STELLAR EVOLUTION. I: LIFE ON THE MAIN SEQUENCE

12.1   Introduction

In lectures 9 and 10 we developed stellar models for stars in hydrostatic equilibrium, and found that a star of a given mass and composition has a unique, fully determined structure. However, the nuclear reactions which power a star by fusing four hydrogen nuclei into a helium nucleus deep in the stellar core change the initial chemical composition of the star over time. Furthermore, convection may set in at some radius, and mix processed and unprocessed gas. As we saw, the equations of pressure, opacity and nuclear power density all depend on the chemical composition of the
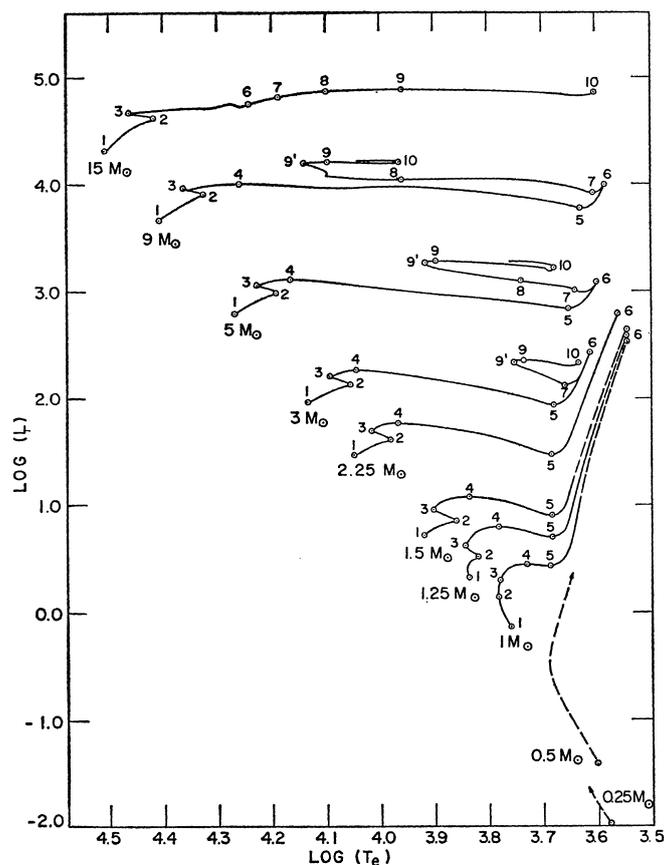
Figure 12.1: Theoretical evolutionary tracks for stars of different masses calculated by Iko Iben in 1967 (ARAA, 5, 571). The numbered stages are referred to in the text. The y-axis units are $L_\odot$.

gas. It is therefore inevitable that stars evolve with time. In the next few lectures we shall look in some detail at the process of stellar evolution.

Stellar evolution, as opposed to equilibrium, can be reproduced in our computers by solving a series of equilibrium stellar models—normally referred to as a stellar evolution track—in which one updates in steps the gradual enrichment by elements heavier than hydrogen at different radii within the star. One such set of tracks is shown in Figure 12.1.

The net effect of stellar evolution is to move a star away from the main sequence onto other regions in the H-R diagram. Observations immediately tell us that stars spend most of their lives on the main sequence, and that the subsequent evolution is relatively short-lived. This is evident as soon as one plots on the H-R diagram all of the stars within a given volume (see Figure 12.2): the number of stars away from the main sequence is much smaller than the number of stars on the main sequence.

How do we know that our computer models of stellar evolution are realistic? The requirement of course is that they must match the data, in this case provided by the H-R diagrams of clusters, where all the stars are assumed to
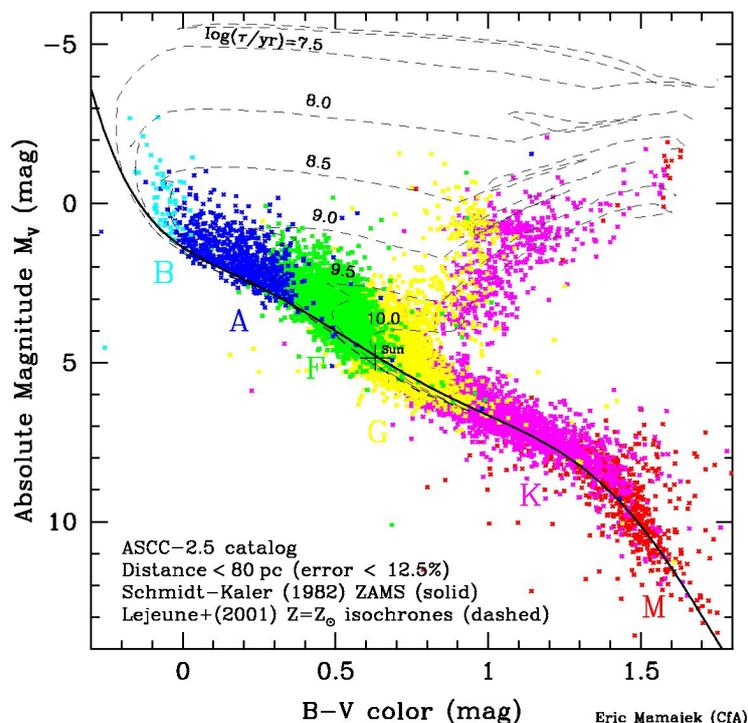
Figure 12.2: Colour-magnitude diagram ($B - V$ vs. M_V) for stars within 80 pc of the Sun, with colour-coding by spectral type. Most stars are found on the main sequence.
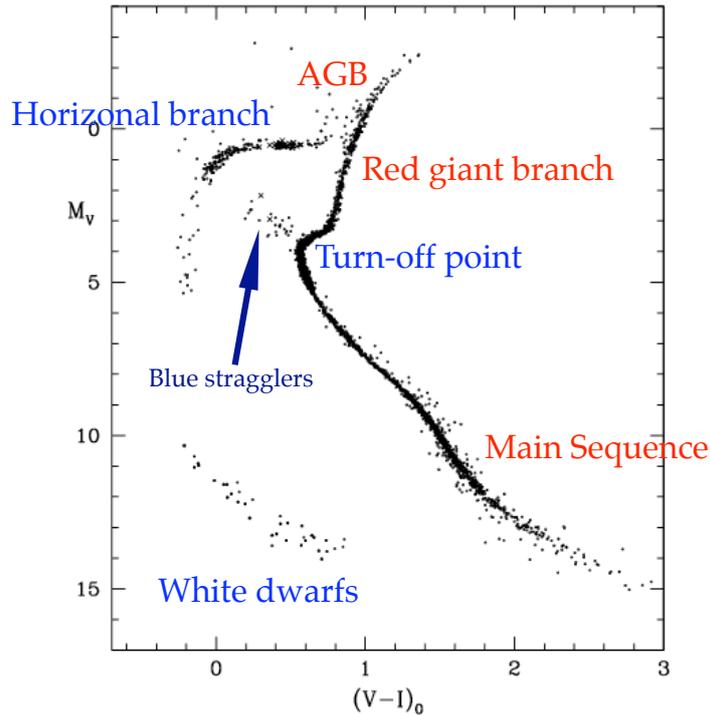
Figure 12.3: Hertzsprung–Russell diagram of a typical globular cluster.

have formed more or less at the same time. This is an adequate assumption in most cases, given the long timescales of stellar evolution, $t = \tau_{\mathrm{nuclear}}$, compared to the star formation timescales, $t \approx \tau_{\mathrm{KH}}$, as we saw in Lecture 7.1 and 7.2.

Particularly useful in this respect are the H-R diagrams of halo globular clusters which, being among the oldest stellar systems known, give us a view of the late stages in the evolution of long-lived stars with masses comparable to that of the Sun (see Figure 12.3).

The combination of computer modelling and observations has led to a well defined picture of stellar evolution. As you may well imagine (at least if you have been following these lectures so far!), mass is the dominant parameter determining not only the length of time a star spends on the main sequence, but also the path that it follows once it evolves off the main sequence. Figure 12.4 gives a schematic summary of the most important evolutionary stages of stars of different masses together with some of their observational characteristics.
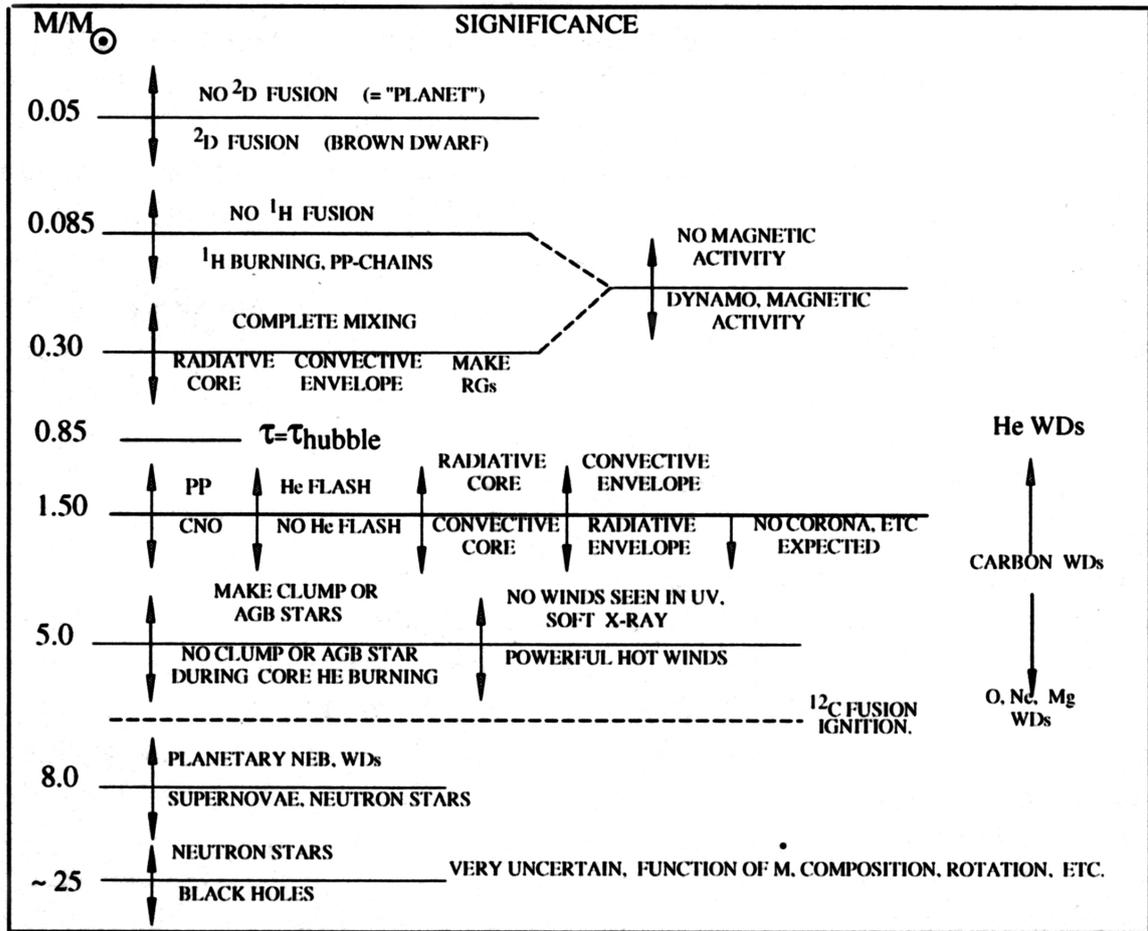
Figure 12.4: Diagram showing the main evolutionary stages of stars in various mass ranges. (Reproduced from Hansen et al. 2004, Stellar Interiors, Springer).

12.2 Life on the Main Sequence

The main sequence has an intrinsic width, even after accounting for errors in the photometry, reddening correction, distance and other factors which may be of importance. The finite width arises because the internal structure of a star does change (slowly) even while it is on the hydrogen-burning main sequence. We look at the factors involved using a $1 M_\odot$ star as an example.

As H is converted into He deep in the stellar core via the p-p chain, the mean molecular weight μ:

$$\mu = \frac{4}{6X + Y + 2}$$

of the plasma increases (as the He fraction, Y, increases at the expense of the H fraction, X). As a consequence, the pressure would also decrease in

accordance with the ideal gas law:

$$P = \frac{\rho k T}{\mu m_H} \equiv \mathcal{R}\,\frac{\rho T}{\mu} \qquad (12.1)$$

and become insufficient to support the overlying layers of the star, unless either ρ or T, or both increase. The result therefore is that the core will shrink, and its density will increase. In the process, gravitational energy is released, half of which will go into heating the plasma while half of it is radiated away (recall the virial theorem).

In Lecture 7 we saw that rate of energy release via the p-p chain per unit mass of nuclear fuel is:

$$\mathcal{E}_{pp} \propto X^2\,\rho\,T^4 \qquad \mathrm{erg\,s^{-1}\,g^{-1}}\,.$$

The higher values of T and ρ more than offset the decrease in X; the resulting increase in $\mathcal{E}_{pp}$ in turn increases the radius of both the core and the envelope, and therefore the luminosity L, of the star. As a consequence, our $1M_\odot$ star will move upwards on the H-R diagram. Referring to Figure 12.1, the portion of the $1M_\odot$ track between points 1 and 2 corresponds to the evolution just described, which takes place over about half of the main sequence lifetime; thus point 2 corresponds roughly to the location of the Sun today, $\sim 30\%$ more luminous that it was on its arrival on the ZAMS.

12.2.1   H shell burning

Figure 12.5 shows the depletion of H from the core of a $1M_\odot$ star with time. At $t \simeq 1 \times 10^{10}\,\mathrm{yr}$, most of the H burning has become confined to a thick shell surrounding a small but growing He core (point 3 on the $1M_\odot$ track in Figure 12.1). The development of a chemically inhomogeneous structure marks the beginnings of the end of the main sequence life of the star.

The temperature of the He core is insufficient at this stage to ignite He burning. Without nuclear reaction taking place in the He core, the luminosity of the core is $L_{\mathrm{He\,core}} \simeq 0$. Therefore, according to Eddington's equation for radiative equilibrium (eq. 8.1), the temperature gradient vanishes: $dT/dr \simeq 0$—that is, the He core is now essentially isothermal.
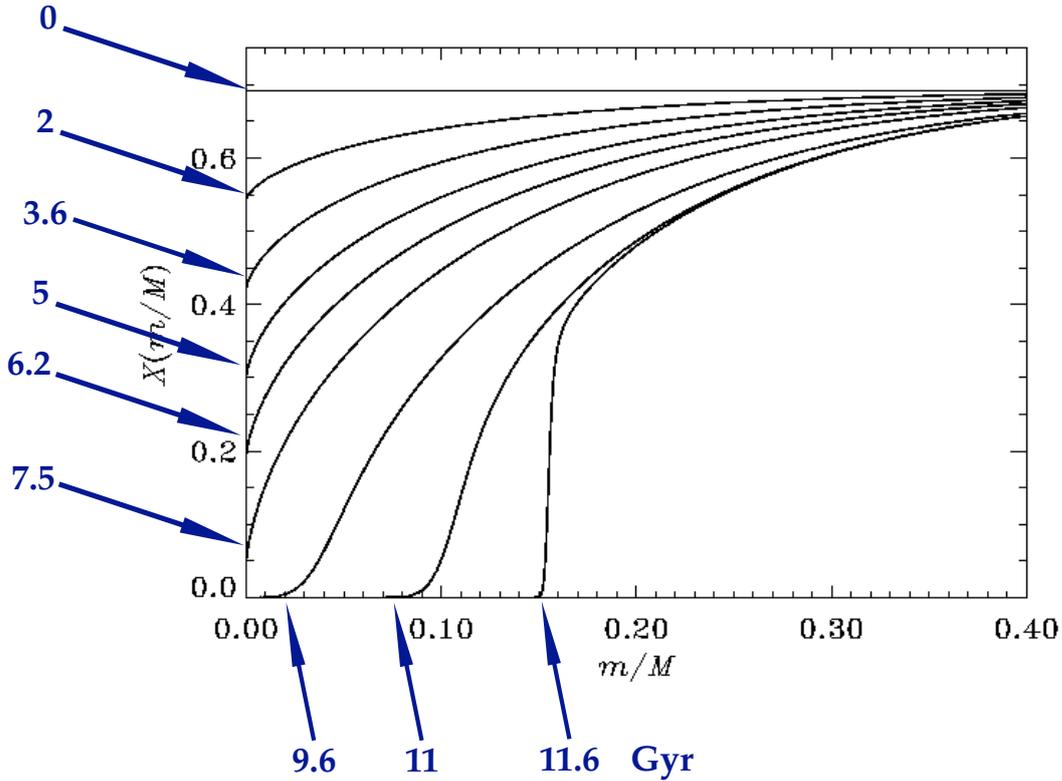
Figure 12.5: Hydrogen profiles showing the gradual exhaustion of H in the core of a $1M_\odot$ star. The homogeneous initial model consists of a mixture of H, He and metals with H mass fraction $X = 0.69$. The plot shows X as a function of stellar mass fraction m/M for nine models at different ages in Gyr after the onset of H burning, as indicated. The model at 5×10^9 yr corresponds roughly to the present-day Sun; the last two models are in the shell hydrogen burning phase. (Reproduced from Christensen-Dalsgaard: Lecture Notes on Stellar Structure and Evolution).

Recall our conclusion in Lecture 8.4.1: in order for a star to be in hydrostatic equilibrium, a negative pressure gradient must exist within the star:

$$\frac{dP}{dr} = -\rho\, g\,. \tag{12.2}$$

Differentiating eq. 12.1, we find:

$$\frac{dP}{dr} = \frac{\mathcal{R}\rho}{\mu} \frac{dT}{dr} + \frac{\mathcal{R}T}{\mu} \frac{d\rho}{dr}\,. \tag{12.3}$$

If $dT/dr \simeq 0$, the entire pressure gradient necessary to support the star in pressure equilibrium must be provided by the density gradient. This can work, provided the isothermal core contains only a small fraction of the total stellar mass. We'll return to this point presently.

At this stage, energy generation is taking place in a thick hydrogen-burning shell. The temperature in the shell increases as the core contracts, leading

to higher rates of energy generation. Some of this energy reaches the surface as radiation, and some goes into a slow expansion of the envelope, causing the effective temperature to decrease (recall that the luminosity of a blackbody is given by $L = 4\pi R^2 \sigma T^4$). The net effect is that the evolutionary track bends to the right in the H-R diagram (see Figure 12.1).

12.3   Main Sequence Evolution of Massive Stars

The evolution of more massive stars on the main sequence differs from that of solar mass stars in one important aspect: stars more massive than the Sun have convective cores, as we saw in Lecture 8.3.1. The timescale for convection, defined as the time it takes for a convective element to travel one mixing length (Lecture 8.5), is considerably shorter than the nuclear timescale. Consequently, material is well mixed in the convective core, keeping its composition nearly homogeneous. So, while hydrogen is being converted to helium more quickly in the center of the star where the temperature is greatest (recall the steep temperature dependence of the energy generation rate in the CNO cycle), hydrogen is depleted more or less uniformly throughout the core. In the case of massive stars with convective cores, the core actually *shrinks* as hydrogen is depleted (see Figure 12.6 and compare it with Figure 12.5).

Why does the core shrink? It is primarily because the *opacity* within the core is decreasing, and the lower the opacity, the easier it is for radiation to escape and convection is no longer needed to transfer energy efficiently. The opacity in the core is due primarily to electron scattering which, as we saw in Lecture 5.5, has the form:

$$\kappa_{es} = \kappa_{0,es} \frac{1}{\mu_e} \simeq \kappa_{0,es} \frac{1}{2}(1+X) \qquad (12.4)$$

where $1/\mu_e$ is the number of electrons per nucleon. As hydrogen is converted into helium, positrons are produced. These annhiliate with electrons, reducing the number of electrons available for scattering and thus reducing the opacity.

It can also be realised from inspection of Figure 12.1 that T_{eff} changes differently in massive stars and low mass stars, as the stars evolve on the main sequence. In massive stars T_{eff} *decreases* while in stars with
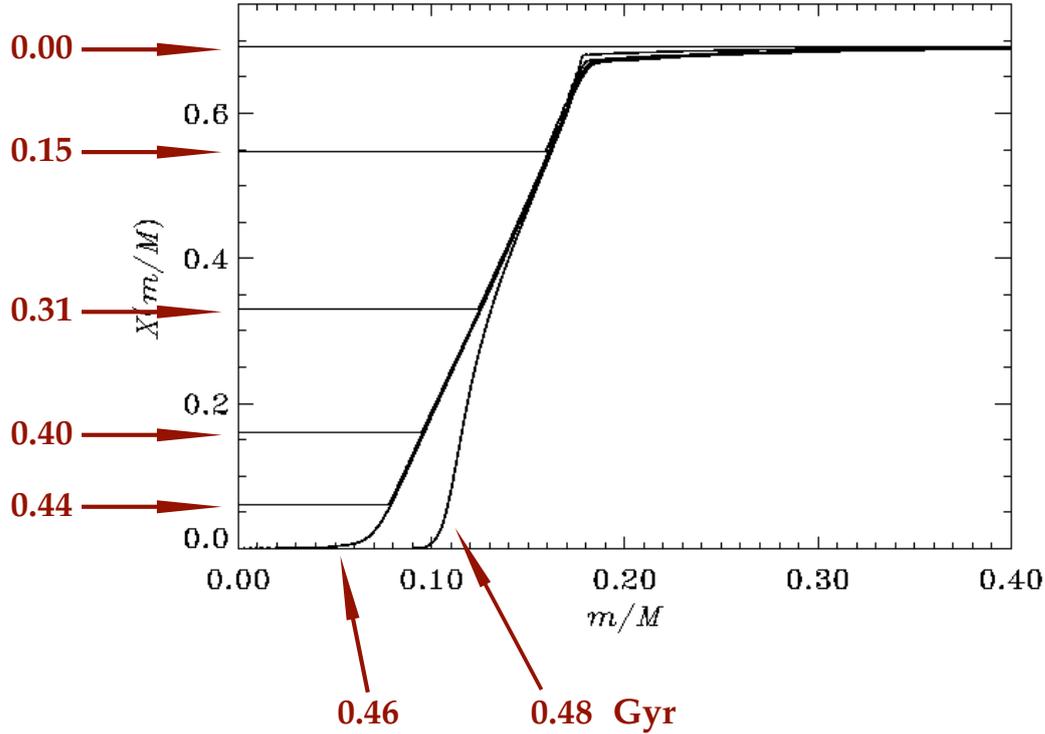
Figure 12.6: Hydrogen profiles showing the gradual exhaustion of H in the core of a $2.5 M_\odot$ star. The homogeneous initial model consists of a mixture of H, He and metals with H mass fraction $X = 0.69$. The plot shows X as a function of stellar mass fraction m/M for nine models at different ages in Gyr after the onset of H burning, as indicated. The last model is in the shell hydrogen burning phase, the helium core having grown substantially beyond the smallest extent of the convective core. (Reproduced from Christensen-Dalsgaard: Lecture Notes on Stellar Structure and Evolution).

$M \leq 1 M_\odot$ T_{eff} *increases* during their main sequence lifetimes. All stars experience expansion of their outer layers throughout their main sequence evolution. Low mass stars experience only a modest expansion, so the energy produced by changes in the core leads to an overall increase in luminosity and surface temperature. More massive stars experience a more rapid expansion. The energy goes into moving the outer layers of the star and results in an overall decrease in effective temperature. Recalling the expression for blackbody luminosity, if the luminosity increases only slowly and the radius increases quickly, the effective temperature must decrease.

In the final stages of H fusion in massive stars, the whole star contracts in an attempt to maintain the energy production by increasing the core temperature. This produces the 'left hook' in the H-R diagram visible in Figures 12.1 and 12.7.
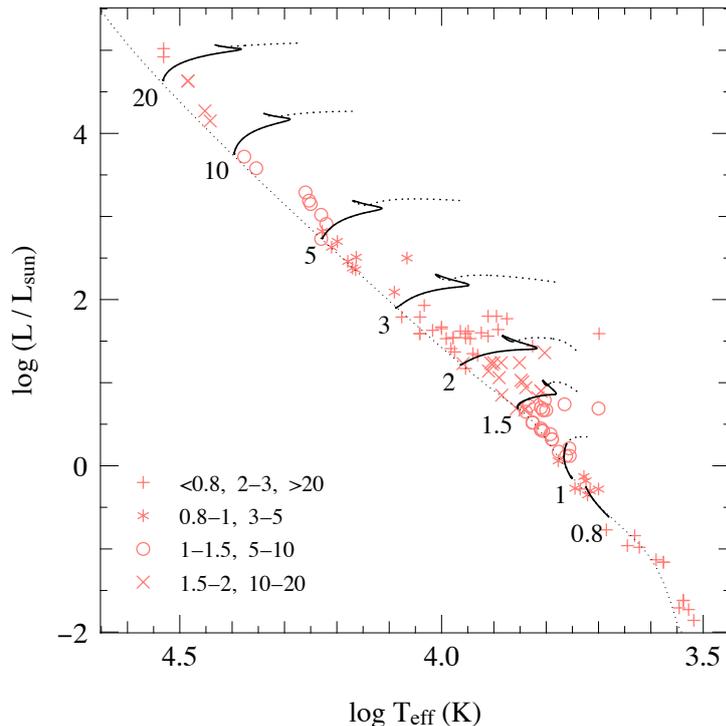
Figure 12.7: Evolutionary tracks in the HR diagram during central hydrogen burning for stars of various masses, as labelled in the legend (in units of $M_\odot$). The thin dotted line is the ZAMS, while the dotted portion of each track shows the continuation of the evolution after central hydrogen exhaustion. The evolution of the $0.8 M_\odot$ star is terminated at an age of $14\,\mathrm{Gyr}$. Symbols show the locations of binary star components with accurately determined masses, luminosities and radii. Note that each symbol is used to indicate more than one mass range.

12.4   The Mirror Principle of Stars with Shell Burning

Whenever a star has a shell-burning source, it appears to act like a mirror:

" When the region within a burning shell contracts, the region outside the shell expands; and when the region inside the shell expands, the region outside the shell contracts."

This is also known as the "shell burning law". It is not a physical law as such, but an empirical observation, supported by the results of numerical simulations. Although the detailed physics is complicated, we can try and understand it in simple terms as follows. If the virial theorem holds and the total stellar energy remains constant, then the gravitational and thermal energy are each conserved. Any contraction of the core must be accompanied by the expansion of the envelope to conserve the gravitational
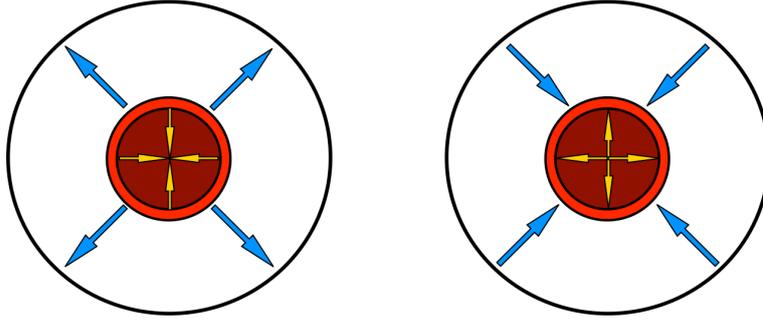
Figure 12.8: The 'mirror principle' of shell-burning stars.

potential energy. At the same time, the heating of the core must result in a cooling of the envelope for the thermal energy to be conserved. If the total energy does not remain constant, but $L_{\mathrm{nuc}} > L$, then the envelope will expand considerably on core contraction.

We can state the above a little more quantitatively, as follows. If we consider timescales shorter than the Kelvin-Helmholtz timescale, then both energy conservation

$$\langle U \rangle + \langle K \rangle = \mathrm{constant}$$

and the virial theorem

$$\langle U \rangle + 2\langle K \rangle = \mathrm{constant}$$

must be satisfied, which means that $\langle U \rangle$ and $\langle K \rangle$ must be conserved separately. For $M_c \gg M_{\mathrm{env}}$, we have:

$$|U| \approx \frac{GM_c^2}{R_c} + \frac{GM_cM_{\mathrm{env}}}{R} \tag{12.5}$$

where R is the radius of the star and we have assumed that the binding energy of the envelope is dominated by the gravity of the core. If we take M_c to be constant (and therefore M_{env} too), then:

$$-\frac{GM_c^2}{R_c^2}\frac{dR_c}{dt} - \frac{GM_cM_{\mathrm{env}}}{R^2}\frac{dR}{dt} = 0 \tag{12.6}$$

which implies:

$$\frac{dR}{dR_c} = -\left(\frac{M_c}{M_{\mathrm{env}}}\right)\left(\frac{R}{R_c}\right)^2 \tag{12.7}$$

i.e. the envelope expands as the core contracts and vice versa.

12.5   The Schönberg-Chandrasekhar Limit

We mentioned earlier that once an isothermal He core has developed, hydrostatic equilibrium is maintained by the density gradient alone (eq. 12.3). In 1942, Schönberg and Chandrasekar (ApJ, 96, 161) showed that there is a limit to the fraction of stellar mass that can be supported by an isothermal core, given by:

$$\left(\frac{M_{\text{ic}}}{M}\right)_{\text{SC}} \simeq 0.37 \left(\frac{\mu_{\text{env}}}{\mu_{\text{ic}}}\right)^2 \approx 0.1 \qquad (12.8)$$

where μ_{env} and μ_{ic} are the mean molecular weights of the envelope and core respectively. If the mass of the core exceeds the Schönberg-Chandrasekhar limit, pressure support from the core can no longer hold the overlying stellar envelope. As a result, the core collapses on a Kelvin-Helmholtz timescale and the star evolves very rapidly relative to the nuclear timescale of main sequence evolution. This occurs at the points labelled 4 in Figure 12.1.

The mass of an isothermal core may exceed the Schönberg-Chandrasekhar limit if an additional source of pressure can be found to supplement the ideal gas pressure. In stars of relatively low mass ($M \lesssim 2M_\odot$), where the helium core is relatively cool and dense, the electrons in the gas start to become *degenerate* and degeneracy pressure can become an important additional source of pressure.

12.6   Electron Degeneracy

Recall that electrons are *fermions*, particles with half-integer spin (angular momentum in units of $\hbar$), as opposed to *bosons* which are particles with integer spin. Fermions and bosons obey very different statistical rules on how they occupy quantum mechanical states.

The state of a free particle in 1-D space may be described by its position x and its momentum p_x; together, these two quantities define a position in a two-dimensional phase space (see Figure 12.9). An area in this 2-D phase space has the dimensions of energy × time (equivalent to momentum × length). Quantum mechanics tells us that, on tiny scales, this phase space is quantized, that is it consists of discrete cells or states, each of area equal

to Planck constant:

Area : $\Delta A = \Delta p_x \Delta x = h$

Only two electrons
permitted in
area $\Delta A = h$

p_x

x
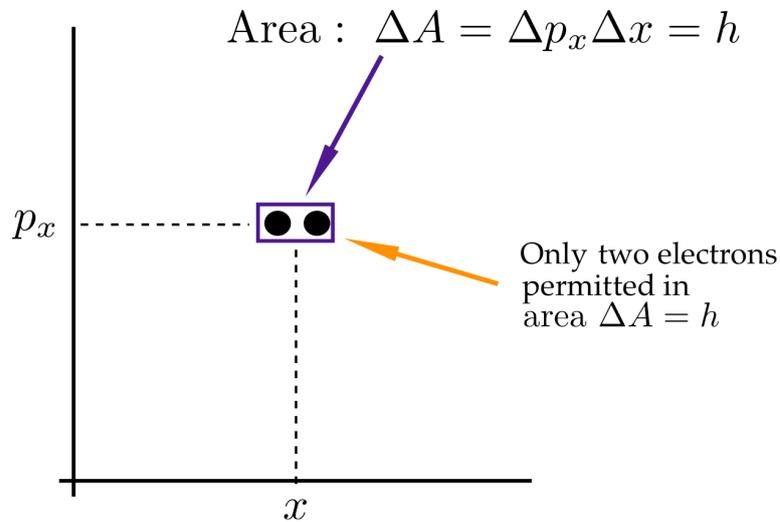
Figure 12.9: Two-dimensional phase space (x, p_x) for a one-dimensional gas showing an area of $h = 6.6 \times 10^{-27}$ erg s (Planck constant). This area is a quantum state that can contain no more than two electrons.

The *Pauli exclusion principle* dictates that no more than one fermion of a given spin can occupy a given state in the x, p_x phase space. Thus, at most two electrons, with opposite spins, can occupy a given state of area h. In 3-D space, phase space has six dimensions: x, y, z, p_x, p_y, p_z, so that a phase space quantum state has a 6-D volume:

$$\Delta x \, \Delta y \, \Delta z \, \Delta p_x \, \Delta p_y \, \Delta p_z = h^3$$

with only two half-spin particles permitted to occupy such a volume element.

A gas of fermions with sufficiently low temperature and/or sufficiently high density will fill all the lowest momentum states. In such degenerate conditions, the momentum (or energy) distribution differs markedly from the Maxwell-Boltzmann distribution of an ideal gas, which describes the situation when the particles occupy only a very small portion of the lowest states. The electrons are forced to occupy higher momentum states than they would normally occupy, resulting in an abnormally high pressure.

The situation is illustrated in Figure 12.10 where panel (a) corresponds to non-degenerate gas. In this case, the greatest density of electrons is in

the lowest momentum state, in accord with the 1-D Maxwell-Boltzmann distribution:

$$\mathcal{P}(p_x)\, dp_x = \left(\frac{1}{2\pi\, mkT}\right)^{1/2} \exp\left(-\frac{p_x^2}{2mkT}\right) dp_x \qquad (12.9)$$

shown on the right side of panel (a). Also indicated in panel (a) is the momentum p_x at which the kinetic energy is at its average value, $p_x^2/2m = kT/2$. The key to the non-degenerate character is that the lowest-momentum states are not all fully occupied—some are empty and some contain only one electron. In reality, in a truly non-degenerate situation, the electrons would occupy only $< 1\%$ of these states, owing to their thermal content.

Consider now what happens if the density is increased significantly (panel b). The gas is 'squeezed' by reducing the length of physical space Δx
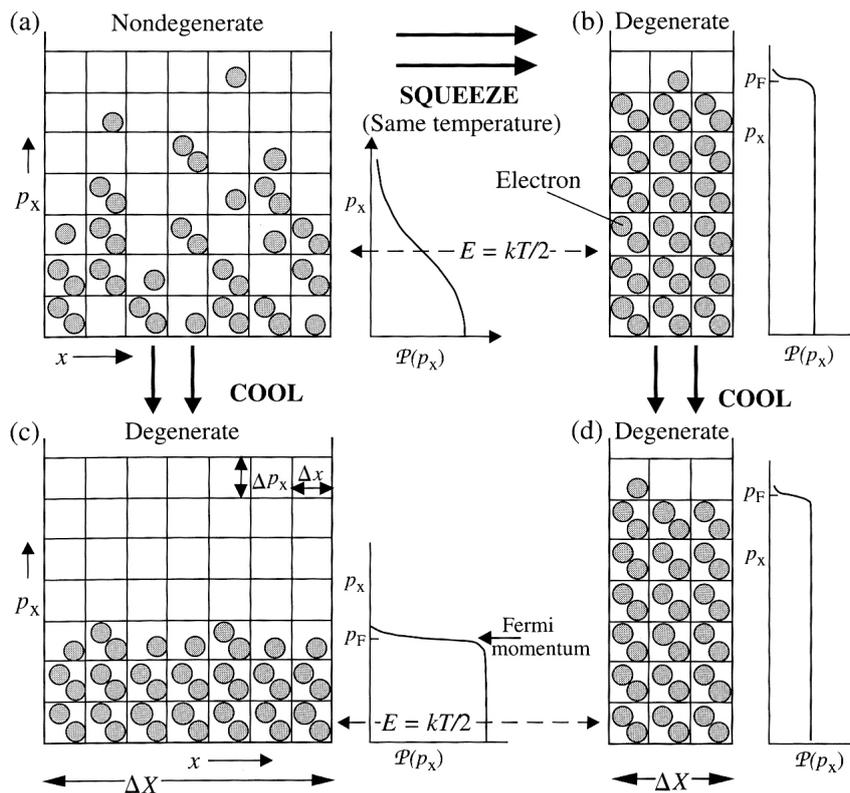
Figure 12.10: The two dimensions of phase space, x and p_x, for a one-dimensional gas of fermions containing 37 particles, for different temperatures and physical lengths ΔX. (a): non-degenerate gas; (b), (c) and (d): fully degenerate gas. The electrons in a degenerate gas are forced to much higher average energies than $kT/2$ because of a shortage of available states at lower energies. (Figure reproduced from Bradt, H., Astrophysics Processes, CUP).

available to it. In the compressed state, very few particles can occupy the lowest-momentum states, because there are not many of them. Electrons are therefore forced to abnormally high energies for a given temperature and the electron gas is said to be *degenerate*. Note that the distribution of particle momenta, $\mathcal{P}(p_x)$, shown on the right of panel (b), is no longer a Gaussian as in eq. 12.9. Rather, electrons are packed solidly up to a maximum value of momentum called the *Fermi momentum*, p_F. It is these high momenta that give rise to degeneracy pressure.

12.6.1   Pressure of a Degenerate Gas

What would be the result if the gas in panel (b) 'cools down' to a lower temperature as in panel (d)? The electrons cannot move to lower energy states, because they are all occupied; they stay very nearly in the same quantum states as before, exerting the same pressure as before. Thus, in a completely degenerate gas, the pressure is independent of temperature and is a function of density only:

$$P \propto \rho^{5/3} \tag{12.10}$$

What does 'cooling down' mean in these circumstances? The answer lies in the presence of heavier particles: protons, neutrons and atomic nuclei. When the gas is non-degenerate, equipartition of energy dictates that the average electron energy, $p_x^2/2m$, will be the same as the average proton energy. But, the mass of a proton is 1836 times greater than the mass of an electron. Consequently, p_x will be $\sqrt{1836} \simeq 43$ times higher for a proton than for an electron. In the core of a star near the end of its main sequence lifetime, p_x will be two orders of magnitude higher for the He nuclei than for the electrons.

The protons are also constrained by the Pauli exclusion principle, but they occupy much higher p_x states than the electrons. Consequently, as cooling or squeezing takes place, the electrons reach degeneracy long before the protons do. The normal Maxwell-Boltzmann distribution of proton energies serves to define the temperature. Thus, a *partly* degenerate gas does in fact have a well-defined thermodynamic temperature: it is simply the temperature reached by a non-degenerate gas (a 'thermometer') that is in contact with it. In the case just described, the thermometer will

indicate a thermodynamic temperature much lower than the high electron momenta would otherwise imply.

Eq. 12.10 applies to a fully degenerate gas. Returning to our stars moving away from the main sequence, if the helium core is only partially degenerate, then some temperature dependence of the pressure remains. The isothermal core of a $1 M_\odot$ star between the points labelled 3 and 4 in Figure 12.1 is partially degenerate; consequently the core mass can reach $\sim 13\%$ of the entire stellar mass before it begins to collapse. Less massive stars exhibit even higher levels of degeneracy on the main sequence and may not exceed the Schönberg-Chandrasekhar limit at all before the next stage of nuclear burning.

We shall return to degenerate gas later in the course, when discussing white dwarfs.

12.7   Low Mass Stars and Brown Dwarfs

As we saw in Lecture 4 (Figure 4.10), the main sequence lifetimes of stars scale approximately as $t_{\mathrm{MS}} \propto M^{-2.5}$. Thus, while the Sun's main sequence lifetime is approximately 10^{10} yr, for stars with masses $M < 0.85\, M_\odot$ $t_{\mathrm{MS}} > t_H$, where $t_H = 13.7\,\mathrm{Gyr}$ is the current best estimate for the age of the Universe (see Figure 12.4). Therefore, stars with $M < 0.85\, M_\odot$ have not yet evolved off the main sequence, with the possible exception of stars in binary systems where the stellar evolution may have been affected by mass transfer from a companion.

We also saw in Lecture 8.3.1 that the size of the convective envelope increases with decreasing stellar mass below $1 M_\odot$. Stars with $M \lesssim 0.3 M_\odot$ are fully convective; they keep fusing and mixing until *all* H is converted to He after 10^{12} yr.

In stars with $M \lesssim 0.085\, M_\odot$, the core never reaches the temperature $T \simeq 4 \times 10^6\,\mathrm{K}$ required to ignite hydrogen burning via the p-p chain (Lecture 10.4). Therefore, this is considered to be the minimum mass of an object for it to be called a 'star'.[1]

---

[1] The minimum stellar mass is a function of composition. The limit $M_{\min} = 0.085\, M_\odot$ applies to gas of solar composition; in a very metal-poor star with less opacity, and therefore more transparent to escaping

$M \simeq 0.085\,M_\odot$ is thus taken to be the limit between H-fusing stars and 'brown dwarfs', which should more appropriately be called 'infrared dwarfs' because their peak emission is at near-IR wavelengths. Indeed brown dwarfs have been discovered primarily by near-IR surveys such as 2MASS (2 Micron All-Sky Survey). While brown dwarfs are too cool to ignite H, they can generate energy via the nuclear reaction:

$$^1_1\text{H} + {}^2_1\text{H} \rightarrow {}^3_2\text{He} + \gamma$$
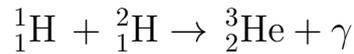
which is the second step in the p-p chain (Figure 7.5). ^2_1H is deuterium, the heavy isotope of hydrogen, created in Big Bang nucleosynthesis with a primordial abundance $(\text{D/H})_\text{p} = 2.5 \times 10^{-5}$ by number which has not changed much to the present day. Even such trace amounts of D are sufficient to slow the contraction and fading of a brown dwarf (see Figure 12.11).
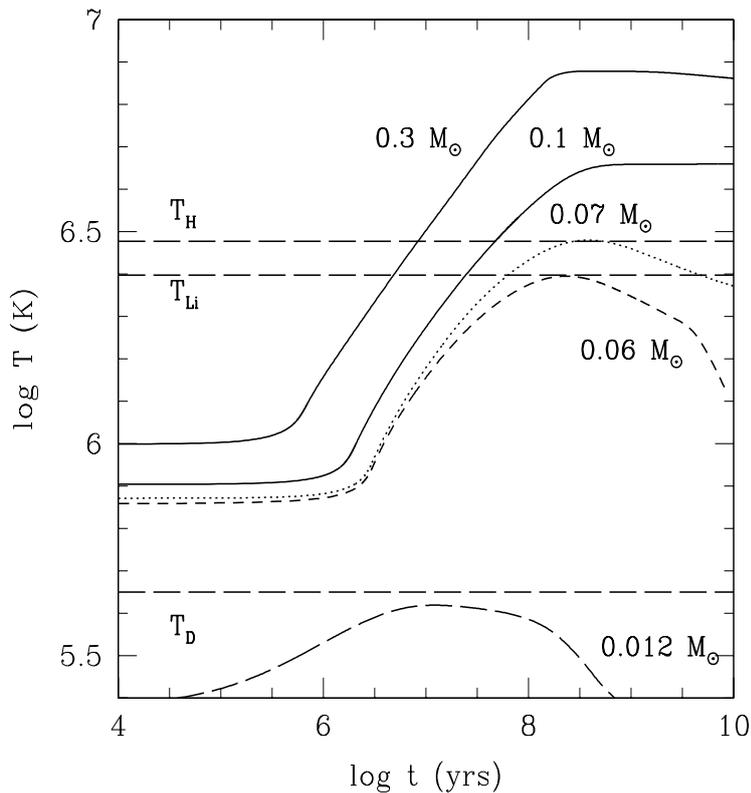
Figure 12.11: Core temperature as a function of age for different masses. T_H, T_Li, and T_D indicate the hydrogen, lithium, and deuterium burning temperatures respectively. (Figure reproduced from Chabrier & Baraffe 2000, ARA&A, 38, 337.)

The first brown dwarfs were discovered in 1995 (Rebolo et al. 1995, Nature, 377, 129; Oppenheimer et al. 1995, Science, 270, 1478). Coincidentally,

radiation, the limit increases to $M_\text{min} > 0.085\,M_\odot$.

16

this was also the year when the discovery of the first extrasolar planet was announced. Hundreds of brown dwarfs are now known. An 'acid' test of the sub-stellar nature of these objects is the presence of methane (CH_4) absorption features in their near-IR spectrum. H-burning stars are far too hot to form methane in their atmospheres, although this is a common molecule in the atmospheres of gas giant planets in our Solar System. Similar test based on the presence of Li and D have also been proposed, although a knowledge of the object age may be required (see Figure 12.11).

Three spectral types have been proposed to classify brown dwarfs: late M (later than M6.5), L and T. Gliese 229B, the prototype T dwarf and the first brown dwarf where methane absorption features were detected, has $T_{\mathrm{eff}} = 950\,K$. Ultracool brown dwarfs of spectral class Y have also been hypothesised. In August 2011 NASA announced the discovery of six Y dwarfs with temperatures as low as the human body, from data collected with the Wide-field Infrared Survey Explorer (WISE) satellite, now decommissioned (see http://www.nasa.gov/mission_pages/WISE/news/wise20110823.html).