 \frac{\partial^2 \delta\rho}{\partial x^2} = 0} \tag{213}$$

This is the one-dimensional **wave equation**, which has solutions that propagate at velocity c_s: $\delta\rho = f(x \pm c_s t)$. The quantity c_s is thus the **speed of sound** in the fluid.

We can think of two possible cases for evaluating c_s. Suppose the fluid was **isothermal** – i.e. at fixed temperature T, so that $P = (\rho/m)kT$, where m is a particle mass. In this case, $c_s^2 = kT/m$, which just says that the sound speed is of order the internal velocity dispersion of the particles. More plausibly, the equation of state will be **adiabatic**, since the fluid may not be able to radiate away extra energy gained when it is compressed. A better assumption is then $P \propto \rho^\gamma$, so that $c_s^2 = \gamma P/\rho$. For a simple gas, $\gamma = 5/3$, so that c_s is $\sqrt{5/3}$ greater than the isothermal case.

16 Shock waves

Despite the above analysis, not all waves in fluids have to travel at the speed of sound. Let's suppose we have a box of gas at rest. We then drive a piston into it, at first slowly: This will set up a pressure gradient which will accelerate the gas ahead of the piston (see Figure 15). The setting-up of the pressure gradient is done by sound waves: they communicate the presence of the piston to the fluid ahead. And so the fluid will be unaltered at $x > c_s t$. If the piston is pushed faster, it will start to catch up to the leading sound wave; as it does so, the gradients of pressure etc. in the fluid ahead become larger and larger. For example, $v(x)$ has to change from the piston velocity to zero over the distance from the piston position to $x = c_s t$, so dv/dx diverges as the piston approaches $x = c_s t$. What will happen if $v_{\rm piston} > c_s$? The sound waves will no longer be able to warn the fluid ahead, and yet somehow the fluid must know about the piston in order to respond. What does the fluid do? It responds by a discontinuous change in the fluid flow variables – a **shock**.

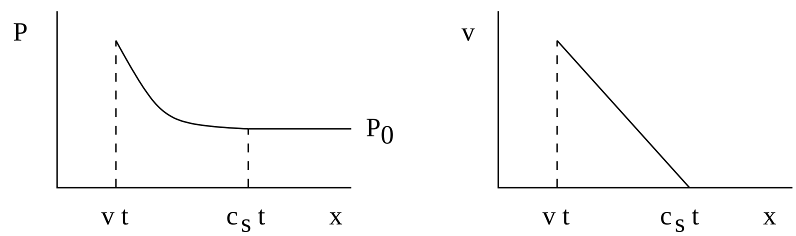

Figure 15. A piston is driven into a fluid at velocity v at $t = 0$, so that the piston position is $x = vt$. Pressure waves can move ahead of the piston at the speed of sound, c_s, so that there are gradients in pressure and velocity ahead of the piston, matching onto the undisturbed conditions (velocity zero; pressure P_0) for $x > c_s t$. As the piston's speed approaches $v = c_s$, these gradients become infinite, and a shock front is formed.

The laws of conservation must still apply, so we can relate the fluid variables on either side of the shock discontinuity, or shock front. Consider a shock front moving at speed $v_{\rm shock}$ into a stationary gas: it is simpler to look at this in the reference frame in which the shock front is stationary, so that upstream fluid arrives with velocity u_0 $(= v_{\rm shock})$, and streams away from the shock with velocity u_1 (strictly, u_0 and u_1 are negative, but it is convenient to reverse the x axis in the shock frame and they are treated as positive hereafter).

If we consider a steady-state shock moving at constant velocity, the conservation laws in the shock frame are very simple: the amounts of mass, momentum and energy arriving per unit time per unit area of the shock from upstream, plus any extra amount generated at the shock, must equal what is transported away downstream.

Mass conservation The flux density of mass is density times velocity. No extra mass is generated at the shock, so that

$$\boxed{\rho_0 u_0 = \rho_1 u_1,} \tag{214}$$

in the shock frame, or $\rho_0 v_{\rm shock} = \rho_1 \left(v_{\rm shock} - v_1\right)$ in the lab frame.

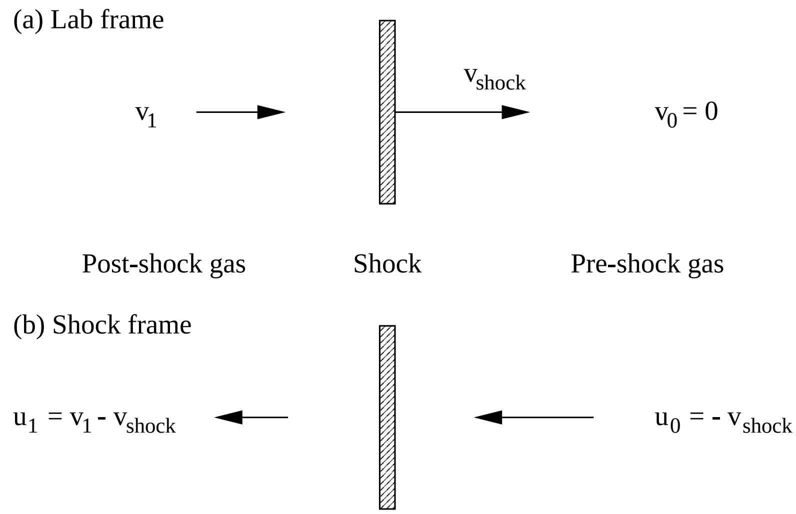

Figure 16. A plane shock front viewed in two reference frames: (a) the frame of the unshocked upstream fluid; (b) the frame in which the shock front is stationary

Momentum conservation The flux density of momentum is momentum density (ρv) times velocity. Momentum is not conserved at the shock, however, since there will be a difference in pressure between upstream and downstream. This also generates momentum (because pressure equals rate of change of momentum per unit area). Thus, the momentum conservation law is

$$\boxed{\rho_0 u_0^2 + P_0 = \rho_1 u_1^2 + P_1} \tag{215}$$

in the shock frame, or $\rho_0 v_{\rm shock}^2 + P_0 = \rho_1 (v_{\rm shock} - v_1)^2 + P_1$ in the lab frame.

Energy conservation The flux density of energy is velocity times the density of energy, which is the density of kinetic energy ($\rho v^2/2$) plus the internal energy density, ϵ. Energy is not conserved at the shock, since the pressure does work on the fluid as it flows. The rate of working per unit area is pressure times velocity. The overall conservation equation in the shock frame is therefore

$$\boxed{u_0(\rho_0 u_0^2/2 + \epsilon_0) + P_0 u_0 = u_1(\rho_1 u_1^2/2 + \epsilon_1) + P_1 u_1} \tag{216}$$

We can simplify this; first note that $\epsilon = 3nkT/2 = 3P/2$ for a gas of simple particles, so that $\epsilon + P = 5P/2$. Secondly, dividing by the equation of conservation of mass ($\rho_0 u_0 = \rho_1 u_1$), gives

$$\frac{1}{2}u_0^2 + \frac{5}{2}\frac{P_0}{\rho_0} = \frac{1}{2}u_1^2 + \frac{5}{2}\frac{P_1}{\rho_1}. \tag{217}$$

In the lab frame, this becomes

$$\frac{1}{2}v_{\rm shock}^2 + \frac{5}{2}\frac{P_0}{\rho_0} = \frac{1}{2}\left(v_{\rm shock} - v_1\right)^2 + \frac{5}{2}\frac{P_1}{\rho_1} \tag{218}$$

We thus have three relations for the three unknowns, u_1, ρ_1, P_1 in terms of u_0, ρ_0, P_0. This can be straightforwardly solved, but the expressions are cumbersome. For our purposes we will make the simplifying assumption that the shock is **strong**:

$$u_0^2 \gg \frac{P_0}{\rho_0} \quad \Rightarrow \quad u_0 \gg c_s(\text{upstream}), \tag{219}$$

so that the shock is **hypersonic**. The last conclusion follows from our previous discussion about sound speeds: $c_s^2 = \gamma P/\rho$ for adiabatic waves, or the same without the γ factor for isothermal waves. We then find the following key relations for a strong shock:

$$\boxed{\frac{\rho_1}{\rho_0} = 4; \quad P_1 = \frac{3}{4}\,\rho_0 u_0^2; \quad u_1 = \frac{1}{4}\,u_0.} \tag{220}$$

In terms of v_1 and $v_{\rm shock}$, noting $u_0 = -v_{\rm shock}$ and $u_1 = v_1 - v_{\rm shock}$, we obtain

$$P_1 = \frac{3}{4}\,\rho_0 v_{\rm shock}^2; \quad v_1 = \frac{3}{4} v_{\rm shock}, \tag{221}$$

plus an alternative form of the condition for a shock to be strong:

$$\frac{P_1}{P_0} = \frac{3}{4}\,\frac{v_{\rm shock}^2}{P_0/\rho_0} \gg 1. \tag{222}$$

In summary, there are three equivalent criteria that determine whether or not a shock is strong: (1) it is hypersonic; (2) it has a large pressure jump ($P_1/P_0 \gg 1$); (3) the **ram pressure** ($\rho_0 u_0^2$) greatly exceeds the upstream thermal pressure, P_0.

These results present a paradox: the shock compresses the gas, and our first thought might be that this process would be adiabatic, implying

$$\frac{P_1}{P_0} = \left(\frac{\rho_1}{\rho_0}\right)^{5/3} = 4^{5/3}, \tag{223}$$

whereas we have just shown that the pressure ratio across a hypersonic shock is $\gg 1$. The solution is that the compression is **irreversible** and hence not adiabatic: entropy is generated through viscous dissipation at the shock (nevertheless, confusingly, these are called **adiabatic shocks**, in order to distinguish them from the **isothermal shocks** studied later). The presence of viscosity is neglected in the above equations, but this is not correct where the fluid gradients are very high. The consequence of this is that the shock front is not really an abrupt discontinuity, but a continuous transition with a width of order the collisional mean free path.

16.1 Isothermal shocks

Often the gas radiates so strongly it maintains nearly a constant temperature on passing through the shock (see Figure 17). In this case, $P_1/\rho_1 = P_0/\rho_0 = c_0^2$ by the ideal gas law (constant T; c_0 is the initial sound speed). Continuity gives $\rho_1 u_1 = \rho_0 u_0$, and conservation of momentum gives $P_1 = \rho_0 u_0^2 - \rho_1 u_1^2$. Here, we again assume the strong-shock condition $\rho_0 u_0^2 \gg P_0$: since $P_0/\rho_0 = c_0^2$, this says that a strong shock is hypersonic, as in the adiabatic case. Now the condition of isothermality has replaced the energy jump condition.

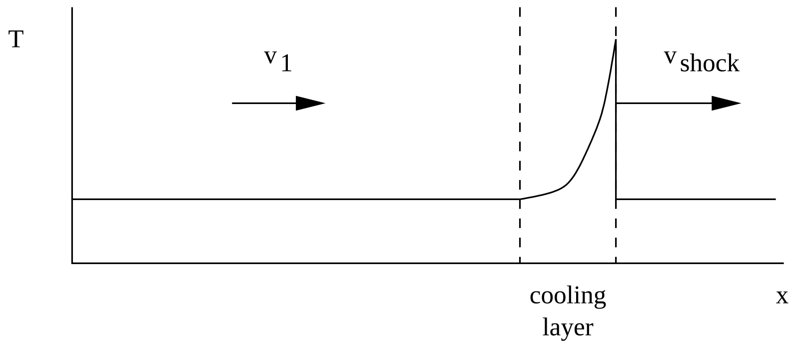

Figure 17. An isothermal shock is one where the post-shock temperature calculated according to the usual adiabatic shock formulae is so high that radiative cooling is very rapid. The temperature can then return to virtually the pre-shock temperature within a short distance. This 'cooling layer' is treated as being negligibly thin, and constituting an effective shock front for an isothermal shock.

To solve these equations, divide the momentum equation by the continuity equation, which gives

$$\frac{P_1}{\rho_1 u_1} + u_1 = u_0. \tag{224}$$

Dividing by u_1, we get

$$\frac{u_0}{u_1} = 1 + \frac{P_1/\rho_1}{u_1^2} = 1 + \frac{c_0^2}{u_0^2}\left(\frac{u_0}{u_1}\right)^2; \tag{225}$$

$c_0 = (P_0/\rho_0)^{1/2}$ is the isothermal sound speed of the pre-shocked gas (which is the same as the post-shock sound speed, via the isothermal assumption). This is a quadratic equation for u_0/u_1. We want the root in which this ratio is > 1 (the shock slows the flow down), so

$$\frac{u_0}{u_1} = \frac{1 + \sqrt{1 - 4c_0^2/u_0^2}}{2\,c_0^2/u_0^2} \quad \to \quad \left(\frac{u_0}{c_0}\right)^2 \quad \text{as} \quad \frac{u_0}{c_0} \to \infty. \tag{226}$$

Using continuity again, the density ratio is

$$\boxed{\frac{\rho_1}{\rho_0} \simeq \left(\frac{v_{\rm shock}}{c_0}\right)^2 \gg 1.} \tag{227}$$

Isothermal shocks can thus be extremely compressive, in contrast to adiabatic shocks. We now find $P_1 \simeq \rho_0 u_0^2$, instead of $(3/4)\rho_0 u_0^2$ previously, and can also write

$$\frac{P_1}{P_0} = \frac{\rho_1}{\rho_0} = \left(\frac{v_{\rm shock}}{c_0}\right)^2 \gg 1, \tag{228}$$

as the criterion for a strong shock.

17 The impact of stellar winds on the ISM

17.1 Eddington limit

Massive O & B stars produce **winds** with velocities as high as $v_* = 2000\,\mathrm{km\,s^{-1}}$. Typical mass-loss rates are $\dot{M}_* \simeq 10^{-6} M_\odot\,\mathrm{yr}^{-1}$, giving a typical **mechanical luminosity** of

$$\dot{E}_* \equiv \frac{1}{2}\dot{M}_* v_*^2 \simeq 10^{29}\,\mathrm{W}. \tag{229}$$

The reason such outflows exist is that the outer layers of high-mass stars are unstable to **radiation pressure**: the momentum transferred to material that intercepts photons can be high enough to overcome gravity. A given photon, of energy E, has a momentum E/c. Therefore, the rate at which a scattering particle with cross-section σ acquires momentum is σ/c times energy flux density:

$$\dot{p} = \frac{\sigma L}{4\pi r^2 c}. \tag{230}$$

This rate of momentum transfer is a force, and it cannot be allowed to exceed the gravitational force, otherwise the surface layers of the star will be blown away:

$$\sigma L/(4\pi r^2 c) < GMm/r^2. \tag{231}$$

Since both these depend on $1/r^2$, there is a maximum luminosity for a given mass, which is the **Eddington luminosity**:

$$L_{\mathrm{Edd}} = 4\pi G M \frac{mc}{\sigma} \tag{232}$$

If the opacity is formed mainly by electron scattering, then the appropriate cross-section is σ_{T} and we get the maximum rate of radiation

$$\boxed{L_{\mathrm{Edd}} = 4\pi G M \frac{m_p c}{\sigma_{\mathrm{T}}} = 10^{4.5}\left(\frac{M}{M_\odot}\right) L_\odot} \tag{233}$$

(the particle mass is m_p rather than m_e because the protons supply effectively all the inertia, even though the electrons do all the scattering). As a side note: humans violate this limit by many powers of 10. The reason we aren't blown apart is that the analysis neglects other forces, particularly the interatomic bonds that hold us together. Since stars at the upper end of the main sequence satisfy

$$L/L_\odot \simeq \left(\frac{M}{M_\odot}\right)^3, \tag{234}$$

we see that stars cannot exceed about $100\,M_\odot$ without blowing themselves apart. This is very much an upper limit, since ionic opacity adds to Thomson scattering. In practice, O stars with $M \gtrsim 30\,M_\odot$ lose a large fraction of their mass through these radiation-driven winds in the course of their main-sequence lifetimes.

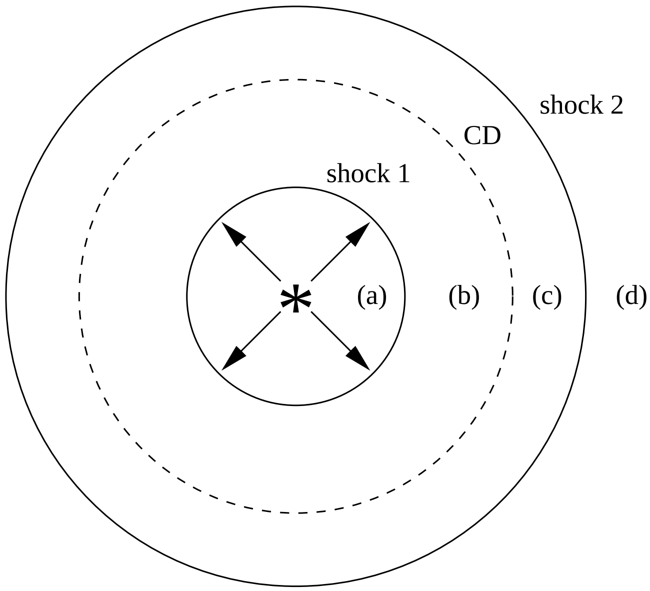

Figure 18. The geometry of a stellar wind. The central star (*) sends out a high-velocity wind in region (a). The overall effect is to drive a shock into the ISM (region d). This is shock 2. Behind this shock, there is shocked ISM, which is separated from wind material that passes through an inner shock (shock 1) by a contact discontinuity (CD).

17.2 Shock structure for a wind

These winds produce a hot bubble, even hotter than an HII region. In general, such winds are confined to the ionized interior of the HII region also produced by the stars. The high velocity gas drives a shockwave into the ISM. The wind itself forms a second shock against the post-shock ISM gas, so that four distinct regions may be defined:

(a) Unshocked stellar wind. S_1: shockfront of wind
(b) Post-shock stellar wind. CD: **contact discontinuity**: the pressure is smooth across CD.
(c) Post-shock ISM
(d) Undisturbed ISM

For this system, we can treat the spherical shocks as locally plane. The (adiabatic) shock jump conditions are

$$P_b = \frac{3}{4}\rho_a v_*^2; \quad v_b = \frac{1}{4}v_*; \quad \frac{\rho_b}{\rho_a} = 4; \quad kT_b = \frac{P_b}{\rho_b}\frac{m_{\rm H}}{2} = \frac{3}{32}m_{\rm H}v_*^2. \tag{235}$$

The post-shock stellar wind (region b) forms a high pressure region which compresses the post-shock ISM (region c) into a very thin layer. We may thus consider CD, (b) and S_2 to coincide at a distance R from the star, defining the surface of the wind bubble.

How fast does the wind bubble expand into the ISM? Consider conservation of momentum and energy.

(1) Conservation of momentum:

$$\frac{d}{dt}\left(\frac{4}{3}\pi R^3 \rho_0 \dot{R}\right) = 4\pi R^2 P_b; \tag{236}$$

in other words, rate of change of $M \times v$ is due to the net force on the ISM from the bubble. P_b is the wind pressure; ρ_0 is the undisturbed ISM density.

(2) Conservation of energy:

$$\frac{d}{dt}\left[\frac{4}{3}\pi R^3 \left(\frac{3}{2}P_b\right)\right] = \dot{E}_* - P_b \frac{d}{dt}\left(\frac{4}{3}\pi R^3\right); \tag{237}$$

or, in shorthand, $du/dt = \dot{E}_* - PdV/dt$: the rate of change of the total thermal energy of bubble ($\gg$ kinetic energy) equals the **mechanical luminosity** of wind ($\dot{E}_* = \dot{M}_* v_*^2/2$) minus the rate at which the hot bubble does work during its expansion.

We could solve these two equations directly, but it's much simpler to use dimensional analysis. First, note that P_b can be eliminated, leaving a relation between R, t, $\dot{E}_*$, and ρ_0 alone. Now

$$\left[\dot{E}_*\right] = \frac{M\,L^2}{t^3}; \quad [\rho_0] = \frac{M}{L^3}. \tag{238}$$

Clearly R should increase with $\dot{E}_*$ and decrease with ρ_0. To cancel m, we must take $\dot{E}_*/\rho_0$, which gives

$$\left[\frac{\dot{E}_*}{\rho_0}\right] = \frac{L^5}{t^3}. \tag{239}$$

We must therefore take the 1/5th root, leaving $t^{-3/5}$, and so

$$\boxed{R = \left(\frac{125}{154\pi}\right)^{1/5}\left(\frac{\dot{E}_*}{\rho_0}\right)^{1/5} t^{3/5}} \tag{240}$$

(where the coefficient of proportionality from a full solution has been inserted). The typical figures $\dot{E}_* = 10^{29}\,\mathrm{W}$ and $n_0 = 10^{10}\,\mathrm{m}^{-3}$ give $R = 10^{-3}\, t_{\rm yr}^{3/5}$ pc.

We can differentiate $R(t)$ to find the rate of advance of the bubble, which clearly declines as $t^{-2/5}$. For our reference figures, it is $\dot{R} = 650 t_{\rm yr}^{-2/5}\,\mathrm{km\,s^{-1}}$. Once $\dot{R} < c_s$, the wind dies. This occurs at $t \simeq 34,000$ yr at which point $R = 0.5$ pc.

18 Supernova Remnants

18.1 Types of supernovae

A more dramatic input of energy into the IGM comes from supernovae (SNe for short). These colossal explosions occur once every 30 years or so in a typical galaxy, and can briefly outshine all the stars in the galaxy that hosts them. Supernovae come in two-and-a-bit varieties, SNe Ia, Ib and II, distinguished according to whether or not they display absorption and emission lines of hydrogen. The SNe II do show hydrogen; they are associated with massive stars at the endpoint of their evolution, and are rather heterogeneous in their behaviour. The former, especially SNe Ia, are much more homogeneous in their properties; there is a characteristic rise to maximum, followed by a symmetric fall over roughly 30 days, after which the light decay becomes less rapid. Type Ib SNe are a complication to the scheme; they do not have the characteristic light curve, and also lack hydrogen lines.

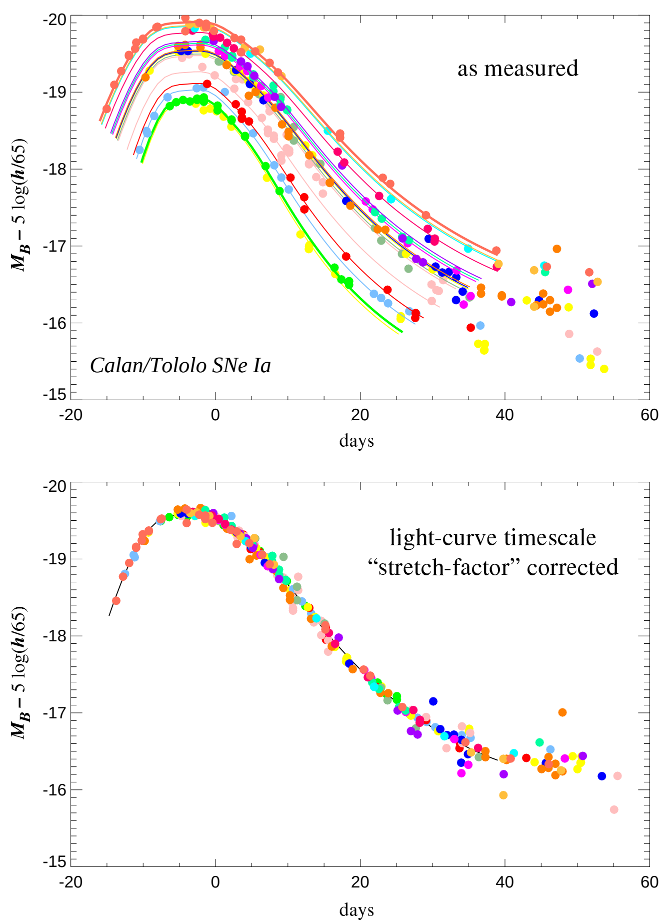

Figure 19. Illustrating the universal lightcurve of SNe Ia. The top panel shows raw results; the lower panel shows how all curves coincide after applying a 'stretch' to the time axis, which correlates with a shift in the luminosity axis (approximately the 1.7 power of the time stretch factor). After this calibration, the SNe are remarkably standard objects. Plot courtesy of Kim & the Supernova Cosmology Project.

Type Ia SNe are especially important in cosmology, since they are very nearly standard objects: their luminosity at maximum light is correlated with the speed of decline beyond maximum (more luminous objects shine for longer). By measuring the duration of the **lightcurve**, we can calibrate the luminosity, which empirically has a scatter of only 10%. Thus, the relative brightness of different supernovae can be used to infer their relative distances to a precision of 5%. This method has been the most important contributor to our knowledge of the extragalactic distance scale.

The reason that SNe Ia are standard objects is still somewhat uncertain. The favoured model is that the explosion results from a white dwarf that has accreted material from a companion star. Since white dwarfs are all close to the Chandrasekhar mass, the explosions tend to be similar. The lack of hydrogen in a white dwarf would account for the absence of hydrogen lines in the spectra of the SNe. In detail, different amounts of energy and mass are ejected by the explosion. The larger explosions expand for longer before the hot ejecta can become **optically thin** to scattering, allowing the radiation to escape. This is the origin of the correlation between luminosity and lightcurve width.

18.2 Initial phase: blastwave

A supernova ejects about half its mass in the initial explosion, propelling its outer shells into the ISM with typical speeds of $0.01c$. The total kinetic energy of the ejecta is $10^{43} - 10^{44}$ J $\simeq M_{\rm ej} v_s^2/2$ for $M_{\rm ej} \simeq 4M_\odot$, where v_s is the initial velocity of the supernova shock.

Let ρ_0 be the density of the gas surrounding the supernova. For an adiabatic shock, the post shock pressure is

$$P_1 = \frac{3}{4}\rho_0\, v_s^2, \tag{241}$$

so that the thermal energy per unit mass is

$$\epsilon_{\rm T} = \frac{3}{2}\frac{P_1}{\rho_1} = \frac{9}{8}\frac{\rho_0}{\rho_1}v_s^2 = \frac{9}{32}v_s^2 \tag{242}$$

(since $\rho_1/\rho_0 = 4$). The kinetic energy per unit mass is

$$\epsilon_{\rm K} = \frac{1}{2}\left(\frac{3}{4}v_s\right)^2 = \frac{9}{32}v_s^2 \tag{243}$$

(since $v_1 = 3v_s/4$). The amount of mass swept up by the shock is

$$M_s = \frac{4}{3}\pi\rho_0 R^3, \tag{244}$$

so the total energy is

$$E_{\rm tot} = \frac{4}{3}\pi R^3 \rho_0\,(\epsilon_{\rm T} + \epsilon_{\rm K}) = \frac{3}{4}\pi\rho_0 R^3 v_s^2. \tag{245}$$

Now, $v_s = \dot{R}$, and $E_{\rm tot} = E_{\rm SN}$ = blast energy, so

$$E_{\rm SN} = E_{\rm tot} = \frac{3\pi}{4}\rho_0 R^3 \dot{R}^2. \tag{246}$$

We have an equation for R in terms of $E_{\rm SN}$, ρ_0, and t. Once again, dimensional analysis gives us

$$[E_{\rm SN}] = \frac{M\,L^2}{t^2}; \quad [\rho_0] = \frac{M}{L^3}, \tag{247}$$

so

$$\boxed{R = \text{constant}\left(\frac{E_{\rm SN}}{\rho_0}\right)^{1/5} t^{2/5}.} \tag{248}$$

Solving the equations in detail gives

$$\text{constant} = \left(\frac{25}{3\pi}\right)^{1/5} = 1.2. \tag{249}$$

Knowing $R(t)$, we differentiate to get the shock speed:

$$v_s = \dot{R} = \frac{2}{5}\left(\frac{25}{3\pi}\right)^{1/5}\left(\frac{E_{\rm SN}}{\rho_0}\right)^{1/5} t^{-3/5}. \tag{250}$$

Taking $E_{\rm SN} = 10^{44}$ J, $n_0 = 10^6\,{\rm m}^{-3}$, we get

$$\begin{aligned} R &\simeq 0.35\, t_{\rm yr}^{2/5}\ {\rm pc} \\ \dot{R} &\simeq 140,000\, t_{\rm yr}^{-3/5}\,{\rm km\,s}^{-1} \\ M_{\rm shell} &= \frac{4}{3}\pi R^3 \rho_0 \simeq 0.004\, t_{\rm yr}^{6/5}\, M_\odot \end{aligned} \tag{251}$$

(because $\rho_0 \simeq 0.025 M_\odot\,{\rm pc}^{-3}$ for $n_{\rm H} = 10^6\,{\rm m}^{-3}$). Note that this solution must break down at early times, as it predicts a velocity that diverges as $t \to 0$. In fact, the similarity solution only starts to apply at about 600 yr, where $v_s \simeq 0.01c$ is predicted (the initial ejection velocity). At this time, the predicted shell mass is of order the mass of the original ejecta: another way of looking at this is that the similarity solution sets in once the blastwave has swept up much more mass than was originally ejected.

18.3 Radiative phase

Initially the post-shock gas temperature is

$$kT_1 = \frac{3}{32} m_{\rm H} v_s^2 \tag{252}$$

(allowing for $\mu = 1/2$: e^- & H^+), so $T_1 \simeq 10^8$ K for $v_s = 3000\,{\rm km\,s}^{-1}$. The temperature decreases with time like v_s^2:

$$T_1 = \frac{3}{32}\frac{m_{\rm H}}{k}\frac{4}{25}\left(\frac{25}{3\pi}\right)^{2/5}\left(\frac{E_{\rm SN}}{\rho_0}\right)^{2/5} t^{-6/5} \simeq 5\times 10^7 t_{1000\,{\rm yrs}}^{-6/5}\ {\rm K}. \tag{253}$$

This assumes an adiabatic shock – i.e. that radiative cooling behind the shock can be neglected. To see whether this is true, we have to consider the interstellar **cooling curve**, shown in figure

20. Here, one expresses the volume emissivity, ϵ (power radiated from unit volume), in terms of a constant times the square of the hydrogen density:

$$\boxed{\epsilon = \Lambda(T)\, n_{\rm H}^2.} \tag{254}$$

The density-squared law applies to the collisional processes that we studied previously. It also applies to **bremsstrahlung**, or free-free radiation: radiation emitted from electrons accelerated by the electrostatic field of ions. This process dominates at high temperatures, $T \gtrsim 10^8$ K. At lower temperatures, collisional processes dominate (especially from oxygen), and the cooling curve has a sharp peak at $T \simeq 3\times10^5$ K. Once the post-shock material reaches this temperature, it will very rapidly radiate away all its internal energy, to reach a temperature of order 10^4 K – similar to much of the undisturbed ISM, for the same reason. This occurs at $t = 70,000$ yrs, $R \simeq 30$ pc. At this time $v_s \simeq 170\,{\rm km\,s^{-1}}$.

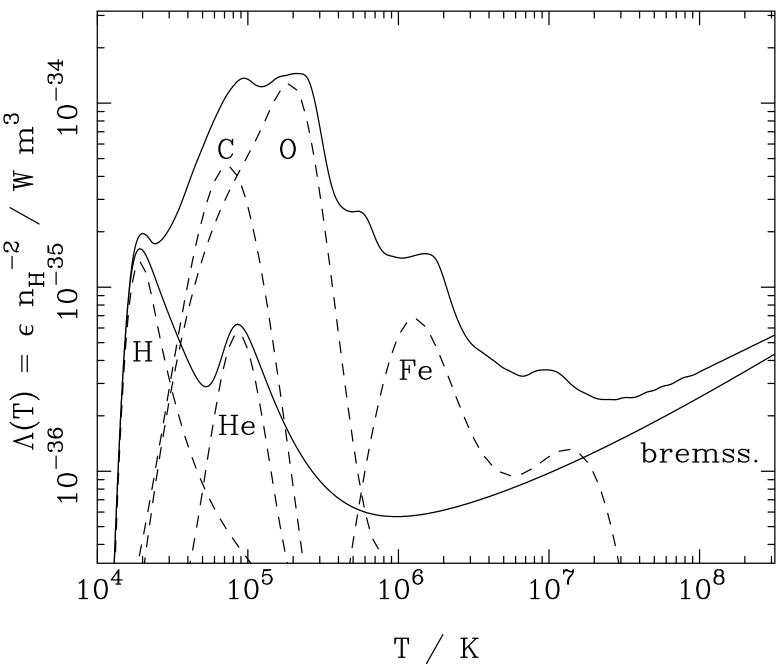

Figure 20. The cooling curve, defined such that the volume emissivity of the plasma, ϵ is $\epsilon \equiv \Lambda(T) n_{\rm H}^2$. The two curves show results for hydrogen+helium only, and including Solar metals. The emissivity falls below 10^4 K, as the plasma becomes neutral. Above that temperature, collisional excitation of ions dominates, with bremsstrahlung or free-free radiation taking over for $T \gtrsim 10^8$ K.

The **cooling time** is

$$t_{\rm cool} = \frac{\frac{5}{2}(2n_{\rm H})kT}{n_{\rm H}^2\Lambda(T)} \simeq 2\times10^{11}\,{\rm s} = 6500\,{\rm yr} \tag{255}$$

(using the fact that $\Lambda(T) \simeq 10^{-34}\,{\rm W\,m^3}$ at the peak of the cooling curve, and taking $n = n_p = n_e = 10^6\,{\rm m^{-3}}$). The shock moves a distance

$$\text{cooling length} \equiv \ell_{\rm cool} = v_s\, t_{\rm cool} \simeq 1\,{\rm pc} \tag{256}$$

in this time, picking up an additional amount of material

$$\boxed{M_{\rm cool} = 4\pi R^2\, \ell_{\rm cool}\, \rho_0 \simeq 300 M_\odot \gg M_{\rm ej}.} \tag{257}$$

All the mass of the **supernova remnant (SNR)** is in the shell, which has now radiated away all its pressure.

18.4 Momentum-conserving ('snowplough') phase

The total mass in the SNR at this time is

$$M_{\rm SNR} = \frac{4\pi}{3} R^3 \rho_0 \simeq 3000 M_\odot \tag{258}$$

(for $R = 30$ pc and $n_0 = 10^6\,{\rm m}^{-3}$). Although the blastwave pressure has been radiated away, the SNR continues to expand by momentum conservation:

$$\frac{4\pi}{3} R^3 \rho_0 \dot{R} = \mathcal{M}_0 = \text{constant}. \tag{259}$$

If we wait long enough that the size and mass of the SNR are much greater than the size and mass at the end of the radiative cooling phase, then dimensional analysis gives

$$[\mathcal{M}_0] = \frac{M\,L}{t} \quad \text{and} \quad [\rho_0] = \frac{M}{L^3}, \tag{260}$$

so that

$$\boxed{R = \text{constant} \left(\frac{\mathcal{M}_0}{\rho_0}\right)^{1/4} t^{1/4}} \tag{261}$$

and

$$\dot{R} = \frac{1}{4} \text{constant} \left(\frac{\mathcal{M}_0}{\rho_0}\right)^{1/4} t^{-3/4} \simeq 4000\, t_{1000\,{\rm yr}}^{-3/4}\,{\rm km\,s^{-1}}. \tag{262}$$

Eventually, $\dot{R}$ reaches the ISM sound speed, and the SNR dissipates. For $c_s^2 = (5/3)kT/\bar{m}$, and taking $T = 10^4$ K and $n_0 = 10^6\,{\rm m}^{-3}$, we get $c_s \simeq 17\,{\rm km\,s^{-1}}$, so the SNR dies after about 10^6 years.

18.5 Filling factor of SNe

What fraction of the disk of the Milky Way is in a SNR? A typical SNR radius is 100 pc and will last about 10^6 years before dissipating. We estimate about 1 SN every 30 yrs, so in 10^6 yrs

$$N_{\rm SNR} = \frac{10^6}{30} = 3.3 \times 10^4 \tag{263}$$

SNR at any one time. The volume occupied by these SNR is then

$$V_{\rm SNR} = N_{\rm SNR} \times \frac{4\pi}{3} R^3_{\rm SNR} = 140\ {\rm kpc}^3. \tag{264}$$

The radius of the Galaxy is $\simeq$ 15 kpc, and its thickness is about 200 pc = 0.2 kpc, so its volume is $\simeq \pi(15\,{\rm kpc})^2 \times 0.2\,{\rm kpc} = 140\,{\rm kpc}^3$.

The mean number of SNR at any given point is thus:

$$\frac{V_{\rm SNR}}{V_{\rm Galaxy}} \simeq 1. \tag{265}$$

Since SNR can overlap, the volume of the Galaxy in SNR is actually smaller: Poisson statistics $\Rightarrow$ probability of no SNR at a given point is $\exp(-v_{\rm SNR}/v_{\rm Galaxy})$, and so the fraction of the Galaxy filled by one or more SNR is

$$f_{\rm SNR} = 1 - \exp(-v_{\rm SNR}/v_{\rm Galaxy}) \simeq 0.6 \tag{266}$$

So most of the Galaxy, including probably us, is in a SNR.

This sum only counts the *active* SNR (i.e. those that exploded in the last 10^6 years). Since the galaxy is of order 10^{10} years old, we see that our neighbourhood has been affected by SNe at least 1000 times. One of the most important functions of supernovae in the ecology of the galaxy is to distribute the heavy elements that are made in the precursor star, thus **enriching** the ISM. This is the reason that the Sun contains metals even though it is young: it formed from gas that had been enriched by supernova explosions. The same also applies to the planets, which only form because the pre-Solar nebula is already rich in metals.