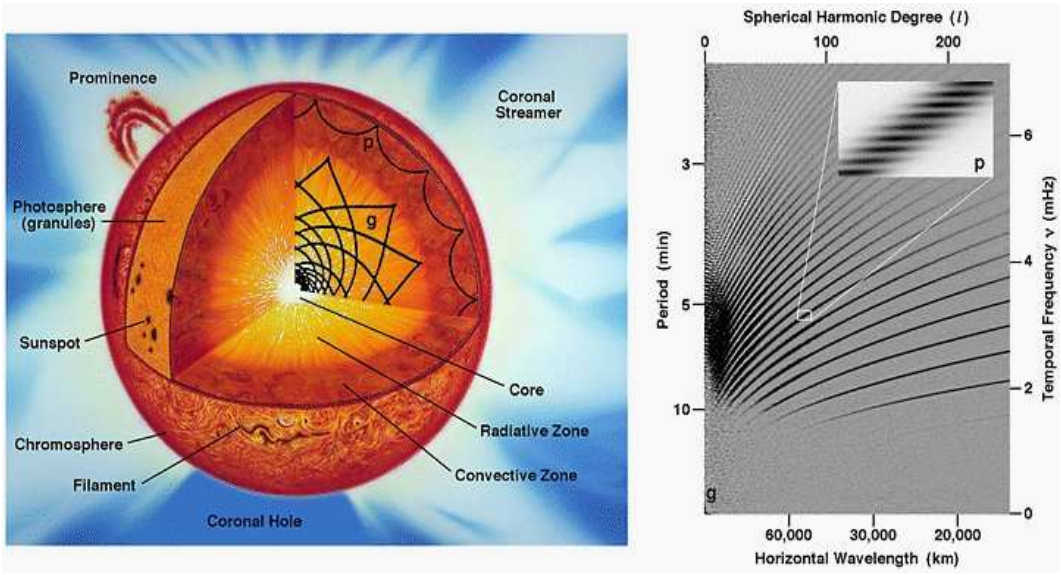
A Concise Introduction to Astrophysics

– Lecture Notes for FY2450 –

M. Kachelrieß



M. Kachelrieß
Institutt for fysikk
NTNU, Trondheim
Norway
email: Michael.Kachelriess@ntnu.no

Contents

Astrophysics—some introductory remarks

- Astronomy is with mathematics one of the oldest branches of science. It has served as basis for calendars, navigation, has been an important input for religions and was for a long time intertwined with astrology.
- Some of the most important steps in modern astronomy were:
  - Galileo performed 1609 the first astronomical studies using a telescope. He discovered among others four Saturn moons and sun spots.
  - Kepler (1571-1630) developed his three laws of planetary motions, based on observations of Tycho Brahe.
  - Newton established 1687 his laws of motion and gravitation.
  - The measurement of the distance to Venus 1761 and 1769 during its transits of the Sun with the help of the first global measurement campaign and to the nearest stars 1838 by Bessel using trigonometric parallaxes established the first rungs in the "cosmic distance ladder."
  - Fraunhofer discovered around 570 spectral lines in the solar light in 1814 and catalogued them. This opened together with the spectral analysis of Kirchhoff and Bunsen (1859) the way to study the physical properties of stars.
  - Einstein's general theory of relativity (1916) provided the first consistent basis to study cosmology.
  - The "Great Debate" in 1920 was concerned about the question "Does the Milky Way represents the whole Universe or is it just one island among many others?" Öpik, Shapley and Hubble showed that the latter is true.
  - Hubble discovered 1926 that galaxies are recessing and that their velocity is increasing with distance: The universe is expanding and, extrapolating this expansion back in time led to the idea of the "Big Bang."
  - Nuclear fusion was suggested in 1920 by Eddington as source for stellar energy, the main principles were worked out after the advent of quantum mechanics by Bethe and v. Weizsäcker in the 1930s.
  - The discovery of the cosmic 2.7 Kelvin background radiation 1964 by Penzias and Wilson gave credit to the Big Bang theory.
- Few other examples for the interconnection of astrophysics and physics, where astronomical observations were an important input for fundamental physics, are:
  - Olaf C. Römer (1644-1710) showed 1676 that the speed of light is finite by observations of Jupiter moons: Light needs around 20 min to cross the Earth orbit.
  - The 1919 solar eclipse was the first crucial test passed by the theory of General Relativity of Einstein, while a binary system of two pulsars discovered by Hulse and Taylor in 1974 became the first experimental evidence for the existence of gravitational waves.

  - Observation of neutrinos from the Sun and produced by cosmic rays in the Earth's atmosphere gave in the 1990's first firm evidence that neutrinos have non-zero masses.
  - The need for a new form of "dark matter" to describe correctly the formation and dynamics of galaxies requires a yet unknown extension of the current standard model of elementary particle physics. The same holds true for a new form of "dark energy" required for the explanation of the accelerated expansion of the universe.

- Astronomy as a purely observational science is unique among natural sciences; all others are based on experiments. Since the observation time is much smaller than the typical time scale for the evolution of astronomical objects, we see just a snapshot of the universe. Nevertheless it is possible to reconstruct, e.g., the evolution of stars by studying large samples. On the extragalactic scales, we can use that looking far away means looking into the past because of the finite speed of light, while on the cosmological scale we can use relics formed soon after the Big Bang as testimonies for the state of the early universe.

- The "cosmological principle" is based on the belief that the mankind and the Earth have no special role. Thus physical laws derived on Earth are valid everywhere and at all times.

- Astronomers and especially cosmologists are said to live these days in a "golden age": There has been a tremendous increase of knowledge in the last 15 years: Telescopes and detectors on satellites explore new wavelength ranges, while new automatized ways to analyse data allow astronomers the comparative study of e.g. millions of galaxies.

Astrophysics needs input of practically all sub-disciplines of physics and thus a course on astrophysics cannot be self-contained. However, the course should be accessible to students with just a general introduction to physics. Few sections of the text that are somewhat more advanced and that can be omitted are marked by stars.

I will be glad to receive feedback from the readers of these notes. If you find errors or have anysuggestions, send me an email!

Part I

Stellar astrophysics

1 Continuous radiation from stars

Practically all information we know about stars and galaxies comes from observing electromagnetic radiation, more precisely from two small windows where the Earth atmosphere is (nearly) transparent: The first one is around the range of visible light (plus some windows in the infra-red (IR)), while the second one is the radio window in the wave-length range $1\,\mathrm{cm} \lesssim \lambda \lesssim 30\,\mathrm{m}$, cf. Fig. 1.1.

1.1 Brightness of stars

Hipparcos and Ptolemy (≈ 150 B.C.) divided stars into six classes of brightness, called (apparent) magnitudes m: Stars with $m = 6$ were the faintest objects visible by eye, stars with $m = 1$ the brightest stars on the sky. Thus the magnitude scale m increases for fainter objects, a counter-intuitive fact that should be kept in mind.

The response of the human eye is not linear but close to logarithmic to ensure a large dynamical range. With the use of photographic plates, the magnitude scale could be defined quantitatively. Pogson found 1856 that the magnitude difference $\Delta m = 5$ corresponds to a ratio of energy fluxes of $\mathcal{F}_2/\mathcal{F}_1 \approx 100$, where the energy flux $\mathcal{F}$ is defined as the energy E going through the area A per time t, $\mathcal{F} = E/(At)$. This ratio was then used as definition, together with $m = 2.12$ for the polar star as reference point. Thus the ratio of received energy per time and area from two stars is connected to their magnitudes by

$$\frac{\mathcal{F}_2}{\mathcal{F}_1} = 100^{(m_1-m_2)/5} = 10^{(m_1-m_2)/2.5} \tag{1.1}$$

or[1]

$$m_1 - m_2 = -2.5 \log \frac{\mathcal{F}_1}{\mathcal{F}_2}\,. \tag{1.2}$$

The modern definition of the magnitude scale does not agree completely with the original Greek one: For instance, the brightest star visible from the Earth, Sirius, has the magnitude $m = -1.5$ instead of the Ptolemic $m = 1$. Note also that in particular astronomers use often the equivalent name (apparent) brightness b for the energy flux received on Earth from a star or a galaxy.

Ex.: The largest ground-based telescopes can detect stars or galaxies with magnitude $m = 26$. How much fainter are they than a $m = 1$ star like Antares?

$$\frac{\mathcal{F}_1}{\mathcal{F}_2} = 10^{-25/2.5} = 10^{-10}\,. \tag{1.3}$$

Thus the number of photons received per time and frequency interval from such a faint object is a factor 10^{-10} smaller than from Antares. Since the energy flux $\mathcal{F}$ scales with distance r approximately

[1] We denote logarithms to the base 10 with log, while we use ln for the natural logarithm.

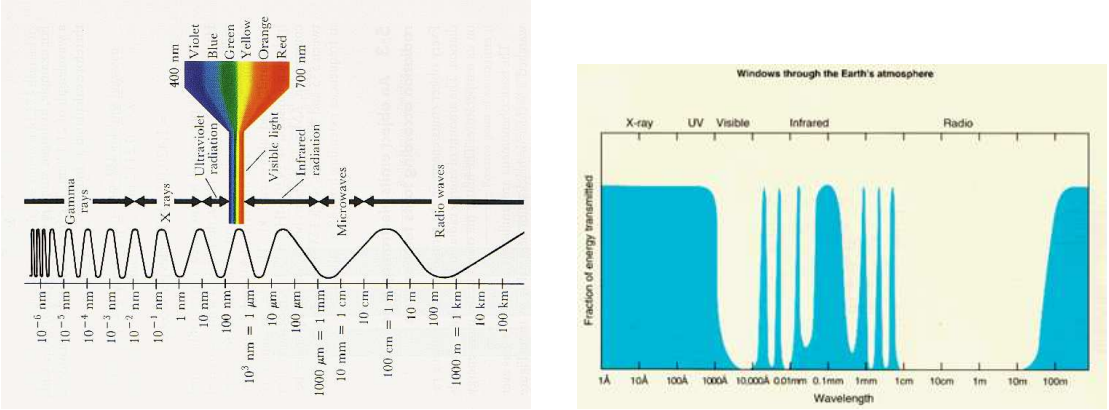


Figure 1.1: Left: The electromagnetic spectrum. Right: Transmission of the Earth's atmosphere.

as $\propto 1/r^2$, one can say equivalently that with the help of such a telescope one would see a factor 10^5 further than with the bare human eye.

This magnitude scale corresponds to the energy flux at Earth, i.e. the scale is not corrected for the different distance of stars. They are therefore called "apparent magnitudes."

1.2 Color of stars

A look at Fig. 1.2 shows that the color of stars and galaxies varies considerably. A quantitative way to measure the color of stars is to use filters which sensitivity is centered at different wavelengths λ_i and compare their relative brightness,

$$m_{\lambda_1} - m_{\lambda_2} = -2.5 \log \frac{\mathcal{F}_{\lambda_1}}{\mathcal{F}_{\lambda_2}} \,. \tag{1.4}$$

Then $m_{\lambda_1} - m_{\lambda_2}$ is called the color or the color index of the object. Often used are filters for visible (V), ultraviolet (U), and blue (B) light. For instance, one denotes briefly with B-V the difference in magnitudes measured with a filter for blue and visible light,

$$\mathrm{B} - \mathrm{V} \equiv m_B - m_V = -2.5 \log \frac{\mathcal{F}_B}{\mathcal{F}_V} \,. \tag{1.5}$$

It is intuitively clear that the color of a star is connected to its temperature—this connection will be made more precise in the next section. The color magnitudes are normalized such that their differences $m_{\lambda_1} - m_{\lambda_2}$ are zero for a specific type of stars with surface temperature $T \approx 10,000\,\mathrm{K}$ (cf. Sec. 2.2).

1.3 Black-body radiation

A black body is the idealization of a medium that absorbs all incident radiation. This idealization is very useful, because many objects close to thermal equilibrium, in particular stars, emit approximately blackbody radiation. In practise, black-body radiation can be approximated by radiation escaping a small hole in a cavity whose walls are in thermal equilibrium. Inside such a cavity, a gas of photons in thermal equilibrium exists.

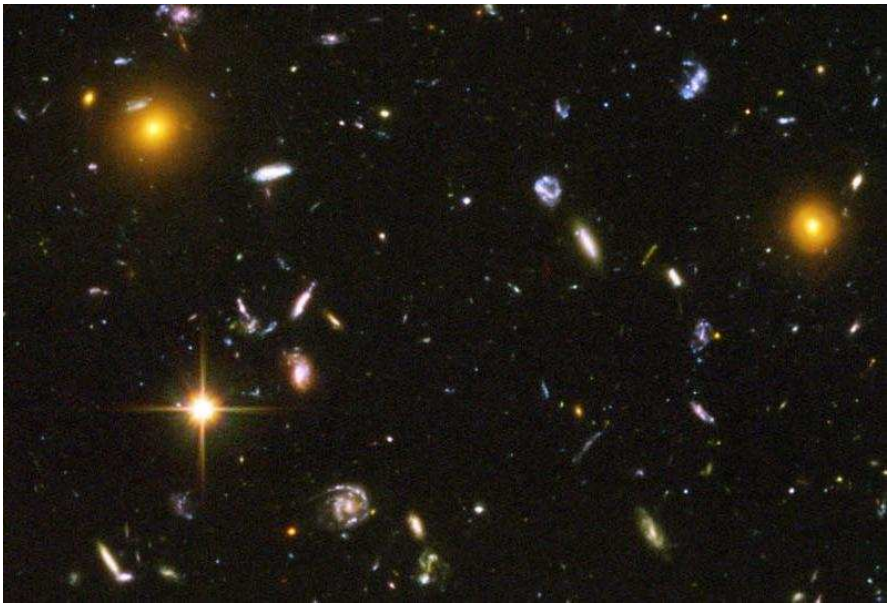


Figure 1.2: An excerpt from the Deep Field View taken by the Hubble Space Telescope.

1.3.1 Kirchhoff-Planck distribution

Planck found empirically 1900 the so-called Kirchhoff-Planck (or simply Planck) distribution

$$B_\nu \mathrm{d}\nu = \frac{2h\nu^3}{c^2} \frac{1}{\exp(\frac{h\nu}{kT}) - 1} \, \mathrm{d}\nu \tag{1.6}$$

describing the amount of energy emitted into the frequency interval $[\nu, \nu + \mathrm{d}\nu]$ and the solid angle $\mathrm{d}\Omega$ per unit time and area by a body in thermal equilibrium. The function B_ν depends only on the temperature T of the body (apart from the natural constants k, c and h) and is shown in Fig. 1.4 for different temperature T. The dimension of B_ν in the cgs system of units is

$$[B_\nu] = \frac{\mathrm{erg}}{\mathrm{Hz\ cm^2\ s\ sr}} \, . \tag{1.7}$$

In general, the amount of energy per frequency interval $[\nu, \nu + \mathrm{d}\nu]$ and solid angle $\mathrm{d}\Omega$ crossing the perpendicular area $A_\perp$ per time is called the (differential) intensity I_ν, cf. Fig. 1.3,

$$I_\nu = \frac{\mathrm{d}E}{\mathrm{d}\nu \, \mathrm{d}\Omega \, \mathrm{d}A_\perp \, \mathrm{d}t} \, . \tag{1.8}$$

For the special case of the blackbody radiation, the differential intensity is given by the Kirchhoff-Planck distribution, $I_\nu = B_\nu$.

Equation (1.6) gives the spectral distribution of black-body radiation as function of the frequency ν. In order to obtain the corresponding formula as function of the wavelength λ, one has to re-express both B_ν and $\mathrm{d}\nu$: With $\lambda = c/\nu$ and thus $\mathrm{d}\lambda = -c/\nu^2 \mathrm{d}\nu$, it follows

$$B_\lambda \mathrm{d}\lambda = \frac{2hc^2}{\lambda^5} \frac{1}{\exp(\frac{hc}{\lambda kT}) - 1} \, \mathrm{d}\lambda \, . \tag{1.9}$$

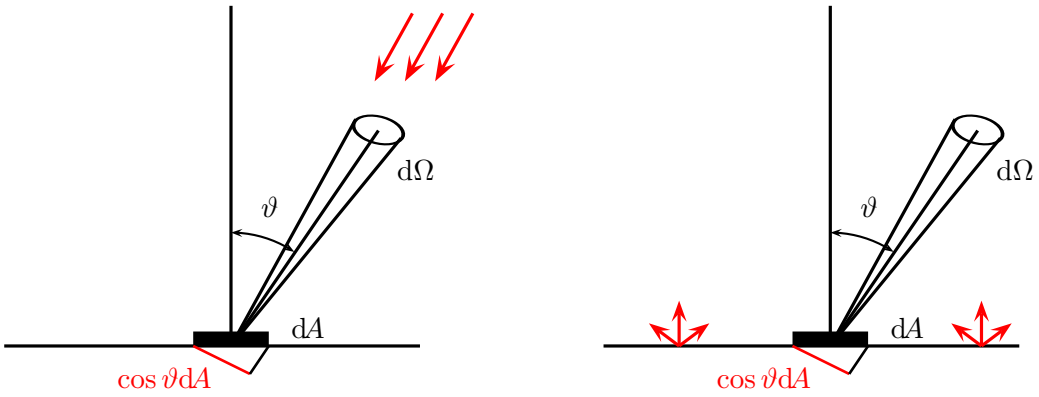


Figure 1.3: Left: A detector with surface element $\mathrm{d}A$ on Earth measuring radiation coming from a direction with zenith angle ϑ. Right: An imaginary detector on the surface of a star measuring radiation emitted in the direction ϑ.

The Kirchhoff-Planck distribution contains as its two limiting cases Wien's law for high-frequencies, $h\nu \gg kT$, and the Rayleigh-Jeans law for low-frequencies $h\nu \ll kT$. In the former limit, $x = h\nu/(kT) \gg 1$, and we can neglect the -1 in the denominator of the Planck function,

$$B_\nu \approx \frac{2h\nu^3}{c^2} \exp(-h\nu/kT)\,. \tag{1.10}$$

Thus the number of photons with energy $h\nu$ much larger than kT is exponentially suppressed. In the opposite limit, $x = h\nu/(kT) \ll 1$, and $\mathrm{e}^x - 1 = (1 + x - \ldots) - 1 \approx x$. Hence Planck's constant h disappears from the expression for B_ν, if the energy $h\nu$ of a single photon is small compared to the thermal energy kT and one obtains,

$$B_\nu \approx \frac{2\nu^2 kT}{c^2}\,. \tag{1.11}$$

The Rayleigh-Jeans law shows up as straight lines left from the maxima of B_ν in Fig. 1.4.

1.3.2 Wien's displacement law

We note from Fig. 1.4 two important properties of B_ν: Firstly, B_ν as function of the frequency ν has a single maximum. Secondly, B_ν as function of the temperature T is a monotonically increasing function for all frequencies: If $T_1 > T_2$, then $B_\nu(T_1) > B_\nu(T_2)$ for all ν. Both properties follow directly from taking the derivative with respect to ν and T. In the former case, we look for the maximum of $f(\nu) = \frac{c^2}{2h} B_\nu$ as function of ν. Hence we have to find the zeros of $f'(\nu)$,

$$3(\mathrm{e}^x - 1) - x \exp^x = 0 \qquad \text{with} \qquad x = \frac{h\nu}{kT}\,. \tag{1.12}$$

The equation $\mathrm{e}^x(3 - x) = 3$ has to be solved numerically and has the solution $x \approx 2.821$. Thus the intensity of thermal radiation is maximal for $x_{\max} \approx 2.821 = h\nu_{\max}/(kT)$ or

$$\frac{cT}{\nu_{\max}} \approx 0.50\mathrm{K\,cm} \qquad \text{or} \qquad \frac{\nu_{\max}}{T} \approx 5.9 \times 10^{10}\mathrm{Hz/K}\,. \tag{1.13}$$

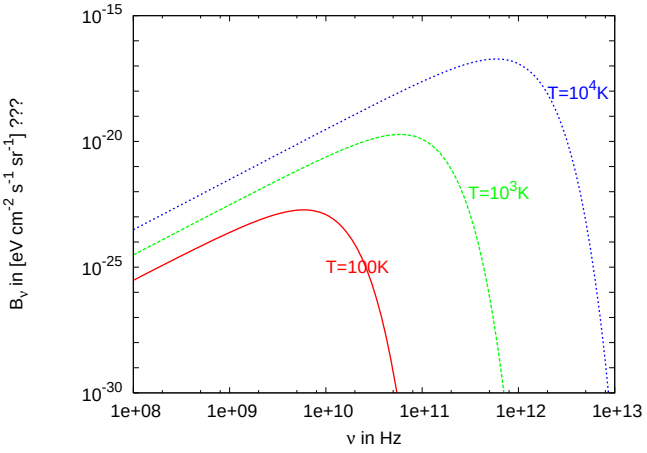


Figure 1.4: The Planck distribution B_ν as function of the frequency ν for three different temperatures.

Similarly one derives, using the expression for B_λ, the value $\lambda_{\max}$ where B_λ peaks as

$$\lambda_{\max} T \approx 0.29\mathrm{K\,cm}\,. \tag{1.14}$$

This equation is called Wien's displacement law. Thus determining the color of a star tells us its temperature: A hot star is bluish and a cool star reddish. In practise, the maximum of the Planck distribution lies often outside the range of visible light and the measurement of a color index like U − V is an easier method to determine the temperature T of a star.

Ex.: Surface temperature of the Sun:
We can obtain a first estimate of the surface temperature of the Sun from the know sensitivity of the human eye to light in the range 400–700 nm. Assuming that the evolution worked well, i.e. that the human eye uses optimal the light from the Sun, and that the atmosphere is for all frequencies in the visible range similarly transparent, we identify the maximum in Wien's law with the center of the frequency range visible for the human eye. Thus we set $\lambda_{\max,\odot} \approx 550$ nm, and obtain $T_\odot = 2.9 \times 10^6\,\mathrm{nm\,K}/550\,\mathrm{nm} \approx 5270\,\mathrm{K}$ for the surface temperature of the Sun.
This simple estimate should be compared to the precise value of 5781 K for the "effective temperature" of the solar photosphere defined in the next subsection.

1.3.3 Stefan-Boltzmann law

Integrating Eq. (1.6) over all frequencies and possible solid angles gives the energy flux $\mathcal{F}$ emitted per surface area A by a black body. The angular integral consists of the solid angle $\mathrm{d}\Omega = \mathrm{d}\vartheta \sin\vartheta \mathrm{d}\phi$ and the factor $\cos\vartheta$ taking into account that only the perpendicular area $A_\perp = A\cos\vartheta$ is visible,

$$\int \mathrm{d}\Omega\, \cos\vartheta = \int_0^{2\pi} \mathrm{d}\phi \int_0^{\pi/2} \mathrm{d}\vartheta \sin\vartheta\, \cos\vartheta = \pi \int_0^{\pi/2} \mathrm{d}\vartheta \sin 2\vartheta = \pi\,. \tag{1.15}$$

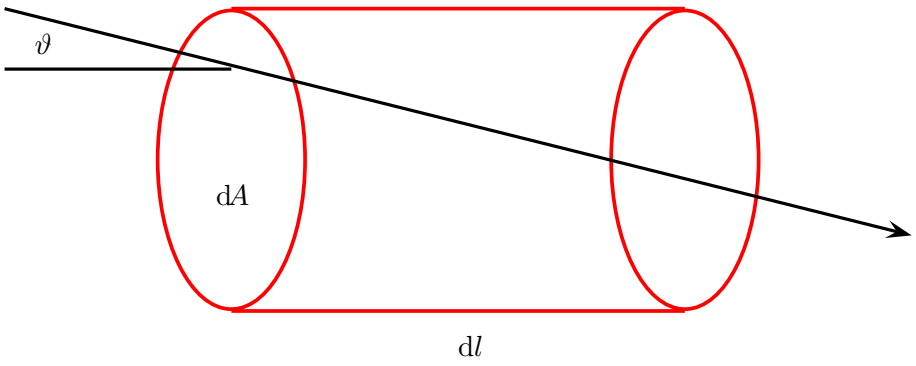


Figure 1.5: The connection between the contained energy density u and the intensity I of radiation crossing the volume $\mathrm{d}V$.

Next we perform the integral over $\mathrm{d}\nu$ substituting $x = h\nu/(kT)$,

$$\mathcal{F} = \pi \int_0^\infty d\nu B_\nu = \frac{2\pi}{c^2 h^3} (kT)^4 \int_0^\infty \frac{x^3 \mathrm{d}x}{\mathrm{e}^x - 1} = \sigma T^4 \,, \tag{1.16}$$

where we used $\int_0^\infty \frac{x^3 \mathrm{d}x}{\mathrm{e}^x - 1} = \pi^4/15$ and introduced the Stefan-Boltzmann constant,

$$\sigma = \frac{2\pi^5 k^4}{15 c^2 h^3} = 5.670 \times 10^{-5} \frac{\mathrm{erg}}{\mathrm{cm^2 K^4 s}} \,. \tag{1.17}$$

If the Stefan-Boltzmann law is used to define the temperature of a body that is only approximately in thermal equilibrium, T is called effective temperature T_e.

1.3.4 Spectral energy density of a photon gas

A light ray crosses an infinitesimal cylinder with volume $\mathrm{d}V = \mathrm{d}A\mathrm{d}l$ in the time $\mathrm{d}t = \mathrm{d}l/(c\cos\vartheta)$, cf. Fig. 1.5. Hence the spectral energy density u_ν of photons, i.e. the number of photons per volume and energy interval, in thermal black-body radiation is

$$u_\nu = \frac{\mathrm{d}E}{\mathrm{d}V\,\mathrm{d}\nu} = \frac{\mathrm{d}E}{c\cos\vartheta \mathrm{d}t \mathrm{d}A \mathrm{d}\nu} = \frac{1}{c} \int \mathrm{d}\Omega\, B_\nu = \frac{4\pi}{c} B_\nu \,. \tag{1.18}$$

The (total) energy density of photons u follows as

$$u = \frac{4\pi}{c} \int_0^\infty \mathrm{d}\nu B_\nu = aT^4 \,, \tag{1.19}$$

where we introduced the radiation constant a. Comparing with Eqs. (1.16) and (1.17) shows that $a = 4\sigma/c$.

Ex.: Calculate the density and the mean energy of the photons in thermal black-body radiation at temperature T.

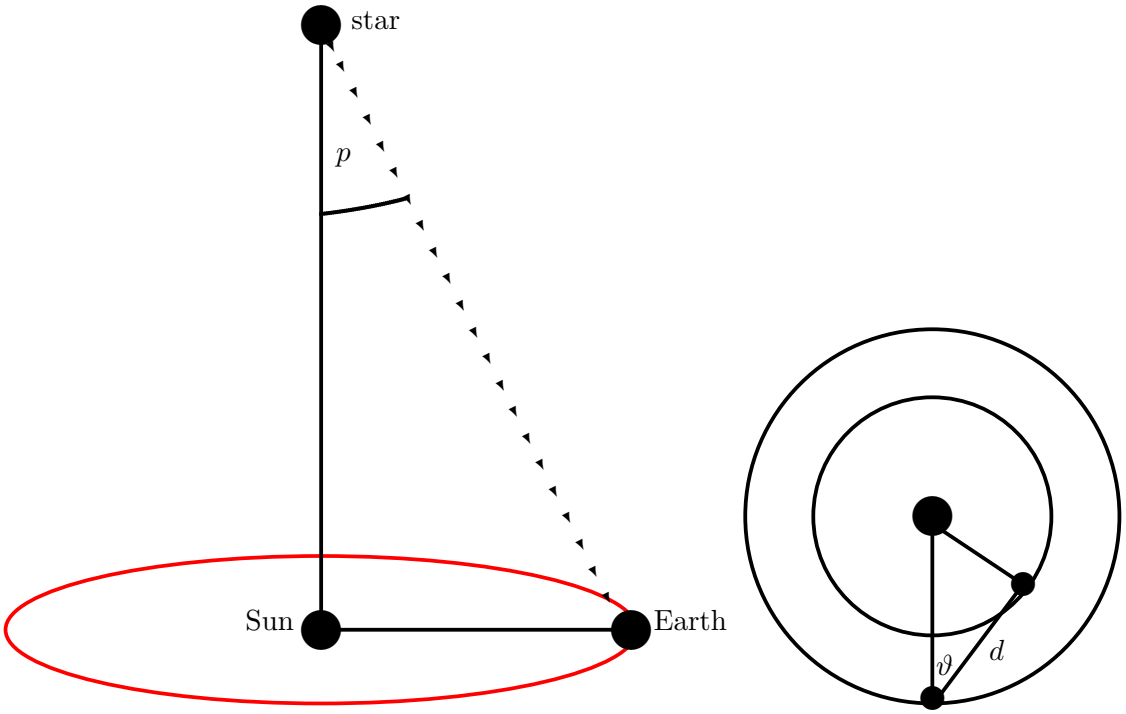


Figure 1.6: Measurement of the trigonometric parallax. Left: The apparent movement of a nearby star during a year: Half of this angular difference is called the parallax angle or simply the parallax p. Right: Indirect determination of the Earth-Sun distance.

The number density of black-body photons is connected to the energy density by the replacement $B_\nu \to B_\nu/(h\nu)$,

$$n = \frac{4\pi}{c}\int_0^\infty d\nu\,\frac{B_\nu}{h\nu} = \frac{8\pi}{c^3}\int_0^\infty \mathrm{d}\nu\,\frac{\nu^2}{\exp(\frac{h\nu}{kT})-1} = \frac{8\pi}{c^3}(kT/h)^3\int_0^\infty \mathrm{d}x\,\frac{x^2}{\mathrm{e}^x-1} = \frac{2\zeta(3)}{\pi^2}\left(\frac{kT}{\hbar c}\right)^3 ,$$

where we looked up the result for the last integral in Table A.1. The mean energy of photons at temperature T follows then as

$$\langle E\rangle = \langle h\nu\rangle = \frac{u}{n} = \frac{\pi^4}{30\zeta(3)}\,kT \approx 2.701kT\,.$$

1.4 Stellar distances

Relatively nearby stars are seen at slightly different positions on the celestial sphere (i.e. the background of stars that are "infinitely" far away) as the Earth moves around the Sun. Half of this angular difference is called the parallax angle or simply the parallax p. From Fig. 1.6, one can relate the mean distance AU of the Earth to the Sun, the parallax p and the distance d to the star by

$$\tan p = \frac{\mathrm{AU}}{d}\,. \tag{1.20}$$

With $\tan p \approx p$ for $p \ll 1$, one has $d \propto 1/p$. Because of this simple and for observations important relation, a new length unit is introduced: The parsec is defined to be the distance from the Earth to a star that has a parallax of one arcsecond[2]. Thus

$$d[\mathrm{pc}] = \frac{1}{p[''']} \tag{1.21}$$

i.e. a star with a parallax angle of n arcseconds has a distance of $1/n$ parsecs. Since one arcsecond is $1/(360 \times 60 \times 60) = 1/206265$ fraction of 2π, a parsec corresponds to 206 265 AU.

To express finally the unit pc in a known unit like cm, we have to determine first the Earth-Sun distance, the astronomical unit AU. As first step one measures the distance d to an inner planet at the time of its greatest elongation (i.e. the largest angular distance ϑ between the Sun and the planet). Nowadays, one uses a radar signal to Venus. Then $\mathrm{AU} = d/\cos\vartheta$, cf. the right panel of Fig. 1.6, and one finds $1\mathrm{AU} = 1.496 \times 10^{13}\,\mathrm{cm}$. As result, one determines one parsec as $1\,\mathrm{pc} = 206,265\,\mathrm{AU} = 3.086 \times 10^{18}\,\mathrm{cm} = 3.26\,\mathrm{lyr}$.

1.5 Stellar luminosity and absolute magnitude scale

The total luminosity L of a star is given by the product of its surface $A = 4\pi R^2$ and the radiation emitted per area σT^4,

$$L = 4\pi R^2 \sigma T^4\,. \tag{1.22}$$

Since the brightness or energy flux $\mathcal{F}$ was defined as $\mathcal{F} = E/(At) = L/A$, we recover the inverse-square law for the energy flux at the distance $r > R$ outside of the star,

$$\mathcal{F} = \frac{L}{4\pi r^2}\,. \tag{1.23}$$

The validity of the inverse-square law $\mathcal{F}(r) \propto 1/r^2$ relies on the assumptions that no radiation is absorbed and that relativistic effects can be neglected. The later condition requires in particular that the relative velocity of observer and source is small compared to the velocity of light.

Ex.: Luminosity and effective temperature of the Sun.
The energy flux received from the Sun at the distance of the Earth, $d = 1$ AU, is equal to $\mathcal{F} = 1365\ \mathrm{W/m^2}$. (This energy flux is also called "Solar Constant.") The solar luminosity $L_\odot$ follows then as

$$L_\odot = 4\pi d^2 \mathcal{F} = 4 \times 10^{33}\mathrm{erg/s}$$

and serves as a convenient unit in stellar astrophysics. The Stefan-Boltzmann law can then be used to define with $R_\odot \approx 7 \times 10^{10}$ cm the effective temperature of the Sun, $T_\mathrm{e}(\mathrm{Sun}) \equiv T_\odot \approx 5780$ K.

The brightness discussed in Sec. 1.1 is based on the observed energy fluxes on Earth and does not account for the different distance of stars. It is therefore called "apparent magnitude". To compare the intrinsic luminosities of stars, one defines the absolute magnitude M as the one a star would have at the standard distance $d = 10$ pc.

[2] A degree, i.e. the 1/360 part of a circle is divided into 60 arcminutes and 60×60 arcseconds, $360^\circ = 360 \times 60' = 360 \times 60 \times 60''$.

Using that $\mathcal{F} \propto 1/d^2$, the relation between m and M is from Eq. (1.2)

$$m - M = 2.5 \log \left(\frac{d}{10\text{pc}} \right)^2 = 5 \log \frac{d}{10\text{pc}} \,. \tag{1.24}$$

The quantity $m - M$ is also called distance modulus.

Measuring the received energy flux $\mathcal{F}$ from a star with known absolute magnitude establishes also the conversion between M and usual cgs or SI units,

$$L = 3.02 \times 10^{35} \, \frac{\text{erg}}{\text{s}} \times 10^{-0.4M} = 3.02 \times 10^{28} \text{W} \times 10^{-0.4M} \,. \tag{1.25}$$

Exercises

1. Show that the mean distance between black-body photons corresponds approximately to the wave-length of photons from the maximum of the Planck distribution, $n_\gamma \approx 1/\lambda_{\text{max}}^3$.

2. Derive Wien's law both for frequency and wavelength. Hint: The equations $\text{e}^x(5-x) = 5$ and $\text{e}^y(3-y) = 3$ have the solutions $x = 4.965$ and $y = 2.821$. Show that $\lambda_{\text{max}} \neq c/\nu_{\text{max}}$.

3. Show that $B_\nu(T_1) > B_\nu(T_2)$ for all ν, if $T_1 > T_2$.

4. Derive the Earth-Sun distance D assuming that both are perfect blackbodies. Use $R_\odot = 7 \times 10^{10}$ cm, $T_\odot = 5780$ K and $T_\oplus = 288$ K.

2 Spectral lines and their formation

Classical astronomy was mainly concerned with the determination of planetary orbits – an understanding of the physical properties of the Sun or stars in general seemed to be impossible. The breakthrough that paved the way to stellar astrophysics was the spectral analysis of stellar light: Already Fraunhofer identified one prominent of the dark lines in the continuous spectrum of the Sun with the one emitted by salt sprinkled in a flame.

The field of spectral analysis was then established by Kirchhoff and Bunsen in 1859: They realized that each element produces its own pattern of spectral lines and can be thereby identified. Thus one could show that the solar light contains strong absorption lines of calcium and of previously unknown elements as helium. However, a proper understanding which factors determine the strength of various spectral lines in the spectrum of the Sun and other stars become only possible in the 1920s, after the advent of quantum mechanics. Thus we start with a brief review of the simplest atom, the hydrogen atom, in a semi-classical picture.

2.1 Bohr-Sommerfeld model for hydrogen-like atoms

Bohr postulated the existence of discrete, stationary orbits for electrons in atoms. More precisely, he required that the angular momentum $J = m_e v r$ of an electron is a multiple of Planck's constant $\hbar \equiv h/(2\pi)$,

$$J = m_e v r = n\hbar, \qquad n = 0, 1, 2, \ldots \tag{2.1}$$

If the electron is on a bound orbit, the electric force is balanced by the centrifugal force,

$$\frac{m_e v^2}{r} = \frac{e^2}{r^2} \tag{2.2}$$

and the quantization condition becomes

$$m_e v^2 = \frac{n^2\hbar^2}{m_e r_n^2} = \frac{e^2}{r_n}\,. \tag{2.3}$$

Solving for r_n gives the allowed electron orbits in the Bohr model,

$$r_n = \frac{n^2\hbar^2}{m_e e^2}, \qquad n = 1, 2, \ldots \tag{2.4}$$

The total energy of an electron moving in the Coulomb field of an (infinitely heavy) nucleus with electric charge e is the sum of its kinetic and potential energy,

$$E = E_{\rm kin} + E_{\rm pot} = \frac{1}{2} m_e v^2 - \frac{e^2}{r} = -\frac{e^2}{2r}\,, \tag{2.5}$$

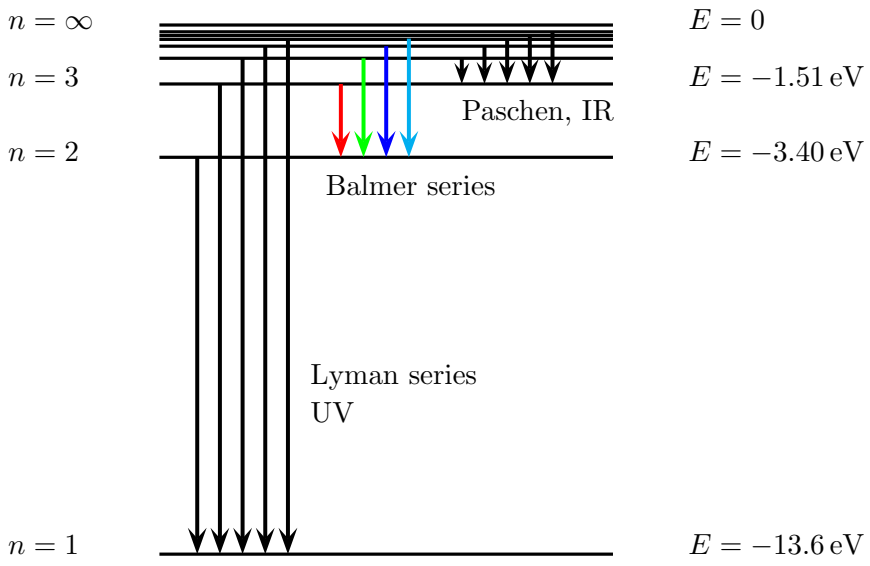


Figure 2.1: The energy levels of the hydrogen spectrum with some lines of the first three spectral series.

where again Eq. (2.5) was used. Inserting r_n gives the allowed stationary energy levels in a hydrogen atom,

$$E_n = -\frac{m_e e^4}{2n^2\hbar^2}, \qquad n = 1, 2, \ldots \tag{2.6}$$

In transition between different levels, only photons with discrete energies

$$h\nu = |E_n - E_m| \tag{2.7}$$

can be absorbed or emitted. An empirical formula describing the wave-length of visible hydrogen spectral lines was found by Balmer already in 1885,

$$\frac{1}{\lambda} = R\left(\frac{1}{4} - \frac{1}{n^2}\right), \tag{2.8}$$

where R is the Rydberg constant, $1/R \approx 91.2$ nm. Rewriting the general formula (2.6) for wavelengths $\lambda = c/\nu$ and comparing it with the Balmer formula gives

$$\frac{hc}{\lambda} = |E_n - E_m| = \frac{m_e e^4}{2\hbar^2}\left|\frac{1}{n^2} - \frac{1}{m^2}\right| \tag{2.9}$$

with $n = 2$ and $R = \pi m_e e^4/(c\hbar^3)$. Thus the Balmer series corresponds to the transitions from a level $m > 2$ to $n = 2$. Transition to the ground level $n = 1$ are called Lyman series and are in the UV, cf. also Fig. 2.1.

Apart from transition between two discrete levels, a free electron with $E > 0$ can become bound or a bound electron can be emitted by an atom. These are called free-bound or bound-free transition. In this case, the energy of the emitted or absorbed photon depends on the kinetic energy of the free electron and can be arbitrary.

Ex.: Find the energy and radius of the ground state of a hydrogen atom.
Insert $n = 1$ in (2.6) and (2.4). Simplest in natural unit ($\hbar = c = 1$) with $m_e = 512$ keV and $\mathrm{GeV}^{-1} = 5.06 \times 10^{13}$ cm (cf. Appendix B), and Sommerfeld's fine structure constant $\alpha = e^2/(\hbar c) \approx 1/137$, [in SI units $\alpha = e^2/(4\pi\varepsilon_0\hbar c) \approx 1/137$]

$$E_1 = -\frac{me^4}{2\hbar^2} = -\frac{\alpha^2}{2}\, m_e c^2 = -\frac{512\mathrm{keV}}{2 \times 137^2} \approx 13.6\mathrm{eV}\,,$$

$$r_1 = \frac{\hbar^2}{me^2} = \frac{1}{\alpha}\,\frac{\hbar}{m_e c} = \frac{137\,\mathrm{cm}}{0.512 \times 10^{-3} \times 5.06 \times 10^{13}} \approx 5.3 \times 10^{-9}\mathrm{cm}\,.$$

Kirchhoff's laws Using Bohr's model we can understand the laws deduced by Kirchhoff empirically:

- A dense gas or a solid body emits continuous radiation without spectral lines. The atoms and the radiation in the body are in thermal equilibrium, and the emitted radiation follows the Planck distribution.
- A hot, diffuse gas produces emission lines. The lines correspond to the energy difference of atomic levels $n \to m < n$, photons are emitted.
- A cool, diffuse gas in front of a continuous sources produces absorption lines. Photons which energy correspond to the the energy difference of an appropriate atomic level are absorbed.

2.2 Formation of spectral lines

A prominent line in the solar spectrum corresponds to the $n = 2 \to n = 3$ transition and is called Balmer-α or Hα. The observation of this line means that a certain fraction of hydrogen in the solar atmosphere must be in the second, excited state. How are the populations of the different excited states governed?

For a classical gas in thermal equilibrium, the population n_i of the state i with energy E_i and degeneracy g_i is given by a Boltzmann distribution, with

$$n_i = \mathcal{Z}^{-1} g_i \exp(-E_i/kT)\,, \tag{2.10}$$

where $\mathcal{Z}$ can be thought of as a normalization constant. Excitations with energy $E \gg kT$ are exponentially suppressed.

For the population of an excited stated $n > 1$ relative to the ground state n_1 it follows

$$\frac{n_i}{n_1} = \frac{g_i}{g_1} \exp\left(-\frac{E_i - E_1}{kT}\right)\,. \tag{2.11}$$

Thus if $(E_2 - E_1) \gg kT$, then the state $n = 2$ is only weakly populated and the Hα line is faint. Increasing the temperature so that $(E_2 - E_1) \sim kT$, the relative occupation of the level $n = 2$ and the strength of the Hα line increases too. Increasing the temperature even further, electrons populate higher and higher levels, reducing the relative population of the $n = 2$ level and in turn the strength of the Hα line. Finally, hydrogen becomes ionized, when the temperature is higher than its binding energy. Thus the following picture for the strength of the Hα line emerges:

Spectral class	Temperature T_e/K	main spectral feature
O	30000	hydrogen faint, ionized helium
B	20000	hydrogen moderate, neutral helium moderate
A	10000	hydrogen strong, neutral helium very faint
F	7000	hydrogen moderate, single ionized heavy elements
G	6000	hydrogen relatively faint, single ionized heavy elements
K	4000	hydrogen faint, single ionized heavy elements
M	3000	hydrogen very faint, neutral atoms, molecules

Table 2.1: Important spectral features of main-sequence stars.

- For a cold star, $kT \ll E_1 - E_2$, most hydrogen is in the ground state; the Hα line is very weak.
- Increasing the temperature, $kT \approx (E_2 - E_1)$, the $n = 2$ level becomes populated; the Hα line becomes strong.
- Increasing further the temperature, $kT \gtrsim (E_2 - E_1)$, higher levels become populated and $n = 2$ level less; the Hα line becomes weaker.
- Increasing even further the temperature, $kT \gtrsim E_1$, Hydrogen becomes ionized; the Hα line disappears.

Stars were classified according to the strength of the Hα lines as A, B, F, G, K, M, O. These classes are additionally subdivided into A0, A1, ... A9, B0,.... Ordered by (decreasing) temperature, the spectral classes form the sequence O, B, A, F, G, K, M, cf. also table 2.1. (A mnemonic is "oh, be a fine girl/guy, kiss me".)

For the visible $n = 2$ transitions, it follows

$$E_2 - E_1 = -(1/4 - 1) \times 13.6\,\mathrm{eV} \quad \text{or} \quad (E_2 - E_1)/k \approx 1.2 \times 10^5\,\mathrm{K}.$$

Thus in the solar atmosphere with $T_e \approx 5700\,\mathrm{K}$ only the tiny fraction $2/8 \exp(-20.8) \approx 10^{-10}$ of all hydrogen atoms is in the $n = 2$ level. Therefore the Balmer lines in the solar spectrum are relatively faint, although the Sun consists mainly of hydrogen. On the other hand, the ionization energy of calcium is 6.11 eV, and thus a factor 2 smaller than the one hydrogen. Since the population of the different levels depends exponentially on the energy, already a factor 2 leads to large differences: Practically all calcium atoms in the solar atmosphere are single ionized and in the ground state, able to absorb a photon in the visible wave-length range. More generally, the single ionized lines of relatively heavy elements as calcium are prominent in F, G and K stars, cf. Table 2.1.

2.3 Hertzsprung-Russel diagram

Between 1905 and 1913, Hertzsprung and Russel examined the distribution of stellar luminosities for different spectral classes. The latter quantity characterizes the surface temperature T, and since the luminosity is proportional to T^4 and the stellar surface $\propto R^2$, one expects a wide distribution in this diagram. As shown in Fig. 2.2, this is not the case and stars appear only in well-defined regions of the Hertzsprung-Russel (HR) diagram. Thus arbitrary combinations of T and L, and thus of R and L are not allowed.

Most stars are found in a narrow band from the left-upper to the right-lower corner, the so-called main sequence. Stars of the main sequence of the same temperature have approximately the same luminosity and thus also radius. This correlation has to be explained by any theory of stellar astrophysics.

Apart from the main sequence, there are two other star populations. Above the main sequence, there are the giant and super giant stars: They are more luminous for the same temperature as main sequence stars, hence their radius has to be larger. Much below the main sequence stars (10 magnitudes fainter) and with roughly white spectral color, there are the so called white dwarf stars.

There exists various variations of the original Hertzsprung-Russel diagram. For instance, the color index can replace the temperature or the apparent magnitude can be used as measure for the luminosity, if one discusses only stars from the same cluster.

The color indices in Eq. (1.4) are normalized so that their differences $m_{\lambda_1} - m_{\lambda_2}$ are zero for stars of the spectral class A0. Note also the colors U, B, V,... denote apparent magnitudes, although astronomers use capital letters for them.

Exercises

1. Consider a gas of neutral hydrogen gas at temperature T, described by a Boltzmann distribution $n_i = \mathcal{Z}^{-1} g_i \exp(-E_i/kT)$ with $g_n = 2n^2$.
   i) At what temperature T will equal number of atoms have electrons in the ground state and in the second excited state ($n = 3$)?
   ii) What predicts the Boltzmann distribution for $T \to \infty$? What happens in reality?

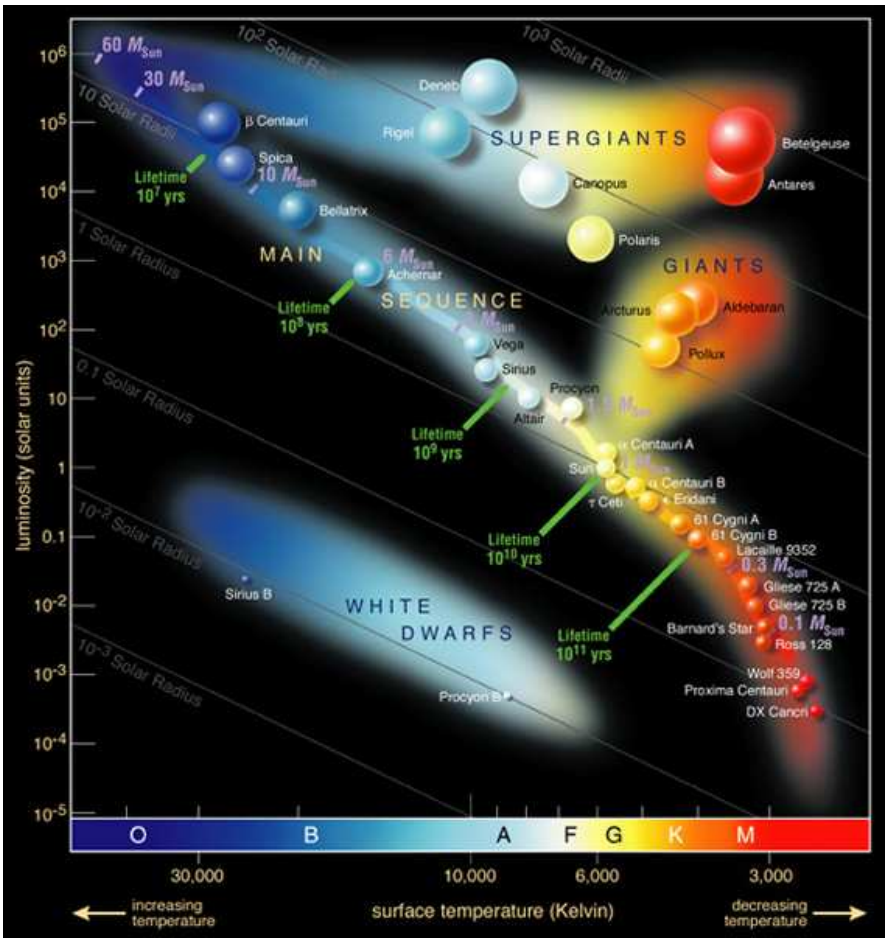


Figure 2.2: A Hertzsprung-Russel diagram. The position of a star along the main-sequence defines uniquely all important stellar parameters: Mass, radius, temperature, lifetime.

3 Telescopes and other detectors

Progress in astronomy has been driven largely by the exploration of new wavelengths ranges of the electromagnetic spectrum. The first discovery of an "extraterrestrial" signal beyond the visible light was the observation of "cosmic rays" by Victor Hess 1912, whose origin is still unresolved. In 1935, Karl Jansky discovered an extraterrestrial radio signal and could associate it to the center and the plane of the Milky Way. After world war II, the radio band has been studied systematically. Starting from the seventies, X- and gamma-ray astronomy as well as the IR wave-lengths have been explored with satellites. Since the last decade, high-energy air Cherenkov telescopes are detecting photons with energies $E \gtrsim 100\,\mathrm{GeV}/c^2$. This extension of the explored wave-length range made it also possible to study non-thermal radiation.

At the same time there has been a continuous improvement of light receivers: The naked eye was replaced first by photo plates and then by charge-coupled devices (CCD) similar to those used in present days digital photo cameras. Another important improvement has been the use of computers for data correction and analysis.

Further progress is expected by attempts to go beyond electromagnetic radiation: There are large efforts underway to use charged particles (cosmic rays), neutrinos (from the Sun, Galactic supernovae or other sources), and gravitational waves as new means to study the Universe.

3.1 Optical telescopes

Optical telescopes can be divided into two principal classes: i) refracting telescopes using lenses and ii) reflecting telescopes using mirrors (plus lenses). From the refracting telescopes two principal subclasses exist: Kepler's version with two concave lenses and Galileo's version with one concave and one convex lens.

3.1.1 Characteristics of telescopes

Magnification V: Many objects (binary stars, galaxies) that appear point-like to the naked eye show structure viewed by a telescope. From the left panel of Fig. 3.1, it follows that the magnification V depends on the ratio of the focal lengths f_i and the apertures D_i as

$$V = \frac{\phi_2}{\phi_1} = \frac{f_1}{f_2} = \frac{D_1}{D_2}\,. \tag{3.1}$$

If an objects subtends the angle α to the unaided eye, a telescope magnifies this angle by the factor V.

Resolution: The smallest angular distance at which two object can be separated in the image is called resolving power or resolution of the telescope. A lower limit is set by diffraction, i.e. the interference of light rays from different positions in the aperture plane of the telescope.

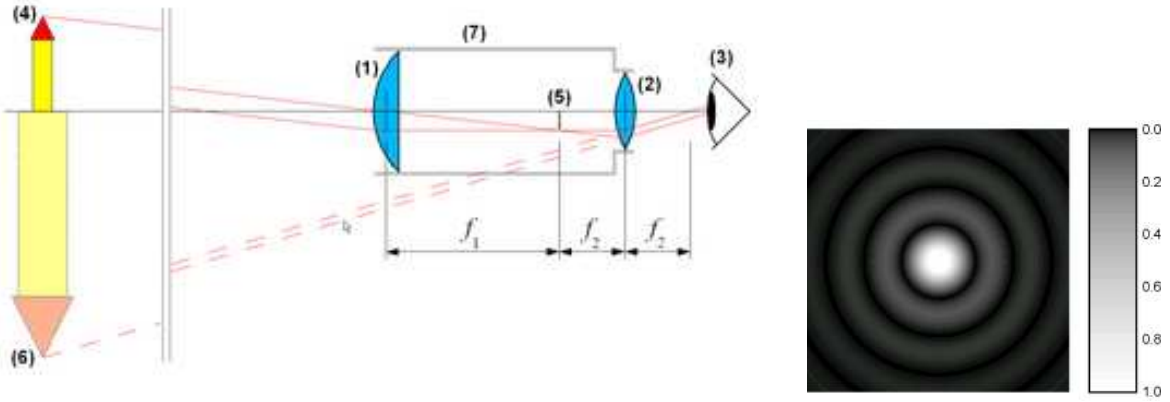


Figure 3.1: Left: Schematic view of a refracting telescope (Keplerian version). Right: Diffraction pattern.

Consider for an one-dimensional example, when destructive interference happens: We divide the aperture into $2n$ segments of length $d = D/(2n)$ and pair light rays separated by d. The phase difference between the pairs is $l \sin\vartheta = D/(2n) \sin\vartheta$ and thus destructive interference happens for

$$\frac{D}{2n} \sin\alpha = \frac{\lambda}{2} \qquad \text{or} \qquad \alpha_n \approx \sin\alpha_n = n\,\frac{\lambda}{D}\,, \qquad n = 1, 2, \ldots \tag{3.2}$$

For a circular aperture D, the first diffraction minimum is at $\alpha_1 \approx 1.22\lambda/D$. Two objects can be resolved, if their angular distance is larger than α_1 ("Rayleigh criteria"). For instance, a 5 m optical telescope has a resolution of

$$\alpha_1 \approx 1.22\,\frac{\lambda}{D} \approx 1.22\,\frac{550\times10^{-9}\mathrm{m}}{5\,\mathrm{m}} \approx 1.34\times10^{-7}\,\frac{60^2\times180''}{\pi} \approx 0.03''\,, \tag{3.3}$$

while the human eye resolves objects with $1'$ angular distance.

Light gathering power The aperture or diameter D of the telescope determines the collecting area for light that is then focused on the receptor surface. Compared to the human eye with $D \approx 5$mm, a 5 m telescope allows to see fainter objects with magnitude difference

$$m_1 - m_2 = 2.5\log\left(\frac{D_1^2}{D_2^2}\right) = 2.5\log\left(\frac{25}{25\times10^{-6}}\right) = 15\,, \tag{3.4}$$

i.e. with a 5 m telescope stars are visible down to $m = 6 + 15 = 21$. The light gathering power is a measure of how strongly light is concentrated on the receptor surface and depends on the kind of object considered. The apparent brightness b of the image of

- a point-like object scales as $b \propto D^2$: The amount of light gathered increases $\propto D^2$ and is still collected on a single pixel of a CCD or on one nerve of the human eye.
- an extended objects is constant: The increase of collected light is compensated by the larger picture of the object.

Exposure and quantum efficiency With photo plates or photoelectric detectors one can collect light for hours, while the naked eye takes exposures of $\sim 1/20$ seconds. Larger exposure extend thus additionally the limiting magnitude of an telescope.

While the eye or a photo plate has a quantum efficiency of about 1%, i.e. 1% of all photons are detected, modern detectors called charge-coupled devices (CCD) come close to a quantum efficiency of about 100%. These are semiconducting detectors where a photon creates (similar as in the photoelectric effect) electrons.

Spectrographs Simplest example is a prism that has typically a wave-length dependent refractive index, $n \propto \lambda^2$. Positive interference occurs in a **diffraction grating** with grating constant $1/D$ when the path length difference between neighboring rays is

$$D(\sin\phi_i + \sin\phi_f) = m\lambda, \qquad m = \pm 1, \pm 2, \ldots \tag{3.5}$$

3.1.2 Problems and limitations

Chromatic aberration Since the refractive index n of a lens (as any material) is wavelength dependent, a picture of a refracting telescopes cannot be focused on a single common point for different wavelengths. Together with the easier production and support of mirrors, this is the main reason for the dominance of reflecting telescopes.

Surface quality: The surface of mirror or lenses has to be accurate generally to $\lambda/20$. For an optical telescope, this corresponds to $\lambda/20 \sim 25\,\text{nm}$, i.e. around 250 atom layers. Similarly, deviations of the various telescope parts due to thermal stress or gravity should be smaller than $\lambda/20$.

Atmospheric seeing: For large ground-based telescopes, the resolution is limited by atmospheric seeing: Turbulence in the air leads to time-dependent refractive index along different lines-of-sight, $n(x) + \delta n(x,t)$, and causes the images of point-sources to break up into speckle patterns, which change very rapidly with time. This limitation can be overcome partially by placing telescopes at high-altitude observatories such as Mauna Kea or La Palma and using adaptive optics or image correcting with the help of computers for ground-based telescopes. Alternatively, one can go above the atmosphere, i.e. to space.

3.2 Other wave-length ranges

Radio telescopes The simplest realization would be a $\lambda/2$ antenna where electromagnetic waves induces resonance currents in it. To increase the sensitivity, one builds large parabolic antenna with the radio receiver in its focal point. Then the signal is electronically amplified.

Since the wavelengths is in the cm to meter range, the parabolic antenna can be made as a simple metallic grid. This allows the construction of large dishes up to 100 m. On the other hand, the resolution of a radio telescope is according to Rayleigh's criteria $\propto \lambda$ and thus very poor compared to optical telescopes. This problem can be partially overcome by combining the information from telescopes distributed in an array or even on different continents (Very Long Baseline interferometry).

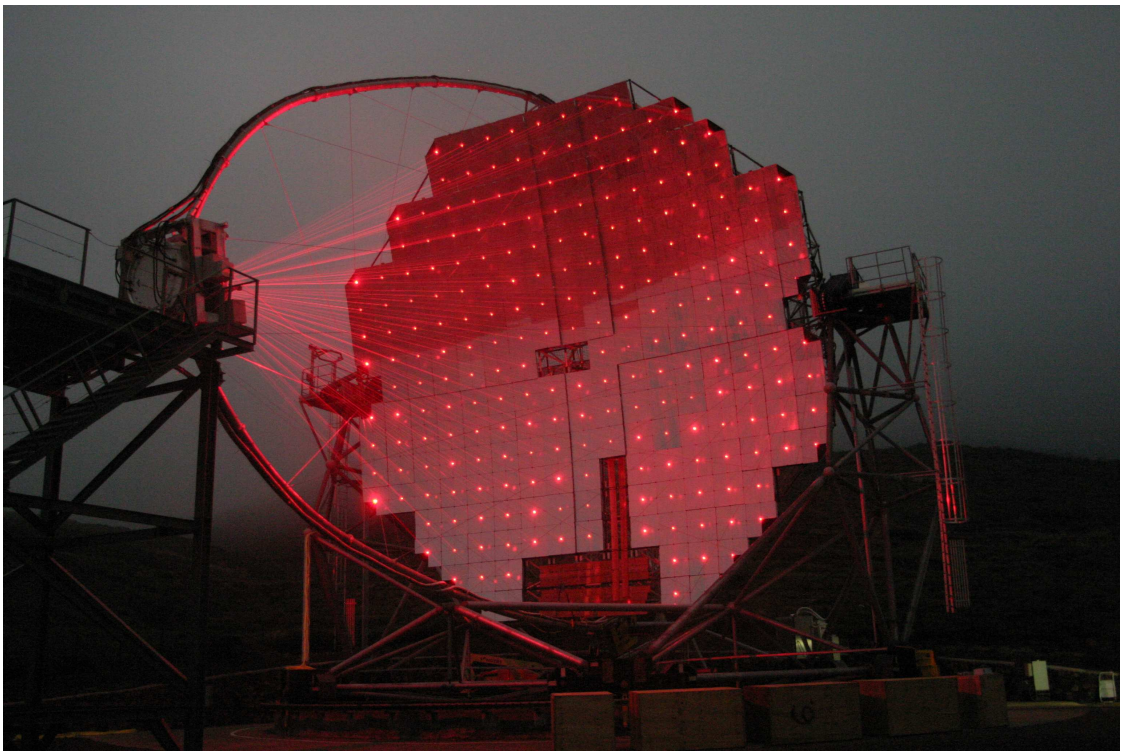


Figure 3.2: The MAGIC air Cherenkov telescope: lasers are used as control for the adaptive optics.

Instruments for high-energy astronomy Extending observations to the UV and higher energies meets three main obstacles: First, the Earth's atmosphere is opaque to photons with energy above 10 eV, meaning that to observe photons in the UV and beyond directly requires the placement of a detector above the Earth atmosphere. A major turning point in X-ray and gamma-ray astronomy was therefore the launch of the first satellite-borne telescope, SAS-2, in 1972. As an additional obstacle, geometrical optic ceases to work for photons in the X-ray range ($0.1\,\mathrm{keV} < E < 100\,\mathrm{keV}$) and above. Finally, the flux of photons decreases with energy strongly.

Small angle reflection is still possible for X-rays, if the incidence angle is very small, and thus one uses a special type of grazing incidence mirror telescopes. Gamma-rays ($E >$ MeV and higher energies) can produce e^+e^- pairs in the field of a nucleus. These electrons and positrons are either detected directly or, especially if $E \gg$ MeV, produce secondary photons forming thereby an electromagnetic cascade. Main limitation of satellite experiments is their small collection area, which limits their use to energies $\lesssim 100\,\mathrm{GeV}$. On the other hand, 100 GeV is the limit where the electromagnetic cascade in the earth's atmosphere from the initial photon can be detected. While the cascade dies out high in the atmosphere below 10 TeV, showers are still detectable via the Cherenkov[1] emission of relativistic electrons and muons. An example for such an atmospheric Cherenkov telescope is shown in Fig. 3.2.

[1] A charged particle that moves in a medium with refractive index $n > 1$ faster than the speed of light, $v > c/n$, emits electromagnetic "Cherenkov radiation."

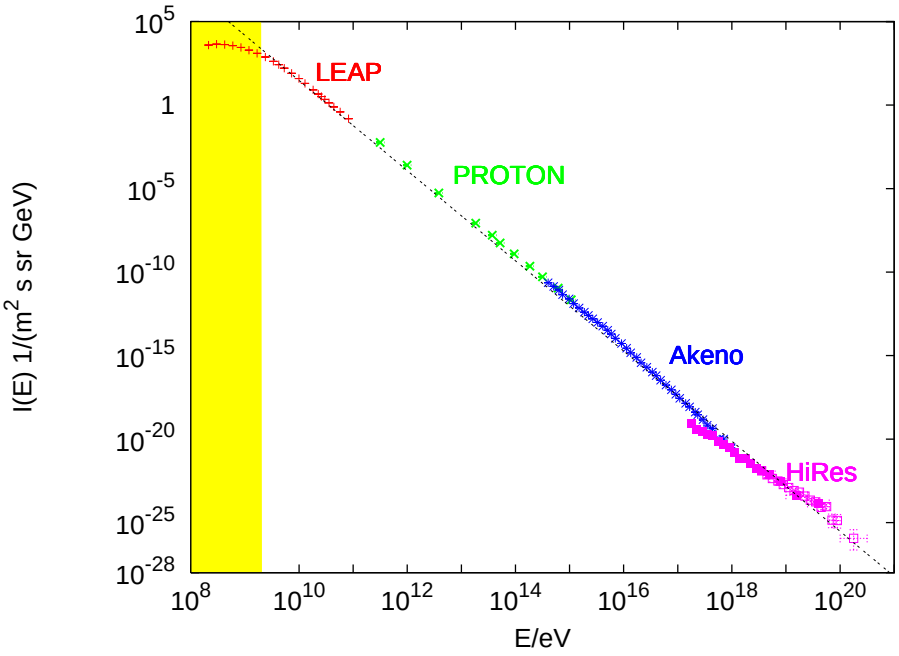


Figure 3.3: The differential intensity $I(E)$ of cosmic rays measured by four different experiments as function of kinetic energy E. The energy region influenced by the Sun is marked yellow and an $1/E^{2.7}$ power-law is also shown as dashed line.

3.3 Going beyond electromagnetic radiation

In principle, all stable particle types may be used to obtain information about the universe: Additionally to photons, these are the charged elementary particles (electrons e^- and protons p) and the stable nuclei. Moreover, the known stable neutral particles are neutrinos ν and the graviton.

Charged particles—Cosmic rays Victor Hess discovered 1912 on a balloon flight that ionizing radiation increases with altitude. He performed observations also during a solar eclipse, but found no decrease. Hence he concluded that the Sun can hardly be considered the source. Later, the study of their deflections in the geomagnetic field showed that most cosmic rays are protons.

The cosmic ray spectrum, a nearly featureless power-law extending over eleven decades in energy up to a few $\times\, 10^{20}$ eV, is shown in Fig. 3.3. The power-law form of the cosmic ray spectrum indicates that they are produced via non-thermal processes, in contrast to all other radiation sources like e.g. stars or (super-) novae known until the 1950's. Since our galaxy, the Milky Way, contains a magnetic field with strength $\sim 3\mu$G in the Galactic disc of thickness $d \sim 100\,\mathrm{pc}$, charged particles are deflected: In a homogeneous magnetic field, particle with charge Ze gyrate with Larmor radius

$$R_L = \frac{cp}{ZeB} = \frac{\mu\mathrm{G}}{B}\,\frac{E}{Z \times 3 \times 10^{15}\mathrm{eV}}\,\mathrm{pc}\,. \tag{3.6}$$

Thus one cannot associate sources to the observed arrival direction of cosmic rays, except perhaps at the highest energies, $E \gtrsim 10^{19}\,\mathrm{eV}$ or the last energy decade shown in Fig. 3.3.

3.3.1 Neutrinos

The main advantage of neutrinos compared to photons is the extreme weakness of their interactions. For instance, a neutrino produced inside the Sun interacts only with a probability 10^{-9} on its way out. Thus neutrinos can provide information about the interior of the stars, while photons inform us only about their surface. Neutrinos are produced in weak interactions like nuclear beta decay $n \rightarrow p + e^- + \bar{\nu}_e$ that are part of the fusion processes in stars as we will discuss later in detail. Therefore they are copiously in the interior of the Sun or a supernova and are a key tool to test our theories of stellar structure and evolution.

The weakness of their interactions is also their main disadvantage compared to photons: They are extremely difficult to detect. Solar neutrinos are observed since the mid-1960s, and in 2002 Raymond Davis Jr. and Masatoshi Koshiba won the Nobel Prize in Physics "for the detection of cosmic neutrinos".

3.3.2 Gravitational waves

Analogously as Maxwell's equations predict electromagnetic waves, general relativity predicts gravitational waves as a distortion of space-time. Passing of a gravitational wave leads (similar to a Lorentz contraction) to a time-dependent change of length scales, $h(t) = \delta L(t)/L$. This change h is proportional to the amplitude of the gravitational wave, and thus the distortion is inversely proportional to the distance, $h \propto 1/d$. Increasing the sensitivity of a gravitational wave experiment by a factor ε increases the number of potential sources therefore as ε^3, in contrast to an increase as $\varepsilon^{3/2}$ for photon or neutrino experiments.

Since gravity is very weak, the expected deviations are tiny, $h \lesssim 10^{-20}$. There are several laser interferometer aiming at detecting gravitational waves.

4 Basic ideas of special relativity

4.1 Time dilation – a Gedankenexperiment

Consider a light clock, i.e. a light pulse traveling between two mirrors separated by the distance L. The time for a round trip is $t_0 = 2L/c$. Now consider the clock, i.e. the two mirrors, moving with velocity v, cf. Fig. 4.1. Then the light needs for bouncing once back and forth the time $t > t_0$,

$$t^2 = \frac{4}{c^2}\left[L^2 + (vt/2)^2\right] = \underbrace{\frac{4L^2}{c^2}}_{t_0^2}\left[1 + \Big(\underbrace{\frac{vt}{2L}}_{vt/ct_0}\Big)^2\right] = t_0^2 + (v/c)^2 t^2 \,. \tag{4.1}$$

Solving for t gives

$$t = \frac{t_0}{\left[1 - (v/c)^2\right]^{1/2}} \,, \tag{4.2}$$

i.e. the famous time dilation formula. Is this just a property of a badly constructed clock or of fundamental importance? The answer given by experiments is that time for fast moving observers indeed slows down.

4.2 Lorentz transformations and four-vectors

4.2.1 Galilean transformations

Newton's gravitational law as well as the Coulomb law are examples for an instantaneous action,

$$\mathbf{F}(\mathbf{x}) = K \sum_i \frac{\mathbf{x} - \mathbf{x}_i}{|\mathbf{x} - \mathbf{x}_i|^3} \,. \tag{4.3}$$

The force $\mathbf{F}(\mathbf{x}, t)$ depends on the distance $\mathbf{x}(t) - \mathbf{x}_i(t)$ to all sources i (electric charges or masses) at the same time t, i.e. the force needs no time to be transmitted from $\mathbf{x}_i$ to $\mathbf{x}$.

The force law is invariant under changes $\mathbf{x}' = R\mathbf{x} + \mathbf{d}$, if $|\mathbf{r}'| = |\mathbf{r}|$. The translation d could be time-dependent via $d = vt$. The condition that the norm of the vector $\mathbf{r}$ is invariant, $\mathbf{r} = \mathbf{r}'$, restricts the possible transformation matrices R,

$$\mathbf{r}'^T\mathbf{r}' = \mathbf{r}^T R^T R\mathbf{r} = \mathbf{r}^T\mathbf{r} \,, \tag{4.4}$$

as $R^T R = 1$. (T denotes the transposed vector or matrix). Hence rotations R and translations d keep a force law of type (4.3) invariant. In two dimensions, the rotation matrix R can be chosen as

$$R = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} . \tag{4.5}$$

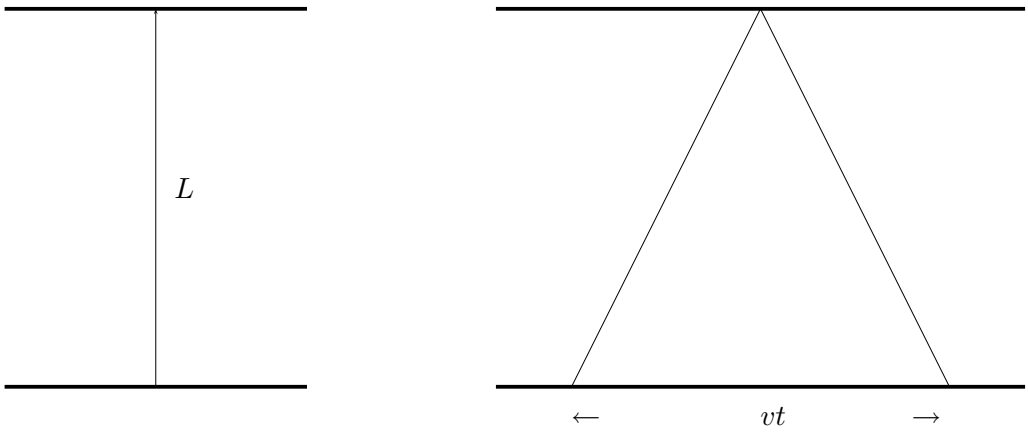


Figure 4.1: A clock consisting of a light pulse travelling between two mirrors; left at rest and right moving with velocity v.

4.2.2 *** Lorentz transformations ***

Comparing the non-relativistic Poisson equation $\Delta\phi = 0$ and Maxwell's wave equation, $(\Delta - (1/c^2)\partial_t^2)A_\mu = 0$ suggests that the latter are invariant under coordinate transformations that keep the space-time interval

$$\Delta s^2 \equiv \mathbf{x}^2 - c^2t^2 = \mathbf{x}'^2 - c^2t'^2 \tag{4.6}$$

invariant. This means that both the time and the space difference between two events measured in the coordinate system K and K' may be different, but that the difference $\Delta s^2 \equiv \mathbf{x}^2 - c^2t^2$ is the same for all inertial systems. Here we call "inertial system" all coordinate systems that move force-free.

We want now to determine the equivalent to the rotation matrices R that are called Lorentz transformations L. If we replace t by it in Δs^2, the difference between two space-time events becomes the normal euclidean distance. Similarly, the identity $\cos^2\alpha + \sin^2\alpha = 1$ for an imaginary angle $\psi = i\alpha$ becomes $\cosh^2\psi - \sinh^2\psi = 1$. Thus a close correspondence exists between rotations R in euclidean space that leave $\Delta\mathbf{x}^2$ invariant and Lorentz transformations L that leave Δs^2 invariant. We try therefore as a guess

$$x = ct'\sinh\psi + x'\cosh\psi\,, \tag{4.7}$$
$$y = y'\,, \qquad z = z'\,, \tag{4.8}$$
$$ct = ct'\cosh\psi + x'\sinh\psi\,. \tag{4.9}$$

Compared to normal coordinate rotation in euclidean space, the trigonometric functions are replaced by hyperbolic ones.

Consider now in system K the origin of the system K'. Then $x' = 0$ and

$$x = ct'\sinh\psi\,, \tag{4.10}$$
$$ct = ct'\cosh\psi\,. \tag{4.11}$$

Dividing the two equations gives

$$\frac{x}{ct} = \tanh\psi\,. \tag{4.12}$$

Since x/t is the relative velocity of the two systems, we have identified the physical meaning of the imaginary "rotation angle ψ" as

$$\tanh\psi = \frac{v}{c} \equiv \beta\,. \tag{4.13}$$

Using the following identities

$$\sinh x = \frac{\tanh x}{\sqrt{1-\tanh^2 x}} = \frac{\beta}{\sqrt{1-\beta^2}} \tag{4.14}$$

$$\cosh x = \frac{1}{\sqrt{1-\tanh^2 x}} = \frac{1}{\sqrt{1-\beta^2}} \equiv \gamma \tag{4.15}$$

together with (4.13) in (4.7) gives the standard form of the Lorentz transformations,

$$x = \frac{x' + vt'}{\sqrt{1-(v/c)^2}} = \gamma(x' + \beta ct') \tag{4.16}$$

$$ct = \frac{ct' + v/cx'}{\sqrt{1-(v/c)^2}} = \gamma(ct' + \beta x')\,. \tag{4.17}$$

4.2.3 Energy and momentum

Another important example for a four-tupel consisting (from a non-relativistic point of view) of a vector and a scalar and transforming like $(ct, \mathbf{x})$ under Lorentz transformations, is $(E, c\mathbf{p})$. The Lorentz transformation written in this case is

$$E = \gamma(E' + \beta c p'_x) \tag{4.18}$$

$$cp_x = = \gamma(cp'_x + \beta E')\,. \tag{4.19}$$

If we choose as system K' the rest system of the particle, then $p'_x = 0$ and $\gamma = 1$. If we would set the rest energy $E' \equiv E_0$ equal to zero, then the energy E would be zero for arbitrary γ. The correct value of E_0 can be obtained by considering the non-relativistic limit,

$$\gamma = \frac{1}{(1-\beta^2)^{1/2}} \approx 1 + \frac{\beta^2}{2} \tag{4.20}$$

and

$$\gamma E_0 \approx E_0 + \frac{E_0 v^2}{2c^2} \stackrel{!}{=} \text{const.} + \frac{mv^2}{2}\,. \tag{4.21}$$

Thus we obtain the correct non-relativistic limit, if a particle at rest has the energy $E_0 = mc^2$.

4.2.4 Doppler effect

We consider as last example how the wave-vector of a photon $(\omega/c, \mathbf{k})$ transforms under a Lorentz transformation,

$$\frac{\omega'}{c} = \frac{\frac{\omega}{c} - \frac{V}{c}k_x}{[1-(V/c)^2]^{1/2}}\,. \tag{4.22}$$

Setting $k_x = k\cos\alpha = \frac{\omega}{c}\cos\alpha$, where α the angle between the wave vector and the direction of the movement of the source, we obtain the relativistic Doppler formula

$$\frac{\omega}{\omega'} = \frac{\left[1-(V/c)^2\right]^{1/2}}{1-(V/c)\cos\alpha}\,. \tag{4.23}$$

For small relative velocities, $v \ll c$, the quadratic term in the nominator can be neglected. Thus one reproduces the non-relativistic Doppler formula where only the radial relative velocity enters.

⁂ **Aberration and beaming** ⁂ Classical aberration was discovered 1725 by James Bradley: He discovered and later explained an apparent motion[1] of celestial objects caused by the finite speed of light and the motion of the Earth in its orbit around the Sun. Assume e.g. to observe a star at the moment when it is right in the zenith. Since the Earth moves, you have to point the telescope a bit "ahead" and thus the position of stars changes in the course of one year.

To obtain the relativistically correct form for the aberration formula, we rewrite first the Lorentz transformations for x and t in differential form,

$$\mathrm{d}x = \gamma(\mathrm{d}x' + \beta c\mathrm{d}t')\,, \tag{4.24}$$

$$\mathrm{d}y = \mathrm{d}y'\,, \qquad \mathrm{d}z = \mathrm{d}z'\,, \tag{4.25}$$

$$c\mathrm{d}t = \gamma(c\mathrm{d}t' + \beta\mathrm{d}x')\,. \tag{4.26}$$

Then we obtain the transformation law for velocities $\mathbf{u}$ as

$$u_\parallel = \frac{\mathrm{d}x_\parallel}{\mathrm{d}t} = \frac{u'_\parallel + v}{1 + vu'_\parallel/c^2} \tag{4.27}$$

$$u_\perp = \frac{\mathrm{d}x_\perp}{\mathrm{d}t} = \frac{u'_\perp}{\gamma(1 + vu'_\perp/c^2)}\,. \tag{4.28}$$

The so-called aberration formula connects the angle between the velocities $\mathbf{u}$ and $\mathbf{u}'$ in the two frames,

$$\tan\vartheta = \frac{u_\perp}{u_\parallel} = \frac{u'\sin\vartheta'}{\gamma(u'\cos\vartheta' + v)}\,. \tag{4.29}$$

The most interesting case is the aberration of light. We set $\vartheta' = \pi/2$, i.e. we ask how light emitted into one half-sphere appears in a different inertial system. With $u = u' = c$ it follows then

$$\tan\vartheta = \frac{c}{\gamma v} \tag{4.30}$$

or $\vartheta \approx 1/\gamma$ for $v \to c$. Thus a fast moving source emits photons mainly in a cone of opening angle $\vartheta \approx 1/\gamma$ around its forward direction.

[1] This motion depends on the velocity of the observer perpendicular to the line-of-sight to the star, but is independent of the distance to the star. The deviation induced by classical aberration is much larger than the parallax we discussed already.

5 Binary stars and stellar parameters

We have discussed up to now the (surface) temperature T, the luminosity L, and the radius R as basic properties of a star. The remaining one, the stellar mass M, can be only determined by its gravitational influence on another body. Fortunately, 50%–80% of all stars are gravitationally bound in multiple star systems, most of them in binary systems.

For instance, 50 binaries, 14 triple and three systems with four stars were known up to a distance of 10 pc from the Sun in the year 1980. For solar-type stars, the ratio of single:double:triple:quadruple stars is 45:46:8:1. On the other hand, most dwarf stars that are more difficult to detect are isolated stars.

Classification of binary stars:

- Apparent or optical binaries are not really binary stars, but random coincidences along the same line-of-sight.
- A visual binary is a bound system that can be resolved into two stars by a telescope. William Herschel discovered 1803 the first example when he noted that the fainter component of Castor (Castor B) had moved relative to the brighter component (Castor A). This was the first indication that gravity acts outside the solar system.
- In an astrometric binary system only the brighter star is visible. Its position "wobbles" back and forth, because the star is moving on a periodic orbit.
- In a spectroscopic binary, the wavelength of spectral lines oscillates periodically due to Doppler shift. The first case was detected 1889 by Pickering who found that the spectral lines of the brighter component of Mizar (Mizar A) were usually double and the spacing varied with a regular 104 day period.
- An eclipsing binary is a system that becomes periodically brighter and fainter. It is dimmed when the faint companion passes in front of the main star. Thus we are aligned to the plane of the orbit. An example visible by eye is Algol: Algol fades at intervals of 2d 20h 49m to a third of its normal brightness, then after a few hours returns to normal brightness.
- In a composite spectrum binary are lines of two stars with different spectral type mixed.

Before we can use binary star systems as tools to determine stellar masses and radii, we have to review the most prominent problem of classical mechanics and astronomy, namely the gravitational two-body or Kepler problem.

5.1 Kepler's laws

Kepler developed his three laws of planetary motions empirically from the observations of Tycho Brahe. The laws are

1. Planets orbit the Sun in ellipses, with the Sun in one of the two focuses.

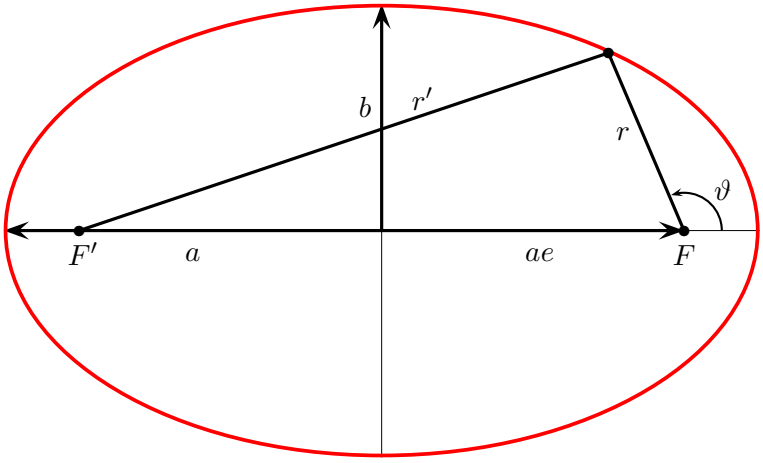
2. The line connecting the Sun and a planet sweeps out equal area in equal time.

3. The "Harmonic Law" states the squared orbital period P of planets measured in years equals to the third power of their major axis measured in astronomical units, $(P/\mathrm{yr})^2 = (a/\mathrm{AU})^3$.

Kepler noted already that his laws describe also the motion of the Saturn moons, if in the Harmonic Law the appropriate units are used. Newton used later the Harmonic Law to derive the $1/r^2$ dependence of the gravitational force. We will follow the opposite way and discuss how Kepler's laws follow from Newton's law for gravitation.

Ellipses:
An ellipse may be defined by the condition $r + r' = 2a$, i.e. as the set of points with a constant sum $2a$ of the distances r and r' to the two focal points F and F'. Additionally to its major axis a, an ellipse is characterized either by its minor axis b or its eccentricity e. The latter two quantities can be connected by considering the two points at the end of the minor axis b: Then $r = r' = a$ and $r^2 = b^2 + (ae)^2$ or

$$b^2 = a^2(1 - e^2)\,. \tag{5.1}$$



Any point of an ellipse can be specified by the distance r to one of its focal points and an angle ϑ that is measured counter-clockwise beginning from the major axis. From the figure, one obtains immediately (with $\cos(180° - \vartheta) = -\cos(\vartheta)$)

$$r'^2 = r^2 + (ae + a)^2 + 2r(2ae + a)\cos\vartheta\,. \tag{5.2}$$

Eliminating r' with the help of $r + r' = 2a$ and solving for r one obtains

$$r = \frac{a(1 - e^2)}{1 + e\cos\vartheta}\,. \tag{5.3}$$

As starting point, we recall how a two-body problem can be reduced to an one-body problem in the case of a central force.

Denoting the positions and the masses of two stars with m_i and r_i, $i = 1, 2$, respectively, the equations of motions are

$$m_1\ddot{\mathbf{r}}_1 = -f(|\mathbf{r}_1 - \mathbf{r}_2|)(\mathbf{r}_1 - \mathbf{r}_2) \tag{5.4}$$
$$m_2\ddot{\mathbf{r}}_2 = +f(|\mathbf{r}_1 - \mathbf{r}_2|)(\mathbf{r}_1 - \mathbf{r}_2)\,. \tag{5.5}$$

Adding Eqs. (5.4) and (5.5),

$$(m_1\ddot{\mathbf{r}}_1 + m_2\ddot{\mathbf{r}}_2) = (m_1 + m_2)\ddot{\mathbf{R}} = 0\,, \tag{5.6}$$

i.e. we see that the center of mass (cm) moves freely. In the second step, we introduced relative coordinates $\mathbf{r}' = \mathbf{r} - \mathbf{R}$ and cm coordinates,

$$\mathbf{R} = \frac{m_1\mathbf{r}_1 + m_2\mathbf{r}_2}{m_1 + m_2}\,. \tag{5.7}$$

Multiplying Eq. (5.4) by m_2 and (5.5) by m_1 and subtracting the two equations, we obtain

$$\mu\ddot{\mathbf{r}} = f(r)\mathbf{r}\,, \tag{5.8}$$

where we introduced the reduced mass

$$\mu = \frac{m_1 m_2}{m_1 + m_2}\,. \tag{5.9}$$

Thus we can solve an one-body problem for the reduced mass μ moving with the distance r in the gravitational field of the mass $M = m_1 + m_2$.

5.1.1 Kepler's second law: The area law

Consider the movement of a body under the influence of a central force,

$$\mu\ddot{\mathbf{r}} = f(r)\frac{\mathbf{r}}{r}\,. \tag{5.10}$$

Since $\mathbf{r} \times \mathbf{r} = 0$, vectorial multiplication with $\mathbf{r}$ should simplify Eq. (5.10). Doing it, we obtain

$$\mu\mathbf{r} \times \ddot{\mathbf{r}} = 0\,, \tag{5.11}$$

that looks already similar to a conservation law. Since

$$\frac{\mathrm{d}}{\mathrm{d}t}\,(\mathbf{r} \times \dot{\mathbf{r}}) = \underbrace{\dot{\mathbf{r}} \times \dot{\mathbf{r}}}_{=0} + \mathbf{r} \times \ddot{\mathbf{r}} \tag{5.12}$$

we obtain indeed the conservation of angular momentum $\mathbf{L} = \mu\mathbf{r} \times \dot{\mathbf{r}}$ for the movement in a central potential,

$$\mu\mathbf{r} \times \ddot{\mathbf{r}} = \frac{\mathrm{d}}{\mathrm{d}t}\,(\mu\mathbf{r} \times \dot{\mathbf{r}}) = \frac{\mathrm{d}}{\mathrm{d}t}\,\mathbf{L} = 0\,. \tag{5.13}$$

There are two immediate consequences: First, the movement is always in the plane perpendicular to $\mathbf{L}$. Second, the area swept out by the vector $\mathbf{r}$ is

$$\mathrm{d}\mathbf{A} = \frac{1}{2}\,\mathbf{r} \times \mathbf{v}\mathrm{d}t = \frac{1}{2\mu}\,\mathrm{d}\mathbf{L} \tag{5.14}$$

and thus also constant.

5.1.2 Kepler's first law

We introduce the unit vector $\hat{\mathbf{r}} = \mathbf{r}/r$ and rewrite the definition of the angular momentum L as

$$\mathbf{L} = \mu \mathbf{r} \times \dot{\mathbf{r}} = \mu r \hat{\mathbf{r}} \times \frac{\mathrm{d}}{\mathrm{d}t}(r\hat{\mathbf{r}}) = \mu r \hat{\mathbf{r}} \times \left(\dot{r}\hat{\mathbf{r}} + r\frac{\mathrm{d}\hat{\mathbf{r}}}{\mathrm{d}t} \right) = \mu r^2 \hat{\mathbf{r}} \times \frac{\mathrm{d}\hat{\mathbf{r}}}{\mathrm{d}t} \,. \tag{5.15}$$

The first term in the parenthesis vanishes, because of $\hat{\mathbf{r}} \times \hat{\mathbf{r}} = 0$. Next we take the cross product of the gravitational acceleration,

$$\mathbf{a} = -\frac{GM}{r^2}\,\hat{\mathbf{r}} \tag{5.16}$$

with the angular momentum,

$$\mathbf{a} \times \mathbf{L} = -\frac{GM}{r^2}\,\hat{\mathbf{r}} \times \left(\mu r^2 \hat{\mathbf{r}} \times \frac{\mathrm{d}\hat{\mathbf{r}}}{\mathrm{d}t} \right) = -GM\mu\,\hat{\mathbf{r}} \times \left(\hat{\mathbf{r}} \times \frac{\mathrm{d}\hat{\mathbf{r}}}{\mathrm{d}t} \right) \,. \tag{5.17}$$

With the identity $\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = (\mathbf{A} \cdot \mathbf{C})\mathbf{B} - (\mathbf{A} \cdot \mathbf{B})\mathbf{C}$, it follows

$$\mathbf{a} \times \mathbf{L} = -GM\mu \left[\left(\hat{\mathbf{r}} \cdot \frac{\mathrm{d}\hat{\mathbf{r}}}{\mathrm{d}t} \right) \hat{\mathbf{r}} - (\hat{\mathbf{r}} \cdot \hat{\mathbf{r}}) \frac{\mathrm{d}\hat{\mathbf{r}}}{\mathrm{d}t} \right] \,. \tag{5.18}$$

Because $\hat{\mathbf{r}}$ is a unit vector, $\hat{\mathbf{r}} \cdot \hat{\mathbf{r}} = 1$ and $\mathrm{d}(\hat{\mathbf{r}} \cdot \hat{\mathbf{r}})/\mathrm{d}t = 0$, and thus

$$\mathbf{a} \times \mathbf{L} = GM\mu \frac{\mathrm{d}\hat{\mathbf{r}}}{\mathrm{d}t} \,. \tag{5.19}$$

Since $\mathbf{L}$ and $GM\mu$ are constant, we can rewrite this as

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{v} \times \mathbf{L}) = \frac{\mathrm{d}}{\mathrm{d}t}(GM\mu\hat{\mathbf{r}}) \,. \tag{5.20}$$

Integrating gives

$$\mathbf{v} \times \mathbf{L} = GM\mu\hat{\mathbf{r}} + \mathbf{C} \,, \tag{5.21}$$

where the integration constant $\mathbf{C}$ is a constant vector.

Taking now the dot product with $\mathbf{r}$, we obtain

$$\mathbf{r} \cdot (\mathbf{v} \times \mathbf{L}) = GM\mu r \hat{\mathbf{r}} \cdot \hat{\mathbf{r}} + \mathbf{r} \cdot \mathbf{C} \,. \tag{5.22}$$

Applying next the identity $\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = (\mathbf{A} \times \mathbf{B}) \cdot \mathbf{C}$, it follows

$$(\mathbf{r} \times \mathbf{v}) \cdot \mathbf{L} = GM\mu r + rC\cos\vartheta = GM\mu r \left(1 + \frac{C\cos\vartheta}{GM\mu} \right) \,. \tag{5.23}$$

Expressing $\mathbf{r} \times \mathbf{v}$ as $\mathbf{L}/\mu$, defining $e = C/(GM\mu)$ and solving for r, we obtain finally the equation for a conic section, i.e. Kepler's first law,

$$\boxed{r = \frac{L^2/\mu^2}{GM(1 + e\cos\vartheta)} \,.} \tag{5.24}$$

In the case of an ellipse,

$$r = \frac{a(1 - e^2)}{1 + e\cos\vartheta} \tag{5.25}$$

and thus we obtain as angular momentum

$$L = \mu\sqrt{GMa(1 - e^2)} \,. \tag{5.26}$$

5.1.3 Kepler's third law

We integrate Kepler's second law in the form of Eq. (5.14) over one orbital period P,

$$A = \pi ab = \frac{L}{2\mu} P \,. \tag{5.27}$$

Squaring and solving for P, it follows

$$P^2 = 4\pi^2 \frac{(ab\mu)^2}{L^2} \,. \tag{5.28}$$

With $b^2 = a^2(1-e^2)$ and Eq. (5.26) for the angular momentum L, we obtain Kepler's third or harmonic law,

$$P^2 = \frac{4\pi^2}{G(m_1+m_2)} a^3 \,. \tag{5.29}$$

5.2 Determining stellar masses

Visual binaries The period of a visual binary system can be determined directly, if it is observed long enough. Note that periods are typically long, from years up to 100–1000 yr, because the angular separation of the two stars should be large. The plane containing the orbit may be inclined relative to the plane orthogonal to the line of sight. This angle is called inclination i and is zero when the two planes coincide and $i = 90°$ when the orbit is seen edge-on.

The inclination and the true orbit can be reconstructed from the apparent orbit, if good enough observations exist: A tilted ellipse projected on the celestial sphere remains an ellipse, but with a changed eccentricity. Therefore the stars seem to move not around the foci, and this discrepancy can be used to find the "true" ellipse. In order to convert the observed angular distances into true lengths, the distance to the binary system has to be known. The total mass follows then from Kepler's 3. law. Moreover, the ratio of masses can be determined from the ratio r_1/r_2 (or via Doppler shift from v_1/v_2, if the system is also a spectroscopic binary).

Spectroscopic binaries and Doppler shifts If two stars orbit their cms closely (AU) and fast ($P \approx$ hours-months) and moreover the inclination i is not close to zero, the Doppler shift of their spectral lines can be detected. The latter condition arises, because in the non-relativistic limit $v \ll c$ or $\gamma \approx 1$

$$\frac{\omega_0}{\omega} = \frac{\left[1-(v/c)^2\right]^{1/2}}{1-(v/c)\cos\alpha} \approx \frac{1}{1-(v/c)\cos\alpha} = \frac{1}{1-v_r/c} \,, \tag{5.30}$$

where v_r is the radial velocity and α was the angle between the wave vector/propagation direction and the direction of movement. Rearranging this equation and introducing $\Delta\omega = \omega - \omega_0$ it follows

$$\frac{\Delta\omega}{\omega} = -\frac{v_r}{c} = -\frac{\Delta\lambda}{\lambda} \,. \tag{5.31}$$

If the orbit is circular, then the curve will be a sine curve. For an elliptical orbit, the shape of the radial velocity curve depends on the eccentricity and the inclination of the ellipse.

For simplicity, we assume orbits with small eccentricity. Then from $r_i = P/(2\pi)v_i$, and $a = r_1 + r_2 = P/(2\pi)(v_1 + v_2)$ and thus Kepler's 3. law becomes

$$2\pi G(m_1 + m_2) = P(v_1 + v_2)^3 = \frac{P(v_{1r} + v_{2r})^3}{\sin^3 i} \,. \tag{5.32}$$

with $v_r = v \sin i$.

If the inclination i is not known, one can derive at least a lower limit on $m_1 + m_2$ using $i = 90°$. Larger samples of spectroscopic binaries allow one to improve these limits by performing a statistical analysis. One assume an isotropic distribution of the inclination angle and calculates then the expectation value of $\sin^3 i$. Since there exists an observational bias to discover values of i close to 90°, the estimation has to be corrected for such a selection effect.

Often only the Doppler shift of one of the two components can be measured. Eliminating e.g. v_2 using the cm condition (assuming again for simplicity circular orbits) gives

$$v_2 = \frac{m_1}{m_2} v_1 \,. \tag{5.33}$$

Then

$$v_1 + v_2 = v_1 \left(1 + \frac{m_1}{m_2}\right) = \frac{v_1}{m_2} \left(m_1 + m_2\right) \,. \tag{5.34}$$

Now Kepler's law can be rewritten as

$$\frac{P}{2\pi G} \, v_{1r}^3 = \frac{m_2^3 \sin^3 i}{(m_1 + m_2)^2} \,. \tag{5.35}$$

The RHS is called mass function and is again studied with statistical methods.

Eclipsing binaries An eclipsing binary star is a binary system in which the orbit plane of the two stars lies nearly in the line-of-sight of the observer, $\sin(i) \approx 1$, so that the components undergo mutual eclipses. Thus they appear as variable stars, with periodic drops in intensity when the less luminous star obscures the other.

5.2.1 Mass-luminosity relation

For any binary system with derived mass also the distance is known. Thus, the luminosity can be derived from the apparent magnitude. When the observed luminosity for stars in binary systems is plotted as function of their mass, the so-called mass-luminosity relationship is obtained. Eddington derived 1924 for main-sequence stars theoretically using a rather crude model

$$\frac{L}{L_\odot} = \left(\frac{M}{M_\odot}\right)^\alpha \tag{5.36}$$

with $\alpha \sim 3$. This means that the main-sequence in the Hertzsprung-Russel diagram can be indeed characterized by essentially a single parameter, e.g. their mass. From observations, one finds a broken power law with

$$\alpha \approx 1.8 \quad \text{for} \quad M < 0.3 M_\odot \,, \tag{5.37}$$

$$\alpha \approx 4.0 \quad \text{for} \quad 0.3 M_\odot < M < 3 M_\odot \,, \tag{5.38}$$

$$\alpha \approx 2.8 \quad \text{for} \quad M > 3 M_\odot \,. \tag{5.39}$$

From observations one tries to measure the number of stars as function of their mass, correcting for the bias to see more easily massive and thus brighter stars. From this, taking into account that massive stars live shorter, one can construct the initial mass function that is an important theoretical ingredient in tests of stellar evolution models.

5.3 Stellar radii

Only for the Sun, the radius can be directly determined from its angular radius and the distance. Other stars cannot be resolved as extended objects from the ground because of atmospheric seeing. If the temperature and the luminosity of a star are known, the Stefan-Boltzmann law can be used to define the "luminosity radius". Eclipsing binaries offer another possibility to measure stellar radii, since their light-curve depends directly on the radii of the two stars. As result, one finds that the MS stars become larger with increasing temperature, cf. Fig. 2.2.

6 Stellar atmospheres and radiation transport

6.1 The Sun as typical star

The Sun is a rather typical main-sequence star, distinguished only by its proximity to the observing astronomers on Earth.
Solar radius $R_\odot$: From the distance D and the apparent radius of the solar disk, $\delta_\odot = 15'59.63''$, one obtains $R_\odot = D\delta_\odot = 696000\,\mathrm{km}$.
Solar mass and density: Kepler's Third Law relates the semi-major axis a of their orbits, the period P and the total mass of two gravitating bodies,

$$\frac{a^3}{P^2} = \frac{G}{4\pi^2}(m_1 + m_2)\,. \tag{6.1}$$

This gives with $M_\oplus \ll M_\odot$, $M_\odot = 1.989 \times 10^{33}$ g. The average solar density follows as $\rho = \frac{M_\odot}{4\pi/3 R_\odot^3} = 1.409\,\mathrm{g/cm^3}$.
Solar temperature: Measuring the Solar Constant, i.e. the energy flux of the Sun at the Earth's position outside the atmosphere, one can use the Stefan-Boltzmann law to define the effective solar temperature, $T = 5762\,\mathrm{K}$.
Magnetic field, sunspots: Magnetic field strengths can be determined spectroscopically via the Zemann effect: The energy levels of atoms split with $\Delta E = g m_J \hbar \omega_L$, where $\omega_L = eB/(2m_e)$ is the larmor frequency, g the Lande factor and $|m_J| \leq J$ is the magnetic quantum number. The Sun seems to have no smooth large-scale magnetic field; magnetic fields are concentrated in thin flux tubes. Sunspots are the ends of such flux tubes at the solar atmosphere and have field strength up-to one $\mathrm{T} = 10^4$ G. The abundance of sunspots varies with an average period of 11 years. With the same period, the magnetic field of the Sun reverses sign.
Rotation period: Observing sunspots, the rotation period P of the Sun can be determined. It is found that the Sun rotates differentially, at the equator with $P = 24.8$ d and e.g. at latitude 70° with $P \approx 31$ d.
Structure: Those layers that give rise to the continuous radiation and the major part of the Fraunhofer lines are called photosphere. During total solar eclipses, one can observe the higher layer of the atmosphere. This layer is called chromosphere and is characterized by the absence of continuum and absorption lines, showing only emission lines. Afterwards the corona follows stretching outwards into the interplanetary medium.

6.2 Radiation transport

Let us idealize atoms as black discs of radius R. Then any photon hitting the disc with area πR^2 is absorbed. In other words, the cross-section σ for the absorption of a photon by a single atom is $\sigma = \pi R^2$.

Consider now a cylinder of length l and area A, filled with N atoms. Their number density n is $n = N/(Al)$. Let us assume first for simplicity that $N\sigma \ll A$, i.e. the atoms do not overlap much and are uniformly distributed. Then the fraction τ of incoming radiation absorbed in the cylinder is simply

$$\frac{N\sigma}{A} = nl\sigma \equiv \tau \tag{6.2}$$

and defines the optical depth $\tau \equiv nl\sigma$. Our assumption $N\sigma \ll A$ corresponds to $\tau \ll 1$, an "optical thin" or transparent source in contrast to an "optical thick" source with $\tau \gg 1$. For arbitrary τ, we have to calculate how much radiation is absorbed passing the infinitesimal distance $\mathrm{d}l$,

$$\mathrm{d}I = -I\mathrm{d}\tau = -In\sigma\mathrm{d}l \tag{6.3}$$

or

$$\frac{\mathrm{d}I}{I} = -n\sigma\mathrm{d}l\,. \tag{6.4}$$

Integrating gives

$$\int_{I_0}^{I} \frac{\mathrm{d}I'}{I} = -\int_0^l \mathrm{d}l' n\sigma \equiv -\tau \quad \Rightarrow \quad \ln\frac{I}{I_0} = -\tau \tag{6.5}$$

and finally

$$I = I_0\exp(-\tau) \qquad \text{or} \qquad I(l) = I_0\exp\left(-\int \mathrm{d}l n\sigma\right)\,. \tag{6.6}$$

For small τ, expansion of the exponential reproduces our old result: From $I_0\exp(-\tau) \approx I_0(1-\tau+\ldots)$, one sees that the intensity is reduced by $I_0\tau$, i.e. the fraction of absorbed photons is given by τ. If $n\sigma$ does not depend on the distance l, the optical depth τ is simply obtained by multiplying $n\sigma$ with the path length, as in our original definition (6.2). In general, $n\sigma$ has to be integrated along the path length l. This equation can be seen as definition and measurement description for the cross section σ.

What is the typical size of the cross section σ for the absorption of photons? A reasonable estimate for the excitation of an atom is $\sigma \sim \pi r_1^2$ choosing $r_1 = \hbar^2/(m_e\alpha)$ as the first Bohr radius, while the cross section for scattering pf photons on free electrons can be estimated as $\sigma \sim \pi r_0^2$, where $r_0 = \alpha\hbar/(m_e c)$ is the classical electron radius. A theory of the radiation transport in stars needs as input scattering cross-section of photons with atoms and ions.

One introduces often the opacity $\kappa = \tau/l = n\sigma$. Then, Eq. (6.3) becomes

$$\mathrm{d}I = -I\kappa\mathrm{d}l\,. \tag{6.7}$$

Photons can not only be absorbed, but have to be also re-emitted by atoms. Hence, we should add a source term S to the previous equation,

$$\frac{1}{\kappa}\frac{\mathrm{d}I}{\mathrm{d}l} = -I + S\,, \tag{6.8}$$

where S accounts e.g. for the de-excitation of atoms. For instance, a photon can be absorbed exciting an atom to the level n, de-exciting then in several steps emitting several photons of lower energy.

Ex.: A beam of light with intensity I_0 enters a gas layer with constant opacity κ and source function S. Derive $I(l)$ and sketch $I(l)$ as function of $\tau = \kappa l$ for $S = 2I_0$ and $S = I_0/2$.
We have solved already Eq. (6.8) for $S = 0$, where we obtained $I(l) = I_0\exp(-\kappa l)$. The general solution

for $S \neq 0$ is obtained by "variation of constants", i.e. if we replace I_0 by the unknown function $f(l)$. Inserting

$$\frac{\mathrm{d}I}{\mathrm{d}l} = \frac{\mathrm{d}f}{\mathrm{d}l}\exp(-\kappa l) - f\kappa\exp(-\kappa l)$$

into Eq. (6.8) gives

$$\frac{\mathrm{d}f}{\mathrm{d}l} = S\exp(\kappa l)$$

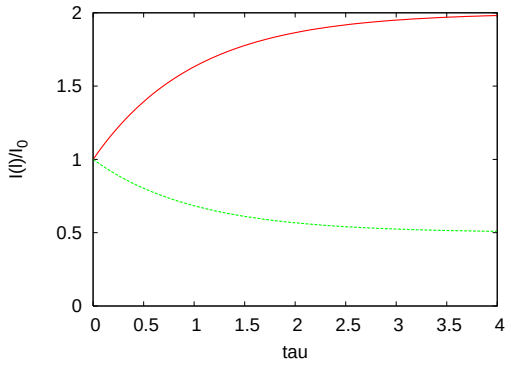
or

$$f(l) = S\int \mathrm{d}l\exp(\kappa l) = S\kappa\exp(\kappa l) + c\,.$$

The integration constant c has to be chosen such that $I(0) = I_0$, $c = I_0 - S$. Hence the solution is

$$I(l) = I_0\exp(-\kappa l) + S[1 - \exp(\kappa l)]\,.$$

The figure shows how the intensity $I(l)$ of the photon beam reaches asymptotically the value S of the source function. If $I_0 < S$, more photons are replaced by the gas than absorbed and $I(l)$ approaches S from below for $\tau \to \infty$. The typical distance for a change of I is τ.



6.3 Diffusion and random walks

What is the fate of a photon created in the center of the Sun? If we estimate its mean free-path $\ell = 1/(n\sigma)$ by assuming that most of hydrogen in the Sun is ionized, and thus the mean density of free electron is $n_e = \rho_\odot/m_H \approx 8.4 \times 10^{23}/\mathrm{cm}^3$, and use as cross section $\sigma \sim 10^{-25}\mathrm{cm}^2$, then we obtain as estimate $\ell \sim 10\,\mathrm{cm}$. In each scattering process, the photon is approximately isotropically emitted: Thus the path of a photon from the center of the Sun to the surface is not a straight-forward travel, but resembles a random-walk. The time a photon needs to travel the distance $R_\odot$ can be estimated in two ways:

After n steps $\mathbf{l_i}$ of the same size $|\mathbf{l_i}| = l$ a photon that started at zero is at the position $\mathbf{d} = \sum_{i=1}^{n}\mathbf{l}_i$. The scalar product of $\mathbf{d}$ with itself is

$$\mathbf{d}\cdot\mathbf{d} = \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{l}_i\cdot\mathbf{l}_j \tag{6.9}$$

and splitting the sum into the diagonal and the off-diagonal terms, we obtain

$$d^2 = nl^2 + 2l^2\sum_{i=1}^{n}\sum_{j<i}^{n}\cos\vartheta_{ij} \approx nl^2\,. \tag{6.10}$$

By assumption, the angles ϑ_{ij} between $\mathbf{l}_i$ and $\mathbf{l}_j$ are chosen randomly and thus the off-diagonal terms cancel against each other.

A more formal argument goes as follows: The probability $p(m,n)$ to end after n steps of unit size at the distance $d = m$ is given in the Table as

m	-3	-2	-1	0	1	2	3
$n=0$				1			
$n=1$			1/2		1/2		
$n=2$		1/4		1/2		1/4	
$n=3$	1/8		3/8		3/8		1/8

The denominator of $p(m,n)$ is the sum of all possible paths and thus equal to 2^n. The numerator is the sum of all paths ending at distance m after n steps. Each entry in the table is given by the sum of the entries to the left and right in the line above and thus equal to the binomial coefficient (remember Pascal's triangle). Thus

$$p(m,n) = 2^{-n}\binom{n}{m}. \tag{6.11}$$

Since $p(m,n) = p(-m,n)$, the average distance is zero. A more useful quantity is the root mean square deviation $d = \langle k^2\rangle^{1/2}$,

$$d = \left(\frac{\sum_k k^2 p(k,n)}{2^n}\right)^{1/2} = n^{1/2} \tag{6.12}$$

If the step size is l, then we obtain again $d = n^{1/2}l$.

6.4 Photo-, chromosphere and corona

Photosphere: The photosphere gives rise to the continuous radiation and the major part of the Fraunhofer lines. Main process for continuum absorption is the bound-free transition $\mathrm{H}^- + \gamma \rightarrow \mathrm{H} + e^-$, for emission the reverse process. Since the electron is free, the energy of the photon is not fixed.

The surface of the Sun is not smooth but shows granules of $\sim$ 100km size: Lighter areas are surrounded by darker (i.e. cooler) ones, providing evidence for convection.

Center limb darkening: An area element of a surface has the same brightness for any viewing angle ϑ ("Lambert's cosine law"): Although the emitted power from the area element into the direction ϑ is reduced by $\cos\vartheta$, the observed size of the area element is also reduced by that same amount. Consequently, the brightness of the solar disk should be constant, if the Sun is an ideal black-body.

The observed darkening of the Sun towards the limb (cf. right panel of Fig. 6.1) can be explained by the finite extension of the photosphere and a decrease of the solar temperature towards its edge. We see by definition to $\tau = 1$. At the center, a certain value $r_* = r(\tau = 1)$ is defined by this condition, while this r_* corresponds to $\tau \sim 1/\cos\vartheta$ at the limb. Thus $\tau = 1$ corresponds to layers lying more outside towards the limb that are in turn cooler. Since I depends exponentially on τ, this explains also the sharpness of the solar atmosphere.

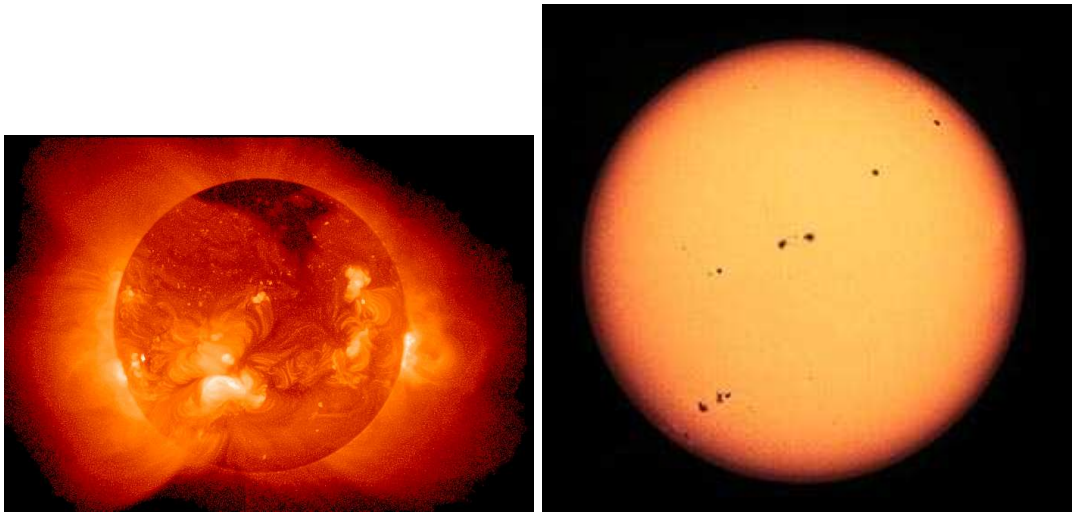


Figure 6.1: Left: Solar corona in X-rays. Right: Photosphere with sunspots and limb-darkening clearly visible.

Chromosphere: At most wavelengths the chromosphere is optically thin, and we see through it to the photosphere. Only using special filters or during a solar eclipse, the chromosphere becomes visible.
Corona: A very hot, low-density plasma that is far from thermal equilibrium. Its energy density is dominated by magnetic fields. The heating mechanism of the corona is not yet understood.

7 Main sequence stars and their structure

7.1 Equations of stellar structure

We look for spherical symmetric, static solutions of the equations of stellar structure. This requires that rotation, convection, magnetic fields B, and other effects that break rotational symmetry have only a minor influence on the star. This assumption is in most cases very well justified.

7.1.1 Mass continuity and hydrostatic equilibrium

We denote by $M(r)$ the mass enclosed inside a sphere with radius r and density $\rho(r)$,

$$M(r) = 4\pi \int_0^r \mathrm{d}r' r'^2 \rho(r') \tag{7.1}$$

or

$$\boxed{\frac{\mathrm{d}M(r)}{\mathrm{d}r} = 4\pi r^2 \rho(r)\,.} \tag{7.2}$$

Although trivial, this equation constitutes the first (the "continuity equation") of the five equations needed to describe the structure and evolution of stars. An important application of the continuity equation is to express physical quantities not as function of the radius r but of the enclosed mass $M(r)$. This facilitates the computation of the stellar properties as function of time, because the mass of a star remains nearly constant during its evolution, while the stellar radius can change considerably.

A radial-symmetric mass distribution $M(r)$ produces according Gauß' law the same gravitational acceleration, as if it would be concentrated at the center $r = 0$. Therefore the gravitational acceleration produced by $M(r)$ is

$$g(r) = -\frac{GM(r)}{r^2}\,. \tag{7.3}$$

If the star is in equilibrium, this acceleration is balanced by a pressure gradient from the center of the star to its surface. Since pressure is defined as force per area, $P = F/A$, a pressure change along the distance $\mathrm{d}r$ corresponds to an increment

$$\mathrm{d}F = \mathrm{d}A\,P - (P + \mathrm{d}P)\mathrm{d}A = -\underbrace{\mathrm{d}A\,\mathrm{d}P}_{\text{force}} = -\underbrace{\rho(r)\mathrm{d}A\mathrm{d}r}_{\text{mass}}\ \underbrace{a(r)}_{\text{acceleration}} \tag{7.4}$$

of the force F produced by the pressure gradient $\mathrm{d}P$. For increasing r, the gradient $\mathrm{d}P < 0$ and the resulting force $\mathrm{d}F$ is positive and therefore directed outward. Hydrostatic equilibrium, $g(r) = -a(r)$, requires then

$$\boxed{\frac{\mathrm{d}P}{\mathrm{d}r} = \rho(r)g(r) = -\frac{GM(r)\rho(r)}{r^2}\,.} \tag{7.5}$$

If the pressure gradient and gravity do not balance each other, the layer at position r is accelerated,

$$a(r) = \frac{GM(r)}{r^2} + \frac{1}{\rho(r)} \frac{\mathrm{d}P}{\mathrm{d}r} \,. \tag{7.6}$$

In general, we need an equation of state, $P = P(\rho, T, Y_i)$, that connects the pressure P with the density ρ, the (not yet) known temperature T and the chemical composition Y_i of the star. For an estimate of the central pressure $P_c = P(0)$ of a star in hydrostatic equilibrium, we integrate (7.5) and obtain with $P(R) \approx 0$,

$$P_c = \int_0^R \frac{\mathrm{d}P}{\mathrm{d}r} \,\mathrm{d}r = G \int_0^M \mathrm{d}M \, \frac{M}{4\pi r^4} \,, \tag{7.7}$$

where we used the continuity equation to substitute $\mathrm{d}r = \mathrm{d}M/(4\pi r^2 \rho)$ by $\mathrm{d}M$. If we replace furthermore r by the stellar radius $R \geq r$, we obtain a lower limit for the central pressure,

$$P_c = G \int_0^M \mathrm{d}M \, \frac{M}{4\pi r^4} > G \int_0^M \mathrm{d}M \, \frac{M}{4\pi R^4} = \frac{M^2}{8\pi R^4} \,. \tag{7.8}$$

Inserting values for the Sun, it follows

$$P_c > \frac{M^2}{8\pi R^4} = 4 \times 10^8 \mathrm{bar} \left(\frac{M}{M_\odot} \right)^2 \left(\frac{R_\odot}{R} \right)^4 \,. \tag{7.9}$$

The value obtained integrating the hydrostatic equation using the "Solar Standard Model" is $P_c = 2.48 \times 10^{11}$ bar, i.e. a factor 500 larger.

7.1.2 Gas and radiation pressure

A (relativistic or non-relativistic) particle in a box of volume L^3 collides per time interval $\Delta t = 2L/v_x$ once with the yz-side of the box, if the x component of its velocity is v_x. Thereby it exerts the force $F_x = \Delta p_x / \Delta t = p_x v_x / L$. The pressure produced by N particles is then $P = F/A = N p_x v_x/(LA) = n p_x v_x$ or for an isotropic distribution with $\langle v^2 \rangle = \langle v_x^2 \rangle + \langle v_y^2 \rangle + \langle v_z^2 \rangle = 3 \langle v_x^2 \rangle$

$$P = \frac{1}{3} n v p \,. \tag{7.10}$$

If the particles have a distribution n_p of momenta with

$$N = V \int_0^\infty dp \, n_p = V \int_0^\infty \mathrm{d}v \, n_v \,, \tag{7.11}$$

then we obtain the so-called pressure integral

$$P = \frac{1}{3} \int_0^\infty \mathrm{d}p \, n_p v p \,. \tag{7.12}$$

Although the derivation assumed classical trajectories of the particles, the result holds for any kind of non-interacting particles, in particular also if quantum effects are important (cf. Exercise 7.1).

The two most important cases in astrophysics are a classical, non-relativistic gas of atoms and a gas of photons. In the first case, we can derive the momentum distribution n_p noting

that the states describing a free particle are labelled by the continuous three-momentum $\mathbf{p}$. Thus the sum over discrete quantum numbers in the Boltzmann factor is replaced by an integration over the momenta $\mathrm{d}^3p/(2\pi)^3$ and the volume d^3x occupied by the system,

$$\sum_i \exp(-E/kT) \to \int \frac{\mathrm{d}^3p}{(2\pi)^3} \exp\left(-\frac{mv^2}{2kT}\right) = \frac{1}{(2\pi)^3} \int_0^\infty \exp\left(-\frac{mv^2}{2kT}\right) 4\pi m^3 v^2 \mathrm{d}v\,. \quad (7.13)$$

If we compare the RHS with Eq. (7.11) we see that we need only to normalize correctly n_v. The integral can be evaluated by substituting $\alpha = m/(2kT)$ and noting that

$$-\frac{\partial}{\partial\alpha}\left\{\int_{-\infty}^{\infty} \mathrm{d}x \exp(-\alpha x^2)\right\} = \int_{-\infty}^{\infty} \mathrm{d}x x^2 \exp(-\alpha x^2) = -\frac{\partial}{\partial\alpha}\sqrt{\frac{\pi}{\alpha}} = \frac{1}{2\alpha}\sqrt{\frac{\pi}{\alpha}}\,. \quad (7.14)$$

Multiplying the integrand with $4(\alpha/\pi)^{3/2}$, we obtain the Maxwell-Boltzmann distribution of velocities for a classical gas,

$$n_v \mathrm{d}v = n\left(\frac{m}{2\pi kT}\right)^{2/3} \exp\left(-\frac{mv^2}{2kT}\right) 4\pi v^2 \mathrm{d}v\,. \quad (7.15)$$

Because of $n_p \mathrm{d}p = n_v \mathrm{d}v$, we can insert now n_v into the pressure integral (7.12),

$$P = \frac{1}{3}\int_0^\infty \mathrm{d}v\, n_v v p = n\left(\frac{\alpha}{\pi}\right)^{2/3} \int_0^\infty \mathrm{d}x\, x^4 \exp\left(-\alpha x^2\right) = nkT\,. \quad (7.16)$$

The integral $\int \mathrm{d}x\, x^4 \exp(-\alpha x^2)$ has been calculated with the same trick, but now differentiating twice the Gaussian integral with respect to α. Since we use generally the mass density ρ instead of the particle number density n, it is more convenient to introduce the gas constant $R = k/m_H$ and the mean atomic weight μ defined by $n = \rho/(\mu m_H)$. Then the ideal gas law becomes

$$\boxed{P = nkT = R\rho T/\mu\,.} \quad (7.17)$$

A fully ionized plasma consisting mainly of hydrogen has $\mu \approx 1/2$.

The second important example is the pressure P_{rad} of radiation, i.e. the pressure of a photon gas. With $p = h\nu/c$ and $n_\nu \mathrm{d}\nu = n_p \mathrm{d}p$ it follows

$$P_{\mathrm{rad}} = \frac{1}{3}\int_0^\infty \mathrm{d}\nu\, n_\nu h\nu\,. \quad (7.18)$$

Remembering that $u_\nu \mathrm{d}\nu = 4\pi/cB_\nu \mathrm{d}\nu$, it follows

$$\boxed{P_{\mathrm{rad}} = aT^4/3\,,} \quad (7.19)$$

where we introduced the radiation constant $a = 4\sigma/c$.

7.1.3 Virial theorem

The virial theorem is an important link between the (gravitational) potential energy and the internal (kinetic) energy of any system in equilibrium. In order to derive it for the special case of a star, we multiply both sides of the hydrostatic equilibrium Eq. (7.5) with $4\pi r^3$ and integrate over r,

$$\int_0^R \mathrm{d}r\, 4\pi r^3 P' = -\int_0^R \mathrm{d}r\, 4\pi r^3 \frac{GM(r)\rho}{r^2}\,. \quad (7.20)$$

Next we insert $\mathrm{d}M(r) = 4\pi r^2 \rho \mathrm{d}r$ on the RHS and integrate partially the LHS with $V(0) = 0$ and $P(R) = 0$,

$$-3\int_0^R \mathrm{d}r\, 4\pi r^2 P = -\int_0^M \mathrm{d}M\, \frac{GM(r)}{r}\,. \tag{7.21}$$

The RHS is the gravitational potential energy $U_{\rm pot}$ of the star. We can rewrite the LHS as

$$-3\int_0^M \mathrm{d}M\, \frac{P}{\rho} = U_{\rm pot}\,. \tag{7.22}$$

For the special case of an ideal gas, $P = nkT = \frac{2}{3}U_{\rm kin}/V$, and we obtain the virial theorem,

$$\boxed{-2U_{\rm kin} = U_{\rm pot}\,.} \tag{7.23}$$

Hence the average energy $E = U/N$ of a single atom or molecule of the gas is $\langle E_{\rm kin}\rangle = -\frac{1}{2}\langle E_{\rm pot}\rangle$. This is the same result as for a free hydrogen atom, indicating that only the shape not the strength of the potential $V(r) \propto r^{-\alpha}$ determines the ratio of kinetic and potential energy.

Ex.: Estimate the central temperature of the Sun with the virial theorem.
We approximate the Sun as a homogeneous sphere. Then the gravitational potential energy of a proton at the center of the Sun is

$$\langle E_{\rm pot}\rangle = -\frac{GM_\odot m_p}{R_\odot} \approx -3.2\mathrm{keV}/c^2\,.$$

For a thermal velocity distribution of a Maxwell-Boltzmann gas we obtain

$$\langle E_{\rm kin}\rangle = \frac{3}{2}kT = -\frac{1}{2}\langle E_{\rm pot}\rangle \approx 1.6\,\mathrm{keV}/c^2\,.$$

Hence our estimate for the central temperature of the Sun is $T_{\rm c} \approx 1.1\,\mathrm{keV}/c^2 \approx 1.2\times10^7\,\mathrm{K}$ – compared to $T_{\rm c} = 1.57\times10^7\,\mathrm{K}$ in the Solar Standard Model.

7.1.4 *** Stability of stars ***

We want to generalize the virial theorem to a gas with an equation of state such that the pressure P can be expressed as function only of the density ρ. Such an equation of state is called polytropic, and normally written as $P = K\rho^\gamma$ with polytropic index γ. To do so, we have to express the energy density as function of the pressure and the index γ. Combining $\mathrm{d}P/P = -\gamma\mathrm{d}\rho/\rho$ and $\mathrm{d}\rho/\rho = -\mathrm{d}V/V$, we obtain

$$V\mathrm{d}P = -\gamma P\mathrm{d}V\,. \tag{7.24}$$

Next we add $p\mathrm{d}V$ to both sides,

$$\mathrm{d}(VP) = V\mathrm{d}P + P\mathrm{d}V = -(\gamma-1)P\mathrm{d}V\,. \tag{7.25}$$

or

$$\mathrm{d}\left(\frac{VP}{\gamma-1}\right) = -P\mathrm{d}V = \mathrm{d}U\,. \tag{7.26}$$

Hence the pressure and the (kinetic) energy density of a polytrope are connected by

$$P = (\gamma - 1)\frac{U}{V}\,. \tag{7.27}$$

For an ideal gas, $U/V = 3/2kT$ and $P = nkT$, and the adiabatic index follows as $\gamma = 5/3$, while for radiation, $U/V = aT^4$ and $P = aT^4/3$, and thus $\gamma = 4/3$.

The relation $P = (\gamma - 1)U/V$ allows us to re-express the LHS of Eq. (7.21) as

$$-3(\gamma - 1)U_{\rm kin} = U_{\rm pot}\,. \tag{7.28}$$

A star can be only stable, if its total energy $U_{\rm tot} = U_{\rm kin} + U_{\rm pot}$ is smaller than zero,

$$U_{\rm tot} = (4 - 3\gamma)U_{\rm kin} = \frac{3\gamma - 4}{3\gamma - 3}\,U_{\rm pot} < 0\,. \tag{7.29}$$

For $\gamma = 5/3$, we obtain back our old result for an ideal gas. A star with $\gamma = 4/3$ has zero energy and marks the border of matter that can become gravitationally bound. Adding an arbitrary small amount of energy would disrupt such a star, while subtraction would lead to its collapse. Important examples for matter with $\gamma = 4/3$ are apart from radiation relativistic fermions, e.g. relativistic electrons, protons and neutrons.

Stars or more generally all gravitationally bound systems, have surprising thermodynamical properties: Consider e.g. the heat capacity, $C_V = \partial U_{\rm kin}/\partial V \propto \partial U_{\rm kin}/\partial R$. Since the temperature T increases for decreasing radius R, the heat capacity of a star is negative.

7.1.5 Energy transport

There exist three different mechanism for the transport of energy: i) radiative energy transfer, i.e. energy transport by photons, ii) conduction, e.g. the transport of energy by valence electrons or phonons in solids, and iii) macroscopic matter flows. Conduction plays a prominent role as energy transport only for dense systems, and is therefore only relevant in the dense, final stages of stellar evolution.

Radiative energy transport

For the energy flux $\mathcal{F}$ emitted per area and time by an layer of the star at radius r and temperature T, a transport equation similar to Eq. (6.8) for the intensity I holds,

$$\mathrm{d}\mathcal{F} = -\sigma n \mathcal{F}\mathrm{d}r = -\kappa\rho\mathcal{F}\mathrm{d}r\,. \tag{7.30}$$

Here we introduced also the opacity κ being the cross section per mass of a certain material. Absorption of radiation implies also a transfer of momentum to the medium. Since the momentum of photons is $p = E/c$ and $\mathcal{F} = E/(At)$, a slab of matter absorbs the momentum $\mathcal{F}/c$ per area and time. According to Newton's second law, $F = \dot{p}$, the absorbed momentum has to be equal to the net force applied to the layer. This force is simply the difference in radiation pressure $\mathrm{d}P_{\rm rad}$ times the area. Thus

$$\frac{1}{c}\,\mathrm{d}\mathcal{F} = \frac{\mathrm{d}p}{\mathrm{d}A\,\mathrm{d}t} = \mathrm{d}P_{\rm rad} \tag{7.31}$$

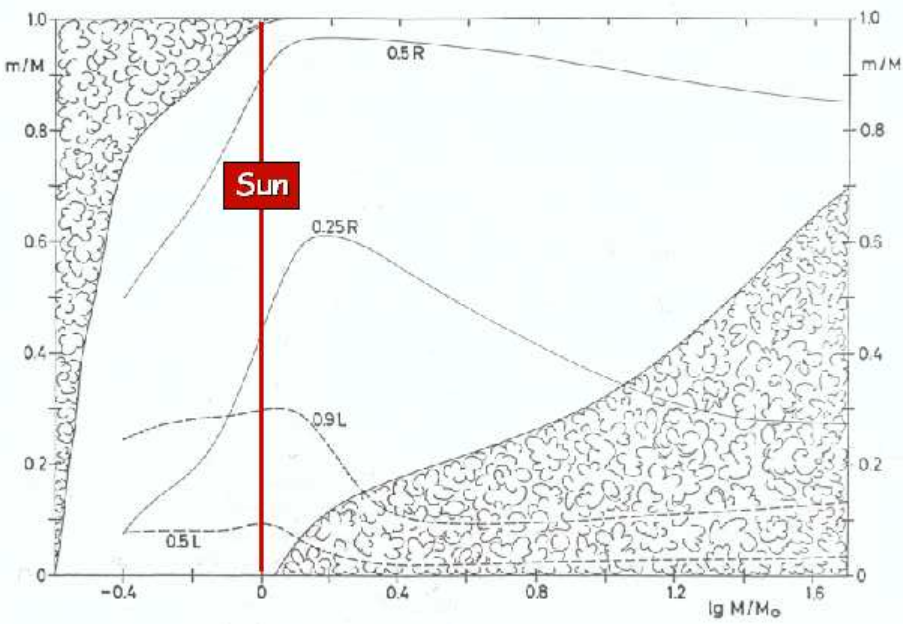


Figure 7.1: In the shadowed regions convection is important for Main Sequence stars, the x axis labels the total mass of the stars, $x = \log(M/M_\odot)$, while the y axis labels the position in the star, $y = M(r)/M$.

and

$$\boxed{\mathcal{F} = -\frac{1}{\kappa\rho}\frac{\mathrm{d}\mathcal{F}}{\mathrm{d}r} = -\frac{c}{\kappa\rho}\frac{\mathrm{d}P_{\mathrm{rad}}}{\mathrm{d}r}\,.} \tag{7.32}$$

The pressure of radiation is $P_{\mathrm{rad}} = aT^4/3$ and hence

$$\mathcal{F} = -\frac{c}{3\kappa\rho}\frac{\mathrm{d}(aT^4)}{\mathrm{d}r} = -\frac{4acT^3}{3\kappa\rho}\frac{\mathrm{d}T}{\mathrm{d}r}\,. \tag{7.33}$$

The luminosity of a shell at radius r and temperature T is thus connected with a temperature gradient $\mathrm{d}T/\mathrm{d}r$ as

$$L(r) = 4\pi r^2\mathcal{F}(r) = -\frac{16\pi r^2\, acT^3}{3\kappa\rho}\frac{\mathrm{d}T}{\mathrm{d}r}\,. \tag{7.34}$$

Convection

Convection is a cyclic mass motion carrying energy outwards, if the temperature gradient in a star becomes too large – a phenomenon familiar to everybody from water close to the boiling point. In the shadowed regions of Fig. 7.1 convection is important for main-sequence stars of various masses. In the case of the Sun, convection takes places in its outer layer, $M(r) > 0.98M_\odot$.

7.1.6 Thermal equilibrium and energy conservation

Thermal equilibrium and energy conservation require that the energy density ϵ produced per time and mass by all possible processes corresponds to an increase of the luminosity L,

$$\frac{\mathrm{d}L}{\mathrm{d}r} = 4\pi r^2 \epsilon \rho \,. \tag{7.35}$$

The energy production rate per time and mass unit, $\epsilon = \mathrm{d}E/(\mathrm{d}t)$, consists of three terms,

$$\epsilon = \epsilon_{\mathrm{grav}} + \epsilon_{\mathrm{nuc}} - \epsilon_{\nu} \,, \tag{7.36}$$

where ϵ_{nuc} accounts for the energy production by nuclear processes and ϵ_{ν} for the energy carried away by neutrinos. Both effects will be discussed in the next chapter in more detail. The term ϵ_{grav} is the only one that can be both positive (contraction) or negative (expansion of the star) and is therefore crucial for the stability of a star.

7.2 Eddington luminosity and convective instability

Our first aim in this section is to derive an upper limit on the luminosity of any stable object. We divide first Eqs. (7.32) and (7.5) and obtain

$$\frac{\mathrm{d}P_{\mathrm{rad}}}{\mathrm{d}P} = \frac{\kappa L(r)}{4\pi c G M(r)} \,. \tag{7.37}$$

Since the total pressure P is the sum of radiation and gas pressure, $P = P_{\mathrm{rad}} + P_{\mathrm{gas}}$, and both decrease outwards, it follows that $\mathrm{d}P_{\mathrm{rad}}$ and $\mathrm{d}P_{\mathrm{gas}}$ have the same sign. Hence, $\mathrm{d}P_{\mathrm{rad}} < \mathrm{d}P$ and

$$L(r) < \frac{4\pi c G M(r)}{\kappa} \,. \tag{7.38}$$

This bound relies on Eq. (7.32), i.e. it assumes that the energy is transported only by radiation, and is therefore not valid in zones of the star with convection. However, for $r = R$, energy transport should be purely radiative if the star is stable. As noted by Eddington, Eq. (7.38) represents therefore a critical luminosity,

$$L < L_{\mathrm{Edd}} = \frac{4\pi c G M}{\kappa} = 3.2 \times 10^4 \frac{M}{M_{\odot}} \frac{\kappa_{\mathrm{es}}}{\kappa} L_{\odot} \,. \tag{7.39}$$

If this luminosity is exceeded, the star ejects its outer layers in a stellar wind. To obtain the numerical estimate, we inserted the opacity κ_{es} for Thompson scattering on free electrons[1] in the last step. Next we consider what happens if for $r < R$ the condition (7.38) is violated. We rewrite first the energy transport equation as equation for $\mathrm{d}P_{\mathrm{rad}}/\mathrm{d}r$,

$$\frac{\mathrm{d}P_{\mathrm{rad}}}{\mathrm{d}r} = -\frac{\kappa \rho}{c} \frac{L(r)}{4\pi r^2} \,. \tag{7.40}$$

Then we solve the hydrostatic equilibrium equation,

$$\frac{\mathrm{d}P_{\mathrm{rad}}}{\mathrm{d}r} = -\frac{G M(r) \rho}{r^2} \,, \tag{7.41}$$

[1] Since the Thompson cross section $\sigma_{\mathrm{th}} \propto 1/m^2$, photons interact much more often with electron than with protons.

with $P = P_{\rm gas} + P_{\rm rad}$ for $\mathrm{d}P_{\rm gas}/\mathrm{d}r$,

$$\frac{\mathrm{d}P_{\rm gas}}{\mathrm{d}r} = \frac{\mathrm{d}P}{\mathrm{d}r} - \frac{\mathrm{d}P_{\rm rad}}{\mathrm{d}r} = -\frac{GM(r)\rho}{r^2} + \frac{\kappa\rho}{c}\,\frac{L(r)}{4\pi r^2} = -\frac{GM(r)\rho}{r^2}\left(1 - \frac{\kappa L(r)}{4\pi c G M(r)}\right). \quad (7.42)$$

Thus for $L > L_{\rm Ed}$, the expression in the parenthesis changes sign and the gas pressure increases outwards: The stratification of the gas layers becomes unstable and convection starts. However, the criterion $L(r) > L_{\rm edd}(r)$ is only a necessary condition for convection.

7.3 Eddington or standard model

An important test for any model for stellar structure is to derive the mass-luminosity relation it predicts and to compare it to observations, Eqs. (5.36).

7.3.1 Heuristic derivation of $L \propto M^3$

Converting the hydrostatic equilibrium equation Eq. (7.5) into a difference equation, $\mathrm{d}P \to \Delta P$ and $\mathrm{d}r \to \Delta r$, gives

$$P \propto M\rho/R\,. \quad (7.43)$$

For an ideal gas $P \propto \rho T$, hence $P \propto M\rho/R \propto \rho T$ or $T \propto M/R$ with $\rho \propto M/R^3$. The radiation transport equation gives in the same way

$$L \propto \frac{RT^4}{\kappa\rho} \propto \frac{R^4T^4}{\kappa M}\,. \quad (7.44)$$

Inserting $T \propto M/R$ shows that the luminosity is only determined by the stellar mass and the opacity,

$$L \propto \frac{M^3}{\kappa}\,.$$

The opacity in turn is controlled by the chemical composition of the star.

7.3.2 *** Analytical derivation of $L \propto M^3$ ***

Let us define

$$\eta\frac{L}{M} = \frac{L(r)}{M(r)} \quad (7.45)$$

and insert it into Eq. (7.37),

$$\frac{\mathrm{d}P_{\rm rad}}{\mathrm{d}P} = \frac{\kappa\eta}{4\pi cG}\,\frac{L}{M}\,. \quad (7.46)$$

At the surface, $\eta = 1$ by definition. In general, κ increases for small r, while η decreases ($L(r) \approx$ const. and $M(r) \to 0$ for $r \to 0$). To proceed, Eddington made the assumption that the product $\kappa\eta$ is approximately independent from the radius r, $\kappa\eta \equiv \kappa_s =$ const. Then we can integrate Eq. (7.46) immediately,

$$P_{\rm rad} = \frac{\kappa_s L}{4\pi cGM}\,P\,. \quad (7.47)$$

Defining β as the fraction the gas contributes to the total pressure, $P_{\rm rad} = (1-\beta)P$ and $P_{\rm gas} = \beta P$, we have

$$P = \frac{P_{\rm rad}}{1-\beta} = \frac{P_{\rm gas}}{\beta}\,. \tag{7.48}$$

Assuming an ideal gas law, $P_{\rm gas} = R\rho T/\mu$, and inserting $P_{\rm rad} = aT^4/3$, we obtain

$$\frac{aT^4}{3(1-\beta)} = \frac{R}{\beta\mu}\,\rho T\,. \tag{7.49}$$

Now we can express the temperature as function of the density,

$$T = \left(\frac{3R(1-\beta)}{a\beta\mu}\right)^{1/3} \rho^{1/3}\,. \tag{7.50}$$

The total equation of state, $P = P_{\rm gas}/\beta = (R\rho/\beta\mu)T$, is therefore

$$P = \underbrace{\left(\frac{3R^4(1-\beta)}{a\beta^4\mu^4}\right)^{1/3}}_{K} \rho^{4/3}\,, \tag{7.51}$$

where the factor K is a constant. Recall that an equation of state with $P = K\rho^\gamma$ is called polytropic with index $\gamma = 1 + 1/n$.

For such an equation of state, the continuity and the hydrostatic equation decouple and can be solved separately from the energy transport and the energy conservation equation. We multiply the hydrostatic equation by r^2/ρ and differentiate it then with respect to r,

$$\frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{r^2}{\rho}\frac{\mathrm{d}P}{\mathrm{d}r}\right) = -G\,\frac{\mathrm{d}M(r)}{\mathrm{d}r} = -4\pi G\rho(r)\,, \tag{7.52}$$

where we inserted the continuity equation in the last step. This second order differential equation was studied already a century ago. The result for the total mass M as function of the radius R is

$$\left(\frac{GM}{M_n}\right)^{n-1}\left(\frac{R}{R_n}\right)^{3-n} = \frac{[(n+1)K]^n}{4\pi G}\,. \tag{7.53}$$

For the case of interest, $n = 3$, the mass is independent of the radius and is determined only by K,

$$M = 4\pi M_3\left(\frac{K}{\pi G}\right)^{3/2}\,, \tag{7.54}$$

where $M_3 \approx 2.02$. Inserting K gives

$$M \propto \frac{(1-\beta)^{1/2}}{\mu^2\beta^2}\,. \tag{7.55}$$

Squaring and inserting the numerical values of the constants, we obtain "Eddington's quartic equation",

$$1-\beta = 0.003\left(\frac{M}{M_\odot}\right)^2 \mu^4\beta^4\,. \tag{7.56}$$

Its solution is shown in Fig. 7.2. What can we learn about the structure and evolution of stars from this result?

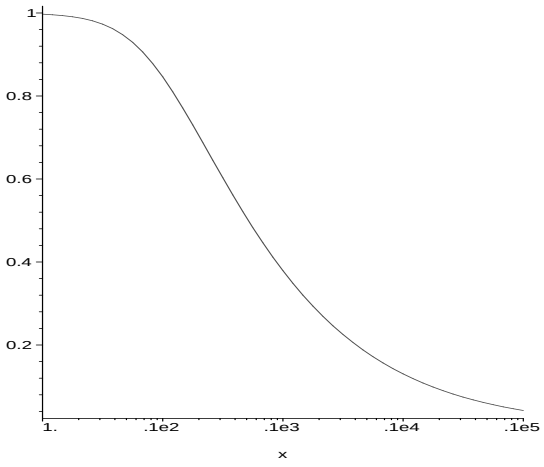


Figure 7.2: The solution β of Eddington's quartic equation as function of $x = \mu^2 M/M_\odot$.

- Remember the meaning of $\beta = P_{\rm gas}/P$. Thus $\beta \to 0$ corresponds to a free gas of photons, $\beta \to 1$ to a "primordial", cold cloud of gas. Only in the small range of $\mu^2 M$ where β has an intermediate value stars can exist.
- Inserting
$$L = \frac{4\pi c G M}{\kappa_s}(1-\beta) \tag{7.57}$$
into Eq. (7.56), we obtain
$$\frac{L}{L_\odot} = \frac{4\pi c G M_\odot}{\kappa_s L_\odot}\mu^4\beta^4\left(\frac{M}{M_\odot}\right)^3 , \tag{7.58}$$
i.e. again a power-law with exponent $\alpha = 3$ that agrees reasonably well with the observations of main-sequence stars with intermediate mass. The variation in μ can explain the observed scatter in the L-M plot.
- For stars of given composition (i.e. fixed μ), β increases as M increases. Thus, radiation pressure is more important in case of massive stars.
- Nuclear reaction cause a gradual increase of μ and therefore an decrease of β. Hence, radiation pressure becomes more important when stars become older. In their late stages, stars may eject part of their envelope, an effect called stellar winds.

7.3.3 Lifetime on the Main-Sequence

If $L/L_\odot \propto (M/M_\odot)^\alpha$ with $\alpha \sim 3.5$, then the life-time τ of stars on the main-sequence relative to the one of the Sun is
$$\frac{t}{t_\odot} = \frac{M}{M_\odot}\,\frac{L_\odot}{L} = \left(\frac{M}{M_\odot}\right)^{-\alpha+1} . \tag{7.59}$$
Hence a $10M_\odot$ star has approximately a 300 times shorter lifetime than the Sun.

7.4 Stability of stars

The longevity of stars on the main-sequence, with only small changes in radius and luminosity, means that a star has a heat-thermostat built in. This self-regulation of the nuclear burning is provided by the virial theorem together with the fact that the nuclear reaction rates $\varepsilon_{\rm nuc}$ are increasing with temperature.

In order to understand this mechanism, let us consider a small perturbation of a star, compressing it. Half of the potential energy released by the contraction is radiated away, while the other half leads because of $E_{\rm kin} = -1/2 E_{\rm pot} \propto 1/R$ to its heating. Since nuclear reaction rates increase strongly with temperature, energy production by nuclear reaction increases and therefore also the pressure. Thus the star expands until it reaches again an equilibrium configuration.

7.5 Variable stars – period-luminosity relation of Cepheids

Cepheids are named after δ Cephei, a star with a period of 5.4 days and $\Delta m = 0.7$ mag. Generally, the periods of Cepheids vary between $P \sim$ 1–100 days. H. Leavitt studied 1908–1912 Cepheids in the Small and Large Magellanic Clouds, i.e. at a fixed distance. She discovered a relation between the period and the luminosity of Cepheids,

$$M = -2.81 \log(P/\mathrm{d}) - 1.43\,. \tag{7.60}$$

Cepheids are our first example of a "standard candle". As soon as this relation is calibrated (e.g. by measuring the distance to the LMC or SMC by other means), it can be used to obtain distance to any other galaxy with Cepheides.

The spectral lines of Cepheids are Doppler shifted with same period as the light-curve. This suggest radial oscillations of the star as a reason for the periodic luminosity change.

Let us consider a small radial perturbation of a stellar layer close to the surface. If the radius of this layer is decreased by a small amount, its density and thus the pressure increase. The increased pressure drives the layer back, it overshots its equilibrium position, and as a result the star starts to oscillate.

Next we consider the effect of a change in the opacity: The opacity κ of the outer layer of a normal star decreases for increasing temperatures. Thus if the radius of a layer is again decreased by a small perturbation, the density and thus the pressure and also the temperature increase. The resulting decrease of the opacity eases the radiation transport and counteracts the increase in temperature and pressure. Hence the oscillations are damped.

The envelopes of Cepheid stars are characterized by a narrow range of temperatures and densities where the opacity increases with temperature. Then the temperature dependence of the opacity does not quench the oscillations, and the amplitude of the oscillations can be stable and large.

7.6 Exercises

1. Show that the relation (7.12) for the pressure of non-interacting particles holds also for a quantum gas. [Hint: Derive first $P = -(\partial U/\partial V)_S$ and consider then how the energy of a free particle quantized in a box changes, if you change the volume of the box.]

8 Nuclear processes in stars

8.1 Possible energy sources of stars

The origin of the radiation energy emitted by the Sun was questioned already 1846, soon after the establishment of the law of energy conservation, by J.R. Meyer. It remained mysterious for ninety years. Since the temperature on the Earth was approximately constant during the last $\tau \approx 4 \times 10^9$ years, the solar luminosity should be also roughly constant. Thus we can bound the minimal energy output of the Sun as $\tau L_\odot \gtrsim 6 \times 10^{50}$ erg.

8.1.1 Gravitational energy

The mass $\mathrm{d}M(r)$ contained in the shell between r and $r + \mathrm{d}r$ feels only the gravitational attraction of the enclosed mass $M(r)$. Finding the total gravitational energy of a spherical star amounts therefore to integrate $\mathrm{d}U_{\rm pot} = -GM(r)\mathrm{d}M(r)/r$,

$$U_{\rm pot}(R) = -G \int_0^R \frac{M(r) dM(r)}{r} \,. \tag{8.1}$$

We assume that the star is homogeneous and use $M(r) = (4\pi/3)r^3\rho$ as well as $\mathrm{d}M(r) = 4\pi r^2 \rho \mathrm{d}r$. Then

$$U_{\rm pot}(R) = -G\frac{(4\pi)^2}{3}\rho^2 \int_0^R \mathrm{d}r\, r^4 = -\frac{3G}{5} \underbrace{\left(\frac{4\pi}{3}\right)^2}_{M^2/R^6} \rho^2 R^5 = -\frac{3}{5}\,\frac{GM^2}{R} \,. \tag{8.2}$$

According to the virial theorem, $-2U_{\rm kin} = U_{\rm pot}$, the total energy U gained by the Sun contracting from infinity is $U = -U_{\rm pot}/2 = 3GM^2/(10R)$ while the other half is radiated away.

Mayer proposed meteorites falling into the Sun, Helmholtz and Kelvin the contraction of the Sun itself as energy source. In the latter case $U \approx (3GM_\odot^2)/(10R_\odot) \approx 1 \times 10^{48}$erg, and this energy would be consumed within the time $U/L_\odot \approx 2 \times 10^{14}\,\mathrm{s} = 1 \times 10^7\,\mathrm{yr}$ compared to the age of the Earth of about 4 Gyr.

8.1.2 Chemical reactions

The typical energy scale of chemical reactions that corresponds to a fraction of the binding energy of molecules, is of the order of 1 eV. (Remember the binding energy of hydrogen is 13.6 eV.) The maximal number of molecules in the Sun is $M_\odot/(2m_p) \sim 2 \times 10^{30}\mathrm{kg}/4 \times 10^{-27}\mathrm{kg} \approx 0.5 \times 10^{57}$, or the total energy $0.5 \times 10^{48}\mathrm{GeV} \sim 8 \times 10^{44}$ erg and thus even less that in gravitational energy.

	n	p	2p+2n	^{4}He
m/u	1.0090	1.0081	4.0342	4.0039

Table 8.1: The masses of nucleons and ^{4}He in atomic mass units u.

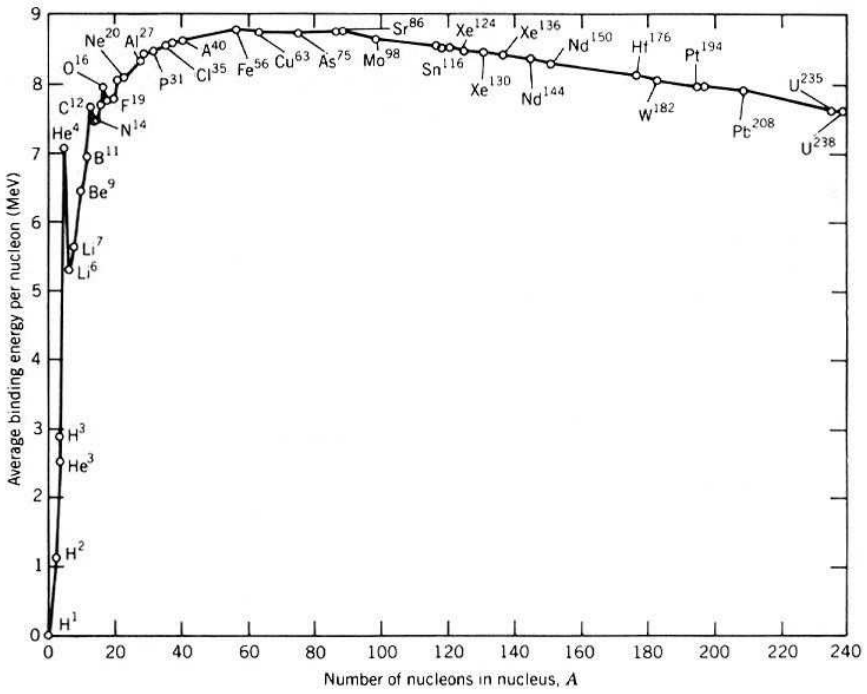


Figure 8.1: The binding energy per nucleon E_b/A as function of the nucleon number A.

8.1.3 Nuclear fusion

The total mass $m(Z, N)$ of a nucleus with Z protons and N neutrons is because of its binding energy E_b somewhat reduced compared to its constituent mass,

$$E_b/c^2 = Zm_H + Nm_n - m(Z, N)\,. \tag{8.3}$$

Measured values of the binding energy per nucleon E_b/A as function of nucleon number A are shown in Fig. 8.1.

From Fig. 8.1, one recognizes that the binding energy per nucleon E_b/A has its maximum at $A \sim 56$, i.e. iron ^{56}Fe is the most stable element. Thus there are two possible ways to release energy by nuclear reactions: Two light nuclei can be fused or a heavy one can be 'broken' up.

While E_b/A is a rather smooth function of A for $A > 20$, there several peaks visible for small A: ^{4}He, ^{12}C, ^{14}N, ^{16}O and ^{20}Ne are energetically much more favourable than their neighbouring elements. The bound states of nucleons in nuclei have a similar shell structure as electrons in atoms. Nuclei with filled shells are especially stable, as noble gases are especially stable atoms.

interaction	gravitation	electromagnetic	weak	strong
force	$G_N m_p^2/r^2$	e^2/r^2	$G_F E^2 e^{-r/\lambda_W}/r^2$	$\alpha_s e^{-r\lambda_\pi}/r^2$
coupling	$G_N = (\hbar c)/M_{\rm Pl}^2$	$\alpha_{\rm em} = e^2/(\hbar c) \approx 1/137$	$G_F = \sqrt{2}g^2/(8M_W^2)$	$\alpha_s = g_s^2/(\hbar c) \approx 1$
range	∞	∞	$\lambda_W = h/(M_W c)$	$\lambda_\pi = h/(M_\pi c)$
particle	graviton	photon	W and Z-boson	pions
mass	0	0	80 and 91 GeV	140 MeV
decay process		$\pi \to 2\gamma$	$\mu^- \to e^+ + \bar{\nu}_e + \nu_\mu$	$\rho \to 2\pi$
decay time		$\tau \sim 10^{-16}$s	$\tau \sim 10^{-6}$s	$\tau \sim 10^{-22}$s

Table 8.2: Some properties of the four fundamental forces.

In Tab. 8.1, the masses of ^{4}He and its constituent nucleons are compared. These numbers imply that the fusion of four protons to ^{4}He releases 26.2 MeV energy, a factor 10^7 more than in our estimate for chemical reactions. Converting a solar mass into helium releases $M_\odot/(4m_p) \times 26.2\,\text{MeV} \approx 1.25 \times 10^{52}\,\text{erg}$ and thus around 5% of the Sun have been already converted into helium.

As shown by the failure of using fusion for the energy production on Earth and the longevity of stars fusion is a non-trivial process: All four interactions are involved in the energy release by nuclear fusion. The strong interaction leads to the binding of nucleons in nuclei, the Coulomb repulsion has to be overcome to combine them, and the weak interactions convert half of the protons into neutrons. Finally, gravitation is responsible for the confinement of matter, heating it up the proto-star to the necessary "start temperature" and serving then as a heat regulator.

8.2 Excursion: Fundamental interactions

Bosons and forces We know four fundamental forces in nature, that are all transmitted by the exchange of bosons, i.e. particles with integer spin. Two of them are the gravitational and electromagnetic force we have already encountered. These two forces fall-off like $1/r^2$ and are important on macroscopic scales. By contrast, the weak and the strong force are exponentially suppressed beyond a distance which corresponds to the Compton wavelength $\lambda = h/(mc)$ of the particles that mediate these forces at energies relevant in astrophysics. A comparison of some properties of the four fundamental forces is given in Table 8.2.

The first line shows the force law characterising the different interactions. The strength of an interactions depends both on how strong the charges or couplings are and how large the Compton wavelength λ of the exchange particles is. In a simple analogy, one can imagine the cross section for a specific reaction as a grey disc: Its area is given by $\pi\lambda^2$ and its greyness depends on the magnitude of the coupling constant.

Fermions and the family structure Fermions carry half-integer spin and obey the Pauli principle, a principle that is crucial for the stability of matter. Protons and neutrons differ from electrons and neutrinos in that they have also strong interactions.

The discovery of the muon happened accidentally in 1936-1937. The realization that it is an exact copy of an electron having only a different mass came as a real surprise. Nowadays we know that there exist three different "families" or "flavors" of fermions, – but still not

why three. For instance, the electron has additionally to the muon with mass $m_\mu = 207\, m_e$ an even heavier copy, the tau with $m_\tau = 3480\, m_e$. In the same way, three different neutrinos exist. They are called ν_e, ν_μ and ν_τ, depending on if they are generated together with an electron, muon or tau.

charged leptons	e^-	μ^-	τ^-
neutrinos	ν_e	ν_μ	ν_τ

8.3 Thermonuclear reactions and the Gamov peak

Coulomb barrier-classically Let us consider the forces between two protons: For $r \gg \lambda_\pi = h/(m_\pi c)$, the Coulomb force dominates over the strong force. The size of a nucleon, $r_N \approx 10^{-13}\,\mathrm{cm}$, is comparable to the Compton wave-length of the pion and we will not distinguish between them in the following. Thus nuclei should have classically the energy $V \approx Z_1 Z_2 e^2/r_N$ to cross the "Coulomb barrier" and to reach another nucleus of size r_N. For a thermal plasma of particles with reduced mass μ and charges $Z_i e$, this condition reads

$$\frac{1}{2}\mu\langle v^2\rangle = \frac{3}{2}kT = \frac{Z_1 Z_2 e^2}{r}\,. \tag{8.4}$$

Specifically, we obtain for protons with $Z_1 = Z_2 = 1$ and size $r_N \approx 10^{-13}\,\mathrm{cm}$ that the temperature should be above

$$T \gtrsim \frac{2e^2}{3kr_N} \approx 10^{10}\,\mathrm{K}\,. \tag{8.5}$$

On the other hand, we have estimated the central temperature of the Sun as $T_c \approx 10^7\,\mathrm{K}$. Hence we should expect that only the tiny fraction of protons with $v^2 \gtrsim 1000\langle v^2\rangle$ is able to cross the Coulomb barrier.

Coulomb barrier-tunneling Quantum mechanically, tunneling though the Coulomb barrier is possible. The wave-function of a particle with $E - V < 0$ is non-zero, but exponentially suppressed. In order to avoid a too strong suppression, we require that $\lambda = h/p \approx r_N$,

$$\frac{h^2}{2\mu\lambda^2} = \frac{Z_1 Z_2 e^2}{\lambda}\,. \tag{8.6}$$

Inserting $\lambda = h^2/(2Z_1 Z_2 e^2 \mu)$ for r_N in Eq. (8.5) gives

$$T \gtrsim \frac{4Z_1^2 Z_2^2 e^4 \mu}{3h^2 k} \approx 10^7\mathrm{K}\,. \tag{8.7}$$

Hence we expect that quantum effects lead to a not too strong suppression of fusion rates.

Next we want to make this statement more precise: We can estimate the probability $p(x)$ that a nuclei is inside the classically forbidden region as

$$p(x)\mathrm{d}x \propto \exp(-Ax/\lambda)\mathrm{d}x \propto \exp(-Axmv/h)\mathrm{d}x\,. \tag{8.8}$$

If the two particles come closer than r_N, the nuclear forces dominate and they fuse. Therefore the probability for a fusion process to happen is proportional to $p(r_N)$ with $r_N = 2Z_1 Z_2 e^2/(\mu v^2)$ from Eq. (8.4), or

$$p(r_N)\mathrm{d}x \propto \exp(-bZ_1 Z_2 e^2/E^{1/2})\mathrm{d}x\,. \tag{8.9}$$

Reaction rates We introduced the interaction depth $\tau = nl\sigma$ as the probability that a particle interacts once travelling the distance l through targets with density n. The rate Γ of such interactions, i.e. the number of reactions per time follows then with $l = vt$ as $\Gamma \equiv n\sigma v$, if the particle moves with velocity v. Since the energies of the particles are not uniform, but distributed according a non-relativistic Maxwell-Boltzmann distribution,

$$n_v dv = \left(\frac{m}{2\pi kT}\right)^{3/2} \exp\left(-\frac{mv^2}{2kT}\right) 4\pi v^2 \mathrm{d}v \propto \mathrm{e}^{-E/kT} E^{1/2} \mathrm{d}E\,, \tag{8.10}$$

we should average over the distribution n_v. Then the rate follows as

$$\Gamma \propto \int dE\, \sigma(E) E\, \mathrm{e}^{-b/E^{1/2}} \mathrm{e}^{-E/kT}\,. \tag{8.11}$$

Cross section for strong interactions are of geometrical nature, $\sigma(E) \approx \pi\lambda^2$ and with $\lambda = h/p$ it follows $\sigma(E) \propto 1/E$. Therefore it is convenient to introduce the so-called S-factor $S(E) = E\sigma(E)$ of the reaction,

$$\Gamma \propto \int dE\, S(E) \mathrm{e}^{-b/E^{1/2}} \mathrm{e}^{-E/kT}\,. \tag{8.12}$$

If the cross-section behaves indeed as $\sigma(E) \propto 1/E$, then $S(E)$ is a slowly varying function. The reminder of the integrand is sharply peaked ("Gamov peak") in the region around 10 keV: For higher energies, the integrand is suppressed because of the Boltzmann factor $\exp(-E/kT)$, while for lower energies the tunneling probability $\exp(-b/E^{1/2})$ goes to zero.

8.4 Main nuclear burning reactions

8.4.1 Hydrogen burning: pp-chains and CNO-cycle

The pp-chains are shown in detail in in the left panel of Fig. 8.2. Its main chain uses three steps:
Step 1: $p + p \rightarrow d + e^+ + \nu_e$
Step 2: $p + d \rightarrow {}^3\mathrm{He} + \gamma$
Step 2: ${}^3\mathrm{He} + {}^3\mathrm{He} \rightarrow {}^4\mathrm{He} + p + p$.

The CNO-cyle is shown in detail in the right panel of Fig. 8.2. The small inlet compares the temperature dependence of the pp-chain and the CNO-cycle: For solar temperatures, the contribution of the pp-chains to the solar energy production is four order of magnitudes more important than the CNO-cycle.

Radius-mass relation of main-sequence stars Hydrogen is burned at nearly fixed temperature T. Via the virial theorem, also the gravitational potential is nearly the same for all main-sequence stars and thus $GM/R \approx$ const. As a result, the radius of main-sequence stars increases approximate linearly with the stellar mass.

8.4.2 Later phases

The increasing Coulomb barrier for heavier nuclei means that the fusion of heavier nuclei requires higher and higher temperatures. Therefore the different fusion phases – hydrogen, helium, carbon,... burning – never coexist, but follow each-other. Since the temperature decreases outwards, the fraction of the core participating in fusion reactions becomes smaller in each new burning phase, cf. Fig. 8.3.

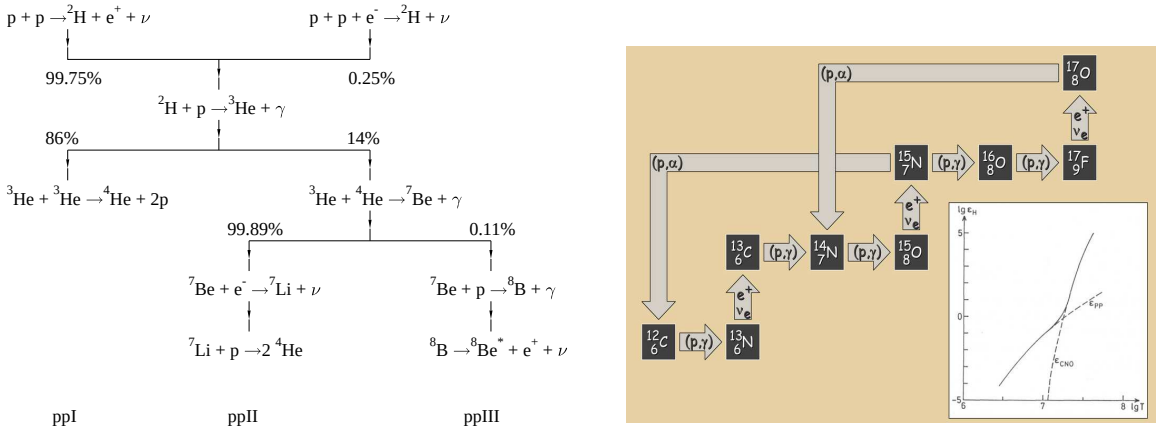


Figure 8.2: Left: proton-proton chains Right: CNO cycle.

Burning Phase	Dominant Process	T_c [keV]	ρ_c [g/cm^3]	L_γ [10^4 L_{sun}]	L_ν/L_γ	Duration [years]
Hydrogen	H → He	3	5.9	2.1	-	1.2×10^7
Helium	He → C, O	14	1.3×10^3	6.0	1.7×10^{-5}	1.3×10^6
Carbon	C → Ne, Mg	53	1.7×10^5	8.6	1.0	6.3×10^3
Neon	Ne → O, Mg	110	1.6×10^7	9.6	1.8×10^3	7.0
Oxygen	O → Si	160	9.7×10^7	9.6	2.1×10^4	1.7
Silicon	Si → Fe, Ni	270	2.3×10^8	9.6	9.2×10^5	6 days

Figure 8.3: Burning phases of a 15 $M_\odot$ star.

- Hydrogen burning $4\mathrm{p}+2e^- \rightarrow {}^4\mathrm{He}+2\nu_e$
  - proceeds by pp chains and CNO cycle.
  - no heavier elements are formed, because no stable isotopes with mass number $A = 8$ exist.
  - neutrinos are produced by weak reaction in $p \rightarrow n$ conversions.
  - the typical temperature is 10^7 K (~ 1 keV).
- Helium burning ${}^4\mathrm{He}+{}^4\mathrm{He}+{}^4\mathrm{He} \leftrightarrow {}^8\mathrm{Be}+{}^4\mathrm{He} \rightarrow {}^{12}\mathrm{C}$
  - triple alpha reaction builds up Be with a concentration $\sim 10^{-9}$ which acts a catalyzer for the production of ^{12}C:
    ${}^{12}\mathrm{C}+{}^4\mathrm{He}\rightarrow{}^{16}\mathrm{O}$
    ${}^{16}\mathrm{O}+{}^4\mathrm{He}\rightarrow{}^{20}\mathrm{Ne}$.
  - the typical temperature is 10^8 K (~ 10 keV).
- Carbon burning
  - many reactions like ${}^{12}\mathrm{C}+{}^{12}\mathrm{C}\rightarrow{}^{20}\mathrm{Ne}+{}^4\mathrm{He}$ etc.
  - the typical temperature is 10^9 K (~ 100 keV).

8.5 Standard solar model and helioseismology

A "standard solar model" uses the i) conservation laws together with ii) an equation of state, iii) energy production by fusion processes and iv) energy transport equation to evolve the Sun in small time-step from some initial conditions to its present age of 4.5×10^9 yr. Agreement with present-day luminosity and radius is enforced by tuning unknown pre-solar helium abundance and a technical parameter determining convection.

The structural changes in the Sun, as it evolves, are caused by the nuclear reactions occurring in the central regions of the Sun. The transmutation of four hydrogen atoms into one helium reduces the number density of particles in the central regions which decreases the pressure. The pressure decrease does not actually occur because the surrounding layers respond to the force imbalance by contracting in the central regions. Half of the gravitational energy released from the contraction goes to raising the temperature of the central regions (the other half, according to the virial theorem, is radiated away). The increased temperature, by the ideal gas law, increases the pressure of this region and restores the balance between the pressure and gravitational forces. The larger mean molecular weight increases the luminosity of the star, and the rate of nuclear reactions. Also, while the central layers contract, the outer regions expand, in a sense, compensating for the steepening temperature gradients in the central regions. Therefore, as the Sun evolves from the zero age both its luminosity and radius increase.

The solar standard model can be tested by **helioseismology:** Inside the Sun, pressure waves with periods between 2 min and 1 h have been observed by Doppler shifts of spectral lines on the solar surface. (The amplitude is a few meters: The solar surface is rising and falling a few meters every few minutes.) The oscillation pattern depends on the sound speed in the solar interior and allows therefore to test predictions of the solar standard model in the interior.

8.6 Solar neutrinos

Solar neutrinos flux From $L_\odot = 4 \times 10^{33}erg/s= 2 \times 10^{39}$MeV/s and the energy release of 26.2 MeV per reaction, the minimal number of neutrinos produced is $\dot{N}_\nu = 2 \times 10^{38}$/s. As we

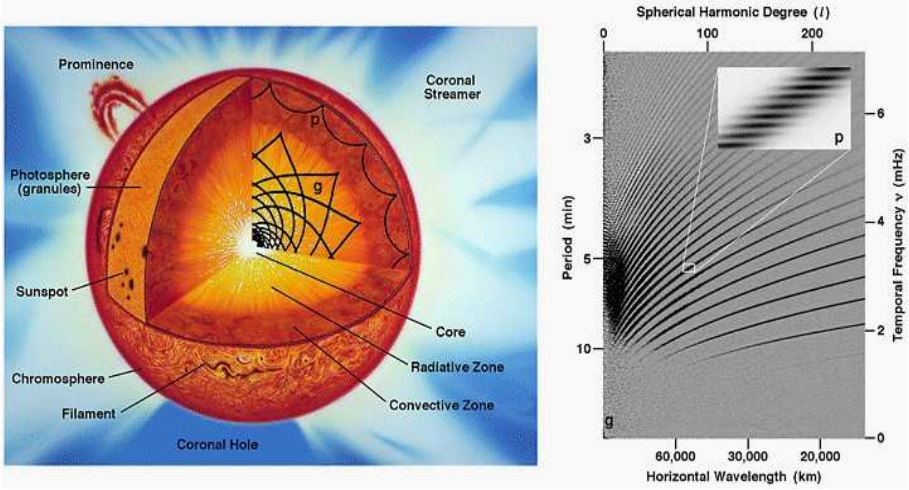


Figure 8.4: Schematic picture of the Sun and different oscillation modes (left) together with the measured oscillation frequencies (right).

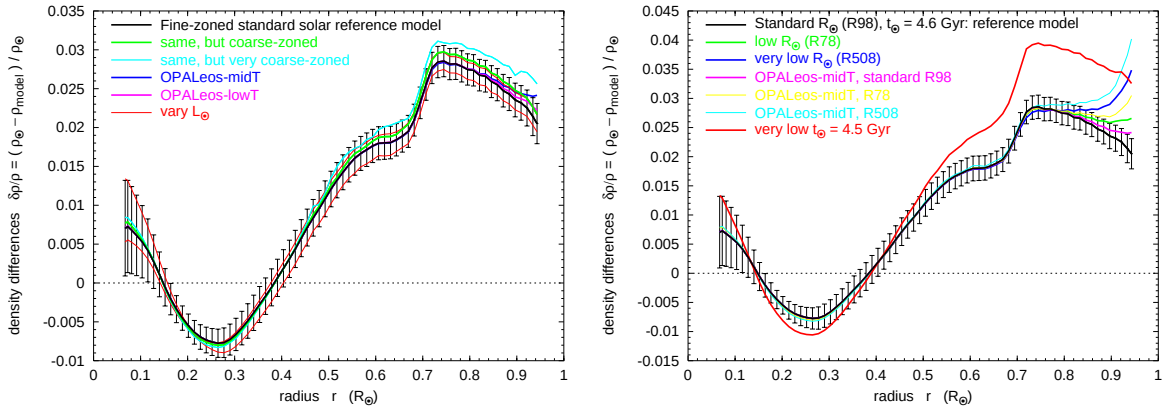


Figure 8.5: Relative density difference between the solar standard model or its variations and the deduced values from helioseismology as function of the solar radius.

have seen, photon perform a random-walk. Neutrinos have much smaller interactions,

$$\sigma_\nu = 10^{-43}\mathrm{cm}^2 \frac{E_\nu}{\mathrm{MeV}} \tag{8.13}$$

and thus they can escape from the Sun: The interaction depth for a neutrino in the Sun is approximately

$$\tau = \sigma_\nu n R_\odot = 10^{-9}\,. \tag{8.14}$$

At the distance of the Earth, this corresponds to a neutrino flux of

$$\phi_\nu = \frac{\dot{N}_\nu}{4\pi D^2} = 7 \times 10^{10} \frac{1}{\mathrm{cm}^2\mathrm{s}}\,. \tag{8.15}$$

(Alternatively, one can derive ϕ_ν from the known value of the Solar constant S via $\phi_\nu = 2S/(26.2\,\mathrm{MeV})$.)

Weak interactions in the Sun produce always electron neutrinos, i.e. $p \to n + e^+ + \nu_e$, but not $p \to n + \mu^+ + \nu_\mu$ or $p \to n + \tau^+ + \nu_\tau$, because the energy released in nuclear reaction and the temperature is too small too produce a μ or τ. Similarly, only ν_e neutrinos are detected in radio-chemical reactions via "inverse beta-decay", while all type of neutrinos can be detected in elastic scattering on electrons.

Solar neutrino experiments Radio-chemical experiments detect neutrinos by "inverse beta-decay" in suitable nuclei. The historically first isotope used was chlorine, $\nu_e + {}^{37}\mathrm{Cl} \to {}^{37}\mathrm{Ar} + e^-$, i.e. changing a neutron inside a ^{37}Cl nuclei into a proton, thereby converting it into a ^{37}Ar. The disadvantage of this reaction was its high energy threshold, $E_\nu \geq 0.814$MeV, that made this experiment only sensitive to 9% of all solar neutrinos, cf. the energy spectrum of solar neutrinos in Fig. 8.6.

The experiment consisted of 615 tons of $C_2\ Cl_4$ solutions in a mine 1500 m underground. After exposure of around 2-3 months, a few Ar atoms were produced. They were chemically extracted and counted by their subsequent decays (halftime 35 days). Starting from first data in 1968, a deficit appeared relative to theoretical expected fluxes: only 30% of predicted event number is measured. This deficit was dubbed "solar neutrino problem." Finding the solution to this problem required more than 30 years of intensive experimental and theoretical work.

Starting from 1991, 2 Gallium experiments $\nu_e + {}^{71}\mathrm{Ga} \to {}^{71}\mathrm{Ge} + e^-$ with threshold $E_\nu \geq$ 233keV took data. They found 55% of the expected neutrino flux, corresponding to about 9 atoms of ^{71}Ge in 30 tons of solution containing 12 tons ^{71}Ga, after three weeks of run time.

What are plausible solutions to the solar neutrino problem?

- The experimental results might be wrong: For instance, the extraction efficiency of the few Argon or Gallium atoms might have been not 100%.
- The nuclear cross sections used as input for the calculation of the solar standard model were measured at higher energies and had to be extrapolation to the energy range of the Gamov peak. This extrapolation might be wrong, if an unknown resonance lies close to the Gamov peak.
- The solar standard model itself and in particular the central temperature T_c might have a larger error than estimated.
- Finally, the particle physics properties of neutrinos might be different than expected and ν_e may not survive the travel to the Earth.

As explained in the next subsection, the last possibility is the correct explanation: A large fraction of electron neutrinos oscillate into muon and tau neutrinos that do not participate in inverse-beta reactions.

8.6.1 *** Solar neutrino problem and neutrino oscillations ***

We can consider the time-evolution of a neutrino analogous to the quantum mechanical problem with n states. If we denote the eigenstates of H by $|n\rangle$, $H|n\rangle = n|n\rangle$, then their time-evolution is simply

$$|n(t')\rangle = |n(t)\rangle \mathrm{e}^{-iE_n(t'-t)} \,. \tag{8.16}$$

An arbitrary (normalized) state $|\alpha(t)\rangle$ is connected by an unitary transformation with the eigenstates $|n\rangle$,

$$|\alpha(t')\rangle = \sum_n U_{\alpha n} |n(t)\rangle \mathrm{e}^{-iE_n(t'-t)} \,. \tag{8.17}$$

Thus the probability to measure the eigen-value n will depend on time, if the state is not an eigenstate of H.

For a two-level system and the unitarity condition $UU^\dagger = UU^{-1} = 1$, it follows

$$U = \begin{pmatrix} \cos\vartheta & \sin\vartheta \\ -\sin\vartheta & \cos\vartheta \end{pmatrix} \,. \tag{8.18}$$

By definition we call the flavour of a neutrino according to the produced charged leptons in weak interaction, ν_α with $\alpha = e, \mu, \tau$. These states are not necessary eigenstates of the free Hamiltonian H describing the propagation of particles.

Analogous to Eq. (8.17), the time-evolution of a state describing an electron-neutrino, $\nu_e(t)$, follows as

$$|\nu_e(t)\rangle = \sum_n U_{en} |n\rangle \mathrm{e}^{-iE_n t} = \cos\vartheta |\nu_1\rangle \mathrm{e}^{-iE_1 t} + \sin\vartheta |\nu_2\rangle \mathrm{e}^{-iE_2 t} \tag{8.19}$$

Thus the probability to find a neutrino in a certain flavour α will depend in general ($\vartheta \neq 0$) on time.

The survival probability $P_{\nu_e \to \nu_e}$ of an ν_e is

$$P_{\nu_e \to \nu_e}(t) = |\langle |\nu_e(t)|\nu_e(0)\rangle|^2 = 1 - \sin^2\left[(E_2 - E_1)t/2\right] \sin^2(2\vartheta) \,. \tag{8.20}$$

For ultra-relativistic neutrinos, $E_i = (m_i^2 + p^2)^{1/2} \approx p^2 + m_i^2/(2p)$, and replacing the travel time t by the distance L,

$$P_{\nu_e \to \nu_e}(L) = 1 - \sin^2\left[\frac{\Delta m^2}{4E} L\right] \sin^2(2\vartheta) \,. \tag{8.21}$$

The two-neutrino survival probability depends on

- the vacuum oscillation length

$$l_{\mathrm{osc}} = 4\pi \frac{E}{\Delta m^2} \approx 2.5 \text{ m} \left(\frac{E}{\mathrm{MeV}}\right) \left(\frac{\mathrm{eV}^2}{\Delta m^2}\right)$$

- mass squared difference, no sensitivity to the absolute mass

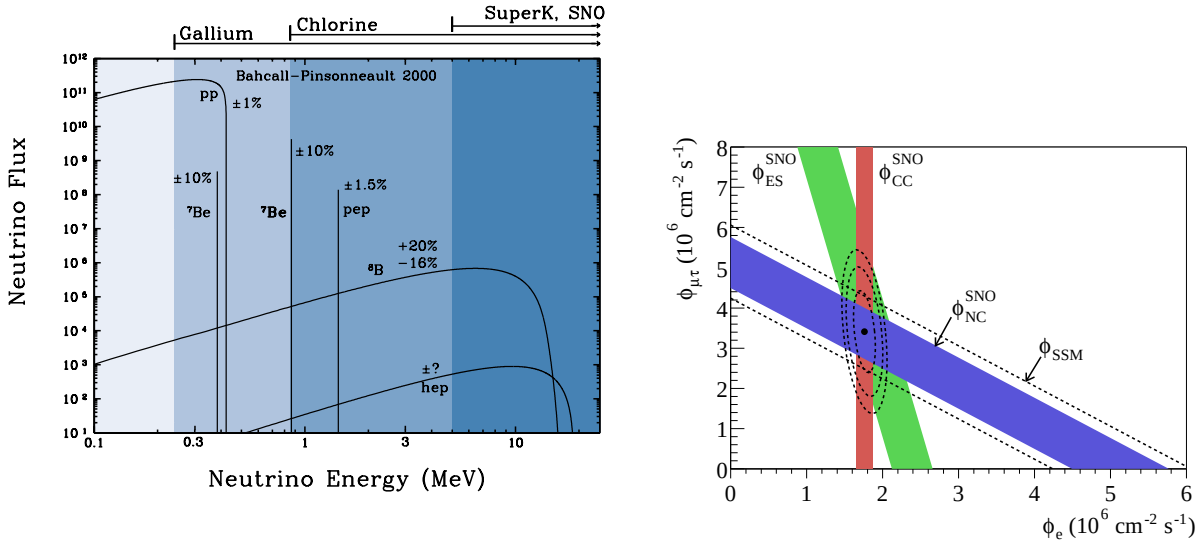


Figure 8.6: Left: The solar standard model predicts the neutrino flux and thus also the number of events that should be measured. Right: Results of the SNO experiment.

- the mixing angle ϑ.

The measurements shown in the right panel of Fig. 8.6 demonstrated that neutrino flavor conversions take place: The red band shows the flux of ν_e, measured by inverse-beta decay reactions. This flux is just 36% of the expected flux from the solar standard model. However, the experiment was also able to measure the flux of ν_μ and ν_τ, cf. the green and the blue band. Summing all three favors up, the three bands cross and one obtains the value predicted by the solar standard model. Hence neutrinos oscillate, changing their flavor on the way from the solar core to the Earth.

9 End points of stellar evolution

9.1 Observations of Sirius B

Sirius is with a distance of 2.6 pc the fifth closest stellar system to the Sun. Analyzing the motions of Sirius 1833–1844, F.W. Bessel concluded that it had an unseen companion, with a orbital period $P \sim 50$ yr. In 1862, A. Clark discovered this companion, Sirius B, at apastron or the time of maximal separation of the two components of the binary system. Following-up observations showed that the mass of Sirius B equals approximately the one of the Sun, $M \approx M_\odot$. Sirius B's peculiar properties were not established until the next apastron, 1915, by W.S. Adams. He noted the high temperature of Sirius B, which requires together with its small luminosity an extremely small radius and thus a large density of this star.

Ex.: Find the mean density of Sirius B from its apparent magnitude, $m = 8.5$, and its surface temperature $T = 25,000\,\mathrm{K}$.
We convert first the apparent into an absolute magnitude,

$$M = m - 5\log(d/10\mathrm{pc} = 11.4\,,$$

and express then the absolute magnitude as a luminosity,

$$L = 3.02 \times 10^{28} W \times 10^{-0.4M} = 3.84 \times 10^{26} W\,.$$

Now we can use the Stefan-Boltzmann law and obtain

$$\frac{R}{R_\odot} = \left(\frac{L}{L_\odot}\right)^{1/2} \left(\frac{T}{T_\odot}\right)^2 \approx 10^{-2}\,.$$

Therefore the mean density of Sirius B is a factor 10^6 higher than the one of the Sun. More precisely one finds $\rho = 3 \times 10^6 \mathrm{g/cm^3}$.

We now apply the lower limit for the central pressure of a star in hydrostatic equilibrium, Eq. (7.9), to Sirius B,

$$P_c > \frac{M^2}{8\pi R^4} = 4 \times 10^8 \mathrm{bar} \left(\frac{M}{M_\odot}\right)^2 \left(\frac{R_\odot}{R}\right)^4 = 4 \times 10^{16}\,\mathrm{bar}\,. \tag{9.1}$$

What would be the central temperature T_c needed, if the pressure is dominated by an ideal gas? From the ideal gas law, we find

$$T_c = \frac{P_c}{nk} \sim 10^2 T_{c,\odot} \approx 10^9\,\mathrm{K}\,. \tag{9.2}$$

For such a high central temperature, the temperature gradient $\mathrm{d}T/\mathrm{d}r$ in Sirius B would be a factor 10^4 larger than in the Sun. This would in turn require a larger luminosity $L(r)$ and a larger energy production rate ε than in a main sequence star. The solution to this puzzle is that the main source of pressure in such compact stars has a different origin.

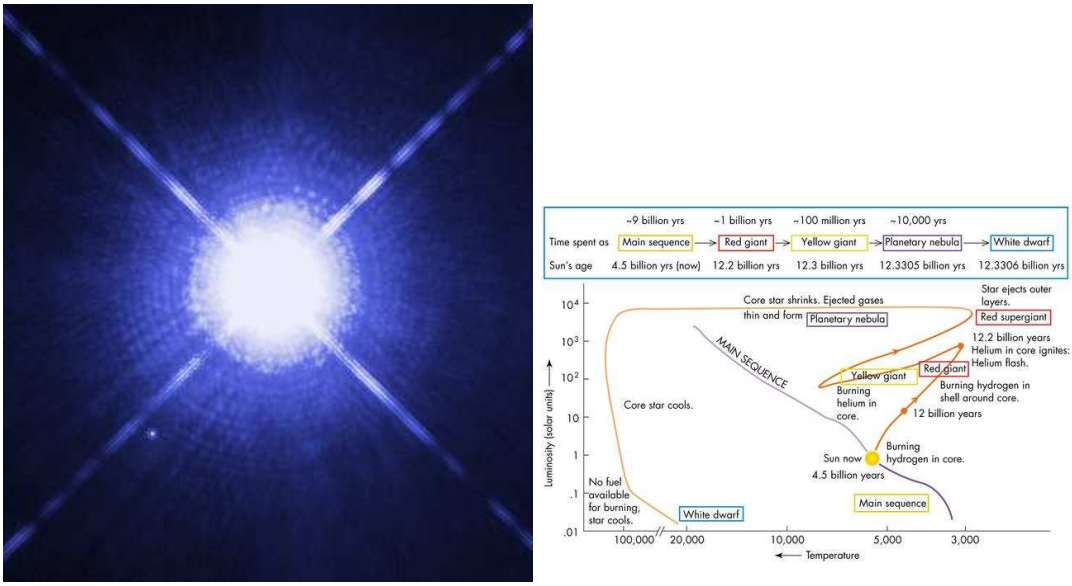


Figure 9.1: Left: Hubble Space Telescope image showing Sirius A, the brightest star in our nighttime sky, along with its faint, tiny stellar companion, Sirius B. Right: Evolution of the Sun.

9.2 Pressure of a degenerate fermion gas

For a classical gas, $P = nkT$, and thus in the limit of zero temperature, also the pressure inside a star goes to zero. How can a star be stabilized after the fusion processes and thus energy production stopped?

The Pauli principle forbids that fermions can occupy the same quantum state. In statistical mechanics, Heisenberg's uncertainty principle $\Delta x \Delta p \geq \hbar$ together with Pauli's principle imply that each phase-space volume $(1/\hbar)dxdp$ can be occupied by only one fermionic state.

If we use $\Delta x = n^{-1/3}$ and $\Delta p \approx \hbar/\Delta x \approx \hbar n^{1/3}$ together with $v = p/m$ valid for a non-relativistic particles, we obtain for the pressure of a degenerate fermion gas

$$P \approx nvp \approx \frac{\hbar^2 n^{5/3}}{m} \qquad \text{or} \qquad P \propto \rho^{5/3}\,. \tag{9.3}$$

For relativistic particles, we can obtain an estimate for the pressure inserting $v = c$,

$$P \approx ncp \approx \frac{c\hbar^2 n^{4/3}}{m} \qquad \text{or} \qquad P \propto \rho^{4/3}\,. \tag{9.4}$$

Note the following important points:

1. Both the non-relativistic and the relativistic pressure laws are polytropic equations of state, $P = K\rho^\gamma$.
2. A non-relativistic degenerate Fermi gas has the same adiabatic index, $\gamma = 5/3$, as an ideal gas, while a relativistic degenerate Fermi gas has the same adiabatic index, $\gamma = 4/3$, as radiation. Since $\gamma = 4/3$ marks the border between stable and unstable systems (cf. Eq. (7.29)), we can expect that there exists a critical mass for cold matter: Increasing the mass of a star beyond the level that its constituents become relativistic should "unbind" its constituents.

3. The pressure is inversely proportional to the fermion mass, $P \propto 1/m$. Thus the degeneracy pressure will become important first for electrons.

Let us compute the pressure of a degenerate non-relativistic electron gas and check if it is consistent for Sirius B with the lower limit for the central pressure. First, we verify that electrons are non-relativistic: With $n_e \approx \rho/(2m_p)$ and $\rho \approx 10^6 \mathrm{g/cm}^3$, where we assume that $N_p \sim N_n$, it follows

$$n_e \approx \frac{1}{2}\,\frac{10^6 \mathrm{g/cm}^3}{1.67 \times 10^{-24} \mathrm{g}} \approx 3 \times 10^{29} \mathrm{cm}^{-3} \ll 2 \times 10^{31} \mathrm{cm}^{-3} = (m_e/\hbar c)^3\,. \tag{9.5}$$

Thus we can use the equation for the pressure of a non-relativistic degenerate electron gas,

$$P \approx \frac{\hbar^2 n_e^{5/3}}{m_e} \approx \frac{(1.05 \times 10^{-27} \mathrm{erg\,s})^2}{9.11 \times 10^{-28} \mathrm{g}} \left(\frac{10^6 \mathrm{g/cm}^3}{2 \times 1.67 \times 10^{-24} \mathrm{g}}\right)^{5/3} \approx 10^{23} \mathrm{dyn/cm}^2\,. \tag{9.6}$$

This pressure corresponds with $10^6 \mathrm{dyn/cm}^2 = 1\mathrm{bar}$ to $P = 10^{17}\mathrm{bar}$, and is consistent with our lower limit for central pressure of Sirius B.

9.3 White dwarfs and Chandrasekhar limit

Stars like Sirius B that are supported by the pressure of a degenerated electron gas are called white dwarf stars. They have very long cooling times because of their small surface luminosity. This type of stars is rather numerous: The mass density of man-sequence stars in the solar neighborhood is $0.04 M_\odot/\mathrm{pc}^3$ compared to $0.015 M_\odot/\mathrm{pc}^3$ in white dwarfs. The typical mass of white dwarfs lies in the range $0.4 - 1 M_\odot$, peaking at $0.6 M_\odot$.

If we combine $P_c \propto \frac{GM^2}{R^4}$ and $P = Kn^{5/3} \propto (M/R^3)^{5/3} = M^{5/3}/R^5$, we obtain

$$\frac{GM^2}{R^4} = \frac{KM^{5/3}}{R^5} \qquad \text{or} \qquad R = \frac{M^{\frac{10-12}{6}}}{K} = \frac{1}{KM^{1/3}}\,. \tag{9.7}$$

Thus there exists a unique relation between mass and radius, if we neglect differences due to small variations in their chemical composition. The radius of white dwarf stars decreases for increasing masses, suggesting also that there exists a maximal mass.

In order to derive this maximal mass, we consider now the energy of a star where the pressure is dominated by a non-relativistic degenerate Fermi gas as function of its radius. The total kinetic energy is $U_{\mathrm{kin}} = N\,p^2/(2m_e)$, where $n \sim N/R^3$ and $p \sim \hbar n^{1/3}$. Thus

$$U_{\mathrm{kin}} \sim N\,\frac{\hbar^2 n^{2/3}}{2m} \sim \frac{\hbar^2\,N^{(3+2)/3}}{2mR^2} = \frac{\hbar^2\,N^{5/3}}{2mR^2}\,. \tag{9.8}$$

For the potential gravitational energy we use again the approximation $U_{\mathrm{pot}} = \alpha GM^2/R$ with $\alpha = 1$. Hence

$$U(R) = U_{\mathrm{kin}} + U_{\mathrm{pot}} \sim \frac{\hbar^2\,N^{5/3}}{2mR^2} - \frac{GM^2}{R}\,. \tag{9.9}$$

For small R, the positive term dominates and therefore a stable minimum R_{min} exists for each M.

If the Fermi gas in the star becomes however relativistic, then $U_{\rm kin} = N\,cp$, or

$$U_{\rm kin} \sim N\,c\hbar n^{1/3} \sim \frac{c\hbar\,N^{4/3}}{R} \tag{9.10}$$

and

$$U(R) = U_{\rm kin} + U_{\rm pot} \sim \frac{c\hbar\,N^{4/3}}{R} - \frac{GM^2}{R}\,. \tag{9.11}$$

Now both terms scale like $1/R$. For a fixed chemical composition, $N/M = \text{const.}$, the negative term increases faster than the first one, if M is increased. Hence, it exists a critical M so that U becomes negative and can be made arbitrary small by decreasing the radius of the star: The star collapses.

This critical mass is called Chandrasekhar mass $M_{\rm Ch}$ and is obtained by solving Eq. (9.11) for $U = 0$ and using $M = N_N m_N$. Then $c\hbar N_{\rm max}^{4/3} = G N_{\rm max}^2 m_N^2$, or

$$N_{\rm max} \sim \left(\frac{c\hbar}{Gm_p^2}\right)^{3/2} \sim \left(\frac{M_{\rm Pl}}{m_p}\right)^3 \sim 2\times 10^{57} \tag{9.12}$$

and

$$M_{\rm Ch} = N_{\rm max} m_p \sim 1.5 M_\odot\,. \tag{9.13}$$

The critical size can be determined from the condition the gas becomes relativistic, $U_{\rm kin} \lesssim Nmc^2$ together with $N = N_{\rm max}$. Thus

$$N_{\rm max} mc^2 \gtrsim \frac{c\hbar\,N_{\rm max}^{4/3}}{R} \tag{9.14}$$

and finally

$$mc^2 \lesssim \frac{c\hbar}{R}\,\left(\frac{c\hbar}{Gm_N^2}\right)^{1/2} \tag{9.15}$$

or

$$R \gtrsim \frac{\hbar}{mc}\left(\frac{c\hbar}{Gm_N^2}\right)^{1/2}\,. \tag{9.16}$$

We have two options: We can

1. set $m = m_e$ and obtain $R \gtrsim 5\times 10^8$cm. This corresponds to the radii found for white dwarf stars.
2. set $m = m_N$ and obtain $R \gtrsim 3\times 10^5$cm. Since already Sirius B was difficult to detect, the question arises if and how these extremely small stars can be observed. Moreover, which factors decide if a star ends as a white dwarf or as a neutron star?

Note that $M_{\rm Ch}$ is for both cases the same, since the stellar mass is in both cases given by the sum of the nucleon masses – only the main source of pressure (electron or neutrons) differs.

9.4 Supernovae

Novae and supernovae were characterized empirically according to their luminosity and their spectral lines. Novae show a smaller luminosity increase than supernovae with peak luminosities between 10 and 10^6 times their average luminosity. They are recursive events with periods in the range $P \sim 1\,\text{h}$–$10^5\,\text{yr}$. By contrast, supernova are singular events. Supernovae are further divided into type II and type I, depending on the presence of Balmer lines in their spectrum, respectively.

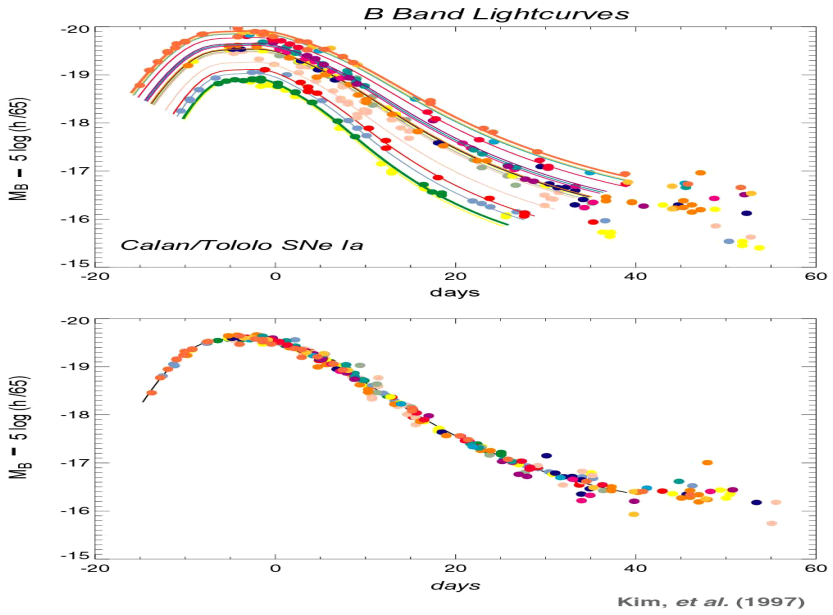


Figure 9.2: Lightcurves of type Ia supernovae; top observed ones, bottom after rescaling.

Novae and type Ia supernovae The physical interpretation of novae and supernovae does not follow precisely this division: Novae and a subset of all type I supernovae called type Ia are variations of accreting binary systems. In the case of a nova, the accreted material starts the CNO cycle in a thin layer on the stellar surface when the pressure is large enough, expelling the outermost shell of the star, but not destroying the star. In a type Ia SN, a white dwarf is driven by accretion beyond M_{Ch} and explodes. More precisely, one finds the following picture from numerical simulations:

- For $M \to 1.44 M_\odot$, carbon in a white dwarf core starts to fuse into Fe, Co, and Ni. The burning expands outwards from the core, synthesizing lighter elements such as Mg, Si, Su.
- If the burning front reaches around $0.7R$, the pressure of the outer layers is not sufficient to confine the burning process. The white dwarf star explodes.
- The main part of optical luminosity comes from the decay of nickel-56, with a half-life of 6.1 days to Cobalt-56 ($\tau = 77.1$d) to iron-56. The decays produce gamma-rays, heating the ejected material.

Since the white dwarf stars explode crossing the Chandrasekhar limit, $M \gtrsim 1.4 M_\odot$, the released total energy should vary not to much. Thus one may wonder if they are possible standard candles? This requires that the spread in luminosity, $L = \mathrm{d}E/\mathrm{d}t$, is not big. As the upper panel of Fig. 9.2, shows more luminous supernovae evolve slower. If this effect is taken into account, as done in the lower panel, the spread in luminosity is small enough to use these events as standard candles.

Core collapse or type-II SN Type II or core collapse supernovae occur at the end of the fusion process in very massive stars, $M \gtrsim (5-8) M_\odot$. Theses stars have an onion-like structure with a degenerate iron core. When the core is completely fused to iron, no further processes

releasing energy are possible. Instead, photo-disintegration destroys the heavy nuclei, e.g. via $\gamma + {}^{56}\mathrm{Fe} \rightarrow {}^{4}\mathrm{He} + 4n$, and removes the thermal energy necessary to provide pressure support and the star collapses.

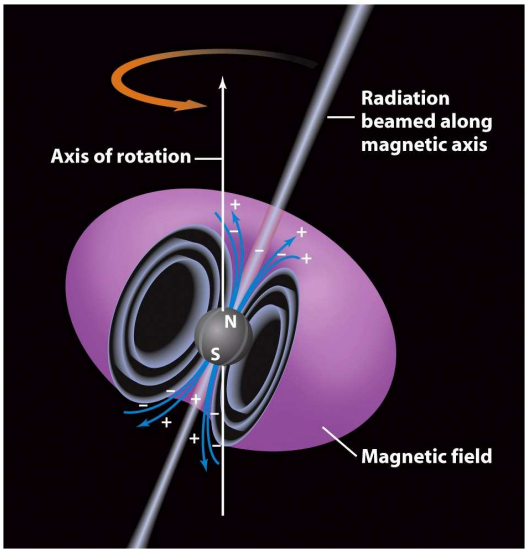
During the collapse of the core, the density increases and the free electrons are forced together with protons to form neutrons via inverse beta decay, $e^- + p \rightarrow n + \nu_e$. A proto-neutron star is formed. When core density reaches nuclear density, the equation of state stiffens suddenly and the infalling material is "reflected".

If the supernova is successful, a neutron star is left over. Otherwise a black hole is formed. The released energy goes mainly into neutrinos (99%), kinetic energy (1%); only 0.01% into photons.

9.5 Pulsars

Generalities of compact stars White dwarf and neutron stars have in common that their radius is strongly increased with respect to a main sequence star, $R_{\mathrm{WD}}/R_\odot \sim 10^{-2}$ and $R_{\mathrm{NS}}/R_\odot \sim 10^{-5}$. Using two conservation laws involving the stellar radius, we can derive immediately two important properties of these compact stars:

- Conservation of angular momentum, $L = I\omega = \text{const.}$ with $I = \alpha M R^2$, implies (initially) fast rotation of a compact star: A star like the sun, rotating once per month, $P \approx 10^6$s, would rotate with a ms period when contracted down to 10 km in size.
- The magnetic flux, $\phi_B = BA = \text{const.}$, is a conserved quantity for an ideal conductor, where magnetic field lines are frozen in the plasma. As the core collapses, the magnetic field lines are pulled more closely together, intensifying the magnetic field by a factor $(R_\odot/R_{\mathrm{NS}})^2 \sim 10^{10}$. Magnetic A stars have surface fields up to 10^4 G, observed field strengths of white dwarfs and of neutron stars are of the order of $B \approx (10^6 - 10^8)$ G and $B \approx (10^{12} - 10^{12})$ G, respectively.



Discovery and identification of pulsars In 1967, Hewish and collaborators detected an object emitting a radio signal with period $P = 1.377$s. They called the object "Pulsar". Only one year later, Gold argued that pulsars are rotating neutron stars. He predicted an increase of the period as pulsars, because they should loose energy via electromagnetic radiation. The slow-down of the Crab pulsar was indeed discovered in 1969.

The key observations that helped to identify pulsars with rotating neutron stars were:

1. The smallness of the observed periods, $P \sim 1ms - 5s$.
2. The extreme stability of the periods, $\Delta P/P = 10^{-13}$.
3. The periods increase slowly.

Additional information from later observations include that the source distribution is peaked in the Galactic

plane, implying a Galactic origin. Moreover, several pulsars like the Crab and Vela pulsars were found in remnants of type II supernovae.

Since the light travel time ct is 500 km for 1.5 ms, we can use this distance as an upper limit for the emission region (that has to be causally connected to act as good clock). This implies that a pulsar has to be a compact object and we have to choose between white dwarfs, neutron stars or black holes as possible sites for the radio emission. Possible clock mechanisms are rotation or pulsation of a single star or the orbital movement of a binary system. We now consider the different options:

Black holes are surrounded by an event horizon. Radiation is emitted by the accretion disk surrounding a black hole. However, both accretion and radiation is very irregular process, excluding the surrounding of a black holes as source of a very regular pulsed radio signal.

The shortest period for a rotating system corresponds to rotation at the break-up velocity,

$$R\omega^2 = \frac{GM}{R^2} \tag{9.17}$$

($v = \omega R$) or in terms of the mean density $\rho \sim M/R^3$

$$\omega^2 = G\rho\,. \tag{9.18}$$

Using as extreme value for the density of a white dwarf $\rho = 10^8 \mathrm{g/cm^3}$ yields $P = 2\pi/\omega \gtrsim 1\,\mathrm{s}$. This excludes rotating white dwarfs.

What about a pulsating white dwarf? The estimate $\omega^2 = G\rho$ holds also for the fundamental frequency of pulsations. Higher harmonics as explanation are excluded because a superposition of frequencies would destroy the observed periodicity. Moreover, energy losses lead generally to an increase of the period in case of oscillations. In case of a binary white dwarf systems we can replace the radius R of the star by the distance $d \gg R$ of the system. Thus the resulting periods are certainly even larger.

Finally, we consider the last option left over, neutron stars. If they are pulsating, the fundamental frequency would be with $\rho_{\mathrm{NS}} \sim 10^6 \rho_{\mathrm{WD}}$ typically too short, $P \sim 10^{-3}\,\mathrm{ms}$. A binary system of neutron stars is a copious emitter of gravitational waves as we will discuss in the next chapter. For such small distances as required by the ms periods, the life-time of these systems would be too short. In summary, the only remains option is a rotating neutron star.

Rotating dipole model A simple model for a pulsar is depicted in the figure above as a neutron star with a non-aligned rotation and magnetic field axes. Physically, on can model a pulsar as a rotating sphere endowed with a magnetic dipole moment. The energy of a rotating sphere is $E = 1/2 I\omega^2$ and the energy change due to a change in the rotation velocity ω is

$$\dot{E} = I\omega\dot{\omega}\,. \tag{9.19}$$

The energy loss of a rotating magnetic dipole due to the emission of electromagnetic radiation is

$$\dot{E} = -\frac{B^2 R^6 \omega^4 \sin^2\alpha}{6c^3}\,, \tag{9.20}$$

where α denotes the angle between rotation and dipole axis. If we define as time-scale τ characterizing the slow-down of a pulsar,

$$\tau = -\frac{\omega}{\dot{\omega}} = -\frac{I\omega^2}{\dot{E}} = \frac{6Ic^3}{B^2 R^6 \omega^2 \sin^2\alpha}\,, \tag{9.21}$$

then we can compare the estimated life-time, e.g. of the Crab nebula, $\tau \sim 1243\,\mathrm{yr}$ with its true age, $1972–1054 = 918\,\mathrm{yr}$. Also, the derived energy output $\dot{E} \sim 6 \times 10^{38}\mathrm{erg}$ is comparable to one derived from observations.

Using pulsars as tool

- Dispersion measure and Galactic electron density:
  The velocity v of electromagnetic waves in a medium is different from c, $n = v/c \neq 1$. The refractive index n is a function of the wave-length λ and thus there is a dispersion between the arrival times of a pulse at different wave-lengths,

$$\Delta t = \frac{L}{c}(n(\lambda_1) - n(\lambda_2))\,. \tag{9.22}$$

  The deviation of the refractive index n from one is proportional to the number density of electrons, $\Delta n \propto n_e$. (Since n_e is not constant, one should integrate along the line of sight.) Thus the time-delay Δt measures the integrated number density of electron along the line of sight to a pulsar. This was used originally as a check that pulsars are indeed Galactic objects. If the distance to a specific pulsar is known, one can the method as a tool to measure the electron density n_e in the Galaxy.
- The Faraday effect is used as measure for the Galactic magnetic field.
- Binary systems of pulsars have been used as test for the emission of gravitational waves as predicted by general relativity, as we will discuss in the next chapter.

10 Black holes

10.1 Basic properties of gravitation

1. In classical mechanics, the equality of gravitating mass $m_g = F/g$ and inertial mass $m_i = F/a$ is a puzzle noticed already by Newton. Knowing more forces, this puzzle becomes even stronger. Contrast the acceleration in a gravitational field to the one in a Coulomb field: In the latter, two independent properties, namely the charge q giving the strength of the electric force and the mass m_i, the inertia of the particle, are needed. The equivalence of gravitating and inertial mass has been tested starting from Bessel, comparing e.g. the period of a pendulum of different materials,

$$P = 2\pi\sqrt{\frac{m_i l}{m_g g}}\,. \tag{10.1}$$

   While $m_i = m_g$ can be achieved for one material by a convenient choice of units, there should be deviations for test bodies with other composition. Current limits are $\Delta a_i/a < 10^{-12}$.

2. Newton's law postulates as the Coulomb law instantaneous an interaction. This is in contradiction to special relativity. Thus, as interactions with electromagnetic fields replace the Coulomb law, a corresponding description should be found for gravity. Moreover, the equivalence of mass and energy found in special relativity requires that, in a loose sense, energy not only mass should couple to gravity: Imagine a particle-antiparticle pair falling down a gravitational potential wall, gaining energy and finally annihilating into two photons moving the gravitational potential wall outwards. If the two photons would not loose energy climbing up the gravitational potential wall, a perpetum mobile could be constructed. If all forms of energy act as sources of gravity, then the gravitational field itself is gravitating. Thus the theory is non-linear and its mathematical structure much more complicated as Maxwell's equations.

3. Gravity can be switched-off locally, just by cutting the rope of an elevator. Inside a freely falling elevator, one does not feel the effect of gravity.

Motivated by 2., Einstein used 1., the principle of equivalence, and 3. to derive general relativity, a theory that describes the effect of gravity as deformation of the space-time known from special relativity.

10.2 Schwarzschild metric

Because Einstein's theory has a rather more complicated mathematical structure than Newton's, no analytical solution to the two-body problem is known. Instead, we are looking for the effect of a finite point-mass on the surrounding space-time.

10.2.1 Heuristic derivation

Consider a freely falling elevator in the gravitational field of a radial-symmetric mass distribution with total mass M. Since the elevator is freely falling, no effects of gravity are felt inside and the space-time coordinates from $r = \infty$ should be valid inside. Let us call these coordinates K_∞ with x_∞ (parallel to movement), y_∞, z_∞ (transversal) and t_∞. The elevator has the velocity v at the distance r from the mass M , measured in the coordinate system $K = (r, \phi, \vartheta, t)$ in which the mass M is at $r = 0$ at rest.

We assume that special relativity can be used for the transformation between K at rest and K_∞ moving with $v = \beta c$, as long as the gravitational field is weak. We shall see shortly what "weak" means in this context. For the moment, we presume that the effect of gravity are small, if the velocity of the elevator that was at rest at $r = \infty$ is still small, $v \ll c$. Then

$$\begin{aligned} \mathrm{d}t_\infty &= \mathrm{d}t\sqrt{1-\beta^2} && (10.2)\\ \mathrm{d}x_\infty &= \frac{\mathrm{d}r}{\sqrt{1-\beta^2}} && (10.3)\\ \mathrm{d}y_\infty &= r\mathrm{d}\vartheta && (10.4)\\ \mathrm{d}z_\infty &= r\sin\vartheta\mathrm{d}\phi\,. && (10.5) \end{aligned}$$

Thus the line-element of special relativity, i.e. the infinitesimal distance between two space-time events,

$$\mathrm{d}s^2 = \mathrm{d}x_\infty^2 + \mathrm{d}y_\infty^2 + \mathrm{d}z_\infty^2 - c^2\mathrm{d}t_\infty^2 \tag{10.6}$$

becomes

$$\mathrm{d}s^2 = \frac{\mathrm{d}r^2}{1-\beta^2} + r^2(\mathrm{d}\vartheta^2 + \sin^2\vartheta\mathrm{d}\phi^2) - (1-\beta^2)\mathrm{d}t^2\,. \tag{10.7}$$

Next, we want to relate the factor $1-\beta^2$ to the quantities M and r. Consider the energy of the elevator with rest mass m,

$$(\gamma-1)mc^2 - \frac{G\gamma mM}{r} = 0\,, \tag{10.8}$$

where the first term is the kinetic energy and the second the Newtonian expression for the potential energy. According to 2), we made here the crucial assumption that gravity couples not only to the mass of the elevator but to its total energy. Dividing by γmc^2 gives

$$\left(1-\frac{1}{\gamma}\right) - \frac{GM}{rc^2} = 0\,. \tag{10.9}$$

Remembering the definition $\gamma = 1/\sqrt{1-\beta^2}$ and introducing $\alpha = GM/c^2$, we have

$$\sqrt{1-\beta^2} = 1 - \frac{\alpha}{r} \tag{10.10}$$

or

$$1-\beta^2 = 1 - \frac{2\alpha}{r} + \frac{\alpha^2}{r^2} \approx 1 - \frac{2\alpha}{r}\,. \tag{10.11}$$

In the last step, we neglected the term $(\alpha/r)^2$, since we attempt only an approximation for large distances, where gravity is still weak. Inserting this expression into Eq. (10.7), we obtain the metric describing the gravitational field produced by a radial symmetric mass distribution,

$$\mathrm{d}s^2 = \frac{\mathrm{d}r^2}{1-\frac{2\alpha}{r}} + r^2(\mathrm{d}\vartheta^2 + \sin^2\vartheta\mathrm{d}\phi^2) - c^2\mathrm{d}t^2\left(1-\frac{2\alpha}{r}\right)\,. \tag{10.12}$$

Surprisingly, this agrees with the exact result found by Karl Schwarzschild 1916.

10.2.2 Interpretation and consequences

Gravitational redshift As in special relativity, the line-element $\mathrm{d}s$ determines the time and spatial distance between two space-time events. The time measured by an observer called the proper-time $\mathrm{d}\tau$ is given by $\mathrm{d}\tau = c\mathrm{d}s$. In particular, the time difference between two events at the same point is given by setting $\mathrm{d}x^i = 0$. If we choose two static observers at the position r and r' in the Schwarzschild metric, then we find with $\mathrm{d}r = \mathrm{d}\phi = \mathrm{d}\vartheta = 0$,

$$\frac{\mathrm{d}\tau(r)}{\mathrm{d}\tau(r')} = \frac{\sqrt{g_{00}(r)}\,\mathrm{d}t}{\sqrt{g_{00}(r')}\,\mathrm{d}t} = \sqrt{\frac{g_{00}(r)}{g_{00}(r')}}\,. \tag{10.13}$$

The time-intervals $\mathrm{d}\tau(r')$ and $\mathrm{d}\tau(r)$ are different and thus the time measured by clocks at different distances r from the mass M will differ too. In particular, the time measured τ_∞ by an observer at infinity will pass faster than the time experienced in a gravitational field,

$$\tau_\infty = \frac{\tau(r)}{\sqrt{1-2\alpha/r}} < \tau(r)\,. \tag{10.14}$$

Since frequencies are inversely proportional to time, the frequency or energy of a photon traveling from r to r' will be affected by the gravitational field as

$$\frac{\nu(r')}{\nu(r)} = \sqrt{\frac{1-2\alpha/r}{1-2\alpha/r'}}\,. \tag{10.15}$$

Hence an observer at $r' \to \infty$ will receive photons with frequency

$$\nu_\infty = \sqrt{1-\frac{2GM}{rc^2}}\nu(r) \tag{10.16}$$

if these photons were emitted by a source at position r with frequency $\nu(r)$: The frequency of a photon is redshifted by a gravitational field, in agreement with our argument against a perpetuum mobile in 2).

The size of this effect is of order Φ/c^2, where $\Phi = -GM/r$ is the Newtonian gravitational potential. We are now in position to specify more precisely what "weak gravitational fields" means: As long as $|\Phi|/c^2 \ll 1$, the deviation of $g_{tt} = 1 - \frac{2GM}{rc^2} \approx 1 - 2\Phi(r)/c^2$ from the Minkowski value $g_{tt} = 1$ is small and Newtonian gravity is a sufficient approximation.

Schwarzschild radius What is the meaning of $r = R_S \equiv 2\alpha$? At

$$R_S = \frac{2GM}{c^2} = 3\,\mathrm{km}\,\frac{M}{M_\odot} \tag{10.17}$$

the coordinate system (10.12) becomes ill-defined. However, this does not mean necessarily that at $r = R_S$ physical quantities like tidal forces become infinite. Instead $r = R_S$ is an event horizon: We cannot obtain any information about what is going on inside R_S, if the gravitating mass is concentrated inside a radius smaller than R_s[1]. An object smaller than its Schwarzschild radius is called a black hole. The black hole is fully characterized by its mass M (and possibly its angular momentum L and electric charge q). To understand this better, we consider next what happens to a photon crossing the event horizon at $r = R_S$ as seen from an observer at $r = \infty$.

[1] Recall that in Newtonian gravity only the enclosed mass $M(r)$ contributes to the gravitational potential outside r for a spherically symmetric system. Thus, e.g. the Sun is not a black hole, since for all r the enclosed mass is $M(r) < rc^2/2G$.

Approaching a black hole Light rays are characterized by $\mathrm{d}s^2 = 0$. Choosing a light ray in radial direction with $\mathrm{d}\phi = \mathrm{d}\vartheta = 0$, the metric (10.12) becomes

$$\frac{\mathrm{d}r}{\mathrm{d}t} = \left(1 - \frac{2\alpha}{r}\right) c\,. \tag{10.18}$$

Thus light travelling towards the star, as seen from the outside, will travel slower and slower as it comes closer to the Schwarzschild radius $r = 2\alpha$. In fact, for an observer at infinity the signal will reach $r = 2\alpha$ only asymptotically for $t \to \infty$. Similarly, the communication with a freely falling space ship becomes impossible as it reaches $r = R_s$. A more detailed analysis shows that indeed no signal can cross the surface at $r = R_s$.

Perihelion precession Ellipses are solutions only for a potential $V(r) \propto 1/r$. General relativity generate corrections to the Newtonian $1/r$ potential, and as a result the perihelion of ellipses describing the motion of e.g. planets process. This effect is largest for Mercury, where $\Delta\phi/\Delta t \approx 43''/\mathrm{yr}$. This was the main known discrepancy of the planetary motions in the solar system with Newtonian gravity at the time, when Einstein and others worked on relativistic theories of gravity.

Light bending The factors $(1 - 2\alpha/r)$ in the line-element $\mathrm{d}s$ will lead to the bending of light in gravitational fields. The measurement of the deflection of light by the Sun during the solar eclipse 1919 was the first crucial test for general relativity. Nowadays, one distinguishes three different cases of gravitational lensing, depending on the strength of the lensing effect:

1. Strong lensing occurs when the lens is very massive and the source is close to it: In this case light can take different paths to the observer and more than one image of the source will appear, either as multiple images or deformed arcs of a source. In the extreme case that a point-like source, lens and observer are aligned the image forms an "Einstein ring".
2. Weak Lensing: In many cases the lens is not strong enough to form multiple images or arcs. However, the source can still be distorted and its image may be both stretched (shear) and magnified (convergence). If all sources were well known in size and shape, one could just use the shear and convergence to deduce the properties of the lens.
3. Microlensing: One observes only the usual point-like image of the source. However, the additional light bent towards the observer leads to brightening of the source. Thus microlensing is only observable as a transient phenomenon, when the lens crosses approximately the axis observer-source.

10.3 Gravitational radiation from pulsars

An accelerated system of electric charges emits dipole radiation with luminosity

$$L_{\mathrm{em}} = \frac{2}{3c^3}|\ddot{\mathbf{d}}|^2\,, \tag{10.19}$$

where the dipole moment of a system of N charges at position $\mathbf{x}_i$ is $\mathbf{d} = \sum_{i=1}^{N} q_i \mathbf{x}_i$. One might guess that for the emission of gravitational radiation the replacement $q_i \to G m_i$ works.

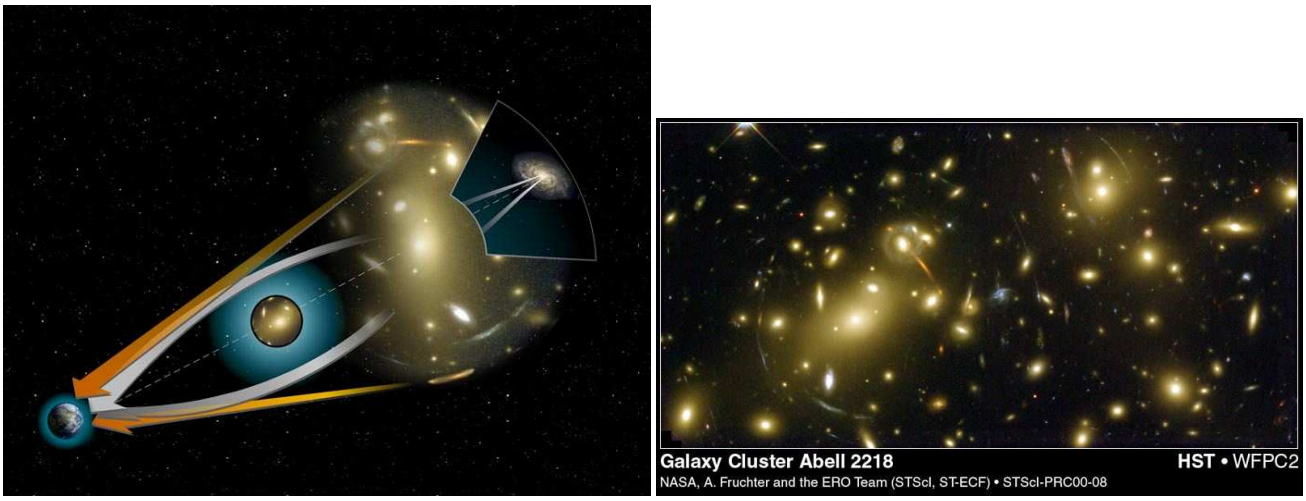


Figure 10.1: Left: Strong gravitational lensing of. Right: Weak gravitational lensing of Abell2218.

But since $\sum m_i \mathbf{x}_i = \mathbf{p}_{\rm tot} = \text{const.}$, momentum conservation means that there exists no gravitational dipole radiation. Thus one has to go to the next term in the multipole expansion, the quadrupole term,

$$Q_{ij} = \sum_{k=1}^{N} m^{(k)}(x_i x_j - \frac{1}{3}\delta_{ij} r^2)\,, \tag{10.20}$$

and finds then for the luminosity emitted into gravitational waves

$$L_{\rm gr} = \frac{G}{5c^5} \sum_{i,j} |\dddot{Q}_{ij}|^2\,. \tag{10.21}$$

Ex.: Derive $L_{\rm em} \propto d^2/c^3$ and $L_{\rm gr} \propto GQ^2/c^5$ using dimensional analysis.
In the case of electromagnetic dipole radiation, we use that $e^2 = \hbar c\alpha$ and the dimension of the electric charge squared is $[e^2] = \text{erg} \times \text{cm}$. Then we compare

$$[L_{\rm em}] = \frac{\text{erg}}{\text{s}} \stackrel{!}{=} \text{erg} \times \text{cm} \times \frac{\text{s}^3}{\text{cm}^3} \times \frac{\text{cm}^2}{\text{s}^4} = \frac{\text{erg}}{\text{s}}\,.$$

In the same way, we can show that the luminosity of gravitational quadrupole radiation is proportional to c^a with $a = -5$: The unique solution for a from

$$[L_{\rm gr}] = \frac{\text{erg}}{\text{s}} = \frac{\text{g cm}^2}{\text{s}^3} \stackrel{!}{=} \frac{\text{cm}^3}{\text{g s}^2} \times \frac{\text{cm}^{\rm a}}{\text{s}^{\rm a}} \times \frac{\text{g}^2\,\text{cm}^4}{s^6}$$

is $a = -5$.

The emission of gravitational radiation is negligible for all systems where Newtonian gravity is a good approximation. One of the rare examples where general relativistic effects can become important are close binary systems of compact stars. The first such example was found 1974 by Hulse and Taylor who discovered a pulsar in a binary system via the Doppler-shift of its radio pulses. The extreme precision of the periodicity of the pulsar signal makes this binary system to an ideal laboratory to test various effect of special and general relativity:

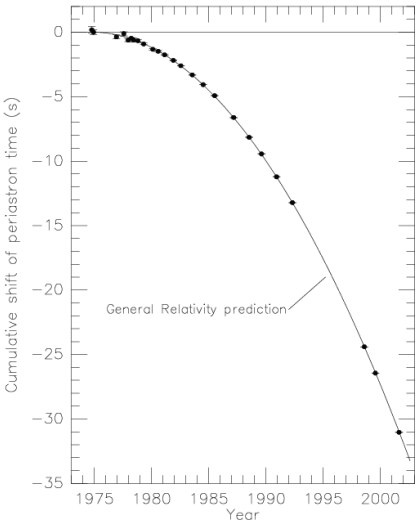


Figure 10.2: Measured change of the period of the orbital movement for the Hulse-Taylor binary compared to the prediction of general relativity.

- The pulsar's orbital speed changes by a factor of four during its orbit and allows us to test the usual Doppler effect.
- At the same time, the gravitational field alternately strengthens at periastron and weakens at apastron, leading to variable gravitational redshift of the pulse.
- The small size of the orbit leads to a precession of the Perihelion by 4.2°/yr.
- The system emits gravitational waves and looses thereby energy. As a result the orbit of the binary shrinks by 4mm/yr, in excellent agreement with the prediction of general relativity, cf. Fig. 10.2.

10.4 ∗∗∗ Thermodynamics and evaporation of black holes ∗∗∗

Black holes thermodynamics Classically, the mass and therefore the radius of a black hole can only increase with time. The only other quantity in physics with the same property is the entropy, $\mathrm{d}S \geq 0$. This suggests a connection between a quantity characterizing the size of the black hole and its entropy. To derive this relation, we apply the first law of thermodynamics $\mathrm{d}U = T\mathrm{d}S$ to a black hole. Its internal energy U is given by $U = Mc^2$ and thus

$$\mathrm{d}U = \mathrm{d}Mc^2 \stackrel{!}{=} T\mathrm{d}S\,. \tag{10.22}$$

A Schwarschild black hole is fully characterised by only one quantity, e.g. R_s. Hence both the temperature T and the entropy S of a black hole should be determined solely by R_s.

Our experience with the thermodynamics of non-gravitating systems suggests that the entropy is an extensive quantity and thus proportional to the volume, $S \propto V$. In this case, $\mathrm{d}V \propto R_s^2\mathrm{d}R_s \propto R_s^2\mathrm{d}M$ and $\mathrm{d}M \propto R_s^{-2}\mathrm{d}V \propto T^2\mathrm{d}V$. On the other hand, if we assume that the entropy is proportional to the area of the event horizon, $\mathrm{d}A \propto R_s\mathrm{d}R_s \propto R_s\mathrm{d}M$ and $\mathrm{d}M \propto R_s^{-1}\mathrm{d}A \propto T\mathrm{d}V$. Surprisingly, several arguments show independently that the second alternative is correct; one we will sketch below.

Thus

$$\mathrm{d}U = T\mathrm{d}S = dMc^2 = \frac{c^4}{16\pi G R_s}\mathrm{d}A = T\mathrm{d}\left(\frac{kc^3 A}{4\hbar G}\right) \tag{10.23}$$

and, as first suggested by Bekenstein, the entropy of a black hole is given by

$$S = \frac{kc^3}{4\hbar G}\, A\,. \tag{10.24}$$

If this analogue is more than a formal coincidence, a black hole should behave similar as other thermodynamical systems we know. In particular, if its surrounded by a cooler medium, it should emit radiation and heat up the environment. Hawking could indeed show 1974 that a black hole in vacuum emits black-body radiation ("Hawking radiation") with temperature

$$kT = \frac{\hbar\, c^3}{8\pi G M}\,. \tag{10.25}$$

Evaporation of Black Holes The radiation of a BH is a quantum process, as indicated by the presence of Planck's constant $\hbar$ in Eq. (10.25). We can understand the basic mechanism and the magnitude of the black hole temperature by considering a quantum fluctuation near its event horizon at R_S: Heisenberg's uncertainty principle allows the creation of a virtual electron-positron pair with total energy E for the time $\Delta t \lesssim \hbar/E$. This pair can become real, if the gravitational force acting on them during the time Δt can supply the energy E. More precisely, the work is done by the tidal force

$$\mathrm{d}F = \frac{\partial F}{\partial r}\,\mathrm{d}r = -\frac{GME/c^2}{2r^3}\,\mathrm{d}r\,, \tag{10.26}$$

because the electron-positron pair is freely falling.

After the time Δt, the pair is separated by the distance $\Delta r = c\Delta t \sim c\hbar/E$. The energy gain by the tidal force is

$$F\Delta r \sim (E/c^2)GM/r^3(\Delta r)^2\,. \tag{10.27}$$

If we require that $F\Delta r > E$ and identify the energy E of the emitted pair with the temperature T of the black hole, we find

$$kT \sim E \lesssim \hbar(GM/r^3)^{1/2}\,. \tag{10.28}$$

The emission probability is maximal close to the event horizon, $r = R_S$, and thus we obtain indeed $kT \sim E \sim \hbar c^3/(GM)$. This result supports the idea that the entropy of a black hole is indeed non-extensive and proportional to the surface of its event horizon. Since entropy counts the possible micro-states of a system, any promising attempt to combine gravity and quantum mechanics should provide a microscopic picture for these results.

Exercises

1. Assume that the surface $A = 4\pi R_S^2$ of a Black Hole is emitting black-body radiation with temperature $kT = \hbar c/(4\pi R_S)$, where R_S is its Scharzschild radius, $R_S = 2GM/c^2$.
   i) Derive the luminosity L of the Black Hole.
   ii) Derive with $E = Mc^2$ and $-\mathrm{d}E/\mathrm{d}t = L$ the lifetime of a Black Hole with initial mass M.
   iii) Find the temperature and the luminosity of a black hole with $M = M_\odot$.

2. Estimate the total entropy of the observable universe of radius 4 Gpc, comparing the entropy of the cosmic microwave radiation ($T = 2.7\,\mathrm{K}$ and $S = (4\pi^2/45)\,T^3V$) with the entropy of all black holes assuming that each galaxy occupying the volume $10\,\mathrm{Mpc}^3$ contains a supermassive black hole with mass $M = 10^6 M_\odot$.

3. The Global Positioning System (GPS) consists of 24 satellites carrying with them atomic clocks at an altitude of about 20,000 km from the ground. A GPS receiver determines its current position by comparing the time signals it receives from a number of the GPS satellites (usually 4 or more) and triangulating on the known positions of each satellite. The attempted precision is below 1 m. Estimate special and general relativistic effects that have to be accounted for to achieve this precision.

Part II

Galaxies

11 Interstellar medium and star formation

11.1 Interstellar dust

Dust becomes visible by its blocking effect of star light. The combined effect of scattering and absorption of light, mainly by dust grains, is called extinction. Clearly, extinction may affects the properties like the luminosity or the distance of a star or galaxy that we deduce from its observed spectra. Surprisingly, this possibility has been ignored until the 1930s by astronomers.

R.J. Trümpler found the first clear evidence for extinction by dust examining cluster of stars in our galaxy. He plotted the angular diameter D of these star clusters versus their distance r. For the derivation of their distance he assumed $\mathcal{F} \propto 1/r^2$. The distribution $n(D/r, r)$ should not depend on r, but he found a systematic increase of the linear size of clusters with distance. There are only two interpretations possible: i) The Sun is at a special place in the galaxy where the size of star clusters has a minimum. ii) Some light is absorbed and thus the energy flux decreases faster than $1/r^2$.

If the dust between us and the source has the optical depth τ, then the observed intensities with (I) and without (I_0) absorption are connected by $I = I_0 \exp(-\tau)$. We can relate the optical depth τ to the extinction A measured in magnitudes via

$$A \equiv m_0 - m = 2.5 \log(I_0/I) \tag{11.1}$$

or

$$A = 2.5 \log(\exp(-\tau)) = 2.5\tau \log(e) \approx 1.09\tau \,. \tag{11.2}$$

Thus the extinction A equals approximately the optical depth.

If the spectral type and distance of a star is known, then the extinction between us and the star can be determined: The spectral type determines the absolute magnitude M, while the apparent magnitude m is observed directly. Then

$$m = M + 5 \log(r/(10\text{pc})) + A \,, \tag{11.3}$$

and we can find A for known m, M and r. On the other hand, distance measurements using spectroscopic parallaxes or standard candles like Cepheides or Supernovae of type Ia are affected by extinction.

Reddening and extinction curves The scattering cross section of light on dust is wavelength dependent. In general, light with shorter wavelengths is more efficiently scattered. In classical electrodynamics, one can roughly distinguish three scattering regimes,

$$\lambda \ll R: \quad \sigma \sim \pi R^2 \qquad \text{geometrical,} \tag{11.4}$$

$$\lambda \sim R: \quad \sigma \propto \lambda^{-1} \qquad \text{Mie,} \tag{11.5}$$

$$\lambda \gg R: \quad \sigma \propto \lambda^{-4} \qquad \text{Rayleigh.} \tag{11.6}$$

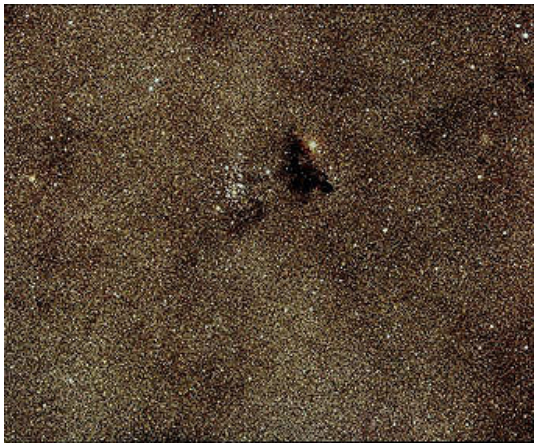


Figure 11.1: A dark cloud, Barnard 86, is silhouetted against a starry background.

Measuring extinction at different wave-lengths allows astronomers to determine the properties of interstellar dust. One finds roughly $A \propto 1/\lambda$, thus the typical size of dust grains is comparable to the wave-length of visible light. If the wave-length dependence of extinction for "typical dust" is known, the absolute value of the extinction in the spectrum of a certain source can be estimated by observing it at different wave-lengths.

Polarization Light passing dust clouds becomes polarized and the degree of its polarization is found to be proportional to the extinction. Using classical electrodynamics one can show that this requires a non-spherical shape of dust grains. Since the polarization builds up scattering on different dust grains, the dust grains have to be moreover partially aligned. The observed polarization is most likely explained, if dust grains have a magnetic moment and become aligned by magnetic fields.

Composition and sources The absorption features visible in the IR of extinction curves correspond to vibrational transitions of silicates and ice, while extinction features in the UV can be connected to the presence of carbon. This leads to the idea that dust grains are mainly "dirty ice balls" containing some graphite. The composition suggests that dust originates from massive stars: In their red giant phase they loose part of their envelope; silicon and carbon condenses cooling down, and a layer of hydrogen is bound around the silicon and graphite core.

11.2 Interstellar gas

Interstellar gas was first detected by its absorption of background light. Since gas clouds are much cooler than the surface of stars, their absorption lines are narrower than those of stars, have smaller excitation energies, and can thus be distinguished from stellar absorption lines.

One expects that gas clouds consist mainly of hydrogen. Since they are cool, hydrogen is in the ground state. Therefore the main absorption feature is the Lyman-α line that is in the UV and thus not observable with a ground telescope. Instead, early studies concentrated on absorption lines of simple molecules like CH and CN and of atoms like Ca and So.

Nowadays, one observes gas clouds mainly with radio observation of the 21cm line of hydrogen. This wavelength corresponds to the temperature $T = ch/k\lambda \approx 0.07\,\mathrm{K}$ and can thus be excited even in cool clouds. The 21cm line is caused by transitions between the two hyperfine levels of the hydrogen 1s ground state. The ground state of hydrogen can be occupied by an electrons with spin up or down. These two stated have the same energy, apart from a small correction due to the interaction of the electron spin with the spin of the proton. The 21cm emission line corresponds to a flip of the electron spin from parallel to anti-parallel to the spin of the proton.

The 21cm emission line is an extremely useful tool to study gas clouds in our and nearby galaxies. Since their movement can be measured using the Doppler effect, they serve as tracers for the velocity distribution of matter in galaxies. Radio observations are not hampered by dust, and thus one can observe gas clouds close to the galactic center and beyond, regions that are obscured to optical observations. The Zeeman effect allows one to measure magnetic fields in the clouds and thereby to trace the galactic magnetic field. In summary, these studies conclude that gas clouds fill $\sim 5\%$ of the galaxy; they have an average density $n_H \sim (1-10)\,\mathrm{cm}^{-3}$.

11.3 Star formation

11.3.1 Jeans length and mass

A cloud of gas will collapse, if the gravitational attraction dominates over thermal pressure. A bound system has negative energy $E = E_{\mathrm{pot}} + E_{\mathrm{kin}} \leq 0$, or

$$\frac{3}{5}\,\frac{GM^2}{R} \geq \frac{3M}{2m}\,kT \qquad \text{or} \qquad \frac{M}{R} \geq \frac{5}{2}\,\frac{kT}{Gm}\,. \tag{11.7}$$

Here we assumed a homogeneous distribution of gas atoms of mass m, obeying the ideal gas law, in a cloud of radius R and total mass M. The smallest mass M (for fixed density ρ and temperature T) which collapses is called Jeans mass $M_J = 4/3\pi R_J^3 \rho$ with Jeans length R_J. Inserting the expression for M_J, we find

$$\frac{4\pi}{3}\,R_J^2\rho = \frac{5}{2}\,\frac{kT}{Gm} \qquad \text{or} \qquad R_J = \left(\frac{15kT}{8\pi G\rho m}\right)^{1/2}\,. \tag{11.8}$$

The acceleration a for a homogeneous cloud is

$$a = \frac{GM(r)}{r^2} = \frac{4\pi}{3}\,G\rho r\,. \tag{11.9}$$

Since the density increases as the cloud collapses, we can derive an upper limit on the time required for a complete collapse, the so called free-fall time τ_{ff}, assuming the density is constant,

$$\tau_{\mathrm{ff}} \sim \frac{1}{(G\rho_0)^{1/2}}\,. \tag{11.10}$$

Since the sound speed of an ideal gas is $v_s^2 = 5kT/(3m)$, the Jeans length can be expressed also as

$$R_J \propto v_s/\sqrt{G\rho}\,. \tag{11.11}$$

This suggests another intuitive interpretation of the Jeans criterion: It compare the timescale of gravitational collapse $\tau_{\rm ff} \sim (G\rho_0)^{-1/2}$ with the one pressure can react, $\tau_p \sim R/v_s$. For $\tau_p \gg \tau_{\rm ff}$, or $R \gg v_s/(G\rho_0)^{1/2}$, gravitational collapse occurs.

The time-scale $\tau_{\rm ff}$ sets the time-scale of growth of density perturbations. However, without detailed analysis we do not know the functional dependence, i.e. if e.g. $\rho(t) \sim \exp(t/\tau_{\rm ff})$ or $\rho(t) \sim (t/\tau_{\rm gr})^\alpha$. The same idea applies to the formation of structures on all scales in the Universe. Since gravitation can enhance only already existing density fluctuation, a mechanism in the early Universe should exist that produced initial density fluctuations.

Density dependence If the initial density of the cloud is ρ_0, then

$$\rho = \rho_0 \left(\frac{r_0}{r}\right)^3 \tag{11.12}$$

or differentiating

$$\mathrm{d}\rho = -\frac{3\rho_0}{r}\left(\frac{r_0}{r}\right)^3 \mathrm{d}r = -\frac{3\rho}{r}\mathrm{d}r \qquad \text{or} \quad \frac{\mathrm{d}\rho}{\rho} = -\frac{3\mathrm{d}r}{r}\,. \tag{11.13}$$

Thus for smaller r, the density increases faster. The free-fall time will increase towards the center, the outer material will lag behind the material closer towards the center, enhancing the difference between core and outer layers.

Rotation The acceleration in radial direction is now the difference between gravitational and centrifugal acceleration,

$$a_r = \frac{GM(r)}{r^2} - r\omega^2\,. \tag{11.14}$$

Let us estimate by how much the cloud collapses until the two forces on the RHS balance each other, $a = 0$.

$$GM(r) = r^3\omega^2 = \underbrace{r^3\omega_0^2}_{r^3 v_0^2/r_0^2}\left(\frac{r_0}{r}\right)^4 = \frac{r_0^2 v_0^2}{r}\,, \tag{11.15}$$

where we used angular momentum conservation $L = I\omega = \text{const.}$ with $I = \alpha M r^2$ and thus $\omega/\omega_0 = (r_0/r)^2$. We now introduce the speed $v_0 = \omega_0 r$ and solve for the fraction r/r_0 the cloud collapses,

$$\frac{r}{r_0} = \frac{r_0 v_0^2}{GM(r)}\,. \tag{11.16}$$

Ex.: Consider the collapse of a cloud with $n_H = 10^5\mathrm{cm}^{-3}$, $T = 50\mathrm{K}$ and initial rotation velocity $v_0 = 1\mathrm{km/s}$.
The Jeans radius is 0.2pc, the Jeans mass $M_J = 76 M_\odot$, and thus $r/r_0 = 0.6$. Thus the cloud should expand for the chosen parameters. Angular momentum conservation is an important obstacle to the formation of dense structures, leading to fragmentation in sub-clouds and to the formation of discs perpendicular to **L**.

11.3.2 Protostars

Luminosity of collapsing clouds Let us apply once again the virial theorem, $-E_{\rm pot} = 2E_{\rm kin}$, for the evolution of a collapsing cloud. Its total energy is $E = -E_{\rm kin} = E_{\rm pot}/2$ and as it collapses $|E_{\rm pot}|$ increases, but only half remains as kinetic energy in the cloud. The other half has to be radiated away.

The total energy $E = -(3/10)GM^2/R$ changes in time as

$$\frac{{\rm d}E}{{\rm d}t} = \frac{3}{10}\frac{GM^2}{R^2}\frac{{\rm d}R}{{\rm d}t} \tag{11.17}$$

or

$$-\frac{1}{E}\frac{{\rm d}E}{{\rm d}t} = \frac{1}{R}\frac{{\rm d}R}{{\rm d}t}\,. \tag{11.18}$$

Thus the relative decrease of the radius is directly connected to the relative energy release of the cloud. If energy cannot be emitted efficiently, the collapse is slowed down.

The formation of a proto-star can be divided into the following stages: Initially, the cloud consists of molecular and atomic hydrogen plus atomic helium. Half of the energy deliberated by the collapse is transferred into the internal energy of the proto-star. In the beginning, the energy is not used to heat the gas, but for its ionization. Only after that, the proto-star heats up and its pressure increases, slowing-down the collapse.

Ex.: For a $1M_\odot$ proto-star with an initial radius $R = 500R_\odot$, find the free-fall time $\tau \approx (G\rho)^{1/2}$, the energy released and the average luminosity in the last 100 years of collapse.
We use our standard approximation of a spherically symmetric and homogeneous gas cloud,

$$E = \frac{3}{10}\,\frac{GM^2}{R} \approx 2\times 10^{45}{\rm erg}\,.$$

The average luminosity is $\langle L\rangle = \Delta E/\Delta t$, or

$$L = \frac{2\times 10^{45}{\rm erg}}{100\times\pi 10^7{\rm s}} \approx 7\times 10^{35}{\rm erg/s} \approx 170L_\odot\,.$$

Once the luminosity reached typical stellar luminosities, the cloud is called proto-star. For increasing density, the interior becomes more and more opaque to radiation, and finally only the surface can radiate. The central temperature is rising until fusion processes start. At same point, the pressure gradient is large enough to stop the collapse and a stable main sequence star is born.

12 Cluster of stars

A look at the night-sky with binoculars or a small telescope shows that stars are not uniformly distributed on the sky. They are concentrated in groups of stars like the Plejades and a thin band across the sky. The thin band with a milky appearance is the disc of our own galaxy, the Milky Way, that we will discuss in the next chapter. In this chapter, we discuss cluster of stars and in particular their evolution.

12.1 Overview

Types and properties There are two main types of star clusters, *galactic* or *open clusters* and *globular clusters*. Figure 12.1 shows in the left panel the open cluster M45, also called Plejades, and in the right one the globular cluster M80.

1. Galactic or open clusters have typically $< 10^3$ stars and a diameter $D \lesssim 10$ pc. They are close to the galactic plane and contain often dust and interstellar gas.
2. Globular clusters consist typically of 10^4–10^6 stars within a diameter of $D \sim 20$–100pc. The globular clusters in our galaxy form a spherical distribution with center around 8 kpc away, cf. Fig. 12.2. Assuming that the center of this distribution coincides with the center of our galaxy allowed Harlow Shapely 1918 to determine for the first time the distance of the Sun to the galactic center.
   Globular clusters contain, if at all, only small amounts of dust and interstellar gas. The fraction of heavier elements of stars in a globular cluster is larger than in stars of a galactic cluster.

Different star populations Because of the continuous fusion in stars of hydrogen to helium and heavier elements, collectively called "metals" in astronomy, the initial chemical composition of stars differs. This effect can be roughly accounted for by dividing stars into two different populations. Population I stars are relatively young stars, which are found mainly in the galactic disc. Population II stars on the other hand are older stars, which are found e.g. in the galactic halo and globular clusters. Typical values for their initial chemical composition are

$$X \approx 0.70 - 0.75, \qquad Y \approx 0.24 - 0.28, \qquad Z \approx 0.005 - 0.04, \qquad \text{Pop I} \tag{12.1}$$
$$X \approx 0.75 - 0.77, \qquad Y \approx 0.23 - 0.25, \qquad Z \approx 10^{-5} - 10^{-4}, \qquad \text{Pop II}, \tag{12.2}$$

where X, Y and Z denote relative mass fractions of hydrogen, helium and heavier elements, respectively. Population III stars are a hypothetical population of extremely massive and hot stars with virtually no metal content which are believed to have been formed in the early universe as the "zeroth" generation of stars. They have not yet been observed directly, but their existence is suggested to account for the fact that "heavy" elements, which could not

Figure 12.1: Left: M45 or the "Pleiades stars", an open cluster of ≈ 500 mostly faint stars spread over a 2 degree (four times the diameter of the Moon) field. Right: The globular cluster M80 containing several 100,000s of stars.

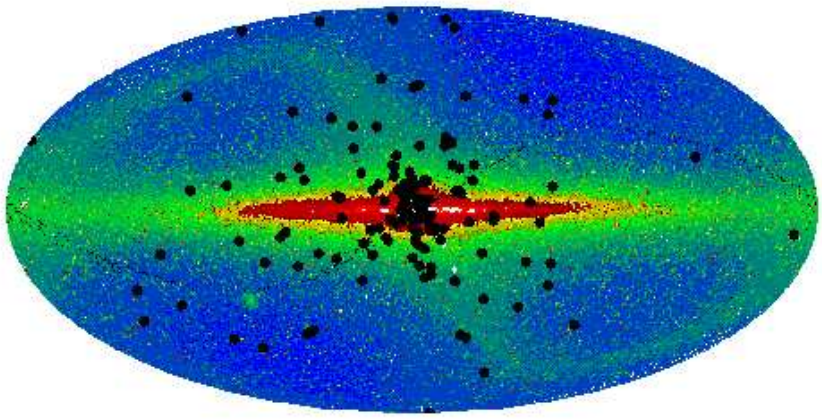


Figure 12.2: Distribution of globular clusters in the Milky Way.

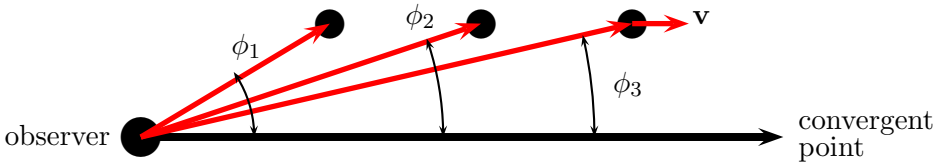


Figure 12.3: The angle ϕ between $\mathbf{v}$ and the line-of-sight approaches zero for $t \to \infty$ and the line-of-sight converges against the "convergent point."

have been created in the "Big Bang", are observed in the emission spectra of very early, powerful galaxies.

A natural explanation for these differences is that star formation started in the whole galaxy. The primordial material from which the first, i.e. Pop. II, stars were formed contained only a tiny contribution of metals. Outside of the galactic disc, most gas has been used up to form this first generation of stars. Now, the density of interstellar gas is too small in the galactic halo and globular clusters, and star formation has stopped there. By contrast, the galactic disc still contains dense regions and star formation continues there. However, the interstellar material contains the ashes of old massive stars, and therefore the metallicity of this second generation of stars (= Pop I) is larger.

Proper motion and distances Some nearby stars move with a sufficiently large velocity such that their motion over the sky becomes visible over the years. The velocity v can be split into its radial and transverse component,

$$v^2 = v_r^2 + v_t^2 = (v \sin\phi)^2 + (v \cos\phi)^2, \qquad \tan\phi = \frac{v_t}{v_r}\,. \tag{12.3}$$

The radial part can be measured via Doppler shift, while the transverse part is responsible for the proper motion $\mu = v_t/D$ of the cluster at the distance D across the celestial sphere. As long as we cannot determine ϕ, we cannot determine v_t and D separately from a single measurement.

A solution is to observe the star (or the cluster of stars) over a longer time period. If the cluster moves away, ϕ decreases and therefore the proper motion becomes slower: the star heads towards a definite point on the celestial sphere, called convergent point. From Fig. 12.3 we see that the angle between the line of sight and the line to the convergent point is ϕ. This method is an important way to establish distances to clusters containing Cepheides.

12.2 Evolution of a globular cluster

The crucial points in our derivation of the virial theorem $2\langle U_{\rm kin}\rangle = -\langle U_{\rm pot}\rangle$ for a star was the assumption of a gravitationally bound system in equilibrium. Thus it holds also for any other system like a cluster of stars or galaxies, if this system fulfills the two conditions to be i) gravitationally bound, and ii) in equilibrium. The first condition can be checked directly by the virial theorem, while the second one needs a more detailed discussion. In particular, we will see that a gravitationally bound system evolves and thus can be only approximately

"in equilibrium." In this context, we define a system as "in equilibrium" or "dynamically relaxed", if the interchange of energy between the members of a cluster is fast compared to the evolution of the cluster.

Rms and escape velocity from the virial theorem: The total kinetic energy of cluster with mass $M = Nm$ is

$$U_{\rm kin} = \frac{1}{2} M \langle v^2 \rangle \,, \tag{12.4}$$

where $v_{\rm rms} \equiv \langle v^2 \rangle^{1/2}$ is the root mean square (rms) velocity. Applying the virial theorem and using $U_{\rm pot} = 3GM^2/(5D)$, it follows

$$\langle v^2 \rangle = \frac{3GM}{5D} \,. \tag{12.5}$$

Ex.: Find the typical rms velocity of stars in a spherical cluster with size $D = 5$ pc that consists of 10^6 stars with average mass $m = 0.5 M_\odot$.

$$\langle v^2 \rangle = \frac{3GM}{5D} = 2.5 \times 10^{12} {\rm cm}^2/{\rm s}^2 \tag{12.6}$$

or $v_{\rm rms} \approx 16$km/s. This value can be compared with observations. If the observed velocities are significantly higher, the cluster cannot be gravitationally bound or its total mass has to be larger.

Crossing time: The crossing time $t_{\rm cr}$ is the typical time required for a star in the cluster to travel the characteristic size D of the cluster (typically taken to be the half-mass radius). Thus, $t_{\rm cr} \sim D/v$ or $t_{\rm cr} \sim 3 \times 10^5$ yr for the values of our example.

Relaxation time: The relaxation time $t_{\rm rel}$ is the typical time in which the star's velocity changed an amount comparable to its original velocity by gravitational encounters. Thus one might think of it as the time-scale after which the velocity distribution of stars bound in a cluster has reached an equilibrium distribution by the exchange energy and momentum with each other. But since part of the stars will have velocities $v > v_{\rm esc}$ and escape, the velocity distribution of a cluster is not stationary: High-velocity stars escape, the cluster contracts and its core is heated-up.

The depth τ for collisions of stars with each-other is

$$\tau = \frac{1}{\sigma n l} = \frac{1}{\sigma n v t} \,, \tag{12.7}$$

when n denotes the density of stars and $l = vt$ the path traveled by a singe star. Let us call the relaxation time $t_{\rm rel}$ the time for which $\tau = 1$. Thus $t_{\rm rel} = 1/(\sigma n v)$.

What should we use for R (as function of v) in $\sigma = \pi R^2$? Stars are certainly gravitationally interacting with each other, if they are bound to each other. Therefore we can estimate the effective interaction range R from $E_{\rm kin} = E_{\rm pot}$, $mv^2/2 = Gm^2/R$ or $R = 2Gm/v^2$. Then

$$t_{\rm rel} = \frac{1}{\pi R^2 n v} = \frac{v^3}{4\pi n (Gm)^2} \,. \tag{12.8}$$

Inserting $1/n = (m/M)(4\pi/3)D^3$ (with $M = Nm$ as total mass of the cluster) gives

$$t_{\rm rel} = \frac{v^3 D^3}{3G^2 mM}\,. \tag{12.9}$$

If the cluster is dynamically relaxed and the virial theorem applies, then $v^2 = 3GM/(5D)$ and thus

$$t_{\rm rel} = \frac{D}{v}\frac{M}{m}\,\frac{v^4 D^2}{3G^2M^2} \sim \frac{ND}{v} = Nt_{\rm cross}\,. \tag{12.10}$$

Note that $t_{\rm rel} \gg t_{\rm cross}$, in striking contrast to an ordinary gas.

We should take into account how much momentum per collision is exchanged: In a collision at small impact parameter b the momentum transfer is larger than in one at large b. Moreover, one cannot treat relaxation just as a two-body process: Because of the infinite range of the gravitational force, N-body processes with $N > 2$ are not negligible. Formalizing this, the relaxation time given by Eq. (12.10) becomes reduced by a logarithmic term, $\approx 12\ln(N/2)$, or

$$t_{\rm rel} \approx 0.1 t_{\rm cross}\frac{N}{\ln(N)}\,. \tag{12.11}$$

Thus, $t_{\rm rel} \sim 2\times 10^9$ yr for the values of our example.

Evaporation time: The evaporation time for a cluster is the time required for the cluster to dissolve through the gradual loss of stars that gain sufficient velocity through encounters to escape its gravitational potential.

Assuming an isolated cluster with negligible stellar evolution, the evaporation time $t_{\rm ev}$ can be estimated by assuming that a constant fraction α of the stars in the cluster is evaporated every relaxation time. Thus, the rate of loss is $\mathrm{d}N/\mathrm{d}t = -\alpha N/t_{\rm rel} = -N/t_{\rm ev}$. The value of α can be determined by noting that the escape speed $v_{\rm esc}$ at a point x is related to the gravitational potential $E_{\rm pot}(x)$ at that point by $v_{\rm esc}^2 = -2E_{\rm pot}(x)$. (The total energy of a particle able to escape has to be equal or larger than zero, i.e. $E_{\rm kin} + E_{\rm pot} \geq 0$ or $v_{\rm esc}^2 \geq 2GM/R$.) If the system is virialized (as we would expect after one relaxation time), then also $\langle v^2\rangle = 3GM/(5R)$.

Thus, stars with speeds above twice the RMS speed will evaporate. Assuming a Maxwellian distribution of speeds, the fraction of stars with $v > 2v_{\rm rms}$ is $\alpha = 7.4\times 10^{-3}$. Therefore, the evaporation time is

$$t_{\rm ev} = \frac{t_{\rm rel}}{\alpha} \approx 136 t_{\rm rel}\,. \tag{12.12}$$

Stellar evolution and tidal interactions with the galaxy tend to shorten the evaporation time. Using a typical $t_{\rm rel}$ for a globular cluster, we see that $t_{\rm ev} \sim 10^{10}$ yr, which is comparable to the observed age of globular clusters.

$***$ **Emergence of order vs. the 2.nd law of thermodynamics** $***$ The second law of thermodynamics, $\mathrm{d}S \geq 0$, implies generally that the order in a closed system decreases: Heat flows from a hotter to a cooler subsystem, and particles diffuse from denser to under-dense regions. The standard example for this behavior are two subsystems of an ideal gas with different temperature and density. The entropy of each system is

$$S_i = kN_i\left(\frac{3}{2}\ln T_i - \ln\rho_i + C\right)\,. \tag{12.13}$$

An exchange of energy $\delta U_1 = -\delta U_2$ or particles $\delta N_1 = -\delta N_2$ between the two systems leads to a change in the total entropy, $\delta S = \delta S_1 + \delta S_2$. With $U = (3/2)NkT$ and thus

$$\delta S_i = \frac{\partial S_i}{\partial U_i}\delta U_i + \frac{\partial S_i}{\partial N_i}\delta N_i = \frac{\delta U_i}{T_i} + k\left(\frac{3}{2}\ln T_i - \ln \rho_i\right)\delta N_i\,, \tag{12.14}$$

the total change of the entropy $\delta S_{\rm tot} = S_1 + S_2$ follows as

$$\delta S_{\rm tot} = \left(\frac{1}{T_1} - \frac{1}{T_2}\right)\delta U_1 + k\left(\frac{3}{2}\ln\frac{T_1}{T_2} - \ln\frac{\rho_1}{\rho_2}\right)\delta N_1\,. \tag{12.15}$$

Therefore the second law of thermodynamics, $\mathrm{d}S \geq 0$, implies the decrease of temperature and density inhomogeneities.

The crucial difference gravity introduces is a connection between the temperature, density and size of the system. From the virial theorem we obtain

$$NT \propto U_{\rm kin} \propto -U_{\rm pot} \propto \frac{M^2}{R} \propto \frac{N^2}{V^{1/3}} \tag{12.16}$$

or $V \propto N^3/T^3$. Thus we can eliminate ρ and obtain

$$S = kN\left(-\frac{3}{2}\ln T + 2\ln N + C'\right)\,. \tag{12.17}$$

Note the changed signs of the $\ln T$ and $\ln N$ terms, indicating that the inhomogeneities in a system dominated by gravity increase.

12.3 Virial mass

For a dynamically relaxed cluster, we can use

$$M = \frac{5R\langle v^2\rangle}{3G} \tag{12.18}$$

to estimate the cluster mass M. Here, $\langle v^2\rangle$ as before refers to the "thermal" motions of stars. Thus the center-of-mass velocity of the cluster has to be subtracted. Using Doppler shift measurement, the radial component $\langle v_r^2\rangle$ can be measured. Since $\langle v^2\rangle = 3\langle v_i^2\rangle$ for an arbitrary cartesian component of v, we have also $\langle v^2\rangle = 3\langle v_r^2\rangle$. Thus

$$M = \frac{5R\langle v_r^2\rangle}{G}\,. \tag{12.19}$$

The main advantage of this method to determine masses is that from a representative sample of v_r measurements the total cluster mass including invisible stuff can be determined. Its main disadvantage is the inherent assumption of a relaxed, bound system that has to be checked, e.g. by comparing the cluster age with $t_{\rm rel}$; also there is a weak dependence of the viral mass on the exact density profile $\rho(r)$. This method can be applied to other bound systems as cluster of galaxies as well.

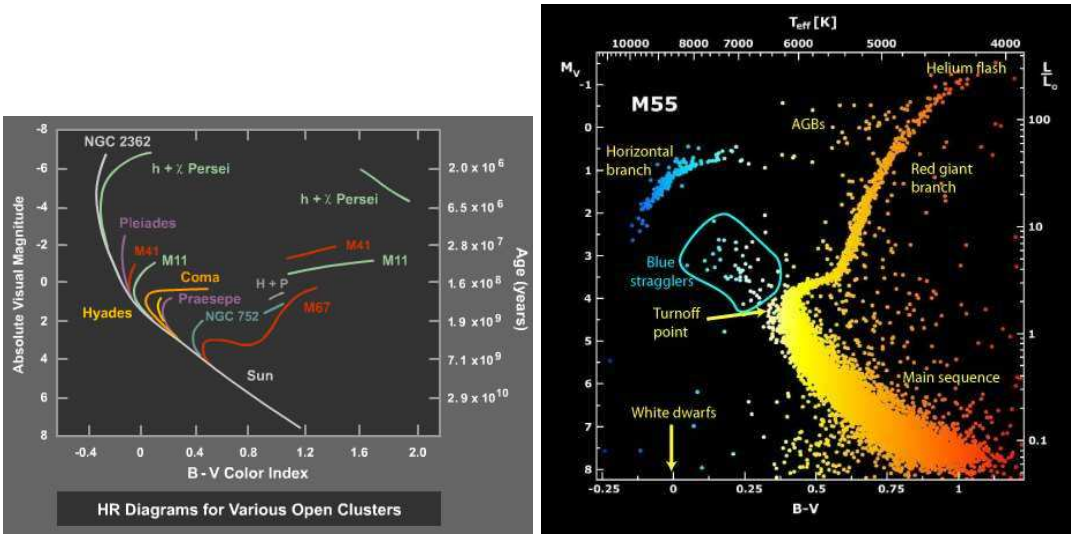


Figure 12.4: Hertzsprung-Russell diagram for several open clusters (left) and the globular cluster M55 (right).

12.4 Hertzsprung-Russell diagrams for clusters

Stars in a cluster have not only a common distance but were most likely also formed at the same time and from material with the same composition. Studying the Hertzsprung-Russell diagram of stars of the same cluster is therefore possible without knowing the distance to the individual stars. Moreover, a comparison of Hertzsprung-Russell diagrams of different clusters informs us about the evolution of stars.

Main sequence fitting A Hertzsprung-Russell diagram for stars of the same cluster can be made using the apparent magnitude m of the stars. In general, there is a shift up or down necessary so that the mean-sequence lies correctly in the absolute magnitude scale M. The shift gives directly the distance modulus $m - M$, or the cluster distance $d/\mathrm{pc} = 10^{(m-M+5)/5}$.

Turn-off point Hertzsprung-Russell diagrams of several clusters are shown in Fig. 12.4, left panel. For all clusters, a turn-off point is visible above which no mean-sequence stars are found. Thus the more luminous stars that are hotter and have a shorter life-time are missing in the clusters: They evolved already into white dwarfs or neutron stars. Since stars in a cluster were formed at the same time, the turn-off point fixes the cluster age. Table B.7 gives the life-time of stars on the mean-sequence. The age of the oldest globular cluster, $t \sim 12$ Gyr, gives a lower limit to the age of the Universe.

13 Galaxies

13.1 Milky Way

In visible light we see an accumulation of stars as a band across the sky. In the infrared or far-infrared the structure of the Milky Way consisting of a disk and a bulge starts to be revealed. The Milkyway is an example for a spiral galaxy; it can be subdivided into (cf. also Fig. 13.2):

- The galactic bulge with a radius of 3 kpc around the center of the galaxy. It is composed both of old (pop II) and young (pop I) stars.
- The galactic disk with a radius of 20 kpc and a thickness 300 pc contains also both pop I and pop II stars together with gas and dust. The disk appears not uniform, but has probably four spiral arms and a central bar. The reason why the arms of spiral galaxies are so prominent is that the brightest stars are found in the spiral arms. Spiral arms are the major regions of star formation in spiral galaxies and this is where most of the major nebulae are found.
- The galactic halo with a radius of 25 kpc includes globular clusters (consisting only of old stars (population II))
- A dark halo with an extension of order 100 kpc contains 80%-95% of the mass of the galaxy (dark matter).

The number of stars in the Milky Way is $\sim 10^{11}$, while its total mass is estimated to be $10^{12} M_\odot$.

13.1.1 Rotation curve of the Milkyway

In the disk of the Milky Way, stars and other matter is rotating around the center in a regular pattern, as revealed by Doppler effects. In the galactic halo and the galactic bulge, the motion is largely random.

Determination of the rotation curve For simplicity, we assume a spherical mass distribution and use again the enclosed mass $M(r)$. For circular orbits, $GM(r)m/r^2 = mv^2(r)/r$, or

$$M(r) = \frac{rv^2(r)}{G}\,. \tag{13.1}$$

Measurements of $v(r)$ are mainly done using the 21cm line of gas clouds as tracer. For orbits inside the Sun's orbit, the radial velocity is largest, when the distance along the line-of-sight to the Galactic center is minimal: Hence the maximal Doppler shift along one line-of-sight determines $v(r)$ at this minimal distance $r_{\min}$.

If the visible stars dominate the mass distribution of the Milky Way, then we expect for small r inside the bulge, $M(r) \propto r^3$ and thus $v(r) \propto r$. Outside the disc, for $r \gtrsim 20\,\text{kpc}$, $M(r) = \text{const.}$ and gas clouds should follow Keplerian orbits with $v(r) \propto 1/r^{1/2}$.

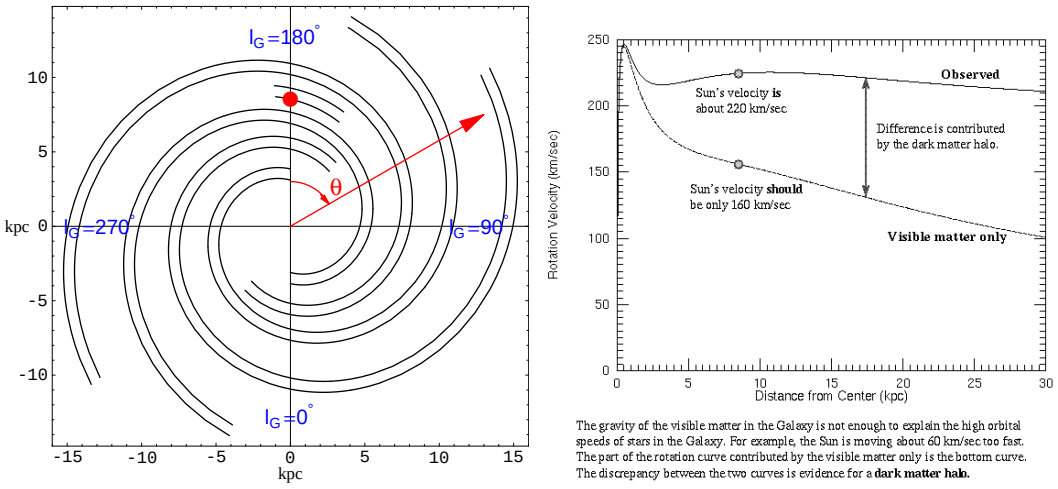


Figure 13.1: Left: Schematic view of the spiral structure of the Milky Way; the position of the Sun is marked by a red dot. Right: Rotation curve of the Milky Way.

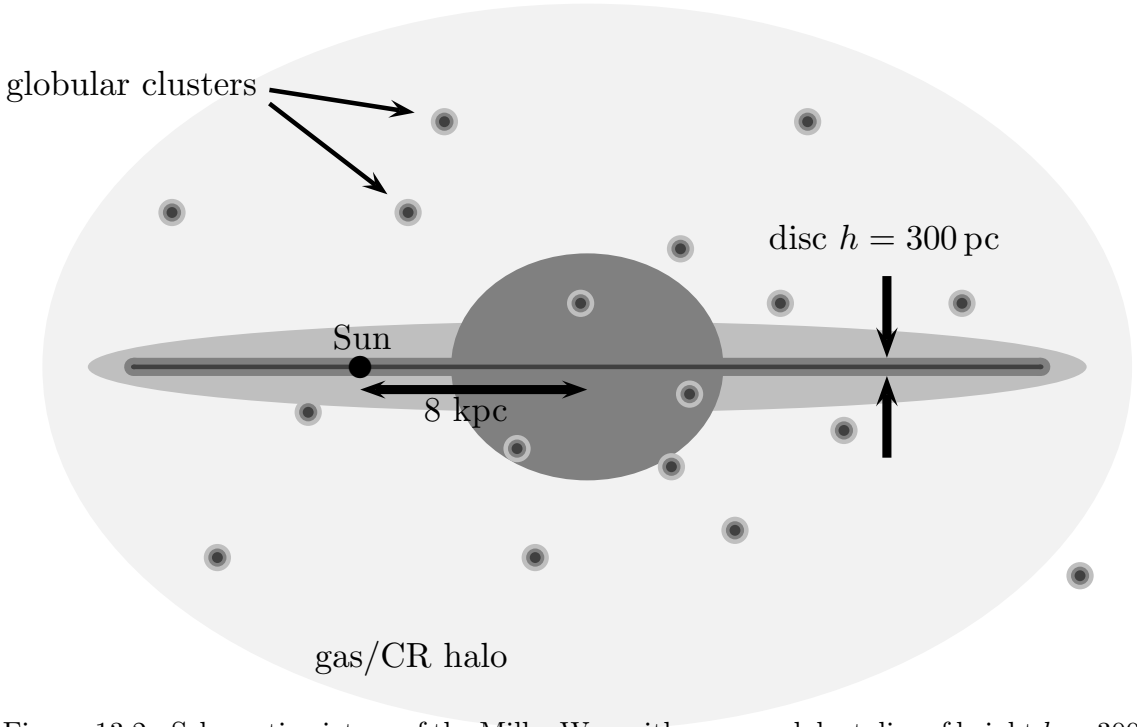


Figure 13.2: Schematic picture of the Milky Way with a gas and dust disc of height $h = 300$pc.

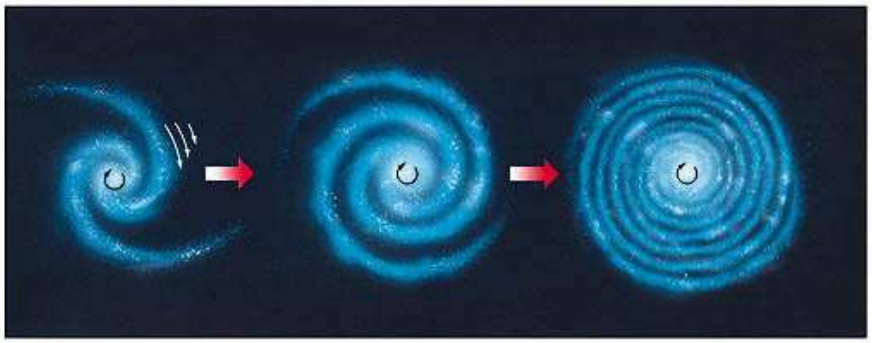


Figure 13.3: Winding-up of the spiral arms due to differential rotation.

Figure 13.1 compares the observed rotation curve with the one obtained theoretically from the observed distribution $M(r)$ of visible matter as stars and gas. The discrepancy indicates that a large fraction, around 90%, of the mass in the Milky Way is non-visible. What $\rho(r)$ corresponds to $v(r) = \text{const.}$? We have $v^2(r) = GM(r)/r = \text{const.} \propto \rho r^2$ or $\rho \propto 1/r^2$. Thus each radial shell of thickness dr contains the same amount of dark matter, d$M \propto \rho(r) r^2 dr \propto$ const. (This makes the definition of the "size" or edge of the DM halo somewhat arbitrary.) We shall discuss what could be the explanation for this unseen matter later, in Sec. 13.3.2.

Rotation and the spiral structure The rotation of stars around the center of the Milky Way is differential: The observed constant rotation velocity, $v(r) \approx$ const., means that the rotation period increases linearly with the distance to the Galactic center, $P \propto r$. Thus the spiral arms should wind up as shown schematically in Fig. 13.3. For instance, the rotation period of the Sun around the Galactic center is 2×10^8 yr, so the Sun completed 20 turns around the Milky Way. How does the observed spiral pattern survives?

The wind-up picture assumes that the spiral arms are a denser region of stars. However, we noted already that spiral arms are instead the major regions of star formation in spiral galaxies. Star formation is triggered by gravitational perturbations, in particular during encounters with neighbor galaxies, which compress interstellar matter. The velocity with which these density perturbations travel has a priori nothing to do with the rotation velocity of individual stars.

13.1.2 Black hole at the Galactic center

Supermassive black holes (SMBH) are supposed to be in the center of each galaxy, and that includes our own galaxy, the Milky Way. One way to show the existence of a SMBH is to deduce first the enclosed mass from rotation curves around the supposed BH. Then one has to show that no object with such a mass has a sufficient long lifetime.

Observationally, it is crucial to observe stars as close as possible to the BH. The large amount of dust in the Galactic bulge requires excellent IR observations. In spring 2002, a star was passing with $v = 5000$ km/s at 17 light hours distance – or 3AU – the point of closest approach to the black hole. One could determine a unique Keplerian orbit (highly elliptical (eccentricity 0.87), semimajor axis of 5.5 light days, a period of 15.2 years and an inclination of 46 degrees). Kepler's 3rd law gives as enclosed mass $M(r) = (3.7 \pm 1.5) \times 10^6 M_\odot$, cf.

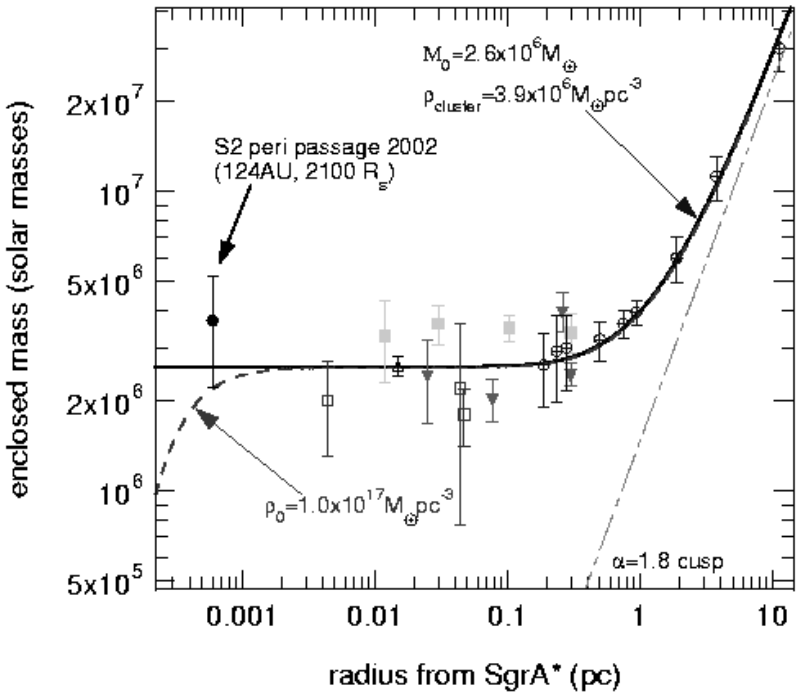


Figure 13.4: Enclosed mass as function of the distance to Sgr A (an object at the galactic center); the dashed-dotted line gives for comparison the enclosed mass for a DM halo with $\rho(r) \propto r^{-\alpha}$ and $\alpha = 1.8$.

Fig. 13.4. It is not possible to explain this result with a dense cluster of dark astrophysical objects because such a cluster would have the extremely short lifetime of at most a few hundred thousand years.

13.2 Normal and active galaxies

Are nebula as e.g. Andromeda galactic objects as globular clusters? Or is our galaxy just one of many in the Universe?

Ernst Julius Öpik estimate 1922 the distance to Andromeda as follows: The rotation velocity at the edge of Andromeda was known from Doppler measurements. Expressing then the radius R of Andromeda as $R = \vartheta d$, where d is its distance and ϑ the angular diameter, in the virial theorem, he obtained

$$\frac{GM}{(\vartheta d)^2} = \frac{v^2}{\vartheta d}\,. \tag{13.2}$$

The luminosity L and the observed flux $\mathcal{F}$ are connected by $\mathcal{F} = L/4\pi d^2$. Öpik did not know L, but he assumed that M/L is quite universal. Thus he used $M/L \sim 3$ determined as average value for the Milky Way also for Andromeda. Inserting $d^2 = \frac{L}{4\pi\mathcal{F}}$ on the LHS of

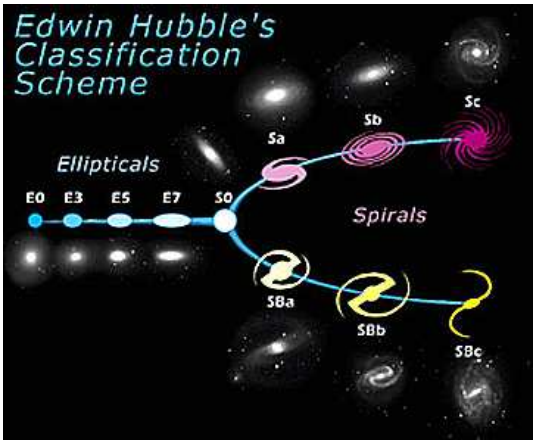


Figure 13.5: Hubble's classification scheme for galaxies.

Eq. (13.2) gives then

$$d \approx \frac{v^2 \vartheta L}{4\pi \mathcal{F} G M} \approx 450\,\mathrm{kpc}\,. \tag{13.3}$$

This distance is much larger than the size of the galactic stellar disc. Edwin Hubble measured 1924 the distance to three spiral nebula including Andromeda by observing Cepheides, confirming that these are extragalactic objects.

13.3 Normal Galaxies

The total energy emitted by a normal galaxies is the sum of the emission from each of the stars found in the galaxy. Thus the emission is roughly thermal and mainly in the infrared, visible and ultraviolet bands.

13.3.1 Hubble sequence

Spiral galaxies are two-dimensional objects, with or without bar. They make up 2/3 of all bright galaxies. They contain gas and dust; young and old stars. The stars move regularly in the disc. The surface luminosity of spiral galaxies decreases exponentially for large radii, $L(r) = L_0 \exp(-r/r_0)$ with $r_0 \sim 5\,\mathrm{kpc}$.

Elliptical galaxies have a three-dimensional, elliptical appearance. Most common are dwarf elliptical galaxies, with sizes of only a few kpc. But there exist also giant elliptical galaxies with an extension up to 100 kpc. Their low gas content ($< 1\%$) excludes the idea of an evolution from elliptical to spiral galaxies by flattening.

Elliptical galaxies contain no young stars. Stars are moving randomly. Their surface luminosity decreases as $L(r) = L_0 \exp(-r/r_0)^{1/4}$ with a large variation of r_0.

S0 and irregular galaxies are the remaining ones.

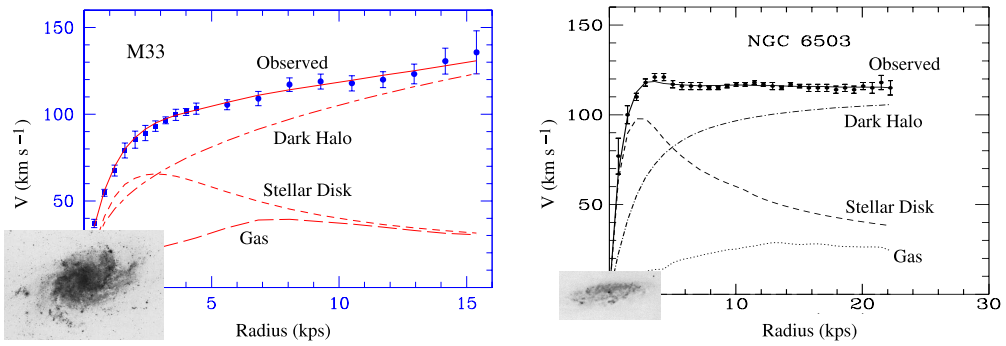


Figure 13.6: Rotation curves of two galaxies, superimposed are optical images.

13.3.2 Dark matter in galaxies

Flat rotation curves as far as luminous matter extends are found in practically all galaxies. Hence, as for the Milky way, $v(r) = \text{const.}$ corresponds to $\rho \propto 1/r^2$, compared to an exponential fall-off of luminous matter.

What are potential explanations for this discrepancy?

- Dust and gas clouds are seen by their blocking effect on light or in radio. Their contribution to the mass density has been accounted for.
- Most of the mass contained in galaxies could consist of non-luminous objects, so-called MAssive Compact Halo Object (MACHOs) like brown dwarfs, Jupiter-like objects, or back holes. This possibility can be tested by microlensing, i.e. the occasional amplification of light from extragalactic stars by the gravitational lens effect. Searches for microlensing towards the Large Magellanic Cloud exclude MACHOs as main component of dark matter in the Milky Way.
- The missing matter is composed of a new stable, neutral particle: "dark matter."
- Gravity has to be modified.

Modification of gravity (MOND) Newton's law of inertia $F = ma$ could be wrong for the very small accelerations that are relevant for the dynamics of galaxies. Such an effect could be incorporated by introducing a function μ in Newton's law,

$$m\mu(a/a_0)\mathbf{a} = \mathbf{F} \tag{13.4}$$

with $\mu(x) \to 1$ for $x \to \infty$ and $\mu(x) \to x$ for $x \to 0$. Choosing $a_0 \sim 10^{-12}\text{cm/s}^2$, all Newtonian standard results and observations are reproduced. Alternatively, Newtonian gravity could be modified,

$$\mu(g/a_0)\mathbf{g} = \mathbf{g_N}\,. \tag{13.5}$$

As long as gravity is the only force relevant for the dynamics, the two alternatives cannot be distinguished. MOND describes a large amount of data successful and also a relativistic generalization of MOND is possible. However, MOND fits many cosmological observations (that we shall discuss later) worse than the hypothesis that a large fraction of the mass of the universe is in form of a new dark matter particle.

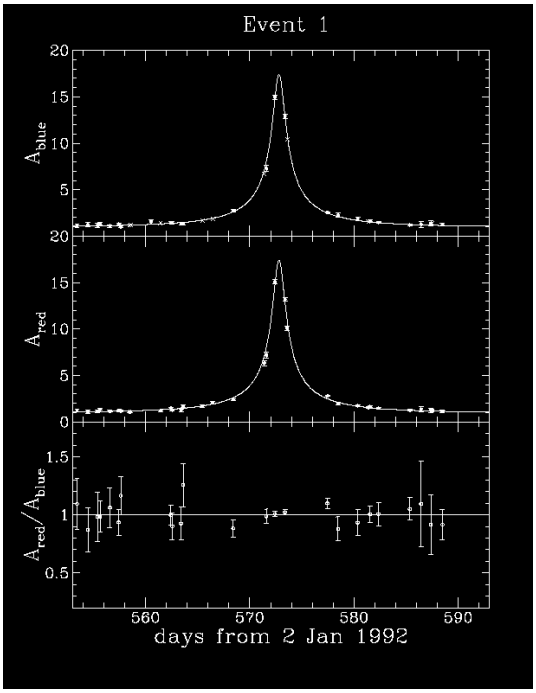


Figure 13.7: A potential MACHO microlensing event.

Dwarf galaxies and neutrino as DM If the mass of galaxies consists indeed mainly of an elementary particle different to protons, then our only candidate for the dark matter particle we know are neutrinos. As neutrino oscillations show, they are massive and they are most likely stable even on time-scales much larger than the age of the universe. However, we can use an argument similar to the one we used to derive the degeneracy pressure of fermions to exclude this possibility. Since neutrinos are fermions, they cannot be packed arbitrarily dense. Together with the upper mass limit $m_\nu < 3$ eV $\approx 6 \times 10^{-6} m_e$ from beta-decay experiments, the Pauli principle excludes neutrinos as main component of dark matter in galaxies.

Ex.: Neutrinos as dark matter.
Consider cold neutrinos with arbitrary mass m as explanation for dark matter in spiral and dwarf galaxies. (Use $M \sim 10^{11} M_\odot$ and $R = 50$kpc for spiral and $M \sim 10^7 M_\odot$ and $R = 0.1$kpc for dwarf galaxies.) Derive a lower limit for the neutrino mass m using that neutrinos are fermions. (Hint: connect the escape velocity from the galaxy with the maximal momentum of a degenerate Fermi gas.)
A neutrino bound to a galaxy has energy $E = GMm/R + mv^2/2 \leq 0$, hence its momentum should be smaller than $p_{\rm esc} = mv_{\rm esc} = \sqrt{2GM/R}$. On the other hand, the number N of possible neutrino states inside the volume $V = 4\pi R^3/3$ up to $p_{\rm esc}$ can be estimated with Heisenberg's uncertainty principle: In one dimension $\Delta x p_x/\hbar \gtrsim 1$ and thus

$$N \approx (1/\hbar^3) \int d^3x d^3p = \frac{4\pi}{3} R^3 \, \frac{4\pi}{3} (p_{\rm esc}/\hbar)^3 \, .$$

Since the Pauli principle forbids that 2 fermions occupy the same state, N is equal to the maximal number of neutrinos in the galaxy. The total mass in neutrinos follows as

$$M = Nm \leq \left(\frac{4\pi}{3}\right)^2 m^4 (2GMR/\hbar^2)^{3/2}$$

or neglecting numerical factors

$$m \gtrsim (G^3 R^3 M/\hbar^6)^{-1/8} \,.$$

For typical spiral galaxies, $m \gtrsim 10$ eV, while for typical dwarf galaxies $m \gtrsim 500$ eV. The current experimental upper limit is $m \lesssim$ few eV and excludes therefore neutrinos as main component of DM.

13.3.3 Galactic evolution

Galaxies are not isolated objects that are well-separated from each other. Instead they are bound together forming larger clusters, collide with each other and may merge. We discuss just one example for what might happen to a smaller body moving in the dark matter cloud of a larger one. For a concrete realization one can think at a dwarf galaxy like the Magellanic clouds orbiting around the Milky Way.

Dynamical friction What happens when a test body of mass M moves through a background of matter? Our test body will attract matter towards his position, creating a high-density "wake" trailing with him. The energy transferred to the surrounding material has to be supplied by the kinetic energy of our test body. As a result, a force F know as gravitational drag or dynamical friction acts on M.

We can obtain an understanding of this effect using just dimensional analysis. Since the force F can depend only on GM, v and ρ, the unique combination with the dimension of a force is (cf. Exercise 13.1)

$$F \approx C\,\frac{\rho (GM)^2}{v^2}\,, \tag{13.6}$$

where C is a dimensionless constant. This constant has to be determined from numerical simulations and varies between 10–100.

Inserting the density distribution appropriate for a dark matter halo, $\rho = v^2/(4\pi r^2 G)$, the force acting on a star cluster of mass M moving around a galaxy is

$$F \approx C\,\frac{GM^2}{4\pi r^2}\,. \tag{13.7}$$

The angular momentum of the cluster is $L = Mvr$. Since F acts tangentially to the obit (assumed to be circular), the torque $\tau = \mathrm{d}\mathbf{L}/\mathrm{d}t = \mathbf{r} \times \mathbf{F}$ is simply

$$\frac{\mathrm{d}L}{\mathrm{d}t} = -rF\,. \tag{13.8}$$

Using $v(r) = \text{const.}$ motivated by the observed rotation curves, we obtain

$$Mv\frac{\mathrm{d}r}{\mathrm{d}t} = -C\,\frac{GM^2}{4\pi r}\,. \tag{13.9}$$

Separating the variables and integrating from the initial distance R until zero gives the lifetime τ of a cluster moving through background matter with density ρ,

$$\int_R^0 \mathrm{d}r\, r = -\frac{CGM}{4\pi v}\int_0^\tau \mathrm{d}t \tag{13.10}$$

or

$$\tau = \frac{2\pi v R^2}{CGM}\,. \tag{13.11}$$

Ex.: i) The age of the oldest globular cluster in the Milky Way is $t = 13$ Gyr. Estimate with $C = 76$, $M = 5 \times 10^6 M_\odot$ and $v = 220$ km/s the maximal distance from which globular clusters could have been spiraled into the center of the galaxy.
ii) The large Magellanic Cloud (LMC), which has $M = 2\times10^{10} M_\odot$, orbits the Milky Way at a distance of 51 kpc. Assuming that the Milky Way's dark matter halo and flat rotation curve extends out to the LMC, estimate how much time it will take for the LMC to spiral into the Milky Way ($C = 23$).
For both questions we use

$$R = \left(\frac{C\tau GM}{2\pi v}\right)^{1/2}$$

solved either for R or τ. Solution to ii) is $R = 4$ kpc and to iii) $\tau = 1.7$ Gyr. Thus inside 4 kpc there should be no globular clusters, because they spiraled already into the center of the Milky Way and were dissolved, while the LMC has still some time to go...

13.4 Active Galaxies and non-thermal radiation

Active galaxies is a common name for all galaxies with unusual, mostly non-thermal emission not associated with stars. The modern view is that the source of the non-thermal radiation is connected to the SMBH in their center. We will therefore first discuss the two most important mechanisms to generate non-thermal photons.

13.4.1 Non-thermal radiation

Synchrotron radiation An electron in a homogeneous magnetic field moves on a Larmor orbit with radius $r_L = v_\perp mc/(eB)$. Because of its acceleration, the electron emits electromagnetic radiation. For typical magnetic field strengths found in galaxies, $B \sim \mu$G, and relativistic electrons, the emitted radiation is in the radio range.

We start by considering non-relativistic electrons. Then the Lorentz force is

$$\mathbf{F}_L = -\frac{e}{c}\,\mathbf{v}\times\mathbf{B} = m\mathbf{a}\,. \tag{13.12}$$

Hence the motion parallel to $\mathbf{B}$ is force-free ($\mathbf{v}\times\mathbf{v} = 0$), while the acceleration in the plane perpendicular to $\mathbf{B}$ is constant, $a_r = (e/m)v_\perp B$. Thus the electron moves along a helix.

Equating the Lorentz and the centripetal force,

$$(e/c)v_\perp B = \frac{mv_\perp^2}{r}\,, \tag{13.13}$$

and introducing the angular frequency $\omega = 2\pi\nu = v_\perp/r$, the cyclotron frequency ω_0 follows as

$$\omega_0 = \frac{eB}{mc} \qquad \text{or} \quad \nu_0/\text{Mhz} = 2.8B/\text{G}\,. \tag{13.14}$$

As long as electrons are non-relativistic, i.e. their total energy is dominated by their rest energy mc^2, they emit cyclotron radiation with the same frequency, independent of their velocity.

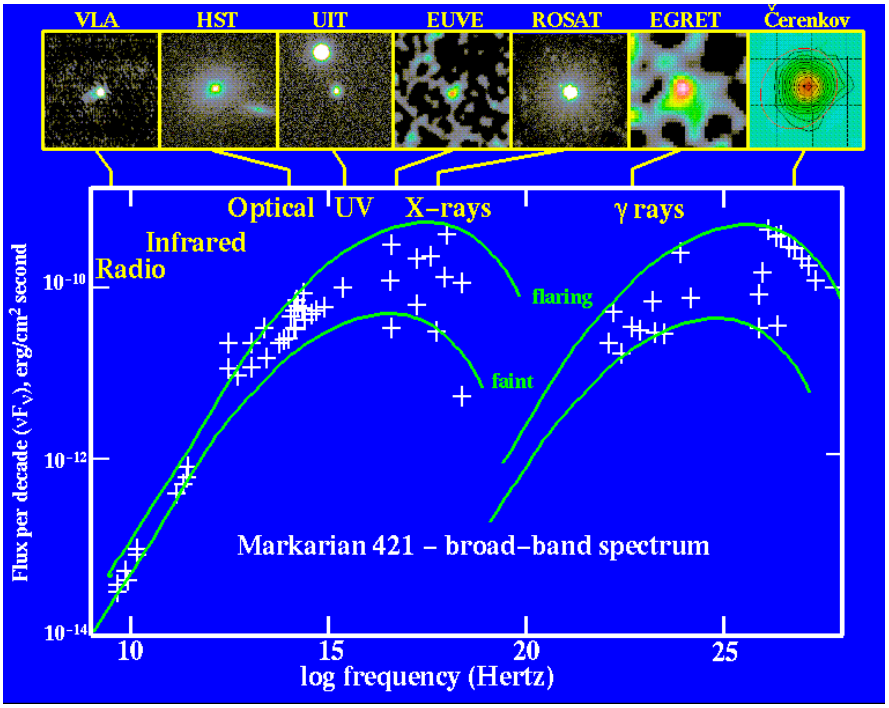


Figure 13.8: The emitted flux $\nu\mathcal{F}$ from the BL Lacartae Markarian 421 as function of frequency ν, in a quiet (lower line) and a flaring state (upper line).

For a relativistic electron, we have to replace $m\mathbf{a}$ by $\gamma m\mathbf{a}$ and thus the cyclotron frequency becomes $\omega_c = \frac{eB}{\gamma mc} = \frac{eB}{E/c}$. However, there are two additional effects. First, the radiation is beamed and emitted in a small cone of opening angle $\Delta\vartheta \approx 1/\gamma$. Second, there is a time-dilation between the frame of the electron and the observer (compare with Eq. (13.16)),

$$\frac{\mathrm{d}t}{\mathrm{d}\tau} = 1 - \beta\cos(\Delta\vartheta) \approx 1 - \beta + \beta(\Delta\vartheta)^2/2 \approx \gamma^{-2}\,. \tag{13.15}$$

As consequence, all frequencies from ω_0 to $\gamma^2\omega_0$ are emitted, although most energy goes into the frequency range around $\omega_{\rm cr} = (3/2)\gamma^2\omega_0$.

Ex.: Determine the Larmor radius r_L, its orbital period P, the opening angle ϑ and the frequency $\nu_{\rm cr}$ in which the radiation is mainly emitted, for an electron with $\gamma = 10^4$ in a magnetic field with strength typical for galaxies, $B = 3\mu$G.
The Larmor radius is $r_L = 6.0 \times 10^{12}$cm, the orbital period with $P = 1/\nu = \gamma/\nu_0 = 20$min, the opening angle $\vartheta = 0.005°$ and the frequency $\nu_{\rm cr} = 890$Mhz.

Inverse Compton scattering The usual text book set-up for Compton scattering assumes an energetic photon hitting an electron at rest, transferring part of its energy to the electron. In an astrophysical environments, the opposite situation is realized more often: A fast electron hits a low-energy photon and transfers a large fraction of its energy to the photon. ("Inverse Compton scattering"). Thus charged electron are accelerated by electromagnetic fields, emit synchrotron radiation in the radio range as well as high-energy photons via inverse Compton

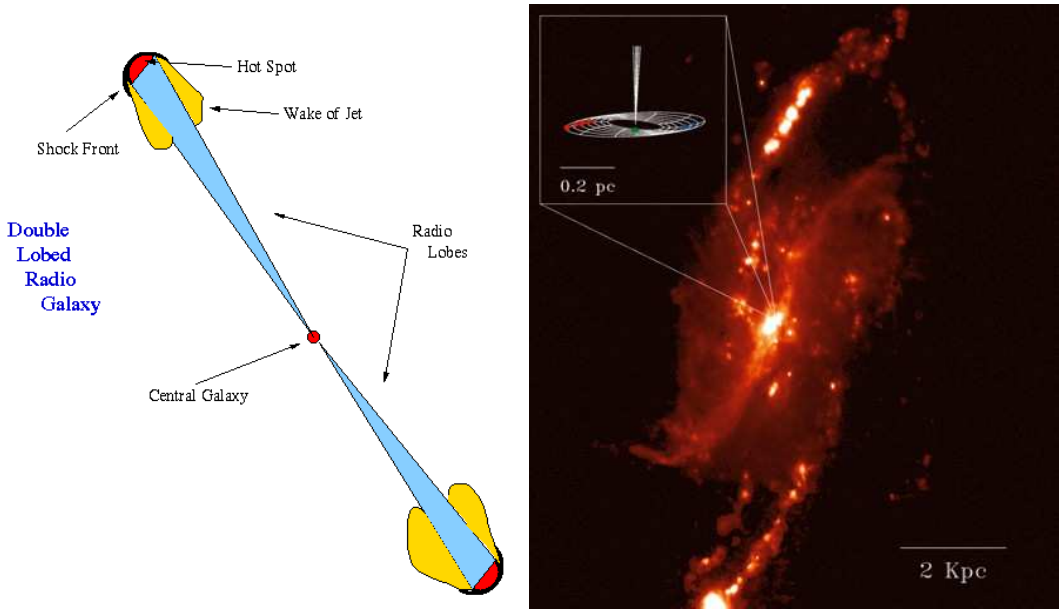


Figure 13.9: Left:Schematic picture of a radio galaxy. Right: NGC4258

scattering. An example where this two different components of the electromagnetic radiation are visible is shown in Fig. 13.8.

13.4.2 Radio galaxies

Radio galaxies emit typically 10^6 times more energy in the radio range than normal galaxies. The main emission mechanism is synchrotron radiation of electrons. The radio emission comes mainly from two *radio lobes* separated by up to a distance of order 10 Mpc, cf. Fig. 13.9. Additionally, there is a weaker radio source in center of the galaxy. Radio galaxies appear in the visible light as giant elliptical galaxies.

The lobes are caused by the interaction of a jet of high energy particles and the intergalactic medium: The medium slows down the jet, material is dragged along with the jet, while higher velocity particle are coming from behind. Thus a shocks forms and creates a "Hot Spot," which is the site of most intense emission, see Fig. 13.9.

Superluminal motion Consider a source moving with velocity v from A to B towards an observer. Denote the difference in radial distance to the observer by $x_\parallel$, the perpendicular one by $x_\perp$. Then a light signal from B needs $x_\parallel/c$ less time than a signal from A. Thus, the apparent time for the object to travel from A to B the distance r is

$$t_{\rm app} = \frac{r}{v} - \frac{x_\parallel}{c} = \frac{r}{v} - \frac{r}{c}\cos\vartheta = \frac{r}{v}(1-\beta\cos\vartheta)\,, \tag{13.16}$$

where r is the distance between A and B. The apparent velocity across the sky is then

$$v_{\rm app} = \frac{x_\perp}{t_{\rm app}} = \frac{r\sin\vartheta}{\frac{r}{v}(1-\beta\cos\vartheta)} = \frac{v\sin\vartheta}{1-\beta\cos\vartheta}\,. \tag{13.17}$$

For $v \ll c$, we obtain the expected result $v_{\rm app} \approx v \sin\vartheta$, while for $\beta \to 1$ the apparent velocity can exceed the speed of light. We determine the maximal value of the apparent velocity by first deriving the maximum of $v_{\rm app}$ as function of ϑ,

$$\frac{{\rm d}v_{\rm app}}{{\rm d}\vartheta} \propto \cos\vartheta_0(1-\beta\cos\vartheta_0) - \beta\sin^2\vartheta_0 = 0 \tag{13.18}$$

or $\cos\vartheta_0 = \beta$. Inserting the latter condition into Eq. (13.17) gives then

$$\left(\frac{v_{\rm app}}{v}\right)_{\rm max} = \frac{1}{(1-\beta^2)^{1/2}} = \gamma\,. \tag{13.19}$$

Hence the apparent velocity is maximally increased by the Lorentz factor γ of the source towards the direction $\cos\vartheta = \beta$.

Jet beaming Another important relativistic effect is jet beaming. Consider an emission process that is isotropic in the rest frame of the emitting source. Let us use the Lorentz transformation between energies in the rest system of the source (primed) and the observer system (unprimed). In the observer system,

$$E = \gamma(E' + \beta p'\cos\vartheta) \tag{13.20}$$

where ϑ is the angle between the velocity β of the source and the emitted photon. The maximal/minimal values of E follow directly as $E = \gamma(E' \pm \beta p')$ for $\cos\vartheta = \pm 1$, i.e. if the photons are emitted parallel and anti-parallel to the direction of the source. For a relativistic source, $\beta \to 1$, and the observed energy of photons in the backward direction goes to zero. Moreover, a similar effect leads also to an increase of the number of photons observed in the direction parallel to the source velocity. As a result, the emitted luminosity of a relativistic source (e.g. the hot spot of a radio galaxy) moving towards us can be increased by a large factor, 10^3–10^4, compared to isotropic emission.

13.4.3 Other AGN types and unified picture

Seyfert galaxies In 1943, Carl Seyfert noticed that certain nearby spiral galaxies have very bright, pinpoint nuclei. The spectra of these galaxies show very strong, often broad, emission lines. The brightness of the cores of Seyfert galaxies fluctuates: The light from the central nucleus varies in less than a year, which implies that the emitting region must be less than one light year across. They do not have radio lobes. Most are powerful sources of infrared radiation. In addition, some emit intensely in the radio, X ray, and gamma ray regimes. Approximately 2% of all spiral galaxies are Seyfert galaxies.

Seyfert galaxies are divided into two classes, based upon the widths of their spectral emission features. Seyfert 1 galaxies have hydrogen emission features with very large widths, indicating that the gas in the galaxy's central regions is moving with velocities of several thousand km/sec (Seyfert 1 galaxies show velocities up to almost 0.1c). Seyfert 2 galaxies have much narrower emission features implying much lower velocities (note that the Seyfert 1 in Fig. 13.10 shows narrow features as well).

BL Lacertae Main characteristics of BL Lacertae is their fast variability, with night-to-night variations of 10-30%, or a factor 100 within weeks if the source flares. Their spectra show no emission lines, and their appearance in the optical is point-like.

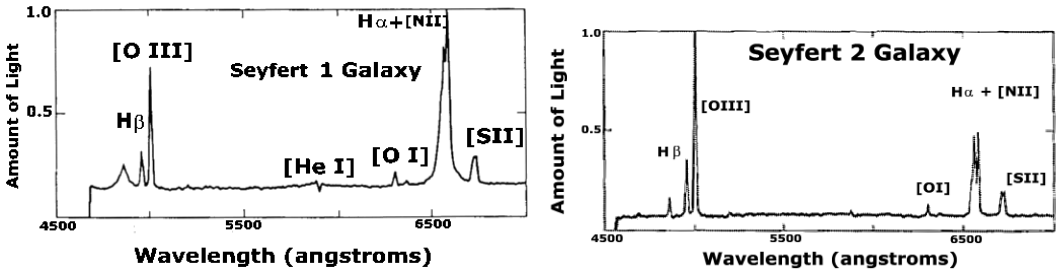


Figure 13.10: The spectra of Seyfert 1 (left) and Seyfert 2 (right) galaxies.

Quasars Quasi-stellar objects were discovered in late 1963, when a star-like object with $m = 13$ was identified with a strong radio source. Emission line as Balmer lines, but displaced by $\Delta\lambda/\lambda = 0.15$ were found. If interpreted as redshift due to expansion of Universe, quasars are at "cosmological distance." For the particular case of the quasi-stellar objects found 1963, the distance is 640 Mpc. This requires enormous luminosities: The distance modulus is

$$m - M = 5\log(d/10pc) = 39 \tag{13.21}$$

of $M = -26$. For comparison, the absolute magnitude of the Sun is 5, a difference of 31 corresponds to a brightness ratio of 10^{12}. In the radio, the source emits even stronger.

Energy gain from accretion onto SMBHs: The maximal energy gain is $E_{\max} = GmM/R_S$, where $R_S = 2GM/c^2$, and thus $E_{\max} = mc^2/2$. Modelling the accretion process gives a maximal efficiency of $\epsilon = 10\%$–20%. Thus the luminosity from accretion is

$$L = \frac{\epsilon c^2 \mathrm{d}m}{2\mathrm{d}t} \tag{13.22}$$

For $\mathrm{d}m/\mathrm{d}t = 1M_\odot/\mathrm{yr}$, one obtains $6\times10^{45}\mathrm{erg/s}$ or $L \sim 10^{12}L_\odot$. Thus the large energy output of AGN can be, at least in principle, easily explained by accretion onto an supermassive black hole.

Unified picture The fast variability and the large energy output point to accretion on SMBH as main energy source of the AGN activity. We observe different types of AGN due to their time evolution (e.g. from Quasars to Seyfert), different angles of view and different stages (active vs. quiet due to changes in $\mathrm{d}m/\mathrm{d}t$). A cartoon sketching the different types of AGN is shown in Fig. 13.11.

Exercises

1. Dynamical friction. Show that the unique combination with the dimension of GM, v and ρ with the dimension of a force is $F \propto \rho(GM)^2/v^2$.

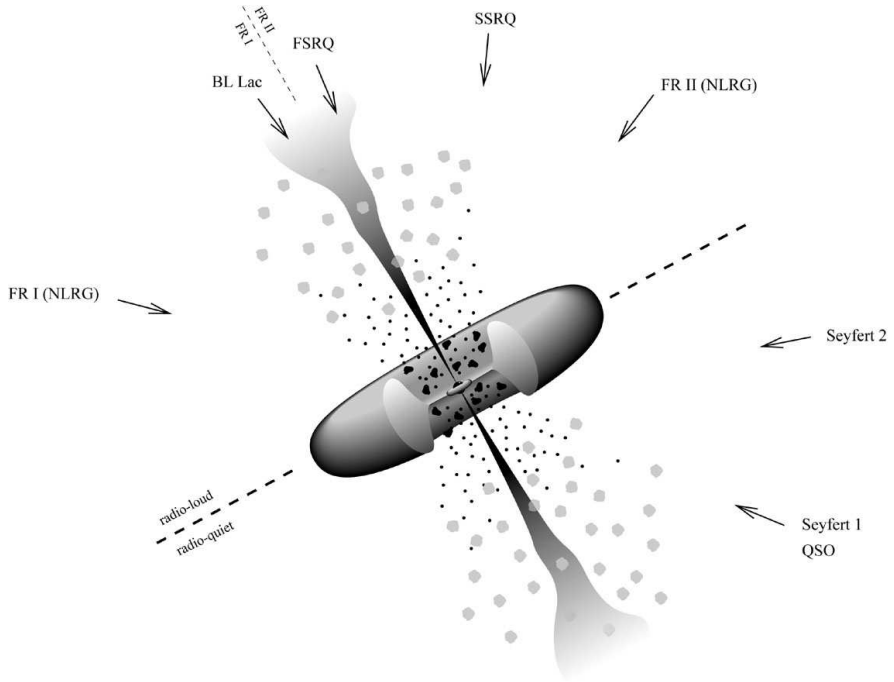


Figure 13.11: The unified picture for AGN explains the large AGN variety mainly as differences in the viewing angle relative to the jet axis.

Part III

Cosmology

14 Overview: Universe on large scales

14.1 Problems of a static, Newtonian Universe

Newton's law of universal gravitation changed together with the Copernican principle, i.e. the assumption that all bodies in the Universe obey the same laws as measured on Earth, cosmology from a purely philosophical subject to a sub-discipline of physics. In Newtonian gravity, a static Universe has to be infinite. However, a static infinite Universe would be an unstable solution and over-dense regions should collapse: As we discussed for the example of star clusters, gravitationally bound systems do not have an equilibrium solution and the second law of thermodynamics, $\mathrm{d}S \geq 0$, requires their collapse. Physicists around the year 1850 were concerned about the opposite faith of the Universe: A state without temperature or density gradients called "heat death."

Olber's paradox addresses the question "why is the night sky dark?" The number of stars in shell at distance r is proportional to $4\pi r^2 \mathrm{d}r$, and thus the observed flux $\mathcal{F} \propto 1/r^2$ from each shell is constant. If the Universe is infinite and has no beginning, the total flux diverges. Thus the darkness of the night sky is inconsistent with an Universe extending infinitely in space and time.

14.2 Einstein's cosmological principle

Einstein postulated that the Universe is homogeneous and isotropic (at each moment of its evolution). An important question is then at which scale the hierarchical structure of the universe (galaxies, cluster of galaxies, superclusters, ...?) stops.

In the extreme case of a fractal Universe an infinite sequence n of structures exists. Imagine then the case that a structure of order n contains p^n stars of mass M and occupies the volume kq^n. Then the density ρ_n averaged over structures of order n is

$$\rho_n \propto \frac{M}{k}\left(\frac{p}{q}\right)^n$$

with limit $\rho_n = 0$ for $p < q$ and $n \to \infty$. Hence, physical quantities O like $\langle 0 \rangle_R = (4\pi/V)\int_0^R dr r^2\, O(r)$ depend on the averaging scale R and the simple picture of an universe with, e.g., "the" average density ρ would be not possible.

Observations show that the hierarchical structure of the real universe stops at the scale of superclusters – the Universe is homogeneous on scales larger than a few 100 Mpc. There is just a factor ~ 30 between the size of the largest structures and size of the observable universe.

Hubble's $\mathcal{F}^{-3/2}$ test If we correct for extinction, we can use $\mathcal{F} = L/(4\pi r^2)$ assuming an euclidean geometry of space. If furthermore all galaxies would have the same luminosity L,

then all galaxies brighter than $> \mathcal{F}$ would be closer than r. Thus the total number of galaxies brighter than a certain limiting value is

$$N(>\mathcal{F}) = nV = \frac{4\pi n}{3}\left(\frac{L}{4\pi\mathcal{F}}\right)^{3/2} \propto \mathcal{F}^{-3/2}\,. \tag{14.1}$$

Around 1920, it was known that star counts vary less rapidly, i.e. the edge of the Milky Way became "visible". Hubble applied the same argument 1926 to galaxies and found agreement with the $\mathcal{F}^{-3/2}$ prediction. This was the first evidence for an homogeneity of the universe on large scales.

14.3 Expansion of the Universe: Hubble's law

Hubble's law Hubble found empirically that the spectral lines of "distant" galaxies are redshifted, $z = \Delta\lambda/\lambda_0 > 1$, with a rate proportional to their distance d,

$$cz = H_0 d\,. \tag{14.2}$$

If this redshift is interpreted as Doppler effect, $z = \Delta\lambda/\lambda_0 = v_r/c$, then the recession velocity of galaxies follows as

$$v = H_0 d\,. \tag{14.3}$$

The restriction "distant galaxies" means more precisely that $H_0 d \gg v_{\rm pec} \sim$ few $\times\, 100\,$km/s. In other words, the peculiar motion of galaxies caused by the gravitational attraction of nearby galaxy clusters should be small compared to the Hubble flow $H_0 d$. Note that the interpretation of v as recession velocity is problematic. The validity of such an interpretation is certainly limited to $v \ll c$.

The parameter H_0 is called Hubble constant and has the value $H_0 \approx 71^{+4}_{-3}$ km/s/Mpc. We will see soon that the Hubble law Eq. (14.3) is an approximation valid for $z \ll 1$. In general, the Hubble constant is not constant but depends on time, $H = H(t)$, and we will call it therefore Hubble parameter for $t \neq t_0$.

Hubble's law does not imply that we live at a special place of the Universe: Imagine a balloon that is inflated during the time dt so that all distances increase by the factor $(1+\mathrm{d}a)$. Then original distances l_0 are changed to $(1+\mathrm{d}a)l_0$ and the recession velocity is

$$v = \frac{\mathrm{d}a}{\mathrm{d}t} l_0 = \dot{a} l_0\,. \tag{14.4}$$

If we write $l(t) = l_0 a(t)$, then $a(t)$ is a scale factor describing how distances or space itself is stretched and

$$v = \dot{a} l_0 = \frac{\dot{a}}{a}\, l\,. \tag{14.5}$$

Comparing Eq. (14.3) and (14.5) we can identify H as $H = \dot{a}/a$, i.e. the Hubble parameter determines the relative change of distance per time.

The Hubble constant has the dimension of an inverse time. If v would be constant, then all galaxies and space would have been expanded from one single point to its present size in the time

$$\frac{1}{H_0} \approx (70\mathrm{km/s/Mpc})^{-1} \approx 4\times 10^{17}\mathrm{s} \approx 13\mathrm{Gyr}\,. \tag{14.6}$$

objects	methods	basis of calibration
nearby stars	trigonometric parallaxes	radar determination of AU
Galactic clusters	main sequence fitting	main sequence fixed by nearby stars trig. parallaxes/moving cluster
MS stars in Galaxy	spectroscopic parallax	trig. parallaxes/color-magnitude diagrams
Cepheides	period-luminosity relation	color-magnitude diagrams
Type Ia supernovae	standard candle	SN nearby in galaxies
spiral galaxies	Tully-Fisher relation	nearby galaxies
galaxies	Hubble's law	nearby galaxies

Table 14.1: Several distance indicators together with the underlying calibration method.

How precise might be $t_0 = H_0^{-1}$ as our first estimate for the age of the Universe? Since gravity should decelerate the expansion of the Universe, we might anticipate that a proper calculation should lead to a smaller age of the Universe. On the other hand, the age of the oldest stars in the Universe is estimated as 13 Gyr—thus there is some tension between a decelerating universe and a large value of H_0.

Note also that the actual size of the balloon at time t is not directly connected to the size of the "observable" part c/H_0 of the balloon. The former can be much larger than c/H_0 or even infinite.

Hubble's law as consequence of homogeneity Consider Hubble's law as a vector equation with us at the center of the coordinate system,

$$\mathbf{v} = H\mathbf{d}\,. \tag{14.7}$$

What sees a different observer at position $\mathbf{d}'$? He has the velocity $\mathbf{v}' = H\mathbf{d}'$ relative to us. We are assuming that velocities are small and thus

$$\mathbf{v}'' \equiv \mathbf{v} - \mathbf{v}' = H(\mathbf{d} - \mathbf{d}') = H\mathbf{d}''\,, \tag{14.8}$$

where $\mathbf{v}''$ and $\mathbf{d}''$ denote the position relative to the new observer. A linear relation between v and d as Hubble law is the only relation compatible with homogeneity and thus the "cosmological principle".

14.3.1 Cosmic distance ladder

The Table 14.1 summarizes a few methods to measure astronomical distances. The chain of overlapping methods by which astronomers establish distance scales in the universe is called "cosmic distance ladder:" Every extension of the distance ladder inherits all the uncertainties of the previous steps it is based on. It is therefore important to establish methods to measure cosmological distances calibrated with as few intermediate steps as possible.

Hubble's law For many galaxies measurements of their redshift has been performed. Since the conversion from redshift to true distances depends on e.g. the numerical value of the Hubble parameter, astronomers often uses the redshift to indicate the extragalactic distances.

Tully-Fisher relation An empirical relation published 1977 between the luminosity of a spiral galaxies and the width of its spectral lines, especially the 21cm line. Since $M \propto v^2$ and there is red- and blue-shifting on opposite sides, such a relation is expected. The detailed form has to derived empirically, since it depends on the distributions $v(r)$ and $L(r)$. One uses then the distance modulus to find the distance from the luminosity and the apparent magnitude.

15 Cosmological models for an homogeneous, isotropic universe

15.1 Friedmann-Robertson-Walker metric for an homogeneous, isotropic universe

The spatial part $\mathrm{d}l$ of the line-element of special relativity,

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - (\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2) = c^2\mathrm{d}t^2 - \mathrm{d}l^2\,, \tag{15.1}$$

corresponds to the one of an usual euclidean three-dimensional space. Such a space is flat, static, homogeneous and isotropic. An expanding universe means that at different times t, the distance $\mathrm{d}l$ between two observes at rest changes,

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - a^2(t)\left(\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2\right) = c^2\mathrm{d}t^2 - a^2(t)\mathrm{d}l^2\,. \tag{15.2}$$

Our postulate of homogeneity requires that the dimensionless function $a(t)$, the scale factor, depends only on t.

Do other homogeneous, isotropic metrics exist? In three dimensions, the sphere S^2 is another example for an homogeneous and isotropic space. This space has constant positive curvature and its line-element is $\mathrm{d}l^2 = r^2\mathrm{d}\vartheta^2 + r^2\sin^2\vartheta\mathrm{d}\phi^2$. Let us try to find the equivalent expression for $\mathrm{d}l$ in 3+1 dimensions by considering a sphere S^3 embedded in a (fictitious!) 4-dimensional,

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = R^2\,, \tag{15.3}$$

euclidean space,

$$\mathrm{d}l^2 = \mathrm{d}x_1^2 + \mathrm{d}x_2^2 + \mathrm{d}x_2^2 + \mathrm{d}x_4^2\,. \tag{15.4}$$

We choose x_1, x_2, x_3 as the three coordinates describing the sphere and eliminate x_4 using Eq. (15.3). Forming the differential of Eq. (15.4) gives $x_1\mathrm{d}x_1 + \ldots + x_4\mathrm{d}x_4 = 0$ and allows us thus to eliminate also $\mathrm{d}x_4$,

$$\mathrm{d}l^2 = \mathrm{d}x_1^2 + \mathrm{d}x_2^2 + \mathrm{d}x_2^2 + \frac{(x_1\mathrm{d}x_1 + x_2\mathrm{d}x_2 + x_3\mathrm{d}x_3)^2}{R^2 - x_1^2 - x_2^2 - x_3^3}\,. \tag{15.5}$$

This expression becomes more transparent in spherical coordinates,

$$x_1 = r\cos\phi\sin\vartheta\,, \tag{15.6}$$

$$x_2 = r\sin\phi\sin\vartheta\,, \tag{15.7}$$

$$x_3 = r\cos\vartheta\,. \tag{15.8}$$

We calculate

$$\mathrm{d}x_i = \frac{\partial x_i}{\partial r}\,\mathrm{d}r + \frac{\partial x_i}{\partial \phi}\,\mathrm{d}\phi + \frac{\partial x_i}{\partial \vartheta}\,\mathrm{d}\vartheta\,, \tag{15.9}$$

i.e.

$$\mathrm{d}x_1 = \cos\phi \sin\vartheta \mathrm{d}r - r\sin\phi\sin\vartheta\mathrm{d}\phi + r\cos\phi\cos\vartheta\mathrm{d}\vartheta,$$

etc., and insert them into $\mathrm{d}l^2$. After somewhat lengthy but elementary simplifications we obtain

$$\mathrm{d}l^2 = \frac{\mathrm{d}r^2}{1-r^2/R^2} + r^2(\sin^2\vartheta\mathrm{d}\phi^2 + \mathrm{d}\vartheta^2) = \frac{\mathrm{d}r^2}{1-r^2/R^2} + r^2\mathrm{d}\Omega\,. \tag{15.10}$$

Note that the angular part $\mathrm{d}\Omega$ corresponds to the usual expression for spherical coordinates, while the radial distance $\mathrm{d}r$ is modified by the factor $(1-r^2/R(t)^2)^{-1}$.

The only other form of a homogeneous, isotropic space is a hyperbolic plane with constant negative curvature. Formally, a hyperbolic plane can be identified with a sphere with imaginary radius R, i.e. it is obtained by replacement $1-r^2/R^2 \to 1+r^2/R^2$.

Combining the three cases (and making the radial coordinate dimensionless, $r \to r/R(t)$), the Friedmann-Robertson-Walker (FRW) metric in its most commonly used representation follows as

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - R^2(t)\left[\frac{\mathrm{d}r^2}{1-kr^2} + r^2(\sin^2\vartheta\mathrm{d}\phi^2 + \mathrm{d}\vartheta^2)\right] \tag{15.11}$$

with $k = \pm 1$ (positive/negative curvature) or $k = 0$ (flat three-dimensional space).

Geometry of the FRW spaces

Let us consider a sphere of fixed radius at fixed time, $\mathrm{d}r = \mathrm{d}t = 0$. The line-element $\mathrm{d}s$ simplifies then to $R^2(t)r^2(\sin^2\vartheta\mathrm{d}\phi + \mathrm{d}\vartheta^2)$, which is the usual line-element of a sphere S^2 with radius $rR(t)$. Thus the area of the sphere is $A = 4\pi(rR(t))^2$ and the circumference of a circle is $L = 2\pi rR(t)$, while $rR(t)$ has the physical meaning of a length.

By contrast, the radial distance between two points (r,ϑ,ϕ) and $(r+\mathrm{d}r,\vartheta,\phi)$ is $\mathrm{d}l = R(t)\mathrm{d}r/\sqrt{1-kr^2}$. Thus the radius of a sphere centered at $r = 0$ is

$$l = R(t)\int_0^r \frac{\mathrm{d}r'}{\sqrt{1-kr'^2}} = R(t)\times\begin{cases}\arcsin(r) & \text{for}\quad k=1\,,\\ r & \text{for}\quad k=0\,,\\ \mathrm{arcsinh(r)} & \text{for}\quad k=-1\,.\end{cases} \tag{15.12}$$

Hence for $k = 0$, i.e a flat space, one obtains the usual result $L/l = 2\pi$, while for $k = 1$ (spherical geometry) $L/l = 2\pi r/\arcsin(r) < 2\pi$ and for $k = -1$ (hyperbolic geometry) $L/l = 2\pi r/\mathrm{arcsinh(r)} > 2\pi$.

For $k = 0$ and $k = -1$, l is unbounded, while for $k = +1$ there exists a maximal distance $l_{\max}(t)$. Hence the first two case correspond to open spaces with an infinite volume, while the latter is a closed space with finite volume.

Lemaître's redshift formula

A light-ray propagates with $v = c$ or $\mathrm{d}s^2 = 0$. Assuming a galaxy at $r = 0$ and an observer at r, i.e. light rays with $\mathrm{d}\phi = \mathrm{d}\vartheta = 0$, we rewrite the FRW metric as

$$\frac{\mathrm{d}t}{R} = \frac{\mathrm{d}r}{\sqrt{1-kr^2}}\,. \tag{15.13}$$

We integrate this expression between the emission and absorption times t_1 and t_2 of the first light-ray,

$$\int_{t_1}^{t_2}\frac{\mathrm{d}t}{R} = \int_0^r \frac{\mathrm{d}r}{\sqrt{1-kr^2}} \tag{15.14}$$

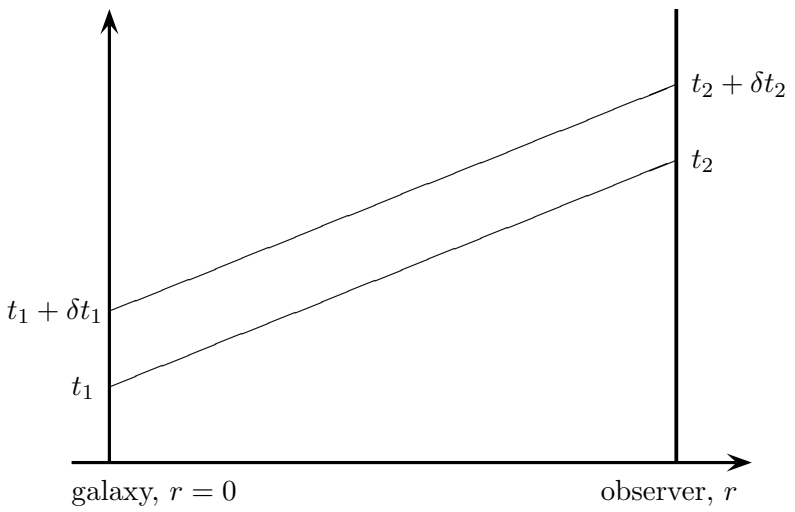


Figure 15.1: World lines of a galaxy emitting light and an observer at comoving coordinates $r = 0$ and r, respectively.

and between $t_1 + \delta t_1$ and $t_2 + \delta t_2$ for the second light-ray (see also Fig. 15.1),

$$\int_{t_1+\delta t_1}^{t_2+\delta t_2} \frac{\mathrm{d}t}{R} = \int_0^r \frac{\mathrm{d}r}{\sqrt{1-kr^2}}\,. \tag{15.15}$$

The RHS's are the same and thus we can equate the LHS's,

$$\int_{t_1}^{t_2} \frac{\mathrm{d}t}{R} = \int_{t_1+\delta t_1}^{t_2+\delta t_2} \frac{\mathrm{d}t}{R}\,. \tag{15.16}$$

We change the integration limits, subtracting the common interval $[t_1 + \delta t_1 : t_2]$ and obtain

$$\int_{t_1}^{t_1+\delta t_1} \frac{\mathrm{d}t}{R} = \int_{t_2}^{t_2+\delta t_2} \frac{\mathrm{d}t}{R}\,. \tag{15.17}$$

Now we choose the time intervals δt_i as the time between two wave crests separated by the wave lengths λ_i of an electromagnetic wave. Since these time intervals are extremely short compared to cosmological times, $\delta t_i = \lambda_i/c \ll t_i$, we can assume $R(t)$ as constant performing the integrals and obtain

$$\frac{\delta t_1}{R_1} = \frac{\delta t_2}{R_2} \qquad \text{or} \qquad \frac{\lambda_1}{R_1} = \frac{\lambda_2}{R_2}\,. \tag{15.18}$$

The redshift z of an object is defined as the relative change in the wavelength between emission and detection,

$$z = \frac{\lambda_2 - \lambda_1}{\lambda_1} = \frac{\lambda_2}{\lambda_1} - 1 \tag{15.19}$$

or

$$1 + z = \frac{\lambda_2}{\lambda_1} = \frac{R_2}{R_1}\,. \tag{15.20}$$

This result is intuitively understandable, since the expansion of the universe stretches all lengths including the wave-length of a photon. For a massless particle like the photon, $\nu = c\lambda$

and $E = cp$, and thus its frequency (energy) and its wave-length (momentum) are affected in the same way. By contrast, the energy of a non-relativistic particle with $E \approx mc^2$ is nearly fixed. A similar calculation as for the photon shows that indeed not the energy but the momentum $p = \hbar/\lambda$ of massive particles is red-shifted.

To be able to compare Eq. (15.20) with Hubble's law, $v = Hd$ with $v = cz$ and $H = \dot{R}/R$, we have to understand next how the scale factor R evolves with time.

15.2 Friedmann equation from Newton's and Hubble's laws

The Friedmann equation connects the expansion rate H of the Universe with its energy density. Since we assume the cosmological principle to be valid, i.e. we assume an homogeneous and isotropic Universe, this equation is a single ordinary differential equation. We can derive this equation in a straightforward manner using Hubble's law together with the Newtonian expression for the energy of a gravitating system plus the insight of special relativity that gravity should couple also to energy.

15.2.1 Friedmann equation

Let us consider the (Newtonian) energy of a galaxy of mass m at the distance R in a Universe that is very close to homogeneity and isotropy,

$$E = E_{\rm kin} + E_{\rm pot} = \frac{1}{2} m\dot{R}^2 - \frac{GmM(R)}{R} \,. \tag{15.21}$$

According to Hubble's law we can express the recession velocity $\dot{R}$ of the galaxy as $\dot{R} = HR$, while the mass enclosed within R is given by $M(R) = (4\pi/3)R^3\rho_m$. Hence

$$E = \frac{1}{2} m(HR)^2 - \frac{4\pi}{3} Gm\rho_m R^2 \tag{15.22}$$

or

$$\frac{2E}{mR^2} = H^2 - \frac{8\pi}{3}\, G\rho_m \,. \tag{15.23}$$

Since we assume that the universe is homogeneous, H and ρ_m cannot be functions of R. Therefore also the LHS, $\frac{2E}{mR^2}$, cannot depend on the chosen distance R to the coordinate center. However, the value of $2E/(mR^2)$ is time-dependent, because the distance between us and the galaxy will change as the Universe expands. Since the mass m of our test galaxy is arbitrary, we can choose it such that $|2E/(mc^2)| = 1$ holds at an arbitrary moment as long as $E \neq 0$. For different times, the LHS scales as $1/R^2$ and thus we can rewrite Eq. (15.23) with $k = 0, \pm 1$ as

$$\frac{kc^2}{R^2} = H^2 - \frac{8\pi}{3} G\rho_m \,. \tag{15.24}$$

Since E is constant, k is constant too. Finally, we account for the equivalence of mass and energy by including in ρ not only the mass but also the energy density, $\rho = \rho_m c^2 + \rho_{\rm rad} + \ldots$:

$$H^2 \equiv \left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi}{3} G\rho/c^2 - \frac{kc^2}{R^2} \,. \tag{15.25}$$

The result is the "Friedmann equation" without cosmological constant and agrees exactly with the one derived from general relativity[1]. The three cases $k = -1, 0, +1$ in the Friedmann equation correspond to the same three cases in Eq. (15.11). Hence they distinguish the geometry of space, i.e. the geometry of 3-dimensional hypersurfaces with $t = \text{const}$. This geometry (k and R) can be determined by measuring two simple numbers, H_0 and ρ_0.

We set from now on $c = 1$. Quantities evaluated at the present time are labelled with a zero, $t_0, H_0, \ldots$. The value of H_0 fixes the present density ρ_0 for $k = 0$ as $\rho_0(k = 0) \equiv \rho_{\rm cr} = \frac{3H_0^2}{8\pi G}$. Most cosmological quantities like $\rho_{\rm cr}$ will depend on the actual value of H_0. One "hides" this dependence by introducing h,

$$H_0 = 100h\text{km/s/Mpc}\,.$$

Then one can express the critical density as function of h,

$$\rho_{\rm cr} = 2.77 \times 10^{11} h^2 M_\odot/\text{Mpc}^3 = 1.88 \times 10^{-29} h^2 \text{g/cm}^3 = 1.05 \times 10^{-5} h^2 \text{GeV/cm}^3\,.$$

Thus a flat universe with $H_0 = 100h$km/s/Mpc requires an energy density of ~ 10 protons per cubic meter.

We define the abundance Ω_i of the different players in cosmology as their energy density relative to $\rho_{\rm cr}$, $\Omega_i = \rho/\rho_{\rm cr}$. The ones we know already are non-relativistic or cold dark matter (CDM with Ω_m) and relativistic matter or radiation ($\Omega_{\rm rad}$).

Cosmological constant In general relativity, the absolute value of the energy matters not only energy differences. Thus we should add a constant (the "cosmological constant" Λ) to the RHS of the Friedmann equation to account for a possible intrinsic energy and pressure of the vacuum,

$$H^2 \equiv \left(\frac{\dot R}{R}\right)^2 = \frac{8\pi}{3} G\rho - \frac{k}{R^2} + \frac{\Lambda}{3}\,. \tag{15.26}$$

With $\Lambda \neq 0$, the simple relation $\rho = \rho_{\rm cr} \leftrightarrow k = 0$ still holds, if the contribution of Λ to the energy density ρ is included viz.

$$\frac{8\pi}{3} G\rho_\Lambda = \frac{\Lambda}{3}\,. \tag{15.27}$$

Thereby one recognizes also that the cosmological constant acts as a constant energy density $\rho_\Lambda = \frac{\Lambda}{8\pi G}$, or $\Omega_\Lambda = \frac{\Lambda}{3H_0^2}$.

For $\Lambda \neq 0$, an observable additional to ρ_0 and H_0 is needed to fix the geometry of the three-dimensional spatial hypersurface: We obtain this observable by a Taylor expansion of $R(t)$,

$$\begin{aligned} R(t) &= R(t_0) + (t - t_0)\dot R(t_0) + \frac{1}{2}(t - t_0)^2 \ddot R(t_0) + \ldots & (15.28)\\ &= R(t_0)\left[1 + (t - t_0)H_0 - \frac{1}{2}(t - t_0)^2 q_0 H_0^2 + \ldots\right]\,, & (15.29)\end{aligned}$$

where $q \equiv -\ddot R(t_0)R(t_0)/\dot R^2(t_0)$ is called deceleration parameter: If the expansion is slowing down, $\ddot R < 0$ and $q_0 > 0$.

[1] In general relativity, a generalization of Gauss' theorem called Birkhoff's theorem is valid. Thus only the enclosed mass and energy enter the Friedmann equation, and choosing R sufficiently small, the Newtonian approximation becomes arbitrarily precise.

We show now that Hubble's law is indeed an approximation for small redshift. For not too large time-differences, we can use the expansion Eq. (15.28) and write

$$1 - z \approx \frac{1}{1+z} = \frac{R(t)}{R_0} \approx 1 + (t - t_0)H_0 \,. \tag{15.30}$$

Hence Hubble's law, $z = (t_0 - t)H_0 = d/cH_0$, is valid as long as $z \approx H_0(t_0 - t) \ll 1$. Deviations from its linear form arises for $z \gtrsim 1$ and can be used to determine q_0.

15.2.2 Local energy conservation and acceleration equation

Additionally to the Friedmann equation we need one equation that describes how the energy content of the Universe is affected by its expansion.

Local energy conservation The first law of thermodynamics becomes with $\mathrm{d}Q = 0$ (no heat exchange to the outside, since no outside exists) simply

$$\mathrm{d}U = T\mathrm{d}S - P\mathrm{d}V = -P\mathrm{d}V \tag{15.31}$$

or

$$\mathrm{d}(\rho R^3) = -P\mathrm{d}(R^3) \,. \tag{15.32}$$

Dividing by $\mathrm{d}t$,

$$R\dot{\rho} + 3(\rho + P)\dot{R} = 0 \,, \tag{15.33}$$

we obtain

$$\dot{\rho} = -3(\rho + P)H \,. \tag{15.34}$$

Thus the expansion decreases the energy density both by dilution and by the work required to expand a gas with pressure $P \geq 0$.

Ex.: Derive how the expansion of the Universe changes the particle number density n
The Hubble flow $v = Hr$ induces the flux vn through the surface $4\pi r^2$ of a sphere with radius r, and thus $\dot{N} = 4\pi r^2 vn$. These particles are missing inside the sphere containing the total number of particles $N = Vn$. Hence

$$\dot{N} = -4\pi r^2 vn = \frac{4\pi}{3} r^3 \dot{n} \tag{15.35}$$

or $\dot{n} = -3vn/r = -3Hn$. Thus the first term on the RHS of Eq. (15.34) corresponds indeed due the dilution of the particle number by the expansion of the Universe. More formally, we can use that $n \propto 1/R^3$ and thus

$$\frac{\mathrm{d}n}{\mathrm{d}t} = \frac{\mathrm{d}n}{\mathrm{d}R}\frac{\mathrm{d}R}{\mathrm{d}t} = -3\,\frac{n}{R}\,\dot{R} \,.$$

Acceleration equation An equivalent, but often more useful equation is obtained by first multiplying the Friedmann equation with R^2,

$$\dot{R}^2 = \frac{8\pi G}{3}\rho R^2 - k + \frac{\Lambda R^2}{3}\,. \tag{15.36}$$

Taking then the time derivative,

$$2\dot{R}\ddot{R} = \frac{8\pi G}{3}\left[\dot{\rho}R^2 + \rho(2R\dot{R})\right] + \frac{2\Lambda R\dot{R}}{3} \tag{15.37}$$

inserting $\dot{\rho}R = -3(\rho + P)\dot{R}$ into $\dot{\rho}R^2$ and dividing by $2\dot{R}$ gives

$$\ddot{R} = \frac{4\pi G}{3}\left[(-3\rho - 3P) + 2\rho\right]R + \frac{\Lambda}{3}R \tag{15.38}$$

and finally the "acceleration equation"

$$\frac{\ddot{R}}{R} = \frac{\Lambda}{3} - \frac{4\pi G}{3}(\rho + 3P)\,. \tag{15.39}$$

This equation determines the (de-) acceleration of the Universe as function of its matter and energy content. "Normal" matter is characterized by $\rho > 0$ and $p \geq 0$. Thus a static solution is impossible for a universe with $\Lambda = 0$. Such a universe is decelerating and since today $\dot{R} > 0$, $\ddot{R}$ was always negative and there was a "big bang".

We can understand better the physical properties of the cosmological constant by replacing Λ by $(8\pi G)\rho_\Lambda$. Now we can compare the effect of normal matter and of the Λ term on the acceleration,

$$\frac{\ddot{R}}{R} = \frac{8\pi G}{3}\rho_\Lambda - \frac{4\pi G}{3}(\rho + 3P) \tag{15.40}$$

Thus Λ is equivalent to matter with an E.o.S. $w_\Lambda = P/\rho = -1$. This property can be checked using only thermodynamics: With $P = -(\partial U/\partial V)_S$ and $U_\Lambda = \rho_\Lambda V$, it follows $P = -\rho$.

The borderline between an accelerating and decelerating universe is given by $\rho = -3P$ or $w = -1/3$. The condition $\rho < -3P$ violates the so-called energy condition for "normal" matter in equilibrium. An accelerating universe requires therefore a positive cosmological constant or a dominating form of matter that is not in equilibrium.

Note that the energy contribution of relativistic matter, photons and possibly neutrinos, is much smaller than the one of non-relativistic matter (stars and cold dark matter). Thus the pressure term in the acceleration equation can be neglected at the present epoch. Measuring $\ddot{R}/R$, $\dot{R}/R$ and ρ fixes therefore the geometry of the universe.

15.3 Scale-dependence of different energy forms

The dependence of different energy forms as function of the scale factor R can derived from energy conservation, $\mathrm{d}U = -P\mathrm{d}V$, if an E.o.S. $P = P(\rho) = w\rho$ is specified. For $w = \text{const.}$, it follows

$$\mathrm{d}(\rho R^3) = -3PR^2\mathrm{d}R \tag{15.41}$$

or eliminating P

$$\frac{\mathrm{d}\rho}{\mathrm{d}R}R^3 + 3\rho R^2 = -3w\rho R^2\,. \tag{15.42}$$

Separating the variables,

$$-3(1+w)\frac{\mathrm{d}R}{R} = \frac{\mathrm{d}\rho}{\rho}\,, \tag{15.43}$$

we can integrate and obtain

$$\rho \propto R^{-3(1+w)} = \begin{cases} R^{-3} & \text{for matter} \quad (w=0) \\ R^{-4} & \text{for radiation} \quad (w=1/3) \\ \text{const.} & \text{for } \Lambda \quad (w=-1) \end{cases} \tag{15.44}$$

This result can be understood also from heuristic arguments:

- (Non-relativistic) matter means that $kT \ll m$. Thus $\rho = nm \gg nT = P$ and non-relativistic matter is pressure-less, $w = 0$. The mass m is constant and $n \propto 1/R^3$, hence ρ is just diluted by the expansion of the universe, $\rho \propto 1/R^3$.
- Radiation is not only diluted but the energy of each single photon is additionally red-shifted, $E \propto 1/R$. Thus the energy density of radiation scales as $\propto 1/R^4$. Alternatively, one can use that $\rho = aT^4$ and $T \propto \langle E \rangle \propto 1/R$.
- Cosmological constant Λ: From $\frac{8\pi}{3}G\rho_\lambda = \frac{\Lambda}{3}$ one obtains that the cosmological constant acts as an energy density $\rho_\lambda = \frac{\Lambda}{8\pi G}$ that is constant in time, independent from a possible expansion or contraction of the universe.
- Note that the scaling of the different energy forms is very different. It is therefore surprising that "just today", the energy in matter and due to the cosmological constant is of the same order ("coincidence problem").

15.4 Cosmological models with one energy component

We consider a flat universe, $k = 0$, with one dominating energy component with E.o.S $w = P/\rho = \text{const.}$. With $\rho = \rho_0(R/R_0)^{-3(1+w)}$, the Friedmann equation becomes

$$\dot{R}^2 = \frac{8\pi}{3}G\rho R^2 = H_0^2 R_0^{3+3w}\, R^{-(1+3w)}\,, \tag{15.45}$$

where we inserted the definition of $\rho_{\mathrm{cr}} = 3H_0^2/(8\pi G)$. Separating variables we obtain

$$R_0^{-(3+3w)/2}\int_0^{R_0} \mathrm{d}R\, R^{(1+3w)/2} = H_0 \int_0^{t_0} \mathrm{d}t = t_0 H_0 \tag{15.46}$$

and hence the age of the Universe follows as

$$t_0 H_0 = \frac{2}{3+3w} = \begin{cases} 2/3 & \text{for matter} \quad (w=0) \\ 1/2 & \text{for radiation} \quad (w=1/3) \\ \to\infty & \text{for } \Lambda \quad (w=-1) \end{cases} \tag{15.47}$$

Models with $w > -1$ needed a finite time to expand from the initial singularity $R(t=0) = 0$ to the current size R_0, while a Universe with only a Λ has no "beginning".

In models with a hot big-bang, $\rho, T \to \infty$ for $t \to 0$, and we should expect that classical gravity breaks down at some moment t_*. As long as $R \propto t^\alpha$ with $\alpha < 1$, most time elapsed during the last fractions of $t_0 H_0$. Hence our result for the age of the universe does not depend on unknown physics close to the big-bang as long as $w > -1/3$.

If we integrate (15.46) to the arbitrary time t, we obtain the time-dependence of the scale factor,

$$R(t) \propto t^{2/(3+3w)} = \begin{cases} t^{2/3} & \text{for matter} \quad (w=0) \\ t^{1/2} & \text{for radiation} \quad (w=1/3) \\ \exp(t) & \text{for } \Lambda \quad (w=-1) \end{cases} \tag{15.48}$$

15.5 Determining Λ and the curvature R_0 from $\rho_{m,0}$, H_0, q_0

General discussion: We apply now the Friedmann and the acceleration equation to the present time. Thus $\dot{R}_0 = R_0 H_0$, $\ddot{R} = -q_0 H_0^2 R_0$ and we can neglect the pressure term in Eq. (15.39),

$$\frac{\ddot{R}_0}{R_0} = -q_0 H_0^2 = \frac{\Lambda}{3} - \frac{4\pi G}{3}\rho_{m,0}\,. \tag{15.49}$$

Thus we can determine the value of the cosmological constant from the observables $\rho_{m,0}$, H_0 and q_0 via

$$\Lambda = 4\pi G \rho_{m,0} - 3 q_0 H_0^2\,. \tag{15.50}$$

Solving next the Friedmann equation (15.26) for k/R_0^2,

$$\frac{k}{R_0^2} = \frac{8\pi G}{3}\rho_{m,0} + \frac{\Lambda}{3} - H_0^2\,, \tag{15.51}$$

we write $\rho_{m,0} = \Omega_m \rho_{\rm cr}$ and insert Eq. (15.50) for Λ. Then we obtain for the curvature term

$$\frac{k}{R_0^2} = \frac{H_0^2}{2}\left(3\Omega_m - 2q_0 - 2\right)\,. \tag{15.52}$$

Hence the sign of $3\Omega_m - 2q_0 - 2$ decides about the sign of k and thus the curvature of the universe. For a universe without cosmological constant, $\Lambda = 0$, equation (15.50) gives $\Omega_m = 2q_0$ and thus

$$\begin{aligned} k=-1 \quad &\Leftrightarrow \quad \Omega_m < 1 \quad \Leftrightarrow \quad q_0 < 1/2\,, \\ k=0 \quad &\Leftrightarrow \quad \Omega_m = 1 \quad \Leftrightarrow \quad q_0 = 1/2\,, \\ k=+1 \quad &\Leftrightarrow \quad \Omega_m > 1 \quad \Leftrightarrow \quad q_0 > 1/2\,. \end{aligned} \tag{15.53}$$

For a flat universe with $\Lambda = 0$, $\rho_{m,0} = \rho_{\rm cr}$ and $k = 0$,

$$0 = 4\pi G\,\frac{3H_0^2}{8\pi G} + H_0^2(q_0 - 1) = H_0^2\left(\frac{3}{2} + q_0 - 1\right)\,, \tag{15.54}$$

and thus $q_0 = 1/2$. In this special case, $q_0 < 1/2$ means $k = -1$ and thus an infinite space with negative curvature, while a finite space with positive curvature has $q > 1/2$.

Ex.: Comparison with observations: Use the Friedmann equations applied to the present time to derive central values of Λ and k, R_0 from the observables $H_0 \approx (71 \pm 4)$ km/s/Mpc and $\rho_0 = (0.27 \pm 0.04)\rho_{\rm cr}$, and $q_0 = -0.6$. Discuss the allowed range and significance of the values.
We determine first

$$H_0^2 \approx \left(\frac{7.1 \times 10^6 \text{cm}}{\text{s}\; 3.1 \times 10^{24}\text{cm}}\right)^2 \approx 5.2 \times 10^{-36}\text{s}^{-2}$$

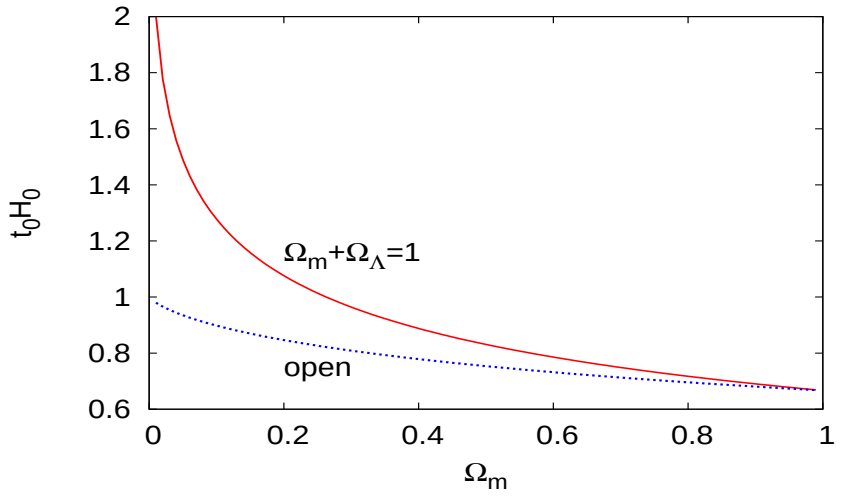


Figure 15.2: The product $t_0 H_0$ for an open universe containing only matter and for a flat cosmological model with $\Omega_\Lambda + \Omega_m = 1$.

The value of the cosmological constant Λ follows as

$$\Lambda = 4\pi G\rho_{m,0} - 3q_0 H_0^2 = 3H_0^2\left(\frac{\rho}{2\rho_{\rm cr}} - 3q_0\right) \approx 3H_0^2 \times (\frac{1}{2} \times 0.27 + 0.6) \approx 0.73 \times 3H_0^2$$

or $\Omega_\Lambda = 0.73$.
The curvature radius R follows as

$$\frac{k}{R_0^2} = 4\pi G\rho_{m,0} - H_0^2(q_0+1) = 3H_0^2\left(\frac{\rho}{2\rho_{\rm cr}} - \frac{q_0+1}{3}\right) \quad (15.55)$$

$$= 3H_0^2\,(0.135 \pm 0.02 - 0.4/3) = 3H_0^2\,(0.002 \pm 0.02) \quad (15.56)$$

thus a flat universe ($k = 0$) is consistent with observations.

Age problem of the universe. The age of a matter-dominated universe is (expanded around $\Omega_0 = 1$)

$$t_0 = \frac{2}{3H_0}\left[1 - \frac{1}{5}(\Omega_0 - 1) + \ldots\right]. \quad (15.57)$$

Globular cluster ages require $t_0 \geq 13$ Gyr. Using $\Omega_0 = 1$ leads to $H_0 \leq 2/3 \times 13\text{Gyr} = 1/19.5\text{Gyr}$ or $h \leq 0.50$. Thus a flat universe with $t_0 = 13$ Gyr without cosmological constant requires a too small value of H_0. Choosing $\Omega_m \approx 0.3$ increases the age by just 14%.

The age t_0 of a flat Universe with $\Omega_m + \Omega_\Lambda = 1$ is given by

$$\frac{3t_0 H_0}{2} = \frac{1}{\sqrt{\Omega_\Lambda}} \ln \frac{1+\sqrt{\Omega_\Lambda}}{\sqrt{1-\Omega_\Lambda}}. \quad (15.58)$$

Requiring $H_0 \geq 65$ km/s/Mpc and $t_0 \geq 13$ Gyr means that the function on the RHS should be larger than $3 \times 13\text{Gyr} \times 0.65/(2 \times 9.8\text{Gyr}) \approx 1.3$ or $\Omega_\Lambda = 0.55$.

15.6 The ΛCDM model

We consider a flat Universe containing as its only two components pressure-less matter and a cosmological constant, $\Omega_m + \Omega_\Lambda = 1$. Then the curvature term in the Friedmann equation and the pressure term in the deceleration equation play no role and we can hope to solve these equations for $a(t)$. Multiplying the deceleration equation (15.39) by two and adding it to the Friedmann equation (15.26), we eliminate ρ_m,

$$2\,\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 = \Lambda\,. \tag{15.59}$$

Next we rewrite first the LHS and then the RHS as total time derivatives: With

$$\frac{\mathrm{d}}{\mathrm{d}t}(a\dot{a}^2) = \dot{a}^3 + 2a\dot{a}\ddot{a} = \dot{a}a^2\left[\left(\frac{\dot{a}}{a}\right)^2 + 2\,\frac{\ddot{a}}{a}\right]\,, \tag{15.60}$$

we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}(a\dot{a}^2) = \dot{a}a^2\Lambda = \frac{1}{3}\,\frac{\mathrm{d}}{\mathrm{d}t}(a^3)\Lambda\,. \tag{15.61}$$

Integrating is now trivial,

$$a\dot{a}^2 = \frac{\Lambda}{3}\,a^3 + C\,. \tag{15.62}$$

The constant C can be determined most easily by setting $a(t_0) = 1$ and comparing the Friedmann equation (15.26) with (15.62) for $t = t_0$ as $C = 8\pi G\rho_{m,0}/3$.

Next we introduce the new variable $x = a^{3/2}$. Then

$$\frac{\mathrm{d}a}{\mathrm{d}t} = \frac{\mathrm{d}x}{\mathrm{d}t}\,\frac{\mathrm{d}a}{\mathrm{d}x} = \frac{\mathrm{d}x}{\mathrm{d}t}\,\frac{2x^{-1/3}}{3}\,, \tag{15.63}$$

and we obtain as new differential equation

$$\dot{x}^2 - \Lambda x^2/4 + C/3 = 0\,. \tag{15.64}$$

Inserting the solution $x(t) = A\sinh(\sqrt{\Lambda}t/2)$ of the homogeneous equation fixes A as $A = \sqrt{3C/\Lambda}$. This can be expressed by the current values of Ω_i as $A = \Omega_m/\Omega_\Lambda = (1-\Omega_\Lambda)/\Omega_\Lambda$. Hence the time-dependence of the scale factor is

$$a(t) = A^{1/3}\sinh^{2/3}(\sqrt{3\Lambda}t/2)\,. \tag{15.65}$$

The time-scale of the expansion is set by $t_\Lambda = 2/\sqrt{3\Lambda}$.

The present age t_0 of the universe follows by setting $a(t_0) = 1$ as

$$t_0 = t_\Lambda \mathrm{arctanh}(\sqrt{\Omega_\Lambda})\,. \tag{15.66}$$

The deceleration parameter $q = -\ddot{a}/aH^2$ is an important quantity for observational tests of the ΛCDM model. We calculate first the Hubble parameter

$$H(t) = \frac{\dot{a}}{a} = \frac{2}{3t_\Lambda}\coth(t/t_\Lambda)\,. \tag{15.67}$$

and then

$$q(t) = \frac{1}{2}[1 - 3\tanh^2(t/t_\Lambda)\,. \tag{15.68}$$

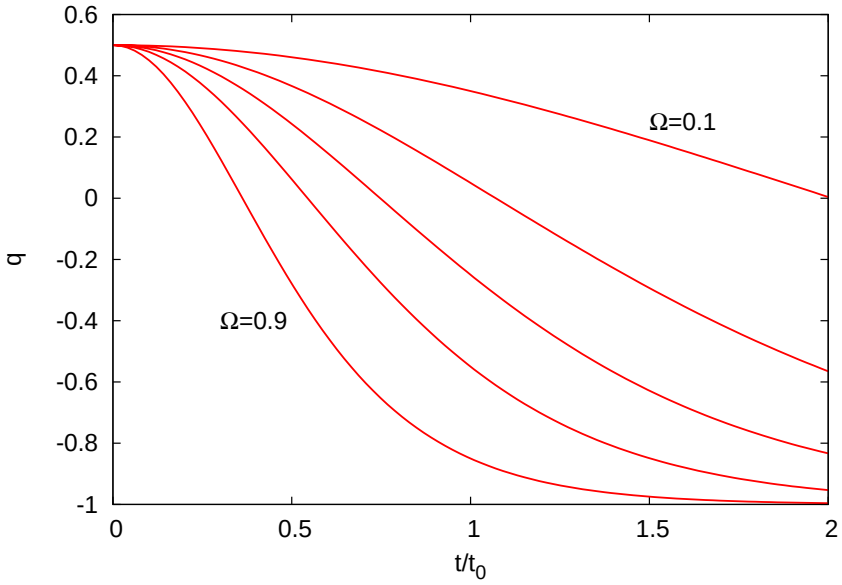


Figure 15.3: The deceleration parameter q as function of t/t_0 for a ΛCDM model and various values for Ω_Λ (0.1, 0.3, 0.5, 0.7 and 0.9 from the top to the bottom).

The limiting behavior of q corresponds with $q = 1/2$ for $t \to 0$ and $q = -1$ for $t \to \infty$ as expected to the one of a flat $\Omega_m = 1$ and a $\Omega_\Lambda = 1$ universe. More interesting is the transition region and, as shown in Fig. 15.3, the transition from a decelerating to an accelerating universe happens for $\Omega_\Lambda = 0.7$ at $t \approx 0.55 t_0$. This can easily converted to redshift, $z_* = a(t_0)/a(t_*) - 1 \approx 0.7$, that is directly measured by Supernova observations.

Exercises

1. Derive the relation between temperature and time in the early (radiation dominated) universe using $\rho = gaT^4 = g\pi^2 T^4/30$ as expression for the energy density of a gas with g relativistic degrees of freedom in the Friedmann equation. What is the temperature at $t = 1$ s? [Hints: The expression $\rho = g\pi^2 T^4/30$ is valid for $k = c = \hbar = 1$. Then one can measure temperatures in MeV/GeV and use $s^{-1} = 6.6 \times 10^{-25}$ GeV and $G^{-1/2} = M_{\rm Pl} = 1.2 \times 10^{19}$ GeV.]

16 Early universe

16.1 Thermal history of the Universe - Time-line of important dates

Different energy form today. Let us summarize the relative importance of the various energy forms today. The critical density $\rho_{\rm cr} = 3H_0^2/(8\pi G)$ has with $h = 0.7$ today the numerical value $\rho_{\rm cr} \approx 7.3 \times 10^{-6}$ GeV/cm^3. This would corresponds to roughly 8 protons per cubic meter. However, main player today is the cosmological constant with $\Omega_\Lambda \approx 0.73$. Next comes (pressure-less) matter with $\Omega_m \approx 0.27$. The energy density of cosmic microwave background (CMB) photons with temperature $T = 2.7\,{\rm K} = 2.3 \times 10^{-4}\,{\rm eV}$ is $\rho_\gamma = aT^4 = 0.4$ eV/cm^3 or $\Omega_\gamma \approx 5 \times 10^{-5}$.

Different energy forms as function of time The scaling of Ω_i with redshift z, $1 + z = R_0/R(t)$ is given by

$$\begin{aligned} H(z)^2/H_0^2 &= \Omega_{m,0}(1+z)^3 + \Omega_{\rm rad,0}(1+z)^4 \\ &+ \Omega_{\Lambda,0} - \underbrace{(\Omega_{\rm tot,0} - 1)(1+z)^2}_{\approx 0} . \end{aligned}$$

Thus the relative importance of the different energy forms changes: Going back in time, one enters first the matter-dominated and then the radiation-dominated epoch.

The cosmic triangle shown in Fig. 16.1 illustrates the evolution in time of the various energy components and the resulting coincidence problem: Any universe with a non-zero cosmological constant will be driven with time to a fix-point with $\Omega_m, \Omega_k \to 0$. The only other constant state is a flat universe containing only matter—however, this solution is unstable. Hence, the question arises why we live in an epoch where all energy components have comparable size.

Temperature increase as $T \sim 1/R$ has three main effects: Firstly, bound states like atoms and nuclei are dissolved when the temperature reaches their binding energy, $kT \gtrsim E_b$. Secondly, particles with mass m_X can be produced, when $kT \gtrsim 2m_X$, in reactions like $\gamma\gamma \to \bar{X}X$. Thus the early Universe consists of a plasma containing more and more heavier particles that are in thermal equilibrium. Finally, reaction rates $\Gamma = n\sigma v$ increases, since $n \propto T^3$. Therefore, reactions that have became ineffective today were important in the early Universe.

Matter-radiation equilibrium $z_{\rm eq}$: The density of matter decreases slower than the energy density of radiation. Going backward in time, there will be therefore a time when the density of matter and radiation were equal. Before that time with redshift $z_{\rm eq}$, the universe was *radiation-dominated*,

$$\Omega_{\rm rad,0}(1+z_{\rm eq})^4 = \Omega_{\rm m,0}(1+z_{\rm eq})^3 \tag{16.1}$$

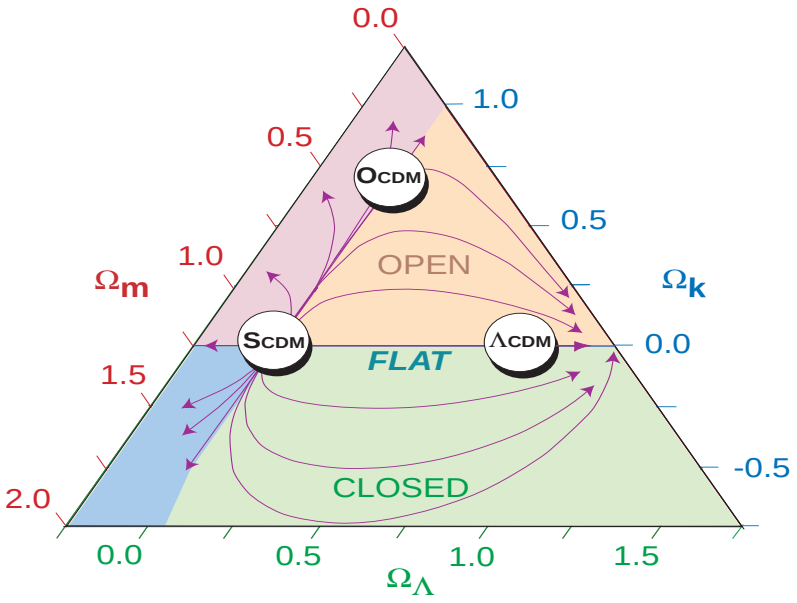


Figure 16.1: The cosmic triangle showing the time evolution of the various energy components.

or

$$z_{\rm eq} = \frac{\Omega_{\rm m,0}}{\Omega_{\rm rad,0}} - 1 \approx 5400\,. \tag{16.2}$$

This time is important, because i) the time-dependence of the scale factor changes from $R \propto t^{2/3}$ for a matter to $R \propto t^{1/2}$ for a radiation dominated universe, ii) the E.o.S. and thus the speed of sound changed from $w \approx 1/3$, $v_s^2 = (\partial P/\partial\rho)_S = c^2/3$ to $w \approx 0$, $v_s^2 = 5kT/(3m) \ll c^2$. The latter quantity determines the Jeans length and thus which structures in the Universe can collapse.

Recombination $z_{\rm rec}$**:** Today, hydrogen and helium in the interstellar and intergalactic medium is neutral. Increasing the temperature, the fraction of ions and free electron increases, i.e. the reaction $H + \gamma \leftrightarrow H^+ + e^-$ that is mainly controlled by the factor $\exp(-E_b/T)$ will be shifted to the right. By definition, we call recombination the time when 50% of all atoms are ionized. A naive estimate gives $T \sim E_b \approx 13.6$ eV$\approx 160.000K$ or $z_{\rm rec} = 60.000$. However, there are many more photons than hydrogen atoms, and therefore recombination happens latter: A more detailed calculation gives $z_{\rm rec} \sim 1000$.
Since the interaction probability of photons with neutral hydrogen is much smaller than with electrons and protons, recombination marks the time when the Universe became transparent to light.

Big Bang Nucleosynthesis At $T_{ns} \sim \Delta \equiv m_n - m_p \approx 1.3$ MeV or $t \sim 1$ s, part of protons and neutrons forms nuclei, mainly ^{4}He. As in the case of recombination, the large number of photons delays nucleosynthesis relative to the estimate $T_{ns} \approx \Delta$ to $T_{ns} \approx 0.1\,$MeV.

Quark-hadron or QCD transition Above $T \sim m_\pi \sim 100$ MeV, hadrons like protons, neutrons or pions dissolve into their fundamental constituents, "quarks" q.

Baryogenesis The fraction baryons (=protons, neutrons, and nuclei) contribute to the critical density is certainly limited by $\Omega_m \sim 0.27$. All the matter observed in the Universe consists of matter (protons and electrons), and not of anti-protons (anti-protons an positrons). Thus the baryon-to-photon ratio is

$$\eta = \frac{n_b - n_{\bar{b}}}{n_\gamma} \lesssim \frac{8 \times 10^{-6}}{400} \sim 10^{-8}\,. \tag{16.3}$$

The early $q\bar{q}$ plasma contained a tiny surplus of quarks. After all anti-matter annihilated with matter, only the small surplus of matter remained. The tiny surplus can be explained by interactions in the early Universe that were not completely symmetric with respect to an exchange of matter-antimatter.

16.2 Big Bang Nucleosynthesis

Nuclear reactions in main sequence stars are supposed to produce all the observed heavier elements up to iron. However, stellar reaction can explain at most a fraction of 5% of ^{4}He, while the production of the weakly bound deuterium and Lithium-7 in stars is impossible. Thus the light elements up to Li-7 are primordial. Their abundance Y is $Y(D) = \text{few} \times 10^{-5}$, $Y(^3H) = \text{few} \times 10^{-5}$ $Y(^4\text{He}) \approx 0.25$, $Y(^7Li) \approx (1-2) \times 10^{-7}$. Observational challenge is to find as "old" stars/gas clouds as possible and then to extrapolate back to the primordial values.

Ex.: Estimate of ^{4}He production by stars:
The binding energy of ^{4}He is $E_b = 28.3$ MeV. If 25% of all nucleons were fussed into ^{4}He during $t \sim 10$ Gyr by stellar fusion, the resulting luminosity-mass ratio would be

$$\frac{L}{M_b} = \frac{1}{4}\,\frac{E_b}{4m_p t} \approx 5\,\frac{\text{erg}}{\text{g s}} \approx 2.5\frac{L_\odot}{M_\odot}\,.$$

The observed luminosity-mass ratio is however only $\frac{L}{M_b} \leq 0.05 L_\odot/M_\odot$. Assuming a roughly constant luminosity of stars, they can produce only $0.05/2.5 \approx 2\%$ of the observed ^{4}He.

Big Bang Nucleosynthesis (BBN) is controlled by two parameters: The mass difference between protons and neutrons, $\Delta \equiv m_n - m_p \approx 1.3$ MeV and the freeze-out temperature T_f of reactions converting protons into neutrons and vice versa. We will treat BBN as a two-step process: First, the weak processes freeze-out and the n/p ratio becomes fixed. Still, the temperature is high compared to E_b and nuclei cannot form. Meanwhile, neutrons decay. After the temperature dropped from T_f to T_NS, most remaining neutrons are built into ^{4}He nuclei.

Gamov criterion We introduced the optical depth $\tau = nl\sigma$ as the probability that a photon interacts in a slab of length l filled with targets of density n. If $\tau \gg 1$, photons and targets interact efficiently and are in thermal equilibrium. We can apply the same criteria to the Universe: We say a particle species A is in thermal equilibrium, as long as $\tau = nl\sigma =$

$n\sigma v\tau \gg 1$. The time τ corresponds to the typical time-scale for the expansion of the universe, $\tau = (\dot{R}/R)^{-1} = H^{-1}$. (Since $T \propto 1/R$ and thus $\Rightarrow \dot{T}/T = -\dot{R}/R = -H$, the time-scale for changes of the expansion rate and of the temperature are the same.) Hence we can rewrite this condition as

$$\Gamma \equiv n\sigma v \gg H\,. \tag{16.4}$$

A particle species "goes out of equilibrium" when its interaction rate Γ becomes smaller than the expansion rate H of the universe.

Decoupling of neutrinos The cross section of reactions involving neutrinos in processes like $n \leftrightarrow p + e^- + \nu_e$ or $e^+e^- \leftrightarrow \bar{\nu}\nu$ is $\sigma \sim G_F^2 E^2$. If we approximate the energy of all particle species by their temperature T, their velocity by c and their density by $n \sim T^3$, then the interaction rate of weak processes is

$$\Gamma \approx \langle v\sigma n_\nu \rangle \approx G_F^2 T^5\,. \tag{16.5}$$

The early universe is radiation-dominated, $\rho_{\rm rad} \propto 1/R^4$, and its curvature k/R^2 is negligible. Thus the Friedmann equation simplifies to $H^2 = (8\pi/3)G\rho$ with $\rho = g_* a T^4$, where g counts the number of relativistic degrees of freedom (photons, neutrinos, electrons) at that time. Requiring $\Gamma(T_{\rm fr}) = H(T_{\rm fr})$ gives as freeze-out temperature $T_{\rm fr}$ of weak processes $T_{\rm fr} \approx 1\,\text{MeV}$, at time $t_{\rm fr} \approx 0.3\,\text{s}$.

Thus the time-sequence is as follows

- At $T_{\rm fr} \approx 1$ MeV, the neutron-proton ratio freezes-in and can be approximated by the ratio of their equilibrium distribution in the non-relativistic limit.
- As the universe cools down from $T_{\rm fr}$ to $T_{\rm ns}$, neutrons decay with half-live $\tau_n \approx 886\,\text{s}$.
- At $T_{\rm ns} \approx 0.1\,\text{MeV}$, practically all neutrons are bound to ^{4}He, with only a small admixture of other elements.

Proton-neutron ratio Above $T_{\rm fr}$, reactions like $\nu_e + n \leftrightarrow p + e^-$ keep nucleons in thermal equilibrium. As we have seen, $T_{\rm fr} \sim 1\,\text{MeV}$ and thus we can treat nucleons in the non-relativistic limit. Then their relative abundance is given by the Boltzmann factor $\exp(-\Delta/T)$ for $T \gtrsim T_{\rm fr}$. Hence for $T_{\rm fr}$,

$$\left.\frac{n_n}{n_p}\right|_{t=t_{\rm fr}} = \exp\left(-\frac{\Delta}{T_f}\right) \approx \frac{1}{6}\,. \tag{16.6}$$

As the universe cools down to $T_{\rm ns}$, neutrons decay with half-live $\tau_n \approx 886\,\text{s}$, and thus

$$\left.\frac{n_n}{n_p}\right|_{t=t_{\rm ns}} \approx \frac{1}{6}\exp\left(-\frac{t_{\rm ns}}{\tau_n}\right) \approx \frac{2}{15}\,. \tag{16.7}$$

Estimate of helium abundance The synthesis if ^{4}He proceeds though a chain of reactions, $pn \to d\gamma$, $dp \to {}^3\text{He}\gamma$, $d\,{}^3\text{He} \to {}^4\text{He}p$.

Let us assume that ^{4}He formation takes place instantaneously. Moreover, we assume that all neutrons are bound in ^{4}He. We need two neutrons to form one helium atom, $n(^4\text{He}) = n_n/2$, and thus

$$Y(^4\text{He}) \equiv \frac{M(^4\text{He})}{M_{\rm tot}} = \frac{4m_N \times n_n/2}{m_N(n_p + n_n)} = \frac{2n_n/n_p}{1 + n_n/n_p} = \frac{4/15}{15/15 + 2/15} = \frac{4}{17} \approx 0.235\,. \tag{16.8}$$

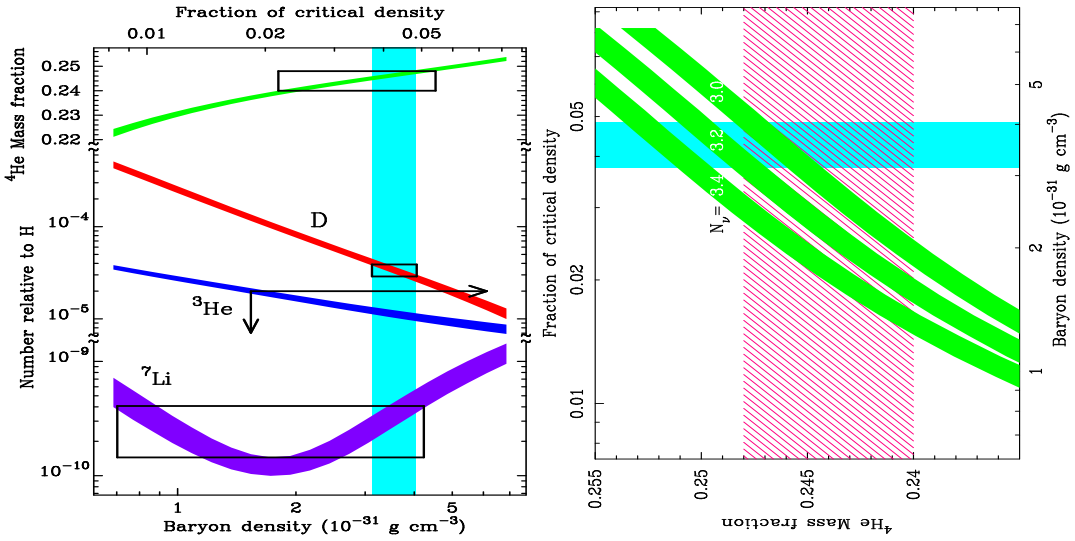


Figure 16.2: Abundances of light elements as function of η (left) and of the number of light neutrino species (right).

Our simple estimate is not too far away from $Y \approx 0.245$.

The abundance of $Y(^4\mathrm{He})$ depends mainly on $\exp(-\Delta/T_f)$: The freeze-out temperature T_{fr} depends in turn on the number of relativistic particles at $t \sim 1\,\mathrm{s}$ and was used as a method to count the number of different light neutrino flavors, cf. right panel of Fig. 16.2. Additionally, there is a weaker dependence on the start of nucleosynthesis T_{ns} and thus η_b or Ω_b.

Results from detailed calculations Detailed calculations predict not only the relative amount of light elements produced, but also their absolute amount as function of e.g. the baryon-photon ratio η. Requiring that the relative fraction of helium-4, deuterium and lithium-7 compared to hydrogen is consistent with observations allows one to determine η or equivalently the baryon content, $\Omega_b h^2 = 0.019 \pm 0.001$. Note that $\Omega_b \ll \Omega_m$ and hence BBN shows that dark matter cannot be in form of baryons (as e.g. the MACHOs hypothesis assumes).

Although the binding energy per nucleon of Carbon-12 and oxygen-16 is higher than the of ^{4}He, these heavier elements are not produced in BBN: At the time of ^{4}He production, the Coulomb barrier prevents already fusion with such large electric charges. Also, a stable intermediate element with $A = 5$ is missing.

16.3 Structure formation

The left panel of Figure 16.3 shows the distribution of matter obtained in a numerical simulation, while the right panel shows the observed distribution of galaxies from an astronomical survey. On both panels, galaxies are distributed in a honeycomb-like structure: There are voids visible that are essentially free of matter, while most galaxies are concentrated at the corners of a web-like structure and along the connections between the corners, so-called filaments.

How can we transform these visual impressions into testable mathematical expressions?

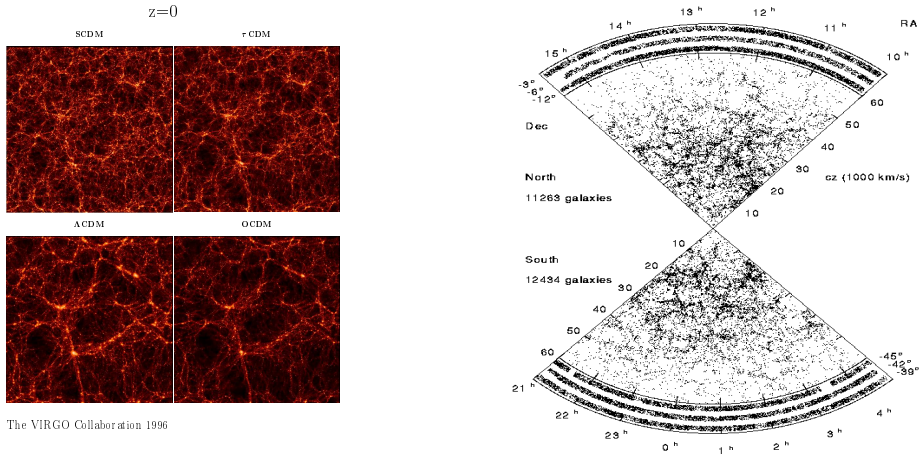


Figure 16.3: Distribution of galaxies found in numerical simulations for four different cosmological scenarios (left) and from observations (right).

Or more precisely, how can we test which one of the four different numerical simulation on the left agrees best with observations? In order to make such a comparison, one employs a statistical description, i.e. one asks questions like how big is the fraction of voids with volume between V_1 and V_2, or how likely is it to find a second galaxies separated by distance r from a first one?

A brief summary of the method is as follows:

- Structure formation operates via gravitational instability, as we discussed for star formation, and needs as starting point a seed of primordial fluctuations.
- One derives from observations the power spectrum $P(k) = |\delta_k|^2$ of density fluctuations today, $\delta(x) \equiv (\rho(x) - \bar{\rho})/\bar{\rho}$:

$$\delta_k \propto \int d^3x \, e^{-ikx} \delta(x) \, .$$

- Assuming initial fluctuations $P_i(k)$, one can calculate how they are transformed by gravitational instability and diffusion of different particle species into a power-spectrum at present time,

$$P(k) = T(k) P_i(k) \, .$$

  Then one can check for which cosmological parameters (H_0, Ω_i,...) and initial spectrum $P_i(k)$ one finds agreement between observations and predictions.
- The Jeans radius of radiation, $P = \rho/3$ equals with $v_s^2 = c^2/3$ approximately the Hubble radius. Since fluctuations with wave-length shorter than the Jeans radius oscillate ("acoustic oscillations"), and only modes with longer wave lengths can grow, structure formations becomes effective only in the matter-dominated epoch.

16.4 Cosmic microwave background

Blackbody radiation in an expanding universe

We discussed earlier that the temperate of photons decreases as $T \sim 1/R$ in an expanding universe. Now we want to justify that the expansion preserves the thermal spectrum in the absence of particle interactions. If reactions like $\bar{f}f \to 2\gamma$ are absent or negligible, the total number N of photons is conserved and the number density $n = N/V$ behaves as $n \propto 1/R^3$.

From the Kirchhoff-Planck distribution

$$B_\nu d\nu = \frac{2h\nu^3}{c^2} \frac{1}{\mathrm{e}^{\frac{h\nu}{kT}} - 1} d\nu \tag{16.9}$$

describing the energy of photons emitted per time and area by a body in thermal equilibrium, we obtain the number density n of photons as $n(\nu, T)d\nu = 4\pi/(ch\nu)B_\nu(\nu, T)d\nu$.

Assume that at a certain time with scale factor R the distribution of photons was initially thermal. Then the number $\mathrm{d}N$ of photons in the volume V and the frequency interval $[\nu : \nu + d\nu]$ is given by

$$\mathrm{d}N = V \, \frac{4}{ch\nu} \, B_\nu d\nu = \frac{8V}{c^3} \frac{h\nu^2 \, d\nu}{\exp\left(\frac{h\nu}{kT}\right) - 1} \,. \tag{16.10}$$

At a different time with scale factor R', the frequency of a single photon scales as $\nu' \propto 1/R'$ and thus also the frequency interval $d\nu$ as $d\nu' \propto 1/R'$. Thus the expression for $\mathrm{d}N$ becomes

$$\mathrm{d}N' = \frac{8V(R'/R)^3}{c^3} \frac{h\nu^2 (R/R')^2 \, d\nu (R/R')}{\exp\left(\frac{h\nu(R/R')}{kT'}\right) - 1} = \frac{8V}{c^3} \frac{h\nu^2 \, d\nu}{\exp\left(\frac{h\nu(R/R')}{kT'}\right) - 1} \,. \tag{16.11}$$

The last expression agrees with Eq. (16.10) as required by photon number conservation, $\mathrm{d}N = \mathrm{d}N'$, if

$$\frac{1}{T} = \frac{R}{R'} \, \frac{1}{T'} \qquad \text{or} \quad T = \frac{R'}{R} \, T' \,. \tag{16.12}$$

Thus a distribution of photons remains thermal, but changes its temperature as $T \propto 1/R$.

CMB and its dipole anisotropy

A background of thermal photons left over from the big-bang was first predicted by Gamov and others. In 1964, the cosmic microwave background was detected and subsequently its isotropy and deviations from a thermal spectrum were searched for. Apart from a dipole anisotropy induced by the relative motion of the Sun with 370 km/h relative to the CMB, the temperature differences are $\Delta T/T \sim 10^{-5}$ in different directions of the sky. In each direction, one measures a perfect blackbody spectrum.

Causality and the isotropy of the CMB

The CMB is to a high degree isotropic, i.e. two points on the sky have the same temperature. Let us estimate the maximal angular separation of a region that was in causally connected at the time of recombination. Two points with angular distance ϑ on the surface of last scattering are separated today by the linear distance $l = d_{\mathrm{ls}}\vartheta$, where $d_{\mathrm{ls}} = c(t_0 - t_{\mathrm{ls}}) \approx ct_0$

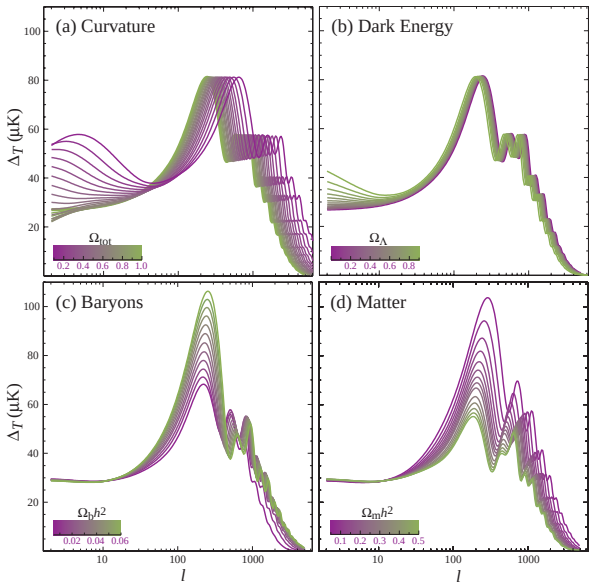


Figure 16.4: The influence of several cosmological parameters on the angular power spectrum of the CMB.

is the distance a photon travelled freely after its last scattering at $t_{\rm ls}$. Thus the maximal angular separation of a causally connected points is with $l = ct_{\rm ls}(1 + z_{\rm ls})$

$$\vartheta = \frac{(1 + z_{\rm ls})t_{\rm ls}}{t_0} \approx 0.02 \approx 1^\circ \tag{16.13}$$

The reason for this "causality problem" is that the universe expands slower than light travels: As the age of the universe increases, the part observable to us increases linearly, $\propto ct$, while the scale factor increases only with $t^{2/3}$ or $(t^{1/2})$. Thus we see more and more regions that were never in causal contact for a radiation or matter-dominated universe.

The sound horizon has approximately the same angular size, because of $v_s \approx c/\sqrt{3}$. The exact size depends among other on the cosmological model: The sound horizon serves as a ruler at fixed redshift z_{ls} to measure the geometry of space-time. Moreover, the fluid of photons and nucleons performs acoustic oscillations with its fundamental frequency connected to the sound horizon plus higher harmonics. The relative size of peaks and locations gives information about cosmological parameters. Figure 16.4 shows the influence of several cosmological parameters on the angular power spectrum as function of $\ell \sim \pi/\vartheta$. The first panel shows that the first peak sits indeed at $\ell \approx 100$ (or $\vartheta \sim 1^\circ$) for a flat Universe, as we have found in our simple estimate (16.13). Observations by the WMAP satellite confirm with high significance the value for Ω_b from BBN, for Ω_Λ from type Ia supernovae, and that we live in a flat Universe.

16.5 Inflation

Shortcomings of the standard big-bang model

- Causality or horizon problem: why are even causally disconnected regions of the universe homogeneous, as we discussed for CMB?

The horizon grows like t, but the scale factor in radiation or matter dominated epoch only as $t^{2/3}$ or $t^{1/2}$, respectively. Thus for any scale l contained today completely inside the horizon, there exists a time $t < t_0$ where it crossed the horizon. A solution to the horizon problem requires that R grows faster than the horizon t. Since $R \propto t^{2/[3(1+w]}$, we need $w < -1/3$ or ($q < 0$, accelerated expansion of the universe).

- Flatness problem: the curvature term in the Friedmann equation is k/R^2. Thus this term decreases slower than matter ($\propto 1/R^3$) or radiation ($1/R^4$), but faster than vacuum energy. Let us rewrite the Friedmann equation as

$$\frac{k}{R^2} = H^2\left(\frac{8\pi G}{3H^2}\rho_0 + \frac{\Lambda}{3H^2} - 1\right) = H^2\left(\Omega_{\mathrm{m+rad}} + \Omega_{\mathrm{rad}} + \Omega_\Lambda - 1\right) = H^2\left(\Omega_{\mathrm{tot}} - 1\right) \tag{16.14}$$

  The LHS scales ar $(1+z)^2$, the RHS for MD $(1+z)^3$ as and for RD as $(1+z)^4$. General relativity is supposed to be valid until the energy scale M_{Pl}. Most of time was RD, so we can estimate $1 + z_{\mathrm{Pl}} = (t_0/t_{\mathrm{Pl}})^{1/2} \sim 10^{30}$ ($t_{\mathrm{Pl}} \sim 10^{-43}$s). Thus if today $|\Omega_{\mathrm{tot}} - 1| \lesssim 1\%$, then the deviation had too be extremely small at t_{Pl}, $|\Omega_{\mathrm{tot}}-1| \lesssim 10^{-2}/(1+z_{\mathrm{Pl}})^2 \approx 10^{-62}$!

  Taking the time-derivative of

$$|\Omega_{\mathrm{tot}} - 1| = \frac{k}{\dot{R}^2} = \frac{|k|}{H^2R^2} \tag{16.15}$$

  gives

$$\frac{d}{dt}\,|\Omega_{\mathrm{tot}} - 1| = \frac{d}{dt}\,\frac{|k|}{\dot{R}^2} = -\frac{2|k|\ddot{R}}{\dot{R}^3} < 0 \tag{16.16}$$

  for $\ddot{R} > 0$. Thus $\Omega_{\mathrm{tot}} - 1$ increases if the universe decelerates, i.e. $\dot{R}$ decreases (radiation/matter dominates), and decreases if the universe accelerates , i.e. $\dot{R}$ increases (or vacuum energy dominates).
- The standard big-bang model contains no source for the initial fluctuations required for structure formation.

Solution by inflation Inflation is a modification of the standard big-bang model where a phase of accelerated expansion in the very early universe is introduced. For the expansion a field called inflaton with E.o.S $w < -1/3$ is responsible. We discuss briefly how the inflation solves the short-comings of standard big-bang model for the special case $w = -1$:

- Horizon problem: In contrast to the radiation or matter-dominated phase, the scale factor grows during inflation faster than the horizon scale, $R(t_2)/R(t_1) = \exp[(t_2 - t_1)H] \gg t_2/t_1$. Thus one can blow-up a small, at time t_1 causally connected region, to superhorizon scales.
- Flatness problem: During inflation $\dot{R} = R\exp(Ht)$ and thus

$$\Omega_{\mathrm{tot}} - 1 = \frac{k}{\dot{R}^2} = k\exp(-2Ht) \tag{16.17}$$

  Thus $\Omega_{\mathrm{tot}} - 1$ exponentially drives towards zero.

- Inflation blows-up quantum fluctuation to astronomical scales, generating initial fluctuation without scale, $P_0(k) = k_s^n$ with $n_s \approx 1$, as required by observations.

A Some formulae

A.1 Mathematical formulae

The following integrals frequently appear in the context of calculations involving particle reactions in thermal media, where ζ refers to the Riemann zeta function.

Table A.1: Thermal integrals.

	Maxwell–Boltzmann	Fermi–Dirac	Bose–Einstein
	$\int_0^\infty \frac{x^n \,\mathrm{d}x}{e^x}$	$\int_0^\infty \frac{x^n \,\mathrm{d}x}{e^x+1}$	$\int_0^\infty \frac{x^n \,\mathrm{d}x}{e^x-1}$
$n=2$	2	$\frac{3}{2}\zeta_3 \simeq 1.8031$	$2\zeta_3 \simeq 2.40411$
$n=3$	6	$\frac{7\pi^4}{120} \simeq 5.6822$	$\frac{\pi^4}{15} \simeq 6.4939$

A.2 Some formulae from cosmology

Redshift-time relations

$$\frac{dz}{dt} = -(1+z)H$$

matter-dominated universe:

$$t = \frac{2}{3H_0}(1+z)^{-3/2}\Omega_0^{-1/2}$$

radiation-dominated universe:

$$t = \frac{1}{2H_0}(1+z)^{-2}\Omega_0^{-1/2}$$

B Units and useful constants

B.1 SI versus cgs units

Doing only mechanics, the difference between the SI and the cgs system is trivial: The first one uses kg, m and s as basic units, while the latter is based on g, cm and s. Thus derived units like energy, power, etc. differ just by powers of ten. As an example compare the energy units Joule and erg in the two systems: $\mathrm{J} = 1\,\mathrm{Nm} = 1\,\mathrm{kgm}^2/\mathrm{s}^2 = 10^3\mathrm{g}10^4\mathrm{cm}^2/\mathrm{s} = 10^7\,\mathrm{erg}$.

Electromagnetism adds as a fundamental property of a particle its charge e. However, one can choose if one expresses the charge through the mechanical units or adds a fourth independent basic unit. This freedom can be seen e.g. in the Coulomb law between two charges,

$$F = k\,\frac{e^2}{r^2}\,, \tag{B.1}$$

where clearly only the product ke^2 is measurable. In the Gaussian cgs system one sets $k = 1$ and uses Eq. (B.1) as definition of the electric charge: Two charges with the "electrostatic charge unit esu" equal to one at the distance $r = 1$cm repel each other with the force 1 dyn.

In the SI system, one defines the current via the force between two parallel, infinitely long conductors: They attract/repel each other being at the distance 1 m with the force 2×10^7 N, if the current 1 Ampere flows in each of them. The charge 1 C follows as Ampere times second. The value of k has to be fixed experimentally and is normally expressed through $\epsilon_0 = 1/(4\pi k_0) \approx 8.85 \times 10^{-12} C^2/(Nm^2)$. Hence, the Coulomb law in SI units is

$$F = \frac{1}{4\pi\epsilon_0}\,\frac{e^2}{r^2}\,. \tag{B.2}$$

The two charge units are connected by

$$1\,\mathrm{C} = 2.998 \times 10^9\,\mathrm{e.s.u.} \tag{B.3}$$

As a last complication, there exist two variants of the cgs system, the Gauß and the Lorentz-Heaviside systems. They differ by the appearance of the factor 4π in Maxwell's equations (Gauß) or of $k = 1/4\pi$ in the Coulomb law.

Finally, note that it is sufficient to know the numerical value of Sommerfeld's finestructure constant, $\alpha \approx 1/137$. In Gauß cgs units, $\alpha = e^2$, in Lorentz-Heaviside cgs units (used in this notes) $\alpha = e^2/(4\pi)$ and in SI units $\alpha = e^2/(4\pi\epsilon_0\hbar c)$.

B.2 Natural units

Special relativity, the wave-particle dualism of quantum mechanics, and statistical mechanics relate many units. For instance, a particle with mass $m = 1$ g has the rest energy $E = mc^2 = (3 \times 10^{10}\mathrm{cm/s})^2 \times 1g = 9 \times 10^{20}$ erg. The associated temperature is $E = kT$, the length scales

is the Compton wave length $\lambda_C = \hbar/mc$, time λ_C/c, etc.
In any calculations, it's possible to set $c = \hbar = k = 1$ and in the end either to the right powers of this constant by dimensional analysis, or or to use directly the values derived above.

B.3 Physical constants and measurements

Gravitational constant	$G = 6.674 \times 10^{-11} \mathrm{m^3\,kg^{-1}\,s^{-2}} = 6.674 \times 10^{-8} \mathrm{cm^3\,g^{-1}\,s^{-2}}$
Planck's constant	$\hbar = h/(2\pi) = 1.055 \times 10^{-27} \mathrm{erg\,s}$
velocity of light	$c = 2.998 \times 10^{10}$ cm/s
Boltzmann constant	$k = 1.381 \times 10^{-23}$ J/K $= 1.38 \times 10^{-16}$ erg/K
electron mass	$m_e = 9.109 \times 10^{-28}\,\mathrm{g} = 0.5110\,\mathrm{MeV}/c^2$
proton mass	$m_p = 1.673 \times 10^{-24}\,\mathrm{g} = 938.3\,\mathrm{MeV}/c^2$
Fine-structure constant	$\alpha = e^2/(\hbar c) \approx 1/137.0$
Fermi's constant	$G_F/(\hbar c)^3 = 1.166 \times 10^{-5}$ GeV^{-2}
Stefan-Boltzmann constant	$\sigma = (2\pi^5 k^4)/(15c^2h^3) \approx 5.670 \times 10^{-5}$ erg s^{-1} cm^{-2} K^{-4}
Radiation constant	$a = 4\sigma/c \approx 7.566 \times 10^{-15}$ erg cm^{-3} K^{-4}
Rydberg constant	$R_\infty = 1.10 \times 10^5$ cm^{-1}
Thomson cross-section	$\sigma_T = 8\pi\alpha_{\mathrm{em}}^2/(3m_e^2) = 6.652 \times 10^{-25}$ cm^2

B.4 Astronomical constants and measurements

Astronomical Unit	$\mathrm{AU} = 1.496 \times 10^{13}\,\mathrm{cm}$
Parsec	$\mathrm{pc} = 3.086 \times 10^{18}\,\mathrm{cm} = 3.261\,\mathrm{ly}$
Tropical year	$\mathrm{yr} = 31\,556\,925.2\,\mathrm{s} \approx \pi \times 10^7\,\mathrm{s}$
Solar radius	$R_\odot = 6.960 \times 10^{10}$ cm
Solar mass	$M_\odot = 1.998 \times 10^{33}$ g
Solar luminosity	$L_\odot = 3.84 \times 10^{33}$ erg/s
Solar apparent visual magnitude	$m = -26.76$
Earth equatorial radius	$R_\oplus = 6.378 \times 10^5$ cm
Earth mass	$M_\oplus = 5.972 \times 10^{27}$ g
Age of the universe	$t_0 = (13.7 \pm 0.2)$ Gyr
present Hubble parameter	$H_0 = 73\,\mathrm{km/(s\,Mpc)} = 100\,h\,\mathrm{km/(s\,Mpc)}$
present CMB temperature	$T = 2.725$ K
present baryon density	$n_\mathrm{b} = (2.5 \pm 0.1) \times 10^{-7}$ cm^3
	$\Omega_b = \rho_b/\rho_\mathrm{cr} = 0.0223/h^2 \approx 0.0425$
dark matter abundance	$\Omega_\mathrm{DM} = \Omega_m - \Omega_b = 0.105/h^2 \approx 0.20$

B.5 Other useful quantities

cross section	1 mbarn $= 10^{-27}$ cm^2
flux conversion	$L = 3.02 \times 10^{28}\mathrm{W} \times 10^{-0.4M}$

B.6 Abbreviations:

B	Kirchoff-Planck function
E	energy of a single object
$\mathcal{F}$	energy flux $E/(At)$, the brightness of a star observed on Earth
J	angular momentum $J = mrv$
L	luminosity, emitted energy per time
m	mass (of a single object), apparent magnitude
M	mass of a system of objects, absolute magnitude
p	momentum
P	pressure
σ	cross section
T	temperature
U	energy of a system of objects

B.7 Properties of main-sequence stars

The spectral class, absolute visual magnitude M_V color index B-V, effective surface temperature radius T, lifetime on the main sequence, and the fraction of the spectral class out of all stars is given in the following table:

SK	M_V	B-V	T/K	lifetime/yr	fraction
O5	-6	-0.45	35000	$< 10^6$	10^{-5}
B0	-3.7	-0.31	21000	3×10^6	10^{-3}
B5	-0.9	-0.17	13500		
A0	0.7	0.0	9700	4×10^8	0.01
A5	2.0	0.16	8100		
F0	2.8	0.30	7200	4×10^9	0.02
F5	3.8	0.45	6500		
G0	4.6	0.57	6000	1×10^{10}	7%
G5	5.2	0.70	5400		
K0	6.0	?	4700	6×10^{10}	15%
K5	7.4	1.11	4000		
M0	8.9	1.39	3300	$> 10^{11}$	75%
M5	12.0	1.61	2600		

Index