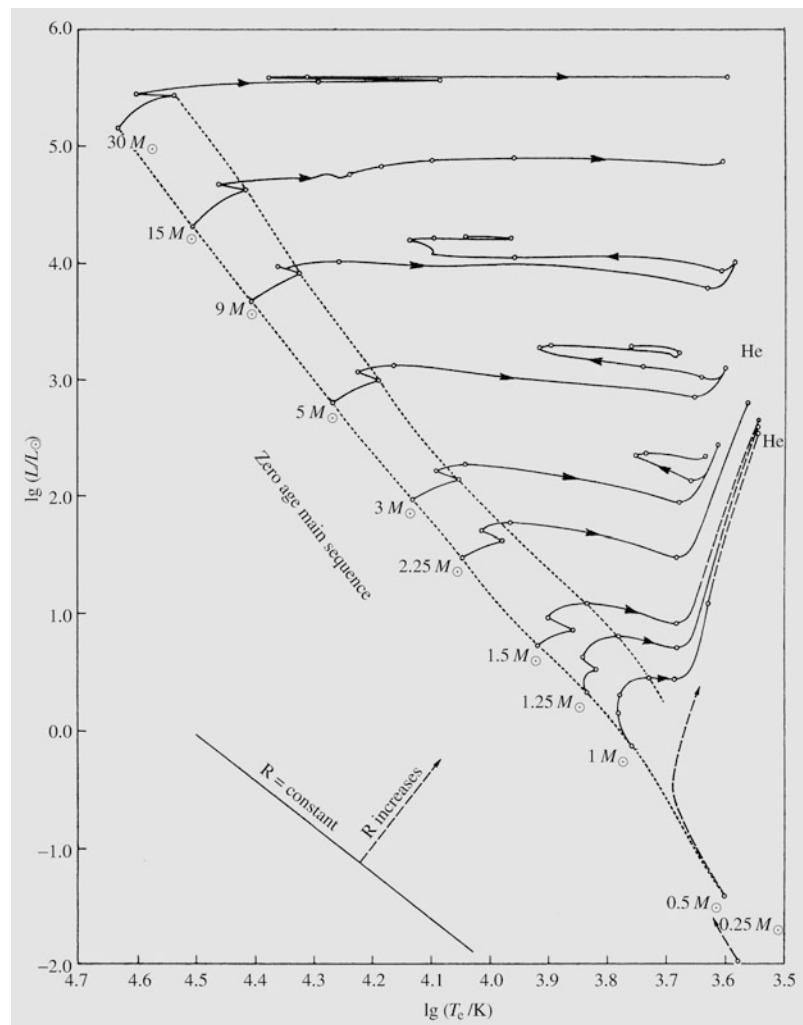


Fig. 12.3 Stellar evolutionary paths in the HR diagram at the main sequence phase and later. On the main sequence, bounded by dashed curves, the evolution is on the nuclear time scale. The post-main sequence evolution to the red giant phase is on the thermal time scale. The point marked He corresponds to helium ignition and in low-mass stars the helium flash. The straight line shows the location of stars with the same radius. (Iben, I. (1967): Annual Rev. Astron. Astrophys. **5**, 571; data for $30 M_{\odot}$ from Stothers, R. (1966): Astrophys. J. **143**, 91)



Outside the core, there is *radiative equilibrium*, i.e. the energy is carried by radiation and there are no nuclear reactions. Between the core and the envelope, there is a transition region where the hydrogen abundance decreases inwards.

The mass of the convective core will gradually diminish as the hydrogen is consumed. In the HR diagram the star will slowly shift to the upper right as its luminosity grows and its surface temperature decreases (Fig. 12.2). When the central hydrogen supply becomes exhausted, the core of the star will begin to shrink rapidly. The surface temperature will increase and the star will quickly move to the upper left. Because of the contraction

of the core, the temperature in the hydrogen shell just outside the core will increase. It rapidly becomes high enough for hydrogen burning to set in again.

The Lower Main Sequence On the lower main sequence, the central temperature is lower than for massive stars, and the energy is generated by the pp chain. Since the rate of the pp chain is not as sensitive to temperature as that of the CNO cycle, the energy production is spread over a larger region than in the more massive stars (Fig. 12.3). In consequence, the core never becomes convectively unstable, but remains radiative.

In the outer layers of lower main sequence stars, the opacity is high because of the low temperature. Radiation can then no longer carry all the energy, and convection will set in. The structure of lower main sequence stars is thus opposite to that of the upper main sequence: the centre is radiative and the envelope is convective. Since there is no mixing of material in the core, the hydrogen is most rapidly consumed at the very centre, and the hydrogen abundance increases outwards.

As the amount of hydrogen in the core decreases, the star will slowly move upwards in the HR diagram, almost along the main sequence (Fig. 12.2). It becomes slightly brighter and hotter, but its radius will not change by much. The evolutionary track of the star will then bend to the right, as hydrogen in the core nears its end. Eventually the core is almost pure helium. Hydrogen will continue to burn in a thick shell around the core.

Stars with masses between $0.08 M_{\odot}$ and $0.26 M_{\odot}$ have a very simple evolution. During their whole main sequence phase they are fully convective, which means that their entire hydrogen content is available as fuel. These stars evolve very slowly toward the upper left in the HR diagram. Finally, when all their hydrogen has burned to helium, they contract to become white dwarfs.

12.4 The Giant Phase

The main-sequence phase of stellar evolution ends when hydrogen is exhausted at the centre. The star then settles in a state in which hydrogen is burning in a shell surrounding a helium core. As we have seen, the transition takes place gradually in lower main-sequence stars, giving rise to the *Subgiant Branch* in the HR diagram, while the upper main-sequence stars make a rapid jump at this point.

The mass of the helium core is increased by the hydrogen burning in the shell. This leads to the expansion of the envelope of the star, which moves almost horizontally to the right in the HR diagram. As the convective envelope becomes more extensive, the star approaches the Hayashi track. Since it cannot pass further to the right, and

since its radius continues to grow, the star has to move upwards along the Hayashi track towards larger luminosities (Fig. 12.2). The star has become a red giant.

In low-mass stars ($M \leq 2.3 M_{\odot}$), as the mass of the core grows, its density will eventually become so high that it becomes degenerate. The central temperature will continue to rise. The whole helium core will have a uniform temperature because of the high conductivity of the degenerate gas. If the mass of the star is larger than $0.26 M_{\odot}$ the central temperature will eventually reach about 100 million degrees, which is enough for helium to burn to carbon in the triple alpha process.

Helium burning will set in simultaneously in the whole central region and will suddenly raise its temperature. Unlike a normal gas, the degenerate core cannot expand, although the temperature increases (cf. (11.16)), and therefore the increase in temperature will only lead to a further acceleration of the rate of the nuclear reactions. When the temperature increases further, the degeneracy of the gas is removed and the core will begin to expand violently. Only a few seconds after the ignition of helium, there is an explosion, the *helium flash*.

The energy from the helium flash is absorbed by the outer layers, and thus it does not lead to the complete disruption of the star. In fact the luminosity of the star drops in the flash, because when the centre expands, the outer layers contract. The energy released in the flash is turned into potential energy of the expanded core. Thus after the helium flash, the star settles into a new state, where helium is steadily burning to carbon in a nondegenerate core.

After the helium flash the star finds itself on the horizontal giant branch in the HR diagram. The exact position of a star on the horizontal branch after the helium flash is a sensitive function of its envelope mass. This in turn depends on the amount of mass lost by the star in the helium flash, which can vary randomly from star to star. While the luminosity does not vary much along the horizontal branch, the effective temperatures are higher for stars with less mass in the envelope. The horizontal branch is divided into

a blue and a red part by a gap corresponding to the pulsational instability leading to RR Lyrae variables (see Sect. 14.2). The form of the horizontal branch for a collection of stars depends on their metal-abundance, in the sense that a lower metal abundance is related to a more prominent blue horizontal branch. Thus the blue horizontal branch in globular clusters with low metal-abundances is strong and prominent (Sect. 16.3). For stars with solar element abundances the horizontal branch is reduced to a short stump, the *red clump*, where it joins the red giant branch.

In intermediate-mass stars ($2.3 M_{\odot} \leq M \leq 8 M_{\odot}$), the central temperature is higher and the central density lower, and the core will therefore not be degenerate. Thus helium burning can set in non-catastrophically as the central regions contract. As the importance of the helium burning core increases, the star first moves away from the red giant branch towards bluer colours, but then loops back towards the Hayashi track again. An important consequence of these *blue loops* is that they bring the star into the strip in the HR diagram corresponding to the cepheid instability (Sect. 14.2). This gives rise to the classical cepheid variables, which are of central importance for determining distances in the Milky Way and to the nearest galaxies.

In the most massive stars helium burning starts before the star has had time to reach the red giant branch. Some stars will continue moving to the right in the HR diagram. For others this will produce a massive stellar wind and a large mass loss. Stars in this evolutionary phase, such as P Cygni and η Carinae, are known as *luminous blue variables*, *LBV*, and are among the brightest in the Milky Way. If the star can retain its envelope it will become a red supergiant. Otherwise it will turn back towards the blue side of the HR diagram, ending up as a Wolf–Rayet star.

The Asymptotic Giant Branch The evolution that follows core helium burning depends strongly on the stellar mass. The mass determines how high the central temperature can become and the degree of degeneracy when heavier nuclear fuels are ignited.

When the central helium supply is exhausted, helium will continue to burn in a shell, while

the hydrogen burning shell is extinguished. In the HR diagram the star will move towards lower effective temperature and higher luminosity. This phase is quite similar to the previous red giant phase of low-mass stars, although the temperatures are slightly hotter. For this reason it is known as the *asymptotic giant branch*, *AGB*.

After the early phase, when the helium shell catches up with the extinguished hydrogen shell, the AGB star enters what is known as the *thermally pulsing phase*, where hydrogen and helium shell burning alternate. A configuration with two burning shells is unstable, and in this phase the stellar material may become mixed or matter may be ejected into space in a shell, like that of a planetary nebula.

The thermally pulsing AGB continues until radiation pressure has led to the complete expulsion of the outer layers into a planetary nebula. Low- and intermediate-mass giants ($M \leq 8 M_{\odot}$) never become hot enough to ignite carbon burning in the core, which remains as a carbon–oxygen white dwarf (Fig. 12.6).

The End of the Giant Phase After the end of helium burning the evolution of a star changes character. This is because the nuclear time scale at the centre becomes short compared to the thermal time scale of the outer layers. Secondly, the energy released in nuclear reactions will be carried away by neutrinos, instead of being deposited at the centre. In consequence, while the thermonuclear burning follows the same pattern as hydrogen and helium burning, the star as a whole does not have time to react immediately.

In stars with masses around $10 M_{\odot}$ either carbon or oxygen may be ignited explosively just like helium in low-mass stars: there is a *carbon or oxygen flash*. This is much more powerful than the helium flash, and may make the star explode as a supernova (Sects. 12.5 and 13.3).

For even larger masses the core remains non-degenerate and burning will start non-catastrophically as the core goes on contracting and becoming hotter. First carbon burning and subsequently oxygen and silicon burning (see Sect. 10.3) will be ignited. As each nuclear fuel

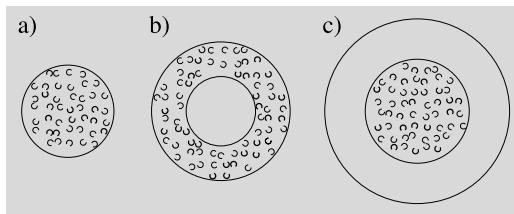


Fig. 12.4 Energy transport in the main sequence phase. (a) The least massive stars ($M < 0.26 M_{\odot}$) are convective throughout. (b) For $0.26 M_{\odot} < M < 1.5 M_{\odot}$ the core is radiative and the envelope convective. (c) Massive stars ($M > 1.5 M_{\odot}$) have a convective core and a radiative envelope

is exhausted in the centre, the burning will continue in a shell. The star will thus contain several nuclear burning shells. At the end the star will consist of a sequence of layers differing in composition, in massive stars (more massive than $15 M_{\odot}$) all the way up to iron.

The central parts of the most massive stars with masses larger than $15 M_{\odot}$ burn all the way to iron ^{56}Fe . All nuclear sources of energy will then be completely exhausted. The structure of a 30 solar mass star at this stage is schematically shown in Fig. 12.4. The star is made up of a nested sequence of zones bounded by shells burning silicon ^{28}Si , oxygen ^{16}O and carbon ^{12}C , helium ^4He and hydrogen ^1H . However, this is not a stable state, since the end of nuclear reactions in the core means that the central pressure will fall, and the core will collapse. Some of the energy released in the collapse goes into dissociating the iron nuclei first to helium and then to protons and neutrons. This will further speed up the collapse, just like the dissociation of molecules speeds up the collapse of a protostar. The collapse takes place on a dynamical time scale, which, in the dense stellar core, is only a fraction of a second. The outer parts will also collapse, but more slowly. In consequence, the temperature will increase in layers containing unburnt nuclear fuel. This will burn explosively, releasing immense amounts of energy in a few seconds, principally in the form of neutrinos.

The final stages of stellar evolution may be described as an implosion of the core, which is

briefly halted every time a new source of nuclear fuel becomes available for burning. It is still an open problem how exactly the energy released in this collapse is transformed into the disruption of the entire star and the ejection of its outer layers. It is also still unclear whether in a given case the remnant will be a neutron star or a black hole.

Although the exact mechanism is not yet understood, the end-point of the evolution of stars more massive than about $8 M_{\odot}$ is that the outer layers explode as a supernova. In the dense central core, the protons and electrons combine to form neutrons. The core will finally consist almost entirely of neutrons, which become degenerate because of the high density. The degeneracy pressure of the neutrons will stop the collapse of a small mass core. However, if the mass of the core is large enough, a black hole will probably be formed.

12.5 The Final Stages of Evolution

The endpoints of stellar evolution can be seen from Fig. 12.7. This shows the relation between mass and central density for a body at zero temperature, i.e. the final equilibrium when a massive body has cooled. There are two maxima on the curve. The mass corresponding to the left-hand maximum is called the *Chandrasekhar mass*, $M_{\text{Ch}} \approx 1.2\text{--}1.4 M_{\odot}$, and that corresponding to the right-hand one, the *Oppenheimer-Volkoff mass*, $M_{\text{OV}} \approx 1.5\text{--}2 M_{\odot}$.

Let us first consider a star with mass less than M_{Ch} . Suppose the mass does not change. When the nuclear fuel is exhausted, the star will become a white dwarf, which will gradually cool down and contract. In Fig. 12.5 it moves horizontally to the right. Finally it will reach zero temperature and end up on the left-hand rising part of the equilibrium curve. Its final equilibrium is a completely degenerate black dwarf.

If the mass of the star is larger than M_{Ch} but smaller than M_{OV} , it can continue cooling until it reaches the right-hand rising section of the curve. Again there is a stable final state, this time corresponding to a completely degenerate neutron star.

An even more massive star with mass larger than M_{OV} will go on contracting past the den-

sity corresponding to a neutron star. There is then no longer any known possible stable equilibrium, and the star must go on contracting to form a black hole.

The only endpoints of stellar evolution predicted by theory are the two stable states of Fig. 12.7 and the two extreme possibilities, collapse to a black hole or explosive disruption.

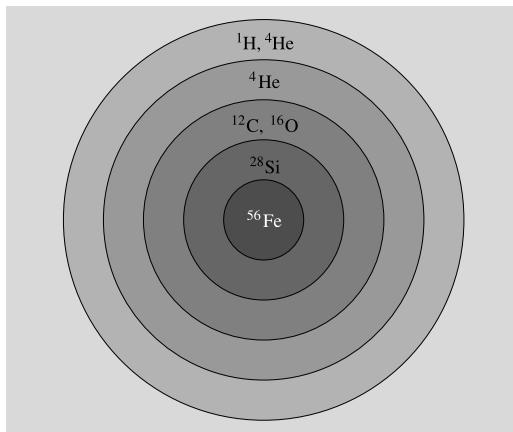


Fig. 12.5 The structure of a massive star ($30 M_{\odot}$) at a late evolutionary stage. The star consists of layers with different composition separated by nuclear burning shells

The preceding considerations are purely theoretical. The final evolutionary stages of real stars involve many imperfectly known factors, which may affect the final equilibrium. Perhaps most important is the question of mass loss, which is very difficult to settle either observationally or theoretically. For example, in a supernova explosion the whole star may be disrupted and it is very uncertain whether what remains will be a neutron star, a black hole or nothing at all. (The structure of compact stars will be discussed in Chap. 15.)

A summary of the various evolutionary paths is given in Fig. 12.8.

12.6 The Evolution of Close Binary Stars

If the components of a binary star are well separated, they do not significantly perturb one another. When studying their evolution, one can regard them as two single stars evolving independently, as described above. However, in close binary pairs, this will no longer be the case.

Close binary stars are divided into three classes, as shown in Fig. 12.9: *detached*, *semidetached* and *contact binaries*. The figure-eight

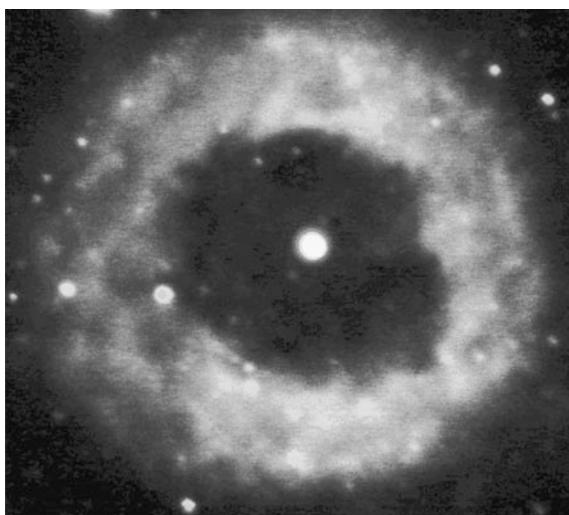
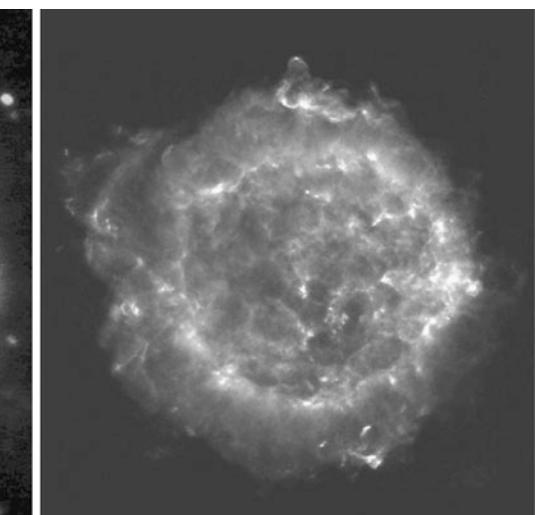


Fig. 12.6 The usual endpoint for the development of a star with a mass of less than three solar masses, is a white dwarf, with an expanding planetary nebula around it. *On the left*, the planetary nebula NGC 6369, photographed with the 8-meter Gemini South tele-



scope. For a massive star, the life ends with a supernova explosion. *On the right*, the supernova remnant Cassiopeia A on radio wavelengths. The image was created by the VLA telescope. (Images Gemini Observatory/Abu Team/NOAO/AURA/NSF and NRAO/AUI)

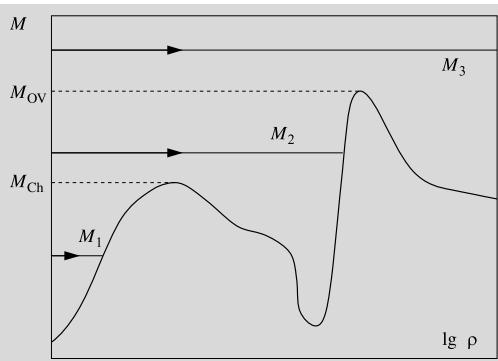


Fig. 12.7 The evolutionary end points of stars with different masses shown as a function of central density. The curve shows the behaviour of the central density of completely degenerate ($T = 0$ K) bodies. The Chandrasekhar mass M_{Ch} and the Oppenheimer–Volkoff mass M_{OV} correspond to maxima on this curve

curve drawn in the figure is an equipotential surface called the *Roche surface*. If the star becomes larger than this surface, it begins to lose mass to its companion through the waist of the Roche surface.

During the main sequence phase the stellar radius does not change much, and each component will remain within its own Roche lobe. When the hydrogen is exhausted, the stellar core will rapidly shrink and the outer layers expand, as we have seen. At this stage a star may exceed its Roche lobe and mass transfer may set in.

Close binary stars are usually seen as eclipsing binaries. One example is Algol in the constellation Perseus. The components in this binary system are a normal main sequence star and a subgiant, which is much less massive than the main sequence star. The subgiant has a high luminosity and thus has apparently already left the main sequence. This is unexpected, since the components were presumably formed at the same time, and the more massive star should evolve more rapidly. The situation is known as the *Algol paradox*: for some reason, the less massive star has evolved more rapidly.

In the 1950's a solution to the paradox proposed that the subgiant was originally more massive, but that it had lost mass to its companion during its evolution. Since the 1960's mass transfer in close binary systems has been much stud-

ied, and has turned out to be a very significant factor in the evolution of close binaries.

As an example, let us consider a close binary, where the initial masses of the components are 1 and 2 solar masses and the initial orbital period 1.4 days (Fig. 12.10). After evolving away from the main sequence the more massive component will exceed the Roche limit and begin to lose mass to its companion. Initially the mass will be transferred on the thermal time scale, and after a few million years the roles of the components will be changed: the initially more massive component has become less massive than its companion.

The binary is now semidetached and can be observed as an Algol-type eclipsing binary. The two components are a more massive main sequence star and a less massive subgiant filling its Roche surface. The mass transfer will continue, but on the much slower nuclear time scale. Finally, mass transfer will cease and the less massive component will contract to a $0.6 M_{\odot}$ white dwarf.

The more massive $2.4 M_{\odot}$ star now evolves and begins to lose mass, which will accumulate on the surface of the white dwarf. The accumulated mass may give rise to *nova outbursts*, where material is ejected into space by large explosions. Despite this, the mass of the white dwarf will gradually grow and may eventually exceed the Chandrasekhar mass. The white dwarf will then collapse and explode as a type I supernova.

As a second example, we can take a massive binary with the initial masses 20 and $8 M_{\odot}$ and the initial period 4.7 days (Fig. 12.11). The more massive component evolves rapidly, and at the end of the main sequence phase, it will transfer more than $15 M_{\odot}$ of its material to the secondary. The mass transfer will occur on the thermal time scale, which, in this case, is only a few ten thousand years. The end result is a *helium star*, having as a companion an unevolved main sequence star. The properties of the helium star are like those of a *Wolf–Rayet star* (Fig. 12.12).

Helium continues to burn to carbon in the core of the helium star, and the mass of the carbon core will grow. Eventually the carbon will be explosively ignited, and the star will explode as a su-

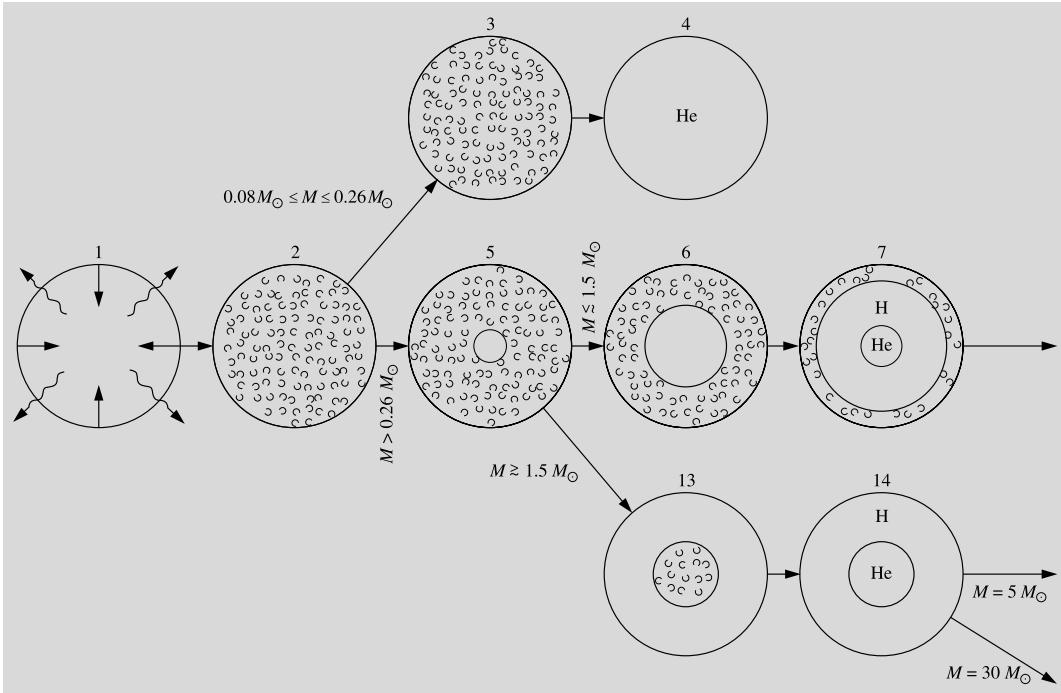


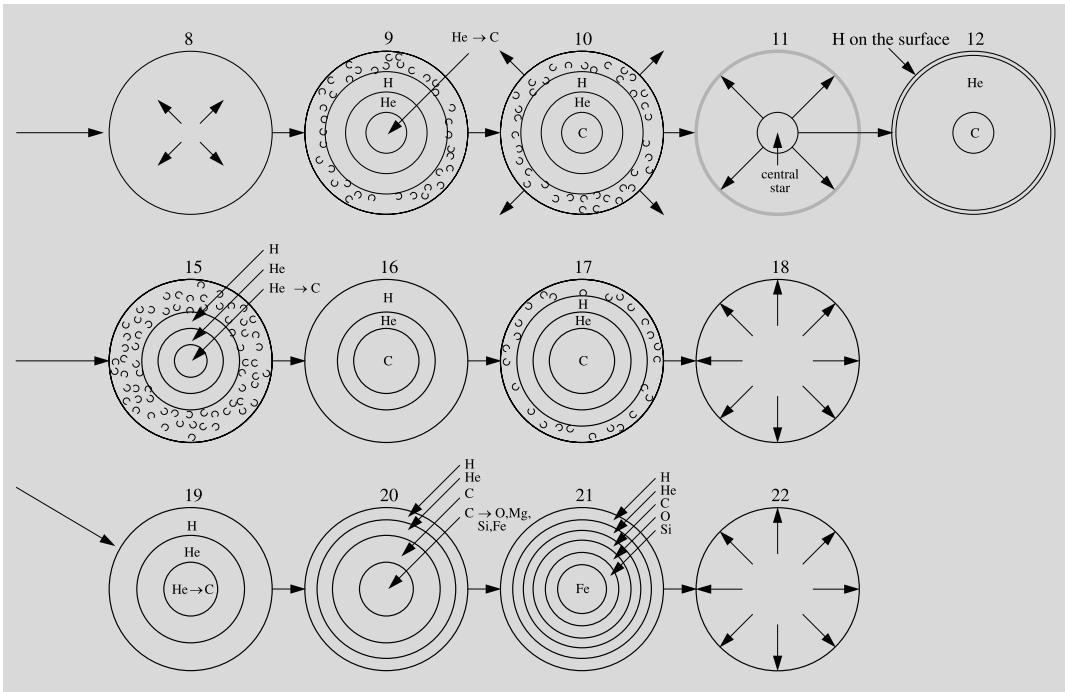
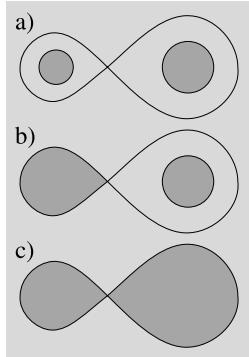
Fig. 12.8 Evolution schemes for stars with different masses. The radius is scaled to be the same in all drawings. In reality, there are vast differences in the sizes of different stars and different phases of evolution. In the beginning (*1*) a gas cloud is contracting rapidly in free fall. Because the gas is quite rarefied, radiation escapes easily from the cloud. As the density increases, radiation transport becomes more difficult, and the released energy tends to warm up the gas. The contraction lasts until the gas is completely ionised, and the star, which has become a protostar, is in hydrostatic equilibrium (*2*). The star is convective throughout its interior.

Now evolution continues on a thermal time scale. The contraction is much slower than in the free-fall phase. The phases of further evolution are determined by the mass M of the star. For $M < 0.08 M_{\odot}$ the temperature in the centre does not rise high enough for hydrogen burning, and these stars contract to planetlike brown dwarfs. Stars with $M \geq 0.08 M_{\odot}$ start hydrogen burning when the temperature has reached about 4×10^6 K. This is the beginning of the main sequence phase. In the main sequence, the lowest-mass stars with $0.08 M_{\odot} \leq M \leq 0.26 M_{\odot}$ are entirely convective, and thus they remain homogeneous (*3*). Their evolution is very slow, and after all the hydrogen has been burnt to helium, they contract to white dwarfs (*4*). The increasing temperature makes the stars with $M > 0.26 M_{\odot}$ radiative in the centre as the opacity decreases (*5*). The low-mass stars with $0.26 M_{\odot} \leq M \leq 1.5 M_{\odot}$ remain radiative in the centre during the main sequence phase (*6*) as they burn their hydrogen through the pp chain. The outer part is convective. At the end of the main sequence phase, hydrogen burning continues in a shell surrounding the helium core (*7*).

The outer part expands, and the giant phase begins. The contracting helium core is degenerate and warms up. At about 10^8 K, the triple alpha process begins and leads immediately to the helium flash (*8*). The explosion is damped by the outer parts, and helium burning goes on in the core (*9*). Hydrogen is still burning in an outer shell. As the central helium is exhausted, helium burning moves over to a shell (*10*). At the same time, the outer part expands and the star loses some of its mass. The expanding envelope forms a planetary nebula (*11*). The star in the centre of the nebula becomes a white dwarf (*12*).

In the upper main sequence with $M \geq 1.5 M_{\odot}$ energy is released through the CNO cycle, and the core becomes convective, while the outer part is radiative (*13*). The main sequence phase ends as the hydrogen in the core is exhausted, and shell burning begins (*14*). The helium core remains convective and nondegenerate, and helium burning begins without perturbations (*15* and *19*). Afterwards, helium burning moves over to a shell (*16* and *20*). For stars with $3 M_{\odot} \leq M \leq 15 M_{\odot}$ the carbon in the core is degenerate, and a carbon flash occurs (*17*). This leads to a supernova explosion (*18*) and possibly to the complete destruction of the star.

For the most massive stars with $M \geq 15 M_{\odot}$ the carbon core remains convective, and carbon burns to oxygen and magnesium. Finally, the star consists of an iron core surrounded by shells with silicon, oxygen, carbon, helium and hydrogen (*21*). The nuclear fuel is now exhausted, and the star collapses on a dynamical time scale. The result is a supernova (*22*). The outer parts explode, but the remaining core continues to contract to a neutron star or a black hole.

**Fig. 12.8** (Continued)**Fig. 12.9** The types of close binary systems: (a) detached, (b) semidetached and (c) contact binary

pernova. The consequences of this explosion are not known, but let us suppose that a $2 M_{\odot}$ compact remnant is left. As the more massive star expands, its stellar wind will become stronger, giving rise to strong X-ray emission as it hits the compact star. This X-ray emission will only cease when the more massive star exceeds its Roche surface.

The system will now rapidly lose mass and angular momentum. A steady state is finally reached when the system contains a $6 M_{\odot}$ helium star in addition to the $2 M_{\odot}$ compact star. The helium star is seen as a Wolf-Rayet star, which, after about a million years, explodes as a supernova. This will probably lead to the breakup of the binary system. However, for certain values of the mass, the binary may remain bound. Thus a binary neutron star may be formed.

12.7 Comparison with Observations

The most important direct support for the theoretical evolutionary models is obtained from the properties of observed HR diagrams. If the theoretical models are correct, the observed number of stars should reflect the duration of the various evolutionary phases. These are given for stars of different masses in Table 12.1. The stars are most numerous along the main sequence. Giants are also common and, in addition to these, there are white dwarfs, subgiants, etc. The sparsely populated region to the right of the main sequence, the

Hertzsprung gap, is explained by the rapid transition from the main sequence to the giant phase.

The cepheids provide an important test for the evolutionary models. The pulsations and the relation between period and luminosity for the cepheids can be understood on the basis of theoretical stellar models.

The evolutionary models can also explain the HR diagrams of star clusters. Let us assume that all the stars in a cluster were formed at the same time. In the youngest systems, the associations, the stars will mainly be found on the upper main

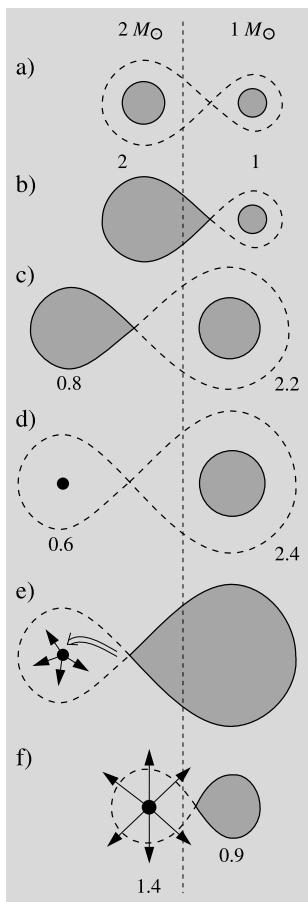


Fig. 12.10 Evolution of a low-mass binary: (a) both components on the main sequence; (b) mass transfer from the more massive component; (c) light subgiant and massive main sequence star; (d) white dwarf and main sequence star; (e) mass transferred to the white dwarf from the more massive component leads to nova outbursts; (f) the white dwarf mass exceeds the Chandrasekhar mass and explodes as a type I supernova

sequence, since the most massive stars evolve most rapidly. To the right of the main sequence, there will be less massive T Tauri stars, which are

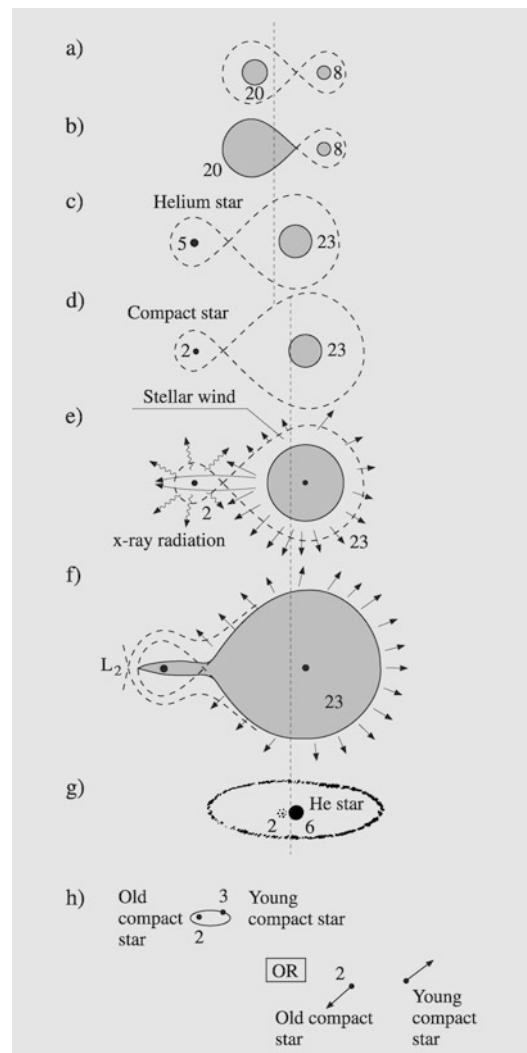


Fig. 12.11 Evolution of a massive binary. It has been assumed that the supernova explosion of a $5 M_{\odot}$ helium star leaves a $2 M_{\odot}$ compact remnant (neutron star or black hole). (a) Main sequence phase; (b) beginning of the first mass transfer phase; (c) end of the first mass transfer phase; the first Wolf-Rayet phase begins; (d) the helium star (Wolf-Rayet star) has exploded as a supernova; (e) the $23 M_{\odot}$ component becomes a supergiant; the compact component is a strong X-ray source; (f) beginning of the second mass transfer phase; the X-ray source is throttled and large-scale mass loss begins; (g) second Wolf-Rayet phase; (h) the $6 M_{\odot}$ helium star has exploded as a supernova; the binary may or may not be disrupted, depending on the remaining mass

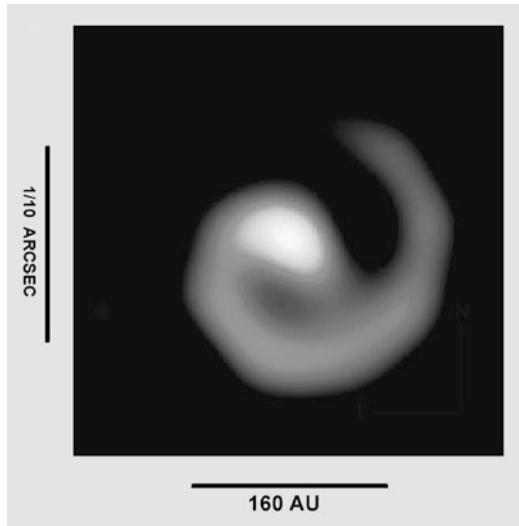


Fig. 12.12 The Wolf-Rayet star WR 104 photographed by the 10-m Keck telescope. The spiral is dust and gas which is thrown out from the rotating binary system. The spiral “pinwheel” is seen to make a full revolution in about 220 days. (Photo U.C. Berkeley Space Sciences Laboratory/W.M. Keck Observatory)

still contracting. In intermediate age open clusters, the main sequence will be well developed and its upper end should bend to the right, since the most massive stars will already have begun to evolve off the main sequence. In the old globular clusters, the giant branch should increase in importance in the older clusters. These predictions are confirmed by the observations, which will be further discussed in Chap. 16 on star clusters.

Of course, the most detailed observations can be made of the Sun, which is therefore a crucial point of comparison for the theoretical models. If a star of one solar mass with an initial composition of 71 % hydrogen, 27 % helium and 2 % heavier elements is allowed to evolve for 5000 million years, it will be very similar to our present Sun. In particular, it will have the same radius, surface temperature and luminosity. According to calculations, about half of the Sun’s supply of hydrogen fuel has been consumed. The Sun will go on shining like a normal main sequence star for another 5000 million years, before there will be any dramatic change.

Some problems remain in regard to the observations. One is the solar neutrino problem. The

neutrinos produced by solar nuclear reactions have been observed since the beginning of the 1970’s by the techniques described in Sect. 3.7. Only the neutrinos formed in the relatively rare ppIII reaction are energetic enough to be observed in this way. Their observed number is too small: whereas the models predict about 5 units, the observations have consistently only registered 1–2.

The discrepancy may be due to a fault in the observational technique or to some unknown properties of the neutrinos. However, if the solar models are really in error, the central temperature of the Sun would have to be about 20 % lower than thought, which would be in serious conflict with the observed solar luminosity. One possibility is that some of the electron neutrinos change to other, unobservable particles during their passage to Earth. (See also Sect. 12.1.)

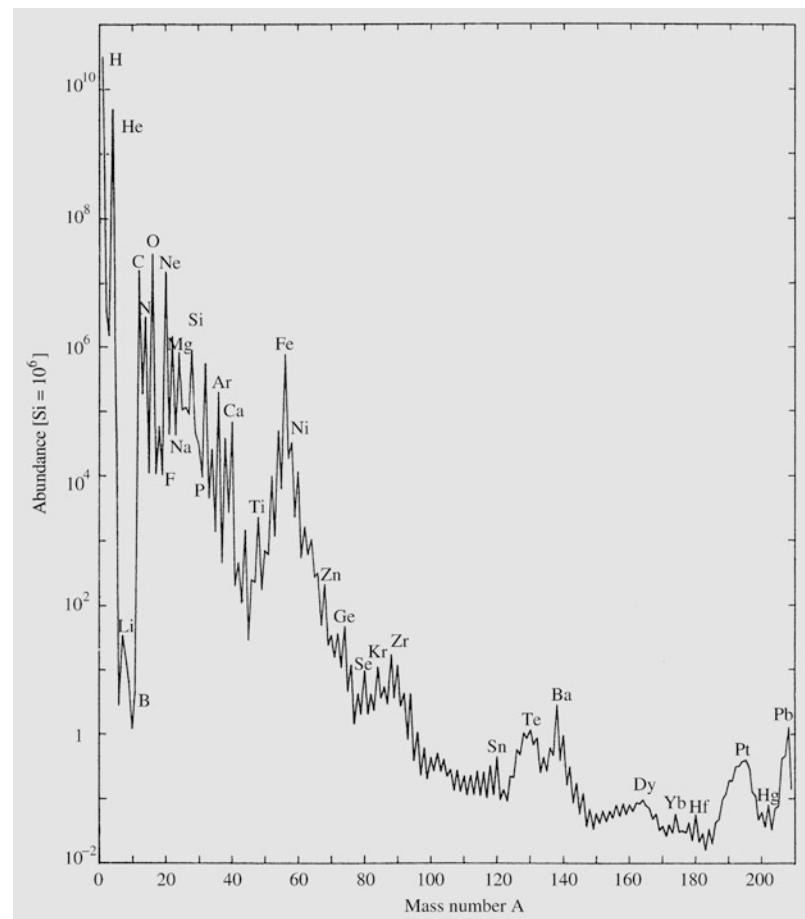
A second problem is the observed abundance of lithium and beryllium. The solar surface contains a normal abundance of beryllium, but very little lithium. This should mean that during its contraction, the Sun was still fully convective when the central temperature was high enough to destroy lithium (3×10^6 K), but not beryllium (4×10^6 K). However, according to the standard solar evolution models, convection ceased in the centre already at a temperature of 2×10^6 K. One suggested explanation is that the convection has later carried down lithium to layers where the temperature is high enough to destroy it.

12.8 The Origin of the Elements

There are just under a hundred naturally occurring elements, and about 300 isotopes in the solar system (Fig. 12.13). In Sect. 12.4, we have seen how the elements up to iron are produced when hydrogen burns to helium and helium further to carbon, oxygen and heavier elements.

Almost all nuclei heavier than helium were produced in nuclear reactions in stellar interiors. In the oldest stars, the mass fraction of heavy elements is only about 0.02 %, whereas in the youngest stars it is a few per cent. Nevertheless, most of the stellar material is hydrogen and

Fig. 12.13 Element abundances in the solar system as a function of the nuclear mass number. The abundance of Si has been normalised as 10^6



helium. According to the standard cosmological model, those were formed in the early stages of the Universe, when the temperature and density were suitable for nuclear reactions. (This will be discussed in Chap. 20.) Although helium is produced during the main sequence stellar evolution, very little of it is actually returned into space to be incorporated into later stellar generations. Most of it is either transformed into heavier elements by further reactions, or else remains locked up inside white dwarf remnants. Therefore the helium abundance does not increase by much due to stellar processes.

The most important nuclear reactions leading to the build-up of the heavy nuclei up to iron were presented in Sect. 11.3. The probabilities of the various reactions are determined either by experiments or by theoretical calculations. When they

are known, the relative abundances of the various nuclei produced can be calculated.

The formation of elements heavier than iron requires an input of energy, and thus they cannot be explained in the same manner. Still heavy nuclei are continually produced. In 1952 technetium was discovered in the atmosphere of a red giant. The half-life of the most longlived isotope ^{98}Tc is about 1.5×10^6 years, so that the observed technetium must have been produced in the star.

Most of the nuclei more massive than iron are formed by *neutron capture* (Fig. 12.14). Since the neutron does not have an electric charge, it can easily penetrate into the nucleus. The probability for neutron capture depends both on the kinetic energy of the incoming neutron and on the mass number of the nucleus. For example, in the solar system the abundances of isotopes show maxima at the mass numbers $A = 70\text{--}90, 130,$

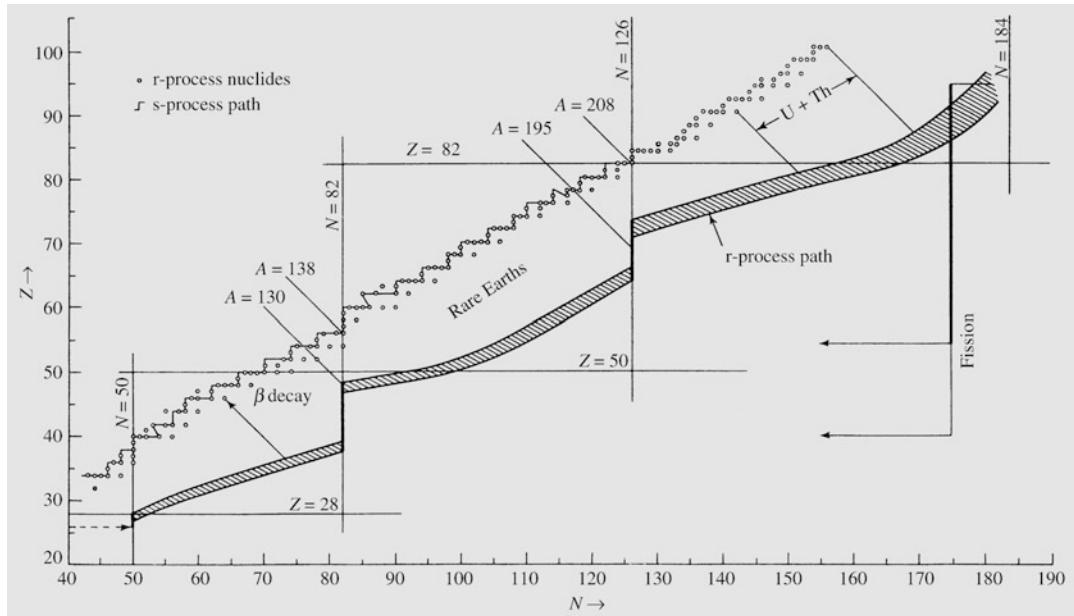
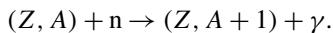


Fig. 12.14 Neutron capture paths for the s-process (from left to right). The s-process follows a path along the line of beta stability. The stable r-process nuclei (small circles) result from beta decay of their neutron rich progenitors on the shaded path shown lower. Beta decay occurs along straight lines $A = \text{const}$. The closed neutron

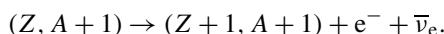
shells in nuclei at $N = 50, 82$ and 126 correspond to abundance peaks in s-process nuclei at $A = 88, 138$ and 208 , and in r-process nuclei at $A = 80, 130$ and 195 . (Seeger, P.A., Fowler, W.A., Clayton, D.D. (1965): *Astrophys. J. Suppl.* **11**, 121)

138, 195 and 208. These mass numbers correspond to nuclei with closed neutron shells at the neutron numbers $N = 50, 82$, and 126 . The neutron capture probability for these nuclei is very small. The closed shell nuclei thus react more slowly and are accumulated in greater abundances.

In a neutron capture, a nucleus with mass number A is transformed into a more massive nucleus:



The newly formed nucleus may be unstable to β decay, where one neutron is transformed into a proton:



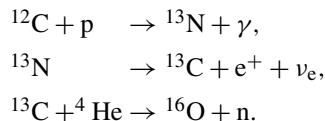
Two kinds of neutron capture processes are encountered, depending on the value of the neutron flux. In the slow *s*-process, the neutron flux is

so small that any β decays have had time to occur before the next neutron capture reaction takes place. The most stable nuclei up to mass number 210 are formed by the *s*-process. These nuclei are said to correspond to the β stability valley. The *s*-process explains the abundance peaks at the mass numbers 88, 138 and 208.

When the neutron flux is large, β decays do not have time to happen before the next neutron capture. One then speaks of the rapid *r*-process, which gives rise to more neutron-rich isotopes. The abundance maxima produced by the *r*-process lie at mass numbers about ten units smaller than those of the *s*-process.

A neutron flux sufficient for the *s*-process is obtained in the course of normal stellar evolution. For example, some of the carbon and oxygen burning reactions produce free neutrons. If there is convection between the hydrogen and helium burning shells, free protons may be carried into the carbon-rich layers. Then the following neutron-producing reaction chain becomes im-

portant:

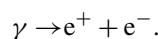


The convection can also carry the reaction products nearer to the surface.

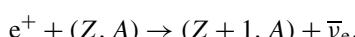
The neutron flux required for the r-process is about 10^{22} cm^{-3} , which is too large to be produced during normal stellar evolution. The only presently known site where a large enough neutron flux is expected is near a neutron star forming in a supernova explosion. In this case, the rapid neutron capture leads to nuclei that cannot capture more neutrons without becoming strongly unstable. After one or more rapid β decays, the process continues.

The r-process stops when the neutron flux decreases. The nuclei produced then gradually decay by the β -process towards more stable isotopes. Since the path of the r-process goes about ten mass units below the stability valley, the abundance peaks produced will fall about ten units below those of the s-process. This is shown in Fig. 12.11. The most massive naturally occurring elements, such as uranium, thorium and plutonium, are formed by the r-process.

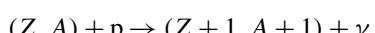
There are about 40 isotopes on the proton-rich side of the β stability valley that cannot be produced by neutron capture processes. Their abundances are very small, relative to the neighbouring isotopes. They are formed in supernova explosions at temperatures higher than 10^9 K by reactions known as the *p-process*. At this temperature, pair formation can take place:



The positron may either be annihilated immediately or be consumed in the reaction



Another reaction in the p-process is



Finally, the *fission* of some heavier isotopes may give rise to p-process nuclei. Examples of this are

the isotopes ${}^{184}\text{W}$, ${}^{190}\text{Pt}$ and ${}^{196}\text{Hg}$ formed by the fission of lead.

All the preceding reaction products are ejected into the interstellar medium in the supernova explosion. Collisions between cosmic rays and heavy nuclei then finally give rise to the light elements lithium, beryllium and boron. Thus the abundances of essentially all naturally occurring isotopes can be explained.

During succeeding generations of stars the relative abundance of heavy elements increases in the interstellar medium. They can then be incorporated into new stars, planets—and living beings.

12.9 Example

Example 12.1 An interstellar cloud has a mass of one solar mass and density of 10^{10} hydrogen atoms per cm^3 . Its rotation period is 1000 years. What is the rotation period after the cloud has condensed into a star of solar size?

The angular momentum is $L = I\omega$, where ω is the angular velocity and I is the moment of inertia. For a homogeneous sphere

$$I = \frac{2}{5}MR^2,$$

where M is the mass and R the radius. From the conservation of the angular momentum we get

$$L = I_1\omega_1 = I_2\omega_2$$

$$\Rightarrow \frac{I_1 2\pi}{P_1} = \frac{I_2 2\pi}{P_2}$$

$$\Rightarrow P_2 = P_1 \frac{I_2}{I_1} = P_1 \frac{\frac{2}{5}MR_2^2}{\frac{2}{5}MR_1^2} = P_1 \left(\frac{R_2}{R_1} \right)^2,$$

where P_1 and P_2 are the rotation periods before and after the collapse. The mass of the cloud is

$$\begin{aligned} M &= \frac{4}{3}\pi R^3 \rho \\ &= \frac{4}{3}\pi R^3 \times 10^{16} \times 1.6734 \times 10^{-27} \text{ kg} \\ &= 1 M_{\odot} = 1.989 \times 10^{30} \text{ kg}. \end{aligned}$$

Solving for the radius we get $R = 3 \times 10^{13}$ m. The rotation period after the collapse is

$$P_2 = 1000 \text{ a} \times \left(\frac{6.96 \times 10^8 \text{ m}}{3 \times 10^{13} \text{ m}} \right)^2 \\ = 5.4 \times 10^{-7} \text{ a} = 17 \text{ s.}$$

This is several orders of magnitude shorter than the actual period. Somehow the star has to get rid of most of its angular momentum during the process.²

12.10 Exercises

Exercise 12.1 Find the free fall time scale for a hydrogen cloud, if the density of H₂ molecules is 3000 cm⁻³. Assume that stars condense from such clouds, there are 100 clouds in the Galaxy,

the mass of each cloud is $5 \times 10^4 M_{\odot}$, and 10 % of the mass is converted into stars. Also assume that the average mass of a star is $1 M_{\odot}$. How many stars are born in one year?

Exercise 12.2 The mass of Vega (spectral class A0 V) is $2 M_{\odot}$, radius $3 R_{\odot}$, and luminosity $60 L_{\odot}$. Find its thermal and nuclear time scales.

Exercise 12.3 Assume that a star remains 10^9 years in the main sequence and burns 10 % of its hydrogen. Then the star will expand into a red giant, and its luminosity will increase by a factor of 100. How long is the red giant stage, if we assume that the energy is produced only by burning the remaining hydrogen?

The Sun is our nearest star. It is important for astronomy because many phenomena which can only be studied indirectly in other stars can be directly observed in the Sun (e.g. stellar rotation, starspots, the structure of the stellar surface). Our present picture of the Sun is based both on observations and on theoretical calculations. Some observations of the Sun disagree with the theoretical solar models. The details of the models will have to be changed, but the general picture should remain valid.

13.1 Internal Structure

The Sun is a typical main sequence star. Its principal properties are:

Mass	$m = M_{\odot}$	=	$1.989 \times 10^{30} \text{ kg}$
Radius	$R = R_{\odot}$	=	$6.960 \times 10^8 \text{ m}$
Mean density	$\bar{\rho}$	=	1409 kg/m^3
Central density	ρ_c	=	$1.6 \times 10^5 \text{ kg/m}^3$
Luminosity	$L = L_{\odot}$	=	$3.9 \times 10^{26} \text{ W}$
Effective temperature	T_e	=	5785 K
Central temperature	T_c	=	$1.5 \times 10^7 \text{ K}$
Absolute bolometric magnitude	M_{bol}	=	4.72
Absolute visual magnitude	M_V	=	4.79
Spectral class			G2 V
Colour indices	$B - V$	=	0.62
	$U - B$	=	0.10

Surface chemical composition	X	=	0.71
	Y	=	0.27
	Z	=	0.02
Rotational period			
at the equator			25 d
at latitude 60°			29 d

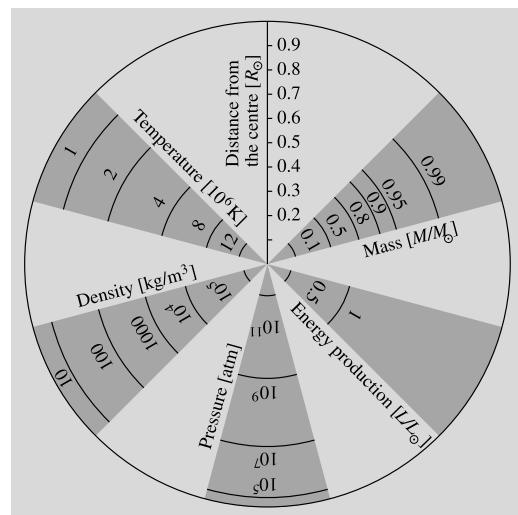
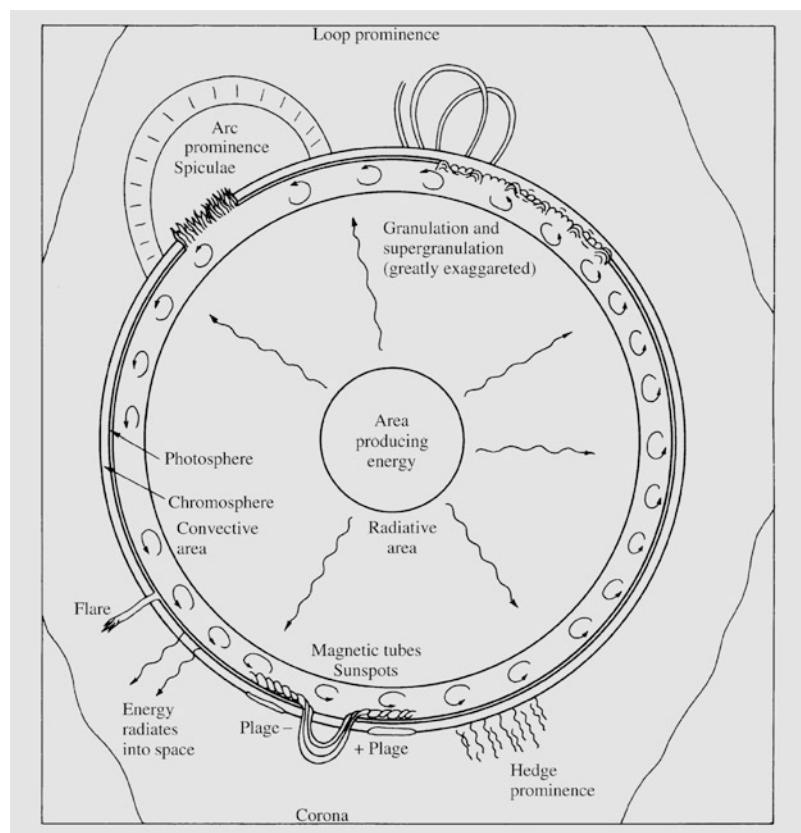


Fig. 13.1 The distribution of temperature, pressure, energy production and mass as functions of radius in the Sun

On the basis of these data, the solar model shown in Fig. 13.1 has been calculated. The energy is produced by the pp chain in a small central region. 99 % of the solar energy is produced within a quarter of the solar radius.

The Sun produces energy at the rate of $4 \times 10^{26} \text{ W}$, which is equivalent to changing about

Fig. 13.2 The interior and surface of the Sun. The various kinds of solar phenomena are schematically indicated. (Based on Van Zandt, R.P. (1977): *Astronomy for the Amateur*, Planetary Astronomy, Vol. 1, 3rd edn. (published by the author, Peoria, Ill.))



four million tonnes of mass into energy every second. The mass of the Sun is so large, about 330,000 times that of the Earth, that during the whole main sequence lifetime of the Sun less than 0.1 % of its mass is turned into energy.

When the Sun formed about 5000 million years ago, its composition was the same everywhere as its present surface composition. Since energy production is concentrated at the very centre, hydrogen is consumed most rapidly there. At about a quarter of the radius the hydrogen abundance is still the same as in the surface layers, but going inwards from that point it rapidly decreases. In the central core only 40 % of the material is hydrogen. About 5 % of the hydrogen in the Sun has been turned into helium.

The radiative central part of the Sun extends to about 70 % of the radius. At that radius the temperature has dropped so much that the gas is no longer completely ionised. The opacity of the solar material then strongly increases, inhibit-

ing the propagation of radiation. In consequence, convection becomes a more efficient means of energy transport. Thus the Sun has a convective envelope (Fig. 13.2).

The Solar Neutrino Problem The central nuclear reactions produce neutrinos at several of the steps in the pp chain (see Fig. 11.5). These neutrinos can propagate freely through the outer layers, and thus give direct information about conditions near the centre of the Sun. When neutrinos from the Sun were first observed in the 1970's, their number was found to be only about a third of what was predicted. This disagreement is called the *solar neutrino problem*.

In the first experiments only neutrinos from the ppII and ppIII branches were observed (Sect. 11.3). Since only a small fraction of the solar luminosity is produced in these reactions, it was not clear what were the consequences of these results for solar models. In the 1990's

neutrinos produced in the ppI branch, the main branch of the pp chain, were observed. Although the disagreement with the standard models was slightly smaller in these observations (about 60 % of the predicted flux was observed), the neutrino problem still remained.

Perhaps the most popular explanation for the solar neutrino problem was based on *neutrino oscillations*. According to this explanation, if neutrinos have a small mass (about 10^{-2} eV), an electron neutrino could change into a μ or a τ neutrino as it passed through the outer parts of the Sun. In the early experiments only electron neutrinos were observed, representing only part of the total number of neutrinos produced.

In 2001 the Canadian Sudbury neutrino observatory (SNO) announced results that seemed to solve the problem. The SNO can detect the flux of all different neutrinos and the fraction of the electron neutrinos. The measurements showed that the total neutrino flux was indeed consistent with the predictions of the solar models, but only 35 % of the flux consisted of electron neutrinos. Thus 65 % of the solar electron neutrinos had changed to μ or τ neutrinos while travelling from the Sun to the Earth. Similar observations were also made in a Japanese neutrino observatory. Among the pioneers of the neutrino astronomy are Raymond Davis and Masatoshi Koshiba who shared the Nobel prize in physics in 2002.

The solar neutrino problem can now be considered to be solved. The solution is a great success for the standard solar model. But it has also revealed the existence of neutrino oscillations, proving that neutrinos have a small but non-zero rest mass. This shows that the standard model of particle physics needs to be revised in some respects.

The Solar Rotation As soon as telescopes were introduced, it was observed from the motions of sunspots that the Sun is rotating with a rotational period of about 27 days. As early as 1630 *Christoph Scheiner* showed that there was *differential rotation*: the rotational period near the poles was more than 30 days, while it was only 25 days at the equator. The rotational axis of the

Sun is inclined at 7° with respect to the plane of the ecliptic, so that the North Pole of the Sun is best visible from the Earth in September.

The motions of sunspots still give the best information on the rotation near the surface of the Sun. Other surface features also have been used for this purpose. The rotational velocity has also been measured directly from the Doppler effect. The angular velocity is usually written

$$\Omega = A - B \sin^2 \psi, \quad (13.1)$$

where ψ is the latitude with respect to the equator. The measured values of the coefficients are $A = 14.5$ and $B = 2.9$ degrees/day.

The rotational velocity deeper down in the Sun cannot be directly observed. In the 1980's a method to estimate the rotation in the interior became available, when it became possible to measure the frequencies of solar oscillations from the variations in spectral lines. These oscillations are essentially sound waves produced by turbulent gas motions in the convection zone. These sound waves have calculable oscillation periods (about 3–12 minutes), which depend on the conditions in the solar interior. By comparing the observed and theoretical values one can get information about the conditions deep inside the Sun. The idea of the method is the same as that used when studying the interior of the Earth by means of waves from earthquakes, and it is therefore called *helioseismology*.

Using helioseismology, models for the solar rotation throughout the convection zone have been deduced. It appears that the angular velocity in the whole convection zone is almost the same as at the surface, although it decreases slightly with radius near the equator, and increases near the poles. The angular velocity of the radiative core is still uncertain, but there are indications that the core is rotating as a solid body with approximately the average surface angular velocity. At the bottom of the convection zone there is a thin layer known as the *tachocline*, where the angular velocity changes rapidly with radius. The internal solar rotation according to the helioseismological studies is shown in Fig. 13.3.

The solar differential rotation is maintained by gas motions in the convection zone. Explaining the observed behaviour is a difficult problem that is not yet completely understood.

13.2 The Atmosphere

The solar atmosphere is divided into the *photosphere* and the *chromosphere*. Outside the actual atmosphere, the *corona* extends much further outwards.

The Photosphere The innermost layer of the atmosphere is the photosphere, which is only about 300–500 km thick. The photosphere is the visible surface of the Sun, where the density rapidly increases inwards, hiding the interior from sight. The temperature at the inner boundary of the photosphere is 8000 K and at the outer

boundary 4500 K. Near the edge of the solar disk, the line of sight enters the photosphere at a very small angle and never penetrates to large depths. Near the edges one therefore only sees light from the cooler, higher layers. For this reason, the edges appear darker; this phenomenon is known as *limb darkening*. Both the continuous spectrum and the absorption lines are formed in the photosphere, but the light in the absorption lines comes from higher layers and therefore the lines appear dark.

The solar convection is visible on the surface as the *granulation* (Fig. 13.4), an uneven, constantly changing granular pattern. At the bright centre of each granule, gas is rising upward, and at the darker granule boundaries, it is sinking down again. The size of a granule seen from the Earth is typically 1'', corresponding to about 1000 km on the solar surface. There is also a larger scale convection called *supergranulation* in the photosphere. The cells of the supergranulation may be about 1' in diameter. The observed velocities in the supergranulation are mainly directed along the solar surface.

The Chromosphere Outside the photosphere there is a layer, perhaps about 500 km thick, where the temperature increases from 4500 K to about 6000 K, the chromosphere. Outside this layer, there is a transition region of a few thousand kilometres, where the chromosphere gradually goes over into the corona. In the outer parts of the transition region, the kinetic temperature is already about 10^6 K.

Normally the chromosphere is not visible, because its radiation is so much weaker than that of the photosphere. However, during total solar eclipses, the chromosphere shines into view for a few seconds at both ends of the total phase, when the Moon hides the photosphere completely. The chromosphere then appears as a thin reddish sickle or ring.

During eclipses the chromospheric spectrum, called the *flash spectrum*, can be observed (Fig. 13.5). It is an emission line spectrum with more than 3000 identified lines. Brightest among these are the lines of hydrogen, helium and certain metals.

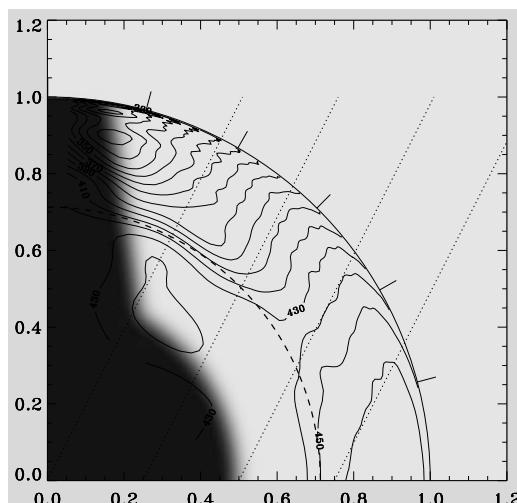


Fig. 13.3 The rotation rate of the Sun inferred from helioseismological observations. The equator is at the horizontal axis and the pole is at the vertical axis, both axes being labelled by fractional radius. Some contours are labelled in nHz, and, for clarity, selected contours are shown as bold. (430 nHz is about 26.9 days.) The dashed circle is at the base of the convection zone and the tick marks at the edge of the outer circle are at latitudes 15°, 30°, 45°, 60°, 75°. The shaded area indicates the region in the Sun where no reliable inference can be made with present data. The slanted dotted lines are at an angle of 27° with the rotation axis. (Adapted from Schou et al. 1998.) (J. Christensen-Dalsgaard 2007, astro-ph/0610942, Fig. 2)

Fig. 13.4 The granulation of the solar surface. The granules are produced by streaming gas. Their typical diameter is 1000 km. The picture was taken in May 13, 2005 with the Swedish one metre solar telescope on La Palma. (Photograph Tom Berger, Royal Swedish Academy of Sciences, ISP/RSAS)

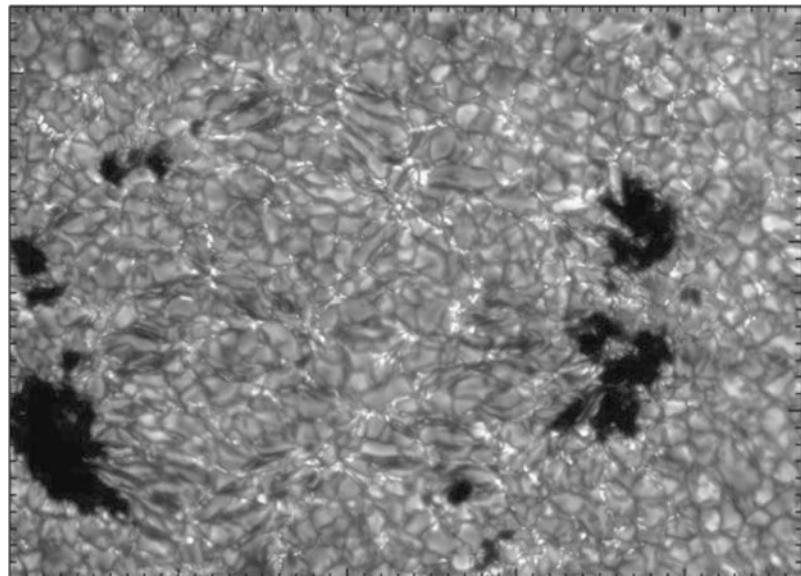
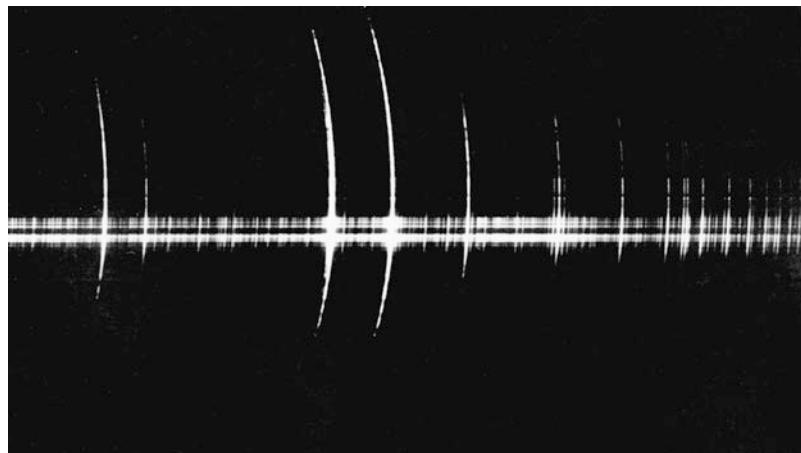


Fig. 13.5 Flash spectrum of the solar chromosphere, showing bright emission lines



One of the strongest chromospheric emission lines is the hydrogen Balmer α line (Fig. 13.6) at a wavelength of 656.3 nm. Since the H_α line in the normal solar spectrum is a very dark absorption line, a photograph taken at this wavelength will show the solar chromosphere. For this purpose, one uses narrow-band filters letting through only the light in the H_α line. The resulting pictures show the solar surface as a mottled, wavy disk. The bright regions are usually the size of a supergranule, and are bounded by *spicules* (Fig. 13.7). These are flamelike structures rising up to 10,000 km above the chromosphere, and lasting for a few minutes. Against

the bright surface of the Sun, they look like dark streaks; at the edges, they look like bright flames.

The Corona The chromosphere gradually goes over into the corona. The corona is also best seen during total solar eclipses (Fig. 13.8). It then appears as a halo of light extending out to a few solar radii. The surface brightness of the corona is about that of the full moon, and it is therefore difficult to see next to the bright photosphere.

The inner part of the corona, the K corona, has a continuous spectrum formed by the scattering of the photospheric light by electrons. Fur-

Fig. 13.6 The solar surface in the hydrogen H_α line. Active regions appear bright; the dark filaments are prominences. Limb darkening has been removed artificially, which brings to light spicules and prominences above the limb. The photograph was taken in October 1997. (Photograph Big Bear Solar Observatory/NJIT)

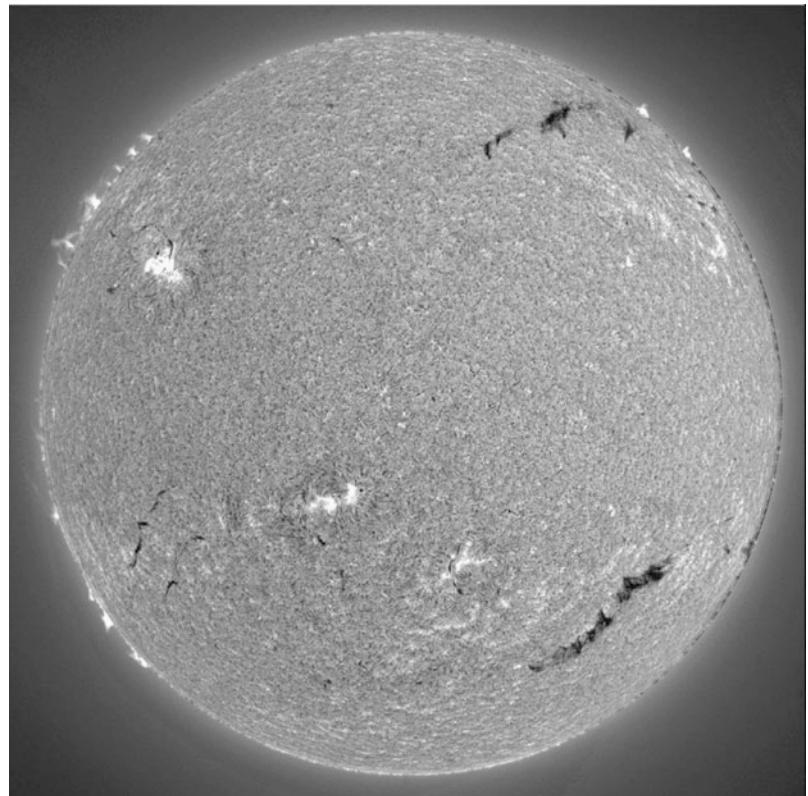
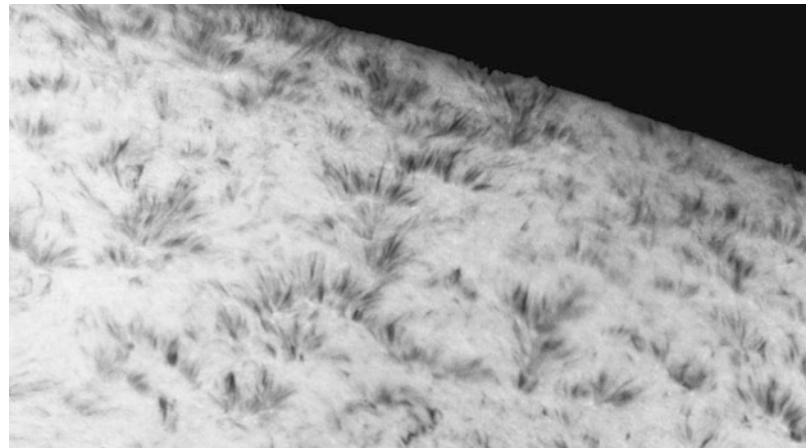


Fig. 13.7 Spicules, flamelike uprisings near the edge of the solar disc. (Photograph Big Bear Solar Observatory)



ther out, a few solar radii from the surface, is the F corona, which has a spectrum showing Fraunhofer absorption lines. The light of the F corona is sunlight scattered by dust.

In the latter part of the 19th century strong emission lines, which did not correspond to those of any known element, were discovered in the

corona (Fig. 13.9). It was thought that a new element, called coronium, had been found—a little earlier, helium had been discovered in the Sun before it was known on Earth. About 1940, it was established that the coronal lines were due to highly ionised atoms, e.g. thirteen times ionised iron. Much energy is needed to remove so many

Fig. 13.8 Previously, the corona could be studied only during total solar eclipses. The picture is from the eclipse on March 7, 1970. Nowadays the corona can be studied continuously using a device called the coronagraph

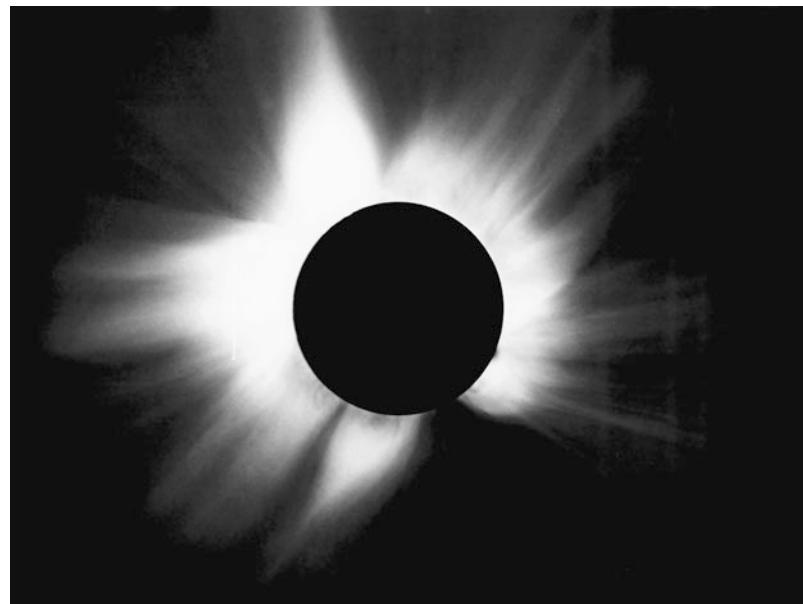
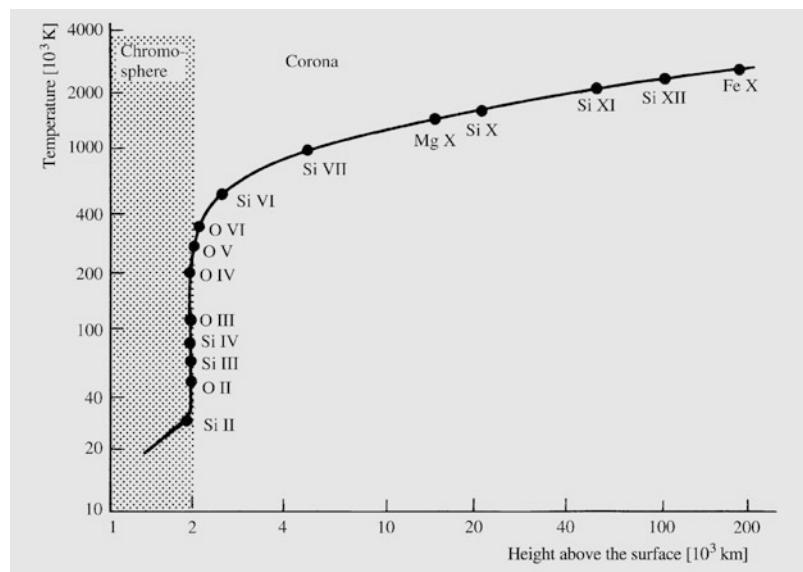


Fig. 13.9 The presence of lines from highly ionised atoms in the coronal spectrum shows that the temperature of the corona has to be very high



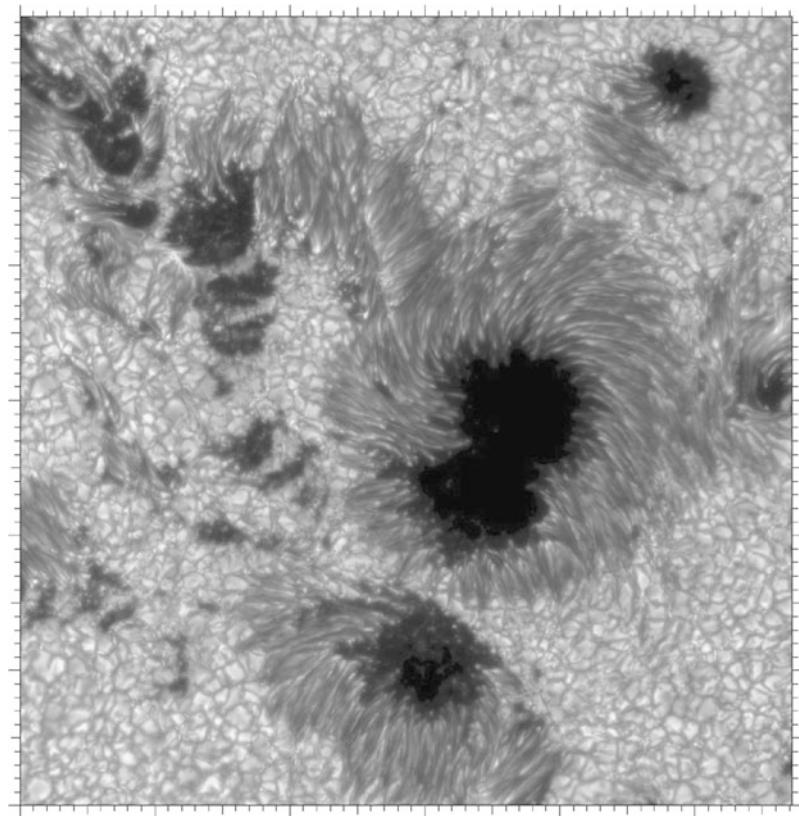
electrons from the atoms. The entire corona has to have a temperature of about a million degrees.

A continuous supply of energy is needed in order to maintain the high temperature of the corona. According to earlier theories, the energy came in the form of acoustic or magnetohydrodynamic shock waves generated at the solar surface by the convection. Most recently, heating by elec-

tric currents induced by changing magnetic fields has been suggested. Heat would then be generated in the corona almost like in an ordinary light bulb.

In spite of its high temperature the coronal gas is so diffuse that the total energy stored in it is small. It is constantly streaming outwards, gradually becoming a *solar wind*, which carries a flux of particles away from the Sun. The gas lost in

Fig. 13.10 The sunspots are the form of solar activity that has been known for the longest time. The photograph was taken with the Swedish 1-meter Solar Telescope in July 2002. (Photograph Royal Swedish Academy of Sciences)



this way is replaced with new material from the chromosphere.

13.3 Solar Activity

Sunspots The clearest visible sign of solar activity are the *sunspots*. The existence of sunspots has been known for long (Fig. 13.10), since the largest ones can be seen with the naked eye by looking at the Sun through a suitably dense layer of fog. More precise observations became available beginning in the 17th century, when Galilei started to use the telescope for astronomical observations.

A sunspot looks like a ragged hole in the solar surface. In the interior of the spot there is a dark *umbra* and around it, a less dark *penumbra*. By looking at spots near the edge of the solar disk, it can be seen that the spots are slightly depressed with respect to the rest of the surface. The surface temperature in a sunspot is about 1500 K below

that of its surroundings, which explains the dark colour of the spots.

The diameter of a typical sunspot is about 10,000 km and its lifetime is from a few days to several months, depending on its size. The larger spots are more likely to be long-lived. Sunspots often occur in pairs or in larger groups. By following the motions of the spots, the period of rotation of the Sun can be determined.

The variations in the number of sunspots have been followed for almost 250 years. The frequency of spots is described by the Zürich sunspot number Z :

$$Z = C(S + 10G), \quad (13.2)$$

where S is the number of spots and G the number of spot groups visible at a particular time. C is a constant depending on the observer and the conditions of observation.

In Fig. 13.11, the variations in the Zürich sunspot number between the 18th century and

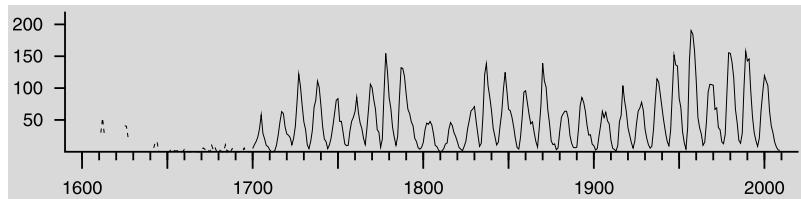


Fig. 13.11 The Zürich sunspot number from 1700 to 2001. Prior to 1700 there are only occasional observations. The number of sunspots and spot groups varies with a period of about 11 years

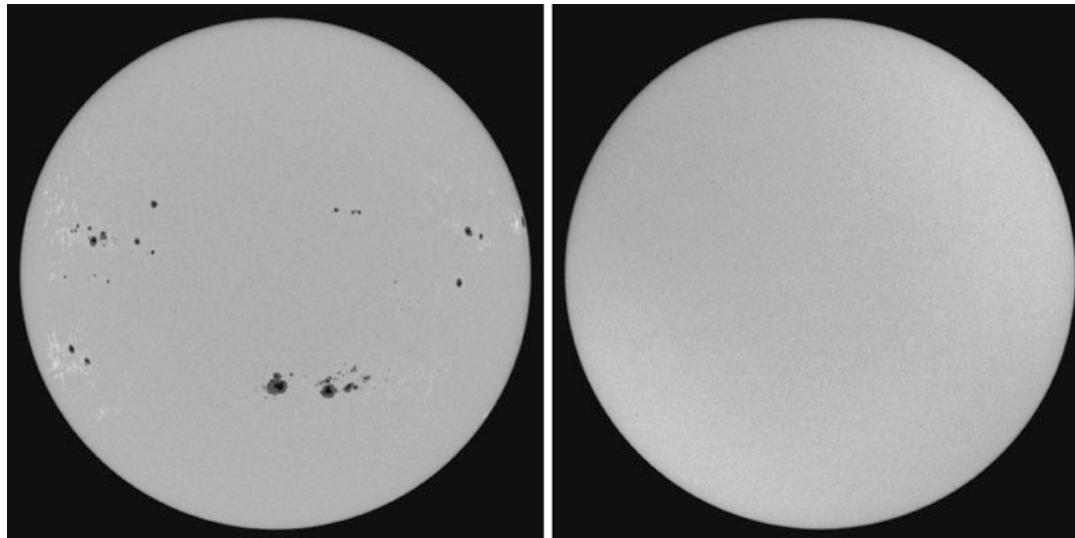


Fig. 13.12 *Left:* During a sunspot maximum (Sept 27, 2001) the Sun is dotted by numerous spots. *Right:* Exactly seven years later no spots are seen, although minimum phase should already be over. (Photos SOHO/MIDI)

the present are shown. Evidently the number of spots varies with an average period of about 11 years. The actual period may be between 7 and 17 years. In the past decades, it has been about 10.5 years. Usually the activity rises to its maximum in about 3–4 years, and then falls off slightly more slowly (Fig. 13.12). The period was first noted by *Samuel Heinrich Schwabe* in 1843.

The variations in the number of sunspots have been fairly regular since the beginning of the 18th century. However, in the 17th century there were long intervals when there were essentially no spots at all. This quiescent period is called the *Maunder minimum*. The similar *Spörer minimum* occurred in the 15th century, and other quiet intervals have been inferred at earlier epochs. The mechanism behind these irregular variations in solar activity is not yet understood.

The magnetic fields in sunspots are measured on the basis of the Zeeman effect, and may be as large as 0.45 tesla. (The magnetic field of the Earth at the equator is 0.03 mT.) The strong magnetic field inhibits convective energy transport, which explains the lower temperature of the spots.

Sunspots often occur in pairs where the components have opposite polarity. The structure of such *bipolar groups* can be understood if the field rises into a loop above the solar surface, connecting the components of the pair. If gas is streaming along such a loop, it becomes visible as a *loop prominence* (Fig. 13.13).

The periodic variation in the number of sunspots reflects a variation in the general solar magnetic field. At the beginning of a new activity cycle spots first begin to appear at latitudes of about

Fig. 13.13 In pairs of sunspots the magnetic field lines form a loop outside the solar surface. Material streaming along the field lines may form loop prominences. Loops of different size can be seen in this image, which the Trace satellite took in 1999. (Photo Trace)

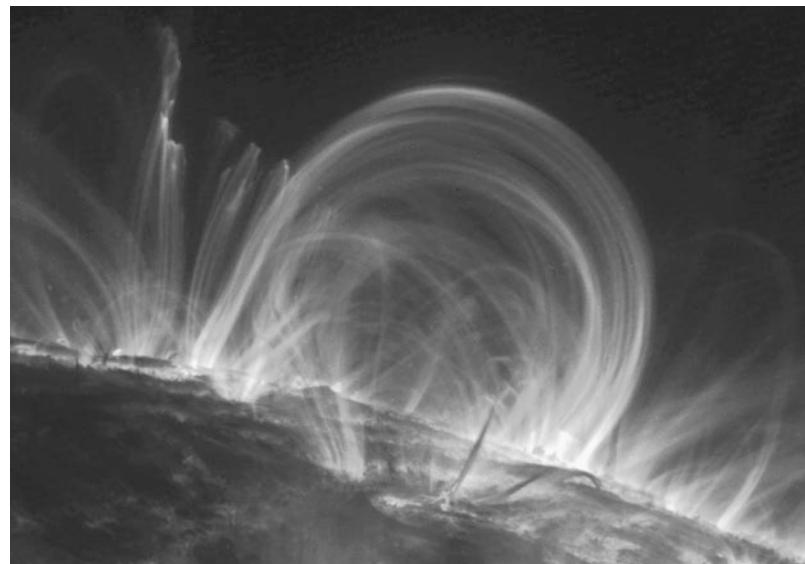
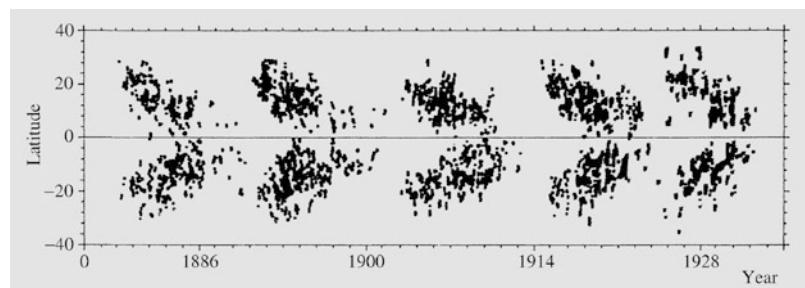


Fig. 13.14 At the beginning of an activity cycle, sunspots appear at high latitudes. As the cycle advances the spots move towards the equator. (Diagram by H. Virtanen, based on Greenwich Observatory observations)



$\pm 40^\circ$. As the cycle advances, the spots move closer to the equator. The characteristic pattern in which spots appear, shown in Fig. 13.14, is known as the *butterfly diagram*. Spots of the next cycle begin to appear while those of the old one are still present near the equator. Spots belonging to the new cycle have a polarity opposite to that of the old ones. (Spots in opposite hemispheres also have opposite polarity.) Since the field is thus reversed between consecutive 11 year cycles the complete period of solar magnetic activity is 22 years.

The following general qualitative description of the mechanism of the solar cycle was proposed by Horace W. Babcock. Starting at a solar minimum, the field will be of a generally dipolar character. Because a conducting medium, such as the outer layers of the Sun, cannot move across the field lines, these will be frozen into the

plasma and carried along by it. Thus the differential rotation will draw the field into a tight spiral (Fig. 13.15). In the process the field becomes stronger, and this amplification will be a function of latitude.

When the subsurface field becomes strong enough, it gives rise to a “magnetic buoyancy” that lifts ropes of magnetic flux above the surface. This happens first at a latitude about 40° , and later at lower latitudes. These protruding flux ropes expand into loops forming bipolar groups of spots. As the loops continue expanding they make contact with the general dipolar field, which still remains in the polar regions. This leads to a rapid reconnection of the field lines neutralising the general field. The final result when activity subsides is a dipolar field with a polarity opposite the initial one.

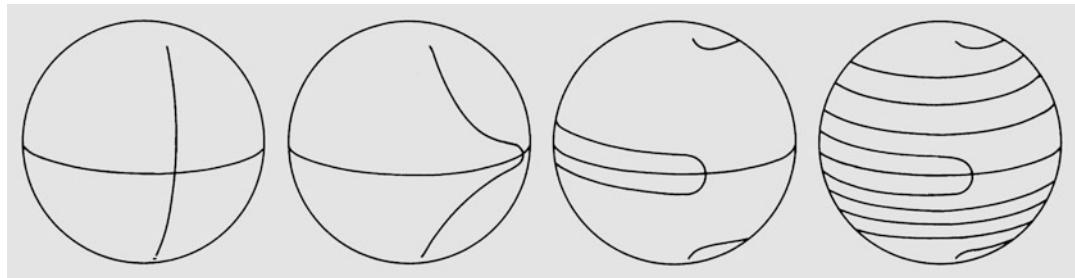


Fig. 13.15 Because the Sun rotates faster at the equator than at the poles, the field lines of the solar magnetic field are drawn out into a tight spiral

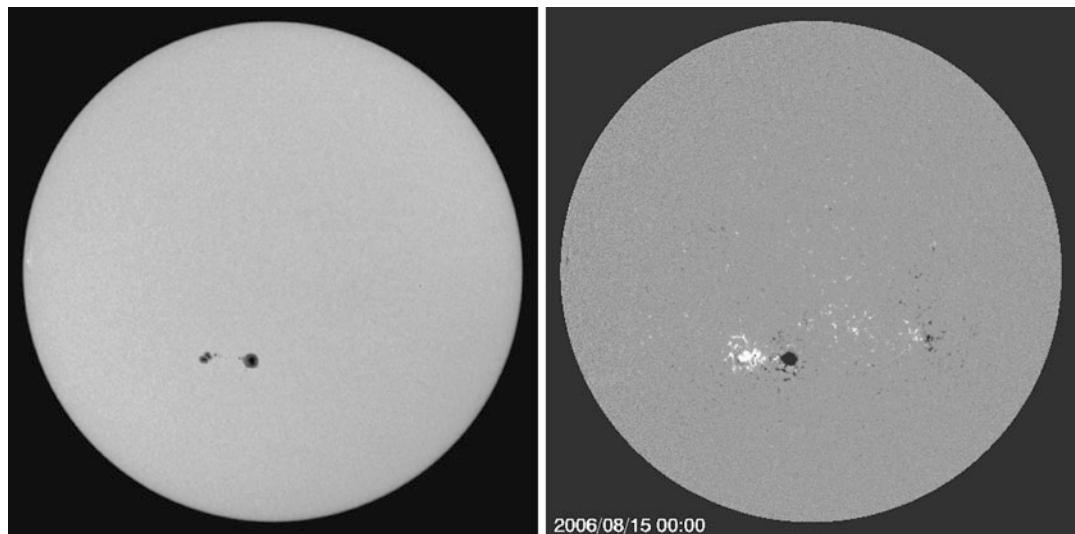


Fig. 13.16 A quiet Sun in August 2006 around the last sunspot minimum. Both pictures were taken by the Michelson Doppler Imager on the SOHO satellite. *On the*

left the Sun in visible light, on the right a magnetogram, which shows the opposite polarities of the magnetic fields as black and white. (Photo SOHO/NASA/ESA)

Thus the Babcock model accounts for the butterfly diagram, the formation of bipolar magnetic regions and the general field reversal between activity maxima. Nevertheless, it remains an essentially phenomenological model, and alternative scenarios have been proposed. In *dynamo theory* quantitative models for the origin of magnetic fields in the Sun and other celestial bodies are studied. In these models the field is produced by convection and differential rotation of the gas. A completely satisfactory dynamo model for the solar magnetic cycle has not yet been found. For example, it is not yet known whether the field is produced everywhere in the convection zone, or just in the boundary layer between the convective

and radiative regions, as some indications suggest.

Other Activity The Sun shows several other types of surface activity: *faculae* and *plages*; *prominences*; *flares*.

The faculae and plages are local bright regions in the photosphere and chromosphere, respectively. Observations of the plages are made in the hydrogen H_α or the calcium K lines (Fig. 13.16). The plages usually occur where new sunspots are forming, and disappear when the spots disappear. Apparently they are caused by the enhanced heating of the chromosphere in strong magnetic fields.

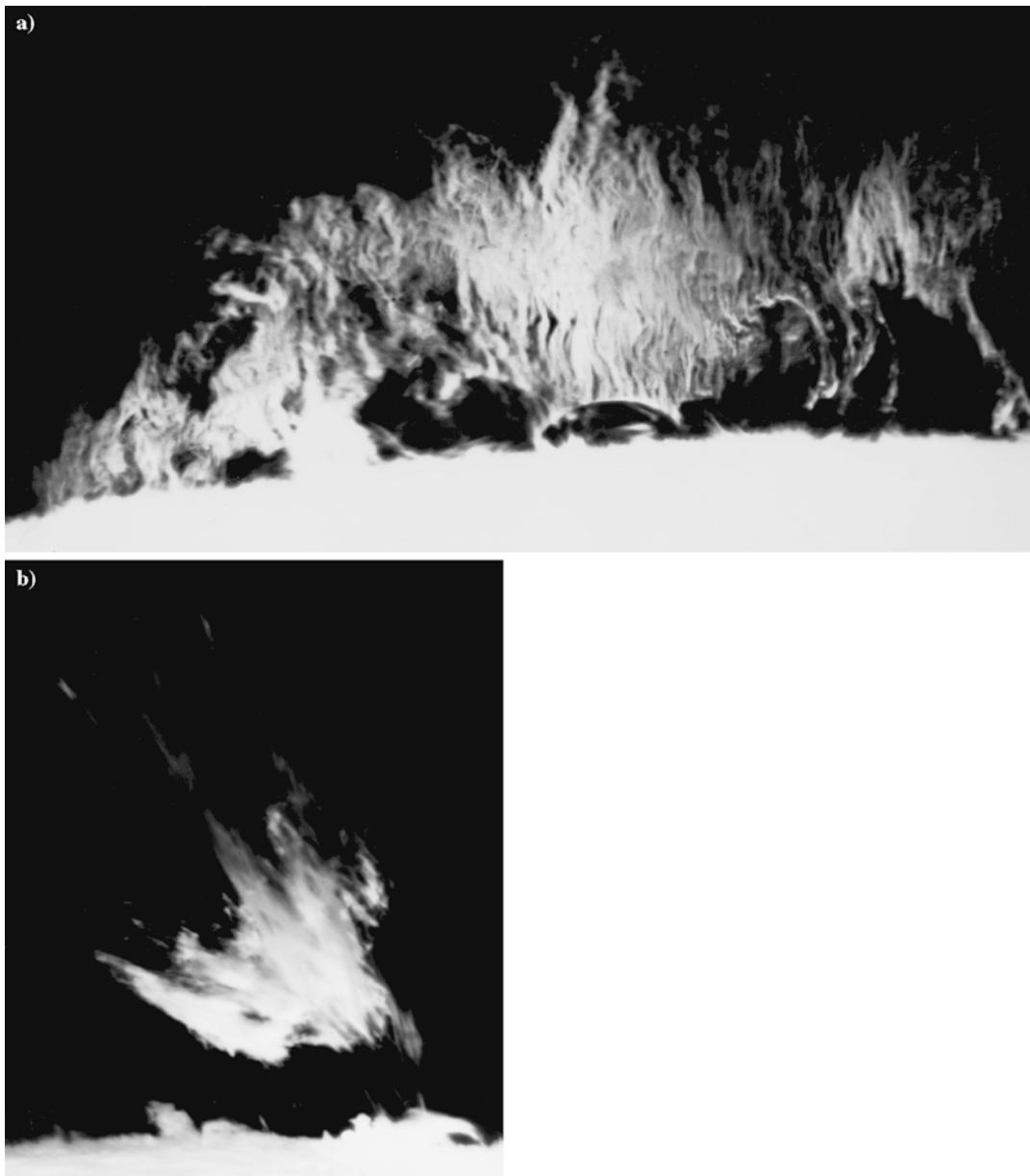


Fig. 13.17 (a) Quiescent “hedgerow” prominence. (Photograph Sacramento Peak Observatory.) (b) Larger eruptive prominence. (Photograph Big Bear Solar Observatory)

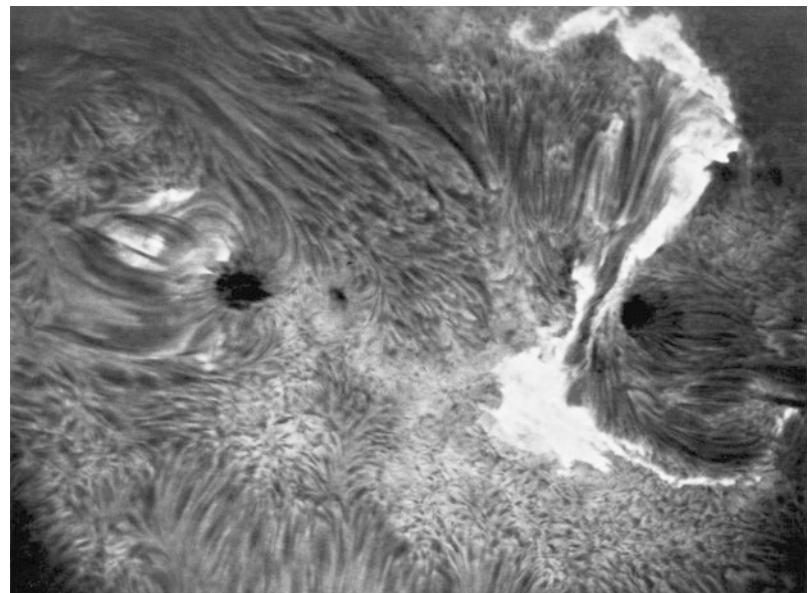
The prominences are among the most spectacular solar phenomena. They are glowing gas masses in the corona, easily observed near the edge of the Sun. There are several types of prominences (Fig. 13.17): the quiescent prominences, where the gas is slowly sinking along the magnetic field lines; loop prominences, connected with magnetic field loops in sunspots; and the

rarer eruptive prominences, where gas is violently thrown outwards.

The temperature of prominences is about 10,000–20,000 K. In H_α photographs of the chromosphere, the prominences appear as dark filaments against the solar surface (Fig. 13.6).

The flare outbursts are among the most violent forms of solar activity (Fig. 13.18). They appear

Fig. 13.18 A violent flare near some small sunspots. (Photograph Sacramento Peak Observatory)



as bright flashes, lasting from one second to just under an hour. In the flares a large amount of energy stored in the magnetic field is suddenly released. The detailed mechanism is not yet known.

Flares can be observed at all wavelengths. The hard X-ray emission of the Sun may increase hundredfold during a flare. Several different types of flares are observed at radio wavelengths. The emission of solar cosmic ray particles also rises.

Prominences and flares are often accompanied with *coronal mass ejections*. Fast moving (500–2000 km/s) clouds cause shock waves accelerating particles to very high velocities. Particles in flares are the fastest ($v \approx 0.3c$). Particles in a magnetic cloud travelling at the velocity of the mass ejection reach the Earth in a couple of days. Particles activated by the shock wave arrive continuously affecting the space weather and causing magnetic storms.

The flares and coronal mass ejections give rise to disturbances on the Earth. The X-rays cause changes in the ionosphere, which affect short-wave radio communications. The particles give rise to strong auroras when they enter the Earth's magnetic field a few days after the outburst.

Solar Radio Emission The Sun is the strongest radio source in the sky and has been observed

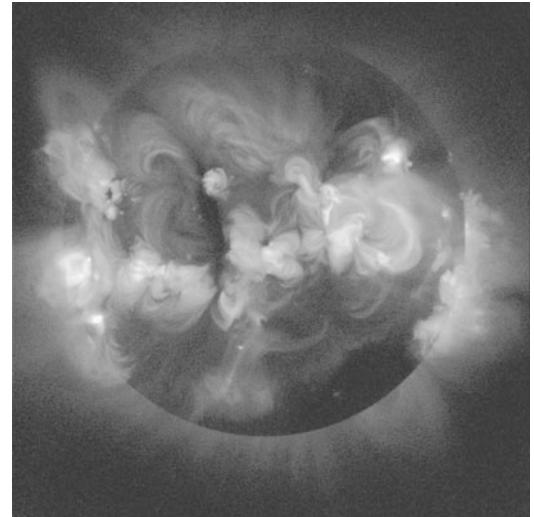
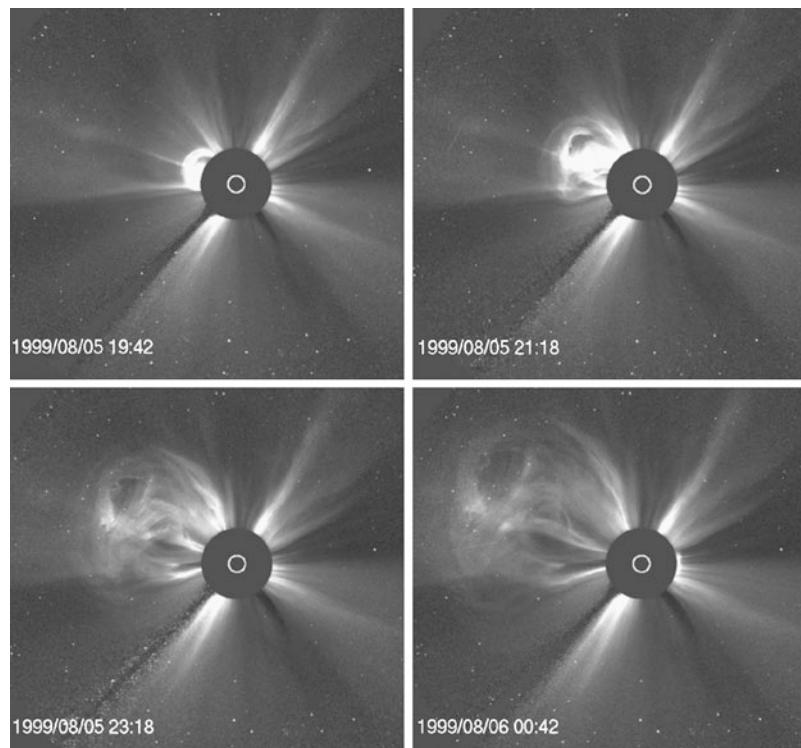


Fig. 13.19 An X-ray picture of the active Sun, taken by the Japanese Yohkoh satellite in 1999, around the last maximum of sunspot activity. (Photo JAXA)

since the 1940's. In contrast to optical emission the radio picture of the Sun shows a strong *limb brightening*. This is because the radio radiation comes from the upper layers of the atmosphere. Since the propagation of radio waves is obstructed by free electrons, the high electron density near the surface prevents radio radiation from getting out. Shorter wavelengths can propa-

Fig. 13.20 The SOHO (Solar and Heliospheric Observatory) satellite keeps a constant watch on the Sun and its surroundings in many wavelengths. Here the LASCO (Large Angle and Spectrometric Coronagraph) instrument sees a large Coronal Mass Ejection erupting from the Sun. The surface of the Sun is covered by a disk, and the size and position of the Sun is indicated by the white circle. (Photo SOHO/NASA/ESA)



gate more easily, and thus millimetre-wavelength observations give a picture of deeper layers in the atmosphere, whereas the long wavelengths show the upper layers. (The 10 cm emission originates in the upper layers of the chromosphere and the 1 m emission, in the corona.)

The Sun looks different at different wavelengths. At long wavelengths the radiation is coming from the largest area, and its electron temperature is about 10^6 K, since it originates in the corona.

The radio emission of the Sun is constantly changing according to solar activity. During large storms the total emission may be 100,000 times higher than normal. Especially the motion of shock waves can be followed by the radio emission, since the electrons accelerated by the shock generate radio emission (type II radio bursts).

X-ray and UV Radiation The X-ray emission of the Sun is also related to active regions (Fig. 13.19). Signs of activity are bright *X-ray regions* and smaller *X-ray bright points*, which last for around ten hours. The inner solar corona also

emits X-rays. Near the solar poles there are *coronal holes*, where the X-ray emission is weak.

Ultraviolet pictures of the solar surface show it as much more irregular than it appears in visible light. Most of the surface does not emit much UV radiation, but there are large active regions that are very bright in the ultraviolet.

Several satellites have made observations of the Sun at UV and X-ray wavelengths, for example Soho (Solar and Heliospheric Observatory, 1995–, Fig. 13.20). These observations have made possible detailed studies of the outer layers of the Sun. Observations of other stars have revealed coronae, chromospheres and magnetic variations similar to those in the Sun. Thus the new observational techniques have brought the physics of the Sun and the stars nearer to each other.

13.4 Solar Wind and Space Weather

A continuous stream of charged particles is coming from the Sun as the solar wind, varying with the solar activity. The coronal mass ejec-

tions are considered as “disturbances”. Around a minimum mass ejections may occur a few times a week; around a maximum there may be several ejections every day. The solar wind consists mainly of electrons and protons. At the distance of the Earth there are typically 5–10 particles per cubic centimetre. There are also some nuclei of helium atoms. The velocity is the solar wind close to the poles of the Sun is about 800 km/s but around the equator only about 300 km/s. At the distance of the Earth the velocity of the particles is about 500 km/s on the average. The Sun loses $2 - 3 \times 10^{-14} M_{\odot}$ of its mass every year as solar wind.

Magnetic fields of planets direct the motions of the particles of the solar wind (Sect. 7.7). The auroras are the most outstanding phenomenon on the Earth. Currents induced by the charged particles can also have considerable negative effects. They can damage satellites and even electric networks. The most serious accident this far happened in 1989 in Quebec when a high voltage network was damaged and millions of people had to survive without electricity for several hours.

Space weather mean the interaction of the solar wind and the magnetic environment of the Earth. Because of its effects it is now followed actively. The space weather affects the upper atmosphere but it is still debated whether it is connected to variations of the weather and climate.

13.5 Example

Example 13.1 Assume that the Sun converts 0.8 % of its mass into energy. Find an upper limit

for the age of the Sun, assuming that its luminosity has remained constant.

The total amount of energy released is

$$\begin{aligned} E &= mc^2 = 0.008 M_{\odot} c^2 \\ &= 0.008 \times 2 \times 10^{30} \text{ kg} \times (3 \times 10^8 \text{ m s}^{-1})^2 \\ &= 1.4 \times 10^{45} \text{ J}. \end{aligned}$$

The time needed to radiate this energy is

$$\begin{aligned} t &= \frac{E}{L_{\odot}} = \frac{1.4 \times 10^{45} \text{ J}}{3.9 \times 10^{26} \text{ W}} \\ &= 3.6 \times 10^{18} \text{ s} \approx 10^{11} \text{ years}. \end{aligned}$$

13.6 Exercises

Exercise 13.1 The solar constant, i.e. the flux density of the solar radiation at the distance of the Earth is 1370 W m^{-2} .

- (a) Find the flux density on the surface of the Sun, when the apparent diameter of the Sun is $32'$.
- (b) How many square metres of solar surface is needed to produce 1000 megawatts?

Exercise 13.2 Some theories have assumed that the effective temperature of the Sun 4.5 billion years ago was 5000 K and radius 1.02 times the current radius. What was the solar constant then? Assume that the orbit of the Earth has not changed.

Stars with changing magnitudes are called *variables* (Fig. 14.1). Variations in the brightness of stars were first noted in Europe at the end of the 16th century, when Tycho Brahe's supernova lit up (1572) and the regular light variation of the star *o* Ceti (Mira) was observed (1596). The number of known variables has grown steadily as observational precision has improved (Fig. 14.2). The most recent catalogues contain about 40,000 stars known or suspected to be variable.

Strictly speaking, all stars are variable. As was seen in Chap. 12, the structure and brightness of a star change as it evolves. Although these changes are usually slow, some evolutionary phases may be extremely rapid. In certain evolutionary stages, there will also be periodic variations, for example pulsations of the outer layers of a star.

Small variations in stellar brightness are also caused by hot and cool spots on a star's surface, appearing and disappearing as it rotates about its axis. The luminosity of the Sun changes slightly because of the sunspots. Probably there are similar spots on almost all stars.

Initially stellar brightnesses were determined visually by comparing stars near each other. Later on, comparisons were made on photographic plates. At present the most accurate observations are made photoelectrically or using a CCD camera. The magnitude variation as a function of time is called the *lightcurve* of a star (Fig. 14.3). From it one obtains the *amplitude* of the magnitude variation and its *period*, if the variation is periodic.

The basic reference catalogue of variable stars is the *General Catalogue of Variable Stars* by the Soviet astronomer Boris Vasilyevich Kukarkin. New, supplemented editions appear at times; the fourth edition published in 1985–1987, edited by P.N. Kholopov, contains about 32,000 variables of the Milky Way galaxy.

14.1 Classification

When a new variable is discovered, it is given a name according to the constellation in which it is located. The name of the first variable in a given constellation is R, followed by the name of the constellation (in the genitive case). The symbol for the second variable is S, and so on, to Z. After these, the two-letter symbols RR, RS, ... to ZZ are used, and then AA to QZ (omitting I). This is only enough for 334 variables, a number that has long been exceeded in most constellations. The numbering therefore continues: V335, V336, etc. (V stands for variable). For some stars the established Greek letter symbol has been kept, although they have later been found to be variable (e.g. δ Cephei).

The classification of variables is based on the shape of the lightcurve, and on the spectral class and observed radial motions. The spectrum may also contain dark absorption lines from material around the star. Observations can be made outside the optical region as well. Thus the radio emission of some variables (e.g. flare stars) increases strongly, simultaneously with their optical brightness. Examples of radio and X-ray vari-

Fig. 14.1 The variables are stars changing in brightness. Two variables in Scorpius, R and S Sco. (Photograph Yerkes Observatory)

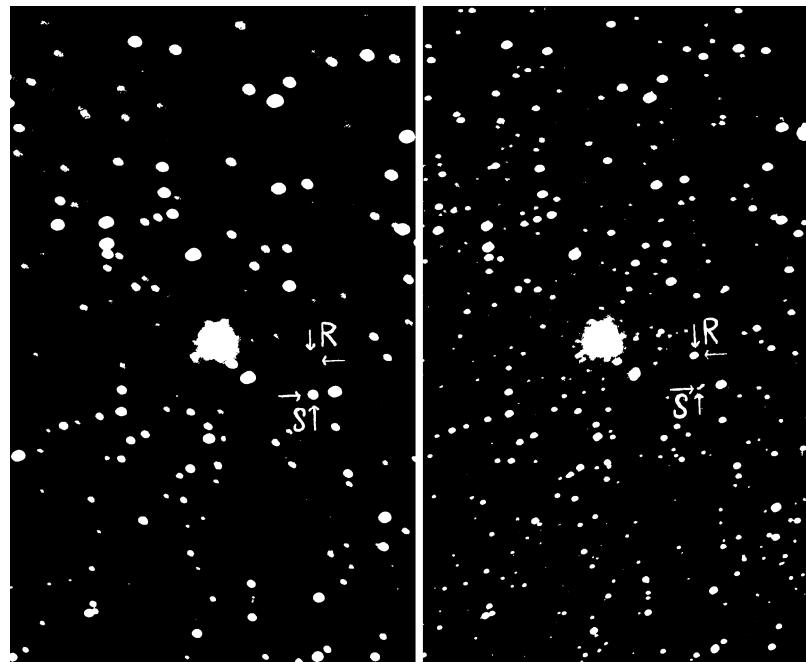
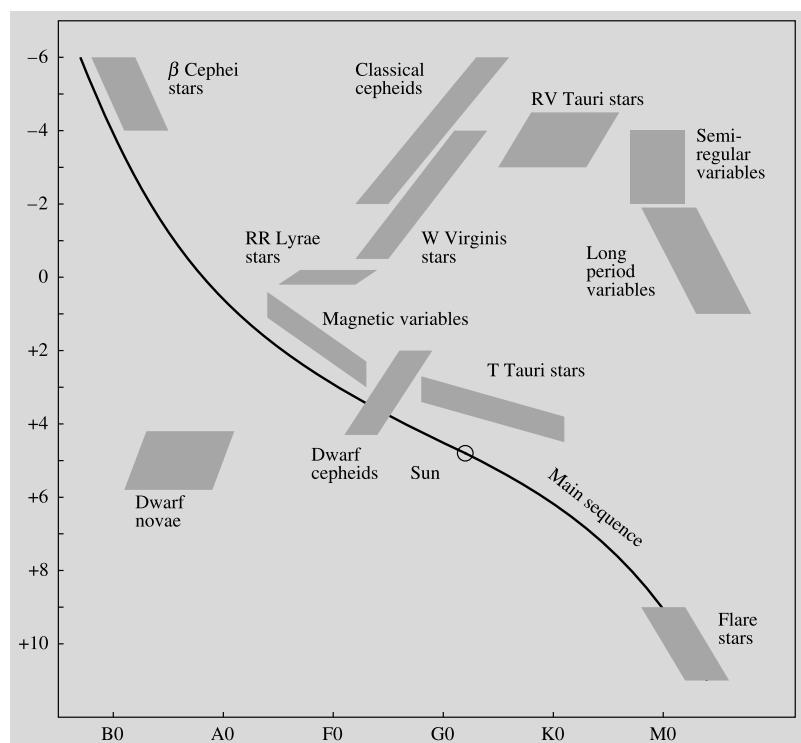


Fig. 14.2 The location of variables in the HR diagram



ables are the radio and X-ray pulsars, and the X-ray bursters.

Variables are usually divided into three main types: *pulsating*, *eruptive* and *eclipsing variables*.

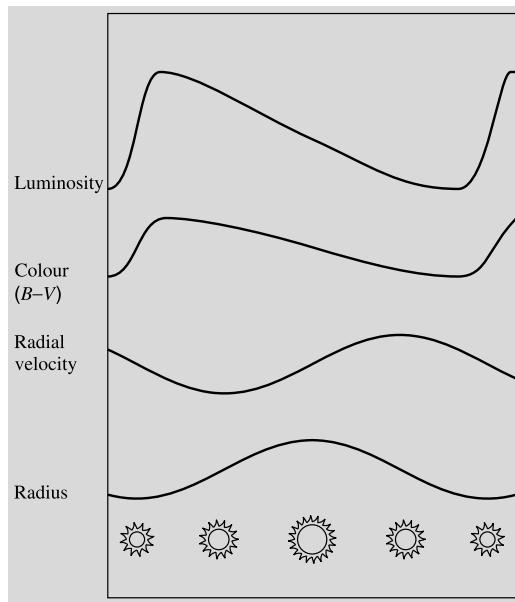


Fig. 14.3 The variation of brightness, colour and size of a cepheid during its pulsation

The eclipsing variables are binary systems in which the components periodically pass in front of each other. In these variables the light variations do not correspond to any physical change in the stars. They have been treated in connection with the binary stars. In the other variables the brightness variations are intrinsic to the stars. In the pulsating variables the variations are due to the expansion and contraction of the outer layers. These variables are giants and supergiants that have reached an unstable stage in their evolution. The eruptive variables are usually faint stars ejecting mass. They are mostly members of close binary systems in which mass is transferred from one component to the other.

In addition a few *rotating variables* are known, where the brightness variations are due to an uneven temperature distribution on the surface, starspots coming into sight when the star rotates. Such stars may be quite common—after all, our Sun is a weak rotating variable. The most prominent group of rotating variables are the magnetic A stars (e.g. the α^2 Canum Venaticorum stars). These stars have strong magnetic fields that may be giving rise to starspots. The periods of rotating

Table 14.1 The main properties of pulsating variables (N , number of stars of the given type in Kukarkin's catalogue, P , pulsation period in days, Δm , pulsation amplitude in magnitudes)

Variable	N	P	Spectrum	Δm
Classical cepheids (δ Cep, W Vir)	800	1–135	F–K I	$\lesssim 2$
RR Lyrae	6100	<1	A–F8	$\lesssim 2$
Dwarf cepheids (δ Scuti)	200	0.05–7	A–F	$\lesssim 1$
β Cephei	90	0.1–0.6	B1–B3 III	$\gtrsim 0.3$
Mira variables	5800	80–1000	M–C	$\gtrsim 2.5$
RV Tauri	120	30–150	G–M	$\lesssim 4$
Semiregular	3400	30–1000	K–C	$\lesssim 4.5$
Irregular	2300	–	K–M	$\lesssim 2$

variables range from about 1 day to 25 d, and the amplitudes are less than 0.1 mag.

14.2 Pulsating Variables

The wavelengths of the spectral lines of the pulsating variables change along with the brightness variations (Table 14.1). These changes are due to the Doppler effect, showing that the outer layers of the star are indeed pulsating. The observed gas velocities are in the range of 40–200 km/s.

The period of pulsation corresponds to a *proper frequency* of the star. Just like a tuning fork vibrates with a characteristic frequency when hit, a star has a fundamental frequency of vibration. In addition to the fundamental frequency other frequencies, “overtones”, are possible. The observed brightness variation can be understood as a superposition of all these modes of vibration. Around 1920, the English astrophysicist Sir Arthur Eddington showed that the period of pulsation P is inversely proportional to the square root of the mean density,

$$P \propto \frac{1}{\sqrt{\rho}}. \quad (14.1)$$

The diameter of the star may double during the pulsation, but usually the changes in size are minor. The main cause of the light variation is the periodic variation of the surface temperature. We have seen in Sect. 5.7 that the luminosity of a star

depends sensitively on its effective temperature, $L \propto T_e^4$. Thus a small change in effective temperature leads to a large brightness variation.

Normally a star is in stable hydrostatic equilibrium. If its outer layers expand, the density and temperature decrease. The pressure then becomes smaller and the force of gravity compresses the gas again. However, unless energy can be transferred to the gas motions, these oscillations will be damped.

The flux of radiative energy from the stellar interior could provide a source of energy for the stellar oscillations, if it were preferentially absorbed in regions of higher gas density. Usually this is not the case but in the *ionisation zones*, where hydrogen and helium are partially ionised, the opacity in fact becomes larger when the gas is compressed. If the ionisation zones are at a suitable depth in the atmosphere, the energy absorbed during compression and released during expansion of an ionisation zone can drive an oscillation. Stars with surface temperatures of 6000–9000 K are liable to this instability. The corresponding section of the HR diagram is called the cepheid instability strip.

Cepheids Among the most important pulsating variables are the cepheids, named after δ Cephei (Fig. 14.3). They are population I supergiants (stellar populations are discussed in Sect. 18.2) of spectral class F–K. Their periods are 1–50 days and their amplitudes, 0.1–2.5 magnitudes. The shape of the light curve is regular, showing a fairly rapid brightening, followed by a slower fall off. There is a relationship between the period of a cepheid and its absolute magnitude (i.e. luminosity), discovered in 1912 by *Henrietta Leavitt* from cepheids in the Small Magellanic Cloud. This *period–luminosity relation* (Fig. 14.4) can be used to measure distances of stars and nearby galaxies.

We have already noted that the pulsation period is related to the mean density. On the other hand the size of a star, and hence its mean density, is related to its total luminosity. Thus one can understand why there should be a relation between the period and the luminosity of a pulsating star.

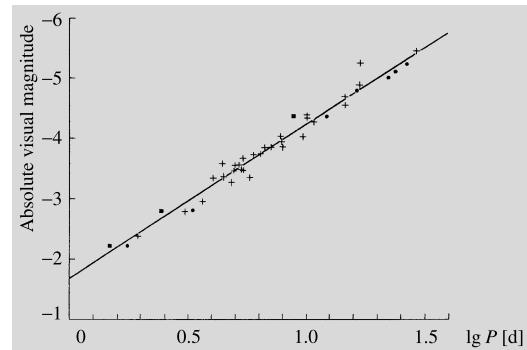


Fig. 14.4 The period–luminosity relation for cepheids. The black points and squares are theoretically calculated values, the crosses and the straight line represent the observed relation. (Drawing from Novotny, E. (1973): *Introduction to Stellar Atmospheres and Interiors* (Oxford University Press, New York) p. 359)

The magnitudes M and periods P of classical cepheids are shown in Fig. 14.4. The relation between M and $\log P$ is linear. However, to some extent, the cepheid luminosities also depend on colour: bluer stars are brighter. For accurate distance determinations, this effect needs to be taken into consideration.

W Virginis Stars In 1952 *Walter Baade* noted that there are in fact two types of cepheids: the classical cepheids and the W Virginis stars. Both types obey a period–luminosity relation, but the W Vir stars of a given period are 1.5 magnitudes fainter than the corresponding classical cepheids. This difference is due to the fact that the classical cepheids are young population I objects, whereas the W Vir stars are old stars of population II. Otherwise, the two classes of variables are similar.

Earlier, the W Vir period–luminosity relation had been used for both types of cepheids. Consequently the calculated distances to classical cepheids were too small. For example, the distance to the Andromeda Galaxy had been based on classical cepheids, since only these were bright enough to be visible at that distance. When the correct period–luminosity relation was used, all extragalactic distances had to be doubled. Distances within the Milky Way did not have to be changed, since their measurements were based on other methods.

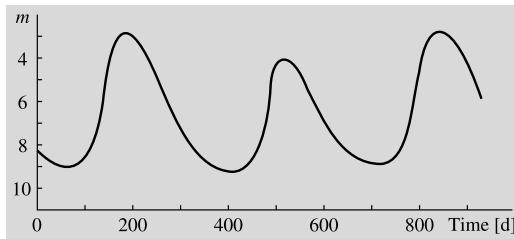


Fig. 14.5 The lightcurve of a long period Mira variable

RR Lyrae Stars The third important class of pulsating variables are the *RR Lyrae stars*. Their brightness variations are smaller than those of the cepheids, usually less than a magnitude. Their periods are also shorter, less than a day. Like the W Vir stars, the RR Lyrae stars are old population II stars. They are very common in the globular star clusters and were therefore previously called cluster variables.

The absolute magnitudes of the RR Lyrae stars are about $M_V = 0.6 \pm 0.3$. They are all of roughly the same age and mass, and thus represent the same evolutionary phase, where helium is just beginning to burn in the core. Since the absolute magnitudes of the RR Lyrae variables are known, they can be used to determine distances to the globular clusters.

Mira Variables (See Fig. 14.5.) The Mira variables (named after Mira Ceti) are supergiants of spectral classes M, S or C, usually with emission lines in their spectrum. They are losing gas in a steady stellar wind. Their periods are normally 100–500 days, and for this reason, they are also sometimes called long period variables. The amplitude of the light variations is typically about 6 magnitudes in the visual region. The period of Mira itself is about 330 days and its diameter is about 2 au. At its brightest, Mira has the magnitude 2–4, but at light minimum, it may be down to 12. The effective temperature of the Mira variables is only about 2000 K. Thus 95 % of their radiation is in the infrared, which means that a very small change in temperature can cause a very large change in visual brightness.

Other Pulsating Variables One additional large group of pulsating stars are the *semiregular*

and *irregular variables*. They are supergiants, often very massive young stars with unsteady pulsations in their extended outer layers. If there is some periodicity in the pulsations, these variables are called semiregular; otherwise they are irregular. An example of a semiregular variable is Betelgeuse (α Orionis). The pulsation mechanism of these stars is not well understood, since their outer layers are convective, and the theory of stellar convection is still poorly developed.

In addition to the main types of pulsating variables, there are some smaller separate classes, shown in Fig. 14.2.

The *dwarf cepheid* and the δ Scuti stars, which are sometimes counted as a separate type, are located below the RR Lyrae stars in the cepheid instability strip in the HR diagram. The dwarf cepheids are fainter and more rapidly varying than the classical cepheids. Their light curves often show a beating due to interference between the fundamental frequency and the first overtone.

The β Cephei stars are located in a different part of the HR diagram than the other variables. They are hot massive stars, radiating mainly in the ultraviolet. The variations are rapid and of small amplitude. The pulsation mechanism of the β Cephei stars is unknown.

The *RV Tauri* stars lie between the cepheids and the Mira variables in the HR diagram. Their period depends slightly on the luminosity. There are some unexplained features in the light curves of the RV Tauri stars, e.g. the minima are alternately deep and shallow.

14.3 Eruptive Variables

In the eruptive variables there are no regular pulsations. Instead sudden outbursts occur in which material is ejected into space. Nowadays such stars are divided into two main categories, *eruptive* and *cataclysmic variables*. Brightness changes of eruptive variables are caused by sudden eruptions in the chromosphere or corona, the contributions of which are, however, rather small in the stellar scale. These stars are usually surrounded by a gas shell or interstellar matter participating in the eruption. This group includes e.g. *flare stars*, various kinds of *nebular variables*,

Table 14.2 Main properties of eruptive variables (N , number of stars of the given type in Kukarkin's catalogue, Δm , change in brightness in magnitudes. The velocity is the expansion velocity in km/s, based on the Doppler shifts of the spectral lines)

Variable	N	Δm	Velocity
Supernovae	7	$\gtrsim 20$	4000–10,000
Ordinary novae	210	7–18	200–3500
Recurrent novae		$\lesssim 10$	600
Nova-like stars (P Cygni, symbiotic)	80	$\lesssim 2$	30–100
Dwarf novae (SS Cyg = U Gem, ZZ Cam)	330	2–6	(700)
R Coronae Borealis	40	1–9	–
Irregular (nebular variables, T Tau, RW Aur)	1450	$\lesssim 4$	(300)
Flare stars (UV Ceti)	750	$\lesssim 6$	2000

and *R Coronae Borealis stars*. Eruptions of the cataclysmic variables are due to nuclear reactions on the stellar surface or interior. Explosions are so violent that they can even destroy the whole star. This group includes *novae* and *nova-like stars*, *dwarf novae* and *supernovae* (Table 14.2).

Flare Stars The *flare* or *UV Ceti stars* are dwarf stars of spectral class M. They are young stars, mostly found in young star clusters and associations. At irregular intervals there are flare outbursts on the surface of the stars similar to those on the Sun. The flares are related to disturbances in the surface magnetic fields. The energy of the outbursts of the flare stars is apparently about the same as in solar flares, but because the stars are much fainter than the Sun, a flare can cause a brightening by up to 4–5 magnitudes. A flare lights up in a few seconds and then fades away in a few minutes (Fig. 14.6). The same star may flare several times in one day. The optical flare is accompanied by a radio outburst, like in the Sun. In fact, the flare stars were the first stars to be detected as radio sources.

Nebular Variables In connection with bright and dark interstellar clouds e.g. in the constellations of Orion, Taurus and Auriga, there are

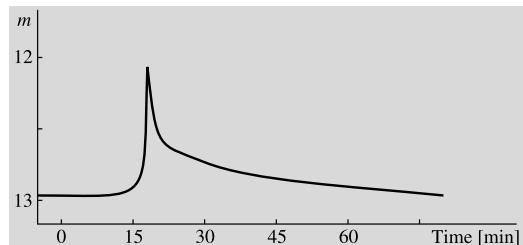


Fig. 14.6 The outbursts of typical flare stars are of short duration

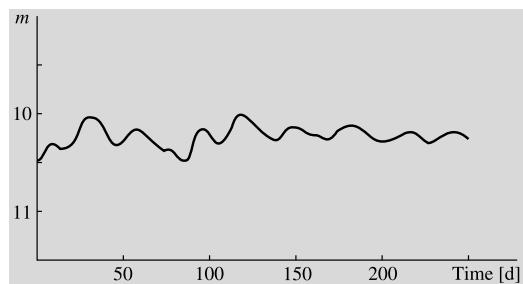


Fig. 14.7 Light curve of a T Tauri variable

variable stars. The *T Tauri stars* are the most interesting of them. These stars are newly formed or just contracting towards the main sequence. The brightness variations of the T Tauri stars are irregular (Fig. 14.7). Their spectra contain bright emission lines, formed in the stellar chromosphere, and forbidden lines, which can only be formed at extremely low densities. The spectral lines also show that matter is streaming out from the stars.

Since the T Tauri stars are situated inside dense gas clouds, they are difficult to observe. However, this situation has improved with the development of radio and infrared techniques.

Stars in the process of formation may change in brightness very rapidly. For example, in 1937, FU Orionis brightened by 6 magnitudes. This star is a strong source of infrared radiation, which shows that it is still enveloped by large quantities of interstellar dust and gas. A similar brightening by six magnitudes was observed in 1969 in V1057 Cygni (Fig. 14.8). Before its brightening, it was an irregular T Tauri variable; since then, it has remained a fairly constant tenth-magnitude AB star.

Stars of the *R Coronae Borealis* type have “inverse nova” light curves. Their brightness may drop by almost ten magnitudes and stay low for years, before the star brightens to its normal luminosity. For example, R CrB itself is of magnitude 5.8, but may fade to 14.8 magnitudes. Figure 14.9 shows its recent decline, based on observations by Finnish and French amateurs. The R CrB stars are rich in carbon and the decline is produced when the carbon condenses into a circumstellar dust shell.

One very interesting variable is η Carinae (Fig. 14.10). At present it is a six magnitude star surrounded by a thick, extensive envelope of dust and gas. In the early 19th century η Carinae was the second brightest star in the sky after Sirius. Around the middle of the century it rapidly dimmed to magnitude 8, but during the 20th century it has brightened somewhat. η Carinae is a so called bright blue variable. Its mass is estimated to be of the order of 100 solar masses. The circumstellar dust cloud is the brightest infrared source in the sky outside the solar system. The en-

ergy radiated by η Carinae is absorbed by the nebula and re-radiated at infrared wavelengths. The exact reason for the enormous brightness variations is not known. It is believed that the stability of the star is disturbed by the radiation pressure of the huge energy produced by a very massive star. When the nucleus will collapse at the end of the lifespan of the star, η Carinae will explode as a supernova.

Novae One of the best known types of eruptive variables are the *novae*. They are classified into several subtypes: *ordinary novae*, *recurrent novae* and *nova-like variables*. The *dwarf novae* (Fig. 14.11) are nova-like rather frequently eruptive stars; although the lightcurves are similar to those of novae, the mechanism is different.

The outbursts of all novae are rapid. Within a day or two the brightness rises to a maximum, which may be 7–16 magnitudes brighter than the normal luminosity. This is followed by a gradual decline, which may go on for months or years. The light curve of a typical nova is shown in Fig. 14.13. This light curve of Nova Cygni 1975 has been composed from hundreds of observations, mostly by amateurs.

In recurrent novae, the brightening is somewhat less than 10 magnitudes and in dwarf novae, 2–6 magnitudes. In both types there are repeated outbursts. For recurrent novae the time between outbursts is a few decades and for the dwarf novae 20–600 days. The interval depends on the strength of the outburst: the stronger the outburst, the longer the time until the next one. The bright-

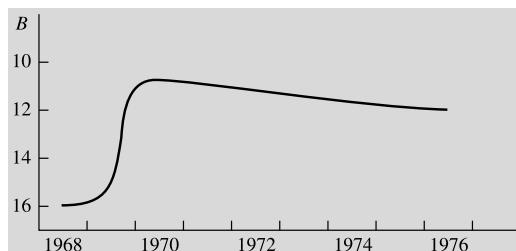


Fig. 14.8 In 1969–1970, the star V1057 Cygni brightened by almost 6 magnitudes

Fig. 14.9 The decline of R Coronae Borealis in 1977–1978; observations by Finnish and French amateur astronomers. (Kellomäki, Tähdet ja Avaruuus 5/1978)

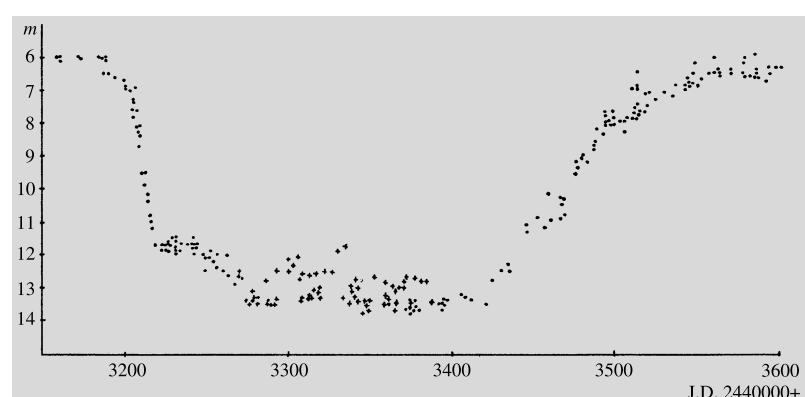


Fig. 14.10 In the 19th century, η Carinae was one of the brightest stars in the sky; since then it has dimmed considerably. In an outburst in 1843 the star ejected an expanding nebula, which has been called “Homunculus”. The Hubble photograph has been printed negative, to show finer details in the bipolar outflow.
 (Photograph
 NASA/HST/University of
 Minnesota)

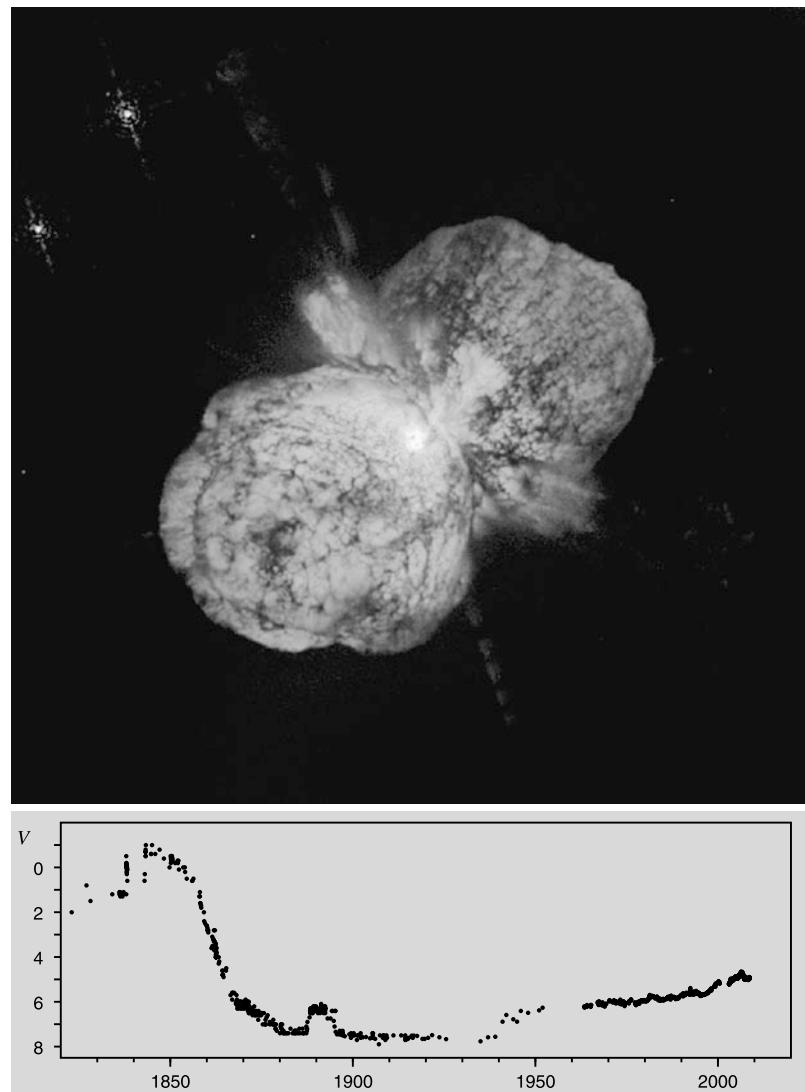
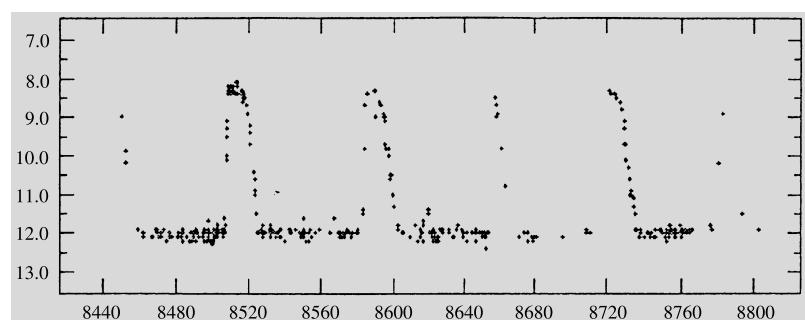


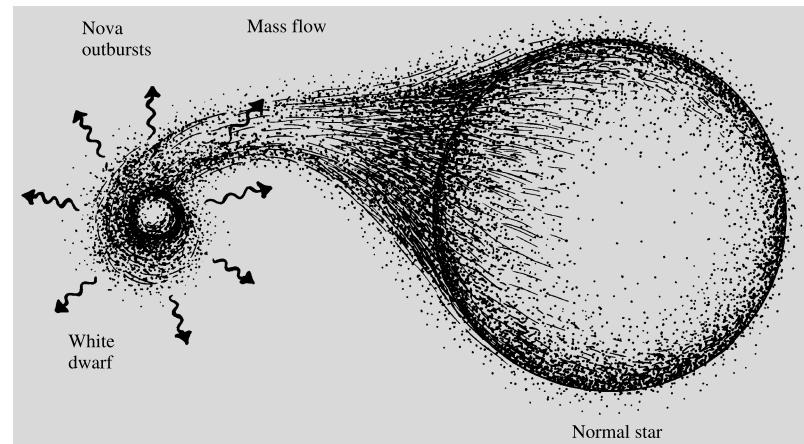
Fig. 14.11 The light curve of the dwarf nova SS Cygni in the beginning of 1966. (Drawing by Martti Perälä is based on observations by Nordic amateurs)



ening in magnitudes is roughly proportional to the logarithm of the recharging interval. It is pos-

sible that ordinary novae obey the same relationship. However, their amplitude is so much larger

Fig. 14.12 The novae are thought to be white dwarfs accreting matter from a nearby companion star. At times, nuclear reactions burning the accreted hydrogen are ignited, and this is seen as the flare-up of a nova



that the time between outbursts should be thousands or millions of years.

Observations have shown all novae and dwarf novae to be members of close binary systems. One component of the system is a normal star and the other is a white dwarf surrounded by a gas ring. (The evolution of close binary systems was considered in Sect. 12.6, where it was seen how this kind of system might have been formed.) The normal star fills its Roche surface, and material from it streams over to the white dwarf. When enough mass has collected on the surface of the white dwarf, the hydrogen is explosively ignited and the outer shell is ejected. The brightness of the star grows rapidly. As the ejected shell expands, the temperature of the star drops and the luminosity gradually decreases. However, the outburst does not stop the mass transfer from the companion star, and gradually the white dwarf accretes new material for the next explosion (Fig. 14.12).

The emission and absorption lines from the expanding gas shell can be observed in the spectrum of a nova. The Doppler shifts correspond to an expansion velocity of about 1000 km/s. As the gas shell disperses, its spectrum becomes that of a typical diffuse emission nebula. The expanding shell around a nova can also sometimes be directly seen in photographs.

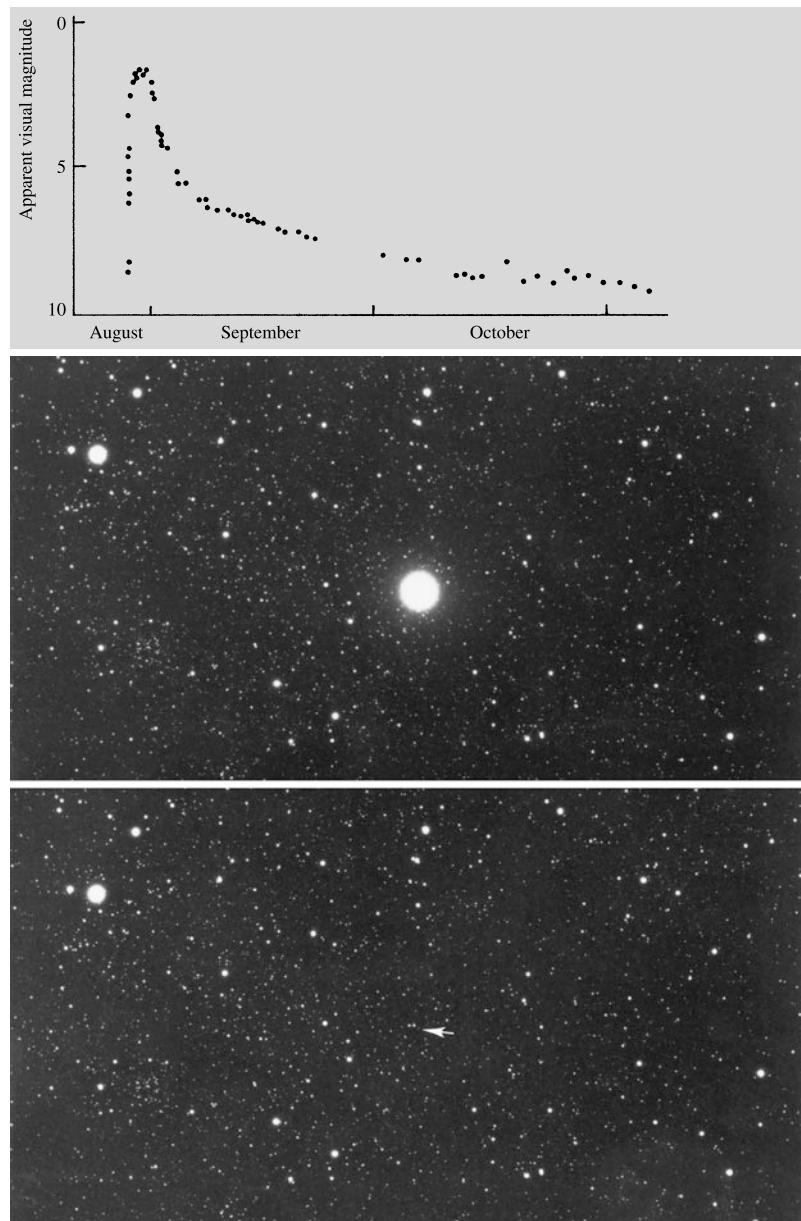
In dwarf novae the energy does not come from nuclear reactions but mainly from the potential energy of the matter falling into the white dwarf. An eruption occurs when the density of the accre-

tion disk formed by the matter flowing from the companion star exceeds a certain critical limit, becomes unstable and is strongly heated.

A considerable fraction of the novae in our Galaxy are hidden by interstellar clouds and their number is therefore difficult to estimate. In the Andromeda Galaxy, observations indicate 25–30 nova explosions per year. The number of dwarf novae is much larger. In addition there are nova-like variables, which share many of the properties of novae, such as emission lines from circumstellar gas and rapid brightness variations. These variables, some of which are called *symbiotic stars*, are close binaries with mass transfer. Gas streaming from the primary hits a gas disk around the secondary in a hot spot, but there are no nova outbursts.

The supersoft stars (SSS) are a somewhat peculiar subclass of cataclysmic variables. Their X-ray radiation is much stronger and much softer, i.e. has longer wavelength, than that of the ordinary cataclysmic variables. In the SSS objects the component losing mass is more massive than the compact star (white dwarf), in contrast with ordinary interacting binaries. Therefore the mass transfer is an unstable self sustained process transferring more material to the surface of the white dwarf than in the normal cataclysmic variables. On the surface of the white dwarf a continuous fusion reaction resembling nova eruptions is producing a lot of soft X-ray radiation. Due to their unstable character the SSS objects are short lived and therefore relatively rare.

Fig. 14.13 In 1975 a new variable, Nova Cygni or V1500 Cygni, was discovered in Cygnus. In the upper photograph the nova is at its brightest (about 2 magnitudes), and in the lower photograph it has faded to magnitude 15. (Photographs Lick Observatory)



14.4 Supernovae

Explosions of stars as supernovae are among the most energetic phenomena of the universe. Within a couple of weeks the star becomes so bright that its luminosity corresponds to over a billion Suns or a whole small galaxy. Thus they can be observed even at cosmological distances, and they are useful standard candles for measur-

ing the dimensions of the universe and cosmological quantities. Supernovae are also important sources of heavy elements. Most elements heavier than iron have originated in supernova explosions.

The maximum is followed by a slow decay, and the nearest supernovae can be observed still years after the explosion. In the explosion a gas shell expanding with a velocity of thousands of

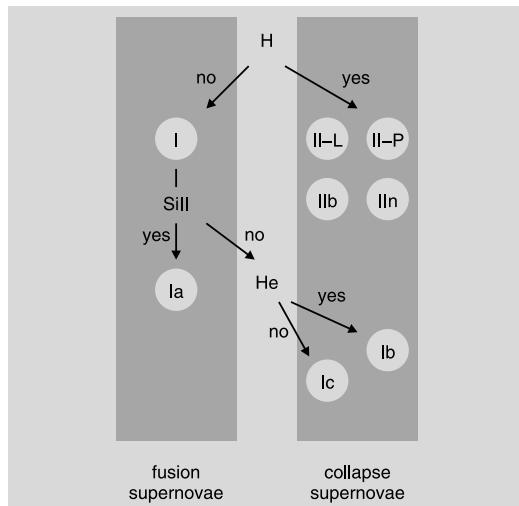


Fig. 14.14 The type of a supernova can be deduced from different spectral lines in its spectrum and the shape of its lightcurve

kilometres per second is ejected. Over 200 of such supernova remnants have been discovered in the Milky Way. Their ages vary from a few hundred to tens of thousands of years.

At least six supernova explosions have been observed in the Milky Way. Best known are the “guest star” seen in China in 1054 (whose remnant is the Crab nebula), Tycho Brahe’s supernova in 1572 and Kepler’s supernova in 1604. On the basis of observations of other Sb–Sc-type spiral galaxies, the interval between supernova explosions in the Milky Way is predicted to be about 50 years. Some will be hidden by obscuring material, especially near the centre of the galaxy. In particularly dusty starburst galaxies almost all supernovae remain undetected due to the extinction of the dust. Still the 400 years’ interval since the last observed supernova in the Milky Way is unusually long.

Supernovae in other galaxies are detected quite frequently. For example, both in 2006 and 2007 over 500 new supernovae were found. In the near future the number will increase rapidly thanks to new all sky surveys.

Earlier supernovae were divided just to two main categories, I and II, based on whether their spectra showed evidence of hydrogen. Later these main classes have been divided into several sub-

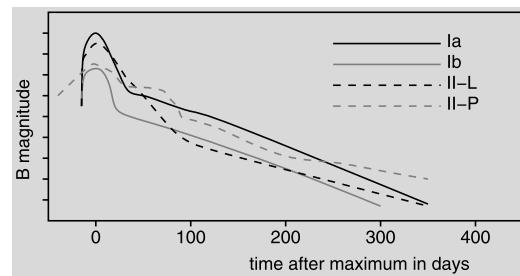


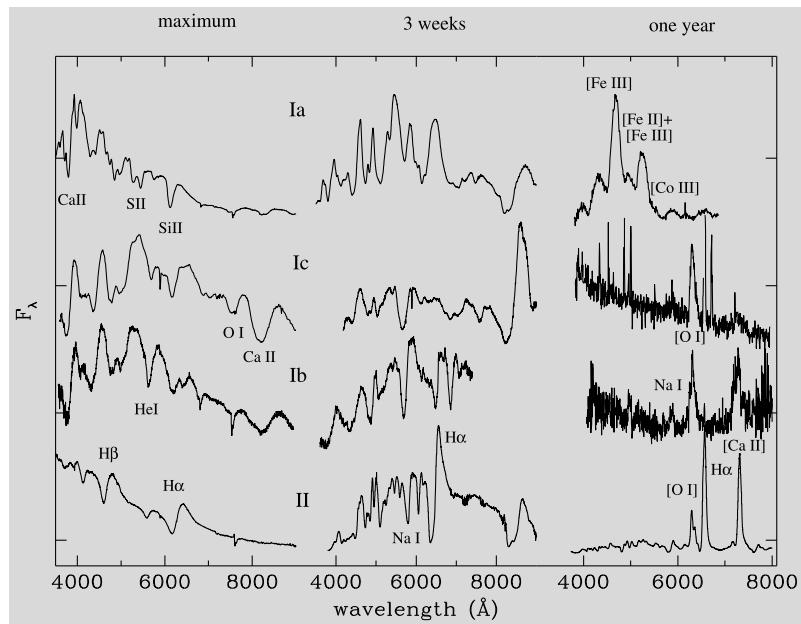
Fig. 14.15 Typical shapes of the lightcurves of different supernovae

classes depending on other spectral features and their lightcurves (Figs. 14.14, 14.15 and 14.16).

Type Ia supernovae are recognised by the silicon absorption line (615 nm), which is strong in the spectra of young novae. It is thought that these supernovae originate in binaries, at least a billion years old, consisting of a white dwarf and its companion. In such a close binary material can flow from the companion to the white dwarf until its mass exceeds the limit (about $1.4M_{\odot}$). Then the pressure of the degenerate of the electron gas inside the white dwarf cannot overcome the gravitation and star will collapse. Increase of the density and temperature will ignite an explosive fusion reaction that can destroy the white dwarf. The kinetic energy released in the process is of the order of 10^{44} J causing part of the gas to expand even at a velocity of $0.1c$. The origin of the radiation energy (around 10^{42} J) is the fission of the radioactive nickel isotope 56 created in the explosion (typically about $0.5M_{\odot}$) into radioactive cobalt and further into stable iron. The more nickel is produced in the explosion the brighter the supernova shines at its maximum; the absolute magnitude will reach -19 – -20 . The shape of the lightcurve depends on the brightness of the supernova: the brighter the supernova the broader the top of the lightcurve. Therefore the maximum brightness of a type Ia supernova can be determined precisely from its lightcurve. This way they can be used as standard candles to determine dimensions of the universe and cosmological parameters (Chap. 20).

In Chap. 12 we discussed the final stages of stellar evolution and how the collapse of a massive star will lead to an explosion. It is believed

Fig. 14.16 Spectra of different types of supernovae at the maximum, three weeks after the maximum and after one year. (Turatto: Supernovae and Gamma-Ray Bursters, ed. K. Weiler, Lecture Notes in Physics, vol. 598, pp. 21–36)



that all supernovae except type Ia are caused by the collapse of massive and short lived stars. At the end of its lifespan such a star has an iron core that will collapse under its own gravity after exceeding the Chandrasekhar limit. In the stellar core the matter will reach the density of atomic nuclei and the outer layers bounce back sending an outbound shock wave. The process creates a huge number of neutrinos. Since they interact only weakly with the outer layers of the star they rush to the surrounding space. The remaining core may become a neutron star or, in the case of a very massive star, a black hole. Most of the released energy (about 99 %) escapes with the neutrinos. The outer layers of the star receive a kinetic energy of about 10^{44} J, part of which may be released as radiation when the expanding matter hits the material surrounding the star and later the interstellar matter. Typically the supernova itself emits radiation about 10^{42} J within a few months after the explosion, reaching an absolute magnitude between -16 and -20 at the maximum. Later the brightness is sustained by the fission of the radioactive nickel 56, born in the explosion, to other elements just as in the case of the type Ia supernovae.

If the spectrum of a supernova does not contain strong hydrogen or silicon lines its type is

either Ib or Ic. The type Ib supernovae are recognised by strong helium lines that are not present in the spectra of type Ic. These supernovae originate in stars that don't have any more hydrogen in their outer layer. The stars exploding as type Ic supernovae have also lost their outer layer of helium. Such stars that have lost their outer layers are called Wolf–Rayet stars.

All type II supernovae have hydrogen lines in their spectra. They are divided into main types II-P and II-L based on the flat (plateau) and linear shape of the lightcurve. Recently several stars that later exploded as supernovae have been identified in high resolution pictures taken e.g. by the Hubble space telescope and the VLT telescope before the explosion (Fig. 14.17). These observations have shown that the more common type II-P supernovae originate in red supergiant stars as predicted by theories of stellar evolution. Initially the masses of such stars have been at least 8 solar masses. Smaller stars are not believed to explode in a collapse. They end up as white dwarfs after more quiescent evolution.

Type IIn supernovae are recognised by their hydrogen emission lines that are much narrower than in the spectra of other supernovae. The lines originate in the matter around the star. The spectra of young type IIn supernovae have hydrogen

lines, which, however, disappear in a few weeks, and afterwards the spectrum resembles the type Ib spectra. It is assumed that these supernovae were originally massive stars that still had a small amount of hydrogen in their outer layer before the explosion. Thus they are intermedia between the types II and Ib. The observed properties and types of supernovae depend on the properties of their progenitors, and hence individual supernovae can of intermediate types. Hence it makes more sense

to classify supernovae according to the explosion mechanism simply to two different types, *thermonuclear* and *core collapse* supernovae. About 30 % of the supernovae found in nearby galaxies are thermonuclear and the remaining 70 % core collapse supernovae.

On February 23, 1987 the first burst of light from a supernova in the Large Magellanic Cloud, the small companion galaxy of the Milky Way, reached the Earth (Fig. 14.18). This supernova,

Fig. 14.17 *Left:* The supernova 2003gd of type II-P after the explosion. *Right:* The same region before the explosion. The exploded red supergiant (progenitor) was located in the image taken before the explosion. Both images were taken with the Hubble space telescope. (Smartt et al., 2004, Science, vol. 303, 5657, 499)

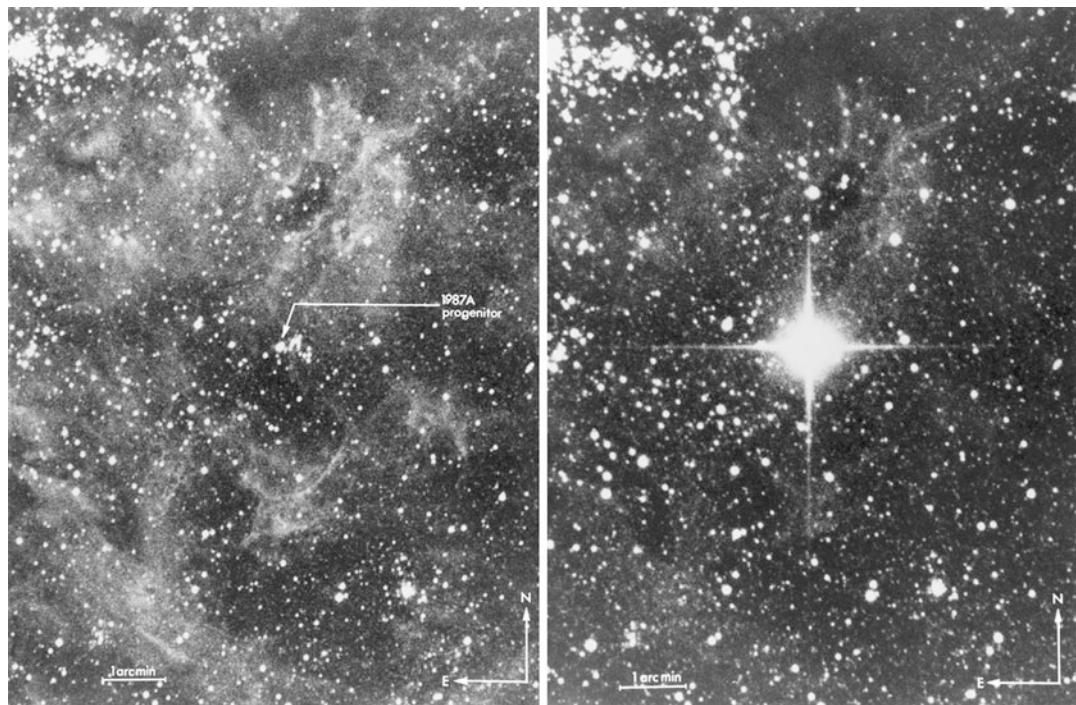
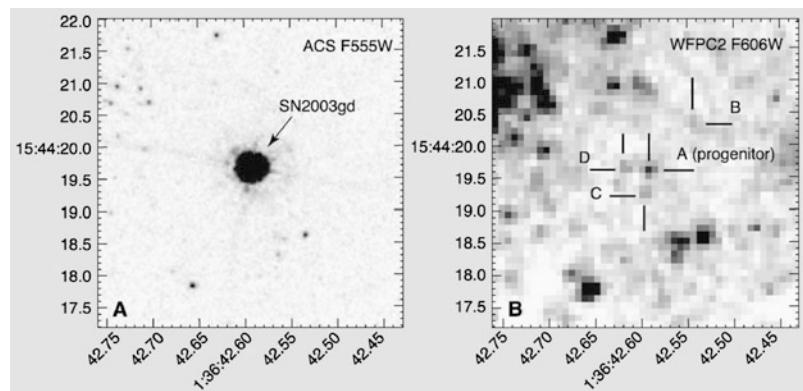


Fig. 14.18 Supernova 1987A in the Large Magellanic Cloud before and after the explosion. (Photographs ESO)

SN 1987A, was of type II, and was the brightest supernova for 383 years. After its first detection, SN 1987A was studied in great detail by all available means. Although the general ideas of Sects. 11.4 and 11.5 on the final stages of stellar evolution have been confirmed, there are complications. Thus e.g. the progenitor star was a blue rather than a red giant as expected, perhaps because of the lower abundance of heavy elements in the Large Magellanic Cloud compared to that in the Milky Way. The collapse of its core released a vast amount of energy as a pulse of neutrinos, which was detected in Japan and the USA. The amount of energy released indicates that the remnant is a neutron star.

14.5 Examples

Example 14.1 The observed period of a cepheid is 20 days and its mean apparent magnitude $m = 20$. From Fig. 14.4, its absolute magnitude is $M \approx -5$. According to (4.12), the distance of the cepheid is

$$\begin{aligned} r &= 10 \times 10^{(m-M)/5} = 10 \times 10^{(20+5)/5} \\ &= 10^6 \text{ pc} = 1 \text{ Mpc}. \end{aligned}$$

Example 14.2 The brightness of a cepheid varies 2 mag. If the effective temperature is 6000 K at the maximum and 5000 K at the minimum, how much does the radius change?

The luminosity varies between

$$\begin{aligned} L_{\max} &= 4\pi R_{\max}^2 \sigma T_{\max}^4, \\ L_{\min} &= 4\pi R_{\min}^2 \sigma T_{\min}^4. \end{aligned}$$

In magnitudes the difference is

$$\begin{aligned} \Delta m &= -2.5 \lg \frac{L_{\min}}{L_{\max}} = -2.5 \lg \frac{4\pi R_{\min}^2 \sigma T_{\min}^4}{4\pi R_{\max}^2 \sigma T_{\max}^4} \\ &= -5 \lg \frac{R_{\min}}{R_{\max}} - 10 \lg \frac{T_{\min}}{T_{\max}}. \end{aligned}$$

This gives

$$\begin{aligned} \lg \frac{R_{\min}}{R_{\max}} &= -0.2 \Delta m - 2 \lg \frac{T_{\min}}{T_{\max}} \\ &= -0.4 - 2 \lg \frac{5000}{6000} = -0.24, \end{aligned}$$

whence

$$\frac{R_{\min}}{R_{\max}} = 0.57.$$

14.6 Exercises

Exercise 14.1 The absolute visual magnitude of RR Lyrae variables is 0.6 ± 0.3 . What is the relative error of distances due to the deviation in the magnitude?

Exercise 14.2 The bolometric magnitude of a long period variable varies by one magnitude. The effective temperature at the maximum is 4500 K.

- (a) What is the temperature at the minimum, if the variation is due to temperature change only?
- (b) If the temperature remains constant, what is the relative variation in the radius?

Exercise 14.3 In 1983 the radius of the Crab nebula was about $3'$. It is expanding $0.21''$ a year. Radial velocities of 1300 km s^{-1} with respect to the central star have been observed in the nebula.

- (a) What is the distance of the nebula, assuming its expansion is symmetric?
- (b) A supernova explosion has been observed in the direction of the nebula. Estimate, how long time ago?
- (c) What was the apparent magnitude of the supernova, if the absolute magnitude was a typical -18 ?

In astrophysics those stars in which the density of matter is much larger than in ordinary stars are known as compact objects. These include white dwarfs, neutron stars, and black holes. In addition to a very high density, the compact objects are characterised by the fact that nuclear reactions have completely ceased in their interiors. Consequently they cannot support themselves against gravity by thermal gas pressure. In the white dwarfs and neutron stars, gravity is resisted by the pressure of a degenerate gas. In the black holes the force of gravity is completely dominant and compresses the stellar material to infinite density.

Compact stars in binary systems give rise to a variety of striking new phenomena. If the companion star is losing mass by a stellar wind or a Roche lobe overflow, the gas that is shed may be accreted by the compact object. This will release gravitational energy that can be observable in the form of X-ray emission and strong and rapid brightness variations.

15.1 White Dwarfs

As was mentioned in Sect. 11.2, in ordinary stars the pressure of the gas obeys the equation of state of an ideal gas. In stellar interiors the gas is fully ionised, i.e. it is plasma consisting of ions and free electrons. The partial pressures of the ions and electrons together with the radiation pressure important in hot stars comprise the total pressure balancing gravitation. When the star runs out of its nuclear fuel, the density in the interior increases, but the temperature does not change

much. The electrons become degenerate, and the pressure is mainly due to the pressure of the degenerate electron gas, the pressure due to the ions and radiation being negligible. The star becomes a *white dwarf*.

As will be explained in Box 15.1 the radius of a degenerate star is inversely proportional to the cubic root of the mass. Unlike in a normal star the radius decreases as the mass increases.

The first white dwarf to be discovered was Sirius B, the companion of Sirius (Fig. 15.1). Its exceptional nature was realised in 1915, when it was discovered that its effective temperature was very high. Since it is faint, this meant that its radius had to be very small, slightly smaller than that of the Earth. The mass of Sirius B was known to be about equal to that of the Sun, so its density had to be extremely large.

The high density of Sirius B was confirmed in 1925, when the gravitational redshift of its spectral lines was measured. This measurement also provided early observational support to Einstein's general theory of relativity.

White dwarfs occur both as single stars and in binary systems. Their spectral lines are broadened by the strong gravitational field at the surface. In some white dwarfs the spectral lines are further broadened by rapid rotation. Strong magnetic fields have also been observed.

White dwarfs have no internal sources of energy, but further gravitational contraction is prevented by the pressure of the degenerate electron gas. Radiating away the remaining heat, white dwarfs will slowly cool, changing in colour from

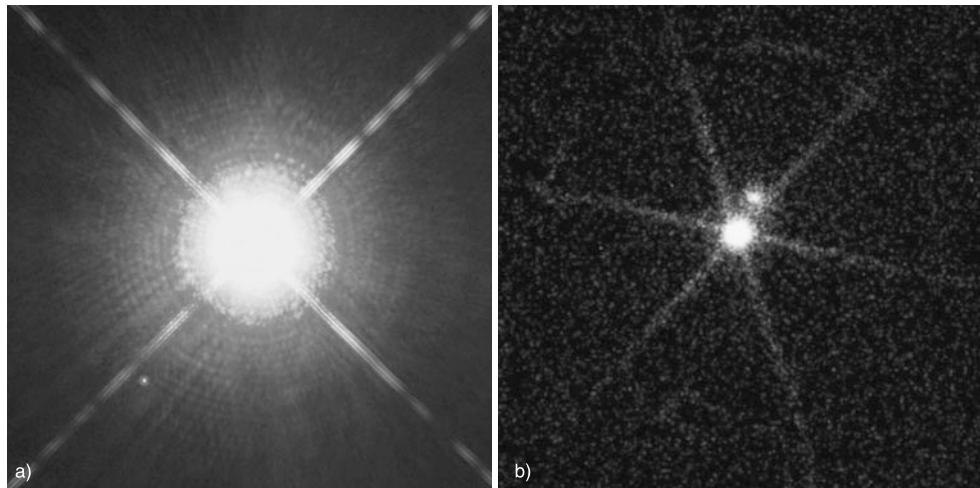


Fig. 15.1 Two views of the best-known white dwarf Sirius B, the small companion to Sirius. *On the left*, a picture in visible light by the Hubble Space Telescope. Sirius B is the tiny white dot on lower left from the overexposed image of Sirius. *On the right*, an X-ray picture of the pair

taken by the Chandra X-ray observatory. Sirius B is now the brighter source, and Sirius is weaker, because its surface is much cooler than the surface of the white dwarf. (Photos NASA / HST and Chandra)

white to red and finally to black. The cooling time is comparable to the age of the Universe, and even the oldest white dwarfs should still be observable. Looking for the faintest white dwarfs has been used as a way to set a lower limit on the age of the Universe.

Cataclysmic Variables When a white dwarf is a member of a close binary system, it can accrete mass from its companion star. The most interesting case is where a main sequence star is filling its *Roche lobe*, the largest volume it can have without spilling over to the white dwarf. As the secondary evolves, it expands and begins to lose mass, which is eventually accreted by the primary. Binary stars of this kind are known as *cataclysmic variables*.

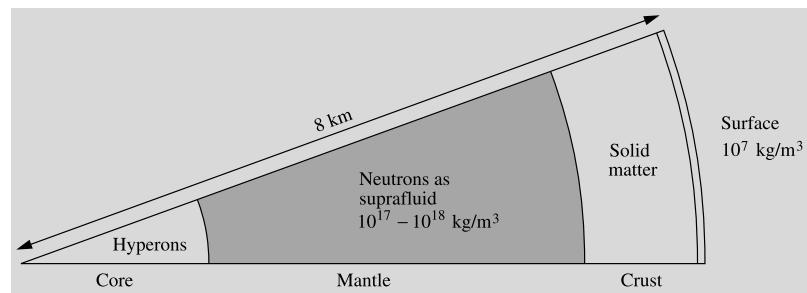
The present definition of the class of cataclysmic variables has gradually evolved, and in consequence many types of systems that were earlier viewed as separate are now collected under this heading. In principle, even type Ia supernovae should be included. The *classical novae*, whose eruptions are caused by the sudden ignition of hydrogen that has collected on the surface of the white dwarf, have been described in Sect. 14.3. In the eruptions most of the accreted

gas is expelled in a shell, but if the mass transfer continues in the system further eruptions may occur, giving rise to *recurrent novae*. Finally, cataclysmic variables without eruptions, for example pre-novae or post-novae, are classified as *nova-like variables*.

The *dwarf novae*, also described in Sect. 14.3, are produced by a quite different mechanism. In their case the outbursts are not caused by thermonuclear reactions, but by instabilities in the accretion flow around the white dwarf. Although the details of the outburst mechanism are still not completely clear, the basic picture is that the disk has two possible states, a hot and a cool one, available. Under some conditions the disk cannot remain permanently in either of these states, and has to jump repeatedly between the hot outburst state and the cool quiescent state.

A special type of nova-like variables are the magnetic cataclysmic variables. In the *polars* the magnetic field is so strong that the accreted gas cannot settle into an accretion disk. Instead it is forced to follow the magnetic field lines, forming an *accretion column*. As the gas hits the surface of the white dwarf it is strongly heated giving rise to bright X-ray emission, which is a characteristic feature of polars. Systems with a slightly

Fig. 15.2 The structure of a neutron star. The crust is rigid solid material and the mantle a freely streaming superfluid



weaker magnetic field are called *intermediate polars*. These systems exhibit both X-ray emission and variations due to an accretion disk.

15.2 Neutron Stars

If the mass of a star is large enough, the density of matter may grow even larger than in normal white dwarfs. The equation of state of a classical degenerate electron gas then has to be replaced with the corresponding relativistic formula. In this case decreasing the radius of the star no longer helps in resisting the gravitational attraction. Equilibrium is possible only for one particular value of the mass, the Chandrasekhar mass M_{Ch} , already introduced in Sect. 12.5. The value of M_{Ch} is about $1.4 M_{\odot}$, which is thus the upper limit to the mass of a white dwarf. If the mass of the star is larger than M_{Ch} , gravity overwhelms the pressure and the star will rapidly contract towards higher densities. The final stable state reached after this collapse will be a *neutron star* (Fig. 15.2). On the other hand, if the mass is smaller than M_{Ch} , the pressure dominates. The star will then expand until the density is small enough to allow an equilibrium state with a less relativistic equation of state.

When a massive star reaches the end of its evolution and explodes as a supernova, the simultaneous collapse of its core will not necessarily stop at the density of a white dwarf. If the mass of the collapsing core is larger than the Chandrasekhar mass ($\gtrsim 1.4 M_{\odot}$), the collapse continues to a neutron star.

An important particle reaction during the final stages of stellar evolution is the *URCA process*, which was put forward by Schönberg and Gamow in the 1940's and which produces a large

neutrino emission without otherwise affecting the composition of matter. (The URCA process was invented in Rio de Janeiro and named after a local casino. Apparently money disappeared at URCA just as energy disappeared from stellar interiors in the form of neutrinos. It is claimed that the casino was closed by the authorities when this similarity became known.) The URCA process consists of the reactions

$$(Z, A) + e^- \rightarrow (Z-1, A) + \nu_e,$$

$$(Z-1, A) \rightarrow (Z, A) + e^- + \bar{\nu}_e,$$

where Z is the number of protons in a nucleus; A the mass number; e^- an electron; and ν_e and $\bar{\nu}_e$ the electron neutrino and antineutrino. When the electron gas is degenerate, the latter reaction is suppressed by the Pauli exclusion principle. In consequence the protons in the nuclei are transformed into neutrons. As the number of neutrons in the nuclei grows, their binding energies decrease. At densities of about $4 \times 10^{14} \text{ kg/m}^3$ the neutrons begin to leak out of the nucleus, and at 10^{17} kg/m^3 the nuclei disappear altogether. Matter then consists of a neutron "porridge", mixed with about 0.5 % electrons and protons.

Neutron stars are supported against gravity by the pressure of the degenerate neutron gas, just as white dwarfs are supported by electron pressure. The equation of state is the same, except that the electron mass is replaced by the neutron mass, and that the mean molecular weight is defined with respect to the number of free neutrons. Since the gas consists almost entirely of neutrons, the mean molecular weight is approximately one.

The typical diameters of neutron stars are about 10 km. Unlike ordinary stars they have

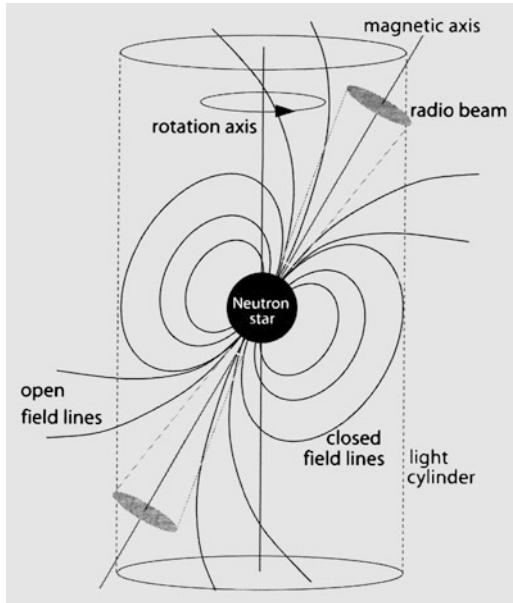


Fig. 15.3 A rotating neutron star is surrounded by a strong magnetic field which drags electrons from the surface and accelerates them to relativistic speeds over the magnetic poles. When the electrons are accelerated along the magnetic field lines, they radiate so called curvature radiation in a narrow beam. Since the magnetic axis is misaligned with the rotation axis, the beams sweep around the sky like in a lighthouse. (Lorimer–Kramer 2005, Handbook of Pulsar Astronomy, Cambridge University Press, p. 55)

a well-defined solid surface. The atmosphere above it is a few centimetres thick. The upper crust is a metallic solid with the density growing rapidly inwards. Most of the star is a neutron superfluid, and in the centre, where the density exceeds 10^{18} kg/m^3 , there may be a solid nucleus of heavier particles (hyperons), or of quark matter, where the quarks that normally constitute neutrons have become unconfined.

A neutron star formed in the explosion and collapse of a supernova will initially rotate rapidly, because its angular momentum is unchanged while its radius is much smaller than before. In a few hours the star will settle in a flattened equilibrium, rotating several hundred times per second. The initial magnetic field of the neutron star will also be compressed in the collapse, so that there will be a strong field coupling the star to the surrounding material. The angular momentum of the neutron star is steadily de-

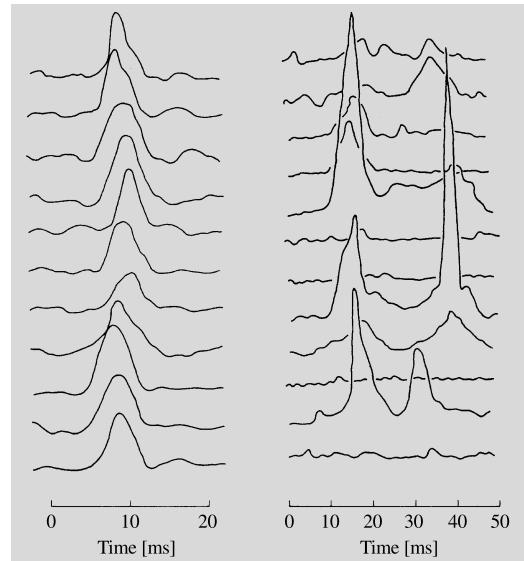


Fig. 15.4 Consecutive radio pulses at 408 MHz from two pulsars. To the left PSR 1642-03 and to the right PSR 1133+16. Observations made at Jodrell Bank. (Picture from Smith, F.G. (1977): Pulsars (Cambridge University Press, Cambridge) pp. 93, 95)

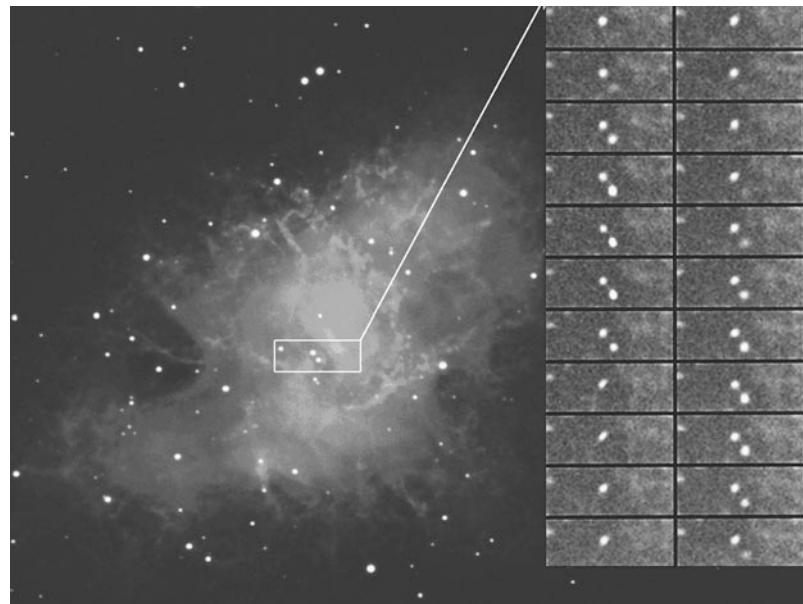
creased by the emission of electromagnetic radiation, neutrinos, cosmic ray particles and possibly gravitational radiation. Thus the angular velocity decreases. The rotation can also break the star into several separate objects. They will eventually recombine when the energy of the system is reduced. In some cases the stars can remain separated, resulting e.g. in a binary neutron star.

The theory of neutron stars was developed in the 1930's, but the first observations were not made until the 1960's. At that time *the pulsars*, a new type of rapidly pulsating radio sources, were discovered and identified as neutron stars. In the 1970's neutron stars were also seen as *X-ray pulsars*, *X-ray bursters* and *magnetars*.

Pulsars The pulsars were discovered in 1967, when *Anthony Hewish* and *Jocelyn Bell* in Cambridge, England, detected sharp, regular radio pulses coming from the sky. Since then about 1500 pulsars have been discovered (Fig. 15.4). Their periods range from 0.0016 s (for the pulsar 1937 + 214) up to 20 minutes.

In addition to the steady slowing down of the rotation, sometimes small sudden jumps in the

Fig. 15.5 A time-sequence of the pulsation of the Crab pulsar in visible light. The pictures were taken once about every millisecond; the period of the pulsar is about 33 milliseconds. (Photos N.A. Sharp/NOAO/AURA/NSF)



period are observed. These might be a sign of rapid mass movements in the neutron star crust (“starquakes”) or in its surroundings.

The origin of the radio pulses can be understood if the magnetic field is tilted at an angle of 45° – 90° with respect to the rotation axis. The field is so strong that it drags electrons from the surface and accelerates them to relativistic speeds over the magnetic poles. When the electrons are accelerated along the magnetic field lines, they radiate so called curvature radiation which is related to synchrotron radiation (Fig. 15.3). In the direction of the magnetic poles two thin beams of radio radiation are emitted. The beams sweep around the sky, and if the Earth happens to be in the path of the beam a pulsar is seen.

The best-known pulsar is located in the Crab nebula (Figs. 15.5 and 15.7). This small nebula in the constellation Taurus was noted by the French astronomer *Charles Messier* in the middle of the 18th century and became the first object in the Messier catalogue, M1. The Crab nebula was found to be a strong radio source in 1948 and an X-ray source in 1964. The pulsar was discovered in 1968. In the following year it was observed optically and was also found to be an X-ray emitter.

Neutron stars are difficult to study optically, since their luminosity in the visible region is very

small (typically about $10^{-6} L_\odot$). For instance the Vela pulsar has been observed at a visual magnitude of about 25. In the radio region, it is a very strong pulsating source.

A few pulsars have been discovered in binary systems; the first one, PSR 1913+16, in 1974. In 1993 *Joseph Taylor* and *Russell Hulse* were awarded the Nobel prize for the detection and studies of this pulsar. The pulsar orbits about a companion, presumably another neutron star, with the orbital eccentricity 0.6 and the period 8 hours. The observed period of the pulses is altered by the Doppler effect, and this allows one to determine the velocity curve of the pulsar. These observations can be made very accurately, and it has therefore been possible to follow the changes in the orbital elements of the system over a period of several years. For example, the periastron direction of the binary pulsar has been found to rotate about 4° per year. This phenomenon can be explained by means of the general theory of relativity; in the solar system, the corresponding rotation (the minor fraction of the rotation not explained by the Newtonian mechanics) of the perihelion of Mercury is 43 arc seconds per century.

The binary pulsar PSR 1913+16 has also provided the first strong evidence for the existence of

gravitational waves. During the time of observation the orbital period of the system has steadily decreased. This shows that the system is losing orbital energy at a rate that agrees exactly with that predicted by the general theory of relativity. The lost energy is radiated as gravitational waves.

Magnetars The energy emitted by common pulsars has its origin in the slowing down of their rotation. In some neutron stars, the magnetars, the magnetic field is so strong that the energy released in the decay of the field is the main source of energy. Whereas in ordinary pulsars the magnetic field is typically 10^9 T, in magnetars a typical value may be 10^9 – 10^{11} T.

Magnetars were first invoked as an explanation of the soft gamma repeaters (SGR), X-ray stars that irregularly emit bright, short (0.1 s) repeating flashes of low-energy gamma rays. Later a second class of mysterious objects, the anomalous X-ray pulsars (AXP), were identified as magnetars. AXP are slowly rotating pulsars, with a rotation period of 6 to 12 seconds. Despite this they are bright X-ray sources, which can be understood if their energy is of magnetic origin.

It is thought that magnetars are the remnants of stars that were more massive and rapidly rotating than those giving rise to ordinary pulsars, although the details are still subject to debate. A magnetar first appears as a SGR. During this phase, which only lasts about 10,000 years, the very strong magnetic field is slowing down the rate of rotation. At the same time the field is drifting with respect to the neutron star crust. This causes shifts in the crust structure, leading to powerful magnetic flares and the observed outbursts. After about 10,000 years the rotation has slowed down so much that the outbursts cease, leaving the neutron star observable as an AXP.

Gamma Ray Bursts For a long time the *gamma ray bursts* (GRB), very short and sharp gamma ray pulses first discovered in 1973, remained a mystery. Unlike the much less common SGR, the GRB never recurred, and they had no optical or X-ray counterparts. A first major advance was made when satellite observations with the Compton Gamma Ray Observatory showed

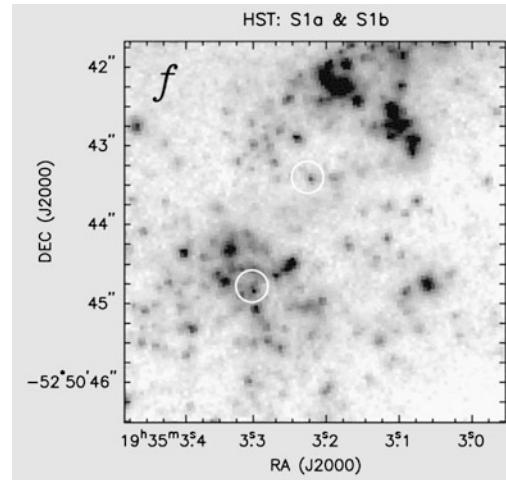


Fig. 15.6 The location of the peculiar type Ibc supernova SN 1998bw at redshift $z = 0.0085$ is in the circle to the lower left. This was also the position of the faint gamma-ray burst GRB 980425, the first GRB to be connected with a supernova. The circle on the upper right marks an ultraluminous X-ray source. (C. Kouveliotou et al. 2004, ApJ 608, 872, Fig. 1)

that the gamma ray bursts are almost uniformly distributed in the sky, unlike the known neutron stars.

The nature of the gamma ray bursts is now becoming clear thanks to dedicated observing programmes that have used burst detections by gamma and X-ray satellites such as Beppo-SAX and, in particular, Swift rapidly to look for afterglows of the GRB at optical wavelengths. The detection of these afterglows has made it possible to determine distances to the bursts and their location in their host galaxies (see Fig. 15.6).

It has become clear that there are at least two kinds of bursts, with the self-descriptive names long soft bursts, and short hard bursts. The long soft gamma ray bursts, lasting longer than 2 seconds, have now been convincingly shown to be produced in the explosions of massive stars at the end of their life, specifically supernovae of types Ib and Ic (Sect. 14.3). Only a small fraction of all type Ibc supernovae give rise to a GRB. The explosions that produce GRB have been called hypernovae, and are among the brightest objects in the Universe. A gamma ray burst observed in late 2005 took place when the Universe was only 900 million years old, making it one of the most

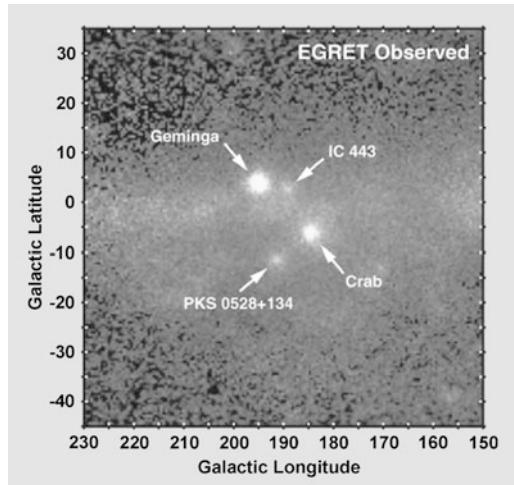


Fig. 15.7 Some pulsars shine brightly in gamma-rays. In the center the Crab pulsar and on the upper left the gamma source Geminga, which was identified in 1992 to be the nearest pulsar with a distance of about 100 pc from the Sun. (Photo by Compton Gamma Ray Observatory)

distant objects ever observed. The conditions required for hypernova explosions are still not certain.

The nature of the systems giving rise to short gamma ray bursts, lasting less than 2 seconds, have been more difficult to ascertain. The most popular theory has been that they are produced in compact binary systems consisting of two neutron stars or a neutron star and a black hole. These systems lose energy by gravitational radiation, and eventually the two components should merge, producing a burst of gamma radiation. This theory has now received strong support when the afterglow of a few short bursts has been detected in the outer parts of their host galaxies. Since the stars in these regions are all old and no longer give rise to core-collapse supernovae, the neutron star merger hypothesis appears most likely. However, it is still also possible that some of the short bursts are exceptionally bright magnetar flares.

Box 15.1 (The Radius of White Dwarfs and Neutron Stars) The mass of a white dwarf or a neutron star determines its radius. This follows from the equation of hydrostatic equilibrium and from the pressure-density relation for

a degenerate gas. Using the hydrostatic equilibrium equation (11.1)

$$\frac{dP}{dr} = -\frac{GM_r\rho}{r^2}$$

one can estimate the average pressure P :

$$\left| \frac{dP}{dr} \right| \approx \frac{P}{R} \propto \frac{M \times M/R^3}{R^2} = \frac{M^2}{R^5}.$$

Here we have used $\rho \propto M/R^3$. Thus the pressure obeys

$$P \propto M^2/R^4. \quad (1)$$

In the nonrelativistic case, the pressure of a degenerate electron gas is given by (11.16):

$$P \approx (h^2/m_e)(\mu_e m_H)^{-5/3} \rho^{5/3}$$

and hence

$$P \propto \frac{\rho^{5/3}}{m_e \mu_e^{5/3}}. \quad (2)$$

By combining (1) and (2) we obtain

$$\frac{M^2}{R^4} \propto \frac{M^{5/3}}{R^5 m_e \mu_e^{5/3}}$$

or

$$R \propto \frac{1}{M^{1/3} m_e \mu_e^{5/3}} \propto M^{-1/3}.$$

Thus the smaller the radius of a white dwarf is, the larger its mass will be. If the density becomes so large that the relativistic equation of state (11.17) has to be used, the expression for the pressure is

$$P \propto \rho^{4/3} \propto \frac{M^{4/3}}{R^4}.$$

As the star contracts, the pressure grows at the same rate as demanded by the condition for hydrostatic support (1). Once contraction has begun, it can only stop when the state of matter changes: the electrons and protons combine into neutrons. Only a star that is massive enough can give rise to a relativistic degenerate pressure.

The neutrons are fermions, just like the electrons. They obey the Pauli exclusion principle,

and the degenerate neutron gas pressure is obtained from an expression analogous to (2):

$$P_n \propto \frac{\rho^{5/3}}{m_n \mu_n^{5/3}},$$

where m_n is the neutron mass and μ_n , the molecular weight per free neutron. Correspondingly, the radius of a neutron star is given by

$$R_{ns} \propto \frac{1}{M^{1/3} m_n \mu_n^{5/3}}.$$

If a white dwarf consists purely of helium, $\mu_e = 2$; for a neutron star, $\mu_n \approx 1$. If a white dwarf and a neutron star have the same mass, the ratio of their radii is

$$\begin{aligned} \frac{R_{wd}}{R_{ns}} &= \left(\frac{M_{ns}}{M_{wd}} \right)^{1/3} \left(\frac{\mu_n}{\mu_e} \right)^{5/3} \frac{m_n}{m_e} \\ &\approx 1 \times \left(\frac{1}{2} \right)^{5/3} \times 1840 \approx 600. \end{aligned}$$

Thus the radius of a neutron star is about 1/600 of that of a white dwarf. Typically R_{ns} is about 10 km.

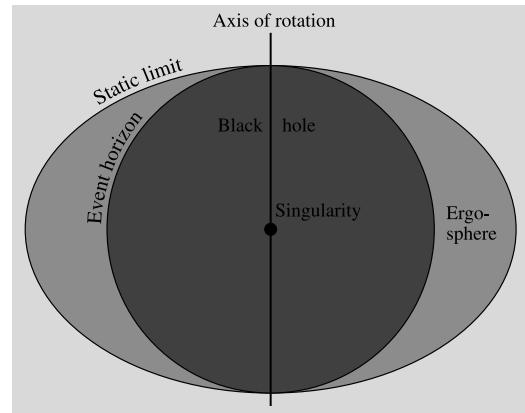


Fig. 15.8 A black hole is surrounded by a spherical event horizon. In addition to this a rotating black hole is surrounded by a flattened surface inside which no matter can remain stationary. This region is called the ergosphere

The same value for the critical radius, the *Schwarzschild radius*, is obtained from the general theory of relativity. For example, for the Sun, R_S is about 3 km; however, the Sun's mass is so small that it cannot become a black hole by normal stellar evolution. Because the mass of a black hole formed by stellar collapse has to be larger than M_{OV} the radius of the smallest black holes formed in this way is about 5–10 km.

The properties of black holes have to be studied on the basis of the general theory of relativity, which is beyond the scope of this book. Thus only some basic properties are discussed qualitatively.

An *event horizon* is a surface through which no information can be sent out, even in principle. A black hole is surrounded by an event horizon at the Schwarzschild radius (Fig. 15.8). In the theory of relativity each observer carries with him his own local measure of time. If two observers are at rest with respect to each other at the same point their clocks go at the same rate. Otherwise their clock rates are different, and the apparent course of events differs, too.

Near the event horizon the different time definitions become significant. An observer falling into a black hole reaches the centre in a finite time, according to his own clock, and does not notice anything special as he passes through the event horizon. However, to a distant observer he never seems to reach the event horizon; his ve-

15.3 Black Holes

If the mass of a star exceeds M_{OV} (Sect. 12.5), and if it does not lose mass during its evolution it can no longer reach any stable final state. The force of gravity will dominate over all other forces, and the star will collapse to a black hole. A black hole is black because not even light can escape from it. Already at the end of the 18th century Laplace showed that a sufficiently massive body would prevent the escape of light from its surface. According to classical mechanics, the escape velocity from a body of radius R and mass M is

$$v_e = \sqrt{\frac{2GM}{R}}.$$

This is greater than the speed of light, if the radius is smaller than the critical radius

$$R_S = 2GM/c^2. \quad (15.1)$$

lacity of fall seems to decrease towards zero as he approaches the horizon.

The slowing down of time also appears as a decrease in the frequency of light signals. The formula for the gravitational redshift can be written in terms of the Schwarzschild radius as (Appendix B)

$$\nu_{\infty} = \nu \sqrt{1 - \frac{2GM}{rc^2}} = \nu \sqrt{1 - \frac{R_s}{r}}. \quad (15.2)$$

Here, ν is the frequency of radiation emitted at a distance r from the black hole and ν_{∞} the frequency observed by an infinitely distant observer. It can be seen that the frequency at infinity approaches zero for radiation emitted near the event horizon.

Since the gravitational force is directed towards the centre of the hole and depends on the distance, different parts of a falling body feel a gravitational pull that is different in magnitude and direction. The tidal forces become extremely large near a black hole so that any material falling into the hole will be torn apart. All atoms and elementary particles are destroyed near the central point, and the final state of matter is unknown to present-day physics. The observable properties of a black hole do not depend on how it was made.

Not only all information on the material composition disappears as a star collapses into a black hole; any magnetic field, for example, also disappears behind the event horizon. A black hole can only have three observable properties: mass, angular momentum and electric charge.

It is improbable that a black hole could have a significant net charge. An electrically charged black hole would attract particles with opposite charge until it became neutral. Rotation, on the other hand, is typical to stars, and thus black holes, too, must rotate. Since the angular momentum is conserved, stars collapsed to black holes must rotate very fast.

In 1963 *Roy Kerr* managed to find a solution of the field equations for a rotating black hole. In addition to the event horizon a rotating hole has another limiting surface, an ellipsoidal *static limit* (Fig. 15.8). Objects inside the static limit cannot be kept stationary by any force, but they must orbit the hole. However, it is possible to escape

from the region between the static limit and the event horizon, called the *ergosphere*. In fact it is possible to utilise the rotational energy of a black hole by dropping an object to the ergosphere in such a way that part of the object falls into the hole and another part is slung out. The outgoing part may then have considerably more kinetic energy than the original object.

At present the only known way in which a black hole could be directly observed is by means of the radiation from gas falling into it. For example, if a black hole is part of a binary system, gas streaming from the companion will settle into a disk around the hole. Matter at the inner edge of the disk will fall into the hole. The accreting gas will lose a considerable part of its energy (up to 40 % of the rest mass) as radiation, which should be observable in the X-ray region.

Some rapidly and irregularly varying X-ray sources of the right kind have been discovered. The first strong evidence for black hole in an X-ray binary was for Cygnus X-1 (Fig. 15.9). Its luminosity varies on the time scale of 0.001 s, which means that the emitting region must be only 0.001 light-seconds or a few hundred kilometres in size. Only neutron stars and black holes are small and dense enough to give rise to such high-energy processes. Cygnus X-1 is the smaller component of the double system HDE 226868. The larger component is an optically visible supergiant with a mass $20\text{--}25 M_{\odot}$. The mass of the unseen component has been calculated to be $10\text{--}15 M_{\odot}$. If this is correct, the mass of the secondary component is much larger than the upper limit for a neutron star, and thus it has to be a black hole.

Today 20 such systems are known, where the compact component has a mass larger than $3 M_{\odot}$, and therefore is probably a black hole. As shown in Fig. 15.10 these can be of very different sizes. Nearly all of them have been discovered as X-ray novae.

Many frightening stories about black holes have been invented. It should therefore be stressed that they obey the same dynamical laws as other stars—they are not lurking in the darkness of space to attack innocent passers-by. If the Sun became a black hole, the planets would continue in their orbits as if nothing had happened.



Fig. 15.9 The arrow shows the variable star V1357 Cyg. Its companion is the X-ray source Cygnus X-1, suspected to be a black hole. Cyg X-1 itself is so faint that it can be observed by its X-ray radiation only. The bright star to the lower right of V1357 is η Cygni, one of the brightest stars in the constellation Cygnus

So far we have discussed only black holes with masses in the range of stellar masses. There is however no upper limit to the mass of a black hole. Many active phenomena in the nuclei of galaxies can be explained with supermassive black holes with masses of millions or thousands of millions solar masses (see Sect. 19.7).

15.4 X-ray Binaries

Close binaries where a neutron star or a black hole is accreting matter from its companion, usually a main sequence star, will be visible as strong X-ray sources. They are generally classified as *high-mass X-ray binaries* (HMXB), when the companion has a mass larger than about $10 M_{\odot}$, and *low-mass X-ray binaries* (LMXB) with a companion mass smaller than $1.2 M_{\odot}$. In HMXBs the source of the accreted material

is a strong stellar wind. LMXBs are produced by Roche-lobe overflow of the companion star, either because the major axis of the binary decreases due to angular momentum loss from the system, or else because the radius of the companion is increasing as it evolves.

Because of the rapid evolution of the massive component in HMXBs these systems are young and short-lived, 10^5 – 10^7 a. In LMXBs the lifetime is determined by the mass-transfer process, and may be longer, 10^7 – 10^{10} a. In many respects they are similar to cataclysmic variables, and may give rise to analogous phenomena.

Many kinds of variable X-ray sources have been discovered since they were first observed in the 1970's. Among these, the X-ray pulsars and the X-ray bursters can only be neutron stars. In other types of X-ray binaries it can be difficult to determine whether the primary is a neutron star or a black hole.

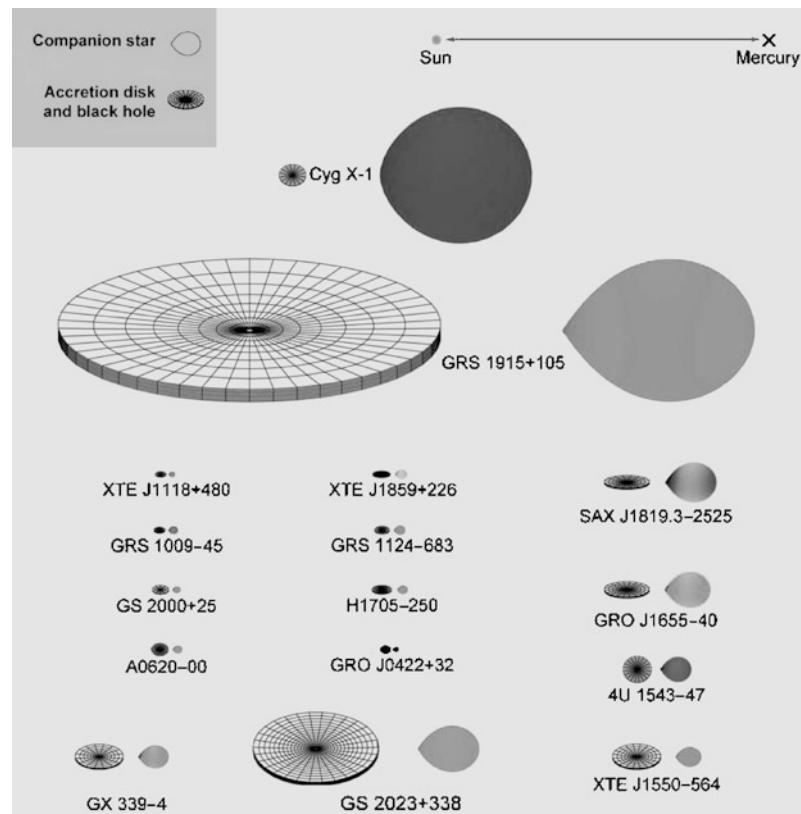
Neutron stars and black holes are formed in supernova explosions, and in a binary system the explosion would normally be expected to disrupt the binary. An X-ray binary will only form under special conditions. Some examples are shown in Sect. 12.6.

X-ray Pulsars X-ray pulsars always belong to binary systems and may be either HMXBs or LMXBs. The pulse periods of X-ray pulsars in high-mass systems are significantly longer than those of radio pulsars, from a few seconds to tens of minutes. In contrast to radio pulsars, the period of the pulsed emission of these pulsars decreases with time.

The characteristic properties of X-ray pulsars can be understood from their binary nature. A neutron star formed in a binary system is first seen as a normal radio pulsar. Initially, the strong radiation of the pulsar prevents gas from falling onto it. However, as it slows down, its energy decreases, and eventually the stellar wind from the companion can reach its surface. The incoming gas is channelled to the magnetic polar caps of the neutron star, where it emits strong X-ray radiation as it hits the surface. This produces the observed pulsed emission.

In low-mass systems, the angular momentum of the incoming gas speeds up the rotation of

Fig. 15.10 Scale drawings of 16 black-hole binaries in the Milky Way (courtesy of J. Orosz). The Sun–Mercury distance (0.4 AU) is shown at the top. The estimated binary inclination is indicated by the tilt of the accretion disk. (R.A. Remillard, J.E. McClintock 2006, ARAA 44, 54)



the pulsar. The maximum possible rotation rate of a neutron star before centrifugal forces start to break it up corresponds to a period of about a millisecond. A few *millisecond pulsars* with periods of this order are known, both in the radio and in the X-ray region. It is thought that these are (or, in the radio case, have once been) members of binary systems.

The emission curve of a typical fast X-ray pulsar, Hercules X1, is shown in Fig. 15.11. The period of the pulses is 1.24 s. This neutron star is part of an eclipsing binary system, known from optical observations as HZ Herculis. The orbital properties of the system can therefore be determined. Thus e.g. the mass of the pulsar is about one solar mass, reasonable for a neutron star.

X-ray Bursters X-ray bursters are irregular variables, showing sudden brightenings, known as type I X-ray bursts, at random times (Fig. 15.12). The typical interval between outbursts is a few hours or days, but more rapid

bursters are also known. The strength of the outburst seems to be related to the recharging time.

Type I X-ray bursts are analogous to the eruptions of classical novae. However, the source of radiation in X-ray bursters cannot be the ignition of hydrogen, since the maximum emission is in the X-ray region. Instead, gas from the companion settles on the surface of the neutron star, where hydrogen burns steadily to helium. Then, when the growing shell of helium reaches a critical temperature, it burns to carbon in a rapid helium flash. Since, in this case, there are no thick damping outer layers, the flash appears as a burst of X-ray radiation.

X-ray Novae The X-ray pulsars and bursters have to be neutron stars. Other X-ray binaries may be either neutron stars or black holes. All compact X-ray sources are variable to some extent. In the *persistent sources* the variations are moderate, and the sources always visible. The majority of sources are *transient*.

Fig. 15.11 The pulses of the X-ray pulsar Hercules X1 have the period 1.24 s. The best-fitting curve has been superimposed on the observations. (Tananbaum, H. et al. (1972): *Astrophys. J. (Lett.)* **174**, L143)

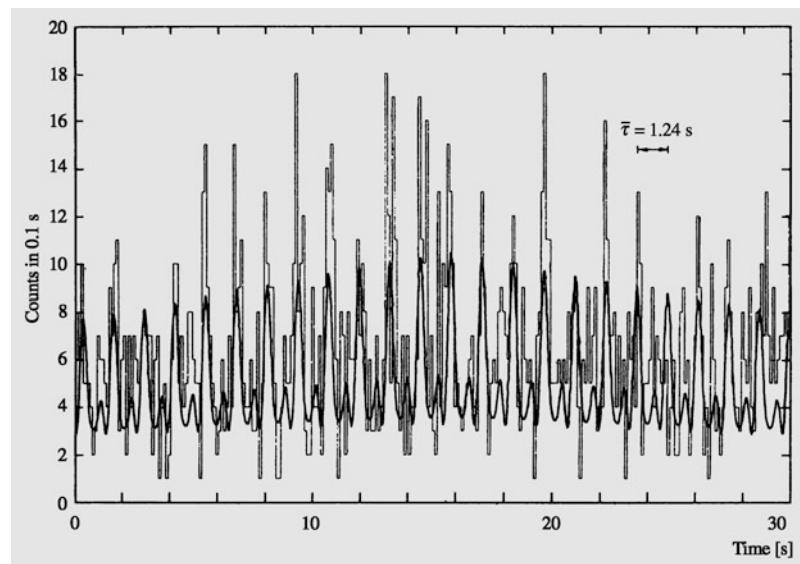
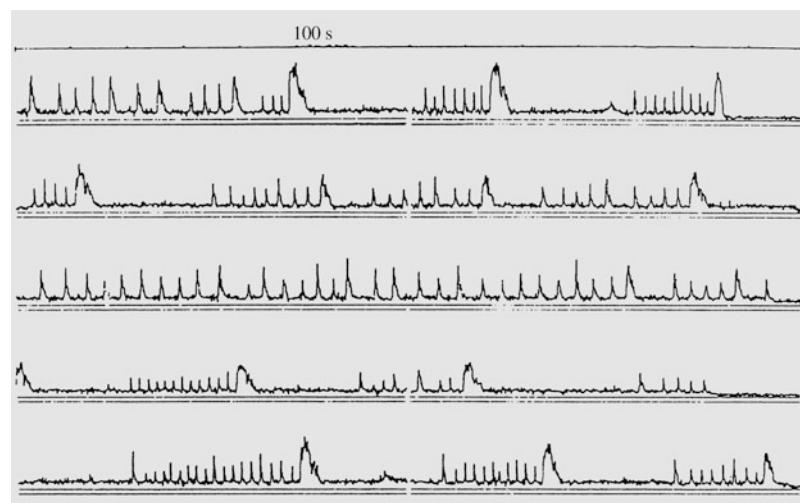


Fig. 15.12 The variations of the rapid X-ray burster MXB 1730–335. An 100 second interval is marked in the diagram. (Lewin, W.H.G. (1977): *Ann. N.Y. Acad. Sci.* **302**, 310)



If the X-ray bursters correspond to classical novae, the counterparts of dwarf novae are the *X-ray novae*, also known as *soft X-ray transients* (SXT). Quantitatively there are large differences between these types of systems. Dwarf novae have outbursts lasting for a few days at intervals of a few months, for SXTs the outbursts happen at decade-long intervals and last for months. A dwarf nova brightens by a factor about 100 during outbursts, a SXT by a factor of 10^6 . The light-curves of neutron-star and black-hole SXTs are compared in Fig. 15.13.

The SXTs are alternating between (at least) two states: During the high state thermal radiation from the accretion disk dominates, whereas in the low state the X-ray have a higher energy, and are produced by Compton scattering by hot electrons in a disk corona or a jet.

Microquasars One interesting aspect of X-ray binaries is their connection to models of active galactic nuclei (AGN, Sect. 19.7). In both systems a black hole, which in the case of AGN may

have a mass in the range $10^6\text{--}10^{10} M_{\odot}$, is surrounded by an accretion disk.

In an X-ray binary there is similarly an accretion disk surrounding a compact object, a stellar-mass black hole. It will exhibit phenomena in many respects similar to those in AGN. Since the galactic sources are much nearer, and vary on much shorter time-scales, they may allow more detailed observations of these phenomena.

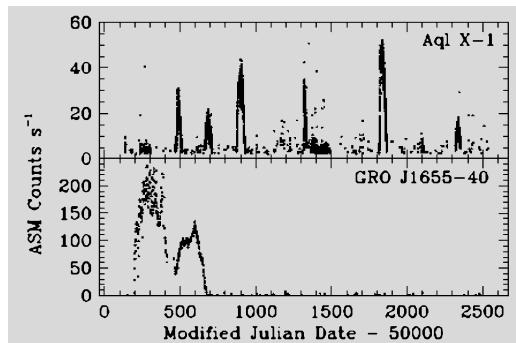


Fig. 15.13 Light-curves of a neutron-star (Aql X-1) and a black-hole (GRO J1655–40) transient source, as observed by the All Sky Monitor on RXTE. (D. Psaltis 2006, in Compact Stellar X-ray Sources, ed. Lewin, vdKlis, CUP, p. 16, Fig. 1.9)

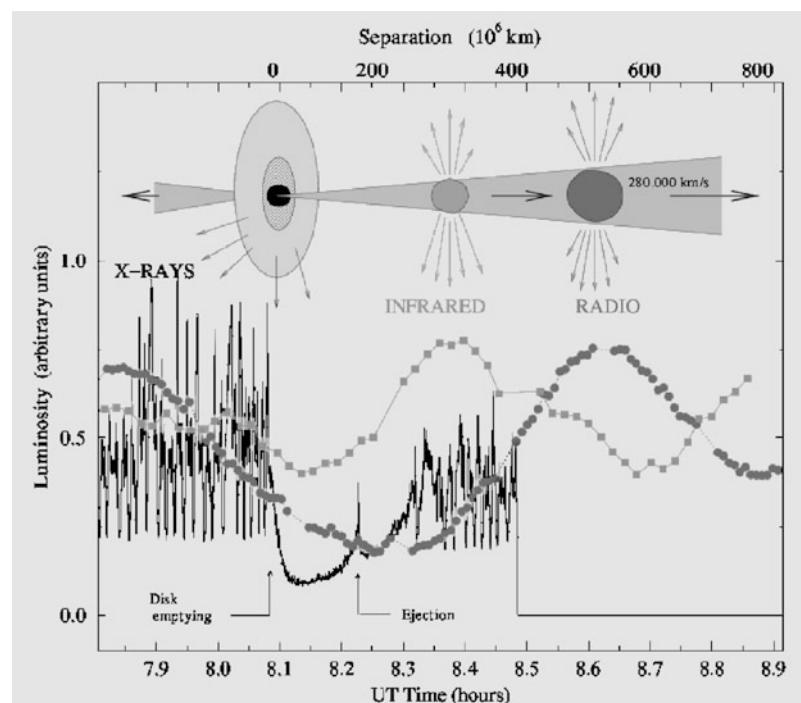
Fig. 15.14 Observed outburst of the microquasar GRS 1915+105 on 9 September, 1997. The disappearance of the internal part of the accretion disc (decrease in the X-ray flux) is followed by an ejection of relativistic plasma clouds (oscillation in the infrared and radio). (S. Chaty, astro-ph/0607668)

For example, relativistic jets perpendicular to the disk are common in AGN, and they can also be expected in X-ray binaries. A few examples of such *microquasars* have been discovered, see Fig. 15.14.

Furthermore, in AGN the jet may sometimes be pointing straight at us. Relativistic effects will then lead to a brightening of the source. In a microquasar there might be a similar effect, which would provide one explanation for the *ultraluminous X-ray sources* (ULX), sources which appear to be too luminous to be produced by ordinary stellar-mass black holes. This is important, because according to an alternative model ULXs contain an *intermediate mass black hole* with a mass about $10^3 M_{\odot}$. The origin of such intermediate mass black holes, if they exist, is an intriguing problem.

15.5 Examples

Example 15.1 Assume that the Sun collapses into a neutron star with a radius of 20 km. (a) What will be the mean density of the neutron star? (b) What would be its rotation period?



(a) The mean density is

$$\rho = \frac{M_{\odot}}{\frac{4}{3}\pi R^3} = \frac{2 \times 10^{30} \text{ kg}}{\frac{4}{3}\pi(20 \times 10^3)^3 \text{ m}^3} \approx 6 \times 10^{16} \text{ kg/m}^3.$$

One cubic millimetre of this substance would weigh 60 million kilos.

(b) To obtain an exact value, we should take into account the mass distributions of the Sun and the resulting neutron star. Very rough estimates can be found assuming that both are homogeneous. Then the moment of inertia is $I = \frac{2}{5}MR^2$, and the angular momentum is $L = I\omega$. The rotation period is then obtained as in Example 12.1:

$$\begin{aligned} P &= P_{\odot} \left(\frac{R}{R_{\odot}} \right)^2 \\ &= 25 \text{ d} \left(\frac{20 \times 10^3 \text{ m}}{6.96 \times 10^8 \text{ m}} \right)^2 = 2.064 \times 10^{-8} \text{ d} \\ &\approx 0.0018 \text{ s}. \end{aligned}$$

The Sun would make over 550 revolutions per second.

Example 15.2 What should be the radius of the Sun if the escape velocity from the surface were to exceed the speed of light?

The escape velocity exceeds the speed of light if

$$\sqrt{\frac{2GM}{R}} > c$$

or

$$R < \frac{2GM}{c^2} = R_S.$$

For the Sun we have

$$\begin{aligned} R_S &= \frac{2 \times 6.67 \times 10^{-11} \text{ m}^3 \text{ s}^{-2} \text{ kg}^{-1} \times 1.989 \times 10^{30} \text{ kg}}{(2.998 \times 10^8 \text{ m s}^{-1})^2} \\ &= 2950 \text{ m}. \end{aligned}$$

15.6 Exercises

Exercise 15.1 The mass of a pulsar is $1.5 M_{\odot}$, radius 10 km, and rotation period 0.033 s. What is the angular momentum of the pulsar? Variations of 0.0003 s are observed in the period. If they are due to radial oscillations (“starquakes”), how large are these oscillations?

Exercise 15.2 In *Dragon’s Egg* by Robert L. Forward a spaceship orbits a neutron star at a distance of 406 km from the centre of the star. The orbital period is the same as the rotation period of the star, 0.1993 s.

- (a) Find the mass of the star and the gravitational acceleration felt by the spaceship.
- (b) What is the effect of the gravitation on a 175 cm tall astronaut, if (s)he stands with her/his feet pointing towards the star? And if (s)he is lying tangential to the orbit?

Exercise 15.3 A photon leaves the surface of a star at a frequency ν_e . An infinitely distant observer finds that its frequency is ν . If the difference is due to gravitation only, the change in the energy of the photon, $h\Delta\nu$, equals the change in its potential energy. Find the relation between ν and ν_e , assuming the mass and radius of the star are M and R . How much will the solar radiation redshift on its way to the Earth?

Although most of the mass of the Milky Way Galaxy is condensed into stars, interstellar space is not completely empty. It contains *gas* and *dust* in the form both of individual clouds and of a diffuse medium. Interstellar space typically contains about one gas atom per cubic centimetre and 100 dust particles per cubic kilometre.

Altogether, about 10 % of the mass of the Milky Way consists of interstellar gas. Since the gas is strongly concentrated in the galactic plane and the spiral arms, in these regions there are many places where the quantities of stars and interstellar matter are about equal. The dust (a better name would be “smoke”, since the particle sizes are much smaller than in terrestrial dust) constitutes about one percent of the gas. High-energy cosmic ray particles are mixed with the gas and dust. There is also a weak, but still very important, galactic magnetic field.

At present the most important observations of the interstellar medium are made at radio and infrared wavelengths, since the peak of the emission often lies at these wavelengths. But many forms of interstellar matter (such as solid bodies with diameters larger than 1 mm) would be almost impossible to detect on the basis of their emission or absorption. In principle, the mass of these forms of matter might be larger than the observed mass of all other forms put together. However, an upper limit on the total mass of interstellar matter, regardless of its form, can be derived on the basis of its gravitational effects. This is the *Oort limit*. The galactic gravitational field is determined by the distribution of matter. By ob-

serving the motions of stars perpendicular to the galactic plane, the vertical gravitational force and hence the amount of mass in the galactic plane can be determined. The result is that the local density within 1 kpc of the Sun is $(7.3\text{--}10.0) \times 10^{-21} \text{ kg m}^{-3}$. The density of known stars is $(5.9\text{--}6.7) \times 10^{-21} \text{ kg m}^{-3}$ and that of known interstellar matter about $1.7 \times 10^{-21} \text{ kg m}^{-3}$. Thus there is very little room for unknown forms of mass in the solar neighbourhood. However, the limit concerns only the dark matter concentrated in the galactic plane. There are indications that the Milky Way is surrounded by a spherical halo of dark matter (Chap. 18).

16.1 Interstellar Dust

The first clear evidence for the existence of interstellar dust was obtained around 1930. Before that, it had been generally thought that space is completely transparent and that light can propagate indefinitely without extinction.

In 1930 *Robert Trumpler* published his study of the space distribution of the open clusters. The absolute magnitudes M of the brightest stars could be estimated on the basis of the spectral type. Thus the distance r to the clusters could be calculated from the observed apparent magnitudes m of the bright stars:

$$m - M = 5 \lg \frac{r}{10 \text{ pc}}. \quad (16.1)$$

Trumpler also studied the diameters of the clusters. The linear diameter D is obtained from the

apparent angular diameter d by means of the formula

$$D = dr, \quad (16.2)$$

where r is the distance of the cluster.

It caught Trumpler's attention that the more distant clusters appeared to be systematically larger than the nearer ones (Fig. 16.1). Since this could hardly be true, the distances of the more distant clusters must have been overestimated. Trumpler concluded that space is not completely transparent, but that the light of a star is dimmed by some intervening material. To take this into account, (16.1) has to be replaced with (4.17)

$$m - M = 5 \lg \frac{r}{10 \text{ pc}} + A, \quad (16.3)$$

where $A \geq 0$ is the extinction in magnitudes due to the intervening medium. If the opacity of the medium is assumed to be the same at all distances and in all directions, A can be written

$$A = ar, \quad (16.4)$$

where a is a constant. Trumpler obtained for the average value of a in the galactic plane, $a_{\text{pg}} = 0.79 \text{ mag/kpc}$, in photographic magnitudes. At present, a value of 2 mag/kpc is used for the average extinction. Thus the extinction over a 5 kpc path is already 10 magnitudes.

Extinction due to dust varies strongly with direction. For example, visible light from the

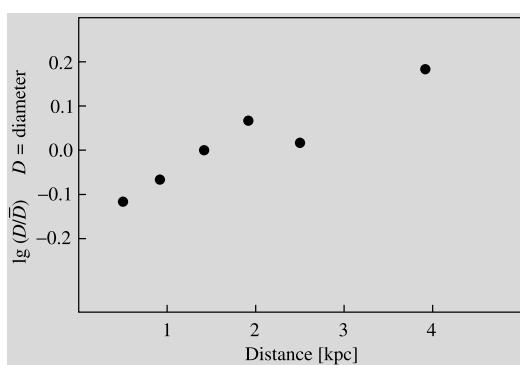


Fig. 16.1 The diameters of open star clusters calculated with the distance given by the formula (15.1) according to Trumpler (1930). The increase of the diameter with distance is not a real phenomenon, but an effect of interstellar extinction, which was discovered in this way

galactic centre (distance 8–9 kpc) is dimmed by 30 magnitudes. Therefore the galactic centre cannot be observed at optical wavelengths.

Extinction is due to dust grains that have diameters near the wavelength of the light. Such particles scatter light extremely efficiently. Gas can also cause extinction by scattering, but its scattering efficiency per unit mass is much smaller. The total amount of gas allowed by the Oort limit is so small that scattering by gas is negligible in interstellar space. (This is in contrast with the Earth's atmosphere, where air molecules make a significant contribution to the total extinction).

Interstellar particles can cause extinction in two ways:

1. In *absorption* the radiant energy is transformed into heat, which is then re-radiated at infrared wavelengths corresponding to the temperature of the dust particles.
2. In *scattering* the direction of light propagation is changed, leading to a reduced intensity in the original direction of propagation.

An expression for interstellar extinction will now be derived. The size, index of refraction and number density of the particles are assumed to be known. For simplicity we shall assume that all particles are spheres with the same radius a and the geometrical cross section πa^2 . The true extinction cross section of the particles C_{ext} will be

$$C_{\text{ext}} = Q_{\text{ext}} \pi a^2, \quad (16.5)$$

where Q_{ext} is the extinction efficiency factor.

Let us consider a volume element with length dl and cross section dA , normal to the direction of propagation (Fig. 16.2). It is assumed that the particles inside the element do not shadow each other. If the particle density is n , there are $n dl dA$ particles in the volume element and they will cover the fraction $d\tau$ of the area dA , where

$$d\tau = \frac{n dA dl C_{\text{ext}}}{dA} = n C_{\text{ext}} dl.$$

In the length dl the intensity is thus changed by

$$dI = -I d\tau. \quad (16.6)$$

On the basis of (16.6) $d\tau$ can be identified as the optical depth.

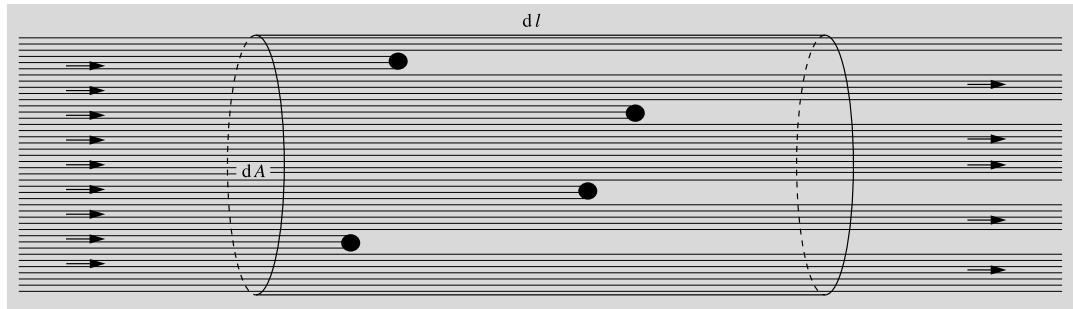


Fig. 16.2 Extinction by a distribution of particles. In the volume element with length dl and cross section dA , there are $n dA dl$ particles, where n is the particle density in the medium. If the extinction cross section of one particle is

The total optical depth between the star and the Earth is

$$\tau(r) = \int_0^r d\tau = \int_0^r n C_{\text{ext}} dl = C_{\text{ext}} \bar{n} r,$$

where \bar{n} is the average particle density along the given path. According to (4.18) the extinction in magnitudes is

$$A = (2.5 \lg e) \tau,$$

and hence

$$A(r) = (2.5 \lg e) C_{\text{ext}} \bar{n} r. \quad (16.7)$$

This formula can also be inverted to calculate \bar{n} , if the other quantities are known.

The extinction efficiency factor Q_{ext} can be calculated exactly for spherical particles with given radius a and refractive index m . In general,

$$Q_{\text{ext}} = Q_{\text{abs}} + Q_{\text{sca}},$$

where

Q_{abs} = absorption efficiency factor,

Q_{sca} = scattering efficiency factor.

If we define

$$x = 2\pi a/\lambda, \quad (16.8)$$

where λ is the wavelength of the radiation, then

$$Q_{\text{ext}} = Q_{\text{ext}}(x, m). \quad (16.9)$$

C_{ext} , the total area covered by the particles is $n dA dl C_{\text{ext}}$. Thus the fractional decrease in intensity over the distance dl is $dI/I = -n dA dl C_{\text{ext}}/dA = -n C_{\text{ext}} dl$

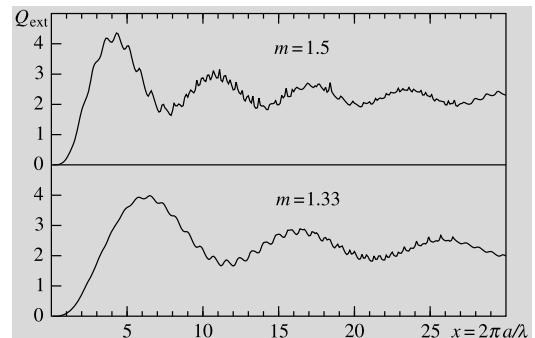


Fig. 16.3 Mie scattering: the extinction efficiency factor for spherical particles for the refractive indices $m = 1.5$ and $m = 1.33$ (refractive index of water). The horizontal axis is related to the size of the particle according to $x = 2\pi a/\lambda$, where a is the particle radius and λ , the wavelength of the radiation

The exact expression for Q_{ext} is a series expansion in x that converges more slowly for larger values of x . When $x \ll 1$, the process is called *Rayleigh scattering*; otherwise it is known as *Mie scattering*. Figure 16.3 shows Q_{ext} as a function of x for $m = 1.5$ and $m = 1.33$. For very large particles, ($x \gg 1$) $Q_{\text{ext}} = 2$, as appears from Fig. 16.3. Purely geometrically one would have expected $Q_{\text{ext}} = 1$; the two times larger scattering efficiency is due to the diffraction of light at the edges of the particle.

Other observable phenomena, apart from extinction, are also caused by interstellar dust. One of these is the *reddening* of the light of stars. (This should not be confused with the redshift of spectral lines.) Reddening is due to the fact

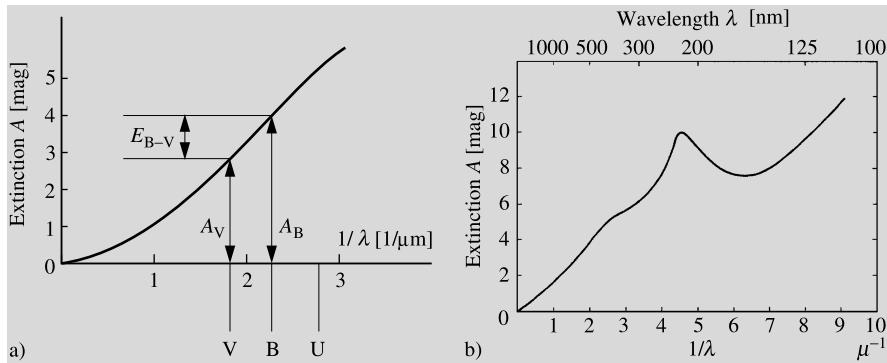


Fig. 16.4 (a) Schematic representation of the interstellar extinction. As the wavelength increases, the extinction approaches zero. (Drawing based on Greenberg, J.M. (1968): “Interstellar Grains”, in *Nebulae and Interstellar Matter*, ed. by Middlehurst, B.M., Aller, L.H., Stars

and Stellar Systems, Vol. VII (The University of Chicago Press, Chicago) p. 224). (b) Measured extinction curve, normalised to make $E_{B-V} = 1$. (Hoyle, F., Narlikar, J. (1980): *The Physics-Astronomy Frontier* (W.H. Freeman and Company, San Francisco) p. 156. Used by permission)

that the amount of extinction becomes larger for shorter wavelengths. Going from red to ultraviolet, the extinction is roughly inversely proportional to wavelength. For this reason the light of distant stars is redder than would be expected on the basis of their spectral class. The spectral class is defined on the basis of the relative strengths of the spectral lines which are not affected by extinction.

According to (4.20), the observed colour index $B - V$ of a star is

$$\begin{aligned} B - V &= M_B - M_V + A_B - A_V \\ &= (B - V)_0 + E_{B-V}, \end{aligned} \quad (16.10)$$

where $(B - V)_0$ is the *intrinsic colour* of the star and E_{B-V} the *colour excess*. As noted in Sect. 4.5 the ratio between the visual extinction A_V and the colour excess is approximately constant:

$$R = \frac{A_V}{E_{B-V}} = \frac{A_V}{A_B - A_V} \approx 3.0. \quad (16.11)$$

R does not depend on the properties of the star or the amount of extinction. This is particularly important in photometric distance determinations because of the fact that the colour excess E_{B-V} can be directly determined from the difference between the observed colour index $B - V$ and the intrinsic colour $(B - V)_0$ known from the spec-

tral class. One can then calculate the extinction

$$A_V \approx 3.0 E_{B-V} \quad (16.12)$$

and finally the distance. Since the interstellar medium is far from homogeneous, the colour excess method gives a much more reliable value than using some average value for the extinction in (4.18).

The wavelength dependence of the extinction, $A(\lambda)$, can be studied by comparing the magnitudes of stars of the same spectral class in different colours. These measurements have shown that $A(\lambda)$ approaches zero as λ becomes very large. In practice $A(\lambda)$ can be measured up to a wavelength of about two micrometres. The extrapolation to zero inverse wavelength is then fairly reliable. Figure 16.4(a) shows $A(\lambda)$ as a function of inverse wavelength. It also illustrates how the quantities A_V and E_{B-V} , which are needed in order to calculate the value of R , are obtained from this *extinction or reddening curve*. Figure 16.4(b) shows the observed extinction curve. The points in the ultraviolet ($\lambda \leq 0.3$ m) are based on rocket measurements.

It is clear from Fig. 16.4(b) that interstellar extinction is largest at short wavelengths in the ultraviolet and decreases for longer wavelengths. In the infrared it is only about ten percent of the optical extinction and in the radio region it is vanishingly small. Objects that are invisible in the opti-

cal region can therefore be studied at infrared and radio wavelengths.

Another observed phenomenon caused by dust is the *polarisation* of the light of the stars. Since spherical particles cannot produce any polarisation, the interstellar dust particles have to be non-spherical in shape. If the particles in a cloud are aligned by the interstellar magnetic field, they will polarise the radiation passing through the cloud. The degree of polarisation and its wavelength dependence give information on the properties of the dust particles. By studying the direction of polarisation in various directions, one can map the structure of the galactic magnetic field.

In the Milky Way interstellar dust is essentially confined to a very thin, about 100 pc, layer in the galactic plane. The dust in other spiral galaxies has a similar distribution and is directly visible as a dark band in the disk of the galaxy (Fig. 19.17 bottom). The Sun is located near the central plane of the galactic dust layer, and thus the extinction in the direction of the galactic plane is very large, whereas the total extinction towards the galactic poles may be less than 0.1 magnitudes. This is apparent in the distribution of galaxies in the sky: at high galactic latitudes, there are many galaxies, while near the galactic plane, there is a 20° zone where hardly any galaxies are seen. This empty region is called the *zone of avoidance*.

If a homogeneous dust layer gives rise to a total extinction of Δm magnitudes in the vertical direction, then according to Fig. 16.5, the total extinction at galactic latitude b will be

$$\Delta m(b) = \Delta m / \sin b. \quad (16.13)$$

If the galaxies are uniformly distributed in space, then in the absence of extinction, the number of galaxies per square degree brighter than the magnitude m would be

$$\lg N_0(m) = 0.6m + C, \quad (16.14)$$

where C is a constant (see Exercise 17.1). However, due to extinction, a galaxy that would otherwise have the apparent magnitude m_0 will have

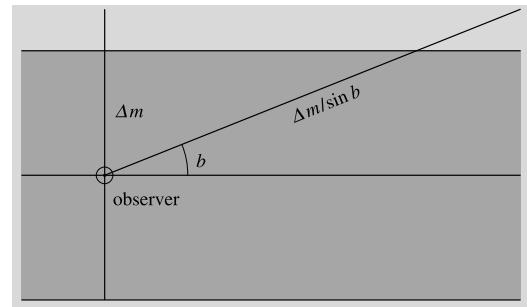


Fig. 16.5 In a homogeneous medium the extinction in magnitudes is proportional to the pathlength traversed. If the extinction in the direction of the galactic pole is Δm , then the extinction at the galactic latitude b will be $\Delta m / \sin b$

the magnitude

$$m(b) = m_0 + \Delta m(b) = m_0 + \Delta m / \sin b, \quad (16.15)$$

where b is the galactic latitude. Thus the observable number of galaxies at latitude b will be

$$\begin{aligned} \lg N(m, b) &= \lg N_0(m - \Delta m(b)) \\ &= 0.6(m - \Delta m(b)) + C \\ &= \lg N_0(m) - 0.6\Delta m(b) \end{aligned}$$

or

$$\lg N(m, b) = C' - 0.6 \frac{\Delta m}{\sin b}, \quad (16.16)$$

where $C' = \lg N_0(m)$ does not depend on the galactic latitude. By making galaxy counts at various latitudes b , the extinction Δm can be determined. The value obtained from galaxy counts made at Lick Observatory is $\Delta m_{pg} = 0.51$ mag.

The total vertical extinction of the Milky Way has also been determined from the colour excesses of stars. These investigations have yielded much smaller extinction values, about 0.1 mag. In the direction of the north pole, extinction is only 0.03 mag. The disagreement between the two extinction values is probably largely due to the fact that the dust layer is not really homogeneous. If the Sun is located in a local region of low dust content, the view towards the galactic poles might be almost unobstructed by dust.



Fig. 16.6 The Coalsack is a dark nebula next to the Southern Cross. (Photograph K. Mattila, Helsinki University)

Dark Nebulae Observations of other galaxies show that the dust is concentrated in the spiral arms, in particular at their inner edge. In addition dust is concentrated in individual clouds, which appear as star-poor regions or *dark nebulae* against the background of the Milky Way. Examples of dark nebulae are the Coalsack in the southern sky (Fig. 16.6) and the Horsehead nebula in Orion. Sometimes the dark nebulae form extended winding bands, and sometimes small, almost spherical, objects. Objects of the latter type are most easy to see against a bright background, e.g. a gas nebula (see Fig. 16.19). These objects have been named *globules* by Bart J. Bok, who

put forward the hypothesis that they are clouds that are just beginning to contract into stars.

The extinction by a dark nebula can be illustrated and studied by means of a *Wolf diagram*, shown schematically in Fig. 16.7. The diagram is constructed on the basis of star counts. The number of stars per square degree in some magnitude interval (e.g. between magnitudes 14 and 15) in the cloud is counted and compared with the number outside the nebula. In the comparison area, the number of stars increases monotonically towards fainter magnitudes. In the dark nebula the numbers first increase in the same way, but beyond some limiting magnitude (10 in the figure)

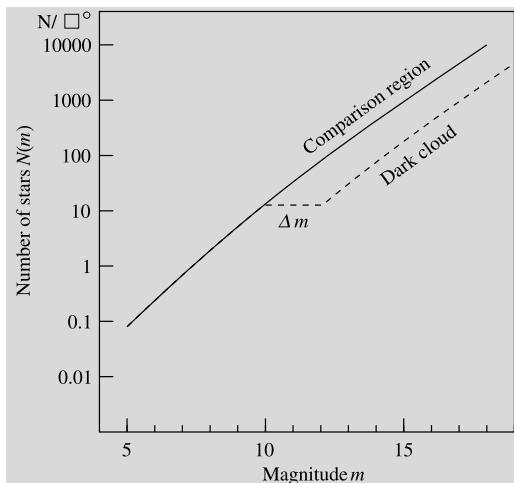


Fig. 16.7 Wolf diagram. The horizontal coordinate is the magnitude and the vertical coordinate is the number of stars per square degree in the sky brighter than that magnitude. A dark nebula diminishes the brightness of stars lying behind it by the amount Δm

the number of stars falls below that outside the cloud. The reason for this is that the fainter stars are predominantly behind the nebula, and their brightness is reduced by some constant amount Δm (2 magnitudes in the figure). The brighter stars are mostly in front of the nebula and suffer no extinction.

Reflection Nebulae If a dust cloud is near a bright star, it will scatter, i.e. reflect the light of the star. Thus individual dust clouds can sometimes be observed as bright *reflection nebulae*. Some 500 reflection nebulae are known.

The regions in the sky richest in reflection nebulae are the areas around the Pleiades and around the giant star Antares. Antares itself is surrounded by a large red reflection nebula. This region is shown in Fig. 16.8. Figure 16.9 shows the reflection nebula NGC 2068, which is located near a large, thick dust cloud a few degrees northwest of Orion's belt. It is one of the brightest reflection nebulae and the only one included in the Messier catalogue (M78). In the middle of the nebula there are two stars of about 11 magnitudes. The northern star illuminates the nebula, while the other one is probably in front of the nebula. Figure 16.10 shows the reflection nebula

NGC 1435 around Merope in the Pleiades. Another bright and much-studied reflection nebula is NGC 7023 in Cepheus. It, too, is connected with a dark nebula. The illuminating star has emission lines in its spectrum (spectral type Be). Infrared stars have also been discovered in the area of the nebula, probably a region of star formation.

In 1922 *Edwin Hubble* published a fundamental investigation of bright nebulae in the Milky Way. On the basis of extensive photographic and spectroscopic observations, he was able to establish two interesting relationships. First he found that emission nebulae only occur near stars with spectral class earlier than B0, whereas reflection nebulae may be found near stars of spectral class B1 and later. Secondly Hubble discovered a relationship between the angular size R of the nebula and the apparent magnitude m of the illuminating star:

$$5 \lg R = -m + \text{const.} \quad (16.17)$$

Thus the angular diameter of a reflection nebula is larger for a brighter illuminating star. Since the measured size of a nebula generally increases for longer exposures, i.e. fainter limiting surface brightness, the value of R should be defined to correspond to a fixed limiting surface brightness. The value of the constant in the Hubble relation depends on this limiting surface brightness. The Hubble relation for reflection nebulae is shown in Fig. 16.11, based on measurements by *Sidney van den Bergh* from Palomar Sky Atlas plates. Each point corresponds to a reflection nebula and the straight line represents the relation (16.17), where the value of the constant is 12.0 (R is given in arc minutes).

The Hubble relation can be derived theoretically, if it is assumed that the illumination of a dust cloud is inversely proportional to the square of the distance to the illuminating star, and that the dust clouds are uniformly distributed in space. The theoretical Hubble relation also gives an expression for the constant on the right-hand side, which involves the albedo and the phase function of the grains.

The observations of reflection nebulae show that the albedo of interstellar grains must be quite

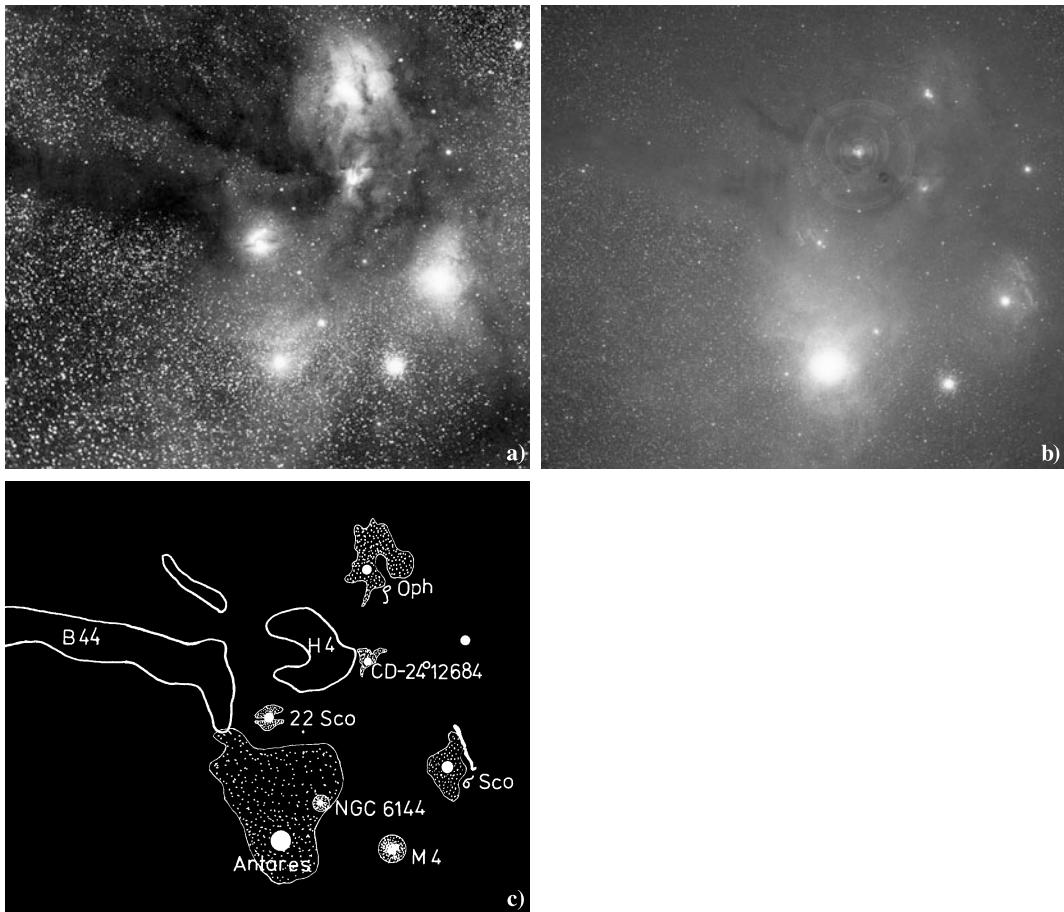


Fig. 16.8 Bright and dark nebulae in Scorpius and Ophiuchus. Photograph (a) was taken in the blue colour region, $\lambda = 350\text{--}500 \text{ nm}$, and (b) in the red colour region, $\lambda = 600\text{--}680 \text{ nm}$. (The sharp rings in (b) are reflections of Antares in the correction lens of the Schmidt camera.) The nebulae located in the area are identified in drawing (c). B44 and H4 are dark nebulae. There is a large reflection nebula around Antares, which is faintly visible in the blue (a), but bright in the red (b) regions. Antares is very red (spectral class M1) and therefore the reflec-

tion nebula is also red. In contrast, the reflection nebulae around the blue stars ρ Ophiuchi (B2), CD-24° 12684 (B3), 22 Scorpis (B2) and σ Scorpis (B1) are blue and are visible only in (a). In (b) there is an elongated nebula to the right of σ Scorpis, which is invisible in (a). This is an emission nebula, which is very bright in the red hydrogen H_α line (656 nm). In this way reflection and emission nebulae can be distinguished by means of pictures taken in different wavelength regions. (Photograph (a) E. Barnard, and (b) K. Mattila)

high. It has not yet been possible to obtain its precise numerical value in this way, since the distances between the nebulae and their illuminating stars are not known well enough.

One may also consider the surface brightness of dark nebulae that are not close enough to a star to be visible as reflection nebulae. These nebulae will still reflect the diffuse galactic light from all the stars in the Milky Way. Calculations show that if the dust grains have a large albedo, then the

reflected diffuse light should be bright enough to be observable, and it has indeed been observed. Thus the dark nebulae are not totally dark. The diffuse galactic light constitutes about 20–30 % of the total brightness of the Milky Way.

Dust Temperature In addition to scattering the interstellar grains also absorb radiation. The absorbed energy is re-radiated by the grains at infrared wavelengths corresponding to their

Fig. 16.9 The reflection nebula NGC 2068 (M78) in Orion. In the middle of the nebula there are two stars of about magnitude 11. The northern one (*at the top*) is the illuminating star, while the other one probably lies in the foreground.
(Photography Lunar and Planetary Laboratory, Catalina Observatory)

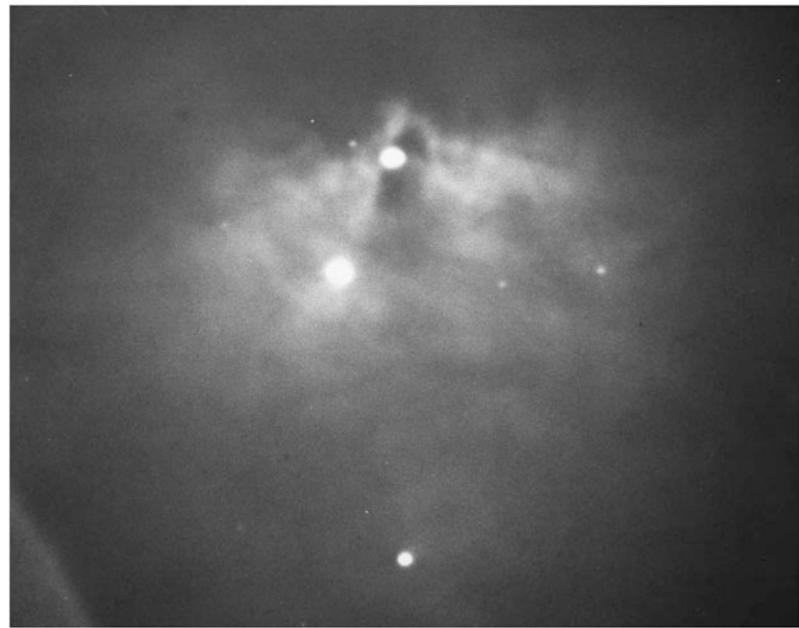


Fig. 16.10 The reflection nebula NGC 1435 around Merope (23 Tau, spectral class B6) in the Pleiades. This figure should be compared with Fig. 16.1, where Merope is visible as the lowest of the bright stars in the Pleiades.
(National Optical Astronomy Observatories, Kitt Peak National Observatory)



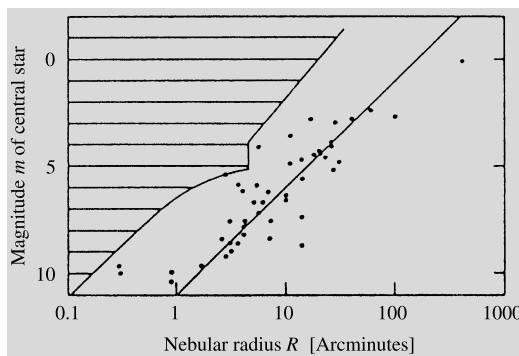
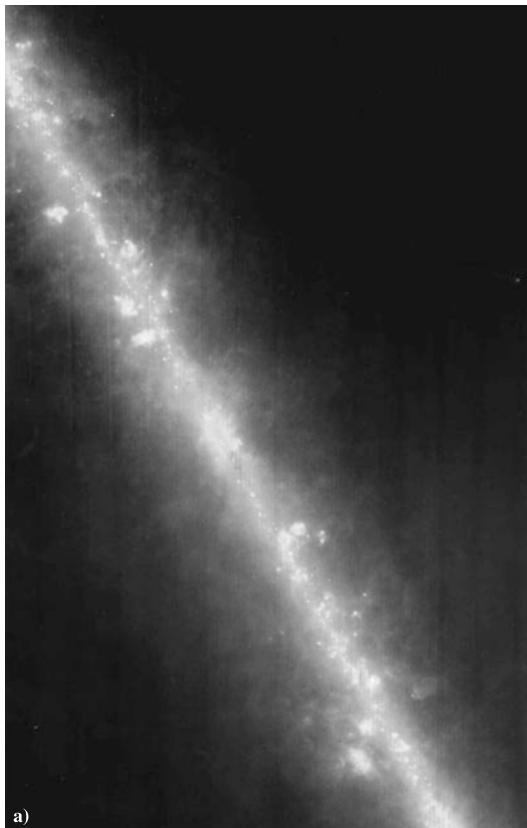


Fig. 16.11 The Hubble relation for reflection nebulae. The horizontal axis shows the radius R of the nebulae in arc minutes and the vertical axis, the (blue) apparent magnitude m of the central star. No measurements were made in the hatched region. (van den Bergh, S. (1966): Astron. J. **71**, 990)

temperatures. The temperature of dust in interstellar space (including dark nebulae) is about 10–20 K. The corresponding wavelength according to Wien's displacement law (5.21) is 300–150 μm . Near a hot star the temperature of the dust may be 100–600 K and the maximum emission is then at 30–5 μm . In H II regions the dust temperature is about 70–100 K.

The rapid development of infrared astronomy in the 1970's has brought all the above-mentioned dust sources within the reach of observations (Fig. 16.12). In addition infrared radiation from the nuclei of normal and active galaxies is largely thermal radiation from dust. Thermal dust emission is one of the most important sources of infrared radiation in astronomy.



a)

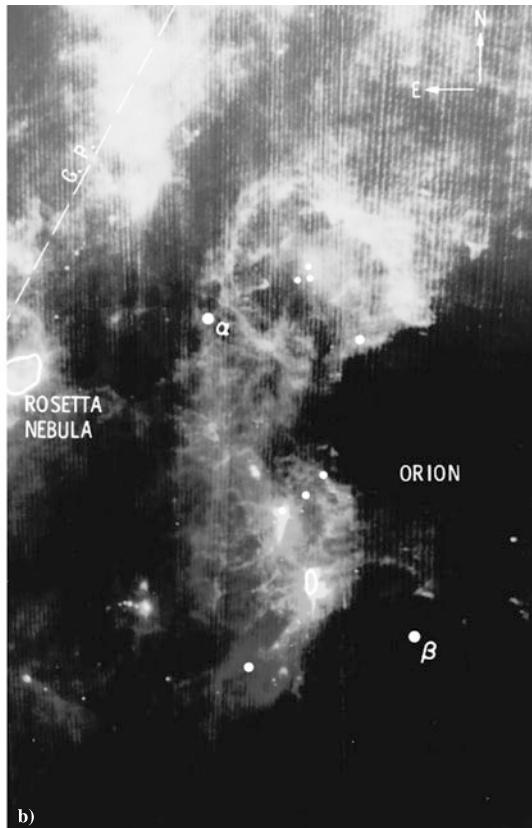


Fig. 16.12 Interstellar dust is best seen in infrared wavelengths. Two examples of the images by the IRAS satellite. (a) In a view towards the Galactic centre, the dust is seen to be concentrated in a narrow layer in the galactic plane. Several separate clouds are also seen. (b) Most of the constellation Orion is covered by a complex area of interstellar matter. The densest concentrations of dust below the centre of the image, are in the region of the Horsehead nebula and the Orion nebula. (Photos NASA)

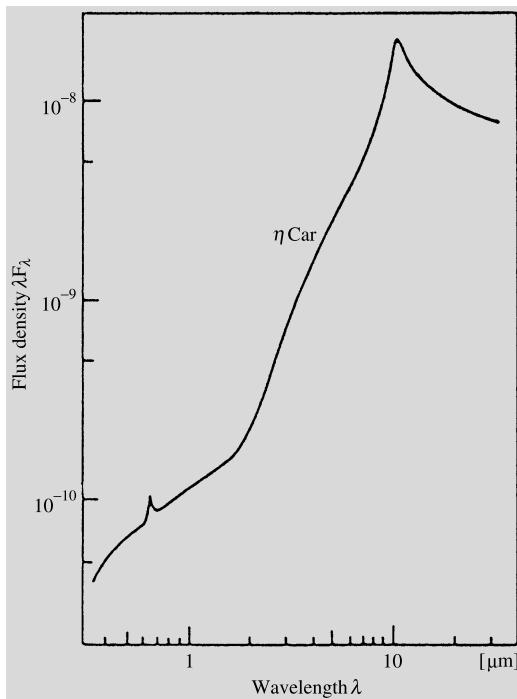


Fig. 16.13 More than 99 % of the radiation from the η Carinae nebula (Fig. 13.10) is in the infrared. The peak in the visual region is from the hydrogen H α line (0.66 μm). In the infrared, the silicate emission from dust is evident at 10 μm . (Allen, D.A. (1975): *Infrared, the New Astronomy* (Keith Reid Ltd., Shaldon) p. 103)

One of the strongest infrared sources in the sky is the nebula around the star η Carinae. The nebula consists of ionised gas, but infrared radiation from dust is also clearly visible in its spectrum (Fig. 16.13). In even more extreme cases, the central star may be completely obscured, but revealed by the infrared emission from hot dust.

Composition and Origin of the Dust (Tables 16.1 and 16.2). From the peaks in the extinction curve, it may be concluded that interstellar dust contains water ice and silicates, and probably graphite as well. The sizes of the grains can be deduced from their scattering properties; usually they are smaller than one micrometre. The strongest scattering is due to grains of about 0.3 μm but smaller particles must also be present.

Dust grains are formed in the atmospheres of stars of late spectral types (K, M). Gas condenses

into grains just as water in the Earth's atmosphere may condense into snow and ice. The grains are then expelled into interstellar space by the radiation pressure. Grains may also form in connection with star formation and possibly directly from atoms and molecules in interstellar clouds as well.

16.2 Interstellar Gas

The mass of gas in interstellar space is a hundred times larger than that of dust. Although there is more gas, it is less easily observed, since the gas does not cause a general extinction of light. In the optical region it can only be observed on the basis of a small number of spectral lines.

The existence of interstellar gas began to be suspected in the first decade of the 20th century, when in 1904 *Johannes Hartmann* observed that some absorption lines in the spectra of certain binary stars were not Doppler shifted by the motions of the stars like the other lines. It was concluded that these absorption lines were formed in gas clouds in the space between the Earth and the stars. In some stars there were several lines, apparently formed in clouds moving with different velocities. The strongest lines in the visible region are those of neutral sodium and singly ionised calcium (Fig. 16.14). In the ultraviolet region, the lines are more numerous. The strongest one is the hydrogen Lyman α line (121.6 nm).

On the basis of the optical and ultraviolet lines, it has been found that many atoms are ionised in interstellar space. This ionisation is mainly due to ultraviolet radiation from stars and, to some extent, to ionisation by cosmic rays. Since the density of interstellar matter is very low, the free electrons only rarely encounter ions, and the gas remains ionised.

About thirty elements have been discovered by absorption line observations in the visible and ultraviolet region. With a few exceptions, all elements from hydrogen to zinc (atomic number 30) and a few additional heavier elements have been detected (Table 16.3). Like in the stars, most of

Table 16.1 Main properties of interstellar gas and dust

Property	Gas	Dust
Mass fraction	10 %	0.1 %
Composition	H I, H II, H ₂ (70 %) He (28 %) C, N, O, Ne, Na, Mg, Al, Si, S, ... (2 %)	Solid particles $d \approx 0.1\text{--}1 \mu\text{m}$ H ₂ O (ice), silicates, graphite + impurities
Particle density	$1 / \text{cm}^3$	$10^{-13} / \text{cm}^3 = 100 / \text{km}^3$
Mass density	10^{-21} kg/m^3	10^{-23} kg/m^3
Temperature	100 K (H I), 10 ⁴ K (H II) 50 K (H ₂)	10–20 K
Method of study	Absorption lines in stellar spectra. Optical: Ca I, Ca II, Na I, K I, Ti II, Fe I, CN, CH, CH ⁺ Ultraviolet: H ₂ , CO, HD Radio lines: hydrogen 21 cm emission and absorption; H II, He II, C II recombination lines; molecular emission and absorption lines OH, H ₂ CO, NH ₃ , H ₂ O, CO, H ₂ C ₂ HCN, C ₂ H ₅ OH	Absorption and scattering of starlight. Interstellar reddening Interstellar polarisation Thermal infrared emission

the mass is hydrogen (about 70 %) and helium (almost 30 %). On the other hand, heavy elements are significantly less abundant than in the Sun and other population I stars. It is thought that they have been incorporated into dust grains, where they do not produce any absorption lines. The element abundances in the interstellar medium (gas + dust) would then be normal, although the interstellar gas is depleted in heavy elements. This interpretation is supported by the observation that in regions where the amount of dust is smaller than usual, the element abundances in the gas are closer to normal.

Atomic Hydrogen Ultraviolet observations have provided an excellent way of studying interstellar *neutral hydrogen*. The strongest interstellar absorption line, as has already been mentioned, is the hydrogen Lyman α line (Fig. 16.15). This line corresponds to the transition of the electron in the hydrogen atom from a state with principal quantum number $n = 1$ to one with $n = 2$.

The conditions in interstellar space are such that almost all hydrogen atoms are in the ground state with $n = 1$. Therefore the Lyman α line is a strong absorption line, whereas the Balmer absorption lines, which arise from the excited initial state $n = 2$, are unobservable. (The Balmer lines are strong in stellar atmospheres with temperatures of about 10,000 K, where a large number of atoms are in the first excited state.)

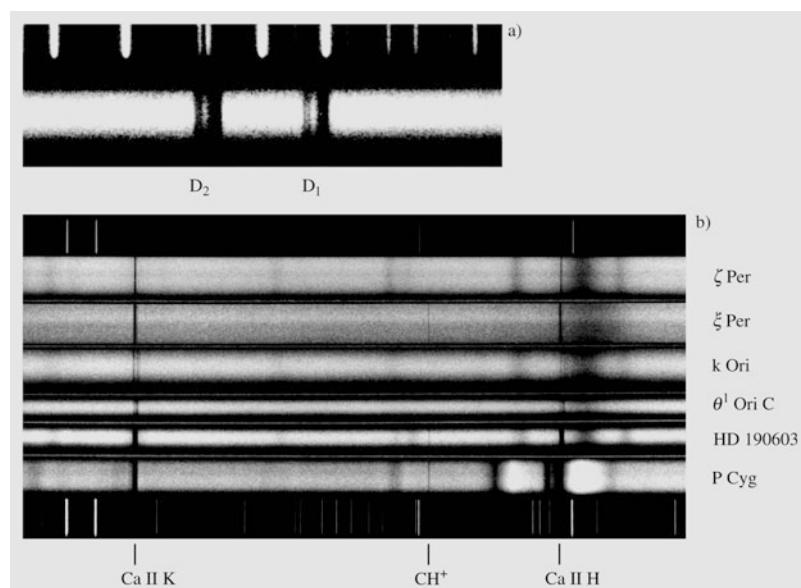
The first observations of the interstellar Lyman α line were made from a rocket already in 1967. More extensive observations comprising 95 stars were obtained by the OAO 2 satellite. The distances of the observed stars are between 100 and 1000 parsecs.

Comparison of the Lyman α observations with observations of the 21 cm neutral hydrogen line have been especially useful. The distribution of neutral hydrogen over the whole sky has been mapped by means of the 21 cm line. However, the distances to nearby hydrogen clouds are difficult to determine from these observations. In the

Table 16.2 Phenomena caused by the interstellar medium

Observable phenomenon	Cause
Interstellar extinction and polarisation	Non-spherical dust grains aligned by magnetic field
Dark nebulae, uneven distribution of stars and galaxies	Dust clouds
Interstellar absorption lines in stellar spectra	Atoms and molecules in the interstellar gas
Reflection nebulae	Interstellar dust clouds illuminated by nearby stars
Emission nebulae or H II regions (optical, infrared and radio emission)	Interstellar gas and dust cloud, where a nearby hot star ionises the gas and heats the dust to 50–100 K
Optical galactic background (diffuse galactic light)	Interstellar dust illuminated by the integrated light of all stars
Galactic background radiation:	
(a) short wavelength ($\lesssim 1$ m)	Free–free emission from hot interstellar gas
(b) long wavelength ($\gtrsim 1$ m)	Synchrotron radiation from cosmic ray electrons in the magnetic field
Galactic 21 cm emission	Cold (100 K) interstellar neutral hydrogen clouds (HI regions)
Molecular line emission (extended)	Giant molecular clouds (masses even 10^5 – $10^6 M_\odot$), dark nebulae
Point-like OH, H ₂ O and SiO sources	Maser sources near protostars and long-period variables

Fig. 16.14 (a) The D lines D_1 and D_2 of interstellar sodium (rest wavelengths 589.89 and 589.00 nm) in the spectrum of the star HD 14134. Both lines consist of two components formed in the gas clouds of two spiral arms. The radial velocity difference of the arms is about 30 km/s. (Mt. Wilson Observatory). (b) The interstellar absorption lines of ionised calcium Ca II and ionised methylidyne CH⁺ in the spectra of several stars. The emission spectrum of iron is shown for comparison in (a) and (b). (Lick Observatory)



Lyman α observations one usually knows the distance to the star in front of which the absorbing clouds must lie.

The average gas density within about 1 kpc of the Sun derived from the Lyman α observations is 0.7 atoms/cm³. Because the interstellar Lyman α line is so strong, it can be observed even

in the spectra of very nearby stars. For example, it has been detected by the Copernicus satellite in the spectrum of Arcturus, whose distance is only 11 parsecs. The deduced density of neutral hydrogen between the Sun and Arcturus is 0.02–0.1 atoms/cm³. Thus the Sun is situated in a clearing in the interstellar medium, where the

Table 16.3 Element abundances in the interstellar medium towards ζ Ophiuchi and in the Sun. The abundances are given relative to that of hydrogen, which has been defined to be 1,000,000. An asterisk (*) means that

the abundance has been determined from meteorites. The last column gives the ratio of the abundances in the interstellar medium and in the Sun

Atomic number	Name	Chemical symbol	Interstellar abundance	Solar abundance	Abundance ratio
1	Hydrogen	H	1,000,000	1,000,000	1.00
2	Helium	He	85,000	85,000	≈ 1
3	Lithium	Li	0.000051	0.00158*	0.034
4	Beryllium	Be	<0.000070	0.000012	<5.8
5	Boron	B	0.000074	0.0046*	0.016
6	Carbon	C	74	370	0.20
7	Nitrogen	N	21	110	0.19
8	Oxygen	O	172	660	0.26
9	Fluorine	F	–	0.040	–
10	Neon	Ne	–	83	–
11	Sodium	Na	0.22	1.7	0.13
12	Magnesium	Mg	1.05	35	0.030
13	Aluminium	Al	0.0013	2.5	0.00052
14	Silicon	Si	0.81	35	0.023
15	Phosphorus	P	0.021	0.27	0.079
16	Sulfur	S	8.2	16	0.51
17	Chlorine	Cl	0.099	0.45	0.22
18	Argon	Ar	0.86	4.5	0.19
19	Potassium	K	0.010	0.11	0.094
20	Calcium	Ca	0.00046	2.1	0.00022
21	Scandium	Sc	–	0.0017	–
22	Titanium	Ti	0.00018	0.055	0.0032
23	Vanadium	V	<0.0032	0.013	<0.25
24	Chromium	Cr	<0.002	0.50	<0.004
25	Manganese	Mn	0.014	0.26	0.055
26	Iron	Fe	0.28	25	0.011
27	Cobalt	Co	<0.19	0.032	<5.8
28	Nickel	Ni	0.0065	1.3	0.0050
29	Copper	Cu	0.00064	0.028	0.023
30	Zinc	Zn	0.014	0.026	0.53

density is less than one tenth of the average density.

If a hydrogen atom in its ground state absorbs radiation with a wavelength smaller than 91.2 nm, it will be ionised. Knowing the density of neutral hydrogen, one can calculate the expected distance a 91.2 nm photon can propagate before being absorbed in the ionisation of a hydrogen atom. Even in the close neighbourhood of the Sun, where the

density is exceptionally low, the mean free path of a 91.2 nm photon is only about a parsec and that of a 10 nm photon a few hundred parsecs. Thus only the closest neighbourhood of the Sun can be studied in the extreme ultraviolet (XUV) spectral region.

The Hydrogen 21 cm Line The spins of the electron and proton in the neutral hydrogen

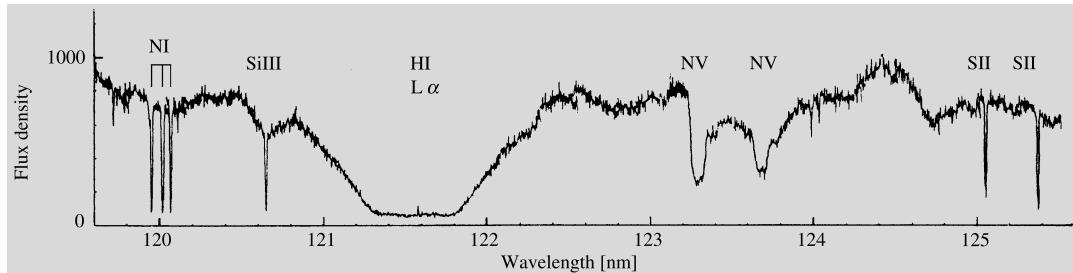


Fig. 16.15 Interstellar absorption lines in the ultraviolet spectrum of ζ Ophiuchi. The strongest line is the hydrogen Lyman α line (equivalent width, more than 1 nm).

The observations were made with the Copernicus satellite. (Morton, D.C. (1975): *Astrophys. J.* **197**, 85)

atom in the ground state may be either parallel or opposite. The energy difference between these two states corresponds to the frequency of 1420.4 MHz. Thus transitions between these two hyperfine structure energy levels will give rise to a spectral line at the wavelength of 21.049 cm (Fig. 5.8). The existence of the line was theoretically predicted by *Hendrick van de Hulst* in 1944, and was first observed by *Harold Ewen* and *Edward Purcell* in 1951. Studies of this line have revealed more about the properties of the interstellar medium than any other method—one might even speak of a special branch of 21 cm astronomy. The spiral structure and rotation of the Milky Way and other galaxies can also be studied by means of the 21 cm line.

Usually the hydrogen 21 cm occurs in emission. Because of the large abundance of hydrogen, it can be observed in all directions in the sky. Some observed 21 cm line profiles are shown in Fig. 16.16. Rather than frequency or wavelength, the radial velocity calculated from the Doppler formula is plotted on the horizontal axis. This is because the broadening of the 21 cm spectral line is always due to gas motions either within the cloud (turbulence) or of the cloud as a whole. The vertical axis is mostly plotted in terms of the antenna temperature T_A (see Chap. 5), the usual radio astronomical measure of intensity. The brightness temperature of an extended source is then $T_b = T_A/\eta_B$, where η_B is the beam efficiency of the antenna.

For the 21 cm line $h\nu/k = 0.07$ K, and thus $h\nu/kT \ll 1$ for all relevant temperatures. One

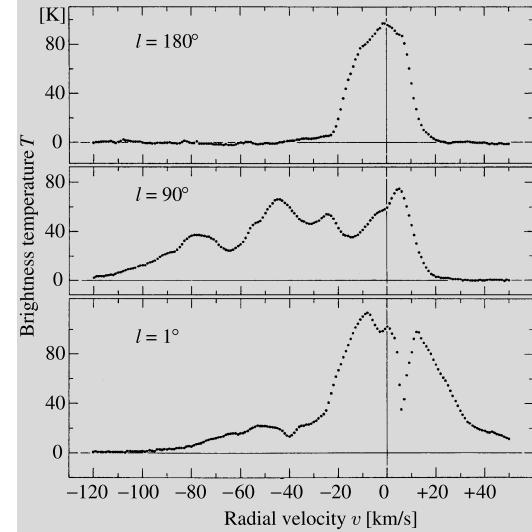


Fig. 16.16 Hydrogen 21 cm emission line profiles in the galactic plane at longitude 180° , 90° and 1° (in the direction $l = 0^\circ$ there is strong absorption). The horizontal axis gives the radial velocity according to the Doppler formula, the vertical axis gives the brightness temperature. (Burton, W.B. (1974): “The Large Scale Distribution of Neutral Hydrogen in the Galaxy”, in *Galactic and Extra-Galactic Radio Astronomy*, ed. by Verschuur, G.L., Kellermann, K.I. (Springer, Berlin, Heidelberg, New York) p. 91)

may therefore use the Rayleigh–Jeans approximation (5.24)

$$I_\nu = \frac{2\nu^2 k T}{c^2}. \quad (16.18)$$

In the solution of the equation of radiative transfer (5.42) the intensity can thus be directly related to a corresponding temperature. By definition, I_ν is related to the brightness temperature T_b , and

the source function S_ν is related to the excitation temperature T_{exc} , i.e.

$$T_b = T_{\text{exc}}(1 - e^{-\tau_\nu}). \quad (16.19)$$

In certain directions in the Milky Way there is so much hydrogen along the line of sight that the 21 cm line is optically thick, $\tau_\nu \gg 1$. In that case

$$T_b = T_{\text{exc}}, \quad (16.20)$$

i.e. the brightness temperature immediately yields the excitation temperature of the cloud. This is often referred to as the *spin temperature* T_S .

The excitation temperature need not always agree with the kinetic temperature of the gas. However, in the present case the population numbers of the hyperfine levels are determined by mutual collisions of hydrogen atoms: the time between collisions is 400 years on the average, whereas the time for spontaneous radiative transitions is 11 million years; thus the excitation temperature will be the same as the kinetic temperature. The observed temperature is $T \approx 125$ K.

The distance to a source cannot be obtained directly from the observed emission. Thus one can only study the number of hydrogen atoms in a cylinder with a 1 cm^2 base area extending from the observer to outside the Milky Way along the line of sight. This is called the *projected* or *column density* and is denoted by N . One may also consider the column density $N(v) dv$ of atoms with velocities in the interval $[v, v + dv]$.

It can be shown that if the gas is optically thin, the brightness temperature in a spectral line is directly proportional to the column density N of atoms with the corresponding radial velocity. Hence, if the diameter L of a cloud along the line of sight is known, the gas density can be determined from the observed line profile:

$$n = N/L.$$

The diameter L can be obtained from the apparent diameter, if the distance and shape of the cloud are assumed known.

The distances of clouds can be determined from their radial velocities by making use of the rotation of the Milky Way (Sect. 18.3). Thus if

the observed peaks in the 21 cm line profiles (Fig. 16.16) are due to individual clouds, their distances and densities can be obtained. Since radio observations are not affected by extinction, it has been possible in this way to map the density distribution of neutral hydrogen in the whole galactic plane. The resulting distribution, based on observations at Leiden and Parkes, is shown in Fig. 16.17. It appears that the Milky Way is a spiral galaxy and that the interstellar hydrogen is concentrated in the spiral arms. The average density of interstellar hydrogen is 1 atom/cm^3 , but the distribution is very inhomogeneous. Typically the hydrogen forms denser regions, a few parsecs in size, where the densities may be $10\text{--}100 \text{ atoms/cm}^3$. Regions where the hydrogen is predominantly neutral are known as *H I regions* (in contrast to H II regions of ionised hydrogen).

The hydrogen 21 cm line may also occur in absorption, when the light from a bright radio source, e.g. a quasar, passes through an intervening cloud. The same cloud may give rise to both an absorption and an emission spectrum. In that case the temperature, optical thickness and hydrogen content of the cloud can all be derived.

Like interstellar dust hydrogen is concentrated in a thin disk in the galactic plane. The thickness of the hydrogen layer is about twice that of the dust or about 200 pc.

H II Regions In many parts of space hydrogen does not occur as neutral atoms, but is ionised. This is true in particular around hot O stars, which radiate strongly in the ultraviolet. If there is enough hydrogen around such a star, it will be visible as an emission nebula of ionised hydrogen. Such nebulae are known as *H II region* (Figs. 16.18 and 16.19).

A typical emission nebula is the great nebula in Orion, M42. It is visible even to the unaided eye, and is a beautiful sight when seen through a telescope. In the middle of the nebula there is a group of four hot stars known as the Trapezium, which can be distinguished inside the bright nebula, even with a small telescope. The Trapezium stars emit strong ultraviolet radiation, which keeps the gas nebula ionised.

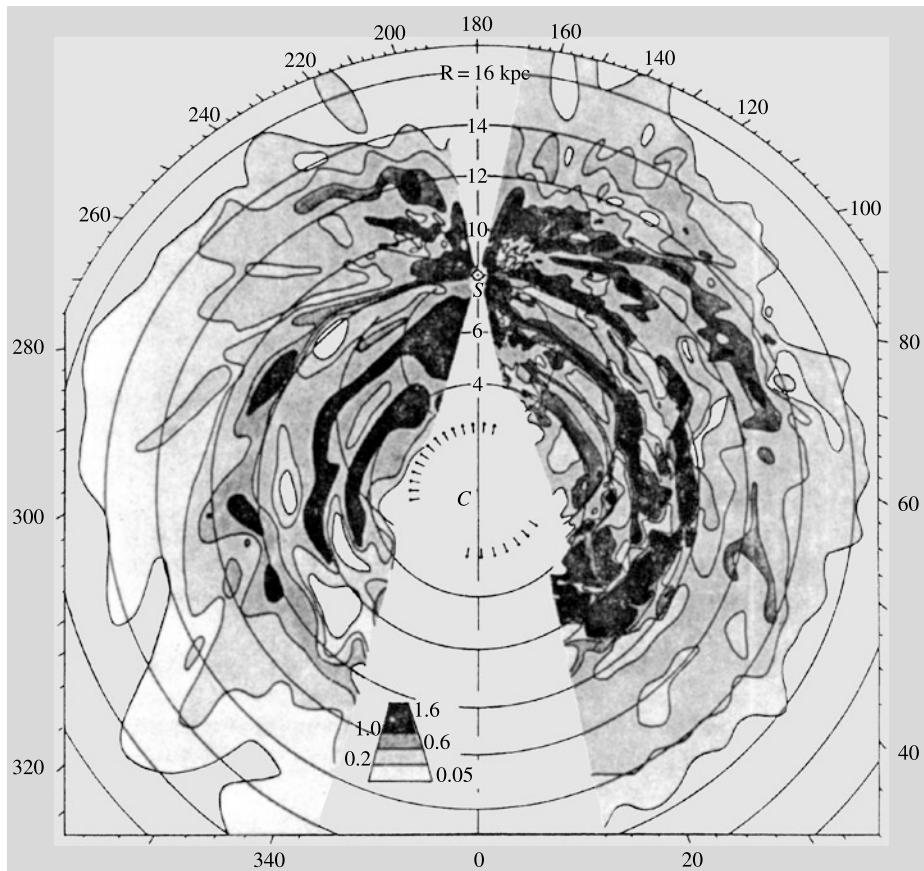


Fig. 16.17 The distribution of neutral hydrogen in the galaxy from the Leiden and Parkes surveys. The density is given in atoms/cm³. (Oort, J.H., Kerr, P.T., Westerhout, G.L. (1958): Mon. Not. R. Astron. Soc. **118**, 379)

Unlike a star a cloud of ionised gas has a spectrum dominated by a few narrow emission lines. The continuous spectrum of H II regions is weak. In the visible region the hydrogen Balmer emission lines are particularly strong. These are formed when a hydrogen atom recombines into an excited state and subsequently returns to the ground state via a sequence of radiative transitions. Typically a hydrogen atom in a H II region remains ionised for several hundred years. Upon recombination it stays neutral for some months, before being ionised again by a photon from a nearby star.

The number of recombinations per unit time and volume is proportional to the product of the densities of electrons and ions,

$$n_{\text{rec}} \propto n_e n_i. \quad (16.21)$$

In completely ionised hydrogen, $n_e = n_i$, and hence

$$n_{\text{rec}} \propto n_e^2. \quad (16.22)$$

Most recombinations will include the transition $n = 3 \rightarrow 2$, i.e. will lead to the emission of a H_α photon. Thus the surface brightness of a nebula in the H_α line will be proportional to the *emission measure*,

$$EM = \int n_e^2 dl, \quad (16.23)$$

where the integral is along the line of sight through the nebula.

The ionisation of a helium atom requires more energy than that of a hydrogen atom, and thus regions of ionised helium are formed only around

Fig. 16.18 The great nebula in Orion (M42, NGC 1976). The nebula gets its energy from newly formed hot stars. The dark regions are opaque dust clouds in front of the nebula. Radio and infrared observations have revealed a rich molecular cloud behind the nebula (Fig. 16.20). In the upper part of this picture is the gas nebula NGC 1977, in the lower part the bright star τ Orionis. (Lick Observatory)



the hottest stars. In these cases, a large H II region will surround a smaller central He^+ or He^{++} region. The helium lines will then be strong in the spectrum of the nebula.

Although hydrogen and helium are the main constituents of clouds, their emission lines are not always strongest in the spectrum. At the beginning of this century it was even suggested that some strong unidentified lines in the spectra of nebulae were due to the new element nebulium. However, in 1927 *Ira S. Bowen* showed that they were *forbidden lines* of ionised oxygen and nitrogen, O^+ , O^{++} and N^+ . Forbidden lines are extremely difficult to observe in the laboratory, because their transition probabilities are so small that at laboratory densities the ions are de-

excited by collisions before they have had time to radiate. In the extremely diffuse interstellar gas, collisions are much less frequent, and thus there is a chance that an excited ion will make the transition to a lower state by emitting a photon.

Because of interstellar extinction, only the nearest H II regions can be studied in visible light. At infrared and radio wavelengths much more distant regions can be studied. The most important lines at radio wavelengths are recombination lines of hydrogen and helium; thus the hydrogen transition between energy levels 110 and 109 at 5.01 GHz has been much studied. These lines are also important because with their help radial velocities, and hence (using the galactic rotation

Fig. 16.19 The Lagoon nebula (M8, NGC 6523) in Sagittarius. This H II region contains many stars of early spectral types and stars that are still contracting towards the main sequence. Small, round dark nebulae, globules, are also visible against the bright background. These are presumably gas clouds in the process of condensation into stars. (National Optical Astronomy Observatories, Kitt Peak National Observatory)



law), distances of H II regions can be determined, just as for neutral hydrogen.

The physical properties of H II regions can also be studied by means of their continuum radio emission. The radiation is due to bremsstrahlung or free-free emission from the electrons. The intensity of the radiation is proportional to the emission measure EM defined in (16.23). H II regions also have a strong infrared continuum emission. This is thermal radiation from dust inside the nebula.

H II regions are formed when a hot O or B star begins to ionise its surrounding gas. The ionisation steadily propagates away from the star. Because neutral hydrogen absorbs ultraviolet radiation so efficiently, the boundary between the H II region and the neutral gas is very sharp. In a homogeneous medium the H II region around a single star will be spherical, forming a *Strömgren sphere*. For a B0 V star the radius of the Strömgren sphere is 50 pc and for an A0 V star only 1 pc.

The temperature of a H II region is higher than that of the surrounding gas, and it therefore tends to expand. After millions of years, it will have become extremely diffuse and will eventually merge with the general interstellar medium.

16.3 Interstellar Molecules

The first *interstellar molecules* were discovered in 1937–1938, when molecular absorption lines were found in the spectra of some stars. Three simple diatomic molecules were detected: *methylidyne* CH, its positive ion CH^+ and *cyanogen* CN. A few other molecules were later discovered by the same method in the ultraviolet. Thus *molecular hydrogen* H_2 was discovered in the early 1970's, and *carbon monoxide*, which had been discovered by radio observations, was also detected in the ultraviolet. Molecular hydrogen is the most abundant interstellar molecule, followed by carbon monoxide.

Molecular Hydrogen The detection and study of molecular hydrogen has been one of the most important achievements of UV astronomy. Molecular hydrogen has a strong absorption band at 105 nm, which was first observed in a rocket experiment in 1970 by *George R. Carruthers*, but more extensive observations could only be made with the *Copernicus* satellite. The observations showed that a significant fraction of interstellar hydrogen is molecular, and that this fraction increases strongly for denser clouds with higher ex-

Fig. 16.20 Radio map of the distribution of carbon monoxide $^{13}\text{C}^{16}\text{O}$ in the molecular cloud near the Orion nebula. The curves are lines of constant intensity. (Kutner, M.L., Evans II, N.J., Tucker, K.D. (1976): *Astrophys. J.* **209**, 452)



tinction. In clouds with visual extinction larger than one magnitude essentially all the hydrogen is molecular.

Hydrogen molecules are formed on the surface of interstellar grains, which thus act as a chemical catalyst. Dust is also needed to shield the molecules from the stellar UV radiation, which would otherwise destroy them. Molecular hydrogen is thus found where dust is abundant. It is of interest to know whether gas and dust are well mixed or whether they form separate clouds and condensations.

UV observations have provided a reliable way of comparing the distribution of interstellar gas and dust. The amount of dust between the observer and a star is obtained from the extinction of the stellar light. Furthermore, the absorption lines of atomic and molecular hydrogen in the ultraviolet spectrum of the same star can be observed. Thus the total amount of hydrogen (atomic + molecular) between the observer and the star can also be determined.

Observations indicate that the gas and dust are well mixed. The amount of dust giving rise to one magnitude visual extinction corresponds to 1.9×10^{21} hydrogen atoms (one molecule is counted as two atoms). The mass ratio of gas and dust obtained in this way is 100.

Radio Spectroscopy Absorption lines can only be observed if there is a bright star behind the molecular cloud. Because of the large dust extinction, no observations of molecules in the densest clouds can be made in the optical and ultraviolet spectral regions. Thus only radio observations are possible for these objects, where molecules are especially abundant.

Radio spectroscopy signifies an immense step forward in the study of interstellar molecules. In the early 1960's, it was still not believed that there might be more complicated molecules than diatomic ones in interstellar space. It was thought that the gas was too diffuse for molecules to form and that any that formed would be destroyed by ultraviolet radiation. The first molecular radio line, the hydroxyl radical OH, was discovered in 1963. Many other molecules have been discovered since then. By 2002, about 130 molecules had been detected, the heaviest one being the 13-atom molecule HC₁₁N.

Molecular lines in the radio region may be observed either in absorption or in emission. Radiation from diatomic molecules like CO (see Fig. 16.20) may correspond to three kinds of transitions. (1) *Electron transitions* correspond to changes in the electron cloud of the molecule. These are like the transitions in single atoms, and

Table 16.4 Some molecules observed in the interstellar medium

Molecule	Name	Year of discovery
<i>Discovered in the optical and ultraviolet region:</i>		
CH	methylidyne	1937
CH ⁺	methylidyne ion	1937
CN	cyanogen	1938
H ₂	hydrogen molecule	1970
CO	carbon monoxide	1971
<i>Discovered in the radio region:</i>		
OH	hydroxyl radical	1963
CO	carbon monoxide	1970
CS	carbon monosulfide	1971
SiO	silicon monoxide	1971
SO	sulfur monoxide	1973
H ₂ O	water	1969
HCN	hydrogen cyanide	1970
NH ₃	ammonia	1968
H ₂ CO	formaldehyde	1969
HCOOH	formic acid	1975
HCCNC	isocyanoacetylene	1991
C ₂ H ₄ O	vinyl alcohol	2001
H ₂ CCCC	cumulene carbene	1991
(CH ₃) ₂ O	dimethyl ether	1974
C ₂ H ₅ OH	ethanol	1975
HC ₁₁ N	cyanopentacetylene	1981

their wavelengths lie in the optical or ultraviolet region. (2) *Vibrational transitions* correspond to changes in the vibrational energy of the molecule. Their energies are usually in the infrared region. (3) Most important for radio spectroscopy are the *rotational transitions*, which are changes in the rotational energy of the molecule. Molecules in their ground state do not rotate, i.e. their angular momentum is zero, but they may be excited and start rotating in collisions with other molecules. For example, carbon sulfide CS returns to its ground state in a few hours by emitting a millimetre region photon.

A number of interstellar molecules are listed in Table 16.4. Many of them have only been detected in the densest clouds (mainly the Sagittarius B2 cloud at the galactic centre), but others are very common. The most abundant molecule H₂ cannot be observed at radio wavelengths, because

it has no suitable spectral lines. The next most abundant molecules are carbon monoxide CO, the hydroxyl radical OH and ammonia NH₃, although their abundance is only a small fraction of that of hydrogen. However, the masses of interstellar clouds are so large that the number of molecules is still considerable. (The Sagittarius B2 cloud contains enough ethanol, C₂H₅OH, for 10²⁸ bottles of vodka.)

Both the formation and survival of interstellar molecules requires a higher density than is common in interstellar clouds; thus they are most common in dense clouds. Molecules are formed in collisions of atoms or simpler molecules or catalysed on dust grains. Molecular clouds must also contain a lot of dust to absorb the ultraviolet radiation entering from outside that otherwise would disrupt the molecules. The most suitable conditions are thus found inside dust and molecular clouds near dense dark nebulae and H II regions.

Most of the molecules in Table 16.4 have only been detected in dense molecular clouds occurring in connection with H II regions. Almost every molecule yet discovered has been detected in Sagittarius B2 near the galactic centre. Another very rich molecular cloud has been observed near the H II region Orion A. In visible light this region has long been known as the Orion nebula M42 (Fig. 16.18). Inside the actual H II regions there are no molecules, since they would be rapidly dissociated by the high temperature and strong ultraviolet radiation. Three types of molecular sources have been found near H II regions (Fig. 16.21):

1. Large gas and dust envelopes around the H II region.
2. Small dense clouds inside these envelopes.
3. Very compact OH and H₂O maser sources.

The large envelopes have been discovered primarily by CO observations. OH and H₂CO have also been detected. Like in the dark nebulae the gas in these clouds is probably mainly molecular hydrogen. Because of the large size and density ($n \approx 10^3\text{--}10^4 \text{ molecules/cm}^3$) of these clouds, their masses are very large, 10⁵ or even 10⁶ solar

Table 16.5 The five phases of interstellar gas

		T [K]	n [cm $^{-3}$]
1.	Very cold molecular gas clouds (mostly hydrogen H ₂)	20	$\gtrsim 10^3$
2.	Cold gas clouds (mostly atomic neutral hydrogen)	100	20
3.	Warm neutral gas enveloping the cooler clouds	6000	0.05–0.3
4.	Hot ionised gas (mainly H II regions around hot stars)	8000	>0.5
5.	Very hot and diffuse ionised coronal gas, ionised and heated by supernova explosions	10^6	10^{-3}

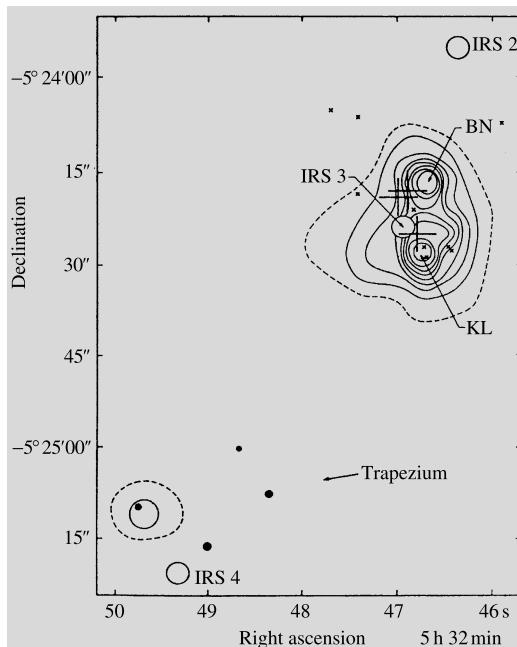


Fig. 16.21 Infrared map of the central part of the Orion nebula. In the lower part are the four Trapezium stars. Above is an infrared source of about 0.5" diameter, the Kleinmann–Low nebula (KL). BN is an infrared point source, the Becklin–Neugebauer object. Other infrared sources are denoted IRS. The large crosses indicate OH masers and the small crosses H₂O masers. On the scale of Fig. 16.18 this region would only be a few millimetres in size. (Goudis, C. (1982): The Orion Complex: A Case Study of Interstellar Matter (Reidel, Dordrecht) p. 176)

masses (Sgr B2). They are among the most massive objects in the Milky Way. The dust in molecular clouds can be observed on the basis of its thermal radiation. Its peak falls at wavelengths of 10–100 μm, corresponding to a dust temperature of 30–300 K.

Some interstellar clouds contain very small maser sources. In these the emission lines of OH,

H₂O and SiO may be many million times stronger than elsewhere. The diameter of the radiating regions is only about 5–10 au. The conditions in these clouds are such that radiation in some spectral lines is amplified by stimulated emission as it propagates through the cloud. Hydroxyl and water masers occur in connection with dense H II regions and infrared sources, and appear to be related to the formation of protostars. In addition maser emission (OH, H₂O and SiO) occurs in connection with Mira variables and some red supergiant stars. This maser emission comes from a molecule and dust envelope around the star, which also gives rise to an observable infrared excess.

16.4 The Formation of Protostars

The mass of the Milky Way is about 10^{11} solar masses. Since its age is about 10^{10} years, stars have been forming at the average rate of $10 M_{\odot}$ per year. This estimate is only an upper limit for the present rate, because earlier the rate of star formation must have been much higher. Since the lifetime of O stars is only about a million years, a better estimate of the star formation rate can be made, based on the observed number of O stars. Accordingly, it has been concluded that at present, new stars are forming in the Milky Way at a rate of about three solar masses per year.

Stars are now believed to form inside large dense interstellar clouds mostly located in the spiral arms of the Galaxy. Under its own gravity, a cloud begins to contract and fragment into parts that will become protostars. The observations seem to indicate that stars are not formed individually, but in larger groups. Young stars are

found in open clusters and in loose associations, typically containing a few hundred stars which must have formed simultaneously.

Theoretical calculations confirm that the formation of single stars is almost impossible. An interstellar cloud can contract only if its mass is large enough for gravity to overwhelm the pressure. As early as in the 1920's, James Jeans calculated that a cloud with a certain temperature and density can condense only if its mass is high enough. If the mass is too small the pressure of the gas is sufficient to prevent the gravitational contraction. The limiting mass is the Jeans mass (Sect. 6.11), given by

$$M_J \approx 3 \times 10^4 \sqrt{\frac{T^3}{n}} M_{\odot},$$

where n is the density in atoms/m³ and T the temperature.

In a typical interstellar neutral hydrogen cloud $n = 10^6$ and $T = 100$ K, giving the Jeans mass, $30,000 M_{\odot}$. In the densest dark clouds $n = 10^{12}$ and $T = 10$ K and hence, $M_J = 1 M_{\odot}$.

It is thought that star formation begins in clouds of a few thousand solar masses and diameters of about 10 pc. The cloud begins to contract, but does not heat up because the liberated energy is carried away by radiation. As the density increases, the Jeans mass thus decreases. Because of this, separate condensation nuclei are formed in the cloud, which go on contracting independently: the cloud *fragments*. Fragmentation is further advanced by the increasing rotation velocity. The original cloud has a certain angular momentum which is conserved during the contraction; thus the angular velocity must increase.

This contraction and fragmentation continues until the density becomes so high that the individual fragments become optically thick. The energy liberated by the contraction can then no longer escape, and the temperature will begin to rise. In consequence the Jeans mass begins to increase, further fragmentation ceases and the rising pressure in existing fragments stops their contraction. Some of the protostars formed in this way may still be rotating too rapidly. These may split

into two, thus forming double systems. The further evolution of protostars has been described in Sect. 11.2.

Although the view that stars are formed by the collapse of interstellar clouds is generally accepted, many details of the fragmentation process are still highly conjectural. Thus the effects of rotation, magnetic fields and energy input are very imperfectly known. Why a cloud begins to contract is also not certain; one theory is that passage through a spiral arm compresses clouds and triggers contraction (see Sect. 17.4). This would explain why young stars are predominantly found in the spiral arms of the Milky Way and other galaxies. The contraction of an interstellar cloud might also be initiated by a nearby expanding H II region or supernova explosion.

Star formation can be observed particularly well in the infrared, since the temperatures of the condensing clouds and protostars are of the order 100–1000 K and the infrared radiation can escape even the densest dust clouds. For example, in connection with the Orion nebula there is a large cloud of hydrogen, found in radio observations, containing small infrared sources. E.g. the Becklin–Neugebauer object has a temperature of a couple of hundred kelvins but a luminosity that is thousandfold compared with the Sun. It is a strong H₂O maser source, located next to a large H II region.

16.5 Planetary Nebulae

Bright regions of ionised gas do not occur only in connection with newly formed stars, but also around stars in late stages of their evolution. The *planetary nebulae* are gas shells around small hot blue stars. As we have seen in connection with stellar evolution, instabilities may develop at the stage of helium burning. Some stars begin to pulsate, while in others the whole outer atmosphere may be violently ejected into space. In the latter case, a gas shell expanding at 20–30 km/s will be formed around a small and hot (50,000–100,000 K) star, the core of the original star.

The expanding gas in a planetary nebula is ionised by ultraviolet radiation from the central

Fig. 16.22 The Helix nebula (NGC 7293). The planetary nebulae are formed during the final stages of evolution of solar-type stars. The centrally visible star has ejected its outer layers into space. (National Optical Astronomy Observatories, Kitt Peak National Observatory)



star, and its spectrum contains many of the same bright emission lines as that of an H II region. Planetary nebulae are, however, generally much more symmetrical in shape than most H II regions, and they expand more rapidly. For example, the well-known Ring nebula in Lyra (M57) has expanded visibly in photographs taken at 50-year intervals. In a few ten thousand years, the planetary nebulae disappear in the general interstellar medium and their central stars cool to become white dwarfs.

The planetary nebulae were given their name in the 19th century, because certain small nebulae visually look quite like planets such as Uranus. The apparent diameter of the smallest known planetary nebulae is only a few arc seconds, whereas the largest ones (like the Helix nebula) may be one degree in diameter (Fig. 16.22).

The brightest emission lines are often due to forbidden transitions, like in H II regions. For example, the green colour of the central parts of the Ring nebula in Lyra is due to the forbidden lines of doubly ionised oxygen at 495.9 and 500.7 nm. The red colour of the outer parts is due to the hydrogen Balmer α line (656.3 nm) and the forbidden lines of ionised nitrogen (654.8 nm, 658.3 nm).

The total number of planetary nebulae in the Milky Way has been estimated to be 50,000. About 2000 have actually been observed.

16.6 Supernova Remnants

In Chap. 11 we have seen that massive stars end their evolution in a supernova explosion. The collapse of the stellar core leads to the violent ejection of the outer layers, which then remain as an expanding gas cloud.

About 120 *supernova remnants* (SNR's) have been discovered in the Milky Way. Some of them are optically visible as a ring or an irregular nebula (e.g. the Crab nebula; see Fig. 16.23), but most are detectable only in the radio region (because radio emission suffers no extinction).

In the radio region the SNR's are extended sources similar to H II regions. However, unlike H II regions the radiation from SNR's is often polarised. Another characteristic difference between these two kinds of sources is that whereas the radio brightness of H II regions grows or remains constant as the frequency increases, that of SNR's falls off almost linearly (in a $\log I_\nu - \log \nu$ diagram) with increasing frequency (Fig. 16.24).

These differences are due to the different emission processes in H II regions and in SNR's.

Fig. 16.23 The Crab nebula (M1, NGC 1952) in Taurus is the remnant of a supernova explosion observed in 1054. The photograph was taken at red wavelengths. The nebula is also a strong radio source. Its energy source is the central rapidly rotating neutron star, pulsar, which is the collapsed core of the original star. (Palomar Observatory)

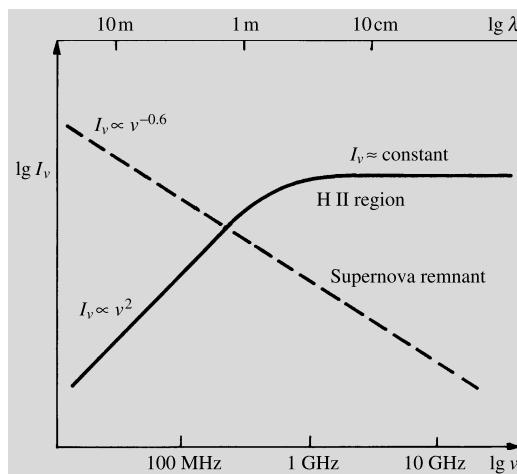


Fig. 16.24 The radio spectra of typical H II regions and supernova remnants. The radiation of H II regions is thermal and obeys the Rayleigh–Jeans law, $I \propto v^2$, at wavelengths larger than 1 m. In supernova remnants the intensity decreases with increasing frequency. (After Scheffler, H., Elsässer, H. (1987): *Physics of the Galaxy and the Interstellar Matter* (Springer, Berlin, Heidelberg, New York))

In an H II region, the radio emission is free–free radiation from the hot plasma. In a SNR it is *synchrotron radiation* from relativistic electrons moving in spiral orbits around the magnetic field

lines. The synchrotron process gives rise to a continuous spectrum extending over all wavelength regions. For example, the Crab nebula looks blue or green in colour photographs because of optical synchrotron radiation.

In the Crab nebula red filaments are also visible against the bright background. Their emission is principally in the hydrogen H_α line. The hydrogen in a SNR is not ionised by a central star as in the H II regions, but by the ultraviolet synchrotron radiation.

The supernova remnants in the Milky Way fall into two classes. One type has a clearly ring-like structure (e.g. Cassiopeia A or the Veil nebula in Cygnus; see Fig. 16.25); another is irregular and bright at the middle (like the Crab nebula). In the remnants of the Crab nebula type there is always a rapidly rotating pulsar at the centre. This pulsar provides most of the energy of the remnant by continuously injecting relativistic electrons into the cloud. The evolution of this type of SNR reflects that of the pulsar and for this reason has a time scale of a few ten thousand years.

Ring-like SNR's do not contain an energetic pulsar; their energy comes from the actual supernova explosion. After the explosion, the cloud expands at a speed of 10,000–20,000 km/s. About

Fig. 16.25 The Veil nebula (NGC 6960 at the right, NGC 6992 at the left) in Cygnus is the remnant of a supernova explosion which occurred several ten thousand years ago. (Mt. Wilson Observatory)

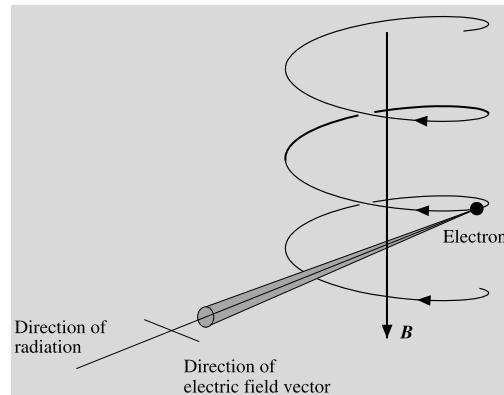


50–100 years after the explosion the remnant begins to form a spherical shell as the ejected gas starts to sweep up interstellar gas and to slow down in its outer parts. The swept-up shell expands with a decreasing velocity and cools until, after about 100,000 years, it merges into the interstellar medium. The two types of supernova remnants may be related to the two types (I and II) of supernovae.

Box 16.1 (Synchrotron Radiation) Synchrotron radiation was first observed in 1948 by *Frank Elder, Robert Langmuir and Herbert Pollack*, who were experimenting with an electron synchrotron, in which electrons were accelerated to relativistic energies in a magnetic

field. It was observed that the electrons radiated visible light in a narrow cone along their momentary direction of motion. In astrophysics synchrotron radiation was first invoked as an explanation of the radio emission of the Milky Way, discovered by *Karl Jansky* in 1931. This radiation had a spectrum and a large metre-wave brightness temperature (more than 10^5 K) which were inconsistent with ordinary thermal free-free emission from ionised gas. In 1950 *Hannes Alfvén and Nicolai Herlofson* as well as *Karl-Otto Kiepenheuer* proposed that the galactic radio background was due to synchrotron radiation. According to Kiepenheuer the high-energy cosmic ray electrons would emit radio radiation in the weak galactic mag-

netic field. This explanation has turned out to be correct. Synchrotron radiation is also an important emission process in supernova remnants, radio galaxies and quasars. It is a *non-thermal* radiation process, i.e. the energy of the radiating electrons is not due to thermal motions.



The emission of synchrotron radiation. A charged particle (electron) propagating in a magnetic field moves in a spiral. Because of the centripetal acceleration, the particle emits electromagnetic radiation

The origin of synchrotron radiation is schematically shown in the figure. The magnetic field forces the electron to move in a spiral orbit. The electron is thus constantly accelerated and will emit electromagnetic radiation. According to the special theory of relativity, the emission from a relativistic electron will be concentrated in a narrow cone. Like the beam from a lighthouse, this cone sweeps across the observer's field of vision once for each revolution. Thus the observer sees a sequence of radiation flashes of very short duration compared with their interval. (In the total emission of a large number of electrons, separate flashes cannot be distinguished.) When this series of pulses is represented as a sum of different frequency components (Fourier transform), a broad spectrum is obtained with a maximum at

$$\nu_{\max} = a B_{\perp} E^2,$$

where B_{\perp} is the magnetic field component perpendicular to the velocity of the electron, and E its energy, a is a constant of proportionality.

The table gives the frequency and wavelength of the maximum as functions of the electron energy for the typical galactic field strength 0.5 nT:

λ_{\max}	ν_{\max} [Hz]	E [eV]
300 nm	10^{15}	6.6×10^{12}
30 μm	10^{13}	6.6×10^{11}
3 mm	10^{11}	6.6×10^{10}
30 cm	10^9	6.6×10^9
30 m	10^7	6.6×10^8

To produce even radio synchrotron radiation, very energetic electrons are required, but these are known to be present in the cosmic radiation. In the optical galactic background radiation, the contribution from synchrotron radiation is negligible, but, for example, in the Crab nebula, a significant part of the optical emission is due to this mechanism.

16.7 The Hot Corona of the Milky Way

As early as 1956 Lyman Spitzer showed that the Milky Way has to be surrounded by a large envelope of very hot gas (Fig. 16.26). Almost two decades later the Copernicus satellite, whose scientific program was directed by Spitzer, found evidence for this kind of gas, which began to be called *galactic coronal gas*, in analogy with the solar corona. The satellite observed emission lines of e.g. five times ionised oxygen (O VI), four times ionised nitrogen (N V) and triply ionised carbon (C IV). The formation of these lines requires a high temperature (100,000–1,000,000 K), and a high temperature is also indicated by the broadening of the lines.

Galactic coronal gas is distributed through the whole Milky Way and extends several thousand parsecs from the galactic plane. Its density is only of the order of 10^{-3} atoms/cm³ (recall that the mean density in the galactic plane is 1 atom/cm³). Thus coronal gas forms a kind of background sea, from which the denser and cooler forms of interstellar matter, such as neutral hydrogen and molecular clouds, rise as islands. In the early 1980's the IUE satellite also

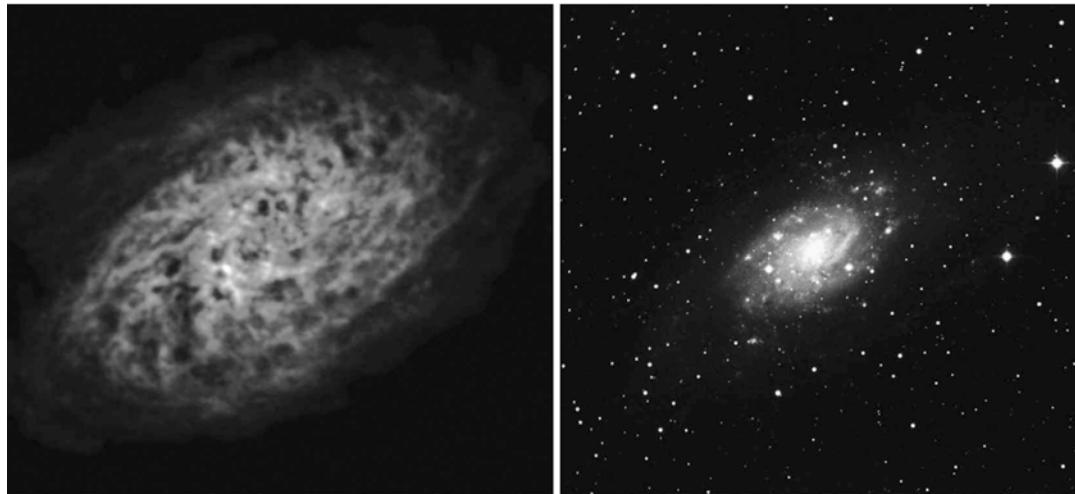


Fig. 16.26 Galaxy's hot corona. NGC 2403 is a spiral galaxy, similar to our Milky Way. *On the right*, it is photographed in visual light. *On the left*, in a VLA radio image, on the same scale with the optical image,

a large hydrogen corona is seen around the galaxy. Large holes created by supernova explosions are seen in the gas corona. (Image NRAO/AUI and Tom Oosterloo, Astron, The Netherlands)

detected similar coronae in the Large Magellanic Cloud and in the spiral galaxy M100. Coronal gas is probably quite a common and important form of matter in galaxies.

Supernova explosions are probably the source of both coronal gas and its energy. When a supernova explodes, it forms a hot bubble in the surrounding medium. The bubbles from neighbouring supernovae will expand and merge, forming a foamlke structure. In addition to supernovae, stellar winds from hot stars may provide some of the energy of the coronal gas.

therefore gives no information about their place of origin. The most important properties of cosmic rays that can be observed from the Earth are their particle composition and energy distribution. As noted in Sect. 3.6, these observations have to be made in the upper atmosphere or from satellites, since cosmic ray particles are destroyed in the atmosphere.

The main constituent of the cosmic rays (about 90 %) is hydrogen nuclei or protons. The second most important constituent (about 9 %) is helium nuclei or α particles. The rest of the particles are electrons and nuclei more massive than helium.

Most cosmic rays have an energy smaller than 10^9 eV. The number of more energetic particles drops rapidly with increasing energy. The most energetic protons have an energy of 10^{20} eV, but such particles are very rare—the energy of one such proton could lift this book about one centimetre. (The largest particle accelerators reach “only” energies of 10^{12} eV.)

The distribution of low-energy (less than 10^8 eV) cosmic rays cannot be reliably determined from the Earth, since solar “cosmic rays”, high-energy protons and electrons formed in solar flares fill the solar system and strongly affect the motion of low-energy cosmic rays.

16.8 Cosmic Rays and the Interstellar Magnetic Field

Cosmic Rays Elementary particles and atomic nuclei reaching the Earth from space are called *cosmic rays*. They occur throughout interstellar space with an energy density of the same order of magnitude as that of the radiation from stars. Cosmic rays are therefore important for the ionisation and heating of interstellar gas.

Since cosmic rays are charged, their direction of propagation in space is constantly changed by the magnetic field. Their direction of arrival

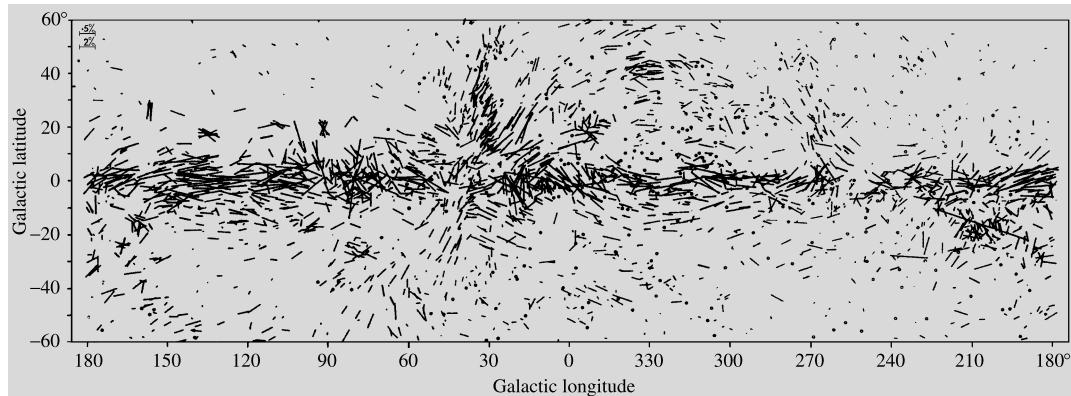


Fig. 16.27 The polarisation of starlight. *The dashes give the direction and degree of the polarisation. The thinner dashes correspond to stars with polarisation smaller than 0.6 %; the thicker dashes to stars with larger polarisation.*

The scale is shown in the upper left-hand corner. Stars with polarisation smaller than 0.08 % are indicated by a small circle. (Mathewson, D.S., Ford, V.L. (1970): Mem. R.A.S. **74**, 139)

The distribution of cosmic rays in the Milky Way can be directly inferred from gamma-ray and radio observations. The collisions of cosmic ray protons with interstellar hydrogen atoms gives rise to pions which then decay to form a gamma-ray background. The radio background is formed by cosmic ray electrons which emit synchrotron radiation in the interstellar magnetic field.

Both radio and gamma-ray emission are strongly concentrated in the galactic plane. From this it has been concluded that the sources of cosmic rays must also be located in the galactic plane. In addition there are individual peaks in the backgrounds around known supernova remnants. In the gamma-ray region such peaks are observed at e.g. the Crab nebula and the Vela pulsar; in the radio region the North Polar Spur is a large, nearby ring-like region of enhanced emission.

Apparently a large fraction of cosmic rays have their origin in supernovae. An actual supernova explosion will give rise to energetic particles. If a pulsar is formed, observations show that it will accelerate particles in its surroundings. Finally the shock waves formed in the expanding supernova remnant will also give rise to relativistic particles.

On the basis of the relative abundances of various cosmic ray nuclei, it can be calculated how far they have travelled before reaching the Earth. It has been found that typical cosmic ray protons

have travelled for a period of a few million years (and hence also a distance of a few million light-years) from their point of origin. Since the diameter of the Milky Way is about 100,000 light-years, the protons have crossed the Milky Way tens of times in the galactic field.

The Interstellar Magnetic Field The strength and direction of the *interstellar magnetic field* are difficult to determine reliably. Direct measurements are impossible, since the magnetic fields of the Earth and the Sun are much stronger. However, using various sources it has been possible to deduce the existence and strength of the field.

We have already seen that interstellar grains give rise to interstellar polarisation. In order to polarise light, the dust grains have to be similarly oriented; this can only be achieved by a general magnetic field. Figure 16.27 shows the distribution of interstellar polarisation over the sky. Stars near each other generally have the same polarisation. At low galactic latitudes the polarisation is almost parallel to the galactic plane, except where one is looking along a spiral arm.

More precise estimates of the strength of the magnetic field can be obtained from the rotation of the plane of polarisation of the radio radiation from distant sources. This Faraday rotation is proportional to the strength of the magnetic field and to the electron density. Another method is to

measure the Zeeman splitting of the 21 cm radio line. These measurements have fairly consistently given a value of 10^{-10} – 10^{-9} T for the strength of the interstellar magnetic field. This is about one millionth of the interplanetary field in the solar system.

16.9 Examples

Example 16.1 Estimate the dust grain size and number density in the galactic plane.

Let us compare the interstellar extinction curve in Fig. 16.4(b) with the Mie scattering curves in Fig. 16.3. We see that the leftmost parts of the curves may correspond to each other: the interval $0 < x < 5$ with $m = 1.5$ in Fig. 16.3 matches the interval $0 < 1/\lambda < 5 \mu\text{m}^{-1}$ in Fig. 16.4(b). Remembering that $x = 2\pi a/\lambda$, this suggests a constant grain radius a , given by $2\pi a \approx 1 \mu\text{m}$, or $a \approx 0.16 \mu\text{m}$.

In the blue wavelength region ($\lambda = 0.44 \mu\text{m}$), $x = 2.3$ and, according to the upper Fig. 16.3, $Q_{\text{ext}} \approx 2$. Using $A = 2$ mag for the interstellar extinction at $r = 1$ kpc, we get, substituting (16.5) into (16.7), $\bar{n} \approx 4 \times 10^{-7} \text{ m}^{-3}$. This should give the order of magnitude of the interstellar dust density.

As a summary we could say that a considerable fraction of the interstellar extinction might be due to grains of diameter $0.3 \mu\text{m}$ and particle density of the order of $10^{-7} \text{ m}^{-3} = 100 \text{ km}^{-3}$.

Example 16.2 Estimate the time interval between successive collisions of a hydrogen atom in interstellar gas.

Two atoms will collide if the separation between their centres becomes less than $2r$, where r is the radius of the atom. Thus, the microscopic cross section for the collision is $\sigma = \pi(2r)^2 = 4\pi r^2$. The macroscopic cross section, or the number of collisions of an H atom per unit length, is then $\Sigma = n\sigma$, where n is the number density of the H atoms. The mean free path l of an atom is the inverse of the macroscopic cross section, $l = 1/\Sigma$, and the time between two collisions is $t = l/v$, where v is the velocity of the atom.

Considering the numerical values, the Bohr radius of an H atom is $r = 5.3 \times 10^{-11} \text{ m}$. Taking

$n = 1 \text{ cm}^{-3}$ we get $l = 2.8 \times 10^{13} \text{ m} \approx 0.0009 \text{ pc}$. The average velocity is not far from the root mean square velocity at $T = 125 \text{ K}$, given by (5.33):

$$v = \sqrt{\frac{3kT}{m}} = 1760 \text{ m s}^{-1}.$$

These values of l and v give $t = l/v = 510$ years for the collision interval. Taking into account the velocity distribution in the gas, the mean free path appears to be shorter by a factor of $1/\sqrt{2}$, which reduces the time to about 400 years.

Example 16.3 Consider the lowest rotational transition of the CO molecule. For ^{12}CO the frequency of this line is $\nu(^{12}\text{CO}) = 115.27 \text{ GHz}$, and for ^{13}CO , $\nu(^{13}\text{CO}) = 110.20 \text{ GHz}$. Estimate the optical thickness of each line in a molecular cloud, where the observed brightness temperatures of the lines are $T_b(^{12}\text{CO}) = 40 \text{ K}$ and $T_b(^{13}\text{CO}) = 9 \text{ K}$.

For the ^{12}CO line, $h\nu/k = 5.5 \text{ K}$. Thus, the Rayleigh–Jeans approximation is valid if the temperature is considerably higher than 5 K. This is not always the case, but the measured value of $T_b(^{12}\text{CO})$ suggests that the approximation can be used.

Ignoring the background, (16.19) gives

$$T_b = T_{\text{exc}}(1 - e^{-\tau_\nu}).$$

The optical thickness τ_ν is proportional to the opacity or the absorption coefficient α_ν [see (4.16)], and α_ν is evidently proportional to the number of CO molecules present. Other differences between the lines are small, so we can write

$$\frac{\tau_\nu(^{12}\text{CO})}{\tau_\nu(^{13}\text{CO})} \approx \frac{n(^{12}\text{CO})}{n(^{13}\text{CO})}.$$

Adopting the terrestrial value $n(^{12}\text{CO})/n(^{13}\text{CO}) = 89$, we set

$$\tau_\nu(^{12}\text{CO}) = 89\tau_\nu(^{13}\text{CO}).$$

Assuming the excitation temperatures equal and denoting $\tau_\nu(^{12}\text{CO})$ by τ , we get

$$T_{\text{exc}}(1 - e^{-\tau}) = 40,$$

$$T_{\text{exc}}(1 - e^{-\tau/89}) = 9.$$

The solution of this pair of equations is

$$\tau_\nu(^{12}\text{CO}) = 23, \quad \tau_\nu(^{13}\text{CO}) = 0.25,$$

$$T_{\text{exc}} = 40 \text{ K}.$$

Thus, the ^{12}CO line seems to be optically thick, and $T_{\text{exc}} = T_b(^{12}\text{CO})$. If also the ^{13}CO line were optically thick, the brightness temperatures would be practically equal, and the optical thicknesses could not be determined.

16.10 Exercises

Exercise 16.1 Two open clusters, which are seen near each other in the galactic plane, have angular diameters α and 3α , and distance moduli 16.0 and 11.0, respectively. Assuming their

actual diameters are equal, find their distances and the interstellar extinction coefficient a in (16.4).

Exercise 16.2 Estimate the free fall velocity on the surface of a spherical gas cloud contracting under the influence of its own gravity. Assume $n(\text{H}_2) = 10^3 \text{ cm}^{-3}$ and $R = 5 \text{ pc}$.

Exercise 16.3 The force \mathbf{F} exerted by a magnetic field \mathbf{B} on a charge q moving with velocity \mathbf{v} is $\mathbf{F} = q\mathbf{v} \times \mathbf{B}$. If \mathbf{v} is perpendicular to \mathbf{B} , the path of the charge is circular. Find the radius of the path of an interstellar proton with a kinetic energy of 1 MeV. Use $B = 0.1 \text{ nT}$ for the galactic magnetic field.

Several collections of stars can be picked out in the sky, even with the naked eye. Closer study reveals that they really do form separate clusters in space. E.g. the Pleiades in Taurus and the Hyades around Aldebaran, the brightest star in Taurus, are such *open star clusters*. Almost the whole of the constellation Coma Berenices is also an open star cluster. Many objects appearing as nebulous patches to the unaided eye, when looked at with a telescope, turn out to be star clusters, like Praesepe in the constellation Cancer, or double cluster in Perseus (Fig. 17.1). In addition to open clusters some apparently nebulous objects are very dense *globular clusters*, such as those in Hercules and in Canes Venatici (Fig. 17.2).

The first catalogue of star clusters was prepared by the French astronomer *Charles Messier* in 1784. The first version contained only 45 objects, but later Messier himself and Pierre Méchain expanded the catalogue to include 103 objects. Later seven more objects were added, possibly observed by Messier. Messier was interested in comets, not in those fuzzy objects, and the reason for the catalogue was to avoid false alarms. The catalogue contains a mixture of very different objects, like about 30 globular clusters and the same number of open clusters, gas nebulae and galaxies.

A larger catalogue, published in 1888, was the *New General Catalogue of Nebulae and Clusters of Stars* prepared by the Danish astronomer *John Louis Emil Dreyer*. The catalogue numbers of objects in this list are preceded by the initials NGC. For example, the large globular cluster in

Hercules is object M13 in the Messier catalogue, and it is also known as NGC 6205. The NGC catalogue was supplemented with the *Index Catalogue* in 1895 and 1910. The objects of this catalogue are given the initials IC.

With a small telescope stars of distant clusters and galaxies cannot be distinguished as separate objects; instead the target looks nebulous. Therefore their true nature was only little by little with spectroscopy and large telescopes. That's why those old catalogues contain all kinds of different objects.

Even now there is no catalogue of open cluster complete to some limiting magnitude. The data are collected from several catalogues published by various astronomers, and therefore the nomenclature is not consistent. The problem is that open clusters concentrate close the plane of Milky Way, and there are a lot of background stars. Thus identifying the cluster members may require rather detailed studies of the properties of the stars. Also the members may already have dispersed to a very wide area making it difficult to decide which of them belong to the same cluster.

17.1 Associations

In 1947 the Soviet astronomer *Viktor Amazaspovich Ambartsumyan* discovered that there are groups of young stars scattered over so large regions of the sky that they would be very difficult to identify merely on the basis of their appearance. These *associations* may have a few tens of

members. One association is found around the star ζ Persei, and in the region of Orion, there are several associations.

Associations are groups of very young stars. They are usually identified on the basis either of absolutely bright main sequence stars or of T Tauri stars. According to the type, one speaks of OB associations and T Tauri associations. The most massive stars of spectral class O stay on the main sequence for only a few million years, and therefore associations containing them are necessarily young. The T Tauri stars are even younger stars that are in the process of contracting towards the main sequence.

Studies of the internal motions in associations show that they are rapidly dispersing. There are

so few stars in an association that their gravity cannot hold them together for any length of time. The observed motions have often confirmed that the stars in an association were very close together a few million years ago (Fig. 17.3).

Large amounts of interstellar matter, gas and dust nebulae often occur in connection with associations, supplying information about the connection between star formation and the interstellar medium. Infrared observations have shown that stars are now forming or have recently formed in many dense interstellar clouds.

Associations are strongly concentrated in the spiral arms in the plane of the Milky Way. Both in the Orion region and in the direction of Cepheus, three generations of associations have been iden-

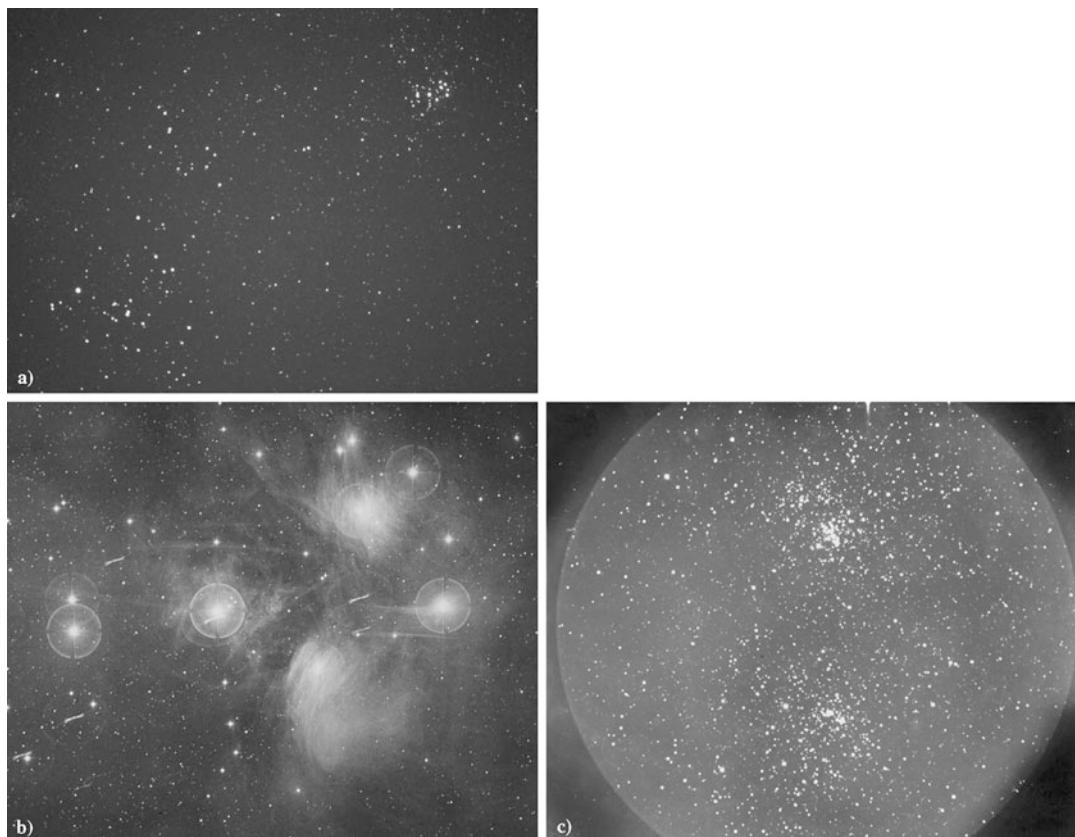


Fig. 17.1 Open clusters. (a) The Hyades slightly to the lower left in the photograph. Above them to the right the Pleiades. (Photograph M. Korpi.) (b) The Pleiades photographed with the Metsähovi Schmidt camera. The diameter of the cluster is about 1° . Reflection nebulae are visible around some of the stars. (Photograph M. Pouta-

nen and H. Virtanen, Helsinki University.) h and χ Persei, the double cluster in Perseus. The separation between the clusters is about $25'$. Picture taken with the Metsähovi 60-cm Ritchey Chrétien telescope. (Photograph T. Markkanen, Helsinki University)

Fig. 17.2 The globular cluster ω Centauri. The picture was taken with the Danish 1.5-m telescope at La Silla, Chile. Thanks to the excellent seeing, one can see through the entire cluster in some places. (Photograph T. Korhonen, Turku University)

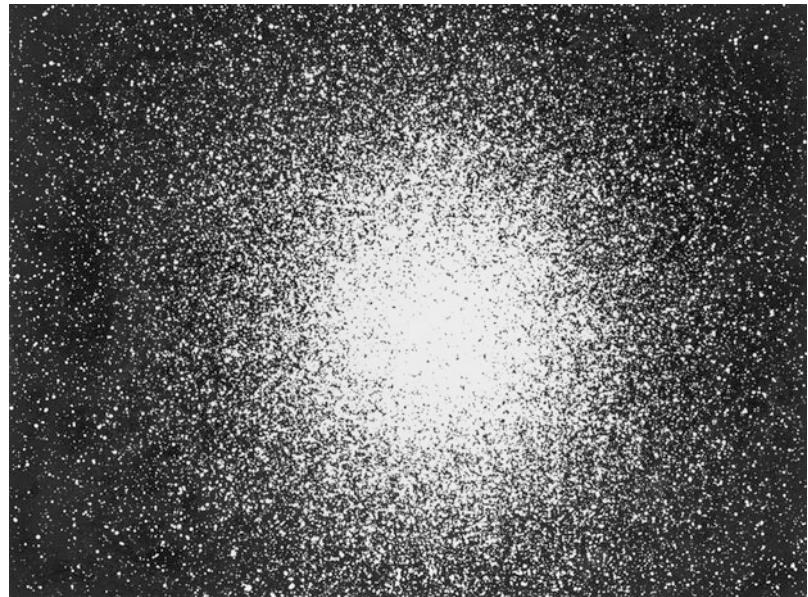
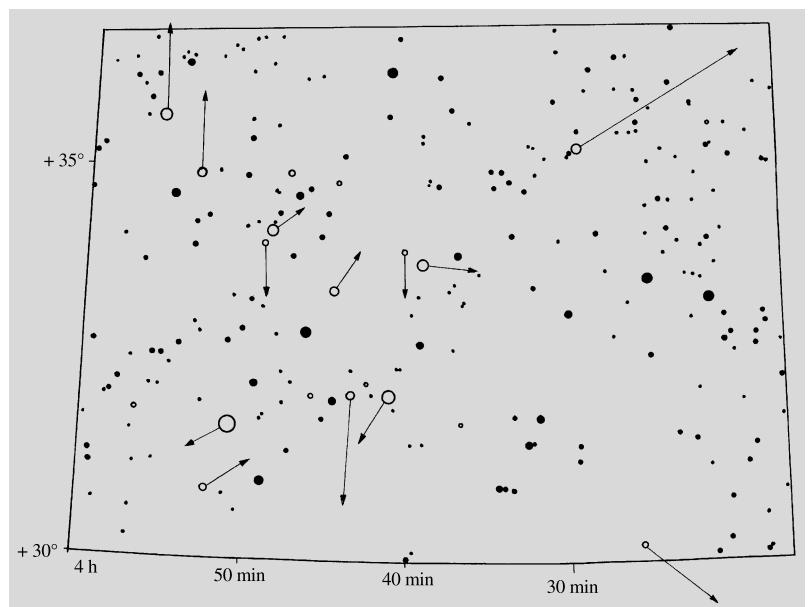


Fig. 17.3 ζ Persei association. O and B stars are shown as *open circles*. The proper motion vectors show the movements of the stars in the next 500,000 years



tified, the oldest ones being most extended and the youngest ones, most dense.

17.2 Open Star Clusters

Open clusters usually contain from a few tens to a few hundreds of stars. The kinetic energy of the cluster members, the differential rotation of

the Milky Way (Sect. 17.3) and external gravitational disturbances tend to gradually disperse the open clusters. Still, many of them are fairly permanent; for example, the Pleiades is many hundreds of millions of years old, but nevertheless, quite a dense cluster.

The distances of star clusters—and also of associations—can be obtained from the photo-

metric or spectroscopic distances of their brightest members. For the nearest clusters, in particular for the Hyades, one can use the method of *kinematic parallaxes*, which is based on the fact that the stars in a cluster all have the same average space velocity with respect to the Sun. The proper motions in the Hyades are shown in Fig. 17.4. They all appear to be directed to the same point. Figure 17.5 explains how this convergence can be understood as an effect of perspective, if all cluster members have the same velocity vector with respect to the observer. Let θ be the angular distance of a given star from the convergence point. The angle between the velocity of the star and

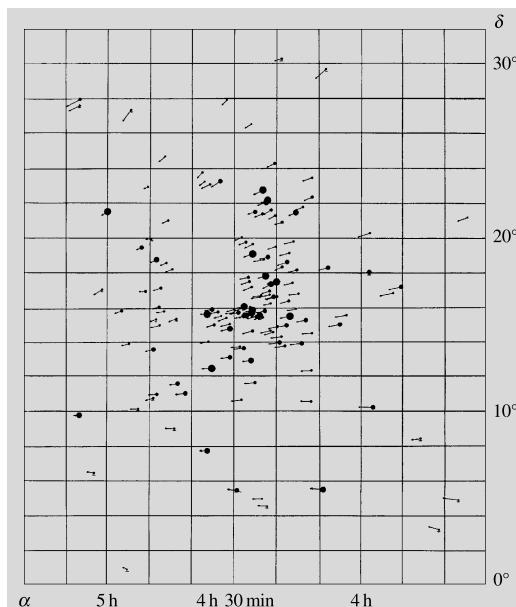


Fig. 17.4 Proper motions of the Hyades. The vectors show the movement of the stars in about 10,000 years. (van Bueren, H.G. (1952): Bull. Astr. Inst. Neth. **11**)

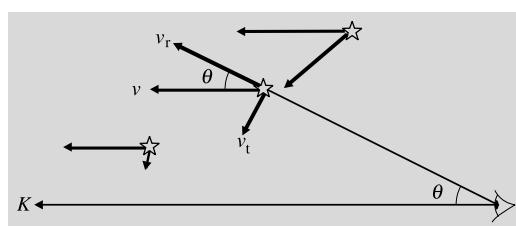


Fig. 17.5 If all stars move in the same direction, their tangential velocity components appear to be directed towards the convergence point K

the line of sight will then also be θ . The velocity components along the line of sight and at right angles to it, v_r and v_t , are therefore given by

$$\begin{aligned} v_r &= v \cos \theta, \\ v_t &= v \sin \theta. \end{aligned} \quad (17.1)$$

The radial velocity v_r can be measured from the Doppler shift of the stellar spectrum. The tangential velocity v_t is related to the proper motion μ and the distance r :

$$v_t = \mu r. \quad (17.2)$$

Thus the distance can be calculated:

$$r = \frac{v_t}{\mu} = \frac{v \sin \theta}{\mu} = \frac{v_r}{\mu} \tan \theta. \quad (17.3)$$

By means of this method, the distances of the individual stars can be determined from the motion of the cluster as a whole. Since the method of (ground-based) trigonometric parallaxes is reliable only out to a distance of 30 pc, the moving cluster method is an indispensable way of determining stellar distances. The distance of the Hyades obtained in this way is about 40 pc. The distance obtained from the trigonometric parallaxes measured directly by the Hipparcos satellite in the 1990's is 46 pc. The Hyades is the nearest open cluster.

The observed HR diagram or the corresponding colour-magnitude diagram of the Hyades and other nearby star clusters show a very well-defined and narrow main sequence (Fig. 17.6). Most of the cluster members are main sequence stars; there are only a few giants. There are quite a few stars slightly less than one magnitude above the main sequence. These are apparently binary stars whose components have not been resolved. To see this, let us consider a binary, where both components have the same magnitude m and the same colour index. If this system is unresolved, the colour index will still be the same, but the observed magnitude will be $m - 0.75$, i.e. slightly less than one magnitude brighter.

The main sequences of open clusters are generally located in the same section of the HR or colour-magnitude diagram (Fig. 17.7). This is because the material from which the clusters formed

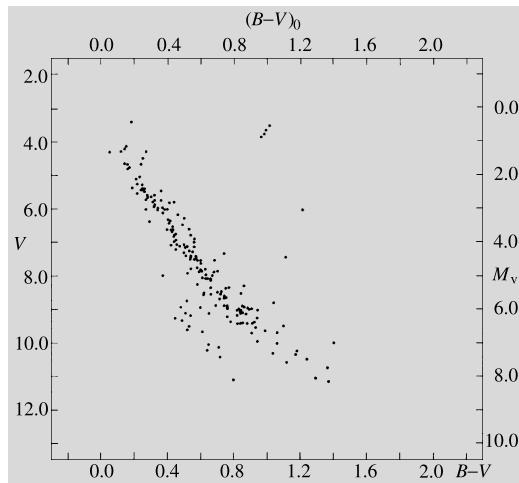


Fig. 17.6 Colour-magnitude diagram of the Hyades. Apparent visual magnitude on the left-hand vertical axis; absolute visual magnitude on the right-hand one

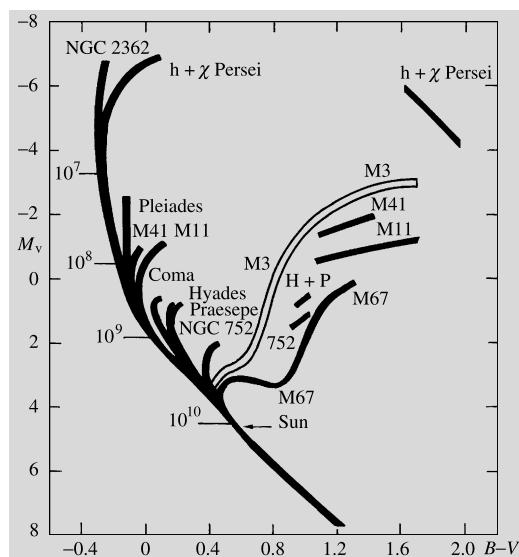


Fig. 17.7 Schematic colour-magnitude diagrams of star clusters. M3 is a globular cluster; the others are open clusters. Cluster ages are shown along the main sequence. The age of a cluster can be told from the point where its stars begin to turn off the main sequence. (Sandage, A. (1956): Publ. Astron. Soc. Pac. **68**, 498)

has not varied much, i.e. their initial chemical composition has been fairly constant. In younger clusters the main sequence extends to brighter and hotter stars and earlier spectral types. Usually one can clearly see the point in the diagram

where the main sequence ends and bends over towards the giant branch. This point will depend very strongly on the age of the cluster. It can therefore be used in determining the ages of open clusters. Star clusters are of central importance in the study of stellar evolution.

The colour-magnitude diagrams of star clusters can also be used to determine their distances. The method is called *main sequence fitting*. By means of multicolour photometry the reddening due to interstellar dust can be removed from the observed colours $B - V$ of the stars, yielding the intrinsic colours $(B - V)_0$. Most star clusters are so far away from us that all cluster members can be taken to be at the same distance. The distance modulus

$$m_{V_0} - M_V = 5 \lg \frac{r}{10 \text{ pc}} \quad (17.4)$$

will then be the same for all members. In (17.4), m_{V_0} is the apparent, M_V the absolute visual magnitude of a star, and r the distance. It has been assumed that the extinction due to interstellar dust A_V has been determined from multicolour photometry and its effect removed from the observed visual magnitude m_V :

$$m_{V_0} = m_V - A_V.$$

When the observed colour-magnitude diagram of the cluster is plotted using the apparent magnitude m_{V_0} rather than the absolute magnitude M_V on the vertical axis, the only change will be that the position of the main sequence is shifted vertically by an amount corresponding to the distance modulus. The observed $(m_{V_0}, (B - V)_0)$ diagram may now be compared with the Hyades $(M_V, (B - V)_0)$ diagram used as a standard. By demanding that the main sequences of the two diagrams agree, the distance modulus and hence the distance can be determined. The method is very accurate and efficient. It can be used to determine cluster distances out to many kiloparsecs.

17.3 Globular Star Clusters

Globular star clusters usually contain about 10^5 stars. The distribution of the stars is spherically

symmetric, and the central densities are about ten times larger than in open clusters. Stars in globular clusters are among the oldest in the Milky Way, and therefore they are of great importance for studies of stellar evolution. There are about 150–200 globular clusters in the Milky Way.

The colour-magnitude diagram of a typical globular cluster is shown in Fig. 17.8. The main sequence only contains faint red stars; there is a prominent giant branch, and the horizontal and asymptotic branches are clearly seen. The main sequence is lower than that of the open clusters, because the metal abundance is much lower in the globular clusters.

The horizontal branch stars have a known absolute magnitude, which has been calibrated using principally RR Lyrae type variables. Because the horizontal branch stars are bright, they can be observed even in distant clusters, and thus using them the distances of globular clusters can be well determined.

Using the known distances, the linear sizes of globular clusters can be calculated. It is found that most of the mass is concentrated to a central core with a radius of about 0.3–10 pc. Outside this there is an extended envelope with a radius that may be 10–100 times larger. At even larger radii stars will escape from the cluster because of the tidal force of the Galaxy.

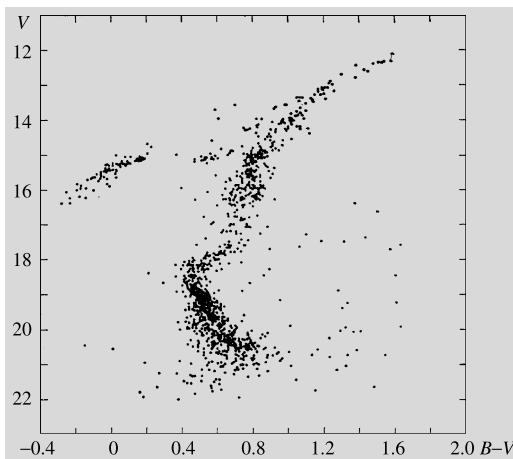


Fig. 17.8 Colour-magnitude diagram of the globular cluster M5. In addition to the main sequence one can see the giant branch bending to the right and to its left the horizontal branch. (Arp, H. (1962): *Astrophys. J.* **135**, 311)

The masses of globular clusters can be roughly estimated from the virial theorem, if the stellar velocities in the cluster have been measured. More precise values are calculated by fitting theoretical models to the observed density and velocity distributions. In this way masses in the range 10^4 – $10^6 M_{\odot}$ have been obtained.

The globular clusters in the Milky Way fall into two classes. In the classification given in Table 18.1 these correspond to intermediate and halo population II. The disk globular clusters are concentrated towards the centre and the plane of the Milky Way and they form a system that is rotating with the general rotation of the Milky Way. In contrast, the halo clusters are almost spherically distributed in an extensive distribution reaching out to at least 35 kpc. The system of halo clusters does not rotate, but instead the velocities of individual clusters are uniformly distributed in all directions. The abundance of heavy elements is also different in the two classes of clusters. For disk clusters it is typically about 30 % of the solar value, for halo clusters it is only about 1 %. In some clusters the heavy element abundances are only 10^{-3} times the solar value. The smallest values, as low as 10^{-4} – 10^{-5} have been detected in some field stars of the halo. They therefore give important information about the production of elements in the early Universe and during the formation of the Milky Way.

All globular clusters are old, and the halo clusters are among the oldest known astronomical objects. Determining a precise age is difficult, and requires both accurate observations of the turn-off point of the main sequence in the HR diagram, as well detailed theoretical stellar evolution models. The ages obtained have been about 13×10^9 years. This age is close to the age of the Universe calculated from its rate of expansion (see Chap. 20).

17.4 Example

Example 17.1 Assume that a globular cluster has a diameter of 40 pc and contains 100,000 stars of one solar mass each.

- (a) Use the virial theorem to find the average velocity of the stars. You can assume that the

- average distance between stars equals the radius of the cluster.
- (b) Find the escape velocity.
- (c) Comparing these velocities, can you tell something about the stability of the cluster?
- (a) First, we have to estimate the potential energy. There are $n(n - 1)/2 \approx n^2/2$ pairs of stars in the cluster, and the average distance of each pair is R . Thus the potential energy is about

$$U = -G \frac{m^2 n^2}{R^2},$$

where $m = 1 M_{\odot}$. The kinetic energy is

$$T = \frac{1}{2} m v^2 n,$$

where v is the root mean square velocity. According to the virial theorem we have $T = -1/2U$, whence

$$\frac{1}{2} m v^2 n = \frac{1}{2} G \frac{m^2 n^2}{R^2}.$$

Solving for the velocity we get

$$\begin{aligned} v^2 &= \frac{Gmn}{2R} \\ &= \frac{6.7 \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2} \times 2.0 \times 10^{30} \text{ kg} \times 10^5}{40 \times 3.1 \times 10^{16} \text{ m}} \\ &= 1.1 \times 10^7 \text{ m}^2 \text{ s}^{-2}, \end{aligned}$$

which gives $v \approx 3 \text{ km s}^{-1}$.

- (b) The escape velocity from the edge of the cluster is

$$\begin{aligned} v_e &= \sqrt{\frac{2Gmn}{R}} \\ &= \sqrt{4v^2} = 2v = 6 \text{ km s}^{-1}. \end{aligned}$$

- (c) No. The average velocity seems to be smaller than the escape velocity, but it was derived from the virial theorem assuming that the cluster is stable.

17.5 Exercises

Exercise 17.1 A globular cluster consists of 100,000 stars of the solar absolute magnitude. Calculate the total apparent magnitude of the cluster, if its distance is 10 kpc.

Exercise 17.2 If the apparent magnitude of the cluster of the previous exercise is 10, what is its distance if the interstellar absorption in the direction of the cluster is 1.5 mag/kpc?

Exercise 17.3 The Pleiades open cluster contains 230 stars within 4 pc. Estimate the velocities of the stars in the cluster using the virial theorem. For simplicity, let the mass of each star be replaced by $1 M_{\odot}$.

On clear, moonless nights a nebulous band of light can be seen stretching across the sky. This is the Milky Way (Fig. 18.1). The name is used both for the phenomenon in the sky and for the large stellar system causing it. The Milky Way system is also called the Galaxy—with a capital letter. The general term galaxy is used to refer to the countless stellar systems more or less like our Milky Way.

The band of the Milky Way extends round the whole celestial sphere. It is a huge system consisting mostly of stars, among them the Sun. The stars of the Milky Way form a flattened disk-like system. In the direction of the plane of the disk huge numbers of stars are visible, whereas relatively few are seen in the perpendicular direction. The faint light of distant stars merges into a uniform glow, and therefore the Milky Way appears as a nebulous band to the naked eye. A long-exposure photograph reveals hundreds of thousands of stars (Fig. 18.2).

In the early 17th century *Galileo Galilei*, using his first telescope, discovered that the Milky Way consists of innumerable stars. Already in the late 18th century *William Herschel* attempted to determine the size and shape of the Milky Way by means of star counts. Only early in the 20th century did the Dutch astronomer *Jacobus Kapteyn* obtain the first estimate of the size of the Milky Way. The true size of the Milky Way and the Sun's position in it became clear finally in the 1920's from *Harlow Shapley*'s studies of the space distribution of globular clusters.

In studying the structure of the Milky Way, it is convenient to choose a spherical coordinate system so that the fundamental plane is the symmetry plane of the Milky Way. This is defined as the symmetry plane of the distribution of neutral hydrogen, and it agrees quite closely with the symmetry plane defined by the distribution of stars in the solar neighbourhood (within a few kpc).

The basic direction in the fundamental plane has been chosen to be the direction of the centre of the Milky Way. This is located in the constellation Sagittarius ($\alpha = 17^{\text{h}} 45.7 \text{ min}$, $\delta = -29^{\circ} 00'$, epoch 2000.0) at a distance of about 8.5 kpc. The galactic latitude is counted from the plane of the Galaxy to its pole, going from 0° to $+90^{\circ}$, and to the galactic south pole, from 0° to -90° . The galactic coordinate system is shown in Fig. 18.3 (see also Sect. 2.8).

18.1 Methods of Distance Measurement

In order to study the structure of the Milky Way, one needs to know how various kinds of objects, such as stars, star clusters and interstellar matter, are distributed in space. The most important ways of measuring the distances will first be considered.

Trigonometric Parallaxes The method of *trigonometric parallaxes* is based on the apparent yearly back-and-forth movement of stars in the sky, caused by the orbital motion of the Earth.

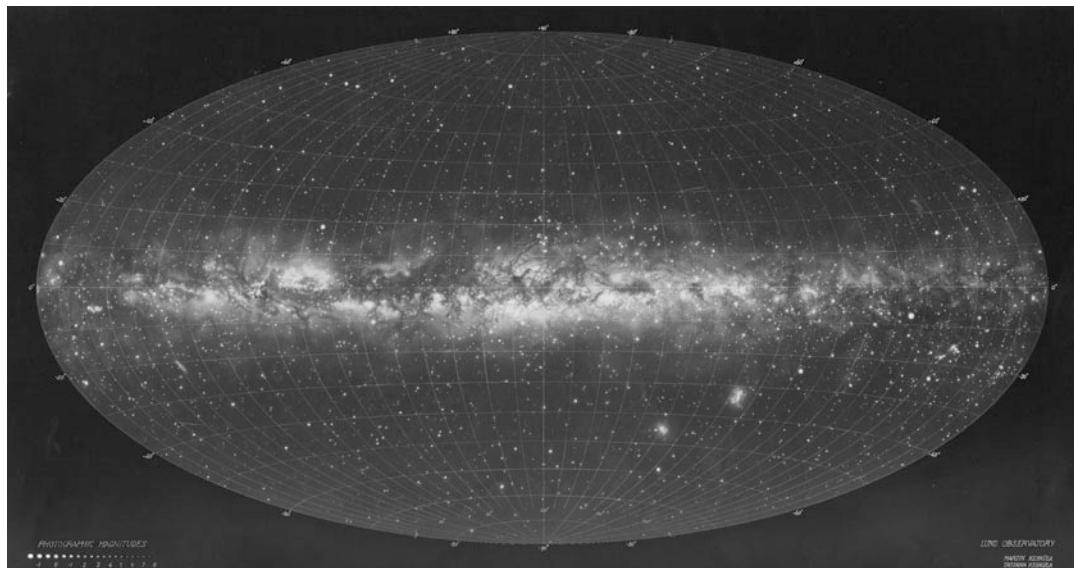


Fig. 18.1 The nebulous band of the Milky Way stretches across the entire sky. (Photograph M. and T. Kesäkula, Lund Observatory)

From Earth-based observations the trigonometric parallaxes can be reliably measured out to a distance of about 30 pc; beyond 100 pc this method is no longer useful. The situation is, however, changing. The limit has already been pushed to a few hundred parsecs by the Hipparcos satellite, and Gaia will mean another major leap in the accuracy.

The Motion of the Sun with Respect to the Neighbouring Stars. The Local Standard of Rest

The motion of the Sun with respect to the neighbouring stars is reflected in their proper motions and radial velocities (Fig. 18.4). The point towards which the Sun's motion among the stars seems to be directed is called the *apex*. The opposite point is the *antapex*. The stars near the apex appear to be approaching; their radial velocities have the largest negative values, on the average. In the direction of the antapex the largest (positive) radial velocities are observed. On the great circle perpendicular to the apex-antapex direction, the radial velocities are zero on the average, but the proper motions are large. The average proper motions decrease towards the apex and the antapex, but always point from the apex towards the antapex.

In order to study the true motions of the stars, one has to define a coordinate system with respect to which the motions are to be defined. The most practical frame of reference is defined so that the stars in the solar neighbourhood are at rest, on the average. More precisely, this *local standard of rest* (LSR) is defined as follows.

Let us suppose the velocities of the stars being considered are distributed at random. Their velocities with respect to the Sun, i.e. their radial velocities, proper motions and distances, are assumed to be known. The local standard of rest is then defined so that the mean value of the velocity vectors is opposite to the velocity of the Sun with respect to the LSR. Clearly the mean velocity of the relevant stars with respect to the LSR will then be zero. The motion of the Sun with respect to the LSR is found to be:

Apex coordinates	$\alpha = 18 \text{ h } 00 \text{ min} = 270^\circ$	$l = 56^\circ$
	$\delta = +30^\circ$	$b = +23^\circ$
Solar velocity	$v_0 = 19.7 \text{ km s}^{-1}$	

The apex is located in the constellation of Hercules. When the sample of stars used to determine the LSR is restricted to a subset of all the

Fig. 18.2 A section of about 40° of the Milky Way between the constellations of Cygnus and Aquila. The brightest star at the upper right is Vega (α Lyrae). (Photograph Palomar Observatory)



stars in the solar neighbourhood, e.g. to stars of a given spectral class, the sample will usually have slightly different kinematic properties, and the coordinates of the solar apex will change correspondingly.

The velocity of an individual star with respect to the local standard of rest is called the *peculiar motion* of the star. The peculiar velocity of a star is obtained by adding the velocity of the Sun with respect to the LSR to the measured velocity. Naturally the velocities should be treated as vectors.

The local standard of rest is at rest only with respect to a close neighbourhood of the Sun. The Sun and the nearby stars, and thus also the LSR,

are moving round the centre of the Milky Way at a speed that is ten times greater than the typical peculiar velocities of stars in the solar neighbourhood (Fig. 18.5).

Statistical Parallaxes The velocity of the Sun with respect to neighbouring stars is about 20 km s^{-1} . This means that in one year, the Sun moves about 4 au with respect to the stars.

Let us consider a star S (Fig. 18.6), whose angular distance from the apex is ϑ and which is at a distance r from the Sun. In a time interval t the star will move away from the apex at the angular velocity $u/t = \mu_A$ because of the solar motion. In the same time interval, the Sun will move the

distance s . The sine theorem for triangles yields

$$r \approx r' = \frac{s \sin \vartheta}{\sin u} \approx \frac{s \sin \vartheta}{u}, \quad (18.1)$$

because the distance remains nearly unchanged and the angle u is very small. In addition to the component μ_A due to solar motion, the observed proper motion has a component due to the peculiar velocity of the star. This can be removed

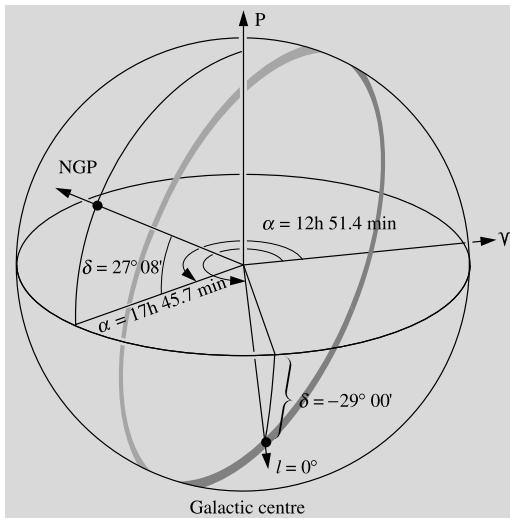


Fig. 18.3 The directions to the galactic centre and the North galactic pole (NGP) in equatorial coordinates. The galactic longitude l is measured from the galactic centre along the galactic plane. The coordinates of the Galactic centre are precessed from the defining equinox 1950 and are not very accurate (see A.P. Lane (1979), PASP, 91, 405)

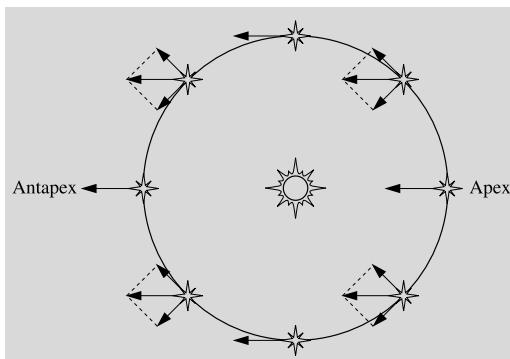


Fig. 18.4 Because of the motion of the Sun towards the apex, the average radial velocity of the nearby stars appears largest in the apex and antapex directions

by taking an average of (18.1) for a sample of stars, since the peculiar velocities of the stars in the solar neighbourhood can be assumed to be randomly distributed. By observing the average proper motion of objects known to be at the same distance one thus obtains their actual distance. A similar statistical method can be applied to radial velocities.

Objects that are at the same distance can be found as follows. We know that the distance modulus $m - M$ and the distance r are related according to:

$$m - M = 5 \lg(r/10 \text{ pc}) + A(r), \quad (18.2)$$

where A is the interstellar extinction. Thus objects that have the same apparent and the same absolute magnitude will be at the same distance. It should be noted that we need not know the absolute magnitude as long as it is the same for all stars in the sample. Suitable classes of stars are

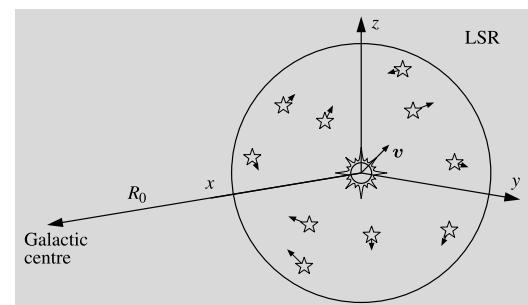


Fig. 18.5 The local standard of rest, defined by the stars in the solar neighbourhood, moves with respect to the galactic centre. However, the average value of the stellar peculiar velocities with respect to the LSR is zero

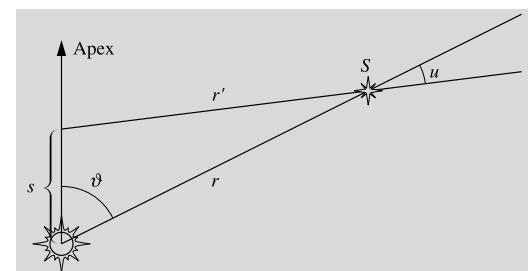


Fig. 18.6 When the Sun has moved the distance s towards the apex, the direction to the star S appears to have changed by the angle u

main sequence A4 stars, RR Lyrae variables and classical cepheids with some given period. The stars in a cluster are also all at the same distance. This method has been used, for example, to determine the distance to the Hyades as explained in Sect. 17.2.

Parallaxes based on the peculiar or apex motion of the Sun are called statistical or secular parallaxes.

Main Sequence Fitting If the distance of a cluster is known, it is possible to plot its HR diagram with the absolute magnitude as the vertical coordinate. Another cluster, whose distance is to be determined, can then be plotted in the same diagram using the apparent magnitudes as the vertical coordinate. Now the vertical distance of the main sequences tells how much the apparent magnitudes differ from the absolute ones. Thus the distance modulus $m - M$ can be measured. This method, known as the *main sequence fitting*, works for clusters whose stars are roughly at the same distance; if the distances vary too much, a clear main sequence cannot be distinguished.

Photometric Parallaxes The determination of the distance directly from (18.2) is called the photometric method of distance determination and the corresponding parallax, the *photometric parallax*. The most difficult task when using this method usually involves finding the absolute magnitude; there are many ways of doing this. For example, the two-dimensional MKK spectral classification allows one to determine the absolute magnitude from the spectrum. The absolute magnitudes of cepheids can be obtained from their periods. A specially useful method for star clusters is the procedure of main sequence fitting. A condition for the photometric method is that the absolute magnitude scale first be calibrated by some other method.

Trigonometric parallaxes do not reach very far. For example, even with the Hipparcos satellite, only a few cepheid distances have been accurately measured by this method. The method of statistical parallaxes is indispensable for calibrating the absolute magnitudes of bright objects. When this has been done, the photometric

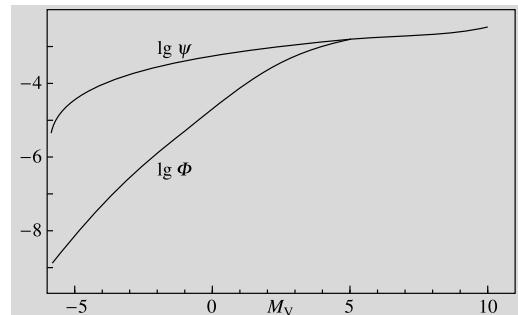


Fig. 18.7 The observed luminosity function $\Phi(M_V)$ and the initial luminosity function $\Psi(M_V)$ for main sequence stars in the solar neighbourhood. The functions give the number of stars per cubic parsec in the magnitude interval $[M_V - 1/2, M_V + 1/2]$; they are actually the products $D\Phi$ and $D\Psi$, where D is the stellar density function (in the solar neighbourhood)

method can be used to obtain distances of objects even further away.

Other examples of indicators of brightness, luminosity criteria, are characteristic spectral lines or the periods of cepheids. Again, their use requires that they first be calibrated by means of some other method. It is a characteristic feature of astronomical distance determinations that the measurement of large distances is based on knowledge of the distances to nearer objects.

18.2 Stellar Statistics

The Stellar Luminosity Function By systematically observing all stars in the solar neighbourhood, one can find the distribution of their absolute magnitudes. This is given by the luminosity function $\Phi(M)$ (Fig. 18.7), which gives the relative number of main sequence stars with absolute magnitudes in the range $[M - 1/2, M + 1/2]$. No stars appear to be forming at present in the region of space where the luminosity function has been determined. The age of the Milky Way is 10–15 Ga, which means that all stars less massive than $0.9 M_\odot$, will still be on the main sequence. On the other hand, more massive stars, formed early in the history of the Milky Way, will have completed their evolution and disappeared. Low-mass stars have accumulated in the luminosity function for many generations of star formation,

whereas bright, high-mass stars are the result of recent star formation.

By taking into account the different main sequence lifetimes of stars of different masses and hence of different magnitudes, one can determine the initial luminosity function $\Psi(M)$, which gives the brightness distribution at the time of star formation, the zero age main sequence luminosity function. The relation between the function Ψ and the observed luminosity function is

$$\Psi(M) = \Phi(M)T_0/t_{\text{E}}(M), \quad (18.3)$$

where T_0 is the age of the Milky Way and $t_{\text{E}}(M)$ is the main sequence lifetime of stars of magnitude M . Here we assume that the birth rate of stars of magnitude M has remained constant during the lifetime of the Milky Way. The initial luminosity function is shown in Fig. 18.7.

The Fundamental Equation of Stellar Statistics. The Stellar Density A crucial problem for studies of the structure of the Milky Way is to find out how the density of stars varies in space. The number of stars per unit volume at a distance r in the direction (l, b) from the Sun is given by the stellar density $D = D(r, l, b)$.

The stellar density cannot be directly observed except in the immediate neighbourhood of the Sun. However, it can be calculated if one knows the luminosity function and the interstellar extinction as a function of distance in a given direction. In addition the number of stars per unit solid angle (e.g. per square arc second) can be determined as a function of limiting apparent magnitude by means of star counts (Fig. 18.8).

Let us consider the stars within the solid angle ω in the direction (l, b) and in the distance range $[r, r + dr]$ (Fig. 18.9). We let their luminosity function $\Phi(M)$ be the same as in the solar neighbourhood and their unknown stellar density D . The absolute magnitude M of the stars of apparent magnitude m is, as usual,

$$M = m - 5 \lg(r/10 \text{ pc}) - A(r).$$

The number of stars in the apparent magnitude interval $[m - 0.5, m + 0.5]$ in the volume element

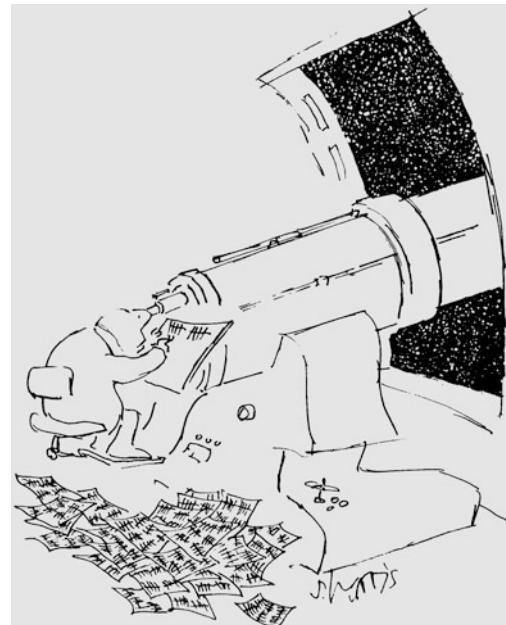


Fig. 18.8 The stellar density is determined by means of star counts. In practice, the counting is done on photographic plates. (Cartoon S. Harris)

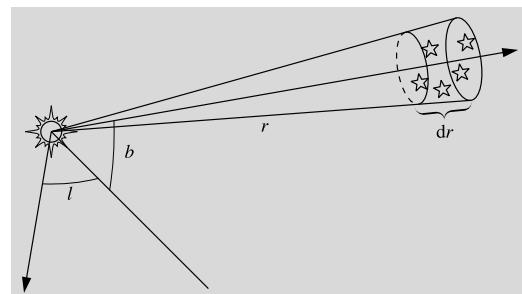


Fig. 18.9 The size of the volume element at distance r in the direction (l, b) is $\omega r^2 dr$

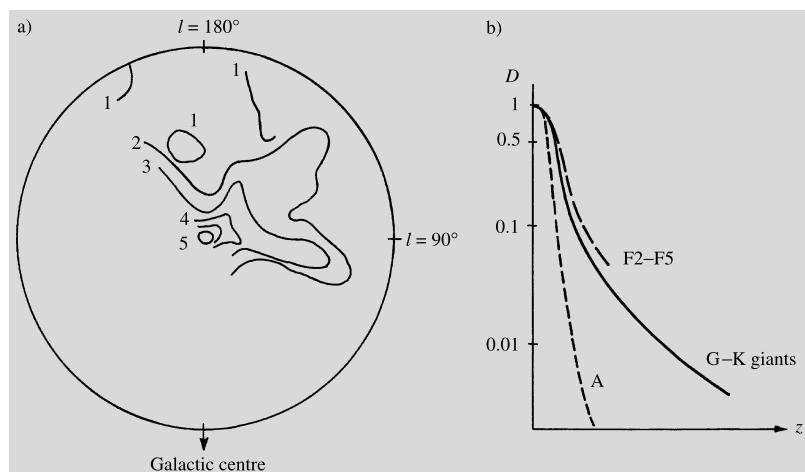
$$dV = \omega r^2 dr \text{ at distance } r \text{ is (Fig. 18.9)}$$

$$dN(m) = D(r, l, b)$$

$$\times \Phi \left[m - 5 \lg \frac{r}{10 \text{ pc}} - A(r) \right] dV. \quad (18.4)$$

The stars of apparent magnitude m in the given area of the sky will in reality be at many different distances. In order to obtain their total number $N(m)$, one has to integrate $dN(m)$ over all

Fig. 18.10 The stellar density near the Sun.
(a) The stellar density of spectral classes A2–A5 in the galactic plane, according to S.W. McCuskey. The numbers next to the isodensity curves give the number of stars in $10,000 \text{ pc}^3$. **(b)** The distribution of different spectral classes perpendicularly to the galactic plane according to T. Elvius. The density in the galactic plane has been normalised to one



distances r :

$$N(m) = \int_0^\infty D(r, l, b) \times \Phi \left[m - 5 \lg \frac{r}{10 \text{ pc}} - A(r) \right] \omega r^2 dr. \quad (18.5)$$

Equation (18.5) is called the *fundamental equation of stellar statistics*. Its left-hand side, the number of stars in the apparent magnitude interval $[m - 0.5, m + 0.5]$ in the solid angle ω , is obtained from the observations: one counts the stars of different magnitudes in a chosen area of a photographic plate. The luminosity function is known from the solar neighbourhood. The extinction $A(r)$ can be determined for the chosen areas, for instance, by means of multicolour photometry. In order to solve the integral equation (18.5) for $D(r, l, b)$, several methods have been developed, but we shall not go into them here.

Figure 18.10(a) shows the stellar density in the solar neighbourhood in the plane of the Milky Way, and Fig. 18.10(b) in the direction perpendicular to the plane. There are several individual concentrations, but e.g. spiral structure cannot be observed in such a limited region of space.

The Distribution of Bright Objects Using stellar statistical methods, one can only study the close neighbourhood of the Sun, out to about 1 kpc at the most. Absolutely faint objects can-

not be observed very far. Since the solar neighbourhood appears to be fairly representative of the general properties of the Milky Way, its study is naturally important, giving information e.g. on the distributions and luminosity functions of stars of various spectral types. However, in order to get an idea of the larger-scale structure of the Milky Way, one has to make use of objects that are as absolutely bright as possible, and which can be observed even at large distances.

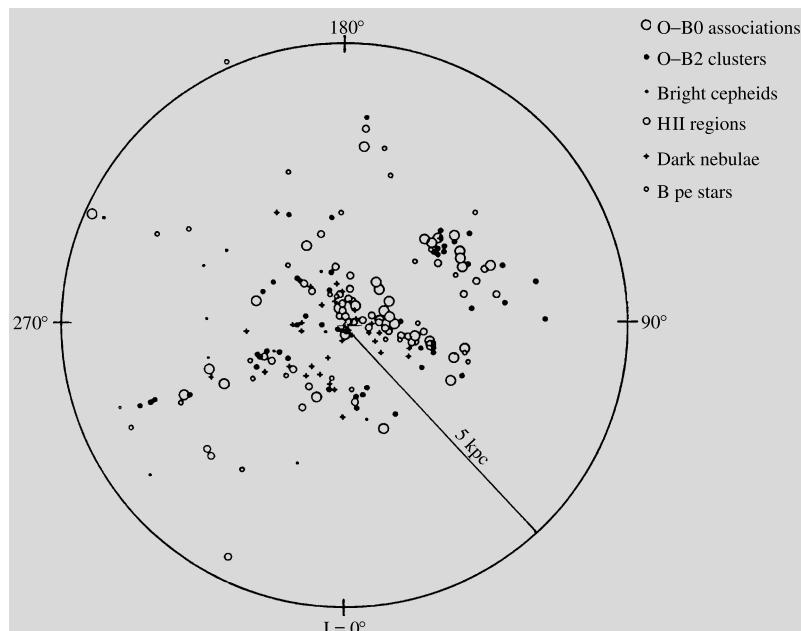
Examples of suitable objects are stars of early spectral types, HII regions, OB associations, open star clusters, cepheids, RR Lyrae stars, supergiants and giants of late spectral types, and globular clusters. Some of these objects differ greatly in age, such as the young OB associations, on the one hand, and the old globular clusters, on the other. Any differences in their space distribution tell us about changes in the general structure of the Milky Way.

The young optical objects, the HII regions, OB associations and open clusters, are strongly concentrated in the plane of the Milky Way (Table 18.1). Figure 18.11 shows that they also appear to be concentrated in three drawn-out bands, at least within the observed region. Since these types of objects in other galaxies are known to be part of a spiral structure, the observed bands in the Milky Way have been interpreted as portions of three spiral arms passing through the solar neighbourhood. Stars of later spectral types seem to be much more evenly distributed. Apart

Table 18.1 Populations of the Milky Way; z is the vertical distance from the galactic plane, and v_r the velocity component perpendicular to the galactic plane

Population	Typical objects	Average age [10^9 a]	z [pc]	v_r [k/s]	Metal abundance
Halo population II	Subdwarfs, globular clusters RR Lyr ($P > 0.4$ d)	14–12	2000	75	0.001
Intermediate population II	Long period variables	12–10	700	25	0.005
Disc population	Planetary nebulae, novae bright red giants	12–2	400	18	0.01–0.02
Old population I	A stars, Me dwarfs classical cepheids	2–0.1	160	10	0.02
Young population I	Gas, dust, supergiants, T Tau stars	0.1	120		0.03–0.04

Fig. 18.11 The distribution of various objects in the galactic plane. Three condensations can be discerned: the Sagittarius arm (*lowest*), the local arm near the Sun and (*outermost*) the Perseus arm



from a few special directions, interstellar dust limits observations in the galactic plane to within 3–4 kpc.

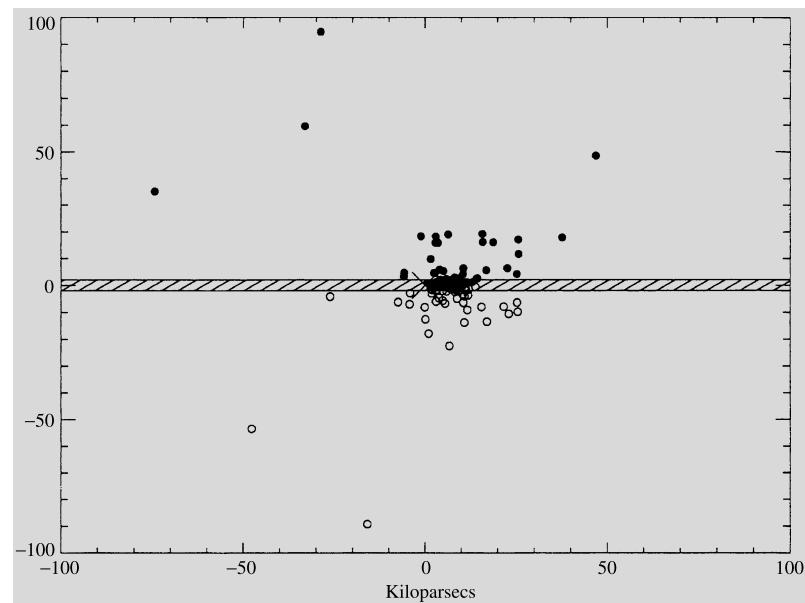
Old objects, particularly the globular clusters, have an almost spherical distribution about the centre of the Milky Way (Fig. 18.12). The space density of old objects increases towards the galactic centre. They can be used to determine the distance of the Sun from the galactic centre; the value of this distance is about 8.5 kpc. More recent measurements using other methods have given a distance of 8 kpc (26,000 light years).

Stellar Populations Studies of the motions of the stars in the Milky Way have revealed that the

orbits of stars moving in the galactic plane are almost circular. These stars are also usually young, a few hundred million years at the most. They also contain a relatively large amount of heavy elements, about 2–4 %. The interstellar material similarly moves in the galactic plane in almost circular orbits. On the basis of their motions and their chemical composition, the interstellar medium and the youngest stars are collectively referred to as *population I*.

Outside the plane of the Milky Way, an almost spherically symmetric *halo* extends out to over 50 kpc and even further out there is a *corona*. The stellar density is largest near the galactic centre and decreases outwards. The halo contains very

Fig. 18.12 The distribution of globular clusters. (From S.R. Majewski, *Stellar populations and the Milky Way*, in C. Martínez Roger, I. Pérez Fournón, F. Sánchez (Eds.) *Globular Clusters*, Cambridge University Press, 1999)



little interstellar matter, and its stars are old, perhaps up to 13×10^9 years. These stars are also very metal-poor. Their orbits may be very eccentric and show no preference for the galactic plane. On the basis of these criteria, one defines stars of *population II*. Typical population II objects are the globular clusters, and the RR Lyrae and W Virginis stars.

While the Sun and the closest stars move much like the stars of population I, the stars of population II have large velocities with respect to the local standard of rest, up to more than 300 km s^{-1} . In reality their velocities at the solar distance from the galactic centre are quite small and may sometimes be opposite to the direction of motion of the LSR. The large relative velocities only reflect the motion of the LSR with a velocity of about 220 km s^{-1} round the galactic centre.

Between these two extremes, there is a sequence of intermediate populations. In addition to populations I and II, one generally also speaks of a disk population, including the Sun, for instance. The typical motions, chemical composition and age of the various populations (Table 18.1) contain information about the evolution of our Galaxy and about the formation of its stars.

18.3 The Rotation of the Milky Way

Differential Rotation. Oort's Formulas The flatness of the Milky Way is already suggestive of a general rotation about an axis normal to the galactic plane. Observations of the motions both of stars and of interstellar gas have confirmed this rotation and shown it to be differential. This means that the angular velocity of rotation depends on the distance from the galactic centre (Fig. 18.13). Thus the Milky Way does not rotate like a rigid body. Near the Sun, the rotational velocity decreases with radius.

The observable effects of the galactic rotation were derived by the Dutch astronomer *Jan H. Oort*. Let us suppose the stars are moving in circular orbits about the galactic centre (Fig. 18.14). This approximation is acceptable for population I stars and gas. The star S , seen from the Sun \odot at galactic longitude l at distance r , has circular velocity V at a distance R from the centre. Similarly for the Sun the galactic radius and velocity are R_0 and V_0 . The relative radial velocity v_r of the star with respect to the Sun is the difference between the projections of the circular velocities on the line of sight:

$$v_r = V \cos \alpha - V_0 \sin l, \quad (18.6)$$

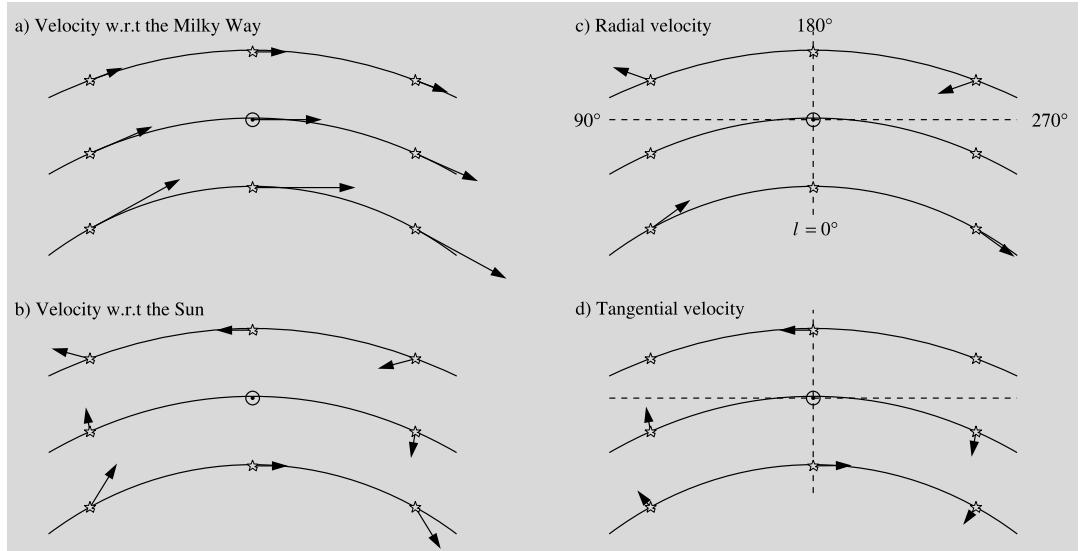


Fig. 18.13 The effect of differential rotation on the radial velocities and proper motions of stars. (a) Near the Sun the orbital velocities of stars decrease outwards in the Galaxy. (b) The relative velocity with respect to the Sun is obtained by subtracting the solar velocity from the ve-

locity vectors in (a). (c) The radial components of the velocities with respect to the Sun. This component vanishes for stars on the same orbit as the Sun. (d) The tangential components of the velocities

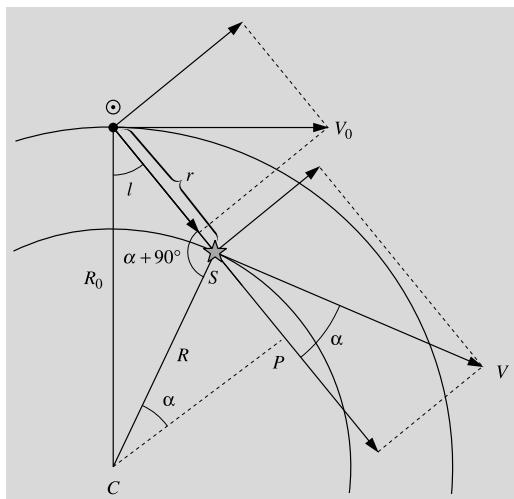


Fig. 18.14 In order to derive Oort's formulas, the velocity vectors of the Sun and the star S are divided into components along the line $\odot S$ and normal to it

where α is the angle between the velocity vector of the star and the line of sight. From Fig. 18.14 the angle $CS\odot = \alpha + 90^\circ$. By applying the sine

theorem to the triangle $CS\odot$ one obtains

$$\frac{\sin(\alpha + 90^\circ)}{\sin l} = \frac{R_0}{R}$$

or

$$\cos \alpha = \frac{R_0}{R} \sin l. \quad (18.7)$$

Denoting the angular velocity of the star by $\omega = V/R$ and that of the Sun by $\omega_0 = V_0/R_0$, one obtains the observable radial velocity in the form

$$v_r = R_0(\omega - \omega_0) \sin l. \quad (18.8)$$

The tangential component of the relative velocity of the Sun and the star is obtained as follows. From Fig. 18.14,

$$v_t = V \sin \alpha - V_0 \cos l = R \omega \sin \alpha - R_0 \omega_0 \cos l.$$

The triangle $\odot CP$ gives

$$R \sin \alpha = R_0 \cos l - r,$$

and hence

$$v_t = R_0(\omega - \omega_0) \cos l - \omega r. \quad (18.9)$$

Oort noted that in the close neighbourhood of the Sun ($r \ll R_0$), the difference of the angular velocities will be very small. Therefore a good approximation for the exact equations (18.8) and (18.9) is obtained by keeping only the first term of the Taylor series of $\omega - \omega_0$ in the neighbourhood of $R = R_0$:

$$\omega - \omega_0 = \left(\frac{d\omega}{dR} \right)_{R=R_0} (R - R_0) + \dots$$

Using $\omega = V/R$ and $V(R_0) = V_0$, one finds

$$\omega - \omega_0 \approx \frac{1}{R_0^2} \left[R_0 \left(\frac{dV}{dR} \right)_{R=R_0} - V_0 \right] (R - R_0).$$

For $R \approx R_0 \gg r$, the difference $R - R_0 \approx -r \cos l$. One thus obtains an approximate form

$$v_r \approx \left[\frac{V_0}{R_0} - \left(\frac{dV}{dR} \right)_{R=R_0} \right] r \cos l \sin l$$

or

$$v_r \approx Ar \sin 2l, \quad (18.10)$$

where A is a characteristic parameter of the solar neighbourhood of the Galaxy, the *first Oort constant*:

$$A = \frac{1}{2} \left[\frac{V_0}{R_0} - \left(\frac{dV}{dR} \right)_{R=R_0} \right]. \quad (18.11)$$

For the tangential relative velocity, one similarly obtains, since $\omega r \approx \omega_0 r$:

$$v_t \approx \left[\frac{V_0}{R_0} - \left(\frac{dV}{dR} \right)_{R=R_0} \right] r \cos^2 l - \omega_0 r.$$

Because $2 \cos^2 l = 1 + \cos 2l$, this may be written

$$v_t \approx Ar \cos 2l + Br, \quad (18.12)$$

where A is the same as before and B , the *second Oort constant*, is

$$B = -\frac{1}{2} \left[\frac{V_0}{R_0} + \left(\frac{dV}{dR} \right)_{R=R_0} \right]. \quad (18.13)$$

The proper motion $\mu = v_t/r$ is then given by the expression

$$\mu \approx A \cos 2l + B. \quad (18.14)$$

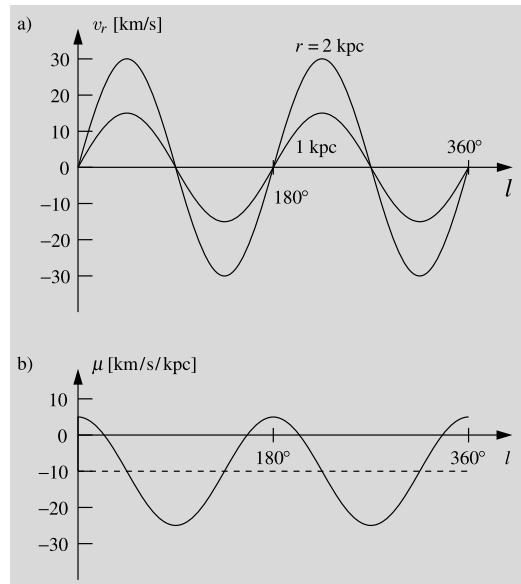


Fig. 18.15 The velocity components due to differential rotation according to Oort's formulas as functions of galactic longitude. (a) Radial velocities for objects at a distance of 1 and 2 kpc. (Compare with Fig. 18.13.) Strictly, the longitude at which the radial velocity vanishes depends on the distance. Oort's formulas are valid only in the close vicinity of the Sun. (b) Proper motions

Equation (18.10) says that the observed radial velocities of stars at the same distance should be a double sine curve as a function of galactic longitude. This has been confirmed by observations (Fig. 18.15(a)). If the distance to the stars involved is known, the amplitude of the curve determines the value of the Oort constant A .

Independently of distance, the proper motions of the stars form a double sine wave as a function of galactic longitude, as seen in Fig. 18.15(b). The amplitude of the curve is A and its mean value, B .

In 1927 on the basis of this kind of analysis, Oort established that the observed motions of the stars indicated a differential rotation of the Milky Way. Taking into account an extensive set of observational data, the International Astronomical Union IAU has confirmed the present recommended values for the Oort constants:

$$A = 15 \text{ km s}^{-1} \text{ kpc}^{-1}, \quad B = -10 \text{ km s}^{-1} \text{ kpc}^{-1}.$$

The Oort constants obey some interesting relations. By subtracting (18.13) from (18.11), one

obtains

$$A - B = \frac{V_0}{R_0} = \omega_0. \quad (18.15)$$

Adding (18.13) and (18.11) gives

$$A + B = -\left(\frac{dV}{dR}\right)_{R=R_0}. \quad (18.16)$$

Knowing the values of A and B , one can calculate the angular velocity $\omega_0 = 0.0053''/\text{year}$, which is the angular velocity of the local standard of rest around the galactic centre.

The circular velocity of the Sun and the LSR can be measured in an independent way by using extragalactic objects as a reference. In this way a value of about 220 km s^{-1} has been obtained for V_0 . Using (18.15) one can now calculate the distance of the galactic centre R_0 . The result is about 8.5 kpc , in good agreement with the distance to the centre of the globular cluster system. The direction to the galactic centre obtained from the distribution of radial velocities and proper motions by means of (18.10) and (18.14) also agrees with other measurements.

The orbital period of the Sun in the Galaxy according to these results is about $2.5 \times 10^8 \text{ years}$. Since the Sun's age is nearly $5 \times 10^9 \text{ years}$, it has made about 20 revolutions around the galactic centre. At the end of the previous revolution, the Carboniferous period had ended on Earth and the first mammals would soon appear.

The Distribution of Interstellar Matter Radio emission from interstellar gas, in particular that of neutral hydrogen, is not strongly absorbed or scattered by interstellar dust. It can therefore be used to map the structure of the Milky Way on large scales. Radio signals can be detected even from the opposite edge of the Milky Way.

The position of a radio source, for example an HI cloud, in the Galaxy cannot be directly determined. However, an indirect method exists, based on the differential rotation of the Galaxy.

Figure 18.16 is a schematic view of a situation in which gas clouds on the circles P_1, P_2, \dots are observed in the direction l ($-90^\circ < l < 90^\circ$). The angular velocity increases inwards, and therefore the greatest angular velocity along the line of

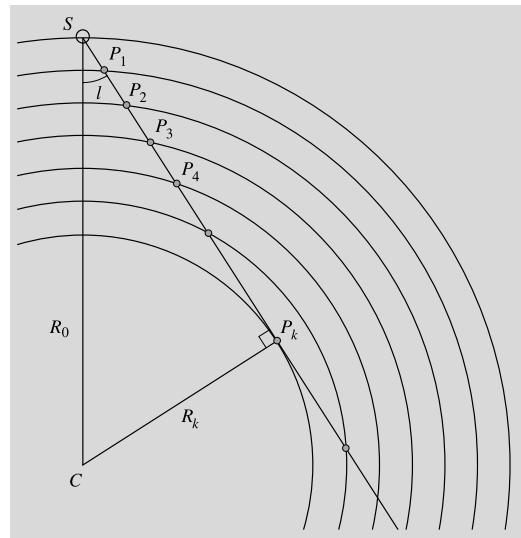


Fig. 18.16 Clouds P_1, P_2, \dots seen in the same direction at various distances

sight is obtained at the point P_k , where the line of sight is tangent to a circle. This means that the radial velocity of the clouds in a fixed direction grows with distance up to the maximum velocity at cloud P_k :

$$v_{r,\max} = R_k(\omega - \omega_0), \quad (18.17)$$

where $R_k = R_0 \sin l$. The distance of cloud P_k from the Sun is $r = R_0 \cos l$. When r increases further, v_r decreases monotonically. Figure 18.17 shows how the observed radial velocity in a given direction varies with distance r , if the gas moves in circular orbits and the angular velocity decreases outwards.

The neutral hydrogen 21 cm line has been particularly important for mapping the Milky Way. Figure 18.18 gives a schematic view of how the hydrogen spectral line is made up of the radiation of many individual concentrations of neutral hydrogen, clouds or spiral arms. The line component produced by each cloud has a wavelength which depends on the radial velocity of the cloud and a strength depending on its mass and density. The total emission is the sum of these contributions.

By making observations at various galactic longitudes and assuming that the clouds form at least partly continuous spiral arms, the distribu-

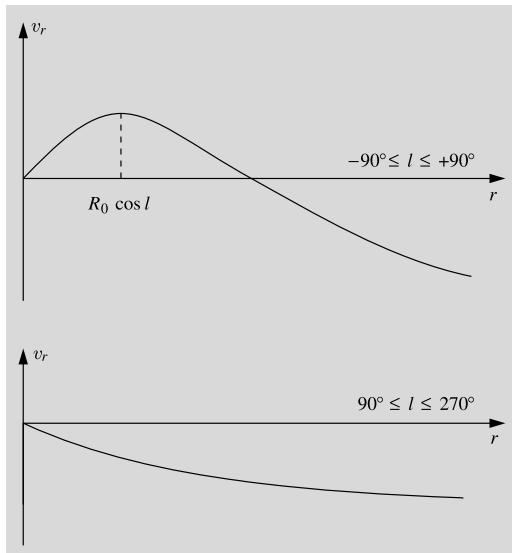


Fig. 18.17 The radial velocity as a function of distance (shown schematically)

tion of neutral hydrogen in the galactic plane can be mapped. Figure 18.17 shows a map of the Milky Way obtained from 21 cm line observations of neutral hydrogen. It appears that the neutral hydrogen is concentrated in spiral arms. However, interpretation of the details is difficult because of the uncertainties of the map. In order to obtain the distances to the gas clouds, one has to know the *rotation curve*, the circular velocity as a function of the galactic radius. This is determined from the same radial velocity observations and involves assumptions concerning the density and rotation of the gas. The interpretation of the spiral structure obtained from radio observations is also still uncertain. For example, it is difficult to fit the radio spiral structure to the one obtained near the Sun from optical observations of young stars and associations.

The Rotation, Mass Distribution and Total Mass of the Milky Way In (18.17) the galactic longitude l gives the galactic radius R_k of the clouds with maximum radial velocity. By making observations at different longitudes, one can therefore use (18.17) to determine the angular velocity of the gas for various distances from the galactic centre. (Circular motions must be assumed.) In this way, the rotation curve $\omega = \omega(R)$

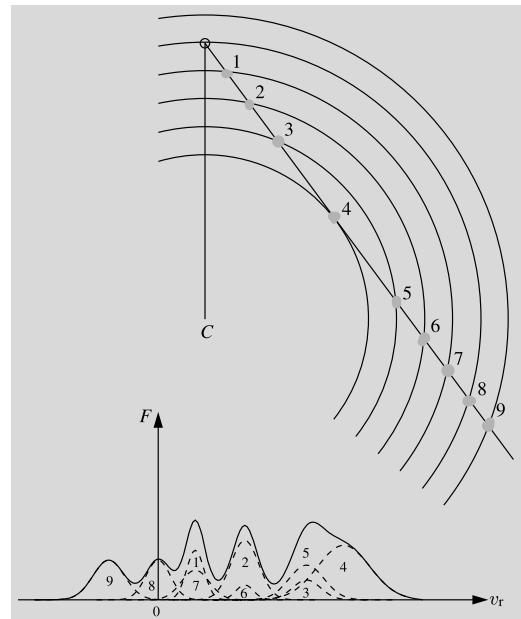


Fig. 18.18 Clouds at different distances have different velocities and therefore give rise to emission lines with different Doppler shifts. The observed flux density profile (continuous curve) is the sum of the line profiles of all the individual line profiles (dashed curves). The numbers of the line profiles correspond to the clouds in the upper picture

and the corresponding velocity curve $V = V(R)$ ($= \omega R$) are obtained.

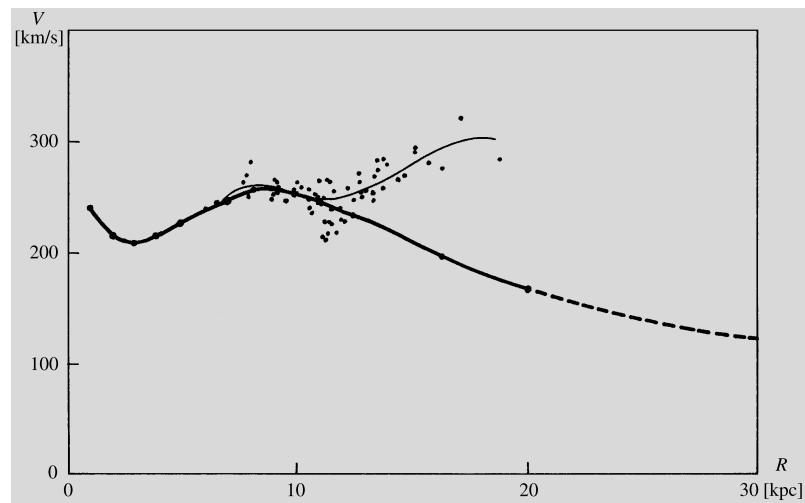
Figure 18.19 shows the rotation curve of the Milky Way. Its central part rotates like a rigid body, i.e. the angular velocity is independent of the radius. Outside this region, the velocity first drops and then begins to rise gradually. A maximum velocity is reached at about 8 kpc from the centre. Near the Sun, a little further from the centre, the rotational velocity is about 220 km s^{-1} . According to earlier opinions, the velocity continues to decrease outwards. This would mean that most of the mass is inside the solar radius. This mass could then be determined from Kepler's third law. According to (6.34),

$$M = R_0 V_0^2 / G.$$

Using the values $R_0 = 8.5 \text{ kpc}$ and $V_0 = 220 \text{ km s}^{-1}$, one obtains

$$M = 1.9 \times 10^{41} \text{ kg} = 1.0 \times 10^{11} M_\odot.$$

Fig. 18.19 Rotation curve of the Milky Way based on the motions of hydrogen clouds. Each point represents one cloud. The thick line represents the rotation curve determined by Maarten Schmidt in 1965. If all mass were concentrated within the radius 20 kpc, the curve would continue according to Kepler's third law (broken line). The rotation curve determined by Leo Blitz on the basis of more recent observations begins to rise again at 12 kpc



The escape velocity at radius R is

$$V_e = \sqrt{\frac{2GM}{R}} = V\sqrt{2}. \quad (18.18)$$

This gives an escape velocity near the Sun $V_e = 310 \text{ km s}^{-1}$. One therefore should not see many stars moving in the direction of galactic rotation, $l = 90^\circ$, with velocities larger than 90 km s^{-1} with respect to the local standard of rest, since the velocity of such stars would exceed the escape velocity. This has been confirmed by observations.

The preceding considerations have been based on the assumption that near the Sun, the whole mass of the Galaxy can be taken to be concentrated in a central point. If this were true, the rotation curve should be of the Keplerian form, $V \propto R^{-1/2}$. That this is not the case can be established from the values of the Oort constants.

The derivative of the Keplerian relation

$$V = \sqrt{\frac{GM}{R}} = \sqrt{GM} R^{-1/2}$$

yields

$$\frac{dV}{dR} = -\frac{1}{2} \sqrt{GM} R^{-3/2} = -\frac{1}{2} \frac{V}{R}.$$

Using the properties (18.15) and (18.16) of the Oort constants, one finds

$$(A - B)/(A + B) = 2 \quad (18.19)$$

for a Keplerian rotation curve. This disagrees with the observed value and thus the assumed Keplerian law does not apply.

The mass distribution in the Milky Way can be studied on the basis of the rotation curve. One looks for a suitable mass distribution, such that the correct rotation curve is reproduced. Recently distant globular clusters have been discovered, showing that the Milky Way is larger than expected. Also, observations of the rotation curve outside the solar circle suggest that the rotational velocity might begin to rise again. These results suggest that the mass of the Galaxy might be as much as ten times larger than had been thought.

18.4 Structural Components of the Milky Way

We have seen how the structure of the Milky Way can be globally described by means of an almost spherical halo of old stars and a disk of gas and young and middle-aged stars. Due to its structure the Milky Way belongs to disk galaxies, which are discussed in the next chapter. In a more detailed picture more small-scale features can also be distinguished.

The Thick Disk In the traditional scheme dividing the stars of the Milky Way into a series of populations it was left undecided whether

the populations should be considered as qualitatively different classes or merely steps along a continuous sequence. As the quality and quantity of observations have improved, it has become clear that what was defined as the intermediate population II really represents a separate component of the Milky Way with a pattern of element abundances stellar motions that separate it clearly from the old (thin) disk. This population is now referred to as the thick galactic disk. A thick disk has also been detected in some other galaxies, but it does not appear to be a universal feature of all disk galaxies.

The Galactic Bar As will be seen in the next chapter, Sect. 19.1, a large fraction of all disk galaxies are *barred*, with an elongated light distribution at the centre. The first indication that this might also be the case for the Milky Way was found in velocity measurements of neutral hydrogen, which were incompatible with gas moving along circular orbits. In 1971 W.W. *Shane* showed that the motions of the gas could be explained, if there is a central bar pointing about 20° away from the direction of the Galactic centre.

It is more difficult to detect a bar in observations of the stars. This was first done using the COBE satellite, which (apart from mapping the cosmic microwave background, see Sect. 19.7) also made a map of the sky at the infrared wavelengths dominated by the light of old stars. Because of perspective, the nearer end of the bar at positive galactic longitude will look slightly different than the farther end. Such an asymmetry was present in the infrared map, consistent with a bar with an axial ratio of 0.6.

Later confirmation of the existence of the bar has come from mapping of the central distribution of old stars using near-infrared photometric distances, see Fig. 18.20.

Spiral Structure As mentioned earlier, the Milky Way appears to be a *spiral galaxy* (Fig. 18.21), but there is no general agreement on the detailed form of the spiral pattern. For example, in 1976 Y.M. *Georgelin* and Y.P. *Georgelin* determined the distances of H II regions by radio and optical observations. In the optical region

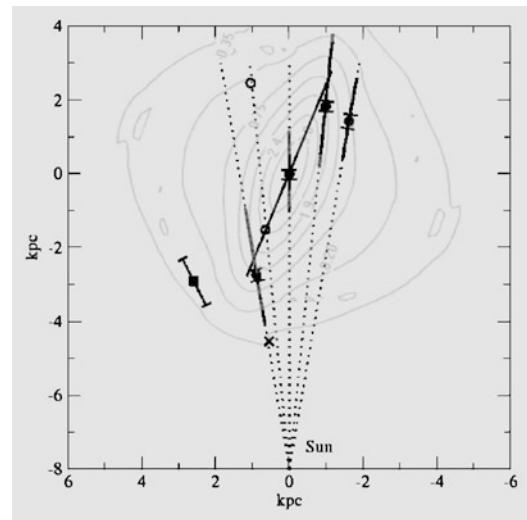


Fig. 18.20 The Milky Way bar viewed from the North Galactic pole. *Solid symbols* indicate the mean positions of red clump giants in given directions, and *the thick grey lines* represent the range of their distances. *The line* through the mean distances illustrates a bar of 3 kpc length inclined 22.5° to the direction to the Galactic centre. *The contour map* shows a model of the bar derived from infrared observations. (C. Babusiaux, G. Gilmore 2005, MNRAS 358, 1309, Fig. 6)

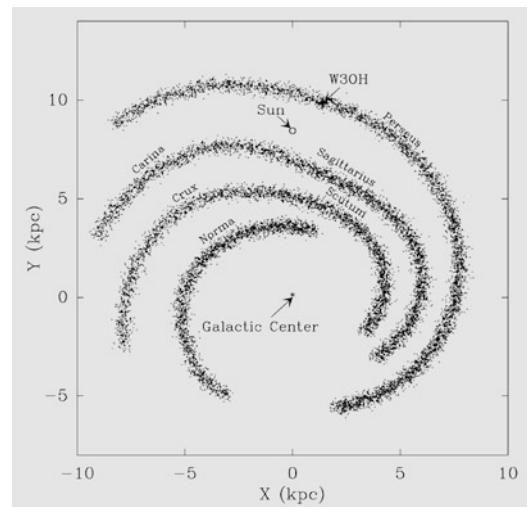


Fig. 18.21 General view of the spiral pattern of the Milky Way. Different tracers of spiral arms lead to somewhat different patterns, but they tend to agree that a four-armed pattern like the one indicated here gives the best overall representation. The names of the arms are those most commonly used. See also Fig. 18.11. (Y. Xu et al. 2006, Science 311, 54)

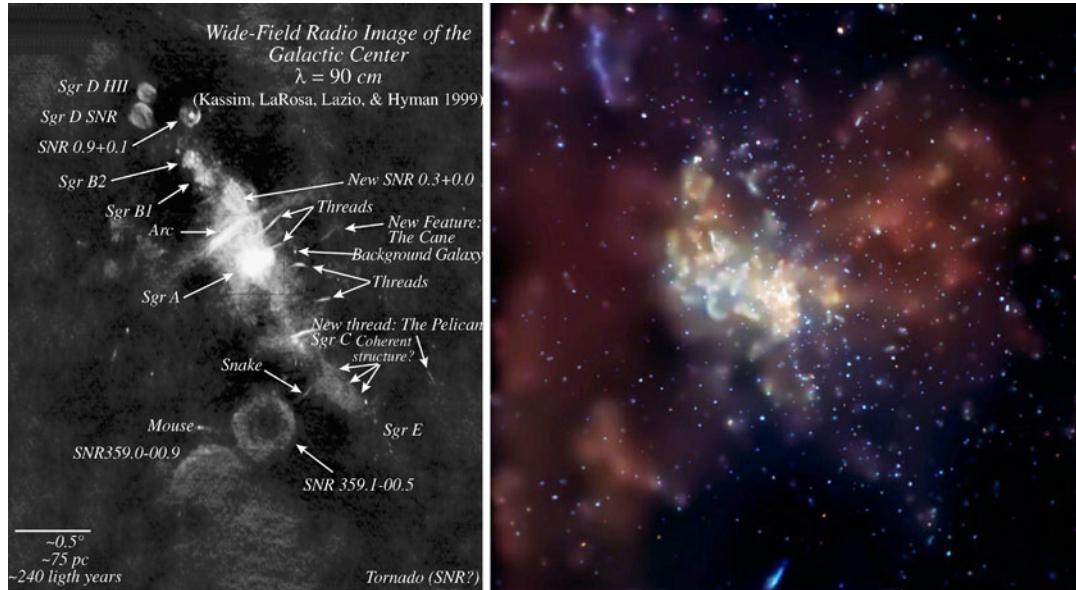


Fig. 18.22 *Left:* The central parts of the Milky Way in a radio picture. The observations were made with the VLA telescope. (Image Kassim et al., Naval Research Laboratory.) *Right:* The image by the Chandra satellite shows the

surroundings of the supermassive black hole SgrA* at the centre of the Milky Way. The width of the picture is $8.4'$, corresponding 20 parsecs at the distance of the centre of the Milky Way. (NASA/CXC/MIT/F.K. Baganoff et al.)

their method is independent of assumptions about the galactic rotation law. They then fitted four spiral arms through the H II regions.

Later investigations using a variety of methods, both optical and radio have confirmed that a four-armed pattern gives the best description of the spiral structure in the Sun's vicinity (Fig. 18.22). The pitch angle of the spiral in this model is about 11.3° . Three of the arms start at the position of the galactic bar.

The cause of the spiral structure is a long-standing problem. A small perturbation in the disk will quickly be stretched into a spiral shape by differential rotation. However, such a spiral would disappear in a few galactic revolutions, a few hundred million years.

An important step forward in the study of the spiral structure was the *density wave theory* developed by *Chia-Chiao Lin* and *Frank H. Shu* in the 1960's. The spiral structure is taken to be a wavelike variation of the density of the disk. The spiral pattern rotates as a solid body with an angular velocity smaller than that of the galactic rotation, while the stars and gas in the disk pass through the wave.

The density wave theory explains in a natural way why young objects, like molecular clouds, H II regions and bright young stars are found in the spiral arms. As gas passes through the wave, it is strongly compressed. The internal gravity of the gas clouds then becomes more important and causes them to collapse and form stars.

It takes about 10^7 years for the material to pass through a spiral arm. By that time, the hot, bright stars have finished their evolution, their ultraviolet radiation has ceased and the H II regions have disappeared. The less massive stars formed in the spiral arms are spread out in the disk by their peculiar velocities.

It is not yet clear what gives rise to the spiral wave. For some further discussion of spiral structure, see Sect. 19.4.

The Galactic Centre Our knowledge of the centre of the Milky Way is mostly based on radio and infrared observations (Fig. 18.22). In the optical region the view towards the centre is blocked already by the dark clouds in the Sagittarius spiral arm about 2 kpc from us. The galactic centre is interesting because it may be a small-scale version

of the much more violently active nuclei of some external galaxies (see Sect. 19.7). It therefore provides opportunities to study at close hand phenomena related to active galaxies. At the galactic centre there is a supermassive black hole with a mass of nearly $5 \times 10^6 M_\odot$.

As one approaches the galactic centre the stellar density continues to rise towards a sharp central peak. In contrast, the galactic gas disk has a central hole of radius about 3 kpc. This may be due to the galactic bar, which will channel gas into the galactic nucleus leaving a gas-free zone at larger radii.

Inside the central hole there is a dense gas disk. Its radius is about 1.5 kpc in neutral hydrogen, but most of its mass is molecular and concentrated within 300 pc of the nucleus. In this region the mass of molecular gas is about $10^8 M_\odot$, or 5 % of the total molecular mass of the Milky Way. The molecular clouds are probably confined by the pressure from surrounding very hot ($T \approx 10^8$ K) gas. This hot gas may then expand vertically, forming a galactic wind. Gas lost to a wind or to star formation is replenished with infalling gas from larger radii.

The central 10 pc are dominated by the radio continuum source Sgr A and a dense star cluster observed in the infrared. There is also molecular gas with complex motions and signs of star formation activity. Within Sgr A there is a unique point-like radio continuum source known as Sgr A*. The position of Sgr A* agrees to within $1''$ with the centre of a cluster of stars that is much denser than anything observed in the galactic disk.

The stellar density of the central star cluster can be as high as 3×10^{14} stars per cubic parsec, while around the Sun it is just one star per cubic parsec. The stars of the cluster orbit the central black hole. Motions of the innermost stars have been followed since the end of the 1990's. A couple of stars have already managed to make a full revolution around the centre (Fig. 18.23), the periods being 16 and 11.5 years. When the period is known and the size of the orbit is seen the mass of the central body can be calculated. The researchers working with the Keck telescope got in 2014 a value of $4.6 \pm 0.7 \times 10^6$ solar masses.

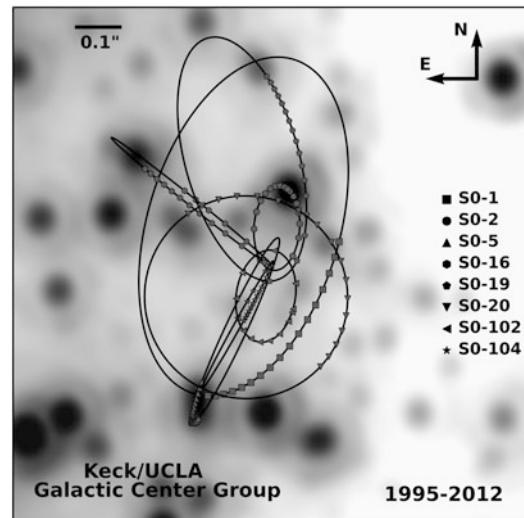


Fig. 18.23 Motions of stars near the centre of the Milky Way during over 15 years. The background image was taken in 2014 with the Keck telescope. (Kuva Andrea Ghez/UCLA/W.M. Keck Telescopes)

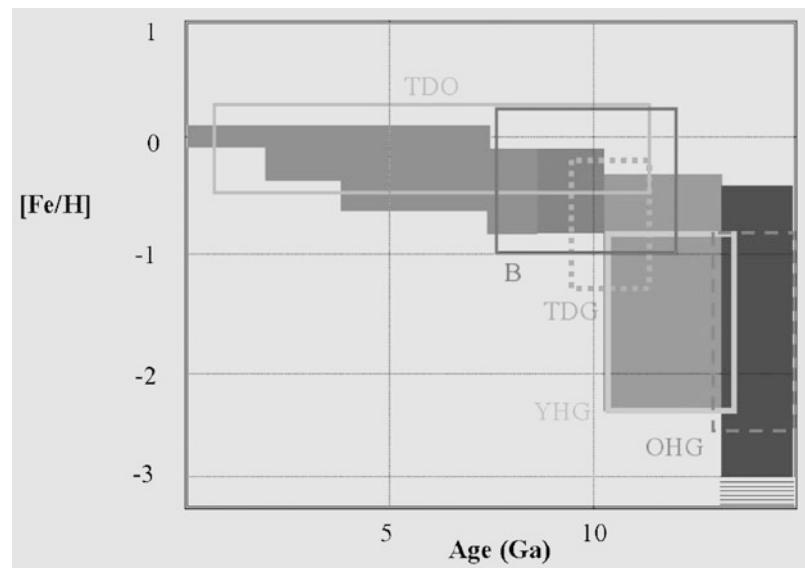
According to VLBI measurements the diameter of the accretion disk surrounding the black hole is less than 10 au. Eruptions observed at infrared wavelengths, lasting only about twenty minutes, point to the same size. It can be calculated that the diameter of the black hole is about 0.2 au.

18.5 The Formation and Evolution of the Milky Way

Like all galaxies the Milky Way is supposed to have formed through the collapse of a region of higher than average density in the Universe. The general view of how this happened will be presented in the next two chapters, in particular in Sects. 19.8 and 20.7. Some traces of the formation process will be preserved in the properties of local stars of different ages. These represent information about the formation of the Milky Way that is not available for other galaxies.

The Ages of Stars The most direct way of studying the evolution of the Milky Way is by means of the observed ages of stars. The traditional sequence of populations described in Sect. 18.2 is also a sequence corresponding to

Fig. 18.24 The age–metallicity relation of the Milky Way for different components. TDO: thin disc open clusters; TDG: thick disc globulars; B: bulge; YHG: young halo globulars; OHG: old halo globulars. The shaded areas correspond to thin disc field stars, thick disc field stars, and halo field stars, in order of increasing age. (K. Freeman, J. Bland-Hawthorn 2002, ARAA 40, 487, Fig. 2)



stars of different ages. The oldest component, the stellar halo forms an almost spherical distribution of stars with ages 12–14 Ga, representing the oldest part of the Milky Way.

In contrast, the old and young population I consists of stars with ages less than 10 Ga, which were initially formed in a very thin layer that has been thickened by encounters with spiral arms and molecular cloud complexes.

Intermediate between these two, as we have noted, is the thick disk with an age about 10–12 Ga. Another intermediate population in the Milky Way is the central bulge including the galactic bar, and containing stars of ages 7–11 Ga.

Chemical Enrichment The history of the formation of the Milky Way is preserved in the properties of its older stars, above all in their “chemical” composition, i.e. the abundances of elements and isotopes heavier than helium. (All these heavier elements are often referred to as “metals” in the astrophysical jargon.)

Only hydrogen and helium were present as the first stars formed. As successive generations of stars evolved, nuclear reactions produced heavy elements, some of which were returned to the interstellar gas in supernova explosions or stellar winds. These heavy elements were then incorporated into the following generation of stars, grad-

ually increasing the metal abundance in the interstellar medium.

Some of the stars formed were of low mass, and had life-times long enough that they are still present. Their chemical composition will reflect the abundances in the interstellar medium at the time they were formed. Studying the chemical abundances of stars of different ages therefore gives information on the star formation history of the Milky Way, both on the rate at which stars have formed at a given epoch and on the masses and other characteristics of those stars when they formed.

One commonly used indicator of the metal abundance in stars is the mass of iron relative to hydrogen, in logarithmic units and measured relative to the solar values, $[Fe/H]$. Figure 18.24 shows the value of $[Fe/H]$ for different types of objects of various ages. The general picture is that the metal abundance rose rapidly during the first billion years and has afterwards grown only slowly. The lowest values of $[Fe/H]$ discovered in old halo stars are about -5 .

Many models have been developed describing the chemical evolution of the Milky Way and other galaxies, incorporating prescriptions for star formation, and infall of gas from the outside. In particular the rapid early rise of the metal abundance is difficult to explain in the sim-

plest models. This difficulty, the scarcity of old metal-poor disk stars, has become known as the “*G dwarf problem*”, since the oldest stars still on the main series are of spectral class G. The most direct way that the G dwarf problem can be avoided is to suppose that a large portion of the interstellar gas was accreted after the formation of the oldest stars.

Formation of the Milky Way Galaxies are thought to have formed when gas clouds of higher than average density in the Universe collapsed under their own gravity. As the gas is compressed stars are born in it. After the end of the collapse the cloud settles into a quasi-stationary state. Evolution continues at a slower pace. Evolving stars return chemically enriched gas to interstellar space, where it is mixed with remaining, unprocessed gas and star formation continues.

There are two competing pictures of how galaxy formation proceeds. In the *monolithic* collapse model it is assumed that large galaxies form coherently in the collapse of a massive cloud containing the bulk of the material constituting the galaxy. In the *hierarchical* picture most of the stars form in much smaller clouds, which later agglomerate together to form the galaxies we now observe. The evidence for these pictures in other galaxies will be considered in the next chapter.

The case of the Milky Way already shows that the contrast of the monolithic and the hierarchical picture is oversimplified. Some aspects of galactic structure fit more naturally into a monolithic theory. For example, the rapid collapse of the halo, followed by a more gradual build-up of the disk fits with this picture. The pattern of chemical abundances are also homogeneous in a way that is more compatible with a pattern of star formation that was uniform throughout galactic history.

Other observations are indicative of a hierarchical formation history. For example, stellar abundance patterns in the thick disk differ from those in the old thin disk. The most natural way of explaining this is that the thick disk formed by the accretion of one or more small satellite galaxies with different histories of star formation. Another sign of the importance of the infall of satellite galaxies is the existence of systems such as the

Sagittarius dwarf galaxy, which appears to be in the course of being disrupted by the Milky Way.

There is about five times as much dark matter as visible matter. Recent simulations have confirmed that this dark matter has an essential role in the formation of galaxies. During the first billion years of the universe the dark matter formed gravitational condensates to which the visible matter was concentrated to form initial galaxies. Formation of the current galaxies has been a complicated process that involved effects of the dark matter, interactions, collisions and mergers of galaxies.

18.6 Examples

Example 18.1 Show that if the stars are uniformly distributed in space and there is no interstellar extinction, the number of stars brighter than apparent magnitude m is

$$N_0(m) = N_0(0) \times 10^{0.6m}.$$

Let us suppose first that all stars have the same absolute magnitude M . The distance in parsecs of those stars having the apparent magnitude m is

$$r = 10 \times 10^{0.2(m-M)}.$$

In order to appear brighter than m , a star has to be within a sphere of radius r . Because the stellar density is constant, the number of such stars is proportional to the volume:

$$N_0(m) \propto r^3 \propto 10^{0.6m}.$$

The result does not depend on the absolute magnitudes of the stars, so that the same result still applies when the magnitudes are not equal, as long as the luminosity function does not depend on distance. Thus the equation is generally valid under the stated conditions.

Example 18.2 (The Estimation of Distances by Means of Oort’s Formulas) An object in the galactic plane at longitude $l = 45^\circ$ has the radial velocity of 30 km s^{-1} with respect to the LSR. What is its distance?

According to (18.10),

$$v_r = Ar \sin 2l.$$

Thus

$$r = \frac{v_r}{A \sin 2l} = \frac{30 \text{ km s}^{-1}}{15 \text{ km s}^{-1} \text{kpc}^{-1}} = 2 \text{ kpc}.$$

In practice, the peculiar velocities are so large that this method cannot be used for distance determination. Oort's formulas are mostly suitable only for statistical studies.

Example 18.3 (Discussion of the Gravitational Field of a Uniform Disk) It can be shown that the gravitational field of a homogeneous infinite thin disk is constant and directed towards the plane of the disk. If the mass per unit area of the disk is σ , the gravitational field is

$$g = 2\pi G\sigma.$$

A test particle located outside the plane will therefore get a constant acceleration towards the plane. Taking a numerical example, assume a mass of $10^{11} M_\odot$ distributed uniformly on a circular disk 20 kpc in diameter. The mass per unit area will be a

$$\begin{aligned}\sigma &= \frac{10^{11} \times 2 \times 10^{30} \text{ kg}}{\pi (10^4 \times 3.086 \times 10^{16} \text{ m})^2} \\ &= 0.67 \text{ kg m}^{-2}.\end{aligned}$$

The corresponding gravitational field is

$$g = 2.8 \times 10^{-10} \text{ m s}^{-2}.$$

Let a star be located at $d = 1$ kpc above the plane, initially at rest (not very near the edge of the disk, in order to keep our approximation valid). The disk will pull it towards the plane, and when the star crosses the plane, it has acquired a velocity given by

$$v = \sqrt{2gd} = 130 \text{ km s}^{-1}.$$

The time required to reach the plane is

$$t = v/g = 15 \times 10^6 \text{ a.}$$

18.7 Exercises

Exercise 18.1 Assume that the Sun and a star move around the Galaxy at the same speed in the same circular orbit in the galactic plane. Show that the proper motion of the star is independent of its distance. How big is this proper motion?

Exercise 18.2 (a) A cepheid has a radial velocity of 80 km s^{-1} , and its galactic longitude is 145° . What is the distance of the cepheid?

(b) The period of the cepheid is 3.16 d, and the apparent visual magnitude is 12.3. What is the distance derived from this information? Are the distances consistent?

Exercise 18.3 (a) How many of the nearest stars (Table C.15) are also among the brightest stars (Table C.16)? Explain.

(b) If the stellar density were constant, how many stars would there be within the distance of Canopus?

Exercise 18.4 (a) Assume that the Galaxy is a homogeneous disk and the Sun lies in the central plane of the disk. The absolute magnitude of a star is M , galactic latitude b , and distance from the central plane z . What is the apparent magnitude of the star, if the extinction inside the Galaxy is $a \text{ mag kpc}^{-1}$?

(b) Assume that the thickness of the galactic disk is 200 pc. Find the apparent magnitude of a star with $M = 0.0$, $b = 30^\circ$, distance $r = 1$ kpc, and $a = 1 \text{ mag kpc}^{-1}$.

The galaxies are the fundamental building blocks of the Universe. Some of them are very simple in structure, containing only normal stars and showing no particular individual features. There are also galaxies that are almost entirely made of neutral gas. On the other hand, others are complex systems, built up from many separate interacting components—stars, neutral and ionised gas, dust, molecular clouds, magnetic fields, cosmic rays.... The galaxies may form small groups or large clusters in space. At the centre of many galaxies, there is a compact nucleus that may sometimes be so bright that it overwhelms all the normal radiation of the galaxy.

The luminosity of the brightest normal galaxies may correspond to 10^{12} solar luminosities, but most of them are much fainter—the smallest ones that have been discovered are about $10^5 L_\odot$. Since galaxies do not have a sharp outer edge, to some extent their masses and radii depend on how these quantities are defined. If only the bright central parts are included, a giant galaxy may typically have a mass of about $10^{13} M_\odot$, and a radius of 30 kpc, and a dwarf, correspondingly, $10^7 M_\odot$, and 0.5 kpc.

When the existence of dark matter was confirmed the whole field of galactic research was changed. It was only the dark matter that helped to give a consistent model for the origin and evolution of galaxies.

The fraction of dark matter of the total mass of galaxies varies considerably. In big galaxies there is about five times as much dark matter as ordinary visible matter. In the smallest galaxies the

ratio of the dark matter to the visible matter can be hundreds or even thousands.

When smaller and smaller galaxies have been found the boundary between star clusters and galaxies has become obscure. The smallest galaxies are similar to dense star clusters. Therefore some researchers have redefined galaxies: a star system is a galaxy if in addition to stars it contains also dark matter.

19.1 The Classification of Galaxies

A useful first step towards an understanding of galaxies is a classification based on their various forms. Although such a morphological classification must always be to some extent subjective, it provides a framework within which the quantitative properties of galaxies can be discussed in a systematic fashion. However, it should always be remembered that the picture thus obtained will be limited to those galaxies that are large and bright enough to be easily visible in the sky. An idea of the consequent limitations can be obtained from Fig. 19.1, showing the radii and magnitudes of normal galaxies. One sees that only within a narrow region of this diagram can galaxies be easily found. If a galaxy has too large a radius for its magnitude (small surface brightness), it will disappear in the background light from the night sky. On the other hand, if its radius is too small, it looks like a star and is not noticed on a photographic plate. In the following, we shall mainly be concerned with bright galaxies that fit within these limits.

If a classification is to be useful, it should at least roughly correspond to important phys-

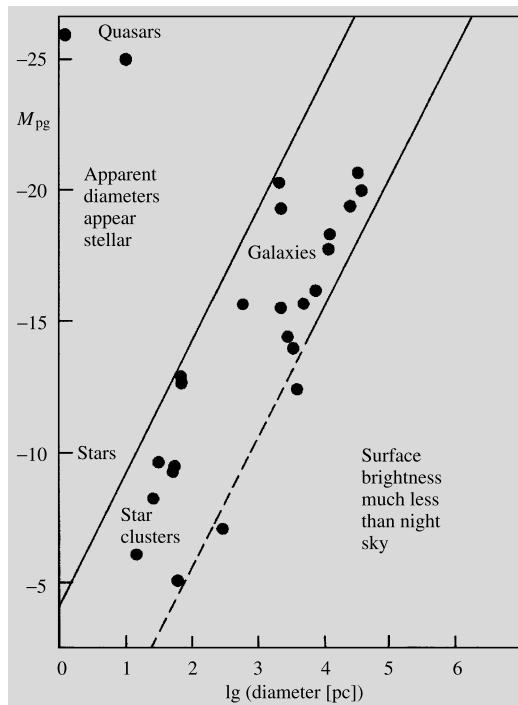


Fig. 19.1 Magnitudes and diameters of observable extragalactic objects. Objects to the upper left look like stars. The quasars in this region have been discovered on the basis of their spectra. Objects to the lower right have a surface brightness much smaller than that of the night sky. In recent years large numbers of low surface brightness galaxies have been discovered in this region. (Arp, H. (1965): *Astrophys. J.* **142**, 402)

ical properties of the galaxies. Most classifications accord in their main features with the one put forward by *Edwin Hubble* in 1926. Hubble's own version of the *Hubble sequence* is shown in Fig. 19.2. The various types of galaxies are ordered in a sequence from early to late types. There are three main types: *elliptical*, *lenticular*, and *spiral* galaxies. The spirals are divided into two sequences, *normal* and *barred* spirals. In addition, Hubble included a class of *irregular galaxies*. In lenticular and spiral galaxies the matter is concentrated in a flat disk and thus they can be called collectively as *disk galaxies*.

The elliptical galaxies (Fig. 19.6) appear in the sky as elliptical concentrations of stars, in which the density falls off in a regular fashion as one goes outwards. Usually there are no signs of interstellar matter (dark bands of dust, bright young stars). The ellipticals differ from each other only in shape and on this basis they are classified as E0, E1, ..., E7. If the major and minor axes of an elliptical galaxy are a and b , its type is defined to be En, where

$$n = 10 \left(1 - \frac{b}{a} \right). \quad (19.1)$$

An E0 galaxy thus looks circular in the sky. The apparent shape of an E galaxy depends on the direction from which it is seen. In reality an E0 galaxy may therefore be truly spherical or it may be a circular disk viewed directly from above.

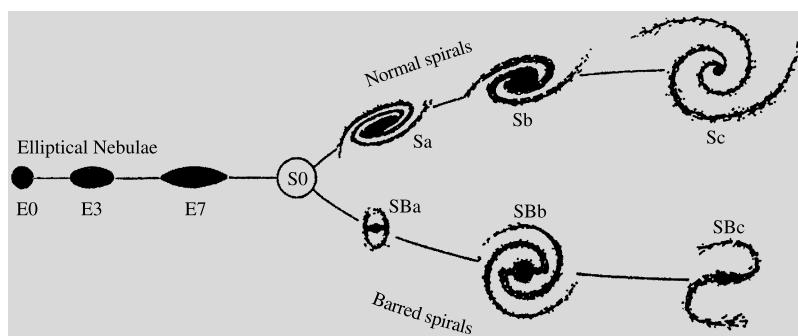


Fig. 19.2 The Hubble sequence in Hubble's 1936 version. At this stage the existence of type S0 was still doubtful. Photographs of the Hubble types are shown in Figs. 18.6 and 18.15 (E); 18.3 and 18.4 (S0 and S); 18.12

(S and Irr II); 18.5 (Irr I and dE). (Hubble, E.P. (1936): *The Realm of the Nebulae* (Yale University Press, New Haven))

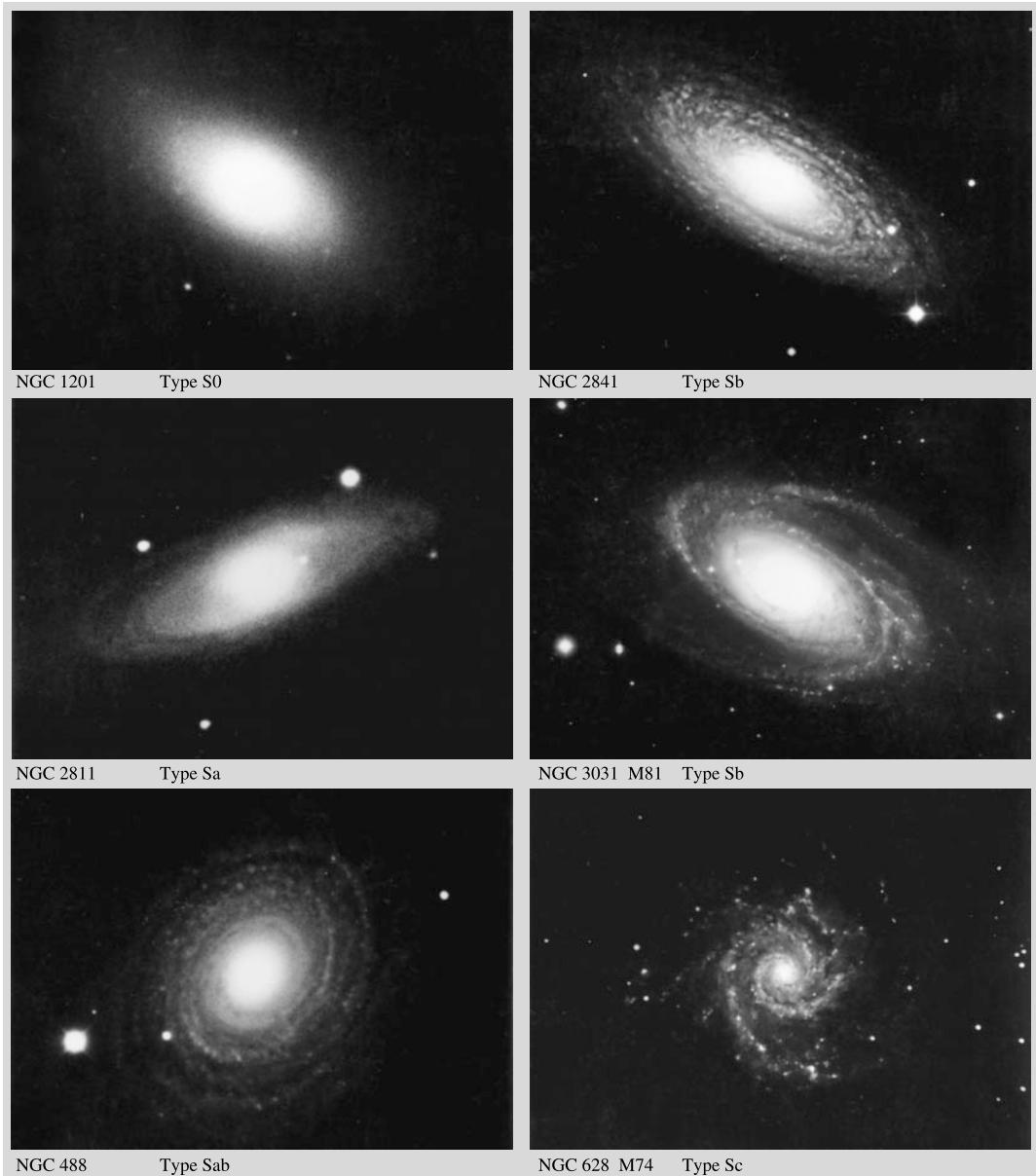


Fig. 19.3 The classification of normal spiral and S0 galaxies. (Mt. Wilson Observatory)

A later addition to the Hubble sequence is a class of *giant elliptical galaxies* denoted cD. These are generally found in the middle of clusters of galaxies. They consist of a central part looking like a normal elliptical surrounded by an extended fainter halo of stars.

In the Hubble sequence the lenticulars or S0 galaxies are placed between the elliptical and

the spiral types. Like the ellipticals they contain only little interstellar matter and show no signs of spiral structure. However, in addition to the usual elliptical stellar component, they also contain a flat disk made up of stars. In this respect they are like spiral galaxies (Figs. 19.3, 19.4).

The characteristic feature of spiral galaxies is a more or less well-defined spiral pattern in the

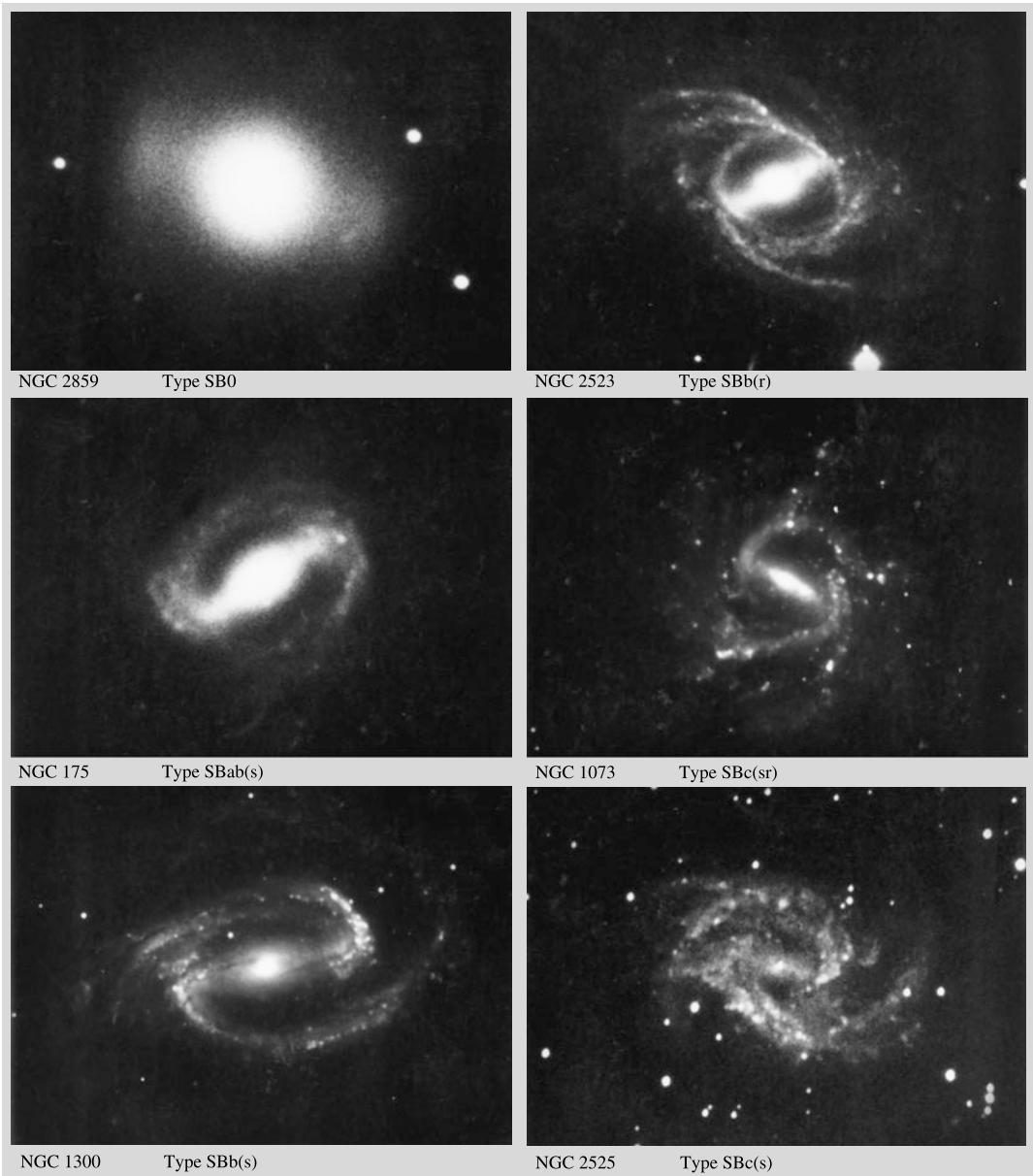
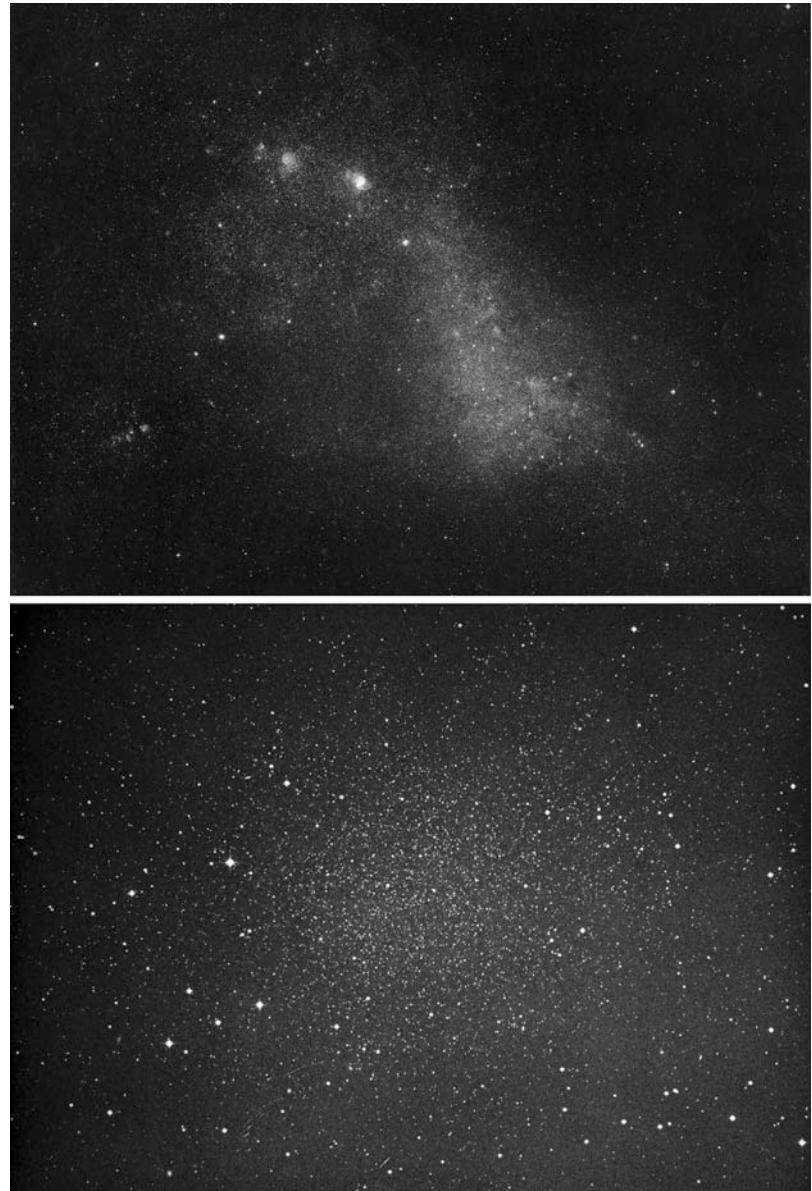


Fig. 19.4 Different types of SB0 and SB galaxies. The type (r) or (s) depends on whether the galaxy has a central ring or not. (Mt. Wilson Observatory)

disk. Spiral galaxies consist of a central *bulge*, which is structurally similar to an E galaxy, and of a stellar disk, like in an S0 galaxy. In addition to these, there is a thin disk of gas and other interstellar matter, where young stars are being born, forming the spiral pattern. There are two sequences of spirals, normal Sa–Sb–Sc, and barred

SBa–SBb–SBc spirals. In the barred spirals the spiral pattern ends at a central bar, whereas in the normal spirals the spiral pattern may end at an inner ring or continue all the way to the centre. The position of a galaxy within the spiral sequence is determined on the basis of three criteria (which are not always in agreement): later types

Fig. 19.5 *Above:* The Small Magellanic Cloud (Hubble type Irr I), a dwarf companion of the Milky Way (Royal Observatory, Edinburgh). *Below:* The Sculptor Galaxy, a dE dwarf spheroidal. (ESO)



have a smaller central bulge, more narrow spiral arms and a more open spiral pattern. The Milky Way Galaxy is thought to be of type SABbc (intermediate between Sb and Sc, and between normal and barred spirals).

The classical Hubble sequence is essentially based on bright galaxies; faint galaxies have been less easy to fit into it (Fig. 19.5). For example, the irregular galaxies of the original Hubble sequence can be divided into the classes Irr I and

Irr II. The Irr I galaxies form a continuation of the Hubble sequence towards later types beyond the Sc galaxies. They are rich in gas and contain many young stars. Type Irr II are somewhat irregular small ellipticals containing only few young stars but a lot of interstellar dust.

The smallest irregular galaxies are classified as dIrr (dwarf irregulars). Another large group of dwarf galaxies are the spherical or elliptic dwarf galaxies dSph (dwarf spheroidals) and dE

Fig. 19.6 M32 (type E2), a small elliptical companion of the Andromeda Galaxy. (NOAO/Kitt Peak National Observatory)



(dwarf ellipsoids). They are much smaller than ordinary elliptic galaxies or the stellar density is much lower. Recently new kinds of even smaller galaxies have been found. These include *ultrafaint dwarfs* (uFd), *blue compact dwarfs* (BCD) and *ultracompact dwarfs* (UCD) (Fig. 19.7).

In addition to the precise morphological classification a simpler classification based on colours was introduced at the beginning of the 2000's. The idea to classify fast the millions of new galaxies found by modern digital survey programs as well as the very faint galaxies whose shape is difficult to distinguish. When galaxies are plotted like stars in a colour-magnitude diagram they are seen to form two distinct regions: *red galaxies* (red sequence) and *blue galaxies* (blue sequence) (Fig. 19.8). The red galaxies are mainly ellipticals with old stars and blue galaxies mainly spirals with younger stars.

19.2 Luminosities and Masses

Distances In order to determine the absolute luminosities and linear dimensions of galaxies one needs to know their distances. Distances are also needed in order to estimate the masses of galaxies, because these estimates depend on the absolute linear size. Distances within the Local Group



Fig. 19.7 The ultracompact dwarf M60-UCD1 is close to the galaxy M60. The mass of the dwarf galaxy is 200×10^6 solar masses but the radius is only 100 lightyears. (Photo NASA, ESA, CXC, J. Strader (Michigan State University))

can be measured by the same methods as inside the Milky Way, most importantly by means of variable stars. On the very large scale (beyond 50 Mpc), the distances can be deduced on the basis of the expansion of the Universe (see

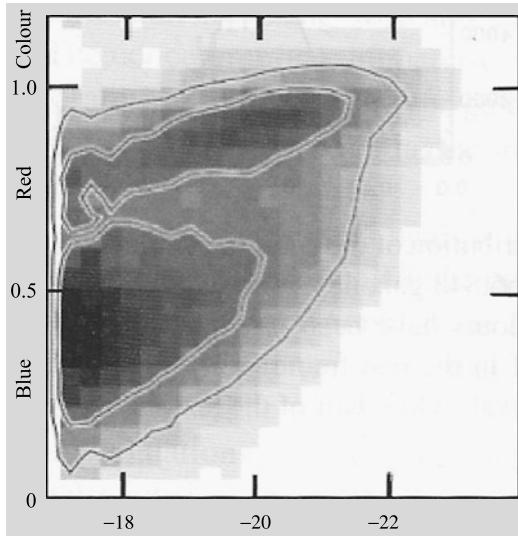


Fig. 19.8 A colour-magnitude graph contains 144,000 galaxies from the Sloan digital sky survey (SDSS). Brightness increases from left to right. Galaxies are clearly divided into two regions, red and blue. (Blanton et al., 2003, ApJ 592, 819)

Sect. 20.1). In order to connect these two regions one needs methods of distance determination based on the properties of individual galaxies.

To some extent local distances can be determined using structural components of galaxies, such as the sizes of HII regions or the magnitudes of globular clusters. However, to measure distances of tens of megaparsecs, one needs a distance-independent method to determine the absolute luminosities of entire galaxies. Several such methods have been proposed. For example, a luminosity classification has been introduced for late spiral types by *Sidney van den Bergh*. This is based on a correlation between the luminosity of a galaxy and the prominence of its spiral pattern. Nowadays this method is not considered accurate enough.

Other distance indicators are obtained if there is some intrinsic property of the galaxy, which is correlated with its total luminosity, and which can be measured independently of the distance. Such properties are the colour, the surface brightness and the internal velocities in galaxies. All of these have been used to measure distances to both spiral and elliptical galaxies.

For example, the absolute luminosity of a galaxy should depend on its mass. The mass, in turn, will be reflected in the velocities of stars and gas in the galaxy. Accordingly there is a relationship between the absolute luminosity and the velocity dispersion (in ellipticals) and the rotational velocity (in spirals). Since rotational velocities can be measured very accurately from the width of the hydrogen 21-cm line, the latter relationship (known as the *Tully–Fisher relation*) is perhaps the best distance indicator currently available.

The luminosity of the brightest galaxies in clusters has been found to be reasonably constant. This fact can be used to measure even larger distances, providing a method which is important in cosmology.

Luminosities The definition of the total luminosity of a galaxy is to some extent arbitrary, since galaxies do not have a sharp outer edge. The usual convention is to measure the luminosity of a galaxy out to a given value of the surface brightness, e.g. to 26.5 mag/sq.arcsec. For a given Hubble type, the total luminosity L may vary widely.

As in the case of stars, the distribution of galaxy luminosities is described by the *luminosity function* $\Phi(L)$ (Fig. 19.9). This is defined so that the space density of galaxies with luminosities between L and $L + dL$ is $\Phi(L) dL$. It can be determined from the observed magnitudes of galaxies, once their distances have been estimated in some way. In practice, one assumes some suitable functional form for $\Phi(L)$, which is then fitted to the observations. One common form is *Schechter's luminosity function*,

$$\Phi(L) dL = \Phi^* \left(\frac{L}{L^*} \right)^\alpha e^{-L/L^*} d\left(\frac{L}{L^*} \right). \quad (19.2)$$

The values of the parameters Φ^* , L^* , α are observationally determined for different types of objects; in general, they will be functions of position.

The shape of the luminosity function is described by the parameters α and L^* . The relative number of faint galaxies is described by α . Since its observed value is about -1.1 , the density of galaxies grows monotonically as one goes towards fainter luminosities. The luminosity function falls off steeply above the luminosity L^* ,

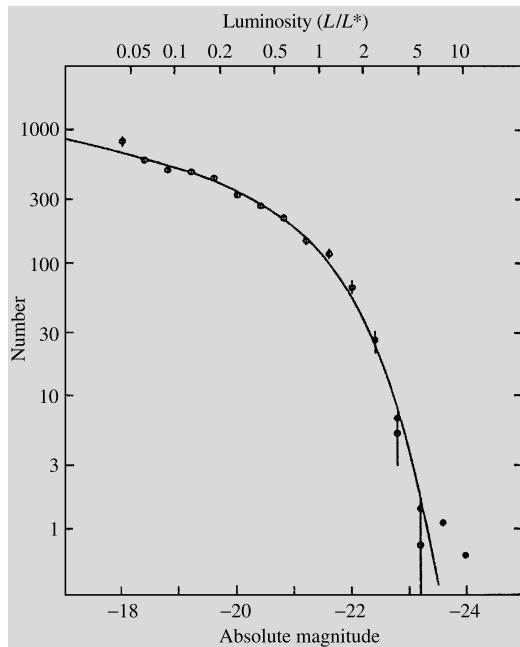


Fig. 19.9 Compound luminosity function of thirteen clusters of galaxies. The open symbols have been obtained by omitting the cD galaxies. The distribution is then well described by (19.2). The cD galaxies (filled symbols) cause a deviation at the bright end. (Schechter, P. (1976): *Astrophys. J.* **203**, 297)

which therefore represents a characteristic luminosity of bright galaxies. The observed L^* corresponds to an absolute magnitude $M^* = -21.0$ mag. The corresponding magnitude for the Milky Way Galaxy is probably -20.2 mag. The cD giant galaxies do not obey this brightness distribution; their magnitudes may be -24 mag and even brighter.

The parameter Φ^* is proportional to the space density of galaxies and is therefore a strong function of position. Since the total number density of galaxies predicted by relation (19.2) is infinite, we define $n^* = \text{density of galaxies with luminosity } > L^*$. The observed average value of n^* over a large volume of space is $n^* = 3.5 \times 10^{-3} \text{ Mpc}^{-3}$. The mean separation between galaxies corresponding to this density is 4 Mpc. Since most galaxies are fainter than L^* , and since, in addition, they often belong to groups, we see that the distances between normal galaxies are generally not much larger than their diameters.

Masses The distribution of mass in galaxies is a crucial quantity, both for cosmology and for theories of the origin and evolution of galaxies. Observationally it is determined from the velocities of the stars and interstellar gas. Total masses of galaxies can also be derived from their motions in clusters of galaxies. The results are usually given in terms of the corresponding mass-luminosity ratio M/L , using the solar mass and luminosity as units. The value measured in the solar neighbourhood of the Milky Way is $M/L = 3$. If M/L were constant, the mass distribution could be determined from the observed luminosity distribution by multiplying with M/L . Unfortunately, there is no universally valid value for the ratio M/L .

The masses of elliptical galaxies may be obtained from the stellar velocity dispersion given by the broadening of spectral lines. The method is based on the virial theorem (see Sect. 6.10), which says that in a system in equilibrium, the kinetic energy T and the potential energy U are related according to the equation

$$2T + U = 0. \quad (19.3)$$

Since ellipticals rotate slowly, the kinetic energy of the stars may be written

$$T = Mv^2/2, \quad (19.4)$$

where M is the total mass of the galaxy and v the velocity width of the spectral lines. The potential energy is

$$U = -GM^2/2R, \quad (19.5)$$

where R is a suitable average radius of the galaxy that can be estimated or calculated from the light distribution. Introducing (19.4) and (19.5) into (19.3) we obtain:

$$M = 2v^2R/G. \quad (19.6)$$

From this formula the mass of an elliptical galaxy can be calculated when v^2 and R are known. Some observations of velocities in ellipti-

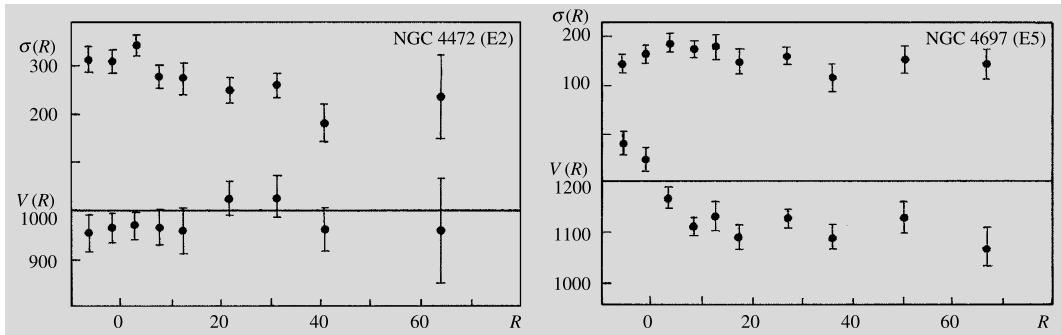
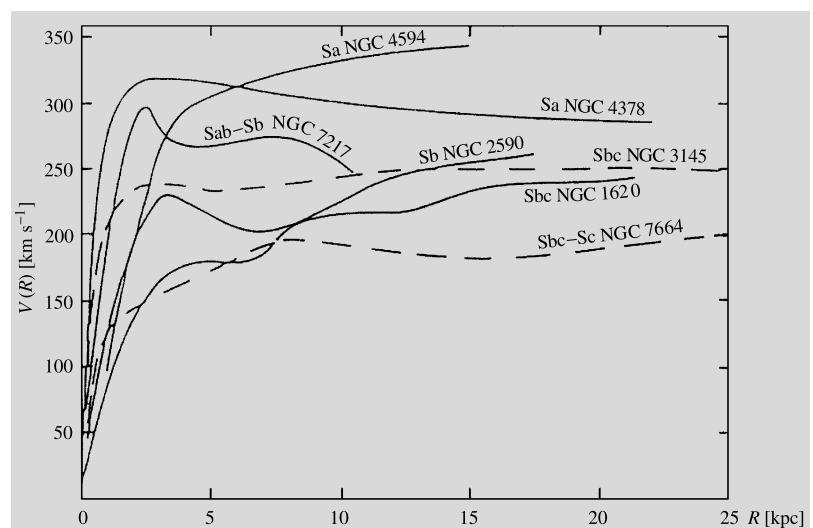


Fig. 19.10 Velocity of rotation $V(R)$ [km s^{-1}] and velocity dispersion $\sigma(R)$ [km s^{-1}] as functions of radius [kpc] for types E2 and E5. The latter galaxy is rotating,

the former is not. (Davies, R.L. (1981): Mon. Not. R. Astron. Soc. **194**, 879)

Fig. 19.11 Rotation curves for seven spiral galaxies. (Rubin, V.C., Ford, W.K., Thonnard, N. (1978): Astrophys. J. (Lett.) **225**, L107)



cal galaxies are given in Fig. 19.10. These will be further discussed in Sect. 19.4. The value of M/L derived from such observations is about 10 within a radius of 10 kpc. The mass of a bright elliptical might thus be up to $10^{13} M_\odot$.

The masses of spiral galaxies are obtained from their *rotation curve* $v(R)$, which gives the variation of their rotational velocity with radius. Assuming that most of the mass is in the almost spherical bulge, the mass within radius R , $M(R)$, can be estimated from Kepler's third law:

$$M(R) = Rv(R)^2/G. \quad (19.7)$$

Some typical rotation curves are shown in Fig. 19.11. In the outer parts of many spirals,

$v(R)$ does not depend on R . This means that $M(R)$ is directly proportional to the radius—the further out one goes, the larger the interior mass is. Since the outer parts of spirals are very faint, at large radii the value of M/L is directly proportional to the radius. For the disk, one finds that $M/L = 8$ for early and $M/L = 4$ for late spiral types. The largest measured total mass is $2 \times 10^{12} M_\odot$.

In order to measure the mass at even larger radii where no emission can be detected, motions in systems of galaxies have to be used. One possibility is to use pairs of galaxies. In principle, the method is the same as for binary stars. However, because the orbital period of a binary galaxy is about 10^9 years, only statistical information can

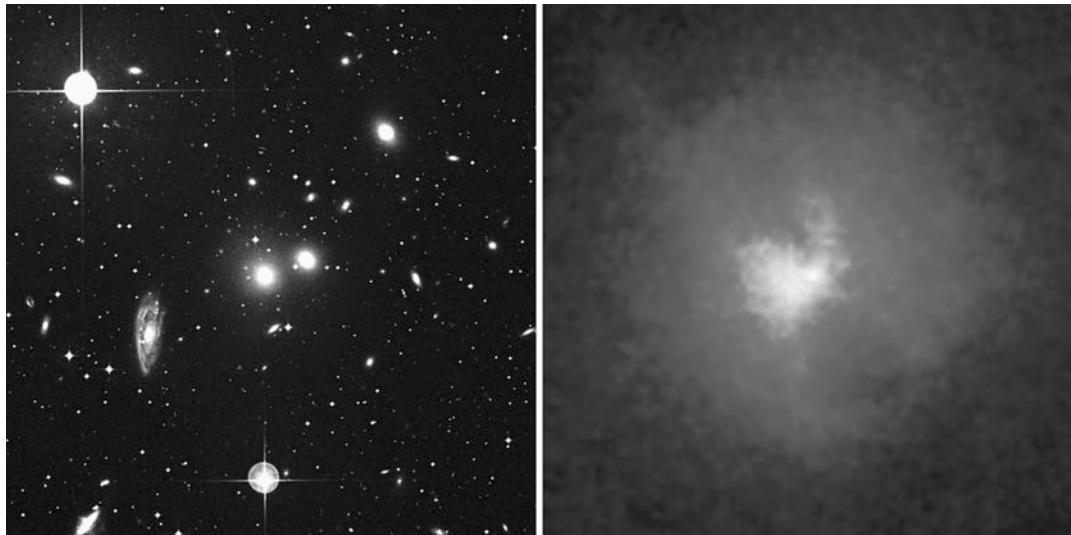


Fig. 19.12 Stars and hot gas in the Hydra galaxy cluster. *Left:* Optical image of the centre of the cluster. *Right:* An X-ray image of the same area in the same scale by

the Chandra satellite. The temperature of the gas is about 4×10^7 K. (Optical image UK Schmidt/DSS1/ESO, SRC, X-ray image NASA/CXC/SAO)

be obtained in this way. The results are still uncertain, but seem to indicate values of $M/L = 20\text{--}30$ at pair separations of about 50 kpc.

A fourth method to determine galaxy masses is to apply the virial theorem to clusters of galaxies, assuming that these are in equilibrium. The kinetic energy T in (19.4) can then be calculated from the observed redshifts and the potential energy U , from the separations between cluster galaxies. If it is assumed that the masses of galaxies are proportional to their luminosities, it is found that M/L is about 200 within 1 Mpc of the cluster centre. However, there is a large variation from cluster to cluster.

There are also two newer and more accurate methods for mass determination. The fifth method utilises the X-ray radiation coming from the hot gas around galaxies or galaxy clusters (Fig. 19.12). The temperature of the gas around galaxy clusters is as high as 10^7 , even 10^8 K. Since the gas seems to be more or less stable there must be a massive gravitational well preventing the gas from escaping. The dependence between the total mass of the galaxy cluster and the gas temperature is almost linear: the hotter the gas the more mass there is. The total mass of the richest galaxy

clusters is over 10^{15} solar masses, and only about 1 % of this is in the form of visible matter.

Gravitational lenses (Sect. 19.7) are the most accurate method. A light ray passing by a galaxy is bent and the amount of bending depends on the mass of the galaxy. The method is generally used to determine masses of galaxy clusters but it can also be used for individual galaxies if the lensing galaxy forms several images of the background object or clearly distorts its shape. Gravitational lenses have finally confirmed that only about one sixth of the mass of galaxies is ordinary matter (like stars or gas) and the rest dark matter.

Dark Matter the most precise value for the dark matter has been obtained by the WMAP and Planck satellites which have studied the details of the cosmic background radiation. As it will be mentioned in Chap. 20, about 31 % of the total content (mass + energy) of the universe is matter. The final results of the Planck satellite published in 2015 show that the contribution of ordinary matter is 5 % and of cold dark matter 26 %.

The distribution of the dark matter can nowadays be determined by weak and strong gravitational lenses. The result is that the dark matter



Fig. 19.13 Distribution of dark matter around the galaxy cluster MACS J0717.5+3745. The background image was taken by the Hubble telescope. Superimposed is the distribution of dark matter calculated using strong and weak gravitational lenses. (NASA, ESA, Harald Ebeling (University of Hawaii at Manoa) & Jean-Paul Kneib (LAM))

appears as wide halos around galaxies and galaxy clusters (Fig. 19.13).

The nature of the dark matter is not yet known. It interacts with baryons almost only by its gravity. It is known to be “cold”, which means that the velocities of the particles are low unlike with neutrinos that always travel almost at the speed of light.

It is assumed that the dark matter consists of unknown elementary particles. Possible candidates are light axions or heavy particles predicted by supersymmetry theories (WIMP, weakly interacting massive particle) like neutralinos. Their mass might be of the order 100 GeV–1 TeV, or hundreds of proton masses. Underground scintillation detectors and the LHC accelerator at the CERN are used to search for particles of the dark matter.

19.3 Galactic Structures

Ellipticals and Bulges In all galaxies the oldest stars have a more or less round distribution. In the Milky Way this component is represented by the population II stars. Its inner parts are called the

bution of dark matter calculated using strong and weak gravitational lenses. (NASA, ESA, Harald Ebeling (University of Hawaii at Manoa) & Jean-Paul Kneib (LAM))

bulge, and its outer parts are often referred to as the halo. There does not appear to be any physically significant difference between the bulge and the halo. The population of old stars can be best studied in ellipticals, which only contain this component. The bulges of spiral and S0 galaxies are very similar to ellipticals of the same size.

The surface brightness distribution in elliptical galaxies essentially depends only on the distance from the centre and the orientation of the major and minor axis. If r is the radius along the major axis, the surface brightness $I(r)$ is well described by *de Vaucouleurs' law*:

$$\log \frac{I(r)}{I_e} = -3.33 \left[\left(\frac{r}{r_e} \right)^{1/4} - 1 \right]. \quad (19.8)$$

The constants in (19.8) have been chosen so that half of the total light of the galaxy is radiated from within the radius r_e and the surface brightness at that radius is I_e . The parameters r_e and I_e are determined by fitting (19.8) to observed brightness profiles. Typical values for elliptical, normal spiral and S0 galaxies are in the ranges $r_e = 1\text{--}10$ kpc and I_e corresponds to 20–23 magnitudes per square arc second.

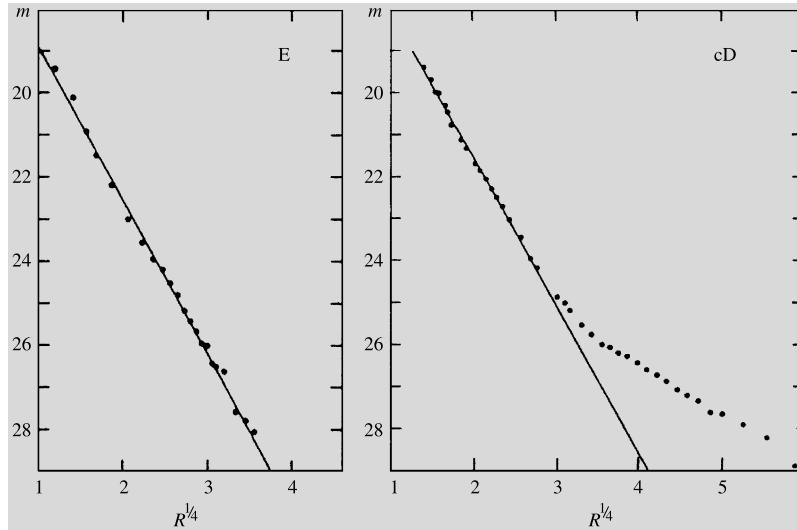
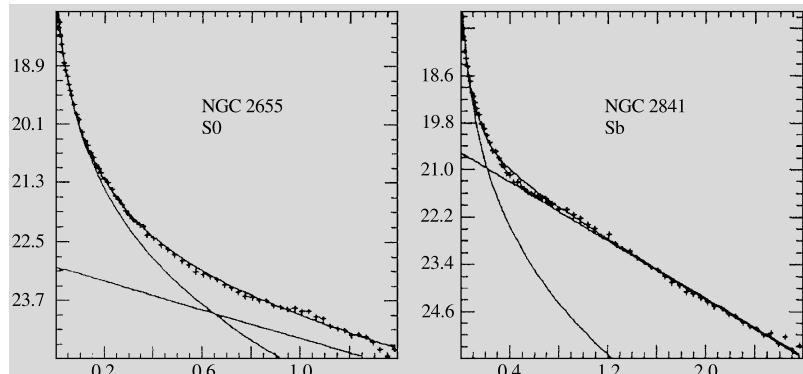


Fig. 19.14 The distribution of surface brightness in E and cD galaxies. *Ordinate:* surface magnitude, mag/sq.arcsec; *abscissa:* $(\text{radius [kpc]})^{1/4}$. Equation (18.8) corresponds to a straight line in this representation. It fits well with an E galaxy, but for type cD the luminosity falls off more

slowly in the outer regions. Comparison with Fig. 19.15 shows that the brightness distribution in S0 galaxies behaves in a similar fashion. cD galaxies have often been erroneously classified as S0. (Thuan, T.X., Romanishin, W. (1981): *Astrophys. J.* **248**, 439)

Fig. 19.15 The distribution of surface brightness in types S0 and Sb. *Ordinate:* mag/sq.arc sec; *abscissa:* radius [arc sec]. The observed surface brightness has been decomposed into a sum of bulge and disc contributions. Note the larger disc component in type Sb. (Boroson, T. (1981): *Astrophys. J. Suppl.* **46**, 177)



Although de Vaucouleurs' law is a purely empirical relation, it still gives a remarkably good representation of the observed light distribution. However, in the outer regions of elliptical galaxies, departures may often occur: the surface brightness of dwarf spheroidals often falls off more rapidly than (19.8), perhaps because the outer parts of these galaxies have been torn off in tidal encounters with other galaxies. In the giant galaxies of type cD, the surface brightness falls off more slowly (see Fig. 19.14). Such galaxies are usually at the centre of dense galaxy

clusters, and during billions of years they have absorbed several smaller galaxies. The stars of the disrupted galaxies usually remain orbiting the giant galaxy in its outskirts increasing the brightness of the outer parts. cD galaxies have been called the cannibals of the world of galaxies.

Although the isophotes in elliptical galaxies are ellipses to a good approximation, their ellipticities and the orientation of their major axes may vary as a function of radius. Different galaxies differ widely in this respect, indicating that

the structure of ellipticals is not as simple as it might appear. In particular, the fact that the direction of the major axis sometimes changes within a galaxy suggests that some ellipticals may not be axially symmetric in shape. A probable reason is that many galaxies of different shapes rotating in different directions have assimilated to the current galaxies.

From the distribution of surface brightness, the three-dimensional structure of a galaxy may be inferred as explained in Box 19.1. The relation (19.8) gives a brightness profile which is very strongly peaked towards the centre. The real distribution of axial ratios for ellipticals can be statistically inferred from the observed one. On the (questionable) assumption that they are rotationally symmetric, one obtains a broad distribution with a maximum corresponding to types E3–E4. If the true shape is not axisymmetric, it cannot even statistically be uniquely determined from the observations.

Disk A bright, massive stellar disk is characteristic for S0 and spiral galaxies, which are therefore called *disk galaxies*. There are indications that in some ellipticals there is also a faint disk hidden behind the bright bulge. In the Milky Way the disk is formed by population I stars.

The distribution of surface brightness in the disk is described by the expression

$$I(r) = I_0 e^{-r/r_0}. \quad (19.9)$$

Figure 19.11 shows how the observed radial brightness distribution can be decomposed into a sum of two components: a centrally dominant bulge and a disk contributing significantly at larger radii. The central surface brightness I_0 typically corresponds to 21–22 mag./sq.arcsec, and the radial scale length $r_0 = 1\text{--}5$ kpc. In Sc galaxies the total brightness of the bulge is generally slightly smaller than that of the disk, whereas in earlier Hubble types the bulge has a larger total brightness. The thickness of the disk, measured in galaxies that are seen edge-on, may typically be about 1.2 kpc. Sometimes the disk has a sharp outer edge at about $4r_0$.

The Interstellar Medium Elliptical and S0 galaxies contain very little interstellar gas. However, in some ellipticals neutral hydrogen amounting to about 0.1 % of the total mass has been detected, and in the same galaxies there are also often signs of recent star formation. In some S0 galaxies much larger gas masses have been observed, but the relative amount of gas is very variable from one galaxy to another. The lack of gas in these galaxies is rather unexpected, since during their evolution the stars release much more gas than is observed.

The relative amount of neutral hydrogen in spiral galaxies is correlated with their Hubble type. Thus Sa spirals contain about 2 %, Sc spirals 10 %, and Irr I galaxies up to 30 % or more.

The distribution of neutral atomic hydrogen has been mapped in detail in nearby galaxies by means of radio observations. In the inner parts of galaxies the gas forms a thin disk with a fairly constant thickness of about 200 pc, sometimes with a central hole of a few kpc diameter. The gas disk may continue far outside the optical disk, becoming thicker and often warped from the central disk plane.

Most of the interstellar gas in spiral galaxies is in the form of molecular hydrogen. The hydrogen molecule cannot be observed directly, but the distribution of carbon monoxide has been mapped by radio observations. The distribution of molecular hydrogen can then be derived by assuming that the ratio between the densities of CO and H₂ is everywhere the same, although this may not always be true. It is found that the distribution obeys a similar exponential law as the young stars and H II regions, although in some galaxies (such as the Milky Way) there is a central density minimum. The surface density of molecular gas may be five times larger than that of HI, but because of its strong central concentration its total mass is only perhaps two times larger.

The distribution of cosmic rays and magnetic fields in galaxies can be mapped by means of radio observations of the synchrotron radiation from relativistic electrons. The strength of the magnetic field deduced in this way is typically 0.5–1 nT. The observed emission is polarised,



Fig. 19.16 A spiral galaxy NGC 5907 seen sideways. The galaxy is surrounded by star streams that are remnants of small dwarf galaxies merged to the big galaxy. (Photo Gabany, Martinez-Delgado et al., 2010, AJ 140, 962)

showing that the magnetic field is fairly well-ordered on large scales. Since the plane of polarisation is perpendicular to the magnetic field, the large-scale structure of the magnetic field can be mapped. However, the plane of polarisation is changed by Faraday rotation, and for this reason observations at several wavelengths are needed in order to determine the direction of the field. The results show that the field is generally strongest in the plane of the disk, and is directed along the spiral arms in the plane. The field is thought to have been produced by the combined action of rising elements of gas, perhaps produced by supernova explosions, and the differential rotation, in principle in the same way as the production of solar magnetic fields was explained in Chap. 13.

The Outer Parts of Galaxies Even the outermost parts outside the visible disk are not empty.

Recently they have been seen to contain several different kinds of matter. The most abundant is the dark matter extending around the galaxy as a sphere or slightly flattened ellipsoid to a distance of tens of times further than the visible matter.

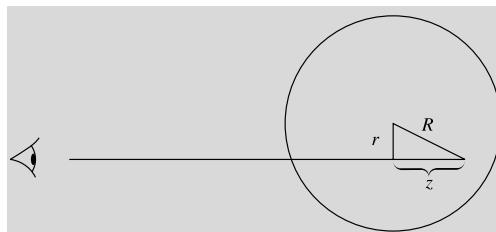
Hot gas is the second largest component. Its high temperature and great abundance were surprising. For example, the Milky Way is surrounded by a region of hot gas extending as a spherical cloud as far as to 150 kiloparsecs. The temperature of the gas is $1\text{--}2.5 \times 10^6$ K and the total mass about the same order of magnitude as the total mass of the stars in the Milky Way.

The third component comprises of separate gas clouds having greater velocities than other matter of the galaxy. Some of the clouds come from the centre of the Milky Way, some from other galaxies.

Yet another component consists of star clusters and star streams. In addition to globular clusters

galaxies are accompanied by star concentrations that are remnants of dwarf galaxies assimilated to the galaxy. For example, the globular cluster M54 is not an original Milky Way cluster but a remnant of the Sagittarius dwarf ellipsoid moving through the central part of the Milky Way. Disrupting dwarf galaxies are sometimes seen as star streams around the galaxy (Fig. 19.16).

Box 19.1 (Three-Dimensional Shape of Galaxies) Equations (19.8) and (19.9) describe the distribution of galactic light projected on the plane of the sky. The actual three-dimensional luminosity distribution in a galaxy is obtained by inverting the projection. This is easiest for spherical galaxies.



Let us suppose that a spherical galaxy has the projected luminosity distribution $I(r)$ (e.g. as in (19.8)). With coordinates chosen according to the figure, $I(r)$ is given in terms of the three-dimensional luminosity distribution $\rho(R)$ by

$$I(r) = \int_{-\infty}^{\infty} \rho(R) dz.$$

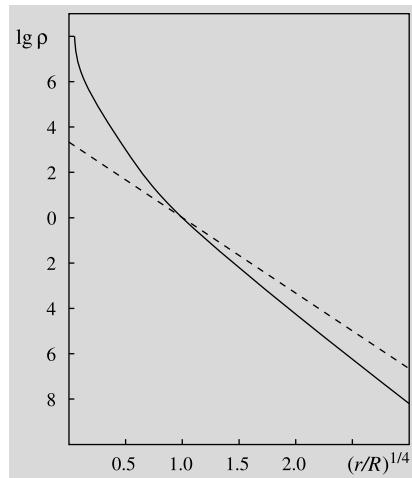
Since $z^2 = R^2 - r^2$, a change of the variable of integration yields

$$I(r) = 2 \int_r^{\infty} \frac{\rho(R) R dR}{\sqrt{R^2 - r^2}}.$$

This is known as an Abel integral equation for $\rho(R)$, and has the solution

$$\begin{aligned} \rho(R) &= -\frac{1}{\pi R} \frac{d}{dR} \int_R^{\infty} \frac{I(r) r dr}{\sqrt{r^2 - R^2}} \\ &= -\frac{1}{\pi} \int_R^{\infty} \frac{(dI/dr) dr}{\sqrt{r^2 - R^2}}. \end{aligned}$$

Introducing the observed $I(r)$ into this expression, one obtains the actual luminosity distribution $\rho(R)$. In the figure the solid curve shows the three-dimensional luminosity distribution obtained from the Vaucouleurs' law (the dashed line).



If the galaxy is not spherical, its three-dimensional shape can only be determined if its inclination with respect to the line of sight is known. Since galactic disks are thin and of constant thickness, the inclination i of a disk galaxy is obtained directly from the axis ratio of its projected image: $\sin i = b/a$.

When the inclination is known, the real axis ratio of the bulge q_0 can be determined from the projected value q . For a rotationally symmetric bulge the relation between q and q_0 is

$$\cos^2 i = \frac{1 - q^2}{1 - q_0^2}.$$

The flattenings of disk galaxy bulges obtained from this relation lie in the range $q_0 = 0.3 - 0.5$. Since the inclinations of ellipticals are generally unknown, only the statistical distribution of q can be determined from that of q_0 .

19.4 Dynamics of Galaxies

We have seen how the masses of galaxies can be derived from observed velocities of stars and gas.

The same observations can be used to study the internal distribution of mass in more detail.

Slowly Rotating Systems The dynamics of elliptical galaxies and disk galaxy bulges are studied by means of the Doppler shifts and broadenings of stellar absorption lines. Since a given absorption line is the sum of contributions from many individual stars, its Doppler shift gives their mean velocity, while its broadening is increased by an amount depending on the dispersion of stellar velocities around the mean. By observing how the wavelengths and widths of spectral lines behave as functions of the radius, one can get some insight into the distribution of mass in the galaxy.

Examples of the observed radial dependence of the rotational velocity and velocity dispersion derived for some ellipticals were given in Fig. 19.10. The observed rotational velocities are often small ($<100 \text{ km s}^{-1}$), while the velocity dispersion may typically be about 200 km s^{-1} . If elliptical galaxies were in fact ellipsoids of revolution, there should be a statistical relation (when projection effects have been taken into account) between flatness, rotational velocity and velocity dispersion. Such a relationship has been observed for fainter ellipticals and for disk galaxy bulges. However, some of the brightest ellipticals rotate very slowly. Therefore their flattening cannot be due to rotation.

The radial dependence of the velocity dispersion gives information on the distribution of mass within the galaxy. Since it also depends on how the shapes of stellar orbits in the galaxy are distributed, its interpretation requires detailed dynamical models.

Some ellipticals have different velocity systems. An extreme case is NGC 4550 classified as E7/S0. Half of its stars orbit the centre in one direction and the rest in the opposite direction. This indicates that the galaxy is a merger of two spiral galaxies that originally rotated in opposite directions. Mutual distances of stars in ellipticals are so large that collisions do not occur and the current system is very stable.

Rotation Curves In spiral galaxies the distribution of mass can be studied directly using the

observed rotational velocities of the interstellar gas. This can be observed either at optical wavelengths from the emission lines of ionised gas in HII regions or at radio wavelengths from the hydrogen 21 cm line. Typical galactic rotation curves were shown in Fig. 19.11.

The qualitative behaviour of the rotation curve in all spiral galaxies is similar to the rotation curve of the Milky Way: there is a central portion, where the rotational velocity is directly proportional to the radius, corresponding to rigid body rotation. At a few kpc radius the curve turns over and becomes flat, i.e. the rotational velocity does not depend on the radius. In early Hubble types, the rotation curve rises more steeply near the centre and reaches larger velocities in the flat region (Sa about 300 km s^{-1} , Sc about 200 km s^{-1}). A higher rotational velocity indicates a larger mass according to (19.7), and thus Sa types must have a larger mass density near the centre. This is not unexpected, since a more massive bulge is one of the defining properties of early type spirals.

A decrease of the rotational velocity at large radii would be an indication that most of the mass is inside that radius. In some galaxies such a decrease has been detected, in others the rotational velocity remains constant as far out as the observations can reach.

Spiral Structure Spiral galaxies (Fig. 19.17) are relatively bright objects. Some have a well-defined, large-scale two-armed spiral pattern, whereas in others the spiral structure is made up of a large number of short filamentary arms. From galaxies where the pattern is seen in front of the central bulge, it has been deduced that the sense of winding of the spiral is trailing with respect to the rotation of the galaxy. However, the rule is not absolute: there are also galaxies with leading arms.

The spiral structure is most clearly seen in the interstellar dust, HII regions, and the OB associations formed by young stars. The dust often forms thin lanes along the inner edge of the spiral arms, with star forming regions on their outside. Enhanced synchrotron radio emission associated with spiral arms has also been detected.

Fig. 19.17 *Above:*

A spiral galaxy from above: M51 (type Sc). The interacting companion is NGC 5195 (type Irr II). (Lick Observatory). *Below:* A spiral galaxy from the side: the Sb spiral NGC 4565. (NOAO/Kitt Peak National Observatory)



A well formed wide spiral pattern is relatively common, and hence it must be a long lasting phenomenon. The spiral pattern is generally thought to be a wave in the density of the stellar disk, as discussed in Sect. 18.4. As the interstellar gas streams through the density wave a shock, marked by the dust lanes, is formed as the interstellar gas is compressed, leading to the col-

lapse of molecular clouds and the formation of stars. The density wave theory predicts characteristic streaming motions within the arm, which have been detected in some galaxies by observations of the HI 21 cm line.

There may be several reasons producing spiral arms. The most common is an external perturbation when another galaxy passes by at a close dis-

tance creating a wave in the galaxy. Other sources of perturbation could be the bar of the galaxy or the gravitational field of a temporary “traffic jam” of stars.

A single close encounter or a traffic jam can explain temporary spiral arms but not more permanent features. The differential rotation of the galaxy tends to wipe out the arms during a few revolutions or in a few hundreds of millions of years. Some renewal mechanism is needed to keep the arms permanent.

Standing waves can explain the permanence. Birth of spiral arms and their motion generate density waves travelling through the whole galaxy. The waves can be reflected or refracted by the central bulge. When wave fronts moving in different directions meet they can either damp or amplify each others. Spiral arms are standing waves, fronts of amplified waves, seen as concentrates of stars and gas clouds. The spiral forms move around the galaxy at constant speed.

19.5 Stellar Ages and Element Abundances in Galaxies

From the Milky Way we know that stars of populations I and II are different not only in respect to their spatial distribution, but also in respect to their ages and heavy element abundances. This fact gives important evidence about the formation of the Milky Way, and it is therefore of interest if a similar connection can be found in other galaxies.

The indicators of composition most easily measured are the variations of colour indices inside galaxies and between different galaxies. Two regularities have been discovered in these variations: First, according to the *colour-luminosity relation* for elliptical and S0 galaxies, brighter galaxies are redder. Secondly, there is a *colour-aperture effect*, so that the central parts of galaxies are redder. For spirals this relationship is due to the presence of young, massive stars in the disk, but it has also been observed for elliptical and S0 galaxies.

Galactic spectra are composed of the spectra of all their stars added together. Thus the colours depend both on the ages of the stars (young stars

are bluer) and on the heavy element abundance Z (stars with larger Z are redder). The interpretation of the observational results thus has to be based on detailed modelling of the stellar composition of galaxies or *population synthesis*.

Stars of different spectral classes contribute different characteristic absorption features to the galaxy spectrum. By observing the strength of various spectral features, one can find out about the masses, ages and chemical composition of the stars that make up the galaxy. For this purpose, a large number of characteristic properties of the spectrum, strengths of absorption lines and broad-band colours are measured. One then attempts to reproduce these data, using a representative collection of stellar spectra. If no satisfactory solution can be found, more stars have to be added to the model. The final result is a population model, giving the stellar composition of the galaxy. Combining this with theoretical stellar evolution calculations, the evolution of the light of the galaxy can also be computed.

Population synthesis of E galaxies show that practically all their stars were formed simultaneously about $13-14 \times 10^9$ years ago, which is close to the age of the universe. Most of their light comes from red giants, whereas most of their mass resides in lower main sequence stars of less than one solar mass.

Since all stars have roughly the same age, the colours of elliptical galaxies are directly related to their metallicities. Thus the colour-luminosity relation indicates that Z in giant ellipticals may be double that in the solar neighbourhood, while it may be smaller by a factor 100 in dwarfs. Similarly, the radial dependence of the colours can be explained if the value of Z at the centre is an order of magnitude larger than it is at larger radii.

The stellar composition of disk galaxy bulges is generally similar to that of ellipticals. The element abundances in the gas in spirals can be studied by means of the emission lines from H II regions ionised by newly formed stars. In this case too, the metallicity increases towards the centre.

The birth histories of galaxies explain partly the different metallicities and age distributions. As we will see later, galaxies collide and merge in the span of billions of years. Most clearly this

is seen in dwarf galaxies having few stars that even a single merger leaves a distinct signature. Figure 19.18 shows the age distribution of two Milky Way's neighbour galaxies. A few of the stars of Leo A were born immediately after the origin of the universe but most of the stars were born during the last 5–6 billion years. The stars of the Carina dwarf galaxy were born during three separate periods, at the age of 1–2, 6–8 and about 11 billion years.

19.6 Systems of Galaxies

The galaxies are not smoothly distributed in space; rather, they form systems of all sizes: galaxy pairs, small *groups*, large *clusters* and *superclusters* formed from several groups and clusters. The larger a given system, the less its density exceeds the mean density of the Universe. On the average, the density is twice the background density for systems of radius 5 Mpc and 10 % above the background at radius 20 Mpc.

Interactions of Galaxies Galaxies do not evolve in isolation; instead, they interact continuously with other galaxies. Each large galaxy has experienced encounters or collisions with several other large galaxies and tens or even hundreds of encounters with much smaller galaxies.

Among the most common distortions are “bridges” and “tails” caused by tidal forces during an encounter (Fig. 19.19). Sometimes a small galaxy breaks down and forms a ring in the equatorial plane of the larger galaxy or sometimes on a polar orbit. The most striking ring galaxies are born when a dense galaxy passes straight through another one (Fig. 19.20). In many cases the effects of collisions are seen as starbursts when the gas and dust clouds of the galaxies meet, condense and start to produce new stars. Sometimes image processing is needed to show faint but sharp rings around a galaxy (Fig. 19.21).

The interactions between galaxies are not always dramatic. For example, the Milky Way has two satellites, the Large and Small Magellanic Clouds (see Fig. 19.5), which are Irr I type dwarf galaxies at about 60 kpc distance. It is thought that approximately 5×10^8 years

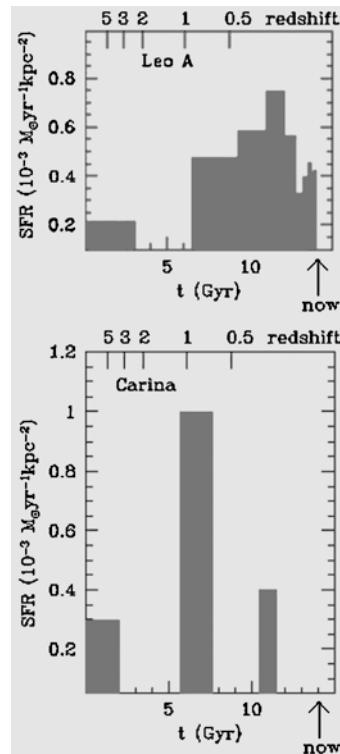


Fig. 19.18 The age distribution of the dwarf galaxies in Leo and Carina shows that the stars of the galaxies were formed in separate collision episodes. (Tolstoy, Hill, Tosi, 2009, Ann. Rev. Astron. Astrophys. 47, 371)

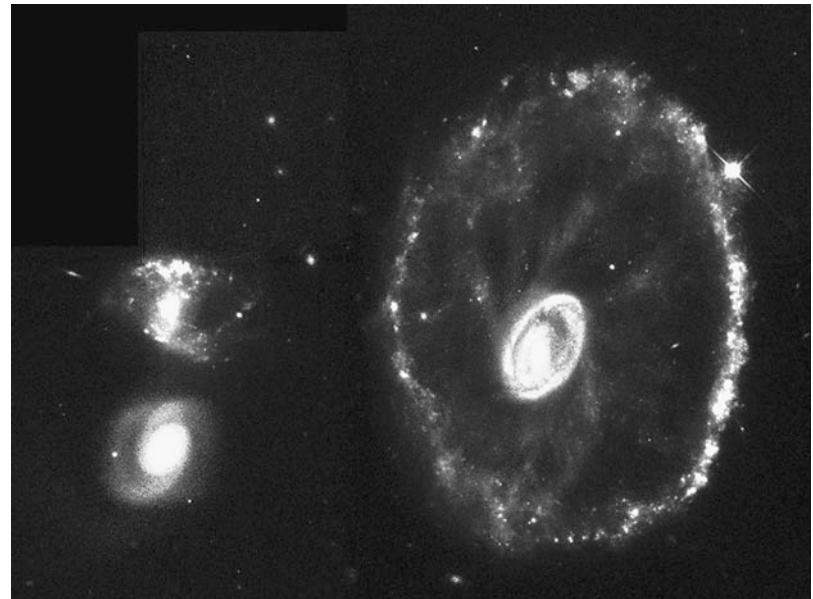
ago, these passed the Milky Way at a distance of about 10–15 kpc, leaving behind the *Magellanic Stream*, a 180° long thin stream of neutral hydrogen clouds. Systems of this type, where a giant galaxy is surrounded by a few small companions, are quite common. Computations show that in many such cases the tidal interactions are so strong that the companions will merge with the parent galaxy at the next close approach. This is likely to happen to the Magellanic Clouds.

During earlier epochs in the Universe, when the density was larger, interactions between galaxies must have been much more common than at present. Thus it has been proposed that a large fraction of bright galaxies have undergone major mergers at some stage in their history. In particular, there are good reasons to believe that the slowly rotating, non-axisymmetric giant ellipticals may have formed by the merger of disk galaxies.

Fig. 19.19 The galaxy pair NGC 4676 is known as the Mice. The galaxies are about to collide and a long gas and star tail protrudes from each of them.
 (Hubble/NASA, H. Ford (JHU), G. Illingworth (UCSC/LO), M. Clampin (STScI), G. Hartig (STScI), the ACS Science Team, ja ESA)



Fig. 19.20 The Cartwheel is the best known ring galaxy. It was born when a small galaxy passed through the centre of the big galaxy. (Photo Hubble, K. Borne (StScI) and Nasa)



Groups The most common type of galaxy systems are small, irregular groups of a few tens of galaxies. A typical example is the *Local Group*, which contains two larger galaxies in addition to the Milky Way—the Andromeda Galaxy M31, an Sb spiral of about the same size as the Milky Way with two dwarf companions, and the smaller Sc spiral M33. The rest of the about 50 members

of the Local Group are dwarfs; about 20 are of type dE and 10 of type Irr I. The diameter of the Local Group is about 1.2 Mpc.

Clusters A system of galaxies may be defined to be a cluster if it contains a larger number (at least 50) of bright galaxies. The number of members and the size of a cluster depend on how they are defined. One way of doing this is to fit the

Fig. 19.21 Rings of stars with sharp boundaries in the outer parts of the galaxy NGC 474. The rings were born when a lenticular galaxy tore a rotating spiral galaxy layer by layer.
(CFHT/Coelum—J.-C. Cuillandre & G. Anselmi)



observed distribution of galaxies within a cluster with an expression of the form (19.8). In this way a characteristic cluster radius of about 2–5 Mpc is obtained. The number of members depends both on the cluster radius and on the limiting magnitude. A large cluster may contain several hundred galaxies that are less than two magnitudes fainter than the characteristic luminosity L^* of (19.2).

Clusters of galaxies can be ordered in a sequence from extended, low-density, irregular systems (sometimes called clouds of galaxies) to denser and more regular structures (Fig. 19.23). The galaxy type composition also varies along this sequence in the sense that in the loose irregular clusters, the bright galaxies are predominantly spirals, whereas the members of dense clusters are almost exclusively type E and S0. The nearest cluster of galaxies is the Virgo Cluster at a distance of about 15 Mpc. It is a relatively irregular cluster, where a denser central region containing early galaxy types is surrounded by a more extended distribution of mainly spiral galaxies. The nearest regular cluster is the Coma Cluster,

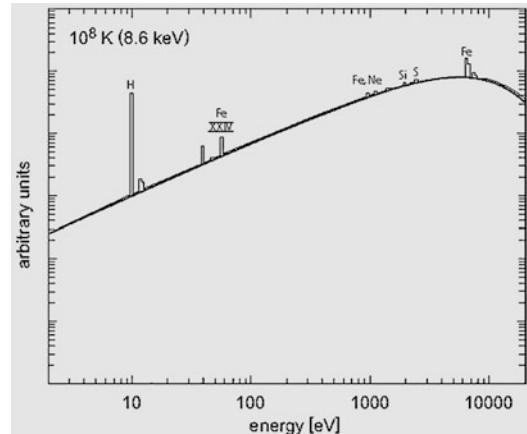


Fig. 19.22 A theoretical X-ray spectrum for plasma with a temperature of 10^8 K. Most of the radiation comes from the bremsstrahlung of electrons, peaking at 8.6 kilo-electronvolts (*thick line*). Superimposed are spectral lines produced by different ions. (Böhringer & Hensler, 1989, A&A 215, 147)

roughly 90 Mpc away. In the Coma Cluster a central pair of giant ellipticals is surrounded by a flat-

Fig. 19.23 *Above:* The irregular Virgo Cluster of galaxies. *Below:* The regular Coma Cluster (ESO and Karl-Schwarzschild-Observatorium)



tened (axis ratio about 2 : 1) system of early type galaxies.

Earlier galaxy groups were studied as conglomerates of visible galaxies. Recently two other, more massive, components have been included.

First of all, galaxy clusters are surrounded by large clouds of hot gas, as mentioned in Sect. 19.2. The gas can be observed by its X-

ray emission. The spectrum shows features typical for bremsstrahlung of electrons and lines of totally ionised metals (like FeXXIV). Temperature measured from X-ray spectra are 10^7 – 10^8 K. The total mass of the gas in the poorest clusters is about 5 times and in the richest clusters even 20 times as big as the mass inside the galaxies. Part of the gas comes from the galaxies (jets of

quasars, supernova explosions), but most of it is hydrogen and helium from the big bang.

Besides in X-ray images the existence of the hot gas is seen also as extra spots in the cosmic background radiation, caused by the *Sunyayev-Zel'dovich effect*. Photons of the background radiation hit electrons of the hot gas and receive some extra energy (inverse Compton scattering). Below 218 GHz the radiation is slightly cooler and above 218 GHz a little brighter. The diameter of the spot is about one arc minute and the temperature difference from the surroundings 0.1–1 millikelvins (Fig. 19.25). One of the main objectives of the Planck satellite was to find gas clouds around galaxy clusters using this phenomenon. The final catalogue of Planck (2015) contains over 1600 objects most of which are previously unknown galaxies with redshifts $z = 0 – 1.5$.

Without a strong source of gravitation the hot gas would quickly disperse into space. The gravitation is caused by dark matter; its mass is about five times as big as that of the ordinary matter (hot gas + visible galaxies) in the cluster. Gravitational lenses have been used to investigate the dark matter in galaxy clusters (Sect. 19.7).

Superclusters Groups and clusters of galaxies may form even larger systems, superclusters. For example, the Local Group belongs to the *Local Supercluster*, a flattened system whose centre is the Virgo Cluster, containing tens of smaller groups and clouds of galaxies. The Coma Cluster is part of another supercluster. The diameters of superclusters are 10–20 Mpc. However, on this scale, it is no longer clear whether one can reasonably speak of individual systems. Perhaps it would be more accurate to think of the distribution of galaxies as a continuous network, where the large clusters are connected by walls and strings formed by smaller systems. Between these there remain empty regions containing very few galaxies, which can be up to 50 Mpc in diameter (Figs. 19.24, 20.8).

19.7 Active Galaxies and Quasars

So far in this chapter we have been concerned with the properties of normal galaxies. In some

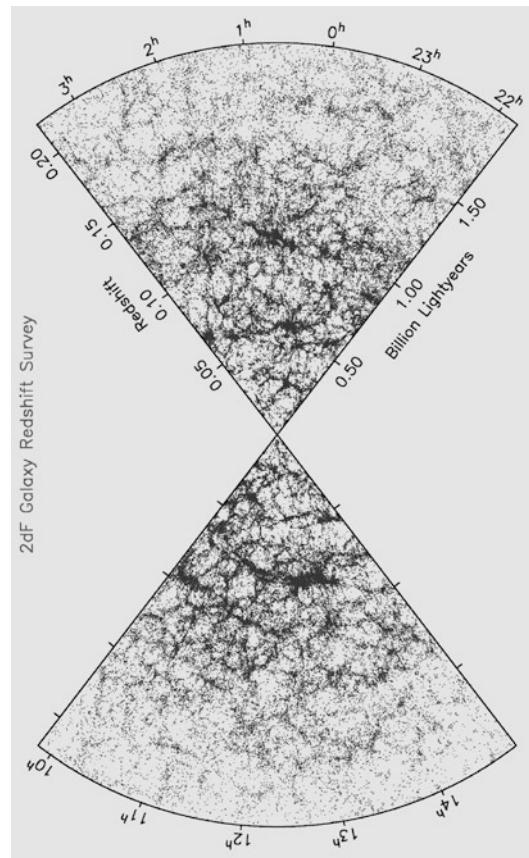


Fig. 19.24 The large-scale space distribution of 245,591 galaxies. The radial co-ordinate is the redshift, which can be translated into a distance using a value for the Hubble constant. The thickness of the slices is about 10° . (2dFGRS Team, <http://www2.ao.au/~TDFgg/>)

galaxies, however, the normal galaxy is overshadowed by violent activity. This activity is produced in the nucleus, which is then called an *active galactic nucleus* (AGN).

The luminosities of active galactic nuclei may be extremely large, sometimes much larger than that of the rest of the galaxy. It seems unlikely that a galaxy could maintain such a large power output for long. For this reason it is thought that active galaxies do not form a separate class of galaxies, but rather represent a passing stage in the evolution of normal galaxies.

Activity appears in many different forms. Some galaxies have an exceptionally bright nucleus similar to a large region of ionised hydrogen. These may be young galaxies, where near

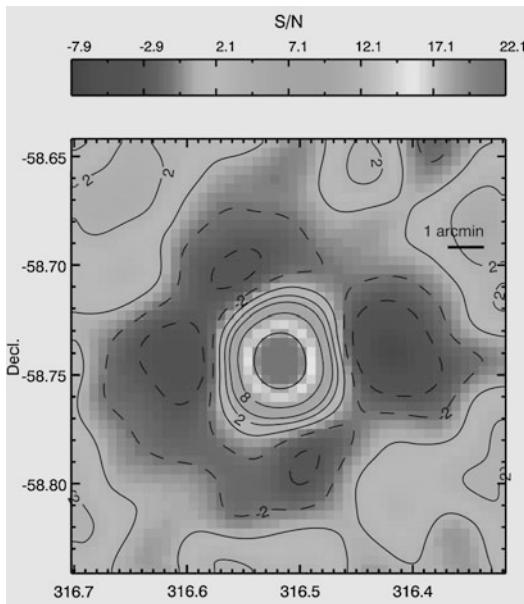


Fig. 19.25 A brightening of the cosmic background radiation caused by the Sunyaev-Zel'dovitch effect. The radio observation in the millimetre band were made with the South pole telescope. The amplitude of the brightening tells that the temperature of the gas cloud is 1.2×10^8 K and the total mass of the galaxy cluster about 10^{15} solar masses. Optical images of the same location show a distant galaxy cluster SPT-CL J2106-5844, which has a redshift $z = 1.13$. (Foley, Andersson, Bazin, de Haan, Ruel et al., 2011, ApJ 731, p. 89)

the centre large numbers of stars are forming and evolving into supernovae (starburst nuclei). In other nuclei the radiation cannot have been produced by stars, and the most plausible source of energy in these nuclei is the gravitational energy of a supermassive black hole (mass $> 10^8 M_\odot$). In some galaxies, the spectral lines are unusually broad, indicating large internal velocities. These may be either rotational velocities near a black hole or due to explosive events in the nucleus. In some galaxies, jets are seen coming out of the nucleus. Many active galaxies radiate a nonthermal spectrum, apparently synchrotron radiation produced by fast electrons in a magnetic field.

The classification of active galaxies has been developed rather unsystematically, since many of them have been discovered only recently, and have not been completely studied. For example, the *Markarian galaxies* catalogued by Benyamin Yerishevich Markarian in the early 1970's are

defined by strong ultraviolet emission. Many Markarian galaxies are Seyfert galaxies; others are galaxies undergoing a burst of star formation. The N galaxies form another class closely similar to the Seyfert galaxies.

Two natural basic classes of active galaxies are the *Seyfert galaxies* and the *radio galaxies*. The former are spirals; the latter are ellipticals. Some astronomers think that the Seyfert galaxies represent the active stage of normal spiral galaxies and the radio galaxies that of ellipticals.

Seyfert Galaxies The Seyfert galaxies are named after *Carl Seyfert*, who discovered them in 1943. Their most important characteristics are a bright, pointlike central nucleus and a spectrum showing broad emission lines. The continuous spectrum has a nonthermal component, which is most prominent in the ultraviolet. The emission lines are thought to be produced in gas clouds moving close to the nucleus with large velocities.

On the basis of the spectrum, Seyfert galaxies are classified as type 1 or 2. In a type 1 spectrum, the allowed lines are broad (corresponding to a velocity of 10^4 km s^{-1}), much broader than the forbidden lines. In type 2, all lines are similar and narrower ($< 10^3 \text{ km s}^{-1}$). Transitions between these types and intermediate cases have sometimes been observed. The reason for the difference is thought to be that the allowed lines are formed in denser gas near the nucleus, and the forbidden lines in more diffuse gas further out. In type 2 Seyfert galaxies, the denser gas is missing or obscured.

Almost all Seyfert galaxies with known Hubble types are spirals; the possible exceptions are of type 2. They are strong infrared sources. Type 1 galaxies often show strong X-ray emission.

The true Seyfert galaxies are relatively weak radio sources. However, there are compact radio galaxies with an optical spectrum that is essentially the same as for Seyfert galaxies. These should probably be classified with the Seyfert galaxies. In general, the stronger radio emission seems to come with a type 2 spectrum.

It is estimated that about 1 % of all bright spiral galaxies are Seyfert galaxies. The luminosities of their nuclei are about $10^{36}\text{--}10^{39}$ W, of the

same order as all the rest of the galaxy. Brightness variations are common.

Radio Galaxies By definition, radio galaxies are galaxies that are powerful radio sources. The radio emission of a radio galaxy is non-thermal synchrotron radiation. The radio luminosity of radio galaxies is typically 10^{33} – 10^{38} W, and may thus be as large as the total luminosity of a normal galaxy. The main problem in explaining radio emission is to understand how the electrons and magnetic fields are produced, and above all, where the electrons get their energy.

The forms and sizes of the radio emitting regions of radio galaxies have been studied ever since the 1950's, when radio interferometers achieved the resolution of optical telescopes. The characteristic feature of a strong radio galaxy is a double structure: there are two large radio emitting regions on opposite sides of the observed galaxy. The radio emitting regions of some radio galaxies are as far apart as 6 Mpc, almost ten times the distance between the Milky Way and Andromeda galaxies. One of the smallest double radio sources is the galaxy M87 (Fig. 19.26), whose two components are only a few kpc distant from each other.

The double structure of radio galaxies appears to be produced by ejections from the nucleus. However, the electrons in the radio lobes cannot be coming from the centre of the galaxy, because they would lose all their energy during such a long transit. Therefore electrons have to be continuously accelerated within the radio-emitting regions. Within the radio lobes there are almost point-like regions, hot spots. These are generally symmetrically placed with respect to the nucleus, and are apparently consequences of nuclear ejections.

“Tailed” radio sources also exist. Their radio emission mainly comes from one side of the galaxy, forming a curved tail, which is often tens of times longer than the diameter of the galaxy. The best examples are NGC 1265 in the Perseus cluster of galaxies and 3C129, which appears to be in an elliptical orbit around a companion galaxy. The tail is interpreted as the trail left by the radio galaxy in intergalactic space.

Another special feature revealed by the radio maps is the presence of *jets*, narrow lines of radio emission, usually starting in the nucleus and stretching far outside the galaxy. The best known may be the M87 jet, which has also been observed as an optical and X-ray jet. The optically observed jet is surrounded by a radio source. A similar radio source is seen on the opposite side of the nucleus, where no optical jet is seen. Our nearest radio galaxy Centaurus A also has a jet extending from the nucleus to near the edge of the galaxy.

VLBI observations of radio jets have also revealed *superluminal motions*: in many compact sources the components appear to be separating faster than the speed of light. Since such velocities are impossible according to the theory of relativity, the observed velocities can only be apparent, and several models have been proposed to account for them.

Quasars The first quasar was discovered in 1963, when Maarten Schmidt interpreted the optical emission lines of the known radio source 3C273 as hydrogen Balmer lines redshifted by 16 %. Such large redshifts are the most remarkable characteristics of the quasars. Properly speaking, the word quasar is an abbreviation for quasistellar radio source, and some astronomers prefer to use the designation QSO (quasistellar object), since not all quasars emit radio radiation.

Optically the quasars appear almost as point sources, although improved observational techniques have revealed an increasing number of quasars located inside more or less normal galaxies (Fig. 19.27). Although the first quasars were discovered by radio observations, only a small fraction of all optically identified quasars are bright radio sources. Most radio quasars are point sources, but some have a double structure like the radio galaxies. Satellite X-ray pictures also show the quasars to be pointlike.

In the visible region the quasar spectra are dominated by spectral lines with rest wavelengths in the ultraviolet. The first observed quasar redshifts were $z = 0.16$ and 0.37 , and later searches have continued to turn up ever larger redshifts.

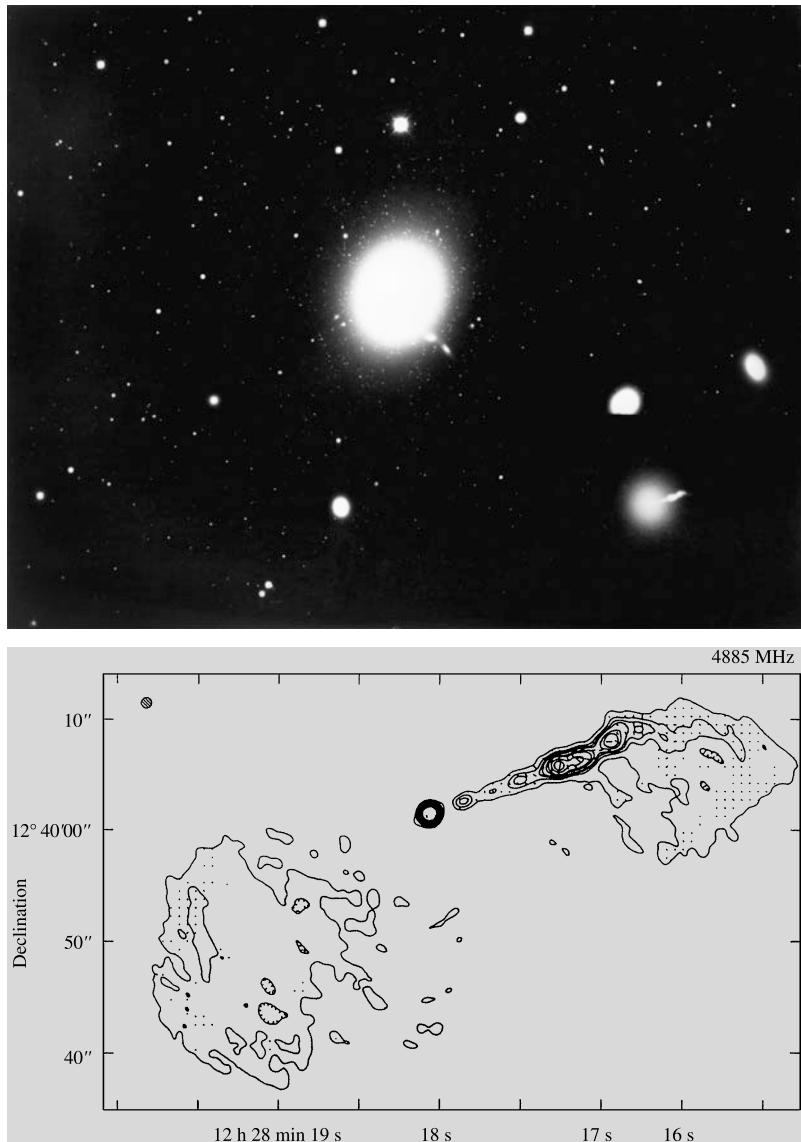


Fig. 19.26 *Above:* The active galaxy M87. In the lower right-hand corner a short exposure of the core region has been inserted (same scale as in the main photograph). One sees a blue jet coming out of the nucleus of a normal E0 galaxy (NOAO/Kitt Peak National Observatory).

Below: In the radio map made using the VLA the jet is observed to be two-sided. The area shown is much smaller than in the upper picture. (Owen, F.N., Hardee, P.E., Bignell, R.C. (1980): *Astrophys. J. (Lett.)* **239**, L11)

The present record is 8.7, but three objects have redshift 10. If these are quasars their light must have been emitted when the age of the universe was only 500 million years. The large inferred distances of the quasars mean that their luminosities have to be extremely large. Typical values lie in the range of 10^{38} – 10^{41} W. The brightness of

quasars may vary rapidly, within a few days or less. Thus the emitting region can be no larger than a few light-days, i.e. about 100 au.

The quasars often have both emission and absorption lines in their spectra. The emission lines are very broad and are probably produced in the quasar itself. Much of the absorption spec-

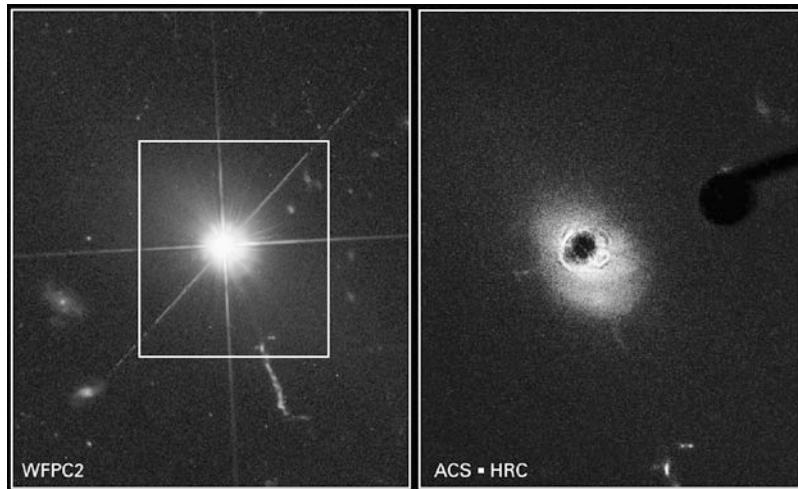


Fig. 19.27 One of the nearest quasars, 3C 273, photographed with two cameras aboard the Hubble Space Telescope. *On the left*, the Wide Field and Planetary Camera sees a bright point-like source, with a jet blasted out from the quasar (towards 5 o'clock). *On the right*, a corona-

graph in the Advanced Camera for Surveys blocks out the brightest parts of the quasar. Spiral arms in the host galaxy can be seen, with dark dust lanes, as well as new details in the path of the jet. (Photos Hubble/NASA/ESA)

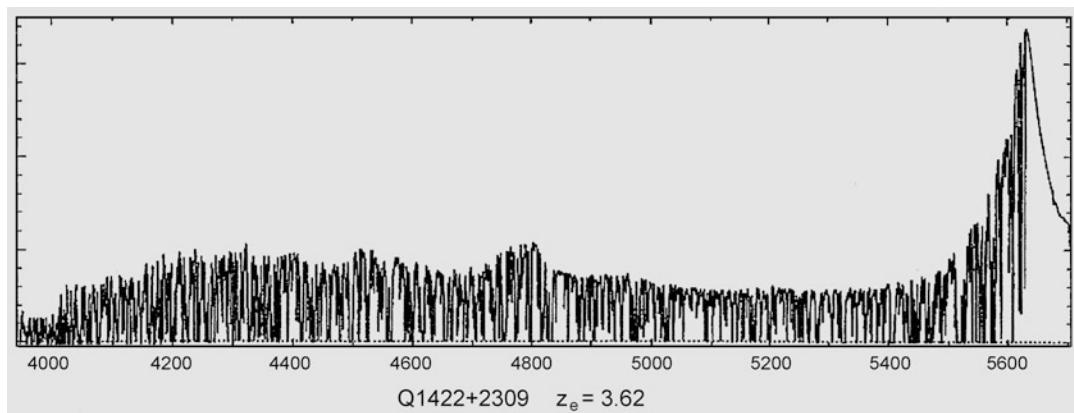


Fig. 19.28 A section of the spectrum of the quasar Q1422+2309 observed with the Keck 10 metre telescope. The Lyman α line of the quasar is near the right edge. Its redshift is 3.62. To the left of the line is a Lyman α for-

est, where each absorption line corresponds to a hydrogen cloud that the radiation had to penetrate. Wavelengths are given in Ångströms. (Kuva Womble, Sargent, Lyons)

trum consists of densely distributed narrow lines that are thought to be hydrogen Lyman α lines formed in gas clouds along the line of sight to the quasar. The clouds producing this “*Lyman α forest*” (Fig. 19.28) are young galaxies or protogalaxies, and they therefore provide important evidence about the formation of galaxies. Lyman α forests can be used to map intergalactic gas clouds up to redshift $z = 6$ (Fig. 19.29). Fur-

ther out Lyman α absorption lines cannot be seen since most of the hydrogen is neutral.

Unified Models Although the forms of galactic activity may at first sight appear diverse, they can be unified within a fairly widely accepted schematic model in such a way that the properties of each individual object depend on a few parameters only.

According to this model, most galaxies contain a compact central nucleus, which is a supermassive black hole, with mass 10^7 – $10^{10} M_\odot$, surrounded by a disk or ring of gas. The source of energy is the gravitational energy released as gas is accreted into the black hole. The disk may also give rise to a jet, where some of the energy is converted into perpendicular motions along the rotational axis. Thus active galactic nuclei are similar to the nucleus of the Milky Way, although the masses of both the black hole and the gas disk may be much larger.

Deducing the mass of the central black hole is difficult and uncertain. However, using a variety of methods involving the motions of stars and gas at the centre of nearby galaxies black hole masses for about 30 galaxies have been determined. The most important result of these studies is that there is a close relationship between the black hole mass and the central velocity dispersion of the galaxy. According to the virial theorem the velocity dispersion is a measure of the bulge mass, and therefore the conclusion is that there is a close relationship between the mass of the bulge and the mass of the central black hole.

The first characteristic parameter of the unified model is obviously the total luminosity. For

example, the only essential difference between Seyfert 1 galaxies and radio-quiet quasars is the larger luminosity of quasars. Another basic parameter is the radio brightness, which may be related to the strength of a jet. On the basis of their radio luminosity one can connect Seyfert galaxies and radio-quiet quasars on one hand, and radio galaxies and radio quasars on the other.

The third important parameter of unified models is the angle from which we happen to view the nuclear disk. For example, if the disk is seen edge-on, the actual nucleus is obscured by the disk. This could explain the difference between Seyfert types 1 and 2: in type 2 we do not see the broad emission lines formed near the black hole, but only the narrower lines from the disk. Similarly a galaxy that looks like a double radio source when seen edge-on, would look like a radio quasar if the disk were seen face-on. In the latter case there is a possibility that we may be seeing an object directly along the jet. It will then appear as a *blazar*, an object with rapid and violent variations in brightness and polarisation, and very weak or invisible emission lines. If the jet is almost relativistic, its transverse velocity may appear larger than the speed of light, and thus superluminal motions can also be understood.

One prediction of the unified model is that there should be a large number of quasars where the nucleus is obscured by the disk as in the Seyfert 2 galaxies. In analogy with the Seyfert galaxies these objects are referred to as type 2 AGN or quasars. Because of the obscuration such sources would not be included in surveys at optical, UV, or soft X-ray wavelengths. In hard X-rays the obscuration is less, and in the far infrared the absorbed energy is re-radiated. Searches for type 2 quasars have been made at these wavelengths with the Chandra X-ray satellite and with the Spitzer Space Telescope. The indications from these searches are that at least 3/4 of all supermassive black holes are heavily obscured.

Gravitational Lenses An interesting phenomenon first discovered in connection with quasars are gravitational lenses. Since light rays are bent by gravitational fields, a mass (e.g. a galaxy)

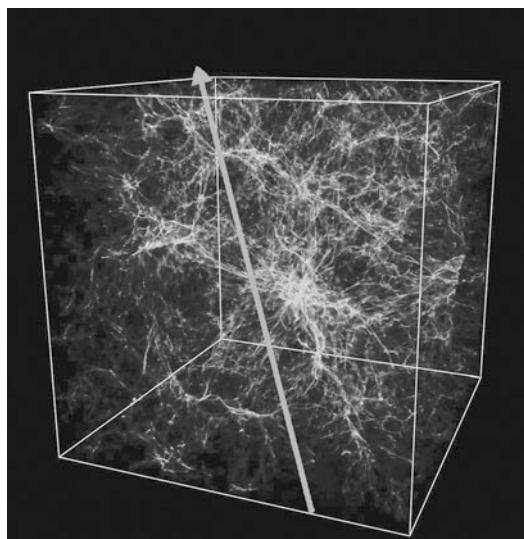
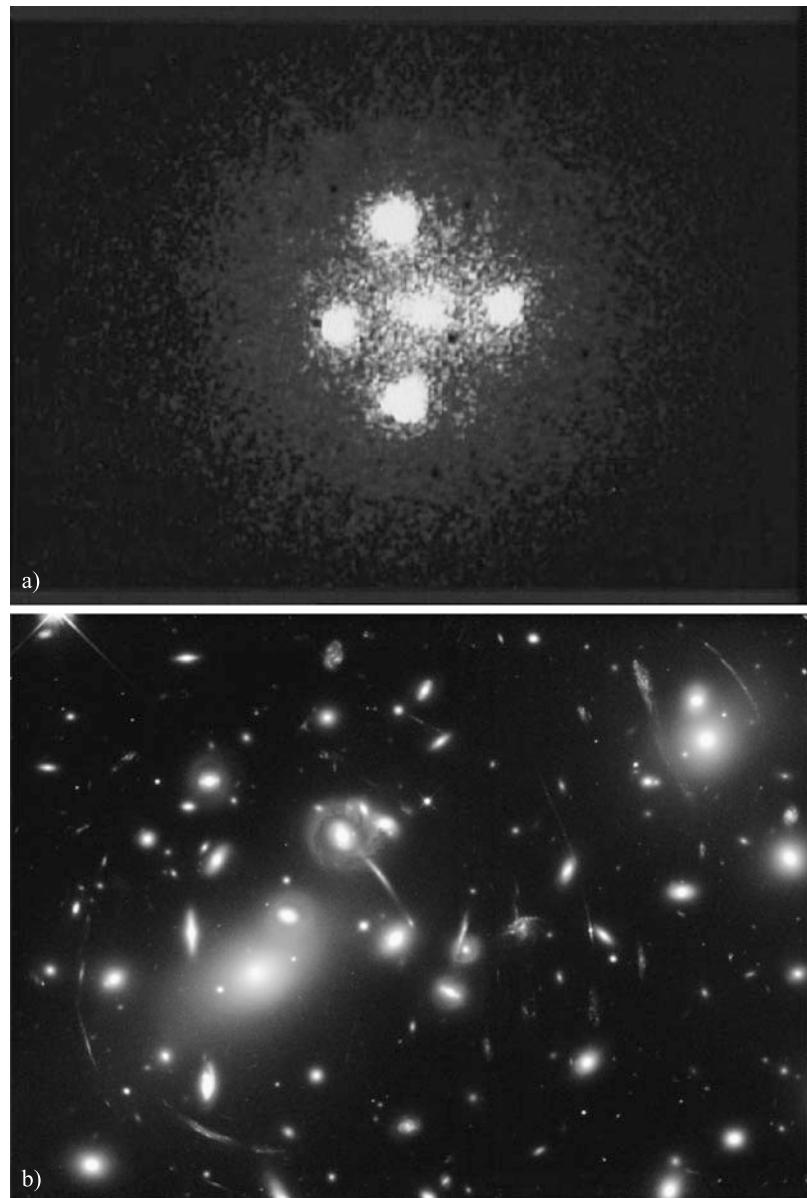


Fig. 19.29 Simulation of hydrogen distribution in space. The light emitted by a quasar goes through different gas clouds creating a Lyman α forest. (Renyue Cen, Princeton University)

Fig. 19.30 (a) The components of the Einstein cross are gravitationally lensed images of the same quasar (ESA/NASA). (b) The massive galaxy cluster Abell 2218 deflects light rays passing through it and acts as a giant gravitational lens. In this Hubble picture from January 2000, dozens of arc-shaped images of distant galaxies can be seen. (Photo A. Fruchter, S. Baggett, R. Hook and Z. Levay, NASA/STScI)



placed between a distant quasar and the observer will distort the image of the quasar. The first example of this effect was discovered in 1979, when it was found that two quasars, $5.7''$ apart in the sky, had essentially identical spectra. It was concluded that the “pair” was really a double image of a single quasar. Since then several other gravitationally lensed quasars have been discovered (Fig. 19.30).

Gravitational lenses have also been discovered in clusters of galaxies. Here the gravitational field of the cluster distorts the images of distant galaxies into arcs around the cluster centre. Currently some two hundred individual galaxies and twenty galaxy clusters are known to produce lensed images of background galaxies or quasars. Such systems are called *strong gravitational lenses* (Fig. 19.31).

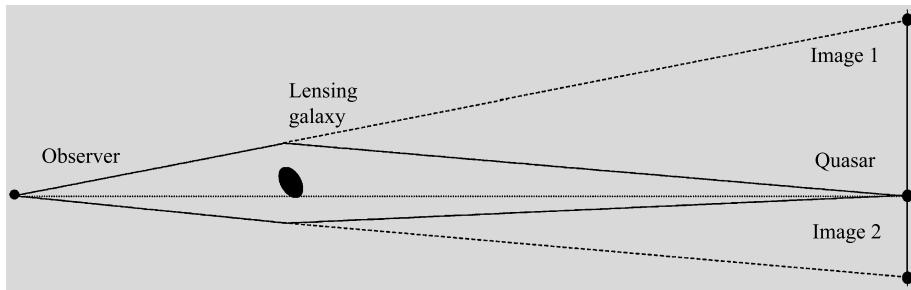


Fig. 19.31 A strong gravitational lens. The light of a quasar is bent by the gravity of a lensing galaxy. The quasar appears as two images on both sides of the lensing galaxy

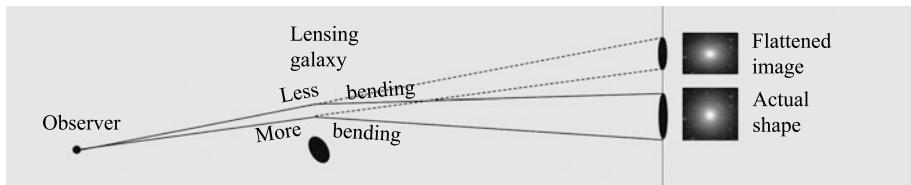


Fig. 19.32 A weak gravitational lens. Light rays coming from the upper and lower edges of a distant galaxy are bent by different amounts. The image of the galaxy becomes slightly flattened

Recently emphasis has moved towards *weak gravitational lenses*, which are much more common (Fig. 19.32). The lens does not produce separate images or arcs from a background but only flattens its shape. Since the axial ratio changes only by 1–2 %, the weak lenses cannot be used to study shapes of individual galaxies because their original shapes are not known. However, if hundreds or thousands of galaxies around a mass concentration are measured, statistical analysis shows the flattening. In the 2010's weak lenses are routinely used to observe the distribution of the invisible matters.

Figure 19.13 showed an example about dark matter clouds around a galaxy cluster.

which therefore always has to be specified when studying the evolution of galaxies.

Evolution of galaxies began when the first stars were born at redshifts 10–20. The universe consisted of neutral hydrogen and helium only, and the temperature of the gas had dropped below 100 kelvins. The first stars were very massive and shortlived. Radiation from the first stars, galaxies and quasars rapidly ionised the interstellar and intergalactic gas. This era is called the time of ionisation. Ionisation was almost complete when the redshift reached $z = 6$, and ever since 99.9 % of the hydrogen has been ionised. Towards the end of the reionisation, at redshifts 6–15, massive galaxies ($M > 10^{10} M_{\odot}$) were already born. Current observations of galaxies go back to redshifts 10–12.

According to the currently widely accepted cosmological models most of the matter in the Universe is in a form that emits no radiation, and is only observable from its gravitational effects. In this *Cold Dark Matter* (CDM) theory (see Sect. 20.7) the first systems to collapse and start forming stars were small, with masses like those of dwarf galaxies. Larger galaxies were formed later as these smaller fragments collected

19.8 The Origin and Evolution of Galaxies

Because the speed of light is finite we see distant galaxies at earlier stages in their life. In the next chapter we show how the age of a galaxy with given redshift can be calculated based on the rate of expansion of the Universe. However, this relationship will depend on the cosmological model,

into larger clumps. This model, where most stars are formed in small galaxies is usually described as the *hierarchical* model.

A large part of our theoretical ideas on galaxy evolution is based on numerical simulations of the collapse of gas clouds and star formation in them. Using some prescription for star formation one can try to compute the evolution of the spectral energy distribution and the chemical abundances in the resulting galaxies. The results of the models can be compared with the observational data presented in the previous sections of this chapter.

The density distribution of dark matter is expected to be very irregular, containing numerous small-scale clumps. The collapse will therefore be highly inhomogeneous, both in the hierarchical and in the monolithic picture, and subsequent mergers between smaller systems should be common. There are additional complicating factors. Gas may be expelled from the galaxy, or there may be an influx of fresh gas. Interactions with the surroundings may radically alter the course of evolution—in dense systems they may lead to the complete merging of the individual galaxies into one giant elliptical. Much remains to be learned about how the formation of stars is affected by the general dynamical state of the galaxy and of how an active nucleus may influence the formation process.

Our observational knowledge of galaxy evolution is advancing very rapidly. Essentially all the relationships described earlier in this chapter have been studied as functions of time. Still, a complete generally accepted description of the way the Universe reached its present state has not yet been established. Here we can only mention a few of the most central aspects of the processes leading to the galaxies we observe to-day.

Density and Luminosity Evolution The most basic way of studying the formation and evolution of galaxies is by counting their numbers, either the number brighter than some given magnitude limit (as was already done by Hubble in the 1930's, see Sect. 20.1) or else the number density as a function of redshift. The counts can be compared with the numbers expected if there is

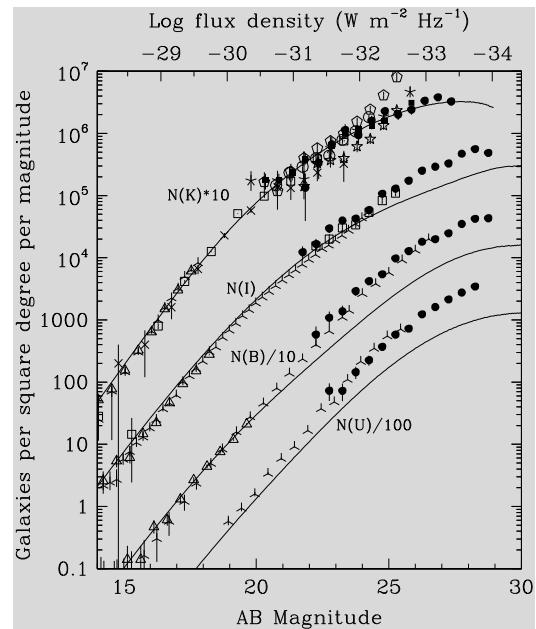


Fig. 19.33 Galaxy counts in the U, B, I, and K wavelength bands. The counts are compared to those in a cosmological model without evolution. The cosmological parameters correspond to the currently preferred “concordance” model (see Sect. 19.5). (H.C. Ferguson et al., 2000, ARAA 38, 667, Fig. 4)

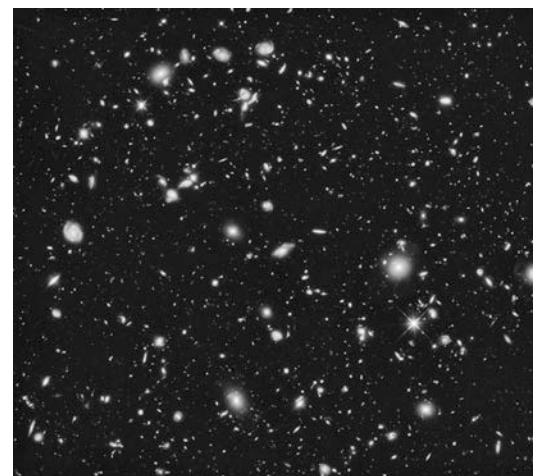
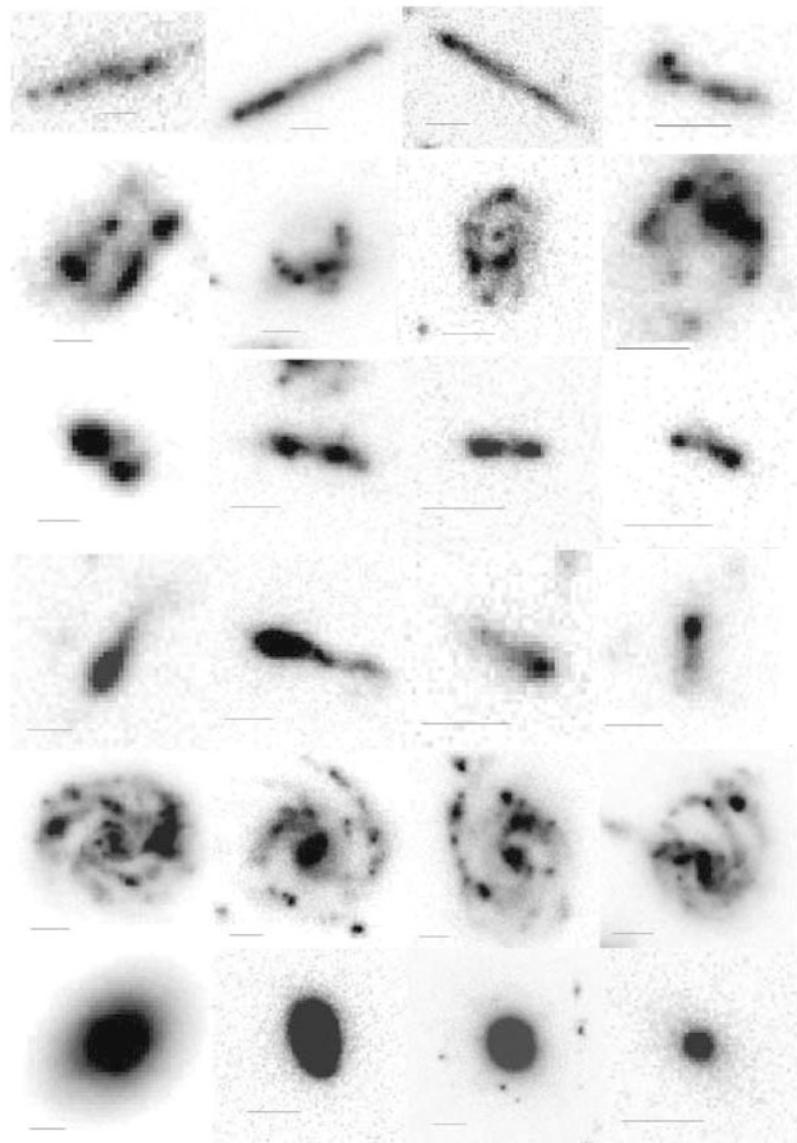


Fig. 19.34 The deepest image taken by the Hubble space telescope and published in 2014. The total exposure time is 27 days. The original picture contains about 15,000 faint galaxies. (NASA, ESA, H. Teplitz, M. Rafelski, A. Koekemoer, R. Windhorst and Z. Levay)

no evolution, which depend on the cosmological model. They can therefore be used either as a cos-

Fig. 19.35 Different early forms of galaxies in the redshift range $z = 2\text{--}4$. *Top line*: chain galaxies; *second line*: clump-clusters; *third line*: double galaxies; *fourth line*: tadpoles; *fifth line*: spiral galaxies being born; *bottom line*: elliptic galaxies. (Elmegreen, Elmegreen, Rubin, Schaffer, 2005, ApJ 631, p. 87)



mological test, or as a test of evolutionary models. However, the present situation is that more reliable cosmological tests are available, and the number counts are mainly used to study the evolution of galaxies.

There are two ways in which the number counts are affected by galaxy evolution. In *density evolution* the actual number of galaxies is changing, whereas in *luminosity evolution* only the luminosity of individual galaxies is evolving. The simplest form of luminosity evolution

is called *passive luminosity evolution*, and is due to the changing luminosity of stars during normal stellar evolution. Pure luminosity evolution is expected to be predominant in the monolithic picture. In the hierarchical picture density evolution will be more prominent, since in this picture smaller galaxies are to a greater extent being destroyed to produce larger more luminous ones.

Figure 19.33 gives an example of the results of number counts. A model without evolution cannot explain these observed counts, and various

models incorporating evolutionary effects have to be introduced. However, a unique model cannot be determined using just the number counts.

Distant Galaxies A more direct approach to galaxy formation is the direct search for the most distant objects visible. In Fig. 19.34 we show the *Hubble Ultra Deep Field* (HUDF), an area of the sky observed with the Hubble Space Telescope. The total exposure time through different filters was 27 days. The image contains over 15,000 galaxies. Over 5000 are seen as dots; their redshift is aver 6. The other ones, seen as extended objects, give information on galaxy evolution at redshifts 1–5.

The most important observation is that with increasing redshift the Hubble classification breaks down and finally disappears. Around redshift $z = 3$ the spiral galaxies resembling the current ones have disappeared almost completely. Instead, there are different irregular and strange galaxies.

Figure 19.35 shows examples of distant galaxies in the range $z = 2$ –4. Over 10 % are “chain galaxies”, about 20 % “clump clusters”, 15 % “doubles” and 10 % “tadpoles”. About 30 % are evolving spirals and 10 % ellipticals.

Another important observation is that in the past galaxies were smaller than nowadays. Even if the mass or luminosity of a galaxy were similar to a modern galaxy its diameter at $z = 3$ is only one tenth. The most probable explanation is that the outer parts grow in several small mergers. Collision of a dwarf galaxy does not considerably increase the mass of the great galaxy but brings more stars to its outskirts increasing its diameter. The observation clearly indicates how important mutual collisions have been in the evolution of galaxies towards their current forms.

Evolution of AGN The first clear indication of cosmic evolution was in the numbers of radio galaxies and quasars. Already in the late 1960’s it was becoming clear that the density of quasars was increasing dramatically towards higher redshifts (Fig. 19.36). Roughly, the number density of quasars increases relative to the present density by a factor 100 out to a broad maximum at

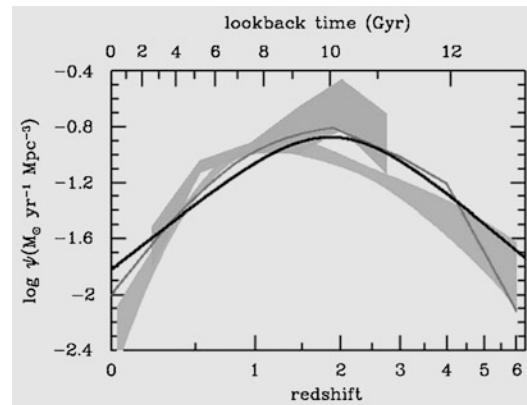


Fig. 19.36 Quasar activity at different redshifts according to X-ray and infrared observations (gray areas and the gray line). The black line corresponds to the stellar birth rate in Fig. 19.37. (Madau ja Dickinson, 2014, Ann. Rev. Astron. Astrophys. 52)

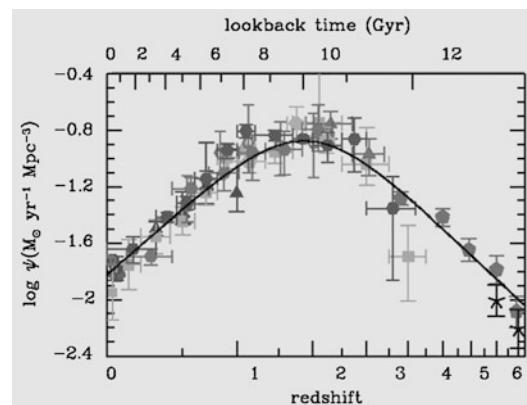


Fig. 19.37 Stellar birth rates at different redshifts. Different symbols correspond to values obtained in different studies and their standard errors. (Madau and Dickinson, 2014, Ann. Rev. Astron. Astrophys. 52)

redshift about 2. The observed behaviour may be due to either density or luminosity evolution. The density of radio galaxies also has a maximum at redshifts about 1.5–3, which is sometimes referred to as the *quasar era*.

The Star Formation History of the Universe Since the Universe contained only neutral gas as the first stars began to form, the most general description of how galaxies came to be is in terms of the rate at which the gas is being turned into stars. The star formation history going back to

a redshift about 8 is shown in Fig. 19.37. The star formation rate was about an order of magnitude larger than its current value at redshift 1–2. For even larger redshifts it seems to have remained nearly constant or decreased slowly.

19.9 Exercises

Exercise 19.1 The galaxy NGC 772 is an Sb spiral, similar to M31. Its angular diameter is $7'$ and apparent magnitude 12.0. The corresponding values of M31 are 3.0° and 5.0. Find the ratio of the distances of the galaxies

- (a) assuming their sizes are equal,
- (b) assuming they are equally bright.

Exercise 19.2 The brightness of the quasar 3C279 has shown changes with the time scale of one week. Estimate the size of the region producing the radiation. The apparent magnitude is 18. If the distance of the quasar is 2000 Mpc, what is its absolute magnitude and luminosity? How much energy is produced per au^3 ?

Cosmology is a science studying the structure and evolution of the whole Universe. Research in cosmology tries to answer questions such as: How large and how old is the Universe? How was it born? How is matter distributed? How were the elements formed? What will be the future of the Universe?

After the demise of the Aristotelian world picture, it took hundreds of years of astronomical observations and physical theories to reach a level at which a satisfactory modern scientific picture of the physical universe could be formed. The decisive steps in the development were the clarification of the nature of the galaxies in the 1920's and the general theory of relativity, which gives a basis for theoretical research of the universe, developed by Einstein in the 1910's. Yet, just a few decades ago there was a lot of room for armwaiving and all kinds of speculations. Only recently the observations made with satellites have given reasonably precise values for some parameters and some alternative models could be excluded. The central tenet of modern cosmology is the model of the expanding universe, getting all the time more supporting observational evidence.

20.1 Cosmological Observations

The Olbers Paradox The simplest cosmological observation may be that the sky is dark at night. This fact was first noted by *Johannes Kepler*, who, in 1610, used it as evidence for a finite universe. As the idea of an infinite space filled

with stars like the Sun became widespread in consequence of the Copernican revolution, the question of the dark night sky remained a problem. In the 18th and 19th centuries *Edmond Halley*, *Loys de Chézeaux* and *Heinrich Olbers* considered it in their writings. It has become known as the *Olbers paradox* (Fig. 20.1).

The paradox is the following: Let us suppose the Universe is infinite and that the stars are uniformly distributed in space. No matter in what direction one looks, sooner or later the line of sight will encounter the surface of a star. Since the surface brightness does not depend on distance, each point in the sky should appear to be as bright as the surface of the Sun. This clearly is not true. The modern explanation of the paradox is that the stars have only existed for a finite time, so that the light from very distant stars has not yet reached us. Rather than proving the world to be finite in space, the Olbers paradox has shown it to be of a finite age.

Extragalactic Space In 1923 *Edwin Hubble* showed that the Andromeda Galaxy M31 was far outside the Milky Way, thus settling a long-standing controversy concerning the relationship between the nebulae and the Milky Way. The numerous galaxies seen in photographs form an extragalactic space vastly larger than the dimensions of the Milky Way. It is important for cosmology that the distribution and motions of the basic components of extragalactic space, the galaxies and clusters of galaxies, should everywhere be the same as in our local part of the



Fig. 20.1 The Olbers paradox. If the stars were uniformly distributed in an unending, unchanging space, the sky should be as bright as the surface of the Sun, since each line of sight would eventually meet the surface of a star.

A two-dimensional analogy can be found in an optically thick pine forest where the line of sight meets a trunk wherever one looks. (Photo M. Poutanen and H. Karttunen)

Universe. Galaxies generally occur in various systems, ranging from small groups to clusters of galaxies and even larger superclusters. The largest structures observed are about 100 Mpc in size (see Sect. 19.6). They are thus significantly smaller than the volume of space (a few thousand Mpc in size) in which the distribution of galaxies has been investigated. One way of studying the large-scale homogeneity of the galaxy distribution is to count the number of galaxies brighter than some limiting magnitude m . If the galaxies are uniformly distributed in space, this number should be proportional to $10^{0.6m}$ (see Example 18.1). For example, the galaxy counts made by Hubble in 1934, which included 44,000 galaxies, were consistent with a galaxy distribution independent of position (homogeneity) and of direction (isotropy). Hubble found no “edge” of the Universe, nor have later galaxy counts found one.

Similar counts have been made for extragalactic radio sources. (Instead of magnitudes, flux densities are used. If F is the flux density, then because $m = -2.5 \lg(F/F_0)$, the number count will be proportional to $F^{-3/2}$.) These counts mainly involve very distant radio galaxies and quasars (Fig. 20.2). The results seem to indicate that the radio sources were either much brighter or much more common at earlier epochs than at present (Sect. 19.8). This constitutes evidence in favour of an evolving, expanding universe.

In general the simple geometric relation between brightness and number counts will only

hold for objects that are uniformly distributed in space. Local inhomogeneities will cause departures from the expected relationship. For more distant sources the geometry of the Universe as well as cosmic evolution will change the basic $10^{0.6m}$ behaviour.

Hubble’s Law (See Fig. 20.3.) In the late 1920’s, Hubble discovered that the spectral lines of galaxies were shifted towards the red by an amount proportional to their distances. If the redshift is due to the Doppler effect, this means that the galaxies move away from each other with velocities proportional to their separations, i.e. that the Universe is expanding as a whole.

In terms of the redshift $z = (\lambda - \lambda_0)/\lambda_0$, Hubble’s law can be written as

$$z = (H/c)r, \quad (20.1)$$

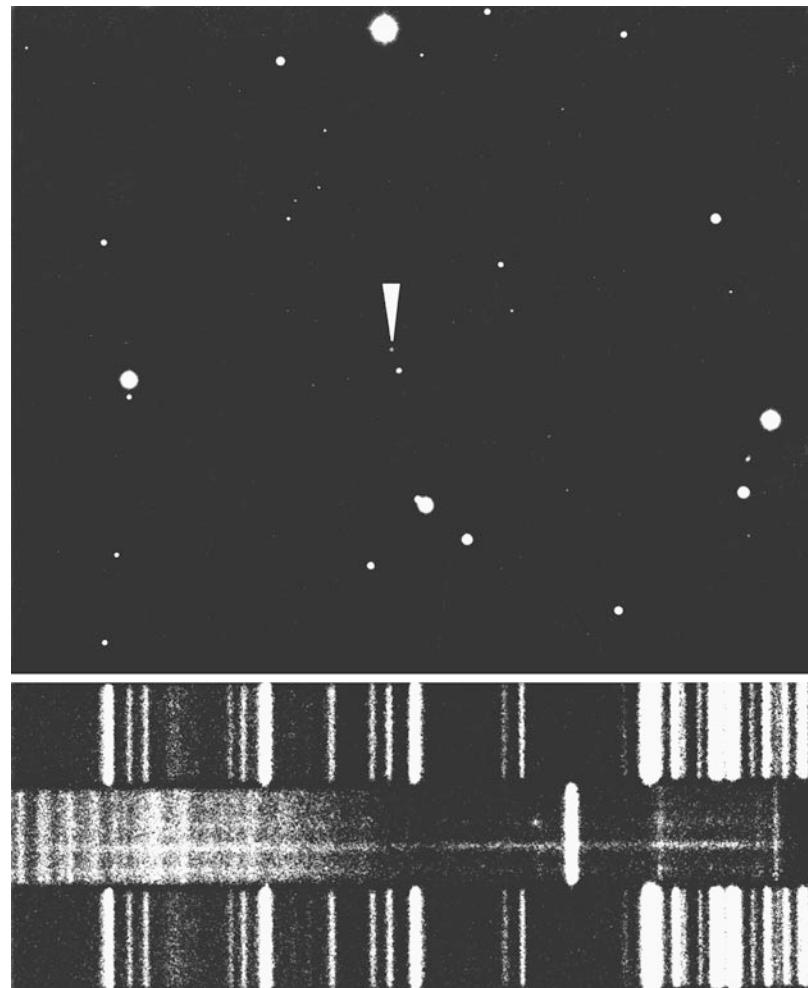
where c is the speed of light, H is the *Hubble constant* and r the distance of the galaxy. For small velocities ($V \ll c$) the Doppler redshift $z = V/c$, and hence

$$V = Hr, \quad (20.2)$$

which is the most commonly used form of Hubble’s law.

For a set of observed “standard candles”, i.e. galaxies whose absolute magnitudes are close to some mean M_0 , Hubble’s law corresponds to

Fig. 20.2 The quasar 3C295 and its spectrum. The quasars are among the most distant cosmological objects. (Photograph Palomar Observatory)



a linear relationship between the apparent magnitude m and the logarithm of the redshift, $\lg z$. This is because a galaxy at distance r has an apparent magnitude $m = M_0 + 5\lg(r/10 \text{ pc})$, and hence Hubble's law yields

$$m = M_0 + 5\lg\left(\frac{cz}{H \times 10 \text{ pc}}\right) = 5\lg z + C, \quad (20.3)$$

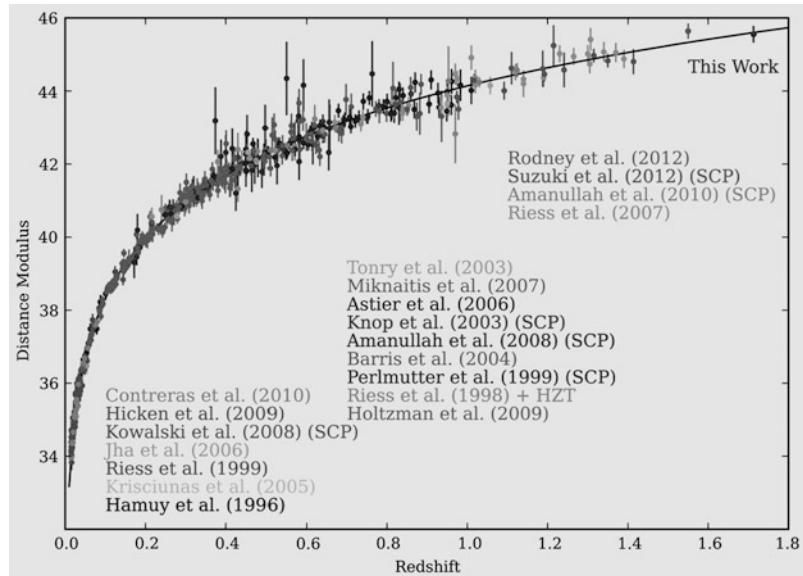
where the constant C depends on H and M_0 . Suitable standard candles are e.g. the brightest galaxies in clusters and Sc galaxies of a known luminosity class. Some other methods of distance determination for galaxies were discussed in Sect. 19.2. Most recently, type Ia supernovae (Sect. 14.4) in distant galaxies have been used to determine distances out to the redshift $z = 1.7$,

where departures from Hubble's law are already detectable.

If the Universe is expanding, the galaxies were once much nearer to each other. If the rate of expansion had been unchanging, the inverse of the Hubble constant, $T = H^{-1}$, would represent the age of the Universe. If the expansion is gradually slowing down, the inverse Hubble constant gives an upper limit on the age of the Universe. In fact, current indications (discussed later in this chapter) are that the rate of expansion is accelerating at present. In that case the age of the Universe may also be larger. However, H^{-1} will still be an estimate for the age of the Universe.

One reason for the difficulty in determining the value of the Hubble constant is the uncer-

Fig. 20.3 Hubble's law for type Ia supernovae. The solid curve represents the “concordance” model, with basic parameters having the values mentioned in Sect. 20.7. Models with a zero cosmological constant cannot explain the observations. (D. Rubin et al., 2013, ApJ 763, 35)



tainty in extragalactic distances. A second problem is that the measured values of the velocity V , corrected to take into account the motion of the Sun within the Local Group, contain a significant component due to the peculiar motions of the galaxies. These peculiar velocities are caused by local mass concentrations like groups and clusters of galaxies. It is possible that the Local Group has a significant velocity towards the centre of the Local Supercluster (the Virgo Cluster). Because the Virgo Cluster is often used to determine the value of H , neglecting this peculiar velocity leads to a large error in H . The size of the peculiar velocity is not yet well known, but it is probably about 250 km s^{-1} .

The most ambitious recent project for determining H used the Hubble Space Telescope in order to measure cepheid distances to a set of nearby galaxies. These distances were then used to calibrate other distance indicators, such as the Tully–Fisher relation and type Ia supernovae. The final result was $H = (72 \pm 8) \text{ km s}^{-1} \text{ Mpc}^{-1}$. The largest remaining source of error in this result is the distance to the Large Magellanic Cloud, used for calibrating the cepheid luminosity.

Nowadays the most accurate value for the Hubble constant is obtained by studying the angular power spectrum of the background radiation (Sect. 20.7). According to the mea-

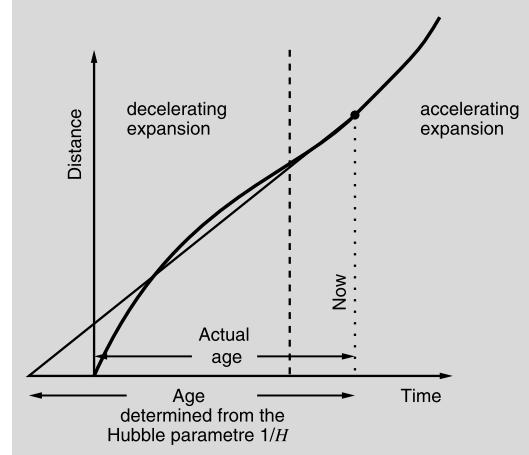


Fig. 20.4 It is nowadays assumed that at the beginning the expansion rate of the Universe was slowing down until about 5 billion years ago the deceleration changed to acceleration. The age of the Universe calculated from the current value of the Hubble constant is bigger than the actual age

surements of the Planck satellite $H = (67.7 \pm 0.5) \text{ km s}^{-1} \text{ Mpc}^{-1}$ and the age of the Universe is $13.80 \pm 0.02 \text{ Ga}$.

The form of Hubble's law might give the impression that the Milky Way is the centre of the expansion, in apparent contradiction with the Copernican principle. Figure 20.5 shows that, in fact, the same Hubble's law is valid at each point

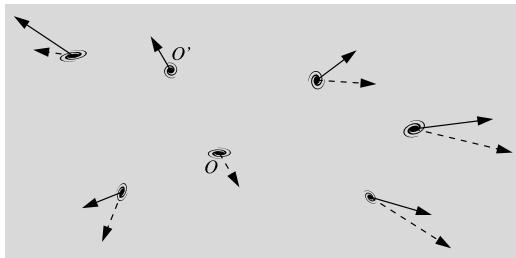


Fig. 20.5 A regular expansion according with Hubble's law does not mean that the Milky Way (O) is the centre of the Universe. Observers at any other galaxy (O') will see the same Hubble flow (dashed lines)

in a regularly expanding universe. There is no particular centre of expansion.

The Thermal Microwave Background Radiation

The most important cosmological discovery since Hubble's law was made in 1965. In that year *Arno Penzias* and *Robert Wilson* discovered that there is a universal microwave radiation, with a spectrum corresponding to that of blackbody radiation (see Sect. 5.6) at a temperature of about 3 K (Fig. 20.7). For their discovery, they received the Nobel prize in physics in 1979.

The existence of a thermal cosmic radiation background had been predicted in the late 1940's by *George Gamow*, who was one of the first to study the initial phases of expansion of the Universe. According to Gamow, the Universe at that time was filled with extremely hot radiation. As it expanded, the radiation cooled, until at present, its temperature would be a few kelvins. After its discovery by Penzias and Wilson, the cosmic background radiation has been studied at wavelengths from 50 cm to 0.5 cm. The first detailed measurements, made from the COBE (Cosmic Background Explorer) satellite showed that it corresponds closely to a Planck spectrum at (2.725 ± 0.002) K (Fig. 20.6). More recently the CMB has been mapped in even greater detail by the WMAP satellite.

The existence of the thermal cosmic microwave background (CMB) gives strong support to the belief that the Universe was extremely hot in its early stages. The background is very nearly isotropic, which supports the isotropic and homogeneous models of the Universe. The COBE

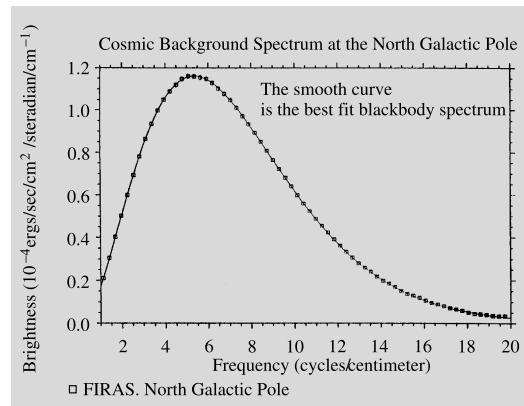


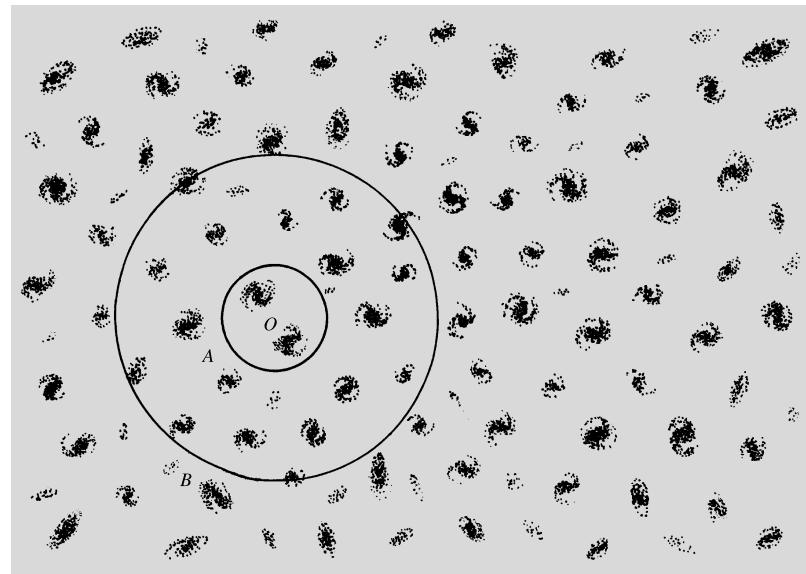
Fig. 20.6 Observations of the cosmic microwave background radiation made by the COBE satellite in 1990 are in agreement with a blackbody law at 2.7 K

and WMAP satellites have also detected temperature variations of a relative amplitude 6×10^{-6} in the background. These fluctuations are interpreted as a gravitational redshift of the background produced by the mass concentrations that would later give rise to the observed structures in the Universe. They are the direct traces of initial irregularities in the big bang, and provide important constraints for theories of galaxy formation. Perhaps even more importantly, the amplitude of the fluctuations on different angular scales have provided crucial constraints on the cosmological model. We shall return to this question in Sect. 20.7.

The Isotropy of Matter and Radiation Apart from the CMB, several other phenomena confirm the isotropy of the Universe. The distribution of radio sources, the X-ray background, and faint distant galaxies, as well as Hubble's law are all isotropic. The observed isotropy is also evidence that the Universe is homogeneous, since a large-scale inhomogeneity would be seen as an anisotropy.

The Age of the Universe Estimates of the ages of the Earth, the Sun and of star clusters are important cosmological observations that do not depend on specific cosmological models. From the decay of radioactive isotopes, the age of the Earth is estimated to be 4600 million years. The age of the Sun is thought to be slightly larger than this.

Fig. 20.7 The cosmological principle. In the small circle (*A*) about the observer (*O*) the distribution of galaxies does not yet represent the large-scale distribution. In the larger circle (*B*) the distribution is already uniform on the average



The ages of the oldest star clusters in the Milky Way are 10–15 Ga.

The values thus obtained give a lower limit to the age of the Universe. In an expanding universe, the inverse Hubble constant gives another estimate of that age. It is most remarkable that the directly determined ages of cosmic objects are so close to the age given by the Hubble constant. This is strong evidence that Hubble's law really is due to the expansion of the Universe. It also shows that the oldest star clusters formed very early in the history of the Universe.

The Relative Helium Abundance A cosmological theory should also give an acceptable account of the origin and abundances of the elements. Even the abundance of the elementary particles and the lack of antimatter are cosmological problems that have begun to be investigated in the context of theories of the early Universe.

Observations show that the oldest objects contain about 25 % by mass of helium, the most abundant element after hydrogen. The amount of helium produced is sensitive to the temperature of the Universe, which is related to that of the background radiation. The computations made for the standard models of the expanding Universe (the Friedmann models) yield a helium abundance of exactly the right size.

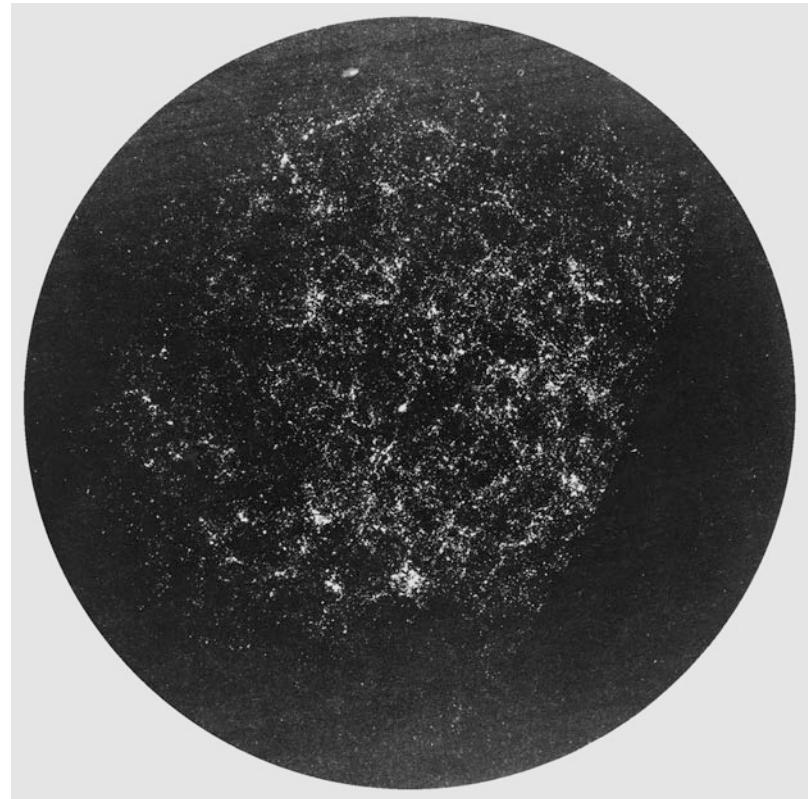
20.2 The Cosmological Principle

One hopes that as ever larger volumes of the Universe are observed, its average properties will become simple and well defined. Figure 20.7 attempts to show this. It shows a distribution of galaxies on a plane. As the circle surrounding the observer *O* becomes larger, the mean density inside the circle becomes practically independent of its size. The same behaviour occurs, regardless of the position of the centre of *O*: at close distances, the density varies randomly (Fig. 20.8), but in a large enough volume, the average density is constant. This is an example of the *cosmological principle*: apart from local irregularities, the Universe looks the same from all positions in space.

The cosmological principle is a far-reaching assumption, which has been invoked in order to put constraints on the large variety of possible cosmological theories. If in addition to the cosmological principle one also assumes that the Universe is isotropic, then the only possible cosmic flow is a global expansion. In that case, the local velocity difference *V* between two nearby points has to be directly proportional to their separation ($V = Hr$); i.e. Hubble's law must apply.

The plane universe of Fig. 20.7 is homogeneous and isotropic, apart from local irregu-

Fig. 20.8 The galaxies seem to be distributed in a “foamlike” way. Dense strings and shells are surrounded by relatively empty regions. (Seldner, M. et al. (1977): Astron. J. **82**, 249)



ities. Isotropy at each point implies homogeneity, but homogeneity does not require isotropy. An example of an anisotropic, homogeneous universe would be a model containing a constant magnetic field: because the field has a fixed direction, space cannot be isotropic.

We have already seen that astronomical observations support the homogeneity and isotropy of our observable neighbourhood, the *metagalaxy*. On the grounds of the cosmological principle, these properties may be extended to the whole of the Universe.

The cosmological principle is closely related to the Copernican principle that our position in the Universe is in no way special. From this principle, it is only a short step to assume that on a large enough scale, the local properties of the metagalaxy are the same as the global properties of the Universe.

Homogeneity and isotropy are important simplifying assumptions when trying to construct *cosmological models* which can be compared

with local observations. They may therefore reasonably be adopted, at least as a preliminary hypothesis.

20.3 Homogeneous and Isotropic Universes

Under general conditions, space and time coordinates in a universe may be chosen so that the values of the space coordinates of observers moving with the matter are constant. It can be shown that in a homogeneous and isotropic universe, the line element (Appendix B) then takes the form

$$\mathrm{d}s^2 = -c^2 \mathrm{d}t^2 + R^2(t) \left[\frac{\mathrm{d}r^2}{1 - kr^2} + r^2(\mathrm{d}\theta^2 + \cos^2 \theta \mathrm{d}\phi^2) \right], \quad (20.4)$$

known as the *Robertson–Walker line element*. (The radial coordinate r is defined to be dimensionless.) $R(t)$ is a time-dependent quantity representing the scale of the Universe. If R increases

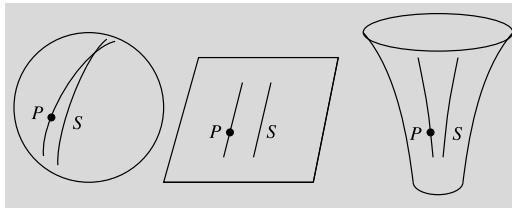


Fig. 20.9 The two-dimensional analogues of the Friedmann models: a spherical surface, a plane and a saddle surface

with time, all distances, including those between galaxies, will grow. The coefficient k may be $+1$, 0 or -1 , corresponding to the three possible geometries of space, the *elliptic* or *closed*, the *parabolic* and the *hyperbolic* or *open* model.

The space described by these models need not be Euclidean, but can have positive or negative curvature. Depending on the curvature, the volume of the universe may be finite or infinite. In neither case does it have a visible edge.

The two-dimensional analogy to elliptical ($k = +1$) geometry is the surface of a sphere (Fig. 20.9): its surface area is finite, but has no edge. The scale factor $R(t)$ represents the size of the sphere. When R changes, the distances between points on the surface change in the same way. Similarly, a three-dimensional “spherical surface”, or the space of elliptical geometry, has a finite volume, but no edge. Starting off in an arbitrary direction and going on for long enough, one always returns to the initial point (unless the distance is increased too much by the expansion).

When $k = 0$, space is flat or Euclidean, and the expression for the line element (20.4) is almost the same as in the Minkowski space. The only difference is the scale factor $R(t)$. All distances in a Euclidean space change with time. The two-dimensional analogue of this space is a plane.

The volume of space in the hyperbolic geometry ($k = -1$) is also infinite. A two-dimensional idea of the geometry in this case is given by a saddle surface or a horn becoming infinitely wider.

In a homogeneous and isotropic universe, many physical quantities will depend on time through the scale factor $R(t)$. For example, from the form of the line element, it is evident that all distances will be proportional to R (Fig. 20.10).

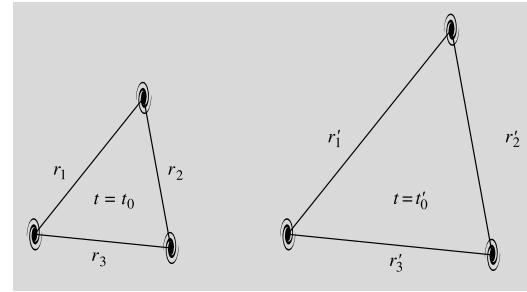


Fig. 20.10 When space expands, all galaxy separations grow with the scale factor R : $r' = [R(t'_0)/R(t_0)]r$

Thus, if the distance to a galaxy is r at time t , then at time t_0 (in cosmology, the subscript 0 refers to the present value) it will be

$$\frac{R(t_0)}{R(t)} r. \quad (20.5)$$

Similarly, all volumes will be proportional to R^3 . From this it follows that the density of any conserved quantity (e.g. mass) will behave as R^{-3} .

It can be shown that the wavelength of radiation in an expanding universe is proportional to R , like all other lengths. If the wavelength at the time of emission, corresponding to the scale factor R , is λ , then it will be λ_0 when the scale factor has increased to R_0 :

$$\frac{\lambda_0}{\lambda} = \frac{R_0}{R}. \quad (20.6)$$

The redshift is $z = (\lambda_0 - \lambda)/\lambda$, and hence

$$1 + z = \frac{R_0}{R}; \quad (20.7)$$

i.e. the redshift of a galaxy expresses how much the scale factor has changed since the light was emitted. For example, the light from a quasar with $z = 1$ was emitted at a time when all distances were half their present values.

For small values of the redshift, (20.7) approaches the usual form of Hubble’s law. This can be seen as follows. When z is small, the change in R during the propagation of a light signal will also be small and proportional to the light travel time t . Because $t = r/c$ approximately, where r is the distance of the source, the redshift will be

proportional to r . If the constant of proportionality is denoted by H/c , one has

$$z = Hr/c. \quad (20.8)$$

This is formally identical to Hubble's law (20.1). However, the redshift is now interpreted in the sense of (20.7).

As the universe expands, the photons in the background radiation will also be redshifted. The energy of each photon is inversely proportional to its wavelength, and will therefore behave as R^{-1} . It can be shown that the number of photons will be conserved, and thus their number density will behave as R^{-3} . Combining these two results, one finds that the energy density of the background radiation is proportional to R^{-4} . The energy density of blackbody radiation is proportional to T^4 , where T is the temperature. Thus the temperature of cosmic background radiation will vary as R^{-1} .

20.4 The Friedmann Models

The results of the preceding section are valid in any homogeneous and isotropic universe. In order to determine the precise time-dependence of the scale factor $R(t)$ a theory of gravity is required.

In 1917 *Albert Einstein* presented a model of the Universe based on his general theory of relativity. It described a geometrically symmetric (spherical) space with finite volume but no boundary. In accordance with the cosmological principle, the model was homogeneous and isotropic. It was also *static*: the volume of space did not change.

In order to obtain a static model, Einstein had to introduce a new repulsive force, the *cosmological term*, in his equations. The size of this cosmological term is given by the *cosmological constant* Λ . Einstein presented his model before the redshifts of the galaxies were known, and taking the Universe to be static was then reasonable. When the expansion of the Universe was discovered, this argument in favour of the cosmological constant vanished. Einstein himself later called it the biggest blunder of his life. Nevertheless, the most recent observations now seem to indicate that a non-zero cosmological constant has to be present.

The St. Petersburg physicist *Alexander Friedmann* and later, independently, the Belgian *Georges Lemaître* studied the cosmological solutions of Einstein's equations. If $\Lambda = 0$, only evolving, expanding or contracting models of the Universe are possible. From the Friedmann models exact formulas for the redshift and Hubble's law may be derived.

The general relativistic derivation of the law of expansion for the Friedmann models will not be given here. It is interesting that the existence of three types of models and their law of expansion can be derived from purely Newtonian considerations, with results in complete agreement with the relativistic treatment. The detailed derivation is given on Box 20.1, but the essential character of the motion can be obtained from a simple energy argument.

Let us consider a small expanding spherical region in the Universe. In a spherical distribution of matter, the gravitational force on a given spherical shell depends only on the mass inside that shell. We shall here assume $\Lambda = 0$.

We can now consider the motion of a galaxy of mass m at the edge of our spherical region. According to Hubble's law, its velocity will be $V = Hr$ and the corresponding kinetic energy,

$$T = mV^2/2. \quad (20.9)$$

The potential energy at the edge of a sphere of mass M is $U = -GMm/r$. Thus the total energy is

$$E = T + U = mV^2/2 - GMm/r, \quad (20.10)$$

which has to be constant. If the mean density of the Universe is ρ , the mass is $M = (4\pi r^3/3)\rho$. The value of ρ corresponding to $E = 0$ is called the *critical density*, ρ_c . We have

$$\begin{aligned} E &= \frac{1}{2}mH^2r^2 - \frac{GMm}{r} \\ &= \frac{1}{2}mH^2r^2 - Gm\frac{4\pi}{3}\frac{r^3\rho_c}{r} \\ &= mr^2\left(\frac{1}{2}H^2 - \frac{4}{3}\pi G\rho_c\right) = 0, \end{aligned} \quad (20.11)$$

whence

$$\rho_c = \frac{3H^2}{8\pi G}. \quad (20.12)$$

The expansion of the Universe can be compared to the motion of a mass launched vertically from the surface of a celestial body. The form of the orbit depends on the initial energy. In order to compute the complete orbit, the mass M of the main body and the initial velocity have to be known. In cosmology, the corresponding parameters are the mean density and the Hubble constant.

The $E = 0$ model corresponds to the Euclidean Friedmann model, the *Einstein-de Sitter* model. If the density exceeds the critical density, the expansion of any spherical region will turn to a contraction and it will collapse to a point. This corresponds to the closed Friedmann model. Finally, if $\rho < \rho_c$, the ever expanding hyperbolic model is obtained. The behaviour of the scale factor in these three cases is shown in Fig. 20.11.

These three models of the universe are called the *standard models*. They are the simplest relativistic cosmological models for $\Lambda = 0$. Models with $\Lambda \neq 0$ are mathematically more complicated, but show the same general behaviour.

The simple Newtonian treatment of the expansion problem is possible because Newtonian mechanics is approximately valid in small regions of the Universe. However, although the resulting equations are formally similar, the interpretation

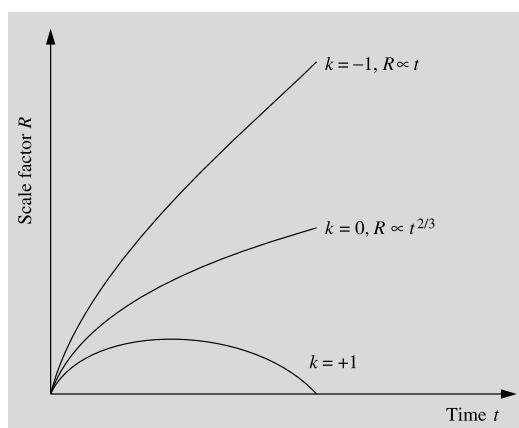


Fig. 20.11 The time dependence of the scale factor for different values of k . The cosmological constant $\Lambda = 0$

of the quantities involved (e.g. the parameter k) is not the same as in the relativistic context. The global geometry of the Friedmann models can only be understood within the general theory of relativity.

Consider two points at a separation r at the time t . Let their relative velocity be V . If the distance at t_0 is r_0 then

$$r = \frac{R(t)}{R(t_0)} r_0 \quad \text{and} \quad V = \dot{r} = \frac{\dot{R}(t)}{R(t_0)} r_0, \quad (20.13)$$

and thus the Hubble constant is

$$H = \frac{V}{r} = \frac{\dot{R}(t)}{R(t)}. \quad (20.14)$$

The deceleration of the expansion is described by the *deceleration parameter* q , defined as

$$q = -R\ddot{R}/\dot{R}^2. \quad (20.15)$$

The deceleration parameter describes the change of the rate of the expansion \dot{R} . The additional factors have been included in order to make it dimensionless, i.e. independent of the choice of units of length and time.

The value of the deceleration parameter can be expressed in terms of the Hubble constant and density. Substitute the definition of H and the expression of \ddot{R} (from Box 20.1) to the definition of q . We get

$$q = \frac{4\pi G}{3} \frac{\rho_0 R_0^3}{R^3 H^2}. \quad (20.16)$$

The density is usually expressed in terms of the *density parameter* Ω defined as $\Omega = \rho/\rho_c$, so that $\Omega = 1$ corresponds to the Einstein-de Sitter model. Thus

$$\Omega = \frac{8\pi G}{3} \frac{\rho_0 R_0^3}{R^3 H^2}. \quad (20.17)$$

Hence there is a simple relation between Ω and q :

$$\Omega = 2q. \quad (20.18)$$

The value $q = 1/2$ of the deceleration parameter corresponds to the critical density $\Omega = 1$. Both quantities are in common use in cosmology. It

should be noted that the density and the deceleration can be observed independently. The validity of (20.18) is thus a test for the correctness of general relativity with $\Lambda = 0$.

20.5 Cosmological Tests

A central cosmological problem is the question of which Friedmann model best represents the real Universe. Different models make different observational predictions. Recently there has been considerable progress in the determination of the cosmological parameters, and for the first time there is now a set of parameters that appear capable of accounting for all observations. In the following, some possible tests will be considered. These tests are related to the average properties of the Universe. Further cosmological constraints can be obtained from the observed structures. These will be discussed in Sect. 20.7.

The Critical Density If the average density ρ is larger than the critical density ρ_c , the Universe is closed. For the Hubble constant $H = 100 \text{ km s}^{-1} \text{ Mpc}^{-1}$, the value of $\rho_c = 1.9 \times 10^{-26} \text{ kg m}^{-3}$, corresponding to roughly ten hydrogen atoms per cubic metre. Mass determinations for individual galaxies lead to smaller values for the density, favouring an open model. However, the density determined in this way is a lower limit, since there may be significant amounts of invisible mass outside the main parts of the galaxies.

If most of the mass of clusters of galaxies is dark and invisible, it will increase the mean density nearer the critical value. Using the virial masses of X-ray clusters of galaxies (Sect. 19.2), one finds $\Omega_0 = 0.3$. Considerations of the observed velocities of clusters of galaxies indicate that the relative amount of dark matter does not increase further on even larger scales.

Neutrinos have a small mass (about 10^{-4} electron mass). A large neutrino background should have been produced in the big bang. In spite of the small suggested mass of the neutrinos, they can still form a part of the mass of the Universe. Nowadays it is known that the mass of the neutrino is too small to explain the dark matter problem. Also, they move at relativistic speeds and are

too “hot” to form the observed structures. Instead much recent work in cosmology has been based on the hypothesis of *Cold Dark Matter* (CDM), i.e. the idea that a significant part of the mass of the Universe is in the form of non-relativistic particles of an unknown kind.

It is understood that the mean density of the Universe cannot be reliably determined from the observations of the current constituents of the Universe. Instead, a very precise value for the mean density is found by studying the details of the cosmic background radiation, since that gives a single picture of the contents of the whole Universe. The measurements show that the density is exactly the critical density, which means that the space is flat. The most recent result of Planck obtained in 2015 gives a curvature of 0.001 ± 0.004 , meaning that the deviation from zero is at most 0.5 %.

The Magnitude-Redshift Test Although for small redshifts, all models predict the Hubble relationship $m = 5 \lg z + C$ for standard candles, for larger redshifts there are differences, depending on the deceleration parameter q . This provides a way of measuring q .

The models predict that galaxies at a given redshift look brighter in the closed models than in the open ones (see Fig. 20.4). Measurements of type Ia supernovae out to redshifts $z = 1.7$ using the Hubble Space Telescope have now shown that the observed q is inconsistent with models having $\Lambda = 0$. Assuming $\Omega_0 = 0.3$ these observations require $\Omega_\Lambda = 0.7$, where Ω_Λ is defined below in Box 20.1.

The Angular Diameter-Redshift Test Along with the magnitude-redshift test, the relation between angular diameter and redshift has been used as a cosmological test. Let us first consider how the angular diameter θ of a standard object varies with distance in static models with different geometries. In a Euclidean geometry, the angular diameter is inversely proportional to the distance. In an elliptical geometry, θ decreases more slowly with distance, and even begins to increase beyond a certain point. The reason for this can be understood by thinking of the surface of a sphere.

For an observer at the pole, the angular diameter is the angle between two meridians marking the edges of his standard object. This angle is smallest when the object is at the equator, and grows towards the opposite pole. In a hyperbolic geometry, the angle θ decreases more rapidly with distance than in the Euclidean case.

In an expanding closed universe the angular diameter should begin to increase at a redshift of about one. This effect has been looked for in the diameters of radio galaxies and quasars. No turnover has been observed, but this may also be due to evolution of the radio sources or to the selection of the observational data. At smaller redshifts, the use of the diameters of clusters of galaxies has yielded equally inconclusive results.

Evolutionary effects are the most important factors for the uncertainty. We know that the sizes of large scale structures are changing while the Universe is evolving. The evolution of the structures has been investigated using cosmological simulations (e.g. Millennium, Illustris Fig. 20.14). The simulations give estimates for the object sizes at different redshift, and these values can be compared with the actual observations. The results are consistent with the cosmological Λ CDM concordance model.

Basically the same idea can be applied to the angular scale of the strongest fluctuation in the cosmic microwave background. The linear size of these depends only weakly on the cosmological model and can therefore be treated as a standard measuring rod. Their redshift is determined by the decoupling of matter and radiation (see Sect. 20.6). Observations of their angular size have provided strong evidence that $\Omega_0 + \Omega_\Lambda = 1$, i.e. the Universe is flat.

Primordial Nucleosynthesis The standard model predicts that 25 % of the mass of the Universe turned into helium in the “big bang”. This amount is not sensitive to the density and thus does not provide a strong cosmological test. However, the amount of deuterium left over from helium synthesis does depend strongly on the density. Almost all deuterons formed in the big bang unite into helium nuclei. For a larger density the collisions destroying deuterium were more frequent.

Thus a small present deuterium abundance indicates a high cosmological density. Similar arguments apply to the amounts of ^3He and ^7Li produced in the big bang. The interpretation of the observed abundances is difficult, since they have been changed by later nuclear processes. Still, present results for the abundances of these nuclei are consistent with each other and with a density corresponding to Ω_0 about 0.04. Note that this number only refers to the mass in the form of baryons, i.e. protons and neutrons. Since the virial masses of clusters of galaxies indicate that Ω_0 is about 0.3, this has stimulated models such as the CDM model, where most of the mass is not in the form of normal baryons.

Ages The ages of different Friedmann models can be compared with known ages of various systems. The age t_0 of a Friedmann model with given Ω_0 and Ω_Λ is obtained by integrating equation (8) of Box 20.1. This gives

$$t_0 = H_0^{-1} \int_0^1 da (\Omega_0 a^{-1} + \Omega_\Lambda a^2 + 1 - \Omega_0 - \Omega_\Lambda)^{-1/2}. \quad (20.19)$$

This age is required to be larger than the ages of the oldest known astronomical objects.

If the density is critical and $\Lambda = 0$, $t_0 H_0 = 2/3$. Thus if $H_0 = 75 \text{ km s}^{-1} \text{ Mpc}^{-1}$, the age is $t_0 = 9 \text{ Ga}$. Larger values of Ω_0 give smaller ages, whereas positive values of Λ lead to larger ages. It has been a source of embarrassment that the best values of H have tended to give an age for the Universe only marginally consistent with the ages of the oldest astronomical objects. With the introduction of a positive cosmological constant and a slight downward revision of stellar ages this problem has disappeared. In 2003 the WMAP results gave an age of 13.7 Ga. A more recent value by the Planck satellite is $13.80 \pm 0.02 \text{ Ga}$.

The best current parameter values give 13–14 Ga for the age of the Universe.

The “Concordance” Model In summary there has been a remarkable recent convergence between different cosmological tests. The resulting model has a positive cosmological constant, and

most of the matter is cold and dark. It is thus referred to as the Λ CDM model. The best parameter values are $H = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_\Lambda = 0.7$, $\Omega_0 = 0.3$, with cold dark matter making up 85 % of the total density.

The concordance model is by no means definitive. In particular the reason for the cosmological constant is a major puzzle. In order to allow for the possibility that Λ is variable it has become customary to refer to it as *dark energy*, saving the term cosmological constant for the case of a constant Λ . Even if some alternative mechanism can produce the same effect as a non-zero Λ , finding at least one set of acceptable parameters, thanks to the WMAP and Planck satellites, is an important step forward.

An additional set of constraints on the cosmological model comes from the large-scale structures of the Universe. These constraints, which contribute further support for the concordance model, will be considered in Sect. 20.7.

20.6 History of the Universe

We have seen how the density of matter and of radiation energy and temperature can be computed as functions of the scale factor R . Since the scale factor is known as a function of time, it is possible to calculate how these quantities change with time.

When we look back in time R is decreasing. Since $\rho \propto R^{-3}$, $T \propto R^{-1}$, densities and temperatures at the beginning were so immense that all theories about the physical processes taking place are highly conjectural. Nevertheless attempts have been made at understanding the most fundamental properties of the Universe on the basis of modern theories of particle physics. For example, no indications of significant amounts of antimatter in the Universe have been discovered. Thus, for some reason, the number of matter particles must have exceeded that of antimatter particles by a factor of 1.000000001. Because of this *symmetry breaking*, when 99.999999 % of the hadrons were annihilated, 10⁻⁷ % was left later to form galaxies and everything else. It has been speculated that the broken symmetry originated in particle processes about 10⁻³⁵ s after the initial time.

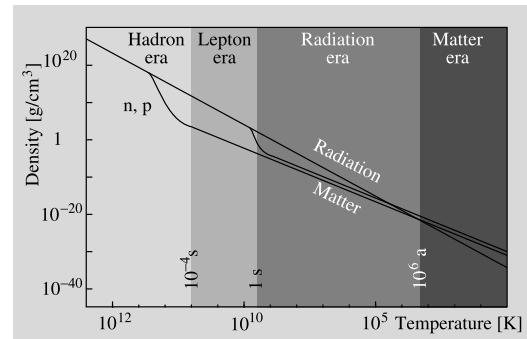


Fig. 20.12 The energy densities of matter and radiation decrease as the Universe expands. Nucleon–antinucleon pairs annihilate at 10^{-4} s; electron–positron pairs at 1 s

The breaking of fundamental symmetries in the early Universe may lead to what is known as *inflation* of the Universe. In consequence of symmetry breaking, the dominant energy density may be the zero-point energy of a quantum field. This energy density will lead to inflation, a strongly accelerated expansion, which will dilute irregularities and drive the density very close to the critical value. One may thus understand how the present homogeneity, isotropy and flatness of the Universe have come about. In the inflationary picture the Universe has to be very nearly flat, $\Omega_0 + \Omega_\Lambda = 1$. The inflationary models also make specific predictions for the form of the irregularities in the CMB. These predictions are in general agreement with what has been observed.

As the Universe expanded, the density and temperature decreased (Fig. 20.12) and conditions became such that known physical principles can be applied. During the hot early stages, photons and massive particles were continually changing into each other: high-energy photons collided to produce particle-antiparticle pairs, which then were annihilated and produced photons. As the Universe cooled, the photon energies became too small to maintain this equilibrium. There is a *threshold temperature* below which particles of a given type are no longer produced. For example, the threshold temperature for *hadrons* (protons, neutrons and mesons) is $T = 10^{12}$ K, reached at the time $t = 10^{-4}$ s. Thus the present building blocks of the atomic nuclei, protons and neutrons, are relics from the time 10^{-8} – 10^{-4} s, known as the *hadron era*.

Even before the hadron era, when the age of the Universe was about 10^{-12} s, the dark matter particles forming about 5/6 of all matter were born. They were decoupled from radiation around 10^{-9} s and started to form structures to which the ordinary matter started to condense hundred million years later.

The Lepton Era In the time period 10^{-4} –1 s, the *lepton era*, the photon energies were large enough to produce light particles, such as electron–positron pairs. Because of matter–antimatter symmetry breaking, some of the electrons were left over to produce present astronomical bodies. During the lepton era *neutrino decoupling* took place. Previously the neutrinos had been kept in equilibrium with other particles by fast particle reactions. As the density and temperature decreased, so did the reaction rates, and finally they could no longer keep the neutrinos in equilibrium. The neutrinos decoupled from other matter and were left to propagate through space without any appreciable interactions. It has been calculated that there are at present 600 such cosmological neutrinos per cubic centimetre, but their negligible interactions make them extremely difficult to observe. After the neutrino decoupling the amount of free neutrons kept decreasing as they decayed to protons and electrons.

The Radiation Era After the end of the lepton era, about 1 s after the initial time, the most important form of energy was electromagnetic radiation. This stage is called the *radiation era*. At its beginning the temperature was about 10^{10} K and at its end, about one million years later, when the radiation energy density had dropped to that of the particles, it had fallen to about 40,000 degrees. At the very beginning of the radiation era within a few hundred seconds helium was produced.

Just before the epoch of helium synthesis, the number ratio of free protons and neutrons was changing because of the decay of the free neutrons. After about 100 s the temperature had dropped to about 10^9 K, which is low enough for deuterons to be formed. All remaining neutrons were then incorporated in deuterons; these,

in turn, were almost entirely consumed to produce helium nuclei. Thus the amount of helium synthesised was determined by the number ratio of protons and neutrons at the time of deuterium production $t = 100$ s. Calculations show that this ratio was about 14 : 2. Thus, out of 16 nucleons, 2 protons and 2 neutrons were incorporated in a helium nucleus. Consequently $4/16 = 25\%$ of the mass turned into helium. This is remarkably close to the measured primordial helium abundance.

Only the isotopes ^2H , ^3He , ^4He and ^7Li were produced in appreciable numbers by nuclear processes in the big bang. The heavier elements have formed later in stellar interiors, in supernova explosions and perhaps in energetic events in galactic nuclei.

Radiation Decoupling. The Matter Era As we have seen, the mass density of radiation (obtained from the formula $E = mc^2$) behaves as R^{-4} , whereas that of ordinary matter behaves as R^{-3} . Thus the radiation mass density decreases more rapidly. At the end of the radiation era it became smaller than the ordinary mass density. The *matter era* began, bringing with it the formation of galaxies, stars, planets and human life. At present, the mass density of radiation is much smaller than that of matter. Therefore the dynamics of the Universe is completely determined by the density of massive particles.

Soon after the end of the radiation era, radiation decoupled from matter. This happened when the temperature had dropped to a few thousand degrees, and the protons and electrons combined to form hydrogen atoms. It was the beginning of the “*dark ages*” at redshifts $z = 1000$ –100, before stars and galaxies had formed, when the Universe only contained dark matter, blackbody radiation, and slowly cooling neutral gas.

At present, light can propagate freely through space. The world is transparent to radiation: the light from distant galaxies is weakened only by the r^{-2} law and by the redshift. Since there is no certain detection of absorption by neutral gas, there must have been a *reionisation* of the Universe. It is thought that this occurred around $z = 5$ –10.

20.7 The Formation of Structure

As we go backward in time from the present, the distances between galaxies and clusters of galaxies become smaller. For example, the typical separation between galaxies is 100 times their diameter. At the redshift $z = 101$ ($R(\text{now})/R(t) = 100$) most galaxies must have been practically in contact. For this reason galaxies in their present form cannot have existed much earlier than about $z = 100$. Since the stars were presumably formed after the galaxies, all present astronomical systems must have formed later than this.

It is thought that all observed structures in the Universe have arisen by the gravitational collapse of small overdensities. Whereas the presently observable galaxies have undergone considerable evolution, which makes it difficult to deduce their initial state, on larger scales the density variations should still be small and easier to study. These are the structures considered in the present section. The later evolution of galaxies is discussed in Sect. 19.8, and the formation of the Milky Way in Sect. 18.5.

The Statistical Description of Large-Scale Structure

Structure The departures from strict homogeneity in the Universe are random in character, and must therefore be described using statistical methods. Perhaps the most straightforward way of doing this is to take regions of a given size, specified in terms of their mass, and give the probability distribution for relative density variations on that scale.

A second method is to consider the spatial separations between individual objects such as galaxies or clusters. The distribution of these separations is used to define the *correlation function*, which is a measure of the clustering of the objects in question.

A third method to describe large-scale fluctuations is by means of the *power spectrum*. Here the density variations (in space or in projection on the sky) are represented as sum of waves. The power spectrum is the squared amplitude of these waves as a function of wavelength.

All three methods are representations of the density variations in the Universe, and they are

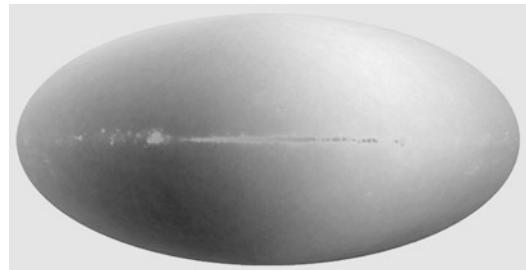


Fig. 20.13 The motion of the Milky Way in relation to the microwave background can be seen in the measurements of WMAP. One side of the sky is darker (colder), and the other side is lighter (warmer). *The horizontal stripe* is the densest part of the Milky Way. (Photo NASA)

theoretically closely related. However, in practice they are observed in different ways, and therefore, which representation is most suitable depends on what kind of observations are being analysed. The density variations are usually described by means of a spectral index n and an amplitude σ_8 to be introduced below.

The Growth of Perturbations In order to describe the growth of structures in the Universe, consider a given region containing the mass M . If its density is slightly larger than the mean density, its expansion will be slightly slower than that of the rest of the Universe, and its relative over-density will grow. The rate of growth as a function of mass depends on the relative importance of the material components of the Universe, dark matter, radiation, and ordinary baryonic matter.

It is assumed that there is an initial distribution of perturbations where the fluctuations with mass M have an amplitude that is proportional to $M^{-(n+3)/6}$. The spectral index n is a cosmological parameter to be determined from observations.

The first step in structure formation is when a given mass comes within the *horizon*, i.e. when there has been enough time since the big bang for light signals to cross the given region. During the radiation era the horizon mass grows proportionally to $t^{3/2}$, and thus the time at which the mass M comes within the horizon will be proportional to $M^{2/3}$. Any perturbation will initially be larger than the horizon mass, and while this is the case it grows in proportion to t . Once a matter

perturbation comes inside the horizon its amplitude will remain constant. This constant amplitude will behave as $t M^{-(n+3)/6}$, which is proportional to $M^{-(n-1)/6}$. If $n = 1$, the perturbations enter the horizon with an amplitude that is independent of mass. We shall see that the observed value of n is in fact very close to 1.

Perturbations of both dark and baryonic matter density will behave as described above during the radiation era. At the end of the radiation era at time t_{EQ} , when the mass densities of radiation and (non-relativistic) matter become equal, the amplitude of the perturbations will be given by the horizon mass $M_{EQ} \approx 10^{16} M_\odot$ for $M \ll M_{EQ}$, and will be proportional to $M^{-(n+3)/6}$ for $M \gg M_{EQ}$. After equality the dark matter perturbations will be free to start growing again as $t^{2/3}$, independent of mass.

Unlike dark matter, ordinary baryonic matter perturbations cannot grow as long as the Universe remains ionised. Instead there is a minimum mass of a collapsing gas cloud given by the *Jeans mass* M_J :

$$M_J \approx \frac{P^{3/2}}{G^{3/2}\rho^2}, \quad (20.20)$$

where ρ and P are the density and pressure in the cloud (see Sect. 6.11). The value of M_J before decoupling was

$$M_J = 10^{18} M_\odot \quad (20.21)$$

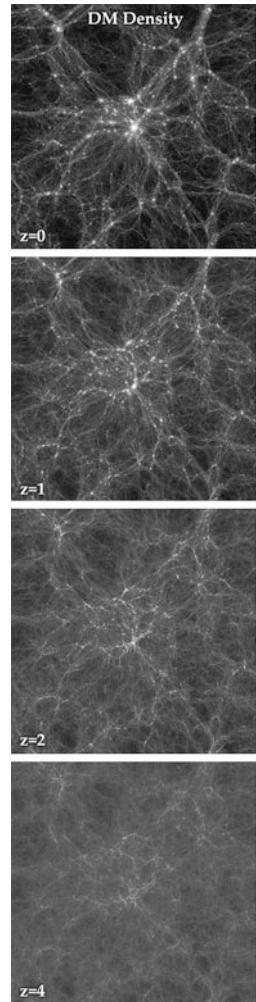
and after decoupling

$$M_J = 10^5 M_\odot. \quad (20.22)$$

The reason for the large difference is that before decoupling, matter feels the large radiation pressure ($P = u/3$, see Box 11.1). After decoupling, this pressure no longer affects the gas.

The large Jeans mass before decoupling means that overdense regions of normal gas cannot start growing before $z = 1000$. Rather than growing they oscillate like sound waves. After decoupling a large range of masses become Jeans unstable. By then density perturbations of dark matter have had time to grow, and the gas will therefore rapidly fall into their potential wells. The first stars will start forming in the collapsing regions, and will reionise the Universe.

Fig. 20.14 Cosmological simulations are used to study the evolution of the structures of the Universe. The international Illustris simulation shows how the network of dark matter evolves from redshift $z = 4$ (bottom frame) to $z = 0$ (topmost frame). (Illustris Collaboration, www.illustris-project.org)



Because the expansion of the Universe works against the collapse, the density of Jeans unstable regions grows rather slowly. In order to produce the observed systems, the density perturbations at decoupling cannot be too small. In models without dark matter the variations in the CMB predicted on this basis tended to be too large. In the CDM model the predicted variations are of the expected amplitude.

In the CDM model the amplitude of fluctuations on scales about M_{EQ} and smaller depends only weakly on mass. This is why the CDM model leads to an hierarchical description of structure formation. In this picture, systems of all masses above $10^5 M_\odot$ begin forming after decoupling. Because smaller systems will collapse

more rapidly, they are the first to form, at redshifts about 20. Once the first sources of light, starbursts or AGNs, had formed, they could reionise the gas. This marked the ending of the dark ages at redshifts $z = 10\text{--}5$.

The redshift of reionisation is still not well known, and is therefore treated as a parameter to be determined in tests based on large-scale structures. It is usually expressed by means of τ , the optical depth to electron scattering of the background radiation. A larger value of τ corresponds to a higher electron density, implying reionisation at a higher redshift. Since galaxies can be seen at redshifts larger than 6, the corresponding value of $\tau = 0.03$ represents a minimum. The final results (2012) of the observations with WMAP gave $\tau = 0.081 \pm 0.012$, which would have corresponded to a redshift 10.1 ± 1.0 . The final results (2015) of Planck gave 8.8 ± 0.1 as the mean redshift of reionisation.

One finally has to ask where the initial perturbations came from. An attractive feature of the inflationary model is that it makes specific predictions for these initial perturbations, deriving them from quantum effects at very early times. In this way the observed properties of the largest astronomical systems contain information about the earliest stages of our Universe.

Fluctuations of the Cosmic Microwave Background One important way of studying the large-scale structure of the early Universe is by means of the irregularities of the cosmic microwave background. The overdensities that were later to give rise to observed structures should also give rise to temperature variations of the CMB.

The temperature variations in the microwave background have been mapped, first by the COBE satellite, then by WMAP (Fig. 20.13) and Planck. A map of the CMB according to Planck is shown in Fig. 20.15. The observed variations are in qualitative agreement with the scenario for structure formation described above.

A more quantitative view of the observations is provided by the power spectrum of the observations shown in Fig. 20.16. This shows the amplitude of the temperature variations as a function of angular scale on the sky.

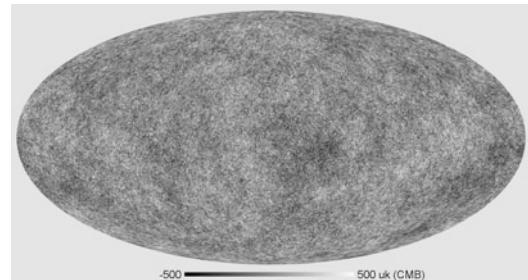
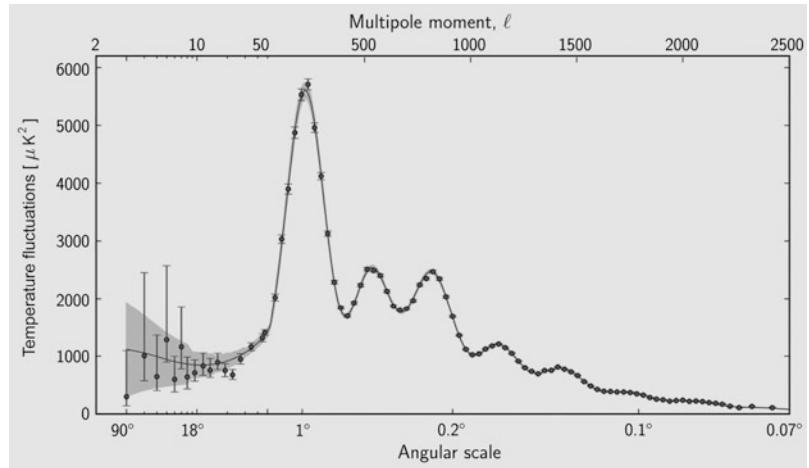


Fig. 20.15 Temperature distribution of the background radiation according to the measurements of the Planck satellite. Dark areas are colder, light areas warmer and more dense. (Planck Collaboration)

The physical processes we have described give rise to the features in the power spectrum. Thus the first peak is produced by perturbations that have just had time to collapse to a maximum density since equality of matter and radiation, before bouncing back. The linear size of this peak does not depend strongly on the exact parameters of the model, and its position (angular scale) can therefore be used as a standard measuring rod in the diameter—redshift test. The second and third peaks in the power spectrum, the “*acoustic peaks*”, correspond to perturbations that have bounced back and collapsed again, respectively. The positions and amplitudes of these features in the power spectrum depend on the cosmological parameters. Thus the WMAP observations of the CMB can be fitted using a model containing six free parameters: the Hubble constant, the densities of dark matter and baryons, the optical depth τ , and the amplitude σ_8 and shape n of the initial perturbations. The curvature of the Universe, $1 - \Omega_0 - \Omega_\Lambda$ can be set equal to zero, which determines the value of Ω_Λ . The results of WMAP and Planck have given values for these parameters in full agreement with the concordance model.

Very Large Scale Structure After decoupling structures on different scales are free to grow. The important dividing line is whether a given overdensity is still expanding with the rest of the Universe or whether it has had time to recollapse. Systems that have recollapsed will fairly rapidly

Fig. 20.16 Angular power spectrum of the temperature variations of the cosmic microwave background from the results of the Planck satellite. This gives the square of the amplitude of the temperature variations on different angular scales. (Planck Collaboration)



virialise, i.e. settle down into a stationary state. Such systems can be considered real astronomical systems.

The largest astronomical systems are clusters of galaxies. For some time there was controversy about whether even larger structures, superclusters, existed. The controversy was settled in the 1970's when it was realised that this largely depended on what was meant by superclusters. Regions of higher density existed on scales larger than clusters of galaxies (a few Mpc), all the way up to 100 Mpc, but they did not really form individual structures in approximate equilibrium like the clusters of galaxies.

Between these two alternatives are structures that are only now turning round and beginning to recollapse. One way of specifying the amplitude of the initial fluctuations is to give the scale of this transition point. This scale corresponds roughly to a present linear size of 8 Mpc, and for this reason the fluctuation amplitude is commonly given by means of σ_8 , the amplitude of the fluctuations at present scale 8 Mpc. Because of the way it has been chosen σ_8 should be close to 1.

Smaller systems will already have collapsed. There are many ways of statistically describing their large-scale distribution. One of the most studied descriptors of clustering is the *correlation function*, $\xi(r)$. Consider two infinitesimal volumes dV_1 and dV_2 separated by the distance r . If there were no clustering, the probability of finding galaxies in these volumes would

be $N^2 dV_1 dV_2$, where N is the number density of galaxies. Because of clustering this probability is actually $N^2(1 + \xi(r)) dV_1 dV_2$. The correlation function thus measures the higher probability of finding galaxies near each other.

Although there are ways of estimating the correlation function from the distribution on the sky, a more reliable estimate can be made by mapping the distribution of galaxies in three dimensions, using redshifts to determine their distances. The distribution from one such survey (SDSS, Sloan Digital Sky Survey) is shown in Fig. 20.17.

As expected, the general form of the correlation function is a power law (close pairs are abundant, distant pairs rarer). It has been known since the 1970's that the correlation function is approximately proportional to r^γ . The constant of proportionality is related to the amplitude σ_8 and the exponent γ is related to the index n of the initial perturbations. The observed value of γ is about 1.8, depending to some extent on the sample of objects studied.

At large separations the curve shows an unexpected maximum, observed for the first time in 2005. This is the same highest acoustic peak seen in the angular power spectrum of Planck. The observation confirms the Λ CDM cosmology and shows that the acoustic waves born around $z = 1000$ have remained till the present time and are seen in the distribution of galaxies. The same peak has been seen later in the larger SDSS survey.

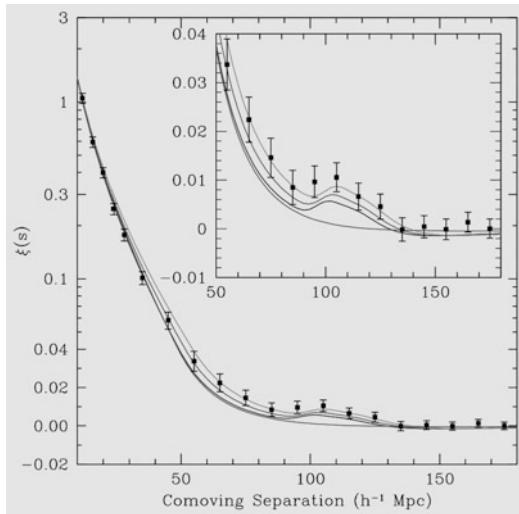


Fig. 20.17 Correlation function for the distribution of 47,000 bright red SDSS galaxies. The inset shows the found “bump” in more detail. The value of h is the Hubble constant divided by 100, or about 0.7. Thus a difference $100h^{-1}$ corresponds to a distance of about 150 Mpc (about 6 degrees on the sky). The bump tells about acoustic oscillations of baryons, i.e. large scale acoustic waves of the visible matter. (Eisenstein et al., 2005, ApJ 633, 560)

In order to determine the cosmological parameters the (three-dimensional) power spectrum of the galaxy distribution is usually used. It is then compared to theoretical power spectra that depend on the parameters in order to find the best-fitting model. The optical depth parameter τ does not affect these tests. The results have been completely consistent with the concordance model.

The Concordance Model Again We have discussed the use of the cosmic microwave background and the large-scale distribution of galaxies in order to determine the cosmological parameters. Both methods independently give consistent results. Combining them produces even more (formally) accurate values. Furthermore, these values are in agreement with the ones obtained from the traditional cosmological tests discussed in Sect. 20.5.

There are eight basic parameters describing the concordance model. According the final results of Planck (2015) they are:

Hubble constant $H_0 6.7 \pm 0.5 \text{ km s}^{-1} \text{ Mpc}^{-1}$, baryon density parameter $\Omega_b 0.049 \pm 0.001$,

mass density parameter $\Omega_0 0.259 \pm 0.006$, redshift of reionisation $z_{\text{ri}} 8.8 \pm 1.1$, spectral index $n 0.967 \pm 0.004$, fluctuation amplitude $\sigma_8 0.82 \pm 0.01$.

Other, less comprehensive tests, such as the gravitational lensing by cosmic structures, the velocities induced by mass concentrations, and the number of clusters of galaxies, have added support to the concordance model. This remarkable agreement between many independent determinations of the cosmological parameters have earned the model its name. Once the parameters necessary for a general description of the Universe are reasonably well known, other components, such as gravitational waves or neutrinos, can be included in the cosmological model. This new development has been called the era of precision cosmology. However, it should still not be forgotten that in the model one has to assume the presence of both dark matter and dark energy, neither of which is based on any evidence apart from their role in cosmology.

20.8 The Future of the Universe

The standard models allow two alternative prospects for the future development of the Universe. Expansion may either go on forever, or it may reverse to a contraction, where everything is eventually squeezed back to a point. In the final squeeze, the early history of the Universe would be repeated backwards: in turn, galaxies, stars, atoms and nucleons would be broken up. Finally the state of the Universe could no longer be treated by present-day physics. It is not known whether the collapse would continue to a single point or would the Universe avoid the singularity and start a new expansion.

However, according to the current observations, this is not the fate of the Universe. Instead, the expansion will continue forever, even at an accelerating pace.

In the open models the future is quite different. The evolution of the stars may lead to one of four end results: a white dwarf, a neutron star or a black hole may be formed, or the star may be completely disrupted. After about 10^{11} years, all

present stars will have used up their nuclear fuel and reached one of these four final states.

Some of the stars will be ejected from their galaxies; others will form a dense cluster at the centre. In about 10^{27} years the central star clusters will become so dense that a black hole is formed. Similarly the galaxies in large clusters will collide and form very massive black holes.

Not even black holes last forever. By a quantum mechanical tunnelling process, mass can cross the event horizon and escape to infinity—the black hole is said to “evaporate”. The rate of this phenomenon, known as the *Hawking process*, is inversely proportional to the mass of the hole. For a galactic-mass black hole, the evaporation time is roughly 10^{98} years. After this time, almost all black holes will have disappeared.

The ever expanding space now contains black dwarfs, neutron stars and planet-size bodies (unless the predictions of a finite proton lifetime of about 10^{31} years are confirmed; in that case all these systems will have been destroyed by proton decay). The temperature of the cosmic background radiation will have dropped to 10^{-20} K.

Even further in the future, other quantum phenomena come into play. By a tunnelling process, black dwarfs can change into neutron stars and these, in turn, into black holes. In this way, all stars are turned into black holes, which will then evaporate. The time required has been estimated to be 10^{1026} years! At the end, only radiation cooling towards absolute zero will remain (Fig. 20.18).

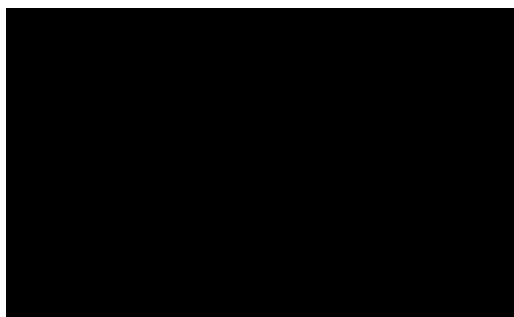


Fig. 20.18 According to current results the future of the Universe looks like this. After a very, very long time all matter will have changed to radiation. The wavelength of the radiation will increase infinitely in an ever expanding and cooling space

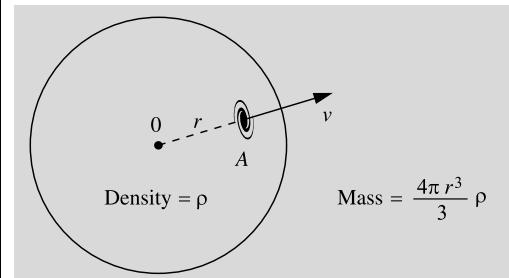
It is of course highly doubtful whether our current cosmological theories really are secure enough to allow such far-reaching predictions. New theories and observations may completely change our present cosmological ideas.

Box 20.1 (Newtonian Derivation of a Differential Equation for the Scale Factor $R(t)$) Let us consider a galaxy at the edge of a massive sphere (see figure). It will be affected by a central force due to gravity and the cosmological force

$$m\ddot{r} = -\frac{4\pi G}{3}\frac{r^3\rho m}{r^2} + \frac{1}{3}m\Lambda r,$$

or

$$\ddot{r} = -\frac{4\pi}{3}G\rho r + \frac{1}{3}\Lambda r. \quad (1)$$



In these equations, the radius r and the density ρ are changing with time. They may be expressed in terms of the scale factor R :

$$r = (R/R_0)r_0, \quad (2)$$

where R is defined to be R_0 when the radius $r = r_0$.

$$\rho = (R_0/R)^3 \rho_0, \quad (3)$$

where the density $\rho = \rho_0$ when $R = R_0$. Introducing (2) and (3) in (1), one obtains

$$\ddot{R} = -\frac{a}{R^2} + \frac{1}{3}\Lambda R, \quad (4)$$

where $a = 4\pi G R_0^3 \rho_0 / 3$. If (4) is multiplied on both sides by \dot{R} , the left-hand side yields

$$\dot{R}\ddot{R} = \frac{1}{2} \frac{d(\dot{R}^2)}{dt},$$

and thus (4) takes the form

$$d(\dot{R}^2) = -\frac{2a}{R^2} dR + \frac{2}{3} \Lambda R dR. \quad (5)$$

Let us define $R_0 = R(t_0)$. Integrating (5) from t_0 to t gives

$$\begin{aligned} \dot{R}^2 - \dot{R}_0^2 &= 2a \left(\frac{1}{R} - \frac{1}{R_0} \right) \\ &\quad + \frac{1}{3} \Lambda (R^2 - R_0^2). \end{aligned} \quad (6)$$

The constants \dot{R}_0 and a can be eliminated in favour of the Hubble constant H_0 and the density parameter Ω_0 . Because $\rho_c = 3H_0^2/8\pi G$,

$$\begin{aligned} 2a &= 8\pi G R_0^3 \rho_0 / 3 \\ &= H_0^2 R_0^3 \rho_0 / \rho_c = H_0^2 R_0^3 \Omega_0, \end{aligned} \quad (7)$$

where $\Omega_0 = \rho_0/\rho_c$. Using expression (6) and $\dot{R}_0 = H_0 R_0$ in (7), and defining $\Omega_\Lambda = \Lambda/(3H_0^2)$, one obtains

$$\begin{aligned} \frac{\dot{R}^2}{H_0^2 R_0^2} &= \Omega_0 \frac{R_0}{R} + \Omega_\Lambda \left(\frac{R}{R_0} \right)^2 \\ &\quad + 1 - \Omega_0 - \Omega_\Lambda \end{aligned} \quad (8)$$

as the basic differential equation governing $R(t)$.

For simplicity we now set $\Omega_\Lambda = 0$. Then the time behaviour of the scale factor R depends on the value of the density parameter Ω_0 . Because $\dot{R}^2 > 0$ always, according to (8)

$$\Omega_0 \frac{R_0}{R} - \Omega_0 + 1 \geq 0,$$

or

$$\frac{R_0}{R} \geq \frac{\Omega_0 - 1}{\Omega_0}. \quad (9)$$

If $\Omega_0 > 1$, this means that

$$R \leq R_0 \frac{\Omega_0}{\Omega_0 - 1} \equiv R_{\max}.$$

When the scale factor reaches its maximum value R_{\max} , then according to (8), $\dot{R} = 0$, and the expansion turns into contraction. If $\Omega_0 < 1$,

the right-hand side of (8) is always positive and the expansion continues forever.

The equation for the time dependence of the scale factor in the general theory of relativity contains the constant k which determines the geometry of space:

$$\dot{R}^2 = \frac{8\pi G R_0^3 \rho_0}{3R} - kc^2. \quad (10)$$

Equations (10) and (6) (or (8)) can be made identical if one chooses

$$H_0^2 R_0^2 (\Omega_0 - 1) = kc^2.$$

Thus, complete agreement between the Newtonian and the relativistic equation for R is obtained. The values of the geometrical constant $k = +1, 0, -1$ correspond respectively to $\Omega_0 > 1, = 1$ and < 1 . More generally, the condition for a flat model, $k = 0$, corresponds to $\Omega_0 + \Omega_\Lambda = 1$.

When $k = 0$, the time dependence of the expansion is very simple. Setting $\Omega_0 = 1$ and using (10) and (7), one obtains

$$\dot{R}^2 = \frac{H_0^2 R_0^3}{R}.$$

The solution of this equation is

$$R = \left(\frac{3H_0 t}{2} \right)^{2/3} R_0. \quad (11)$$

It is also easy to calculate the time from the beginning of the expansion: $R = R_0$ at the time

$$t_0 = \frac{2}{3} \frac{1}{H_0}.$$

This is the age of the Universe in the Einstein-de Sitter model.

Box 20.2 (Three Redshifts) The redshift of a distant galaxy is the result of three different mechanisms acting together. The first one is the peculiar velocity of the observer with respect to the mean expansion: the Earth moves about the Sun, the Sun about the centre of the Milky Way, and the Milky Way and the Local Group

of galaxies is falling towards the Virgo Cluster. The apparatus measuring the light from a distant galaxy is not at rest; the velocity of the instrument gives rise to a Doppler shift that has to be corrected for. Usually the velocities are much smaller than the speed of light. The Doppler shift is then

$$z_D = v/c. \quad (1)$$

For large velocities the relativistic formula has to be used:

$$z_D = \sqrt{\frac{c+v}{c-v}} - 1. \quad (2)$$

The redshift appearing in Hubble's law is the *cosmological redshift* z_c . It only depends on the values of the scale factor at the times of emission and detection of the radiation (R and R_0) according to

$$z_c = R_0/R - 1. \quad (3)$$

The third type of redshift is the *gravitational redshift* z_g . According to general relativity, light will be redshifted by a gravitational field. For example, the redshift of radiation from the surface of a star of radius R and mass M will be

$$z_g = \frac{1}{\sqrt{1 - R_S/R}} - 1, \quad (4)$$

where $R_S = 2GM/c^2$ is the Schwarzschild radius of the star. The gravitational redshift of the radiation from galaxies is normally insignificant.

The combined effect of the redshifts can be calculated as follows. If the rest wavelength λ_0 is redshifted by the amounts z_1 and z_2 by two different processes, so that

$$z = \frac{\lambda_2 - \lambda_0}{\lambda_0} = \frac{\lambda_2}{\lambda_0} - 1 = \frac{\lambda_2}{\lambda_1} \frac{\lambda_1}{\lambda_0} - 1,$$

or

$$(1+z) = (1+z_1)(1+z_2).$$

Similarly, the three redshifts z_D , z_c and z_g will combine to give an observed redshift z , according to

$$1+z = (1+z_D)(1+z_c)(1+z_g). \quad (5)$$

20.9 Examples

Example 20.1 (a) In a forest there are n trees per hectare, evenly spaced. The thickness of each trunk is D . What is the distance of the wood not seen for the trees? (Find the probability that the line of sight will hit a trunk within a distance x .)
 (b) How is this related to the Olbers paradox?

(a) Imagine a circle with radius x around the observer. A fraction $s(x)$, $0 \leq s(x) \leq 1$, is covered by trees. Then we'll move a distance dx outward, and draw another circle. There are $2\pi n x dx$ trees growing in the annulus limited by these two circles. They hide a distance $2\pi x n D dx$ or a fraction $n D dx$ of the perimeter of the circle. Since a fraction $s(x)$ was already hidden, the contribution is only $(1-s(x))n D dx$. We get

$$s(x+dx) = s(x) + (1-s(x))n D dx,$$

which gives a differential equation for s :

$$\frac{ds(x)}{dx} = (1-s(x))n D.$$

This is a separable equation, which can be integrated:

$$\int_0^s \frac{ds}{1-s} = \int_0^x n D dx.$$

This yields the solution

$$s(x) = 1 - e^{-n D x}.$$

This is the probability that in a random direction we can see at most to a distance x . This function s is a cumulative probability distribution. The corresponding probability density is its derivative ds/dx . The mean free path λ is the expectation of this distribution:

$$\lambda = \int_0^\infty x \left(\frac{ds(x)}{dx} \right) dx = \frac{1}{n D}.$$

For example, if there are 2000 trees per hectare, and each trunk is 10 cm thick, we can see to a distance of 50 m, on the average.

- (b) The result can easily be generalised into three dimensions. Assume there are n stars per unit volume, and each has a diameter D and surface $A = \pi D^2$ perpendicular to the line of sight. Then we have

$$s(x) = 1 - e^{-nAx},$$

where

$$\lambda = 1/nA.$$

For example, if there were one sun per cubic parsec, the mean free path would be 1.6×10^4 parsecs. If the universe were infinitely old and infinite in size, the line of sight would eventually meet a stellar surface in any direction, although we could see very far indeed.

Example 20.2 Find the photon density of the 2.7 K background radiation.

The intensity of the radiation is

$$B_\nu = \frac{2h\nu^3}{c^2} \frac{1}{e^{h\nu/(kT)} - 1}$$

and the energy density

$$u_\nu = \frac{4\pi}{c} B_\nu = \frac{8\pi h\nu^3}{c^3} \frac{1}{e^{h\nu/(kT)} - 1}.$$

The number of photons per unit volume is found by dividing the energy density by the energy of a single photon, and integrating over all frequencies:

$$N = \int_0^\infty \frac{u_\nu d\nu}{h\nu} = \frac{8\pi}{c^3} \int_0^\infty \nu^2 \frac{d\nu}{e^{h\nu/(kT)} - 1}.$$

We substitute $h\nu/kT = x$ and $d\nu = (kT/h) dx$:

$$N = 8\pi \left(\frac{kT}{hc} \right)^3 \int_0^\infty \frac{x^2 dx}{e^x - 1}.$$

The integral cannot be expressed in terms of elementary functions (however, it can be expressed as an infinite sum $2 \sum_{n=0}^{\infty} (1/n^3)$), but it can be evaluated numerically. Its value is 2.4041. Thus the photon density at 2.7 K is

$$N$$

$$= 16\pi \left(\frac{1.3805 \times 10^{-23} \text{ J K}^{-1} \times 2.7 \text{ K}}{6.6256 \times 10^{-34} \text{ Js} \times 2.9979 \times 10^8 \text{ ms}^{-1}} \right)^3 \\ \times 1.20206 \\ = 3.99 \times 10^8 \text{ m}^{-3} \approx 400 \text{ cm}^{-3}.$$

20.10 Exercises

Exercise 20.1 The apparent diameter of the galaxy NGC 3159 is 1.3', apparent magnitude 14.4, and radial velocity with respect to the Milky Way 6940 km s^{-1} . Find the distance, diameter and absolute magnitude of the galaxy. What potential sources of error can you think of?

Exercise 20.2 The radial velocity of NGC 772 is 2562 km s^{-1} . Compute the distance obtained from this information and compare the result with Exercise 18.1.

Exercise 20.3 If the neutrinos have nonzero mass, the universe can be closed. What is the minimum mass needed for this? Assume that $\Lambda = 0$, the density of neutrinos is 600 cm^{-3} , and the density of other matter is one tenth of the critical density.

Is there life elsewhere in the universe? And if there is, are there any intelligent beings? How did life emerge on the Earth, and, as a matter of fact, how do we define life and intelligence? These are probably the most interesting unsolved questions in science. During the last few decades a whole new field of astrobiology has evolved around these problems.

21.1 What Is Life?

There have been several science fiction tv-series showing extraterrestrial life forms. Usually the sentient beings have appeared quite humanlike, except of some appendices or other rather trivial features that try to make them unattractive. They may even have sex with humans. However, if the foreign life forms have evolved independently, interbreeding would be totally impossible.

The chemistry of foreign beings could be totally different from ours. Could we even understand that they are living beings? In fact, what is life? It seems that life is an elusive concept, difficult to define in terms of just a few properties. We have only one example of life, and therefore it is difficult to make general conclusions of its properties. However, we can assume that certain properties of the known life forms may be generalised also to foreign life.

Common features of all terrestrial life forms are reproduction and evolution. If living beings produced exact replicas of themselves, there would be no evolution and no adaptation to changing environments. Thus the reproduction

process must be slightly imperfect leading to a variety of descendants. This will give material for the natural selection, ‘survival of the fittest’.

Natural selection is a fairly general principle, working in some sense also outside of biology. Are there any other general principles related to the evolution of life? If we could find even one example of life that evolved independently of ours, this would vastly improve our knowledge.

Energy consumption is also characteristic of life. Life requires increasing order, i.e. decreasing entropy. Local decrease of entropy is not against thermodynamics: it only means that a living being must be able to take energy in some form and utilise it for reproduction, growth, motion or other purposes.

To produce similar offsprings a living being must have the ability to store information and pass it to its descendants. All terrestrial life forms use DNA or RNA molecules composed of nucleotides for storing information (see next section).

Carbon can combine to form very complex molecules. Silicon can also form large molecules but they are not as stable as carbon compounds, and silicon cannot form rings like carbon. Maybe some simple life forms could be based on silicon, or on something quite different that we have not even thought of.

Also a liquid solvent is needed. Our own life would not be possible without water. It remains liquid in a much wider temperature range than most other substances, which makes it a good solvent. Yet in the astronomical sense the tempera-

ture range is rather limited. In a colder environment, methane or ammonium might act as the solvent.

The basic building block of all terrestrial life forms is the cell. It has a membrane surrounding liquid cytoplasm. The cell membrane is semipermeable and functions as a two-way filter that lets certain molecules go in and others come out; this selective transport is mediated via specific proteinaceous channels. There are two kinds of cells, simpler prokaryotic cells and more complex eukaryotic cells. In eukaryotic cells the genetic material, in the form of DNA molecules, is inside a nucleus, surrounded by a nuclear membrane. In the prokaryotic cells there is no separate nucleus, and the DNA floats coiled in the cytoplasm.

Terrestrial life is divided into three domains, *Bacteria*, *Archaea* and *Eukarya*. Both *Bacteria* and *Archaea* contain usually a single prokaryotic cell. *Eukarya* contains all more complex beings, like animals and plants.

According to this scheme viruses are not alive although they have certain properties common to living beings. There are also some other molecules, such as viroids and prions, that are not classified as living; yet they are not quite inanimate.

If even the definition of terrestrial life leads to such borderline cases, a more general definition of all possible kinds of life is truly challenging. Facing this problem we have to restrict our discussion to life that, at least to some extent, resembles our own.

21.2 Chemistry of Life

The set of really important elements is relatively small; it includes hydrogen (H), oxygen (O), nitrogen (N), carbon (C), sulphur (S) and phosphorus (P). The heavier elements can be remembered by the mnemonic SPONC.

The importance of carbon is in its ability to make lots of different very complex molecules, which are essential for life. There are three basic types of molecules that function as common building blocks to all life: lipids for membranes, nucleotides and amino acids.

Amino acids consist of three different components, a carboxyl (COOH), an amine part (NH_2) and a side chain, which can be just a single hydrogen or a more complex structure. Altogether there are dozens of different amino acids, but only 20 of them are used in genetically coded proteins.

Amino acids can join to form more complex molecules, *proteins*. Typically, up to several hundred amino acids are needed for a protein molecule. Proteins have numerous functions: they support structures, they act as catalysts in nearly all biological reactions, in which case they are called *enzymes*, they carry messages as hormones, and so on.

Nucleotides are basic building blocks of the genetic material, DNA and RNA. A nucleotide has three components, a sugar, a phosphate group and a base. The phosphate parts are always the same. All DNA molecules have the same sugar part; also all RNA molecules have the same sugar part, which, however, differs from the sugar of the DNA molecules by having one additional oxygen atom. The base can be one of five different types, adenine (A), guanine (G), cytosine (C), thymine (T), or uracil (U). DNA molecules contain bases A, G, C and T, and RNA molecules A, G, C and U.

Nucleotides join to form long chains. The *deoxyribonucleic acid* or *DNA* (Fig. 21.1) consists of two such chains that are bound together as two intertwined helices. The corresponding bases join together by hydrogen bonds. The bases come always in matching pairs, AT, TA, CG, or GC.

A helix is like a screw that can be either left- or righthanded. All terrestrial DNA molecules have the same handedness or *chirality*: they are all of L-type, or lefthanded. The reason for the asymmetry is not quite understood.

The DNA molecule contains information on how to make proteins. Three consecutive base pairs form a code, called a *codon*, that specifies one amino acid. Usually thousands of such triplets together contain instructions for building a protein.

The basic units of heredity are called *genes*. They are regions of DNA whose final products are either proteins or RNA molecules.

Human cells contain some 25,000 genes, and the DNA consists of 3×10^9 base pairs. In plants

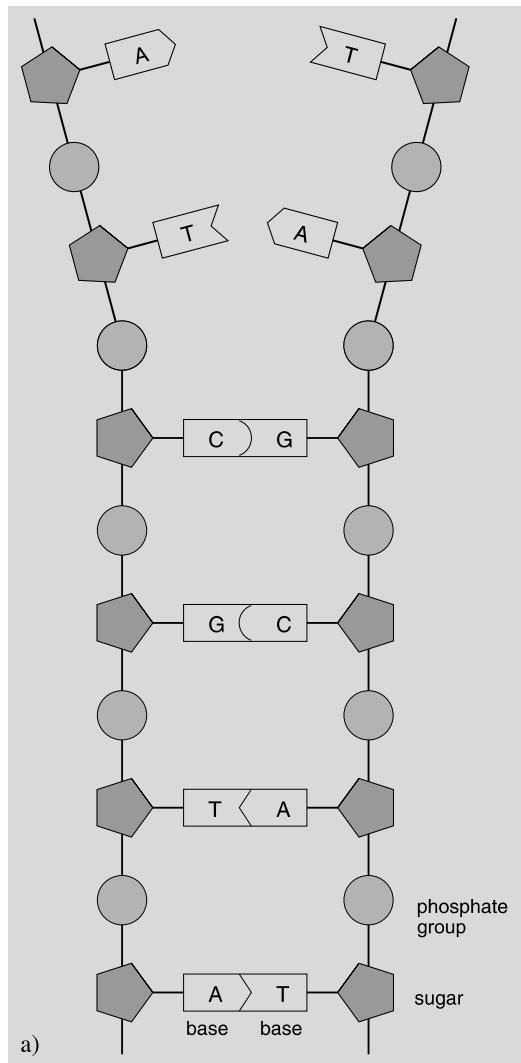


Fig. 21.1 (Continued)

the junk-DNA is highly variable; in some bacteria the amount is very small.

The DNA is the storage of the genetic code, but the code is not functional directly from the DNA. Instead a complex molecular translation machinery is needed to execute the instructions.

First, the instructions encoded in the genetic sequence are copied (or transcribed) into another type of nucleic acid. This *ribonucleic acid* or RNA resembles the DNA but has one more oxygen atom in its sugar. Different RNA molecules serve several roles in the translation process. Messenger RNA (mRNA) carries the information in the DNA to an organ called ribosome, composed of several RNA's and multiple proteins. There another RNA, the transfer RNA (tRNA), brings amino acids into the reaction site, and still another RNA molecule, or one of the ribosomal RNA's (rRNA) forms the linkages (peptide bonds) between the adjacent amino acids.

RNA molecules have to carry instructions for making just one protein, and thus they are much shorter than DNA molecules. Still they contain similar information, and in some simple life

the numbers can be much higher, but the simplest known bacterial genomes have only a few hundred genes. It has been estimated that the minimum number of genes needed for a living being is about 200–300.

Actually, only a part of the DNA contains genetic information. The rest is called junk-DNA since it has no known function. The fraction of

forms like viruses they can act as the storage of the genetic code.

21.3 Prerequisites of Life

Just after the big bang there were hardly any other elements than hydrogen and helium. Heavier elements are needed, both for life itself and for solid planets on which life can evolve. Thus at least some of the earliest stars had to explode as supernovas and eject their fusion products to the interstellar space.

Most stars in elliptic galaxies and globular clusters are very old and have a low metallicity. Therefore they are not very probable locations for life. The most suitable places for life seem to be the disk populations of spiral galaxies, containing young stars with high abundances of heavier elements.

Not all places of the galactic disk are equally profitable. Far out from the galactic centre the star birth rate and consequently the metallicity is low. Close to the centre, metallicity is high but the environment is rather hostile. Star density is very high, and therefore radiation is intense, and nearby stars disturb planetary orbits. A rough estimate is that only about 20 % of the stars of the Milky Way or a similar galaxy lie within the *galactic habitable zone*.

Assume that a star is in the galactic habitable zone. The star has its own habitable zone where habitable planets can exist. This zone is usually defined as the region where the temperature is between the freezing and boiling points of water. For a fast-rotating perfect blackbody ($A = 0$ in Eq. (7.51)) orbiting the Sun this would mean that its distance should be in the range 0.56–1.04 au. Planets, however, are not blackbodies, and the real situation is more complicated. If a planet has a high albedo, it reflects away most of the incident radiation and its temperature will be much lower. But depending on the chemical composition of the atmosphere and possible clouds the greenhouse effect may increase the temperature considerably. Many gases are transparent to visible light, allowing it to heat the surface. Most of the energy is emitted back in the infrared region,

which is effectively absorbed by these greenhouse gases, such as water vapour, carbon dioxide and methane, and thus remains trapped in the atmosphere.

If the star is cool, the habitable zone is very narrow. Hot stars have wider habitable zones, but their main sequence phase is short, giving little time for life to evolve. Thus main sequence stars not too different from the Sun are usually considered the best candidates for having habitable planets.

During the main sequence phase a star will become a little brighter, which will push the habitable zone slightly outwards. Thus the region where the temperature remains suitable for a long period of time is narrower than the habitable zone at any given moment. The continuously habitable zone can be defined as the region that remains habitable for a time that is comparable to the main sequence phase of the star. For the Sun the estimates of the width of this region vary at least from 0.06 au up to 0.2 au. The problem is how to model albedo and greenhouse effects over a very long period of time.

Binary stars are very common, but at least earlier it was thought that they could not have habitable planets, since planetary orbits would be complicated or unstable. However, there are two kinds of possible orbits that might be suitable. If the components of a binary are far away from each other, each component could have a planetary system of its own. Or, if it's a close binary, there can be distant planets orbiting the whole binary system.

21.4 Hazards

Even if life could emerge, there are many hazards that may wipe it out. By looking at the Moon we can see that meteor bombardment was very intense in the young solar system. Collision of a big asteroid or comet could be fatal, the immediate devastation caused by the explosion being only one of the consequences. The collision would eject a lot of dust to the atmosphere cooling the climate for several years. The mass extinction 65 million years ago seems to have been caused by



Fig. 21.2 In the young solar system small objects collided continuously to larger planets, threatening seriously the evolution of life. Later collisions and perturbations by planets have cleaned away most of the potentially dangerous objects. Still collisions take place even nowadays. In 1994 comet Shoemaker-Levy collided to Jupiter after disintegrating into several parts. Traces of the collisions of the fragments were seen as dark spots in the atmosphere of Jupiter. In June 20, 1908, an explosion happened in Tunguska, Siberia, hewing down trees in a couple of thousand square kilometres. The explosion was possibly caused by a comet with a diameter of about hundred metres

such an event. As the comet Shoemaker-Levy hitting Jupiter (Fig. 21.2) in 1994 showed, such collisions are still possible. Fortunately they are not very frequent any more.

Almost all of the currently known exoplanets are Jupiter-like giants. They seem to be neces-

sary for habitable planets, because their perturbations clean the young solar system from debris by ejecting it outside the planetary system. However, many of the known giant planets move on highly eccentric orbits, and they may disturb also the orbits of earthlike planets. Thus it is further required that the giant planets should be on nearly circular orbits and not too close to the star.

Also smaller planets have participated in clearing the regions around their orbits, which is reflected in the new definition of a planet (Sect. 7.1).

Seasons depend on the obliquity of the rotation axis and the eccentricity of the orbit. High values will lead to strong seasonal temperature variations. In the case of the Earth the Moon seems to have a stabilising effect; without the Moon the tilt of the axis would have varied much more, possibly causing more severe ice ages fatal to life. Hence also a relatively big moon seems to be in the shopping list of a habitable planet. However, recently there have been some objections to this requirement.

We have only recently started to understand the rather delicate balance and complex feedback effects working in the atmosphere. Currently the climate is warming due to the increasing greenhouse effect, but the Earth has experienced also quite opposite phases. If the albedo increases, the amount of energy reaching the surface decreases, glaciers and snow cover expand, and the amount of clouds increases till most of the atmospheric humidity is solidified as snow and ice. All this will increase the albedo further, speeding up this icehouse effect. There is geological evidence of global glaciation periods 750–580 million years ago and possibly also 2.3 billion years ago. The Snowball Earth hypothesis assumes that the climate cooled down for millions of years and the whole surface was covered by a layer of ice at least one kilometer thick. During the long cold period most of the living organisms, all of which at that time lived in water, became extinct. Volcanic activity was still going on, adding more carbon dioxide to the atmosphere. Finally the resulting greenhouse effect started to warm the climate.

There are many factors that seem to be crucial for life. Some of them may not look very important, but might still have made it impossible for

life to emerge. However, in many cases we don't know how important they really are or if they are equally crucial for foreign life forms.

21.5 Origin of Life

One way to try to understand the origin of the terrestrial life is to start with the available atoms and molecules and see if they could produce life. During the last decades there has been considerable progress, but the process is very complicated and not yet well understood. Here we can only outline briefly how it might have happened.

In a famous experiment in 1953 Harold Urey and Stanley Miller sent energy in the form of electric sparks through a gas mixture supposed to be similar to the early atmosphere of the Earth, containing methane, ammonia, hydrogen and water vapour. After a few days the solution contained several organic compounds, including some amino acids. At that time it was assumed that the early atmosphere was reducing. More recent studies suggest that this is not quite true, and the earliest atmosphere was rather neutral, containing mostly CO₂, CO, N₂, H₂O and maybe some H₂. Such an atmosphere would have produced organic compounds much slower, if at all.

Some amino acids have been found in meteorites. Thus they seem to have been already present in the nebula from which the planetary system condensed. Complex organic molecules have been found also in interstellar molecule clouds (Sect. 15.3). There have even been claims of detecting the simplest amino acid, glycine, but the results are controversial.

The next step, putting the basic blocks together to form DNA or RNA molecules, is much more difficult. This looks like the chicken and egg paradox: the information contained in the DNA is needed to make proteins, and proteins are needed to catalyse the production of the nucleotides, which are the building blocks of the nucleic acids. So which came first?

In the 1980's Sidney Altman and Thomas Cech found that some RNA molecules can act as catalysts. Since RNA resembles DNA, it can store genetic material to some extent. Thus there is no

need for the DNA and proteins. Even RNA fragments cannot be synthesised easily, but as they act as enzymes and can replicate, it is assumed that the initial chemical evolution first led to short and relatively simple RNA molecules. Eventually some of them combined to more complex ones, some of which were better adapted to the environment either by replicating faster or by being more durable. Thus the natural selection started to produce more complex molecules; this chemical evolution was working already before actual life emerged.

The first cell-like structures could evolve from asymmetric molecules or lipids, one end of which attracts water and the other end repels water. In water such molecules tend to form bi-layered membranes where the hydrophilic or water-attracting end points outwards and hydrophobic or water-repelling end inwards. Further on, such membranes form spontaneously spherical vesicles. If RNA happened to get inside such a membrane, it may have been protected from the environment, and could have been contained within its own chemical environment. In some cases this could have improved its replication, and thus led to further increase its concentration within the vesicle.

A rather common assumption is that the first primitive life forms were RNA life. However, recently this theory has been challenged. RNA has some drawbacks. It is not as stable as DNA, and its replication is not as accurate as the protein mediated replication of DNA. If the life was initially based on RNA, evolution led finally to the appearance of DNA molecules. Since DNA is superior to RNA due to its stability, it soon took over the role of information carrier.

Currently the energy of sunlight is utilised by plants and some bacteria in photosynthesis, which produces carbohydrates from water and carbon dioxide. There are also organisms that do not need sunlight but can use chemical energy to produce organic matter in a process called chemosynthesis. Such organisms have been found e.g. near hydrothermal vents on mid-ocean ridges (Fig. 21.3). These vents eject hot mineral-rich water to the ocean. Even though the temperature can be as high as 400 °C, the

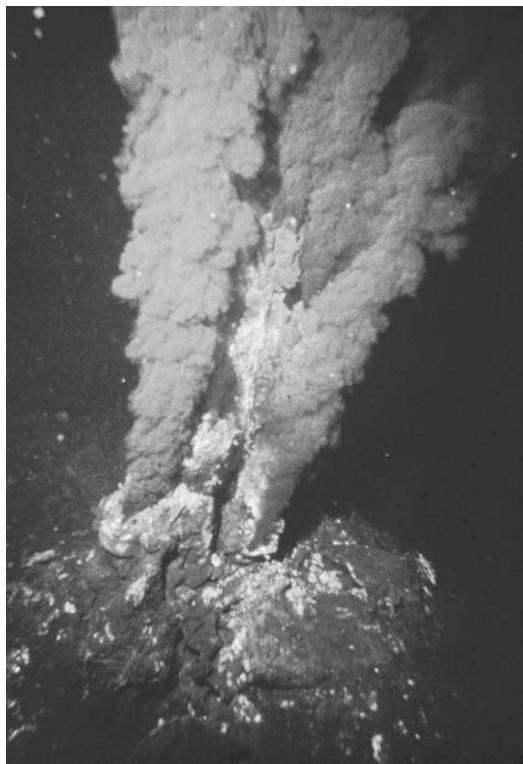


Fig. 21.3 Black smokers in the Mid-Atlantic Ridge are hydrothermal vents that sprout hot mineral-rich water. (Photo: P. Rona, Credit: OAR/National Undersea Research Program (NURP); NOAA)

high pressure prevents the water from boiling. Although this is too hot for life, there are regions around the vents where the temperature is suitable for such thermophiles. They could have been the first life forms, in which case life did not emerge in a Darwinian warm pond but in a hot pressure kettle. However, this is also a matter of debate.

This kind of bottom-up approach tries to build life from the simple constituents already available in the interstellar space. Another approach, the top-down method, tries to trace life back in time as far as possible.

The oldest sediment rocks on the Earth, found in Isua in western Greenland, are 3.8 Ga old. Since they contain sediments, deposited by water, and pillow lavas, formed in water, the temperature at that time could not have a value very different from the current one. The solar luminosity was then lower than nowadays, but the difference

was compensated by a higher amount of decaying radioactive materials and remanent heat of the recently born Earth.

Oldest signs of life are almost as old. These signs are, however, just isotope ratios that can be interpreted as results of bacterial life. The carbon isotope ^{12}C is about 100 times as abundant as the heavier isotope ^{13}C . The lighter isotope is somewhat more reactive and tends to be enriched in living organisms. In the Isua rocks there are sediments with a small excess of ^{12}C , which might indicate some kind of life.

In the Warrawoona Group in Australia there are 3.5 Ga old formations that look like stromatolites, mounds consisting of layers of microbial cells and calcium carbonate. If they are real stromatolites, they may have been formed by cyanobacteria, but this is still a matter of debate.

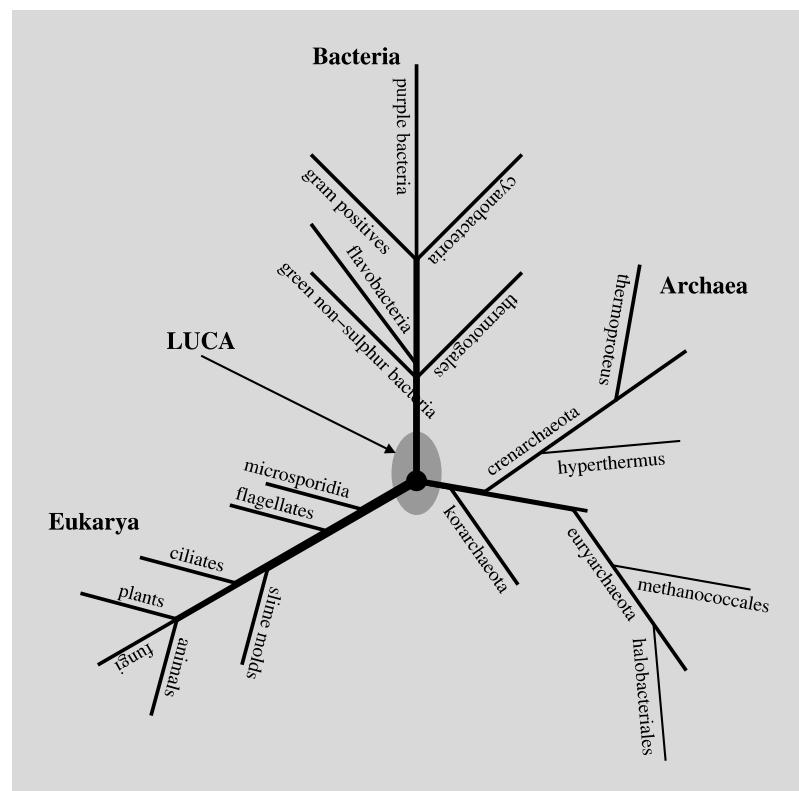
In the early times, at least for a billion years, photosynthesis was non-oxygenic. Cyanobacteria were possibly the first organisms capable of oxygenic photosynthesis. They started to produce oxygen, but initially it was dissolved in water and consumed in different oxidation reactions. Eventually also the amount of atmospheric oxygen started to rise, and 2.2 Ga ago it reached 10 % of the current value, i.e. about 2 % of the total abundance in the atmosphere.

First eukaryotes appeared in the fossil record 2.1 Ga ago and multicellular organisms 1.5 Ga ago. The fossil evidence becomes much clearer towards the end of the Proterozoic era. The Ediacara fauna, which is about 600 million years old, contains the oldest fossils of big and complex animals. These were softbodied animals. At the end of the Cambrian period 543 million years ago traces of the Ediacara fauna disappear and are replaced by a huge variety of new animals, many with protecting shields. This increase in the variety of life forms is called the Cambrian explosion.

All life forms use similar genetic codes, which indicates that they have the same origin. This forefather of all life is called LUCA, the Last Universal Common Ancestor.

Relationships of living beings can be studied by comparing their DNA or RNA. The more the molecules of two species differ, the more distant the species are in the evolutionary sense. These

Fig. 21.4 A simplified phylogenetic tree. A branch is the older the closer it is to the last common ancestor, LUCA. (Adapted from Webb: *Where is everybody?*)



distances can be plotted as a map, called the phylogenetic tree (Fig. 21.4).

The phylogenetic tree, as we now know it, has three branches, the domains of *Archaea*, *Bacteria* and *Eukarya*. The organisms closest to the root are thermophiles that live close to hydrothermal vents or in hot water. Obviously, the LUCA lived in such a hot environment. However, RNA molecules do not remain intact in such hot environments. If the earliest life was RNA life, it would have evolved in a cooler environment. Currently we do not know the real birthplace of life.

Although the phylogenetic tree points to a common origin, there may have been other starts, too, but natural selection has eliminated the other ones that were less competitive.

21.6 Are We Martians?

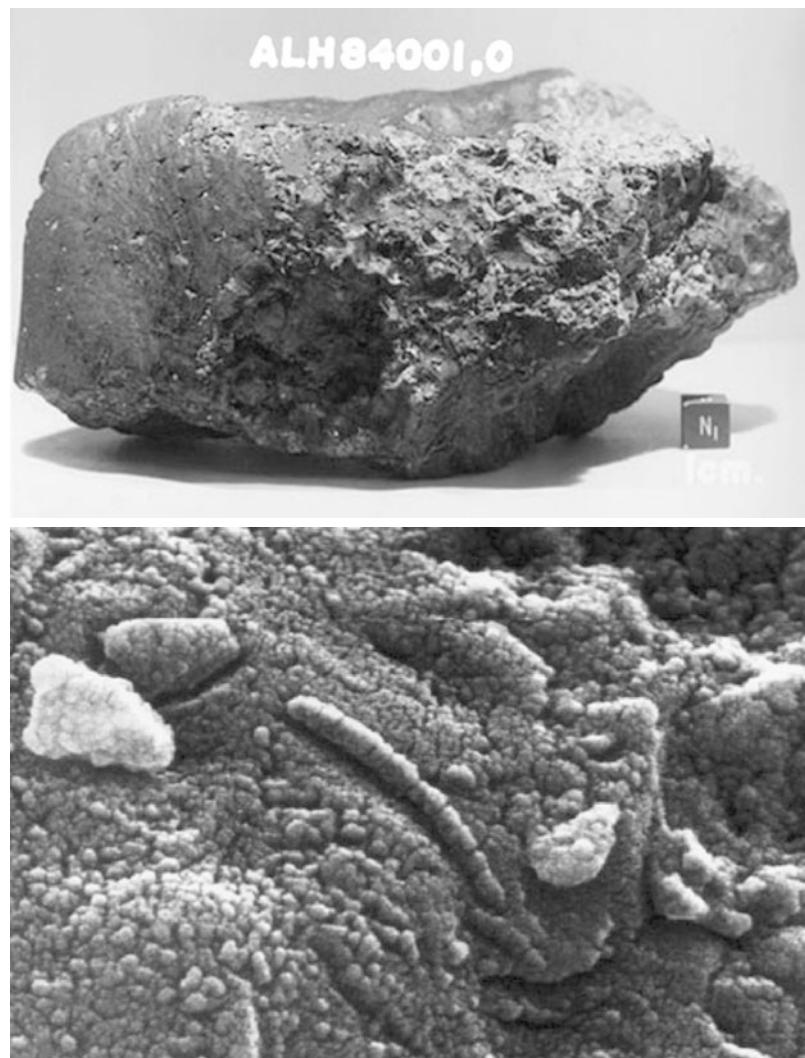
Mars and the Earth have very similar rotation periods and axial tilts. Since the notorious ‘discovery’ of canals, Mars and its inhabitants have been

a popular subject in science fiction stories. Later observations revealed a very thin atmosphere and low temperature, which make Mars a rather hostile place. Finally, the Viking landers showed a marred planet. Yet the possibility of some simple life forms cannot be excluded.

In 1984 a meteorite was found in the Allan Hills region in Antarctica and labelled as ALH 84001 (Fig. 21.5). The piece of rock was estimated to be 3.9 Ga old. Chemistry of the meteorite shows that it had originated on Mars; an impact had thrown it to an orbit that brought it to the Earth.

In 1996 a group of NASA scientists announced that the meteorite contained structures resembling fossilised microbes and compounds that could be products of living organisms, such as polycyclic aromatic hydrocarbons (or PAH) and magnetite. However, they can be produced by other processes, too. Not surprisingly, the results and the implications of Martian life caused a lot of scepticism. Only further Mars expeditions and

Fig. 21.5 The meteorite ALH 84001 found in Antarctica has a chemical composition that indicates its Martian origin. The meteorite contains substances that can be produced by living organisms. The wormlike structures resemble bacteria. It is, however, reasonable to ask whether one single sample can prove that once there was life on Mars. (NASA)



possible *in situ* experiments can decide whether there has been life on Mars.

In case there has really been life on Mars, there are several possibilities:

- Life originated independently on the Earth and Mars.
- Life originated only on the Earth and was then transported to Mars.
- Life originated only on Mars and was transported to the Earth.

It seems that life on the Earth emerged almost as soon as the conditions became favourable. It has been argued that the life appeared even too

quickly. This problem would be solved if life originated on Mars. The surface of the more distant and smaller Mars had cooled down faster to become habitable before the Earth. Thus life would have had more time to evolve on Mars, and was transferred to the Earth when conditions here became suitable. Thus our earliest ancestors could be Martian bacteria. Presently such considerations are, however, just speculations.

The idea of life spreading from one celestial body to another is known as *panspermia*. The idea dates back to the antiquity, but its first serious advocate was the Swedish chemist Svante Ar-

thenius, who published a book on the subject in 1908. Of the later proponents, Sir Fred Hoyle was the most famous. Panspermia fit well to his cosmology: the universe had no beginning, neither did life, but had always existed. Thus the tough problem of the origin of life was neatly avoided.

Now panspermia, in a certain more limited sense, begins to seem a little more possible theory. Primitive life forms can survive inside meteorites in the coldness and lethal radiation of the interplanetary space long enough to travel from one planet to another. Interstellar distances, though, are too long, and the probability of a meteoroid from one planetary system hitting another system is too low. It seems obvious that our life has originated here in our own solar system.

21.7 Life in the Solar System

Once there may have been life on Mars. Although probes have not detected signs of life, it is not impossible that there might still be some microscopic life, but we cannot expect to find any macroscopic life forms. The same is true for other places in our solar system.

Mercury has no atmosphere, Venus is too hot, and the giant planets have no solid surface. Besides the Earth and Mars this leaves only some satellites as possible habitats. It has also been speculated that there might be living things floating in the atmospheres of the giant planets, but the emergence and evolution of such things seems rather improbable.

Europa as well as some other icy satellites are nowadays considered potential places for life. Their surfaces are too cold, but tidal heating keeps the interiors warm enough. Big satellites rotate synchronously, but if the orbit is not perfectly circular, the orbital velocity varies according to Kepler's second law. Thus the satellite librates just like our Moon, which means that the direction of the tidal distortion keeps changing. Also the distance to the planet varies, and therefore the magnitude of the tidal force varies, too. These effects deform the satellite continuously giving rise to tidal heating.

Europa's surface is covered by ice. In some places the ice cover is broken into plates that ob-

viously have moved relative to each other. The rotation period of the surface differs from the rotation period of the magnetic field, which is frozen to the interior. Observations seem to indicate that the ice cover floats on an ocean.

The illumination under the ice is too dim for photosynthesis, but there might be thermal vents as in the oceans of the Earth. Hence the ocean could be a habitat for microbes that can utilise the thermal energy.

Titan is the only satellite with a thick atmosphere. The atmosphere is also rich in organic compounds, like methane. Methane dissociates rapidly, and thus the high methane content means that there must be a source of new methane. Living organisms are one such source, but because of its coldness Titan does not look like a promising place for life. A more plausible theory was that there was a methane ocean on Titan but the Huygens probe revealed a rather dry landscape, which, however, has signs of liquid flows. Radar images sent by the Cassini probe show some dark areas that might be methane lakes (Fig. 8.25).

21.8 Detecting Life

If we find a potentially habitable planet, is there any hope that we can see if life has emerged on it? The question can be answered by studying whether we can find life on the Earth from satellite observations. In 1990 the Galileo probe made just such experiments, and it seems that it is indeed possible to detect life, at least the kind of life we have on the Earth. Similar observations of the Moon showed no traces of life.

The detection is based on spectroscopic observations that can reveal some signatures of life. These signatures are in the infrared part of the spectrum, thus requiring observations made outside the atmosphere.

There are two emission features that are strong indicators of life, ozone and methane. Photosynthesis is the most probable source of molecular oxygen, which is then broken into two oxygen atoms by ultraviolet radiation. The free oxygen atoms join to molecules to form ozone. Methane is also produced by living beings. It is quickly oxidised, and has to be continuously replenished to

keep the level noticeable. However, there can be large reservoirs of methane, particularly in cold environments. Thus methane itself is not a sign of life, but if it is found together with ozone, the evidence becomes more convincing.

Another feature is the infrared reflectance spectrum of the green plants. Chlorophyll absorbs visible light, particularly blue and red, but there is a distinct cutoff called the red-edge, seen as a steep gradient of the spectrum between 690 and 740 nm. Longer wavelengths are very effectively reflected to avoid excessive heating.

21.9 SETI—Detecting Intelligent Life

Mankind has been sending radio transmissions for almost a century. Our radio signals are now filling a sphere with a radius of almost one hundred lightyears. Another civilisation orbiting a nearby star might be able to pick up this transmission with a big radio telescope. Such leakage radiation is, however, very weak. Sensitivity of our own radio receivers has increased enormously since they were invented. Thus it has been possible to reduce the power of the transmitters, and the signals leaking to space have become weaker. Also, more and more signals are sent in cables and optical fibres. If another civilisation has undergone similar development, detecting leakage signals is extremely difficult. Chances are much better if the signal has been sent intentionally towards potential receivers in the hope that somebody will detect it.

One might think that galaxies and star clusters are worth listening, since there are many stars in the narrow beam of the telescope. Unfortunately, it is not quite so. Other galaxies are so far away that the signal could be too weak to be detected. Globular clusters consist of very old stars with low metallicity. Thus the probability of finding a habitable planet is tiny indeed. Open clusters are relatively young, and life may not have had time to evolve to a communicating civilisation. Thus galaxies and star clusters are not the best places to search for signs of life.

What frequency should we use? If the sending and receiving party have developed radio astronomy, they must be aware of certain common frequencies, like the hydrogen 21 cm radiation. Such

a wavelength itself may not be a good choice because of the background noise, but some of its multiples or a sum of two common frequencies might fall in the quiet part of the radio spectrum. A good frequency could be the H₂O maser emission at 22 GHz. Around this frequency the sky is pretty quiet except for a few sources. But it is not enough to listen to those frequencies only, since they are Doppler shifted due to the relative motion of the transmitter and receiver. And if the transmitter and/or receiver are on planets orbiting a star, the Doppler shift will change periodically. Fortunately, current receivers are capable of following millions of frequencies simultaneously.

Radio emission from natural sources can be steady noise or vary in a periodic, quasiperiodic or chaotic manner. If we want to send a signal to be recognised as artificial, it should contain a pattern that cannot arise naturally. It could e.g. contain an increasing number of pulses representing the first few prime numbers.

Although most SETI research concentrates on radio frequencies, also optical wavelengths have recently been considered seriously. Pulsed lasers pack a lot of energy to a short pulse (lasting typically one nanosecond) confined to a very narrow wavelength band and a narrow beam. The flash can be even brighter than the central star. If such a signal were pointed towards us, it should be relatively easy to detect. Such optical SETI research, or OSETI, has already been started, but it is still behind the radioastronomical SETI.

In 1974 Frank Drake used the Arecibo radio telescope to send a message towards the globular cluster M13 (which is not a good place for life). The message contained 1679 pulses (Fig. 21.6). This number has exactly two factors, 23 and 73. Thus the receiver, who obviously must understand some mathematics, could guess that the message contains a two-dimensional picture. If we ever detect such a message, we can be pretty certain of its artificial origin, even if we were not able to interpret the message.

The first serious SETI project (Search for ExtraTerrestrial Civilisations) was carried out in 1960, also by Frank Drake. This project Ozma observed two nearby stars, τ Cet and ε Eri, at the 21 cm wavelength. Since then radio technology

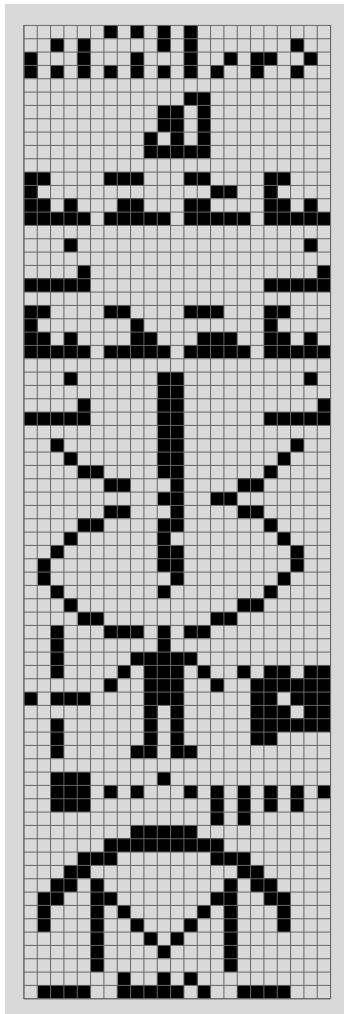


Fig. 21.6 The signal sent with the Arecibo radio telescope consists of 1679 pulses. The number 1679 has only two factors, 23 and 73, so obviously the signal represents a twodimensional image of 23×73 pixels. A lot of information about the sender is coded in the image, but it is difficult to say whether the receiving civilisation could interpret it. In any case, the artificial origin of the signal is indisputable

has improved enormously, and currently we are able to listen to a huge number of channels at the same time.

There are two basic search strategies. In a targeted search we listen to a few objects that are potential candidates for harbouring life. In a wide-sky survey large areas of the sky are scanned. Most current projects belong to the latter category.

Observing time with large telescopes is expensive, and priority is given to projects that can be expected to produce positive results. In some projects, like SERENDIP, this problem is avoided by having the receiver sit piggyback on some other instrument and listen to whatever the telescope happens to be observing. Thus the SETI project does not need any dedicated time of its own. The drawback is, of course, that many of the target areas may not be interesting in the SETI sense.

Detecting a potentially artificial signal among the huge amount of data requires a lot of computing power. The seti@home project has connected millions of computers to an enormous virtual machine to analyse the data. Anyone with a computer connected to the Internet can load a screen-saver program that will automatically fetch packets of data and send back the results.

Thus far not a single message sent by another civilisation has been confirmed. There have been some interesting cases, but they have been single bursts of unknown origin. They have not been detected later, not even with more sensitive instruments.

21.10 Number of Civilisations

Although no extraterrestrial civilisation has been found, we can try to estimate their number. The SETI pioneer Frank Drake suggested a formula for calculating the number of civilisations in the Milky Way capable of communication at a given instant:

$$N = R \times f_p \times f_h \times f_l \times f_i \times f_c \times L, \quad (21.1)$$

where N is the total number of communicating civilisations in the Milky Way, R is the annual birth rate of stars, f_p is the fraction of stars possessing planets, f_h is the fraction of planets being habitable, f_l is the fraction of habitable planets having some kind of life, f_i is the fraction of these planets having intelligent life, f_c is the fraction of intelligent civilisations that have developed means for interstellar communication, and finally L is the time in years that such a civilisation has been communicating. All the f -factors are probabilities that are in the range $[0, 1]$.

The astronomical factors (R , f_p and f_h) are the only ones that are known with any accuracy. The biological factors, f_l and f_i , involve a lot of guesswork. The last two factors, f_c and L , are even harder, since they are related to the socio-logical behaviour of the exocivilisation.

Actually, the formula was intended as the basis of the agenda of an influential SETI meeting held in Green Bank in 1961. It splits the problem nicely into smaller subproblems that can be discussed separately. But using the formula to find the actual number of civilisations is not very meaningful, since so many of the factors are totally unknown. In the most “optimistic” case we could find out that the distance between neighbouring civilisations is just a few parsecs, but giving the probabilities small (possibly more realistic) values, their product might be so minute that we ought to be alone in the Milky Way. At least the formula shows how little we know.

Even if favourable conditions and evolution of communicating civilisations were relatively common, the last factor may turn out to be the limiting one. If the lifetime of a civilisation is short compared to the age of the universe, the chances of hearing a message from another star are poor.

Earlier many astronomers seemed to think that exocivilisations would not be that rare, while biologists showed that the evolution of life had so many obstacles that we should not expect to find other civilisations in our neighbourhood. Now we understand better both the biochemistry of early life and the many problems in having a habitable

planet. Although opinions vary considerably, we might guess that very simple microbe-like life is relatively common, but intelligent, communicating beings might be extremely rare.

21.11 Exercises

Exercise 21.1 Calculate the limits of the habitable zone of the Sun assuming the planet is a fast-rotating blackbody with a Bond albedo of 0.3. What is the continuously habitable zone, if the luminosity of the Sun was originally 0.7 times the current value?

Exercise 21.2 Assume there are n stars in a cubic parsec and a fraction p of them have communicating civilisations. What is the average distance between two neighbouring civilisations? Apply the result to the solar vicinity. The stellar density can be estimated from Table C.17. What is the average distance between nearest civilisations, if the probability of a star having a planet with a civilisation is (a) 0.01, (b) 0.00001?

Exercise 21.3 An asteroid with a diameter of one hundred metres is approaching the Earth. Estimate the minimum value of the kinetic energy released in the collision. Compare the result with the Hiroshima atomic bomb. The energy of the bomb was equivalent to 15 kilotons of TNT. One ton of TNT corresponds to the energy of 4.184×10^9 joules.

The Copernican principle means that the Earth has gradually lost its special status: it is just a planet orbiting a relatively common star in an ordinary galaxy. But how common planetary systems are? Nowadays we think that they belong to the normal life of a star. However, it is difficult to make ground based observations of *exoplanets*, i.e. planets orbiting some other star than our Sun. The light reflected by the planet is so dim that it is swamped with the light of the star. The situation is, however, improving thanks to evolving observation methods.

22.1 Other Planetary Systems

At the beginning of 2016 over 1300 planetary systems and a total of over 2000 planets had been found. The numbers have increased rapidly during the last few years. It seems that the birth of planetary systems is more and more closely related to the birth process of stars.

Due to the observing methods most of the known exoplanets are rather massive. Recently also planets about the size of the Earth have been detected. The next big question is if any of them is suitable for harbouring life.

Origin of planetary systems have been studied using theoretical calculations. Most calculations have dealt with single stars around which stable planetary orbits are easier to find. However, stable orbits are also possible around wide binary stars either close to one of the components, or so far that the planet orbits both components.

The Milky Way contains about 10^{11} stars. Maybe 10^9 – 10^{10} of them have suitable conditions for a planetary system.

The first exoplanet was found in 1992. It was quite a strange object, orbiting a pulsar. Pulsars are final stages of stellar evolution, and planets orbiting the star should be destroyed.

Some earlier observations had given hints about exoplanets but they were not conclusive. The first exoplanet orbiting an ordinary star was found in 1995 around β Pictoris.

In addition to planets orbiting a star there are some indications of *rogue planets* that have escaped from planetary systems. Star density in star birth regions can be so high that the perturbations of other stars throw planets away from their original orbits. Such runaway planets are even more difficult to detect than planets orbiting other stars.

22.2 Observational Methods

Although only a few planets can be observed directly with current methods, the existence of planets can be deduced using several indirect methods.

Astrometric methods are based on perturbations of the proper motion or radial velocity of the star. If the mass of the star is big enough the centre of mass will be clearly displaced from the star. For example, the centre of mass of our solar system is outside the surface of the Sun, and the Sun as well the planets orbit around this point.

Therefore the motion of the Sun through space is not linear but shows particularly the 12 year period of Jupiter's orbital motion.

Earliest observations were based on similar astrometric measurements as with binary stars. The wobble of the star caused by a planet is so minute that the results were not conclusive. The advantage of the method is that it does not depend on the angle between the orbital plane and the line of sight.

Much more precise results have been obtained by studying the periodic changes of the radial velocities of stars using the Doppler effect. The radial velocities of many nearby stars show indeed periodic variations indicating that they have massive invisible companions. The method makes it possible to detect Jupiterlike planets around nearby planets, but smaller ones are too light to cause detectable variations in the radial velocity of the star. This is why most of the currently known exoplanets are more massive than Jupiter.

This method works the better the closer the observer is to the orbital plane of the orbit. If the orbital plane is almost perpendicular to the line of sight no Doppler shift cannot be seen.

A third method is based on transits of planets: when a planet is between the star and the observer it occults a part of the stellar disk. Such events have already been observed with ground based telescopes, but especially the *Kepler* spacecraft sent by NASA in 2009 has investigated them by following about one hundred thousand stars in the galactic habitable zone. The method works, of course, only if the observer is very close to the orbital plane of the planet. Yet over a thousand planets have been found, many of which are about the size of the Earth. Planetary transits are seen as very regular dimming of the star. At the same time the satellite has observed also other kinds of brightness variations, thus advancing studies of variable stars.

A fourth method is gravitational lensing. If a star passes in front of a background star its gravity will bend the light, and the background star will look brighter. If the front star has planets the brightening is slightly stronger than without planets.

Also rogue planets can be detected with the transit or gravitational lens methods. Then the

brightness will change only once and for a very short period. Interpretation of such events is problematic: is it really a planet or a phenomenon of the star itself.

A fifth method is timing, which is well applied to regularly variable stars. A planet causes the distance of the star vary slightly, which changes the periodicity of the brightness variation of the star in a regular manner.

This was the method that gave the first convincing evidence of exoplanets in 1992 when the pulsation of the pulsar PSR B1257+12 showed regular variation. The discovery was strange since a pulsar should not have planets. The objects may be second generation planets that condensed from the material ejected by a supernova explosion. Later the same method has also revealed some planets around variable stars.

Direct imaging of exoplanets is not yet an effective search method. The first images show objects whose existence has already been confirmed with other methods (Fig. 22.1).

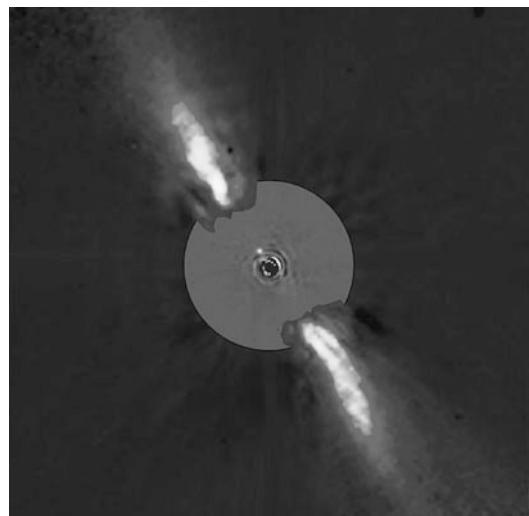


Fig. 22.1 A detailed image of the circumstellar disk around β Pictoris was obtained with the ESO ADONIS adaptive optics system at the 3.6 m telescope at La Silla, Chile, and the Observatoire de Grenoble coronagraph at the wavelength of 1.25 microns in 1996. The full extent of the disk is about 1500 AU. The area around the star masked by the coronagraph is only 24 AU (corresponding to a distance from the Sun to halfway between Uranus and Neptune). No planets are directly seen but their gravitational effects can be detected as a bending of the main plane in the inner part of the disk. (ESO)

Also young recently born stars can indicate planetary systems. Many stars are surrounded by a *dust disk* (Fig. 22.2). The easiest way to explain the infrared radiation of some young stars is to assume a massive disk surrounding the star and interacting with the star. Later the disk can evolve to a planetary system. Such disks have been detected around e.g. Vega and β Pictoris.

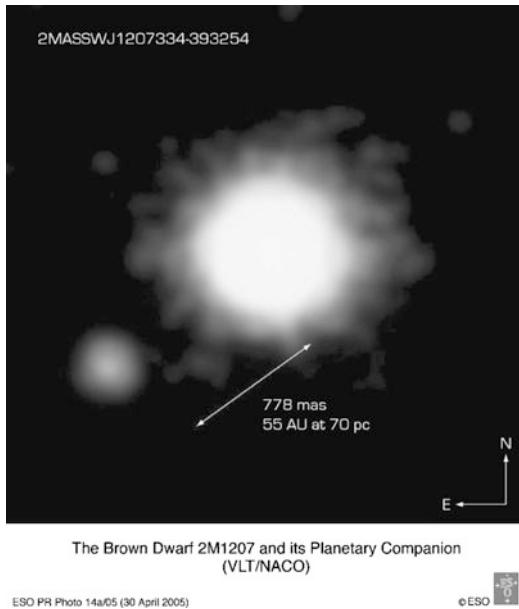


Fig. 22.2 The first image of an exoplanet was obtained in 2004 with the VLT telescope. The mass of the planet is about five times the mass of Jupiter and it orbits a brown dwarf at a distance of 55 AU. (ESO)

Fig. 22.3 Orbit and masses of some extrasolar planets. Open circles are planets orbiting a pulsar. Planetary masses are given in Jupiter's masses. The figure contains some 1200 planets that were known reasonably well at the beginning of 2016. Actually, a lot more planets have been detected, but their data are yet quite uncertain

22.3 Properties of Exoplanets

For practical reasons it is easiest to find big planets orbiting close to a star. They affect the position of the star most strongly, and due to the short orbital period the periodicity of the variations is most easy to detect. Therefore many of the exoplanets are such “hot jupiters”. This is seen in the distribution of Fig. 22.3. A great majority of the exoplanets found this far are clearly more massive than the Earth. This is another example of a statistical bias caused by a selection effect typical in astronomy. There may be a lot of Earth-sized planets but they are extremely difficult to observe.

The planetary orbits of our own solar system are nearly circular. If Mercury is excluded the eccentricity of every orbit is less than 0.1. Or-

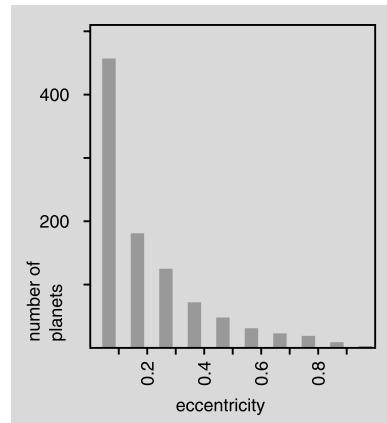
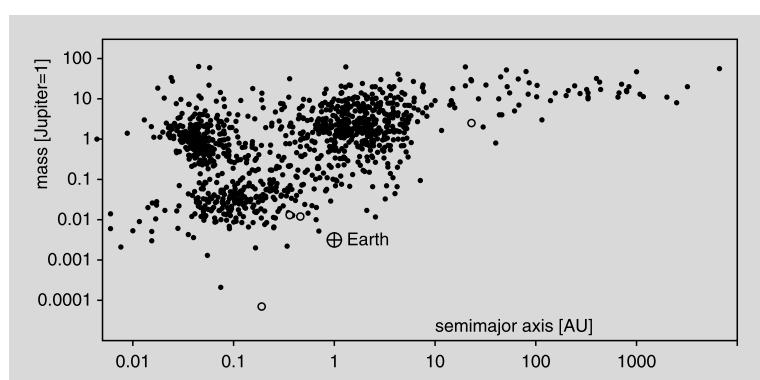


Fig. 22.4 Distribution of the eccentricities of the known exoplanets



bits of the exoplanets, on the other hand, can be very eccentric (Fig. 22.4). Theories concerning the evolution of planetary systems must be able to explain how the eccentricities decrease by, for example, collisions of minor bodies or mutual perturbations of planets.

When also solid earthlike planets can be observed the next question is, if any of them has an environment suitable for life. After that we can investigate if their spectra show features typical for life mentioned in the previous chapter.

22.4 Exercises

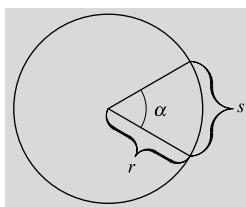
Exercise 22.1 How much the apparent place of the Sun oscillates due to Jupiter's orbital motion as seen from the distance of 10 parsecs.

Exercise 22.2 A distant observer detects the brightness of the Sun changing when Jupiter transits the Sun. How big is this change in magnitudes? And what is the change when the Earth transits the Sun?

A.1 Geometry

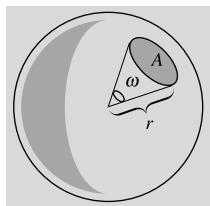
Units of Angle and Solid Angle *Radian* is the angular unit most suitable for theoretical studies. One radian is the angle subtended by a circular arc whose length equals the radius. If r is the radius of a circle and s the length of an arc, the arc subtends an angle

$$\alpha = s/r.$$



Since the circumference of the circle is $2\pi r$, we have

$$2\pi \text{ rad} = 360^\circ \quad \text{or} \quad 1 \text{ rad} = 180^\circ/\pi.$$



In an analogous way we can define a *steradian*, a unit of solid angle, as the solid angle subtended by a unit area on the surface of a unit sphere as seen from the centre. An area A on the

surface of a sphere with radius r subtends a solid angle

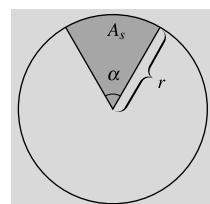
$$\omega = A/r^2.$$

Since the area of the sphere is $4\pi r^2$, a full solid angle equals 4π steradians.

Circle

$$\text{Area } A = \pi r^2.$$

$$\text{Area of a sector } A_s = \frac{1}{2}\alpha r^2.$$



Sphere

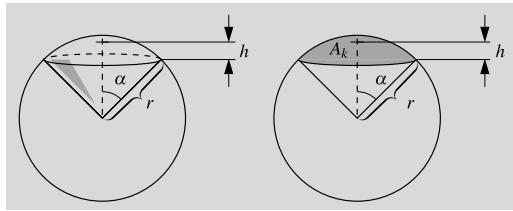
$$\text{Area } A = 4\pi r^2.$$

$$\text{Volume } V = \frac{4}{3}\pi r^3.$$

$$\text{Volume of a sector } V_s = \frac{2}{3}\pi r^2 h$$

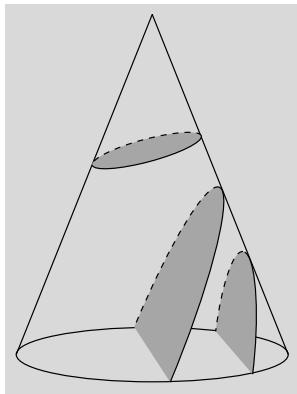
$$= \frac{2}{3}\pi r^3 (1 - \cos \alpha) = V_{\text{sphere}} \operatorname{hav} \alpha.$$

$$\begin{aligned} \text{Area of a segment } A_s &= 2\pi r h = 2\pi r^2 (1 - \cos \alpha) \\ &= A_{\text{sphere}} \operatorname{hav} \alpha. \end{aligned}$$



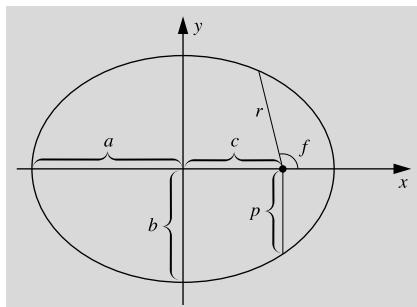
A.2 Conic Sections

As the name already says, conic sections are curves obtained by intersecting circular cones with planes.



Ellipse Equation in rectangular coordinates

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$



a = the semimajor axis,

b = the semiminor axis $b = a\sqrt{1 - e^2}$,

e = eccentricity $0 \leq e < 1$.

Distance of the foci from the centre $c = ea$.

Parameter (semilatus rectum) $p = a(1 - e^2)$.

Area $A = \pi ab$.

Equation in polar coordinates

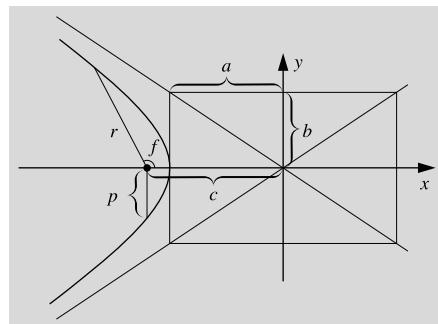
$$r = \frac{p}{1 + e \cos f},$$

where the distance r is measured from one focus, not from the centre.

When $e = 0$, the curve becomes a circle.

Hyperbola Equations in rectangular and polar coordinates

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \quad r = \frac{p}{1 + e \cos f}.$$



Eccentricity $e > 1$.

Semi-minor axis $b = a\sqrt{e^2 - 1}$.

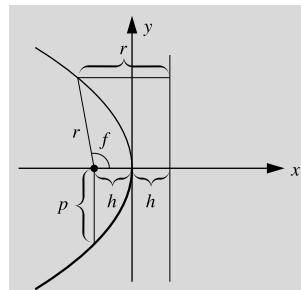
Parameter $p = a(e^2 - 1)$.

Asymptotes $y = \pm \frac{b}{a}x$.

Parabola Parabola is a limiting case between the previous ones; its eccentricity is $e = 1$.

Equations

$$x = -ay^2, \quad r = \frac{p}{1 + \cos f}.$$



Distance of the focus from the apex $h = 1/4a$.

Parameter $p = 1/2a$.

A.3 Taylor Series

Let us consider a differentiable real-valued function of one variable $f : \mathbf{R} \rightarrow \mathbf{R}$. The tangent to the graph of the function at x_0 is

$$y = f(x_0) + f'(x_0)(x - x_0),$$

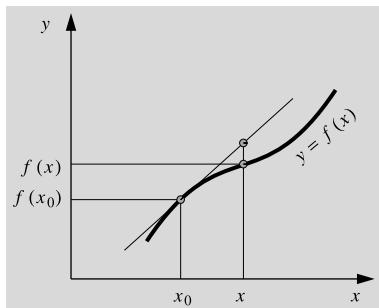
where $f'(x_0)$ is the derivative of f at x_0 . Now, if x is close to x_0 , the graph of the tangent at x will not be very far from the graph of the function itself. Thus, we can approximate the function by

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0).$$

The approximation becomes worse, the more the derivative f' varies in the interval $[x_0, x]$. The rate of change of f' is described by the second derivative f'' , and so on. To improve accuracy, we have to also include higher derivatives. It can be shown that the value of the function f at x is (assuming that the derivatives exist)

$$\begin{aligned} f(x) &= f(x_0) + f'(x_0)(x - x_0) \\ &\quad + \frac{1}{2} f''(x_0)(x - x_0)^2 + \dots \\ &\quad + \frac{1}{n!} f^{(n)}(x_0)(x - x_0)^n + \dots, \end{aligned}$$

where $f^{(n)}(x_0)$ is the n th derivative at x_0 and $n!$ is the n -factorial, $n! = 1 \cdot 2 \cdot 3 \cdots \cdots n$. This expansion is called the *Taylor series* of the function at x_0 .



The following list gives some useful Taylor series (in all these cases we have $x_0 = 0$):

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + \dots$$

converges if $|x| < 1$

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \dots$$

$$\sqrt{1+x} = 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \frac{1}{16}x^3 - \dots$$

$$\sqrt{1-x} = 1 - \frac{1}{2}x - \frac{1}{8}x^2 - \frac{1}{16}x^3 - \dots$$

$$\frac{1}{\sqrt{1+x}} = 1 - \frac{1}{2}x + \frac{3}{8}x^2 - \frac{5}{16}x^3 - \dots$$

$$\frac{1}{\sqrt{1-x}} = 1 + \frac{1}{2}x + \frac{3}{8}x^2 + \frac{5}{16}x^3 + \dots$$

$$e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \dots + \frac{1}{n!}x^n$$

+ ... converges for all x

$$\ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + \dots$$

$x \in (-1, 1]$

$$\sin x = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \dots \quad \text{for all } x$$

$$\cos x = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \dots \quad \text{for all } x$$

$$\tan x = x + \frac{1}{3}x^3 + \frac{2}{15}x^5 + \dots \quad |x| < \frac{\pi}{2}.$$

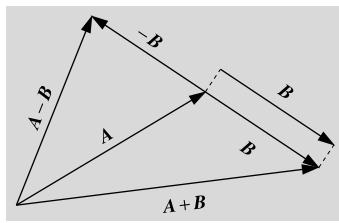
Many problems involve small perturbations, in which case it is usually possible to find expressions having very rapidly converging Taylor expansions. The great advantage of this is the reduction of complicated functions to simple polynomials. Particularly useful are linear approximations, such as

$$\sqrt{1+x} \approx 1 + \frac{1}{2}x, \quad \frac{1}{\sqrt{1+x}} \approx 1 - \frac{1}{2}x, \quad \text{etc.}$$

A.4 Vector Calculus

A *vector* is an entity with two essential properties: *magnitude* and *direction*. Vectors are usually denoted by boldface letters \mathbf{a} , \mathbf{b} , \mathbf{A} , \mathbf{B} etc. The *sum* of the vectors \mathbf{A} and \mathbf{B} can be determined graphically by moving the origin of \mathbf{B} to the tip of \mathbf{A} and connecting the origin of \mathbf{A} to the tip of \mathbf{B} . The vector $-\mathbf{A}$ has the same magnitude as \mathbf{A} , is

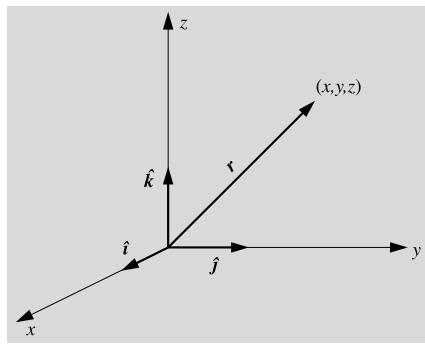
parallel to \mathbf{A} , but points in the opposite direction. The *difference* $\mathbf{A} - \mathbf{B}$ is defined as $\mathbf{A} + (-\mathbf{B})$.



Addition of vectors satisfies the ordinary rules of commutativity and associativity,

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A},$$

$$\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}.$$



A point in a coordinate frame can be specified by giving its *position* or *radius vector*, which extends from the origin of the frame to the point. The position vector \mathbf{r} can be expressed in terms of *basis vectors*, which are usually *unit vectors*, i.e. have a length of one distance unit. In a rectangular xyz -frame, we denote the basis vectors parallel to the coordinate axes by $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$ and $\hat{\mathbf{k}}$. The position vector corresponding to the point (x, y, z) is then

$$\mathbf{r} = x\hat{\mathbf{i}} + y\hat{\mathbf{j}} + z\hat{\mathbf{k}}.$$

The numbers x , y and z are the *components* of \mathbf{r} . Vectors can be added by adding their components. For example, the sum of

$$\mathbf{A} = a_x\hat{\mathbf{i}} + a_y\hat{\mathbf{j}} + a_z\hat{\mathbf{k}},$$

$$\mathbf{B} = b_x\hat{\mathbf{i}} + b_y\hat{\mathbf{j}} + b_z\hat{\mathbf{k}},$$

is

$$\mathbf{A} + \mathbf{B} = (a_x + b_x)\hat{\mathbf{i}} + (a_y + b_y)\hat{\mathbf{j}} + (a_z + b_z)\hat{\mathbf{k}}.$$

The magnitude of a vector \mathbf{r} in terms of its components is

$$\mathbf{r} = |\mathbf{r}| = \sqrt{x^2 + y^2 + z^2}.$$

The *scalar product* of two vectors \mathbf{A} and \mathbf{B} is a real number (scalar)

$$\mathbf{A} \cdot \mathbf{B} = a_x b_x + a_y b_y + a_z b_z = |\mathbf{A}| |\mathbf{B}| \cos(\mathbf{A}, \mathbf{B}),$$

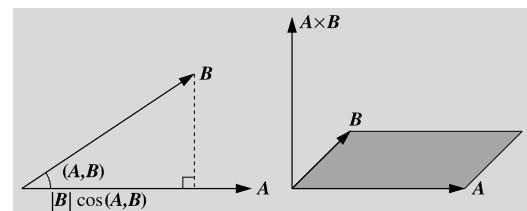
where (\mathbf{A}, \mathbf{B}) is the angle between the vectors \mathbf{A} and \mathbf{B} . We can also think of the scalar product as the projection of, say, \mathbf{A} in the direction of \mathbf{B} multiplied by the length of \mathbf{B} . If \mathbf{A} and \mathbf{B} are perpendicular, their scalar product vanishes. The magnitude of a vector expressed as a scalar product is $A = |\mathbf{A}| = \sqrt{\mathbf{A} \cdot \mathbf{A}}$.

The *vector product* of the vectors \mathbf{A} and \mathbf{B} is a vector

$$\begin{aligned} \mathbf{A} \times \mathbf{B} &= (a_y b_z - a_z b_y)\hat{\mathbf{i}} + (a_z b_x - a_x b_z)\hat{\mathbf{j}} \\ &\quad + (a_x b_y - a_y b_x)\hat{\mathbf{k}} \\ &= \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ a_x & a_y & a_z \\ b_x & b_y & b_z \end{vmatrix}. \end{aligned}$$

This is perpendicular to both \mathbf{A} and \mathbf{B} . Its length gives the area of the parallelogram spanned by \mathbf{A} and \mathbf{B} . The vector product of parallel vectors is a null vector. The vector product is anti-commutative:

$$\mathbf{A} \times \mathbf{B} = -\mathbf{B} \times \mathbf{A}.$$



Scalar and vector products satisfy the laws of distributivity:

$$\mathbf{A} \cdot (\mathbf{B} + \mathbf{C}) = \mathbf{A} \cdot \mathbf{B} + \mathbf{A} \cdot \mathbf{C},$$

$$\begin{aligned} \mathbf{A} \times (\mathbf{B} + \mathbf{C}) &= \mathbf{A} \times \mathbf{B} + \mathbf{A} \times \mathbf{C}, \\ (\mathbf{A} + \mathbf{B}) \cdot \mathbf{C} &= \mathbf{A} \cdot \mathbf{C} + \mathbf{B} \cdot \mathbf{C}, \\ (\mathbf{A} + \mathbf{B}) \times \mathbf{C} &= \mathbf{A} \times \mathbf{C} + \mathbf{B} \times \mathbf{C}. \end{aligned}$$

A scalar triple product is a scalar

$$\mathbf{A} \times \mathbf{B} \cdot \mathbf{C} = \begin{vmatrix} a_x & a_y & a_z \\ b_x & b_y & b_z \\ c_x & c_y & c_z \end{vmatrix}.$$

Here the cross and dot can be interchanged and the factors permuted cyclically without affecting the value of the product. For example $\mathbf{A} \times \mathbf{B} \cdot \mathbf{C} = \mathbf{B} \times \mathbf{C} \cdot \mathbf{A} = \mathbf{B} \cdot \mathbf{C} \times \mathbf{A}$, but $\mathbf{A} \times \mathbf{B} \cdot \mathbf{C} = -\mathbf{B} \times \mathbf{A} \cdot \mathbf{C}$.

A vector triple product is a vector, which can be evaluated using one of the expansions

$$\begin{aligned} \mathbf{A} \times (\mathbf{B} \times \mathbf{C}) &= \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{C}(\mathbf{A} \cdot \mathbf{B}), \\ (\mathbf{A} \times \mathbf{B}) \times \mathbf{C} &= \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{A}(\mathbf{B} \cdot \mathbf{C}). \end{aligned}$$

In all these products, scalar factors can be moved around without affecting the product:

$$\begin{aligned} \mathbf{A} \cdot k\mathbf{B} &= k(\mathbf{A} \cdot \mathbf{B}), \\ \mathbf{A} \times (\mathbf{B} \times k\mathbf{C}) &= k(\mathbf{A} \times (\mathbf{B} \times \mathbf{C})). \end{aligned}$$

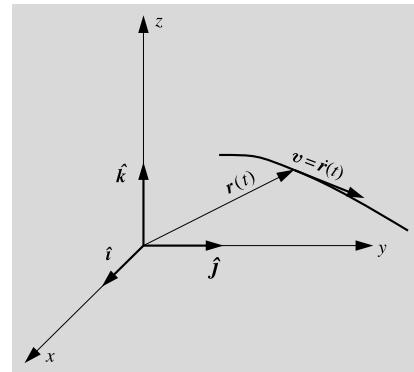
The position vector of a particle is usually a function of time $\mathbf{r} = \mathbf{r}(t) = x(t)\hat{\mathbf{i}} + y(t)\hat{\mathbf{j}} + z(t)\hat{\mathbf{k}}$. The velocity of the particle is a vector, tangent to the trajectory, obtained by taking the derivative of \mathbf{r} with respect to time:

$$\mathbf{v} = \frac{d}{dt}\mathbf{r}(t) = \dot{\mathbf{r}} = \dot{x}\hat{\mathbf{i}} + \dot{y}\hat{\mathbf{j}} + \dot{z}\hat{\mathbf{k}}.$$

The acceleration is the second derivative, $\ddot{\mathbf{r}}$.

Derivatives of the various products obey the same rules as derivatives of products of real-valued functions:

$$\begin{aligned} \frac{d}{dt}(\mathbf{A} \cdot \mathbf{B}) &= \dot{\mathbf{A}} \cdot \mathbf{B} + \mathbf{A} \cdot \dot{\mathbf{B}}, \\ \frac{d}{dt}(\mathbf{A} \times \mathbf{B}) &= \dot{\mathbf{A}} \times \mathbf{B} + \mathbf{A} \times \dot{\mathbf{B}}. \end{aligned}$$



When computing a derivative of a vector product, one must be careful to retain the order of the factors, since the sign of the vector product changes if the factors are interchanged.

A.5 Matrices

Assume we have a vector \mathbf{x} with components (x, y, z) . We can calculate another vector $\mathbf{x}' = (x', y', z')$, the components of which are linear combinations of the original components:

$$\begin{aligned} x' &= a_{11}x + a_{12}y + a_{13}z, \\ y' &= a_{21}x + a_{22}y + a_{23}z, \\ z' &= a_{31}x + a_{32}y + a_{33}z. \end{aligned}$$

This is a linear transform that maps the vector \mathbf{x} to a vector \mathbf{x}' .

We can collect the coefficients to an array, called a *matrix A*:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

A general matrix can consist of an arbitrary number of rows and columns. In this book we need only matrices operating on vectors of a three-dimensional space, and they always have three rows and columns. Two subscripts refer to the different elements of the matrix, the first one giving the row and the second one the column.

When using matrix formalism it is convenient to write vectors in the form of column vectors:

$$\mathbf{A} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

We now define that the product of a matrix and a column vector

$$\mathbf{x}' = \mathbf{Ax}$$

or

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

means just

$$x' = a_{11}x + a_{12}y + a_{13}z,$$

$$y' = a_{21}x + a_{22}y + a_{23}z,$$

$$z' = a_{31}x + a_{32}y + a_{33}z.$$

Comparing these equations we see, for example, that the first component of \mathbf{x}' is obtained by taking the first row of the matrix, multiplying the components of the vector \mathbf{x} by the corresponding components of that row, and finally adding the products.

This definition can easily be generalised to the product of two matrices. The elements of the matrix

$$\mathbf{C} = \mathbf{AB}$$

are

$$c_{ij} = \sum_k a_{ik}b_{kj}.$$

This is easy to remember by noting that we take the row i of the first factor \mathbf{A} and the column j of the second factor \mathbf{B} and evaluate the scalar product of the two vectors. For example

$$\begin{aligned} & \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 1 \\ 1 & 3 & 2 \end{pmatrix} \\ &= \begin{pmatrix} 1+2+1 & 2+1+3 & 0+1+2 \\ 0+2+2 & 0+1+6 & 0+1+4 \\ 1+4+3 & 2+2+9 & 0+2+6 \end{pmatrix} \\ &= \begin{pmatrix} 4 & 6 & 3 \\ 4 & 7 & 5 \\ 8 & 13 & 8 \end{pmatrix}. \end{aligned}$$

When multiplying matrices, we have to be careful with the order of the factors, because usually $\mathbf{AB} \neq \mathbf{BA}$. If we multiply the matrices of

the previous example in the reverse order, we get quite a different result:

$$\begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 1 \\ 1 & 3 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 1 & 2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 3 & 5 \\ 3 & 5 & 7 \\ 3 & 8 & 13 \end{pmatrix}.$$

A *unit matrix* \mathbf{I} is a matrix, which has ones on its diagonal and zeros elsewhere:

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

If a vector or a matrix is multiplied by a unit matrix, it will remain unchanged.

If the product of two matrices is a unit matrix, the two matrices are *inverse matrices* of each others. The inverse matrix of \mathbf{A} is denoted by \mathbf{A}^{-1} . It satisfies the equations

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{AA}^{-1} = \mathbf{I}.$$

In spherical astronomy we need mainly *rotation matrices*, describing the rotation of a coordinate frame. The following matrices correspond to rotations around the x , y and z axes, respectively:

$$\mathbf{R}_x(\alpha) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & \sin \alpha \\ 0 & -\sin \alpha & \cos \alpha \end{pmatrix},$$

$$\mathbf{R}_y(\alpha) = \begin{pmatrix} \cos \alpha & 0 & \sin \alpha \\ 0 & 1 & 0 \\ -\sin \alpha & 0 & \cos \alpha \end{pmatrix},$$

$$\mathbf{R}_z(\alpha) = \begin{pmatrix} \cos \alpha & \sin \alpha & 0 \\ -\sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

If the angle is $\alpha = 0$, only a unit matrix remains.

The elements of a rotation matrix can easily be determined. For example, a rotation around the x axis will leave the x coordinate unaffected, and thus the first row and column must be zeroes, except for the diagonal element, which must be one. This will leave four elements. When the angle is zero, the matrix has to reduce to a unit matrix; thus the diagonal elements must be cosines and the other ones sines. The only problem is to decide, which of the sines will get the minus sign.

This is most easily done by testing the effect of the matrix on some basis vector.

The inverse matrix of a rotation matrix corresponds to a rotation in the opposite direction. Thus it is obtained from the original matrix by replacing the angle α by $-\alpha$. The only change in the matrix is that the signs of the sines are changed.

For example, the *precession matrix* is a product of three rotation matrices. Since the matrix product is not commutative, these rotations must be carried out in the correct order.

A.6 Multiple Integrals

An integral of a function f over a surface A

$$I = \int_A f \, dA$$

can be evaluated as a double integral by expressing the surface element dA in terms of coordinate differentials. In rectangular coordinates,

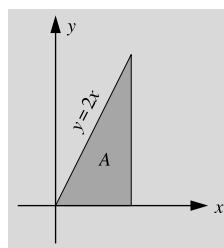
$$dA = dx \, dy$$

and in polar coordinates

$$dA = r \, dr \, d\varphi.$$

The integration limits of the innermost integral may depend on the other integration variable. For example, the function xe^y integrated over the shaded area is

$$\begin{aligned} I &= \int_A xe^y \, dA = \int_{x=0}^1 \int_{y=0}^{2x} xe^y \, dx \, dy \\ &= \int_0^1 [0^{2x} xe^y] \, dx = \int_0^1 (xe^{2x} - x) \, dx \\ &= \left[\frac{1}{2} xe^{2x} - \frac{1}{4} e^{2x} - \frac{1}{2} x^2 \right]_0^1 = \frac{1}{4}(e^2 - 1). \end{aligned}$$



The surface need not be confined to a plane. For example, the area of a sphere is

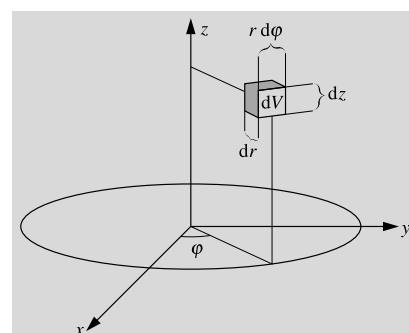
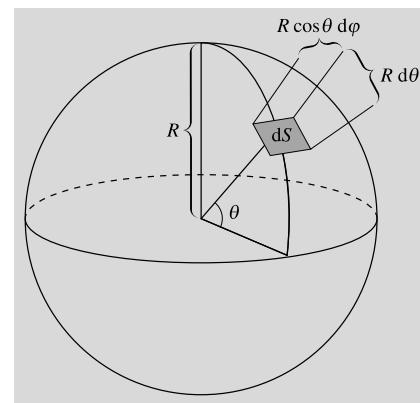
$$A = \int_S dS,$$

where the integration is extended over the surface S of the sphere. In this case the surface element is

$$dS = R^2 \cos \theta \, d\varphi \, d\theta,$$

and the area is

$$\begin{aligned} A &= \int_{\varphi=0}^{2\pi} \int_{\theta=-\pi/2}^{\pi/2} R^2 \cos \theta \, d\varphi \, d\theta \\ &= \int_0^{2\pi} \left[\int_{-\pi/2}^{\pi/2} R^2 \sin \theta \right] d\varphi \\ &= \int_0^{2\pi} 2R^2 \, d\varphi = 4\pi R^2. \end{aligned}$$



Similarly, a volume integral

$$I = \int_V f \, dV$$

can be evaluated as a triple integral. In rectangular coordinates, the volume element dV is

$$dV = dx dy dz;$$

in cylindrical coordinates

$$dV = r dr d\varphi dz,$$

and in spherical coordinates

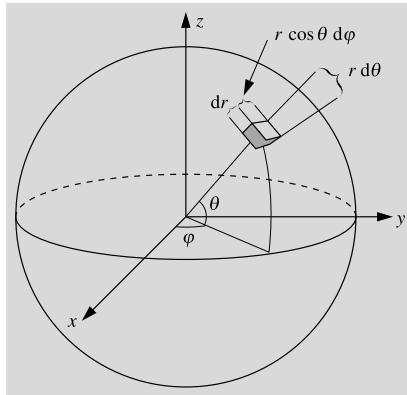
$$dV = r^2 \cos \theta dr d\varphi d\theta$$

(θ measured from the xy plane)

or

$$dV = r^2 \sin \theta dr d\varphi d\theta$$

(θ measured from the z axis).



For example, the volume of a sphere with radius R is

$$\begin{aligned} V &= \int_V dV \\ &= \int_{r=0}^R \int_{\varphi=0}^{2\pi} \int_{\theta=-\pi/2}^{\pi/2} r^2 \cos \theta dr d\varphi d\theta \\ &= \int_0^R \int_0^{2\pi} \left[\int_{-\pi/2}^{\pi/2} r^2 \sin \theta \right] dr d\varphi \\ &= \int_0^R \int_0^{2\pi} 2r^2 dr d\varphi \\ &= \int_0^R 4\pi r^2 dr = \left| \frac{4\pi r^3}{3} \right|_0^R = \frac{4}{3}\pi R^3. \end{aligned}$$

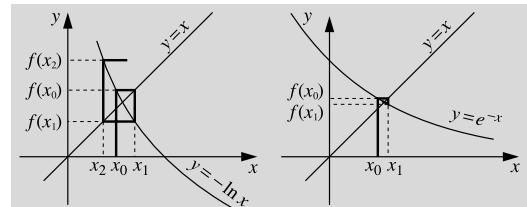
A.7 Numerical Solution of an Equation

We frequently meet equations defying analytical solutions. Kepler's equation is a typical example. If we cannot do anything else, we can always apply some numerical method. Next we shall present two very simple methods, the first of which is particularly suitable for calculators.

Method 1: Direct Iteration We shall write the equation as $f(x) = x$. Next we have to find an initial value x_0 for the solution. This can be done, for example, graphically or by just guessing. Then we compute a succession of new iterates $x_1 = f(x_0)$, $x_2 = f(x_1)$, and so on, until the difference of successive solutions becomes smaller than some preset limit. The last iterate x_i is the solution. After computing a few x_i 's, it is easy to see if they are going to converge. If not, we rewrite the equation as $f^{-1}(x) = x$ and try again. (f^{-1} is the inverse function of f .)

As an example, let us solve the equation $x = -\ln x$. We guess $x_0 = 0.5$ and find

$$x_1 = -\ln 0.5 = 0.69, \quad x_2 = 0.37, \quad x_3 = 1.00.$$



This already shows that something is wrong. Therefore we change our equation to $x = e^{-x}$ and start again:

$$\begin{aligned} x_0 &= 0.5, \\ x_1 &= e^{-0.5} = 0.61, \\ x_2 &= 0.55, \\ x_3 &= 0.58, \\ x_4 &= 0.56, \\ x_5 &= 0.57, \\ x_6 &= 0.57. \end{aligned}$$

Thus the solution, accurate to two decimal places, is 0.57.

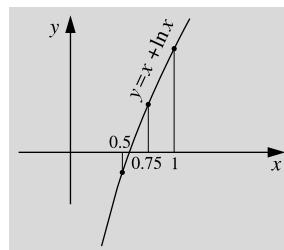
Method 2. Interval Halving In some pathological cases the previous method may refuse to converge. In such situations we can use the foolproof method of interval halving. If the function is continuous (as most functions of classical physics are) and we manage to find two points x_1 and x_2 such that $f(x_1) > 0$ and $f(x_2) < 0$, we know that somewhere between x_1 and x_2 there must be a point x in which $f(x) = 0$. Now we find the sign of f in the midpoint of the interval, and select the half of the interval in which f changes sign. We repeat this procedure until the interval containing the solution is small enough.

We shall try also this method on our equation $x = -\ln x$, which is now written as $f(x) = 0$, where $f(x) = x + \ln x$. Because $f(x) \rightarrow -\infty$, when $x \rightarrow 0$ and $f(1) > 0$, the solution must be in the range $(0, 1)$. Since $f(0.5) < 0$, we know

that $x \in (0.5, 1)$. We continue in this way:

$$\begin{aligned} f(0.75) > 0 &\Rightarrow x \in (0.5, 0.75), \\ f(0.625) > 0 &\Rightarrow x \in (0.5, 0.625), \\ f(0.563) < 0 &\Rightarrow x \in (0.563, 0.625), \\ f(0.594) > 0 &\Rightarrow x \in (0.563, 0.594). \end{aligned}$$

The convergence is slow but certain. Each iteration restricts the solution to an interval which is half as large as the previous one, thus improving the solution by one binary digit.



Albert Einstein published his *special theory of relativity* in 1905 and the *general theory of relativity* ten years later. Especially the general theory, which is essentially a gravitation theory, has turned out to be very important for the theories of the evolution of the Universe. Therefore it is appropriate to consider here some basic principles of relativity theory. A more detailed discussion would require sophisticated mathematics, and is beyond the scope of this very elementary book.

B.1 Basic Concepts

Everyone knows the famous Pythagorean theorem

$$\Delta s^2 = \Delta x^2 + \Delta y^2,$$

where Δs is the length of the hypotenuse of a right-angle triangle, and Δx and Δy are the lengths of the other two sides. (For brevity, we have denoted $\Delta s^2 = (\Delta s)^2$.) This is easily generalised to three dimensions:

$$\Delta s^2 = \Delta x^2 + \Delta y^2 + \Delta z^2.$$

This equation describes the *metric* of an ordinary rectangular frame in a Euclidean space, i.e. tells us how to measure distances.

Generally the expression for the distance between two points depends on the exact location of the points. In such a case the metric must be expressed in terms of infinitesimal distances in order to correctly take into account the curvature of the coordinate curves. (A coordinate curve

is a curve along which one coordinate changes while all the others remain constant.) An infinitesimal distance ds is called the *line element*. In a rectangular frame of a Euclidean space it is

$$ds^2 = dx^2 + dy^2 + dz^2,$$

and in spherical coordinates

$$ds^2 = dr^2 + r^2(d\theta^2 + \cos^2\theta d\phi^2).$$

Generally ds^2 can be expressed as

$$ds^2 = \sum_{i,j} g_{ij} dx_i dx_j,$$

where the x_i 's are arbitrary coordinates, and the coefficients g_{ij} are components of the *metric tensor*. These can be functions of the coordinates, as in the case of the spherical coordinates. The metric tensor of an n -dimensional space can be expressed as an $n \times n$ matrix. Since $dx_i dx_j = dx_j dx_i$, the metric tensor is symmetric, i.e. $g_{ij} = g_{ji}$. If all coordinate curves intersect perpendicularly, the coordinate frame is *orthogonal*. In an orthogonal frame, $g_{ij} = 0$ for all $i \neq j$. For example, the spherical coordinates form an orthogonal frame, the metric tensor of which is

$$(g_{ij}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2 \cos^2\theta \end{pmatrix}.$$

If it is possible to find a frame in which all the components of g are constant, the space is *flat*. In the rectangular frame of a Euclidean space,

we have $g_{11} = g_{22} = g_{33} = 1$; hence the space is flat. The spherical coordinates show that even in a flat space, we can use frames in which the components of the metric tensor are not constant. The line element of a two-dimensional spherical surface is obtained from the metric of the spherical coordinate frame by assigning to r some fixed value R :

$$ds^2 = R^2(d\theta^2 + \cos^2\theta d\phi^2).$$

The metric tensor is

$$(g_{ij}) = \begin{pmatrix} 1 & 0 \\ 0 & \cos^2\phi \end{pmatrix}.$$

This cannot be transformed to a constant tensor. Thus the surface of a sphere is a *curved space*.

If we know the metric tensor in some frame, we can compute a fourth-order tensor, the *curvature tensor* R_{ijkl} , which tells us whether the space is curved or flat. Unfortunately the calculations are slightly too laborious to be presented here.

The metric tensor is needed for all computations involving distances, magnitudes of vectors, areas, and so on. Also, to evaluate a scalar product, we must know the metric. In fact the components of the metric tensor can be expressed as scalar products of the basis vectors:

$$g_{ij} = \hat{\mathbf{e}}_i \cdot \hat{\mathbf{e}}_j.$$

If \mathbf{A} and \mathbf{B} are two arbitrary vectors

$$\mathbf{A} = \sum_i a^i \hat{\mathbf{e}}_i, \quad \mathbf{B} = \sum_i b^i \hat{\mathbf{e}}_i,$$

their scalar product is

$$\mathbf{A} \cdot \mathbf{B} = \sum_i \sum_j a^i b^j \hat{\mathbf{e}}_i \cdot \hat{\mathbf{e}}_j = \sum_i \sum_j g_{ij} a^i b^j.$$

B.2 Lorentz Transformation. Minkowski Space

The special theory of relativity abandoned the absolute Newtonian time, supposed to flow at the same rate for every observer. Instead it required that the speed of light must have the same value c in all coordinate frames. The constancy of the

speed of light follows immediately from the basic assumptions of the special theory of relativity. This is possible only if different observers measure time intervals differently.

Let us send a beam of light from the origin. It travels along a straight line at the speed c . Thus, at the moment t its space and time coordinates satisfy the equation

$$x^2 + y^2 + z^2 = c^2 t^2. \quad (\text{B.1})$$

Next we study what the situation looks like in another frame, $x'y'z'$, moving at a velocity v with respect to the xyz frame. Let us select the new frame so that it coincides with the xyz frame at $t = 0$. Also, let the time coordinate t' of the $x'y'z'$ frame be $t' = 0$ when $t = 0$. And finally we assume that the $x'y'z'$ frame moves in the direction of the positive x axis. Since the beam of light must also travel at the speed c in the $x'y'z'$ frame, we must have

$$x'^2 + y'^2 + z'^2 = c^2 t'^2.$$

If we require that the new (dashed) coordinates be obtained from the old ones by a linear transformation, and also that the inverse transformation be obtained simply by replacing v by $-v$, we find that the transformation must be

$$\begin{aligned} x' &= \frac{x - vt}{\sqrt{1 - v^2/c^2}}, \\ y' &= y, \\ z' &= z, \\ t' &= \frac{t - vx/c^2}{\sqrt{1 - v^2/c^2}}. \end{aligned} \quad (\text{B.2})$$

This transformation between frames moving at a constant speed with respect to each other is called the *Lorentz transformation*.

Because the Lorentz transformation is derived assuming the invariance of (B.1), it is obvious that the *interval*

$$\Delta s^2 = -c^2 \Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2$$

of any two events remains invariant in all Lorentz transformations. This interval defines a metric in

four-dimensional spacetime. A space having such a metric is called the *Minkowski space* or Lorentz space. The components of the metric tensor g are

$$(g_{ij}) = \begin{pmatrix} -c^2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Since this is constant, the space is flat. But it is no longer an ordinary Euclidean space, because the sign of the time component differs from the sign of the space components. In the older literature, a variable ict is often used instead of time, i being the imaginary unit. Then the metric looks Euclidean, which is misleading; the properties of the space cannot be changed just by changing notation.

In a Minkowski space position, velocity, momentum and other vector quantities are described by *four-vectors*, which have one time and three space components. The components of a four-vector obey the Lorentz transformation when we transform them from one frame to another, both moving along a straight line at a constant speed.

According to classical physics the distance of two events depends on the motion of the observer, but the time interval between the events is the same for all observers. The world of special relativity is more anarchistic: even time intervals have different values for different observers.

B.3 General Relativity

The Equivalence Principle Newton's laws relate the acceleration, a , of a particle and the applied force \mathbf{F} by

$$\mathbf{F} = m_i a,$$

where m_i is the *inertial mass* of the particle, resisting the force trying to move the particle. The gravitational force felt by the particle is

$$\mathbf{F} = m_g f,$$

where m_g is the *gravitational mass* of the particle, and f is a factor depending only on other masses. The masses m_i and m_g appear as coefficients related to totally different phenomena. There is no

physical reason to assume that the two masses should have anything in common. However, already the experiments made by Galilei showed that evidently $m_i = m_g$. This has been verified later with very high accuracy.

The *weak equivalence principle*, which states that $m_i = m_g$, can therefore be accepted as a physical axiom. The *strong equivalence principle* generalises this: if we restrict our observations to a sufficiently small region of spacetime, there is no way to tell whether we are subject to a gravitational field or are in uniformly accelerated motion. The strong equivalence principle is one of the fundamental postulates of general relativity.

Curvature of Space General relativity describes gravitation as a geometric property of the spacetime. The equivalence principle is an obvious consequence of this idea. Particles moving through spacetime follow the shortest possible paths, *geodesics*. The projection of a geodesic onto a three-dimensional space need not be the shortest way between the points.

The geometry of spacetime is determined by the mass and energy distribution. If this distribution is known, we can write the *field equations*, which are partial differential equations connecting the metric to the mass and energy distribution and describing the curvature of spacetime.

In the case of a single-point mass, the field equations yield the Schwarzschild metric, the line element of which is

$$\begin{aligned} ds^2 = -\left(1 - \frac{2GM}{c^2 r}\right)c^2 dt^2 + \frac{dr^2}{1 - 2GM/c^2 r} \\ + r^2(d\theta^2 + \cos^2 \theta d\phi^2). \end{aligned}$$

Here M is the mass of the point; r, θ and ϕ are ordinary spherical coordinates. It can be shown that the components cannot be transformed to constants simultaneously: spacetime must be curved due to the mass.

If we study a very small region of spacetime, the curvature has little effect. Locally the space is always a Minkowski space where special relativity can be used. Locality means not only a limited spatial volume but also a limited interval in time.

Unlike in the Newtonian formalism in general relativity there are no equations of motion describing the motion of a particle. It is possible to use the positions of the bodies to compute the metric tensor of the space and then the geodesics representing the orbits of the bodies. When the bodies are moving, the metric of the space keeps changing, and thus this method is very laborious. The *PPN* (Parametrised Post-Newtonian) formalism has been developed for practical calculations, yielding approximate equations of motion. It contains ten constants, the values of which are specific for Einstein's theory of gravity. Alternative theories give different values for these constants. The formalism therefore provides a framework for experimental tests of general relativity.

The PPN formalism is an approximation that can be used if velocities are well below the speed of light and the gravitational field is weak and hence the curvature of the space is low.

B.4 Tests of General Relativity

General relativity gives predictions different from classical physics. Although the differences are usually very small, there are some phenomena in which the deviation can be measured and used to test the validity of general relativity. At present five different astronomical tests have verified the theory.

First of all the orbit of a planet is no longer a closed Keplerian ellipse. The effect is strongest for the innermost planets, whose perihelia should turn little by little. Most of the motion of the perihelion of Mercury is predicted by Newtonian mechanics; only a small excess of 43 arc seconds per century remains unexplained. And it so happens that this is exactly the correction suggested by general relativity.

Secondly a beam of light should bend when it travels close to the Sun. For a beam grazing

the surface, the deviation should be about $1.75''$. Such effects have been observed during total solar eclipses and also by observing pointlike radio sources, just before and after occultation by the Sun.

The third classical way of testing general relativity is to measure the redshift of a photon climbing in a gravitational field. We can understand the redshift as an energy loss while the photon does work against the gravitational potential. Or we can think of the redshift as caused by the metric only: a distant observer finds that near a mass, time runs slower, and the frequency of the radiation is lower. The time dilation in the Schwarzschild metric is described by the coefficient of dt ; thus it is not surprising that the radiation emitted at a frequency ν from a distance r from a mass M has the frequency

$$\nu_\infty = \nu \sqrt{1 - \frac{2GM}{c^2r}}, \quad (\text{B.3})$$

if observed very far from the source. This gravitational redshift has been verified by laboratory experiments.

The fourth test employs the slowing down of the speed of light in a gravitational field near the Sun. This has been verified by radar experiments.

The previous tests concern the solar system. Outside the solar system binary pulsars have been used to test general relativity. An asymmetric system in accelerated motion (like a binary star) loses energy as it radiates *gravitational waves*. It follows that the components approach each other, and the period decreases. Usually the gravitational waves play a minor role, but in the case of a compact source the effect may be strong enough to be observed. The first source to show this predicted shortening of period was the binary pulsar PSR 1913+16.

Tables

C

Table C.1 SI basic units

Quantity	Symbol	Unit	Abbr.	Definition
Length	l, s, \dots	Metre	m	The length of the path travelled by light in vacuum during a time interval of $1/299,792,458$ of a second.
Mass	m, M	Kilogram	kg	Equal to the mass of the international prototype of the kilogram.
Time	t	Second	s	The duration of $9,192,631,770$ periods of the radiation corresponding to the transition between two hyperfine levels of the ground state of the caesium 133-atom.
Electric current	I	Ampere	A	That constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular cross section and placed 1 metre apart in a vacuum, would produce a force equal to 2×10^{-7} Newton per metre of length between these conductors.
Temperature	T	Kelvin	K	The fraction $1/273.16$ of the thermodynamic temperature of the triple point of water.
Amount of substance	n	Mole	mol	The amount of substance of a system which contains as many elementary entities as there are atoms in 0.012 kg of ^{12}C .
Luminous intensity	I	Candela	cd	The luminous intensity in a given direction of a source which emits monochromatic radiation of frequency 540×10^{12} Hz and of which the radiant intensity in that direction is $1/683$ Watt per steradian.

Table C.2 Prefixes for orders of ten

Prefix	Symbol	Multiple	Prefix	Symbol	Multiple
yocto	y	10^{-24}	deca	da	10^1
zepto	z	10^{-21}	hecto	h	10^2
atto	a	10^{-18}	kilo	k	10^3
femto	f	10^{-15}	Mega	M	10^6
pico	p	10^{-12}	Giga	G	10^9
nano	n	10^{-9}	Tera	T	10^{12}
micro	μ	10^{-6}	Peta	P	10^{15}
milli	m	10^{-3}	Exa	E	10^{18}
centi	c	10^{-2}	Zetta	Z	10^{21}
deci	d	10^{-1}	Yotta	Y	10^{24}

Table C.3 Constants and units

Radian	1 rad	$= 180^\circ/\pi = 57.2957795^\circ = 206,264.8''$
Degree	1°	$= 0.01745329 \text{ rad}$
Arc second	$1''$	$= 0.000004848 \text{ rad}$
Velocity of light	c	$= 299,792,458 \text{ m s}^{-1}$
Gravitational constant	G	$= 6.673 \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}$ $= 4\pi^2 \text{ AU}^3 M_{\odot}^{-1} \text{ a}^{-2}$ $= 3,986,005 \times 10^8 \text{ m}^3 M_{\oplus}^{-1} \text{ s}^{-2}$
Planck constant	h	$= 6.6261 \times 10^{-34} \text{ Js}$
	\hbar	$= h/2\pi = 1.0546 \times 10^{-34} \text{ Js}$
Boltzmann constant	k	$= 1.3807 \times 10^{-23} \text{ JK}^{-1}$
Radiation density constant	a	$= 7.5659 \times 10^{-16} \text{ J m}^{-3} \text{ K}^{-4}$
Stefan-Boltzmann constant	σ	$= ac/4 = 5.6705 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$
Atomic mass unit	amu	$= 1.6605 \times 10^{-27} \text{ kg}$
Electron volt	eV	$= 1.6022 \times 10^{-19} \text{ J}$
Electron charge	e	$= 1.6022 \times 10^{-19} \text{ C}$
Mass of electron	m_e	$= 9.1094 \times 10^{-31} \text{ kg} = 0.511 \text{ MeV}$
Mass of proton	m_p	$= 1.6726 \times 10^{-27} \text{ kg} = 938.3 \text{ MeV}$
Mass of neutron	m_n	$= 1.6749 \times 10^{-27} \text{ kg} = 939.6 \text{ MeV}$
Mass of ${}^1\text{H}$ atom	m_H	$= 1.6735 \times 10^{-27} \text{ kg} = 1.0078 \text{ amu}$
Mass of ${}^4_2\text{He}$ atom	m_{He}	$= 6.6465 \times 10^{-27} \text{ kg} = 4.0026 \text{ amu}$
Rydberg constant for ${}^1\text{H}$	R_H	$= 1.0968 \times 10^7 \text{ m}^{-1}$
Rydberg constant for ∞ mass	R_∞	$= 1.0974 \times 10^7 \text{ m}^{-1}$
Gas constant	R	$= 8.3145 \text{ JK}^{-1} \text{ mol}^{-1}$
Normal atmospheric pressure	atm	$= 101,325 \text{ Pa} = 1013 \text{ mbar} = 760 \text{ mmHg}$
Astronomical unit	AU	$= 1.49597870 \times 10^{11} \text{ m}$
Parsec	pc	$= 3.0857 \times 10^{16} \text{ m} = 206,265 \text{ AU} = 3.26 \text{ ly}$
Light-year	ly	$= 0.9461 \times 10^{16} \text{ m} = 0.3066 \text{ pc}$

Table C.4 Units of time

Unit	Equivalent to
Sidereal year	365.2564 d (with respect to fixed stars)
Tropical year	365.2422 d (equinox to equinox)
Anomalistic year	365.2596 d (perihelion to perihelion)
Gregorian calendar year	365.2425 d
Julian year	365.25 d
Julian century	36,525 d
Eclipse year	346.6200 d (with respect to the ascending node of the Moon)
Lunar year	354.367 d = 12 synodical months
Synodical month	29.5306 d (newmoon to newmoon)
Sidereal month	27.3217 d (with respect to fixed stars)
Tropical month	27.3216 d (with respect to the vernal equinox)
Anomalistic month	27.5546 d (perigee to perigee)
Draconic month	27.2122 d (node to node)
Mean solar day	24 h mean solar time = 24 h 03 min 56.56 s sidereal time = 1.00273791 sidereal days
Sidereal day	24 h sidereal time = 23 h 56 min 04.09 s mean solar time = 0.99726957 mean solar days
Rotation period of the Earth (referred to fixed stars)	1.000000097 sidereal days = 23 h 56 min 04.10 s mean solar time

Table C.5 The Greek alphabet

A, α	B, β	Γ, γ	Δ, δ	E, ϵ, ε	Z, ζ	H, η	$\Theta, \theta, \vartheta$
alpha	beta	gamma	delta	epsilon	zeta	eta	theta
I, ι	K, κ	Λ, λ	M, μ	N, ν	Ξ, ξ	O, o	Π, π, ϖ
iota	kappa	lambda	mu	nu	xi	omicron	pi
P, ρ	$\Sigma, \sigma, \varsigma$	T, τ	Υ, υ	Φ, ϕ, φ	X, χ	Ψ, ψ	Ω, ω
rho	sigma	tau	upsilon	phi	chi	psi	omega

Table C.6 The Sun

Property	Symbol	Numerical value
Mass	M_{\odot}	$1.989 \times 10^{30} \text{ kg}$
Radius	R_{\odot}	$6.960 \times 10^8 \text{ m} = 0.00465 \text{ AU}$
Effective temperature	T_e	5785 K
Luminosity	L_{\odot}	$3.9 \times 10^{26} \text{ W}$
Apparent visual magnitude	V	-26.78
Colour indices	$B - V$	0.62
	$U - B$	0.10
Absolute visual magnitude	M_V	4.79
Absolute bolometric magnitude	M_{bol}	4.72
Inclination of equator to ecliptic		7°15'
Equatorial horizontal parallax	π_{\odot}	8.794"
Motion: direction of apex		$\alpha = 270^\circ$
		$\delta = 30^\circ$
velocity in LSR		19.7 km s ⁻¹
Distance from galactic centre		8.5 kpc

Table C.7 The Earth

Property	Symbol	Numerical value
Mass	M_{\oplus}	$= M_{\odot}/332,946 = 5.974 \times 10^{24} \text{ kg}$
Mass, Earth + Moon	$M_{\oplus} + M_{\mathbb{M}}$	$= M_{\odot}/328,900.5 = 6.048 \times 10^{24} \text{ kg}$
Equatorial radius	R_e	= 6,378,137 m
Polar radius	R_p	= 6,356,752 m
Flattening	f	$= (R_e - R_p)/R_e = 1/298.257$
Surface gravity	g	= 9.81 m s ⁻²

Table C.8 The Moon

Property	Symbol	Numerical value
Mass	$M_{\mathbb{M}}$	$= M_{\oplus}/81.30 = 7.348 \times 10^{22} \text{ kg}$
Radius	$R_{\mathbb{M}}$	= 1738 km
Surface gravity	$g_{\mathbb{M}}$	= 1.62 m s ⁻² = 0.17g
Mean equatorial horizontal parallax	$\pi_{\mathbb{M}}$	= 57'
Semimajor axis of the orbit	a	= 384,400 km
Smallest distance from Earth	r_{\min}	= 356,400 km
Greatest distance from Earth	r_{\max}	= 406,700 km
Mean inclination of orbit to ecliptic	ι	= 5.145°

Table C.9 Planets. R_e = equatorial radius, ρ = mean density, τ_{sid} = sidereal rotation period (R indicates retrograde rotation), ε = inclination of the equator with respect to the ecliptic (at the beginning of 2000),

Name	R_e [km]	Mass			ρ [g/cm ³]	Number of known satellites		
		Planet [kg]	Planet + Satellites					
			[M_\oplus]	[M_\odot]				
Mercury	2440	3.30×10^{23}	0.0553	1/6,023,600	5.4	–		
Venus	6052	4.87×10^{24}	0.8150	1/408,523.5	5.2	–		
Earth	6378	5.97×10^{24}	1.0123	1/328,900.5	5.5	1		
Mars	3397	6.42×10^{23}	0.1074	1/3,098,710	3.9	2		
Jupiter	71,492	1.90×10^{27}	317.89	1/1047.355	1.3	67		
Saturn	60,268	5.69×10^{26}	95.17	1/3498.5	0.7	62		
Uranus	25,559	8.66×10^{25}	14.56	1/22,869	1.3	27		
Neptune	24,764	1.03×10^{26}	17.24	1/19,314	1.8	14		

Name	τ_{sid}	ε [°]	f	T [K]	p	V_0
Mercury	58.646 d	0.0	0	130–615	0.106	–
Venus	243.019 d R	177.4	0	750	0.65	–
Earth	23 h 56 min 04.1 s	23.4	0.003353	300	0.367	–
Mars	24 h 37 min 22.6 s	25.2	0.006476	220	0.150	–2.01
Jupiter	9 h 55 min 30 s	3.1	0.06487	140	0.52	–2.70
Saturn	10 h 39 min 22 s	26.7	0.09796	100	0.47	+0.67
Uranus	17 h 14 min 24 s R	97.8	0.02293	65	0.51	+5.52
Neptune	16 h 06 min 36 s	28.3	0.01708	55	0.41	+7.84

Table C.10 Magnitudes of the planets. The table gives the expressions used in the *Astronomical Almanac*. They give the magnitudes as functions of the phase angle α . Interior planets can be observed at relatively large phase angles, and thus several terms are needed to describe their phase curves. Phase angles of the exterior planets are always small, and thus very simple expressions are sufficient. The magnitude of Saturn means the magnitude of the planet only; the total magnitude also depends on the orientation of the rings

	$V(1, 0)$	$V(1, \alpha)$
Mercury	−0.36	$V(1, 0) + 3.80(\alpha/100^\circ)$ − 2.73($\alpha/100^\circ$) ² + 2.00($\alpha/100^\circ$) ³
Venus	−4.29	$V(1, 0) + 0.09(\alpha/100^\circ)$ + 2.39($\alpha/100^\circ$) ² − 0.65($\alpha/100^\circ$) ³
Mars	−1.52	$V(1, 0) + 1.60(\alpha/100^\circ)$
Jupiter	−9.25	$V(1, 0) + 0.5(\alpha/100^\circ)$
Saturn	−8.88	
Uranus	−7.19	$V(1, 0) + 0.28(\alpha/100^\circ)$
Neptune	−6.87	$V(1, 0)$

Table C.11 Osculating elements of planetary orbits on JD 2,451,600.5 (Feb 24, 2000). a = semimajor axis, e = eccentricity, i = inclination, Ω = longitude of ascen-

ding node, ϖ = longitude of perihelion, L = mean longitude, P_{sid} = mean sidereal orbital period (here 1 a means a Julian year, or 365.25 days), P_{syn} = mean synodic period

	a	e	i	Ω	ϖ	L	P_{sid}	P_{syn}
	[AU]	[10^6 km]	[$^\circ$]	[$^\circ$]	[$^\circ$]	[$^\circ$]	[a]	[d]
Mercury	0.387	57.9	0.2056	7.01	48.3	77.5	119.4	0.2408
Venus	0.723	108.2	0.0068	3.39	76.7	131.9	270.9	0.6152
Earth	1.000	149.6	0.0167	0.00	143.9	102.9	155.2	1.0000
Mars	1.524	228.0	0.0934	1.85	49.6	336.1	24.5	1.8807
Jupiter	5.204	778.6	0.0488	1.30	100.5	15.5	39.0	11.8565
Saturn	9.582	1433.4	0.0558	2.49	113.6	89.9	51.9	29.4235
Uranus	19.224	2875.8	0.0447	0.77	74.0	170.3	314.1	83.747
Neptune	30.092	4501.7	0.0112	1.77	131.8	39.5	305.5	163.723
							59,800	367.5

Table C.12 Mean elements of planets with respect to the equator and equinox of J2000.0. The variable t is the time in days since J2000.0 and T is the same time in Julian centuries:
 $t = J - 2,451,545.0$,
 $T = t/36,525$. L is the mean longitude,
 $L = M + \varpi$. The elements do not contain periodic terms, and the accuracy of the positions computed from them is of the order of a few minutes of arc. The values are from the *Explanatory Supplement to the Astronomical Almanac*. The elements of the Earth describe the orbit of the barycentre of the Earth–Moon system

Mercury	$a = 0.38709893 + 0.00000066T$	$e = 0.20563069 + 0.00002527T$
	$i = 7.00487^\circ - 23.51''T$	$\Omega = 48.33167^\circ - 446.30''T$
	$\varpi = 77.45645^\circ + 573.57''T$	$L = 252.25084^\circ + 4.09233880^\circ t$
Venus	$a = 0.72333199 + 0.00000092T$	$e = 0.00677323 - 0.00004938T$
	$i = 3.39471^\circ - 2.86''T$	$\Omega = 76.68069^\circ - 996.89''T$
	$\varpi = 131.53298^\circ - 108.80''T$	$L = 181.97973^\circ + 1.60213047^\circ t$
Earth + Moon	$a = 1.00000011 - 0.00000005T$	$e = 0.01671022 - 0.00003804T$
	$i = 0.00005^\circ - 46.94''T$	$\Omega = -11.26064^\circ - 18,228.25''T$
	$\varpi = 102.94719^\circ + 1198.28''T$	$L = 100.46435^\circ + 0.98560910^\circ t$
Mars	$a = 1.52366231 - 0.00007221T$	$e = 0.09341233 + 0.00011902T$
	$i = 1.85061^\circ - 25.47''T$	$\Omega = 49.57854^\circ - 1020.19''T$
	$\varpi = 336.04084^\circ + 1560.78''T$	$L = 355.45332^\circ + 0.52403304^\circ t$
Jupiter	$a = 5.20336301 + 0.00060737T$	$e = 0.04839266 - 0.00012880T$
	$i = 1.30530^\circ - 4.15''T$	$\Omega = 100.55615^\circ + 1217.17''T$
	$\varpi = 14.75385^\circ + 839.93''T$	$L = 34.40438^\circ + 0.08308676^\circ t$
Saturn	$a = 9.53707032 - 0.00301530T$	$e = 0.05415060 - 0.00036762T$
	$i = 2.48446^\circ + 6.11''T$	$\Omega = 113.71504^\circ - 1591.05''T$
	$\varpi = 92.43194^\circ - 1948.89''T$	$L = 49.94432^\circ + 0.03346063^\circ t$
Uranus	$a = 19.19126393 + 0.00152025T$	$e = 0.04716771 - 0.00019150T$
	$i = 0.76986^\circ - 2.09''T$	$\Omega = 74.22988^\circ + 1681.40''T$
	$\varpi = 170.96424^\circ + 1312.56''T$	$L = 313.23218^\circ + 0.01173129^\circ t$
Neptune	$a = 30.06896348 - 0.00125196T$	$e = 0.00858587 + 0.00002514T$
	$i = 1.76917^\circ - 3.64''T$	$\Omega = 131.72169^\circ - 151.25''T$
	$\varpi = 44.97135^\circ - 844.43''T$	$L = 304.88003^\circ + 0.00598106^\circ t$

Table C.13 Largest satellites of the planets. The giant planets have a large number of satellites, and new ones are found frequently. Also the distinction between a large ring particle and a small satellite is somewhat arbitrary. Therefore this table is not complete. a = semimajor axis, P_{sid} = sidereal period (tropical for Saturn's moons), R means that the motion is retrograde, e = eccentricity, i = inclination with respect to the equator (E = ecliptic), r = radius (calculated), ρ = density (calculated), p = geometric albedo and V_0 mean opposition magnitude

		Discoverer	Year of discovery	a [10^3 km]	[R_p]	P_{sid} [d]	e	i [°]	r [km]	M [M_{planet}]	ρ [g/cm ³]	p	V_0
Earth	Moon			384.4	60.27	27.3217	0.055	18.28–28.58	1737	0.0123	3.34	0.12	-12.74
Mars	Phobos	Hall	1877	9.38	2.76	0.3189	0.015	1.0	$13 \times 11 \times 9$	1.7×10^{-8}	2.0	0.07	11.3
	Deimos	Hall	1877	23.46	6.91	1.2624	0.0005	0.9–2.7	$7 \times 6 \times 5$	3.7×10^{-9}	2.7	0.08	12.4
Jupiter	XVI Metis	Synnott	1979	128	1.79	0.295			20	5×10^{-11}	2.8	0.05	17.5
XV Adrastea	Jewitt	1979	129	1.80	0.298				$13 \times 10 \times 8$	1×10^{-11}	4.4	0.05	19.1
	Danielson												
	Synnott												
V Amalthea	Barnard	1892	181	2.53	0.498	0.003	0.40		$131 \times 73 \times 67$	3.8×10^{-9}	2.7	0.07	14.1
XIV Thebe	Synnott	1979	222	3.11	0.674	0.015	0.8	55 × 45		4×10^{-10}	1.3	0.04	15.7
I Io	Galilei	1610	422	5.90	1.769	0.004	0.04		$1830 \times 1819 \times 1815$	4.7×10^{-5}	3.5	0.63	5.0
II Europa	Galilei	1610	671	9.39	3.551	0.009	0.47	1565		2.5×10^{-5}	3.0	0.67	5.3
III Ganymedes	Galilei	1610	1070	15.0	7.155	0.002	0.21	2634		7.8×10^{-5}	1.9	0.44	4.6
IV Callisto	Galilei	1610	1883	26.3	16.689	0.007	0.51	2403		5.7×10^{-5}	1.9	0.20	5.6
XIII Leda	Kowal	1974	11,094	155.2	238.72	0.148	26.07	5		3×10^{-12}		0.07	20.2
VI Himalia	Perrine	1905	11,480	160.6	250.566	0.158	27.63	85		5×10^{-9}		0.03	14.8
X Lysithea	Nicholson	1938	11,720	163.9	259.22	0.107	29.02	12		4×10^{-11}		0.06	18.4
VII Elara	Perrine	1905	11,737	164.2	259.653	0.207	24.77	40		4×10^{-10}		0.03	16.8
XII Ananke	Nicholson	1951	21,200	297	631 R	0.169	147	10		2×10^{-11}		0.06	18.9
XI Carme	Nicholson	1938	22,600	316	692 R	0.207	164	15		5×10^{-11}		0.06	18.0
VIII Pasiphae	Melotte	1908	23,500	329	735 R	0.378	145	18		1×10^{-10}		0.10	17.0
IX Sinope	Nicholson	1914	23,700	332	758 R	0.275	153	14		4×10^{-11}		0.05	18.3

Table C.13 (Continued)

		Discoverer	Year of discovery	a [10^3 km]	R_p	P_{sid} [d]	e	i [°]	r [km]	M [M_{plan}]	ρ [g/cm ³]	p	V_0	
Saturn														
XVIII	Pan	Showalter	1990	133.58	2.22	0.575			10			0.5		
XV	Atlas	Terrile	1980	137.67	2.28	0.602	0.000	0.3	18 × 17 × 13			0.8	18	
XVI	Prometheus	Collins	1980	139.35	2.31	0.613	0.003	0.0	74 × 50 × 34			0.5	16	
XVII	Pandora	Collins	1980	141.70	2.35	0.629	0.004	0	55 × 44 × 31			0.7	16	
		Carlson												
XI	Epimetheus	Cruikshank	1980	151.42	2.51	0.694	0.009	0.34	69 × 55 × 55	9.5×10^{-10}	0.6	0.8	15	
X	Janus	Pascu	1980	151.47	2.51	0.695	0.007	0.14	97 × 95 × 77	3.4×10^{-9}	0.7	0.9	14	
I	Mimas	W. Herschel	1789	185.52	3.08	0.942	0.020	1.53	209 × 196 × 191	6.6×10^{-8}	1.1	0.5	12.9	
II	Enceladus	W. Herschel	1789	238.02	3.95	1.370	0.005	0.00	256 × 247 × 245	1×10^{-7}	0.9	1.0	11.7	
XIII	Telesto	Smith	1980	294.66	4.89	1.888			15 × 12 × 7			1.0	18.5	
		Larson												
		Reitsema												
III	Tethys	Cassini	1684	294.66	4.89	1.888	0.000	1.86	536 × 528 × 526	1.1×10^{-6}	1.0	0.9	10.2	
XIV	Calypso	Pascu	1980	294.66	4.89	1.888			15 × 8 × 8			1.0	18.7	
		Seidelmann												
		Baum												
		Currie												
IV	Dione	Cassini	1684	377.40	6.26	2.737	0.002	0.02	560	1.9×10^{-6}	1.5	0.7	10.4	
XII	Helene	Laques	1980	377.40	6.26	2.737	0.005	0.0	18 × 16 × 15			0.7	18	
		Lecacheux												
V	Rhea	Cassini	1672	527.04	8.74	4.517	0.001	0.35	764	4.1×10^{-6}	1.2	0.7	9.7	
VI	Titan	Huygens	1665	1221.83	20.3	15.945	0.03	0.33	2575	2.4×10^{-4}	1.9	0.22	8.3	
VII	Hyperion	Bond	1848	1481.1	24.6	21.277	0.10	0.43	180 × 140 × 113	4×10^{-8}	1.4	0.3	14.2	
VIII	Iapetus	Cassini	1671	3561.3	59.1	79.330	0.028	14.72	718	2.8×10^{-6}	1.0	0.5/0.05	11.1	
IX	Phoebe	Pickering	1898	12,952	215	550.48 R	0.163	177 E	110	$7 \times 10^{-10}?$	0.06		16.4	

Table C.13 (Continued)

		Discoverer	Year of discovery	a [10^3 km]	$[R_p]$	P_{std} [d]	e	i [°]	r [km]	M [M_{planet}]	ρ [g/cm 3]	p	V_0
Uranus													
VI	Cordelia	Voyager 2	1986	49.77	1.95	0.335	0.00	0.08	13			0.07	24.1
VII	Ophelia	Voyager 2	1986	53.79	2.10	0.376	0.01	0.10	15			0.07	23.8
VIII	Bianca	Voyager 2	1986	59.17	2.32	0.435	0.00	0.19	21			0.07	23.0
IX	Cressida	Voyager 2	1986	61.78	2.42	0.464	0.00	0.01	31			0.07	22.2
X	Desdemona	Voyager 2	1986	62.68	2.45	0.474	0.00	0.11	27			0.07	22.5
XI	Juliet	Voyager 2	1986	64.35	2.52	0.493	0.00	0.07	42			0.07	21.5
XII	Portia	Voyager 2	1986	66.09	2.59	0.513	0.00	0.06	54			0.07	21.0
XIII	Rosalind	Voyager 2	1986	69.94	2.74	0.558	0.00	0.28	27			0.07	22.5
XIV	Belinda	Voyager 2	1986	75.26	2.94	0.624	0.00	0.03	33			0.07	22.1
XVIII	S/1986 U10	Karkoschka	1999	77.30	3.0	0.637			20			0.07	
XV	Puck	Voyager 2	1985	86.01	3.37	0.762	0.00	0.32	77			0.07	20.2
V	Miranda	Kuiper	1948	129.39	5.06	1.413	0.003	4.2	240 × 234 × 233	8×10^{-7}	1.3	0.27	16.3
I	Ariel	Lassell	1851	191.02	7.47	2.520	0.003	0.3	581 × 578 × 578	1.6×10^{-5}	1.7	0.35	14.2
II	Umbriel	Lassell	1851	266.30	10.42	4.144	0.005	0.36	585	1.4×10^{-5}	1.4	0.19	14.8
III	Titania	W. Herschel	1787	435.91	17.06	8.706	0.002	0.14	789	4.1×10^{-5}	1.7	0.28	13.7
IV	Oberon	W. Herschel	1787	583.52	22.83	13.463	0.001	0.10	761	3.5×10^{-5}	1.6	0.25	13.9
XVI	Caliban	Nicholson	1997	7169	281	579 R	0.08	140 E	30			0.07	22.4
XVII	Sycorax	Nicholson	1997	12,214	477	1289 R	0.5	153 E	60			0.07	20.9
Neptune													
III	Naiad	Voyager 2	1989	48.23	1.95	0.294	0.000	4.74	29			0.06	24.7
IV	Thalassa	Voyager 2	1989	50.07	2.02	0.311	0.000	0.21	40			0.06	23.8
V	Despina	Voyager 2	1989	52.53	2.12	0.335	0.000	0.07	74			0.06	22.6
VI	Galatea	Voyager 2	1989	61.95	2.50	0.429	0.000	0.05	79			0.06	22.3
VII	Larissa	Voyager 2	1989	73.55	2.97	0.555	0.001	0.20	104 × 89			0.06	22.0
VIII	Proteus	Voyager 2	1989	117.65	4.75	1.122	0.000	0.55	218 × 208 × 201			0.06	20.3
I	Triton	Lassell	1846	354.76	14.33	5.877 R	0.000	157.35	1353	2.1×10^{-4}	2.1	0.77	13.5
II	Nereid	Kuiper	1949	5513.4	222.6	360.136	0.751	27.6	170	2×10^{-7}	1.0	0.4	18.7

Table C.14 Some well-known asteroids. a = semimajor axis, e = eccentricity, i = inclination, d = diameter, τ_{sid} = sidereal period of rotation, p = geometric albedo, V_0 = mean opposition magnitude

	Asteroid	Discoverer	Year of discovery	a [AU]	e	i [°]	d [km]	τ_{sid} [h]	p	V_0	Type
1	Ceres	Piazzi	1801	2.77	0.08	10.6	946	9.08	0.07	7.9	C
2	Pallas	Olbers	1802	2.77	0.23	34.8	583	7.88	0.09	8.5	U
3	Juno	Harding	1804	2.67	0.26	13.0	249	7.21	0.16	9.8	S
4	Vesta	Olbers	1807	2.36	0.09	7.1	555	5.34	0.26	6.8	U
5	Astraea	Hencke	1845	2.58	0.19	5.3	116	16.81	0.13	11.2	S
6	Hebe	Hencke	1847	2.42	0.20	14.8	206	7.27	0.16	9.7	S
7	Iris	Hind	1847	2.39	0.23	5.5	222	7.14	0.20	9.4	S
8	Flora	Hind	1847	2.20	0.16	5.9	160	13.60	0.13	9.8	S
9	Metis	Graham	1848	2.39	0.12	5.6	168	5.06	0.12	10.4	S
10	Hygiea	DeGasparis	1849	3.14	0.12	3.8	443	18.00	0.05	10.6	C
243	Ida	Palisa	1884	2.86	0.04	1.1	32	4.63	0.16		
433	Eros	Witt	1898	1.46	0.22	10.8	20	5.27	0.18	11.5	S
588	Achilles	Wolf	1906	5.18	0.15	10.3	70	?	?	16.4	U
624	Hektor	Kopff	1907	5.16	0.03	18.3	230	6.92	0.03	15.3	U
944	Hidalgo	Baade	1920	5.85	0.66	42.4	30	10.06	?	19.2	MEU
951	Gaspra	Neujmin	1916	2.21	0.17	4.1	19	7.04	0.15		
1221	Amor	Delporte	1932	1.92	0.43	11.9	5	?	?	20.4	?
1566	Icarus	Baade	1949	1.08	0.83	22.9	2	2.27	?	12.3	U
1862	Apollo	Reinmuth	1932	1.47	0.56	6.4	2	?	?	16.3	?
2060	Chiron	Kowal	1977	13.64	0.38	6.9	320	?	?	17.3	?
5145	Pholus	Rabinowitz	1992	20.46	0.58	24.7	190				

Table C.15 Principal meteor showers

Shower	Period of visibility	Maximum	Radiant		Meteors per hour	Comet
			α	δ		
Quadrantids	Jan. 1–5	Jan. 3–4	15.5 h	+50°	30–40	
Lyrids	Apr. 19–25	Apr. 22	18.2 h	+34°	10	Thatcher
η Aquarids	May 1–12	May 5	22.4 h	-1°	5–10	Halley
Perseids	Jul. 20–Aug. 18	Aug. 12	3.1 h	+58°	40–50	Swift-Tuttle
κ Cygnids	Aug. 17–24	Aug. 20	19.1 h	+59°	5	
Orionids	Oct. 17–26	Oct. 21	6.3 h	+16°	10–15	Halley
Taurids	Oct. 10–Dec. 5	Nov. 1	3.8 h	+14°, +22°	5	Encke
Leonids	Nov. 14–20	Nov. 17	10.2 h	+22°	10	Tempel-Tuttle
Geminids	Dec. 7–15	Dec. 13–14	7.5 h	+33°	40–50	
Ursids	Dec. 17–24	Dec. 22	13.5 h	+78°	5	

Table C.16 Periodic comets with several perihelion passages observed. N = number of passages observed, τ = time of perihelion passage, P = sidereal period, q = perihelion distance, e = eccentricity, ω = argument of perihelion (1950.0), Ω = longitude of ascending node (1950.0), i = inclination, l = longitude of perihelion, de-

fined here as $l = \Omega + \arctan(\tan \omega \cos i)$, b = latitude of perihelion ($\sin b = \sin \omega \sin i$), Q = aphelion distance. The elements are affected by planetary perturbations as well as reaction forces due to evaporating material the amount of which is difficult to predict

Comet	N	τ	P [a]	q [AU]	e	ω [°]	Ω [°]	i [°]	l [°]	b [°]	Q [AU]
Encke	56	Feb. 9, 1994	3.28	0.331	0.850	186.3	334.0	11.9	160.2	-1.3	4.09
Grigg-Skjellerup	16	Jul. 24, 1992	5.10	0.995	0.664	359.3	212.6	21.1	212.0	-0.3	4.93
Honda-Mrkos-Pajdušáková	8	Sep. 12, 1990	5.30	0.541	0.822	325.8	88.6	4.2	54.5	-2.4	5.54
Tuttle-Giacobini-Kresák	8	Feb. 8, 1990	5.46	1.068	0.656	61.6	140.9	9.2	202.1	8.1	5.14
Tempel 2	19	Mar. 16, 1994	5.48	1.484	0.522	194.9	117.6	12.0	312.1	-3.1	4.73
Wirtanen	8	Sep. 21, 1991	5.50	1.083	0.652	356.2	81.6	11.7	77.9	-0.8	5.15
Clark	8	Nov. 28, 1989	5.51	1.556	0.501	208.9	59.1	9.5	267.7	-4.6	4.68
Forbes	8	Mar. 15, 1993	6.14	1.450	0.568	310.6	333.6	7.2	284.5	-5.4	5.25
Pons-Winnecke	20	Aug. 19, 1989	6.38	1.261	0.634	172.3	92.8	22.3	265.6	2.9	5.62
d'Arrest	15	Feb. 4, 1989	6.39	1.292	0.625	177.1	138.8	19.4	316.0	1.0	5.59
Schwassmann-Wachmann 2	11	Jan. 24, 1994	6.39	2.070	0.399	358.2	125.6	3.8	123.8	-0.1	4.82
Wolf-Harrington	8	Apr. 4, 1991	6.51	1.608	0.539	187.0	254.2	18.5	80.8	-2.2	5.37
Ciacobini-Zinner	12	Apr. 14, 1992	6.61	1.034	0.707	172.5	194.7	31.8	8.3	3.9	6.01
Reinmuth 2	8	Jun. 29, 1994	6.64	1.893	0.464	45.9	295.4	7.0	341.1	5.0	5.17
Perrine-Mrkos	8	Mar. 1, 1989	6.78	1.298	0.638	166.6	239.4	17.8	46.6	4.1	5.87
Arend-Rigaux	7	Oct. 3, 1991	6.82	1.438	0.600	329.1	121.4	17.9	91.7	-9.1	5.75
Borrelly	11	Dec. 18, 1987	6.86	1.357	0.624	353.3	74.8	30.3	69.0	-3.4	5.86
Brooks 2	14	Sep. 1, 1994	6.89	1.843	0.491	198.0	176.2	5.5	14.1	-1.7	5.40
Finlay	11	Jun. 5, 1988	6.95	1.094	0.700	322.2	41.7	3.6	4.0	-2.2	6.19
Johnson	7	Nov. 19, 1990	6.97	2.313	0.366	208.3	116.7	13.7	324.3	-6.4	4.98
Daniel	8	Aug. 31, 1992	7.06	1.650	0.552	11.0	68.4	20.1	78.7	3.8	5.71
Holmes	8	Apr. 10, 1993	7.09	2.177	0.410	23.2	327.3	19.2	349.4	7.4	5.21
Reinmuth 1	8	May 10, 1988	7.29	1.869	0.503	13.0	119.2	8.1	132.0	1.8	5.65
Faye	19	Nov. 15, 1991	7.34	1.593	0.578	204.0	198.9	9.1	42.6	-3.7	5.96
Ashbrook-Jackson	7	Jul. 13, 1993	7.49	2.316	0.395	348.7	2.0	12.5	350.9	-2.4	5.34
Schaumasse	10	Mar. 5, 1993	8.22	1.202	0.705	57.5	80.4	11.8	137.3	10.0	6.94
Wolf	14	Aug. 28, 1992	8.25	2.428	0.406	162.3	203.4	27.5	7.6	8.1	5.74
Whipple	9	Dec. 22, 1994	8.53	3.094	0.259	201.9	181.8	9.9	23.4	-3.7	5.25
Comas Solá	8	Aug. 18, 1987	8.78	1.830	0.570	45.5	60.4	13.0	105.2	9.2	6.68
Väisälä 1	6	Apr. 29, 1993	10.8	1.783	0.635	47.4	134.4	11.6	181.2	8.5	7.98
Tuttle	11	Jun. 27, 1994	13.5	0.998	0.824	206.7	269.8	54.7	106.1	-21.5	10.3
Halley	30	Feb. 9, 1986	76.0	0.587	0.967	111.8	58.1	162.2	305.3	16.4	35.3

Table C.17 Nearest stars. V = apparent visual magnitude, $B - V$ = colour index, r = distance, μ = proper motion, v_r = radial velocity (positive for receding objects)

Name	α_{2000}		δ_{2000}		V	$B - V$	Spectrum	r [pc]	μ ["/a]	v_r [km/s]
	[h]	[min]	[°]	[']						
Sun					-26.8	0.6	G2V			
α Cen C (Proxima)	14	29.7	-62	41	11.0	2.0	M5eV	1.30	3.9	-16
α Cen A	14	39.6	-60	50	-0.0	0.7	G2V	1.33	3.7	-22
α Cen B	14	39.6	-60	50	1.3	0.9	K1V	1.33	3.7	-22
Barnard's star	17	57.8	4	42	9.5	1.7	M5V	1.83	10.3	-108
Wolf 359	10	56.5	7	01	13.5	2.0	M6eV	2.39	4.7	+13
BD+36°2147	11	03.3	35	58	7.5	1.5	M2V	2.54	4.8	-86
α CMa (Sirius) A	6	45.1	-16	43	-1.5	0.0	A1V	2.66	1.3	-8
α CMa (Sirius) B	6	45.1	-16	43	8.4		wdA	2.66	1.3	-8
Luyten 726-8 A	1	39.0	-17	57	12.5		M6eV	2.66	3.3	+29
Luyten 726-8 B (UV Cet)	1	39.0	-17	57	13.0		M6eV	2.66	3.3	+32
Ross 154	18	49.8	-23	50	10.4		M4eV	2.92	0.7	-4
Ross 248	23	41.9	44	11	12.2	1.9	M5eV	3.13	1.6	-81
ε Eri	3	32.9	-9	27	3.7	0.9	K2V	3.26	1.0	+16
Ross 128	11	47.7	0	48	11.1	1.8	M5V	3.31	1.4	-13
Luyten 789-6 A	22	38.6	-15	17	12.8	2.0	M5eV	3.40	3.3	-60
Luyten 789-6 B	22	38.6	-15	17	13.3			3.40	3.3	-60
BD+43°44 A	0	18.4	44	01	8.1	1.6	M3V	3.44	2.9	+13
BD+43°44 B	0	18.4	44	01	11.0	1.8	M6V	3.44	2.9	+20
ε Ind	22	03.4	-56	47	4.7	1.1	K5V	3.45	4.7	-40
BD+59°1915 A	18	42.8	59	38	8.9	1.5	M4V	3.45	2.3	0
BD+59°1915 B	18	42.8	59	38	9.7	1.6	M4V	3.45	2.3	+10
61 Cyg A	21	06.9	38	45	5.2	1.2	K5V	3.46	5.2	-64
61 Cyg B	21	06.9	38	45	6.0	1.4	K7V	3.46	5.2	-64
τ Cet	1	43.1	-15	56	3.5	0.7	G8V	3.48	1.9	-16
CD-36°15693	23	05.9	-35	51	7.4	1.5	M2V	3.51	6.9	+10
α CMi (Procyon) A	7	39.3	5	14	0.4	0.4	F5IV	3.51	1.3	-3
α CMi (Procyon) B	7	39.3	5	14	10.7		wdF	3.51	1.3	
G 51-15	8	29.8	26	47	14.8		M7V	3.62	1.3	
BD+5°1668	7	27.4	5	13	9.8	1.6	M4V	3.76	3.8	+26
Luyten 725-32	1	12.6	-17	00	11.8		M6eV	3.77	1.4	
Kapteyn's star	5	11.7	-45	01	8.8	1.6	M1VI	3.85	8.8	+245
CD-39°14192	21	17.2	-38	52	6.7	1.4	M0eV	3.85	3.5	+21
Krüger 60 A	22	28.0	57	42	9.9	1.6	M3V	3.95	0.9	-26
Krüger 60 B	22	28.0	57	42	11.5	1.8	M4eV	3.95	0.9	-26
Ross 614 A	6	29.4	-2	49	11.2	1.7	M4eV	4.13	1.0	+24
Ross 614 B	6	29.4	-2	49	14.8			4.13	1.0	+24
BD-12°4523	16	30.3	-12	40	10.2	1.6	M5V	4.15	1.2	-13
Wolf 424 A	12	33.3	9	01	13.2	1.8	M6V	4.29	1.8	-5
Wolf 424 B	12	33.3	9	01	13.2			4.29	1.8	-5

Table C.17 (Continued)

Name	α_{2000}		δ_{2000}		V	$B - V$	Spectrum	r [pc]	μ ["/a]	v_r [km/s]
	[h]	[min]	[°]	[']						
van Maanen's star	0	49.2	5	23	12.4	0.6	wdG	4.33	3.0	+54
Luyten 1159–16	2	00.2	13	03	12.2		M5eV	4.48	2.1	
CD–37°15492	0	05.4	–37	21	8.6	1.5	M3V	4.48	6.1	+23
Luyten 143–23	10	44.5	–61	12	13.9		dM	4.48	1.7	
CD–46°11540	17	28.7	–46	54	9.4	1.5	M3	4.52	1.1	
LP 731–58	10	48.2	–11	20	15.6		M7V	4.55	1.6	
Luyten 145–141	11	45.7	–64	50	11.4	0.2	wdA	4.57	2.7	
BD+68°946	17	36.4	68	20	9.1	1.5	M3V	4.63	1.3	–22
CD–49°13515	21	33.6	–49	01	8.7	1.5	M2V	4.63	0.8	+8
BD+50°1725	10	11.3	49	27	6.6	1.4	K2V	4.67	1.5	–26
G 158–27	0	06.7	–07	32	13.7		M5V	4.67	2.1	
BD–15°6290	22	53.3	–14	18	10.2	1.6	M4V	4.69	1.1	+9
CD–44°11909	17	37.1	–44	19	11.0		M5V	4.72	1.1	
G 208–44/45 A	19	53.9	44	25	13.4		M6eV	4.72	0.7	
G 208–44/45 B	19	53.9	44	25	14.3		dM	4.72	0.7	
G 208–44/45 C	19	53.9	44	25	15.5		dM	4.72	0.7	
σ^2 Eri A	4	15.3	–7	39	4.4	0.8	K0V	4.76	4.0	–43
σ^2 Eri B	4	15.3	–7	39	9.5	0.0	wdA	4.76	4.0	–21
σ^2 Eri C	4	15.3	–7	39	11.2	1.7	M4eV	4.76	4.0	–45
BD+20°2465	10	19.6	19	52	9.4	1.5	M4V	4.88	0.5	+11
70 Oph A	18	05.5	2	30	4.2	0.9	K0V	4.98	1.1	–7
70 Oph B	18	05.5	2	30	6.0		K5V	4.98	1.1	–10
BD+44°2051 A	11	05.5	43	32	8.7		M2V	5.00	4.5	+65
BD+44°2051 B	11	05.5	43	32	14.4		M5eV	5.00	4.5	+65
α Aql (Altair)	19	50.8	8	52	0.8	0.2	A7V	5.08	0.7	–26

Table C.18 Brightest stars ($V \leq 2$). V = apparent visual magnitude, $B - V$ = colour index, r = distance. Remarks: b = binary, sb = spectroscopic binary, v = variable

Name		α_{2000}		δ_{2000}		V	$B - V$	Spectrum	r [pc]	Remarks
		[h]	[min]	[°]	[']					
α CMa	Sirius	6	45.2	–16	43	–1.5	0.0	A1V,wdA	2.7	b
α Car	Canopus	6	24.0	–52	42	–0.7	0.2	A9II	60	
α Cen	Rigel	14	39.6	–60	50	–0.3	0.7	G2V,K1V	1.3	b, Proxima
	Kentaurus									2.2° apart
α Boo	Arcturus	14	15.7	19	11	–0.0	1.2	K2IIIIP	11	
α Lyr	Vega	18	36.9	38	47	0.0	0.0	A0V	8	
α Aur	Capella	5	16.7	46	00	0.1	0.8	G2III,G6III	14	b
β Ori	Rigel	5	14.5	–8	12	0.1	–0.0	B8Ia	90	b
α CMi	Procyon	7	39.3	5	14	0.4	0.4	F5IV,wdF	3.5	b
α Eri	Achernar	1	37.7	–57	14	0.5	–0.2	B3Vp	40	

Table C.18 (Continued)

Name		α_{2000}		δ_{2000}		V	$B - V$	Spectrum	r [pc]	Remarks
		[h]	[min]	[°]	[']					
α Ori	Betelgeuze	5	55.2	7	24	0.5	1.9	M2I	200	v 0.4 – 1.3, sb
β Cen	Hadar	14	03.8	-60	22	0.6	-0.2	B1III	60	b
α Aql	Altair	19	50.8	8	52	0.8	0.2	A7V	5.1	
α Cru	Acrux	12	26.6	-63	06	0.8	-0.3	B0.5IV,B1V	120	b 1.6 + 2.1
α Tau	Aldebaran	4	35.9	16	31	0.9	1.5	K5III	20	b, v
α Vir	Spica	13	25.2	-11	10	1.0	-0.2	B1IV	50	sb, several comp.
α Sco	Antares	16	29.4	-26	26	1.0	1.8	M1.5I,B2.5V	50	v 0.9 – 1.8
β Gem	Pollux	7	45.3	28	02	1.2	1.1	K0III	11	
α PsA	Fomalhaut	22	57.6	-29	37	1.2	0.1	A3V	7.0	
α Cyg	Deneb	20	41.4	45	17	1.3	0.1	A2Ia	500	
β Cru	Mimosa	12	47.7	-59	41	1.3	-0.2	B0.5III	150	v, sb
α Leo	Regulus	10	08.4	11	58	1.4	-0.1	B7V	26	b
ϵ CMa	Adhara	6	58.6	-28	58	1.5	-0.2	B2II	170	b
α Gem	Castor	7	34.6	31	53	1.6	0.0	A1V,A2V	14	b
γ Cru	Gacrux	12	31.2	-57	07	1.6	1.6	M3.5III	40	v
λ Sco	Shaula	17	33.6	-37	06	1.6	-0.2	B1.5IV		v
γ Ori	Bellatrix	5	25.1	6	21	1.6	-0.2	B2III	40	
β Tau	Elnath	5	26.3	28	36	1.7	-0.1	B7III	55	
β Car	Miplacidus	9	13.2	-69	43	1.7	0.0	A1III	30	
ϵ Ori	Alnilam	5	36.2	-1	12	1.7	-0.2	B0Ia		
α Gru	Al Na'ir	22	08.2	-46	58	1.7	-0.1	B7IV	20	
ϵ UMa	Alioth	12	54.0	55	58	1.8	-0.0	A0IVp	120	v
γ Vel	Regor	8	09.5	-47	20	1.8	-0.2	WC8,B1IV		b 1.8 + 4.3, each sb
α Per	Mirfak	3	24.3	49	52	1.8	0.5	F5Ib	35	
α UMa	Dubhe	11	03.7	61	45	1.8	1.1	K0III	30	b
ϵ Sgr	Kaus Australis	18	24.2	-34	23	1.9	-0.0	A0II	70	
δ CMa	Wezen	7	08.4	-26	23	1.9	0.7	F8Ia		
ϵ Car	Avior	8	22.5	-59	31	1.9	1.3	K3III,B2V	25	sb
η UMa	Alkaid	13	47.5	49	19	1.9	-0.2	B3V		
θ Sco	Girtab	17	37.3	-43	00	1.9	0.4	F0II	50	
β Aur	Menkalinan	5	59.6	44	57	1.9	0.0	A1IV	30	
ζ Ori	Alnitak	5	40.8	-1	57	1.9	-0.2	O9.5Ib,B0III	45	b 2.1 + 4.2
α TrA	Atria	16	48.7	-69	02	1.9	1.4	K2II–III	40	
γ Gem	Alhena	6	37.7	16	24	1.9	0.0	A1IV	30	
α Pav	Peacock	20	25.7	-56	44	1.9	-0.2	B3V		
δ Vel		8	44.7	-54	43	2.0	0.0	A1V	20	
β CMa	Mirzam	6	22.7	-17	57	2.0	-0.2	B1II–III	70	v
α Hya	Alphard	9	27.6	-8	40	2.0	1.4	K3II–III	60	
α Ari	Hamal	2	07.2	23	28	2.0	1.2	K2III	25	

Table C.19 Some double stars. Magnitudes of the components are m_1 and m_2 , and the angular separation d ; r is the distance of the star

Name		α_{2000}		δ_{2000}		m_1	m_2	Spectrum		d [""]	r [pc]
		[h]	[min]	[°]	[']						
η Cas	Achird	0	49.1	57	49	3.7	7.5	G0V	M0	12	6
γ Ari	Mesarthim	1	53.5	19	18	4.8	4.9	A1p	B9V	8	40
α Psc	Alrescha	2	02.0	2	46	4.3	5.3	A0p	A3m	2	60
γ And	Alamak	2	03.9	42	20	2.4	5.1	K3IIb	B8V,A0V	10	100
δ Ori	Mintaka	5	32.0	-0	18	2.5	7.0	B0III,O9V	B2V	52	70
λ Ori	Meissa	5	35.1	9	56	3.7	5.7	O8e	B0.5V	4	140
ζ Ori	Alnitak	5	40.8	-1	56	2.1	4.2	O9.5Ibe	B0III	2	40
α Gem	Castor	7	34.6	31	53	2.0	3.0	A1V	A5Vm	3	15
γ Leo	Algiedba	10	20.0	19	50	2.6	3.8	K1III	G7III	4	80
ξ UMa	Alula Australis	11	18.2	31	32	4.4	4.9	G0V	G0V	1	7
α Cru	Acrux	12	26.6	-63	06	1.6	2.1	B0.5IV	B1V	4	120
γ Vir	Porrima	12	41.7	-1	27	3.7	3.7	F0V	F0V	3	10
α CVn	Cor Caroli	12	56.1	38	18	2.9	5.5	A0p	F0V	20	40
ζ UMa	Mizar	13	23.9	54	56	2.4	4.1	A1Vp	A1m	14	21
α Cen	Rigel Kentaurus	14	39.6	-60	50	0.0	1.3	G2V	K1V	21	1.3
ϵ Boo	Izar	14	45.0	27	04	2.7	5.3	K0II-III	A2V	3	60
δ Ser		15	34.8	10	32	4.2	5.3	F0IV	F0IV	4	50
β Sco	Graffias	16	05.4	-19	48	2.6	4.9	B1V	B2V	14	110
α Her	Rasalgethi	17	14.6	14	23	3.0-4.0	5.7	M5Ib-II	G5III,F2V	5	120
ρ Her		17	23.7	37	08	4.5	5.5	B9.5III	A0V	4	
70 Oph		18	05.5	2	30	4.3	6.1	K0V	K5V	2	5
ϵ Lyr		18	44.3	39	40	4.8	4.4	A4V,F1V	A8V,F0V	208	50
ϵ^1 Lyr		18	44.3	39	40	5.1	6.2	A4V	F1V	3	50
ϵ^2 Lyr		18	44.4	39	37	5.1	5.3	A8V	F0V	2	50
ζ Lyr		18	44.8	37	36	4.3	5.7	Am	F0IV	44	30
θ Ser	Alya	18	56.2	4	12	4.5	4.9	A5V	A5V	22	30
γ Del		20	46.7	16	07	4.5	5.4	K1IV	F7V	10	40
ζ Aqr		22	28.8	-0	01	4.4	4.6	F3V	F6IV	2	30
δ Cep		22	29.1	58	24	3.5-4.3	7.5	F5Ib-G2Ib	B7IV	41	90

Table C.20 Milky Way Galaxy

Property	Value
Mass	$> 2 \times 10^{11} M_{\odot}$
Disc diameter	30 kpc
Disc thickness (stars)	1 kpc
Disc thickness (gas and dust)	200 pc
Halo diameter	50 kpc
Sun's distance from the centre	8.5 kpc
Sun's orbital velocity	220 km s ⁻¹
Sun's period	240×10^6 a
Direction of the centre (2000.0)	$\alpha = 17 \text{ h } 45.7 \text{ min}$ $\delta = -29^\circ 00'$
Direction of the north pole (2000.0)	$\alpha = 12 \text{ h } 51.4 \text{ min}$ $\delta = +27^\circ 08'$
Galactic coordinates of the celestial north pole	$l = 123^\circ 00'$ $b = +27^\circ 08'$

Table C.21 Members of the Local Group of Galaxies. V = apparent visual magnitude, M_V = absolute visual magnitude, r = distance

	α_{2000}		δ_{2000}		Type	V	M_V	r [kpc]
	[h]	[min]	[$^{\circ}$]	[']				
Milky Way	17	45.7	-29	00	Sbc		-20.9	8
NGC 224 = M31	00	42.7	41	16	Sb	3.2	-21.2	760
NGC 598 = M33	01	33.8	30	30	Sc	5.6	-18.9	790
Large Magellanic Cloud	05	19.6	-69	27	Irr	0.0	-18.5	50
Small Magellanic Cloud	00	52.6	-72	48	Irr	1.8	-17.1	60
NGC 221 = M32	00	42.7	40	52	E2	7.9	-16.5	760
NGC 205	00	40.4	41	41	dE5	8.0	-16.4	760
IC 10	00	20.4	59	17	Irr	7.8	-16.3	660
NGC 6822	19	44.9	-14	48	Irr	7.5	-16.0	500
NGC 185	00	39.0	48	20	dE3	8.5	-15.6	660
IC 1613	01	04.8	02	08	Irr	9.0	-15.3	720
NGC 147	00	33.2	48	30	dE4	9.0	-15.1	660
WLM	00	02.0	-15	28	Irr	10.4	-14.4	930
Sagittarius	18	55.1	-30	29	dE7	3.1	-13.8	24
Fornax	02	39.9	-34	30	dE3	7.6	-13.1	140
Pegasus	23	28.6	14	45	Irr	12.1	-12.3	760
Leo I	10	08.4	12	18	dE3	10.1	-11.9	250
And II	01	16.5	33	26	dE3	12.4	-11.8	700
And I	00	45.7	38	00	dE0	12.7	-11.8	810
Leo A	09	59.4	30	45	Irr	12.7	-11.5	690
Aquarius	20	46.9	-12	51	Irr	13.7	-11.3	1020
SagDIG	19	30.0	-17	41	Irr	15.0	-10.7	1400
Pegasus II = And VI	23	51.7	24	36	dE	14.0	-10.6	830
Pisces = LGS 3	01	03.9	21	54	Irr	14.1	-10.4	810
And III	00	35.3	36	30	dE6	14.2	-10.2	760
And V	01	10.3	47	38	dE	14.3	-10.2	810
Leo II	11	13.5	22	10	dE0	11.5	-10.1	210
Cetus	00	26.1	-11	02	dE	14.3	-10.1	780
Sculptor	01	00.1	-33	43	dE3	10.0	-9.8	90
Phoenix	01	51.1	-44	27	Irr	13.2	-9.8	400
Tucana	22	41.8	-64	25	dE5	15.1	-9.6	870
Sextans	10	13.0	-01	37	dE4	10.3	-9.5	90
Cassiopeia = And VII	23	26.5	50	42	dE	14.7	-9.5	690
Carina	06	41.6	-50	58	dE4	10.6	-9.4	100
Ursa Minor	15	08.8	67	07	dE5	10.0	-8.9	60
Draco	17	20.3	57	55	dE3	10.9	-8.6	80
Ursa Major	158	43.2	51	55	dE	13.2	-6.8	100
Canes Venatici	13	28.0	33	33	dE	13.9	-7.9	220
Boötes	14	00.0	14	30	dE	13.3	-5.7	60
Ursa Major II	08	51.5	63	08	dE	14.3	-3.8	30

Table C.21 (Continued)

	α_{2000}		δ_{2000}		Type	V	M_V	r [kpc]
	[h]	[min]	[$^{\circ}$]	[$'$]				
Coma Berenices	12	27.0	23	54	dE	14.5	-3.7	44
Canes Venatici II	12	57.2	34	19	dE	15.1	-4.8	150
Hercules	16	31.0	12	47.5	dE	14.7	-6.0	140
Leo IV	11	33.0	-0	32	dE	15.9	-5.1	160
And IX	00	52.9	43	12	dE	16.2	-8.3	760
And X	01	06.5	44	48	dE	16.1	-8.1	710
And XI	00	46.3	33	48	dE	17.2	-7.3	760
And XII	00	47.5	34	22	dE	18.1	-6.4	760
And XIII	00	51.8	33	00	dE	17.6	-6.9	760

Table C.22 Optically brightest galaxies. B = apparent blue magnitude, d = apparent diameter, r = distance

Name	α_{2000}		δ_{2000}		Type	B	d [""]	r [Mpc]
	[h]	[min]	[$^{\circ}$]	[$'$]				
NGC 55	0	15.1	-39	13	Sc/Irr	7.9	30 × 5	2.3
NGC 205	0	40.4	41	41	E6	8.9	12 × 6	0.7
NGC 221 = M32	0	42.7	40	52	E2	9.1	3.4 × 2.9	0.7
NGC 224 = M31	0	42.8	41	16	Sb	4.3	163 × 42	0.7
NGC 247	0	47.2	-20	46	S	9.5	21 × 8	2.3
NGC 253	0	47.6	-25	17	Sc	7.0	22 × 5	2.3
Small Magellanic Cloud	0	52.6	-72	48	Irr	2.9	216 × 216	0.06
NGC 300	0	54.9	-37	41	Sc	8.7	22 × 16	2.3
NGC 598 = M33	1	33.9	30	39	Sc	6.2	61 × 42	0.7
Fornax	2	39.9	-34	32	dE	9.1	50 × 35	0.2
Large Magellanic Cloud	5	23.6	-69	45	Irr/Sc	0.9	432 × 432	0.05
NGC 2403	7	36.9	65	36	Sc	8.8	22 × 12	2.0
NGC 2903	9	32.1	21	30	Sb	9.5	16 × 7	5.8
NGC 3031 = M81	9	55.6	69	04	Sb	7.8	25 × 12	2.0
NGC 3034 = M82	9	55.9	69	41	Sc	9.2	10 × 1.5	2.0
NGC 4258 = M106	12	19.0	47	18	Sb	8.9	19 × 7	4.3
NGC 4472 = M49	12	29.8	8	00	E4	9.3	10 × 7	11
NGC 4594 = M104	12	40.0	-11	37	Sb	9.2	8 × 5	11
NGC 4736 = M94	12	50.9	41	07	Sb	8.9	13 × 12	4.3
NGC 4826 = M64	12	56.8	21	41		9.3	10 × 4	3.7
NGC 4945	13	05.4	-49	28	Sb	8.0	20 × 4	3.9
NGC 5055 = M63	13	15.8	42	02	Sb	9.3	8 × 3	4.3
NGC 5128 = Cen A	13	25.5	-43	01	E0	7.9	23 × 20	3.9
NGC 5194 = M51	13	29.9	47	12	Sc	8.9	11 × 6	4.3
NGC 5236 = M83	13	37.0	-29	52	Sc	7.0	13 × 12	2.4
NGC 5457 = M101	14	03.2	54	21	Sc	8.2	23 × 21	4.3
NGC 6822	19	45.0	-14	48	Irr	9.2	20 × 10	0.7

Table C.23

Constellations. The first column gives the abbreviation of the Latin name used to form star names

Abbreviation	Latin name	Genitive	English name
And	Andromeda	Andromedae	Andromeda
Ant	Antlia	Antliae	Air Pump
Aps	Apus	Apodis	Bird of Paradise
Aql	Aquila	Aquilae	Eagle
Aqr	Aquarius	Aquarii	Water-bearer
Ara	Ara	Arae	Altar
Ari	Aries	Arietis	Ram
Aur	Auriga	Aurigae	Charioteer
Boo	Boötes	Boötis	Herdsman
Cae	Caelum	Caeli	Chisel
Cam	Camelopardalis	Camelopardalis	Giraffe
Cnc	Cancer	Cancri	Crab
CMa	Canis Major	Canis Majoris	Great Dog
CMi	Canis Minor	Canis Minoris	Little Dog
Cap	Capricornus	Capricorni	Sea-goat
Car	Carina	Carinae	Keel
Cas	Cassiopeia	Cassiopeiae	Cassiopeia
Cen	Centaurus	Centauri	Centaurus
Cep	Cepheus	Cephei	Cepheus
Cet	Cetus	Ceti	Whale
Cha	Chamaeleon	Chamaeleontis	Chameleon
Cir	Circinus	Circini	Compasses
Col	Columba	Columbae	Dove
Com	Coma Berenices	Comae Berenices	Berenice's Hair
CrA	Corona Australis	Coronae Australis	Southern Crown
CrB	Corona Borealis	Coronae Borealis	Northern Crown
Crv	Corvus	Corvi	Crow
Crt	Crater	Crateris	Cup
Cru	Crux	Crucis	Southern Cross
CVn	Canes Venatici	Canum Venaticorum	Hunting Dogs
Cyg	Cygnus	Cygni	Swan
Del	Delphinus	Delphini	Dolphin
Dor	Dorado	Doradus	Swordfish
Dra	Draco	Draconis	Dragon
Equ	Equuleus	Equulei	Little Horse
Eri	Eridanus	Eridani	Eridanus
For	Fornax	Fornacis	Furnace
Gem	Gemini	Geminorum	Twins
Gru	Grus	Gruis	Crane
Her	Hercules	Herculis	Hercules
Hor	Horologium	Horologii	Clock
Hya	Hydra	Hydrae	Water Serpent
Hyi	Hydrus	Hydri	Water Snake
Ind	Indus	Indi	Indian

Table C.23 (Continued)

Abbreviation	Latin name	Genitive	English name
Lac	<i>Lacerta</i>	<i>Lacertae</i>	Lizard
Leo	<i>Leo</i>	<i>Leonis</i>	Lion
Lep	<i>Lepus</i>	<i>Leporis</i>	Hare
Lib	<i>Libra</i>	<i>Librae</i>	Scales
LMi	<i>Leo Minor</i>	<i>Leonis Minoris</i>	Little Lion
Lup	<i>Lupus</i>	<i>Lupi</i>	Wolf
Lyn	<i>Lynx</i>	<i>Lyncis</i>	Lynx
Lyr	<i>Lyra</i>	<i>Lyrae</i>	Lyre
Men	<i>Mensa</i>	<i>Mensae</i>	Table Mountain
Mic	<i>Microscopium</i>	<i>Microscopii</i>	Microscope
Mon	<i>Monoceros</i>	<i>Monocerotis</i>	Unicorn
Mus	<i>Musca</i>	<i>Muscae</i>	Fly
Nor	<i>Norma</i>	<i>Normae</i>	Square
Per	<i>Perseus</i>	<i>Persei</i>	Perseus
Phe	<i>Phoenix</i>	<i>Phoenicis</i>	Phoenix
Pic	<i>Pictor</i>	<i>Pictoris</i>	Painter
PsA	<i>Piscis Austrinus</i>	<i>Piscis Austrini</i>	Southern Fish
Psc	<i>Pisces</i>	<i>Piscium</i>	Fishes
Pup	<i>Puppis</i>	<i>Puppis</i>	Poop
Pyx	<i>Pyxis</i>	<i>Pyxidis</i>	Compass
Ret	<i>Reticulum</i>	<i>Reticuli</i>	Net
Scl	<i>Sculptor</i>	<i>Sculptoris</i>	Sculptor
Sco	<i>Scorpius</i>	<i>Scorpii</i>	Scorpion
Sct	<i>Scutum</i>	<i>Scuti</i>	Sobieski's Shield
Ser	<i>Serpens</i>	<i>Serpentis</i>	Serpent
Sex	<i>Sextans</i>	<i>Sextantis</i>	Sextant
Sge	<i>Sagitta</i>	<i>Sagittae</i>	Arrow
Sgr	<i>Sagittarius</i>	<i>Sagittarii</i>	Archer
Tau	<i>Taurus</i>	<i>Tauri</i>	Bull
Tel	<i>Telescopium</i>	<i>Telescopii</i>	Telescope
TrA	<i>Triangulum Australe</i>	<i>Trianguli Australis</i>	Southern Triangle
Tri	<i>Triangulum</i>	<i>Trianguli</i>	Triangle
Tuc	<i>Tucana</i>	<i>Tucanae</i>	Toucan
UMa	<i>Ursa Major</i>	<i>Ursae Majoris</i>	Great Bear
UMi	<i>Ursa Minor</i>	<i>Ursae Minoris</i>	Little Bear
Vel	<i>Vela</i>	<i>Velorum</i>	Sails
Vir	<i>Virgo</i>	<i>Virginis</i>	Virgin
Vol	<i>Volans</i>	<i>Volantis</i>	Flying Fish
Vul	<i>Vulpecula</i>	<i>Vulpeculae</i>	Fox
Oct	<i>Octans</i>	<i>Octantis</i>	Octant
Oph	<i>Ophiuchus</i>	<i>Ophiuchi</i>	Serpent-bearer
Ori	<i>Orion</i>	<i>Orionis</i>	Orion
Pav	<i>Pavo</i>	<i>Pavonis</i>	Peacock
Peg	<i>Pegasus</i>	<i>Pegasi</i>	Pegasus

Table C.24 Largest optical telescopes.
 D = diameter of the mirror

Telescope	Location	Completion year	D [m]
GranTeCan	La Palma	2009	10.4
William M. Keck Telescope I	Mauna Kea, Hawaii	1992	10
William M. Keck Telescope II	Mauna Kea, Hawaii	1996	10
Southern African Large Telescope	Sutherland, South Africa	2005	10
Subaru Telescope	Mauna Kea, Hawaii	1999	8.3
Large Binocular Telescope 1	Mt. Graham, Arizona	2005	8.4
Kueyen Telescope (VLT 2)	Cerro Paranal, Chile	1999	8.2
Melipal Telescope (VLT 3)	Cerro Paranal, Chile	2000	8.2
Yepun Telescope (VLT 4)	Cerro Paranal, Chile	2000	8.2
Gemini North Telescope	Mauna Kea, Hawaii	1999	8.1
Gemini South Telescope	Cerro Pachon, Chile	2000	8.1
Multi-Mirror Telescope	Mt. Hopkins, Arizona	1999	6.5
Walter Baade (Magellan 1 Telescope)	Las Campanas, Chile	2000	6.5
Landon Clay (Magellan 2 Telescope)	Las Campanas, Chile	2002	6.5

Table C.25 Largest parabolic radio telescopes. D = diameter of the antenna, λ_{\min} = shortest wavelength

		Completion year	D [m]	λ_{\min} [cm]	Remarks
Arecibo	Puerto Rico, USA	1963	305	5	Fixed disk; limited tracking
Green Bank	West Virginia, USA	2001	100×110	0.3	The largest fully steerable telescope
Effelsberg	Bonn, Germany	1973	100	0.8	
Jodrell Bank	Macclesfield, Great Britain	1957	76.2	10–20	First large paraboloid antenna
Jevpatoria	Crimea	1979	70	1.5	
Parkes	Australia	1961	64	2.5	Innermost 17 m of dish can be used down to 3 mm wavelengths
Goldstone	California, USA		64	1.5	Belongs to NASA deep space network
Tidbinbillla	Australia		64	1.3	NASA
Madrid	Spain		64	1.3	NASA

Table C.26 Millimetre and submillimetre telescopes and interferometers. h = altitude above sea level, D = diameter of the antenna, λ_{\min} = shortest wavelength

Institute	Location	h [m]	D [m]	λ_{\min} [mm]	Remarks; operational since
NRAO, VLA	New Mexico, USA	2124	25	7	27 antennas $d_{\max} = 36.6$ km 1976
NRAO, VLBA	USA	16–3720	25	13	10 antennas 1988–1993
Max-Planck-Institut für Radioastronomie & University of Arizona	Mt. Graham, USA	3250	10	0.3	1994
California Institute of Technology	Mauna Kea, Hawaii	4100	10.4	0.3	1986
Science Research Council England & Holland	Mauna Kea, Hawaii	4100	15.0	0.5	The James Clerk Maxwell Telescope 1986
California Institute of Technology	Owens Valley, USA	1220	10.4	0.5	3 antenna interferometer 1980
Sweden-ESO Southern Hemisphere Millimeter Antenna (SEST)	La Silla, Chile	2400	15.0	0.6	1987
Institut de Radioastronomie Millimetrique (IRAM), France & Germany	Plateau de Bure, France	2550	15.0	0.6	3 antenna interferometer 1990; fourth antenna 1993
IRAM	Pico Veleta, Spain	2850	30.0	0.9	1984
National Radio Astronomy Observatory (NRAO)	Kitt Peak, USA	1940	12.0	0.9	1983 (1969)
University of Massachusetts	New Salem, USA	300	13.7	1.9	radom 1978
University of California, Berkeley	Hat Creek Observatory	1040	6.1	2	3 antenna interferometer 1968
Purple Mountain Observatory	Nanjing, China	3000	13.7	2	radom 1987
Daeduk Radio Astronomy Observatory	Söul, South-Korea	300	13.7	2	radom 1987
University of Tokyo	Nobeyama, Japan	1350	45.0	2.6	1982
University of Tokyo	Nobeyama, Japan	1350	10.0	2.6	5 antenna interferometer 1984
Chalmers University of Technology	Onsala, Sweden	10	20.0	2.6	radom 1976

Table C.27 Some important astronomical satellites and space probes 1980–2002

Satellite		Launch date	Target
Solar Max	USA	Feb. 14, 1980	Sun
Venera 13	SU	Oct. 30, 1981	Venus
Venera 14	SU	Nov. 4, 1981	Venus
IRAS	USA	Jan. 25, 1983	infrared
Astron	SU	Mar. 23, 1983	ultraviolet
Venera 15	SU	Jun. 2, 1983	Venus
Venera 16	SU	Jun. 7, 1983	Venus
Exosat	ESA/USA	May 26, 1983	X-ray
Vega 1	SU	Dec. 15 1984	Venus/Halley
Vega 2	SU	Dec. 21, 1984	Venus/Halley
Giotto	ESA	Jul. 2, 1985	Halley
Suisei	Japan	Aug. 18, 1985	Halley
Ginga	Japan	Feb. 5, 1987	X-ray
Magellan	USA	May 4, 1989	Venus
Hipparcos	ESA	Aug. 8, 1989	astrometry
COBE	USA	Nov. 18, 1989	cosmic background radiation
Galileo	USA	Oct. 18, 1989	Jupiter etc.
Granat	SU	Dec. 1, 1989	gamma ray
Hubble	USA/ESA	Apr. 24, 1990	UV, visible
Rosat	Germany	Jun. 1, 1990	X-ray
Gamma	SU	Jul. 11, 1990	gamma ray
Ulysses	ESA	Oct. 6, 1990	Sun
Compton	USA	Apr. 5, 1991	gamma ray
EUVE	USA	Jun. 7, 1992	extreme UV
Asuka	Japan	Feb. 20, 1993	X-ray
Clementine	USA	Jan. 25, 1994	Moon
ISO	ESA	Nov. 17, 1995	infrared
SOHO	ESA	Dec. 2, 1995	Sun
Near-Shoemaker	USA	Feb. 17, 1996	Mathilde, Eros
BeppoSAX	Italy	Apr. 30, 1996	X-ray
Mars Global Surveyor	USA	Nov. 7, 1996	Mars
Cassini/Huygens	USA/ESA	Oct. 15, 1997	Saturn, Titan
Mars Pathfinder/Sojourner	USA	Dec. 4, 1996	Mars
Lunar Prospector	USA	Jan. 6, 1998	Moon
Nozomi	Japan	Jul. 4, 1998	Mars
Deep Space 1	USA	Oct. 24, 1998	Braille, Borrelly
Stardust	USA	Feb. 7, 1999	Wild 2
Chandra	USA	Jul. 23, 1999	X-ray
XMM-Newton	ESA	Dec. 10, 1999	X-ray
Hete 2	USA	Oct. 9, 2000	gamma ray
Mars Odyssey	USA	Apr. 7, 2001	Mars
MAP	USA	Jun. 30, 2001	cosmic background radiation

Table C.27 (Continued)

Satellite		Launch date	Target
Genesis	USA	Aug. 8, 2001	solar particles
RHESSI	USA	Feb. 5, 2002	Sun
Grace	Germany-USA	Mar. 17, 2002	Earth's gravity
Integral	ESA	Oct. 17, 2002	gamma ray
Galex	USA	Apr. 28, 2003	galaxies
Hayabusa	Japan	May 9, 2003	Itokawa
Mars Express	ESA	Jun. 2, 2003	Mars
Spirit	USA	Jun. 10, 2003	Mars
Opportunity	USA	Jul. 8, 2003	Mars
Spitzer	USA	Jun. 10, 2003	infrared
Smart-1	ESA	Sep. 28, 2003	Moon
Rosetta	ESA	Mar. 2, 2004	Churyumov-Gerasimenko
Gravity Probe B	USA	Apr. 20, 2004	relativity
Messenger	USA	Aug. 3, 2004	Mercury
Swift	USA	Nov. 20, 2004	gamma ray bursts
Deep Impact	USA	Jan. 12, 2005	Tempel 1
Mars Recon Orbiter	USA	Aug. 12, 2005	Mars
Venus Express	ESA	Nov. 9, 2005	Venus
New Horizons	USA	Jan. 19, 2006	Pluto
Akari	Japan	Feb. 22, 2006	infrared

Answers to Exercises

Chapter 2

2.1 The distance is ≈ 7640 km, the northernmost point is 79°N , 45°W , in North Greenland, 1250 km from the North Pole.

2.2 The star can culminate south or north of zenith. In the former case we get $\delta = 65^\circ$, $\phi = 70^\circ$, and in the latter $\delta = 70^\circ$, $\phi = 65^\circ$.

2.3 (a) $\phi > 58^\circ 7'$. If refraction is taken into account, the limit is $57^\circ 24'$. (b) $\phi = \delta = 7^\circ 24'$. (c) $-59^\circ 10' \leq \phi \leq -0^\circ 50'$.

2.4 Pretty bad.

2.5 $\lambda_\odot = 70^\circ 22'$, $\beta_\odot = 0^\circ 0'$, $\lambda_\oplus = 250^\circ 22'$, $\beta_\oplus = 0^\circ 0'$.

2.6 (c) $\Theta_0 = 18$ h.

2.8 $\alpha = 6$ h 45 min 9 s, $\delta = -16^\circ 43'$.

2.9 $v_t = 16.7 \text{ km s}^{-1}$, $v = 18.5 \text{ km s}^{-1}$, after $\approx 61,000$ years. $\mu = 1.62''$ per year, parallax $0.42''$.

Chapter 3

3.1 (a) The flux density in the focal plane as well as the exposure time are proportional to $(D/f)^2$. Thus the required exposure is 3.2 s. (b) 1.35 cm and 1.80 cm. (c) 60 and 80.

3.2 (a) $0.001''$ (note that the aperture is a line rather than circular; therefore the coefficient 1.22 should not be used). (b) 140 m.

Chapter 4

4.1 0.9.

4.2 The absolute magnitude will be -17.5 and apparent 6.7.

4.3 $N(m+1)/N(m) = 10^{3/5} = 3.98$.

4.4 $r = 2.1$ kpc, $E_{B-V} = 0.7$, and $(B-V)_0 = 0.9$.

4.5 (a) $\Delta m = 1.06$ mag, $m = 2.42$. (b) $\tau = -\ln 0.85^6 \approx 0.98$.

Chapter 5

5.2 $n = 166$, which corresponds to $\lambda = 21.04$ cm. Such transitions would keep the population of the state $n = 166$ very high, resulting in downward transitions also from this state. Such transitions have not been detected. Hence, the line is produced by some other process.

5.3 If we express the intensity as something per unit wavelength, we get $\lambda_{\max} = 1.1$ mm. If the intensity is given per unit frequency, we have $\lambda_{\max} = 1.9$ mm. The total intensity is $2.6 \times 10^{13} \text{ W m}^{-2} \text{ sterad}^{-1}$. At 550 nm the intensity is practically zero.

5.4 (a) $L = 1.35 \times 10^{29}$ W. The flux in the given interval can be found by integrating Planck's law numerically. Using the Wien approximation a rather complicated expression can be derived. Both methods give the result that 3.3 % of the radiation is in the visual range, thus $L_V = 4.45 \times 10^{27}$ W. (b) At a distance of 10 pc the observed flux density is 3.7×10^{-9} W m $^{-2}$. (b) 10.3 km.

5.5 $M_{\text{bol}} = 0.87$, whence $R = 2.0 R_{\odot}$.

5.6 $T = 1380$ K. There are several strong absorption lines in this spectral region, reducing the brightness temperature.

5.8 $v_{\text{rms}} \approx 6700$ km s $^{-1}$.

Chapter 6

6.1 $v_a/v_p = (1 - e)/(1 + e)$. For the Earth this is 0.97.

6.2 $a = 1.4581$ AU, $v \approx 23.6$ km s $^{-1}$.

6.3 The period must equal the sidereal rotation period of the Earth. $r = 42,339$ km = $6.64 R_{\oplus}$. Areas within 8.6° from the poles cannot be seen from geostationary satellites. The hidden area is 1.1 % of the total surface area.

6.4 $\rho = 3\pi/(GP^2(\alpha/2)^3) \approx 1400$ kg m $^{-3}$.

6.5 $M = 90^\circ$, $E = 90.96^\circ$, $f = 91.91^\circ$.

6.6 The orbit is hyperbolic, $a = 3.55 \times 10^7$ AU, $e = 1 + 3.97 \times 10^{-16}$, $r_p = 2.1$ km. The comet will hit the Sun.

6.7 The orbital elements of the Earth calculated from Table C.12 are $a = 1.0000$, $e = 0.0167$, $i = 0.0004^\circ$, $\Omega = -11.13^\circ$, $\varpi = 102.9^\circ$, $L = 219.5^\circ$. The geocentric radius vector of the Sun in the ecliptic coordinates is

$$\mathbf{r} = \begin{pmatrix} 0.7583 \\ 0.6673 \\ 0.0 \end{pmatrix}.$$

The corresponding equatorial radius vector is

$$\mathbf{r} = \begin{pmatrix} 0.7583 \\ 0.6089 \\ 0.2640 \end{pmatrix},$$

which gives $\alpha \approx 2 \text{ h } 35 \text{ min } 3 \text{ s}$, $\delta \approx 15.19^\circ$. The exact direction is $\alpha = 2 \text{ h } 34 \text{ min } 53 \text{ s}$, $\delta = 15.17^\circ$.

Chapter 7

7.1 Assuming the orbits are circular, the greatest elongation is $\arcsin(a/1 \text{ AU})$. For Mercury this is 23° and for Venus 46° . The elongation of a superior planet can be anything up to 180° . The sky revolves about 15° per hour, and thus corresponding times for Mercury and Venus are 1 h 30 min and 3 h 5 min, respectively. In opposition Mars is visible the whole night. These values, however, depend on the actual declinations of the planets.

7.2 (a) 8.7° . (b) The Earth must be 90° from the ascending node of Venus, which is the situation about 13 days before vernal and autumnal equinoxes, around March 8 and September 10.

7.3 (a) If the radii of the orbits are a_1 and a_2 , the angular velocity of the retrograde motion is

$$\frac{d\lambda}{dt} = \frac{\sqrt{GM}}{\sqrt{a_1 a_2}(\sqrt{a_1} + \sqrt{a_2})}.$$

(b) In six days Pluto moves about 0.128° corresponding to 4 mm. For a main belt asteroid the displacement is almost 4 cm.

7.4 If the orbital velocity of the planet is v the deviation in radians is

$$\alpha = \frac{v}{c} = \frac{1}{c} \sqrt{\frac{GM_{\odot}}{a}}.$$

This is greatest for Mercury, $\alpha = 0.00016$ rad = $33''$. This planetary aberration must be taken into account when computing accurate ephemerides. The deviation is largest when the planet is in conjunction or opposition and moves almost perpendicularly to the line of sight.

7.5 $p = 0.11$, $q = 2$, and $A = 0.2$. In reality the Moon reflects most of the light directly backwards (opposition effect), and thus q and A are much smaller.

7.6 $\Delta m = 0.9$. The surface brightness remains constant.

7.7 The absolute magnitude is $V(1, 0) = 23$. (a) $m = 18.7$. (b) $m = 14.2$. At least a 15 cm telescope is needed to detect the asteroid even one day before the collision.

7.8 Using the values of Appendix C, the gravitational acceleration is 9.865 m/s^2 at the poles and 9.799 m/s^2 on the equator. On the equator then rotational velocity is 464 m/s and the centrifugal acceleration 0.034 m/s^2 ; thus the total acceleration is 9.765 m/s^2 .

Chapter 8

8.1 $P_{\text{sid}} = 11.9 \text{ a}$, $a = 5.20 \text{ AU}$, $d = 144,000 \text{ km}$. Obviously the planet is Jupiter.

8.2 (a) Hint: If there is a synodic period P there must be integers p and q such that $(n_2 - n_1)P = 2\pi p$ and $(n_3 - n_1)P = 2\pi q$. Sometimes one can see claims that the configuration of the whole planetary system will recur after a certain period. Such claims are obviously nonsense. (b) 7.06 d .

8.3 Apply (7.1), where now $P_1 = 24.62$ and $P_2 = 30.30$. Hence $P_{1,2} = 131.34 \text{ h} = 5 \text{ d } 11 \text{ h } 20 \text{ min}$. One revolution as seen from the surface of Mars takes 131.34 hours. Thus Deimos seems to move $360/131.34 = 2.74$ degrees an hour.

8.4 According to Example 8.1 the distance of the Roche limit is $R \approx 2.5 \times \mathcal{R}$, where \mathcal{R} is the radius of the planet. The following table gives the distances of the nearest moon and the inner edge of the brightest ring of each planet.

Planet	Nearest moon	Radius of the planet	Radius of moon's orbit [km]	Ratio
Earth	Moon	6378	384,400	60
Mars	Phobos	3396	9377	2.8
Jupiter	Metis	71,492	127,690	1.8
–	rengas	–	122,500	1.7
Saturn	Pan	60,268	133,584	2.2
–	ring	–	92,000	1.5
Uranus	Cordelia	25,559	49,751	1.9
–	ring	–	44,718	1.7
Neptune	Naiad	24,764	48,227	1.9
–	ring	–	63,000	2.5

The nearest satellites and inner edges of rings are either close to the Roche limit or inside it. The Earth's Moon is an exception in this respect.

8.5 The densest hexagonal packing gives a volume density of

$$\frac{\pi}{3\sqrt{3}} \times \frac{2r}{\sqrt{8/3r}} = \frac{\pi}{3\sqrt{2}} \approx 0.74.$$

The mass of hydrogen atoms that could be packed in one cubic metre is $(1/5.5 \times 10^{-11})^3 \times 1.67 \times 10^{-27} \times 0.74 = 7428 \text{ kg}$. The density of Jupiter calculated this way would be 7428 kg/m^3 . The actual density is only 1326 kg/m^3 .

Chapter 9

9.1 c g a d f e b; the actual spectral classes from top to bottom are A0, M5, O6, F2, K5, G2, B3.

Chapter 10

10.1 The period is $P = 1/\sqrt{2}$ years and the relative velocity $42,100 \text{ m s}^{-1}$. The maximum separation of the lines is 0.061 nm .

10.3 Substituting the values to the equation of Example 10.1 we get an equation for a . The solution is $a = 4.4 \text{ AU}$. The mass of the planet is $0.0015 M_\odot$.

Chapter 11**11.1** 10.5.**11.2** (a) $9.5 \times 10^{37} \text{ s}^{-1}$. (b) The neutrino production rate is $1.9 \times 10^{38} \text{ s}^{-1}$, and each second 9×10^{28} neutrinos hit the Earth.**11.3** The mean free path is $1/\kappa\rho \approx 42,000 \text{ AU}$.**Chapter 12****12.1** $t_{\text{ff}} = 6.7 \times 10^5 \text{ a}$. Stars are born at the rate $0.75 M_\odot$ per year.**12.2** $t_t \approx 400,000 \text{ a}$, $t_n \approx 3 \times 10^8 \text{ a}$.**12.3** About 900 million years.**Chapter 13****13.1** (a) $6.3 \times 10^7 \text{ W m}^{-2}$. (b) 16 m^2 .**13.2** 807 W m^{-2} .**Chapter 14****14.1** $dr/r = -0.46$, $dM = 0.14$.**14.2** (a) $T = 3570 \text{ K}$. (b) $R_{\min}/R_{\max} = 0.63$.**14.3** (a) 1300 pc. (b) 860 years ago; due to inaccuracies a safe estimate would be 860 ± 100 years. Actually, the explosion was observed in 1054. (c) -7.4 .**Chapter 15****15.1** $L = 2.3 \times 10^{40} \text{ kg m}^2 \text{s}^{-1}$, $dR = 45 \text{ m}$.**15.2** (a) $M = 0.5 M_\odot$, $a = 0.49 \times 10^9 \text{ ms}^{-2} \approx 4 \times 10^7 \text{ g}$. (b) A standing astronaut will be subject to a stretching tidal force. The gravitational acceleration felt by the feet is $3479 \text{ ms}^{-2} \approx 355g$ larger than that felt by the head. If the astronaut lies tangentially, (s)he will experience a compressing force of 177 g .**15.3** $v = v_e(1 - GM/(Rc^2))$. If $\Delta v/v$ is small, we have also $\Delta\lambda = (GM/Rc^2)\lambda_e$. A photon emitted from the Sun is reddened by $2.1 \times 10^{-6}\lambda_e$. In yellow light (550 nm) the change is 0.0012 nm.**Chapter 16****16.1** 2.6 kpc and 0.9 kpc, $a = 1.5 \text{ mag/kpc}$.**16.2** 7 km s^{-1} .**16.3** The velocity of the proton is $v = 0.0462c = 1.38 \times 10^7 \text{ m s}^{-1}$. The radius of the orbit is $r = mv/qB = 0.01 \text{ AU}$.**Chapter 17****17.1** 7.3.**17.2** The potential energy is approximately $U = -G(m^2n^2/(2R))$, where m is the mass of one star, n the number of stars (there are $n(n-1)/2 \approx n^2/2$ pairs), and R the radius of the cluster. The average velocity is $\approx \sqrt{Gmn/(2R)} = 0.5 \text{ km s}^{-1}$.**Chapter 18****18.1** $\mu = 0.0055'' \text{ a}^{-1}$.**18.2** (a) 5.7 kpc. (b) 11 kpc. Possible reasons for the discrepancy include: (1) The distance is so large that the approximations used for deriving Oort's formulae are not very good. (2) Taking into account the interstellar extinction will reduce the distance in (b). (3) The peculiar velocity of the star was neglected.**18.3** (a) 3 (and the Sun). (b) The number is of the order of 100,000. This is a typical selection effect: bright stars are rare but they are visible over long distances.**18.4** (a) If the thickness of the disk is H , the light has to travel a distance $s = \min\{r, (H/2)\sec b\}$ in the interstellar medium. Thus the magnitude will be $m = M + 5\lg(r/10 \text{ pc}) + as$. (b) $s = 200 \text{ pc}$, $m = 10.2$.

Chapter 19

19.1 (a) 26, (b) 25.

19.2 The diameter must be of the order one light-week ≈ 1200 AU. $M = -23.5$. If this is the bolometric magnitude, the luminosity is $L \approx 2 \times 10^{11} L_\odot$, corresponding to $210 L_\odot \text{AU}^{-3}$.

Chapter 20

20.1 $r = v/H = 93$ Mpc (if $H = 75$), diameter is 35 kpc, $M = -20.4$. Potential sources of error include: (1) inaccuracy of the Hubble constant, (2) peculiar velocity of the galaxy, (3) intergalactic extinction, (4) only 2-dimensional projection is observed, and the edge depends on the limiting magnitude used.

20.2 If $H = 50 \text{ km s}^{-1} \text{Mpc}^{-1}$, $r = 51$ Mpc, or 74 times the distance of M31. If H is doubled, the distance is reduced to half of that, but is still higher than the value obtained in Exercise 19.1. A possible explanation is the peculiar velocity of the galaxy.

20.3 $m_v = 1.5 \times 10^{-35}$ kg, or 0.00002 times the mass of the electron.

Chapter 21

21.1 If the greenhouse effect is neglected, the distances corresponding to temperatures 373 K and 273 K are 0.47 AU and 0.87 AU, respectively. The effective temperature of the young Sun was $T = \sqrt[4]{0.7} \times 5785$ K = 5291 K. The corre-

sponding limits then were 0.39 AU and 0.73 AU. Thus the continuously habitable zone would extend from 0.47 to 0.73 AU.

21.2 If there are n stars in a volume V , the distance between the nearest neighbours is of the order $\sqrt[3]{V/n}$. If binary and triple stars in Table C.17 are counted as single objects, there are 47 stars in a volume of 520 cubic parsecs. (a) If a fraction 0.01 of them is inhabited, the average distance between neighbouring civilisations is 10 parsecs; (b) 100 parsecs.

21.3 Since comets are mostly dirty ice, the density is close to that of water. Thus the mass of the comet is about 5×10^8 kg. A freely falling object will reach a velocity that is the same as the escape velocity from the Earth, 11 km/s. The kinetic energy of the comet is at least 3×10^{16} J, corresponding to 7.6 megatons of TNT or 500 Hiroshima bombs.

Chapter 22

22.1 The centre of gravity of the Sun and Jupiter is about 743,000 km from the centre of the Sun, i.e. somewhat outside the solar surface. From a distance of 10 pc this subtends an angle of 0.0005 arcsecond. Thus the total amount of oscillation is one milliarcsecond.

22.2 Jupiter covers a fraction 0.0106 of the solar disc, corresponding to 0.0115 magnitudes. For the Earth the change is 0.0001 magnitudes.

Further Reading

The following list of references is not intended as a complete bibliography. It gives a number of intermediate level or more advanced works, which can serve as starting points for those who wish to learn about some specific topic in more detail or depth.

General Reference Works

- Cox: *Allen's Astrophysical Quantities*, Springer 2000.
Harwit: *Astrophysical Concepts*, 4th edn., Springer 2006.
Kutner: *Astronomy: A Physical Perspective*, 2nd edn., Cambridge University Press 2003.
Lang: *Astrophysical Formulae*, 3rd edn., Springer 1999.
Maran (ed.): *The Astronomy and Astrophysics Encyclopedia*, Van Nostrand–Cambridge University Press 1992.
Meeus: *Astronomical Algorithms*, Willman-Bell 1991.
Schaifers, Voigt (eds.): *Landolt-Börnstein Astronomy and Astrophysics*, Springer 1981–1982.
Shu: *The Physical Universe*, University Science Books 1982.
Unsöld, Baschek: *The New Cosmos*, 5th edn., Springer 2002.

Chapter 2. Spherical Astronomy

- Green: *Spherical Astronomy*, Cambridge University Press 1985.
Seidelmann (ed.): *Explanatory Supplement to the Astronomical Almanac*, University Science Books 1992.

Smart: *Spherical Astronomy*, Cambridge University Press 1931.

Chapter 3. Observations and Instruments

- Evans: *Observation in Modern Astronomy*, English Universities Press 1968.
Hecht, Zajac: *Optics*, Addison-Wesley 1974.
Howell (ed.): *Astronomical CCD Observing and Reduction Techniques*, ASP Conference Series 23, 1992.
King: *The History of the Telescope*, Charles Griffin & Co. 1955, Dover 1979.
Kitchin: *Astrophysical Techniques*, Hilger 1984; 4th edn., Institute of Physics Publishing 2003.
Léna: *Observational Astrophysics*, Springer 1998.
Roth (ed.): *Compendium of Practical Astronomy* 1–3, Springer 1994.
Rutten, van Venrooij: *Telescope Optics*, Willman-Bell 1988.

Chapters 4 and 5. Photometry and Radiation Mechanisms

- Chandrasekhar: *Radiative Transfer*, Dover 1960.
Emerson: *Interpreting Astronomical Spectra*, Wiley 1996; reprinted 1997.
Rybicki, Lightman: *Radiative Processes in Astrophysics*, Wiley 1979.

Chapter 6. Celestial Mechanics

- Brouwer, Clemence: *Methods of Celestial Mechanics*, Academic Press 1960.
Danby: *Fundamentals of Celestial Mechanics*, MacMillan 1962; 2nd edn. Willman-Bell, 3rd revised and enlarged printing 1992.

- Goldstein: *Classical Mechanics*, Addison-Wesley 1950.
- Roy: *Orbital Motion*, John Wiley & Sons 1978; 3rd edn. Institute of Physics Publishing 1988, reprinted 1991, 1994.
- Valtonen, Karttunen: *The Three-Body Problem*, Cambridge University Press 2006.

Chapter 7. The Solar System

- Atreya, Pollack, Matthews (eds.): *Origin and Evolution of Planetary and Satellite Atmospheres*, 1989;
- Bergstrahl, Miner, Matthews (eds.): *Uranus*, 1991;
- Binzel, Gehrels, Matthews (eds.): *Asteroids II*, 1989;
- Burns, Matthews (ed.): *Satellites*, 1986;
- Cruikshank: *Neptune and Triton*, 1996;
- Gehrels (ed.): *Jupiter*, 1976;
- Gehrels (ed.): *Asteroids*, 1979;
- Gehrels (ed.): *Saturn*, 1984;
- Gehrels (ed.): *Hazards due to Comets and Asteroids*, 1994;
- Greenberg, Brahic (eds.): *Planetary Rings*, 1984;
- Hunten, Colin, Donahue, Moroz (eds.): *Venus*, 1983;
- Kieffer, Jakosky, Snyder, Matthews (eds.): *Mars*, 1992;
- Lewis, Matthews, Guerreri (eds.): *Resources of Near-Earth Space*, 1993;
- Morrison (ed.): *Satellites of Jupiter*, 1982;
- Vilas, Chapman, Matthews (eds.): *Mercury*, 1988;
- Wilkening (ed.): *Comets*, 1982;
- The previous items belong to a series of books on planetary astronomy published by Arizona University Press.

- Beatty, Chaikin (eds.): *The New Solar System*, 3rd edn., Sky Publishing 1990.
- Encrenaz et al.: *The Solar System*, 3rd edn., Springer 2004.
- Heiken, Vaniman, French (eds.): *Lunar Sourcebook*, Cambridge University Press 1991.
- Lewis: *Physics and Chemistry of the Solar System*, Academic Press, revised edition 1997.
- Minnaert: *Light and Color in the Outdoors*, Springer 1993.
- Schmadel: *Dictionary of Minor Planet Names*, 5th edn., Springer 2003.

Chapter 9. Stellar Spectra

- Böhm-Vitense: *Introduction to Stellar Astrophysics*, Vols. 1–3, Cambridge University Press 1989–1992.
- Gray: *Lectures on Spectral-Line Analysis: F, G, and K stars*, Arva, Ontario 1988.
- Gray: *The Observation and Analysis of Stellar Photospheres*, 2nd edn., Cambridge University Press 1992.
- Mihalas: *Stellar Atmospheres*, Freeman 1978.
- Novotny: *Introduction to Stellar Atmospheres and Interiors*, Oxford University Press 1973.

Chapter 10. Binary Stars

- Aitken: *The Binary Stars*, Dover 1935, 1964.
- Heinz: *Double Stars*, Reidel 1978.
- Hilditch: *An Introduction to Close Binary Stars*, Cambridge University Press 2001.
- Sahade, Wood: *Interacting Binary Stars*, Pergamon Press 1978.

Chapters 11 and 12. Stellar Structure and Stellar Evolution

- Bowers, Deeming: *Astrophysics I: Stars*, Jones and Bartlett Publishers 1984.
- Böhm-Vitense: *Introduction to Stellar Astrophysics*, Vols. 1–3, Cambridge University Press 1989–1992.
- Clayton: *Principles of Stellar Evolution and Nucleosynthesis*, McGraw-Hill 1968.
- Eggleton: *Evolutionary Processes in Binary and Multiple Stars*, Cambridge University Press 2006.
- Hansen, Kawaler: *Stellar Interiors, Physical Principles, Structure and Evolution*, Springer 1994.
- Harpaz: *Stellar Evolution*, Peters Wellesley 1994.
- Kippenhahn, Weigert: *Stellar Structure and Evolution*, 2nd edn., Springer 1994.
- Phillips: *The Physics of Stars*, Manchester Phys. Ser. 1994, 2nd edn. 1999.
- Salaris, Cassisi: *Evolution of Stars and Stellar Populations*, Wiley 2005.
- Taylor: *The Stars: Their Structure and Evolution*, Cambridge University Press 1994.

Chapter 13. The Sun

- Golub, Pasachoff: *The Solar Corona*, Cambridge University Press 1997.

Priest: *Solar Magnetohydrodynamics*, Reidel 1982.
 Stix: *The Sun*, 2nd edn., Springer 2002.
 Taylor: *The Sun as a Star*, Cambridge University Press 1997.

Chapter 14. Variable Stars

Cox: *Theory of Stellar Pulsations*, Princeton University Press 1980.
 Glasby: *Variable Stars*, Constable 1968.
 Warner: *Cataclysmic Variable Stars*, Cambridge University Press 1995; paperback edn. 2003.

Chapter 15. Compact Stars

Camenzind: *Compact Objects in Astrophysics*, Springer 2007.
 Chandrasekhar: *The Mathematical Theory of Black Holes*, Oxford University Press 1983.
 Frank, King, Raine: *Accretion Power in Astrophysics*, Cambridge Astrophysics Series 21, 2nd edn., Cambridge University Press 1992.
 Glendenning: *Compact Stars*, 2nd edn., Springer 2000.
 Lewin, van Paradijs, van den Heuvel (eds.), *X-ray Binaries*, Cambridge University Press 1995.
 Manchester, Taylor: *Pulsars*, Freeman 1977.
 Smith: *Pulsars*, Cambridge University Press 1977.
 Poutanen, Svensson: *High Energy Processes in Accreting Black Holes*, Astronomical Society of Pacific Conference series, Vol. 161, 1999.
 Shapiro, Teukolsky: *Black Holes, White Dwarfs and Neutron Stars*, Wiley 1983.

Chapter 16. Interstellar Matter

Dopita, Sutherland: *Astrophysics of the Diffuse Universe*, Springer 2003.
 Dyson, Williams: *The Physics of the Interstellar Medium*, Manchester University Press 1980.
 Longair: *High Energy Astrophysics*, Cambridge University Press, 2nd edn. Vols. 1–2 1992, 1994.
 Spitzer: *Physical Processes in the Interstellar Medium*, Wiley 1978.
 Tielens: *The Physics and Chemistry of the Interstellar Medium*, Cambridge University Press 2005.

Chapter 17. Stellar Clusters

Ashman, Zepf: *Globular Cluster Systems*, Cambridge University Press 1998.

Hanes, Madore (eds.): *Globular Clusters*, Cambridge University Press 1980.
 Spitzer: *Dynamical Evolution of Globular Clusters*, Princeton University Press 1987.

Chapter 18. The Milky Way

Binney, Merrifield: *Galactic Astronomy*, Princeton University Press 1998.
 Bok, Bok: *Milky Way*, 5th edn., Harvard University Press 1982.
 Gilmore, King, van der Kruit: *The Milky Way as a Galaxy*, University Science Books 1990.
 Mihalas, Binney: *Galactic Astronomy*, Freeman 1981.
 Scheffler, Elsässer: *Physics of the Galaxy and the Interstellar Matter*, Springer 1988.

Chapter 19. Galaxies

van den Bergh: *Galaxy Morphology and Classification*, Cambridge University Press 1998.
 Binney, Tremaine: *Galactic Dynamics*, Princeton University Press 1987.
 Combes, Boissé, Mazure, Blanchard: *Galaxies and Cosmology*, Springer 1995; 2nd edn. 2001.
 Frank, King, Raine: *Accretion Power in Astrophysics*, Cambridge Astrophysics Series 21, 2nd edn., Cambridge University Press 1992.
 Salaris, Cassisi: *Evolution of Stars and Stellar Populations*, Wiley 2005.
 Krolik: *Active Galactic Nuclei*, Princeton University Press 1999.
 Sandage: *The Hubble Atlas of Galaxies*, Carnegie Institution 1961.
 Sparke, Gallagher: *Galaxies in the Universe*, Cambridge University Press 2000.

Chapter 20. Cosmology

Dodelson: *Modern Cosmology*, Academic Press 2003.
 Harrison: *Cosmology*, Cambridge University Press 1981.
 Kolb, Turner: *The Early Universe*, Perseus Books 1993.
 Peacock: *Cosmological Physics*, Cambridge University Press 1999.
 Peebles: *Physical Cosmology*, Princeton University Press 1971.

- Peebles: *Principles of Physical Cosmology*, Princeton University Press 1993.
- Raine: *The Isotropic Universe*, Hilger 1981.
- Roos: *Introduction to Cosmology*, Wiley 1994, 2nd edn. 1997.
- Weinberg: *Gravitation and Cosmology*, Wiley 1972.

Chapter 21. Astrobiology

- Cassen, Guillot, Quirrenbach: *Extrasolar Planets*, Springer 2006.
- Gargaud, Barbier, Martin, Reisse (eds.): *Lectures in Astrobiology*, Springer Vol. I (2006), Vol II. (2007)
- Gilmour, Sephton: *An Introduction to Astrobiology*, The Open University, Cambridge University Press 2004.
- Rettberg, Horneck (eds.): *Complete Course in Astrobiology*, Wiley–VCH 2007.
- Webb: *Where is Everybody?* Copernicus Books in association with Praxis Publishing 2002.

Physics

- Feynman, Leighton, Sands: *The Feynman Lectures on Physics I–III*, Addison-Wesley 1963.
- Shu: *The Physics of Astrophysics I–II*, University Science Books 1991
- Taylor, Wheeler: *Spacetime Physics*, Freeman 1963.
- Misner, Thorne, Wheeler: *Gravitation*, Freeman 1970.

Maps and Catalogues

- Burnham: *Burnham's Celestial Handbook I, II, III*, Dover 1966, 2nd edn. 1978.
- de Vaucouleurs et al.: *Reference Catalogue of Bright Galaxies*, University of Texas Press 1964, 2nd catalogue 1976.
- Hirshfeld, Sinnott: *Sky Catalogue 2000.0*, Sky Publishing 1985.
- Hoffleit: *Bright Star Catalogue*, Yale University Observatory 1982.
- Kholopov (ed.): *Obshij katalog peremennyh zvezd*, 4th edn., Nauka 1985.
- Luginbuhl, Skiff: *Observing Handbook and Catalogue of Deep-Sky Objects*, Cambridge University Press 1989.
- Ridpath: *Norton's 2000.0*, Longman 1989.

- Rükl: *Atlas of the Moon*, Hamlyn 1991.
- Ruprecht, Baláz, White: *Catalogue of Star Clusters and Associations*, Akadémiai Kiadó (Budapest) 1981.
- Tirion: *Sky Atlas 2000.0*, Sky Publishing 1981.
- Tirion, Rappaport, Lovi: *Uranometria 2000.0*, Willman-Bell 1987.
- Greeley, Batson: *The NASA Atlas of the Solar System*, Cambridge University Press 1997.

Yearbooks

- The Astronomical Almanac*, Her Majesty's Stationery Office.

Useful Internet addresses

Apparent Places of Fundamental Stars
<http://www.ari.uni-heidelberg.de/ariapfs/>

The Astronomical Almanac Online
<http://asa.usno.navy.mil/> and
<http://asa.nao.rl.ac.uk/>

ESO
<http://www.eso.org/>

ESA
<http://www.esa.int/>

H.M. Nautical Almanac Office
<http://www.nao.rl.ac.uk/>

Hubble Space Telescope
<http://oposite.stsci.edu/>

IAU
<http://www.iau.org>

IAU Central Bureau for Astronomical Telegrams
<http://cfa-www.harvard.edu/cfa/ps/cbat.html>

Jet Propulsion Laboratory
<http://www.jpl.nasa.gov/>

Mauna Kea Observatories
<http://www.ifa.hawaii.edu/mko/maunakea.htm>

Nasa
<http://www.nasa.gov/>

Nasa Planetary Photojournal
<http://photojournal.jpl.nasa.gov/>

National Optical Astronomy Observatory
<http://www.noao.edu/>

National Radio Astronomy Observatory
<http://www.nrao.edu/>

National Space Science Data Center
<http://nssdc.gsfc.nasa.gov/>

U.S. Naval Observatory
<http://aa.usno.navy.mil/>

Photograph Credits

We are grateful to the following institutions who kindly gave us permission to use illustrations (abbreviated references in the figure captions)

Anglo-Australian Observatory, photograph by David R. Malin

Arecibo Observatory, National Astronomy and Ionosphere Center, Cornell University

Arp, Halton C., Mount Wilson and Las Campanas Observatories (colour representation of plate by Jean Lorre)

Big Bear Solar Observatory, California Institute of Technology

Catalina Observatory, Lunar and Planetary Laboratory

CSIRO (Commonwealth Scientific and Industrial Research Organisation), Division of Radiophysics, Sydney, Australia

ESA (copyright Max-Planck-Institut für Astronomie, Lindau, Harz, FRG)

European Southern Observatory (ESO)

Helsinki University Observatory

High Altitude Observatory, National Center for Atmospheric Research, National Science Foundation

Karl-Schwarzschild-Observatory Tautenburg of the Central Institute of Astrophysics of the Academy of Sciences of the GDR (Archives)

Lick Observatory, University of California at Santa Cruz

Lowell Observatory

Lund Observatory

Mauna Kea Observatory Hawaii, The James Clerk Maxwell Telescope

Mount Wilson and Las Campanas Observatories, Carnegie Institution of Washington

NASA (National Aeronautics and Space Administration)

National Optical Astronomy Observatories, operated by the Association of Universities for Research in Astronomy (AURA), Inc., under contract to the National Science Foundation

NRL-ROG (Space Research Laboratory, University of Groningen)

Palomar Observatory, California Institute of Technology

Yerkes Observatory, University of Chicago

Colour Supplement

Chapter 2. Spherical Astronomy

Plate 1 A long exposure shows the apparent motion of the celestial sphere. Pekka Parviainen has caught in his image the star trails, a trail of the crescent moon, and a trail of the rising Sun, taken through a 24 stop ND-filter. (Photo www.polarimage.fi)

Plate 2 Atmospheric refraction has distorted and discoloured the image of the setting Sun. A rare *green* segment can be seen above the Sun. (Photo Pekka Parviainen)

Chapter 3. Observations and Instruments

Plate 3 Astronomical observatories are built in high, dry places with a clear sky. The picture shows the summit of Mauna Kea in Hawaii with the twin domes of the world's largest telescopes, the William M. Keck Telescopes. Each of them has a mosaic mirror with a diameter of 10 meters. (Photo W.M. Keck Observatory)

Plate 4 The next generation space telescope, James Webb, is under construction in the U.S. It will be launched about 2013. It will have a 6.5 meter mirror made of 18 segments. The telescope will be situated at the L2 point of the Sun-Earth system, about 1.5 million kilometers from the Earth. (Illustration NASA)

Chapter 7. The Solar System

Plate 5 The transits of Mercury over the disk of the Sun were widely observed in 1999 and 2003. The planet is nearing the limb of the Sun at the end of the latter transit. The photograph was taken with the 1 meter Swedish solar telescope on La Palma. (Photo Royal Swedish Academy of Sciences)

Plate 6 Eclipses are astronomical phenomena that tend to attract the attention of the general public. A composite made from five different exposures shows the extent of the Earth's shadow and the movement of the Moon from lower right to upper left. The eclipse occurred in October 2004. (Photo Pekka Parviainen)

Plate 7 The first and so far only colour pictures from the surface of Venus were obtained in 1982. Panoramic views from Venera 13 (*above*) and Venera 14 (*below*). At the edges, the camera is looking at the horizon and in the centre, at the ground around the base of the space probe. Parts of the base, the protective covers of the camera and a colour map can be seen. (Photo Soviet Academy of Science)

Plate 8 One of the largest impact craters on the surface of the Earth, Lake Manicouagan in northern Canada. The ancient impact structure

is about 70 kilometers in diameter. This natural-colour image of the crater was acquired by Nasa's Terra satellite in June 2001. The impact is thought to have occurred about 212 million years ago. (Photo NASA/JPL/MISR)

Plate 9 The last people to have walked on the Moon were geologist Harrison Schmitt (in the picture) and Eugene Cernan, who took the photo. Schmitt is studying a big boulder at the Taurus-Littrow landing site of Apollo 17 spacecraft in December 1972, with the lunar rover in the foreground. (Photo NASA)

Plate 10 The "Burns Cliff" in the "Endurance Crater" on Mars. The photo was taken by the Opportunity Mars rover in November 2004. Opportunity and its twin Spirit have studied the surface of Mars in unprecedented detail since their landing in early 2002. The mosaic spans about 180 degrees from side to side. (Photo NASA)

Plate 11 Colour-enhanced picture of the surface of Europa. The surface is criss-crossed by lineaments, and the colour differences show that they are younger than the surrounding areas. The surface is mainly water ice. (Photo Galileo/NASA)

Plate 12 A true colour mosaic of Jupiter was taken by Cassini spacecraft in 2000, when Cassini made its closest approach to Jupiter on its way to Saturn. The clouds of the giant planet move in *reddish-brown* and *white* bands, with the Great Red Spot and several smaller white ovals in view. The smallest visible features are about 60 kilometers across. (Photo NASA)

Plate 13 A simulated image of the rings of Saturn, based on radio occultation measurements by Cassini spacecraft in May 2005. *Green* and *blue shades* indicate regions where the particles are smaller than 5 centimeters and 1 centimeter across. *In purple regions*, the particles are mostly larger than 5 centimeters. *The white area* is the thickest part of B-ring, where radio signals were blocked. (Photo NASA)

Plate 14 The first view from the surface of Titan, the largest moon of Saturn, was obtained in

January 2005, when the European Huygens probe made a soft landing. The boulders are made of water ice, and their size is mostly 10–20 centimeters. The surface itself is methane-filled mud and sand. (Photo ESA/NASA)

Plate 15 Details of the icy surface of Triton, the largest satellite of Neptune. The plain in the center is called Tuonela. The albedo of Triton is very similar to that of Pluto, and thus this kind of feature may be seen by the New Horizons spacecraft in July 2015 when it flies by Pluto. (Photo Voyager 2/NASA)

Plate 16 Surface and jets of Comet Wild 2. The Stardust spacecraft flew by the head of comet Wild 2 in January 2004. The best black-and-white photo of the comet's surface has been merged with a long-exposure photo showing the faint dust jets emanating from the surface. The diameter of the comet's head is about 5 km. (Photo NASA)

Chapter 9. Stellar Spectra

Plate 17 Spectra of stars of different spectral classes. The wavelength range is in the visual part, from 400 to 700 nanometers. The O stars at the top are hottest, and the M stars at the bottom, the coolest. (Photo NOAO/AURA/NSF)

Chapter 13. The Sun

Plate 18 Three-colour composite of the solar surface, photographed by the Extreme Ultraviolet Imaging Telescope of SOHO satellite in May 1998. Three ultraviolet images taken at different wavelengths (17.1 nm, 19.5 nm and 28.4 nm) were coded *red*, *yellow* and *blue* and merged into one picture. (Photo SOHO/EIT/NASA/Goddard Space Flight Center)

Chapter 16. The Interstellar Medium

Plate 19 One of the most detailed visual-wavelength pictures of the Orion Nebula was published in 2006. It was taken by the Hubble Space Telescope, which used 105 orbits to photograph the nebula. The original picture has 1000 million

pixels. Thousands of previously unknown stars were found in the picture, and also many new brown dwarfs. Orion Nebula is a large dust and gas region, giving rise to new stars. The width of the picture is about half a degree (the same as full Moon). (Photo NASA/ESA/M. Robberto)

Plate 20 Orion Nebula photographed by the Spitzer Telescope in infrared light in 2006. While the Hubble picture (Plate 19) shows bright gas, with dark dust lanes, Spitzer sees the dust as bright regions. The wavelength of 8 micrometers has been coded *red*, and 5.8 microns, *orange*. They represent dust heated by the stars. Light of 4.5 microns (*green*) shows hot gas and dust, and light of 3.6 microns (*blue*) is from starlight. (Photo NASA)

Plate 21 The dark cloud Barnard 68 in the constellation of Ophiuchus. B68 is a dense molecular cloud, which is just starting to contract and later form new stars. The diameter of the cloud is 7 light-months (0.2 parsecs), and its distance is about 500 light-years. It was photographed by the Antu (VLT 1) Telescope in 1999. (Photo ESO)

Plate 22 The star V838 Monocerotis brightened suddenly in early 2002. The outburst lasted several weeks and sent a sphere of light spreading into the surrounding interstellar medium. The light-echo has illuminated the dust clouds around the star and unveiled never-before-seen patterns. This photo was taken by the Hubble Space Telescope at the end of 2004, nearly three years after the outburst. (Photo NASA/ESA)

Plate 23 The Rosette Nebula in Monoceros belongs to the same extensive dust and gas complex as the Orion Nebula. Within the Rosette Nebula there is a young star cluster called NGC 2244, which gives energy to the cloud. The ultraviolet light from the stars makes the hydrogen glow *red*, and the oxygen, *greenish*. The picture is about 1 degree wide, and it was taken by the MegaPrime Camera at the Canada-France-Hawaii Telescope on Mauna Kea in 2003. (Photo CFHT/J.-C. Cuillandre)

Plate 24 The Helix Nebula is one of the nearest planetary nebulae and also one of the apparently widest. Its diameter is nearly one degree (two full Moons side by side), but its surface brightness is so low it cannot be seen by the naked eye. It is located in the constellation of Aquarius. This composite photograph combines an ultra-sharp mosaic by the Hubble Space Telescope with a wide view by the WIYN 0.9 meter Telescope on Kitt Peak. (Photo NASA/NOAO/ESA)

Plate 25 Radio image of the Cassiopeia A supernova remnant. Cas A is one of the brightest radio sources in the sky. This image was made at three different frequencies: 1.4 GHz, 5.0 GHz and 8.4 GHz, with the Very Large Array radio telescope in New Mexico. The radiation comes mainly from relativistic electrons. (Image NRAO/AUI)

Plate 26 The remnant of the Tycho supernova in X-rays. The supernova exploded in 1572 and was observed by Tycho Brahe. During the last years, the Chandra X-Ray Observatory has made observations of this object. The image shows an expanding bubble of extremely hot gas (in *green* and *red*) inside a shell of rapidly moving electrons (*blue*). (Image NASA/Chandra)

Chapter 19. Galaxies

Plate 27 Part of the Andromeda Galaxy M31, photographed with the Japanese Subaru 8.4 m telescope on Mauna Kea, Hawaii. The nucleus of the galaxy is outside the picture on the upper left. The stars with the vertical lines are foreground stars in our Milky Way. M31 is a sister galaxy to our Milky Way, with similar star clusters, dust and gas clouds and spiral arms. (Photo Subaru Telescope, NOAJ)

Plate 28 The most detailed infrared view of the Andromeda Galaxy so far was obtained by the Spitzer Telescope in 2004. While the Subaru picture (Plate 27) shows the starlight from the galaxy, the Spitzer picture records the infrared radiation from the dust in the spiral arms at the wavelength of 24 micrometers. The spiral arms

with their warm dust and star-forming clouds can be seen much more clearly than at visual wavelengths. (Photo Spitzer/NASA)

Plate 29 The central part of the active galaxy Centaurus A, NGC 5128, seen by the Hubble Space Telescope. The whole galaxy is crossed by a thick dust lane with intricate wisps of warm dust and gas. At the centre of the galaxy, behind the dust shroud, lies a black hole, which sends powerful jets in two directions. The jets can be seen only in X-rays or at radio wavelengths. For a wider view of Centaurus A see the next Plate. (Photo NASA)

Plate 30 New pictures of the outer parts of Centaurus A have revealed several ring-like features, which contain faint blue stars. The rings may have been formed in interactions with other, smaller galaxies, which have collided with Centaurus A and lost part of their stars to this giant galaxy. (Photo Cerro Tololo International Observatory, NOAO/NSF)

Plate 31 A well-known galaxy-quasar pair: in the center the galaxy NGC 4319, and on the upper right, the quasar Markarian 215. Although the pair seems to belong together, in reality this is a case of chance alignment. The distance to the quasar is 14 times larger than the distance to the galaxy. (Photo Hubble Space Telescope/NASA)

Plate 32 A small group of interacting galaxies called Seyfert's Sextet. As the name implies,

there seem to be six galaxies engaging in a cosmic dance. The number of interacting galaxies is however only four. The small well-defined spiral galaxy to the right of the center, lies nearly five times further away than the others. And the right-most “galaxy” isn’t actually a galaxy at all, but a long “tidal tail” of stars torn off from the other galaxies. (Photo Hubble Space Telescope/NASA)

Chapter 20. Cosmology

Plate 33 The Hubble Space Telescope has photographed several deep sky fields to get representative samples of the Universe. This photograph is the Hubble Deep Field South in the constellation Tucana, and it contains thousands of never-before seen galaxies up to the distance of 12 Ga. *The small red dot* in the middle of the photograph is a distant quasar, whose light was studied with the Space Telescope Imaging Spectrograph, to map invisible clouds of hydrogen gas between us and the quasar. (Photo NASA)

Plate 34 Temperature fluctuations in the cosmic microwave background. The observations were made with the Wilkinson Microwave Anisotropy Probe (WMAP) in 2002. From the scale dependence of the fluctuations we can estimate the average density of the Universe (very near the critical density), and the age of the Universe (about 13.7 Ga). (Photo WMAP/NASA)

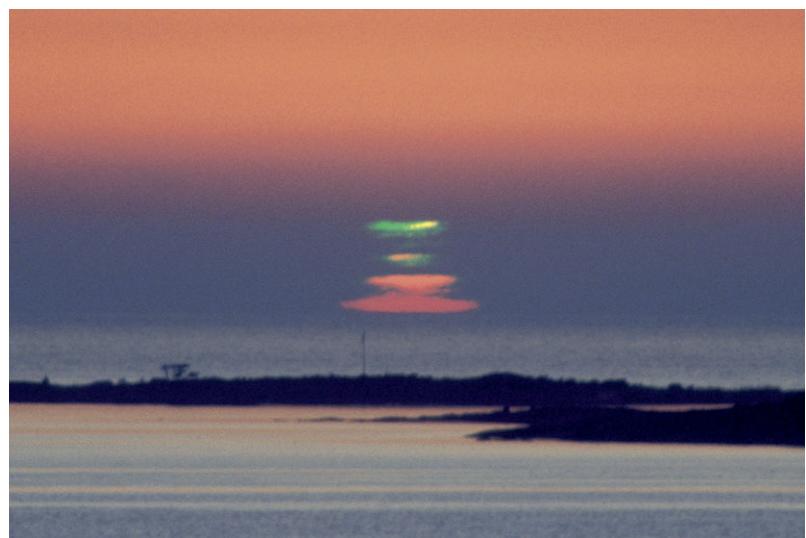
Plate 1**Plate 2**

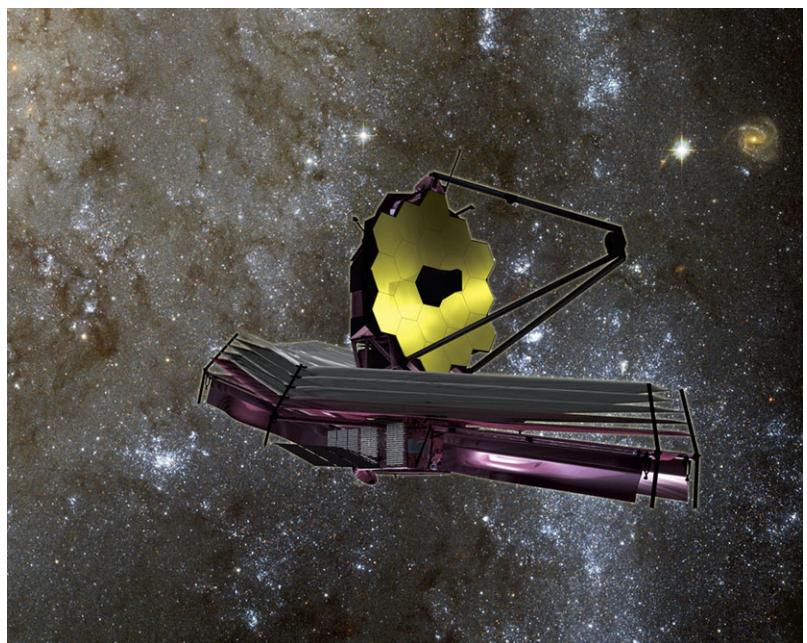
Plate 3**Plate 4**

Plate 5

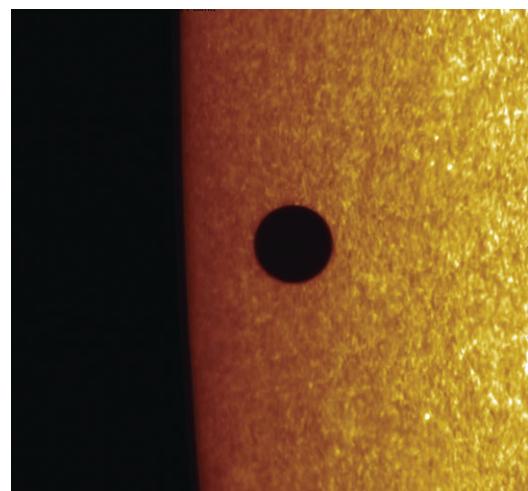
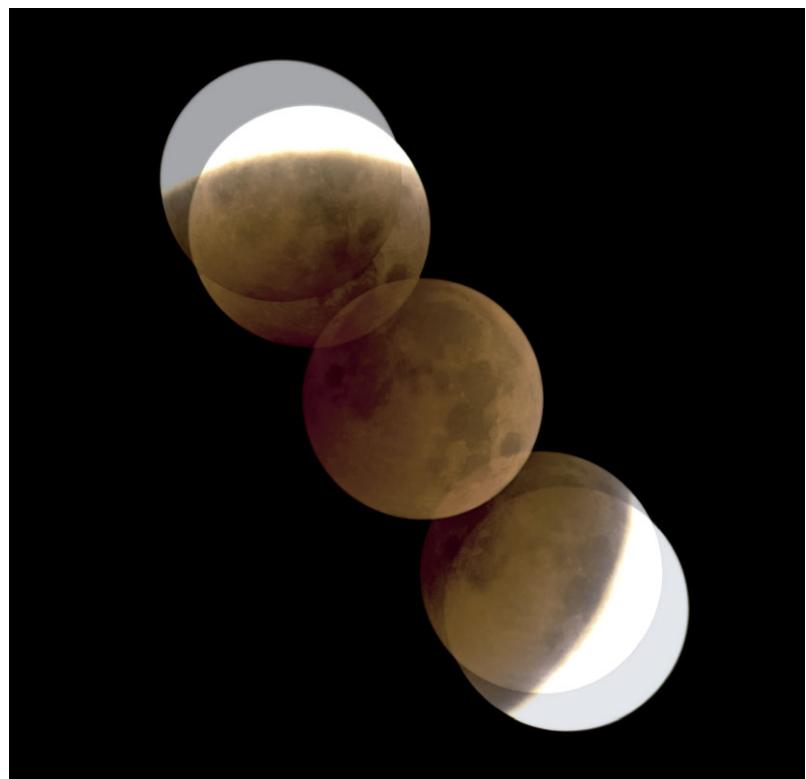


Plate 6



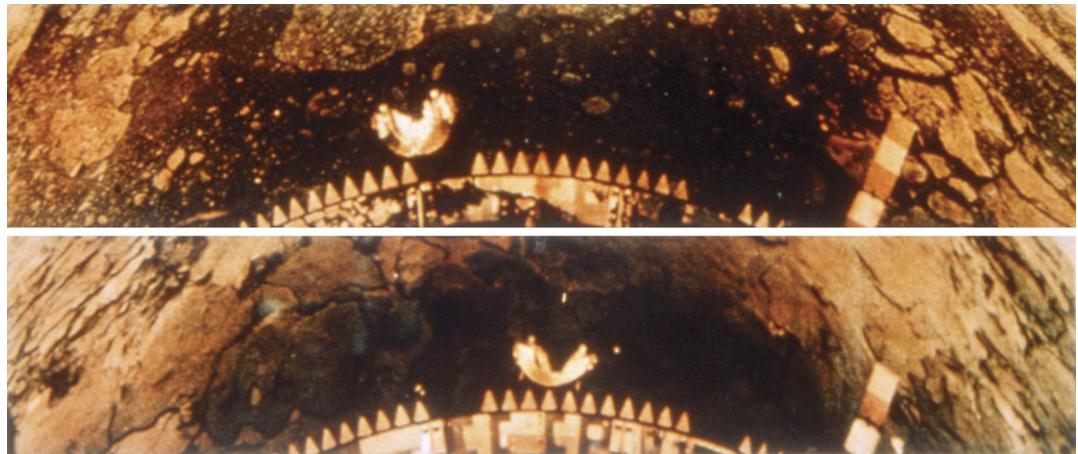


Plate 7

Plate 8



Plate 9

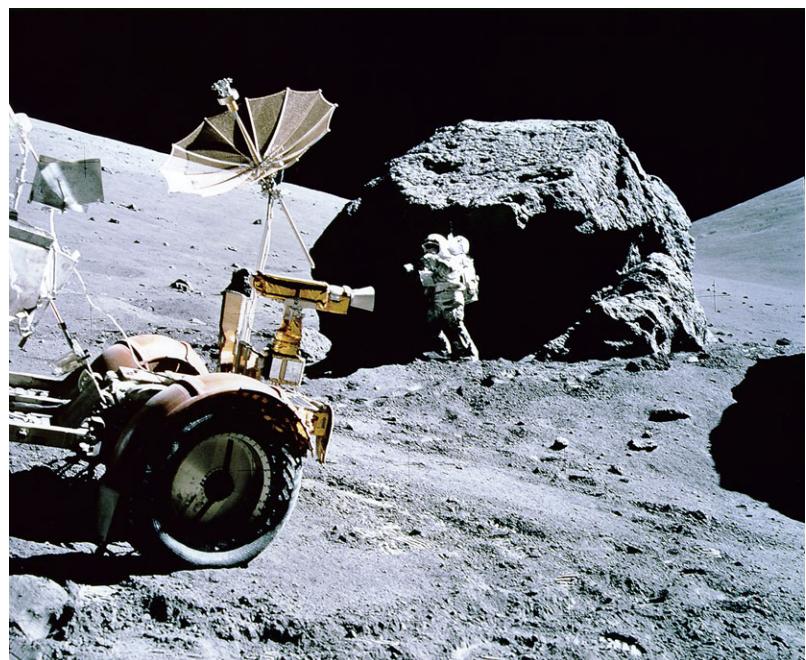
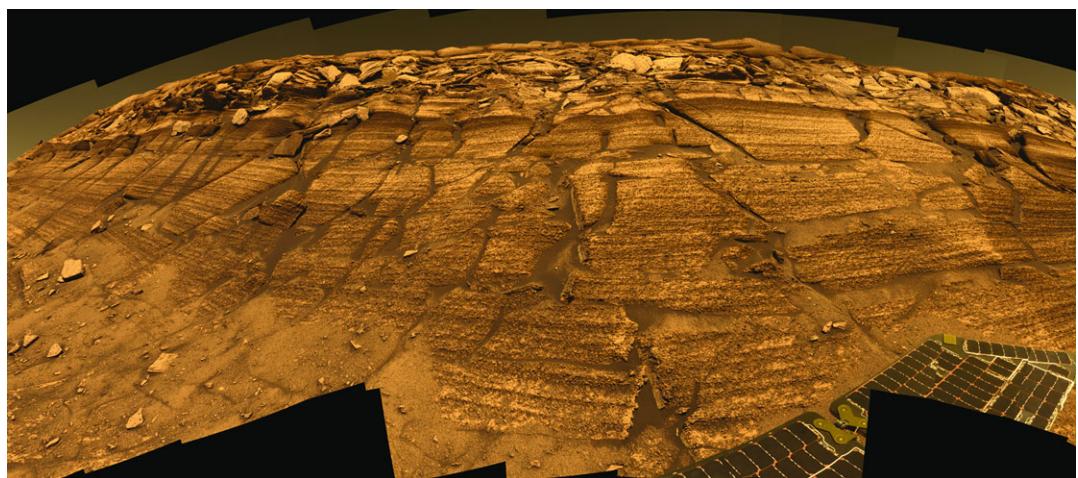


Plate 10



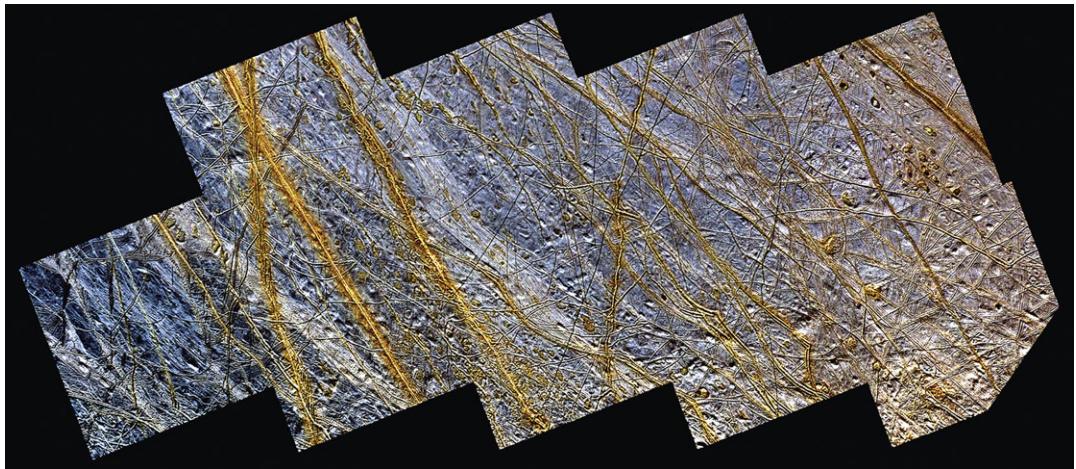


Plate 11



Plate 12

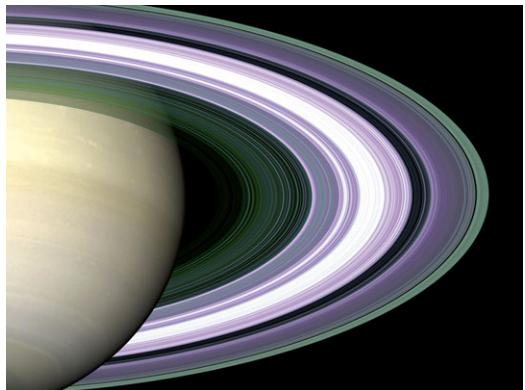


Plate 13

Plate 14

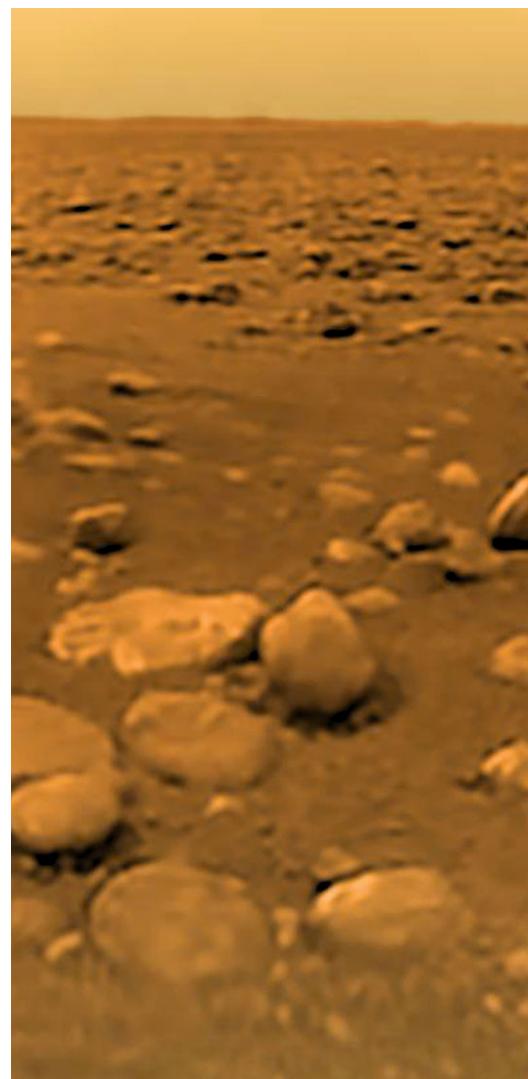
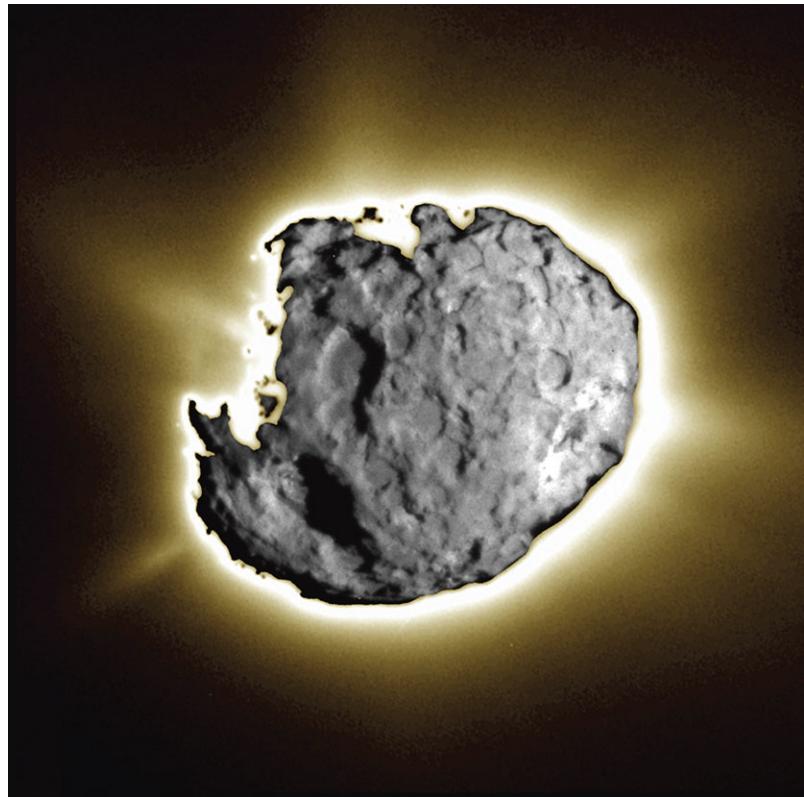
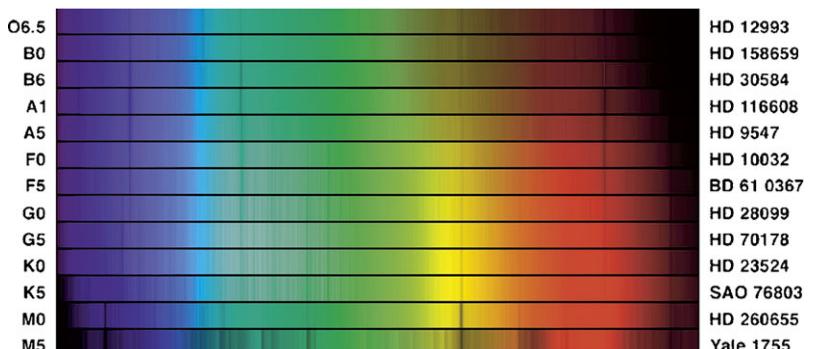


Plate 15



Plate 16**Plate 17**

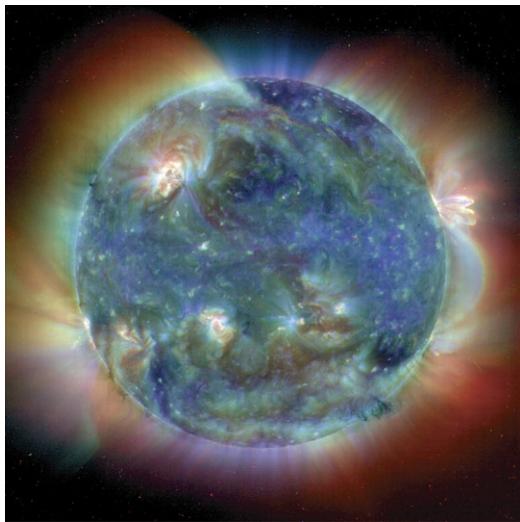
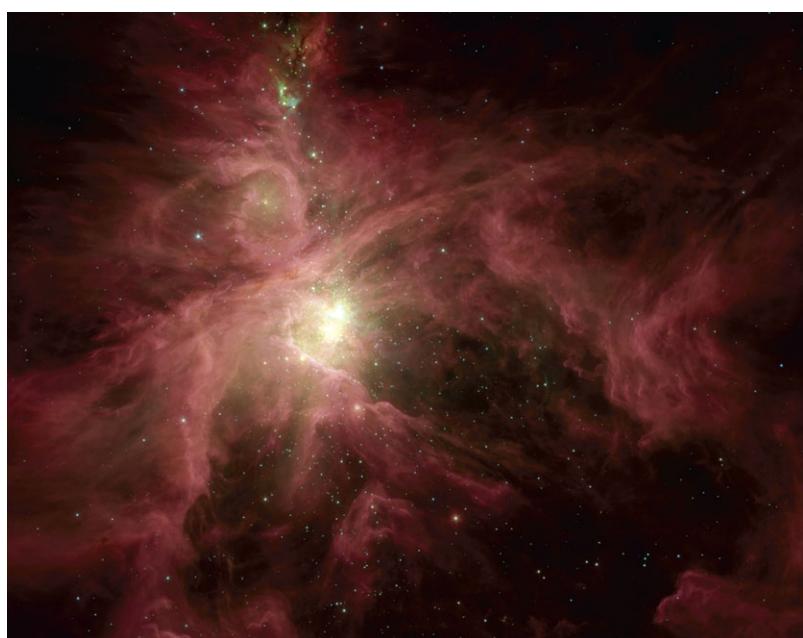


Plate 18



Plate 19

Plate 20



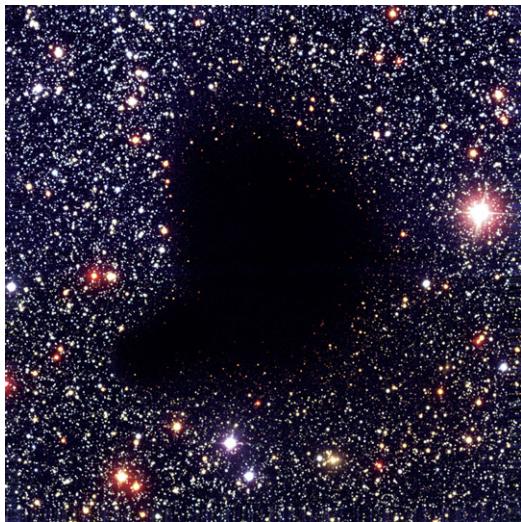


Plate 21



Plate 22

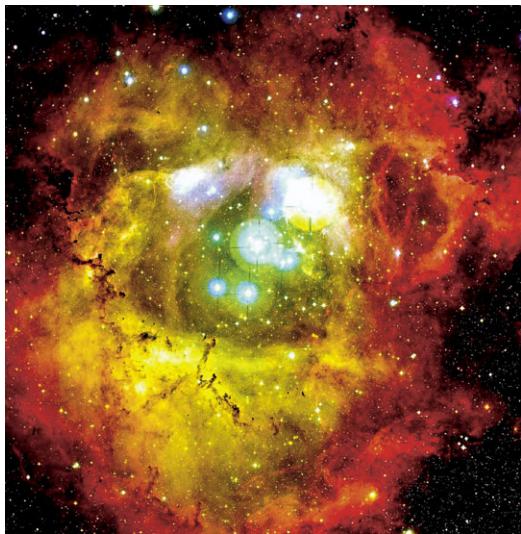


Plate 23

Plate 24

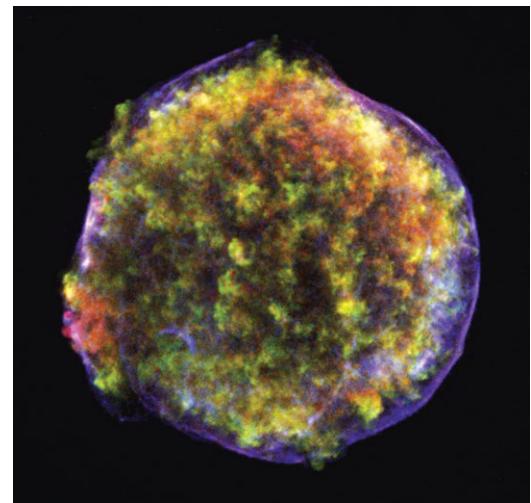
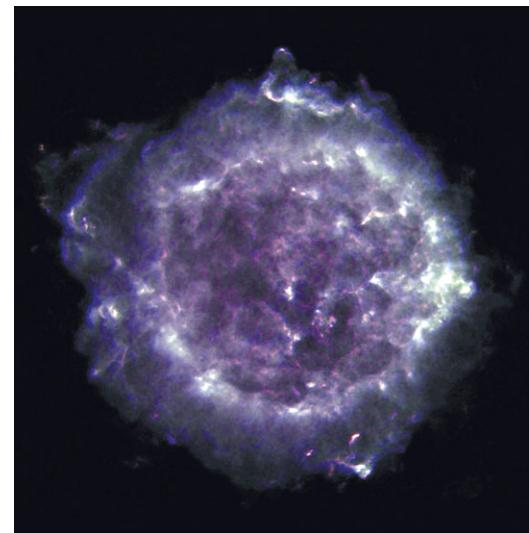
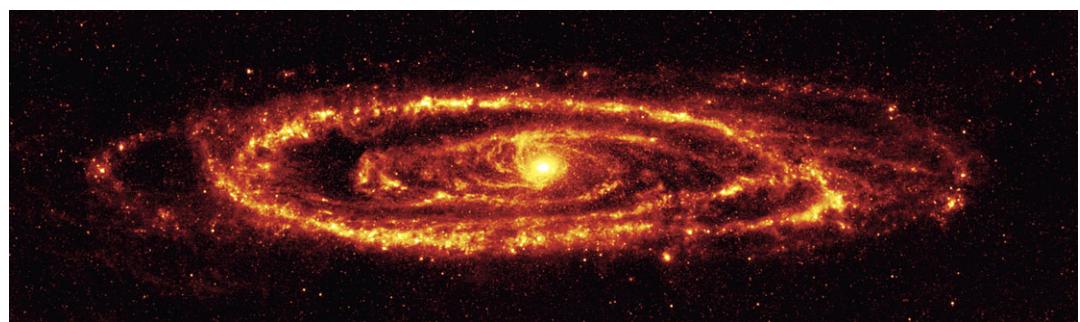
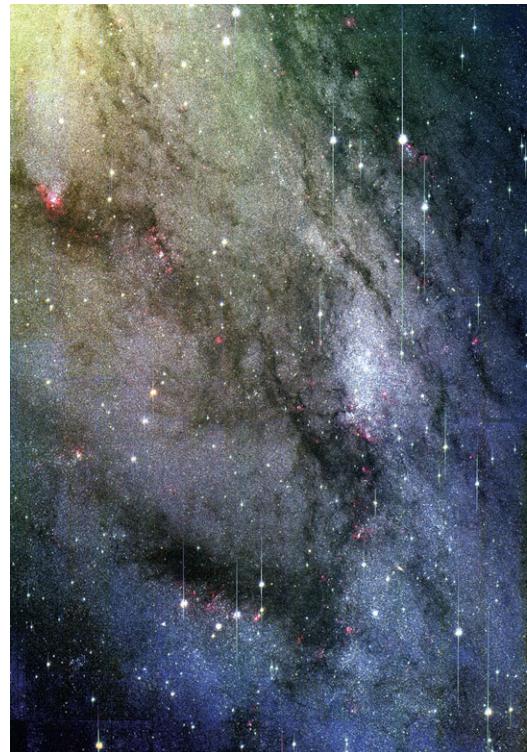


Plate 25

Plate 26

Plate 27**Plate 28**

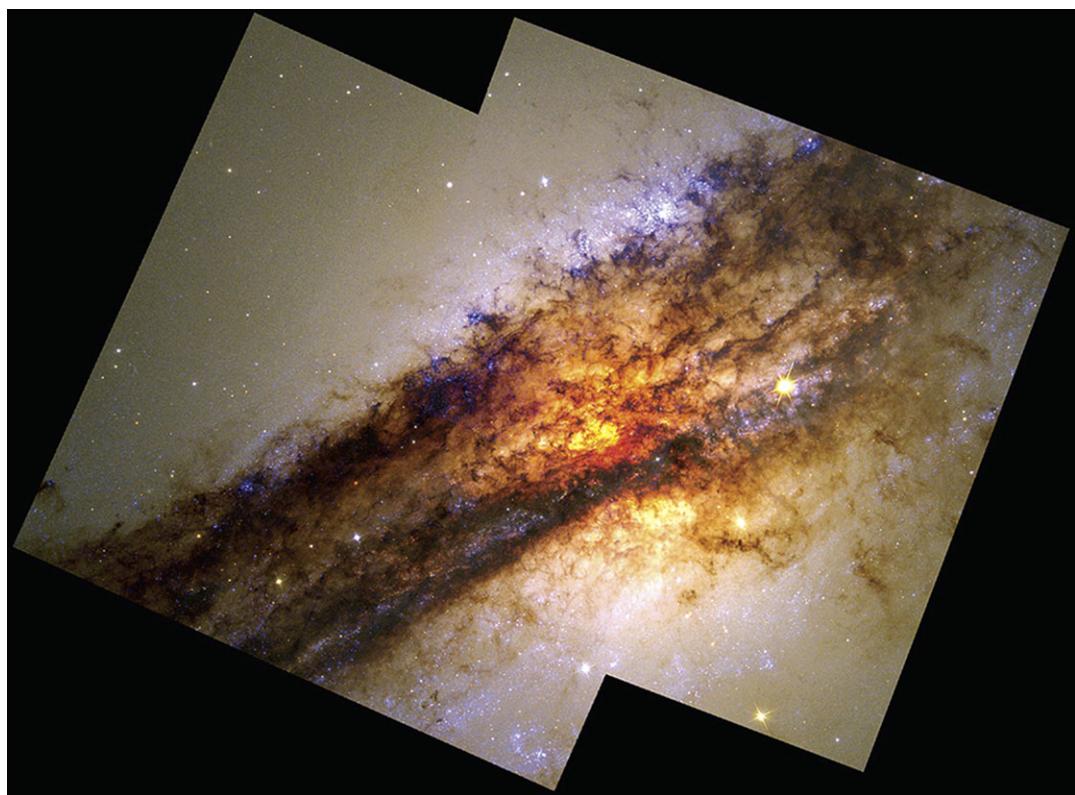


Plate 29

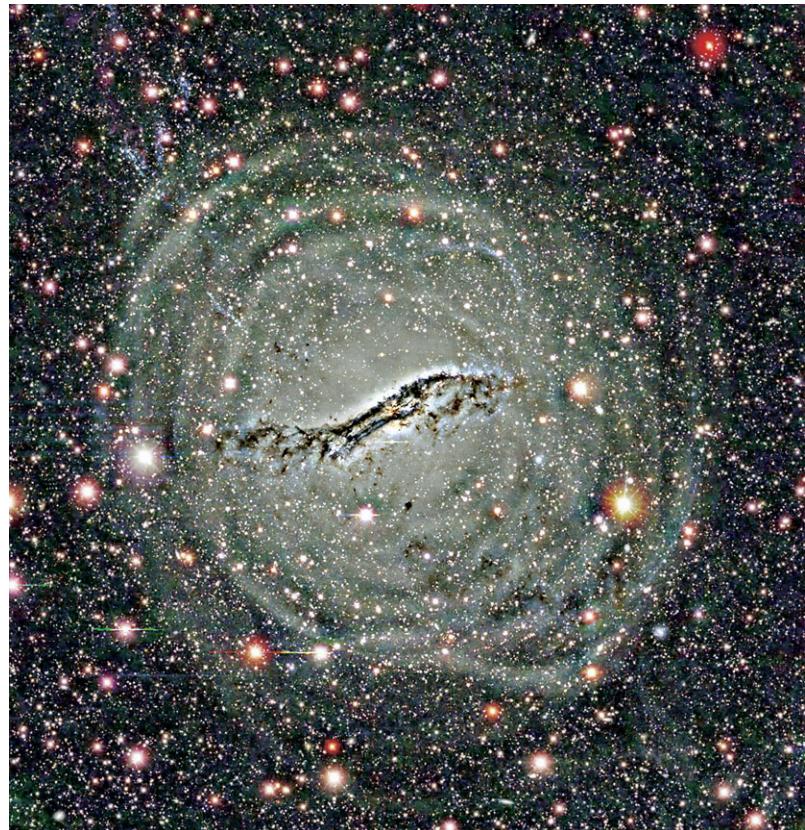
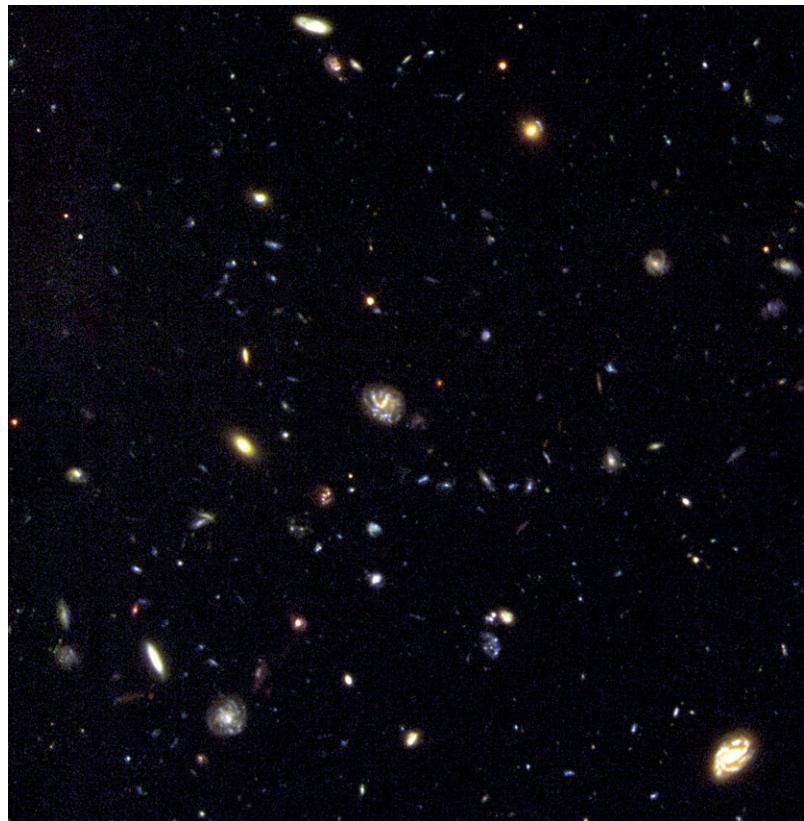
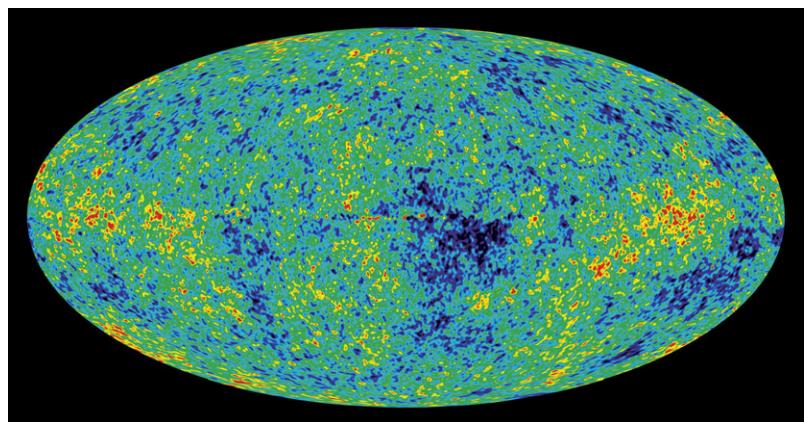
Plate 30**Plate 31**



Plate 32

Plate 33**Plate 34**

Index

Symbols

3C129, 411
3C273, 411, 413
3C295, 423
61 Cygni, 28

A

Aberration, 25
annual, 43
chromatic, *see* Chromatic aberration
diurnal, 25
Aberration constant, 25
Absolute magnitude, 163
Absorption, 53, 103, 104, 328
atmospheric, 51
interstellar, 97
Absorption efficiency factor, 329
Absorption line, 104, 227
Abundance
helium, 426
Accretion, 307, 321
Achromatic lens, 56
Active galactic nucleus, 409
type 2, 414
Active optics, 62
Adams, John C., 211
Adaptive optics, 62
Adiabatic equation of state, 251
Adiabatic exponent, 251
AGN, *see* Active galactic nucleus
Air mass, 99
Aiglow, 194
Albedo, 160
Bond, 160, 163, 167
geometric, 161, 163
of interstellar grains, 333
Aldebaran, 230, 359
Alfvén, Hannes, 352
Algol, 273
Algol paradox, 273
Algol stars, 244, 245
ALH 84001, 452, 453

Almagest, 31
Alpha reactions, 257
Altitude, 17, 26
Altman, Sidney, 450
Am star, 234
Ambartsumyan, Viktor, 359
Amino acid, 446, 450
Ammonia, 347
Andromeda Galaxy, 10
Angle, 463
Ångström unit, 229
Angular diameter-redshift test, 431
Angular momentum, 129
of a contracting cloud, 280
of a planet, 124
of electron, 105, 108, 109
Angular momentum quantum number, 108
Angular velocity
of orbital motion, 128
Annual parallax, 24
Anomalistic month, 148
Anomalistic year, 146
Anomalous X-ray pulsars, 318
Anomaly, 126
eccentric, 132
mean, 133
true, 133
Antapex, 368, 370
Antares, 333, 334
Antenna, 74
dipole, 75
parabolic, 75
Antimatter, 433
Ap star, 234
Aperture, 54
Aperture ratio, 54
Aperture synthesis, 78
Apex, 368, 370
Aphelion, 129
Aphrodite Terra, 186
Apollo 11, 192
Apollo-Amor asteroids, 217

- Apparent motions, 143
 Apparent place, 27, 43
 Aquila, 369
 Arctic Circle, 146
 Arcturus, 339
 Argelander, F.W.A., 32, 34
 Argument of perihelion, 127, 128
 Ariel, 210
 Aries, 17
 Arrhenius, Svante, 453
 Association, 349, 359
 OB, 360, 373
 T Tauri, 360
 Astenosphere, 152
 Asteroids, 142, 214, 486
 collision, 218
 composition, 219
 distances, 215
 formation, 215
 magnitude, 164
 mass, 215
 number of, 215
 orbits, 216
 rotation, 215
 shape, 151
 size, 215
 Asthenosphere, 189
 Astrograph, 57
 Astrometry
 absolute, 27
 relative, 27
 Astronomical Almanac, 35, 39
 Astronomical unit, 24, 28, 130, 141, 151, 478
 Asymptotic branch, 236, 270, 364
 Atlas, 207
 Atmosphere
 planetary, 118
 solar, 286
 stellar, 118, 228, 236, 238
 Atom, 103
 nucleus, 254
 Atomic mass unit (amu), 478
 Atomic number, 254
 Atomic weight, 254
 Au, *see* Astronomical unit
 Aurora, 158, 192, 194, 295, 297
 Autumnal equinox, 21, 146
 Azimuth, 17
- B**
 β Cephei star, 299, 300–303
 β decay, 279
 β Lyrae star, 245
 Baade, Walter, 302
 Babcock, Horace W., 292
 Background radiation, 425
 Balmer, Johann, 107
 Balmer discontinuity, 106
 Balmer lines, 229, 230, 287
 Balmer series, 106, 107
 Barium star, 235
 Barnard, E.E., 218
 Barnard's Star, 28
 Basis vector, 466
 Bayer, Johannes, 31, 34
 Be star, 234
 Beam efficiency, 115
 Becklin–Neugebauer object, 348, 349
 Bell, Jocelyn, 316
 BepiColombo, 184
 Beryllium, 277, 280
 Bessel, F.W., 28
 Betelgeuse, 236, 303
 Bethe, Hans, 254
 Big Dipper, 31
 Binary pulsar, 246
 Binary stars, 241, 491
 astrometric, 241, 242
 close, 241, 272, 275, 307
 contact, 241, 245, 272, 275
 detached, 272, 275
 distant, 241
 evolution, 272, 276
 mass transfer, 273
 masses, 242, 244, 246
 nova, 307
 optical, 241
 photometric, 241, 244
 semidetached, 272, 275
 spectroscopic, 241, 243
 visual, 241, 242
 Binding energy, nuclear, 254, 255
 Black dwarf, 271
 Black hole, 271, 272, 274, 310, 313, 320, 383, 414, 440
 rotating, 320, 321
 supermassive, 322
 Blackbody, 111, 113, 425, 429
 Blazar, 414
 Blue loop, 270
 Blueshift, 30
 Bode, Johann, 208
 Bohr model, 105
 Bok, Bart, 332
 Bolides, 195, 223
 Bolometric correction, 95
 Boltzmann constant, 111, 155, 478
 Boltzmann distribution, 110, 116
 Bond albedo, *see* Albedo, Bond
 Bonner Durchmusterung, 31, 32, 34
 Boron, 280
 Boson, 109
 Bow shock, 158
 Bowell, Edward, 164
 Bowen, Ira S., 344
 Brackett series, 107
 Breccia, 193
 Bremsstrahlung, 105
 Brown dwarfs, 229, 267, 274
 Bulge, galactic, 390, 397
 Bunsen, Robert, 227

- Bureau International des Poids et Mesures, 38
Burning, 254
Butterfly diagram, 292
- C**
C asteroids, 219
Caesar, Julius, 40
Calcium, 229, 293
 interstellar, 337, 339
Calendar, 40
 Gregorian, 41
 Julian, 40
 Roman, 40
Callisto, 202, 203
Caloris Basin, 184
Cannon, Annie Jump, 229
Capella, 31
Carbon, 445, 451
Carbon burning, 257, 270
Carbon cycle, *see* CNO cycle
Carbon flash, 270, 274
Carbon monoxide, 345, 347
Carbon-nitrogen-oxygen cycle, *see* CNO cycle
Carbonaceous chondrites, 224
Carlsberg foundation, 62
Carruthers, George R., 345
Carte du Ciel, 32
Cassini, G., 200
Cassini division, 206
Cassiopeia A, 351
Cataclysmic variables, 303
Catastrophe theories, 168
CCD camera, 70, 95
CDM, *see* Cold dark matter
Cech, Thomas, 450
Celestial mechanics, 123
Celestial pole, 14, 17, 20
Celestial sphere, 16
Cell, eukaryotic, 446, 451
Cell, prokaryotic, 446
Centaurus A, 411
Central angle, 11
Cepheid, 270, 276, 300–302, 312, 371
Čerenkov radiation, 81, 86
Ceres, 213, 214
Chajnantor, 77
Chandler period, 38
Chandrasekhar limit, 309
Chandrasekhar mass, 271, 273, 315
Charge number, 103
Charon, 213
Chemosynthesis, 450
Chile, 53
Chirality, 446
Chiron, 218
Chondrites, 224
Chromatic aberration, 56
Chromosphere, 286, 287, 293, 296
Churyumov–Gerasimenko, 221, 222
Circle, 463
- Circumpolar, 20
Clementine, 194
Cloud
 molecular, 347
Cluster
 globular, 10, 359, 363, 374, 375
 open, 359, 361, 373
Cluster variable, 303
CMB, *see* Cosmic microwave background
CNO cycle, 255, 257, 267, 274
Coalsack, 332
COBE, 425
Codon, 446
Coelostat, 62
Coherent radiation, 104
Cold dark matter, 416, 431, 433
Colour excess, 98, 330
Colour index, 96, 235, 330
Colour-aperture effect, 404
Colour-luminosity relation, 404
Colour-magnitude diagram, 362, 363
 of globular clusters, 364
Column density, 342
Coma, 59, 220
Coma Berenices, 359
Coma Cluster, 407
Comet, 214
Comets
 orbit, 222
 origin, 221
 periodic, 487
 structure, 219
 tail, 220
Compact star, 313
Conduction, 250
Conic section, 125, 126, 464
Conjunction, 144, 147
 inferior, 144
 superior, 145
Constellation, 31, 495
Continental drift, 189
Continuously habitable zone, 448
Continuum, 111
Convection, 251, 265, 269, 271, 279, 286
Convective temperature gradient, 251
Coordinated universal time, *see* Time, coordinated universal
Coordinates
 absolute, 27
 ecliptic, 21
 equatorial, 17, 20, 21, 138
 galactic, 22, 367, 370
 geocentric, 22
 heliocentric, 22
 heliocentric ecliptic, 137
 horizontal, 16
 inertial, 123
 perturbations, 22
 polar, 126, 128
 reduction of, 41

- relative, 27
 terrestrial, 38
 topocentric, 16, 22
 Copernican principle, 427
 Copernicus, 3
 Core, 188
 Core rotation, 189
 Corona, 187, 287, 289, 294, 296
 Coronagraph, 289
 Coronal holes, 296
 Coronal mass ejection, 295
 Coronium, 288
 Correlation function, 435
 Cosine formula, 14
 Cosmic microwave background, 425
 Cosmic ray, 10, 72, 85, 280, 295, 327, 354, 399
 Cosmogony, 167
 Cosmological constant, 429, 432
 Cosmological model, 427
 Cosmological principle, 426
 Cosmological term, 429
 Cosmology, 421
 Coulomb's law, 106
 Crab nebula, 309, 317, 350, 351, 355
 Crab pulsar, 317, 319
 Critical density, 429, 431
 Critical layer, 157
 Crust, 152
 Culmination, 17, 39
 Curie point, 159
 Curvature tensor, 474
 Curve of growth, 229
 Curved space, 474
 Cyanobacteria, 451
 Cyanogen, 345
 Cygnus, 369
 Cygnus X-1, 321, 322
- D**
- δ Scuti star, 303
 - D'' layer, 189
 - Dark current, 71
 - Dark energy, 433
 - Dark matter, 417, 431
 - Dark nebula, 332
 - Davis, R., 285
 - Day, 40, 479
 - length, 35
 - sidereal, 34, 35
 - solar, 35
 - synodic, 35
 - De Buffon, 168
 - De Chézeaux, Loys, 421
 - De Vaucouleurs' law, 397, 401
 - Deceleration parameter, 430
 - Declination, 17, 20, 23, 27, 174
 - Declination axis, 18, 60
 - Degenerate matter, 253, 269, 271, 313, 319
 - Degenerate neutron gas, 315
 - Deimos, 199
 - Deneb, 31
 - Density, 10
 - Density parameter, 430
 - Density wave theory, 382
 - Descartes, R., 168
 - Detectors, 64
 - Deuterium, 432
 - Deuteron, 255
 - Diaphragm, 69
 - Differential rotation
 - Milky Way, 375–377
 - Sun, 285, 292
 - Differentiation, 152, 172
 - Diffraction, 88, 329
 - Diffraction grating, 73
 - Diffuse galactic light, 334
 - Dipole magnetic moment, 159
 - Direct motion, 144
 - Dirty snowball, 220
 - Dispersion, 73, 227
 - Dissociation, 264
 - Distance modulus, 96, 98, 370, 371
 - Distances
 - extragalactic, 302
 - Diurnal parallax, 24
 - Diurnal tides, 175
 - DNA, 446, 447, 450
 - Doodson's tidal constant, 174
 - Doppler broadening, 108
 - Doppler effect, 28, 229, 243, 301, 337, 362, 422
 - Draconic month, 148
 - Drake, Frank, 455, 456
 - Drake formula, 456
 - Draper, Henry, 229
 - Dreyer, J.L.E., 359
 - Dust
 - interplanetary, 214
 - Dust, interstellar, 266, 327
 - Dust disk, 461
 - Dust tail, 220
 - Dwarf cepheid, 300, 301, 303
 - Dwarf nova, 300, 304, 306, 307
 - Dwarf planet, 141, 212
 - Dwarf spheroidal, 391
 - Dwarf stars, 237
 - Dynamo, 159, 293
 - Dynode, 69
 - Dysnomia, 214
- E**
- η Carinae, 270, 305, 306, 337
 - E.T., *see* Equation of time
 - Early spectral class, 229
 - Earth, 14, 142, 188, 480
 - atmosphere, 190
 - core, 188
 - crust, 189
 - density, 188
 - flattening, 15, 175
 - greenhouse effect, 167

- life, 188
magnetic field, 160, 189, 192
magnetic field polarity, 160, 192
mantle, 189
radius, 15, 24
rotation, 34, 38
surface, 153
temperature, 152, 167
water, 190
- Earthquakes, 152
- Ebb, 149
- Eccentricity, 126, 127, 464
- Eclipse, 149
- annular, 149
 - lunar, 149
 - solar, 149
- Eclipsing binary, 273
- Eclipsing variable, 241, 244
- Ecliptic, 21, 128, 146
- Ecliptic latitude, 21
- Ecliptic longitude, 21, 23
- Ecliptic system, 21
- Eddington, Sir Arthur, 301
- Eddington-Barbier approximation, 239
- Einstein, Albert, 421, 429
- Einstein cross, 415
- Einstein-de Sitter model, 430
- Elder, Frank, 352
- Electric dipole transitions, 110
- Electromagnetic radiation, 103
- Electron, 103, 478
- Electron gas, 253
- Electron volt, 478
- Elements
- abundances, 278
 - origin, 277
- Elena, 191
- Elevation, 17
- Ellipse, 126, 464
- Ellipsoidal variables, 244
- Elongation, 145
- Emission, 103, 104, 106
- induced, 104, 116
 - spontaneous, 103
- Emission coefficient, 117
- Emission line, 104, 227, 304
- Emission measure, 343
- Emission nebula, 307
- Enceladus, 162, 207, 208
- Encke's division, 206
- Energy conservation equation, 250
- Energy density, 93, 429
- Energy integral, 126, 127, 133
- Energy level, 103, 104, 108
- Energy production in stars, 250
- Energy transport, 236, 271
- Enzyme, 446
- Ephemeris second, 39
- Epimetheus, 207
- Epoch, 23
- Equation of a planetary orbit, 126
- Equation of equinoxes, 35
- Equation of motion, 123, 130
- Equation of state
- of degenerate electron gas, 253
 - of degenerate gas, 315
 - of perfect gas, 252
- Equation of time, 36
- Equator, 14, 20, 146
- Equatorial system, 17
- Equivalence principle, 475
- Equivalent width, 108, 229, 232
- Ergosphere, 320, 321
- Eris, 213, 214
- Eros, 130
- Erosion, 190
- ESA, 32
- Escape velocity, 133, 320
- from the Galaxy, 380
- ESO, 34, 63
- Ethanol, 347
- Euclidean space, 473
- Europa, 202, 203, 454
- water, 203
- European Southern Observatory, *see* ESO
- European Space Agency, *see* ESA
- EUV, 82
- Evening star, 145, 185
- Event horizon, 320
- Ewen, Harold, 341
- Excitation, 103
- Excitation state, 103
- lifetime, 107, 110, 116
- Exit pupil, 56
- Exoplanet, 459, 461
- Exosphere, 157, 192
- Explorer 1, 158
- Extinction, 53, 97, 98
- atmospheric, 98
 - interstellar, 98, 101, 328–331
- Extinction coefficient, 99
- Extinction curve, 330, 337
- Extinction efficiency, 329
- Extinction efficiency factor, 328, 329
- Extreme ultraviolet, 82
- Eye, 56
- Eyepiece, 54
- F**
- Facula, 293
- Faraday rotation, 105, 355, 400
- Fault, 190
- Fermion, 109
- Field equation, 475
- Filter, 69, 95
- Fission, 280
- FK 6, 32
- Flare, 294, 295
- Flare star, 300, 303, 304

- Flash spectrum, 286, 287
 Flat-field, 71
 Flattening, 151
 Flood, 149
 Flux, 91–93, 98
 total, 92
 Flux density, 91, 93–96, 98, 114
 Focal length, 54
 Focal plane, 54
 Focus
 Cassegrain, 57
 coude, 58
 Nasmyth, 61
 Newton, 57
 of an ellipse, 464
 primary, 57
 Forbidden line, 110, 304, 344
 Forbidden transition, 110, 350
 Four-vector, 475
 Fraunhofer, Joseph, 227
 Free–free radiation, 104, 105, 351
 Friedmann, Alexander, 429
 Friedmann model, 429
 FU Orionis, 304
 Full moon, 147, 149
 Full width at half maximum, *see* FWHM
 Fundamental Katalog, 32
 Fusion of hydrogen, 254
 Fusion reaction, 254
 FWHM, 108
- G**
 Gaia, 368
 Galactic coronal gas, 353
 Galactic habitable zone, 448
 Galactic north pole, 22
 Galaxies, 387
 active, 409
 age, 404
 barred spiral, 381, 390
 blue compact dwarf, 392
 brightest, 494
 cD, 388, 398
 classification, 387
 dE, 391, 392
 disk, 388, 399
 distances, 392, 423
 distribution, 331, 422
 dSph, 391
 dynamics, 401
 element abundances, 404
 elliptical, 388, 397
 evolution, 416
 irregular, 391
 lenticular, 389
 luminosity, 387, 393
 luminosity classification, 393
 mass, 387, 394
 rotation, 402
 shape, 401
 spiral, 389, 390
 spiral structure, 402
 structure, 397
 surface brightness, 397
 ultracompact dwarf, 392
 ultrafaint dwarf, 392
 unified models, 413
 Galaxy cluster, 405, 406
 Galaxy group, 405, 406
 Galilean satellites, 202
 Galilei, Galileo, 3, 51, 202, 206, 290, 367, 475
 Galle, J.G., 211
 Gamma ray bursts, 318
 Gamma rays, 80, 354
 Gamow, George, 315, 425
 Ganymede, 202, 203
 magnetic field, 160
 Gas, interstellar, 266, 327
 Gas constant, 478
 Gauss, J.K.F., 132
 Gegenschein, 224
 Geiger–Müller counter, 82
 Geminga, 319
 Geminides, 223
 Gene, 446
 General catalogue, 32
 General Catalogue of Variable Stars, 299
 General theory of relativity, 182, 421, 429
 Geodesic, 475
 Geoid, 15, 151
 Geometric albedo, *see* Albedo, geometric
 Georgelin, Y.M. and Y.P., 381
 Geoscience, 141
 Giant, 236
 Giant branch, 277
 Giant phase, 269, 274, 276
 Giant planets, 142
 atmosphere, 153
 Giant star, 232
 Globular cluster, 270, 277, 303
 Globule, 332
 Gondwana, 189
 Granulation, 286, 287
 Graphite, 337
 Gravitation, 142
 Newton's law, 123, 136
 Gravitational constant, 123, 130, 478
 Gravitational lens, 414, 460
 Gravitational radiation, 87
 Gravitational waves, 318, 476
 Grazing reflection, 81, 82
 Great circle, 11
 Great Red Spot, 200
 Greenhouse effect, 167, 185, 448
 Greenwich, 14, 36
 Gregory, James, 57
 Gregory XIII, Pope, 41
 Ground state, 103
 GRS-80, 15

H

h and χ Persei, 360
 H_{α} line, 287, 288, 293
Habitable zone, 448
Hadron, 433
Hadron era, 433
Halley, Edmond, 421
Halley's comet, 128, 220
Halo, 194
galactic, 397
Hartmann, Johannes, 337
Haumea, 213
Hawking process, 440
Hayashi track, 265, 269, 270
HDE 226868, 321
Heisenberg uncertainty principle, 107
Helioseismology, 285
Helium, 255, 434
Helium burning, 269
Helium flash, 268, 269, 274
Helium star, 273, 276
Helix nebula, 350
Henry Draper catalogue, 31, 229
Herbig–Haro object, 266, 267
Hercules, 368
Hercules X1, 323, 324
Herlofson, Nicolai, 352
Herschel, William, 208, 367
Hertzsprung, Ejnar, 235
Hertzsprung gap, 275
Hertzsprung–Russell diagram, *see* HR diagram
Hesperos, 185
Hewish, Anthony, 316
HG system, 164
HI region, 342
H II region, 336, 342, 345, 347, 351, 373, 381
Hipparchos, 10, 32, 93
Hirayama families, 216
Homo habilis, 10
Homunculus nebula, 306
Horizon, 16
Horizontal branch, 236, 269, 364
Horizontal refraction, 20
Horizontal system, 16
Horsehead nebula, 332, 336
Hot jupiter, 461
Hour angle, 18, 174
Hour axis, 18, 60
Hour circle, 18
Hoyle, Fred, 453
HR diagram, 235, 258, 265, 268, 371
variables, 300
Hubble, Edwin, 333, 388, 421
Hubble constant, 422, 424, 430
Hubble Guide Star Catalog, 34
Hubble relation for reflection nebulae, 333, 336
Hubble sequence, 388
Hubble Space Telescope, 63, 419, 424, 431
Hubble's law, 422, 428
Hulse, Russell, 317

Huygens, 208

Huygens, C., 206
Hyades, 359, 360, 362
colour–magnitude diagram, 363
distance, 362

Hydra, 214

Hydrogen
21 cm line, 110, 338, 340, 379
interstellar, 110, 338
molecular, 345
neutral, 232
structure of atom, 105
transitions, 106

Hydrogen burning, 263, 266, 267, 274

Hydrostatic equilibrium, 151, 236, 249, 264, 274, 302, 319

Hydroxyl radical, 346, 347

Hyperbola, 126, 127, 464

Hyperion, 207

HZ Herculis, 323

I

Iapetus, 207
IAU, 31, 41, 130, 377
Ice, interstellar, 337
IceCube, 87
Icehouse effect, 449
Image intensifier, 70
Impact crater, 223
Inclination, 127, 128, 243
Index Catalogue, 359
Index of refraction, 26
Inferior conjunction, 144, 181
Inferior planet, 144
Inflation, 433, 437
Infrared radiation, 52, 84
Integral, of equation of motion, 124

Intensity, 91, 99

specific, 91

total, 91

Interferometer, 73, 77

intensity, 74

Michelson, 74, 87

speckle, 74

International Astronomical Union, *see* IAU, 141

Interstellar cloud

fragmentation, 349

Interstellar dust, 327

albedo, 333

density, 356

grain size, 356

origin, 337

properties, 338

temperature, 334

Interstellar gas, 337

atomic hydrogen, 338

hydrogen distribution, 343

properties, 338

Interstellar matter, 327, 360

distribution, 378

- element abundances, 340
in galaxies, 399
- Interstellar molecules, 75, 345, 347
- Interval, 474
- Interval halving, 471
- Intrinsic colour, 98, 330
- Inverse matrix, 468
- Io, 154, 202
volcanic activity, 203
- Ion tail, 220
- Ionisation, 104, 264, 416
- Ionisation zone, 302
- Ionosphere, 53, 74, 191
- IRAS, 336
- Iron, 254, 271
- Iron meteorites, 223
- Irregular variable, 301, 304
- Ishtar Terra, 186
- Isotopes, 277
- Isotropic radiation, 92
- Isotropy, of the Universe, 425
- Isua, 451
- Iteration, 470
- IUE, 353
- IUGG, 15
- J**
- J2000.0, 23
- James Webb Space Telescope, 63
- Jansky, Karl, 74, 352
- Jansky (Jy), 92
- Janus, 207
- Jeans, James, 135, 168, 349
- Jeans length, 136
- Jeans limit, 135
- Jeans mass, 136, 169, 349, 436
- Jet, galactic, 411
- Johnson, Harold L., 95
- Julian century, 38, 41
- Julian date, 38, 41, 44
- Jupiter, 142, 199
atmosphere, 156
belts, 200
composition, 199
helium abundance, 201
magnetic field, 160, 201, 202
moons, 202
radio emission, 202
ring, 202
rotation, 200
structure, 200
temperature, 152, 167
winds, 201
zones, 200
- K**
- Kamiokande neutrino detector, 87
- Kant, I., 168
- Kapteyn, Jacobus, 367
- Katalog der Astronomischen Gesellschaft, 32
- Keenan, Philip C., 232
- Kellman, Edith, 232
- Kepler, 460
- Kepler, Johannes, 3, 421
- Kepler's equation, 133, 138
- Kepler's first law, 127
- Kepler's second law, 129, 132, 147
- Kepler's supernova, 309
- Kepler's third law, 129, 216, 242, 244
- Kerr, Roy, 321
- Kholopov, P.N., 299
- Kiepenheuer, Karl-Otto, 352
- Kinematic parallax, 362
- Kirchhoff, Gustav R., 227
- Kirkwood gaps, 216
- Kleinmann–Low nebula, 348
- Koshiba, M., 285
- Krüger 60, 242
- Kuiper, Gerard, 84, 218, 221
- Kuiper Airborne Observatory, 84
- Kuiper belt, 173, 218, 221
- Kukarkin, Boris, 299
- L**
- Λ CDM model, 433
- La Palma, 53
- Labeyrie, Anton, 74
- Lagoon nebula, 345
- Lagrangian points, 131, 217
- Lambertian surface, 161
- Langmuir, Robert, 352
- Laplace, 168, 320
- Laplace's tidal equation, 175
- Large Magellanic Cloud, 354, 405, 424
supernova, 311
- Laser, 116
- Last Universal Common Ancestor, 451
- Late spectral class, 229
- Latitude, 14, 174
galactic, 331
geocentric, 15
geodetic, 15
geographical, 14
- Le Verrier, U.J.J., 211
- Lead, 280
- Leap day, 41
- Leap second, 39, 146
- Leavitt, Henrietta, 302
- Lemaître, Georges, 429
- Lepton era, 434
- Lexell, Anders, 208
- Libration, 148
- Life, 188, 445
detection of, 454
intelligent, 455
- Light-year, 28, 478
- Lightcurve
of a binary, 244–246
stellar, 299
- LIGO, 87, 88

- Limb brightening, 295
Limb darkening, 209, 237, 245, 286
Lin, Chia-Chiao, 382
Line element, 427, 473
Line profile, 107, 228
Line spectrum, 103
Lithium, 266, 277, 280
Lithosphere, 152, 189
Local arm, 374
Local Group, 406, 493
Local standard of rest, 368, 370, 375
Local Supercluster, 409, 424
Long period comets, 221
Long period tides, 175
Long period variable, 300, 303
Longitude, 14, 126
 ecliptic, 23, 24, 144
 galactic, 22
Longitude of perihelion, 128
Longitude of the ascending node, 127, 128
Lorentz transformation, 474
Love number, 175
Lowell, P., 196, 213
Lowell Observatory, 213
Lower culmination, 19
LUCA, 451, 452
Luminosity, 92, 97, 112, 232
Luminosity class, 232
Luminosity function, 371, 393, 394
 initial, 372
Luminous blue variable, 270
Lumme, Kari, 164
Luna, 192
Lunar eclipse, 149
 partial, 150
 total, 150
Lunar phases, 147
Lunar Prospector, 194
Lyman α line, 337, 338, 341
Lyman α forest, 413
Lyman series, 107
Lyot, Bernard, 185
- M**
- M1, 317, 351
M100, 354
M13, 359, 455
M31, 406, 421
M32, 392
M33, 406
M42, 342, 344, 347
M5, 364
M51, 403
M57, 350
M78, 333, 335
M8, 345
M87, 411, 412
Magellan, 186
Magellanic Stream, 405
- Magnetars, 318
Magnetic field, 105, 109, 117, 142, 157, 189, 353
 galactic, 327, 399
 interstellar, 331, 354, 355
 of neutron stars, 316
 of pulsar, 317
 solar, 291–293
Magnetic quantum number, 109
Magnetic variable, 300
Magnetopause, 158
Magnetosphere, 157, 202
Magnification, 55
Magnitude, 94, 115
 AB, 96
 absolute, 96, 98, 235, 371
 absolute bolometric, 97, 113
 apparent, 94
 bolometric, 95
 of a binary star, 100
 photographic, 94
 photovisual, 95
 UBV, 95
 ugriz, 96
 visual, 94
Magnitude-redshift test, 431
Main sequence, 232, 235, 236, 258, 265–268, 271, 274, 276
 of globular clusters, 364
 of open clusters, 362
Main sequence fitting, 363, 371
Makemake, 213
Maksutov telescope, 60
Mantle, 189
Mare, 192
Mare Imbrium, 193
Mariner 10, 184
Mariner 2, 186
Markarian, Benyamin, 410
Markarian galaxy, 410
Mars, 142, 196, 452
 atmosphere, 196
 canals, 196
 composition, 199
 ice, 198
 life, 199
 magnetic field, 199
 polar caps, 196, 198
 rivers, 197
 structure, 199
 temperature, 167, 198
 volcanoes, 197
Mars Global Surveyor, 196, 199
Mars Odyssey, 198
Mascons, 193
Maser, 116
 interstellar, 347, 348
Mass
 gravitational, 475
 inertial, 475
Mass absorption coefficient, 251, 254, 262

- Mass continuity equation, 250
 Mass distribution of stars, 250
 Mass function, 244
 Mass number, 103
 Mass transfer, 276
 Mass–luminosity relation, 237, 238, 258
 Matrix, 467
 Matrix product, 468
 Matter era, 434
 Mauna Kea, 53, 63, 84, 156
 Maunder minimum, 291
 Maximum magnification, 56
 Maxwell, James Clerk, 186, 206
 Maxwell Montes, 186
 Mean equinox, 35
 Mean free path, 254, 262
 Mean longitude, 128
 Mean motion, 145
 Mean place, 27, 42
 Mean sun, 36
 M'echain, Pierre, 359
 Mercury, 142, 181
 - advance of perihelion, 182, 476
 - atmosphere, 184
 - diameter, 181
 - distance, 182
 - ice, 184
 - length of day, 181, 182
 - magnetic field, 160, 184
 - orbit, 182
 - surface, 184
 - temperature, 167, 182
 - transit, 151
- Meridian, 14, 17, 18
 Meridian circle, 27, 62
 Merope, 333, 335
 Mesopause, 191
 Mesosphere, 191
 Messenger, 184
 Messier, Charles, 317, 359
 Messier catalogue, 359
 Metagalaxy, 427
 Metallic hydrogen, 153, 200, 205
 Metastable state, 116
 Meteor, 195, 222, 223
 Meteor shower, 223, 486
 Meteor streams, 223
 Meteorites, 223
 Meteoroids, 192, 214, 222
 Meteors
 - origin, 224
- Methane, 454
 Methyldyne, 345
 Metric, of space, 473
 Metric tensor, 473
 Michelson, 77
 Micrometeoroids, 222
 Mid-Atlantic ridge, 189
 Mid-ocean ridge, 189
 Midnight Sun, 146
- Mie scattering, 329
 Milanković, Milutin, 146
 Milanković cycle, 147
 Milky Way, 22, 367, 492
 - bar, 381
 - centre, 382
 - corona, 353, 374
 - G dwarf problem, 385
 - halo, 374
 - metallicity, 384
 - rotation, 342, 379, 380
 - spiral arms, 348, 381
 - spiral structure, 373, 379, 381
 - supernovae, 309
 - thick disk, 381
- Miller, Stanley, 450
 Mimas, 207
 Minimum magnification, 56
 Minkowski space, 475
 Minor bodies, 214
 Mira Ceti, 299, 303
 Mira star, 301, 303, 348
 Miranda, 210, 211
 Misam, 360
 Mizar, 243
 MKK spectral classification, 371
 Moho discontinuity, 189
 Mohorovičić, A., 189
 Molecular weight, 252
 Month, 40, 479
 Moon, 147, 188, 192, 480
 - asthenosphere, 193
 - craters, 192
 - first quarter, 147
 - laser ranging, 192
 - last quarter, 147
 - lithosphere, 193
 - nodes, 149
 - orbit, 147, 149
 - orbit change, 149
 - orbital perturbations, 148
 - orbital plane, 148
 - origin, 194
 - perigee, 148
 - phases, 147
 - seismometry, 192
 - water, 194
- Moonquakes, 192
 Morgan, William W., 95, 232
 Morning star, 145, 185
 Mosaic mirror, 62
 Moulton, F.R., 168
 Mount Lemon, 84
 Mount Palomar, 34
 Mounting
 - azimuthal, 60, 61
 - Dobson, 61
 - equatorial, 60
- Mrkos, 221
 Muon, 86

- MXB 1730–335, 324
- N**
- Nadir, 16
- Nautical mile, 16
- NEAR, 219
- Near-Earth asteroids, 218
- Nebula
- dark, 332, 334, 347
 - emission, 334, 342
 - reflection, 333, 334, 360
- Nebular hypothesis, 168
- Nebular variable, 303, 304
- Neptune, 142, 211
- discovery, 211
 - magnetic field, 160, 211
 - moons, 212
 - orbit, 211
 - rings, 212
 - rotation, 211
 - structure, 211
- Neutrino, 310, 431
- Neutrino decoupling, 434
- Neutrino detectors, 86
- Neutrinos, 86
- from pp chain, 255
 - from supernova, 312
 - from URCA process, 315
 - mass, 285
 - solar, 277, 284
- Neutron, 103, 478
- Neutron capture, 257, 278
- Neutron star, 10, 271, 274, 310, 312, 313, 315
- radius, 319
 - rotation, 316
- New General Catalogue of Nebulae and Clusters of Stars, 359
- New moon, 147
- Newton, Isaac, 3, 57, 227
- Newton's laws, 136
- N galaxy, 410
- NGC 1265, 411
- NGC 1435, 333, 335
- NGC 1952, 351
- NGC 1976, 344
- NGC 2068, 333, 335
- NGC 4565, 403
- NGC 5195, 403
- NGC 6523, 345
- NGC 6960, 352
- NGC 6992, 352
- NGC 7023, 333
- NGC 7293, 350
- Nice model, 170
- Nix, 214
- Noctilucent clouds, 194
- Nodal line, 148
- North Polar Spur, 355
- Nova, 229, 304, 305, 307
- dwarf, 305
 - ordinary, 305
 - recurrent, 305
- Nova Cygni, 308
- Nova outburst, 273
- Nova-like star, 304
- Nova-like variable, 305
- Nucleosynthesis, 432
- Nucleotide, 446
- Numerical solution of an equation, 470
- Nutation, 24, 35, 41, 146, 148
- O**
- ω Centauri, 361
- OAO 2, 338
- Objective, 54
- Objective prism, 227, 229
- Obliquity of the ecliptic, 22, 24, 35, 42, 146
- Occultation, 149, 150
- Olbers, Heinrich, 421
- Olbers paradox, 421, 422
- Olympus Mons, 197, 198
- Oort, Jan, 220, 375
- Oort cloud, 173, 218, 221
- Oort constant, 377
- Oort limit, 327
- Oort's formulas, 375, 376, 385
- Opacity, 97, 253, 264, 302
- Open cluster, 277, 327, 349
- Open star cluster, 328
- Ophiuchus, 334
- Oppenheimer–Volkoff mass, 271, 273
- Opposition, 144
- Opposition effect, 163
- Optical depth, 118, 236, 328
- Optical SETI, 455
- Optical thickness, 97, 101
- Optical window, 52
- Orbit determination, 131
- Orbital elements, 124, 127, 136
- Orbital resonance, 202
- Orbital velocity, 134, 137
- Orion, 336
- Orion nebula, 336, 342, 344, 346–348
- Osculating elements, 128, 482
- OSETI, 455
- Oxygen burning, 257, 270
- Oxygen flash, 270
- Ozma, 455
- Ozone, 52, 454
- Ozone layer, 191
- P**
- P waves, 152, 189
- P-process, 280
- PAH, 452
- Pair formation, 280
- Palomar Sky Atlas, 34
- Pancake domes, 187

- Pandora, 207
 Pangaea, 189
 Panspermia, 453
 Parabola, 126, 127, 464
 Parallax, 24, 27
 - annual, 24
 - diurnal, 24
 - of the Moon, 25
 - of the Sun, 25
 - photometric, 371
 - statistical, 369
 - trigonometric, 27, 367, 371
 Parallel of latitude, 14
 Parameter, of a conic section, 127, 464
 Parametrised Post-Newtonian, 476
 Parsec, 28, 478
 Paschen series, 107
 Pauli exclusion principle, 253, 315
 PCygni star, 234
 PCygni, 270
 Peculiar motion, 369
 Peculiar velocity, 424
 Pelican Nebula, 266
 Penumbra, 290
 Penzias, Arno, 425
 Periastron, 125
 Pericentre, 125
 Perigee, 125
 Perihelion, 125, 127, 146
 Period
 - pulsation, 301
 - sidereal, 35, 145
 - synodic, 145
 Period–luminosity relation, 276, 302
 Permanent tide, 175
 Perseids, 223
 Perseus arm, 374
 Perturbations, 130
 Perturbograph, 131
 Phase, 145
 Phase angle, 145, 161, 163, 481
 Phase curve, 163
 Phase function, 161
 Phase integral, 161
 Phase space, 253
 Phobos, 199
 Phosphorus, 185
 Photocathode, 68
 Photodissociation, 257
 Photography, 32, 55, 64, 65
 Photometer, 69, 95
 Photometry, 69
 Photomultiplier, 69
 Photon, 103
 Photonuclear reaction, 257
 Photopolarimeter, 69
 Photosphere, 286, 293
 Photosynthesis, 191, 450
 Phylogenetic tree, 451, 452
 Piazzi, G., 214
 Pico del Teide, 84
 Pico Veleta, 77
 Pillan Patera, 154
 Pioneer 11, 206
 Pioneer Venus 1, 186
 Pixel, 70
 Place, 293
 Planck constant, 103, 105, 478
 Planck's law, 111, 117
 Planet, 141
 - inferior, 142
 - superior, 142
 Planetary nebula, 229, 270, 274, 349
 Planetary surfaces, 153
 Planetesimal, 170
 Planets, 141
 - magnitudes, 162, 481
 - mean elements, 482
 - satellites, 483
 Plasma, 191, 264, 351
 Plate tectonics, 152, 189
 Pleiades, 333, 335, 359–361
 Plutinos, 218
 Pluto, 213
 - atmosphere, 213
 - composition, 213
 - discovery, 213
 - mass, 213
 - orbit, 213
 - rotation, 213
 - size, 213
 Pogson, Norman R., 94
 Polar axis, 60
 Polar variation, 38
 Polaris, 17, 23
 Polarisation, 69, 105, 165, 355, 399
 - interstellar, 331
 Pole, of a spherical triangle, 11
 Pollack, Herbert, 352
 Population, 374
 - I, 302, 374
 - II, 302, 364, 375
 - intermediate, 375
 Population inversion, 116
 Population numbers, 110
 Population synthesis, 404
 Positional astronomy, 27
 Positron, 255
 Power spectrum, 435
 Poynting, J.P., 173
 Poynting–Robertson effect, 173
 pp chain, 255, 256, 266–268, 274, 283
 ppI chain, 255
 ppIII, 255, 277
 PPM, 32
 PPN, 476
 Praesepe, 359
 Precession, 22, 31, 35, 36, 41, 146
 Precession constants, 24

- Pressure
of degenerate electron gas, 253
of degenerate gas, 259
of gas, 258
of relativistic degenerate gas, 253
radiation, 258
- Principal quantum number, 105, 108
- Prometheus, 154, 207
- Prominence, 288, 294
loop, 291, 292, 294
- Proper frequency, 301
- Proper motion, 28, 30, 243, 362, 377
- Proplyd, 169
- Proportional counter, 82
- Protein, 446
- Proton, 103, 440, 478
- Proton-proton chain, *see* pp chain
- Protostar, 264, 274, 348
- Proxima Centauri, 28, 141
- PSR 1133+16, 316
- PSR 1642–03, 316
- PSR 1913+16, 317, 476
- Ptolemy, 31
- Pulsar, 316, 351, 459
binary, 317
- Purcell, Edward, 341
- Q**
- QSO, 411
- Quantum efficiency, 67, 68, 70
- Quantum numbers, 108
- Quasar, 20, 411, 422
- R**
- R Coronae Borealis star, 304, 305
- R-process, 279
- Radial velocity, 28, 30
interstellar clouds, 378
- Radian, 463
- Radiation, 266, 271
blackbody, 117
coherent, 116
non-thermal, 116, 353
- Radiation belts, 158
- Radiation constant, 251, 259
- Radiation decoupling, 434
- Radiation density constant, 478
- Radiation era, 434
- Radiation pressure, 142, 252
- Radiative energy transport, 251
- Radiative temperature gradient, 251
- Radiative transfer, 117, 251, 341
- Radio galaxy, 410, 411, 422
- Radio spectroscopy, 346
- Radio telescope, 74
- Radio window, 53
- Radiometer, 74
- Radius vector, of a planet, 132
- Rainbow, 194
- Random walk, 261
- Rayet, Georges, 234
- Rayleigh, Lord, 53
- Rayleigh scattering, 329
- Rayleigh–Jeans approximation, 113, 114, 341, 351
- Readout noise, 72
- Reber, Grote, 74
- Recombination, 104, 343
- Recurrent nova, 305
- Red clump, 270
- Red giant, 235, 268, 269
- Red giant branch, 236
- Red-edge, 455
- Reddening, 98, 329, 363
- Reddening curve, 330
- Redshift, 30, 411, 422, 428
gravitational, 313, 321, 425, 476
- Reduction of observations, 99, 101
- Reflection nebula, 333
- Reflector, 54, 57
- Refraction, 20, 25, 51
- Refraction angle, 26
- Refractor, 54, 56
- Regolith, 193
- Relativity theory, 320, 473
general, 313, 317, 475
tests of, 476
- Resolving power, 55, 75, 77
- Retrograde motion, 128, 144
- Retrograde rotation, 35
- Right ascension, 17, 20, 23, 27
- Ring nebula, 350
- Rising and setting times, 20, 39
- RNA, 447, 450
- Robertson, H.P., 173
- Robertson–Walker line element, 427
- Roche surface, 245, 273, 275, 307
- Rogue planet, 459
- Rosetta, 222
- Rotation curve, 379, 380, 395, 402
- Rotation matrix, 468
- Rotational transition, 347
- RR Lyrae star, 270, 300, 301, 303, 364, 371, 375
- Russell, Henry Norris, 235
- RV Tauri star, 300, 301, 303
- Rydberg constant, 107, 478
- Ryle, Sir Martin, 78
- S**
- S asteroids, 219
- S waves, 152, 189
- S-process, 279
- Sagittarius arm, 374, 382
- Sagittarius B2, 347, 348
- San Andreas fault, 190
- SAO, 32
- SAO catalogue, 32
- Saros, 150
- Satellite, 142, 499
Beppo-SAX, 318
Chandra, 81, 414

- COBE, 84, 224, 381, 425, 437
 Compton Gamma Ray Observatory, 80, 318
 Copernicus, 339, 345, 353
 Einstein, 81
 EUVE, 82
 Fermi, 80
 Gaia, 34
 GLAST, 80
 Hinode, 81
 Hipparcos, 32, 368, 371
 Integral, 80
 IRAS, 84
 ISO, 84
 IUE, 82
 MAP, 84
 OAO, 82
 OSO 3, 80
 Planck, 396, 409, 437
 RHESSI, 81
 SAS 1, 81
 SOHO, 296
 Spitzer, 84
 Swift, 318
 TD-1, 82
 Uhuru, 81
 WMAP, 396, 425, 437
 XMM, 81, 83
 Yohkoh, 81, 295
 Saturation, 67, 73
 Saturn, 142, 205
 density, 153, 205
 differentiation, 205
 flattening, 151
 helium abundance, 205
 magnetic field, 160
 moons, 207
 origin of rings, 206
 rings, 205
 rotation, 205
 winds, 205
 Scalar product, 466
 Scalar triple product, 467
 Scale, of a telescope, 54
 Scale factor, 428, 430, 440
 Scale height, 156
 Scattering, 53, 105, 328
 Scattering efficiency factor, 329
 Scattering plane, 165
 Schechter's luminosity function, 393
 Scheiner, Christoph, 285
 Schiaparelli, G., 196
 Schmidt, Bernhard, 59
 Schmidt, Maarten, 411
 Schmidt camera, 60
 Schönberg, 315
 Schwabe, Samuel Heinrich, 291
 Schwarzschild metric, 475
 Schwarzschild radius, 320
 Scintillation, 51
 Scintillation detector, 80
 Scorpius, 334
 Sculptor Galaxy, 391
 SDSS, 96, 438
 Seasons, 146
 Second, 38
 Seeing, 51, 55
 Seismic waves, 152
 Selection effect, 236
 Selection rules, 110
 Semi-diurnal tide, 175
 Semilatus rectum, 464
 Semimajor axis, 127, 464
 Semiminor axis, 464
 Semiregular variable, 300, 301
 SERENDIP, 456
 SETI, 455
 Setiathome, 456
 Seyfert, Carl, 410
 Seyfert galaxy, 410
 Sgr A, 383
 Sgr A*, 383
 Shane, W.W., 381
 Shapley, Harlow, 367
 Shell star, 234
 Shoemaker-Levy 9, 220, 222
 Short period comets, 220
 Shu, Frank H., 382
 SI units, 130, 477
 Sidereal month, 147
 Sidereal period, 145
 Sidereal year, 146
 Siding Spring, 34
 Silicates, 337
 Silicon burning, 257, 270
 Sine formula, 14
 Sine-cosine formula, 14
 Sirius, 94, 242, 243
 Sirius B, 236, 243, 313, 314
 Slit spectrograph, 227
 Sloan Digital Sky Survey, 96
 Small circle, 11
 Small Magellanic Cloud, 302, 391, 405
 Smithsonian Astrophysical Observatory, *see* SAO
 SN 1987A, 311, 312
 SNO, 87, 285
 Snowball Earth, 449
 Socorro, 79
 Sodium
 interstellar, 337, 339
 Soft gamma repeaters, 318
 SOHO, 293, 296
 Solar activity, 290
 Solar constant, 100
 Solar eclipse, 149, 286, 287, 289
 annular, 150
 partial, 150
 total, 150
 Solar system
 origin, 167
 Solar wind, 142, 157, 224

- Solid angle, 463
Sonoran desert, 53
Source function, 117, 342
Space
 elliptical, 428
 Euclidean, 428
 flat, 428
 hyperbolic, 428
 Minkowski, 428
Space, curved, 474, 475
Space weather, 297
Spectral class, 229, 235
Spectral classification
 Harvard, 229
 Yerkes (MKK), 232
Spectral line, 111, 227
 natural width, 107
Spectrograph, 73
 objective prism, 73
 slit, 73
Spectrum, 30
 absorption, 103, 104
 blackbody, 112
 continuous, 111, 227
 emission, 103, 104, 234, 286
 luminosity effects, 232, 233
 molecular, 111
 peculiar, 234
 stellar, 227, 228
Sphere, 463
Spherical aberration, 59
Spherical astronomy, 11
Spherical excess, 12
Spherical triangle, 11
Spherical trigonometry, 11
Spicule, 287
Spin, 109, 110
Spitzer, Lyman, 353
Spitzer Space Telescope, 414
Spörer minimum, 291
Spring tide, 149
SS Cygni, 306
SSS, 307
Standard models, of the Universe, 430
Star
 supersoft, 307
Star catalogue, 31
Star map, 34
Stark effect, 111, 232
Starquake, 317
Stars
 binary, *see* Binary stars
 brightest, 489
 central temperature, 258
 chemical composition, 238
 densities, 238
 diameters, 237
 distances, 27, 302
 energy production, 254
 evolution, 263, 274
 luminosity, 113, 238
 masses, 237, 241, 246
 motion, 374
 nearest, 488
 number density, 372
 populations, 258
 radius, 113, 236, 237
 rotation, 238
 structure, 249
 temperature, 113, 238
 variable, *see* Variables
Static limit, 321
Stationary points, 144
Stefan-Boltzmann constant, 112, 478
Stefan-Boltzmann law, 112, 114, 166
Stellar model, 257
Stellar statistics, 371
 fundamental equation, 373
Stellar wind, 267, 303
Steradian, 463
Stone-iron meteorites, 223
Stony meteorites, 223
Stratosphere, 191
Stromatolite, 451
Strömgren sphere, 345
Strömgren's four-colour system, 95
Subduction zone, 190
Subdwarf, 236
Subgiant, 232, 236, 269
Sudbury neutrino detector, 87
Sudbury neutrino observatory, 285
Sun, 141, 263, 283, 480
 activity, 290
 at zenith, 146
 atmosphere, 286
 coordinates, 36
 declination, 146
 density, 283
 energy production, 283, 297
 evolution, 277
 gravitational acceleration, 260
 luminosity, 283
 magnitude, 283
 mass, 283
 mean molecular weight, 261
 motion, 368–370
 orbital period, 378
 pressure, 283
 radio emission, 295
 radius, 283
 rotation, 283, 285
 spectral class, 283
 temperature, 261, 283
 UV radiation, 296
 X-rays, 296
Sundman, Karl F., 131
Sunspot number, 290, 291
Sunspots, 285, 290, 292
Sunyayev-Zel'dovich effect, 409
Supercluster, 405, 409

- Supergiant, 232, 235, 236, 302, 303, 348
 Supergranulation, 286
 Superior conjunction, 145
 Superior planet, 144
 Superluminal motion, 411
 Supernova, 169, 273–276, 280, 304, 308, 355, 423, 424
 - core collapse, 311
 - thermonuclear, 311
 - type I, 273, 276, 309, 318, 431
 - type Ia, 309
 - type Ib, 310
 - type Ic, 310
 - type II, 309, 310
 Supernova remnant, 309, 350
 Surface brightness, 93, 100, 114, 421
 Surface gravity, 232
 Surface velocity, 129
 Symbiotic stars, 307
 Symmetry breaking, 433
 Synchronous rotation, 147
 Synchrotron radiation, 117, 351, 352, 399
 Synodic month, 147
 Synodic period, 145
 Synthetic aperture radar, 186
- T**
- T Tauri stage, 170, 173
 TAI, *see* Time, atomic
 Tangential velocity, 28, 31
 Tautenburg, 60
 Taylor, Joseph, 317
 Taylor series, 465
 TDB, *see* Time, barycentric dynamical
 TDT, *see* Time, terrestrial dynamical
 Technetium, 278
 Telescope
 - ALMA, 77, 79
 - Arecibo, 75
 - Cambridge array, 79
 - Cassegrain, 84
 - catadioptric, 60
 - Effelsberg, 75
 - Green Bank, 75
 - infrared, 84
 - IRAM, 77
 - Jodrell Bank, 75
 - Keck, 62, 74
 - Maksutov, 60
 - MERLIN, 79
 - Newton, 57
 - Nobeyama, 76
 - Ritchey–Chrétien, 60
 - Schmidt, 60
 - Schmidt–Cassegrain, 60
 - VLA, 79
 - VLT, 63, 74
 - Westerbork array, 79
 Telescope mounting, 60
 Telescopes
 - millimetre and submillimetre, 498
 optical, 497
 radio, 497
 Temperature, 114
 - antenna, 114, 341
 - brightness, 114, 115, 341
 - colour, 115, 120
 - effective, 113, 114, 119, 232, 235
 - excitation, 110, 116, 342
 - ionisation, 116
 - kinetic, 115, 136
 - spin, 342
 Temperature gradient in stars, 250
 Terrestrial planets, 142
 - core, 152
 - mantle, 152
 Tethys, 207
 Thermal equilibrium, 167
 Thermal radiation, 112, 166
 Thermodynamic equilibrium, 116, 117
 Thermosphere, 191
 Three-body problem, 130, 131, 217
 Threshold temperature, 433
 Tidal force, 321
 Tidal heating, 454
 Tides, 149
 Time
 - apparent sidereal, 35, 38
 - atomic, 38
 - barycentric dynamical, 39
 - coordinated universal, 39
 - daylight saving, 36
 - dynamical, 38, 39
 - ephemeris, 38, 39
 - Greenwich mean sidereal, 38
 - mean, 36
 - mean sidereal, 35, 38
 - mean solar, 36
 - SI unit, 38
 - sidereal, 18, 19, 34, 38, 48
 - solar, 19, 35, 38
 - terrestrial, 39
 - terrestrial dynamical, 39
 - units, 479
 - universal, 36, 38
 - zonal, 36, 39
 Time of perihelion, 126, 127
 Time scale
 - dynamical, 264, 271
 - nuclear, 263, 267
 - thermal, 263, 265
 Time zone, 36
 Titan, 207, 208, 454
 Titania, 210
 Titius–Bode law, 168, 215
 TM, *see* Time, mean solar
 TNO, 215
 Tombaugh, C., 213
 Trans-Neptunian object, 215
 Transit, 19, 150, 460

- Trapezium, 342, 348
Triple alpha process, 256, 269, 274
Triton, 212
Trojan asteroids, 131, 217
Tropic of Cancer, 146
Tropic of Capricorn, 146
Tropical year, 146
Tropopause, 191
Troposphere, 191
True anomaly, 125, 126
Trumpler, Robert, 327
TT, *see* Time, terrestrial
T Tauri star, 169, 266, 276, 300, 304
Tully–Fisher relation, 393, 424
Tunnelling, 440
Tycho Brahe, 3
Tycho Brahe's supernova, 299, 309
Tycho catalogue, 32
- U**
U.S. Naval Observatory, 62
UBV system, 95
UBVRI passbands, 95
UBVRI system, 95
Ultraviolet catastrophe, 113
Ultraviolet radiation, 52, 82
Umbra, 290
Umbriel, 210
Unit matrix, 468
Unit vector, 466
Universal time, *see* Time, universal
Universe
 age, 424, 425, 432
 expansion, 422, 424
 flat, 433
 future, 439
 geometry, 428
 homogeneous, 427, 433
 isotropic, 426, 427, 433
 scale, 427
 static, 429
 structures, 435
Upper conjunction, 145
Upper culmination, 19
Uranometria, 31
Uranometria Nova, 34
Uranus, 142, 208
 discovery, 208
 magnetic field, 160, 210
 rings, 209, 210
 rotation, 208
URCA process, 315
Urey, Harold, 450
USNO-B1.0 Catalog, 34
UT, *see* Time, universal
UT0, 38
UT1, 38
UTC, *see* Time, coordinated universal
- UV Ceti stars, 304
Uvby system, 95
- V**
V1057 Cygni, 304, 305
V1357 Cyg, 322
V1500 Cygni, 308
Väisälä, Yrjö, 60
Valles Marineris, 196, 197
Van Allen's belts, 158
Van de Hulst, Hendrik, 341
Van den Bergh, Sidney, 333, 393
Van der Waals force, 170
Variable stars, *see* Variables
Variables, 299
 classification, 299
 eclipsing, 300
 eruptive, 300, 303, 304
 irregular, 303
 names, 299
 pulsating, 300, 301
 rotating, 301
 semiregular, 303
Vector, 465
Vector product, 466
Vector triple product, 467
Vega, 23, 96, 230, 369
Veil nebula, 351, 352
Vela pulsar, 317, 355
Velocity
 radial, 362
 tangential, 362
Velocity of light, 478
Venera, 186
Venus, 142, 185
 atmosphere, 185, 188
 atmospheric pressure, 185
 clouds, 188
 greenhouse effect, 167
 impact craters, 187
 lava flows, 187
 phases, 185
 rotation, 185
 surface, 186, 187
 temperature, 167, 185
 transit, 151
 volcanoes, 186
Venus Express, 186
Vernal equinox, 17, 21, 23, 34, 35, 126, 146
Vertical, 17
Vibrational transitions, 347
Vidicon camera, 70
Viking, 199
Virgo Cluster, 407, 408, 424
Virial, 134
Virial theorem, 134, 394
VLBI, 80
Vogt–Russell theorem, 252
Voigt profile, 108

-
- Voyager 1, 206
 Voyager 2, 206
- W**
 Warrawoona Group, 451
 Wavelength, 28, 428
 Weber cylinder, 87
 Weird Terrain, 184
 Weizsäcker, Carl Friedrich von, 254
 Whipple, F., 220
 White dwarf, 235, 236, 269–271, 274, 313
 radius, 313, 319
 Wien approximation, 113, 115
 Wien displacement constant, 113
 Wien displacement law, 113, 118
 Wilson, Robert, 425
 Winter solstice, 146
 Wolf, Charles, 234
 Wolf diagram, 332, 333
 Wolf–Rayet star, 229, 234, 270, 273, 275–277, 310
 W UMa stars, 245
 W Virginis star, 300, 302, 375
- X**
 X-ray burster, 323, 324
 X-ray pulsar, 322, 324
- X-rays, 81, 82, 408
 solar, 295, 296
 Xena, 214
 XUV, 82
- Y**
 Year, 40, 479
 anomalistic, 146
 sidereal, 36, 130, 146
 tropical, 36, 146
 Yerkes Observatory, 57
- Z**
 ζ Persei, 360, 361
 ZAMS, 258
 Zeeman effect, 109, 234, 291, 356
 Zenith, 16
 Zenith angle, 174
 Zenith distance, 17, 26, 99
 Zero age main sequence, *see* ZAMS
 Zero-point energy, 433
 Zodiacal light, 224
 Zone catalogue, 32
 Zone of avoidance, 331