AST320 – intro to astrophysics

Yanqin Wu, expanded from the original notes by Marten van Kerkwijk
This note will be continuously updated as the term proceeds.
Timestamp for the more recent update is noted below.

March 23, 2020

Contents

List of Figures

List of Tables

Table 0.1: Physical and Astronomical Constants in cgs units

Solar mass	$1\,M_\odot$	$= 1.989 \times 10^{33}$ g
Solar luminosity	$1\,L_\odot$	$= 3.826 \times 10^{33}$ erg s^{-1}
Solar radius	$1\,R_\odot$	$= 6.9599 \times 10^{10}$ cm
Solar effective temperature	$T_\odot$	$= 5770$ K
Earth mass	$1\,M_\oplus$	$= 5.974 \times 10^{27}$ g
Earth radius	$1\,R_\oplus$	$= 6.378 \times 10^{8}$ cm
year	$1\,\mathrm{yr}$	3.15576×10^{7} s
Parsec	$1\,pc$	$= 3.0857 \times 10^{18}$ cm
Astronomical Unit	$1\,AU$	$= 1.4960 \times 10^{13}$ cm
Atmospheric pressure on ground	1 bar	$= 10^{6}\,\mathrm{dyne/\,cm^2}$ (cgs) $= 10^{5}$ Pascale (SI)
Gravitational constant	G	$= 6.67259 \times 10^{-8}$ dyne cm^2 g^{-2}
Speed of light	c	$= 2.998 \times 10^{10}$ cm s^{-1}
Planck's constant	h	$= 6.626 \times 10^{-27}$ erg s
	$\hbar$	$= h/2\pi = 1.055 \times 10^{-27}$ erg s
Boltzmann's constant	k	$= 1.38 \times 10^{-16}$ erg K^{-1}
Stefan-Boltzmann constant	σ	$= 5.67 \times 10^{-5}$ erg cm^{-2} s^{-1} K^{-4}
Radiation constant	a	$= 4\sigma/c$
Proton mass	m_p	$= 1.6726231 \times 10^{-24}$ g
Neutron mass	m_n	$= 1.674929 \times 10^{-24}$ g
atmoic mass unit (amu)	u	$= 1.660540 \times 10^{-24}$ g
Electron mass	m_e	$= 9.1093897 \times 10^{-28}$ g
Hydrogen mass	m_H	$= 1.673534 \times 10^{-24}$ g
electric charge	e	$= 4.803204 \times 10^{-10}$ (cgs)
Electron volt	$1\,eV$	$= 1.6022 \times 10^{-12}$ erg
Bohr radius	a_0	$= \hbar^2/m_e e^2 = 5.292 \times 10^{-9}$ cm
Thompson cross-section	σ_T	$= \frac{8\pi}{3}\left(\frac{e^2}{m_e c^2}\right)^2 = 6.652 \times 10^{-25}$ cm^2

Useful references:

- Textbook: Carroll & Ostlie, *An introduction to Modern Astrophysics*, 2006,

- Kippenhahn & Weigert, *Stellar Structure and Evolution*, 1990, graduate-level, detailed and rigorous

- Prialnik, *An introduction to the theory of Stellar Structure and Evolution*, 2009, undergraduate level,

- Pols, on stellar structure and evolution, On-line lecture notes for a graduate level course, last update 2011, detailed and physical

- Cole Miller, University of Maryland, Undergrad cosmology course Note, easy-going to initiate you

- Komissarov Cosmology lecture note, if you feel CO doesn't quench your curiosity

1   Intro

Readings: review CO §*10.1, 2.4, 12.2 (assumed known* §*2.1-2.3, 12.1),* Back

– mass distribution

$$\frac{dM_r}{dr} = 4\pi r^2 \rho \,.$$ (1.1)

– gravitational acceleration for a spherical mass distribution, as if all mass inward of r located at origin,

$$g = -\frac{GM_r}{r^2} \,,$$ (1.2)

$g = \nabla U$ where U is the gravitational potential and depends on how mass is distributed. for constant density, $U = -\frac{3}{5}\frac{GM_r^2}{r}$.

– In spherical symmetry, equation of motion

$$\rho\frac{d^2r}{dt^2} = -\frac{GM_r\rho}{r^2} - \frac{dP}{dr}$$ (1.3)

– free-fall and dynamical timescale, set $dP/dr = 0$, and solve for above differential equation,

$$t_{\text{ff}} \approx \sqrt{\frac{1}{G\bar{\rho}}} \,,$$ (1.4)

–in hydrostatic equilibrium, take $d^2r/dt^2 = 0$,

$$\frac{dP}{dr} = -g\rho = -\frac{GM_r\rho}{r^2} \,.$$ (1.5)

– Virial theorem, an integral form for the force balance

$$KE = -U/2 \,,$$ (1.6)

specialize to rotational supported system, and pressure supported system; Derivation for gaseous spheres: multiply equation of hydrostatic equilibrium by r on both sides, integrate over sphere, and relate pressure to kinetic energy (easiest to verify for ideal gas).

– application in star formation: Jeans mass $KE + U/2 \leq 0$, so

$$M_{\text{J}} = \left(\frac{3}{4\pi}\right)^{1/2} \left(\frac{5k}{G\mu m_{\text{H}}}\right)^{3/2} \frac{T^{3/2}}{\rho^{1/2}} = 29\,M_\odot\,\mu^{-2}\left(\frac{T}{10\,\text{K}}\right)^{3/2}\left(\frac{n}{10^4\,\text{cm}^{-3}}\right)^{-1/2} \,,$$ (1.7)

$$R_{\text{J}} = \left(\frac{3}{4\pi}\frac{5k}{G\mu m_{\text{H}}}\right)^{1/2} \frac{T^{1/2}}{\rho^{1/2}} = 0.30\,\text{pc}\,\mu^{-1}\left(\frac{T}{10\,\text{K}}\right)^{1/2}\left(\frac{n}{10^4\,\text{cm}^{-3}}\right)^{-1/2} \,.$$ (1.8)

– Pulsation time scale

$$t_{\text{puls}} \simeq \frac{R}{c_{\text{s}}} \simeq \sqrt{\frac{1}{\gamma G\rho}} \qquad \left(\text{usually simply: } \sim\sqrt{\frac{1}{G\rho}}\right).$$ (1.9)

– another timescale of relevance when in hydrostatic equilibrium: the thermal timescale (or Kelvin-Helmholtz time)

$$t_{\text{KH}} = -\frac{KE + U}{L} = -\frac{U}{2L} \approx \frac{GM^2}{RL}$$ (1.10)

short of other types of energy supply (except for thermal and gravitational), as thermal support is lost in this timescale, system gradually contracts to a new equilibrium

Food for thought

Consider a proto-stellar cloud with a mass like that of the Sun, but starting with a larger radius and a temperature too cool for nuclear fusion.

- The cloud radiates its heat into space. In what way does the evolution depend on the rate of radiation loss?

- What happens to the total, potential, and kinetic energy of the star as it radiates into space?

- As a result, how do you expect the radius and internal temperature to change?

- When will the object stabilise?

- Eventually, what will happen to this cloud? Recall that any object with $T > 0\,\mathrm{K}$ will radiate energy and lose heat.

For next time

- Remind yourself about thermodynamics, in particular adiabatic processes (bottom of p. 317 to p. 321).

- Remind yourself about pressure integral (§10.2, in particular eq. 10.9), as well as mean molecular weight; and

- Remind yourself of relativistic energy and momentum (§4.4)

2 Equation of state

Readings: CO §4.4, p. 204-209, 10.2, 16.3, Back

Pressure (repulsion) can come from various interactions between particles

–Ideal gas: idealize approximation, air, stellar atmosphere, ISM

–non-ideal effects: appear at either high density or high temperature

- Electron degeneracy: white dwarfs, helium core of stars, low-mass stars, planets

- Neutron degeneracy: neutron stars,

- Coulumb pressure: giant planets (liquid, solid, human body, 'residual')

- Photon pressure: high-mass stars, accretion disk around blackholes

General expressions for number density, pressure, and energy

Given a momentum distribution $n(p)\mathrm{d}p$, then the particle density n, kinetic energy density U, and pressure P are given by

$$n = \int_0^\infty n(p)\mathrm{d}p, \tag{2.1}$$

$$U = \int_0^\infty n(p)\epsilon_p\,\mathrm{d}p, \tag{2.2}$$

$$P = \frac{1}{3}\int_0^\infty n(p)v_p p\,\mathrm{d}p. \tag{2.3}$$

For non-relativistic particles, $v_p = p/m$ and $\epsilon_p = p^2/2m$, while for (extremely) relativistic particles, $v_p \simeq c$, and $\epsilon_p = pc$. Hence,

$$P_{\mathrm{NR}} = \frac{1}{3}\int_0^\infty 2\epsilon_p n(p)\mathrm{d}p \qquad \Rightarrow \qquad P = \frac{2}{3}U, \tag{2.4}$$

$$P_{\mathrm{ER}} = \frac{1}{3}\int_0^\infty \epsilon_p n(p)\mathrm{d}p \qquad \Rightarrow \qquad P = \frac{1}{3}U. \tag{2.5}$$

General momentum distribution

$$n(p)\mathrm{d}p = n(\epsilon)\frac{g}{h^3}4\pi p^2\mathrm{d}p \qquad \text{(where g is the statistical weight).} \tag{2.6}$$

Here, $n(\epsilon)$ is the average number of particles per quantum state, and it depends on the nature of the particles:

$$n(\epsilon) = \begin{cases} \dfrac{1}{\mathrm{e}^{(\epsilon-\mu)/kT}+0} & \text{classical; Maxwell-Boltzmann statistics,} \\[2mm] \dfrac{1}{\mathrm{e}^{(\epsilon-\mu)/kT}+1} & \text{fermions; Fermi-Dirac statistics,} \\[2mm] \dfrac{1}{\mathrm{e}^{(\epsilon-\mu)/kT}-1} & \text{bosons; Bose-Einstein statistics.} \end{cases} \tag{2.7}$$

Here, μ is the chemical potential; one can view the latter as a normalisation term that ensures $\int_0^\infty n(p)\mathrm{d}p = n$.

Case I: Classical: Maxwellian

Using above expression to solve for μ, one recovers the Maxwellian momentum distribution:

$$n(p)\mathrm{d}p = n\frac{4\pi p^2\mathrm{d}p}{(2\pi mkT)^{3/2}}\mathrm{e}^{-p^2/2mkT}, \tag{2.8}$$
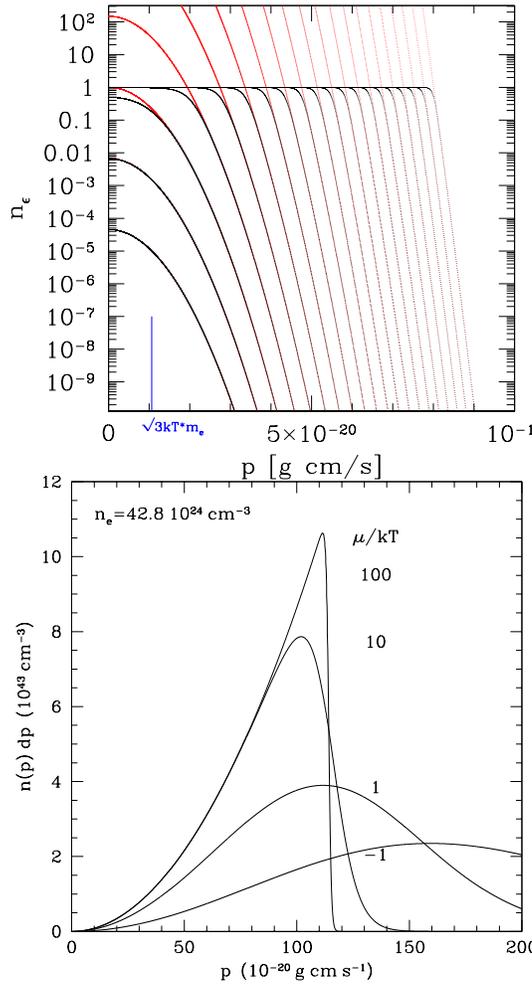
Figure 2.1: The occupation number of fermions per quantum state, as a function of momentum, for an electron gas with $T = 300\,\mathrm{K}$, in black lines. From bottom to top, the number density increases. The red lines are the corresponding value for ideal-gas particles. And the blue line marks out where the typical momentum at that temperature, for ideal-gas particles. Degeneracy is relevant when particles of these momentum are starting to fill up their respective momentum occupation. As density increases further, the Fermi sea is filled up gradually to higher and higher momentum. At sufficiently high density, the maximum momentum is only a function of density, not temperature. These curves approach complete degeneracy (with temperature playing a less and less important role).

Figure 2.2: The distribution of the number of particles $n(p)$ as a function of momentum p for a number of values of μ/kT.

This yields the usual 'ideal gas law', $P = nkT$, or $P = \rho kT/\mu m_H$. For the translational movement of each particle, the average energy is $mv^2/2 = 3/2\,kT$.

Case II: Fermions: application to electrons

In dense material ($n(\epsilon) \ll 1$ fails), each particle can no longer be considered independent. Fermions satisfy Pauli's exclusion principle,

$$n(p)\mathrm{d}p = \frac{g}{h^3}\frac{4\pi p^2 \mathrm{d}p}{\mathrm{e}^{(\epsilon-\mu)/kT}+1}. \tag{2.9}$$

– Complete Degeneracy: at the limit $kT \ll \mu$, cold and very dense,

$$n(\epsilon) = \left\{ \begin{array}{ll} 1 & \text{for} \quad \epsilon < \epsilon_F \\ 0 & \text{for} \quad \epsilon > \epsilon_F \end{array} \right\} \quad \Leftrightarrow \quad n(p) = \left\{ \begin{array}{ll} \frac{g}{h^3}4\pi p^2 \mathrm{d}p & \text{for} \quad p < p_F \\ 0 & \text{for} \quad p > p_F \end{array} \right. . \tag{2.10}$$

where ϵ_F is Fermi energy, and p_F the Fermi momentum that rises with particle density (and not related to temperature)

$$p_F = h\left(\frac{3n}{4\pi g}\right)^{1/3}. \tag{2.11}$$

Higher density material has higher Fermi momentum, and can become very relativistic. Two different equation of states.

- NRCD: non-relativistic complete degeneracy, one has $\epsilon_p = p^2/2m$, $P = \frac{2}{3}U$. Hence,

$$P = \frac{2}{3} \int_0^{p_F} n(p)\epsilon_p \, dp = \frac{1}{20}\left(\frac{3}{\pi}\right)^{2/3}\frac{h^2}{m}n^{5/3}. \tag{2.12}$$

$$\text{For electrons: } P_e = K_1(\rho/\mu_e m_H)^{5/3} \quad \text{with} \quad K_1/m_H^{5/3} = 9.91 \times 10^{12} \text{ (cgs)}. \tag{2.13}$$

where μ_e is the mean molecular weight per free electron, with $n = \rho/\mu_e m_H$

- ERCD: extremely relativistic complete degeneracy, one has $\epsilon_p = pc$, and $P = \frac{1}{3}U$ (Eq. 2.5). Hence,

$$P = \frac{1}{3}\int_0^{p_F} n(p)\epsilon_p \, dp = \frac{1}{8}\left(\frac{3}{\pi}\right)^{1/3}hc\,n^{4/3}. \tag{2.14}$$

$$\text{For electrons: } P_e = K_2(\rho/\mu_e m_H)^{4/3} \quad \text{with} \quad K_2/m_H^{4/3} = 1.231 \times 10^{15} \text{ (cgs)}. \tag{2.15}$$

Case III: Bosons: application to photons

For photons, the normalisation is not by total number of particles, but by energy; one finds $\mu = 0$. The statistical weight is $g = 2$ (two senses of polarisation). With $\epsilon = h\nu$ and $p = h\nu/c$, one finds the following distribution, also called the 'blackbody spectrum',

$$n(\nu)d\nu = n(p)dp = \frac{4\pi\nu^2 d\nu}{c^3}\frac{2}{e^{h\nu/kT}-1}; \tag{2.16}$$

$$U(\nu)d\nu = n(\nu)h\nu d\nu = \frac{8\pi h\nu^3}{c^3}\frac{d\nu}{e^{h\nu/kT}-1}. \tag{2.17}$$

– radiation pressure and radiation energy density

$$P_{\rm rad} = \frac{1}{3}\int \frac{h\nu}{c}n(\nu)d\nu = \frac{1}{3}\int u_\nu d\nu = \frac{1}{3}aT^4\,, \tag{2.18}$$

while energy density $U_{\rm rad} = 3P_{\rm rad} = aT^4$ (eq. 2.5).

Food for thought

- Why do electrons become degenerate at high density? what does temperature have to do with it?

- What determines the boundary between ideal gas and degeneracy, between NRCD and ERCD, ideal gas and radiation domination, in Fig. 2.3? Can you write down one expression for each?

- What non-ideal effects, other than degeneracy, can be important for stars and planets?

- For each equation of state, can you derive the mass-radius scaling?

- Is degeneracy important for daily materials?
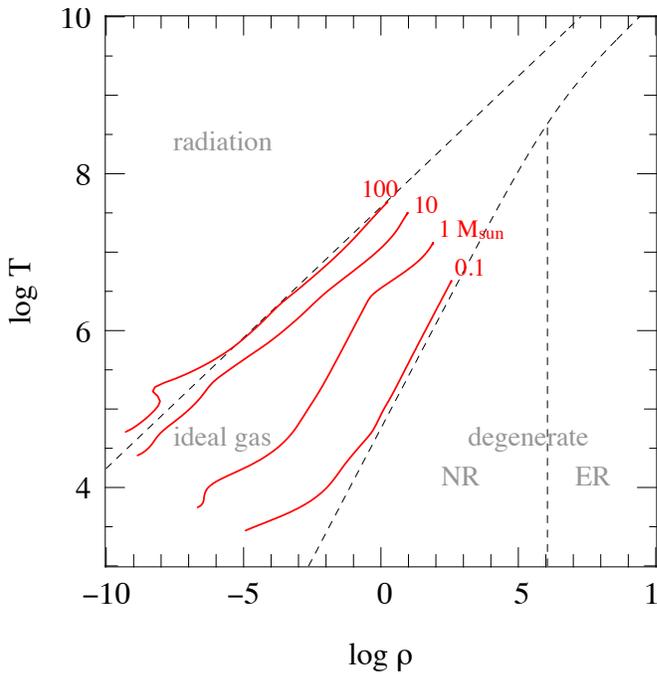
Figure 2.3: The T, ρ diagram for $X = 0.7$ and $Z = 0.02$, with the areas indicated where matter behaves as an ideal gas ($P \propto nT$), non-relativistic degenerate gas ($P \propto n_e^{5/3}$), relativistic degenerate gas ($P \propto n_e^{4/3}$), or radiation-dominated gas. Pressure is the third dimension that is not plotted here but you should try to visualize how it looks like. Note that these are not "sharp" boundaries. Also, at high enough T relativistic effects will become significant at all densities, not just for degenerate matter. Red lines plot the interior structure models for zero-age main-sequence stars of $0.1...100 M_\odot$. The $1 M_\odot$ model is well within the ideal-gas region of the equation of state. In the $0.1 M_\odot$ star electron degeneracy pressure is important, except in the outer layers. In stars more massive than $10 M_\odot$, radiation pressure becomes important, and it dominates in the surface layers of the $100 M_\odot$ model. Taken from Pols lecture notes.
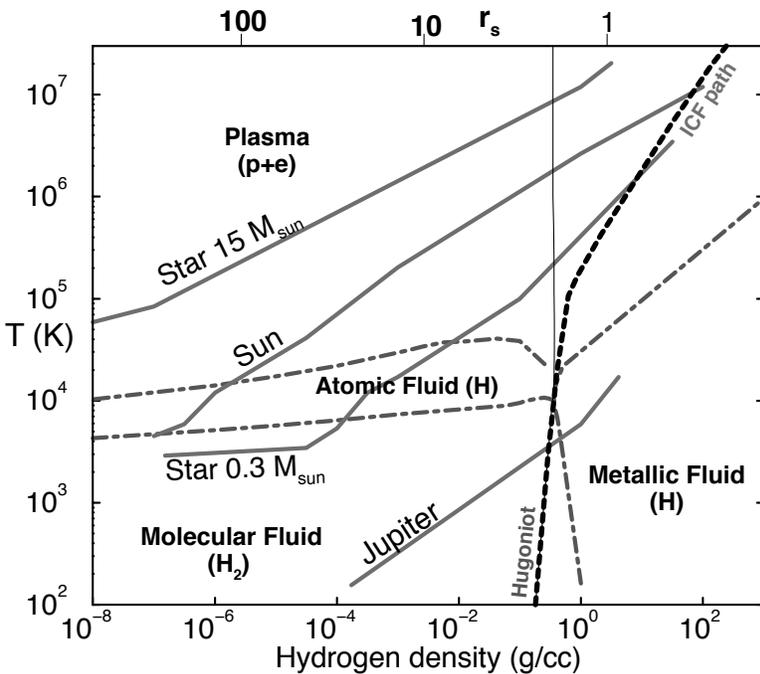
Figure 2.4: Density-temperature phase disagram of hydrogen, applicable both stellar and (giant) planetary interiors. The dash-dotted lines separate the molecular, atomic, metallic, and plasma regimes. The solid lines are isentropes for Jupiter and stars with $0.3, 1$, and $15 M_\odot$. Taken from Fortney et al (2010). For fun: our universe at redshift $z = 1100$ ("recombination") $- \rho \sim 10^{-20}\,\mathrm{g\,cm^{-3}}$ and a temperature of $T \sim 3000\,\mathrm{K}$; center of Jupiter $- \rho \sim 5\,\mathrm{g\,cm^{-3}}$ and $T \sim 10^4\,\mathrm{K}$; human body $- \rho \approx 1\,\mathrm{g\,cm^{-3}}$.
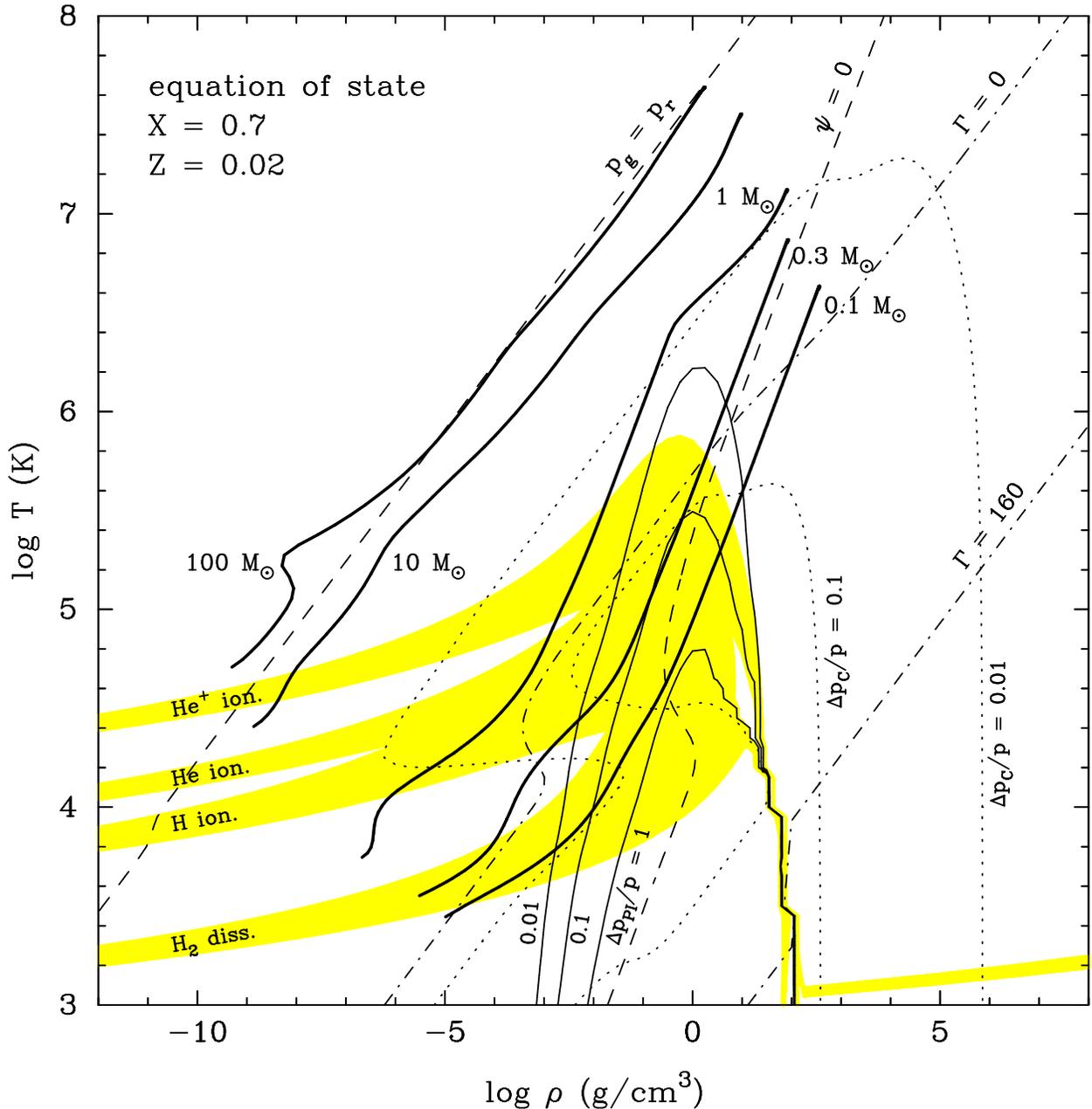
Figure 2.5: T, ρ diagram for $X = 0.7$ and $Z = 0.02$ from Pols et al. (1995). Dashed lines indicates where radiation pressure equals the gas pressure ($P_g = P_r$), and where electron degeneracy becomes important (to the right of the dashed line labelled $\psi = 0$, indicating where $\mu/kT = 1$). The shaded regions indicate transition regions where various ions become dominant (abundances range from 0.1 to 10). None of the other lines were discussed in the text. Dash-dotted lines indicate constant plasma-interaction parameter $\Gamma \equiv \frac{1}{kT} \frac{(Ze)^2}{d}$ (where Z is the ion charge and d the ion separation; the two lines correspond to $\Gamma = 1$ and 160; typically crystallization is considered to occur for $\Gamma \geq 175$); dotted lines constant contribution from Coulomb interactions to the ideal gas pressure (at high density, plasma screening is important, reducing the effect of Coulomb interaction); thin solid lines constant contribution from pressure ionisation (atoms separated by less than the Bohr radius). The thick solid lines indicate the run of density-temperature found in zero-age main sequence (ZAMS) stellar models for several masses.

3 Polytropes

Readings: CO p. 334-340, applications in §16.4, Back

Traditionally, polytropic models are introduced for simple modelling of stars. They still provide useful insights – examples to follow in this class. Let $P = K\rho^\gamma$, with γ being the adiabatic index, one can write down a set of scale-free equations for the star. Combine

$$\left.\begin{aligned} M_r &= -\frac{r^2}{\rho G}\frac{\mathrm{d}P}{\mathrm{d}r} \Rightarrow \frac{\mathrm{d}M_r}{\mathrm{d}r} = -\frac{1}{G}\frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{r^2}{\rho}\frac{\mathrm{d}P}{\mathrm{d}r}\right) \\ \frac{\mathrm{d}M_r}{\mathrm{d}r} &= 4\pi r^2 \rho \\ P &= K\rho^\gamma \end{aligned}\right\} \quad \Rightarrow \quad \frac{1}{\rho r^2}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2 \rho^{\gamma-2}\frac{d\rho}{\mathrm{d}r}\right) = -\frac{4\pi G}{K\gamma}. \tag{3.1}$$

Making the equation dimensionless, we derive the *Lane-Emden equation* of index n,

$$\left.\begin{aligned} \rho &= \rho_c \theta^n \quad \text{with} \quad n = \frac{1}{\gamma-1} \quad \left(\text{i.e.,}\gamma = 1 + \frac{1}{n}\right) \\ r &= \alpha\xi \quad \text{with} \quad \alpha = \left(\frac{n+1}{4\pi G}K\rho_c^{(1/n)-1}\right)^{1/2} \end{aligned}\right\} \quad \Rightarrow \quad \frac{1}{\xi^2}\frac{\mathrm{d}}{\mathrm{d}\xi}\left(\xi^2\frac{\mathrm{d}\theta}{\mathrm{d}\xi}\right) = -\theta^n. \tag{3.2}$$

The boundary conditions are $\theta_c = 1$ and $(\mathrm{d}\theta/\mathrm{d}\xi)_c = 0$.

Solutions of the Lane-Emden equations

In general, the Lane-Emden equation does not have an analytic solution, but needs to be solved numerically. The exceptions are $n = 0$, 1, and 5, for which,

$$\begin{aligned} n = 0 \quad (\gamma = \infty): \quad & \theta = 1 - \frac{\xi^2}{6} & \Rightarrow \quad & \rho = \rho_c, \\ n = 1 \quad (\gamma = 2): \quad & \theta = \frac{\sin\xi}{\xi} & \Rightarrow \quad & \rho = \rho_c\frac{\sin(r/\alpha)}{r/\alpha}, \\ n = 5 \quad (\gamma = 6/5): \quad & \theta = \left(1 + \frac{\xi^2}{3}\right)^{-1/2} & \Rightarrow \quad & \rho = \rho_c\left(1 + \frac{(r/\alpha)^2}{3}\right)^{-5/2}. \end{aligned} \tag{3.3}$$

– Physical Connections

The $n = 0(\gamma = \infty)$ case is 'incompressible', suitable for a star made up of solid. The $n = 1$ case is 'isothermal', or $P \propto \rho$. All polytropes with $n \geq 5$ ($\gamma \leq 6/5$) have radii that extend to infinity (and infinite mass for $n > 5$). Furthermore, polytropes with $n \geq 3$ ($\gamma \leq 4/3$) are dynamically unstable: an adiabatic compression of the entire star leads the central pressure to rise with ρ as $\rho^\gamma \propto R^{-3\gamma}$, but if $\gamma \leq 3/4$, this is less than needed for hydrostatic equilibrium which is $P_c \propto R^{-4}$.

Most realistic stellar models are between $n = 1.5$ and $n = 3$, with higher n (smaller γ) corresponding to more compressible gas and more concentrated centers. For reference, monotomic ideal gas $\gamma = 5/3$ (e.g., ions and atoms), diatomic gas (H_2, CO, ...e.g., Earth atmosphere) $\gamma = 7/5$.

Moreover, real stars have temperature gradients that are determined by heat transfer (next lecture), and their equation of state changes depending on the main constituent particles, hence one expects a spatially varying γ. But typically, high mass stars $n \approx 3$ ($\gamma \approx 4/3$, see, e.g, the Eddington model for radiation-dominated stars); low mass stars $n \approx 1.5$ ($\gamma \approx 5/3$); white dwarfs: $n = 1.5$ ($\gamma = 5/3$) for NRCD and $n = 3$ ($\gamma = 4/3$) for ERCD; Jupiter: $n \approx 2.5$ ($\gamma \approx 7/5$) in the atmosphere; Earth bulk: $n \approx 0$ ($\gamma \to \infty$).

In terms of stability, massive stars, and ERCD white dwarfs tend to be dynamically unstable; and stars with regions of partial ionization (Fig. 5.1) have locally low γ. Stellar explosion or pulsation ensue.

Table 3.1: Constants for the Lane-Emden functions

n	γ	ξ_1	$-\xi^2 \dfrac{\mathrm{d}\theta_n}{\mathrm{d}\xi}\Big|_{\xi_1}$	$\dfrac{\rho_c}{\overline{\rho}}$	$K\dfrac{R^{(n-3)/n}}{GM^{(n-1)/n}}$	$\dfrac{P_c}{GM^2/R^4}$
0.0	∞	2.4494	4.8988	1.0000	. . .	0.119366
0.5	3	3.7528	3.7871	1.8361	2.270	0.26227
1.0	2	3.14159	3.14159	3.28987	0.63662	0.392699
1.5	5/3	3.65375	2.71406	5.99071	0.42422	0.770140
2.0	3/2	4.35287	2.41105	11.40254	0.36475	1.63818
2.5	7/5	5.35528	2.18720	23.40646	0.35150	3.90906
3.0	4/3	6.89685	2.01824	54.1825	0.36394	11.05066
3.5	9/7	9.53581	1.89056	152.884	0.40104	40.9098
4.0	5/4	14.97155	1.79723	622.408	0.47720	247.558
4.5	11/9	31.83646	1.73780	6189.47	0.65798	4922.125
5.0	6/5	∞	1.73205	∞	∞	∞

Taken from Chandrasekar, 1967, Introduction to the study of stellar structure (Dover: New York), p. 96

Mass-Radius

Mass-radius of white dwarfs, neutron stars, main-sequence stars. Could be roughly obtained using Virial theorem, but now can be exact using polytropes.

– The stellar radius

Since one has $r = \alpha\xi$, the stellar radius is given by

$$R = \alpha\xi_1 = \left[\frac{(n+1)K}{4\pi G}\right]^{1/2} \rho_c^{(1-n)/2n}\xi_1, \tag{3.4}$$

where ξ_1 is the value of ξ for which $\theta(\xi)$ reaches its first zero. In Table 3.1, values of ξ_1 are listed for various n.

– the stellar mass

Integration of $\rho(r)$ gives the total mass of the star,

$$M = 4\pi\alpha^3\rho_c \int_0^{\xi_1} \xi^2\theta^n\mathrm{d}\xi = 4\pi\alpha^3\rho_c \int_0^{\xi_1} \mathrm{d}\left(-\xi^2\frac{d\theta}{d\xi}\right) = 4\pi\left[\frac{(n+1)K}{4\pi G}\right]^{3/2}\rho_c^{(3-n)/2n}\left(-\xi^2\frac{d\theta}{d\xi}\right)_{\xi_1} \tag{3.5}$$

where we used the Lane-Emden equation to substitute for θ^n. Values of $(-\xi^2\mathrm{d}\theta/\mathrm{d}\xi)_{\xi_1}$ are again listed in Table 3.1. By combining the relations for the radius and the mass, one also derives a relation between the radius, mass, and K, which, for given K, gives the mass-radius relation. The appropriate numbers are listed in the table.

– central density, pressure and potential energy

We can express the central density ρ_c in terms of the mean density $\overline{\rho} = M/\frac{4}{3}\pi R^3$ using the relations for the mass and radius. Solving K from the expressions for the mass and radius, one can also find the ratio of the central pressure to GM^2/R^4. Values of $\rho_c/\overline{\rho}$ and $P_c/(GM^2/R^4)$ are listed in Table 3.1.

Given the polytropic relation, one can also calculate the total potential energy. We just list the result here:

$$E_{\text{pot}} = -\frac{3}{5-n}\frac{GM^2}{R}. \tag{3.6}$$
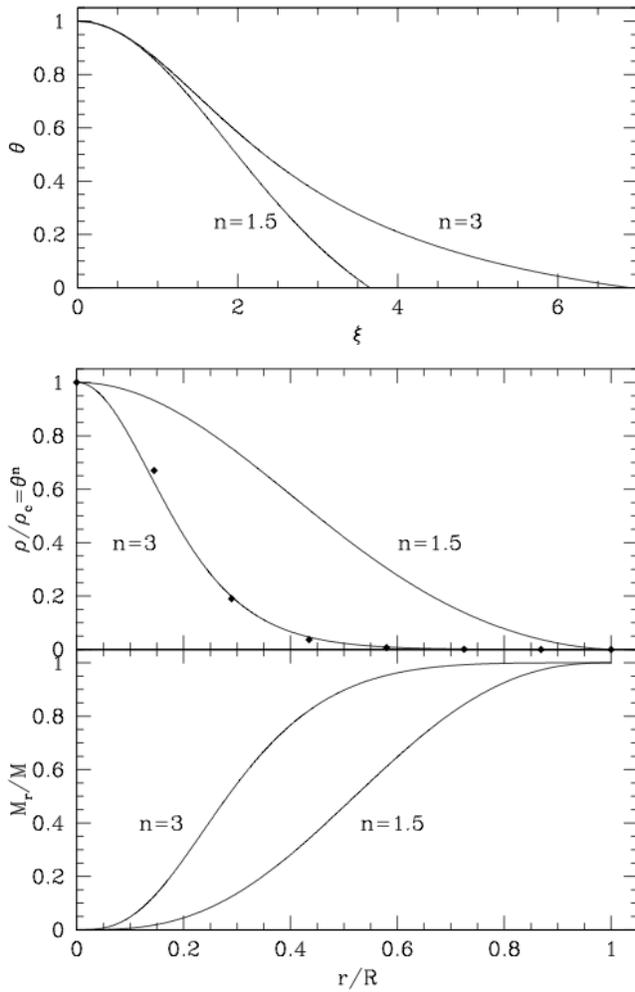
Figure 3.1: (top) Run of $\theta(\xi)$ as a function of ξ for $n = 1.5$ and $n = 3$ (i.e., $\gamma = 5/3$ and $\gamma = 4/3$). Note that $\xi \propto r$ and $\theta^n \propto \rho$. For non-degenerate stars, $T \propto \theta$. (middle) Corresponding run of $\rho(r)/\rho_C$ as a function of r/R. The black dots indicate the values appropriate for the Sun; (bottom) Run of M_r/M as a function of r/R. Note how much more centrally condensed the $n = 3$ polytrope is compared to the $n = 1.5$ one.

Food for thought

- Polytrope model only concerns $P = P(\rho)$. Where does temperature go? why can one approximate the stellar interior as a polytrope?

- At a given n and ρ_c, one can find an infinite series of models with different K. What do you think the K value corresponds to? answers differ depending on whether the gas is degenerate or ideal-gas.

- Derive a mass-radius relation for a given n polytrope, at a constant K value.

- In Eq. (3.5), there is something funny at $n = 3$. What does this mean physically?

- Why are more massive stars more centrally concentrated (higher n value) than lower mass stars?

- Dynamically stable polytropes require $\gamma \geq 4/3$. What happens to polytropes above this limit?

- In Fig. 2.5, what would be the best polytropic index that describes stars of different masses?

4 Radiation

Readings: CO §9.2,8.1, parts of §9.3, 10.4 Back

– review concept: mean-free-path, mean-free-time, random walk, cross-section, diffusion time, opacity, optical depth,

– heat can be transported outward by radiative diffusion (photons), convection (fluid advection) and conduction (electrons, phonons..., in degenerate cores).

4.1 Equation of Radiative Diffusion

– concept of Blackbody and black-body radiation (eq. 2.17)

– the concept of (mass) opacity and mean-free-path,

$$\kappa\rho = n\sigma \,. \tag{4.1}$$

– Equation of radiative diffusion. Deriving it, using the concept of mean-free-path, and photon diffusion

$$F_{\rm rad} = -\frac{1}{3}\frac{c}{\kappa\rho}\frac{{\rm d}U_{\rm rad}}{{\rm d}r} = -\frac{4ac}{3}\frac{T^3}{\kappa\rho}\frac{{\rm d}T}{{\rm d}r}. \tag{4.2}$$

where radiative luminosity $L_{\rm rad} = 4\pi r^2 F_{\rm rad}$

– Rosseland mean opacity: appropriately weighted by the black-body radiation spectrum, with excessive contribution from wavelengths of lower opacity,

$$\frac{1}{\bar{\kappa}} = \frac{1}{\kappa_{\rm R}} \equiv \frac{\pi}{acT^3}\int_0^\infty \frac{1}{\kappa_\nu}\frac{{\rm d}B_\nu}{{\rm d}T}\,{\rm d}\nu \,, \qquad \text{where} \quad B_\nu \equiv \frac{c}{4\pi}U(\nu) = \frac{2h\nu^3}{c^2}\frac{1}{e^{h\nu/kT}-1} \tag{4.3}$$

Since $\int ({\rm d}B_\nu/{\rm d}T){\rm d}\nu = acT^3/\pi$, the Rosseland mean is the harmonic mean of κ_ν weighted by ${\rm d}B_\nu/{\rm d}T$.

Excitation and Ionisation

Opacity depends on the available particle species that photons can interact with. These include, free electrons (electron scattering), ions (free-free), bound electrons (bound-free, in particular, H- opacity, also Rayleigh scattering), molecules (ro-vibrational bands) and dust grains. The following predicts the relative fraction of these particles.

– Review Bohr's model for hydrogen.

Fraction of occupation at different energy level depends on local density and temperature. In general, the different states of ions and atoms will be populated according to the *Boltzmann equation,*

$$\frac{N_b}{N_a} = \frac{g_b}{g_a}e^{-(\chi_b-\chi_a)/kT}. \tag{4.4}$$

Here, $g_{a,b}$ are the statistical weights (e.g., $g = 2n^2$ for level n in Hydrogen), and $\chi_{a,b}$ are the excitation potentials.

Comparing the ground state of one ionisation stage with the ground state of the next one, one has to take into account that the electron can have a range of kinetic energies and associated states. One finds

$$\frac{{\rm d}n_{i+1,0}(p)}{n_{i,0}} = \frac{g_{i+1,0}{\rm d}g_{\rm e}(p)}{g_{i,0}}e^{-(\chi_i+p_{\rm e}^2/2m_{\rm e})/kT} \tag{4.5}$$

where ${\rm d}n_{i+1,0}(p)$ is the number density of atoms in the ground state of ionisation stage $i+1$ with an electron with momentum p, $n_{i,0}$ the number density of atoms in the ground state of ionisation stage i, and $g_{\rm e}(p)$ the statistical weight of the electron at momentum p. The latter is given by

$${\rm d}g_{\rm e}(p) = \frac{2}{h^3}\frac{1}{n_{\rm e}}4\pi p^2\,{\rm d}p. \tag{4.6}$$

Integrating over all possible electron momenta and summing over all possible excitation states n (using the "partition function" $\mathcal{Z} = \sum_n g_n \exp(-\chi_n/kT)$), one finds the *Saha equation*,

$$\frac{n_{i+1}}{n_i} n_{\mathrm{e}} = \frac{\mathcal{Z}_{i+1}}{\mathcal{Z}_i} 2 \frac{(2\pi m_{\mathrm{e}} kT)^{3/2}}{h^3} \mathrm{e}^{-\chi_i/kT}. \tag{4.7}$$

4.2 Opacity Sources

In general, the opacity is a complicated function of density, temperature and abundances. Three main processes dominate the continuum opacity at temperatures typically encountered in stars. For cooler stars (cooler than $\sim 4000\,\mathrm{K}$), molecular opacities start becoming important; for even cooler objects ($T \leq 1500\,\mathrm{K}$), dust opacities dominate.

Electron scattering

Free electrons can scatter photons of all wavelengths equally efficiently (except for relativistic corrections),

$$\sigma_{\mathrm{T}} = \frac{8\pi}{3} \left(\frac{e^2}{m_{\mathrm{e}} c^2} \right)^2 = 6.65\ 10^{-25}\,\mathrm{cm}^2 \quad \Rightarrow \quad \kappa_{\mathrm{es}} = \frac{n_e \sigma_T}{\rho} = \sigma_{\mathrm{T}} \frac{1+X}{2m_{\mathrm{H}}} = 0.2(1+X)\,\mathrm{cm}^2\,\mathrm{g}^{-1}. \tag{4.8}$$

where σ_T is called the Thompson cross-section and the above assumed full ionization of hydrogen and helium plasma (with hydrogen mass fraction X and helium mass fraction Y and $X + Y = 1$). This opacity does not depend on density or temperature. Typically in the hot stellar interior, this provides a base opacity.

Free-free absorption (or bremsstrahlung)

Electrons experience acceleration in the electric field of an ion. This can lead to photon absorption or emission, the so-called 'braking radiation' (bremsstrahlung). It is a 2-body process. The free-free cross section for a certain ion i depends on electron density and is,

$$\sigma_{\nu,i}^{\mathrm{ff}} = \left(\frac{2m_{\mathrm{e}}}{\pi kT} \right)^{1/2} n_{\mathrm{e}} \frac{4\pi}{3\sqrt{3}} \frac{Z_i^2 e^6}{hcm_{\mathrm{e}}^2 \nu^3} g_\nu^{\mathrm{ff}}. \tag{4.9}$$

For a general mixture of ions, one has to add over all constituents and their corresponding Z_i^2:

$$\overline{n_{\mathrm{ion}} Z^2} = \sum \frac{\rho X_i}{m_{\mathrm{H}} A_i} Z_i^2 = \frac{\rho}{m_{\mathrm{H}}} \left[X + Y + \sum_{i \geq 3} \frac{X_i}{A_i} Z_i^2 \right], \tag{4.10}$$

where hydrogen and helium are assumed to be completely ionised.

In the integration over frequency required to calculate the Rosseland mean, one finds that the dependence on ν leads to the introduction of a T^{-3} term. The result is the so-called *Kramers free-free opacity*,

$$\kappa_{\mathrm{ff}} = 3.8\ 10^{22}\,\mathrm{cm}^2\,\mathrm{g}^{-1}\ \rho T^{-7/2} g_{\mathrm{ff}} (1 + X)\,(X + Y + B), \tag{4.11}$$

where B is the sum in eq. (4.10) and the Gaunt factor g_{ff} is a suitably averaged value of g_ν^{ff}.

Bound-free absorption

The semi-classical Kramers cross section for an ion with charge Z_i with an electron in state n is given by

$$\sigma_{\nu,i,n}^{\mathrm{bf}} = \frac{64\pi^4}{3\sqrt{3}} \frac{m_{\mathrm{e}} e^{10}}{ch^6} \frac{Z_i^4}{n^5 \nu^3} g_{\nu,i,n}^{\mathrm{bf}} = 2.82 \times 10^{25}\,\mathrm{cm}^2\ \frac{Z_i^4}{n^5 \nu^3} g_{\nu,i,n}^{\mathrm{bf}}. \tag{4.12}$$

Most of the ions will be in an ionisation state $i + 1$ which cannot be ionised by a typical photon with $h\nu \simeq kT \ll \chi_{i+1}$; the relevant ions for the opacity are the somewhat rarer ions in ionisation state i. Combining the Boltzmann and Saha equations, and writing $n_{i,n}$ explicitly in terms of $n_{i+1,1}$,

$$n_{i,n} = n_{i+1,1} n_{\rm e} \frac{n^2}{2} \left(\frac{h^2}{2\pi m_{\rm e} kT} \right)^{3/2} {\rm e}^{\chi_{i,n}/kT}, \tag{4.13}$$

where the hydrogenic approximation $(g_n = 2n^2)$ was made.

For the Rosseland mean, one needs to add all states of all ions. For stellar interiors, hydrogen and helium will be completely ionised, so the mean opacity will be proportional to the metallicity Z. One finds the *Kramers bound-free opacity*,

$$\kappa_{\rm bf} = 4.3 \; 10^{24} \, {\rm cm}^2 \, {\rm g}^{-1} \, \rho T^{-7/2} \frac{g_{\rm bf}}{t} Z(1 + X) \,, \tag{4.14}$$

where g is a mean Gaunt factor and t the "guillotine" factor that accounts for the number of different ions being available. This is a 1-body process, however, the opacity still depends on density. Think why.

Negative hydrogen ion, H^-

Hydrogen atom has a bound state for a second electron in the field of the proton, though it has a very low ionisation potential (binding energy), $\chi_{\rm H^-} = 0.75\,{\rm eV}$. The number density of negative hydrogen ions will be proportional to the electron density, which, in all but the most metal-poor stars, will be set by ionisation of the metals (which have much lower ionisation potentials that hydrogen and helium). Thus, the H$^-$ opacity will scale as $\kappa_{\rm H^-} \propto \rho XZ$ at low temperatures; H$^-$ is of course easily ionized at higher temperatures, and at *very* low temperatures even metals will not be ionized, so there will be no electrons to form H$^-$ by combining with H.

H^- opacity is important in the solar surface, and in ISM. A very approximate formula in the range $T \sim (36) \times 10^3\,{\rm K}, \rho \sim (10^{10} 105)\,{\rm g/\,cm}^3$ and $0.001 < Z < 0.02$ is

$$\kappa_H \approx 2.5 \times 10^{31} \left(\frac{Z}{0.02} \right) \rho^{1/2} T^9 \, {\rm cm}^2/\,{\rm g} \,. \tag{4.15}$$

Food for thought

- What determines the luminosity of a star? is it the mechanism of energy generation (e.g., nuclear fusion)? Invert Eq. (4.2) to obtain dT/dr, apply it to the whole star, and use virial theorem to estimate T; how does L depend on M and R? Why is it independent of the source of energy?

- For more massive stars, radiation pressure becomes important. How does this affect the above scaling?

- Fig. 2.4 (and 2.5) shows the separating line between ionized and neutral (atomic) hydrogen. It hugs closely to $T = 10^4\,{\rm K}$ and only rises slightly with density. The ionization temperature of Hydrogen at ground state is 13.6 eV, which corresponds to a kinetic temperature of $\sim 10^5\,{\rm K}$. But why is hydrogen already half-ionized at $10^4\,{\rm K}$? Also, physically, what causes the ionization temperture to rise with density?

- Understand for the three following contexts, what is the dominant opacity for photons: 1. $T \gg 10^4\,{\rm K}$ plasma; 2. $T \sim 10^4\,{\rm K}$; 3. $10^3\,{\rm K} \le T \le 10^4\,{\rm K}$; 4. $T \le 10^3\,{\rm K}$.

- Explain the transitions in absorption features in Fig. 4.5, as stellar effective temperature changes.

- Why is electron scattering important in hot stellar interior? what happens to proton scattering? physically, why are electrons so much more important than protons?

- According to Fig. 4.1, which part inside the Sun is the most difficult for photons to pass through (per gram of material-wise)?
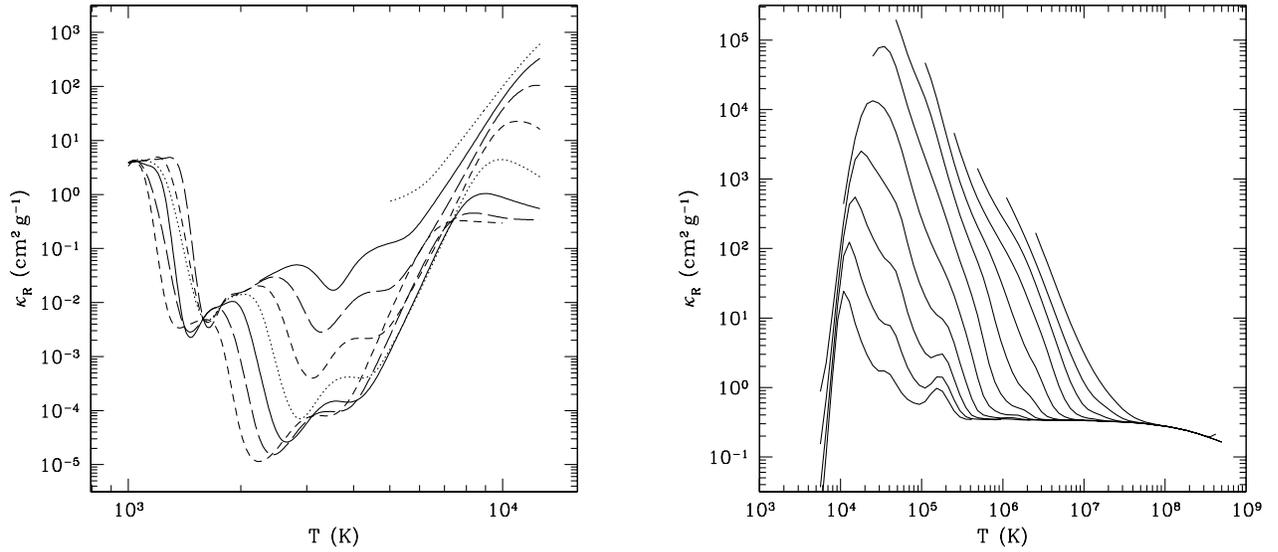
Figure 4.1: Opacities as a function of temperature. *(left)* Low-temperature regime, from Alexander & Ferguson (1994, ApJ 437, 879). Opacities are shown for densities from 10^{-13} to $10^{-6}\,\mathrm{g\,cm^{-3}}$ in factors of ten, with lower densities corresponding to lower opacities. The sequence in line types is short-dashed, long-dashed, solid, dotted. The bump on the left is due to dust, that in the middle mostly to water, and that on the right to H^-. *(right)* High-temperature regime, for densities from 10^{-9} to $10^2\,\mathrm{g\,cm^{-3}}$, from the OPAL group (Iglesias & Rogers, 1996, ApJ 464, 943). The bump at the left is due to bound-free and free-free absorption, and the lower level at the right to electron scattering. Note the difference in scale between the two panels.
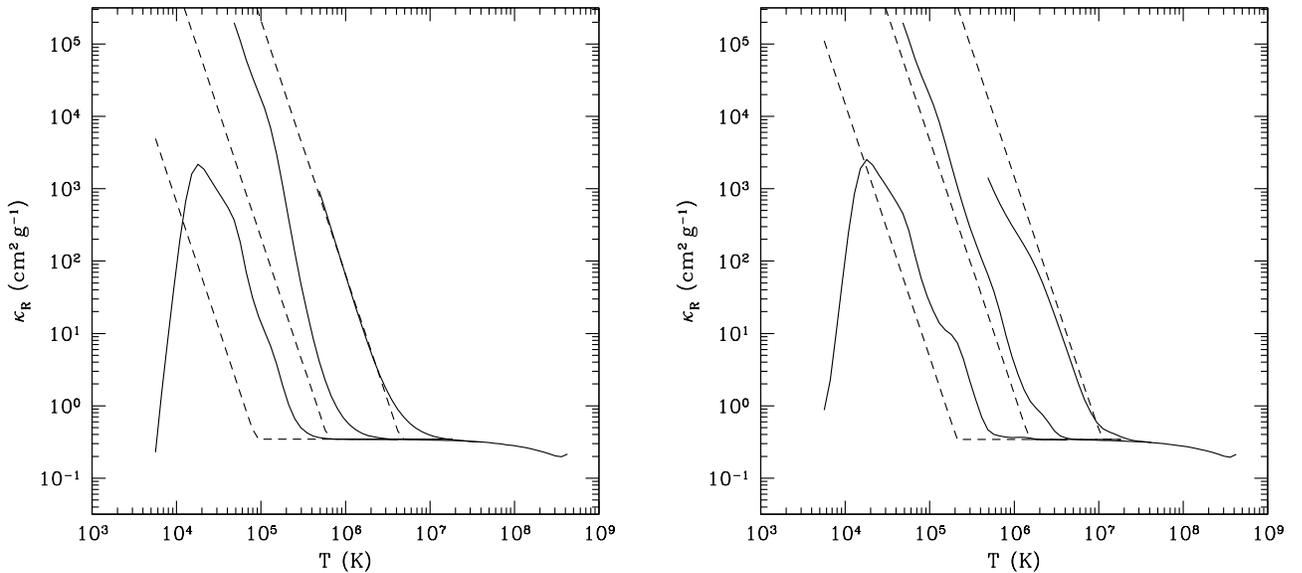
Figure 4.2: Opacities as a function of temperature as estimated with the Kramers formulae (short-dashed lines) compared to those calculated by the OPAL group, for densities 10^{-6}, 10^{-3}, and $1\,\mathrm{g\,cm^{-3}}$. *(left)* $Z = 0$: OPAL vs. the Kramers free-free opacity; *(right)* $Z = 0.02$: OPAL vs. the Kramers bound-free opacity.
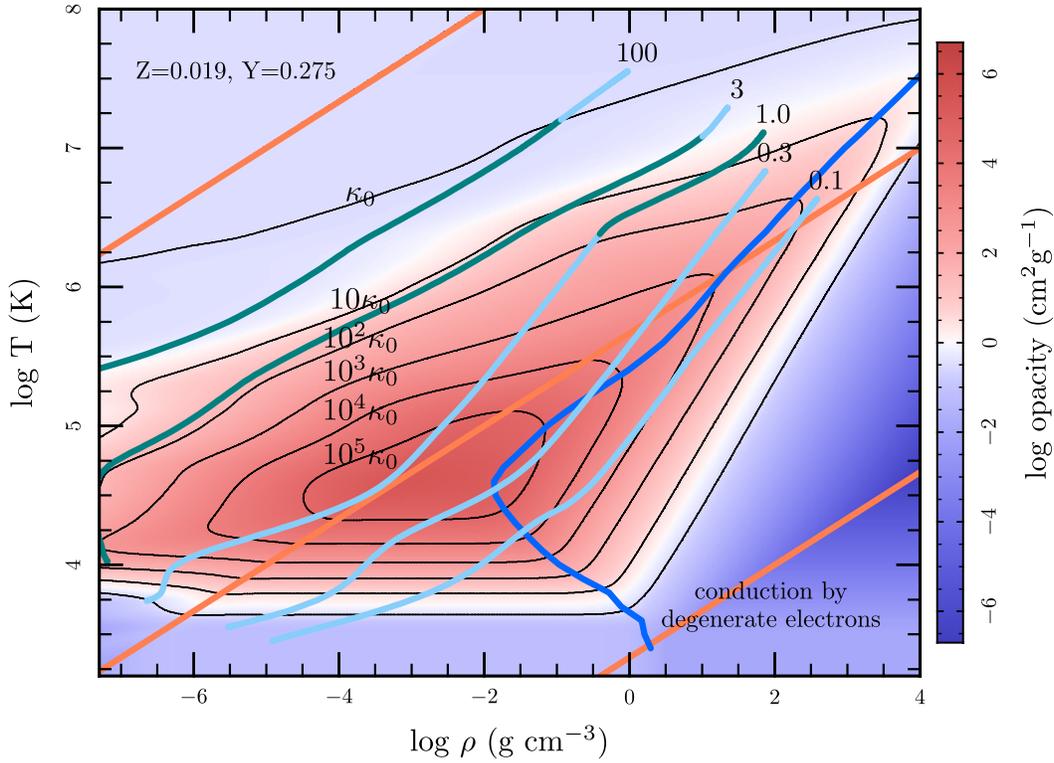
Figure 4.3: Opacities for Z = 0.019, Y = 0.275, taken from the MESA compilation by Paxton et al. (2011). The underlying shades show the value of κ, whereas the contours are in units of the electron scattering opacity, $\kappa_0 = 0.2(1 + X)\text{cm}^2\text{g}^1$. The orange lines show (top to bottom) where $\log R = -8, \log R = 1$ and $\log R = 8$, with a conventional definition of $R \equiv \rho/(T/10^6\text{K})^3$. Stellar interior profiles for main sequence stars of mass $M = 0.1, 0.3, 1.0, 3.0, \& 100 M_\oplus$ are shown by the green (radiative regions) and light blue (convective regions) lines. Electron conduction dominates the opacity to the right of the dark blue line (which is where the radiative opacity equals the conductive opacity).
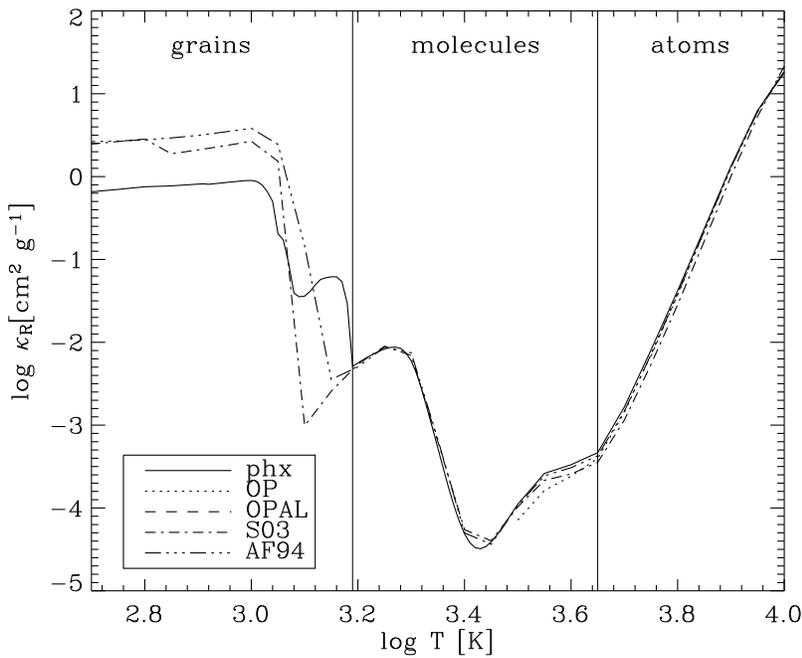
Figure 4.4: Similar to left panel of Fig. 4.1, showing logarithm of the Rosseland mean opacity as a function of temperature at solar abundance and $\log R = -3$. Different curves indicate different computations. Regions where certain species (grains, molecules, and atoms) dominate the opacity are indicated. The opacity has a minimum at $T = 2500\text{K}$ (due to the disappearance of H^- opacity), but as molecules, and then grains form at subsequently cooler temperatures, opacity rises again. Maximum values of opacity are comparable to that from electron scattering ($\kappa \sim$ order unity). Taken from Ferguson et al (2005).
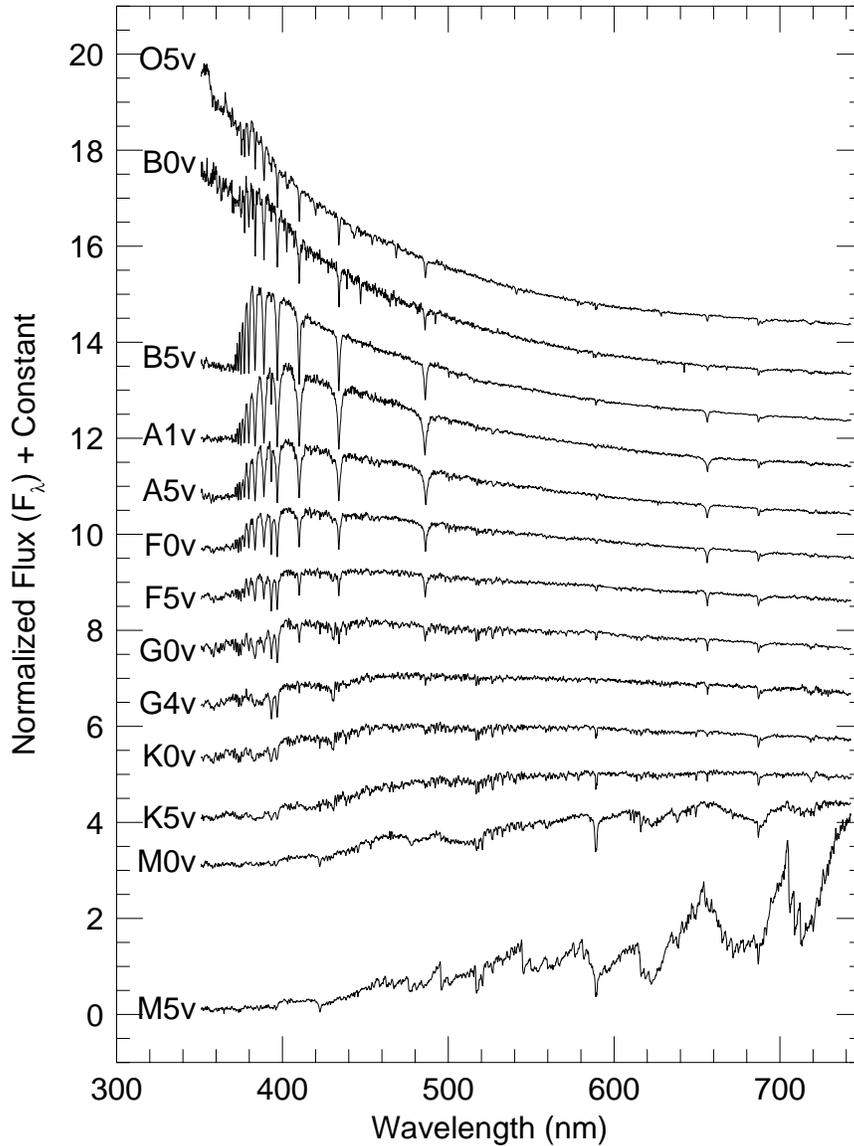
Figure 4.5: Various stellar spectra with different effective temperatures. From bottom to top, the temperature increases and the the blackbody peaks shifts to the left. The dominant absorption features changes from molecular bands (M-type), to metal lines (K-type, G-type, F-type), to Balmer lines (A-type), to HeI lines (B-type), to HeII lines (O-type). The decrement shortward of 365nm (in stars cooler than B5) is the Balmer break. Hydrogen ionization occurs around A0 ($T_{\text{eff}} \approx 10,000\,\text{K}$), also the place where Balmer lines are most prominent.
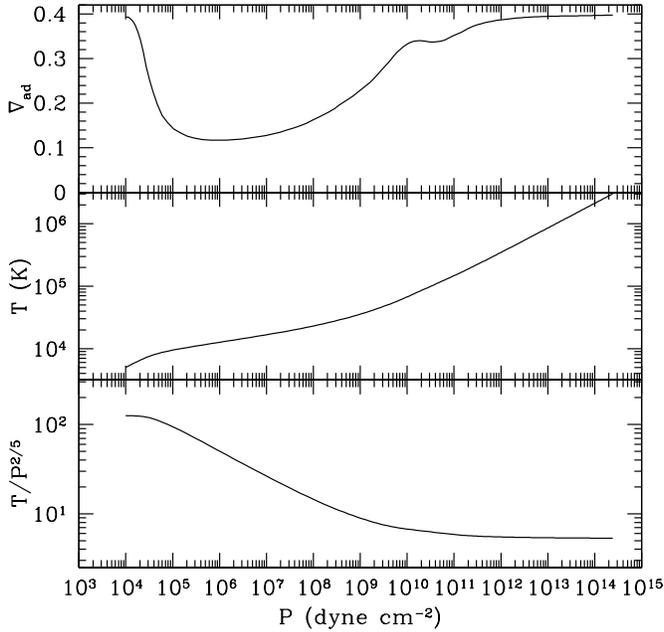
Figure 5.1: Adiabatic gradient *(top)*, temperature *(middle)* and $T/P^{2/5}$ *(bottom)* as a function of pressure, calculated using the OPAL equation of state for a solar mixture. The effect of the hydrogen and helium ionisation zones is clearly seen in the depressions in ∇_{ad} and the changes in slope in the other panels. The reduction in ∇_{ad}, coupled with the higher opacity (which increases ∇_{rad}), always tend to make partial ionization region convective. Moreover, the star is an imperfect polytrope as ∇_{ad} drops below 0.4, the value expected for mono-atomic gas.

5 Convection

*Readings: CO §10.4,*Back

Convection (fluid turbulence) can be very efficient at advecting energy flux, when-ever permitted. Conduction, in contrast, can be ignored except in degenerate stars.

5.1 Convection, Mixing Length Theory

For adiabatically (no heat-exchange) rising and falling gas parcels ($P \propto \rho^{\gamma}$), stability requires a general criterion:

$$-\frac{1}{\gamma}\frac{1}{P}\frac{\mathrm{d}P}{\mathrm{d}r} > -\frac{1}{\rho}\frac{\mathrm{d}\rho}{\mathrm{d}r}\,. \tag{5.1}$$

This specializes into the following two named criteria under special circumstances.

– Schwarzschild instability criterion: assume ideal gas law and no chemical gradient,

$$\left.\frac{\mathrm{d}\ln T}{\mathrm{d}\ln P}\right|_{\text{ad}} < \frac{\mathrm{d}\ln T}{\mathrm{d}\ln P} \quad \Leftrightarrow \quad \nabla_{\text{ad}} < \nabla. \tag{5.2}$$

where $\nabla_{\text{ad}} = (\gamma - 1)/\gamma$. The actual temperature gradient is compared with that required to carry the radiative flux (eq. 4.2) to determine the dominant mode of energy transport,

$$\nabla_{\text{rad}} = \frac{F_{\text{rad}}3\kappa\rho}{4acT^3}\frac{p}{g\rho T}\,. \tag{5.3}$$

– Ledoux instability criterion: ideal gas but with chemical gradient (e.g., Helium in the bottom),

$$\frac{\gamma-1}{\gamma} < \frac{\mathrm{d}\ln T}{\mathrm{d}\ln P} - \frac{(\partial\ln\rho/\partial\ln\mu)}{(-\partial\ln\rho/\partial\ln T)}\frac{\mathrm{d}\ln\mu}{\mathrm{d}\ln P}, \quad \Leftrightarrow \quad \nabla_{\text{ad}} < \nabla - \frac{(\partial\ln\rho/\partial\ln\mu)}{(-\partial\ln\rho/\partial\ln T)}\nabla_{\mu}, \tag{5.4}$$

where we have defined $\nabla_{\mu} = \mathrm{d}\ln\mu/\mathrm{d}\ln P$ to be the changes in μ due to changes in composition X_i only, and where for a fully-ionised ideal gas, the term with the partial derivatives equals unity.

Efficiency of convection

The above discussed fluid local instability creates flow and eddies, which transport heat efficiently by advection, as opposed to slow photon diffusion. The motion is too complicated to study in detail (see the youtube movie on 2-D convection), so we use an approach called the mean-field theory approach, which leads to the following simplified theory, the mixing length theory.

A general expression for the convective flux is

$$F_{\mathrm{conv}} = \rho \bar{v}_{\mathrm{conv}} \Delta q = \rho \bar{v}_{\mathrm{conv}} c_P \Delta T = \rho \bar{v}_{\mathrm{conv}} c_P T \frac{\ell_{\mathrm{mix}}}{2H_P} \left(\nabla - \nabla_{\mathrm{ad}} \right), \tag{5.5}$$

where ℓ_{mix} is the *mixing length*, usually parametrized as a fraction of the scale height, i.e., $\ell_{\mathrm{mix}} \equiv \alpha_{\mathrm{mix}} H_P$, with α_{mix} the *mixing length parameter*.

What is the appropriate v_{conv}? This is likely less than the local speed of sound (think why). To estimate v_{conv}, we use a method different from that used in the textbook: balance buoyancy (force per unit volumn, $g\Delta\rho$) against ram pressure (force per unit area, ρv^2); evaluate velocity at $l_{\mathrm{mix}}/2$

$$v^2_{\mathrm{conv}} = \frac{g}{H_P} \frac{\ell^2_{\mathrm{mix}}}{2} \left(\nabla - \nabla_{\mathrm{ad}} \right) . \tag{5.6}$$

compare this to sound speed and show that this is much smaller, unless $\nabla - \nabla_{\mathrm{ad}} \sim$ unity. The convective flux is given by

$$F_{\mathrm{conv}} = \rho c_P T \alpha^2_{\mathrm{mix}} \sqrt{\frac{gH_P}{8}} \left(\nabla - \nabla_{\mathrm{ad}} \right)^{3/2} . \tag{5.7}$$

Order-of-magnitudely, a very useful relation is that $F_{\mathrm{conv}} \approx \rho v^3_{\mathrm{conv}}$, where we have taken $c_p T \sim g H_p$.

What is the value of α_{mix}? this is harder to answer without accurate simulations. Traditionally this has been inferred from stellar parameters: for the Sun, $\alpha \sim 1.75$; for other stars, see, e.g., this paper.

– nearly adiabatic temperature gradient

For completely convective stars, the temperature gradient needs to be only very slightly superadiabatic for substantial luminosities to be transported. So we can simply write

$$\nabla \sim \nabla_{\mathrm{ad}} \tag{5.8}$$

inside most convection zones. The convective region is largely adiabatic with a very sub-sonic convective movement. This temperature gradient carries a non-zero radiative flux, but the rest of the energy flux is carried by convection, trivially.

The implication is that whatever luminosity the star manages to radiate away, will be brought to the surface without any problem by a corresponding energy flux in the convective regions. Thus, the actual luminosity of the star is determined only in the radiative region in the star, most notably, the photosphere.

5.2 Hayashi Track

Before contraction onto the main sequence, the proto-star is highly luminous and therefore fully convective. More massive stars have shorter pre-MS phase. Convection gradually lifts from the centre of these stars, excepts for stars less massive than $\sim 0.3 M_\odot$. They remain fully convective throughout their main-sequence lifetime.

A completely convective star

To find a solution for the whole star, we need to match a photosphere to the interior solution, where the latter is given by a polytrope $P = K\rho^{5/3}$. Matching the two solutions will set K, and for fixed K one knows how the radius depends on mass. For the run of pressure in the atmosphere, we have

$$\frac{\mathrm{d}P}{\mathrm{d}r} = -\frac{GM}{R^2}\rho \quad \text{or} \quad \frac{\mathrm{d}P}{\mathrm{d}h} = -g\rho,$$

where h is the height above some reference level. For the photosphere, $\tau = \kappa \rho h = \frac{2}{3}$, or

$$h = \frac{2}{3\kappa\rho} \quad \Rightarrow \quad P_{\text{phot}} = \frac{2g}{3\kappa}. \tag{5.9}$$

Now, assume that the opacity is given by a law of the form (this scales with pressure, rather than density, as is usually done, but you can do it differently)

$$\kappa = \kappa_0 P^a T^b, \tag{5.10}$$

where in general a will be a positive number of order unity, while for cool temperatures b will be a relatively large positive number. Given this general opacity law, one has

$$P_{\text{phot}} = \left(\frac{2GM}{3\kappa_0 R^2 T_{\text{eff}}^b}\right)^{1/(1+a)}. \tag{5.11}$$

For the interior, we write the polytropic relation in terms of pressure and temperature, and combine it with the mass-radius relation for polytropes of $n = 1.5$ (Table 3.1),

$$\left.\begin{array}{l} P = K\rho^{5/3} \\ P = \frac{\rho}{\mu m_{\text{H}}}kT \\ K = C_{1.5}GM^{1/3}R \quad \text{with} \quad C_{1.5} = 0.42422 \end{array}\right\} \Rightarrow \left.\begin{array}{l} P = K\left(\frac{P\mu m_{\text{H}}}{kT}\right)^{5/3} \end{array}\right\} \Rightarrow P_{\text{int}} = \frac{M^{-1/2}}{(RC_{1.5}G)^{3/2}}\left(\frac{kT}{\mu m_{\text{H}}}\right)^{5/2}. \tag{5.12}$$

Note that in this equation, if you substitute in $T = T_{\text{eff}}$, you will obtain surface pressure; while if you substitute in $T = T_c$, you obtain central pressure. So let's set $T = T_{\text{eff}}$ and equate P_{int} with P_{phot}, raising to the $2(1+a)$ power. Sorting, one finds

$$\left(\frac{2}{3\kappa_0}\right)^2 G^{1+3a}M^{3+a}R^{-1+3a} = C_{1.5}^{-3-3a}\left(\frac{k}{\mu m_{\text{H}}}\right)^{5+5a}T^{5+5a+2b}. \tag{5.13}$$

Solving for T_{eff},

$$T_{\text{eff}} = \mathcal{C}_R M^{\frac{3+a}{5+5a+2b}}R^{\frac{-1+3a}{5+5a+2b}} \quad \text{with} \quad \mathcal{C}_R = \left[\left(\frac{2}{3\kappa_0}\right)^2 G^{1+3a}C_{1.5}^{3+3a}\left(\frac{k}{\mu m_{\text{H}}}\right)^{-5-5a}\right]^{\frac{1}{5+5a+2b}}. \tag{5.14}$$

For a of order unity and large positive b one thus sees that T_{eff} depends only very weakly on the mass and radius. With $L = 4\pi R^2 \sigma T_{\text{eff}}^4$, we can determine the dependencies on M and L, and thus where the star would be in the HRD. One finds

$$\left(\frac{2}{3\kappa_0}\right)^2 G^{1+3a}M^{3+a}\left(\frac{L}{4\pi\sigma}\right)^{(3a-1)/2} = C_{1.5}^{-3-3a}\left(\frac{k}{\mu m_{\text{H}}}\right)^{5+5a}T^{3+11a+2b}, \tag{5.15}$$

$$T_{\text{eff}} = \mathcal{C}_L M^{\frac{6+2a}{6+22a+4b}}L^{\frac{3a-1}{6+22a+4b}} \quad \text{with} \quad \mathcal{C}_L = \left[\left(\frac{2}{3\kappa_0}\right)^2 G^{1+3a}C_{1.5}^{3+3a}\left(\frac{k}{\mu m_{\text{H}}}\right)^{-5-5a}\right]^{\frac{2}{6+22a+4b}}. \tag{5.16}$$

Again, for a of order unity and large b, T_{eff} depends extremely weakly on the luminosity, and thus one expects nearly vertical lines in the HRD. Given the slight positive dependence on M, one expects the lines to move slightly towards higher temperatures for larger masses.

Most stars on the Hayashi tracks have cool photospheres. The opacity is mainly contributed by H^- absorption which have $a \approx 0.5$ and $b \approx 9$. So $T_{\text{eff}} \propto L^{0.01}M^{0.14}$. Intuitively, a slightly too hot star will have a much lower photospheric opacity, which, given that the photosphere is the only bottleneck for heat loss for a fully convective star, will lead to a much larger luminosity and a faster cooling.
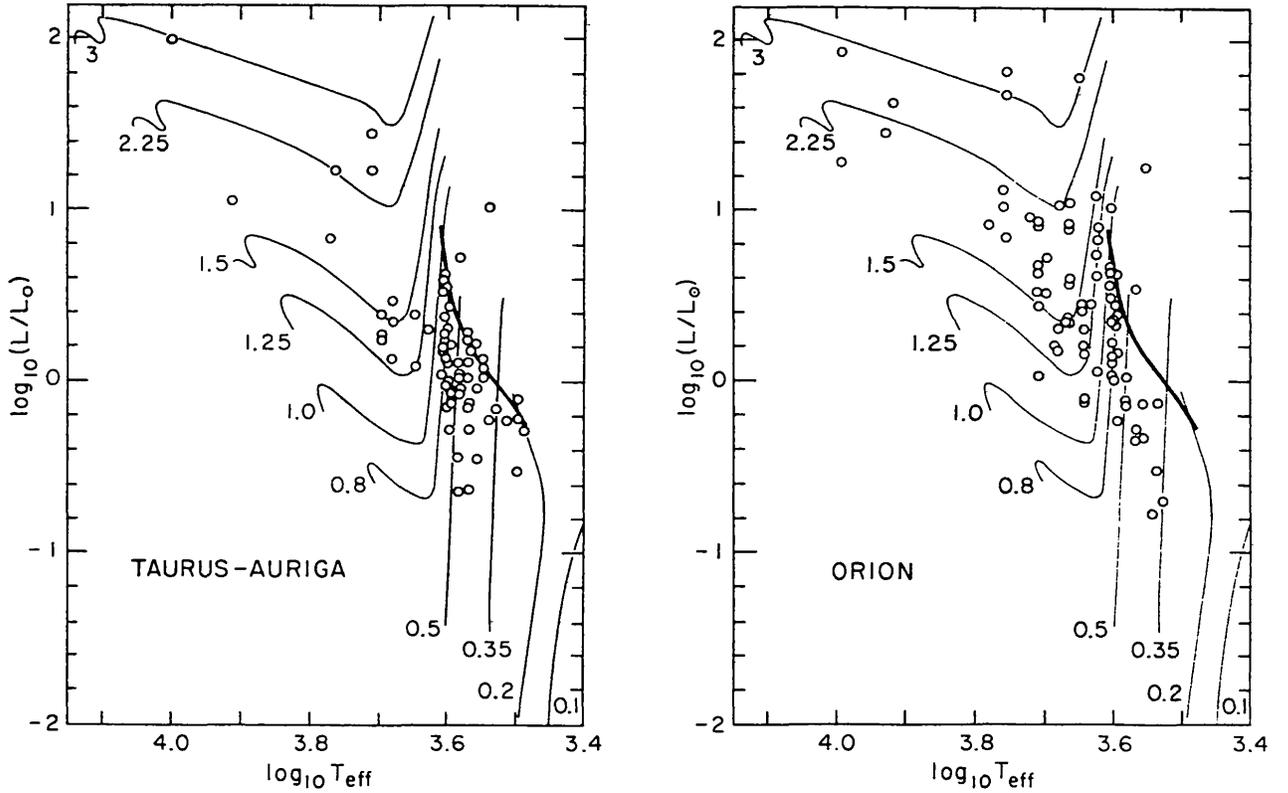
Figure 5.2: Theoretical tracks for the pre-main sequence contraction phase for several different masses (as indicated). Overdrawn are observed temperatures and luminosities for pre-main sequence stars in two star-forming regions with rather different properties. In both, stars first appear along a very similar "birth line" (indicated with the thick line).

Complications

The scaling that one finds from the above relations is reasonable. If one were to calculate numerical values, however, the answers would be very puzzling. The reason is that the assumption of a polytrope breaks down near the surface. Going towards the surface, it first fails in the ionisation zone, where recombination is an additional source of heat. Due to the recombination, the temperature of an adiabatically expanding blob does not decrease as it would otherwise, and therefore, above the ionisation zone the temperatures will be higher than would be the case if recombination were ignored. The effect can be seen Fig. 5.1.

Just below the photosphere, the convective energy transport becomes much less efficient, i.e., the superadiabatic gradient becomes substantial, while in the assumption of a $n = 1.5$ polytrope it is assumed to be negligible. With less efficient energy transport, the temperature will decrease more rapidly than adiabatic. Thus, the substantially superadiabatic region near the photosphere counteracts the effects of the ionisation zone. Net, the ionisation zone is more important.

Contraction along the Hayashi track – turning radiative

The star needs to contract in order to provide the energy it radiates away. Since it is completely convective, the entropy remains constant through the star, but decreases (increasing the entropy of the universe in order not to violate the second law). Since $dq = T ds$, the energy generated per gram is proportional to the local temperature. Therefore, the increase in luminosity in a shell dM_r is $dL_r \propto T dM_r$. With this, and with $P \propto T^{5/2}$, we can estimate whether the radiative gradient decreases towards the surface or towards the centre of the star. We assume again an opacity law of the form $\kappa \propto P^a T^b$,
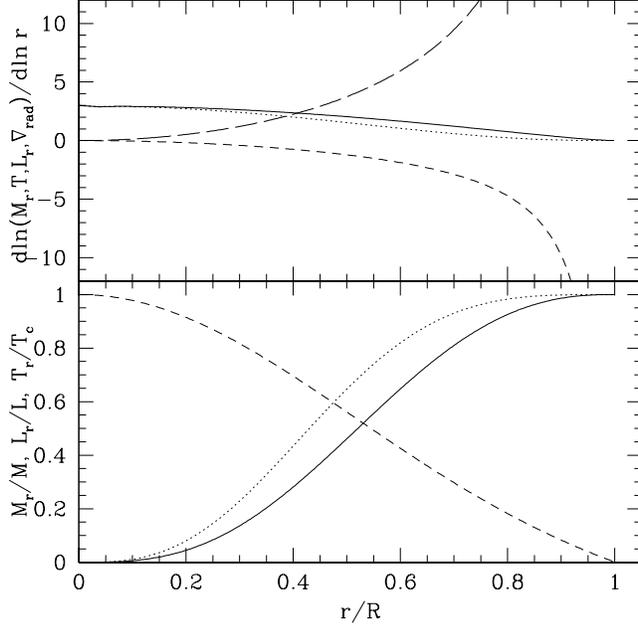
Figure 5.3: *(Bottom)* Run of mass (solid line), luminosity (dotted line), and temperature (dashed line) as a function of radius for a contracting polytrope with $n = 1.5$ (i.e., the local energy generation per unit mass is proportional to temperature). *(Top)* Logarithmic derivatives of mass (solid line), luminosity (dotted line), temperature (short-dashed line), and radiative gradient (long-dashed line) as a function of radius. A Kramers-type opacity law was assumed.

with $a = 1$, $b = -4.5$ for a Kramers-type law. We find

$$
\begin{aligned}
\frac{d \ln \nabla_{\rm rad}}{d \ln r} &= \frac{d \ln L_r}{d \ln r} - \frac{d \ln M_r}{d \ln r} + \frac{d \ln \kappa}{d \ln r} + \frac{d \ln P}{d \ln r} - 4 \frac{d \ln T}{d \ln r} \\
&= \frac{d \ln L_r}{d \ln r} - \frac{d \ln M_r}{d \ln r} + [b - 4 + 2.5(a + 1)] \frac{d \ln T}{d \ln r}.
\end{aligned}
\tag{5.17}
$$

Generally, one has $M_r = \int \rho r^2 dr$ and $L_r \propto \int T \rho r^2 dr$. With the polytropic relations, therefore, $M_r \propto \int \theta^n \xi^2 d\xi$ and $L_r \propto \int \theta^{n+1} \xi^2 d\xi$. Thus, one can use the solution $\theta(\xi)$ for a polytropic star to calculate $d \ln (M_r, L_r, T)/d \ln r$. The result for $n = 1.5$ is shown in Fig. 5.3. Also drawn is $d \ln \nabla_{\rm rad}/d \ln r$, assuming $a = 1$ and $b = -4.5$. One sees that it is always larger than zero, i.e., the radiative gradient decreases inwards. This is true for any reasonable opacity law. In consequence, the interior is always the first part of the star to become radiative.

We can also estimate how the radiative gradient scales with the stellar parameters in the core. There, the temperature hardly varies, and one has $L_r \propto (L/M)T_c M_r$. Furthermore, for any two stars with the same structure, $T_c \propto M/R$ and $P_c \propto M^2/R^4$, with the same constants of proportionality. Taking again $\kappa \propto P^a T^b$, one finds for the radiative gradient in the core,

$$
\nabla_{\rm rad,c} \propto \frac{L_r \kappa P}{M_r T^4} \propto \frac{L}{M} T^{1+b-4} P^{a+1} \propto L M^{-2+b+2a} R^{-1-b-4a} \propto L M^{-4.5} R^{-0.5}
\tag{5.18}
$$

where in the last proportionality we used $a = 1$, $b = -4.5$ (Kramers). From Eqs. 5.14, 5.16, one sees that for given mass, $L \propto R^\alpha$, with $\alpha = (6 + 22a + 4b)/(5 + 5a + 2b)$, where a and b are now the coefficients in the atmospheric opacity law. Generally, $a \simeq 1$ and b large, hence, $\alpha \simeq 2$. Thus, the radiative gradient decreases as one descends the Hayashi track. At constant luminosity, one has $R \propto M^\beta$, with $\beta = (6 + 2a)/(7 - a + 2b) \leq 1$. Hence, the radiative gradient is smaller for larger masses, and more massive stars will become radiative in their core sooner.

5.3 Cooling Giant Planets

Jupiter-like giant planets cannot ignite nuclear burning, but cool and contract over time. They dim with time. Numerical models (Fig. 5.4) show that $L(t) \propto t^{-4/3}$. We devise a simple analytical model to understand this result. We ignore physical constants in the following as we are aiming for a final power-law solution.

- the interior of these bodies remain convective (Fig. 5.4). Study eq. (5.3) to understand why. So planet cools along a sequence of adiabatic models defined by a single parameter s, the entropy.
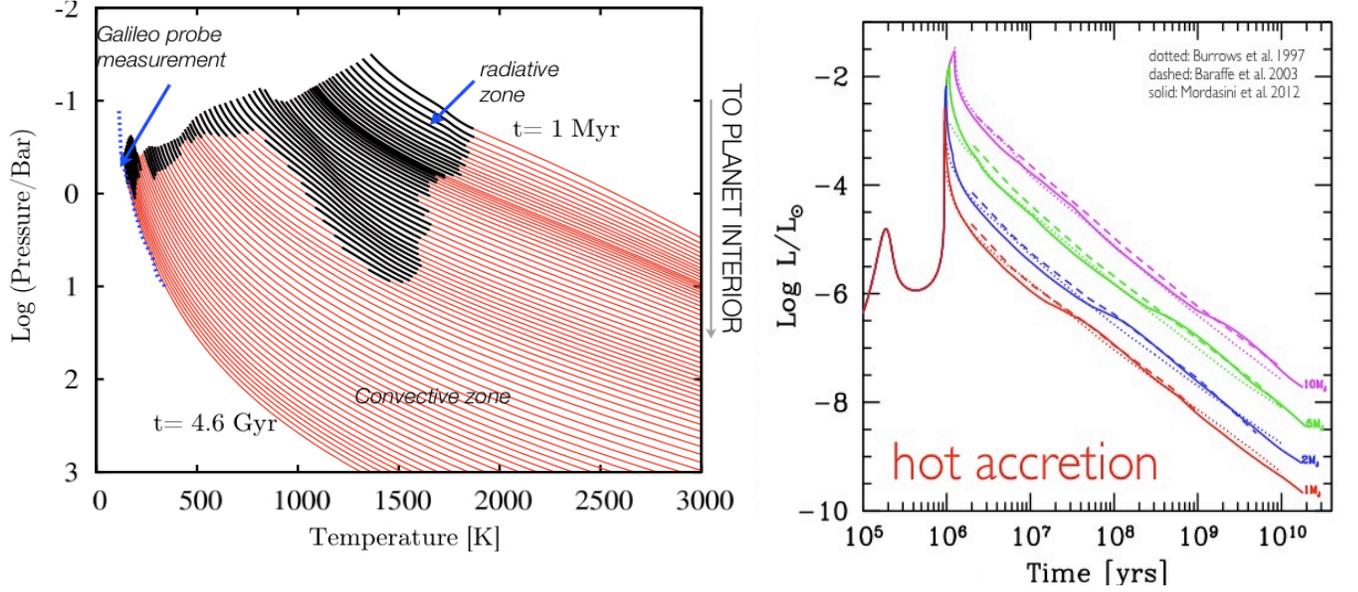
Figure 5.4: *(Left)* Pressure-temperature profile inside a cooling gas giant planet (Jupiter). The model's upper reach is the location of the photosphere. Most of the interior remains convective and the cooling is limited by the radiative diffusion near the surface. *(Right)* The cooling luminosity of such a model as a function of time, for Jupiter and more massive giant planets. After formation, the luminosity fades with time roughly as $t^{-4/3}$, and the Kelvin-Helmholtz time is always of order the age. The dimming poses challenges for direct imaging. From Mordasini et al (2012).

- Assume the models all have the same polytropic index γ: $P \propto K\rho^{\gamma}$ where $K = K(s)$ and s is the overall entropy. The planet radius is linked to entropy (or rather $K(s)$) in a polytrope model by eq. (3.4);

$$R \propto K^{1/2}\rho_c^{\frac{1-n}{2n}} \quad \Rightarrow \quad R \propto K^{\frac{1}{3\gamma-4}} . \tag{5.19}$$

  where we have used the fact that $\rho_c \propto \bar{\rho} \propto M/R^3$ and the polytropic index $n = 1/(\gamma - 1)$.

- Assume the interior is made up of ideal gas, so we can equivalently write

$$P = K_2 T^{\frac{\gamma}{\gamma-1}} \tag{5.20}$$

  where $K_2 = K_2(s) \propto K^{-\frac{1}{\gamma-1}}$. The adiabatic temperature gradient $\nabla_{\text{ad}} \equiv d\ln T/d\ln P|_{\text{adia}} = (\gamma - 1)/\gamma$,

- At the surface of such a model, $p_{\text{phot}} \sim g/\kappa \propto M/R^2/\kappa$, $T = T_{\text{eff}}$. Using eq. (5.20), and taking the surface opacity to be constant – a reasonable assumption at the low temperatures we are interested in (Fig. 4.4) – we obtain,

$$T_{\text{eff}} = \left(\frac{P_{\text{pho}}}{K_2}\right)^{\frac{\gamma-1}{\gamma}} \propto K^{\frac{2-\gamma}{(3\gamma-4)\gamma}} \tag{5.21}$$

- Lastly, all models are linked together as

$$L = \int T\frac{ds}{dt}dM = \frac{ds}{dt}\int TdM \propto \frac{ds}{dt}\frac{1}{R} , \tag{5.22}$$

  where we have used Virial theorem to relate the heat content of the planet ($\propto \int TdM$) to its gravitational energy ($\propto 1/R$). On the other hand, heat loss is limited at the surface by radiative diffusion. In the case where convection zone reaches the photosphere,

$$L = 4\pi R^2 \sigma T_{\text{eff}}^4 \tag{5.23}$$

- Now we just need a relation between K and s to solve for the time dependence. With $Tds = dq + Pdv = c_v dT + PdV$ (where c_v is the specific heat per unit mass at constant volumn and the volumn $V = 1/\rho$ for unit mass), we find,

$$ds = c_v d\ln\left(\frac{P}{\rho^\gamma}\right) = c_v d\ln K \quad \Rightarrow K = \exp\left(\frac{s}{c_v} + \text{constant}\right). \tag{5.24}$$

- Now collect all scalings and solve for the time dependence for K, we find

$$\frac{ds}{dt} = c_v d\ln K \propto R^3 T_{\text{eff}}^4 \quad \Rightarrow \quad K \propto t^{\frac{7\gamma-8}{(3\gamma-4)\gamma}} \tag{5.25}$$

or a time-dependence for the luminosity of

$$L \propto R^2 T_{\text{eff}}^4 \propto t^{\frac{-8-2\gamma}{7\gamma-8}}. \tag{5.26}$$

Substituting in $\gamma = 5/3$, we get $L \propto t^{-14/11} \sim t^{-1.273}$, which is close to the numerical result of $L \propto t^{-4/3}$.

Food for thought

- Think about why a star cannot sit to the right of the Hayashi limit.

- Why are fully convective stars well described by polytropes? What is the appropriate polytropic index for low mass, fully convective stars? What is the physical meaning of the constant K in $P = K\rho^\gamma$?

- In Fig. 2.5, the curves for low-mass stars have wiggles that are coincident with ionisation and dissociation zones. What is the physical reason for this? why are ionization regions typically convective?

- How super-adiabatic is the solar convection zone?

- Convection zone can carry a tremendous amount of flux outward. It is almost a free-pass. So what determines how much luminosity the Sun (which has a surface convection zone) should produce?

- How about a fully convective body (like a pre-main-sequence star or a giant planet)? what determines how much flux is leaking out per second?

6 Nuclear Fusion

*Readings: CO §10.3,*Back

Energy Balance

$$\frac{\mathrm{d}L_r}{\mathrm{d}r} = 4\pi r^2 \rho\, \epsilon \quad \Leftrightarrow \quad \frac{\mathrm{d}L_r}{\mathrm{d}M_r} = \epsilon, \tag{6.1}$$

where ϵ is the energy generated per unit mass. In general,

$$\epsilon = \epsilon_{\mathrm{grav}} + \epsilon_{\mathrm{nuc}} - \epsilon_{\nu}, \tag{6.2}$$

where ϵ_{grav} is the energy liberated or lost by contraction or expansion, ϵ_{nuc} is the energy produced (or lost) in nuclear processes, and ϵ_{ν} is that part of the latter that escapes the star immediately in the form of neutrinos.

– Contraction or expansion

The energy gained or lost in mass movements inside the star can be derived from the first law of thermodynamics, and written in various equivalent forms as

$$\epsilon_{\mathrm{grav}} = -\frac{\mathrm{d}Q}{\mathrm{d}t} = -T\frac{\mathrm{d}S}{\mathrm{d}t} = -\frac{\mathrm{d}u}{\mathrm{d}t} - P\frac{\mathrm{d}\mathcal{V}}{\mathrm{d}t}, \tag{6.3}$$

where $\mathcal{V} \equiv 1/\rho$ and u is the energy density per unit mass.

Nuclear processes

The main source of energy in stars is nuclear fusion, which we will now treat in more detail than in CO, §10.3 (Kippenhahn & Weigert, chapter 18, was used extensively below).

Basic considerations

The energy gained or lost in nuclear processes is related to the mass defect Δm:

$$E = \Delta m\, c^2 = \left(\sum_i m_{\mathrm{init},i} - \sum_j m_{\mathrm{final},j} \right) c^2. \tag{6.4}$$

The mass defect reflects the different binding energies per nucleon in different nuclei,

$$\frac{E_{\mathrm{bind}}}{A} = \frac{1}{A} \left(Z m_{\mathrm{p}} + (A-Z)m_{\mathrm{n}} - m_{\mathrm{nucleus}} \right) c^2 . \tag{6.5}$$

The binding energy per nucleon increases steeply from hydrogen, then flattens out and starts to decrease, having reached a maximum at ^{56}Fe; see Fig. 6.1. Defining hydrogen to have zero binding energy, helium has 7.07 MeV per nucleon, carbon 7.68 MeV, and iron 8.73 MeV.

For fusion, nuclei must be brought close enough together that the short-range strong nuclear force can dominate over the weaker, but long-range repulsive Coulomb force. The range of the strong nuclear force is set by the Compton wavelength of its carrier, the pi meson, $\hbar/m_\pi c = 1.41$ fm. The repulsive Coulomb potential at a distance of ~ 1 fm (10^{-13} cm) is $E_{\mathrm{Coul}} = Z_1 Z_2 e^2/r \simeq 1.44\,\mathrm{MeV} \left(\frac{1\,\mathrm{fm}}{r} \right) Z_1 Z_2$, where Z_1 and Z_2 are the atomic numbers of the colliding nuclei. This should be compared with typical kinetic energy of a particle, of order $kT = 0.86\, T_7$ keV, where T_7 is the temperature in units of 10^7 K. Thus, classically, in the centre of the Sun (where $T_7 \approx 1.5$), particles trying to interact should be turned around by the Coulomb force at $\sim 10^3$ fm; as a result, no reactions would be expected.

From quantum mechanics, however, a particle has a certain finite probability of "tunneling" through the Coulomb barrier (see CO, p. 147–148, which is perhaps more insightful than the motivation on p. 335-338). If a proton reaches within one
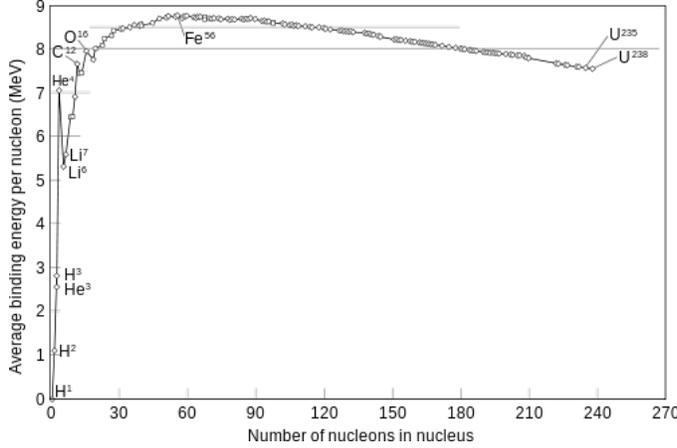
Figure 6.1: Binding energy per nucleon for the different nuclei. Taken from Wiki.

de Broglie wavelength of another proton, there is a certain probability that they are' closer than ~ 1 fm. Requiring that the kinetic energy of a hydrogen nuclei ($m_p v^2/2 = 3/2kT$) has to exceed the repulsive Coulomb potential at 1 de Broglie wavelength e^2/λ with $\lambda = h/p = h/m_p v$, we find

$$\lambda \leq \frac{h^2}{2e^2 m_p} \sim 100\,\text{fm}\,, \tag{6.6}$$

The corresponding temperature (ignition temperature) is then

$$T \geq \frac{m_p e^4}{k h^2} \sim 10^7\,\text{K}\,. \tag{6.7}$$

So hydrogen burning can start at $T_7 \sim 1$. One can repeat the above analysis to find the temperature dependence on the nuclei charge to show that higher temperatures are required to fuse heavier nuclei.

Armed with such an order-of-magnitude understanding, we discuss the nuclear fusion in more detail. Pay particular attention to the concept of "Gamow peak". The reaction cross section per nucleus is usually written as,

$$\sigma(E) = \frac{S(E)}{E} e^{-b/\sqrt{E}} \quad \text{with} \quad b = \frac{1}{h} 2\pi^2 \sqrt{2m'} Z_1 Z_2 e^2 \quad \text{and} \quad m' = \frac{m_1 m_2}{m_1 + m_2}. \tag{6.8}$$

Here, the term $1/E$ reflect the effective area for the interaction (for which one can take $\pi\lambda^2 \propto 1/p^2 \propto 1/E$), and the exponential term the penetration probability; effects from the nuclear force are absorbed into a function $S(E)$ which is, under most conditions, a relatively slowly varying function of the interaction energy E (but see "resonances" below).[1]

The fusion product is at first a compound nucleus in an excited state with positive total energy. Often, this compound nucleus will decay into the same particles that formed it – i.e., the incoming particle is just scattered by the collision. The cases in which the decay products are different define the net reaction rate, the details of which are hidden in $S(E)$. The rates $S(E)$ can be calculated (with great difficulty!), or one can extrapolate from measurements (which are typically done at far larger energies than those relevant to stellar conditions).

In general, the compound nucleus has several discrete bound states at negative energies in the nuclear potential well, the stable ground state of the nucleus and some excited states that can decay into lower-energy states by emission of

---

[1]For those interested in seeing how the above tunnelling probability is derived: consider a wave encountering a potential barrier with potential $V = z_1 z_2 e^2 / r \gg E$. The wavelength now becomes imaginary (evanescent) and has a magnitude of $|k| = \sqrt{2m(V - E)}/h$ (recall that the de Broglie wavelength $\lambda = 2\pi/k = h/\sqrt{2mE}$). Wave propagation changes phase as $\exp(kr)$, but with k being imaginary, this indicates an exponentially decaying propagation. The probability of tunnelling through a barrier V is then $\exp(-\int_a^b |k| dr)$. The exponent

$$\int_a^b |k| dr = \frac{\sqrt{2m} z_1 z_2 e^2}{h} \int_{v_b}^{v_a} \frac{\sqrt{V - E}}{V^2} dV = \frac{\sqrt{2m} z_1 z_2 e^2}{\sqrt{E} h} \int_{u_b}^{u_a} \frac{\sqrt{u - 1}}{u^2} du \tag{6.9}$$

where we have used multiple times the technique of changing integration variables, in particular, $u = V/E$. Taking $u_b = 1$ (where $V = E$) and $u_b = \infty$ (where $V = 0$), we obtain the tunnelling probability as in eq. (6.8), except for factors of order unity which we have glossed over in this simplified derivation. It is interesting to note that most of the physics here is just classical wave physics.
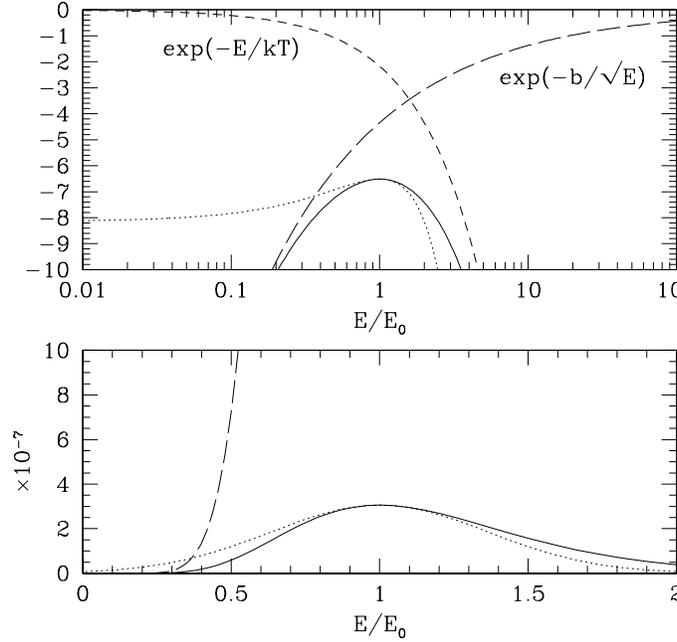
Figure 6.2: Gamow peak resulting from the competing exponential terms: (1) from the Maxwellian (short-dashed line: $\propto \exp(-E/kT)$, with $kT = 0.2\,E_0$ here), and (2) from the penetration probability (long-dashed line: $\propto \exp(-b/\sqrt{E})$, with $b = 10\,\sqrt{E_0}$ here). The solid line indicates the product, and the dotted line the approximating Gaussian discussed in the text. *Upper panel:* logarithmic scale; *lower panel:* linear scale.

photons (γ-rays). These states are similar to the bound states of electrons in an atom, but comprising nucleons instead of electrons. However, the compound nucleus may also have quasi-stable excited states of positive energy (below the top of the Coulomb barrier), which can decay by emission of particles (by quantum tunnelling *outwards* through the Coulomb barrier) as well as by emission of a photon. Incoming particles with "resonant" energy corresponding to such a quasi-stable state can form a compound nucleus much more easily, leading to a greatly enhanced reaction rate.

Given the cross section $\sigma(E)$, the reaction rate between particles of types a and b (at a given energy E) is given by

$$r_{a,b}(E)\,\mathrm{d}E = n_a n_b v \sigma(E) f(E)\,\mathrm{d}E \;, \tag{6.10}$$

where n_a and n_b are the number densities of a and b, v is the relative velocity between a and b (corresponding to energy E), $f(E)$ is the energy probability distribution, and $\sigma(E)$ is the cross section defined above. The factor v accounts for the fact that for larger velocities v, more particles pass each other per unit time. Note that if particles a and b are identical, we need to multiply the above by $\frac{1}{2}$ in order to avoid counting double. Including that in the integrated reaction rate, we find a rate

$$r_{a,b} = \frac{1}{1 + \delta_{a,b}} n_a n_b \langle \sigma v \rangle \;, \qquad \text{where} \quad \langle \sigma v \rangle \equiv \int_0^\infty v(E)\sigma(E)f(E)\,\mathrm{d}E \tag{6.11}$$

is the average reaction rate per pair of particles, i.e., $\langle \sigma v \rangle$ is an effective cross-section.

If the velocity probability distributions are Maxwellian for both particles (i.e., particles have momenta as in Eq. 2.8, divided by n), the distribution of the *relative velocity* of the particles is also Maxwellian, but with $m = m' = m_a m_b/(m_a + m_b)$ [verify this]. We can rewrite the Maxwell distribution in Eq. 2.8 as a function of energy using $p = \sqrt{2mE}$ and $\mathrm{d}p = \frac{1}{2}\sqrt{2m/E}\,\mathrm{d}E$,

$$f(E)\,\mathrm{d}E = \frac{2\pi\sqrt{E}}{(\pi kT)^{3/2}}\mathrm{e}^{-E/kT}\,\mathrm{d}E \;. \tag{6.12}$$

Hence, for the effective cross section (using $v(E) = p/m = \sqrt{2E/m}$),

$$\langle \sigma v \rangle = \left(\frac{8}{m'\pi}\right)^{1/2}\left(\frac{1}{kT}\right)^{3/2}\int_0^\infty S(E)\mathrm{e}^{-E/kT}\mathrm{e}^{-b/\sqrt{E}}\,\mathrm{d}E \;. \tag{6.13}$$

25

Assuming $S(E)$ is a slowly varying function, the integrand will be small everywhere (due to the two exponential terms that are respectively important at different energy limits, *think physically why*), except where the term $h(E) \equiv -E/kT - b/\sqrt{E}$ in the exponential reaches a maximum. This is called the **"Gamow peak"** (Fig. 6.2) and is where most of the contribution to reaction arises. This position E_0 is obtained as,

$$\frac{\mathrm{d}h(E)}{\mathrm{d}E} = \frac{\mathrm{d}}{\mathrm{d}E}(-E/kT - b/\sqrt{E}) = 0 \quad \Rightarrow$$

$$E_0 = \left(\frac{bkT}{2}\right)^{2/3} = 5.665\,\mathrm{keV}\,(Z_1 Z_2)^{2/3}\left(\frac{m'}{m_\mathrm{u}}\right)^{1/3} T_7^{2/3}\,, \tag{6.14}$$

where m_u is the atomic unit mass.

Of relevance for the integral is the width of the Gamow peak. Using a Taylor expansion of $h(E)$ around its maximum,

$$h(E) = h_0 + h_0'(E - E_0) + \frac{1}{2}h_0''(E - E_0)^2 + \ldots \simeq -\tau - \frac{1}{4}\tau\left(\frac{E}{E_0} - 1\right)^2 + \ldots\,, \tag{6.15}$$

where we have used the fact that the first derivative h_0' must be zero (since we are expanding around the maximum), and where we have defined

$$\tau = \frac{3E_0}{kT} = 19.721\,(Z_1 Z_2)^{2/3}\left(\frac{m'}{m_\mathrm{u}}\right)^{1/3} T_7^{-1/3}\,. \tag{6.16}$$

Using this in the integral, the exponential is approximately a Gaussian, as one can see by substituting $\xi = (E/E_0 - 1)\sqrt{\tau}/2$,

$$\int_0^\infty \mathrm{e}^{h(E)}\,\mathrm{d}E = \int_0^\infty \mathrm{e}^{-\tau - \frac{1}{4}\tau(E/E_0 - 1)^2}\,\mathrm{d}E = \frac{2}{3}kT\tau^{1/2}\mathrm{e}^{-\tau}\int_{-\sqrt{\tau}/2}^\infty \mathrm{e}^{-\xi^2}\,\mathrm{d}\xi\,. \tag{6.17}$$

Since τ is relatively large and the main contribution to the integral comes from the range close to E_0 (i.e., $\xi = 0$), the error introduced by extending the integration to $-\infty$ is small, i.e., the integral is approximately $\sqrt{\pi}$. For the Gaussian, the fractional full width at half maximum $\Delta E/E_0$ is

$$\frac{\Delta E}{E_0} = 4\left(\frac{\ln 2}{\tau}\right)^{1/2} = 0.750\,(Z_1 Z_2)^{-1/3}\left(\frac{m'}{m_\mathrm{u}}\right)^{-1/6} T_7^{1/6}\,. \tag{6.18}$$

Doing the integration using the Gaussian and inserting the result in Eq. 6.13 (after taking out the slowly varying $S(E)$), one obtains

$$\langle \sigma v \rangle = \frac{4}{3}\left(\frac{2}{m'}\right)^{1/2}\left(\frac{1}{kT}\right)^{1/2} S_0 \tau^{1/2}\mathrm{e}^{-\tau}\,, \tag{6.19}$$

where $S_0 = S(E_0)$. Since $T \propto \tau^{-3}$ (Eq. 6.16), one thus has that $\langle \sigma v \rangle \propto \tau^2 \mathrm{e}^{-\tau}$. It is the exponential, however, that really determines the reaction speeds. The dependences on Z_1, Z_2, and m' ensure that more massive, more highly charged ions hardly react at all as long as the fusion processes of the lighter elements still are taking place.

It is often useful to know the temperature dependence of the reaction rate, given by

$$\nu \equiv \frac{\partial \ln \langle \sigma v \rangle}{\partial \ln T} = \frac{1}{3}(\tau - 2) = 6.574\,(Z_1 Z_2)^{2/3}\left(\frac{m'}{m_\mathrm{u}}\right)^{1/3} T_7^{-1/3} - \frac{2}{3} \tag{6.20}$$

(note that, for a given reaction, ν usually becomes smaller with increasing temperature). For the fusion of two protons in the centre of the Sun, $Z_1 = Z_2 = 1$, $m' = \frac{1}{2}$, $T_7 \simeq 1.5$, hence $\nu \simeq 4$, which is a relatively mild temperature dependence. For other fusion processes, we will find exponents of $\nu \sim 20$ and above, making these processes among the most strongly varying functions in physics.

Corrections to the above rate formulae

A few corrections are usually made in more detailed derivations. The first is a small correction factor $g_{a,b}$ to account for any temperature dependence of S_0 and for the inaccuracy of approximating the Gamow peak by a Gaussian. The second is more physical, and is a correction $f_{a,b}$ for the effect of electron screening — due to the presence of electrons, the effective potential that two ions see is slightly reduced ("screened"); as a result, the reaction will be faster. This correction is more important at higher densities, and at *very* high densities burning starts to depend more sensitively on the density than on the temperature. (For this case, one speaks of *pycnonuclear reactions*.) Also, separate terms may be added to account for resonances.

Nuclear fusion energy production

The energy production per unit stellar mass can be conviniently expressed as (dimension [erg/s/g])

$$\epsilon_{\mathrm{nuc}} = (\Delta m\, c^2) \frac{1}{t_{\mathrm{mfp}}} \frac{n}{\rho} = (\Delta m\, c^2) \frac{X_1 X_2}{A m_H} n < \sigma v > \quad \sim \quad \epsilon_0 X_1 X_2 \rho T^\nu , \tag{6.21}$$

where X_1, X_2 are the mass abundances of the two reactants and A the atomic mass per nuclei. The temperature scaling ν is of importance in determining stellar structure. In the following, we give details of different sorts of nuclear burning. Pay special attention to the value of ν.

6.1 Hydrogen burning

In principle, many nuclear reactions can occur at the same time. As we saw above, however, the weighting of the exponential with $(Z_1 Z_2)^{2/3}$ strongly inhibits processes involving more massive, more highly charged particles. In combination with the initial abundances of stars, with the largest fraction of the mass being hydrogen, generally only a small number of fusion processes turn out to be relevant in a given evolutionary stage.

P-P chain

In less massive stars ($M \leq 1.2\, M_\odot$), the fusion of hydrogen to helium on the main sequence is mostly by the *proton-proton chain* (p-p chain). The possible variants of the p-p chain are:

$$
\begin{aligned}
{}^1\mathrm{H} + {}^1\mathrm{H} &\rightarrow {}^2\mathrm{D} + \mathrm{e}^+ + \nu \\
{}^2\mathrm{D} + {}^1\mathrm{H} &\rightarrow {}^3\mathrm{He} + \gamma
\end{aligned}
$$

$$
{}^3\mathrm{He} + {}^3\mathrm{He} \rightarrow {}^4\mathrm{He} + 2\,{}^1\mathrm{H} \qquad \textbf{or} \qquad {}^3\mathrm{He} + {}^4\mathrm{He} \rightarrow {}^7\mathrm{Be} + \gamma
$$

$$
\textbf{pp1}
$$

$$
\begin{aligned}
{}^7\mathrm{Be} + \mathrm{e}^- &\rightarrow {}^7\mathrm{Li} + \nu \qquad \textbf{or} \qquad {}^7\mathrm{Be} + {}^1\mathrm{H} \rightarrow {}^8\mathrm{B} + \gamma \\
{}^7\mathrm{Li} + {}^1\mathrm{H} &\rightarrow {}^4\mathrm{He} + {}^4\mathrm{He} \qquad\qquad\qquad\quad {}^8\mathrm{B} \rightarrow {}^8\mathrm{Be} + \mathrm{e}^+ + \nu \\
& \qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad {}^8\mathrm{Be} \rightarrow {}^4\mathrm{He} + {}^4\mathrm{He}
\end{aligned}
$$

$$
\textbf{pp2} \qquad\qquad\qquad\qquad\qquad\qquad\qquad \textbf{pp3}
$$
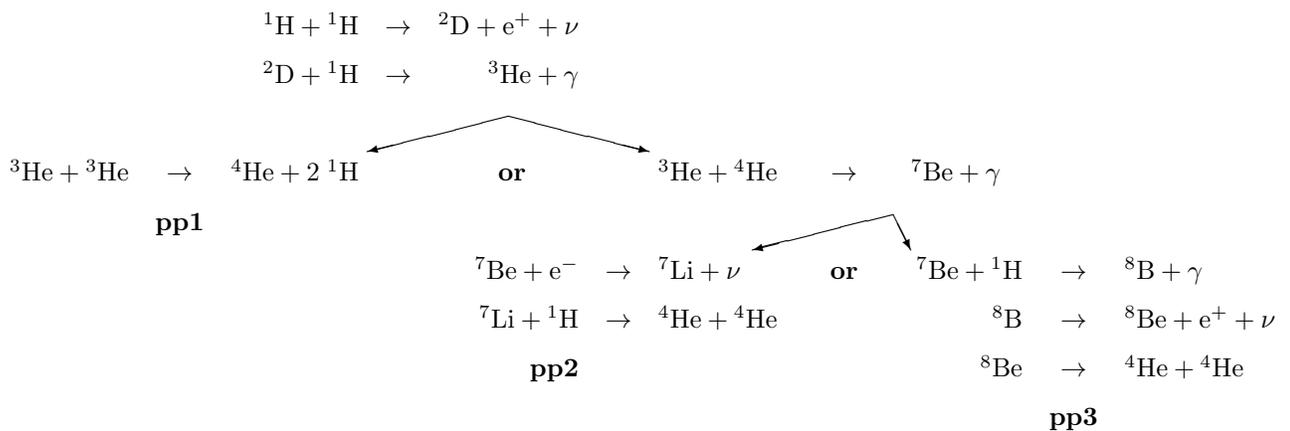
In these chains, the positrons made will meet an electron and annihilate, adding $1.022\,\mathrm{MeV}$ of photon energy. Note that while the total energy released (per ${}^4\mathrm{He}$ produced) for the three chains is equal, the fraction of that energy put in *neutrinos* is not the same. The net energy put into the local medium per ${}^4\mathrm{He}$ nucleus produced is $26.20\,\mathrm{MeV}$ for pp1, 25.67 for pp2, and 19.20 for pp3.

The relative frequency of the branches depends on the temperature, density, and chemical composition. Since the reduced mass is slightly larger for the ${}^3\mathrm{He} + {}^4\mathrm{He}$ reaction than it is for the ${}^3\mathrm{He} + {}^3\mathrm{He}$ reaction, it will have a slightly larger

temperature sensitivity. With increasing temperature, pp2 and pp3 will therefore start to dominate over pp1 if ^{4}He is present in appreciable amounts. Similarly, with increasing temperature, the importance of proton capture on ^{7}Be will start to dominate over the electron capture.

For low temperatures, say $T_7 \leq 0.8$, one has to calculate all the reactions independently and keep track of relative abundances. For higher temperatures, the intermediate reactions will be in equilibrium, and the energy generation can be taken to be proportional to the first step, which is the slowest – a "bottle-neck" for the p-p chain. This is because it involves the weak nuclear force in the decay of a proton to a neutron during the short time the two protons are together. Indeed, in by far most cases, the compound two-proton nucleus that is formed at first, will simply break apart into two protons again. As a result, the effective cross-section is very small, $\sim 10^{-47}\,\mathrm{cm}^2$. The corresponding nuclear fusion energy yield is

$$\epsilon_{\mathrm{pp}} = 2.54\,10^6\,\mathrm{erg\,s^{-1}\,g^{-1}}\;\psi f_{1,1}\,g_{1,1}\,X_1^2\,\rho\,T_6^{-2/3}\,\mathrm{e}^{-33.81/T_6^{1/3}}, \tag{6.22}$$

with an uncertainty of about 5%. Here, $g_{1,1} \simeq 1 + 0.00382 T_6$, $f_{1,1} \simeq 1$ for electron screening, and ψ corrects for the relative contributions of the different chains. At $T_7 \leq 1$, $\psi \simeq 1$, but at $T_7 = 2$, it varies between 1.4 for $Y = 0.3$ to nearly 2 for $Y = 0.9$. At still higher temperatures, when pp3 starts to dominate, it goes to 1.5 almost independent of Y. The temperature dependence of the reaction, as calculated from Eq. 6.20, is relatively mild: $\nu \simeq 4$ (i.e., $\epsilon_{\mathrm{pp}} \propto T^4$, much less steep than we will find below for other reactions).

Deuterium burning Though rare, there is Deuterium in nature, with an interstellar Deuterium abundance is $D/H \sim 2.5 \times 10^{-5}$.[2] Deuterium burning circumvents the first step in p-p chain and can occur at a lower temperature ($T \sim 10^6\,\mathrm{K}$).

This process can temporarily halt the collapse of a proto-star at a lower density and a lower central temperature ($\sim 10^6\,\mathrm{K}$) than otherwise possible, until Deuterium in the convective core is exhausted. This effect is visible in Fig. 8.2. For very low mass stars (brown-dwarfs, $M \leq 80 M_J$), Deuterium burning is all the nuclear fusion they can ever achieve. In contrast, lower mass objects ($M \leq 13 M_J$) can not even ignite Deuterium, and we call these 'planets'.

Since Deuterium is destroyed in pre-main sequence stars, and it is also destroyed by cosmic-ray spallation, one expects the cosmic Deuterium abundance to decrease with time. It is therefore surprising that the measured ISM value is not too different from the prediction from big-bang nucleosynthesis (2.7×10^{-5}).

Lithium burning Also part of the p-p cycle, Li, Be and B have fusion temperatures of $2.5 \times 10^6\,\mathrm{K}$, $4 \times 10^6\,\mathrm{K}$ and $5 \times 10^6\,\mathrm{K}$, respectively. So these are readily destroyed in the cores of main sequence stars, but not in brown dwarfs – the presence of Li in the spectrum identifies brown dwarfs from stars.

There are two interesting Lithium puzzles.

The solar surface exhibits a very low Li abundance – less than a percent of the meteoritic value (Anders & Grevesse, 1989). This implies more depletion than possible in the pre-main sequence phase and is called the "solar lithium problem" (Fig. 6.3). It seems other old stars also exhibit this behaviour (Baumann et al, 2010), with a decreasing surface Lithium abundance with age. This points to some mixing mechanism that continuously transport surface Lithium from the convection zone into the deep radiative interior.

Another is called the "primordial Lithium problem". Big-bang nucleosynthesis makes a precise prediction for the primordial Lithium abundance at $Li/H \sim 5 \times 10^{-10}$. Metal poor stars also exhibit a flat Lithium abundance (see Fig. 6.3), indicating a primordial nature (the so-called 'Spite plateau'). The trouble is, the latter is some three times lower than the former one. There may be unknown physics operating.

CNO cycle

At sufficiently high temperatures, hydrogen can be burned to helium via the CNO cycle, in which carbon, nitrogen, and oxygen act more or less as catalysts (these have to be present, of course). The reactions are split in a main cycle (CN cycle)

---

[2]Similar to that in Jupiter and the outer atmosphere of the Sun. In contrast, the Earth's ocean has $D/H \sim 1.5 \times 10^{-4}$.
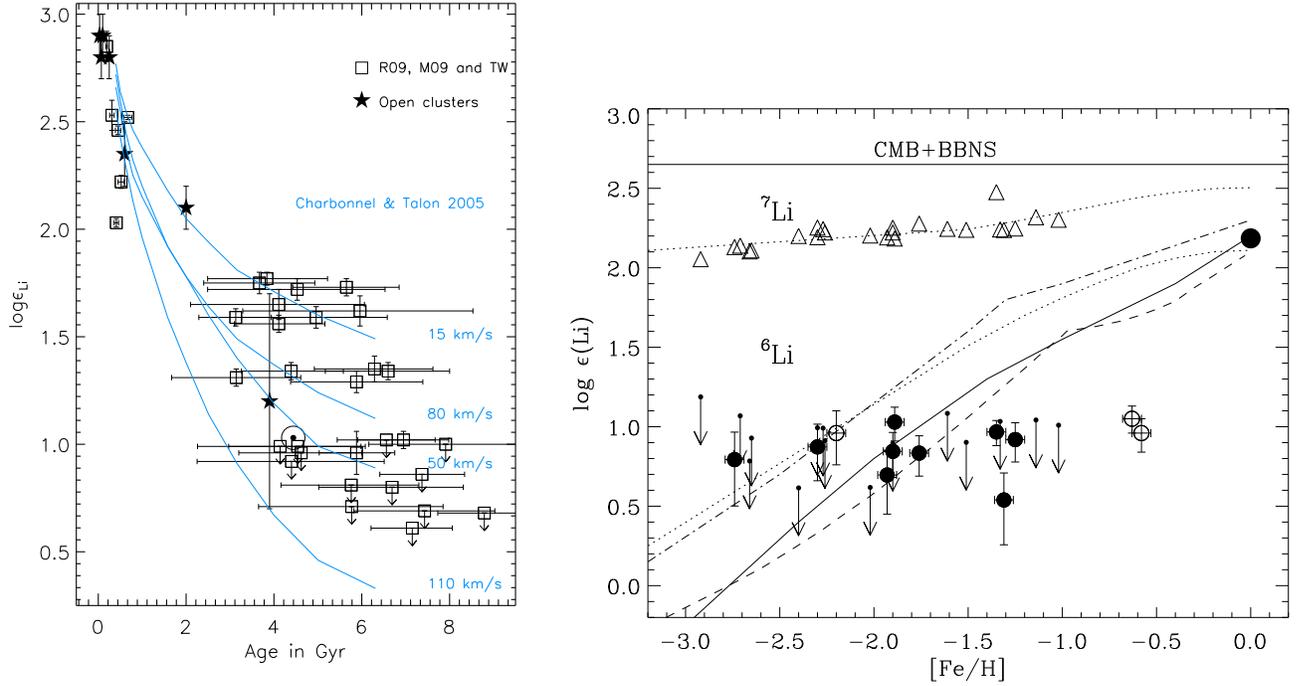
Figure 6.3: (Left) Surface Lithium abundance (in this logarithim scale, Hydrogen is at 12) for a number of solar twins as a function of their ages. The Sun is marked with $\odot$. There appears to be gradual Lithium depletion over Gyrs timescale. The solid lines are the predicted values from the models by Charbonnel & Talon (2005) for various initial rotational velocities. From Baumann et al (2010). *(Right)*: Lithium abundances in selected metal-poor Galactic halo stars from Asplund et al (2006). Fore each star, elemental Li = ^{6}Li + ^{7}Li is plotted at the stars metallicity $[Fe/H] = \log[(Fe/H)_{\rm obs}/(Fe/H)_\odot]$. The flatness of Li vs Fe is the "Spite plateau" and indicates that the bulk of the lithium is unrelated to Galactic nucleosynthesis processes and thus is primordial. The horizontal band gives the BBN+WMAP prediction; the gap between this and the plateau illustrates the ^{7}Li problem. Points below the plateau show ^{6}Li abundances; the apparent plateau constitutes the ^{6}Li problem. More metal rich stars have a spread of Lithium abundances. Taken from the review by Fields (2011).

and a secondary cycle (ON cycle), as follows:

M1.	$^{12}\mathrm{C} + {}^1\mathrm{H}$	$\rightarrow$	$^{13}\mathrm{N} + \gamma$
M2.	$^{13}\mathrm{N}$	$\rightarrow$	$^{13}\mathrm{C} + \mathrm{e}^+ + \nu$
M3.	$^{13}\mathrm{C} + {}^1\mathrm{H}$	$\rightarrow$	$^{14}\mathrm{N} + \gamma$
M4.	$^{14}\mathrm{N} + {}^1\mathrm{H}$	$\rightarrow$	$^{15}\mathrm{O} + \gamma$
M5.	$^{15}\mathrm{O}$	$\rightarrow$	$^{15}\mathrm{N} + \mathrm{e}^+ + \nu$

M6. $\quad ^{15}\mathrm{N} + {}^1\mathrm{H} \quad \rightarrow \begin{cases} ^{12}\mathrm{C} + {}^4\mathrm{He} & \text{and back to line M1 (main CN cycle).} \\ ^{16}\mathrm{O} + \gamma & \text{(secondary ON cycle):} \end{cases}$

S1.	$^{16}\mathrm{O} + {}^1\mathrm{H}$	$\rightarrow$	$^{17}\mathrm{F} + \gamma$
S2.	$^{17}\mathrm{F}$	$\rightarrow$	$^{17}\mathrm{O} + \mathrm{e}^+ + \nu$

S3. $\quad ^{17}\mathrm{O} + {}^1\mathrm{H} \quad \rightarrow \begin{cases} ^{14}\mathrm{N} + {}^4\mathrm{He} & \text{and back to line M4.} \\ ^{18}\mathrm{F} + \gamma & \end{cases}$

S4.	$^{18}\mathrm{F}$	$\rightarrow$	$^{18}\mathrm{O} + \mathrm{e}^+ + \nu$
S5.	$^{18}\mathrm{O} + {}^1\mathrm{H}$	$\rightarrow$	$^{15}\mathrm{N} + {}^4\mathrm{He}$ and back to line M6.
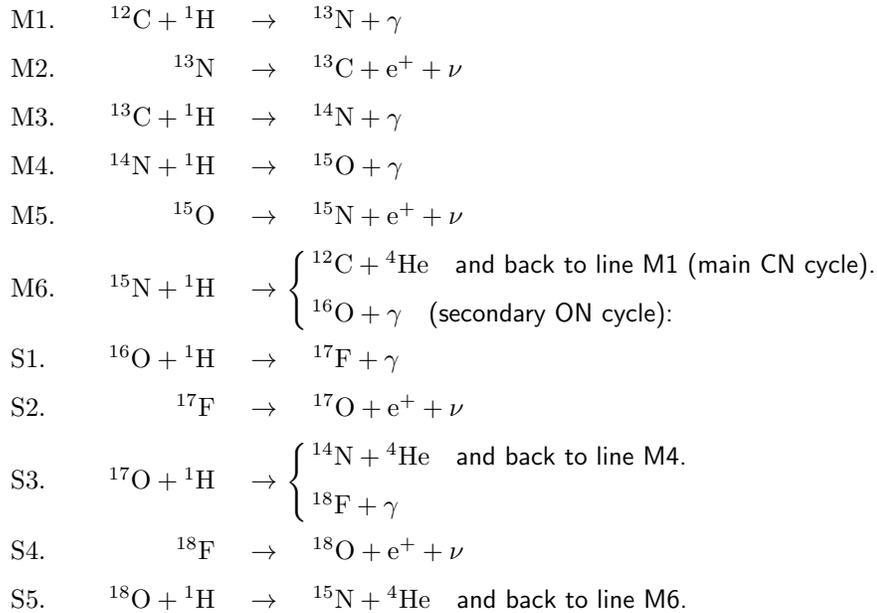
Figure 6.4: Energy generation rates for matter with $\rho = 10\,\mathrm{g\,cm^{-3}}$, $X_1 = 0.7$, $X_{\mathrm{CNO}} = 0.01$, and a range of temperatures. The contributions from the p-p chain (short-dashed) and CNO cycle (long-dashed) are also indicated separately. Notice the much steeper temperature dependence for the latter. Central temperature for the Sun is $T \sim 1.5 \times 10^7\,\mathrm{K}$. More massive stars have (slightly) hotter central temperatures.

The branch to the ON cycle (at line M6) is roughly 10^{-3} to 10^{-4} times less likely than the main branch back to the beginning of the CN cycle. The ON cycle is important, however, since it results in oxygen being converted to nitrogen (which takes part in the CN 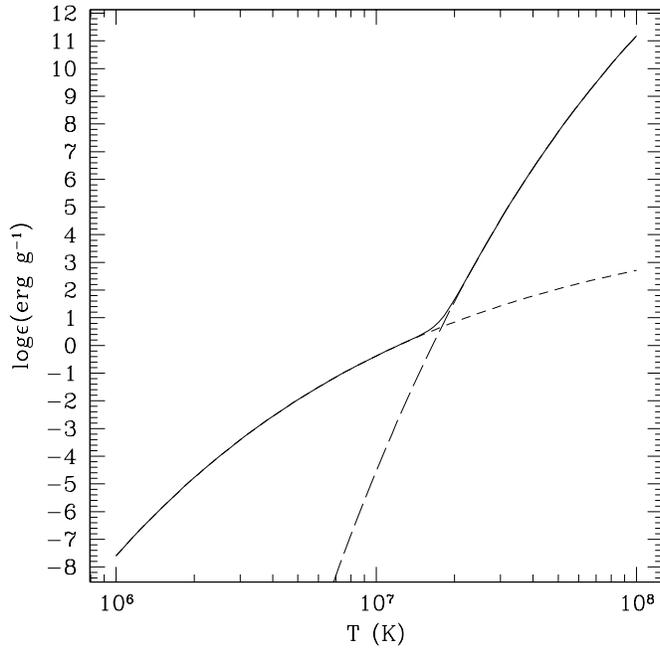cycle) — the branching *inside* the ON cycle (at line S3) does not strongly favor one branch over the other, but both branches lead to the CN cycle. The beta-decay times are of order $10^2 \ldots 10^3$ seconds, *much* shorter than typical nuclear reaction timescales.

Again, for high enough temperatures the reaction cycle will reach equilibrium, and the reaction rate will be set by the slowest link in the CN cycle, which is the proton-capture on ^{14}N. Because of this bottleneck in the CN cycle, and due to the small branching ratio into the ON cycle, most of the CNO originally present will be turned into ^{14}N. The energy gain of the whole cycle, after taking out neutrino losses, is 24.97 MeV, and one finds

$$\epsilon_{\mathrm{CNO}} = 7.48\,10^{27}\,\mathrm{erg\,s^{-1}\,g^{-1}}\,g_{14,1}\,f_{14,1}\,X_{\mathrm{CNO}}\,X_1\,\rho\,T_6^{-2/3}\,\mathrm{e}^{-152.31/T_6^{1/3}-(T_6/800.)^2} \tag{6.23}$$

(with an uncertainty of $\pm 10\%$), where $g_{14,1} \simeq 1 - 0.002T_6$, $f_{14,1} \sim 1$ for electron screening, and $X_{\mathrm{CNO}} = X_{\mathrm{C}} + X_{\mathrm{N}} + X_{\mathrm{O}}$. At somewhat lower temperatures, the CN cycle can reach equilibrium, but the burning of ^{16}O proceeds slowly; Eq. 6.23 is still quite a good approximation, but with $X_{\mathrm{CNO}} = X_{\mathrm{C}} + X_{\mathrm{N}} + |\Delta X_{\mathrm{O}\to\mathrm{N}}(t)|$, where $|\Delta X_{\mathrm{O}\to\mathrm{N}}(t)|$ is the amount of ^{16}O that has been burned to nitrogen as of time t (note that the intermediate ^{17}O stage may also slow down the conversion of ^{16}O to nitrogen, since the reaction rates of ^{16}O and ^{17}O may be comparable).

Inside stars that burn predominantly via the CNO cycle, the nitrogen abundance will be far larger than it normally is, while carbon and oxygen will be correspondingly underabundant. Indeed, such abundance patterns are observed in massive stars which have lost a lot of mass, so that processed material reaches the surface. Examples of these are the ON stars and Wolf-Rayet stars of type WN. (In carbon-rich Wolf-Rayet stars, one even sees the products of helium fusion.) Also, in lower-mass red giants, some CNO-processed material is mixed to the surface.

For the CNO cycle, the temperature sensitivity is high, $\nu = 23 \ldots 13$ for $T_6 = 10 \ldots 50$. As a result, the p-p chain dominates at low temperatures, and the CNO cycle at high temperatures, as is illustrated in Fig. 6.4. Furthermore, because of the steep temperature dependence, the energy production will be highly concentrated towards the centre. Therefore, L_r/r^2 will be large, and thus ∇_{rad} will be large as well. This is why massive stars have convective cores.

6.2 Helium burning

When all the hydrogen has been fused into helium, it is difficult to continue, because until one reaches carbon, the elements following helium have lower binding energy per nucleon (see Fig. 6.1). As a result, the fusion of two helium nuclei leads to a ^{8}Be nucleus whose ground state is nearly $100\,\text{keV}$ lower in energy; therefore, it decays back into two alpha particles in a few $10^{-16}\,\text{s}$. Nevertheless, this is still about 10^5 times longer than the encounter time — in fact, a ^{8}Be abundance of about 10^{-9} builds up in stellar matter. Occasionally, it will happen that another alpha particle comes by so that a carbon nucleus can be formed. This whole process is called the *triple-alpha reaction* because it almost is a three-body interaction. Writing out the reactions,
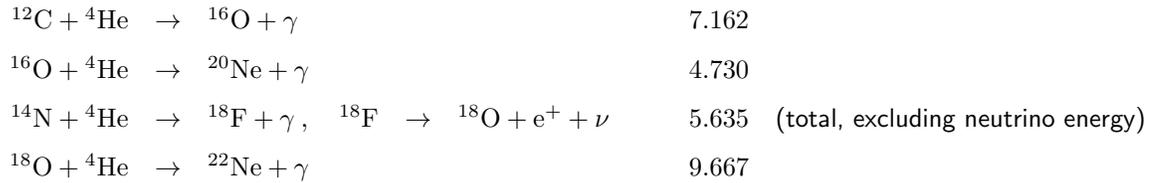
$$
\begin{aligned}
^4\text{He} + {}^4\text{He} &\rightleftharpoons {}^8\text{Be} \\
^8\text{Be} + {}^4\text{He} &\rightarrow {}^{12}\text{C} + \gamma
\end{aligned}
$$
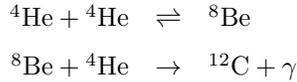
The total energy released per carbon nucleus formed is $7.274\,\text{MeV}$. For these reactions, it is much less straightforward to derive an energy generation rate, because "*resonances*" (as described above) are important for both the above steps. Roughly, the energy generation rate is

$$
\epsilon_{3\alpha} = 4.99\,10^{11}\,\text{erg s}^{-1}\,\text{g}^{-1}\,f_{3\alpha}\,Y^3\,\rho^2\,T_8^{-3}\left(1 + 0.00354T_8^{-0.65}\right)\text{e}^{-43.92/T_8} \tag{6.24}
$$

(with an uncertainty of $\pm 14\%$), where $f_{3\alpha} = f_{4,4}f_{8,4}$ is the combined electron screening factor. For this reaction, the temperature sensitivity is very high, $\nu = 40 \ldots 19$ for $T_8 = 1 \ldots 2$.

Other fusion processes can occur simultaneously (energy gain in MeV is shown to the right):

$$
\begin{aligned}
^{12}\text{C} + {}^4\text{He} &\rightarrow {}^{16}\text{O} + \gamma & 7.162 \\
^{16}\text{O} + {}^4\text{He} &\rightarrow {}^{20}\text{Ne} + \gamma & 4.730 \\
^{14}\text{N} + {}^4\text{He} &\rightarrow {}^{18}\text{F} + \gamma\,, \quad {}^{18}\text{F} \rightarrow {}^{18}\text{O} + \text{e}^+ + \nu & 5.635 \quad \text{(total, excluding neutrino energy)} \\
^{18}\text{O} + {}^4\text{He} &\rightarrow {}^{22}\text{Ne} + \gamma & 9.667
\end{aligned}
$$

The second of these is slow, and for the last two ^{14}N is not very abundant (and thus its product ^{18}O is not very abundant either). The first reaction is therefore the most important one. It is rather complicated (and has an uncertainty of $\pm 40\%$); approximately,

$$
\begin{aligned}
\epsilon_{12,\alpha} \simeq\ & 9.58\,10^{26}\,\text{erg s}^{-1}\,\text{g}^{-1}\,f_{12,4}\,X_{12}\,Y\rho\,T_8^{-2}\left[\left(1 + 0.254T_8 + 0.00104T_8^2 - 0.000226T_8^3\right)\text{e}^{-(T_8/46.)^2}\right. \\
& \left. + \left(0.985 + 0.9091T_8 - 0.1349T_8^2 + 0.00729T_8^3\right)\text{e}^{-(T_8/13.)^2}\right]\text{e}^{-71.361/T_8^{1/3}}\,. \tag{6.25}
\end{aligned}
$$

6.3 Carbon burning and onward

After helium has been exhausted, the next processes to start are those of carbon burning, at temperatures of order $T_9 = 0.5 \ldots 1$. The situation is very complicated, since the excited ^{24}Mg nucleus that is produced is unstable and can decay in a number of different ways:

$$
\begin{aligned}
^{12}\text{C} + {}^{12}\text{C} &\rightarrow {}^{24}\text{Mg} + \gamma & 13.931 \\
&\rightarrow {}^{23}\text{Mg} + \text{n} & -2.605 \\
&\rightarrow {}^{23}\text{Na} + \text{p} & 2.238 \\
&\rightarrow {}^{20}\text{Ne} + \alpha & 4.616 \\
&\rightarrow {}^{16}\text{O} + 2\alpha & -0.114
\end{aligned}
$$
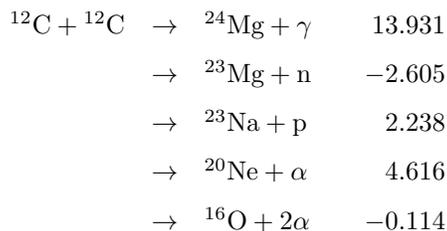
The last column lists the energy gain in MeV. Here, the most probable reactions are those leaving ^{23}Na and ^{20}Ne. The next complication that arises, is that the proton and alpha particle produced in these two reactions immediately fuse with

other particles (since for them, the temperatures are extremely high). As a result of these complications, the energy rate is rather uncertain. For some approximate values, see Kippenhahn & Weigert, p. 167.

For temperatures above 10^9 K, the photon energies become so large that they can lead to the break-up of not-so-tightly bound nuclei. Reaction ra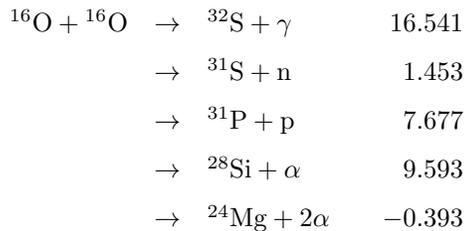tes analogous to the Saha equation for ionization can be written to determine equilibrium conditions. Generally, however, equilibrium will not be reached as time is most definitely running out if a star reaches these stages. A reaction which is important subsequent to Carbon burning is $^{20}\mathrm{Ne} + \gamma \to {}^{16}\mathrm{O} + \alpha$ (the *reverse* of the helium burning reaction). The alpha particles resulting from this photo-disintegration are captured faster by Neon (via $^{20}\mathrm{Ne} + \alpha \to {}^{24}\mathrm{Mg} + \gamma$) than by the Oxygen nuclei, and hence the net reaction is $2\,{}^{20}\mathrm{Ne} + \gamma \to {}^{16}\mathrm{O} + {}^{24}\mathrm{Mg} + \gamma$, with an energy gain of 4.583 MeV. This is called Neon burning.

The next phase is oxygen burning, for which temperatures in excess of 10^9 K are required. As for carbon burning, the reaction can proceed via a number of channels:

$$
\begin{aligned}
{}^{16}\mathrm{O} + {}^{16}\mathrm{O} \;\to\;& {}^{32}\mathrm{S} + \gamma & 16.541 \\
\to\;& {}^{31}\mathrm{S} + \mathrm{n} & 1.453 \\
\to\;& {}^{31}\mathrm{P} + \mathrm{p} & 7.677 \\
\to\;& {}^{28}\mathrm{Si} + \alpha & 9.593 \\
\to\;& {}^{24}\mathrm{Mg} + 2\alpha & -0.393
\end{aligned}
$$

For these reactions, the most frequent product is $^{31}\mathrm{P}$; next most frequent is $^{28}\mathrm{Si}$. Again, the small particles immediately lead to a multitude of other reactions. Among the end products will be a large amount of $^{28}\mathrm{Si}$.

At the end of Oxygen burning, photo-disintegration becomes more and more important. In particular, photo-disintegration of $^{28}\mathrm{Si}$ leads to the ejection of protons, neutrons and alpha particles, which fuse with other $^{28}\mathrm{Si}$ particles to form bigger nuclei that in turn are subjected to photo-disintegration. Still, gradually larger nuclei are built up, up to $^{56}\mathrm{Fe}$. Since iron is so strongly bound, it may survive as the dominant species. The whole process is called silicon burning.

stellar nucleosynthesis yield

All elements past B are not primordial and are instead produced in stars. Nuclear reactions described above determine the abundance pattern of interstellar gas and their offsprings, the stars.

Nuclear reactions depend on the temperature and density. So product yields from AGB stars, Type II SN, Type Ia SN differ. Abundance pattern in the Sun and in two of the most metal-poor stars known (Figs. 6.5-6.6) showcase these differences.

Food for thought

- Nuclear fusion, with its steep temperature dependence, is subject to a 'thermostat' control inside stars. There is no runaway fusion nor dwindling fire. Think why.

- Think about what happens when the core has turned into Iron.

- We have previously derived mass-radius-luminosity for radiative stars, only taking opacity into account. What if the fusion produces too much luminosity for a given star, compared to our results?

- Similarly, what about a convective star?

- What contributes to the very steep temperature dependence for fusion reactions? (e.g., explain eq. [6.22])

- Hydrogen, Deuterium, Lithium, Helium, Carbon... all have a so-called 'ignition temperature'. How does it depend on nucleus charge? and on density? why do different types of fusion have different temperature sensitivity?
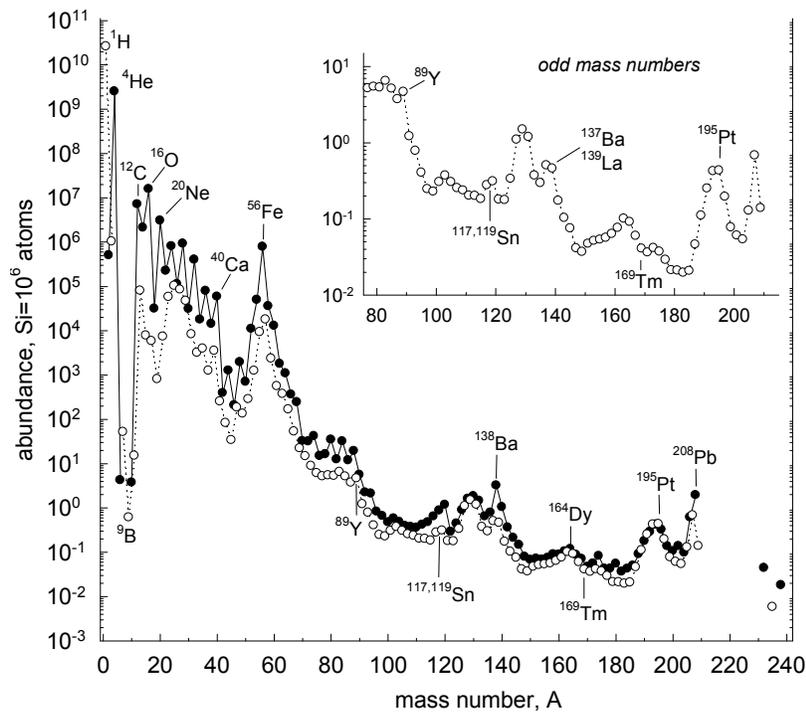
Figure 6.5: Nucleide abundances in the Sun, plotted against atomic mass number. In this logarithimc scale, Silicon is at 6 and Hydrogen at 10.5. Open symbols are for odd mass numbers and full symbols are for even ones. The odd numbered nuclides (inset) are less abundant than the even ones; nucleids that contain integer multiples of 4 (^{12}C, ^{16}O, $^{20}Ne...$, $^{28}Si...$) are called α-elements and are more abundant than others. They are produced in the α-process whereby an α-element fuses with a ^{4}He nuclei, mostly inside Type-II supernova (massive star cores). The prominent Fe-peak is a result of its nuclear stability and its production in Type Ia supernova (low-mass stars). Elements heavier than Fe are either produced by the so-called r-process (rapid neutron capture) in Type II SN or the s-process (slow capture) in AGB stars. Taken from Lodders et al (2009).

Figure 6.6: Logarithmic elmental abundances (compared to Fe) in two of the most metal-poor stars in the Milky way ($[Fe/H] < -5$) in color dots. Notice the extreme shortage of Fe relative to that in the Sun – this suggests that the stars were not yet polluted by SNIa. Superimposed in black curve is the nucleiosynthetic yield from a $25M_\odot$ Pop-III (extremely metal-poor) star (Iwamoto et al 2005).

7 Making a star/Planet

*Readings: CO §10.5, App. H,*Back

I will not have time to cover this chapter in class. But interested people can read on.

The problem

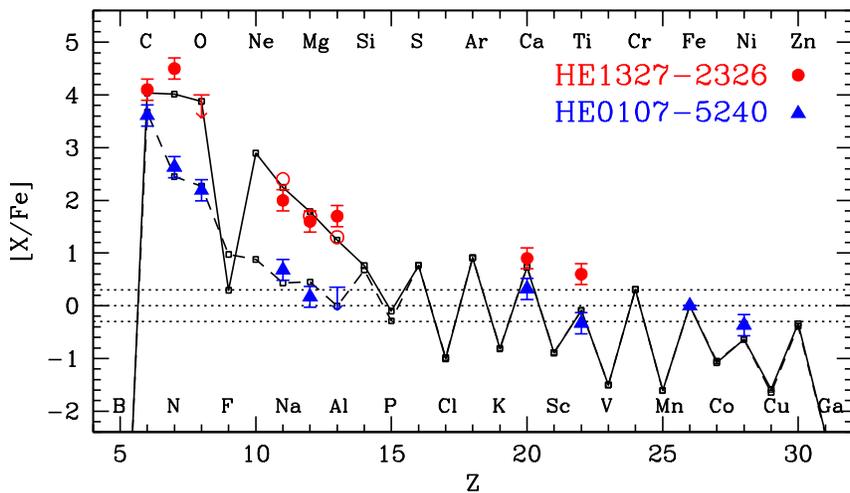To calculate a star's structure, we need to solve the equations of hydrostatic equilibrium, mass continuity, energy balance, and energy transport. It makes most sense to write these in terms of fractional mass M_r rather than fractional radius r (since composition profiles are determined by the position in terms of M_r, which, unlike r, does not change when the star expands or contracts). The mass continuity equation (Eq. **??**) can be used to put the equations into the following form:

$$[\text{mass continuity (Eq. 1.1)}]: \qquad \frac{\mathrm{d}r}{\mathrm{d}M_r} = \frac{1}{4\pi r^2 \rho} \,, \tag{7.1}$$

$$[\text{hydrostatic equilibrium (Eq. 1.5)}]: \qquad \frac{\mathrm{d}P}{\mathrm{d}M_r} = -\frac{GM_r}{4\pi r^4} \,, \tag{7.2}$$

$$[\text{energy balance (Eq. 6.1)}]: \qquad \frac{\mathrm{d}L_r}{\mathrm{d}M_r} = \epsilon_{\mathrm{nuc}} - \epsilon_\nu + \epsilon_{\mathrm{grav}} \,, \tag{7.3}$$

$$[\text{generalized Eddington equation}]: \qquad \frac{\mathrm{d}T}{\mathrm{d}M_r} = -\frac{GM_r T}{4\pi r^4 P}\nabla_* \,. \tag{7.4}$$

In Eq. (7.4), depending on whether the layer is radiative or convective, one has

$$\nabla_* = \begin{cases} \nabla_{\mathrm{rad}} = \dfrac{3}{16\pi acG}\dfrac{\kappa L_r P}{M_r T^4} & (\text{radiative layers}) \,, \\[2mm] \nabla_{\mathrm{ad}} + \nabla_{\mathrm{sa}} & (\text{convective layers}) \,. \end{cases} \tag{7.5}$$

Here, ∇_{sa} is the super-adiabatic part of the gradient (i.e., $\nabla_{\mathrm{sa}} \equiv \nabla_{\mathrm{conv}} - \nabla_{\mathrm{ad}}$); ∇_{sa} can be neglected in the interior (where $\nabla_{\mathrm{conv}} \simeq \nabla_{\mathrm{ad}}$) but not near the surface (where $\nabla_{\mathrm{conv}} > \nabla_{\mathrm{ad}}$). The condition for convection can either be the Ledoux or the Schwarzschild criterion.

Evolution consists of thermal adjustments (via ϵ_{grav}) and changes in the abundances, due to the fusion reactions that proceed with rates $r_{a,b}$ (Eq. 6.11 — note that $\langle\sigma v\rangle$ is a function of T):

$$\frac{\mathrm{d}X_i}{\mathrm{d}t} = \frac{m_i}{\rho}\left(\sum_{j,k} r_{j,k(\to i)} - \sum_{k'}(1+\delta_{i,k'})\,r_{i,k'}\right)\,, \qquad i = 1,\ldots, I \,, \tag{7.6}$$

where i labels all isotopes being considered, $r_{j,k(\to i)}$ are reactions that produce isotope i (from j and k), and $r_{i,k'}$ are reactions that destroy i (and also k'). One of the relations can be replaced by the normalization condition, $\sum_i X_i = 1$ (or this condition can be used to check that you have coded the nuclear reactions correctly!). Furthermore, the abundances should be mixed in convective (and semi-convective) zones, taking account of possible overshooting.

In the above equations, we assume that the equation of state, the opacity, and the nuclear reactions are known functions of composition, temperature, and either density or pressure – these are equivalent, as the usual expression of the equation of state $P = P(\rho, T, X_i)$ can be inverted and expressed as $\rho = \rho(P, T, X_i)$ instead. In other words, as functions of (ρ, T, X_i) or (P, T, X_i), we have:

$$\text{Equation of state:} \quad \{\, P(\rho, T, X_i) \text{ or } \rho(P, T, X_i) \,\}, \; \nabla_{\mathrm{ad}}, \; s, \; C_V, \; C_P, \; \left(\tfrac{\partial \ln P}{\partial \ln T}\right)_\rho, \; \left(\tfrac{\partial \ln P}{\partial \ln \rho}\right)_T$$

$$\text{Opacity (incl. conduction):} \quad \kappa$$

$$\text{Nuclear reaction rates:} \quad r_{j,k}, \; \epsilon_{\mathrm{nuc}}, \; \epsilon_\nu$$

[Note that equation-of-state quantities s, C_V, C_P, $\left(\tfrac{\partial \ln P}{\partial \ln T}\right)_\rho$, and $\left(\tfrac{\partial \ln P}{\partial \ln \rho}\right)_T$ enter into ϵ_{grav} and the formulae that can be used to obtain ∇_{conv} in regions where ∇_{sa} is not negligible.] With the above given, there are as many differential equations as unknowns.

While the equations can be expressed equally well in terms of (ρ, T, X_i), for simplicity, we will assume hereafter that the above are expressed as functions of (P, T, X_i). The unknowns are then $(P, r, L_r, T, X_1, \ldots, X_I)$, whose dependence as a function of M_r and t is to be determined. For this purpose, we need boundary conditions at $M_r = 0$ and $M_r = M$ and initial values for the composition X_i and gravitational energy (e.g., an entropy profile).

Boundary conditions

The inner boundary condition is simple: $r = 0$, $L_r = 0$ for $M_r = 0$. Unfortunately, we cannot put any a priori constraints on P_c and T_c, so that integrating from the centre outwards we have families of two-parameter solutions $r(P_\mathrm{c}, T_\mathrm{c})$ and $L_r(P_\mathrm{c}, T_\mathrm{c})$. For small M_r, we can write these functions as expansions in M_r,

$$r(P_\mathrm{c}, T_\mathrm{c}) = \left(\frac{3}{4\pi\rho_\mathrm{c}}\right)^{1/3} M_r^{1/3} \,, \tag{7.7}$$

$$L_r(P_\mathrm{c}, T_\mathrm{c}) = (\epsilon_\mathrm{nuc,c} - \epsilon_{\nu,\mathrm{c}} + \epsilon_\mathrm{grav,c}) M_r \,, \tag{7.8}$$

where ρ_c and the various ϵ_c are known functions of $(P_\mathrm{c}, T_\mathrm{c})$. These expansions are often more useful than the $M_r = 0$ conditions, since Eqs. (7.1), (7.2), and (7.4) become indeterminate at $M_r = 0$.

At the surface, we will have conditions for P and T, but R and L are unknown a priori, leading to a situation similar to that in the centre: for given M, R, and L, one can calculate $\log g$ and T_eff, which determine the run of pressure and temperature in the atmosphere. Thus, integrating from the surface downwards we have families of two-parameter solutions $P(R, L)$ and $T(R, L)$. Unfortunately, the surface condition is not simple. One could use $P = 0$, $T = 0$ for $M_r = M$, but for convective envelopes this leads to gross errors. Somewhat more elegant is to use the photosphere, where $T_\mathrm{eff} = (L/4\pi R^2 \sigma)^{1/4}$ and $P_\mathrm{phot} = 2g/3\kappa$. The condition for the pressure is derived from requiring $\tau = \frac{2}{3}$ at the photosphere, as was done in the discussion of the Hayashi line (Eq. 5.9); for κ, a suitably chosen average of the opacity above the photosphere has to be used in order to get an accurate value for P_phot (see Fig. 7.1).

The main problem with these simple boundary conditions is that near the surface the assumptions underlying the energy transport equation break down: the photon mean-free path becomes substantial. In these regions, much more detailed radiative transfer calculations are required. One can use a simple "grey atmosphere" approximation (in which one assumes that the opacity κ_ν is equal to the Rosseland value, independent of wavelength) to perform an approximate integral over the atmosphere. An alternate solution to this problem is to leave it to those interested in detailed stellar atmospheres, and use a grid of their results. For given (R, L), one calculates T_eff and $\log g$, and uses this to to interpolate in the (R, L, M) grid of model atmosphere results to find P_*, T_* at the bottom of the atmosphere.

Computational methods

There are several ways one could attempt to calculate stellar models and evolution numerically. First consider the case where $X_i(M_r)$ and $\epsilon_\mathrm{grav}(M_r)$ are known, i.e., where we have to solve just the structure of the star.

In principle, one could simply start integrating from both sides for trial values of $(P_\mathrm{c}, T_\mathrm{c})$ and (R, L), and try to match the two solutions at some intermediate fitting point, by varying the trial values. This is called the *shooting method*. In general, given a good scheme, the solution converges quickly (the program `statstar` in CO, App. H, is a simple example; see NUMERICAL RECIPES, §17.2 for more details). It is not very efficient, however, if one wants to calculate the evolution, in which the star evolves through a series of spatial models which are very similar. For this case, it is better to use a method which uses the spatial model from a previous step as an initial guess and makes small adjustments in order to find the new equilibrium. Most commonly used for this purpose is the *Henyey method*, which is especially well-suited for solving differential equations with boundary conditions on both sides.

The method works as follows. Take a grid of points $M_r^{(j)}$, with $j = 1, \ldots, N$. Then, discretise the differential equations,

35

bring both sides to the left-hand side, and call these $A_i^{(j)}$. Then, a solution will be given by

$$A_i^{(j)} = \frac{y_i^{(j+1)} - y_i^{(j)}}{M_r^{(j+1)} - M_r^{(j)}} - f_i(M_r^{(j+\frac{1}{2})}, y_1^{(j+\frac{1}{2})}, y_2^{(j+\frac{1}{2})}, y_3^{(j+\frac{1}{2})}, y_4^{(j+\frac{1}{2})}) = 0,$$

$$i = 1, \ldots, 4, \quad j = 1, \ldots, N-1 \tag{7.9}$$

where $y_1, \ldots, y_4$ are the four variables of interest (e.g., $y_1 = r$, $y_2 = P$, $y_3 = L_r$, $y_4 = T$), the index i numbers the four equations, and $f_1, \ldots, f_4$ are the right–hand side functions in the differential equations. The superscript $j + \frac{1}{2}$ is meant to indicate that a suitable average of the values at grid points j and $j+1$ is taken (e.g., just a straight mean).

At the inner and outer boundaries, we have

$$\begin{aligned}
B_1^{(\text{in})} &= r^{(1)} - r(P_c, T_c) = y_1^{(1)} - f_1^{(\text{in})}(y_2^{(1)}, y_4^{(1)}) = 0, \\
B_3^{(\text{in})} &= L_r^{(1)} - L_r(P_c, T_c) = y_3^{(1)} - f_3^{(\text{in})}(y_2^{(1)}, y_4^{(1)}) = 0, \\
B_2^{(\text{out})} &= P^{(N)} - P(R, L) = y_2^{(N)} - f_2^{(\text{out})}(y_1^{(N)}, y_3^{(N)}) = 0, \\
B_4^{(\text{out})} &= T^{(N)} - T(R, L) = y_4^{(N)} - f_4^{(\text{out})}(y_2^{(N)}, y_4^{(N)}) = 0,
\end{aligned} \tag{7.10}$$

where we assumed one could determine (P_c, T_c) from the values at the first grid point and (R, L) from those at the last. Note that for the simple case for which $M_r^{(1)} = 0$, the functions $r(P, T)$ and $L_r(P, T)$ are identical to zero. If one choses to work in logarithmic units for $\{\rho, P, r, T\}$, however, the first point cannot be at $M_r = 0$, and therefore the inner boundary conditions are written in their more general form above. Thus, with the above definitions of A, B, a solution for the problem requires $A_i^{(j)} = 0$, $B_i = 0$.

Considering the whole grid, we have $4N$ unknowns $y_i^{(j)}$ and $4(N-1) + 2 + 2 = 4N$ equations. Now suppose that we have a first approximation $y_i^{(j)}(1)$ to the solution. For this initial guess, the constraints will not be met, i.e., $A_i^{(j)}(1) \neq 0$, $B_i(1) \neq 0$, and we need to find corrections $\delta y_i^{(j)}$ such that a second approximation $y_i^{(j)}(2) = y_i^{(j)}(1) + \delta y_i^{(j)}$ does give a solution, i.e., we are looking for changes $\delta y_i^{(j)}$ that imply changes $\delta A_i^{(j)}$, δB_i, such that $A_i^{(j)}(1) + \delta A_i^{(j)} = 0$, $B_i(1) + \delta B_i = 0$, or

$$\begin{aligned}
\delta B_i^{(\text{in})} &= -B_i^{(\text{in})}(1), & i &= 1, 3 \\
\delta A_i^{(j)} &= -A_i^{(j)}(1), & i &= 1, \ldots, 4, \quad j = 1, \ldots, N-1 \\
\delta B_i^{(\text{out})} &= -B_i^{(\text{out})}(1), & i &= 2, 4.
\end{aligned} \tag{7.11}$$

For small enough corrections, we can expand the A and B linearly in $\delta y_i^{(j)}$, and write

$$\sum_{k=1}^{4} \frac{\partial B_i^{(\text{in})}}{\partial y_k^{(1)}} \delta y_k^{(1)} = -B_i^{(\text{in})}, \quad i = 1, 3$$

$$\sum_{k=1}^{4} \frac{\partial A_i^{(j)}}{\partial y_k^{(j)}} \delta y_k^{(j)} + \sum_{k=1}^{4} \frac{\partial A_i^{(j)}}{\partial y_k^{(j+1)}} \delta y_k^{(j+1)} = -A_i^{(j)}, \quad i = 1, \ldots, 4, \quad j = 1, \ldots, N-1 \tag{7.12}$$

$$\sum_{k=1}^{4} \frac{\partial B_i^{(\text{out})}}{\partial y_k^{(N)}} \delta y_k^{(N)} = -B_i^{(\text{out})}, \quad i = 2, 4$$

[we have dropped the (1) numbering the 1$^{\text{st}}$ approximation]. This system has $2 + 4(N-1) + 2 = 4N$ equations which need to be solved for the $4N$ unknown corrections $\delta y_i^{(j)}$. In matrix form,

$$\mathcal{H} \begin{pmatrix} \delta y_1^{(1)} \\ \vdots \\ \delta y_i^{(j)} \\ \vdots \\ \delta y_4^{(N)} \end{pmatrix} = - \begin{pmatrix} B_1^{(\text{in})} \\ \vdots \\ A_i^{(j)} \\ \vdots \\ B_4^{(\text{out})} \end{pmatrix}, \tag{7.13}$$
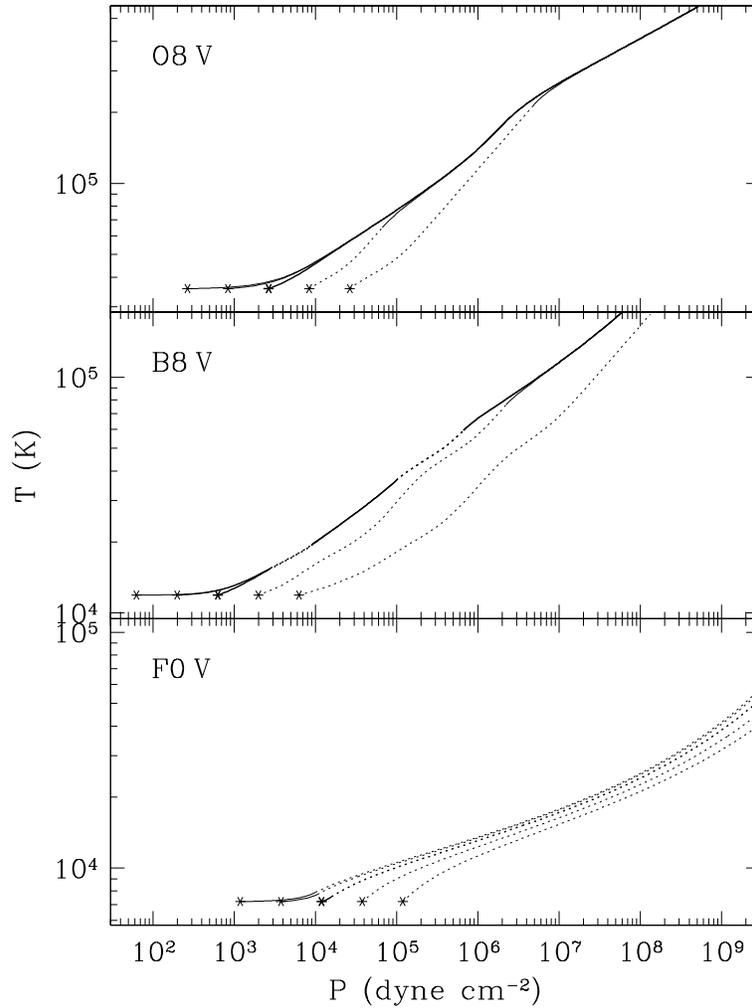
36

where $\mathcal{H}$ is called the *Henyey matrix*. Generally, this matrix equation can be solved ($\det \mathcal{H} \neq 0$), but since we used a first-order expansion, the next approximation $y + \delta y$ will still not fulfill the conditions accurately. Thus, one iterates, until a certain pre-set convergence criterion is met.

Note that Henyey matrix has a relatively simple form, as can be seen by writing out which elements are actually used for the case $N = 3$,

$$
\begin{pmatrix}
\bullet & \bullet & \bullet & \bullet & & & & & & & & \\
\bullet & \bullet & \bullet & \bullet & & & & & & & & \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & & & & \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & & & & \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & & & & \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & & & & \\
& & & & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\
& & & & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\
& & & & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\
& & & & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\
& & & & & & & & \bullet & \bullet & \bullet & \bullet \\
& & & & & & & & \bullet & \bullet & \bullet & \bullet
\end{pmatrix}
\begin{pmatrix}
\delta y_1^{(1)} \\
\delta y_2^{(1)} \\
\delta y_3^{(1)} \\
\delta y_4^{(1)} \\
\delta y_1^{(2)} \\
\delta y_2^{(2)} \\
\delta y_3^{(2)} \\
\delta y_4^{(2)} \\
\delta y_1^{(3)} \\
\delta y_2^{(3)} \\
\delta y_3^{(3)} \\
\delta y_4^{(3)}
\end{pmatrix}
= -
\begin{pmatrix}
B_1^{(\mathrm{in})} \\
B_3^{(\mathrm{in})} \\
A_1^{(1)} \\
A_2^{(1)} \\
A_3^{(1)} \\
A_4^{(1)} \\
A_1^{(2)} \\
A_2^{(2)} \\
A_3^{(2)} \\
A_4^{(2)} \\
B_2^{(\mathrm{out})} \\
B_4^{(\mathrm{out})}
\end{pmatrix}.
$$

Here, the bullets indicate the elements that are used; all others are zero. Because of the simple structure, the solution can be found in a relative straightforward manner. See NUMERICAL RECIPES, §17.3, for details, and for a method that is fast and minimizes storage.

Evolution

So far, we have ignored the chemical evolution and assumed that ϵ_{grav} was a known function. The latter function can be estimated easily once we have made an initial model and want to compute a model one time step later, by approximating

$$
\epsilon_{\mathrm{grav}}^{(j+\frac{1}{2})} = -T^{(j+\frac{1}{2})} \frac{\mathrm{d}}{\mathrm{d}t} s^{(j+\frac{1}{2})} = -\frac{T^{(j+\frac{1}{2})}}{\Delta t} \left(s^{(j+\frac{1}{2})} - s_{\mathrm{prev}}^{(j+\frac{1}{2})} \right), \tag{7.14}
$$

Here, we expressed ϵ_{grav} in terms of the entropy change, but the other expressions in Eq. (6.3) can be used in the same way. The point to note is that s_{prev}, which is the entropy that the element had in the previous model, is known. Since the current entropy is a known function $s(P, T, X_i)$ (from the equation of state), also $\mathrm{d}s/\mathrm{d}t$ is a known function of (P, T, X_i). Thus, ϵ_{grav} is a known function of (P, T, X_i) and can be used without problems in deriving the stellar structure. [A complication arises in convective regions, especially ones that are advancing into regions of different chemical composition. Mixing at constant pressure has no energy cost, but since it is an irreversible process, it results in an increase of entropy (which of course does *not* contribute towards ϵ_{grav}). On the other hand, when one is mixing the products of nuclear burning (i.e., *heavy* nuclei) outwards against gravity while mixing unburned stellar material (i.e., *light* nuclei) downwards, there is an energy cost involved in doing this (which is incurred throughout the region where material mixed upwards has a higher mean molecular weight than material mixed downwards). These effects may need to be accounted for correctly during stages when the star is evolving on a short timescale, which involves some modification of Eq. (6.3) for ϵ_{grav}.]

A scheme like the above for including a variable that changes in time is usually called an *implicit scheme*, since the time derivative is calculated implicitly, using parameters from the new model one is trying to determine. Schemes which rely only on previous model(s) are called *explicit*; these are often easier to code but in order to keep good accuracy small timesteps need to be taken.

For the abundances, an explicit scheme is simpler. In such a scheme, one determine the time derivatives $(\mathrm{d}X_i/\mathrm{d}t)_{\mathrm{prev}}$

Figure 7.1: Effect on the stellar envelope of choosing an incorrect value of P_{phot} in main sequence stars of solar metallicity, for a massive star with a fully radiative envelope (O8 V: $T_{\mathrm{eff}} \approx$ 37,000 K, $M \approx 15\ M_\odot$), an intermediate mass star with very small convective zones in ionization regions (B8 V: $T_{\mathrm{eff}} \approx$ 12,000 K, $M \approx 2.5\ M_\odot$), and a relatively low-mass star with a convective envelope of non-negligible extent (F0 V: $T_{\mathrm{eff}} \approx$ 7,200 K, $M \approx 1.2\ M_\odot$). Star symbols ("$*$") indicate choices for P_{phot} at the relevant T_{eff}, and lines indicate run of T with P inside the photosphere (solid lines indicate radiative regions, dashed lines indicate convective regions). **Heavy symbols and lines indicate the correct models.**

from the previous models according to Eq. (7.6) and then for the next model uses

$$X_i = X_{i,\mathrm{prev}} + \Delta t \left(\frac{\mathrm{d}X_i}{\mathrm{d}t} \right)_{\mathrm{prev}} . \tag{7.15}$$

Note that it is also possible to calculate the chemical evolution using an implicit scheme. For a more detailed but quite readable discussion, see Eggleton (1971, MNRAS 151, 351). In the same reference, another choice of independent grid variable is discussed, which allows one to regrid the model automatically so that fine grid spacing is used where required.
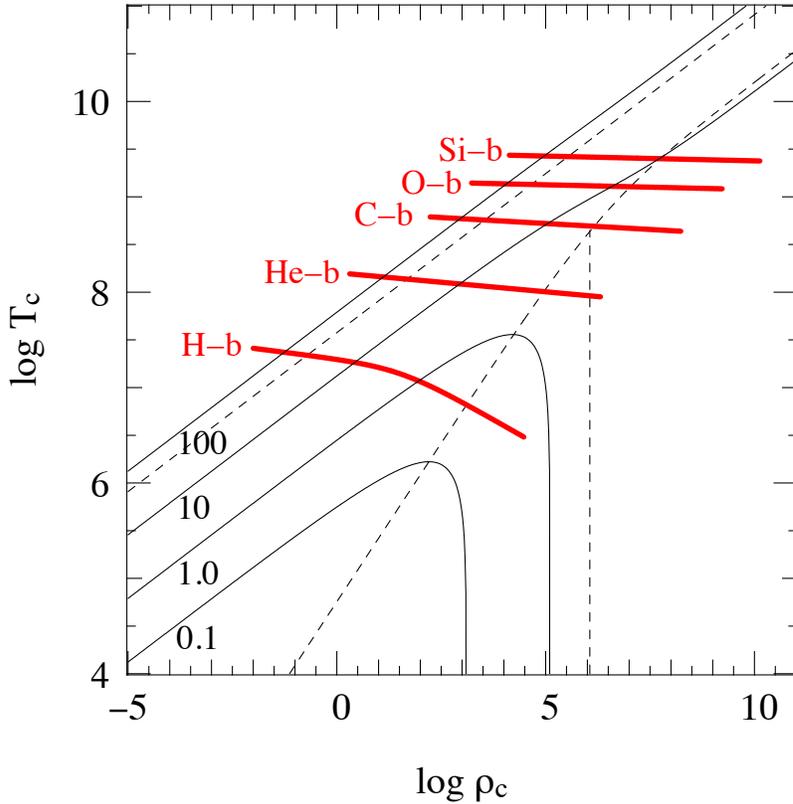
Figure 8.1: Equation of state for a composition of $X = 0.7$ and $Z = 0.02$, with dashed lines indicating regions of different physics (see Fig. 2.3 for labels). Overplotted in solid lines are the central temperature - central density tracks for contracting stars of $0.1 - 100 M_\odot$, assuming simple homologous contraction (i.e., ignore changes in density concentration). The lines are approximately $T_c \propto \rho_c^{1/3}$, in regions where ideal gas matters, but turn vertical when electron degeneracy comes into play. The red curves indicate approximate regions where nuclear fusion (of the indicated kind) are ignited. This figure roughly explains why there are "brown dwarfs" and why lower mass stars do not burn all the way to Fe. From Pols lecture notes.

8   Main-sequence

*Readings: CO §10.6, 13.1 (p. 446-451),*Back

8.1   Pre-main-sequence contraction

Before nuclear burning is initiated, a star contracts in Kelvin-Helmholtz timescales where its surface temperature remains roughly constant, while its luminosity decreases as its radius shrinks (Hayashi track, §5.2). As hydrostatic equilibrium is maintained throughout the contraction, Virial theorem says that half of the gravitational energy gain that gets turn into heat has to be lost to space.

We study how the central temperature changes as a given star contracts. As central pressure scales as M^2/R^4, and central density as $\bar\rho \propto M/R^3$ (§3). If the gas remains ideal gas,

$$T_c \propto \frac{P_c}{\rho_c} \propto \frac{M}{R} \propto \rho_c^{1/3}, \tag{8.1}$$

and the star's center heats up as it contracts. This may eventually lead to ignition temperature. However, if the central gas becomes degenerate, using $P \propto \rho^{5/3}$ (for NRCD), we get instead a central density $\rho_c \propto M^2$, one that is independent of temperature. This explains the vertical tracks in Fig. 8.1. The maximum density a star can be compressed to depends only on its mass.

– Brown dwarfs

So for a given stellar mass, there may be a maximum temperature contraction can heat it up. Insert the above degenerate central density into eq. (8.1), we find $T_{c,max} \propto M^{4/3}$. If the center enters into degeneracy at a temperature lower than that needed to ignite fusion, it is out of luck. Keeping count of all physical constants, and set the ignition temperature to be 10^6 K, we find nuclear fusion fails for stars with $M \leq 0.1 M_\odot$. A more careful calculation gives $M \leq 0.08 M_\odot$. Stars below this mass are the "brown-dwarfs".
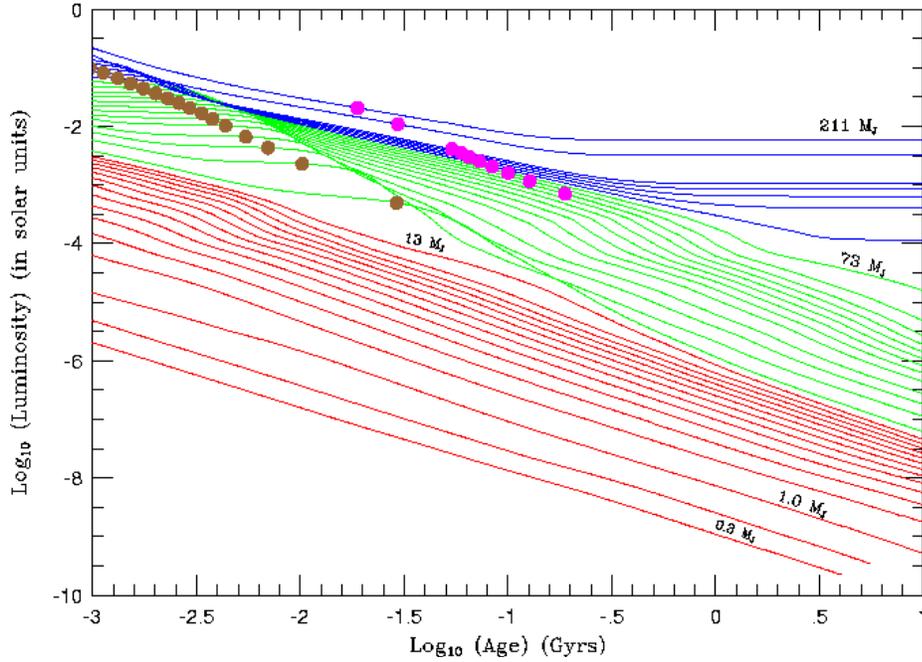
Figure 8.2: Luminosity as a function of time for very low mass stars (*solid lines*) and brown dwarfs (*dashed lines*). The horizontal plateaus in the tracks at upper left show where the period of deuterium burning halts the pre–main-sequence luminosity decline (for a period of up to a few million years) in very low mass stars, as well as in brown dwarfs. Models of mass $< 0.013\ M_\odot$ (i.e., less than about 13 Jupiter masses) have been designated as "planets" (*dot-dashed lines*) in this figure – they are too low in mass to burn Deuterium. The luminosities of brown dwarfs and planets dim with time, while those of main-sequence stars flatten.

Brown dwarfs go through a brief Deuterium-burning stage – see the green lines in Fig. (8.2). A reasonable number of them have been studied, and new spectral classes (e.g., L, T and Y) have been defined to distinguish them via features in their infrared spectra.

– massive end

Stars more massive than $\sim 100 M_\odot$ have problems with stability and mass-loss. The most massive star known to exist has $M \approx 250 M_\odot$ (R136a1).

8.2 Zero-age main sequence

The zero-age main sequence (ZAMS) is defined as the beginning of the long, stable period of core hydrogen burning during the star's lifetime. Stars burn up their (primordial) deuterium via $^2\mathrm{D} + \mathrm{p} \rightarrow {}^3\mathrm{He} + \gamma$ before this point, while they are still contracting towards the main sequence (see Fig. 8.2). Also, the *initial* carbon abundance in stars is much larger than the CN-cycle equilibrium value. For stars of solar metallicity of mass $\geq 1\ M_\odot$, the reactions that convert $^{12}\mathrm{C}$ to $^{14}\mathrm{N}$ (part of the CN-cycle) can supply the star's total luminosity for a brief period at the start of hydrogen-burning. This stage is so short that it is often ignored — e.g., it is not shown in the evolutionary tracks of Fig. 8.4 below. In the pre–main-sequence evolutionary tracks of Fig. 5.2, this $^{12}\mathrm{C} \rightarrow {}^{14}\mathrm{N}$ stage causes the last, small upwards-and-downwards wiggle at the end (at left).

Zero-age main sequence luminosity

For a crude estimate of the luminosity[3], we use the energy transport equation in terms of mass (Eq. 7.4), and apply it at $T \simeq \frac{1}{2}T_\mathrm{c}$, where we assume $L_r \simeq L$ [why?], $r \simeq \frac{1}{4}R$, (see Fig. 3.1 for $n = 3$ and also CO, Fig. 11.4), and take some appropriately averaged opacity $\overline{\kappa}$. Furthermore, we approximate $\mathrm{d}T/\mathrm{d}M_r \simeq T_\mathrm{c}/M$. Thus,

$$\frac{T_\mathrm{c}}{M} \simeq \frac{3}{64\pi^2 ac} \frac{\overline{\kappa}L}{(\frac{1}{4}R)^4(\frac{1}{2}T_\mathrm{c})^3} \simeq \frac{96}{\pi^2 ac} \frac{\overline{\kappa}L}{R^4 T_\mathrm{c}^3} \quad \Rightarrow \quad L \simeq \frac{\pi^2 ac}{96} \frac{R^4 T_\mathrm{c}^4}{\overline{\kappa}M}. \tag{8.2}$$

---

[3]See KW, chapter 20, for somewhat less crude approximations.
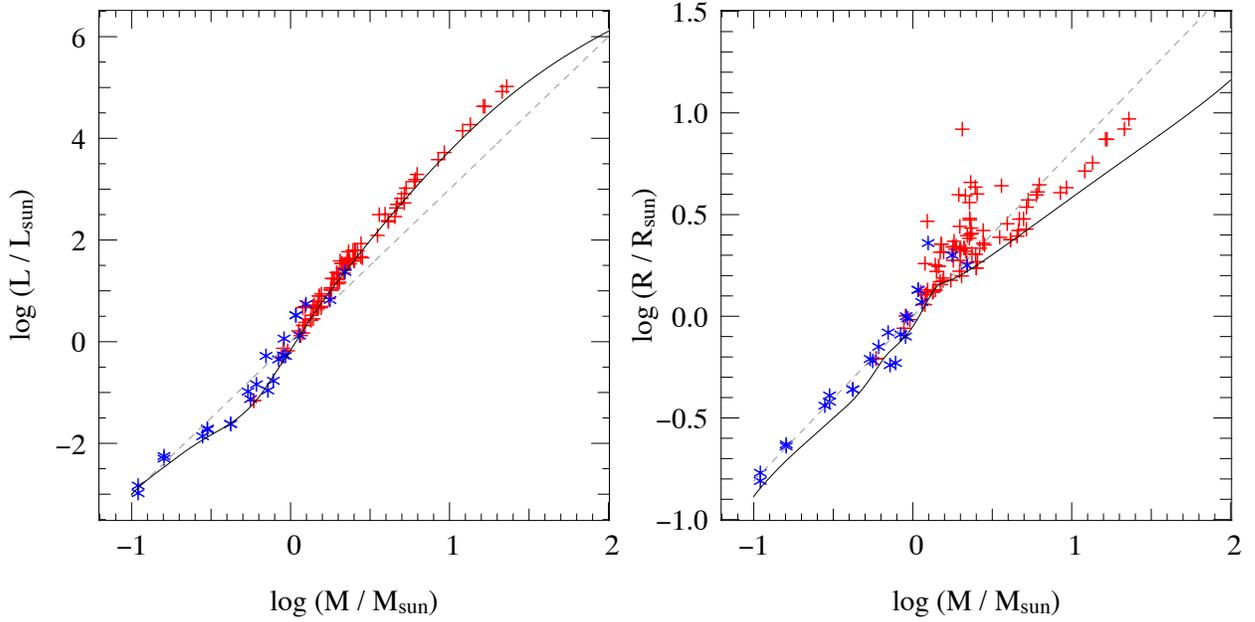
Figure 8.3: ZAMS mass-luminosity (left) and mass-radius (right) relations from detailed structure models with $X = 0.7, Z = 0.02$ (solid lines). The dashed curves indicate $L \propto M^3$ and $R \propto M^{0.81}$. Blue and red symbols indicate components of double-lined eclipsing binaries with accurately measured M, R and L, most of which are MS stars (but not necessarily ZAMS). From Pols lecture notes.

Expressing the central temperature in terms of the central pressure and density using the ideal gas law, and using the expressions for P_c and ρ_c appropriate for a polytrope with $n = 3$,

$$T_c = \frac{\mu m_{\mathrm{H}}}{k} \frac{P_{c,\mathrm{gas}}}{\rho_c} = 1.95\,10^7 \,\mathrm{K}\, \mu\beta \left(\frac{M}{M_\odot} \right) \left(\frac{R}{R_\odot} \right)^{-1}, \tag{8.3}$$

where β was defined as the ratio of the gas pressure and the total pressure. Inserting this in Eq. (8.2),

$$\frac{L}{L_\odot} \simeq 10 \, \frac{\mu^4 \beta^4}{\overline{\kappa}} \left(\frac{M}{M_\odot} \right)^3. \tag{8.4}$$

Hot zero-age main-sequence stars

For a hot star, electron scattering dominates in the interior. Thus, $\overline{\kappa} \simeq 0.2(1 + X)\,\mathrm{cm}^2\,\mathrm{g}^{-1}$ (Eq. 4.8). For a star with solar abundaces which has just arrived on the main sequence, $\mu \simeq 0.613$, and $L \simeq 4\,L_\odot\,(M/M_\odot)^3$. For intermediate-mass stars, this estimate agrees reasonably well with detailed models (see Fig. 8.3). The slope, however ($L \propto M^3$) is slightly too shallow between 2 and 8 $M_\odot$, where the detailed models give $L \propto M^{3.7}$; above 8 $M_\odot$ it is too steep. These effects are due to the presence of a central convection zone and the contribution of radiation pressure. The convection zone increases in size with increasing mass (see Fig. 8.5 and Table 8.1).

Cool zero-age main-sequence stars

For stars with $M \lesssim 1\,M_\odot$, the opacity is dominated by bound-free processes. Inserting the estimate Eq. (4.14) in Eq. (8.4), and using $\rho \simeq \frac{1}{8}\rho_c \simeq 7\,\overline{\rho}$ (for an $n = 3$ polytrope) as well as Eq. (8.3),

$$\frac{L}{L_\odot} \simeq 0.07 \, \frac{\mu^{7.5}}{Z(1 + X)} \left(\frac{M}{M_\odot} \right)^{5.5} \left(\frac{R}{R_\odot} \right)^{-0.5}. \tag{8.5}$$

Thus, given that R depends approximately linearly on M, we find a very steep mass-luminosity relation, much steeper than that observed or inferred from models. Furthermore, the luminosity of the Sun is underestimated ($L \simeq 0.05\,L_\odot$ for
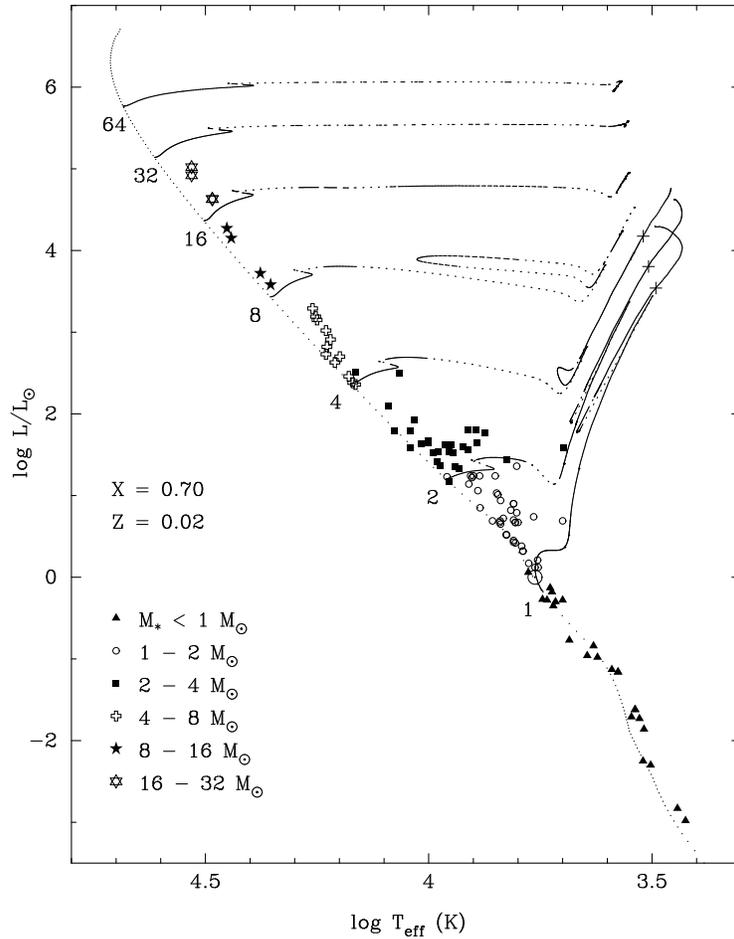
Figure 8.4: HRD for the ZAMS and several evolutionary tracks, calculated with the Eggleton evolutionary code. The labels are masses in solar units. The symbols indicate components of binaries for which the masses, radii, and luminosities were determined observationally. For the tracks, the solid, dotted, and dashed portions indicate where evolution is on a nuclear, thermal, and intermediate time scale, respectively (evolution is upwards and rightwards from the ZAMS; the brief initial $^{12}C \rightarrow ^{14}N$ stage is not shown). For masses $\geq 2\,M_\odot$, the end of the main sequence occurs at the first wiggle in the tracks, a bit to the right of the ZAMS. From Pols et al. (1995, MNRAS 274, 964).

$\mu = 0.613$, $Z = 0.02$, $X = 0.708$). The reason this does not work as well as for the massive stars, is that with decreasing mass, more and more of the outer region becomes convective; see Fig. 8.5. Only $\sim 2\%$ of the Sun's mass is convective (although this is nearly the outer $\sim 30\%$ of the Sun's radius), so a $n = 3$ polytrope is not completely unreasonable, but stars of $M \leq 0.2\,M_\odot$ are completely convective (so a $n = 1.5$ polytrope would be more appropriate). Furthermore, for very low masses, degeneracy becomes important.

Evolution on the main sequence

For both hot and cool stars, the luminosity scales with a high power of the mean molecular weight. As hydrogen is burnt, μ increases, and therefore the luminosity will increase as well, as can be seen in Fig. 8.4. Numbers for parameters at the beginning and end of the main sequence for massive stars are given in Table 8.1.

8.3 The end of the main sequence

Hydrogen exhaustion in the core

For more massive stars, hydrogen exhaustion will happen in a larger region at the same time, while for less massive stars, it will initially just be the centre itself. Since in the core one gets $L_r = 0$, also the temperature gradient $dT/dr = 0$, i.e., the core will become isothermal.

From our discussion of polytropes, it was clear that completely isothermal stars cannot exist ($\gamma = 1$ and $n = \infty$), but is
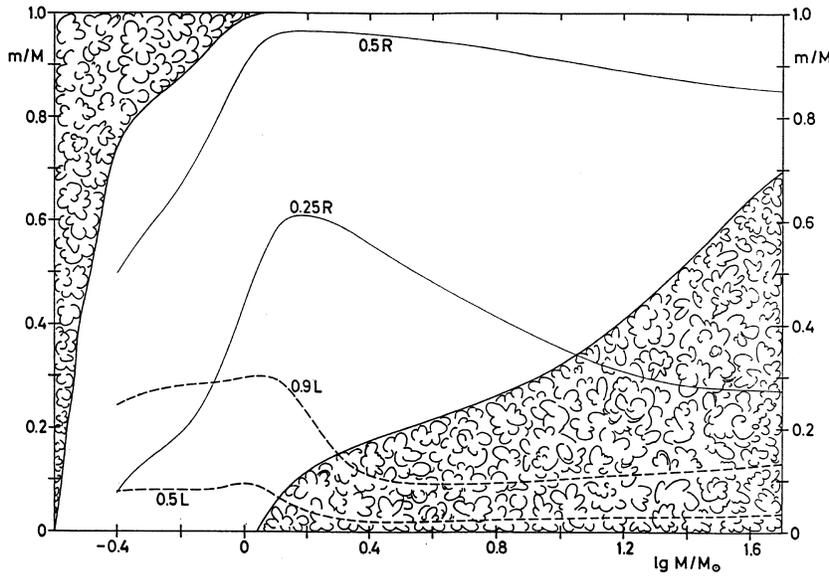
Figure 8.5: Mass fraction $m/M \equiv M_r/M$ as a function of stellar mass M at the ZAMS. Convective regions are shaded. Lower mass stars have progressively deeper surface convection zones – this is caused by the rise of opacity with temperature in the envelopes – with stars less massive than $0.35 M_\odot$ being fully convective. Stars more massive than $1.2 M_\odot$ have progressively larger convective cores – this is both because CNO burning has a steeper temperature dependence, and because radiation pressure in their interior reduces the adiabatic index γ to below $4/3$. The solid lines indicate the fractional masses at which $r/R = 0.25$ and 0.5, and the dashed ones those at which $L_r/L = 0.5$ and 0.9. The solar surface convection zone is about 1% in mass. Taken from KW (their Fig. 22.7).

Table 8.1: Fractional sizes of the convective core for main-sequence stars more massive than the Sun

	.... ZAMS				.... TAMS			
M_*	$\log L$	$\log T_{\text{eff}}$	M_{cc}	M_{cc}/M	t	M_*	M_{cc}	M_{cc}/M
$(M_\odot)$	$(L_\odot)$	(K)	$(M_\odot)$		(yr)	$(M_\odot)$	$(M_\odot)$	
120	6.254	4.739	102.4	0.853	$2.9\ 10^6$	80.9	63.6	0.786
60	5.731	4.693	46.3	0.772	$3.7\ 10^6$	43.0	27.5	0.640
20	4.643	4.552	10.8	0.540	$8.8\ 10^6$	19.1	6.5	0.339
5	2.720	4.244	1.52	0.304	$9.9\ 10^7$	5	0.39	0.078
2	1.177	3.952	0.46	0.229	$1.7\ 10^9$	2	0.13	0.065
1	-0.207	3.732	0	0	$9.7\ 10^9$	1	0	0

it possible to have an isothermal core that is embedded by an external medium? In the context of polytropes, one could rephrase this as the requirement that averaged over the whole star one has $\overline{\gamma} > 1.2$ ($\overline{n} < 5$). The result is that for a star in hydrostatic equilibrium, only a relatively small fraction of its mass can be in an isothermal core.

In fact, such a configuration, an isothermal sphere with an imposed positive pressure boundary, appears in star formation (Bonnor-Ebert sphere), in globular cluster (King's model), ... and here.

Schönberg-Chandrasekhar limit

As Hydrogen is progressively fused from the inside out, the core grows in mass. However, there is a maximum mass for a non-fusing, and therefore isothermal, core, first estimated by Schönberg & Chandrasekar at 1942.

For the isothermal core, one can rederive the virial theorem for the case that the pressure external to the object under consideration is not equal to zero. One finds (CO p. 453-455)

$$2K_{\text{core}} = -U_{\text{core}} + 4\pi R_{\text{core}}^3 P_{\text{core}}, \tag{8.6}$$

where P_{core} is the pressure at the outer boundary of the core.

For an isothermal core (and ideal gas), the thermal energy is simply $K_{\text{core}} = \frac{3}{2} N_{\text{core}} k T_{\text{core}}$, with $N_{\text{core}} = M_{\text{core}}/m_{\text{H}}\mu_{\text{core}}$

the number of particles in the core. The self-gravitational energy $U_{\rm core} = -q_{\rm core}GM_{\rm core}^2/R_{\rm core}$ ($q_{\rm core} = 3/5$ for uniform density sphere) and solving for $P_{\rm core}$, one finds,

$$P_{\rm core} = \frac{3}{4\pi}\frac{kT_{\rm core}}{m_{\rm H}\mu_{\rm core}}\frac{M_{\rm core}}{R_{\rm core}^3} - \frac{q_{\rm core}}{4\pi}\frac{GM_{\rm core}^2}{R_{\rm core}^4}. \tag{8.7}$$

Thus, the difference between thermal pressure ($\sim \overline{\rho}T_{\rm core}$) and self-gravity ($\sim \overline{\rho}/R_{\rm core}$) is balanced by external pressure confinement. The interesting part is that this pressure has a maximum value somewhere. As a growing $R_{\rm core}$, it first rises with $R_{\rm core}$ as the self-gravity weakens; it then drops with $R_{\rm core}$ as the necessary boundary term (PR^3) to balance thermal and the self-gravity diminishs. The maximum pressure can be determined by taking the derivative of Eq. (8.7) with respect to radius[4], and setting it equal to zero. One finds,

$$R_{\rm core} = \frac{4}{9}q_{\rm core}GM_{\rm core}\frac{m_{\rm H}\mu_{\rm core}}{kT_{\rm core}} \quad \Rightarrow \quad P_{\rm core,max} = \frac{3}{16\pi}\left(\frac{9}{4}\right)^3\left(\frac{kT_{\rm core}}{m_{\rm H}\mu_{\rm core}}\right)^4\frac{1}{q_{\rm core}^3 G^3 M_{\rm core}^2}. \tag{8.8}$$

The pressure at the core boundary has to be balanced by that exerted by the envelope, which can be estimated as $P \approx GM^2/R^4$, $\rho \approx M/R^3$, and, since also $P_{\rm env} = kT_{\rm env}\rho_{\rm env}/m_{\rm H}\mu_{\rm env}$, $T_{\rm env} \approx (m_{\rm H}\mu_{\rm env}/k)(GM/R)$. Combining,

$$P_{\rm env} = C_{\rm env}\frac{1}{G^3M^2}\left(\frac{kT_{\rm env}}{m_{\rm H}\mu_{\rm env}}\right)^4, \tag{8.9}$$

where $C_{\rm env}$ is a constant depending on the precise structure of the envelope. At the boundary, $T_{\rm env} = T_{\rm core}$.

So for a given core mass, the core can provide a matching pressure by adjusting its radius, as long as $P_{\rm env} < P_{\rm core,max}$, or,

$$C_{\rm env}\frac{1}{G^3M^2}\left(\frac{kT_{\rm core}}{m_{\rm H}\mu_{\rm env}}\right)^4 < C_{\rm core}\left(\frac{kT_{\rm core}}{m_{\rm H}\mu_{\rm core}}\right)^4\frac{1}{G^3M_{\rm core}^2}. \tag{8.10}$$

Inserting numerical values of $C_{\rm core}$ and $C_{\rm env}$ obtained from more detailed studies, one finds

$$\frac{M_{\rm core}}{M} \leq 0.37\left(\frac{\mu_{\rm env}}{\mu_{\rm core}}\right)^2, \tag{8.11}$$

For a helium core ($\mu_{\rm core} \simeq \frac{4}{3}$) and an envelope with roughly solar abundances[5] ($\mu_{\rm env} \simeq 0.6$), one thus finds a limiting fractional mass $M_{\rm SC} \simeq 0.08M$. Isothermal cores above this mass can no longer sustain the gravity of the envelope and must collapse. This yields heat.

As a function of mass

With the above, we can describe what will happen when hydrogen is exhausted in the core,

- For massive stars ($M \geq 6\,M_\odot$), the convective core at hydrogen exhaustion exceeds 8% of the total mass (see Table 8.1). Thus, an isothermal core cannot form. Instead, the core will contract until helium fusion starts. This happens on a thermal timescale, and causes the star to become a red giant (see next chapter).

- For intermediate-mass stars ($1.4 \leq M \leq 6\,M_\odot$), an isothermal core will form once hydrogen is exhausted in the centre. Around this core, hydrogen burning will continue, leading to growth of the core. This phase of the evolution is called the *sub-giant* branch. It will continue until the mass of the core exceeds 8% of the total mass, at which time the core has to contract, and the star becomes a red giant on the thermal timescale. Since this is much shorter than nuclear timescale, this leads to an observational gap in distribution of stars in HRD (the Hertzsprung gap).

- For low-mass stars ($M \leq 1.4\,M_\odot$), the isothermal core becomes degenerate before the critical mass fraction is reached (see Fig. 8.1), and no rapid phase of contraction occurs. Thus, the star moves to the red-giant branch on the nuclear time scale of the shell around the core.

---

[4]In CO, p. 455, the derivative is taken with respect to mass. This is rather illogical.
[5]In general, some processed material will be present in the envelope as well.
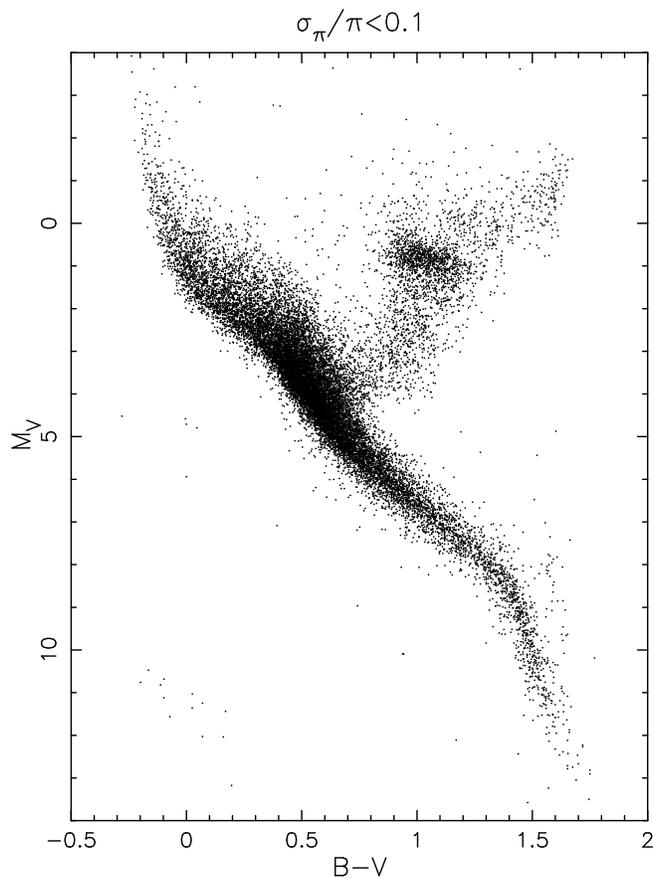
$\sigma_\pi/\pi < 0.1$

Figure 8.6: The HRD of nearby stars, with colours and distances measured by the Hipparcos satellite. The vertical axis is absolute V-magnitude. These stars have a variety of ages and metallicities. The main-sequence stretches from bottom right to top left; the red giant branch sits to the top right; and the red-clumps, at about $M_v = 0$ and in the middle of the red giant branch. These last group are actually metal rich stars undergoing Helium core fusion. The smattering of points to the bottom right are the white dwarfs. The Hertzbrung gap is the empty space, above $M_v \approx 2$, between the main-sequence and the red giants – stars more massive than $2M_\odot$ go through this stage in thermal, not nuclear, timescales. In contrast, stars less massive go through it gradually, producing the so-called 'subgiant' branch.

▲ Cool 1997
⋏ King et al. 1998
□ adapted from Kaluzny 96

$(m-M)_0 = 12.05$ E(B−V)=0.18
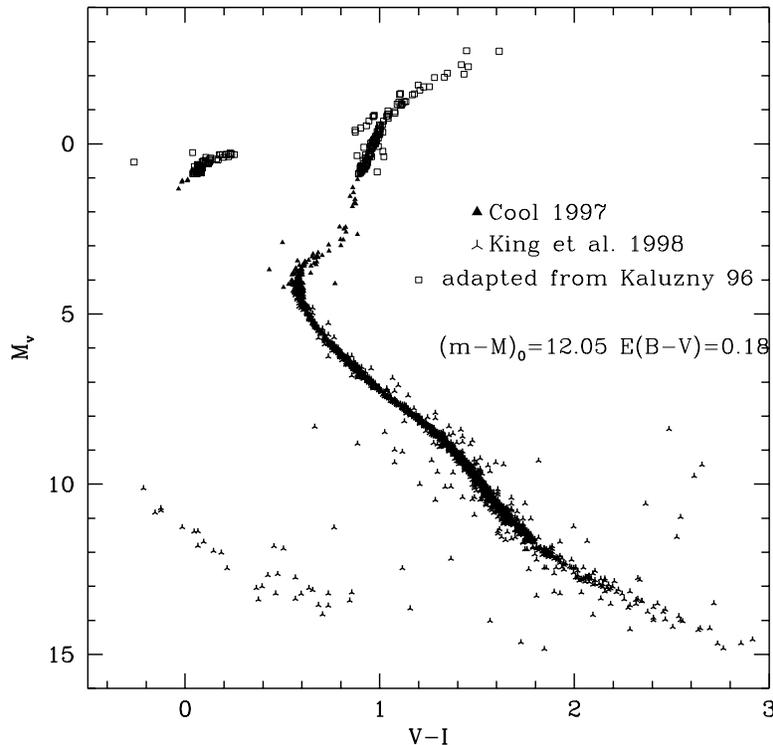
Figure 8.7: Observed HRD of stars in the globular cluster NGC 6397, at a distance of $\sim$ 2kpc. All stars have the same age, ~ 13.4 Gyrs and are very metal poor ([Fe/H] = -1.74). The stars undergoing Helium core fusion (the equivalent of 'red clumps' in the metal-rich case) sit to the top left, and are called 'Horizontal branch'. In this cluster, one observes a real bottom for the main-sequence, occupied by the lowest-mass stars ($M \sim 0.08 M_\odot$). All stars more massive than $0.8 M_\odot$ have left the main-sequence. At the tip of the remaining main-sequence, there is an conspicuous turn of stars to the right – they are entering the 'subgiant' phase. This feature is called the **main-sequence 'turn-off'** and is useful for dating clusters. Taken from D'Antona (1999, in "The Galactic Halo: from Globular Clusters to Field Stars", 35th Liege Int. Astroph. Colloquium).

Food for thought

- What determines the lowest mass of a main-sequence star?

- A number of features in Fig. 8.2 have been discussed before. These include separation between main-sequence stars and brown dwarfs; separation between brown-dwarfs and planets; the Hayashi phase; deuterium burning (the plateau in the green group of curves); the (power-law) cooling of giant planets (the red group of curves). It's very information-dense.

- A lot of attention has also been put on Fig. 8.3. Lots of stellar physics are condensed in the shape for the various broken power-laws.

- Why are more massive stars $(M > 1.2M_\odot)$ convective in the centre? why are less massive stars $(M < 1.2M_\odot)$ convective in envelopes?

- Ensure you understand why stars of different mass behave differently when Hydrogen is exhausted in their cores.
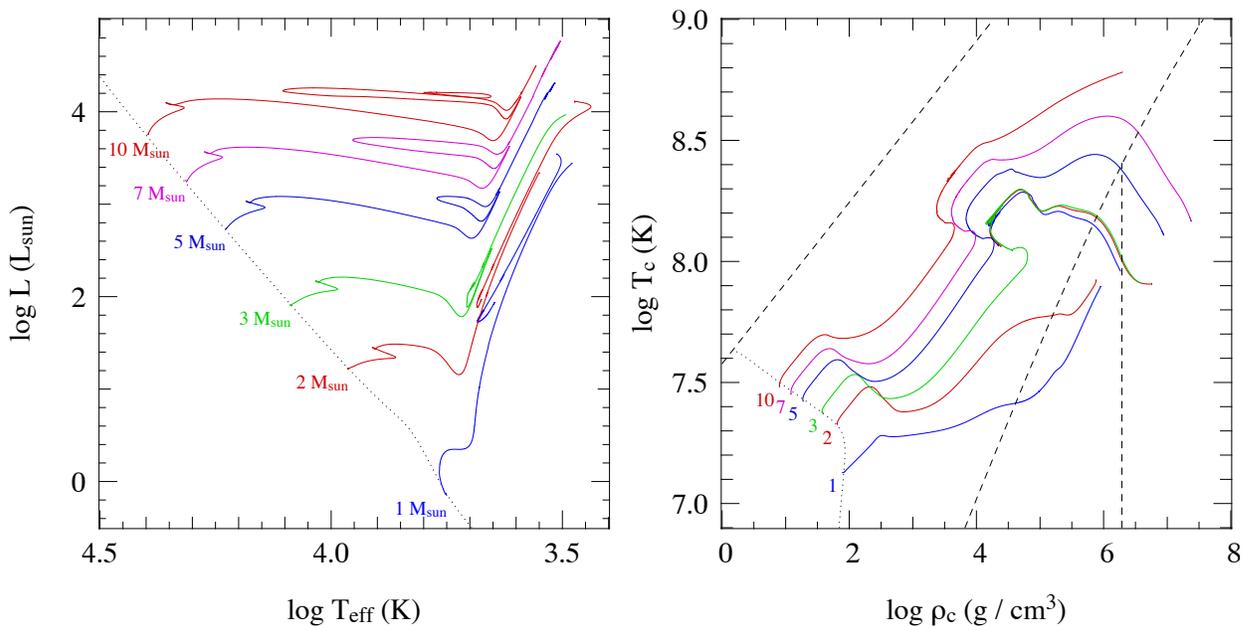
Figure 9.1: Evolution tracks for stars of quasi-solar composition ($X = 0.7, Z = 0.02$) and masses of $1, 2, 3, 5, 7$ and $10 M_\odot$ in the H-R diagram (left panel) and in the central temperature versus density plane (right panel). Dotted lines in both diagrams show the ZAMS, while the dashed lines in the right-hand diagram show the borderlines between equation-of-state regions and is to be compared against the simple model in Fig. (8.1). The $1 M_\odot$ model is characteristic of low-mass stars: the central core becomes degenerate soon after leaving the main sequence and helium is ignited in an unstable flash at the top of the red giant branch. When the degeneracy is eventually lifted, He burning becomes stable and the star moves to the zero-age horizontal branch in the HRD, at $\log L \approx 1.8$. The $2 M_\odot$ model is a borderline case that just undergoes a He flash. The He flash itself is not computed in these models, hence a gap appears in the tracks. The $5 M_\odot$ model is representative of intermediate-mass stars, undergoing quiet He ignition and He burning in a loop in the HRD. The appearance of the 7 and $10 M_\odot$ models in the HRD is qualitatively similar. However, at the end of its evolution the $10 M_\odot$ star undergoes carbon burning in the centre, while the cores of lower-mass stars become strongly degenerate. From Pols lecture notes

9 Post-main-sequence: Giants

Readings: This supplements (and partly replaces) CO §13.2, §15.1 on the fate of high mass stars, and §15.3 on core collapse (up to p. 534). Back

Stellar evolution past main-sequence is relatively complicated. There are a cornucopia of important concepts in this section: **subgiants**, **main-sequence turn-off**, **red giants**, **dredge**-up, **helium core flashes**, **red clump**, **horizontal branch**, **RR Lyrae variables**, **helium shell flash**, **asymptotic giant branch**, **super-wind**, **planetary nebula**, **Cepheid variables**, **Humphreys**-Davidson limit, **luminous blue variables**, **onion**-shells... the goal is to be able to point them out on the HR diagram and explain them quanlitatively.

General considerations

From observations, we see that stars which have left the main sequence, cluster predominantly near low temperatures, but high luminosity. Thus, their radii are large, i.e., they are giants. From observations of globular clusters, one finds that even low-mass stars can become extremely luminous in this phase (see Fig. 8.7). The two basic questions to be addressed are why stars become so cool, and how they can become so much more luminous than they were on the main sequence. Both properties are reproduced in stellar models, but it is not always simple to rationalize why a star behaves as it does. Indeed, even in the 1990s there have been a number of articles with titles like "why stars inflate to and deflate from red giant dimensions" (Renzini et al, 1992) and, in response, "on why intermediate-mass stars become giants after

the exhaustion of hydrogen in their cores" (Iben, 1993). Out of necessity, therefore, the discussion in this chapter will be somewhat vague. To give a framework, schematic evolutionary tracks for a low-mass ($1\,M_\odot$), an intermediate-mass ($5\,M_\odot$), and a high-mass ($25\,M_\odot$) star are shown in Fig. 9.2. One sees that intermediate-mass stars go through the most contorted track.

Regarding the question of the increase in luminosity seen for giants, it is worthwhile to think back to what determines the luminosity on the main sequence. There, in essence, the luminosity is determined by how quickly the envelope can transfer and radiate energy; the star will contract until nuclear fusion generates a matching amount of energy in the core. The principal difference for a giant is that the burning occurs in a shell, whose properties are not just determined by the envelope above, but also by the core below.

As a star's envelope expands and it becomes a red giant (approaching the Hayashi track), the convective envelope eventually comprises most of the tenuous envelope. In other words, the base of the convective envelope moves inwards in mass M_r (though not necessarily in radius), reaching into regions that had been partially processed by nuclear burning on the main sequence (CNO-cycle and p-p chain reactions). This processed material is thus mixed throughout the convective envelope to become visible at the star's surface, a process known as **"first dredge-up"**. This yields reduced $^{12}\mathrm{C}/^{13}\mathrm{C}$ and C/N ratios at the stellar surface; observations of these ratios in stars during this stage of evolution are in fairly good agreement with what is predicted by computational models.

9.1   Low mass giants

For low-mass stars, the contraction of the core after hydrogen exhaustion is stopped by electron degeneracy pressure before the core becomes hot enough for helium ignition. Therefore, the Schönberg-Chandrasekhar limit becomes irrelevant, and the core can grow until something more drastic happens. Since no energy is generated within the core, the temperature in the whole core will equilibrate with that in the surrounding hydrogen-burning shell.

Shell burning around a degenerate core

In the layers near the dense, concentrated core, the pressure structure is dominated by the strong gravitational attraction of the core rather than by the pressure of the overlying envelope. The core becomes more and more dominant as the star evolves, since the core grows in mass and shrinks in size, while the envelope becomes more and more tenuous.

In the limit that the envelope can be considered weightless, and the shell contains a mass much smaller than that of the core (and provided also that the base of the convective envelope does not actually reach *into* the burning shell — as opposed to the so-called "hot bottom burning"), the properties of the shell depend only on the mass $M_{\rm core}$ and radius $R_{\rm core}$ of the core. This implies that the length scale in the shell will be set by $R_{\rm core}$, i.e., that if one compares models for different $(M_{\rm core}, R_{\rm core})$, the run of pressure, density, etc., with $r/R_{\rm core}$ will be very similar (the so-called 'homologous models', or 'self-similar models'). For instance, if in a given model, $P/P_{\rm core} = f(r/R_{\rm core})$, where $P_{\rm core}$ is the pressure at the bottom of the shell (i.e., the outer boundary of the core) and $f(r/R_{\rm core})$ a functional dependence on $r/R_{\rm core}$, one then expects that in another model $P'/P'_{\rm core} = f(r'/R'_{\rm core})$. This expectation is confirmed by real models. Refsdal & Weigert (1970) used such assumptions to derive the dependencies of $\rho(r/R_{\rm core})$, $T(r/R_{\rm core})$, $P(r/R_{\rm core})$, and $L_r(r/R_{\rm core})$ on $M_{\rm core}$ and $R_{\rm core}$ (see also KW, §32.2). They assumed, for the burning shell, the ideal gas law, an opacity law $\kappa = \kappa_0 P^a T^b$, and energy production $\epsilon = \epsilon_0 \rho^{\eta-1} T^\nu$ (via reactions with η reactants, where $\eta = 2$ except for the 3α reaction), and found

$$
\begin{array}{llllll}
\rho(r/R_{\rm core}) & \propto & M_{\rm core}^{\alpha_1} R_{\rm core}^{\alpha_2}, & \alpha_1 = -\frac{\nu-4+a+b}{\eta+1+a}, & \alpha_2 = \frac{\nu-6+a+b}{\eta+1+a}, \\[2mm]
T(r/R_{\rm core}) & \propto & M_{\rm core}^{\beta_1} R_{\rm core}^{\beta_2}, & \beta_1 = 1, & \beta_2 = -1, \\[2mm]
P(r/R_{\rm core}) & \propto & M_{\rm core}^{\gamma_1} R_{\rm core}^{\gamma_2}, & \gamma_1 = 1 - \frac{\nu-4+a+b}{\eta+1+a}, & \gamma_2 = -1 + \frac{\nu-6+a+b}{\eta+1+a}, \\[2mm]
L_r(r/R_{\rm core}) & \propto & M_{\rm core}^{\delta_1} R_{\rm core}^{\delta_2}, & \delta_1 = \nu - \eta\frac{\nu-4+a+b}{\eta+1+a}, & \delta_2 = -\nu + 3 + \eta\frac{\nu-6+a+b}{\eta+1+a}.
\end{array}
\tag{9.1}
$$

One sees that the temperature scaling with $M_{\rm core}/R_{\rm core}$ is independent of details (a, b, n, ν) of the energy generation process and the opacity law (indeed, the scaling follows directly from hydrostatic equilibrium and the ideal gas law). Thus, for a degenerate core with $R_{\rm core} \propto M_{\rm core}^{-1/3}$, one expects $T \propto M_{\rm core}^{4/3}$. The implied strong dependence of the luminosity
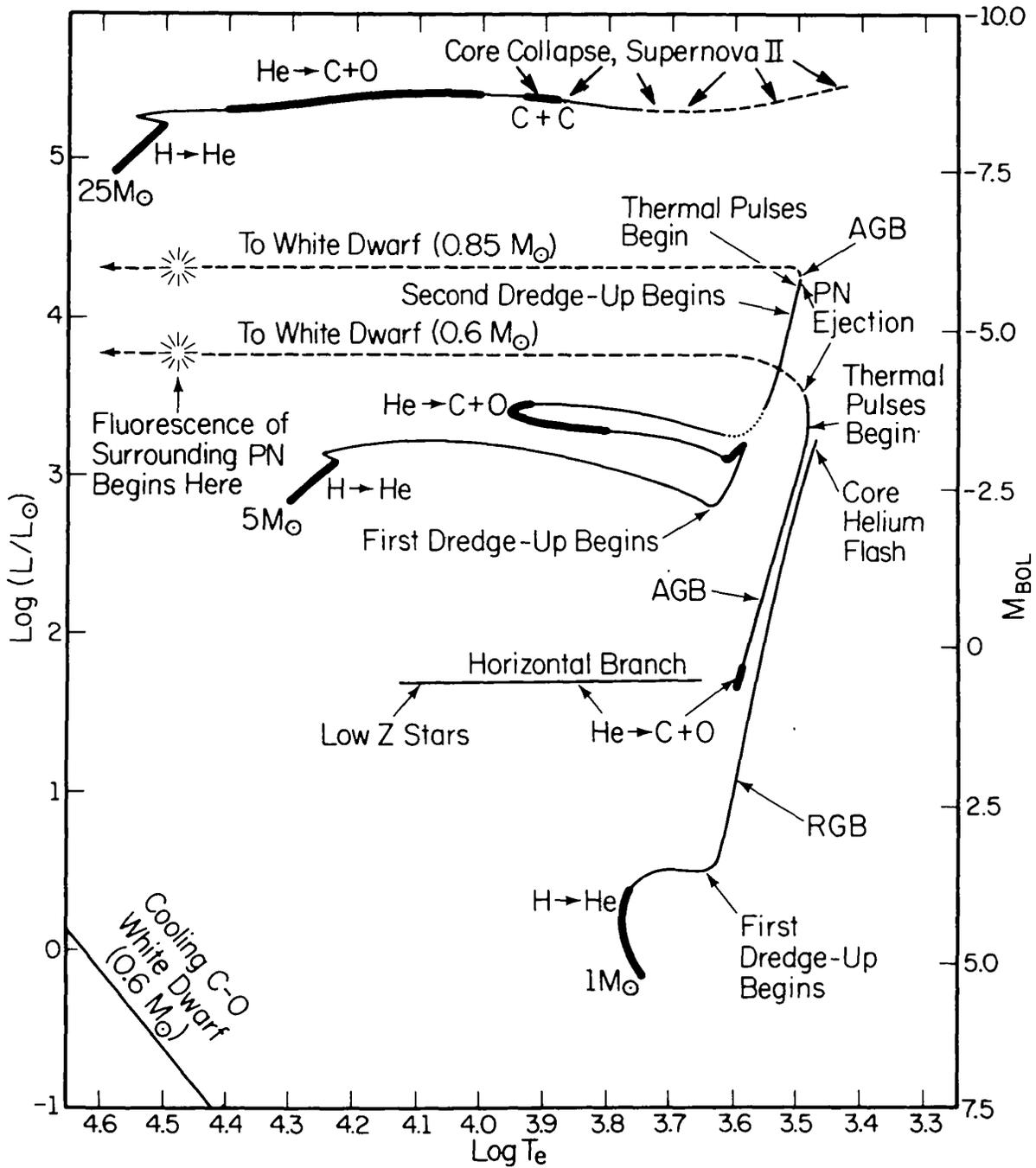
Figure 9.2: Schematic evolutionary tracks in the HR diagram for stars of low $(1M_\odot)$, intermediate $(5M_\odot)$, and high mass $(25M_\odot)$. Heavy portions indicate phases where the evolution proceeds on a slow, nuclear timescale. Indicated also are the first and second "dredge up," phases in which the outer convection zone reaches down to layers with processed material. A third dredge-up occurs during the thermal-pulse phase, which is also indicated. Note that the luminosity at which a star leaves the AGB is a conjecture based on observed white-dwarf masses. The helium cores of low-mass stars are supported by degeneracy pressure, so they have a relatively long-lived red giant branch phase, and have an unstable ignition (helium flash, $0.8 - 2M_\odot$). Intermediate-mass stars $(2 - 8M_\odot)$ develop a helium core that remains non-degenerate, and they ignite helium in a stable manner. After the central He burning phase they form a carbon-oxygen core that becomes degenerate. Both low-mass and intermediate-mass stars shed their envelopes by a strong stellar wind at the end of their evolution and their remnants are CO white dwarfs. Massive stars have masses $(> 8M_\odot)$ ignite carbon in a non-degenerate core. Except for a small mass range $(811M_\odot)$, these stars also ignite heavier elements in the core until an Fe core is formed which collapses. Taken from the seminal review by Iben (1991).
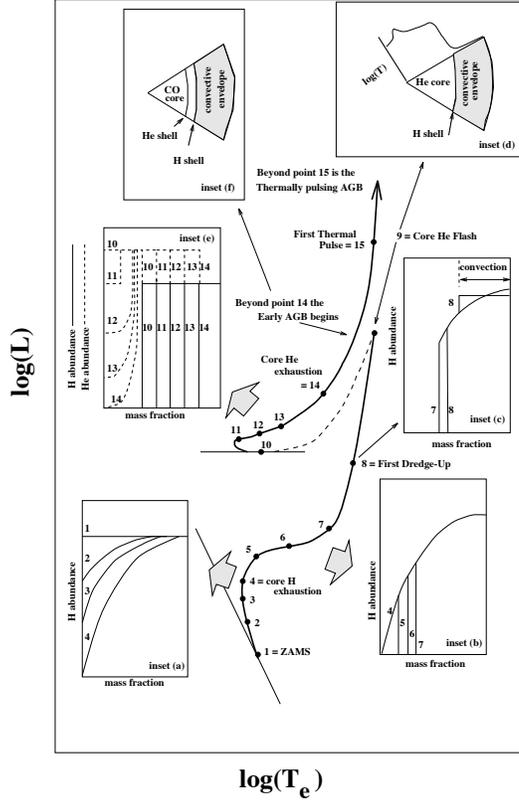
Figure 9.3: Schematic evolution of a $1\ M_\odot$ star. After core hydrogen exhaustion (step 4, at about 9 Gyrs), nuclear energy generation gradually moves out to a thick shell surrounding the isothermal helium core. The core grows in mass and contracts, while the envelope expands. By step 7, the helium core has become degenerate, and the envelope has cooled and become largely convective, with the star finding itself at the base of the red giant branch. From 4 to 7, the star remains in thermal equilibrium and the evolution lasts ~ 2 Gyrs. This long-lived phase is often called **'sub-giants'**. At step 9, the core has reached a mass of $\sim 0.45 M_\odot$ and core helium flash occurs. From Lattanzio & Boothroyd (1995).

Table 9.1: Dependencies of ρ, T, P, and L in a shell on M_{core} and R_{core}.

case	η	ν	α_1	α_2	β_1	β_2	γ_1	γ_2	δ_1	δ_2
CNO, hot	2	13	-3	2.33	1	-1	-2	1.33	7	-5.33
CNO, cool	2	16	-4	3.33	1	-1	-3	2.33	8	-6.33
triple-α	3	22	-4.5	4	1	-1	-3.5	3	8.5	-7

Taken from Refsdal & Weigert (1970) For all cases, it is assumed that electron scattering dominates the opacity (i.e.,
$$a = b = 0).$$

on the core mass is only partly offset by the fact that the pressure and density actually decrease with increasing M_{core}. Indeed, from numerical values (see Table 9.1), one sees that one has stellar luminosity $L \propto M_{\mathrm{core}}^{\sim 9}$ for a shell in which hydrogen is burned via the CNO cycle; this is confirmed by detailed models, which find $L \propto M_{\mathrm{core}}^{\sim 8}$ on the upper RGB (where the envelope is the most extended).

Thus, we see that the luminosity increases very steeply with increasing core mass. Since the envelope is almost completely convective, and the star is close to the Hayashi line, the effective temperature cannot increase much. In the HR diagram, the star therefore moves almost straight up, along the so-called *ascending* or **red giant branch** (RGB) – the hydrogen shell burning stage. As the hydrogen shell burns its way outwards in mass M_r, the convective envelope retreats ahead of it: deepest first dredge-up occurs not far above the base of the RGB (see Fig. 9.3).

On the upper RGB of low mass stars (*subsequent to* first dredge-up), there is evidence of some further CNO-cycle processing of envelope material, in spite of the fact that the base of the convective envelope is at temperatures far too low for such nuclear processing. This indicates that a slow "extra" mixing mechanism is at work (probably driven by rotation effects), mixing some material between the convective envelope and the hydrogen-burning shell. (This is a similar mechanism to that which causes the main-sequence lithium depletion in stars like the Sun.)

Evolution of the degenerate core

While the core grows from surrounding H-shell burning, it remains approximately isothermal, and at the temperature of the shell surrounding it. In principle, the increase in temperature goes towards lifting the electron degeneracy, but this is more than compensated for by the increase in core density, $\overline{\rho}_{\mathrm{core}} \propto M_{\mathrm{core}}/R_{\mathrm{core}}^3 \propto M_{\mathrm{core}}^2$; see Fig. 9.1.

As one increases the density and temperature, however, the helium ions (which are not degenerate) start approaching each other more and more closely during interactions, and will start to fuse when the core mass increases to $0.45\,M_{\odot}$ (and $T_{\mathrm{core}} = T_{\mathrm{shell}} \simeq 10^8\,\mathrm{K}$, eq. [9.1]). [Verify that you understand why this is independent of the total mass of the star.] The fusion will increase the temperature in the core, but will not reduce the density at first, since the pressure exerted by the ions is small compared to the electron degeneracy pressure. With increasing temperature and constant density, energy generation increases exponentially, until finally the thermal pressure becomes high enough to force the core to expand. By this time, the luminosity from the core has become $\sim 10^{11}\,L_{\odot}$, i.e., roughly equal to that from the entire Galaxy! Unfortunately, it does not seem possible to observe this **helium core flash**: the energy is used to expand the envelope. [Compare this scenario to a case where the core is not electron degenerate.]

From detailed models, it turns out that as the degenerate core grows hotter, in its centre the pressure and temperature are sufficiently high that energy is lost in neutrino creation. As a result, the centre will be slightly cooler, and helium core flash ignition will be in a shell around it. Burning will move inwards as the core is heated (possibly in a succession of mini-core-flashes following the main core flash), until degeneracy is lifted throughout the core.

After the helium core flash

The evolution *during* the helium flash is not very well understood, but it appears to be followed by a phase of quiet helium burning in a non-degenerate core. This core will still have $M_{\mathrm{core}} \simeq 0.45\,M_{\odot}$, but its radius will have increased significantly. Thus, one expects that the luminosity contributed by the hydrogen shell will be much smaller, $\sim 100\,L_{\odot}$ (down from $\sim 1000\,L_{\odot}$, see eq. [9.1]). During this time, the position of the star in the HR diagram depends on its metallicity, which determines the opacity in the envelope as well as the efficiency of energy generation in the CNO cycle (via X_{CNO}). For solar metallicity, stars remains near the Hayashi track, in the so-called **red clump** (see Fig. 8.6); for lower metallicities (like in globular clusters), stars will move to the **horizontal branch** (see Figs 8.7, 9.2, and 9.3) The position on the horizontal branch is determined by the envelope mass as well as the metallicity.

Mass loss of order $0.2\,M_{\odot}$ appears to take place between the main sequence and the horizontal branch; possibly there is a mass ejection episode due to the helium core flash (although pure stellar wind mass loss on the RGB has not been ruled out). Some such low mass stars traverse a region of the HR diagram where their outer envelopes are pulsationally unstable, becoming RR Lyrae variables.

After helium is exhausted in the core, the core, now composed of carbon and oxygen, will become degenerate, and burning will continue in a helium shell. This shell will become brighter as the core mass increases, and the star starts to move up the **asymptotic giant branch** (AGB). During the later phases, the burning in the helium shell becomes unstable, leading to so-called **helium shell flashes**. During this phase, the envelope mass is reduced by nuclear burning and mass loss. The latter becomes especially important at very high luminosities, when the envelope becomes pulsationally unstable (becoming, e.g., Mira variables, with large pulsation amplitudes). At that time, a so-called **"super-wind"** starts. Once the hydrogen-rich envelope has dwindled to $\leq 1\%$ of the total mass, it deflates, and the star moves towards the blue at essentially constant luminosity, burning what little material remains. (After the star has left the AGB, there is a period when its surface is hot enough to yield UV radiation that ionizes the material lost most recently, which is then visible as a glowing **"planetary nebula"** — a misnomer, since it has nothing to do with planets). The star will be left with roughly $10^{-2}\,M_{\odot}$ of helium and $10^{-4}\,M_{\odot}$ of hydrogen, around a carbon-oxygen white dwarf. From observations of white dwarfs, one finds masses mostly in the range 0.55–$0.60\,M_{\odot}$. Apparently, the remainder of the envelope mass of low-mass stars is lost in their latest stages.
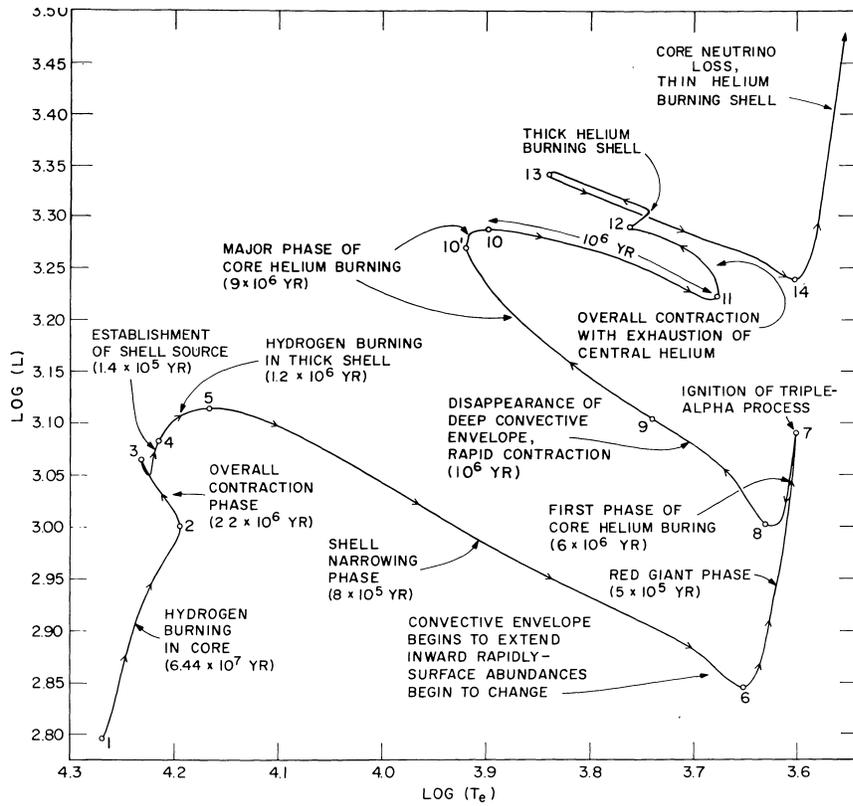
Figure 9.4: Evolutionary track of a 5 $M_\odot$ star in detail, with important physical processes for the different phases indicated (note that the extra loop from point 11 to point 14 is probably spurious, as it does not show up in more recent stellar models such as those in shown in Figs. 9.5 — this track was taken from a less accurate computation, performed several decades earlier). From the review by Iben (1991).
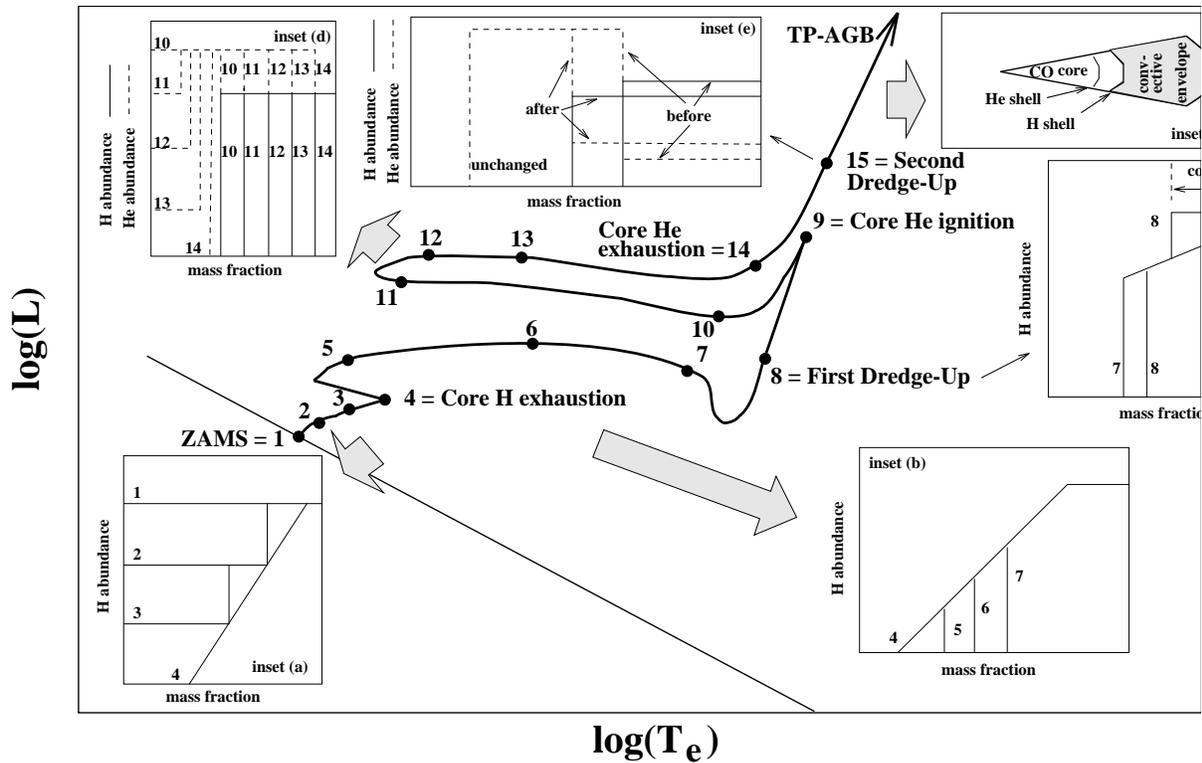

Figure 9.5: Schematic evolution of a 5 $M_\odot$ star, paying special attention to the abundance profiles at various stages. From Lattanzio & Boothroyd (1995).

9.2 Intermediate mass giants

For an intermediate-mass star, after hydrogen is exhausted in the core, burning continues in a thick shell around an isothermal core. This corresponds to the phase between points 4 and 5 in Fig. 9.4. When the isothermal core reaches the Schönberg-Chandrasekhar limit, the core collapses and the envelope expands, the star moves rapidly towards the red. During this phase, the surface luminosity drops, but this is mostly because part of the energy generated in the collapsing core is used for the expansion of the envelope (see below). The star stabilizes again when helium is ignited in the core, and the envelope has become largely convective (point 7 in Fig. 9.4, point 9 in Fig. 9.5 — *this is the point of deepest first dredge-up in intermediate mass stars*).

At this phase, the core (which initially has mass $\sim 0.75\,M_\odot$ in a $5\,M_\odot$ star) hardly notices that there is another $4\,M_\odot$ of shell and envelope around it, and its structure and luminosity are very similar to what they would have been if the core had been an isolated $0.75\,M_\odot$ helium main-sequence star. This reflects the fact that the envelope has become so dilute that it exerts negligible pressure. Like for the low-mass stars, the conditions in the hydrogen-burning shell depend almost completely on the properties of the helium-burning core.

When the helium core evolves, its "effective temperature" will at first, like that of a hydrogen main-sequence star, become slightly lower, and its radius will become slightly bigger. As a result, the hydrogen shell becomes less luminous. Since the shell produces most of the star's luminosity, the luminosity will drop somewhat (between 7 and 8 in Fig. 9.4). The mass of the helium core, however, will increase, and this causes the core to move upward in mass along the helium main-sequence, towards somewhat larger radius and higher temperature. The higher temperature causes an increase in the energy production in the shell, and therewith a rise in the star's luminosity. This corresponds to the increase in luminosity (between 9 and 10 in Fig. 9.4). During portions of these "blueward loops" in the HR diagram, intermediate mass stars may also lie in regions of the HR diagram where their outer envelopes are pulsationally unstable, becoming **Cepheid variables**.

When helium is exhausted in the centre, an isothermal carbon-oxygen core forms, and around it helium is burnt in a thick shell (10 to 11 in Fig. 9.4). When the core reaches the Schönberg-Chandrasekhar limit, it will collapse (note that the mass of the carbon-oxygen core should be measured relative to the mass of the helium star). As a result, the helium shell will become much more luminous, the layers above it will expand, and the hydrogen shell will be extinguished (13-14 in Fig. 9.4) This expansion causes the convective envelope to engulf hydrogen-exhausted material that the hydrogen shell had left behind, in a process known as **"second dredge-up"**. (near point 15 in Fig. 9.5. Low mass stars, where the hydrogen-burning shell is not extinguished, do not experience second dredge-up). The core becomes degenerate, and at first there is only a helium shell around it. As the shell eats outwards, it comes close to the position where second dredge-up has left hydrogen-rich material, and the hydrogen shell is re-ignited.

From here on, the evolution becomes similar to the late evolution of low-mass stars. The helium shell becomes unstable, and near the top of the asymptotic giant branch a super wind sets in, which limits the growth of the degenerate core. When the envelope has become too tenuous, it deflates, the star moves to the blue, and a white dwarf is formed.

9.3 High mass giants

For even more massive stars, after hydrogen exhaustion the core contracts immediately to helium ignition. This slows down, but does not stop the star from moving across the HR diagram. For the $25\,M_\odot$ star shown in Fig. 9.2, helium is exhausted while the star is only midway over to the red-giant branch. At that point, the core contracts further, and carbon is ignited. After that, things move on very fast, and the star soon explodes as a supernova.

The evolution of these massive stars is complicated greatly by mass loss, even on the main sequence. Due to mass loss, the whole hydrogen-rich envelope may disappear, in which case the star becomes a helium star, and moves to high temperatures in the HR diagram. Indeed, for very massive stars, this is virtually unavoidable, as their luminosity on the way to the red giant branch exceeds the Eddington luminosity, and their envelopes are rapidly blown off. This results in an empty region in the top right of the HR diagram, above and to the right of the **Humphreys**-**Davidson limit** (see Fig. 9.7). Stars close to this limit indeed are observed to have extremely large and variable mass-loss rates; these are the so-called **luminous blue variables**.
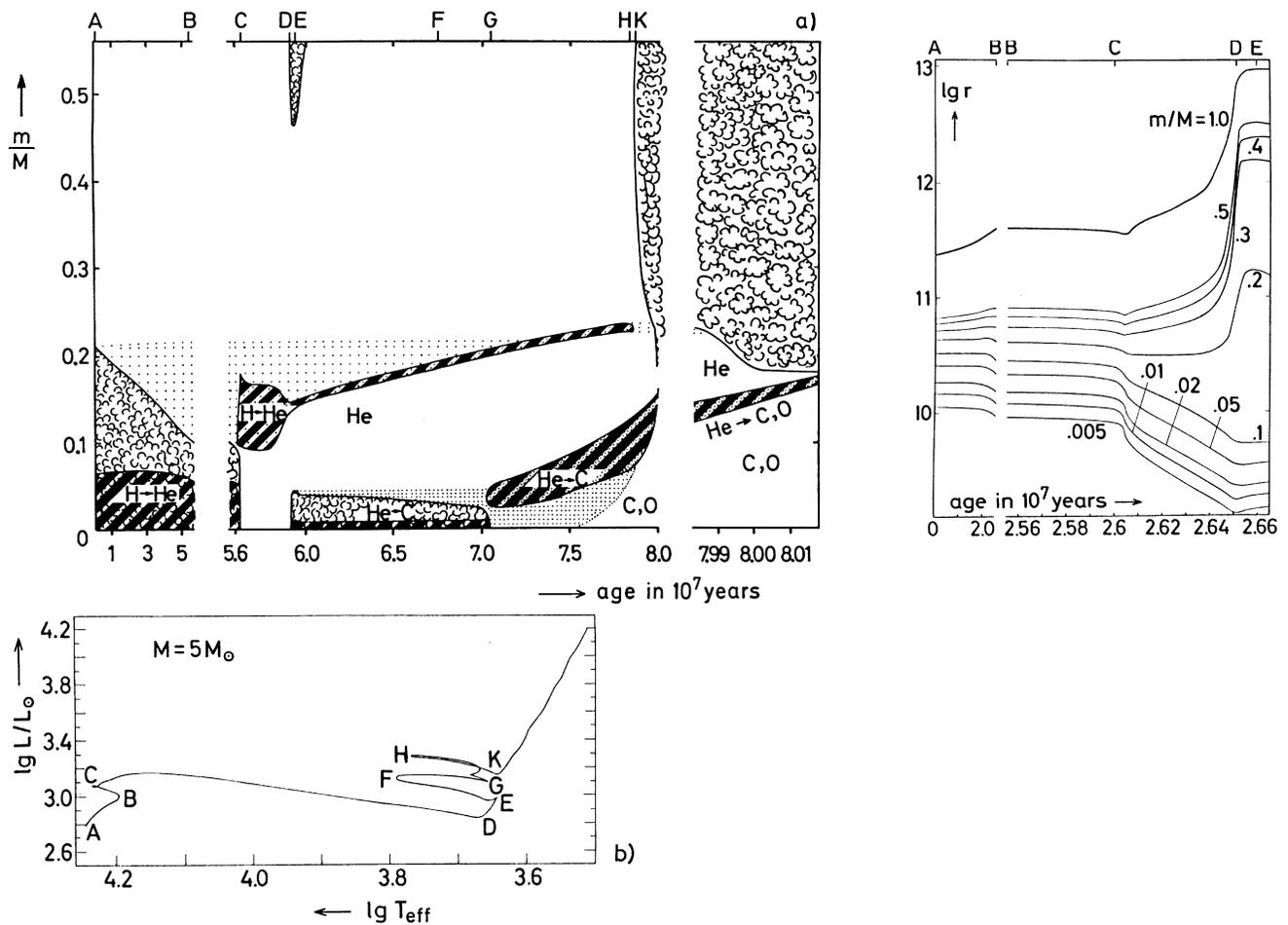
Figure 9.6: *(Upper left)* Interior structure (the famous 'Kippenhahn diagram' way of plotting) of a $5M_\odot$ star during its evolution. "Clouds" indicate convective regions, heavy shading energy generation at rates $\epsilon > 10^3 \, \mathrm{erg\,g^{-1}\,s^{-1}}$, and stippling variable chemical composition. *(Lower left)* Evolutionary track in the HRD for the same model. *(Right)* Radial variation of different mass shells during the evolution of a $7 \, M_\odot$ star. The letters A,...,E correspond to the same evolutionary phases labeled for a $5 \, M_\odot$ star in the left-hand panels. Taken from KW (Fig 31.2)

Why do stars become Giants?

In the above discussion, we have mostly ignored the envelope. This is not unreasonable if it is as tenuous as it has to be when the star has swollen to giant dimensions, but we have not yet addressed why this swelling actually happens. It is clear that real stars do it, and their behavior can be reproduced by models, but it is not so clear what physical mechanism dominates this process. Indeed, as was clear from the beginning of this chapter, this question is still debated (see the references quoted there for more detail).

Partly, it seems it is related to the way the opacity varies with density and opacity. Quite generally, as a star becomes more luminous, its radius increases and effective temperature decreases a little. This in itself is not enough to bring the star over to the red giant regime. As the temperature in the outer layers decreases, however, the opacity there increases quite strongly, since it is dominated by bound-free processes (the lower temperature leads to lower ionization states of the metals, which therefore can absorb photons more easily). Therefore, the luminosity cannot easily be transported anymore, and part of it is trapped, leading to further expansion. At some point, this apparently can become a runaway process, in which the envelope cools more and more, becomes more and more opaque, traps more and more of the luminosity, and expands to larger and larger radii. It only stops when the star reaches the Hayashi line, where the envelope has become almost completely convective, and energy can be transported more easily.
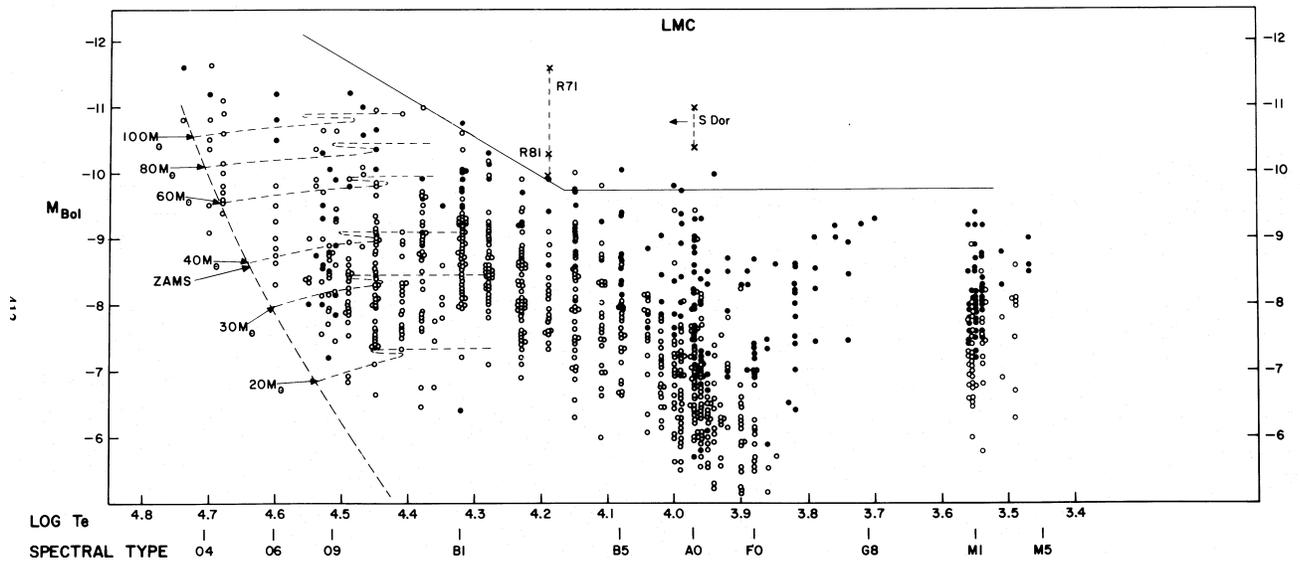
Figure 9.7: HRD of the brightest stars in the Large Megellanic Cloud (and therefore all at roughly the same distance), with observed spectral types and magnitudes transformed to temperatures and luminosities. Overdrawn is the empirical upper limit to the luminosity (solid lines), as well as a theoretical main sequence (dashed curve). Taken from Humphreys & Davidson (1979).

This runaway expansion may be responsible for intermediate-mass stars crossing the HR diagram very fast. Observationally, this results in a lack of stars between the main sequence and the giant branches, in the so-called *Hertzsprung gap* (Fig. 8.6). When the luminosity decreases, it appears the inverse instability can happen, where the envelope heats a little, becomes less opaque, therefore shrinks a little, releasing energy which increases the temperature, etc. This deflation instability might be responsible for the blue loops seen in the evolutionary tracks of intermediate-mass stars (see Fig. 9.4).

During stellar evolution, there appears to be a general rule of thumbs that in shell-burning sources, 'core contraction' leads to 'envelope expansion', and 'core expansion' leads to 'envelope contraction'. [You should now have enough physics to try to make some sense of this.]

Food for thought

- Explain Fig. 8.1 right panel, in particular the curve for $1M_\odot$ star – why does the central temperature/density first rise towards the right? after that why does the central temperature become constant while density keeps on increasing? after that why does the central temperature rise again? what evolutionary stages do these changes correspond to (e.g., point out in left panel of that figure)?

- Why are nuclear burning in degenerate cores always "flashes"?

- What is your understanding on why the Sun becomes a 'giant' after main-sequence?

- A Youtube video that shows both the global properties as well as interior profiles of a $3M_\odot$ star as it evolves, is both entertaining and informative.

- Why don't massive stars go through the same helium core flashes as low mass stars do?

- How does post-main-sequence evolution of different stars relate to the concept of Schonberg-Chandrasekhar limit?
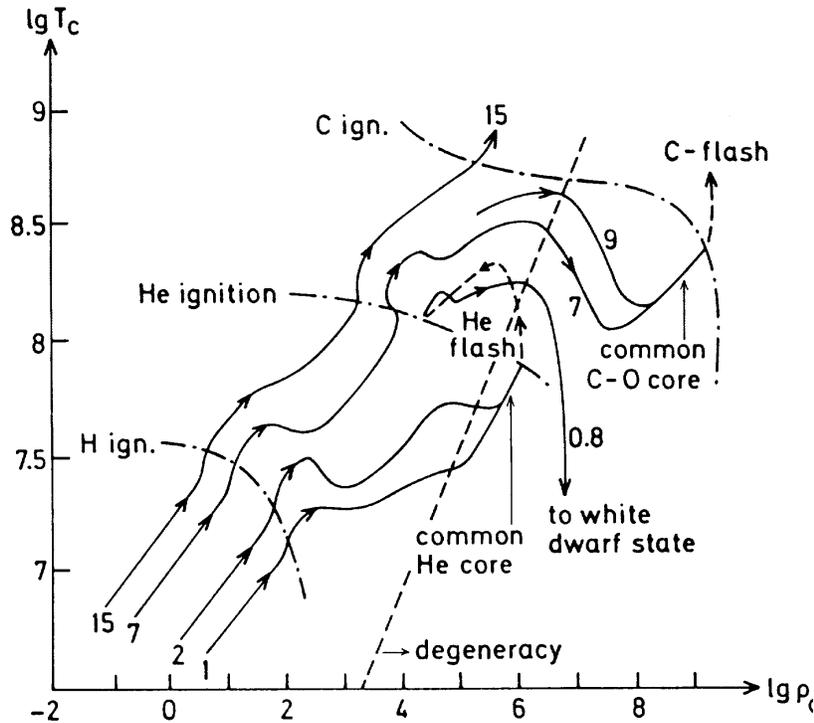
Figure 10.1: Variation of central density and temperature during the course of the evolution of stars of various masses. If the evolution is homologous and if the core is not degenerate, one has $T \propto \rho^{1/3}$, as indeed seen in the lower left corner. The long-dashed line indicates the approximate limit to the right of which the core becomes degenerate and has a slope of $T \propto \rho^{2/3}$. As a result, contraction leads the stellar core toward degeneracy. Dash-dotted lines indicate regimes where hydrogen, helium, and carbon are ignited. The fate toward degeneracy may be (temporarily) prevented if nuclear ignition occurs. (Note that these earlier models showed a $9\,M_\odot$ star's core becoming degenerate, rather than igniting carbon burning non-degenerately. Mass loss appears to terminate AGB evolution before the C-O core reaches the "C-flash" ignition line in this diagram.) In essence, the higher mass stars can move along evolution further because they are less dense. Taken from KW Fig. 33.6.

10 End of Star: Supernova & Degenerate Remnants

Readings: CO §*15.1, 15.3 (up to p. 534; different emphasis), CO* §*16.1, 16.2, 16.3, 16.4, 16.6 for degenerate remnants,*Back

Low mass star – Dwindling into oblivion

A very simple picture of the evolution of a star can be obtained by looking just at the changes in central density and temperature (Fig. 10.1). All stars more massive than $0.08 M_\odot$ can reach Hydrogen ignition. However, for a low-mass star, the core becomes degenerate before Helium is ignited. If the star does not have sufficient mass, the core cannot grow up to $0.45 M_\odot$, and Helium will not be ignited; the star will dwindle, and become a Helium white dwarf.[6]

A degenerate core that is massive enough undergoes Helium flash that lifts it temporarily out of degeneracy. However, the resultant carbon-oxygen core becomes degenerate again after Helium burning. In principle, if the star were massive enough, the core might grow sufficiently due to shell burning to ignite carbon burning. If so, the star would likely explode, leaving no remnant. In practice, however, it seems the super wind intervenes, and no carbon-oxygen cores above $\sim 1.2\,M_\odot$ are formed (Fig. 10.2). In fact, a rather large range of initial stellar masses are turned into white dwarfs of a rather narrow mass range (Fig. 10.2), and the most massive star to form a white dwarf seems to be $\sim 7 M_\odot$.

10.1 High mass stars – going out with a bang

For stars more massive than $\sim 8 M_\odot$, the carbon-oxygen cores do not become degenerate, and core Carbon burning is ignited non-explosively. The cores of these stars can continue to further burning stages.[7] The stages follow each other more and more rapidly, as neutrino losses become more and more important, while the energy gain from the fusion dwindles (Fig. 6.1). This, the author believes but have yet to check against numerical simulations, is due to the fact that lots of

---

[6]Such low mass stars have not yet had time to finish their main-sequence lives even if they were formed very early on — the universe has not lived long enough. In binaries, however, somewhat more massive stars can be "stripped" of their envelopes by mass transfer onto a binary companion while they are on the red giant branch, and these can indeed leave helium white dwarfs.

[7]It is *possible* that stars in a narrow mass range near $8\,M_\odot$ may proceed no farther than carbon burning, and end up as oxygen-neon-magnesium white dwarfs.
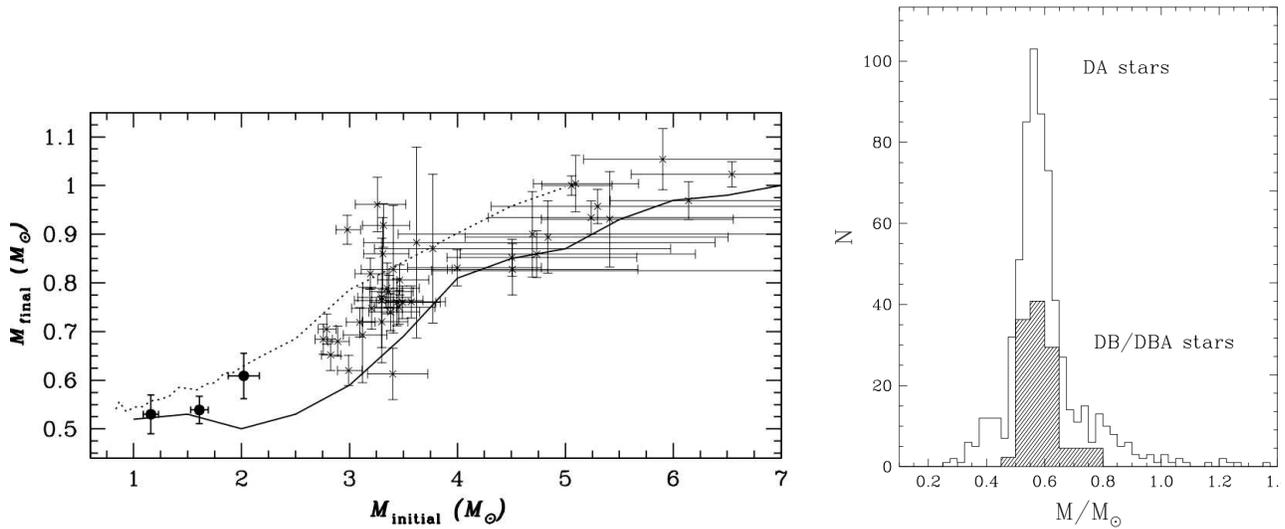
Figure 10.2: (Left:) Relation between the initial and final mass of low- and intermediate-mass stars, from Kalirai et al. (2008). The data points represent white dwarfs observed in open clusters, for which the mass has been determined from their spectra. The age of the cluster and the cooling time of the white dwarf have been used to estimate the initial mass, because the difference of the two corresponds to the lifetime of the progenitor star. The solid line shows model predictions for the core mass of a star at the start of the TP-AGB phase (from Marigo, 2001) for solar metallicity. The dotted line shows the final mass of these models, which is reasonably consistent with the data points. The growth of the core mass on the AGB is severely limited by dredge-up and strong mass loss. As a result, the final white dwarf mass spans a much smaller range than the main-sequence mass. *(Right)*: Observed mass distribution of white dwarfs, for a large sample of DA white dwarfs and a smaller sample of DB white dwarfs (from Bergeron et al. 2007). There is a sharp peak between 0.55 and $0.6M_\odot$, as can be expected from the initial-final mass relation because most white dwarfs come from low-mass stars with $M < 2M_\odot$. From Pols lecture notes.
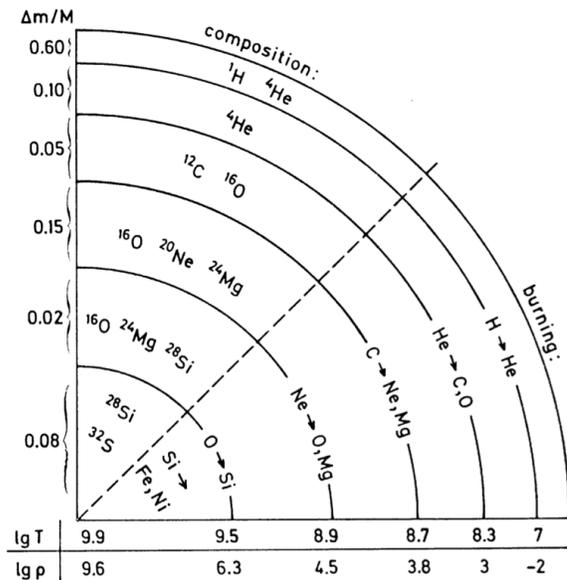
Figure 10.3: Onion-skin structure of a massive star in the very last stages of its life (not to scale). Typical fractional masses, temperatures (K), and densities ($g\,cm^3$) are indicated along the axes. From KW Fig. 33.1.

Table 10.1: Neutrino luminosities and timescales of late burning phases

Burning	.........$15\,M_\odot$.........		.........$25\,M_\odot$.........	
stage	L_ν/L	τ	L_ν/L	τ
	$(L \simeq 10^4\,L_\odot)$	(yr)	$(L \simeq 3\,10^5\,L_\odot)$	(yr)
C	1.0	$6.3\,10^3$	8.3	$1.7\,10^2$
Ne	$1.8\,10^3$	7	$6.5\,10^3$	1.2
O	$2.1\,10^4$	1.7	$1.9\,10^4$	0.51
Si	$9.2\,10^5$	0.017	$3.2\,10^6$	0.004

Taken from KW, their Table 33.1

nuclear reactions are now also buffeted by their reverse reactions, as the core is very hot and dense. Nuclear fusion is not a simple one-way process as in the Solar core, but proceeds more like one step forward and two steps backward. The resultant photons are absorbed and reabsorbed repeatedly, while the resultant neutrinos freely stream away. Some typical numbers are listed in Table 10.1, with neutrino loss exceeds surface luminosity starting from Carbon burning. As a result, there is little indication of the core evolution on the stellar surface.

While the next burning stage starts in the core, the burning of lighter elements will still continue in shells. As a result, the structure of a high-mass star near the end of its life becomes somewhat akin to that of an onion, in which regions with different chemical compositions are separated by burning shells (see Fig. 10.3). Such results have implications for the yield of nucleosynthesis the star returns to the interstellar medium.

When an iron core is formed, no further energy can be gained by fusion. In order to match the neutrino losses, therefore, the core has to shrink. This will cause the temperature to rise, and at $T > 5\,10^9$ K the photons become energetic enough to break up the iron nuclei into α particles, protons and neutrons. These reactions are endothermic and thus cool the core. As a result, the pressure drops, the core shrinks further, more iron becomes disintegrated, etc. At the same time, neutrinos keep on removing energy. Furthermore, as the density increases, electrons are being captured by remaining heavy nuclei (leading to neutronisation, i.e., converting a proton into a neutron, with the emission of a neutrino), thus reducing the pressure further. All these processes quicken the collapse, creating a 'perfect storm'.

At first, the core collapses roughly homologously (i.e., velocity proportional to radius), but soon this would require speeds in excess of the free-fall speed in the outer region. Thus, one has an inner collapsing core, with the outer core following on its own (longer) free-fall time. The latter is of order one second. The collapse of the inner core will stop only when the neutrons become degenerate, at $\rho \geq \rho_{\rm nuc} \simeq 10^{14}\,{\rm g\,cm^{-3}}$. The outer layers are still falling in, however, which leads to the development of a strong shock wave, which will start to move outward. At the same time, the inner core will become more massive and, since it is degenerate, smaller. Apparently, in nature, this eventually leads to a giant explosion with the stellar envelope expelled at high speed.

We now estimate a few energy quantities for the core, taking $M \simeq 1.4\,M_\odot$ and $\bar{\rho} \simeq 10^{14}\,{\rm g\,cm^{-3}}$, and, therefore, $R \simeq 2 \times 10^6\,{\rm cm}$. For these numbers, the potential energy is roughly

$$E_{\rm pot} \simeq \frac{GM^2}{R} \simeq 3 \times 10^{53}\,{\rm erg}. \tag{10.1}$$

Since the core was much larger before the collapse, we see that a couple $10^{53}\,{\rm erg}$ has to be liberated. We can compare this with the energy required to dissociate the iron in the core. For every nucleon, $\epsilon_{\rm diss} \simeq 9\,{\rm MeV} \simeq 1.4 \times 10^{-5}\,{\rm erg}$ is required (see Fig. 6.1). Thus,

$$E_{\rm diss} = \epsilon_{\rm diss}\frac{M}{m_{\rm H}} \simeq 2 \times 10^{52}\,{\rm erg} \tag{10.2}$$

which is substantially less than the total energy available. Next, compare the potential energy with the kinetic energy given to the envelope in a supernova explosion. With an envelope mass of $10\,M_\odot$ and a typical (observed) velocity of

$\sim 10,000\,\mathrm{km\,s^{-1}}$, the total kinetic energy is,

$$E_{\mathrm{kin}} = \frac{1}{2}M_{\mathrm{env}}v_{\mathrm{env}}^2 \simeq 10^{52}\ \mathrm{erg}. \tag{10.3}$$

Thus, there is ample energy available to expel the envelope. The energy emitted in optical light is $\sim 10^{49}\,\mathrm{erg}$, negligible in comparison, but leads to a luminosity similar to that of an entire galaxy during the roughly one month it lasts. By far most of the energy is lost in neutrinos. 24 of these were observed over a period of 13 seconds, from the explosion of SN 1987A.

While there is enough energy to expel the envelope, it has proven very difficult to reproduce the expulsion in models. There are three effects which are conjectured to help start an explosion:

1. The shock. There is enough energy in the shock for expulsion, but a lot of the energy is lost as the shock goes through the relatively dense inner part of the envelope (which is still falling in). This is because material is shock-heated to such an extend that neutrino losses and dissociation become important. From simulations, it seems only a very strong shock could cross through these layers and lead to a *prompt hydrodynamic explosion*.

2. Neutrino radiation pressure. The core is so dense that it is optically thick to neutrinos. As a result, the neutrinos have to diffuse out, and for a few seconds the core is a strong neutrino source (with $L_\nu \simeq 10^{53}\,\mathrm{erg\,s^{-1}}$). Above the "neutrinosphere", a fraction of the neutrinos will still be scattered, causing a radiation pressure term just like that due to photons. By equating the force due to neutrino scattering, $f_\nu = \kappa_\nu(L_\nu/4\pi R^2 c)$, with that due to gravity, $f_{\mathrm{g}} = GM\rho/R^2$, one can define a neutrino equivalent of the Eddington luminosity, $L_{\mathrm{edd},\nu} = 4\pi GMc/\kappa_\nu$. From calculations, it appears that the pressure due to the neutrinos in itself is insufficient to expel the outer layers, but that an explosion can be produced in combination with the shock, via strong heating and convective motion, in the so-called *delayed explosion mechanism*.

3. Thermonuclear reactions. When the shock arrives outside the original iron core, the shock heating will increase the speed of the fusion reactions in those regions dramatically. At the increased temperature, Si-burning results mostly in $^{56}_{28}\mathrm{Ni}$. This is an unstable isotope, which decays to $^{56}_{27}\mathrm{Co}$ through β-decay, with a half-life time of 6.1 d. $^{56}_{27}\mathrm{Co}$ is unstable as well, and decays to $^{56}_{26}\mathrm{Fe}$ (half-life 77.7 d). These and other decay processes keep the supernova bright for a longer time.

However, it remains a 'holy-grail' to witness the explosion happens inside a computer.

The explosion of a massive star core is called a **Type II** supernova (classified as such by virtue of possessing hydrogen lines in its spectrum), or **core-collapse SN**.

10.2   Type Ia Supernova

There are other types of supernova. In fact, the varieties are increasing as our observing capacity expands (Fig. 10.5).

Type I supernovae show no hydrogen lines in their spectra. One of the most famous type is called **Type Ia** supernovae (with a strong Si II line at 6150 Å) and result from the explosion of a carbon-oxygen white dwarf, which has ignited due to accretion from a binary companion, although both the nature of the companion, and the nature of the explosion, still remain largely in dispute. It is clear, however, that this explosion is intimately related to the **Chandrasekhar mass**, the maximum mass of a white dwarf.

Additionally, **Type Ib** supernovae (with helium lines) and **Type Ic** supernovae (with no helium lines) appear to be the explosions of massive stars, similar to those described above, but which have lost their hydrogen envelope, or even their helium envelope, by the time of their explosion, likely due to binary interactions or from very massive stars (Fig. 10.4).

Enrichment of the interstellar medium

Supernovae are a major source of heavy elements in the interstellar medium, contributing some helium, carbon, nitrogen, oxygen, iron, and many other elements. Intermediate mass stars (and even low mass stars) contribute some heavy elements
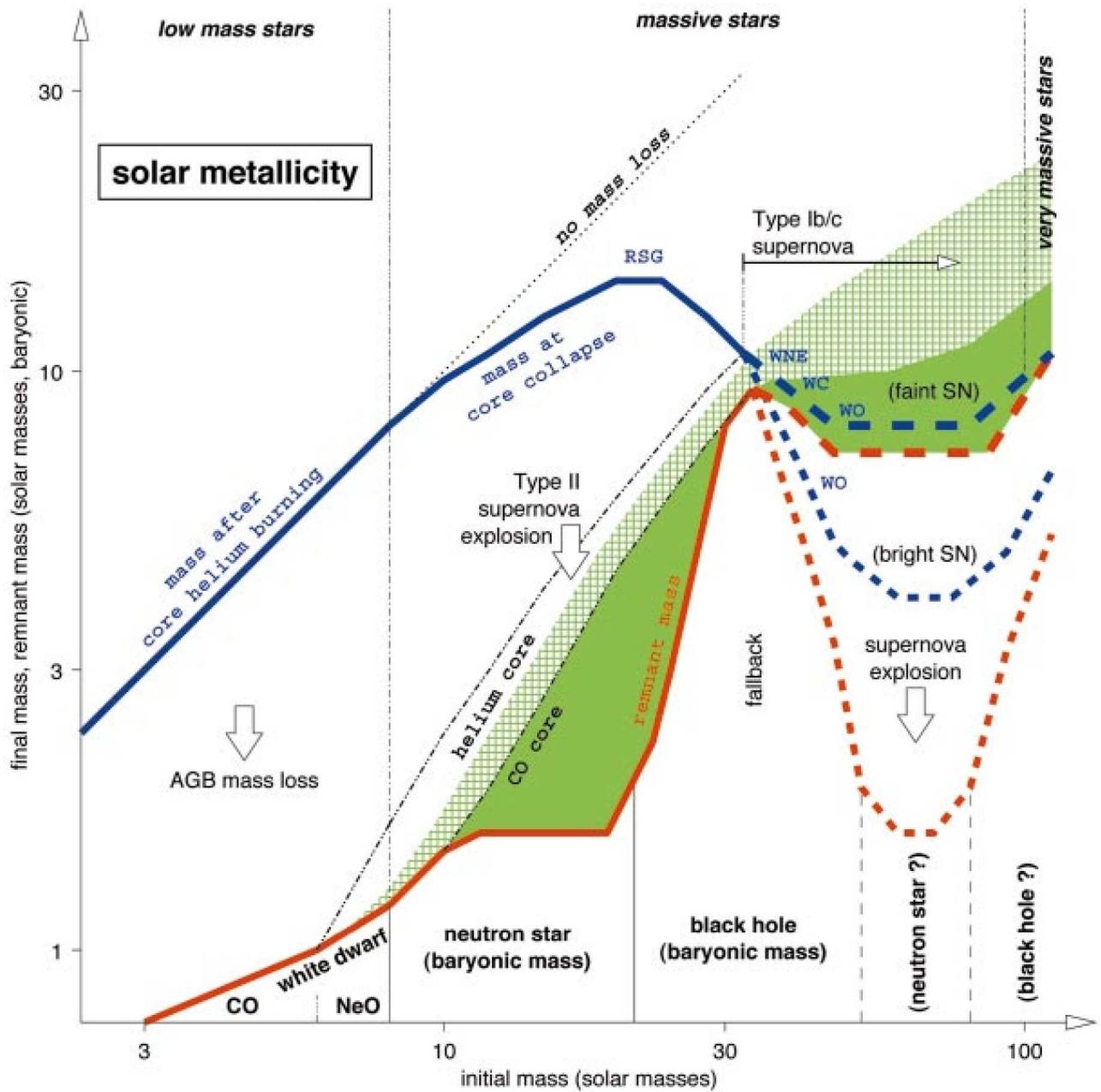
Figure 10.4: Initial-final mass relation for stars of solar composition, across a broad range of masses. The blue line shows the stellar mass after core helium burning, reduced by mass loss during earlier phases. For $M >\sim 30 M_\odot$, the helium core is exposed as a WR star, the dashed line gives two possibilities depending on the uncertain WR mass-loss rates. The red line indicates the mass of the compact stellar remnant, resulting from AGB mass loss in the case of intermediate- mass stars, and ejection of the envelope in a core-collapse supernova for massive stars. The green areas indicate the amount of mass ejected that has been processed by helium burning and more advanced nuclear burning. From Woosley et al (2002).
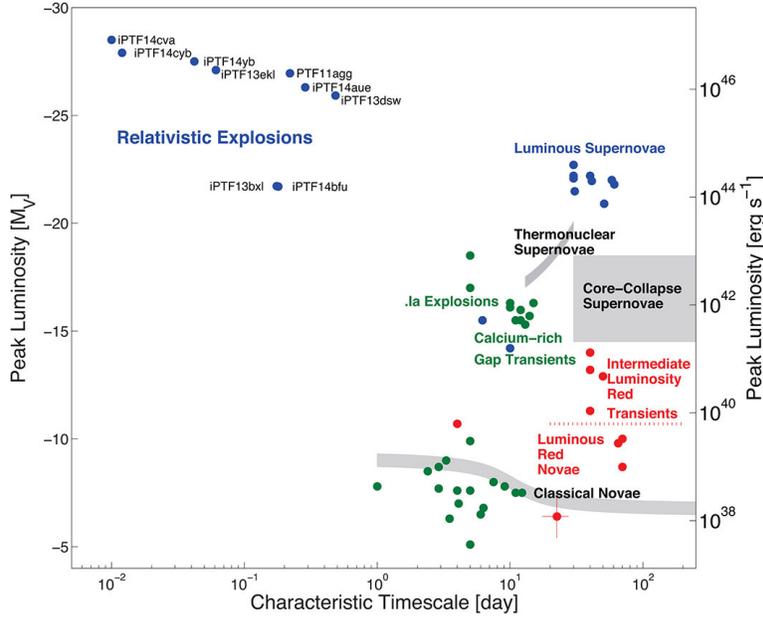
Figure 10.5: Astronomical transient sources, with their peak magnitude plotted as a function of characteristic decay timescale, both measured in optical. SN Type II are results of core-collapse. They have total luminosities of $\sim 10^{49}\,\mathrm{erg}$ spread over about a month. In a Milky-way like galaxy $(L = 10^{10} L_\odot)$, there is roughly one SN Type II per century. SN Type Ia are results of white dwarf collapse and they appears to form a tight sequence with decay time correlating with peak magnitude. This precious (but not yet understood) attribute allows them to be used as comological standard candles. They are a few times rarer than SN Type II. Many new types of transients (which are rarer) are discovered with time, currently mainly by the Palomar Transient Factory (PTF). These include relativistic explosions that have timescales of minutes, and abnormally luminous supernova. Figure by Kasliwal.

due to the mass loss that removes their envelopes, yielding mainly helium, carbon, nitrogen, and s-process elements. There are other sources as well, including novae (recurrent thermonuclear explosions on the surfaces of accreting white dwarfs in binary systems) and cosmic rays (which produce beryllium, boron, and lithium by spallation as they hit heavier nuclei such as carbon in the interstellar medium).

10.3   Degenerate Remnants

White Dwarfs

White dwarfs cannot have arbitrary masses. For non-relativistic, complete degenerate electron gas, the pressure $P = \mathrm{const} \times (\rho/\mu_e)^{5/3}$ (eq. [2.13]), which yields a mass-radius relation of

$$R \propto M^{-1/3}, \tag{10.4}$$

namely, more massive white dwarfs are more compact. However, above a certain density, electron momenta rise so much they are relativistic $(v_e \to c)$, $P = \mathrm{const} \times (\rho/\mu_e)^{4/3}$ (eq. [2.15]), and for this equation of state, the polytropic index is 3 and there is something funny.

At $n = 3$, Eq. (3.5) shows that mass is a constant, independent of radius. Inserting in all physical constants, this unique mass is called the **Chandrasekhar mass**,

$$M_{\mathrm{Ch}} = 5.836 \mu_e^{-2} M_\odot. \tag{10.5}$$

This is first derived by Chandrasekhar (1930). It is interesting that the electron mass never enters this derivation.[8] For helium, carbon or oxygen... typical composition of white dwarfs, $\mu_e = 2$. So $M_{\mathrm{Ch}} = 1.46 M_\odot$. No known white dwarf has a mass that exceeds this limit.

If a sub-Chandrasekhar mass white dwarf accretes enough mass (from a binary companion) to cross the boundary, it may undergo catastrophic collapse. Carbon may be explosively detonated in the interior ("thermonuclear explosion"), leading to a Type Ia supernova. In detail, at high enough density $(\rho \geq 2 \times 10^9\,\mathrm{g/cm^3}$, see Fig. 9.1), carbon burning may start. Due to electron degeneracy, the burning is not controlled by a thermostat but is instead a run-away. Temperature can reach upward of $10^{10}\,\mathrm{K}$, allowing fusion to proceed all the way to Fe-peak (actually ^{56}Ni). For a mass $M \approx 1.4 M_\odot$, this

---

[8]Though the electron mass does enter in the mass-radius relation for white dwarfs.
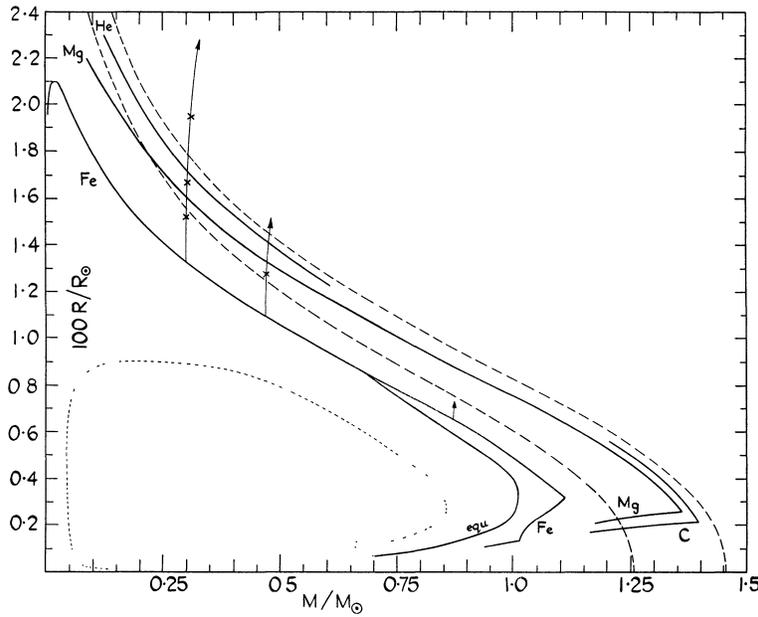
Figure 10.6: Mass-radius relation for white dwarfs of various compositions. The dashed curves indicate Chandrasekhar models for $\mu_e = 2$ (upper) and 2.15 (lower), in which simple estimates like those discussed in class are used, except that the mildly relativistic regime is treated correctly. The models deviate from these idealized curves because the elements are not completely ionized, and at very high densities, inverse beta decay becomes important (the curve labelled 'equ' takes into account the resulting changes in elemental abundances). For both reasons, there are variations in μ_e. The arrows indicate the effects of adding a hydrogen atmosphere. And the down-turn of the curves at $M \sim 1.4 M_\odot$ corresponds to the Chandrasekhar mass limit. The dotted curve is a mass-radius relation for neutron stars. Taken from Hamada & Salpeter (1961). Observational support for these values are presented in Provencal et al (1998).
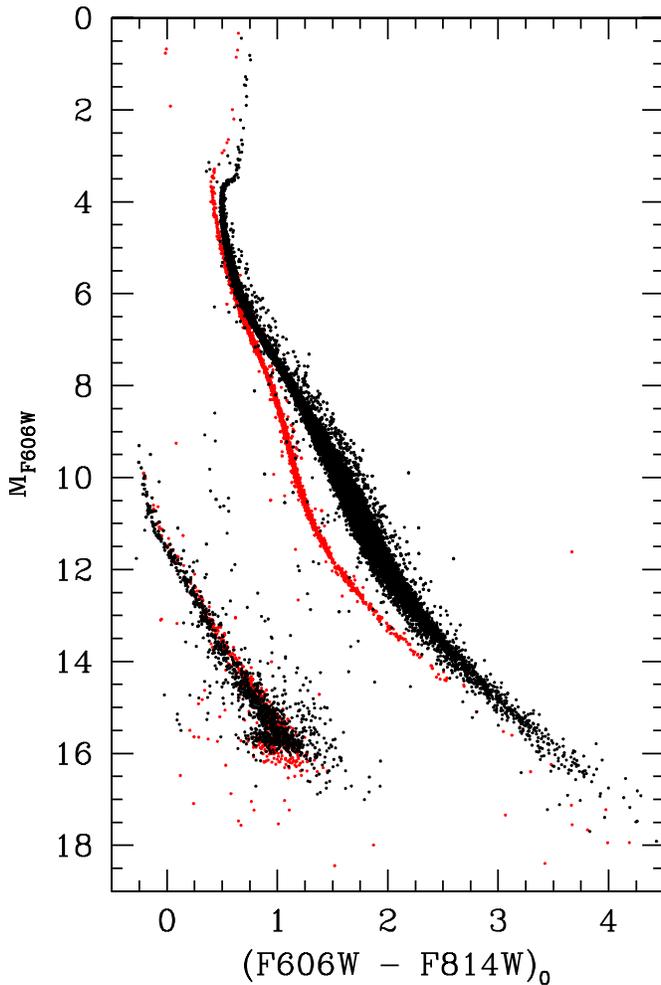
Figure 10.7: Exquisite HRD obtained by HST for two globular clusters NGC 6397 ($[Fe/H] = -2$, ~ 13 Gyrs, red) and 47 Tuc ($[Fe/H] = -0.75$, ~ 10 Gyrs, black). The main-sequences of the clusters do not align while the WD cooling sequences do. Metallicity affects radiative transfer on main-sequence stars, while white dwarfs have very uniform chemical composition and their cooling is only a function of their masses. Taken from Richer et al (2013).
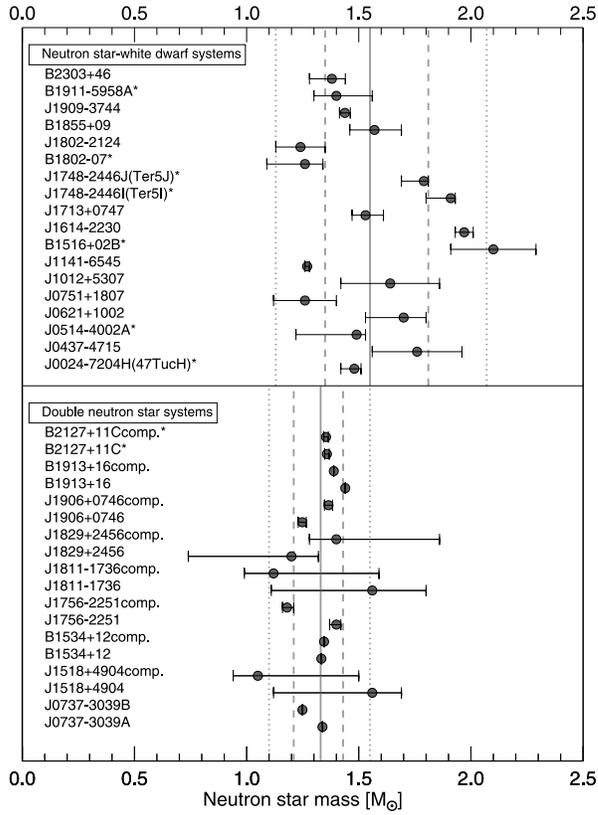
Figure 10.8: Measured masses of radio pulsars. All error bars indicate the central 68% confidence limits. Vertical solid lines are the peak values of the underlying mass distribution for NS-NS ($m = 1.33 M_\odot$) and NSWD ($m = 1.55 M_\odot$) systems. The former may be a result of formation and the latter may indicates some additional accretion. Systems marked with asterisks are found in globular clusters. The highest mass is $2.1 M_\odot$. This establishes a firm lower bound for the maximum NS mass and rules out the majority of strange quark and soft equation of state models as viable configurations for NS matter. It may be a result of evolution rather than the true maximum mass of a neutron star. From Kiziltan et al (2013).

releases an energy of order ($\epsilon_{\mathrm{nuc}} \approx 1$ Mev per nucleon)

$$E_{\mathrm{gain}} \approx \epsilon_{\mathrm{nuc}} \frac{M}{m_H} \approx 3 \times 10^{51} \,\mathrm{erg}\,. \tag{10.6}$$

This is sufficient to unbind the white dwarf, which, for a radius of 5000km, has a gravitational energy of $\sim GM^2/R \sim 10^{51}\,\mathrm{erg}$. How the explosion actually occurs, however, remains another mystery to be cracked.

Though much less energetic than a SN Type II, the radioactive decay of ^{56}Ni (half-life 6 days, first to ^{56}CO, half-life 77 days, then to ^{56}Fe) in the ejecta boosts the optical brightness of a Type Ia over that of a Type II (Fig. 10.5).

– cooling

The cooling (but no contraction) of white dwarfs can be meticulously studied, both observationally and theoretically (Fig. 10.7).

Neutron Stars

– neutron star maximum mass

Neutron stars are made up of degenerate neutrons. What is its maximum mass? Repeat the derivation that leads up to the Chandrasekhar mass, you will find that, for extremely relativistic matter, the Chandrasekhar mass should remain the same (with μ_e now substituted with $\mu_n = 1$ in eq. [10.5]). However, neutron stars are much denser than white dwarfs, as a result, general relativistic correction is important. This is shown to reduce the maximum mass to $0.75 M_\odot$ (and a radius of 9.6km). (Oppenheimer & Volkoff , 1939).

At nuclear density, however, strong force is relevant. Its repelling nature reduces effective gravity and raises the maximum mass. Nuclear equation of state is uncertain at these densities, but typical calculations give a maximum neutron star mass between 1.5 to $3 M_\odot$. So looking for the heaviest neutron star becomes a test for this complicated physics (Fig. 10.8).
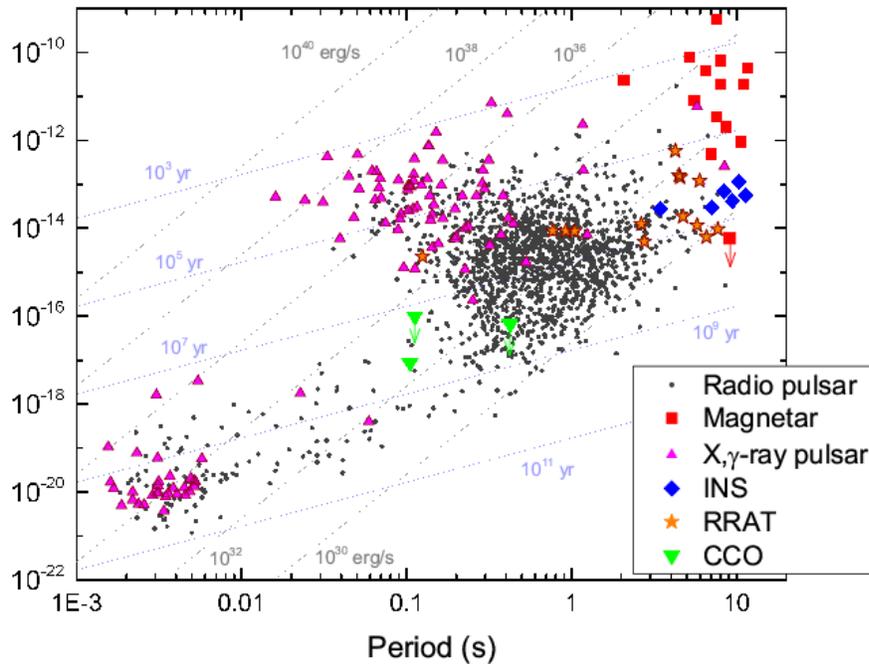
– neutron star magnetism

Figure 10.9: Plot of period vs. period derivative for the presently known rotation-powered pulsars, Isolated Neutron Stars (INS), Compact Central Objects (CCO), Rotating Radio Transients (RRATs) and magnetars Lines of constant characteristic age, $P/2\dot{P}$, and dipole spin-down luminosity, are superposed. From Harding (2013).

Most of the neutron stars we know of are discovered by the fact that they are magnetic and spinning. They lose energy via magnetic dipole radiation and spin down. There are large dynamic ranges in both magnetic field strength and rotation period.

Food for thought

- Why is successive burning of elements in a massive star becoming increasingly faster? what can be observed at the surface?

- So why does the inward collapse of a massive star drive an outward explosion?

- Why do white dwarfs have such a narrow mass spread?

- What kind of stars make black holes?

- Why are neutron stars much smaller than white dwarfs?

- Why don't single star evolution produce Type Ia SN?
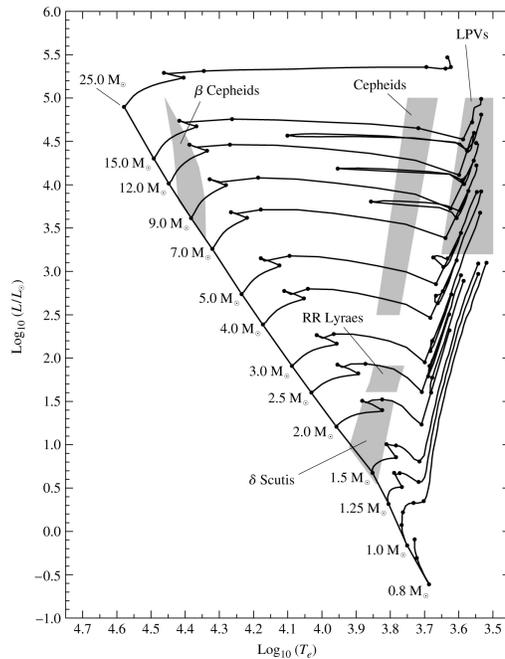
Figure 10.10: Major types of pulsating stars. Notice the slanted strip that includes some of the most prominent pulsating stars (Cepheids, RR Lyraes...). This strip corresponds to stars that have near-surface ionization zones. Such a zone drives instability in stars. Lower density stars have longer pulsational periods so measuring periods yields stellar properties (including its intrinsic luminosity and therefore the distance). Going up the strip, the stars become ever more luminous and Cepheids can be visible out to distances of tens of Mpc (out to the Virgo cluster, the Hubble Key project). In contrast, RR Lyrae stars are more useful for measuring distances within the local group. CO Fig. 14.8.

Table 11.1: Cosmological Distances and Speeds

	distance	peculiar velocity [km/s]	Hubble expansion ($v = Hd$) [km/s]
nearest star	1 pc	20	~ 0.00007
Galactic centre	8 kpc	200	0.6
LMC	50 kpc	200	3.5
M31 (Andromeda)	0.8 Mpc	200	56
Virgo cluster/super-cluster	30 Mpc	2000	2000

11 Distance Ladder

Readings: CO §*27.1, 27.2, 27.3,*Back

– Review AST222 material on cosmology

As we start on our study of cosmology, it is useful to keep this in mind: in modern cosmology, there are a large number of ideas that are firmly established, and there are some that are in the frontier and yet to be confirmed. It is important to make the distinction between the two.

– the Hubble constant $H_0 = v/d$, where $v \approx cz$ is the observed Doppler redshift velocity due to universal expansion, and d the physical distance.

Relevant astronomical distances go from AU to Gpc. For a "pocket map of the universe", see Fig. 7 in this paper. The simplified Table 11.1 gives you some sense of the distance scale and associated speeds. "Peculiar velocity" is the physical motion of galaxies (or else) move around each other under their mutual gravity. This can be much larger than the Hubble expansion at distances small than ~ 50 Mpc (cluster of glaxies, also see Fig. 10.13). In other words, at scales below this value, motion is little affected by the universal expansion.
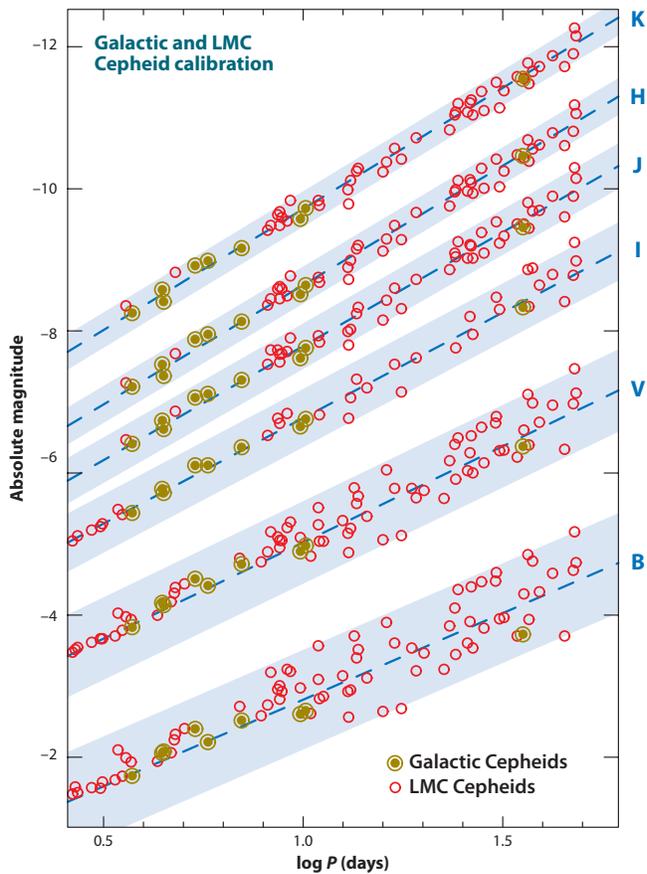
Figure 10.11: Multiwavelength period-luminosity (PL) relations (Leavitt Laws) for Galactic (circled filled dark yellow dots) and Large Magellanic Cloud (LMC) (open red circles, using the LMC distance as measured in Fig. 10.12) Cepheids from the optical (BVI) through the near-IR (JHK). There is a monotonic increase in the slope, coupled with a dramatic decrease in total dispersion of the PL relations as one goes to longer and longer wavelengths. This motivates using the Spitzer infrared space telescope to monitor Cepheids. From Freedman & Madore (2010).
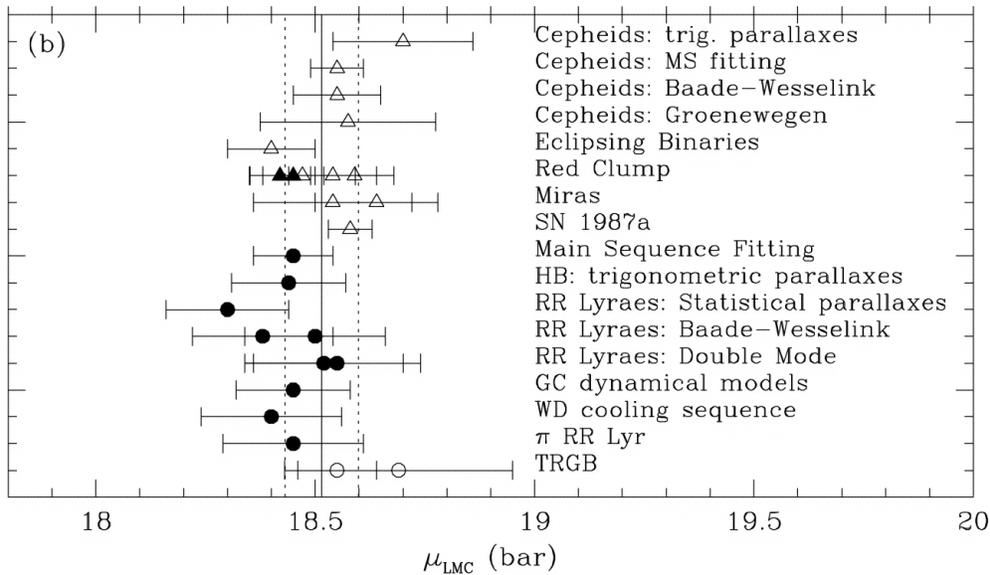
Figure 10.12: Different methods to measure distance to the Large Megellanic cloud, an important rung in the distance ladder with errors that propagates throughout all of extragalactic astronomy. The Distance modulus is related to distance as $\mu = 5\log(\text{distance}[\text{pc}]) - 5$ and LMC is at about 50 kpc. Use your knowledge to guess the rational behind the methods listed here (see, e.g., the Wiki page), and to translate an error bar of $\delta\mu \sim 0.1$ into an uncertainty in distance – compare this to the intrinsic thickness of LMC (with a diameter of ~ 5 kpc).
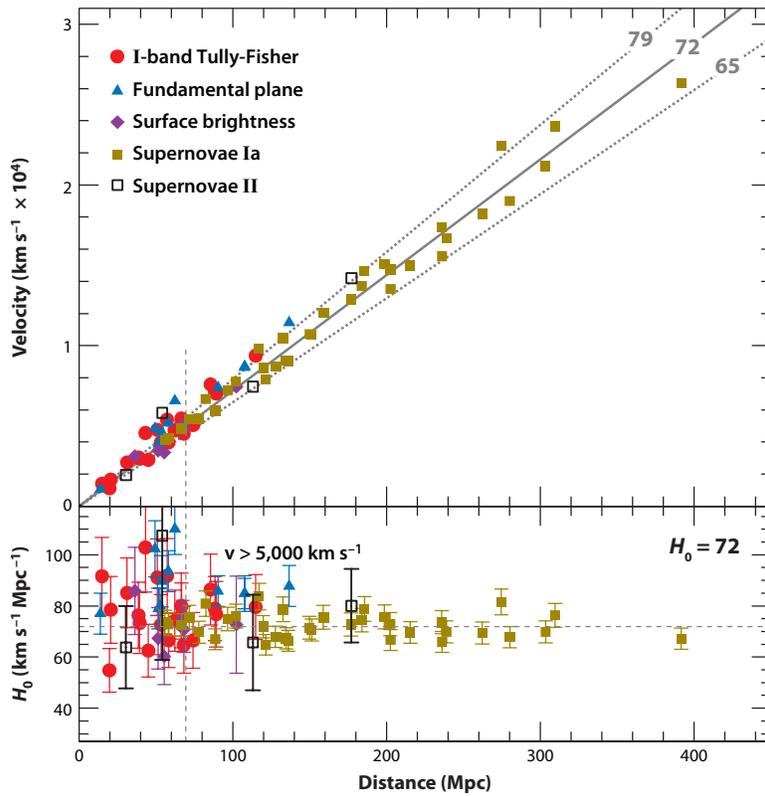
Figure 10.13: Measurements of the Hubble constant. *(Top)*: The Hubble diagram of distance versus recession velocity using secondary distance indicators that are calibrated by Cepheids discovered in the Hubble Key project. Each point stands for one galaxy. The Hubble key project determines the Hubble constant to be $H_0 = 72 \pm 2(\text{statistical}) \pm 6(\text{systematic}) \, \text{km s}^{-1} \text{Mpc}^1$ (shown as solid and dotted gray lines). *(Bottom)*: the Hubble constant obtained from these measurements. There is a large scatter in H_0 for galaxies receding slower than $5,000 \, \text{km s}^1$ (vertical dashed line), a result of their "peculiar velocities". The Type Ia supernovae extend to the furthest distance, among all standard candles. These objects were later to reveal the presence of "dark energy". From Freedman & Madore (2010).
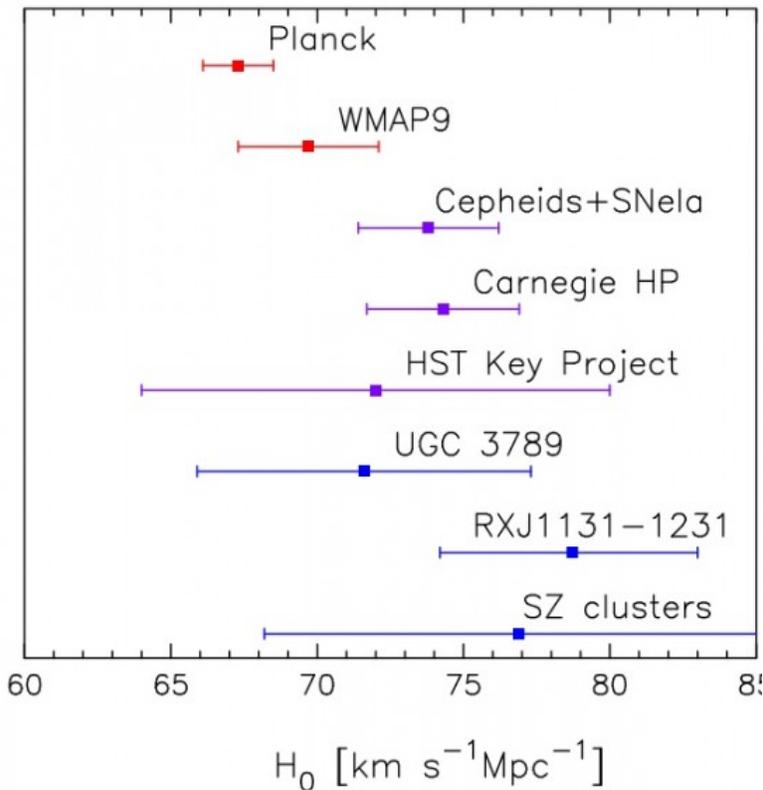

Figure 10.14: Aside from a host of other cosmological parameters, CMB can also be used to measure the Hubble constant. The latest result from Planck mission states $H_0 = 67.3 \pm 1.2 \, \text{km s}^{-1} \text{Mpc}^{-1}$, in slight tension with the other measurements. From Abe et al (2013) (the Planck team).

Food for thought

- How far out in distance can the parallax technique probe? FYI, a major milestone was the Hipparcos satellite (1989-1993) which provided a parallax precision of 0.002 arcsecond; while the current GAIA satellite (2015-) aims for a precision of ~ 10 micro-arcsecond (μas).

- Edwin Hubble, who discovered the Hubble expansion (1924), originally pegged the Hubble constant at $500 \, \mathrm{km \, s^{-1} Mpc^{-1}}$. Since then, this value has been continuously driven down (See here for a brief history). What did Hubble get wrong? Why did he get a much larger H_0?

- How can everything be receding away from us, yet we are not at the centre of the universe?

- Given that Cepheids have luminosities of $L \sim 10^4 L_\odot$, and SN Ia have luminosities of $L \sim 10^9 L_\odot$, how far out can you see these two types of objects? (assuming a magnitude limit of 25 mag)

- SN Type Ia and Type II have comparable peak luminosities (Fig. 10.5). Why aren't the latter used as standard candles? What is the unique quality of Ia that makes it so useful for cosmology?

- If $z \geq 1$, does that mean recession speed is faster than speed of light? can space expands faster than light?

- When one writes down $v = Hd$, which distance is referred to in that expression?

First, I am not as an expert on cosmology as I could be. So to make up for my brief notes and shortage of contents, here are a few good links, in addition to the text book CO:

- NED: NASA's Knowledge base for Extragalactic Astronomy and Cosmology. A wikipedia written by experts.

- Cosmic Calculator by Ned Wright, where you can try your hand to get various cosmological measure-ables, for a given cosmological model.

- introductory undergraduate lecture notes from Caltech, by Prof. George Djorgovski, including many links.

- advanced undergraduate lecture notes from Ohio State University, by Prof. David Weinberg.

12 Newtonian Cosmology

Readings: CO §*29.1,*Back

– **Olbers' paradox** The universe must be finite in age. This resolves the paradox "Why is the sky dark?" – imagine a forest that stretches forever, eventually every sightline will be hitting a tree, even though a far away tree will appear thinner. In a static universe that has existed forever, the night sky should be bright.

– **Cosmological principle** The universe is isotropic and homogeneous (on large scales) – we are not special and no direction is special, at least not in sufficiently large scale.

– **universal expansion, Hubble law** Given the cosmological principle, expansion can only be by way of a Hubble-like law, in which the current positions

$$\vec{x}(t) = R(t)\vec{\varpi} \ , \tag{12.1}$$

with $\vec{\varpi}$ co-moving positions that do not depend on time (on large scales) and $R(t)$ a time-dependent scale factor. As universe expands, both photon wavelength and space-time expands,

$$1 + z = \frac{\lambda_{\text{obs}}}{\lambda_{\text{em}}} \qquad \Rightarrow \qquad \lambda \propto R \tag{12.2}$$

or $R/R(t_0) = 1/(1+z)$. For the Hubble constant,

$$\vec{v}(t) = \dot{\vec{x}}(t) = H(t)\vec{x}(t) \qquad \Rightarrow \qquad H(t) = \frac{\dot{R}(t)}{R(t)} \ . \tag{12.3}$$

Note that $H_0 = H(t_0)$ is the present value of the Hubble constant ($t_0 = $ "now"), and by definition $R(t_0) = 1$. Also interesting to note that $1/H_0$ is time, called the Hubble time, $t_H = 1/H_0 \sim 13.8$ Gyrs.

Friedmann equation

We obtain the expansion of the universe by using a bad Newtonian analogy. This is bad because it considers movement of mass, not expansion of space. Fortuitously, it gives identical mathematical results as the full GR version. Assume flat space-time, mass conservation, zero pressure,[9] and energy conservation, the expansion of the universe is akin to the motion of an apple thrown upward (with initial energy $-1/2 \, kc^2\varpi^2$),

$$\frac{1}{2}v^2 - \frac{GM_r}{r} = -\frac{1}{2}kc^2\varpi^2 \quad \Rightarrow \quad H(t)^2 - \frac{8\pi G}{3}\rho(t) = -\frac{kc^2}{R(t)^2} \qquad \text{with} \quad \begin{cases} k > 0 & \Rightarrow \quad \text{closed (bound)} \\ k = 0 & \Rightarrow \quad \text{flat (parabolic)} \\ k < 0 & \Rightarrow \quad \text{open (escaping)} \end{cases} \tag{12.4}$$

where $\rho = \rho(t) = R^3(t_0)\rho(t_0)/R^3 = \rho_0(1+z)^3$. The **critical density** required to yield a flat universe:

$$\rho_{\text{c}}(t) = \frac{3H(t)^2}{8\pi G} \ , \tag{12.5}$$

Associated is a dimensionless *density parameter*, the density relative to the critical density:

$$\Omega(t) \equiv \frac{\rho(t)}{\rho_{\text{c}}(t)} = \frac{8\pi G\rho(t)}{3H(t)^2} \qquad \text{with} \quad \begin{cases} \Omega > 1 & \Rightarrow \quad \text{closed (bound)} \\ \Omega = 1 & \Rightarrow \quad \text{flat (parabolic)} \\ \Omega < 1 & \Rightarrow \quad \text{open (escaping)} \end{cases} \ . \tag{12.6}$$

The measured Hubble parameter (H_0), and the Newton's constant combine to give $\rho_{\text{c},0} = 10^{-29} \, \text{g/cm}^3$ (or ~ 6 Hydrogen per cubic metre, compared to the estimated ~ 0.2; the latter values includes dark matter contribution, see Fig. 12.1). Or, currently, $\Omega < 1$.

---

[9] For non-relativistic particles, the contribution to energy density by pressure is much smaller than that from rest mass, $P = nkT \ll \rho c^2$.
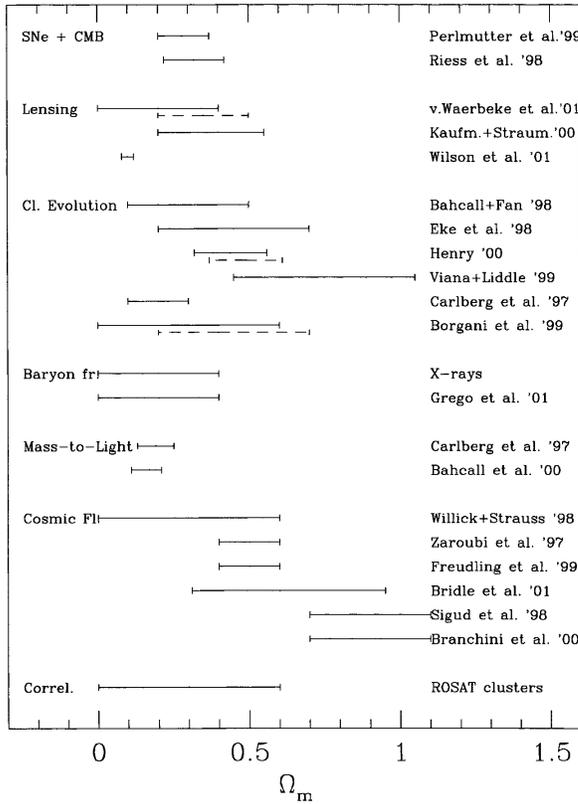
Figure 12.1: Summary of Ω_m values as derived by the different methods (selection only). The methods are listed on the left hand side, the references on the right hand side. If the authors distinguished between open and flat models, the flat models are shown as full lines and the open models as dashed lines. Most methods are in agreement with an $\Omega_m \approx 0.3$, this includes both baryons and dark matter. This was an active area of research (with heated debates) about ten years ago, but by now the dust has mostly settled. From Schindler (2002).

Matter-dominated, flat Universe

$$R = (6\pi G\rho_c)^{1/3}\, t^{2/3} = \left(\frac{3t}{2t_H}\right)^{2/3} \qquad \Rightarrow \qquad \frac{t(z)}{t_H} = \frac{2}{3}\left(\frac{1}{1+z}\right)^{3/2}. \tag{12.7}$$

In this model, age of the flat universe is $2/3\, t_H \sim 9.2$ Gyrs, shorter than the age of some of the oldest known globular clusters (~ 13 Gyrs). The current expansion rate is too fast. There are a number of ways out of this: (1) H_0 value wrong; (1) Universe not flat **and** $\Omega_0 < 1$ (see CO Fig. 29.6, understand why); (2) Matter-domination incorrect, invoke 'pressure'; (3) Newtonian view incorrect, needs rescue by GR. If you are an observer, what experiments should you design to sort this out?

Food for thought:

- Why is 'homogenous' not enough? one needs 'isotropic' for the universe as well? can you imagine an isotropic universe without being homogeneous? or a homogeneous one without being isotropic?

- What does the initial energy $1/2\, kc^2\varpi^2$ in eq. (12.4) correspond to? Why does it have this form?

- Measuring H_0 yields two pieces of information about the universe. What are these?

- How does one measure matter density in the universe?

- If we found some stars or other objects that had ages $t_{\rm star} > t_H$, what would this imply?

- We assume matter-domination, no pressure. For baryons, what about gas pressure? For photons, is it reasonable to ignore pressure?

- The universe is intrinsically general relativistic. What is the value in understanding it using Newtonian method?

13 One Equation for the Universe

Einstein's theory of general relativity (1915) allows us to write down one equation for the universe. See Notes by Weinberg for some background/math.

Robertson-Walker metric According to special relativity, two observers, moving relative to each other, would disagree with the time separation (dt) and spatial separation between two events (dr). For instance, the heartbeat of an astronaut on a moving space-ship, as viewed by the astronaut and someone on Earth, are different in both dt and dr, but both should agree on the spacetime interval, the so-called 'proper distance',

$$(\mathrm{d}s)^2 = \sum_\mu \sum_\nu g_{\mu\nu} dx^\mu dx^\nu = (c\mathrm{d}t)^2 - (\mathrm{d}r)^2 \,, \tag{13.1}$$

and $g_{\mu\nu}$ is the metric tensor and relates how space coordinates are related to lengths. The fact that time dimension and space dimension contribute oppositely to the proper distance is an insight by Poincare and Minkowski, and it leads to Lorentz invariance (physical laws invariant to frame change). Time can be thought of as the fourth spatial dimension, and photons move along the so-called 'geodesics' (straight line in curved space-time) for which $ds = 0$.

The most general metric possible for describing the proper distance between two events (in space-time) in an isotropic and homogeneous universe is called the **Robertson-Walker metric**,

$$(\mathrm{d}s)^2 = g_{\mu\nu} dx^\mu dx^\nu = (c\mathrm{d}t)^2 - R^2(t)\left[\left(\frac{\mathrm{d}\varpi}{\sqrt{1-k\varpi^2}}\right)^2 + (\varpi\mathrm{d}\theta)^2 + (\varpi\sin\theta\,\mathrm{d}\phi)^2\right] \tag{13.2}$$

where the radial vector $r(t) = R(t)\varpi$, and $R(t)$ the scale factor $(R_0 = R(now) = 1)$. The coordinates (ϖ, θ, ϕ) are the so-called co-moving coordinates as they remain constant for observers frozen in the Hubble expansion – if a galaxy is separated from us by 1 Mpc today, it was and will always be so in comoving distance. This metric is homogeneous– no cross-term between time and space, and isotropic – spherical symmetry.

Here, k is the time-independent curvature of the Universe $(k = \pm 1, 0)$, with the instantaneous curvature radius being $R(t)/k$. $k = 0$ recovers the trivial flat-space (Euclidean) distance, $(dr)^2 = R^2(t)\left[(d\varpi)^2 + \varpi^2 d\Omega^2\right]$, where $d\Omega$ is the variation in angular coordinates. A $k = 1$ (positively curved) universe corresponds to a sphere embedded in 4D (our normal sphere is in 3D). In such a universe, initially parallel geodesics converge, as they do on a sphere (e.g., great circles).[10] Imagine yourself as an ant on a 2-sphere to see why the proper distance is larger than that in flat-space.[11]

Friedmann equation

In Newtonian gravity, gravity and its effects are described by the Poisson equation and the equation for gravitational acceleration

$$\nabla^2\Phi = 4\pi G\rho, \qquad\qquad \mathbf{g} = -\nabla\Phi \,, \tag{13.3}$$

where Φ is the scalar gravitational potential. In this picture, space-time is fixed and matter moves according to some 'force' called gravity.

In general relativity, the second equation is recast into the equation for geodesics (not reproduced here), while the first equation is modified into the Einsten field equation,

$$R_{\mu\nu} - \frac{1}{2}R g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} \,, \tag{13.4}$$

where $R_{\mu\nu}$ is the Ricci tensor that describes the space-time and is function of $g_{\mu\nu}$ (through the Christoffel symbols, these are all in introductory GR courses that you can learn about), and R the scalar curvature; $T_{\mu\nu}$ is the stress-energy-momentum

---

[10]For an understanding of curvature of a 3-D space, see this youtube video.

[11]For a geometric understanding of this equation, see Lecture Note by Weinberg.

tensor that describes how matter, energy, momentum... are distributed in space. In the weak field ($\Phi \ll c^2$) and non-relativisit ($v \ll c$) limit, space-time curvature is nearly zero (Minkowski space), and we retrieves the Newtonian gravity.

The catch-phrase for GR is: matter tells space how to bent ($T_{\mu\nu}$ informs $g_{\mu\nu}$), and space tells matter how to glide (the equation for geodesic). The concepts of 'force' and 'acceleration' are useless here. These equations have been applied to motion around blackholes, neutron stars, and the Sun and have been experimentally confirmed to high precision. One makes a bold jump to apply them now to the entire universe, extrapolating boldly in time and spatial scales.

For an isotropic, homogeneous ideal fluid, $T_{\mu\nu} = diag(-\rho, p, p, p)$, where ρ is the mass energy (energy density) and P the pressure (momentum density). The density ρ combines rest mass and the mass-equivalent of energy (e.g., photon energy). Assuming that the universe is indeed isotropic and homogeneous, we can solve the Einstein's field equation to obtain

$$\left[\frac{\dot{R}(t)}{R(t)}\right]^2 - \frac{8\pi G}{3}\rho(t) = -\frac{kc^2}{R(t)^2} . \tag{13.5}$$

Notice the uncanny resemblance to eq. (12.4). The present interpretation for k, however, has a concrete physical foundation. The lecture note here gives a derivation of the Friedman equation.

Present observations of distant supernovae and of the cosmic microwave background indicate an additional, more exotic contribution than the previously-assumed photons, baryonic matter, neutrinos, and dark matter. This exotic *dark energy* appears to be causing the expansion of the universe to accelerate. The simplest (but not the unique) way to obtain such an effect in the Friedmann equation is to insert a *cosmological constant* Λ , as follows,[12]

$$\left[\frac{\dot{R}(t)}{R(t)}\right]^2 - \frac{8\pi G}{3}\rho(t) - \frac{1}{3}\Lambda c^2 = -\frac{kc^2}{R(t)^2} , \tag{13.6}$$

where Λ is, for the following discussion, assumed to be a constant (but it doesn't have to be this way).

It is often useful to write the Friedmann equation in terms of current values of $\Omega_{\rm M}$, Ω_Λ, etc., as follows:

$$\left[\frac{\dot{R}}{R}\right]^2 = H^2 = H_0^2\left(\frac{\Omega_{\rm rad,0}}{R^4} + \frac{\Omega_{\rm M,0}}{R^3} + \frac{\Omega_{k,0}}{R^2} + \Omega_{\Lambda,0}\right) \quad \text{with} \quad \begin{cases} \Omega_{\rm rad,0} &= \dfrac{8\pi G a T_0^4}{3c^2 H_0^2} , \\[2mm] \Omega_{\rm M,0} &= \dfrac{8\pi G \rho_{\rm M,0}}{3 H_0^2} , \\[2mm] \Omega_{k,0} &= \dfrac{-kc^2}{H_0^2} , \\[2mm] \Omega_{\Lambda,0} &= \dfrac{\Lambda c^2}{3 H_0^2} . \end{cases} \tag{13.7}$$

Note that $\Omega_{\rm rad,0} + \Omega_{\rm M,0} + \Omega_{k,0} + \Omega_{\Lambda,0} = 1$. Very oftenly, one defines $\Omega_{\rm tot,0} = \Omega_{\rm rad,0} + \Omega_{\rm M,0} + \Omega_{\Lambda,0}$ and $\Omega_{k,0} = 1 - \Omega_{\rm tot}$. A super-critical universe ($\Omega_{\rm tot,0} > 1$) has $\Omega_{k,0} < 0$ and the universe has a positive curvature (or closed). The different powers of R for the right-hand-side terms can be explained as:

- for non-relativistic particles, the kinetic energy of a particle is negligible compared to its rest mass, and thus,

$$\text{matter:} \qquad \rho_{\rm M} = \rho_{\rm M,0}\, R^{-3} , \qquad \text{where } \rho_{\rm M,0} \text{ is constant.} \tag{13.8}$$

- For photons and extreme-relativistic particles, not only is the number density diluted by $1/R^3$, the wavelength is stretched proportionally to R (as one can see also from $\rho_\gamma = u/c^2$, where the radiation energy density is $u = aT^4$, combined with $T \propto 1/R$). Thus,

$$\text{radiation:} \qquad \rho_\gamma = \rho_{\gamma,0}\, R^{-4} , \qquad \text{where } \rho_{\gamma,0} \text{ is constant} \tag{13.9}$$

Note that the above relations assume that matter and radiation are decoupled, and can have different temperatures.

---

[12]This is allowed by the Einstein equation as one can add an arbitrary normalization term $\Lambda g_{\mu\nu}$ to the left-hand-side.

- Radiation dominates over matter at early enough times. At present, $\rho_\gamma \ll \rho_M$, with ρ_M dominated by *dark matter*. Proposals for dark matter range from mini-blackholes to cold planets to exotic types of elementary particles. The last proposal is the most likely at the moment. These are particles that presently interact only by means of gravity, not electro-magnetically (unlike baryons). They were produced in the early universe where they interact strongly but have since decoupled. Moreover, '**cold dark matter**', where the rest mass energy of the particles (mc^2) dominates over its kinetic energy (in contrast to hot dark matter), is observationally preferred.

- The cosmological constant is sometimes thought of as vacuum energy. If this density doesn't change with time, then it must be negligible at early times (in comparison to matter and radiation). However, to our utter confusion, currently, $\Omega_\Lambda \sim \Omega_M$.

- Similarly, at the very early universe, the curvature term becomes too small to matter. So regardless of cosmology, the early universe is very 'flat'.

- In general the "dark energy" need not be a cosmological constant per se, but something with an exotic "equation of state" which yields a pressure $p_\Lambda = w\rho_\Lambda$, where negative values of w yield acceleration in the universal expansion. A cosmological constant is equivalent to $w = -1$ (which is often simply assumed to be the case), but other values of w (possibly varying with time) correspond to other forms for Λ (varying with time).

Solutions for special cases

The Friedmann equation can often be solved for flat Universes ($k = 0$). Examples are (below, $t_H \equiv 1/H_0$ is the current Hubble time):[13]

- $k = 0$, $\Omega_M = 1$ (flat, matter-dominated): (same as eq. [12.7])

$$R = (6\pi G\rho_c)^{1/3}\, t^{2/3} = \left(\frac{3t}{2t_H}\right)^{2/3} \qquad \Rightarrow \qquad \frac{t}{t_H} = \frac{2}{3}\left(\frac{1}{1+z}\right)^{3/2} \tag{13.10}$$

- $k = 0$, $\Omega_\Lambda = 1$: (flat, cosmological-constant dominated):

$$R = \exp\left((t - t_0)\sqrt{\frac{1}{3}\Lambda c^2}\right) = \exp\left(\frac{t - t_0}{t_H}\right) \tag{13.11}$$

- $k = 0$, $\Omega_M + \Omega_\Lambda = 1$ (flat universe with both matter and cosmological constant):

$$R = \left(\frac{\Omega_{M,0}}{\Omega_{\Lambda,0}}\right)^{1/3} \sinh^{2/3}\left(\frac{3t}{2t_{H,0}}\Omega_{\Lambda,0}^{1/2}\right), \tag{13.12}$$

  where the subscript 0 stands for current values.

Acceleration of the expansion

Friedmann's solution for Einstein's field equation actually contains two, not one equation. The other one is called the acceleration equation and concerns both the mass-energy ρ and the pressure P. It could be manipulated to yield

$$\frac{R(t)\ddot{R}(t)}{\dot{R}(t)^2} = -\frac{1}{2}\Omega_m(t) - \Omega_{\rm rad}(t) + \Omega_\Lambda(t)\,. \tag{13.13}$$

where matter is assumed to have no pressure ("dust"); radiation (or ER particles) satisfies $P_{\rm rad} = 1/3\rho_{\rm rad}c^2$; $P_\Lambda = -\rho_\Lambda c^2$, and $\rho_\Lambda = \Lambda c^2/(8\pi G)$ is a constant in time. A positive Λ energy density actually corresponds to a negative pressure, a result of ρ_Λ being a constant – as the universe expands, the total vacuum energy increases; however, if this corresponds to a positive pressure, it would lose energy by the PdV work it does.

---

[13] In case you want to derive this yourself: start with Eq. (13.7), subsitute $x^2 = R^3\Omega_{\Lambda,0}/\Omega_{M,0}$ and use that $d\,\text{arcsinh}\,x/d\,x = (1+x^2)^{-1/2}$.
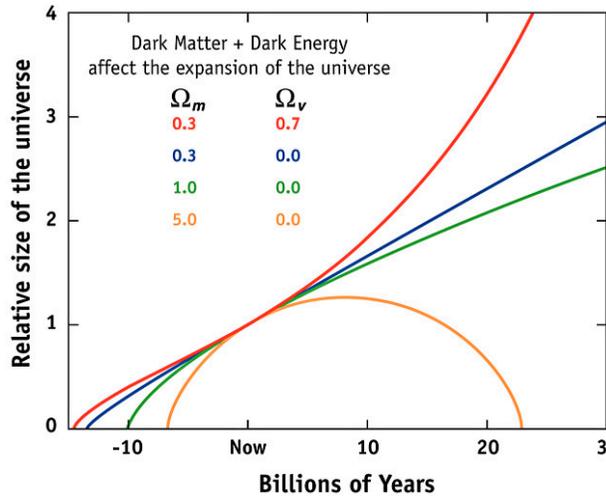
Figure 13.1: The evolution of the scale factor for different universe. Pay attention to the flat ones: green − matter dominated; red − matter + cosmological constant. The former is a power-law solution while the latter approaches an exponential solution at late times. So any universe with even a small amount of Ω_Λ to start with will end up having accelerating expansion. These models also differ in values for the current age.

Food for thought:

- Difference between Newtonian and GR description for the universe? what is the different meaning of k?

- What would dominate the universal expansion in the far future?

- How does one have to tune the various Ω_0 values to obtain: a collapsing universe; an accelerating universe; a closed universe?

- Can a universe be both open and collapsing?

- How does one go about measuring the different Ωs? are they degenerate or can we break the degeneracy?

- Among different flat universes, which one has the largest current age?

- If Λ is an energy density (which, one naively expects, draws the universe back in), how does it make the universe expand faster and faster? (see CO p. 1190-1192)

- Which Ω in eq. (13.7) is most relevant for very early universe? How does the universe expand then?

14 Three pillars of the Big Bang: I – Observational Cosmology

Readings: CO §*29.4,* Back

Our current picture of the universe have names like Big Bang cosmology, Λ-cold-dark-matter universe, etc. This is supported by three major pillars, observational cosmology, CMB and big-bang-nucleosynthesis.

14.1 Primer on Cosmological Distances

To use the measurements of a standard ruler/candle at a certain redshift to constrain cosmology models, we typically need to relate redshift (the observable) with a distance. However, because the universe is expanding and not necessarily flat, (large) distances can be tricky to define. A primer is provided in this paper by Hogg.

Co-moving distance

Recall that redshift is defined as $(1 + z) = R(t_0)/R(t_e)$, where t_0 is the time at observation (typically, now), and t_e time at emission. We take $R(t_o) = 1$ so $R(t_e) = R(z) = 1/(1 + z)$. We further define a function (following Peebles 1993)

$$E(R) = E(z) = \sqrt{\Omega_{M,0}/R^3 + \Omega_{k,0}/R^2 + \Omega_{\Lambda,0}} = \sqrt{\Omega_{M,0}(1 + z)^3 + \Omega_{k,0}(1 + z)^2 + \Omega_{\Lambda,0}}, \tag{14.1}$$

ostensibly ignoring contribution from radiation (since photons from radiation-domination eopch can't propagate information to us). According to the Friedmann equation (eq. [13.7]), the Hubble constant at redshift z is $H(z) = \dot{R}(z)/R(z) = H_o E(z)$.

Co-moving distance is the distance between us and another galaxy if some super-being is able to measure it INSTANTANEOUSLY NOW. Using light (which travels with a constant speed c) to do the same trick, however, one would have to discount the fact that the universe was smaller in the past,

$$D_c \equiv R(t_0) \int_{t_e}^{t_0} \frac{c\,dt}{R(t)} = \int_{R(t_e)}^{R(t_0)} \frac{c\,dR}{R\dot{R}} = \int_{R(t_e)}^{R(t_0)} \frac{c\,dR}{H_0 R^2 E(R)} = \frac{c}{H_0} \int_0^z \frac{dz'}{E(z')}. \tag{14.2}$$

This distance ought to remain constant for two galaxies that are co-moving with the Hubble flow. So it's a nice conserved quantity (e.g., the number of galaxies per comoving volumn is a useful index of galaxy frequencies that remains constant with redshift).

Using the fact that photon path is along geodesic ($ds = 0$) and radial ($d\theta = d\phi = 0$), this distance can also be calculated in the co-moving coordinates as (eq. [13.2])

$$D_c \equiv \int_{t_e}^{t_0} \frac{cdt}{R(t)} = \int_0^\varpi \frac{d\varpi'}{\sqrt{1 - k\varpi'^2}} = \begin{cases} \dfrac{1}{\sqrt{k}} \arcsin(\varpi\sqrt{k}) & (k > 0) \\ \varpi & (k = 0) \\ \dfrac{1}{\sqrt{|k|}} \operatorname{arcsinh}(\varpi\sqrt{|k|}) & (k < 0) \end{cases} \tag{14.3}$$

For two special cases, there are simple analytical expression for D_c at redshift z:

$$k = 0,\ \Omega_M = 1,\ \Omega_\Lambda = 0 \quad \Rightarrow \quad D_c = \frac{2c}{H_0}\left(\sqrt{R(t_0)} - \sqrt{R(t_e)}\right) = \frac{2c}{H_0}\left(1 - \sqrt{\frac{1}{1+z}}\right), \tag{14.4}$$

$$k = 0,\ \Omega_M = 0,\ \Omega_\Lambda = 1 \quad \Rightarrow \quad D_c = \frac{c}{H_0}\left(\frac{1}{R(t_e)} - \frac{1}{R(t_0)}\right) = \frac{cz}{H_0}. \tag{14.5}$$

But regardless of the cosmological model, for all cases, one should obtain $D_c \approx cz/H_0$, the Hubble's original law, in the limit $z \ll 1$.
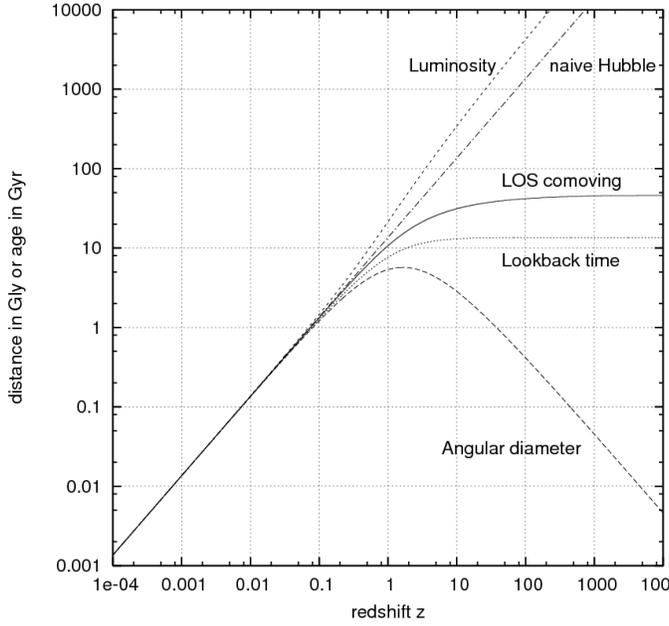
Figure 14.1: A comparison of cosmological distance measures, from redshift 0 to $10,000$. The background cosmology is $H_0 = 72$ km/s/Mpc, $\Omega_{\Lambda,0} = 0.732, \Omega_{M,0} = 0.266$, and $\Omega_{\text{tot}} = 1$. The luminosity distance, the angular diameter distance, etc. are defined in text. From Wiki page.

Horizon distance The maximum co-moving distance an observer (at time t) can see in the universe is defined by a photon that starts travelling toward us since the Big Bang ($z = \infty$):

$$D_{\text{h}}(t) = R(t) \int_0^t \frac{c\text{d}t'}{R(t')} \tag{14.6}$$

For a flat, matter-dominated Universe (i.e., $\Omega_{\text{M}} = 1$, $\Omega_\Lambda = 0$), one finds from eq. (14.5) that $D_{\text{h}}(t) = 2c/H = 3ct$ (using that $t = \frac{2}{3}t_H = 2/3H$). This is larger than ct – the photon appears to make more mileage in an expanding universe. In contrast, for a flat, Λ-dominated Universe, the horizon is at infinity (though also infinitely redshifted). This is because in this model the Universe was very close together for an infinitely long time.

The horizon distance matters for the causal connection between locations. We are not affected by any events beyond our current horizon distance. The universe beyond that can in principle be different. Similarly, two patches on the CMB that are further than the horizon distance at that point should not have causally communicated.

Angular diameter distance and 'standard rulers'

The angular diameter distance D_A is defined as the ratio of a rod's physical transverse size (ℓ) to its apparent angular size ($\Delta\theta$, in radians).

$$D_A \equiv \frac{\ell}{\Delta\theta} \,. \tag{14.7}$$

Consider this situation in the co-moving coordinates: let the rod be at co-moving distance D_c (eq. [14.3]) from us and extends an angle $\Delta\theta$. The physical size $\ell = R(t_e)D_c\Delta\theta$. As a result,

$$D_A = D_c/(1+z) \,. \tag{14.8}$$

So for a Milky Way galaxy with size ℓ but placed at different redshift, assuming no evolution in size, its angular extent goes as

$$k = 0, \ \Omega_{\text{M}} = 1, \ \Omega_\Lambda = 0 \quad \Rightarrow \quad \Delta\theta \ = \ \frac{H_0\ell(1+z)}{2c\left(1 - \sqrt{1/(1+z)}\right)}, \tag{14.9}$$

$$k = 0, \ \Omega_{\text{M}} = 0, \ \Omega_\Lambda = 1 \quad \Rightarrow \quad \Delta\theta \ = \ \frac{H_0\ell(1+z)}{cz} \,. \tag{14.10}$$
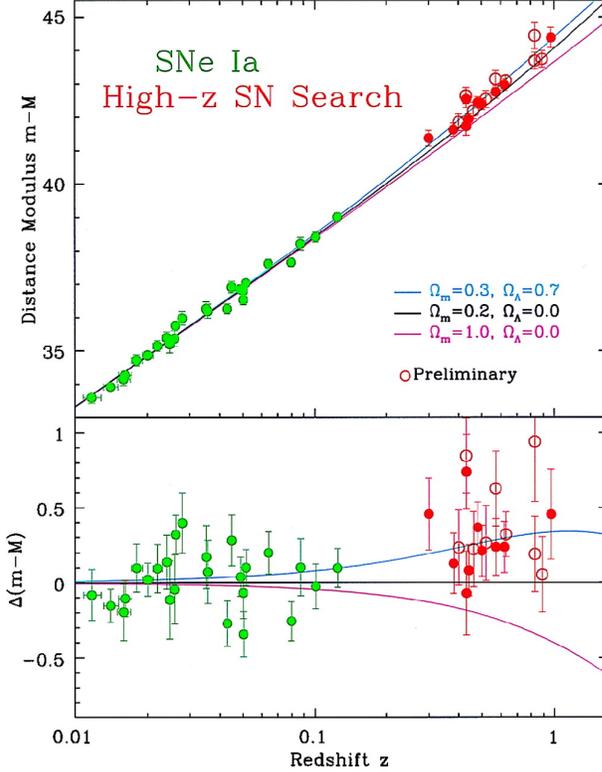
Figure 14.2: Luminosity distances (parameterized here by magnitudes) of supernova vs their redshifts, superimposed on various cosmological models. The high redshift ones are too dim (too large luminosity distances) to be explained by any model without a cosmological constant. These data have been supported by more recent SN observations (see Fig. 14.4). Low panel shows that different cosmologies only change apparent magnitudes by ~ 0.3 at $z = 1$ – a strong requirement on the 'standard candle' calibration. From the review by Kirshner (1999).

In either case, the angular extent does not go to zero as redshift is increased; in particular, for matter-dominated universe at $z \geq 1$, more distant objects actually appear larger in angular size, a bizarre effect of the universal expansion.

Luminosity distance and 'standard candles'

The luminosity distance D_L is defined by the ratio of bolometric (ie, integrated over all frequencies) flux F and bolometric luminosity L:

$$D_L \equiv \sqrt{\frac{L}{4\pi F}}. \tag{14.11}$$

Consider photons propagating isotropically outward and passing through a spherical surface of area A at time t_0. The Robertson-Walker metric (eq. [13.2]) is sufficiently similar to that of a spherical coordinates (it only differs in the radial direction but identical in the angular directions), we can analogously write down the spherical area as $A = 4\pi D_c^2 R(t_0)^2$. However, there are two effects on photons that affect its flux at t_0: photons energy decreases as $h\nu_0 = h\nu_e/(1+z)$; photons arrive less frequently due to the universal stretching, $dt_0 = dt_e(1+z)$.

Combining all, we have

$$D_L = R(t_0)D_c(1+z) = D_c(1+z). \tag{14.12}$$

14.2 Observables

Supernova Type Ia Perhaps the most celebrated method of studying the universe, other than CMB. This is where the first evidence for Ω_Λ was discovered (Nobel prize in physics, 2011, Fig. 14.2). The result of the thermonuclear explosion on an accreting CO white dwarf star approaching the Chandrasekhar mass limit, SN Ia are standardizable candles which explode with nearly the same brightness everywhere in the universe due to the uniformity of the triggering mass and hence the available nuclear fuel. Empirically (but not yet understood), there is a scatter in their intrinsic brightness but it appears
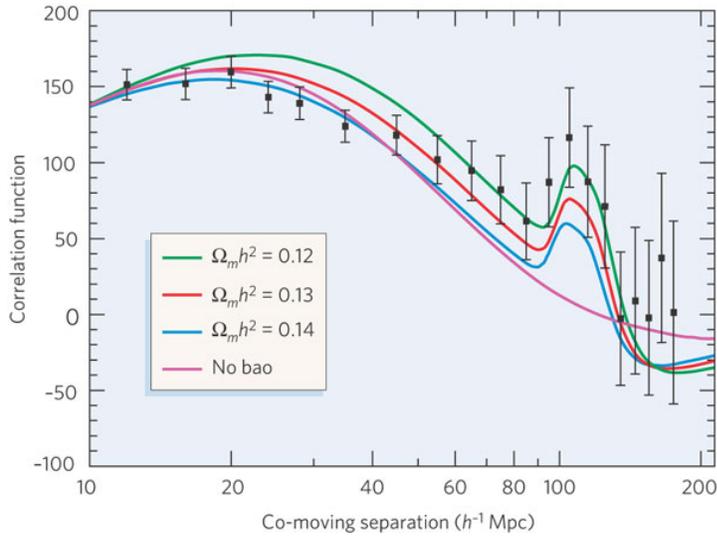
Figure 14.3: The galaxy redshiftspace correlation function reveals baryon acoustic oscillations. The data are from red luminous galaxies in the 3816 square degree Sloan Digital Sky Survey (SDSS). The baryon oscillation signal visible at $\sim 100h^{-1}$ Mpc is detected at 3.4σ. The green, red and blue models have $\Omega_m h^2 = 0.12, 0.13$ and 0.14, respectively. The magenta curve is a CDM model with no baryon oscillations. The acoustic peak gives the ratio of the distances to $z = 0.35$ and $z = 1,100$ to 4% fractional accuracy. The absolute distance to z=0.35 is determined to 5% accuracy. The full correlation function shape gives $\Omega_m h^2$ to 8% and agrees with the value from the CMB. The co-moving sound horizon scale is $\sim 150h^{-1}$ Mpc, similar to the first peak of CMB at $\ell \sim 200$. From the short review by Bennett (2006). Here, h is the standard notation, $h = H_0/(100\text{km/s/Mpc}) \sim 0.7$.
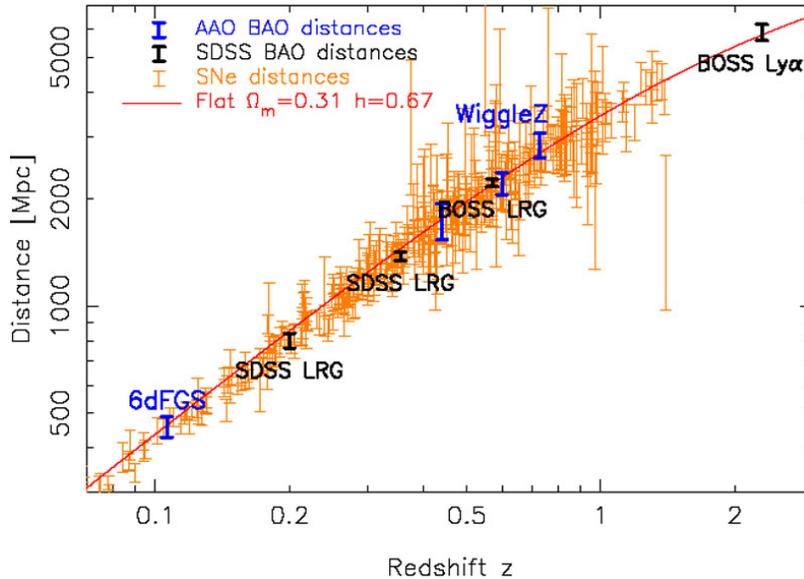

Figure 14.4: The cosmic distance scale with redshift. This modern version of the 'Hubble Diagram' combines data from SN Ia as standard candles and BAO as standard rulers in the LRG SSDS, BOSS, 6dFGRS, and WiggleZ galaxy surveys and from the BOSS Lyman-alpha at high redshift. Taken from Lahav & Liddle (2014).

that slower SNIa are brighter and bluer, so one can correct for this scatter. SNIa then yield luminosity distances as a function of redshifts (eq. [14.11]).

Besides from inferring Ω_Λ, one can combine SNIa with other constraints to infer the equation of state for dark energy, $p_\Lambda = w\rho_\Lambda$. It appears that $w = -1.06 \pm 0.07$ (Sullivan et al 2011, CFHT supernova legacy survey, 472 SNIa), consistent with a cosmological *constant*.

BAO Acoustic oscillations in the baryons is the counterpart to CMB acoustic oscillations, and they show up in the redshift clustering (clustering in velocity space) of galaxies and clusters. Similar to the fact that CMB peaks are standard rulers at redshift $z \sim 1100$, the peaks in BAO provide standard rulers at a much lower redshift ($z \leq 2$). As such, it is mildly sensitive to the curvature of the universe.

The first BAO peak was first detected in 2005 and is now a main-stay of modern cosmology. It provides a particularly sensitive measurement of the matter density (Fig. 14.3). Differing from CMB however, since matter does not resist clumping as radiation does (practically pressure-less), small scale power in BAO have gone highly nonlinear with time. So the power spectrum at higher ℓ are dominated by nonlinear effect, but going to higher redshift should help alleviating this (e.g., studying high redshift Ly-α clouds, Fig. 14.4).
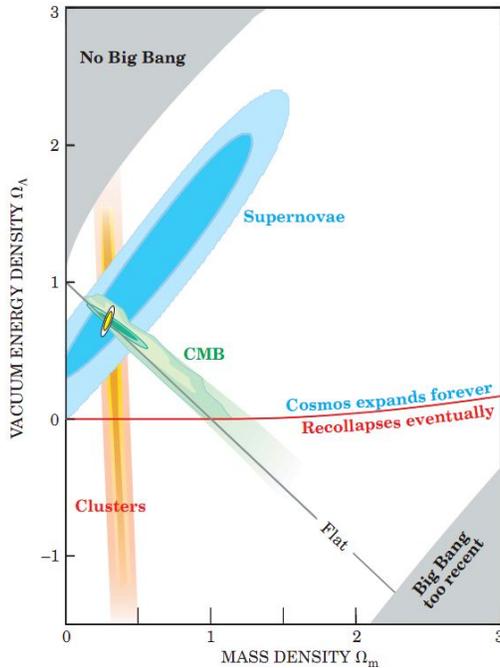
Figure 14.5: Key figure of this chapter. Three independent sets of measurements – high redshift supernova, galaxy cluster inventories, and the cosmic microwave background – converge nicely near $\Omega_{\mathrm{m}} \sim 0.3$ and $\Omega_{\Lambda} \sim 0.7$. The inflationary expectation of $\Omega_{\mathrm{m}} + \Omega_{\Lambda} = 1$ is indicated by the black diagonal line. The little yellow ovals indicate the expected constraints by the originally proposed SNAP satellite (now superceded by WFIRST). The red curve separates an eternally expanding cosmos from one that ends in a "Big Crunch". Credit Saul Perlmutter.

Weak gravitational lensing Images of background galaxies can be (minutely) distorted as their light passes through the gravitational potential of foreground galaxies (galaxy lensing), clusters of galaxies (cluster lensing) and the ubiquitous large scale inhomogeneities (cosmic shear). This can be understood as distortions of the local geodesics (away from the Robertson-Walker metrics) by density perturbations. The magnitude and direction of the distortion measures the integrated mass distribution along the line of sight. So by measuring correlations between shapes of these galaxies, lensing allows one to constrain the matter content in the universe (Fig. **??**). More specifically, it measures the product of $\sigma_8 \Omega_M$. First detected in the early 2000s, this technique is now a strong contender for future missions.

They are nearly orthogonal

Cosmological parameters have now been reasonably well measured, after the past two decades' of efforts – it is remarkable that we converged on the fate of the universe in two short decades.

This is to some degree thanks to the orthogonalities that exist between various measurements. As is shown in Fig. 14.5, the uncertainties in different measurements are largely orthogonal to each other and to that of CMB.

- Cluster abundances (as measured by various techniques, including gravitational lensing) are most sensitive to the matter content, $\Omega_{M,0}$. So it is a nearly vertical line in that figure.

- SN Ia (or other standard candles) are most sensitive to the acceleration of the universe, and therefore to $\Omega_{\Lambda,0} - \Omega_{M,0}/2$ (see eq. 13.13).

- CMB, BAO and other 'standard candles' are most sensitive to the curvature of the universe, or $\Omega_{M,0} + \Omega_{\Lambda,0}$ (see p. 1186-1187 of CO).

14.3   So what is our universe?

Our universe does appear to be homogeneous and isotropic on large scales. It is expanding and has a finite age. The currently favoured model is called the 'ΛCDM' model (CDM stands for cold dark matter), where

- matter (baryon and dark matter) makes up some 30% of the critical density; most of this is dark matter. The dark matter is likely 'cold', or, has a smaller velocity dispersion in the early universe. This is necessary to explain the gravitational structures in the universe (galaxies, clusters, superclusters etc).

- dark energy (the so-called 'Λ') comprises some 70% of the critical density; this is first observed in the expansion history using supernova Ia, but now also confirmed by BAO. Universal expansion has been accelerating over the past 5 Gyrs.

- combing matter and dark energy, the universe appears to be flat. In concordance, the CMB measurements have returned a nearly zero curvature. The universal expansion appears to be eternal and accelerating.

At the moment, there are many candidates for dark matter (elementary particles that are non-baryonic, non-dissipative, un-charged), e.g., axions. Intensive searches are being carried out. We have little idea about the dark energy, on the other hand. It should have an exotic equation of state that provides repulsive gravity. An alternative theory to Einstein's GR remains viable.

Food for thought

- Why are there different 'distances' in an expanding universe?

- Think what could go wrong with the SN Ia method.

- What could be used as standard rulers? standard candles?

- How do you explain the fact that angular size for the Milky Way galaxy actually increases with z as it passes $z = 1$?

- In Fig. 14.1, the comoving distance is not increasing with distance when $z \gg 1$. How do you interpret this?

- In Fig. 14.5, cluster inventories give a nearly vertical error ellipse, meaning it is very good at measuring $\Omega_{M,0}$, but not sensitive to $\Omega_{\Lambda,0}$. Why?

- Why do we need high redshift supernova Ia to measure Ω_Λ?

- What is BAO and how does it give us information about the universe?

- CMB is called 'standard ruler' and is used to measure curvature of the universe. How is this done? Why is the CMB contour in Fig. 14.5 slanted in the way it does?

- What is the fate of our universe? how does it expand in the far future? is the Hubble constant increasing or decreasing in the future?

15 the Hot Big Bang

Readings: CO §29.2, up to p. 1176 Back

The early universe must have been dense, as mass is conserved during expansion. Let us also **assume** that the early universe was **hot**. This is the hot-big-bang model. This model is able to explain a slew of observations and has therefore become the 'standard cosmology'.

Radiation dominated period

For radiation, the equivalent density is $\rho_{\text{rad}} = aT^4/c^2$. The blackbody radiation has a temperature that cools as (eq. [13.9])

$$T = \frac{T_0}{R} = (1+z)T_0 \,. \tag{15.1}$$

From CMB measurements, currently $T_0 \approx 2.7\,\text{K}$ and radiation contains a negligible amount of energy density ($\Omega_{\text{rad}} \sim 10^{-4}$). But at some far enough past, because $\rho_{\text{rad}} \propto R^{-4}$, radiation dominates over matter. The transition occurs when

$$\rho_0(1+z)^3 = \frac{aT_0^4}{c^2}(1+z)^4 \quad \Rightarrow \quad 1+z = \frac{3\Omega_{\text{m,0}}H_0^2 c^2}{8\pi G a T_0^4} \simeq 2.2 \times 10^4 \Omega_{\text{m,0}} \left(\frac{H_0}{75\,\text{km}\,\text{s}^{-1}\,\text{Mpc}^{-1}}\right)^2 , \tag{15.2}$$

or, $z \approx 4000$. Here, $\Omega_{\text{m,0}} \sim 0.3$ is the current mass density (including baryons and dark matter). All other terms in eq.(13.7) are unimportant except for radiation, and

$$\frac{\dot{R}}{R} = \left(\frac{8\pi G}{3}\frac{aT^4}{c^2}\right)^{1/2} . \tag{15.3}$$

This yields $R(t) \propto t^{1/2}$. Contrast this with the matter-dominated expansion of $R(t) \propto t^{2/3}$. Inserting eq. (15.1), we get

$$T(t) = \left(\frac{3c^2}{32\pi Ga}\right)^{1/4} t^{-1/2} \sim 10^{10}\text{K} \left(\frac{t}{1\text{s}}\right)^{-1/2} . \tag{15.4}$$

So the universe was hotter before. Moreover, since matter is in near thermal equilibrium with radiation (very small mean-free-path), matter was also hotter in the past. This can have a range of ramifications. For our discussion below, we rewrite the above equation in energy unit,

$$kT(t) \approx 1\text{MeV}\left(\frac{t}{1\text{s}}\right)^{-1/2} . \tag{15.5}$$

The above description where photons are the only energy carrier is simplified. At earlier times, there are other relativistic particles, not just photons, which have significant energy density. These include neutrinos/anti-neutrinos, electron/positron, and more particles at even earlier times.

For each neutrino family (including neutrino and its anti-particle), the energy density is $u_{\nu,\bar{\nu}} = \frac{7}{8}aT^4$; relativistic electrons or positrons (which have two spin states) have $u_{e+} = u_{e-} = \frac{7}{8}aT^4$. The factor 7/8 differs from photon (factor 1) because these particles are fermions and satisfy a different statistical distribution.

These contribution can be included by substituting a by $a' = a(1 + \frac{7}{8}[N_\nu + 2])$ in eq. (15.3), where $N_\nu = 3$ is the number of known neutrino families; one can include more relativistic particles as they arise at earlier times.

Particle Freeze-out

The universe expands and dilutes. Interaction timescale gets longer and longer. Mean-free-path $l_{\text{mfp}} \approx 1/n\sigma - n$ drops with time, and the cross-section σ, usually a function of energy, also drops with time.
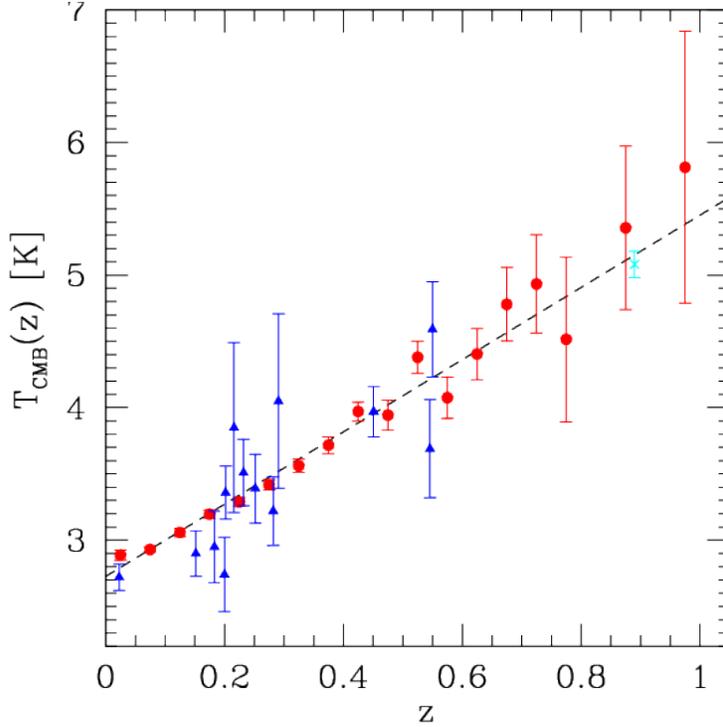
Figure 15.1: Measured redshift evolution of the CMB temperature as compared to the adiabatic prediction (dashed line): $T_{\mathrm{CMB}} \propto (1+z)$. Red circles indicate the Sunyaev-Zeldovich Effect measurements by stacking a sample of clusters detected by Planck; blue triangles show the Sunyaev-Zeldovich Effect measurements for individual clusters; a cyan cross is the measurement using molecular absorption lines. The universe was hotter in the past. Taken from Kitayama 2014. Recently such measurements have been extended to redshift up to 3.5 (see Hurier et al, 2014).

Going back in time, "Freeze-out" occurs whenever mean-free-time for a given particle interacting with the others becomes longer than the lifetime of the universe at that moment, $n\sigma v \leq H$. After this time, they are no longer in thermal equilibrium with the others (same temperature) and instead cool down independently. These 'relics' therefore allow us to probe the condition at their decoupling. Two examples: radiation-matter decoupling, 'relic' being the cosmic microwave background; neutrino-matter decoupling, 'relic' being the cosmic neutrino background.

Going back in time

Taking $T_0 = 2.7 \, \mathrm{K}$ and the current matter density measurements, tracing back in time, here is a physical picture of what must have happened.

- Above $T \approx 3000 \, \mathrm{K}$ ($z \sim 1100, t \sim 3 \times 10^5$ yrs), hydrogen becomes fully ionized (the average photon energy is only ~ 0.3 eV). Electron scattering generates a huge amount of opacity, allowing matter and radiation to equilibrate thermally. The decouping is called the 'CMB', or 'the last scattering surface', or the 'era of recombination', or 'photon decoupling'.

  For simplicity, assume pure Hydrogen. Then, from the Saha equation (Eq. 4.7), scaling to $n_{\mathrm{B}} = \rho_{\mathrm{B},0}(1+z)^3/m_{\mathrm{H}}$ where $\rho_{\mathrm{B},0}$ is the current density of baryons (only a small part of all matter), and writing in terms of the ionized fraction $x \equiv n_{\mathrm{p}}/n_{\mathrm{B}}$,

$$\frac{x^2}{1-x} = \frac{n_p n_e}{n_{H^0} n_{\mathrm{B}}} = \frac{m_{\mathrm{H}}}{\rho_{\mathrm{B},0}(1+z)^3} \frac{(2\pi m_{\mathrm{e}} k T_0(1+z))^{3/2}}{h^3} \, \mathrm{e}^{-\chi/kT_0(1+z)} \, , \tag{15.6}$$

  where n_{H^0} is the number density of neutral hydrogen. Solving this for $\Omega_{\mathrm{B},0} \approx 0.04$ and $x = 0.5$, gives $1+z \approx 10^3$. In reality, recombination will be a little delayed as photons produced in the recombination can excite and ionize other atoms: current estimates yield $1+z \approx 1090$ for the typical redshift at the time a CMB photon last scattered. Moreover, this "surface of last scattering" has a finite thickness $\Delta z \approx 200$. After decoupling, matter is no longer tied to the radiation field and can collapse under gravitational instability.

- Above $T \approx 10^9 \, \text{K}$ ($z \sim 4 \times 10^8, t \sim 3$ mins), average photon energy ≥ 100 keV and becomes comparable to the typical nucleus binding energy of 1 MeV (Fig. 6.1). This is the era of big-bang nucleosynthesis, the 'first 3 minutes' in which the primordial abundances are set in stone.

- Above $T \sim 5 \times 10^9 \, \text{K}$ (~ 500 keV $\approx m_e c^2$, $t \sim 1.3 \, \text{s}$), electrons/positrons are relativistic and behave like photons. The energy density is now shared equitably between photons (with two polarization states), electrons (2 spin states) and positrons (2 spin states). So energy density $u = a'T^4$ with $a' = 11/4 \, a$. The electron-positron pairs are produced freely as they are in thermal equilibrium with photons.

- Above $T \sim 10^{10} \, \text{K}$ (~ 1 MeV, $t \sim 0.5 \, \text{s}$), neutrinos (which only interact via weak interactions) are strongly coupled to the other particles (the interaction rate $n\sigma v > 1/t$). As a result, they participate in the thermal bath ($T_\nu = T_\gamma = T$). Conversely, later than this point, neutrinos start to "freeze out" and evolve independently of the radiation background ("relic"), thereby providing a measurable entity independent of CMB. The cosmic neutrino background is yet to be discovered.

- Depending on the nature of interaction for dark matter particles, they may freeze-out around 1 MeV (their rest mass energy) and produce the observed dark matter relic density $\Omega_{\text{dm},0} = \Omega_{\text{m},0} - \Omega_{\text{B},0}$. These so-called 'cold dark matter' have mass-energy of order MeV (mc^2), much greater than their kinetic energy (they can be bound by galaxy potential, so kinetic energy not more than $\frac{1}{2}m(200 \, \text{km/s})^2$).

- Earlier in time, we observe in term quark-hadron phase transition (~ 1 GeV, where quarks are no longer absorbed into hadrons but make a quark soup), electro-weak phase transition (~ 100 GeV, above which electromagnetic force and weak nuclear force are unified), grand-unification transition (GUT, $\sim 10^{15}$ GeV, where all forces unify into one), and the Planck epoch. The physical laws that applies during the last two epochs still elude us. Baryogenesis (where an asymmetry between matter and anti-matter allows more matter to be produced in our universe) may have occurred somewhere there. So may have inflation. Above 100 GeV, all known particles are relativistic.

Inflation

There are two major issues with the universe we know. The smoothness issue: the CMB radiation is too smooth on scales much larger than the horizon scale at recombination. The flatness problem: the present observed density of the universe being of order unity means this density must be arbitrarily close to unity at earlier times. Who allows communication between different causally connected regions in the universe? and who fine-tunes the density? Both these problems, and more, are "solved" in the inflation hypothesis, in which the universe was assumed to have undergone a short period of exponential expansion, perhaps around the time of GUT transition. But since the physics is not well understood at that stage, the inflationary theory is still somewhat ad-hoc at the moment. However, it is a powerful and imaginative new front for cosmology. (reading: CO §30.1).

Aside from solving the above philosophical questions, inflation is also important for later evolution of the universe. Before inflation, quantum fluctuations have planted seeds of density inhomogeneities in extremely small scales (Planck scale). The huge exponential expanion during inflation stretches these tiny regions to observable sizes (as seen on CMB) and these later undergo gravitational instability to make the galaxies, galaxy clusters we see today.

Food for thought:

- The radiation energy density of the universe varies as R^{-4}, while the mass density varies as R^{-3}. Why the difference?

- Why does the universe recombine at $T \approx 3000\,\mathrm{K}$, as opposed to the $T \approx 10,000\,\mathrm{K}$ for stars?

- What is preventing us from seeing the 'big bang' directly as it happens? if we can't see directly, what can we instead observe?

- What does 'freeze-out' mean?

- one of the freeze-out is matter-radiation decoupling at recombination. What is the impact on matter and radiation separately?

- choose another freeze-out event and think about it. What are the impacts?

- Could the hot big-bang hypothesis be wrong, given current data?

- Was the universe ever degenerate? any foreseeable consequences?

- Inflation is supposed to have occurred around a time of $t \sim 10^{-32}\,\mathrm{s}$. To make sure that the entire universe today is in causal contact back at that time, by what factor inflation must have blown up the scale factor of the universe?
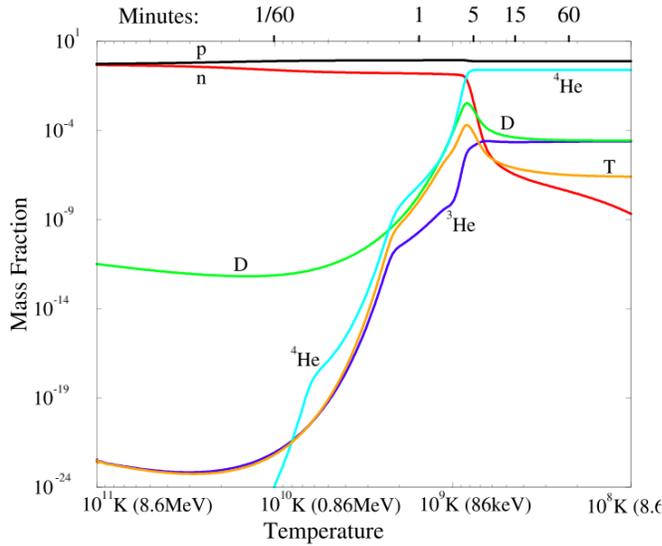
Figure 16.1: Big bang nucleosynthesis: light element abundances (mass fractions, with D for deuterium and T for tritium) as a function of time (upper scale) or temperature (lower scale) in the early uinverse, for a case with $\Omega_{B,0} = 0.05$ and $H_0 = 75$ km s^1 Mpc^1, i.e., the baryon-to-photon ratio $\eta = 7.5 \times 10^{-10}$, just slightly higher than the current best value of $\eta \approx 6 \times 10^{-10}$. Note the sharp peak in the D abundance at $t \approx 4$ minutes ($T \approx 9 \times 10^8$ K), corresponding to the time when the ^{4}He abundance rises to its final value. From Mukhanov (2003), 'Nucleosynthesis without computer'

16 The three pillars: II - big bang nucleosynthesis

Readings: CO P. 1177 − 1181, Back

The typical binding energy of a nuclei is ~ 1MeV. Thus, as the universe cools below that value ($T \sim 10^{10}$ K, $t \sim 1$ s), complex nuclei (made up of protons and neutrons) could now survive the bombardment by energetic particles. From this to $t \sim 3$ minutes, it is the era of big bang nucleosynthesis. Nuclear reactions proceed in earnst and produce light elements, leading to an early universe that is made up mostly of hydrogen, a quarter helium, and trace amounts of deuterium (^{2}H, or D), tritium (^{3}H, or T, unstable), helium-3 (^{3}He), berylium (^{7}Be, unstable) and lithium (^{7}Li). Essentially all heavier elements are made subsequently in stars. The abundances of these primordial elements constrain history during the first 3-minutes. BBN is another success story of the hot-big-bang model.

16.1 The First 3-minutes

Ratio of baryons to photons This turns out to be the single number that determines the bbn yield. The number density of photons is given by,

$$n_\gamma = \int_0^\infty \frac{4\pi\nu^2 \mathrm{d}\nu}{c^3} \frac{2}{\mathrm{e}^{h\nu/kT} - 1} = 8\pi \left(\frac{kT}{hc}\right)^3 \int_0^\infty \frac{x^2 \mathrm{d}x}{\mathrm{e}^x - 1} = 4.2 \times 10^2 \text{ cm}^{-3} \left(\frac{T}{2.736 \text{ K}}\right)^3, \quad (16.1)$$

where we used $\int_0^\infty [x^2/(\mathrm{e}^x - 1)]\,\mathrm{d}x = 2\zeta(3) = 2.404$. Note that going backward in time, unless there is some unknown entropy production post BBN, $n_\gamma \propto T^3 \propto R^{-3}$, just like matter does.

The present baryon number density is

$$n_{\mathrm{B},0} = \frac{\Omega_{\mathrm{B},0}\,\rho_{\mathrm{c},0}}{m_{\mathrm{H}}} = 6.3 \times 10^{-6} \text{ cm}^{-3}\,\Omega_{\mathrm{B},0} \left(\frac{H_0}{75\,\mathrm{km\,s^{-1}\,Mpc^{-1}}}\right)^2. \quad (16.2)$$

Thus, the (present and past) ratio of baryons to photons (valid after e^+e^- annihilation) is

$$\eta \equiv \frac{n_{\mathrm{B},0}}{n_{\gamma,0}} \simeq 1.5 \times 10^{-8}\,\Omega_{\mathrm{B},0} \left(\frac{H_0}{75\,\mathrm{km\,s^{-1}\,Mpc^{-1}}}\right)^2, \quad (16.3)$$

or, there are about a billion photons for every baryon in the universe. This high entropy state produces unique BBN products which then freeze out and become a relic record of that environment. It is remarkable such a relic record is all around us (yes, even inside your body).

Neutron to proton ratio At high T, the reactions changing proton into neutrons and vice versa proceed quickly, and hence the number densities follow their equilibrium ratio:

$$\frac{n_\mathrm{n}}{n_\mathrm{p}} = \mathrm{e}^{-Q/kT} \qquad \text{with} \qquad Q = (m_\mathrm{n} - m_\mathrm{p})c^2 = 1.293 \text{ MeV}. \tag{16.4}$$

This ratio will 'freeze out' at $T \simeq 10^{10}$ K, when the reaction speeds $(n\sigma v)$ become too slow compared to cosmic expansion $(1/H)$. This leaves an initial combo of $n/p \sim 1/6$. Free neutron β-decay is slow (half-life ~ 15 minutes) so is somewhat negligible for the first 3 minutes.

Deuterium formation Protons can fuse with the free neutrons to form Deuterium. Deuterium is formed and broken apart by the reaction

$$\mathrm{p} + \mathrm{n} \leftrightarrow {}^2\mathrm{H} + \gamma \,. \tag{16.5}$$

In equilibrium, the abundance can be described by the Saha equation,

$$\frac{n_\mathrm{p} n_\mathrm{n}}{n_\mathrm{D}} = \frac{g_\mathrm{p} g_\mathrm{n}}{g_\mathrm{D}} \frac{(2\pi kT)^{3/2}}{h^3} \left(\frac{m_\mathrm{p} m_\mathrm{n}}{m_\mathrm{D}}\right)^{3/2} \mathrm{e}^{-B/kT} \,, \quad \text{with} \quad B = (m_\mathrm{p} + m_\mathrm{n} - m_\mathrm{D})c^2 = 2.225 \text{ MeV}. \tag{16.6}$$

The statistical weights are $g_\mathrm{p} = g_\mathrm{n} = 2$, and $g_\mathrm{D} = 3$. Just like Hydrogen ionization proceeds at 10^4 K (not 10^5 K, or blackbody peak ~ 10eV), with a binding energy of 2.225 MeV, Deuterium formation only begins in earnest when temperature has dropped to $T \sim 10^9$ K (blackbody peak $\sim 10^5$ eV) and there aren't too many highly energetic photons to break it apart. By now ($t \sim 2$ minutes), free neutron decay has brought down a bit $n_n/n_p \sim 1/7$, but all free neutrons are essentially absorbed into D.

Formation of light elements Once Deuterium becomes present in significant abundances, it is rapidly burned to Helium which has a binding energy of 28.3 MeV, funnelling all neutrons to ${}^4\mathrm{He}$. One then expects a primordial He mass fraction to be $Y \sim 24\%$ (or $\sim 8\%$ in number), weakly dependant on the value of η.

A few other light elements are also formed similarly from D, via a sequence of 2-body interactions (none of which involves weak interactions which are too slow)
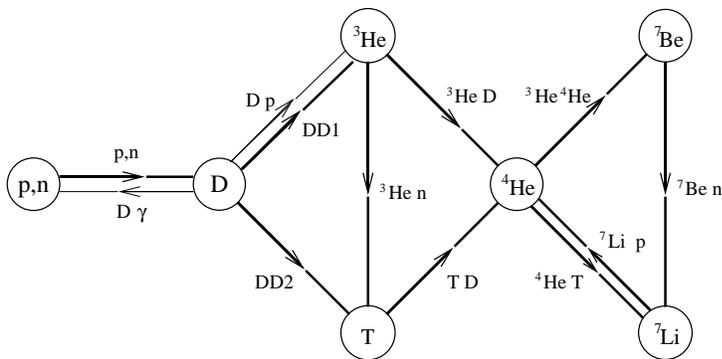
Much later, ${}^7\mathrm{Be}$ has a half-life of 53.28 days toward electron capture and is converted to ${}^7\mathrm{Li}$; left-over tritium also decay into ${}^3\mathrm{He}$ with a half-life 12 yrs. Together with left-over D, we have now ${}^3\mathrm{He}$, ${}^4\mathrm{He}$ and ${}^7\mathrm{Li}$ – the vast majority being ${}^4\mathrm{He}$ – the only nucleosynthesis products from the early universe. Elements above atomic number 8 (the number-8 gap) cannot be made – stars can do it by triple-α reaction, but these 3-body reactions are too slow at these densities and timescales.

Dependency on η The final ${}^4\mathrm{He}$ abundance mostly depends on the ratio of n_n/n_p at D formation, and is fairly independent of η. However, the other trace elements are strongly affected: if n_B is *very* small (small η), then all reactions are too slow, and little of these elements form; if n_B is very big (large η), one still has to wait until the temperature is low enough that D can form. Because the $D + D$ and $D + p$ reactions are faster at high density, less D is left at the end (see Fig. 16.2). Conversely, D abundance is a sensitive probe of the baryon density. ${}^3\mathrm{He}$ and ${}^7\mathrm{Li}$ can also be useful (Fig. 16.2).
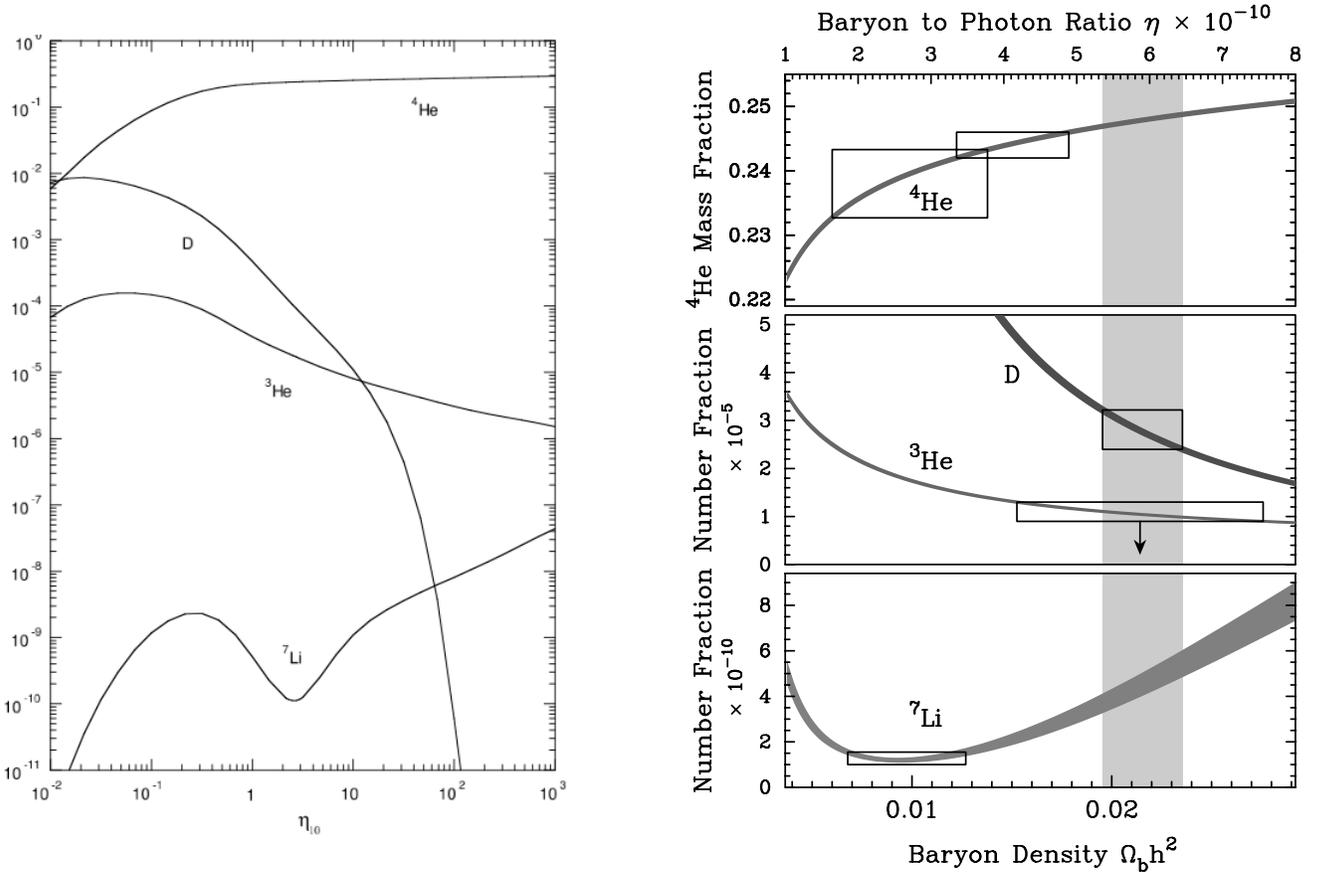
Figure 16.2: *(Left)*: Expected final abundances (mass fractions) for the light elements from big bang nucleosynthesis, as a function of baryon-to-photon ratio η (in units of 10^{10}: $\eta_{10} \equiv \eta/10^{-10}$). From Mukhanov (2003) Fig. 3. *(Right)*: Comparison of predicted and measured abundances of four light nuclei as a function of the baryon density. The figure has three vertical panels each with a different linear scale. The curves show the abundance ratios predicted for BBN, with the top panel for the mass fraction of ^{4}He, the middle and lower panels for the number fractions of D/H, ^{3}He/H ad ^{7}Li/H. The vertical widths of the curves show the uncertainties in the predictions. The five boxes show measurements, with the vertical extension showing the $1 - \sigma$ random error, and the horizontal range is adjusted to overlap with the prediction curves. Not all the data boxes overlap the vertical band that indicates the 2013 CMB (Planck) result of $\Omega_b h^2 = 0.02222 \pm 0.00023$ (the band, made in 2003, is much wider than this new error-bar), especially ^{4}He and ^{7}Li. The authors suggest that this is probably due to systematic effects. From Kirkman et al (2003).

.

16.2 Observations of primordial light elements

The relic abundances of the light elements synthesized in BBN result from the competition of nuclear reactions against universal expansion. As such, they provide independent measures of $\Omega_b h^{-2}$. From CMB (Planck), $\Omega_b h^{-2} = 0.0222 \pm 0.002$. How do the light element abundances compare?

Primordial Helium The ^{4}He mass fraction in the present universe has received contribution from stars. So one should instead look in low-metallicity, extragalactic HII regions (in, e.g., HeI emission lines in blue compact galaxies). If Helium is primordial, one expects the measured Y value is independent of metallicity. However, to obtain Helium abundance, one needs to conduct photoionization models of the HII regions, which can be highly inhomogeneous in temperature and density. This may introduce systematic errors. The current result (depending on whom you ask) is Helium mass fraction $Y = 0.24 \pm 0.01$, a few-σ away from the CMB value.

Primordial Deuterium Deuterium can fuse (at $T \sim 10^6$ K) in pre-main-sequence stars and brown dwarfs, but it cannot be generated. So any Deuterium still lying around (including those in the ocean and in your body) is thought to be primordial and is a lower-bound to the bbn production.
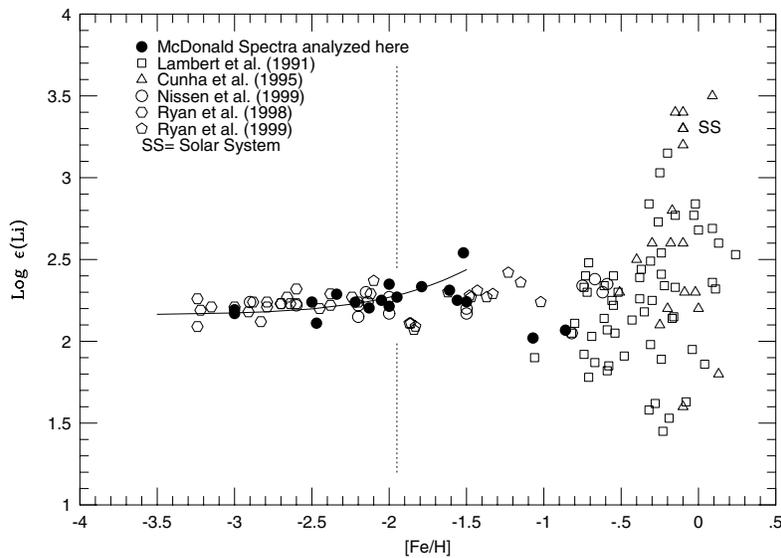
Figure 16.3: Lithium number abundances, $\log \epsilon(Li) \equiv 12 + \log(Li/H)$, in different populations of stasr, are plotted versus stellar metallicties. The solar system is marked out by 'SS'. The solid line is intended to guide the eye to the "Spite Plateau", the flatness of Lithium abundances in metal poor stars indicating that the bulk of their lithium is perhaps unrelated to Galactic nucleosynthesis processes and is instead primordial. More metal rich stars have a spread of Lithium abundances, likely a result of the continuous production and destruction of Lithium during stellar evolution. From the short review by Steigman (2006).

Since Deuterium is destroyed in pre-main sequence stars, and it is also destroyed by cosmic-ray spallation, one expects the cosmic Deuterium abundance to decrease with time. It is therefore surprising that the measured ISM value is not too different from the prediction from big-bang nucleosynthesis (2.7×10^{-5}).

In the Galaxy, the interstellar medium has a Deuterium number abundance of $D/H \sim (2.8 \pm 0.5) \times 10^{-5}$,[14] Studies of quasar absorption line systems at high redshift, despite the small wavelength shift between DI and HI, are able to establish Deuterium abundance in these low metallicity clouds. The results, however, carry a scatter of two orders of magnitude, with the average consistent with CMB value. A similar story runs for ^{3}He.

Primordial Lithium Lithium has a more complicated genesis. While also easily fused inside stars ($T \sim 2.5 \times 10^6$ K), ^{7}Li can also be continuously produced by cosmic ray spallation. And some red giants appear to be Lithium rich, indicating some fusion production (not just pure destruction) inside stars. These new production is consistent with the observed rise of Lithium with stellar metallicity. This can explain why the measured Lithium abundances in metal poor stars exhibit a flat value (the primordial value), while those in metal rich stars have a large scatter (Fig. 16.3). The trouble is, the plateau value is about twice lower than that required by CMB baryon density. This tension is currently unresolved. There could be either unknown physics operating, or larger measurement uncertainties than we allow.

Food for thought:

- What is the difference between p-p chain (§6) and the light-element nucleosynthesis here? why is there no p-p chain in early universe? (hint: BBN doesn't need weak interaction, why not?)

- why did the universe contain more protons than neutrons at early times?

- How is the cosmic density at 3 minutes compared to that of the solar center? what about temperature?

- Why are the baryons mostly H/He rich? why can't BBN make heavier nuclei like carbon and oxygen?

- Why is Helium production only sensitive to proton-to-neutron ratio, but less so on photon number?

- Why is Deuterium abundance decreasing with increasing photon number?

- What is the tension in predicted vs. observed light element abundances?

---

[14]Similar to that in Jupiter and the outer atmosphere of the Sun. In contrast, the Earth's ocean has $D/H \sim 1.5 \times 10^{-4}$.

17 The three pillars: III - CMB

Readings: CO §*29.2,* Back

There are the so-called 'three pillars' on which the hot big-bang cosmology is firmly founded. One is expansion of the universe; the other cosmic microwave background; and the last big-bang nucleosynthesis. Observations on these three separate fronts are important and interesting. In particular, they constrain various parameters in the equation for the universe. More importantly, these parameters, whenever measurable in multiple ways, yield the same answer. So we now have a 'concordant' cosmological model.

CMB has been our foremost tool in precisely measuring the universe. The discussion below largely follows the tutorial by Wayne Hu.

Recombination and Reionization

After recombination, the cosmic microwave background (CMB) radiation streams essentially freely through space from the "surface of last scattering". It therefore preserves essentially unchanged the effects of the temperature and density fluctuations that existed *at that time*.

The first generation of luminous stars and/or quasars at $z \sim 20$ produced enough UV radiation to reionize the intergalactic medium. However, the density at this point had become low enough that the optical depth to electron scattering remained small (the total electron scattering optical path from here to recombination is $\tau_s \sim 0.1$).

CMB anisotropy

The CMB is remarkably uniform. While the first discovery of the 2.7K background radiation was in 1965 (Penzias & Wilson), the first detection of CMB fluctuations by the COBE satellite was not till 1992 (Smoot et al., though the dipole anisotropy was reported earlier). The fluctuation amplitude in temperature is $\sim 10^{-5}$ (but the dipole has an amplitude 10^{-3}). This is in contrast to the highly inhomogeneous distribution of matter, even though the inhomogeneity in matter and the anisotropy in radiation have the same origin.

Fluctuations in the microwave background are generally presented in terms of coefficients C_ℓ for fluctuations of a given angular size on the sky — for (Gaussian) temperature fluctuations as a function of position on the sky $\Delta T/T = \Theta(\theta, \phi)$, the multipole moments of the CMB temperature

$$\Theta_{\ell m} = \int \sin\theta \mathrm{d}\theta \mathrm{d}\phi \, Y_{\ell m}^* \, \Theta(\theta, \phi) \tag{17.1}$$

are fully characterized by their power spectrum

$$\langle \Theta_{\ell m}^* \Theta_{\ell' m'} \rangle = \delta_{\ell\ell'} \delta_{mm'} C_\ell \ . \tag{17.2}$$

Note that a given wavelength on the sky θ corresponds to $2\pi/\ell$.

It is hypothesized that an early phase of inflation leads to identical initial phases for all fourier oscillations in the universe. This then naturally leads to the peaks and troughs in oscillation amplitudes as a function of wavenumber, as observations have now confirmed. These are called the 'acoustic peaks' (see Fig. 17.1 for the early results from *Planck*). These peaks can then be used to measure various cosmological parameters, as shown in Fig. 17.2.

CMB dipole It has been known since the 70s that CMB is slightly warmer in one direction and cooler in the opposite one, by an amplitude of $\Delta T/T \sim 10^{-3}$. This is easiest interpreted as a kinetic velocity $v/c \sim 10^{-3}$.[15] Subtracting orbital motions of the Sun, the Milky Way galaxy, one gets a velocity of $630 \, \mathrm{km/s}$ for the local group, towards the so-called 'the great attractor'– this is also the direction headed by the bulk motion of local galaxies. Subtracting this wopping dipole, we are left with much smaller anisotropies.

---

[15]This interpretation is supported by the recent Planck results that discovers the aberration associated with this velocity.
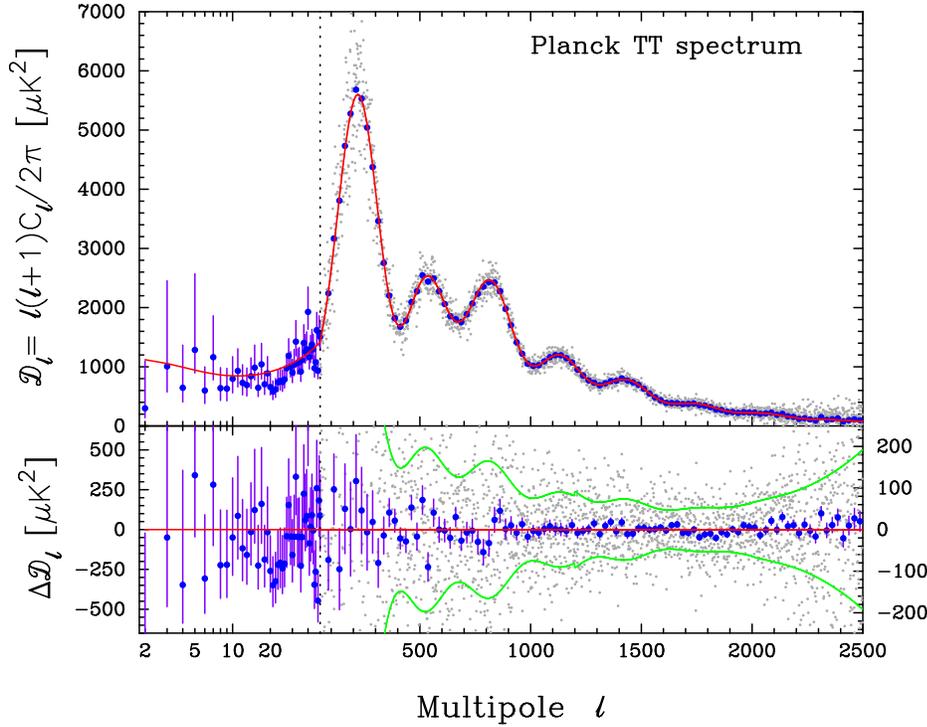
Figure 17.1: The angular power spectrum for temperature fluctuations in the CMB as measured by the Planck satellite (Ade et al, 2014). Error bars include cosmic variance. The horizontal axis is logarithmic up to $\ell = 50$ (the dotted vertical line), and linear beyond. Here, both the binned data (solid points with error bars, from combined measurement and cosmic variance errors), and the unbinned data (gray points) are shown. The much larger ($100\times$) dipole signal ($\ell = 1$) has been removed. In making this spectrum, one has to cut out large areas around the galactic plane and the ecliptic plane where emission from galactic dust and zodiacal dust dominates.

Cosmic variance or sample variance. Note that the C_ℓ are measurements of the *variance* of the CMB fluctuations, a statistical quantity whose accuracy is limited by sample size. For a given ℓ value, the range of m values is $-\ell \leq m \leq \ell$, thus there are only $(2\ell + 1)$ samples at a given ℓ, since we only have one universe to observe. This implies an inevitable error due to "cosmic variance" of

$$\Delta C_\ell = \sqrt{\frac{2}{2\ell + 1}} C_\ell .$$

(17.3)

There is no way we can determine any monopole term ($\ell = 0$), nor can a dipole term be distinguished from the dipole that results from our motion relative to the CMB frame, and other low-ℓ terms have large uncertainties. On the other hand, at higher ℓ one can bin the data, averaging over ℓ in bands of size $\Delta\ell \propto \ell$, in order to obtain final uncertainties at a given ℓ that are of order ℓ^{-1} (provided there are no other sources of uncertainty).

The First Peak The prominent peak at $\ell \sim 200$ reflects the horizon scale of the universe at recombination and shows that the universe is flat.

Quantum fluctuations in the early universe produce density perturbations at all scales. The corresponding potential flutuations pull materials into deeper wells. However, earlier than recombination, perfect matter-radiation coupling means radiation pressure can provide a restoring force against these pulls, the so-called 'sound waves'. Denser regions are necessarily hotter, and vice versa. Shorter wavelength waves will have shorter sound crossing times, and therefore shorter oscillation periods. All waves oscillate independently.

At recombination, radiation is released from matter and can thereafter stream unimpeded (no more oscillation). When photons from different distances reach us, they carry information about the phase of the acoustic oscillation at recombination, or, we can get a frozen snapshot of the acoustic oscillations in terms of angular variations. In other words, CMB angular anisotropy records spatial inhomengeity at recombination.

Define the distance sound can travel by a time t as the sound horizon. Waves that have half an wavelenth comparable to the sound horizon at recombination will appear as the first peak in CMB. Full wavelength as the second peak, $3/2$ wavelength as the third peak, etc.

What happens to waves at other wavelengths? Recall that CMB angular spectrum is obtained by summing over the whole sky (and therefore multiple horizon patches).
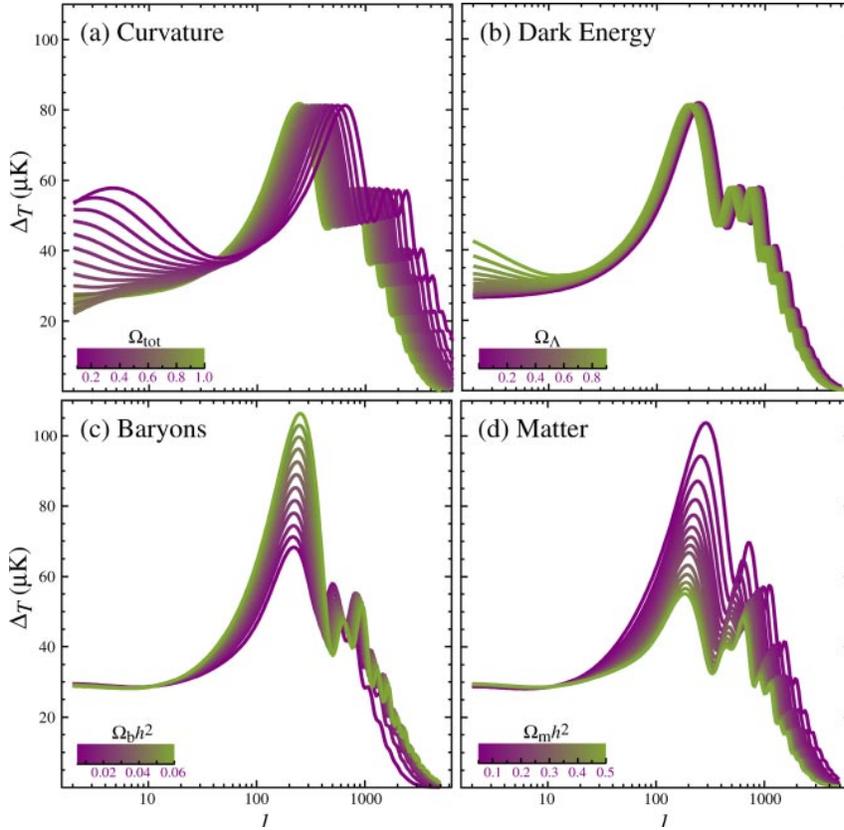
Figure 17.2: Sensitivity of the CMB angular spectrum to four fundamental cosmological parameters. Here, the power spectrum is presented in terms of $\Delta_T^2 \equiv \frac{\ell(\ell+1)}{2\pi} C_\ell T^2$, or roughly the power per logarithmic interval in wavenumber k. (a) The curvature, as quantified by $\Omega_{\rm tot} = 1 - \Omega_k$, affects the position of the first peak; (b) the dark energy, as quantified by the current cosmological constant Ω_Λ (dark energy equation of state with $w = -1$), matters little to the CMB power spectrum; (c) the current baryon density $\Omega_b h^2$, where $h \equiv H_0/(100\,{\rm km\,s^{-1}\,Mpc^{-1}})$, changes the relative amplitudes between even and odd peaks; (d) the current matter density, $\Omega_m h^2$, changes the overall amplitudes of all peaks. All are varied around a fiducial model of $\Omega_{\rm tot} = 1$, $\Omega_\Lambda = 0.65$, $\Omega_b h^2 = 0.02$, and $\Omega_m h^2 = 0.147$, with spectral index $n = 1$. From the review by Hu & Dodelson (2002).

So the position of the first peak ($\ell \sim 200$, or about 1° deg spatial scale) provides a measure of the sound horizon at $z \sim 1100$, and therefore, the curvature (see Fig. 17.2). CMB gives the best indication that the universe is flat: if the curvature of the universe is positive (closed, or $\Omega_k = 1 - \Omega_{\rm tot} < 0$, like on the surface of a sphere), e.g., the sound horizon will appear to us larger in angular scale (smaller value of ℓ).

In a two-component fluid where pressure is supplied by the radiation pressure ($P = aT^4/3$), and density is by both matter and radiation ($\rho = \rho_b + aT^4/c^2$), sound speed for the acoustic waves is,

$$c_s^2 \equiv \frac{\partial P}{\partial \rho} = \frac{\partial P/\partial \ln T}{\partial \rho/\partial \ln T} = \frac{4aT^4/3}{3\rho_b + 4aT^4/c^2} \ . \tag{17.4}$$

Or $c_s \sim c/\sqrt{3}$ if matter density can be ignored. Size of the sound horizon on the other hand, is $s = \int c_s d\eta$, where the conformal time $d\eta = dt/a(t)$. So $s \approx \eta\, c/\sqrt{3}$. This provides a standard ruler to measure the curvature of the universe.

The flat universe ($\Omega_{\rm tot} = 1$), together with the measured matter density in galaxy clusters ($\Omega_M \sim 0.2$), argue that there must be a form of energy that does not cluster gravitationally even at the scale of the largest structure. This, which we call the dark energy, is in concord with the SNIa measurement on the expansion speed of the universe.

High-ℓ peaks Amplitudes of the higher harmonics of the first peak reflect one aspect in the physics of baryon-radiation coupling, called 'baryon loading'. Higher baryon density (Ω_b) will mean more baryons load into gravitational potential minimum, which further enhances the potential contrast. This allows the oscillation to compress slightly deeper than the subsequent rarefraction. This lopsided oscillation reduces the amplitudes in even order peaks.

Damping tails As recombination occurs over a finite period of time, photons can random walk a finite distance despite their tight coupling with matter. The acoustic oscillations are exponentially damped on scales smaller than the distance photons random walk during recombination. This physics provides an additional consistency check on cosmological parameters.

Increasing the baryon density (Ω_b) will decrease the photon mean-free-path and shift the damping wavelength to higher ℓ. Increasing the total matter density (Ω_M), on the other hand, increases the age of the universe at recombination and so decreases the ℓ value for the damping tail.

Parameter	TT+lowP 68 % limits
$\Omega_b h^2$	0.02222 ± 0.00023
$\Omega_c h^2$	0.1197 ± 0.0022
$100\theta_{\mathrm{MC}}$	1.04085 ± 0.00047
τ	0.078 ± 0.019
$\ln(10^{10} A_s)$	3.089 ± 0.036
n_s	0.9655 ± 0.0062
H_0	67.31 ± 0.96
Ω_Λ	0.685 ± 0.013
Ω_m	0.315 ± 0.013
$\Omega_m h^2$	0.1426 ± 0.0020
$\Omega_m h^3$	0.09597 ± 0.00045
σ_8	0.829 ± 0.014
$\sigma_8 \Omega_m^{0.5}$	0.466 ± 0.013
$\sigma_8 \Omega_m^{0.25}$	0.621 ± 0.013
z_{re}	$9.9^{+1.8}_{-1.6}$
$10^9 A_s$	$2.198^{+0.076}_{-0.085}$
$10^9 A_s e^{-2\tau}$	1.880 ± 0.014
Age/Gyr	13.813 ± 0.038
z_*	1090.09 ± 0.42
r_*	144.61 ± 0.49

Figure 17.3: Cosmological parameters for the base ΛCDM model measured by the *Planck* CMB experiment (Ade et al 2015), showing here only results using temperature fluctuations in low multiples. Ω_b is the baryon density today; Ω_c the cold dark matter density today; $100\Theta_{\mathrm{MC}}$ is the $100\times$ approximation to r_*/D_A, where r_* is the sound horizon at recombiation, and D_A the angular distance to recombination; τ is the Thomson scattering optical depth to CMB; n_s the scalar power slope of primordial fluctuations (inflation predicting $n_s = 1$); Ω_Λ the dark energy contribution; Ω_m the matter density; σ_8 measures the RMS matter fluctuations today in linear theory; z_{re} the reionization redshift (between z_{re} and us the universe is largely ionized); z_* the redshift of the last scattering.

Polarization The new frontier of CMB nowadays lies in measuring the polarization. See the tutorial by Wayne Hu for more detail.

Food for thought

- Why does CMB remain so homogenous while matter in the universe has clumped up?

- What happens to sound waves of other wavelengths? Why are they not making a peak in the CMB spectrum?

- Why do we say that the first peak measures the universe's curvature? Imagine we live on 2-D space. If curvature is positive and we live on a sphere, show that the angular extent measured for a fixed linear length is larger than if measured in flat space.

- Why are almost all cosmological parameter measurements have an h^2 in expression? What does the Hubble constant have to do with anything?

- What parameters cannot be measured by CMB alone?

- How does CMB look like for an observer at $t =$, say, 1 Gyrs? What's the BB temperature and where is the first peak?
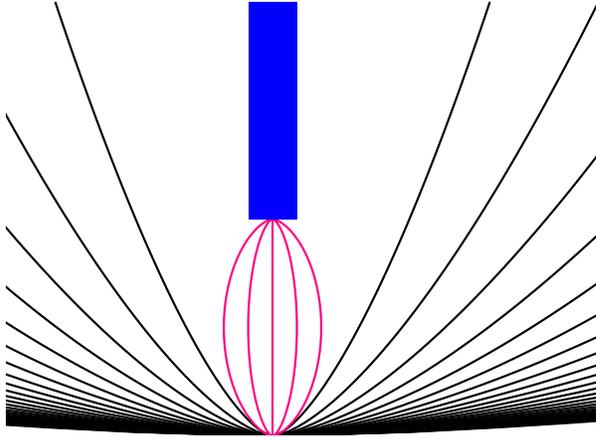
Figure 18.1: A space-time diagram showing a spherical overdense region collapsing and virializing, with time running upward. It initially expands with the rest of the universe, but with a slightly reduced rate due to its overdensity. As $\delta\rho/\rho$ becomes of order unity, it turns around at a size $r_{\max}$, the radius at maximum expansion (KE = 0) and collapses on its own. Instead of collapsing into a perfect singularity, due to the presence of inhomogeneities and non-sphericity, it usually rebounds and forms a virialized cluster of size $r_{\mathrm{vir}} \sim r_{\max}/2$, since KE = - PE/2. From the lecture notes by N. Wright.

18 Structure Formation

*Readings: CO §30.2,*Back

Primordial fluctuations and gravitational instability

The universe was not completely homogeneous at the time of radiation-matter decoupling. CMB records temperature fluctuations of order $\Delta T/T \sim 10^{-4}$, on a range of scales. These arise from the quantum fluctuations in the early universe, stretched by inflation and are therefore roughly scale independent (equal power on all scales). These produce minima and maxima in the gravitational potential. Dark matter, once decoupled from matter and radiation, can 'flow' down the potential and amplify the primordial fluctuations. Baryon matter, after recombination, are no longer coupled to radiation pressure (which gives a strong pressure with a sound speed of order c) and also start to follow these flow. The initially very small fluctuations become of order a percent by time of recombination, and continue to grow slowly – the overdense region expands slightly slower than the remaining of the universe. Consider a matter-dominated flat universe. Consider a spherical region that is slightly overdense, $\rho(t) > \rho_0(t)$. It evolves as $\rho \propto R^{-3}$ and since $R \propto t^{2/3}$, we have

$$\frac{\delta\rho}{\rho} = \left(\frac{\delta\rho}{\rho}\right)_{\mathrm{initial}} \left(\frac{t}{t_{\mathrm{initial}}}\right)^{2/3} . \tag{18.1}$$

Large scale structure in the universe, including superclusters (clusters of clusters), galaxy clusters and individual galaxies, forms when local density fluctuation is large enough ($\delta\rho/\rho \sim 1$) that the region detaches from the Hubble flow and begins to collapse. A patch that has a density $\rho \geq \rho_0$ has a dynamical time of $t_{\mathrm{dyn}} = 1/\sqrt{4\pi G\rho}$. Compare this to the expansion of the universe

$$Ht_{\mathrm{dyn}} = \sqrt{\frac{2}{3}\frac{3H^2}{8\pi G\rho}} = \sqrt{\frac{2}{3}\frac{\rho_0}{\rho}} , \tag{18.2}$$

or, unless $\delta\rho/\rho = 1 - \rho_0/\rho \sim 1$, the e-folding time for the perturbation is not much shorter than the universal expansion and is wiped out. Once perturbation becomes nonlinear, it collapses to of order half the maximum radius and evolution beyond is difficult to trace analytically.

Bottom-up Galaxy Formation and mergers

The presence and nature of dark matter is essential for structure formation. The most compelling proposal for dark matter currently is some kind of 'cold dark matter' ($mc^2 \gg mv^2$) that interacts only via gravity (but not electro-magnetically). Unlike hot dark matter, cold dark matter can accumulate in shallow potentials.
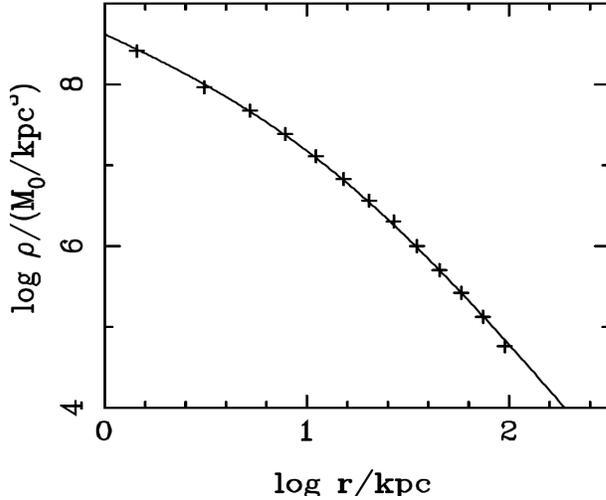
Figure 18.2: The end state of almost any gravitational collapse, independent of details like the small scale structure and hierarchical merging, leads to a universal radial density profile for the final halo, called the NFW (Navarro, Frenk & White, 1997) profile, $\rho(r) \propto 1/[r(1+r/r_s)^2]$ (the solid line), where r_s is the scale radius. The dots show the virialized density profile of a million particle dark matter simulation by Dubinski & Carlberg (1990), one of the first such attempts.

Moreover, given the relative fractions, one can ignore baryon matter to first order when considering structure formation. This can be probed using N-body simulations of collisionless, gravitating particles, starting from realistic and well constrained cosmological initial conditions.

why small scales collapse first

CDM theory predicts that the growth of cosmic structures begins at halo masses that can be, depending on the dark matter particle rest mass (and therefore their decoupling time), lower than that of the Sun (see, e.g., Diemand, Moore & Stadel, 2005). These mini-halos can merge with each other, eventually building up to galaxy clusters that have masses $\sim 10^{15} M_\odot$. There are a few consequences of such a picture. Each resultant halo satisifes a density profile that is rather universal (the NFW profile); many subhalos may persist in a single large halo; larger objects (very massive galaxies) form later in time, but their stellar components (e.g.) can be much older; large halos can only form in denser environments where mergers are frequent. These features can be directly tested by observations.

galaxy: the Milky way galaxy is a generic galaxy, with a total mass of $\sim 10^{12} M_\odot$ (where $\sim 5\%$ is baryon and mostly in stars) and a halo size ~ 100 kpc (the stellar disk radius ~ 10 kpc). The velocity dispersion is $\sim 200 \, \mathrm{km/s}$. The Galaxy is a spiral type and may have grown by accreting many other smaller dwarf galaxies. It is destined to have a major merger with the Andromeda galaxy in a few billion years, perhaps producing a spheroidal galaxy afterwards.

In dense environments, e.g., in cluster centers, one observe BCGs (brightest cluster galaxies) with masses of $\sim 10^{15} M_\odot$. A merger tree for such a galaxy contains many steps (Fig. 18.4).

galaxy cluster: clusters of galaxies have masses $\sim 10^{15} M_\odot$, containing hundreds of galaxies and extending to ~ 10 Mpc. The velocity dispersion is $\sim 1000 \, \mathrm{km/s}$. About 10% of the mass is baryon: most of it in hot intracluster gas, and only $\sim 10\%$ in stars. Strong lensing measures masses of clusters and yields $\Omega_{\mathrm{cluster}} \sim 0.2 - 0.3$, not enough to close the universe by itself.

superclusters Galaxy clusters are not uniformly distributed but form even larger structure like the superclusters. These are the largest structure today in the universe, with a dynamical time of order the current Hubble time. With sizes of ~ 100 Mpcs, they are marginally bound structure and will likely disperse with the Hubble flow as the universe expands. Redshift surveys show that they lie along filaments of overdensities (cosmic web), with giant **voids** in between. These likely have the cosmic complements of baryons and dark matter ($\Omega_b h^2 = 0.02$ and $\Omega_m h^2 = 0.15$).

The highest redshift galaxies and quasars

Observations of the highest redshift galaxy and massive blackholes pose challenging tests for the bottom-up CDM model. The current confirmed records are held by a galaxy at $z = 8.7$ and a quasar at $z = 7.1$ (Wiki page, though a new paper claims a $z = 11.1$ galaxy), corresponding to the time when the universe was 0.57 Gyrs and 0.75 Gyrs old, respectively. Inserting these into eq. (18.1), and requiring that $\delta\rho/\rho = 1$ at these ages, we obtain a minimum estimate for the original
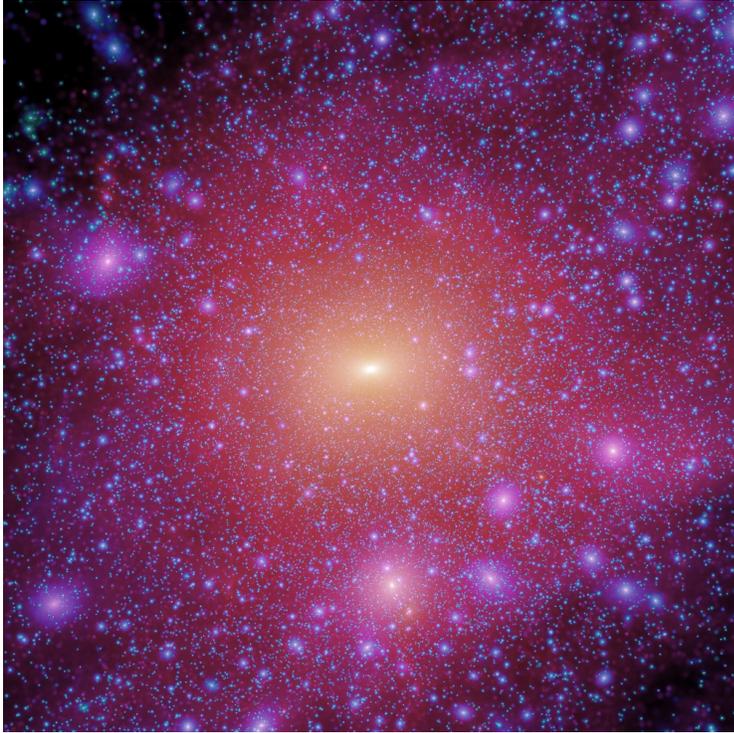
Figure 18.3: High resoution N-body simulations (the Aquarius project) showing the dark-matter halo of a Milky Way-like galaxy, represented in simulation by some 1 billion particles. It has a radial profile that is roughly NFW, but is far from being smooth. The multitude of sub-halos are remains of the cores of dwarf galaxies that have been swallowed in the past – the mergers are not complete, the inner cores of their halos survive. This can be tested by studying stellar dynamics, dark-matter annihilation, tidal stream disturbances... to confirm the CDM nature.
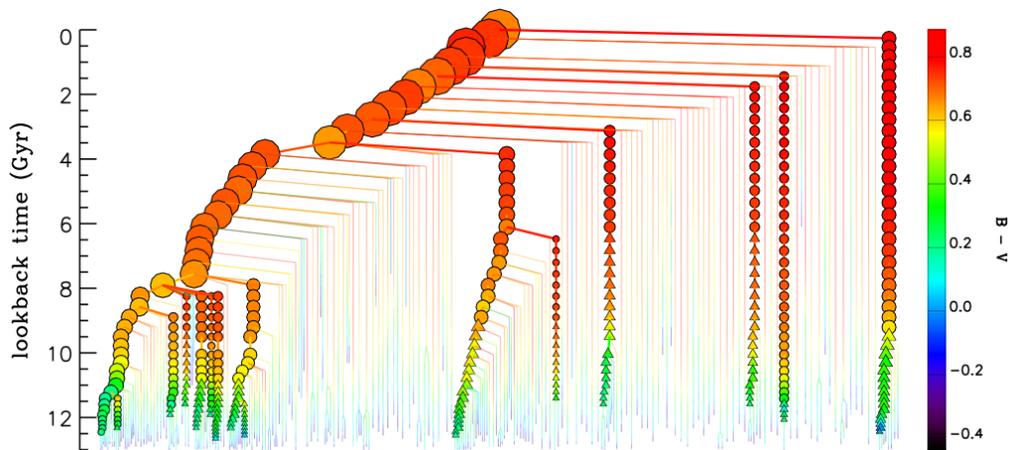

Figure 18.4: The merger tree for a BCG with a total mass of $10^{15} M_{\odot}$. The BCG itself is plotted at the top, and all its progenitors (and also their histories) are plotted backward in time recursively. Galaxies with stellar masses larger than $10^{10} M_{\odot}$ are shown as symbols, while lower mass galaxies are shown as lines. The size of the symbols scales with stellar mass, while the colour scales with the B-V colour of the galaxy itself, which can be considered as a proxy for the age of the stellar population. Model BCGs are therefore old (i.e. their stars formed very early), but their mass assembled quite late. Moreover, the halo of the BCG should contains numerous subhalos, the incomletely swallowed cores of the progenitor galaxies. These may be observationally detected. From De Lucia & Blaizot (2007).
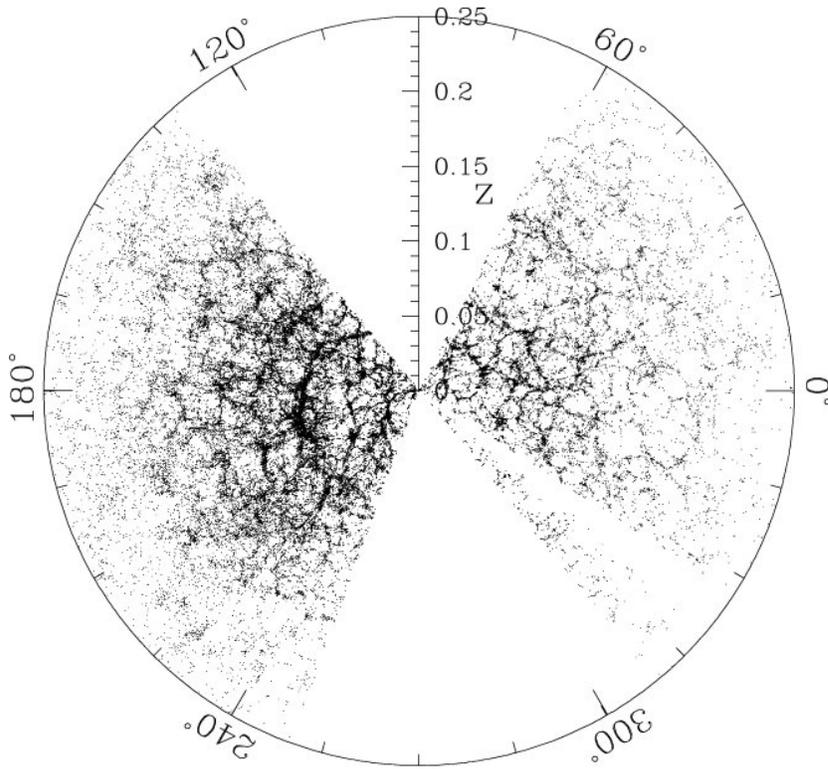
Figure 18.5: One of the largest redshift surveys to date (with about a million galaxies), the Sloan Digitial Sky Survey (SDSS) presents a picture of the large scale structure in redshift space. Each point is a galaxy with a measured redshift, out to redshift $z = 0.25$, corresponding to a comoving distance of ~ 1 Gpc. Only a thin slice of the entire sky is shown. Galaxies are seen to group into clusters (nodes) and to superclusters (filaments). Denser regions tend to contain older (redder) stars. The giant voids must contain matter but they do not form into galaxies (bias). The regions between the wedges are not probed due to galactic extinction.
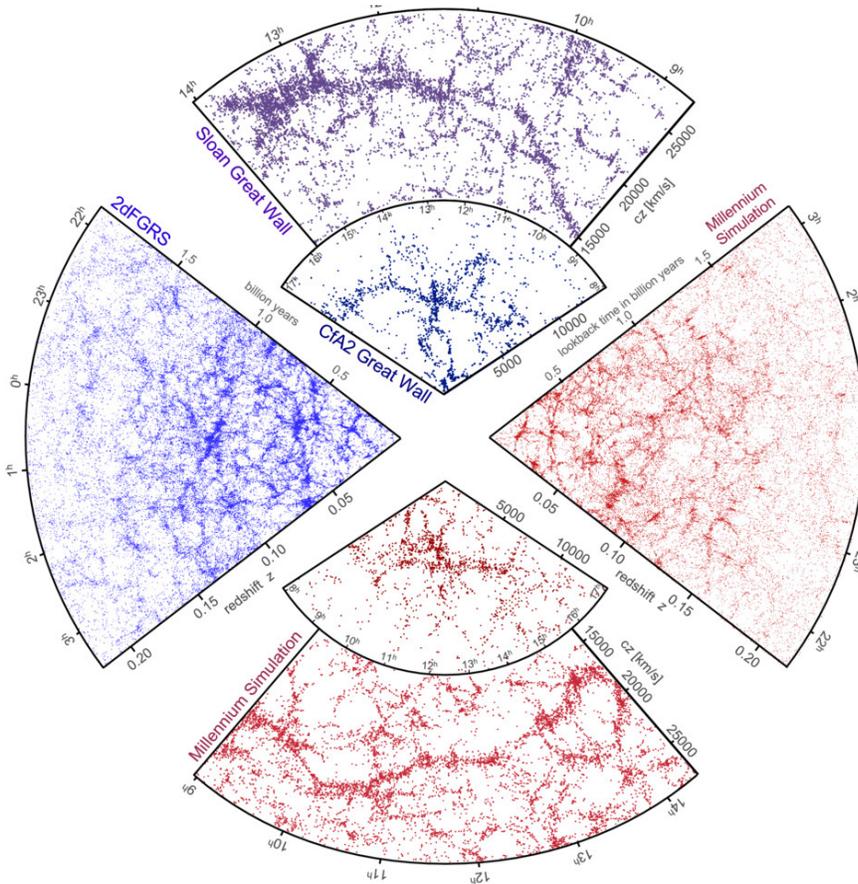

Figure 18.6: Using more than 10 billion particles to trace the evolution of matter in a co-moving cube ~ 1 Gpc on a side, and starting at $z = 127$, the Millenium simulation produced some 20 million galaxies (red dots, so each galaxy is represented in the simulation by ~ 500 particles) which can be compared with data like the SDSS redshift surey (blue and purple dots). Besides from staring at images, one can also compare quantitatively, using, e.g., galaxy-galaxy 2-point correlation function, galaxy mass functions, etc. ΛCDM seems to be a success.

fluctuation at decoupling to be $\delta\rho/\rho \sim 0.0076$ and ~ 0.0063, respectively. These earliest structures likely formed in regions of largest perturbations.

Before these first objects form, the universe was in a 'dark age' with largely neutral hydrogen. The UV radiation from the first stars (see below) and/or AGN re-ionized the universe. This **epoch of reionization** shows up as dense Lyα absorption in the spectra of high-redshift quasars. It seems that reionization was complete by redshift ~ 6, and possibly starting around $z \sim 11$ (according to Planck). It is interesting to know what structures formed so early to reionize the universe, and how the reionization suppresses/promotes structure formation. Probing reionization using the 21-cm hydrogen hyperfine transition is currently a hot topic, as the Canadian CHIME project is set out to do.

The stars that are formed in the early universe, due to a lack of metal (and their associated cooling), are assumed to be much more massive than those formed today. Called **Pop III** stars, they may have long burned out, but may have left their mark in nucleosynthesis, or in reionization. Looking for the most metal poor stars (polluted only by Pop III stars but otherwise pristine) in our galaxy is another interesting venue to constrain the properties of these first generation stars.

Food for thought

- How are fluctuations in the CMB related to large-scale structure we see in the universe today? When were the fluctuations seeded? How does density fluctuations grow in baryons and photons, separately? What happen to these fluctuations after matter-radiation decoupling? How does matter fluctuations lead to gravitational collapse and structure formation?

- Why do we think dark matter is 'cold'? What evidences do we have?

- What is unique about gravitational instability in an expanding universe?

- How can we confirm that structure formation in the universe is bottom-up?