OPTION A

Astrophysics and Cosmology

Preface

During the last thirty or forty years, Astrophysics and Cosmology have undergone very significant and exciting changes. A notable symbol of this fact was the launch of the Hubble Space Telescope. The repair of the telescope's mirror, in space, was itself spectacular. Having now been repaired, the telescope is functioning superbly, producing data that both astonishes and confounds astronomers and cosmologists. It is self-evident that, as an option within a complete A level syllabus, it is only possible to make a very brief and selective incursion into the field of Astrophysics and Cosmology. However, in order to avoid the possibility that these Notes would appear unduly terse and turgid, material that directly relates to the syllabus for the option has been embedded into other contextual material. Students, and teachers, will need to bear this fact in mind when using this text so as to recognise, and possibly pay greater attention to, that material which is specifically included to 'cover' the syllabus.

A1. Contents and Scale of the Universe

1 *(a)* *Candidates should be able to describe the principal contents of the Universe, including stars, galaxies and radiation.*

As discussed more fully later (section *2(j)* p. 41), the currently favoured theory for the origin of the Universe is that of the hot big bang. The essential premise of the big bang theory is that it is believed that some 15 to 20 billion (10^9) years ago, space and time came into being – with the energy content of the Universe being at an exceedingly high temperature and exceedingly high pressure. From its initial creation under these conditions, the Universe began to expand (and has been expanding ever since, though not at a constant rate). At the early stages of the expansion of the Universe, radiation 'dominated' the Universe, the ratio of photons to nucleons being of the order of 10^9 to 1. The early Universe expanded and, as it expanded, it cooled. This expansion and cooling still continues. Relatively shortly after the big bang, atomic nuclei, particularly of hydrogen and helium, formed. Somewhat later, the Universe had cooled to a sufficient extent for atoms of these elements to form (and then molecules of hydrogen). With electrons being bound within atoms, the interactions between electrons and photons of radiation lessened to such a degree that radiation and matter are said to have 'decoupled' and the Universe became transparent.

It is a success of the big bang theory that it is able to explain why the observed ratio of hydrogen atoms to helium atoms formed as a consequence of the big bang should be 3:1, with a lower proportion of deuterium ($^2_1$H) and a much lower proportion of lithium of (slightly) greater proton number also being formed. Another linch-pin for the big bang theory has been the observation of remarkably uniform background radiation which accurately matches black-body radiation corresponding to a temperature of about 3 K. The existence of such radiation is what would be expected on the basis of a very hot Universe that has been expanding and cooling for some 15 to 20 billion years.

For this section, the history of the evolution of the Universe might conveniently be taken from the decoupling of radiation and matter. At this period, the Universe may be regarded, in simple terms, as consisting of hydrogen, of helium, (in the atomic ratio 3:1), of a much smaller proportion of deuterium and lithium, and of radiation. It is from this basic material that the Universe has evolved into its present state.

From a human point of view, the history of Cosmology might, arguably, date from the recognition of some basic observations. These observations are so familiar that it is easy to forget that they *are* observations and to underestimate their significance. These basic observation include:

(i) the existence of the Sun and its apparent motion across the sky;

(ii) the existence of the Moon, its phases and its motion across the sky;

(iii) the existence of stars, their differing seasonal visibility and their apparent motion across the sky.

The Sun, the Earth and the Moon would be sufficient to constitute *a* solar system but other bodies have long been recognised as belonging to *the* Solar System, i.e. the planets (so-called from the Greek for 'wanderer'). Whether the Solar System is to be regarded as a principal member of the Universe is a rather philosophical question. There is, however, reason to believe that the existence of a planetary system around a star might not be uncommon and there is some observational evidence that relatively near, but otherwise 'ordinary', stars have planet-sized bodies associated with them. If this is so, then the assumption is that there must, in all probability, be a great many other stars with planetary systems around them. The possibility of life existing elsewhere in the Universe cannot then be readily discounted.

Stars are not all alike. Their differences include differences in age, in mass and in brightness, both intrinsically and because of differing distances from the Earth. Their differences in colour are discernible to the naked eye, from the blue-white of Sirius (the brightest star in the sky) to the red of the giant star Betelgeuse. Stars are not necessarily of constant brightness, e.g. Mira: they may have bright nebulosity* associated with them, e.g. the nebula* in the Sword of Orion, the Pleiades.

These differing properties of individual stars are of significance but there is another fact that is also particularly noteworthy. The Sun happens to be a single star but this is by no means true of stars in general. Of the few thousand stars visible to the naked eye, about 20% are double stars: of the 250 stars within 10 pc (see sections *1(c)* and *(d)*) of the Sun, more than half occur as double or more highly multiple stars. [A 'double star' is a system in which two stars exist as a relatively close pair of stars gravitationally bound to each other and orbiting each other round their common centre of mass.] It is possible that most stars are formed in groups**, e.g. the Pleiades, Praesepe, which, over time, disperse.

Attention is drawn to three other types of collections of stars, visible to the naked eye, provided that the night sky is suitably dark.

*The word 'nebula' comes from the Greek, meaning 'cloud'. In Astronomy, the term nebula was historically used to describe regions of the night sky from which diffuse and undifferentiated light could be observed. There are several causes of such nebulosity and these terms are still used even when they are no longer strictly applicable. For example, the Andromeda Nebula has now been (at least, partly) resolved into stars and this object is now recognised as being a galaxy in its own right, similar to, but rather larger than, the Milky Way Galaxy. Similarly, globular clusters which appeared nebulous in early telescopes are recognised as aggregates of millions of stars. On the other hand, comets are commonly diffuse in appearance, and some regions of gas and/or dust may glow due to the existence within them of bright stars, e.g. the two objects quoted above.

**These groups must not be confused with constellations which are, in effect, merely arbitrarily defined areas of the sky.

The first type is the bright nebula in Hercules, known as M13, being the thirteenth in the list of nebulous objects compiled by the 18th century French comet-hunter Messier. He produced his list because he was anxious to avoid the risk that he and other comet-hunters would be misled into thinking that he and they had discovered a new comet. This object is a so-called globular cluster, estimated to contain at least 10^5 stars.

The second of these collections visible to the naked eye is the Milky Way, which can be seen as a broad, faint band – but brighter than the rest of the night sky – stretching right across the sky. This band is made up of the stars situated in the disc of the Milky Way Galaxy. This Galaxy (in which the Sun and the Earth are situated) is a typical spiral galaxy and it contains some 2×10^{11} stars, i.e. 200 thousand million, or 200 billion, stars.

Another spiral galaxy which is just visible to the naked eye is the nebula in the constellation Andromeda, known as M31. This spiral galaxy – the nearest to the Earth at a distance of some 2×10^6 light years (see sections *1(c)* and *(d)*) – is similar to the Milky Way Galaxy but somewhat larger. The Milky Way Galaxy and M31 are both members of the Local Group of galaxies.

The third type of collection is only visible from the southern hemisphere. There are two examples, namely the Large and the Small Magellanic Clouds. They are irregular galaxies. Not only are they gravitationally bound to the Milky Way Galaxy, they are physically linked by streamers of hydrogen gas. M31 also has companion galaxies.

Just as there are various types of star, there are also various types of galaxy. Two of these types have already been mentioned: spiral and irregular galaxies. Even these delineations are broad. For example, spirals vary in size and in the openness of their spiral arms: some spirals are ‘barred’: they differ in the degree of the activity of their central regions and in the intensity of emission of radio waves.

Irregular galaxies are, likewise, not a homogeneous set. Some, such as the Magellanic Clouds, are intrinsically irregular whereas others arise as a consequence of ‘collisions’ between galaxies.

The elliptical galaxies constitute a third type of galaxy. As in the case of spirals, ellipticals vary in size, the largest elliptical being larger than the largest spiral. Similarly, the smallest ellipticals are smaller than the smallest spirals. It is thought that small ellipticals may be the commonest type of galaxy. However, this is difficult to establish because small ellipticals have a low surface brightness and may escape detection.

Fig. 1.1 illustrates these various types of galaxy.

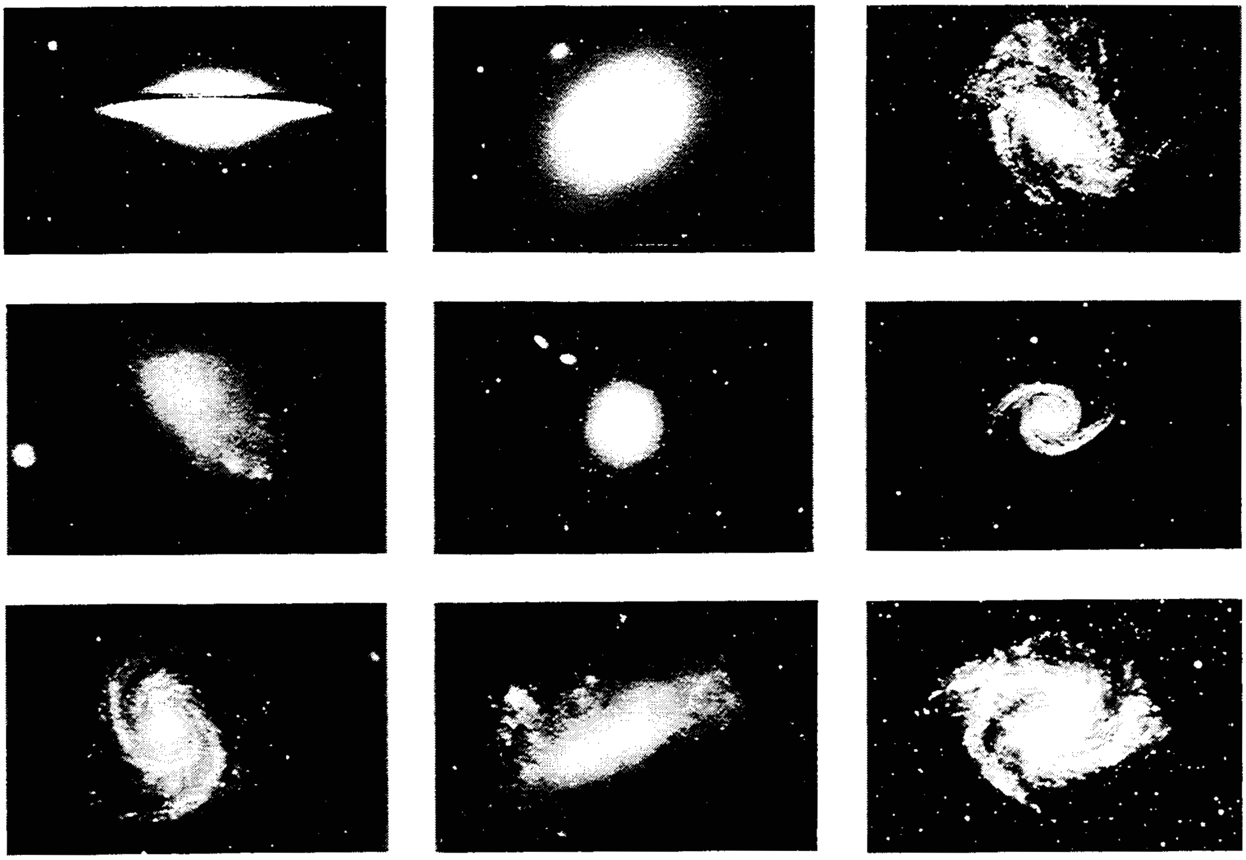


Fig. 1.1

Apart from shape, there is another significant difference between elliptical and spiral galaxies. The former contain little, if any, gas and dust so that star formation has effectively ceased. On the other hand, spiral galaxies contain significant amounts of gas and dust. Accordingly, spiral galaxies are characterised by continuing star formation. That there is dust in such galaxies is well illustrated by the Sombrero Galaxy. It is also evident in the fact that dust clouds obscure the stars lying behind them. It is, indeed, dust that causes the discernible dark patches in the Milky Way. Other well known dark clouds are the 'Horse Head Nebula' in Orion and, for observers in the southern hemisphere, the Coalsack in the Southern Cross.

It is within clouds of hydrogen, helium and, if present, dust within galaxies that new stars begin to form. The energy radiated from young stars may cause the outer reaches of the clouds surrounding them to glow. This explains the bright nebulosity associated with the nebula in the Sword of Orion, M42, and with the cluster of young stars, the Pleiades, M45.

The hydrogen and dust distribution within a spiral galaxy is typically of less thickness than the disc of stars: for example, in the Milky Way Galaxy, the radius of the disc is about 15 kpc, the stellar disc is a few hundreds of parsecs thick and the hydrogen/dust disc is about 10 times less thick (see also sections *1(c)* and *(d)*). The dimensions of the central bulge, which is almost spherical, are of the order of 1 or 2 kpc.

It must not be thought that hydrogen and helium only occur in the interstellar space within a galaxy. A typical galactic density of these elements is 10^5 atoms per cubic metre. The corresponding density in intergalactic space is of the order of a few atoms per cubic metre.

The difference between these densities is remarkable and that of intergalactic space may appear to be surprisingly low. However, it should be remembered that the distances that separate galaxies (both within and between clusters of galaxies) are *very* large. The total mass of hydrogen and helium in the Universe is *not* small!

So far, the discussion has concentrated on three types of the principal constituents of the Universe:

(i) hydrogen and helium,

(ii) stars,

(iii) galaxies.

There are three other constituents to be added to the above list:

(iv) the elements of proton number of up to 92, i.e. uranium,

(v) radiation,

(vi) ‘missing’ or ‘dark’ matter.

As previously mentioned, the proportions of elements beyond helium in the Periodic Table created during the early period after the big bang (section *2(j)*, p. 41) were small. The proportions of these elements remain small. While, therefore, these elements cannot strictly be regarded as principal constituents of the Universe, they are certainly significant constituents. If they had not been formed during the evolution of the Universe, there would be nobody to ask the question as to how they have been formed.

During the major part of their lifetime, all stars owe their existence to the nuclear fusion reaction(s) occurring at their centres. The most important process is the fusion of four hydrogen nuclei to form a helium nucleus. The mass of the latter is less than that of the former. The difference in mass appears as energy, including radiant energy. This is the process that maintains the Sun in its present state. It is also the process that initially maintains stars of significantly greater mass than the Sun. In such massive stars, other fusion reactions are initiated at their centres as they age, namely, from helium to other nuclei of greater proton number, such as carbon, oxygen, silicon and, finally, to iron. The formation of iron within the core of a massive star is, however, the ‘end of the road’ as far as a massive star is concerned. The iron nucleus is energetically the most stable of all elemental nuclei. Fusion of nuclei to form a nucleus yet heavier is energy-absorbing rather than energy-releasing. When a massive star can no longer maintain itself by the formation of iron nuclei in its innermost core, it ‘dies’ as an extremely violent explosion, i.e. a supernova explosion, as in the Crab nebula. It is during such an explosion that nuclei beyond iron are formed. These nuclei are dispersed into interstellar space. Eventually, a cloud of hydrogen and helium, now mixed with these ‘evolved’ elements, can form another star [or stars]. There are, in fact, two (and more) general categories of stars in the Universe. The first, and oldest, generation of stars is called population II stars and they contain few, if any, nuclei of high proton number. The Sun, on the other hand, is much richer in nuclei beyond helium and is typical of a population I star. The implications of this are discussed more fully later.

It might seem a simplistic question to ask how it is known that stars and galaxies exist but it is not as simplistic as all that. Certainly, the stars can be seen. As discussed above, so also can the galaxies – but compare, as an illustration, the professional photograph of M31 with that obtained by an amateur astronomer using a 35 mm camera (see Fig. 1.2).

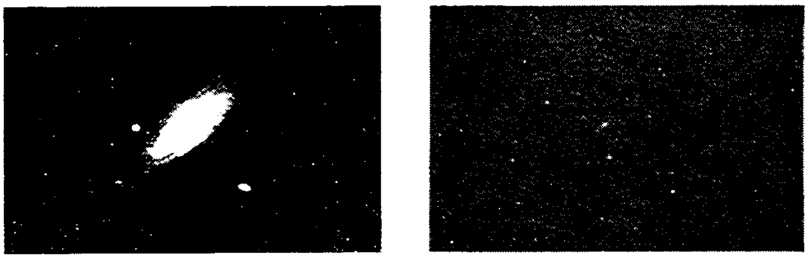
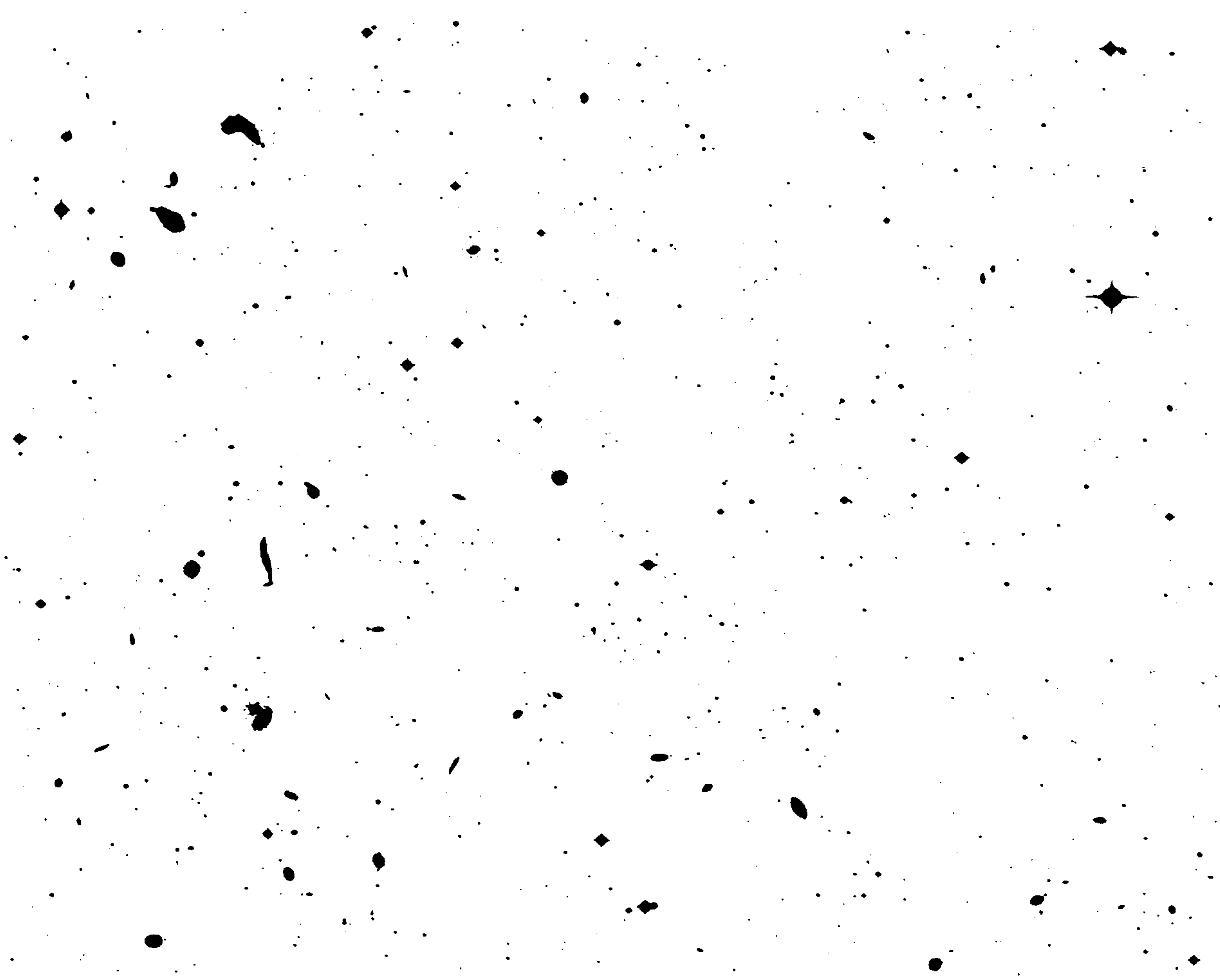


Fig. 1.2

As late as the 1920's, there was debate as to whether galaxies such as M31 were separate entities beyond the Milky Way Galaxy or whether they were bodies within the Galaxy. It was by the use of better telescopes in order to resolve M31, and other similar galaxies, into separate stars and by the use of sophisticated techniques – to determine distances – that the issue was resolved in favour of separate galaxies. It is now known that there are a great many millions of galaxies. It is thought that there are at least as many galaxies as there are stars in a typical galaxy. A cluster of galaxies is shown in Fig. 1.3 opposite.



Each ‘blob’ on this diagram (as against a point-like star image) represents a galaxy.

Fig. 1.3

Most of the knowledge about the Universe has been gained from the analysis of radiation received. Exceptions include knowledge from meteorites and samples of rock brought back from the Moon. However, a crucial factor in the great advances that have, in recent years, been made in astrophysics and cosmology has been the increase in the ranges of wavelengths used to obtain information about the Universe. This increase has been on both sides of the quite narrow range of the visible spectrum (400 nm to 700 nm). These double-sided increases are developments that have occurred over a time-span of some fifty years.

Radio-astronomy is Earth-based, the Earth’s atmosphere being effectively transparent at these longer wavelengths/lower frequencies.

Advances of knowledge on the other side of the visible spectrum have depended on rocket technology to take instruments above the Earth’s atmosphere. [Advances in rocketry have also, of course, been exploited in other ways, e.g. landing men on the Moon, sending probes to the planets, investigating Halley’s comet on its most recent return.]

It is now evident that processes occurring in the Universe are associated with wavelengths right across the electromagnetic spectrum, from the low energy 3 K background radiation to the highly energetic γ-rays. The information gained has revolutionised understanding of astrophysics and cosmology. A case in point is, indeed, information about hydrogen in the Universe. On the Earth, hydrogen exists as an invisible diatomic gas. In space, although it is cold enough for hydrogen to occur as a molecule, this is rather rare because of the low concentration and the fact that there is no mechanism for removing the bond energy released if two hydrogen atoms were to collide and combine; they merely dissociate again. Atomic hydrogen is invisible but observable at a wavelength of about 21 cm, depending on Doppler effects – see section *2(a)*.

One other point about radiation as a principal constituent of the Universe may be made here. As has previously been stated, the ratio of photons to nucleons at the time of the decoupling of radiation and matter was 10^9:1. This ratio has stayed constant since that time. Given that it is now estimated that there are about 400 photons per cubic centimetre, it follows that the estimated mean density of matter in the Universe is of the order of $10^{-26}\,\mathrm{kg\,m^{-3}}$ (note: the density of water is $10^3\,\mathrm{kg\,m^{-3}}$).

There is one other principal constituent of the Universe to be considered, namely 'missing' or 'dark' matter. The Doppler effect (see section *2(a)*) is also relevant in this context. The rotations of individual galaxies, the motions of galaxies within clusters and, indeed, the lifetime of such clusters, are all such that they cannot adequately be explained on the basis of the estimated masses of their visible components. The nature of this supposed 'missing matter' remains an unsolved problem, one which is being actively pursued because of its implications.

It may be appropriate to mention here another unsolved problem, namely, the mechanism by which the present distribution of galaxies came about from the primaeval mixture of hydrogen and helium.

1 (b) Candidates should be able to describe the Solar System in terms of the Sun, planets, planetary satellites and comets. Details of individual planets are not required.

It is not proposed to describe the structure of the Milky Way Galaxy in any great detail but rather to provide a frame in which to draw the picture of the Solar System.

The Milky Way Galaxy contains some 2×10^{11} stars. There is a central bulge in which the concentration of the stars is much greater than elsewhere (see Fig. 1.4). Elsewhere, most stars lie in the disc, concentrated in, but not confined to, the spiral arms. The radius of the disc is 15 000 pc and its thickness is of the order of a few hundred parsecs. Hydrogen is distributed in the spiral arms and between them. The dust disc is much less thick than the stellar disc. (See section *1(c)* for the definition of the distance units commonly used in astronomy.)

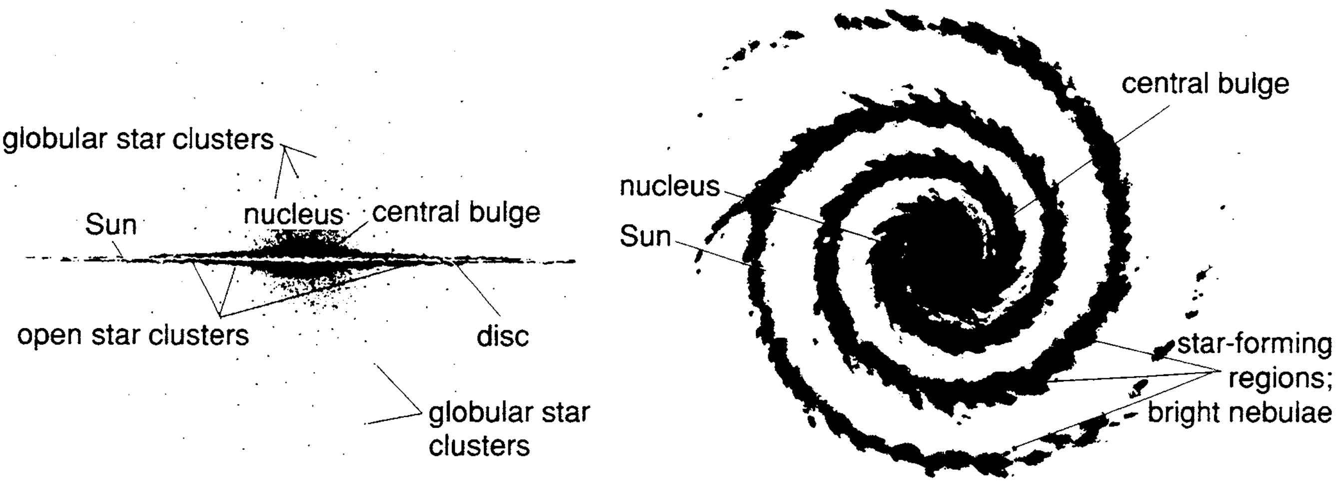


Fig. 1.4(a) and (b)

The Galaxy is, as a whole, rotating about its centre. This rotation, however, differs from that of a solid body. The Sun, situated in the so-called Perseus arm, lies about 10 000 pc from the Galactic centre (and about 15 pc north of the central Galactic plane). The Sun takes 2 x 10^8 years to make one complete revolution of the Galaxy: since it was formed, it has been round the Galaxy some 20 times. The orbital speed of a disc star depends on its distance from the centre of the Galaxy, but not in the same manner as the planets around the Sun. The main reason for this is that, for a given star (X, say), it is the total mass of those stars which are nearer than X to the centre of the Galaxy which provides the gravitational force responsible for the acceleration of X towards the centre of the Galaxy and, thus, the overall motion of X round the centre.

As well as the disc stars, there is a spherical distribution of so-called halo stars. These also orbit the Galactic centre but cross the central plane of the disc. [It may, incidentally, be noted that stellar number densities, i.e. the *number* of stars per unit volume, are so low that collisions between stars are very rare – more rare than collisions between galaxies!] As stated earlier, there are some 250 stars within 10 pc or 32 light-years of the Sun. This corresponds to a number density of 1 star in about 600 cubic light-years. As an approximation, this gives a typical interstellar distance of about 8.5 light-years, which makes Proxima Centauri, the star nearest to the Sun, rather close at 4.3 light-years. Overall, disc stars are, on average, a few light-years apart.

There is also a nearly spherical distribution of over 100 globular star clusters, the radius of this sphere being about 20 000 pc. Globular clusters are themselves commonly spherical and may contain between 50 thousand and 50 million stars. They are thought to be the oldest objects in the Galaxy.

There are some essentially dynamic similarities between the Milky Way Galaxy and the Solar System. There is a concentration of mass at the centre, i.e. the Sun in the Solar System. There is a decidedly narrow disc of bodies orbiting the centre, the sense of movement of these bodies being the same, i.e. the planets all orbit the Sun in an anticlockwise direction when viewed from above the Earth's North pole.

On the other hand, there are significant differences. The Sun is the dominant factor in determining the planetary orbits and the planets have only a minor gravitational effect on each other. The planetary orbits have a relatively simple pattern – well described by

property	Sun	Mercury	Venus	Earth	Mars	Jupiter	Saturn	Uranus	Neptune	Pluto
mass /kg	1.99×10^{30}	3.30×10^{23}	4.87×10^{24}	5.97×10^{24}	6.42×10^{23}	1.90×10^{27}	5.69×10^{26}	8.66×10^{25}	1.03×10^{26}	† (10^{22})
mass relative to Earth	3.33×10^{5}	0.056	0.815	1.000	0.107	318	95.1	14.5	17.3	(2.2×10^{-3})
mean radius of orbit $/10^{9}$ m	—	58	108	150	228	778	1430	2870	4500	5900
mean radius of orbit relative to Earth	—	0.387	0.723	1.000	1.524	5.203	9.539	19.18	30.06	39.44
density /g cm^{-3}	1.41	5.50	5.25	5.52	3.94	1.33	0.71	1.70	1.77	5.5
orbital period /days	—	88	225	365	687	4330	1.07×10^{4}	3.07×10^{4}	6.02×10^{4}	9.05×10^{4}
orbital period relative to Earth	—	0.241	0.615	1.000	1.881	11.86	29.46	84.01	164.8	247.7
** eccentricity of orbit	—	0.206	0.007	0.017	0.093	0.048	0.056	0.047	0.009	0.250
tilt of plane of orbit to ecliptic	—	7.00°	3.39°	—	1.85°	1.31°	2.49°	0.77°	1.77°	17.14°
tilt ¤	7° (⊥ʳ to ecliptic)	0°	177°	23.5°	25°	3°	27°	98°	29°	122°
period of rotation	‡ 25.4 d	58 d	* 243 d	24 h	24.6 h	10 h	10 h	11 h	16 h	6.4 d
number of satellites	—	0	0	1	2	16	17	15	8	1

Fig. 1.5

Newton's laws. (Those planets with their own satellites – or moons – are, in a sense, 'mini-solar systems'.) None of the bodies orbiting the Sun are self-luminous.

Of the 10^{11} stars in the Galaxy, the Sun is no more than a typical star of quite moderate mass. It is thought to be about 5×10^9 years old, with about the same length of time to go before it runs out of fuel. It will then expand into a red giant of a size greater than the Earth-Sun distance, collapse into a white dwarf and eventually radiate all its heat away. At present, it is a yellowish star with a surface temperature of about 6000 K. Other basic physical facts about the Sun, its satellites and their satellites are, for reference only, given in Fig. 1.5. It should be noted that time intervals quoted in the table are in terms of Earth-based periods, e.g. year, day, hour, etc.

The discussion below draws attention to some of the more notable features of the Solar System.

The Sun is much the most massive body in the Solar System, i.e. 1000 times more massive than Jupiter, the largest planet, which itself is 320 times more massive than the Earth. Conversely, the Sun has relatively little angular momentum compared with its satellites.

As well as radiation, the Sun also emits charged particles. This emission is known as the solar wind. The Earth's magnetic field tends to deflect these particles towards the Earth's North and South magnetic poles. Interaction of the solar wind particles with the Earth's atmosphere gives rise to the Aurora Borealis and the Aurora Australis, the so-called Northern and Southern Lights. The intensity of the aurora is variable and the variations are associated with the sunspots (magnetic disturbances) on the Sun's surface. The number of sunspots varies over an 11-year period (or 22 years, if the magnetic polarities of the sunspots is taken into account). Radio-wave propagation on the Earth can be affected by this and other types of activity on the Sun.

Between Mars and Jupiter lies the asteroid belt containing many thousands of rocky bodies of mass up to 10^{21} kg and dimensions up to 700 km (most being very much smaller). The brightest are visible by using binoculars and their paths can readily be traced by short exposure photographs, using a 35 mm camera, say, with the photographs taken at intervals of a few days. Although most asteroids lie between 2 AU and 3 AU from the Sun, there are some asteroids with markedly anomalous orbits, i.e. in terms of how inclined their orbits are to the ecliptic, how close they come to the Sun and how eccentric their elliptical orbits are.

As already indicated, the orbits of the planets lie in a disc, the inclinations of the orbits to the plane of the Earth's Orbit, called the ecliptic, being only a few degrees – except for Pluto, with an inclination of nearly 20°.

Notes

* Venus' rotation about its own axis is retrograde, i.e. in the opposite direction to the other planets.

** Eccentricity is a measure of how far a planet's elliptical orbit differs from being circular. A circular orbit has an eccentricity of zero: the maximum value of eccentricity is 1.

† Some of the properties of Pluto are rather uncertain.

‡ d = day; h = hour.

¤ Angle of equator of planet to plane of the orbit.

The planetary orbits are essentially elliptical but, with two exceptions, differ little from being circular. The Sun lies at one of the foci of the relevant ellipse. Pluto's orbit has the highest eccentricity, i.e. departure from circularity, and the departure is such that for part of its orbit, Pluto is closer to the Sun than is Neptune. There is, in fact, some doubt about whether it is correct to describe Pluto as being a planet rather than its being a satellite of Neptune that has had its orbit about that latter body disrupted.

Mercury's orbit is only slightly less elliptical than Pluto's. In fact, no planetary orbit is strictly elliptical about the Sun. Under the influence of the other planets, the orbit of any one planet precesses, so that the orbit traces a rosette pattern about the Sun (see Fig. 1.6).

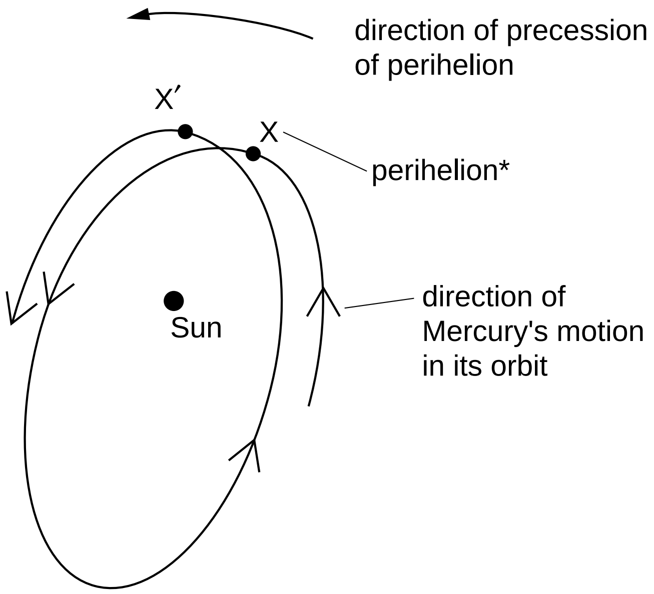


Fig. 1.6

Notes

For clarity:

(i) the ellipticity of Mercury's orbit,

(ii) the shift of the perihelion, i.e. X to X′, per orbit,

are each greatly exaggerated.

* Perihelion is the point of closest approach of a planet to the Sun.

The precession of Mercury's orbit is of particular interest. The observed precession is larger than that predicted from Newton's law of Gravitation. It was a successful test of Einstein's theory of General Relativity that it effectively solved this discrepancy.

A feature of the planets beyond the Earth is their occasional apparent retrograde motion across the sky. Usually, as a planet proceeds along its solar orbit in an anti-clockwise direction (anti-clockwise, that is, if the orbit is viewed from a position that is to the North of the general plane of planetary orbits), it is observed from the Earth to make apparently slow advance across the night sky from West to East. At certain times, however, this easterly motion appears to slow down, come to a stop and the planet then appears to move westerly across the sky. It is this westerly motion that is called ‘retrograde’. After some time, the retrograde motion slows down, stops and then the planet resumes its more orthodox easterly motion, as illustrated in Fig. 1.7(a) and (b).

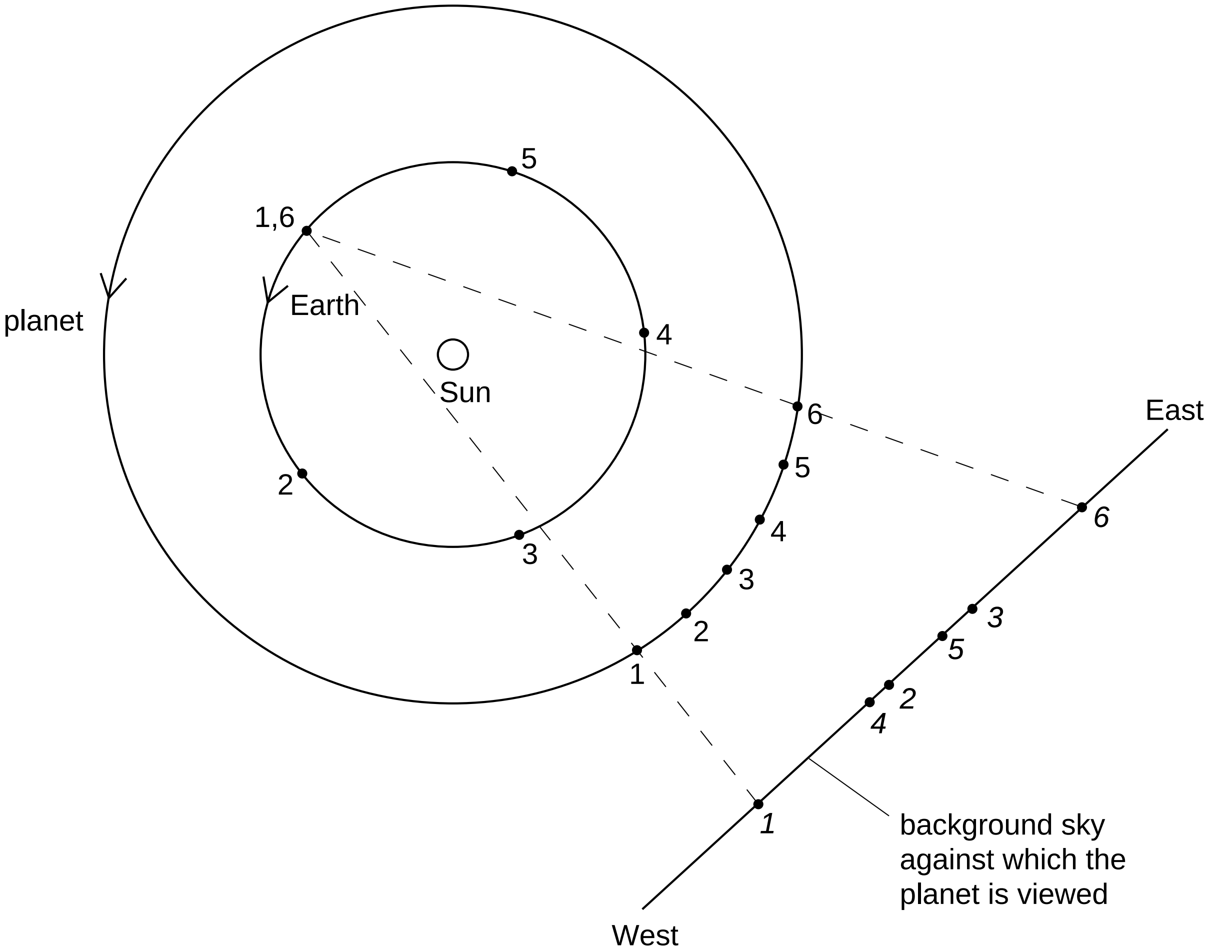


Fig. 1.7(a)

Notes

Points 1 to 6 on the two orbits represent successive (and simultaneous) positions of the planet and the Earth. Points *1* to *6* on the line which represents the ‘background sky’ illustrate the apparent positions of the planet against the night sky.

The relative sizes of the Earth’s orbit and the planet’s orbit are not to scale. Similarly, the relative motions of these two bodies are not to scale. The diagram is purely illustrative.

It should also be realised that the two orbits are not co-planar. Instead of the planet’s retrograde motion appearing to be a linear west to east then east to west oscillation, the motion actually appears as a ‘loop’, see Fig. 1.7(b) below.

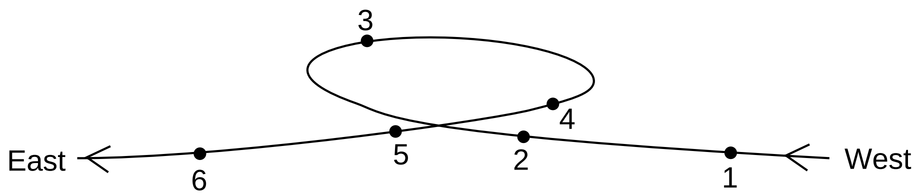


Fig. 1.7(b)

Being next nearest to the Earth as an ‘outer planet’, Mars shows the most pronounced retrograde motion. Whether considered in terms of linear or angular speed, Mars travels more slowly than the Earth. As a result, the Earth completes more than one of its orbits in the time that Mars completes an orbit. In effect, therefore, the Earth sometimes overtakes Mars ‘on the inside’. It is during the overtaking process that the apparent retrograde motion occurs.

The other outer planets also show retrograde motion, with the frequency of its occurrence, its extent (as measured angularly on the sky) and the length of duration depending on the relationship between the parameters of the Earth’s orbit and those of the planet being considered.

As previously mentioned, there are some asteroids with orbits that differ markedly from the commonality of asteroidal orbits. These ‘maverick’ asteroids may have orbits that are more highly elliptic, more inclined to the ecliptic and have a perihelion* much less than average. An example of this latter is the asteroid Icarus, so-named because its closest approach to the Sun lies within the orbit of Mercury. (It is possible that this asteroid may one day suffer its mythological namesake’s fate!) The orbital elements of these unusual asteroids have some similarities with those of the short-period comets (named from the Greek for ‘hairy star’).

The origin and nature of comets is not fully understood. One theory is that comets were formed at the same time as the Solar System but in a cloud some 50 (or more) AU – see section *1(c)* – from the Sun, the so-called Oort Cloud. On this theory, the fact that ‘new’ comets are continually being discovered is explained by supposing that comets in the Oort Cloud are dislodged into orbits about the Sun, thus allowing them to be observed. An alternative theory is that they are formed by ‘gravitational focussing’ by the Sun as, during its Galactic motion, it sometimes passes through a gas/dust cloud. A supporting piece of evidence for this theory is the disposition of cometary orbits.

Two general categories of comet are recognised – the long-period and the short-period comets. The somewhat arbitrary dividing criterion is a period of rotation about the Sun of longer than, or shorter than, 200 years. The shortest period is a few years but the longest period may be measured in thousands of years or even longer. Bearing in mind the eccentricity of their orbits, the implication of the arbitrary 200 years criterion is that the short-period comets spend most of their time within planetary distances of the Sun whereas long-period comets spend most of their time in the distant reaches of the Solar System.

Comets are relatively short-lived. A hundred or so short-period comets are known and 500 or more of the long-period category. In essence, a short-period comet is an ‘evolved’ remnant of a long-period comet. One piece of evidence for this is that the majority of short-period comets (Halley’s comet is an exception) travel in their orbits in the same sense as the planets: this is thought to occur because of the perturbing influence of the planets, Jupiter especially, on the comets. That comets are subject to such influences is well illustrated by the destruction, in 1994, of the comet Shoemaker-Levy by its collision with Jupiter. As a result of perturbations, some cometary orbits are changed into hyperbolic or parabolic paths, with the result that a comet so affected never returns to the Solar System. Only a minority of comets have elliptical orbits such that they repeatedly come back to the vicinity of the Sun. For each of these various ‘conic section’ orbits, the Sun is a focus.

*Perihelion is the term given to the closest approach of a Solar System body to the Sun.

Comets are sometimes described as 'dirty snowballs'. They are thought to have a rocky nucleus (or nuclei) within a dusty matrix of frozen volatile compounds such as water, methane and carbon dioxide. As for other Solar System 'junior' members, comets shine by reflected sunlight. They brighten as they approach the Sun, not only because of inverse law effects but also because the frozen material tends to melt and evaporate, forming a tenuous coma much larger than the nucleus. The most spectacular feature of a comet, not shown by all comets, is the formation of a tail as it comes sufficiently close to the Sun (see Fig. 1.8).

Fig. 1.8

When a tail does form, it always points directly away from the Sun. Two types of tail may form. Radiation pressure on the coma is the cause of one type of tail. For the other type, molecules within the coma may become ionised and these are then repelled by the charged particles present in the solar wind through which the comet is passing.

A consequence of the formation of the coma – and tail(s), if any – is that by close passage around the Sun, a comet loses material and thus eventually decays. In such decay, the nucleus may break up and separate, as in the case of Comet Shoemaker-Levy. Dusty particles released from the matrix surrounding the nucleus may become spread out along the cometary orbit. In certain cases, the Earth may pass through this debris – sometimes on an annual basis. During such a passage, there is an enhanced display of meteors. The dusty particles become incandescent by friction with the Earth's atmosphere and burn out. An example of such a cometary meteor shower is the Perseids which occur annually in mid-August, the constellation Perseus being the region of the sky from which the meteors appear to spread out.

1 (c) *Candidates should be able to define distances measured in astronomical units (*AU*), parsecs (*pc*) and light-years.*

1 (d) *Candidates should be able to recall the approximate magnitudes, in metres, of the astronomical unit, parsec and light-year.*

Knowledge of the distance of an astronomical body from the Earth is of critical importance. For example, the Sun's angular diameter is directly measurable but its actual diameter can only be determined if its distance is known. Then, knowing its mass, temperature, composition, power output etc., understanding of the processes occurring in the Sun can start to be developed. Likewise, sending men to the Moon or probes to the planets is critically dependent on knowing the relevant distances. As a third example, Hubble's law (see section *2(b)*) is only possible as a concept if the distances to the galaxies are known. However, determination of distances, especially for the most remote objects, is fraught with difficulties.

The first difficulty is an obvious one – the distances cannot be measured directly. The second is that the distances are large. The quantities actually measured, e.g. angles, have to be measured carefully because small uncertainties in the values obtained lead to relatively large uncertainties in the calculated distances. Distances to remote objects are obtained from a complex series of stages. An uncertainty in the relevant distance at any one stage is carried forward to (and made greater in) the next stage and subsequent ones.

The first stage in determining astronomical distances is to determine the mean distance from the Earth to the Sun. This distance is known as the *astronomical unit*, symbol AU.

The astronomical unit has been determined from measurement of the periods of Venus and Earth, and radar determinations of the distance between Venus and Earth when these two planets were in different relative positions in their orbits. By combining these two types of measurement, the mean Earth/Sun distance was then calculated. (This distance is shown in Fig. 1.5). The accurate value of the astronomical unit is 149 597 870 km, which may be approximated to 1.50×10^{11} m, i.e. 1 AU = 1.50×10^{11} m.

A human hair is about 50×10^{-6} m thick: the distance to a horizon may well be of the order of 20 km, i.e. 2×10^4 m. This represents a size factor of about 10^8. It is difficult to imagine that the size factor between the distance to a horizon and the distance to the Sun is 10^7. In other words, the Sun is ten *million* times more distant than a typical horizon. Taking the speed of light as 3×10^8 m s^{-1} and the AU as 1.50×10^{11} m, it follows that it takes light some 500 s, or 8.33 min, to travel from the Sun to the Earth.

The next nearest star is Proxima Centauri and it takes 4.3 years for light from this star to reach the Earth. As a rough approximation, there are 10^7 s in a year. Thus, Proxima Centauri is $4.3 \times 10^7/500$ times further away than the Sun, i.e. a further size factor of approximately 10^5: and this for the nearest star to the Solar System!

The SI unit of speed is m s^{-1} and the SI unit of time is s. The product of speed and time obviously has the unit m, namely that of distance. If the speed used is that of light, i.e. 3×10^8 m s^{-1} and the time is the year, then the product, a distance, is large. The accurate value of the speed of light in a vacuum is 2.977925×10^8 m s^{-1}. In 1 year, there are 365.2564 x 24 x 60 x 60 seconds, i.e. 3.1558153×10^7 s. The product of the speed of *light* and one *year* gives a distance of 9.4608976×10^{15} m, i.e. approximately 9.46×10^{15} m. This distance, known as the *light-year* is suitably large to be a convenient unit for astronomical distances.

Distances to the nearest stars are obtained by a method that is, in principle, a standard triangulation technique, but using the longest available base-line of 2 AU, obtained when the Earth is at opposite ends of a diameter of its orbit. The method is called ‘stellar parallax’. The principle of the method can be readily demonstrated as follows. Stand at one wall of a room, facing the opposite wall, and choose a reference line, e.g. the edge of a window or door frame. Now close the left eye and align an outstretched finger with the chosen reference line, using only the right eye. Close the right eye and open the left eye. The finger will appear to shift its position relative to the reference line by ‘jumping’ to the right. On closing the left eye and opening the right, the finger ‘moves back’ to its original position as shown by the reference line. If the demonstration is repeated with the finger held closer to the face, the apparent shift in position when closing one eye and opening the other will be greater. A similar apparent shift occurs for stars. When viewed against a background of the most distant Galactic stars – and, hence, effectively stationary stars – a nearby star appears to shift its position in the sky over intervals of a few months. Imagine a star Y for which the Y – Sun axis is perpendicular to the ecliptic, i.e. the plane of the Earth’s orbit round the Sun (see Fig. 1.9). Now imagine that Y is observed over a period of a year. During these observations, Y will **appear** to trace a (very small) circle in its position relative to the background stars. This effect is shown in Fig. 1.9 below except that it is necessary to use an oblique viewpoint such that the Earth's circular orbit appears as an ellipse: similarly, the apparent motion of Y against the background stars also appears elliptical. Fig. 1.9 is *not to scale*.

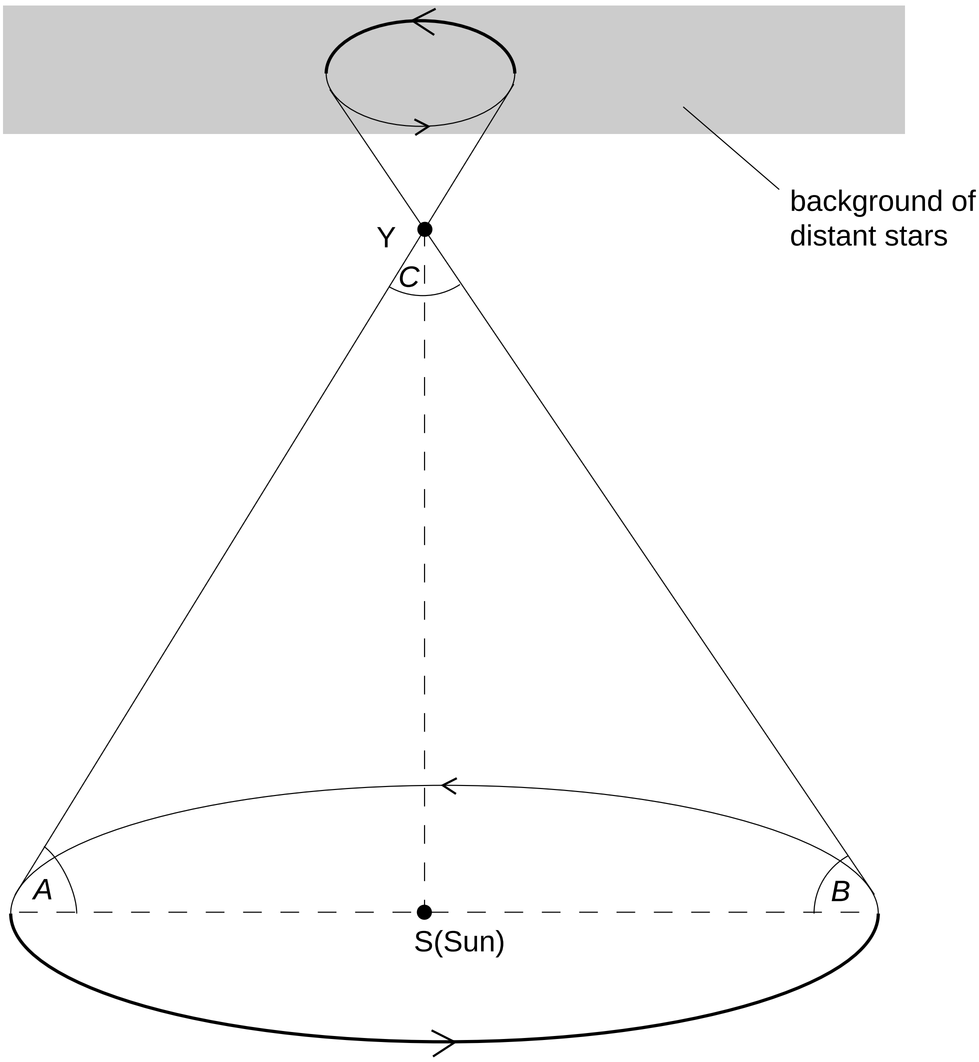


Fig. 1.9 – Stellar parallax (simplified)

Note

The size of the Earth's orbit has been much exaggerated relative to stellar distances.

Fig. 1.9 readily shows, however, that the two angles *A* and *B* can be measured, although the measurements are not easy and require considerable precision. [Indeed, because of the need for precision, the first star for which the parallax was successfully measured was 61 Cygni, by Bessel in 1837.] By simple trigonometry, the angle *C* can be calculated. The stellar parallax is *C*/2. This parallax angle can then be converted into a distance. Again, by simple trigonometry, it can be seen from Fig. 1.9 that

$$\tan C/2 = 1\ \text{AU/distance YS}.$$

However, the angle $C/2$ is so small that the value of its tangent is effectively $C/2$, provided that C is measured in radians. It is now a matter of *definition* that if $C/2$ is an angle of one second of arc, then the distance YS is the standard distance known as a parsec, abbreviated to pc. [This name is a contraction of '*par*allax of one *sec*ond of arc'.] The trigonometric equation above shows that the larger distance YS is, the smaller the parallax angle $C/2$. A second of arc, 1", is 1/60th of a minute of arc, 1', which is itself 1/60th of a degree, there being 360° in a complete circle. A second of arc is, therefore, a very small angle to be measured. Even the nearest star, Proxima Centauri, has a parallax angle of less than 1 second of arc. Because of the observational difficulties, parallax angles smaller than 0.03" become increasingly inaccurate. As a consequence, accurate star distances obtained by this trigonometric parallax method are limited to stars up to about 30 pc away. One other point may be made here. Stars, even nearby ones, do not for convenience situate themselves on a line through the Sun and perpendicular to the ecliptic! It is usually the case that a star for which its parallax angle is measurable will be in a position that is oblique with respect to the Earth, see Fig. 1.10.

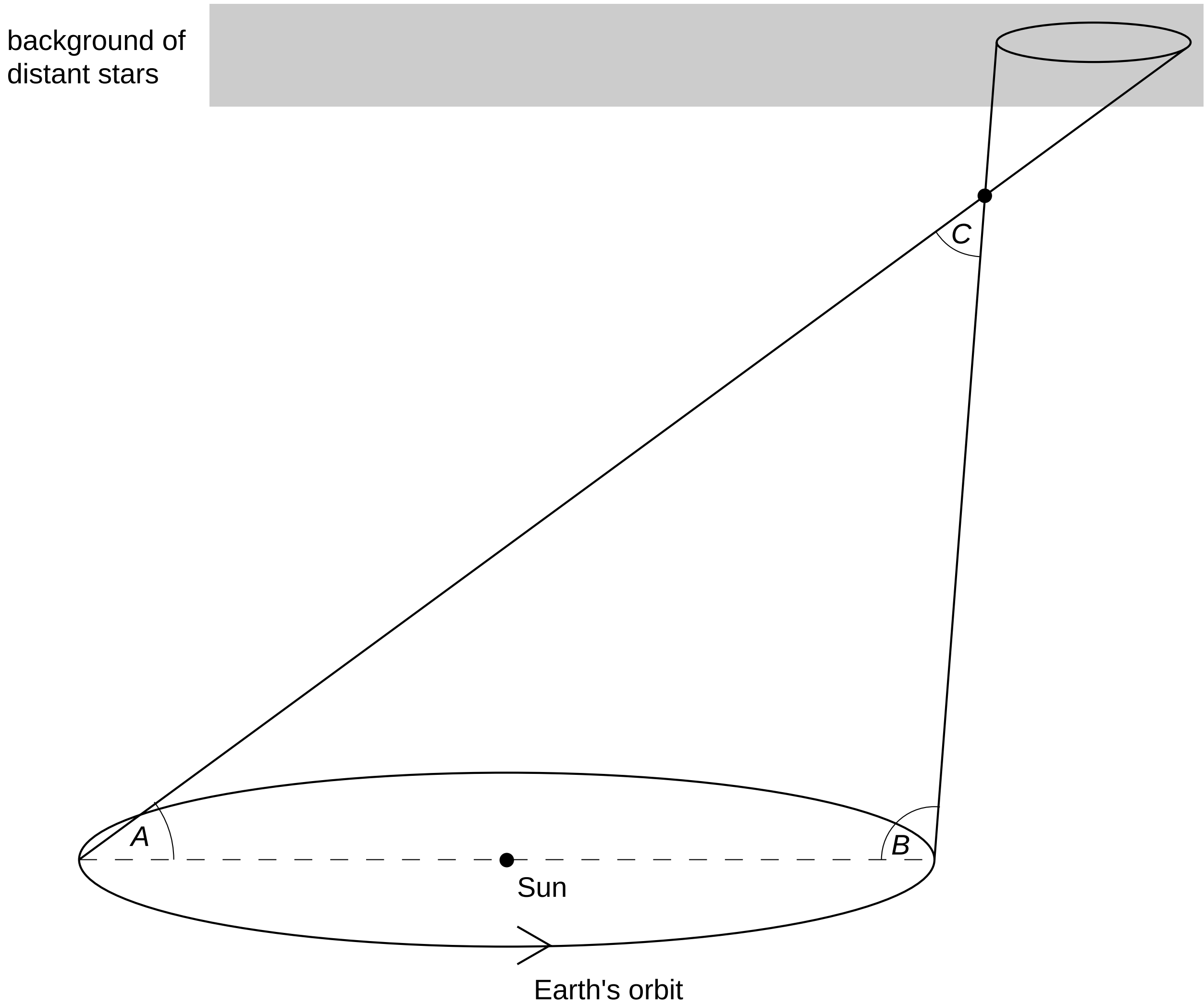


Fig. 1.10

Nevertheless, it remains the case that measurement of angles *A* and *B* will yield angle *C*, from which – with a little more processing – the stellar distance in parsecs, pc, can be deduced.

(It may be noted that this same phenomenon of parallax is a potential source of error when viewing a pointer against a calibrated scale behind it.)

On the definition given earlier, 360° = 2π rad: hence, 1" = 2π /(360 x 60 x 60) rad. For such a small angle, the approximation sin $\theta = \theta$ has negligible error. It follows from the relationship (by definition) between the AU and the pc that

$$1'' = 1\,\text{AU}/1\,\text{pc},$$

that is

$$1\text{ pc} = \{(360 \times 60 \times 60)/2\pi\}\text{ AU} = 2.062648 \times 10^5\text{ AU}.$$

By expressing the AU in metres, the pc can also be converted into metres:

hence, $1\text{ pc} = 3.0856776 \times 10^{16}\text{ m}$
$= 3.09 \times 10^{16}\text{ m}.$

These values may be summarised as follows.

	distance/m	AU	light-year	pc
AU	1.50×10^{11}	1	1.59×10^{-5}	4.85×10^{-6}
light-year	9.46×10^{15}	6.31×10^{4}	1	3.06×10^{-1}
pc	3.09×10^{16}	2.06×10^{5}	3.26	1

An important aspect of establishing the astronomical unit and the parsec as units of astronomical distance is that these units are independent of intrinsic stellar properties, e.g. magnitude.

Proxima Centauri, the nearest star after the Sun, is 1.3 pc away. This is the merest step to the truly enormous distances to the outermost reaches of the observable Universe.

1 (e) Candidates should be able to appreciate the sizes and masses of objects in the Universe.

1 (f) Candidates should be able to appreciate the distances involved between objects in the Universe.

It is not proposed to describe in any detail the complexities of determining the larger astronomical distances. Such determinations need great care and the results of these determinations may be subject to errors that are relatively large compared to the size of error of measurements in other fields of Physics. For example, distance measurements may be in error by a factor of 2.

Nevertheless, a brief outline of methods of measuring astronomic distances is given as a footnote because distance determinations are of such fundamental importance to the understanding of the Universe. By such methods – and others not mentioned – the distances shown in Fig. 1.11 can be determined.

Clusters of stars, i.e. those that are in the same region of the sky and have a common proper motion are, effectively, at the same distance. Observations of their proper motion – i.e. motion perpendicular to the line of sight from Earth – leads to this common distance being determined. By this means and the parallax method, a reasonably large number of stellar distances has been accumulated. Knowing its distance and the apparent magnitude m, i.e. the observed brightness of a star, the latter property can be standardised to give the so-called absolute magnitude M of the star.

There is a category of variable (and intrinsically bright) stars, called Cepheids. The variations in brightness of Cepheids are directly related to their absolute magnitudes, as determined by means such as those just described. Because Cepheids are intrinsically bright, they are detectable over relatively large distances, e.g. in nearby galaxies. The observed brightness, and its variations, of a Cepheid in a galaxy are measured. Then, because the Cepheid's absolute magnitude is deducible from its brightness variations, the distance to the galaxy can be calculated. There are two difficulties worth mentioning.

Firstly, it was later realised that there are two types of Cepheid variable. At a stroke, galaxy distances had to be adjusted by a factor of 2. Secondly, interstellar dust diminishes the brightness of a distant star – or a galaxy for that matter. This is an example of a problem that adds to the uncertainty associated with distances to remote astronomical objects.

By a similar logic, the distance to an even more remote cluster of galaxies can be estimated. The surface brightness (not of individual stars, which cannot be resolved at such distances) of the brightest galaxy in a remote cluster of galaxies is measured. This brightest galaxy is assumed to be as bright, in absolute terms, as a similar galaxy of known distance and known absolute surface brightness: knowledge of all these quantities allows the distance of the remote cluster of galaxies to be estimated.

distance from Earth's centre to	/m	/AU	/pc	/light-years
Earth's surface (i.e. its radius)	6.4×10^{6}	4.3×10^{-5}	2.1×10^{-10}	6.8×10^{-10}
Moon's centre (mean value)	3.5×10^{8}	2.3×10^{-3}	1.1×10^{-8}	3.6×10^{-8} (1 light second)
Sun's centre (mean value)	1.5×10^{11}	1.0	4.9×10^{-6}	1.6×10^{-5} (8.3 light min)
Jupiter (mean value)	7.8×10^{11}	5.2	2.5×10^{-5}	8.1×10^{-5} (0.7 light hr)
Proxima Centauri	4.1×10^{16}	2.7×10^{5}	1.3	4.3
centre of Milky Way Galaxy	3.1×10^{20}	2.1×10^{9}	1.0×10^{4}	3.3×10^{4}
Andromeda Galaxy	2×10^{22}	1×10^{11}	6×10^{5}	2×10^{6}
Virgo cluster of galaxies	6.5×10^{23}	4.3×10^{12}	2.1×10^{7}	6.5×10^{7}
nearby quasar	2×10^{25}	1×10^{14}	6×10^{8}	2×10^{9}
most distant quasar yet observed	$\sim 10^{26}$	$\sim 7 \times 10^{14}$	$\sim 3 \times 10^{9}$	1×10^{10}
observable limit of Universe	$\sim 10^{27}$	$\sim 10^{15}$	$\sim 10^{10}$	$\sim 10^{11}$

Fig. 1.11

As well as distances, determining the masses of astronomical objects is also very important.

The gravitational force due to the Sun on the Earth determines the Earth's orbit. From Newton's laws

$$F = mv^2/r = G\,m\,M/r^2,$$

where m is the mass of the Earth,
v is its orbital speed,
r is the radius of the Earth's orbit,
M is the mass of the Sun,
G is the Universal gravitational constant.

This equation simplifies to $M = v^2r/G$.

Determination of v and of r has already been discussed. The value of G can be determined in the laboratory. Hence, the mass $M_\odot$ of the Sun is determined.

Double stars are useful for determining stellar masses. If their separation and mutual rotational period are known, their masses can be found – using Newton's law in the same way as in the solar example above. Knowledge of stellar masses and their galactic motions gives galactic masses – and, also, of galactic size in terms of their numbers of stars. Some relevant data for astronomical bodies are given in Fig. 1.12.

object	mass	radius/m
Earth	6.0×10^{24} kg	6.4×10^{6}
Moon	7.4×10^{22} kg	1.7×10^{6}
Jupiter	1.9×10^{27} kg	1.4×10^{8}
Sun	2.0×10^{30} kg	7.0×10^{8}
other stars (as a range)	$(0.1 \text{ to } 50) \times 10^{30}$ kg	1.8×10^{8} to 3.6×10^{11}
typical globular cluster in Milky Way Galaxy	3×10^{36} kg	1×10^{18}
Milky Way Galaxy	$10^{11}\ M_\odot$	4.6×10^{20}
other galaxies (as a range)	10^{9} to $10^{13}\ M_\odot$	3×10^{19} to 8×10^{20}
typical cluster of galaxies	$\sim 10^{3}$ galaxies	$\sim 10^{23}$

[$M_\odot$ = mass of the Sun]

Fig. 1.12

The values of these quantities for objects in the Universe are greater than the human mind can readily comprehend. For this reason, it is perhaps helpful to translate these values on to a diagram using logarithmic axes, as shown in Fig. 1.13.

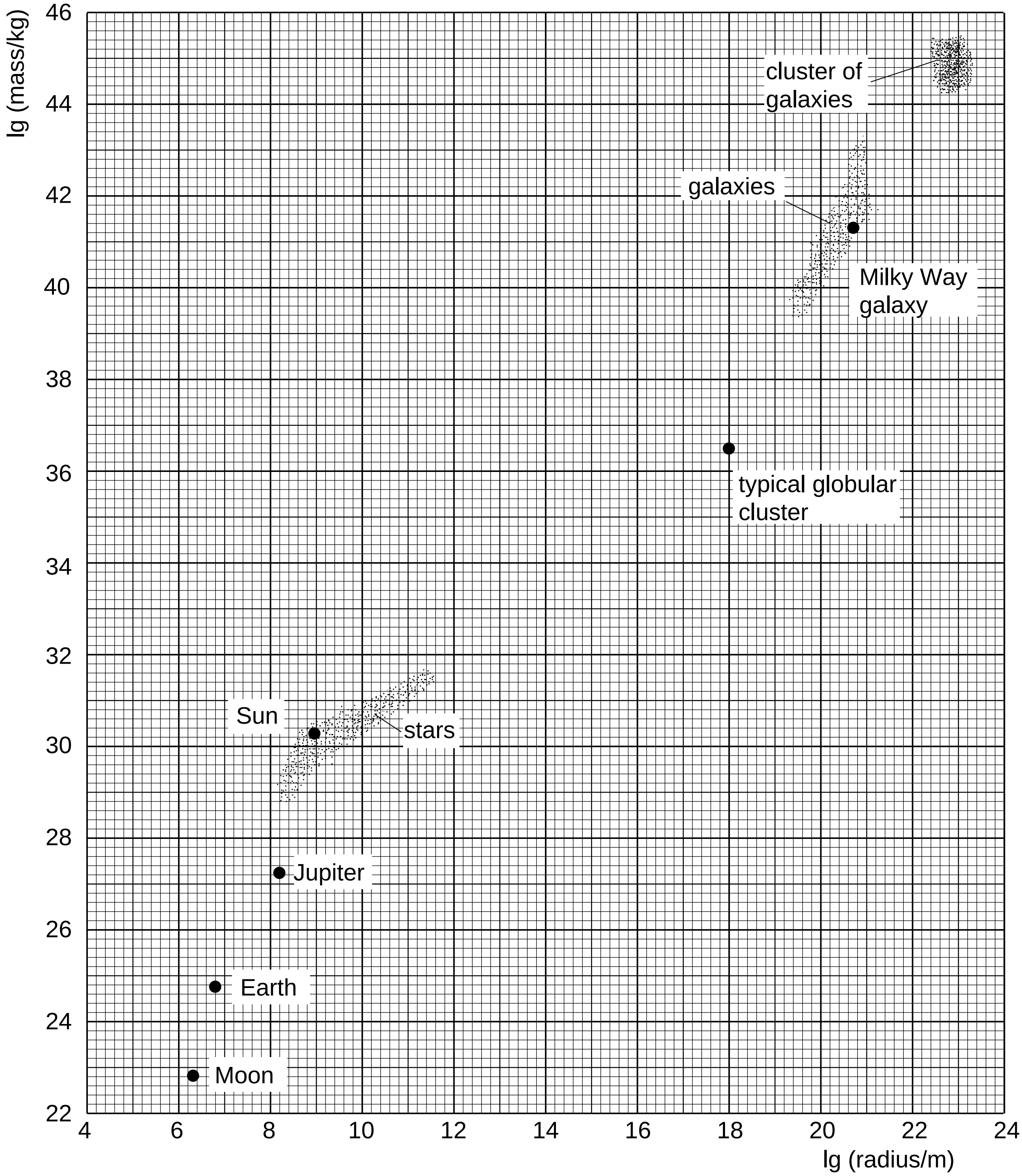

lg (mass/kg)
46
44
42
40
38
36
34
32
30
28
26
24
22
4
6
8
10
12
14
16
18
20
22
24
lg (radius/m)
cluster of
galaxies
galaxies
Milky Way
galaxy
typical globular
cluster
Sun
stars
Jupiter
Earth
Moon

Fig. 1.13

A2. The Standard Model of the Universe

2 (a) *Candidates should be able to describe and interpret Hubble's redshift observations.*

2 (b) *Candidates should be able to recall and interpret Hubble's law.*

As with mass and distance determinations, spectroscopic observations are also of critical importance to astronomy. All stars have absorption lines in their spectra and some have emission lines. The spectral lines of the elements (and their readily formed ions) can be catalogued in an Earth-based laboratory. Observation of the lines in a star's spectrum gives information about the elements present, including their degree of ionisation and relative concentration, in a star's photosphere. (Other information about a star can be obtained from its spectrum but this is not considered here. It is also of interest to note that helium is so called because its spectral lines were discovered in the Sun's spectrum – the Greek for the Sun is helios – before the presence of helium in the Earth's atmosphere was known.)

It was then realised that, in some cases, the lines were shifted towards one end of the spectrum or the other, i.e. towards the red or the blue. If a shift is present in a star's spectrum, then all the lines are shifted in the same direction and to the same relative extent. This is due to relative motion between the source and the Earth-based observer. This shifting of wavelength or frequency due to relative motion between a source and an observer is known as the *Doppler effect*, similar to the effect of, say, the sound of an ambulance's siren dropping in frequency as the ambulance overtakes and recedes from the hearer.

A mathematical treatment of the Doppler effect leads to a relationship between the observed wavelength λ' when the source and the observer are moving relative to one another and the wavelength λ emitted by the source. This relationship may be written as

$$\frac{\lambda' - \lambda}{\lambda} = \frac{v}{c} ,$$

where v is the relative velocity of the source and observer along the line of sight, and c is the speed of light.

It will be noted that a distinction has been made between the relative *velocity* between the source and the observer and the *speed* of light. According to Einstein's General Relativity theory, this latter quantity is a Universal constant that is independent of direction, i.e. it has the same magnitude in all directions and c is, therefore, described as a speed, a scalar quantity. On the other hand, the rate of change of distance between a source and an observer is a vector since direction is involved in quoting the magnitude of v. For example, the source and the observer may be moving further away from each other. In this case, the value of v is, by definition, given a positive value. It follows from the equation above that when v is positive the observed wavelength λ' is greater than the emitted wavelength λ. In the visible spectrum, it is the red end, rather than the blue end, that corresponds to longer wavelengths. This being so, when the source and the observer are moving apart from each other, the electromagnetic radiation emitted by the source is said to show a red shift. Conversely, when a source and observer are approaching each other, the wavelength of the radiation received from the source appears to be shorter than that emitted. In this case, the relative velocity v between the source and observer is given a negative value and the observed radiation is said to show a blue shift.

The above equation can be re-written, partly to take account of the sign of v, as

$$\frac{\Delta\lambda}{\lambda} = \frac{v}{c} \, ,$$

where $\Delta\lambda$ is the short-hand way of showing the shift in wavelength due to relative motion between source and observer.

[It is sometimes convenient to use another form of this equation. Since wavelength is inversely proportional to frequency, the above equation can be modified as

$$\frac{\Delta f}{f} = \frac{v}{c} \, ,$$

where Δf is the shift in frequency.]

It was Hubble who, by resolving the stars in nearby galaxies, helped to establish that galaxies are separate entities, distinct from and at large distances from the Milky Way Galaxy. However, in general, the spectrum of a galaxy is a composite one due to all the stars in it. Hubble also devoted considerable research effort to determining the distances to galaxies – ones that are now known to be the nearer ones. Apart from a few galaxies, Hubble realised that nearly all the galaxies had redshifts in their spectra.

As discussed above, the significance of a *redshift* (as opposed to a blueshift) is that it indicates that the source is moving *away* from the observer. Two qualitative inferences may be drawn. Most galaxies are moving away from the Earth. How, then, are the exceptional, approaching galaxies to be explained? These 'blue-shifted' galaxies are not merely neighbours of the Milky Way Galaxy in their being relatively close but they are also gravitationally associated with each other. Moreover, they are all members of the so-called Local Group of galaxies. [In fact, much of the blue shift of the Andromeda galaxy is due to the Sun's Galactic rotation but there is still a residual blue shift that arises from M31's motion towards the Earth.]

Notwithstanding the fact that Hubble only had data on rather few galaxies and that there was obvious scatter in the data, he postulated quantitative conclusions. Hubble's conclusions have since been vindicated by later research, using larger telescopes and other improved techniques. (The particular benefit of a larger telescope is its improved light-gathering ability, which allows fainter galaxies to be observed. Fainter galaxies are likely to be further away. Moreover, the greater the radius of the Universe that is observable, the greater the volume of space and, hence, the greater the number of galaxies 'brought into view'.)

Fig. 2.1 shows a plot of the recessional speeds of galaxies (calculated from the redshift, using the formula above) against distance from Earth.

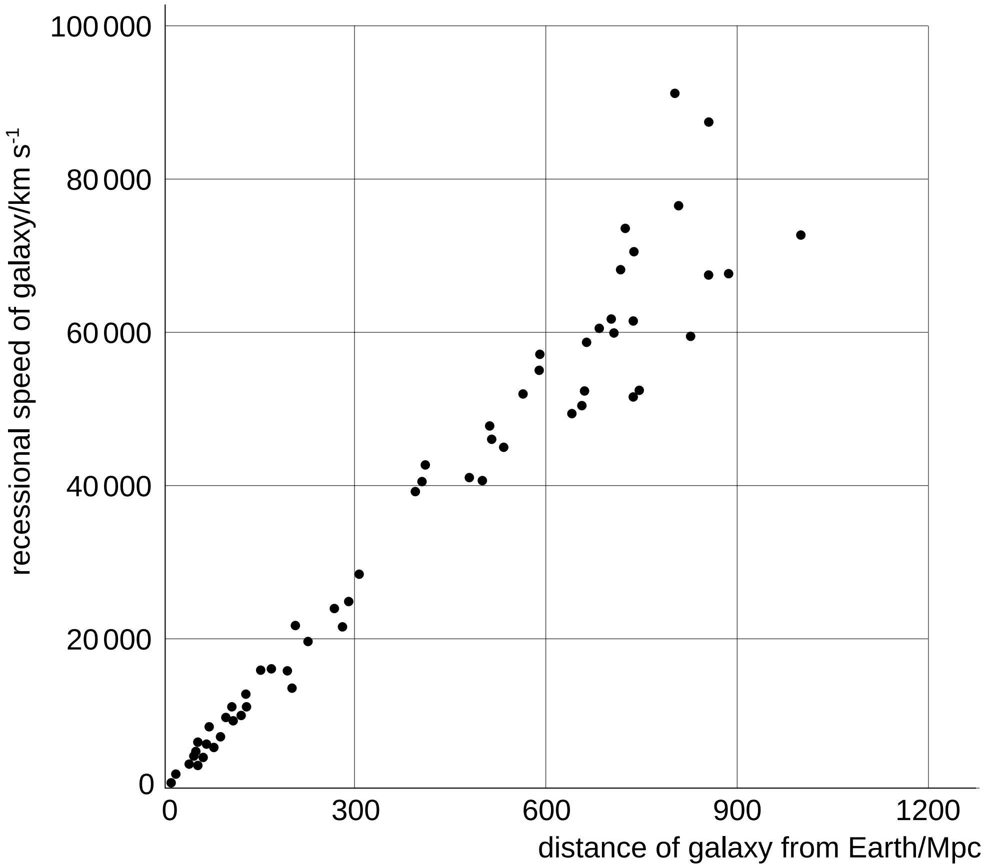


Fig. 2.1

Notes

(a) The gradient of this graph gives the value of the Hubble constant.

(b) The scatter in the top right-hand corner of the graph reflects the difficulty of establishing the distances of remote galaxies.

(c) The square in the bottom left-hand corner shows, approximately, the galaxies on which Hubble originally based his law.

It can be deduced from the graph that a galaxy's recessional speed is proportional to its distance, i.e. the graph is a straight line passing through the origin. As for any measurement(s), there is experimental error – see the earlier comments about the difficulties of measuring galactic distances and, in particular, about interstellar dust. It should be noted that, although the actual value of the gradient of the line is important (see section *2(i)*), a systematic error would not affect the linearity of the graph. The gradient of this graph is, in fact, known as the Hubble constant H_0. From this, the following equation can be written.

$$\text{speed} = H_0 \times \text{distance}$$

A critical inference from the graph is that the Universe is expanding and, apparently, at a constant rate. (See also below.)

As astronomical understanding has increased, the Earth has been progressively displaced from being in a special location. First of all, the Solar System is heliocentric and not geocentric. Secondly, the Sun is not at the centre of the Milky Way Galaxy: it is, rather, no more than a quite ordinary star situated in a quite ordinary position within the Galaxy. Thirdly, the Galaxy itself is no more than a typical spiral galaxy.

It should be clearly understood that the seeming uniformity of Universal expansion away from the Earth does *not* imply that the Earth is the centre of the expansion. Except for the local ‘anomalies’ within clusters of galaxies where the local relative speeds may modify the general expansion, any one galaxy ‘sees’ all other galaxies receding with speeds that are proportional to their separation from the ‘observer’ galaxy. As a consequence, no one galaxy can be regarded as being specially located. A corollary is that the location of the big bang is unknown and unknowable.

Hubble’s law can be expressed by the equation

$$v = H_0 d,$$

where v is the recessional speed between any two galaxies,
d is their distance apart,
H_0 is the Hubble constant.

By convention, galactic speeds v are quoted in $\mathrm{km\,s^{-1}}$. Typical galactic separations d are measured in Mpc ($= 10^6$ pc, i.e. approximately 3 million light-years). Keeping the above equation dimensionally consistent and using these conventional units of v and d means that H_0 is typically quoted in units of $\mathrm{km\,s^{-1}\,Mpc^{-1}}$.

2 (c) *Candidates should be able to convert the Hubble ‘constant’ (H_0) from its conventional units ($\mathrm{km\,s^{-1}\,Mpc^{-1}}$) to SI ($\mathrm{s^{-1}}$).*

It is quite simple to convert the Hubble constant from its conventional units into SI units.

The equation $v = H_0 d$ can be expressed as

$$\text{speed} = H_0 \times \text{distance},$$

$$\text{i.e.} \quad \frac{\text{distance}}{\text{time}} = H_0 \times \text{distance}.$$

It follows that H_0 has dimensions of time^{-1}, the corresponding SI unit of which is $\mathrm{s^{-1}}$.

2 (d) *Candidates should be able to recall Olbers’ paradox.*

2 (e) *Candidates should be able to interpret Olbers’ paradox to explain why it suggests that the model of an infinite, static Universe is incorrect.*

The concept of an expanding Universe is a dramatic departure from the earlier view, that of an infinite, uniform and static Universe. It is perhaps a measure of the innate conservatism of the human mind that even Einstein was minded to add an arbitrary ‘cosmological constant’ to his relativity equations in order to ‘prevent’ the Universe from expanding – as was implied by his equations.

There had been a previously recognised problem with a static model of the Universe. Although not the first to do so, the German astronomer Olbers had pointed out (in 1826) that, if the Universe is infinite, static and uniformly populated with stars, then every line of sight from the Earth would eventually meet the surface of a star. (Remember that the concept of a galaxy was entirely unknown in Olbers' day.) In this case, the night sky should be as bright as daylight. It is not! The implication is that at least one of the postulates underlying Olbers' paradox is wrong. Starting from postulates that were generally accepted in his day, Olbers correctly reasoned and reached a conclusion that conflicted with observation. Given that this conflict is incontrovertible, and that the reasoning is valid, the only remaining fault must be the invalidity of one or more of the postulates. If the Universe is finite, rather than infinite, then the argument that any line, and all lines, of sight meet the surface of a star is invalid. If the Universe is expanding with the speed of recession increasing with increasing distance, then the light from sufficiently distant sources will be red-shifted to such an extent that visible light is shifted into the infra-red and is no longer visible to the human eye. If the Universe is both finite and expanding, then – because the stars, galaxies and the Earth only came into existence somewhat later than the big bang – there are radiant objects so distant that their light has not had time to reach the Earth. The darkness of the night sky can be thought of as the darkness that existed before stars came into being.

In any case, the Universe cannot be 'static'. Stars emit light. Light is a form of energy. A star is finite in size and cannot, therefore, radiate light indefinitely. Stars must evolve and so also the Universe. It is also worth mentioning in this context that the Universe is thought to be only as old as a few stellar lifetimes.

2 (f) Candidates should be able to understand what is meant by the Cosmological Principle.

If the Universe is expanding, it is neither static nor infinite. The third of Olbers' postulates – that the Universe is uniform – is still maintained. It is called the Cosmological Principle. However, the Principle is not intended to apply on a small scale. In this context, 'small scale' is, from an Earth-based point of view, rather large. There are self-evidently 'local' aggregations of matter, from the Earth up to clusters of galaxies and even superclusters of galaxies.

Hubble's law implies that every galaxy 'sees' every other galaxy receding from it. An extension of this idea is that the Universe has the same general appearance irrespective of the vantage point. Figs. 2.2(a) and (b) show the overall distribution of galaxies and clusters thereof. There are volumes of space where galaxies are congregated and other volumes, or voids, of space where galaxies are less common. However, there is no self-evidently preferred direction in which galaxies or voids are to be found. It is this idea that the Cosmological Principle expresses, namely that, on the largest of scales, the Universe is uniform – or, more formally, isotropic and homogeneous.

Fig. 2.2(a)

Notes

(a) The diagram shows the distribution of galaxies within a cone having its apex centred on the centre of the Milky Way Galaxy and its axis perpendicular to the plane of the Galaxy.

(b) It is evident that galaxies are not uniformly distributed but there is no direction which is generally preferred.

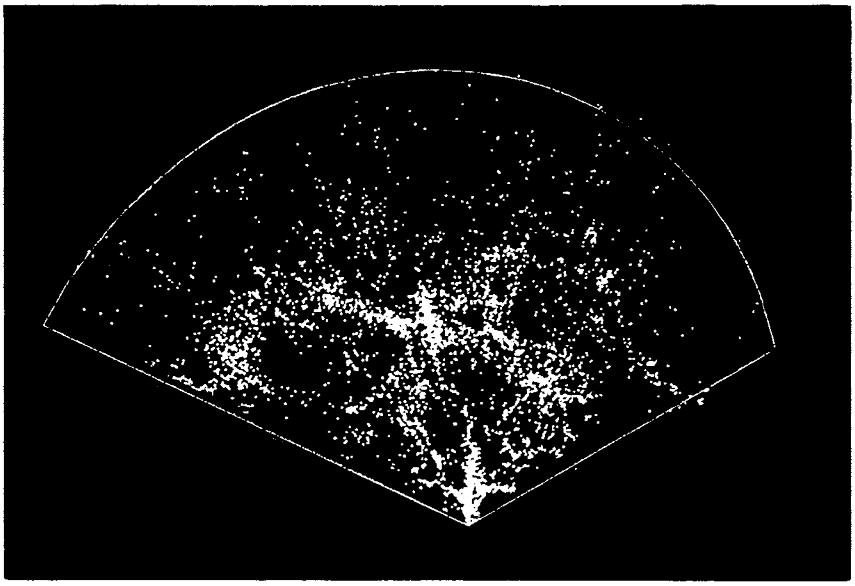


Fig. 2.2(b)

Notes

1 (i) Imagine a fan-shaped area of space, centred on the Sun, with the circular arc of the fan covering 135°.

(ii) The sides of the fan cover distances within which a galaxy has a recessional speed of up to 15 000 km s^{-1}.

(iii) Now imagine an axis through the apex of the fan and in the plane of the fan: the fan is now rotated through some 20° about this axis.

2 The diagram is a composite of the galaxies within this volume of space.

3 As in Fig. 2.2(a), the galaxies are unevenly distributed and appear to mark out quasi-spherical voids containing few galaxies.

2 (g) Candidates should be able to describe, and interpret the significance of, the 3 K microwave background radiation.

2 (h) Candidates should be able to understand that the standard (hot big bang) model of the Universe implies a finite age for the Universe.

The microwave region of the complete electromagnetic spectrum covers wavelengths from about a millimetre to half a metre.

Any hot body emits electromagnetic radiation. The way in which the intensity of this radiation varies with wavelength is well understood and is illustrated in Fig. 2.3. It is not proposed to consider the form of the equation that describes the variation of intensity of emitted radiation with wavelength.

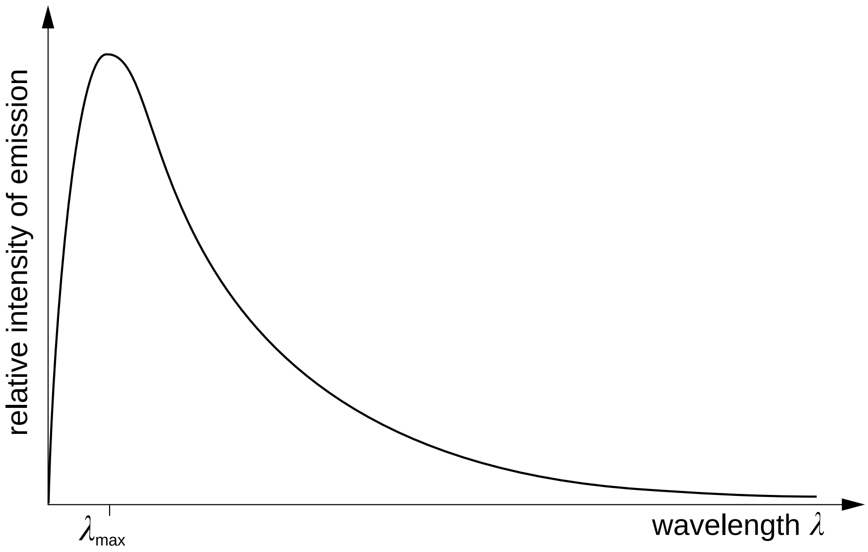


Fig. 2.3

Notes

(a) The general shape of the curve is independent of temperature T.

(b) The wavelength λ_{max} of maximum intensity of radiation varies inversely with thermodynamic temperature.

(c) The accurate value of T for the cosmic microwave background radiation is 2.736 K and the observed microwave background radiation corresponds accurately with that of a black-body radiator.

One simple relationship associated with this curve is that λ_{max}, the wavelength of maximum intensity, varies inversely with thermodynamic temperature T, i.e.

$$\lambda_{max} T = \text{a constant.}$$

This is the simplest relationship from which the temperature associated with the microwave background radiation can be deduced.

It was in 1965, while they were trying to develop a radio antenna for radio astronomy research, that Penzias and Wilson were troubled by a persistent ‘noise problem’ in their system. This ‘problem’ was independent of the direction in which the antenna was pointing, the time of day and the time of year, i.e. the radiation was isotropic. It was then realised that this radiation, which corresponded to a temperature of 2.7 K, was in close accord with what had been predicted by theoretical cosmologists from calculations based on big bang theories.

The observations by Penzias and Wilson were the first ones to be recognised as being associated with the microwave background radiation. The existence of this background radiation has since been fully vindicated and analysed. The isotropy of the radiation is almost too good. For galaxies to have been formed at all, there must have been, it is argued, some inhomogeneities at quite early stages of the big bang. Such inhomogeneities should then be detectable, even today, by some lack of isotropy in the background microwave radiation.

Two recent developments are worth mentioning in this context. The Cosmic Background Explorer satellite, COBE, launched in 1989, has measured the background radiation very accurately. These measurements indicate a black-body temperature of 2.735 K. In addition, 'ripples', i.e. very small intensity variations, in the radiation have been detected. Active research in this field is continuing.

In September 1994, results from the Keck telescope in Hawaii were reported. This telescope, the largest in the world, consists of an array of 36 controllable small mirrors which are equivalent to a single mirror of about 10 m diameter, twice the size of the Palomar reflector. Even with the advantage of increased light-gathering ability of the Keck telescope, some 13 hours of observation time were necessary to obtain the results being looked for. [Modern computers and advanced electronics are needed to be able to keep the telescope accurately pointed at the object under observation and to record the photons being received.] According to the big bang theory, the light now being received at the Earth from the most distant (and, hence, faintest) galaxies was emitted at a time some 2 billion years after the big bang. The present age of the Universe is thought to be about 15 to 20 billion years. The big bang theory also indicates that the Universe would then have been hotter than it is now. By analysis of the light emitted by carbon clouds in these galaxies, evidence has been produced that the background radiation temperature in the region of these galaxies was then 7.4 K, compared with the present 2.7 K.

The significance of the background radiation is that it lends strong support to the idea of a big bang. It is a basic tenet of such theories that the Universe has existed only for a finite time.

There has already been reference to the Cosmological Principle (see section *2(f)*), which indicates that, on the largest scales, the Universe is isotropic and homogeneous. This principle appears to be in conflict with the fact that the Universe is expanding. A theory that offered an alternative explanation to that of the big bang theory was the Steady State theory. In this theory, the Universe was postulated as being of infinite age. However, this theory also had to provide an explanation of the expansion of the Universe. A consequence of this expansion is that there is an 'observable limit' (as opposed to an actual limit) to the Universe. If, in the Hubble constant equation given above, distance d is large enough, then the recessional speed v will become greater than the speed of light. The interpretation of this is that such distant galaxies are so distant that their light cannot reach us. This requires some other explanation for the apparent uniformity of the Universe. The significant idea behind the Steady State theory is that there is a steady, and continuous, creation of hydrogen atoms throughout the Universe such that as galaxies 'disappear' as a result of Universal expansion, they are replaced by galaxies that form from the extra hydrogen atoms that are being continuously created. On this model, there is no 'place' for a background of microwave radiation.

On the other hand, the big bang model of the Universe definitely requires the existence of the background microwave radiation. The expansion of the Universe implies that earlier in the Universe's history, the galaxies were much closer together. The closer the galaxies were, the hotter the Universe was. Hot objects emit radiation. Thus, on the big bang model, when radiation and matter became decoupled, the whole Universe was bathed in black-body radiation corresponding to the temperature of the Universe at the time of this decoupling. As the Universe continued to expand, its mean temperature fell and the temperature of the background radiation also fell. The background radiation retained its black-body variation of intensity with wavelength but the whole spectrum shifted to longer wavelengths – see section *2(a)*). There has been sufficient time for the temperature of the background radiation to fall to 2.7 K (or thereabouts), rather than to absolute zero, 0 K, as would be the case for an infinitely old 'big bang' Universe. In this way, the existence of background radiation of a measurable temperature not only supports a big bang model (as against a Steady State model) but also indicates that the Universe has a finite age.

2 (i) *Candidates should be able to recall and use the expression $t \approx 1/H_0$ to estimate the order of magnitude of the age of the Universe.*

As shown by the dimensional analysis given in section *2(c)*, the SI unit of the Hubble constant H_0 is s^{-1}. The reciprocal, $1/H_0$, has the unit of time. What time is it?

Consider the defining equation for the Hubble constant. In accordance with the Cosmological Principle, the distance d can be that between two receding galaxies and v is then the speed of separation of these galaxies. The expression $1/H_0 = d/v$ can then be regarded as the time it has taken for the galaxies to have become separated by distance d. Given that the expansion of the Universe started with the big bang, the time for the galaxies to separate ($1/H_0$) is a measure of the age of the Universe. It is, however, only an estimate: there are various reasons for this.

The first reason is, in a sense, trivial. As previously indicated, determining the distances of galaxies is difficult and there are uncertainties in the values of d. A consequence is that there are uncertainties in H_0 and, thus, also in the possible age of the Universe.

Another possible reason is that the big bang theory postulates that there was a time delay between the big bang and the formation of the galaxies. The time back to the coalescence of the galaxies is not quite as far back as the big bang itself. This would imply that the reciprocal of the Hubble constant gives an underestimate of the age of the Universe. However, this possible reason is overshadowed by the following considerations.

Because the Universe is expanding, the value of the Hubble constant is changing with time. The separation of galaxies means that work has to be done against the gravitational (attractive) forces between them. This 'work done', which increases the total potential energy of the galaxies, necessarily reduces their kinetic energy. In other words, the expansion of the Universe is slowing down. This effect is illustrated in Fig. 2.4.

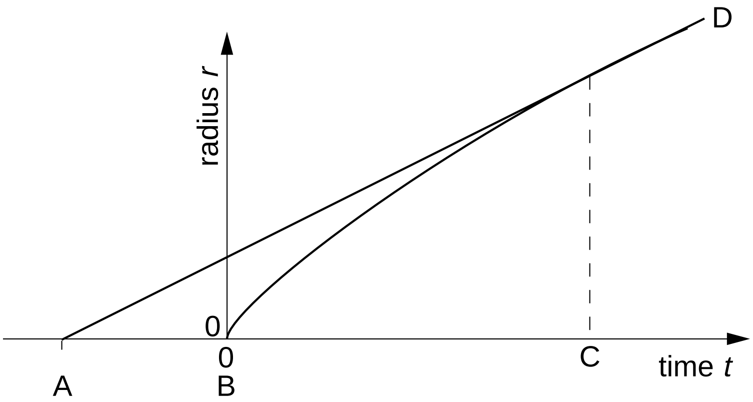


Fig. 2.4

Notes

(*a*) Radius *r* is the radius of any representative spherical volume of space which is taking part in the expansion of the Universe.

(*b*) Point C represents the present time.

(*c*) (i) The line AD represents a tangent (at the present time) to the expansion curve of the Universe.

(ii) The gradient of line AD gives the Hubble Constant.

(iii) The graph shows that the current rate of expansion of the Universe is not changing very rapidly.

(*d*) Point B represents the big bang.

(*e*) The time AC represents the age of the Universe as given by the Hubble Constant.

The 'present time' is indicated on this graph and it can be seen that the expansion of the Universe is now, to a first approximation, linear. However, the graph also shows that the expansion was, at the earliest times, much more rapid. Taking the reciprocal of the Hubble constant is, in effect, extrapolating the gradient of this graph backwards to the intercept on the time axis, i.e. when the radius of the Universe was zero. Doing this gives an overestimate of the age of the Universe.

2 (k) *Candidates should be able to understand that the Universe may be 'open', 'flat' or 'closed', depending on its density.*

2 (l) *Candidates should be able to appreciate that the age of the Universe cannot be determined from the Hubble constant until its density is known accurately.*

2 (m) *Candidates should be able to understand that the ultimate fate of the Universe depends on its density.*

As Fig. 2.1 shows, there is a linear relationship between the recessional speed of a galaxy and its distance from the Earth (or any other vantage point). The furthest known quasar so far observed is at a distance from the Earth of about 10 x 10^9 light-years. The use of the light-year as a distance unit is particularly useful in this context not only because it indicates how long it has taken the light from such galaxies to reach the Earth but also because it indicates how much nearer to the time of the big bang it was that the light now being received started on its journey.

The estimated age of the Universe is 15 to 20 billion years so that this most distant quasar so far observed appears to have been in existence for not less than about half of the current time of existence of the Universe. The recessional speed is about a twentieth of the speed of light. Since no object can travel faster than light, being able to observe these distant galaxies is not all that far from representing an 'observational' (but not 'real') 'edge' or 'horizon' of the Universe.

Galaxies exert gravitational attraction for each other and work has to be done in separating them: this work has to come from the kinetic energies of the galaxies. The consequence of this is that the expansion of the Universe is slowing down, a further consequence being that the Hubble constant is not, in an absolute sense, a constant in that its value must be decreasing with time. However, the Universe is sufficiently old for the time-dependent decrease of the value of the Hubble constant to be, at the present time, very slow (and getting slower). At earlier/the earliest times, the expansion of the Universe must have been very much more rapid than it is at present. The Hubble graph of Fig. 2.4 appears to be linear because the current rate of the decrease in the value of H_0 is sufficiently small (and this is distinct from uncertainties in the value of H_0 as a result of experimental errors in measuring it).

Compared with the current overall age of the Universe, the period of very rapid expansion was exceedingly short. Estimating the age of the Universe by using $t = 1/H_0$ may give an over-estimate of about 50%.

But what of the future? It has already been mentioned that there is clear evidence from galactic rotations that the Universe contains dark matter. There are then the important questions as to how much dark matter there is and what its nature is. The more matter (visible or invisible) there is in the Universe, the greater the average density of matter in the Universe. Likewise, the greater the density of matter in the Universe, the greater the work that has to be done in the expansion; the greater this work is, the more rapidly the expansion slows down.

There are then three possibilities.

(i) The average density of matter in the Universe is too low to cause the rate of expansion to decrease to zero.

Under these conditions, the Universe will continue to expand for ever. The term used to describe this situation is to say that the Universe is 'open'.

(ii) It is also possible that the average density of matter in the Universe has the critical value that would result in the rate of expansion decreasing precisely to zero.

For reasons that need not be dealt with, the description of the expansion becoming precisely zero in an infinite future is that the Universe is 'flat'.

(iii) The third possibility is that the mattter in the Universe is sufficiently dense for the expansion to be brought to a halt and, indeed, reversed.

If this were to be the case, the Universe would at some remote, but not infinite, time start to contract. Just as the expansion is now slowing down, once contraction starts it begins to accelerate. After due time, this would lead to a 'big crunch'. Such a Universe is said to be 'closed'.

These three possibilities are illustrated in Fig. 2.5.

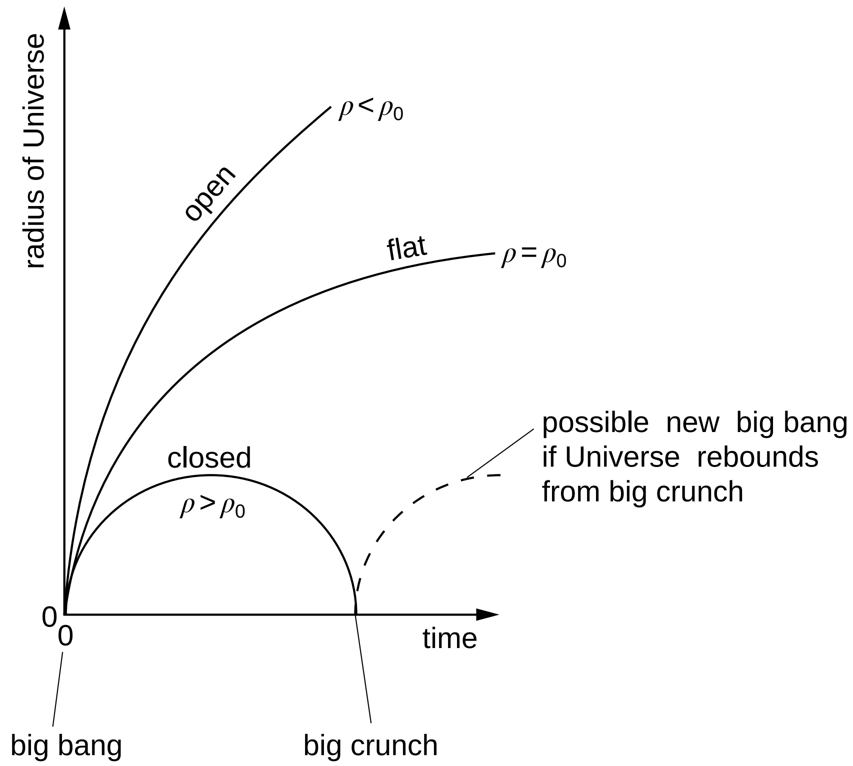


Fig. 2.5

Notes

(a) ρ is actual mean density of matter in the Universe.

(b) ρ_0 is the critical density of matter in the Universe for it to be flat, i.e. to expand to an infinite size at infinite time.

It is tantalising that current knowledge of the rate of expansion of the Universe (and the rate at which it is slowing down) cannot *quite* – because of the experimental difficulties – determine what the future fate of the Universe will be.

[It is also tantalising that techniques – based on gravitational lensing effects (not being considered, but see the footnote) – are currently being developed that might well provide the answer to the question of whether the Universe will expand forever, eventually re-collapse or do neither of these.]

The fate of the Universe depends on the average density of the matter it contains. Likewise, if this average density could be accurately determined, together with an accurate value of the Hubble constant H_0, a more accurate value of the age of the Universe could be found. The density of matter in the Universe governs how the value of H_0 is changing with time and did so in the past. If both ρ and H_0 were accurately known, a better correction to the simple estimate of $1/H_0$ for the age of the Universe could be made. Unfortunately, at present neither of these parameters has been reliably determined.

2 (n) *Candidates should be able to recall that it is currently believed that the density of the Universe is close to, and possibly exactly equal to, the critical density needed for a 'flat' cosmology.*

Establishing accurate values for galactic distances remains an issue. The relevance is that these distances affect the estimation of the density of matter in the Universe.

Based on the currently accepted scale of galactic distances, it is estimated that the stars, i.e. the visible matter, in galaxies give a value of ρ, the Universal average density, which is 1% of the critical density ρ_0 needed for the Universe to be flat. In other words, $\rho = 0.01\ \rho_0$.

On the other hand, estimates based on the primordial abundances of hydrogen and the other light elements give $\rho = 0.03\ \rho_0$. This itself indicates that there is more dark matter than luminous matter in galaxies.

Observation of the rotation of stars about the centres of their parent galaxies also confirms that galaxies must contain dark matter, possibly a spherical dark halo in which the galaxy is wholly embedded. This increases the estimate of ρ up to about 10% to 20% of ρ_0. Similarly, observation of galactic motion within large clusters further increases the estimate of ρ up to about 30% of ρ_0.

In January 1983, the Infra-red Astronomical Satellite (IRAS) was launched into an Earth orbit to carry out a survey of infra-red galactic sources. Because infra-red radiation is not affected by absorption by interstellar dust in the Milky Way, the satellite discovered 'new', more distant galaxies. The survey covered most of the sky. The IRAS results suggest $\rho \approx 0.70\ \rho_0$, the uncertainty in this value being consistent with $\rho = \rho_0$.

A prediction of Einstein's theory of General Relativity was that light passing near a massive body is deflected. There was an early successful test of the theory during the May 1919 total solar eclipse. Those stars which were very nearly in the same direction as the Sun became observable during the eclipse. Ordinarily, of course, stars are 'obliterated' by the intensity of scattered sunlight during the day. These stars appeared to be displaced from their known directions because their light was deflected by the Sun. The observed displacement was in very close agreement with that predicted by Einstein's theory.

In recent years, multiple images of very distant quasars, i.e. starlike images of galaxies at very early stages of their evolution, have been observed. The multiple images occur because the light from such a quasar is deflected by the gravitational effect of a nearer galaxy that happens to be in front of the quasar in question. This deflection is called gravitational lensing. The amount of the deflection allows the total mass of the interfering galaxy to be estimated. This mass, which naturally also includes the mass of dark matter, can be compared with other estimates of the mass of the interfering galaxy, e.g. from its brightness and distance or from its rotational behaviour. Development of the theory of gravitational lensing may well lead to a more reliable value of the average density of matter in the Universe and, hence, to an answer to the Universe's eventual fate.

From these results, it is evident that ρ is close to the critical 'flat' value. This is remarkable in itself, that ρ should apparently merely 'happen' to be so close to ρ_0. Indeed, some astronomers think that this closeness is more than co-incidental, especially in view of the experimental uncertainties. This is perhaps a rather philosophical approach, although it is supported by theoretical -cum- mathematical arguments that, as the Universe ages, any initial departure of ρ from ρ_0 would increase. In other words if ρ is *now* apparently very close to ρ_0, it must have been closer earlier in the Universe's history. These mathematical arguments are not being considered.

However, the arguments which suggest that the Universe is actually 'flat' also relate to the smoothness of the background microwave radiation. It has been a 'relief' that ripples have been detected in this background radiation. The problem is that if the background radiation is too smooth, it is difficult to explain how galaxies could form from such a smooth distribution of matter in the time available since the era of matter/radiation decoupling. Failure to find these ripples would have thrown doubt on the concept of big bang cosmology. As mentioned just above, there are arguments that if the distribution of matter was so smooth at this decoupling era, the degree of smoothness would have been even greater at earlier stages.

2 (o) Candidates should be able to derive, from Newton's law of gravitation, the expression

$$\rho_0 = \frac{3H_0^2}{8\pi G}$$

and recognise that General Relativity is needed for a strict derivation.

2 (p) Candidates should be able to use the expression $\rho_0 = \dfrac{3H_0^2}{8\pi G}$.

It is possible to use Newton's law of Gravitation to derive a relationship between the critical density ρ_0 for the Universe to be flat and the Hubble constant. However, this is a simplified derivation. Einstein demonstrated in his theory of General Relativity that energy and mass are equivalent, i.e. $E = mc^2$. A moving body has kinetic energy, i.e. more energy than when it is at rest. From $E = mc^2$, it follows that increasing the kinetic energy of a body necessarily also increases its mass. A strict derivation of the relationship between ρ_0 and H_0 needs to take this relativistic effect into account but such a derivation is too complex to be treated here. It is sufficient to take the simpler approach based on Newtonian gravitation. [It is quite commonly the case that a Newtonian approach, rather than a relativistic one, is thoroughly satisfactory other than in extreme circumstances, e.g. bodies moving with speeds close to that of light. In this sense, Einstein's once very radical theories 'improve upon' rather than 'overturn' Newtonian ideas.]

Consider a 'test' galaxy of mass m at the present time. This 'test' galaxy is now at a distance r from the site of the big bang. It is subject to the gravitational force of attraction F of all the galaxies (matter) of total mass M that is at a distance *less than* r. From Newton's law

$$F = \frac{GmM}{r^2} .$$

(It may be noted that, notwithstanding the expansion of the Universe, the value of M is fixed.)

The work done ΔW in increasing r to $(r + \Delta r)$ of the 'test' galaxy (where Δr is a small increase in r) is given by the product $F\Delta r$. This is shown by the shaded area in Fig. 2.6.

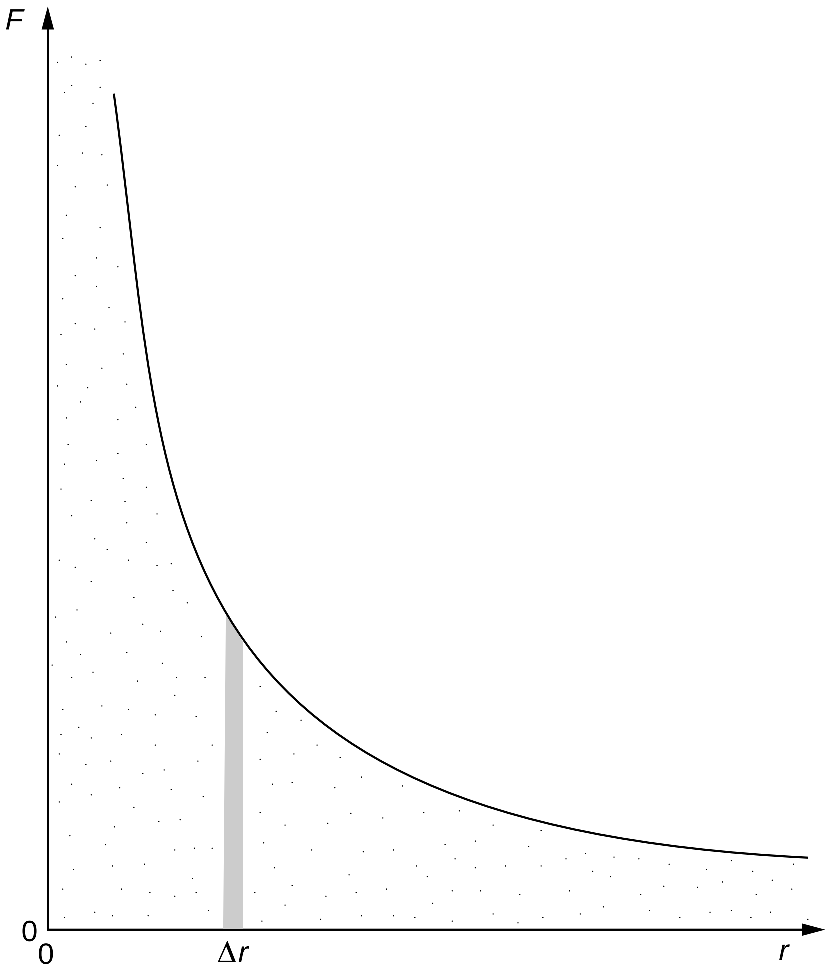


Fig. 2.6

For the 'test' galaxy to become infinitely separated from the galaxies of mass M, the total work done W is given by the total area below the curve in Fig. 2.6, i.e. the area shown stippled. It can be shown (by using calculus) that

$$W = \frac{GmM}{r} .$$

[Alternatively, the potential φ at a point in the gravitational field due a (point) mass M is given by

$$\varphi = \frac{-GM}{r} ,$$

where r is the distance of the point from the mass. The gravitational potential energy of a mass m at this distance r is then the same as that quoted above.]

By definition,

$$\text{density } \rho = \frac{\text{mass}}{\text{volume}},$$

where the mass is M and the volume is $4\pi r^3/3$.

This gives $\rho = \dfrac{3M}{4\pi r^3}$ or $M = \dfrac{4\pi r^3 \rho}{3}$.

Substituting for M in the expression for W gives

$$W = \frac{4\pi}{3} Gmr^2\rho .$$

From Hubble's law, the recessional speed v of the 'test' galaxy is given by

$$v = H_0 r.$$

Substituting for r gives

$$W = \frac{4\pi}{3} G\rho \frac{mv^2}{H_0^2}$$

If, now, the density of the Universe at the present time has the corresponding critical value, i.e. ρ_0, for the Universe to be flat, the work done W is equal to the gain in gravitational potential energy of the 'test' galaxy. This gain must be equal to the loss of kinetic energy of the 'test' galaxy. The overall result is that, at an infinite time in the future, the 'test' galaxy would be at rest and infinitely separated from the galaxies of mass M (and, incidentally, from all other galaxies).

In other words,

$$W = \frac{1}{2} mv^2 = \frac{4\pi}{3} \frac{G\rho_0}{H_0^2} mv^2,$$

and

$$\rho_0 = \frac{3H_0^2}{8\pi G}.$$

2 (j) *Candidates should be able to describe qualitatively the evolution of the Universe from 0.01* s *after the big bang to the present, including the production of excess matter over antimatter, the formation of light nuclei, the recombination of electrons and nuclei and the formation of stars, galaxies and galactic clusters.*

2 (q) *Candidates should be able to appreciate that there is no experimental evidence for the physics involved at the energies prevailing during the evolution of the Universe before about 1* ms*.*

2 (r) *Candidates should be able to outline the difficulties involved in projecting the evolution of the Universe back before 0.01* s*.*

Before embarking on a more detailed discussion of the topics set out above, it is probably advantageous to repeat or summarise certain points and to introduce some quantum ideas from particle physics.

During the first 30 years of the 20th Century, Hubble established that galaxies exist as objects which are external to the Milky Way Galaxy and discovered the expansion of the Universe. This led to different views of the origin of the Universe, namely the big bang cosmology and steady state cosmology. The latter approach was that there was continuous, but very slow, creation of hydrogen in space to replace the galaxies which were receding out of view so as to ensure that the Universe appeared to have an aspect that was, in its essentials, unchanging.

Although there are astronomers who do not accept big bang cosmology, this approach is now dominant following the discovery of the microwave background radiation. Important features of this radiation in relation to cosmology are:

(i) its high degree of homogeneity and isotropy,

(ii) the closeness of its spectral profile (i.e. intensity *v* wavelength) to that of a black-body radiator, suggesting that there was a stage at which matter and radiation were in thermal equilibrium, i.e. during the early stages of the big bang,

(iii) the observed abundances in the Universe of hydrogen, deuterium, helium and lithium agree well with the values predicted by big bang theory.

A major feature of big bang cosmology is that the initial conditions were of extremely high pressure and extremely high temperature. In other words, initially there was a state of very great energy density. It is to such conditions that ideas of quantum and particle physics are particularly relevant.

The ‘world’ of particle physics is rather complex. For example, the proton and the neutron are simultaneously classified as hadrons, baryons and fermions. They are hadrons because they are subject to the ‘strong force’ that holds them together within an atomic nucleus. They are baryons because they are relatively massive, some 2000 times more massive than the electron. They are fermions in that they are considered to be ‘matter particles’.

By contrast, the electron is both a lepton and a fermion (but not a hadron), being both relatively light, and a ‘matter particle’ (but not subject to the ‘strong force’).

For the present discussion, another type of particle to be considered is the boson. Bosons may be described as ‘force carriers’. One member of the family of bosons is the pion which mediates the strong force holding protons and neutrons together in an atomic nucleus. Photons are also bosons in that they are not matter particles and mediate the electromagnetic force.

Another important quantum idea is that every particle has an antiparticle. A particle and its antiparticle have equal masses (and spins). Their other properties, however, are exactly equal but opposite. Perhaps the most familiar (and earliest observed) example of a particle-antiparticle pair is the electron, e^-, and its antiparticle, the antielectron, more commonly known as the positron, e^+.

As a consequence of their having opposite properties, when a matter particle and its antiparticle meet, they mutually annihilate with their combined masses appearing as two identical photons of radiation, i.e. of equal energy and frequency. The total energy of the two photons which are formed as a result of the annihilation can be found from Einstein’s mass-energy relation. For example, if a particle of mass m_{X} and its antiparticle of mass $m_{\overline{\mathrm{X}}}$ mutually annihilate each other, then

$$E = (m_{\mathrm{X}} + m_{\overline{\mathrm{X}}})\, c^2 = 2hf$$

Given that, by definition, $m_{\mathrm{X}} = m_{\overline{\mathrm{X}}}$, this equation can be modified as

$$hf_{\mathrm{X}} = m_{\mathrm{X}}c^2,$$

where f_{X} indicates the minimum frequency that each of a pair of identical photons needs to have to create a particle of mass m_{X} (together with its companion antiparticle).

Certain massless particles, e.g. photons, are in effect their own antiparticles. This means that it is possible, therefore, for the opposite process of matter particle-antiparticle annihilation to occur. In other words, it is possible for two photons to create a matter particle-antiparticle pair.

A further concept needs to be mentioned: protons and neutrons are thought of, in particle physics, as consisting of trios of quarks. For forming these two nucleons (which are, as indicated above, fermions), two types of quark are postulated, each having its own antiquark. The so-called up-quark has a charge of + 2/3 relative to the electronic charge –1, whereas the charge on a down-quark is –1/3.

On this model:

a proton consists of two up-quarks and one down-quark;
a neutron consists of one up-quark and two down-quarks.

It may be noted in passing, however, that although fermions can only be created or destroyed as particle-antiparticle pairs, this is not true of bosons. For example, a single photon is destroyed when its associated energy is absorbed by increasing the amplitude of vibration of a molecule. Similarly, a photon is released when the electron in a hydrogen atom undergoes a transition to a lower energy level.

The corresponding bosons, the force-carriers which hold these quarks together in a nucleon, are called gluons and they are thought to consist of pairs, rather than trios, of quarks. The boson which acts as the ‘glue’ for the three quarks in a proton is made up of an up-quark and an anti-down-quark.

Finally, three energy relationships in particular are relevant to discussion of the evolution of the Universe. At a temperature T, the characteristic particle energy is given by kT, where k is the Boltzmann constant. The other two energy relationships should already be familiar (and are referred to above), i.e. $E = hf$ as the (quantised) energy of a photon and $E = mc^2$.

These three relationships can be linked together as

$$E = kT_x = hf_x = m_x c^2$$

As already mentioned above, a pair of identical photons can form a matter particle-antiparticle pair, each of mass m_x, only if the frequency f_x of the photons is such that

$$hf_x \geqslant m_x c^2.$$

Correspondingly, the temperature T_x must satisfy the inequality

$$kT_x \geqslant m_x c^2$$

$$\text{or } T_x \geqslant \frac{m_x c^2}{k}$$

This minimum temperature T_x for the creation of a matter particle of mass m_x (and equally, its antiparticle) is called the threshold temperature of the particle. On the other hand, particles such as neutral atoms and even atomic nuclei could not exist at the very high temperatures of the very early Universe. In such particles, the forces holding these composite entities together would be too weak to withstand the disruptive effect of collisions with highly energetic photons.

It is now appropriate to start the description of the big bang from its beginning in terms of the standard model*.

The main hypotheses of big bang cosmology are that, between 15 to 20 billion years ago, space and time came into being, the radius (or volume) of the Universe being exceedingly small and the temperature being exceedingly high. The energy density was correspondingly exceedingly high. From its beginning in the big bang, the Universe has been expanding. It should be stressed at this point that the expansion of the Universe comes from space itself expanding: it is not, as in an ordinary explosion, a case of the material of the Universe spreading into a pre-existing but empty space.

*There are certain problems, not being discussed, which are associated with the standard model of big bang cosmology and which are more readily resolved by a so-called inflationary model of the big bang. This latter model, which is closely related to the standard model, postulates that almost immediately after the initiation of the big bang there was a very short, but critical, period of extreme expansion of the Universe. The model is called the inflationary model because of this supposed extreme expansion. After the end of this inflationary period, the two models agree.

As well as expanding, indeed because of its expansion, the Universe cooled after the big bang. For reasons which are referred to later in this Section, the evolution of the Universe is described from a time 0.01 s after the big bang. At this time, the temperature of the Universe was of the order of 10^{11} K, but its volume was still small.

At such a high temperature and high energy density,

(i) the rate of collision between particles was high so that the following opposing processes were taking place rapidly

(identical) photon pair $\rightleftharpoons$ matter particle/antiparticle pair,

(ii) the Universe was effectively at thermal equilibrium.

The abundant particles were those with a threshold temperature of the order of 10^{11} K. Particles (and their antiparticles) of a mass with a threshold temperature much higher than 10^{11} K were no longer abundant. Such pairs of particles annihilated each other but they could not be re-formed because the mean photon energy was too low. On the other hand, composite particles with relatively weak binding forces could not ‘stand the heat’.

In this way, the Universe of this temperature consisted primarily of photons, neutrinos, antineutrinos, electrons and positrons. There were relatively few nucleons, i.e. 1 per 10^9 photons or electrons or neutrinos. Neutrons and protons were approximately equally abundant at a time of 0.01 s or 10^{11} K because the two reversible processes below were – like those in (i) above – proceeding rapidly and were in equilibrium:

neutron + positron $\rightleftharpoons$ proton + antineutrino

neutron + neutrino $\rightleftharpoons$ proton + electron.

Neutrons are slightly more massive than protons. From the equivalence of mass and energy, it follows that the neutron is of greater intrinsic energy than the proton. As a consequence, there is a tendency for the conversion of neutrons into protons to be more rapid than the converse. (Students also studying Chemistry will be familiar with the idea that exothermic reactions are more likely to occur than endothermic ones.) Then, as the Universe continued to expand and cool rapidly, the ratio of neutrons to protons began to shift in favour of the proton.

At this time, the bias of the Universe towards matter rather than antimatter had already occurred but this topic is referred to again a little later in this Section.

When the age of the Universe was between 1 s and 14 s, the temperature had fallen from 10^{10} K to about 3 x 10^9 K. At this lower temperature, the mean photon energy had decreased to such an extent that the rate of formation of electron-positron pairs became lower than the rate of mutual annihilation of these particles. The abundance of these particles decreased therefore. Between these two times, the overall conversion of neutrons into protons continued but became less rapid. By the end of this period, the neutron/proton ratio had fallen to about 0.25, i.e. 1 neutron for every 4 protons.

As the temperature fell to 3 x 10^9 K, the Universe was theoretically cool enough for ${}^{4}_{2}He^{2+}$ ions to remain in existence. However, the formation of this nucleus required other steps involving deuterium nuclei, tritium nuclei and/or helium-3 nuclei. These three latter were not stable enough at this temperature so it was not kinetically possible for helium-4 nuclei to occur.

Then, when the Universe was about 225 s old, the temperature was down to 9×10^8 K and deuterium nuclei became stable. As a consequence, it was then possible for helium-4 nuclei to be formed and the available neutrons became locked mainly into helium nuclei, but with smaller proportions of deuterium and lithium-7 nuclei. As an approximation, the proportion (by mass) of hydrogen nuclei to helium nuclei was 3 to 1.

After about half an hour, the annihilation of positrons by electrons was complete but there was a sufficient surplus of electrons to neutralise the charge on all the positive atomic nuclei in existence. Nevertheless, the temperature was still too high for these nuclei and electrons to form neutral hydrogen or other atoms. The nucleus and electron that momentarily formed a neutral atom were driven apart by the energy density of the Universe, by absorption of suitably energetic photons or by bombardment with matter particles temporarily created from these photons.

However, once the Universe had expanded and cooled further, the photons then no longer had, on average, sufficient energy to cause the ionisation or disruption of hydrogen atoms or other atoms. Since there were no longer free electrons to interact with the photons, the Universe became transparent: radiation and matter had become ‘decoupled’. From this moment, the Universe became dominated by matter rather than by radiation as it had been previously. For this process to occur, the temperature had to fall to about 4000 K. It should be remembered that the rate of expansion and consequential cooling of the Universe was (and is) all the time decreasing.

The process of decoupling radiation and matter did not, it is thought, occur until between a hundred thousand and a million years after the big bang, a value of 3×10^5 years being commonly suggested.

The scene was then set for the formation of stars, galaxies and all the other structures and material bodies found in the Universe. It is perhaps an oddity that the current level of theoretical understanding of the very early periods of the Universe does not, as yet, extend to the galactic structure of the Universe. Debate continues as to whether the process was ‘top down’ or ‘bottom up’. The ‘top down’ mechanism suggests that originally inconceivably vast clouds of hydrogen and helium first divided – through self-gravitation – into clouds of sufficient size to form superclusters of galaxies. Further self-gravitation led to clouds commensurate with the size of a ‘simple’ cluster of galaxies, then down to single galaxy size, with individual stars being last in this downwards, or greater fragmentation, series. Alternatively, the ‘bottom up’ sequence supposes that stars formed first, with subsequent greater and greater aggregation.

The evolution of the Universe, as outlined above, is illustrated in Fig. 2.7.

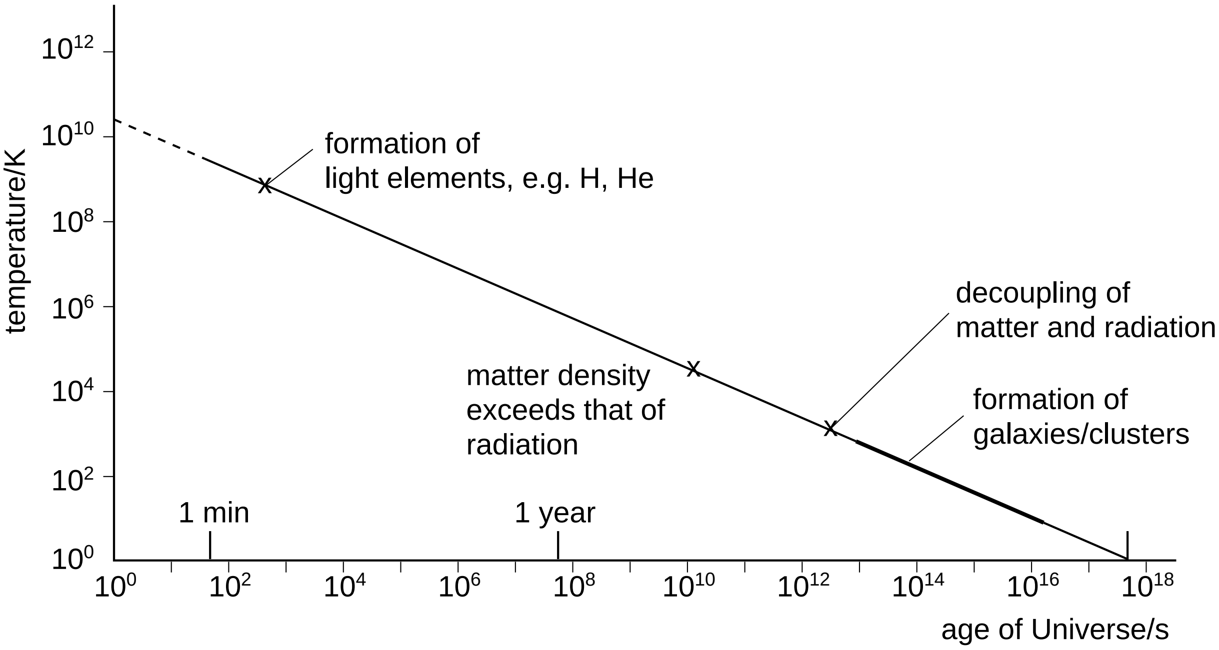


Fig. 2.7

Theorising about the very early stages of a big bang cosmology is supported by observations of the current state of the Universe. This is not yet the case for explaining the formation and distribution of galaxies. This may well change in time as greater understanding of quasars is gained. Quasars are thought to be very distant – and hence very young – galaxies in the process of being formed. (However, the light from these quasars has taken so long to reach the Earth that it is likely that, in fact, the formation has by now occurred but Earth-based observers cannot tell!) There is evidence that quasars may evolve into other types of still 'active' but older galaxies, e.g. galaxies that are very luminous at radio wavelengths and so-called Seyfert galaxies.

This is not to imply, of course, that the earliest stages of big bang cosmology are by any means fully understood. This is evident from the development of the inflationary model of the big bang in place of the 'standard' model and the fact the inflationary approach has not yet been fully developed or recognised. Moreover, there is no direct experimental evidence which relates to processes thought to occur in the evolution of the Universe at times less than about 1 ms. At this age, the temperature of the Universe would have been about 3×10^{11} K. Such extreme temperatures cannot be maintained in bulk in the laboratory. What can be done, however, is to accelerate charged particles, e.g. protons and antiprotons, to speeds that are close to that of the speed of light, and then to cause these particles to collide inelastically. The (kinetic) energy involved in these collisions appears in the form of newly created particles and antiparticles. The machines that achieve these collisions are large, very complex technically and correspondingly expensive to build and run. For example, the new CERN particle accelerator is an underground tunnel with a perimeter measured in kilometres and the highest energies yet achieved in proton-antiproton collisions is of the order of 10^{-8} J per collision. This corresponds to collisions occurring in a plasma, i.e. a gaseous mixture of charged particles, at a temperature of 10^{15} K.

It is thought that, before radiation and matter became decoupled, the temperature of the Universe was inversely proportional to the square root of its age. A temperature of 10^{15} K corresponds to an age of only 10^{-10} s, i.e. one tenth of a billionth of a second or 0.1 ns.

One interaction which is certainly well out of reach of Earth-based experimentation is to test the hypotheses that seek to explain why there is a Universe composed of matter, i.e. one in which an excess of matter over antimatter occurred. As mentioned earlier, the conversion of photons into matter particles is thought to result in the production of equal numbers of such particles and their antiparticles. One theory suggests that the excess of matter over antimatter depends on a very slight asymmetry in the decay rate of the so-called X-boson. This particle is thought to be so massive (as compared to other sub-atomic particles) that it would only have been present in the Universe during the very, very earliest times and at a temperature of about 10^{27} K. The X-boson is thought to have two modes of decay, one being into 2 quarks. Correspondingly, the antiparticle X-boson can decay into 2 antiquarks. The asymmetry is that the X-boson decay is very slightly slower than that of its antiparticle. The difference is of the order of 1 in 10^9, the same ratio as matter particles to photons.

Nevertheless, there is some prospect of this theory being tested experimentally, but only indirectly. Free neutrons decay with a half-life of about 9×10^2 s. Compared with this, protons are extremely stable. The X-boson theory for the excess of matter over antimatter also predicts that protons should decay with a half-life of between 10^{30} years to 10^{32} years – longer than the present age of the Universe! However, if 10^{32} protons are observed, then 1 proton per year might be expected to decay. In 18 kg (18×10^{-3} m^3) of water, there are roughly 10^{24} protons: a thousand tons of water contains about 10^{33} protons. This is not an impossibly large body of water with which to experiment: water is readily obtainable and it is chemically stable. However, other technicalities of incontrovertibly observing the possible decay of a few protons over a few months or years are not easy. No conclusive results have yet been achieved.

There are thought to be four types of force operating in the Universe:

(i) gravity, by far the weakest of the four, which determines the motions of stars and their aggregations in galaxies, clusters of galaxies and superclusters;

(ii) the electromagnetic force, the most readily observable force, which is manifest in everyday physics and chemistry;

(iii) the ‘weak force’, which is important in radioactive decay;

(iv) the ‘strong force’, which is responsible for holding neutrons and protons together in atomic nuclei.

[The latter two types of force have much shorter ranges than the others.]

Progress has been made in developing GUT’s, i.e. Grand Unified Theories, which bring together the latter three forces as manifestations that separate from each other at different times/temperatures of the early Universe. Developing a quantum theory of gravity to ‘marry’ particle physics and Einstein’s theory of General Relativity is proving more difficult. The present state of understanding/theorising is represented by Fig. 2.8 and is illustrated pictorially in Fig. 2.9.

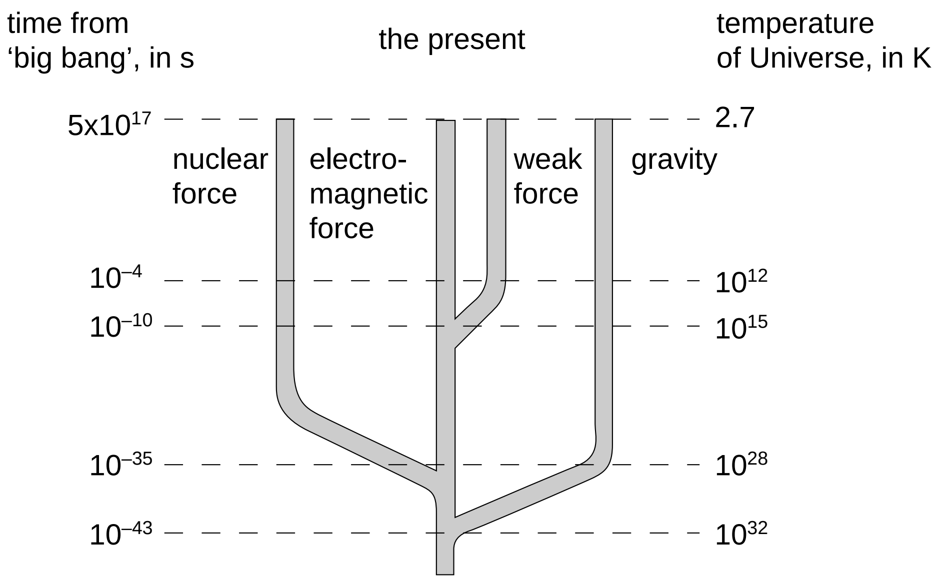

time from
'big bang', in s
the present
temperature
of Universe, in K
5x10^17
2.7
nuclear
force
electro-
magnetic
force
weak
force
gravity
10^–4
10^12
10^–10
10^15
10^–35
10^28
10^–43
10^32

Fig. 2.8

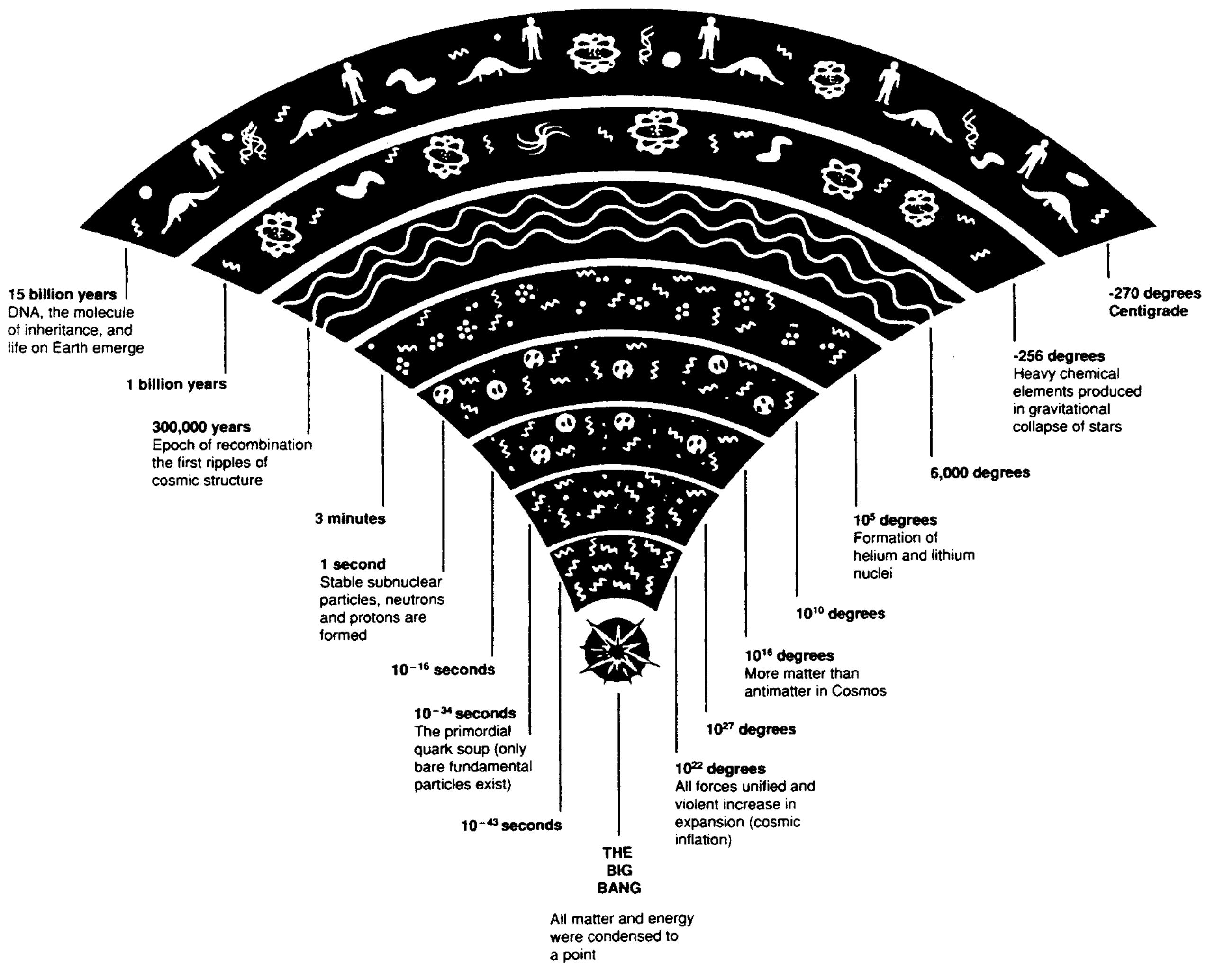

15 billion years
DNA, the molecule
of inheritance, and
life on Earth emerge
1 billion years
300,000 years
Epoch of recombination
the first ripples of
cosmic structure
3 minutes
1 second
Stable subnuclear
particles, neutrons
and protons are
formed
10^-16 seconds
10^-34 seconds
The primordial
quark soup (only
bare fundamental
particles exist)
10^-43 seconds
THE
BIG
BANG
All matter and energy
were condensed to
a point
-270 degrees
Centigrade
-256 degrees
Heavy chemical
elements produced
in gravitational
collapse of stars
6,000 degrees
10^5 degrees
Formation of
helium and lithium
nuclei
10^10 degrees
10^16 degrees
More matter than
antimatter in Cosmos
10^27 degrees
10^32 degrees
All forces unified and
violent increase in
expansion (cosmic
inflation)

Fig. 2.9

A3. Techniques of Observation

3 (a) Candidates should be able to appreciate that stars and galaxies are detected by the electromagnetic radiation which they emit.

Some further discussion of the electromagnetic spectrum in respect of the transparency, or otherwise, of the Earth's atmosphere is given in section *3(c)*. The differing transparency of the Earth's atmosphere to different wavelengths has meant that, from early historical times, the study of astronomy was based on the visible light emitted by stars.

The Sun is a typical yellowish star with a surface temperature of about 6000 K. At this temperature, the dominant radiation happens to be in the yellow region of the electro-magnetic spectrum. Cooler stars, e.g. at about 4000 K, are redder and hotter stars, e.g. at about 10 000 K, are bluer. Given that stars are a significant component of galaxies, the colours of the stars determine the colours of galaxies, either as a whole or in different regions. [It should be realised, however, that the idea that stars are aggregated into galaxies did not become established until the 1920s.]

Although stars and galaxies were detected, i.e. recognised to exist, by their emission of visible light, both objects are detectable by their emissions in other regions of the electro-magnetic spectrum. For example, the Sun is active in the radio-wave region because of the occurrence of sunspots. Both ultraviolet light and X-rays are emitted from the Sun's corona, (i.e. the region lying outside the Sun's photosphere, and which is readily visible during a total eclipse of the Sun).

Other, less common, stars may emit, or cause emissions, in regions other than the visible. For example, stars in the process of being formed may stimulate gas/dust clouds surrounding them to emit in the radio-wave region or to glow, e.g. the Orion nebula, the stars in the Pleiades cluster. The collapse of a red giant into a white dwarf can result in the transitional formation of a so-called planetary nebula, which consists of a spherical shell of ejected material. This shell can be a radio-frequency emitter, stimulated by the radiation absorbed from the central star.

Pulsars, which are rotating neutron stars, can emit a range of wavelengths (due to 'synchrotron' radiation, caused by the motion of electrons of different speeds in the strong magnetic field associated with the neutron star).

In binary star systems with a collapsed star, e.g. a white dwarf or neutron star, the latter may, because of the strength of its gravitational field, draw off material from its companion star. The loss of gravitational potential energy of the material falling on to the collapsed star may result in the emission of ultraviolet light or X-ray radiation.

Other catastrophic events such as novae and supernovae are accompanied by emission both in the visible and other regions of the spectrum.

The events summarised above occur with stars within galaxies and, hence, galactic emission of radiation is not confined to the visible region of the spectrum. Indeed, the galaxies themselves, rather than their stars, emit radiation outside the visible spectrum for other reasons. For example, the dust lane in a spiral galaxy may be hot enough to radiate radio frequencies. The 21 cm line of atomic hydrogen is used to plot galactic rotation, this wavelength being in the radio-wave region. Other galaxies are known as 'active galaxies' because of their radiations in regions other than the visible. Certain large elliptical galaxies have active cores which result in paired radio-emitting lobes which extend across millions of light-years. Such active cores also cause the emission of X-rays and their activity is believed to be due to the presence of massive black holes in their cores. There is more to such galaxies than meets the eye!

3 (b) *Candidates should be able to appreciate that planets are detected by reflected sunlight.*

The visual recognition of galaxies as astronomical objects in their own right is a 20th Century discovery. On the other hand, the Sun, the Moon, stars and the inner planets have been known – visually – from antiquity. However, the recognition of planets as being different from stars was a later discovery.

There is a great difference between stars and planets. The former visibly shine because of the energy-releasing processes that occur within them. Planets, by contrast, are visible only by reflected sunlight. The factors which affect the brightness of a planet are its distance from the Sun and from the Earth and its 'albedo'.

The albedo of a planet is a measure of how well its surface reflects incident sunlight. The albedo of Venus is relatively high because it is cloud-covered. (It is, therefore, a bright planet – well known from its reddish appearance as the 'Evening Star' as the sky darkens.)

However, planets are not totally cold and are, therefore, radiating at electro-magnetic wavelengths commensurate with their (surface) temperature. For example, the surface temperature of Jupiter is typically 125 K and it radiates at radio wavelengths of the order of 25 μm. Both Jupiter and Saturn have an internal source of energy, possibly due to on-going contraction and the consequential release of gravitational potential energy as heat. Jupiter is also X-ray active, having aurorae similar to those that occur on Earth – the magnetic field of Jupiter is much stronger than that of the Earth. The actual patterns and variations of these different radiations gives information about the structure and dynamics of planetary atmospheres.

Nevertheless, it is still the case that, while not totally 'dark' as radiators, it is as reflectors of visible light that planets make their presence readily known in the night sky.

3 (c) *Candidates should be able to describe the transparency of the Earth's atmosphere to different regions of the electromagnetic spectrum from radio waves to X-rays.*

3 (d) *Candidates should be able to explain why the transparency of the Earth's atmosphere has led to observations which are terrestrial, high-altitude, from satellites or from space probes.*

Fig. 3.1 illustrates the relative absorption of the Earth's atmosphere for different wavelengths in the electromagnetic spectrum from radio waves to X-rays.

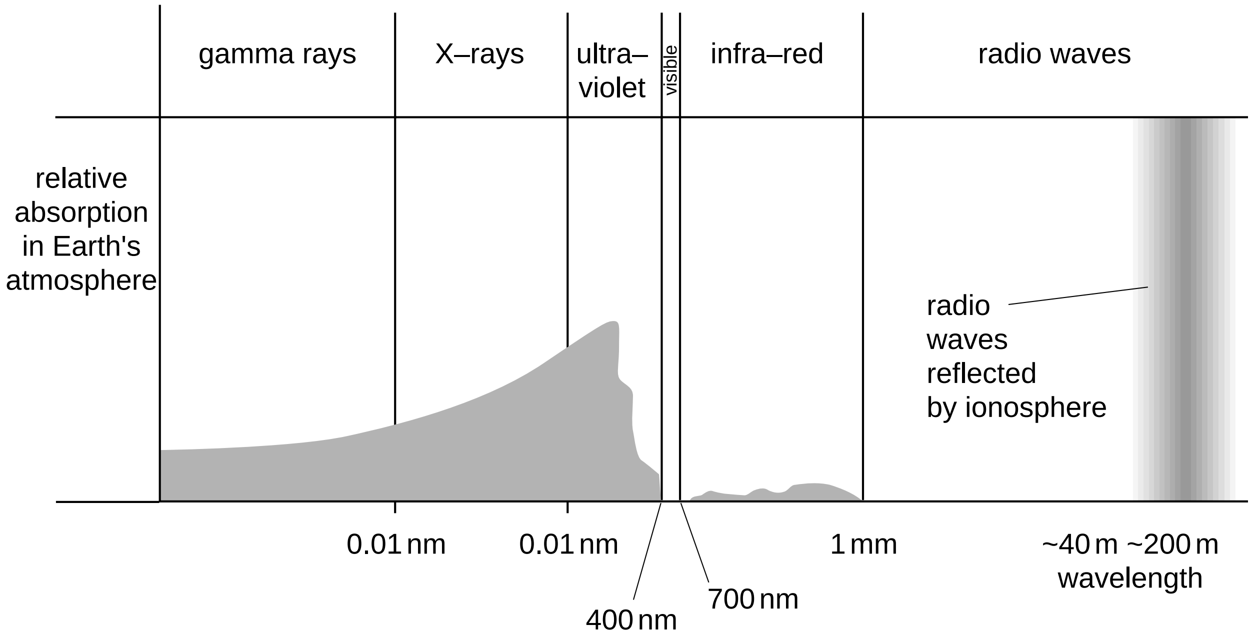


Fig. 3.1

It can be recognised that the Earth's atmosphere behaves as a 'window' only for a limited number of ranges of wavelength or frequency. These are:

(i) for radio waves, of wavelength about a millimetre and longer,

(ii) the visible region from 400 nm (blue) to 700 nm (red),

(iii) the near ultraviolet.

Lack of transparency means, of course, that absorption occurs. This, in turn, implies interaction between the radiation being absorbed and the gases (or dust) present in the atmosphere. Whether or not such interaction occurs depends on the quantum energy of the photon concerned and the corresponding energy change in the absorbing matter. Photons of lower frequency (longer wavelength) have less energy than photons of higher frequency (shorter wavelength).

The quantum energy of a photon of radio frequency is too low to interact with the gaseous molecules in the atmosphere, which is thus transparent to radio waves. [Radio waves do, of course, interact with the 'free' electrons in a radio aerial. Indeed, electromagnetic radiation only interacts with electric charge, whether 'free' electrons in an aerial, the electrons in a chemical bond or exciting an electron within an atom, ion or molecule.]

Increasing the amplitude of vibration in a molecule requires less energy than electronic excitation of a molecule. Lack of transparency in the infra-red and microwave regions is associated with absorption due to molecular vibration, e.g. carbon dioxide and water vapour.

The quantum energy of high frequency ultraviolet rays and X-rays is sufficiently high to cause chemical reaction by breaking bonds. For example absorption, in the u.v., is associated with the photodissociation of ozone in the upper atmosphere (and of the atmospheric pollutants nitrogen dioxide and chlorofluorocarbons, CFCs).

As will be recognised from several previous references, astronomical observations in the past have been confined both to the visible region of the spectrum and to the Earth's surface.

There are, however, disadvantages to terrestrial observations in the visible region. The transparency of the atmosphere to visible radiation is variable. Cloud and fog can either block out visible light, particularly if the light is of low intensity, e.g. from a star, or they can scatter light. (Scattering explains why sunsets appear red. Being of shorter wavelength, blue light is more readily scattered by airborne particles. The red light is less scattered and hence is dominant.) On the other hand, water droplets (as opposed to water vapour) have no effect on radio waves because of the long wavelengths. For this reason, weather conditions do not affect radio astronomy observations.

The scattering of light from terrestrial light sources is disadvantageous for 'visible' astronomy, the effect being called 'light pollution'. Urban lighting is now so intense and widespread that atmospheric scattering reduces the effective transparency for a telescope located near a town or city.

An urban location of a telescope has a further disadvantage. Towns and cities cause an increase in the thermal currents in the atmosphere. Such currents manifest themselves in random fluctuations in the refractive index of the air. The refractive index changes cause refraction effects. As a result, the image of a star as formed by the telescope is degraded: the 'seeing' is said to be poorer. (It is because of natural refractive index fluctuations in the atmosphere that stars appear to twinkle. Being so very far away, stars are, in effect, point objects. The random movement of the image formed by the eye due to random changes in the refractive index of the air through which the starlight passes on its way to the eye is detectable as a twinkle. Planets do not twinkle. Planetary images have an angular diameter sufficiently large to mask the twinkle effect.)

Infra-red astronomy can also be hampered by water vapour and other atmospheric constituents. Radioastronomy can suffer from electrically generated noise created by human activity. Indeed, the Earth is a disproportionately 'bright' object in terms of radio emissions, both deliberate and incidental, from man-made devices. For example, suppressors are commonly fitted to electric motors to avoid television reception being spoilt. Reducing radio interference to an acceptable level is one reason why the well known Cambridge radio telescope was deliberately sited in a shallow, relatively sparsely populated valley. [It may be of interest to note, in passing, that when T.V. broadcasting is finished on a particular channel but the set is still switched on, some of the 'snowflake' noise is due to the cosmic microwave background radiation.]

Terrestrial astronomical observation is, therefore, limited to radio waves, a limited number of i.r. wavelengths, light waves and the near ultraviolet. Particularly for the latter two short-wavelength regions, better observing conditions can be achieved by siting the telescope at a high altitude (e.g. on a mountain peak) – although this can also be beneficial for i.r. astronomy. Such high altitudes offer the advantages of

(i) being less often cloud bound, by being above the typical cloud level,

(ii) clearer skies, being less polluted in respect of dust and of scattered light from nearby built-up areas,

(iii) air patterns which are less disturbed thermally.

Other potential advantages include drier air (water vapour pressure being lower at lower temperatures) and unobstructed horizons.

These advantages are carried over – and even enhanced – for satellite telescopes. Radio astronomy by satellite is less subject to Earth-based radio noise. Because of lack of transparency of the atmosphere, ultraviolet and X-ray observation can, in effect, only be carried out by instruments taken above the Earth's atmosphere.

By being projected above the atmosphere, instruments on board space probes have similar advantages as satellite-borne instruments. Space probes are, however, more expensive than satellites and need more power for transmitting signals back to the Earth. On the other hand, they do have some advantages over satellites. By being carried closer to the target object, space probe instruments produce images of greater resolving power, i.e. the images show finer detail: they can also obtain views, e.g. from the rear, not possible from an observing platform 'tied' to the Earth. Other types of instrument can also be carried, such as for measuring magnetic fields, the solar wind and the degree of polarisation in e.m. sources.

3 (e) Candidates should be able to show an awareness of the conflict between the value of astronomical research and economic considerations.

There is an apochryphal story about Michael Faraday when, in the early 19th Century, he first demonstrated electromagnetic induction. It is said that in reply to the question "What use is it?", he retorted, "What use is a baby?".

This story, true or not, illustrates the tension that exists, and which has become more pronounced in recent years, between pure and applied research. There is, moreover, a third element, namely that of technology. These three elements are mutually interdependent.

Apart from being interrelated (as referred to in section *2(j)*), astrophysics and particle physics have other properties in common as regards research. The research relies heavily on advanced technology and, partly as a consequence of this, it is intrinsically very expensive. Putting even a satellite into orbit, let alone sending a space probe off on its journey, makes heavy demands on resources, whether of manpower, time, equipment or money. An important characteristic of a rocket is its payload. Likewise, the question "What is the pay-off?" may be asked. It is, in some degree, unfortunate that a possible answer is "You never can tell".

Faraday's electromagnetic induction research may be said to have been 'pure', although he was, apparently, well aware of the importance of being able to apply new knowledge. By contrast, present day research into superconductivity has moved from the stage of academic interest into trying to develop materials which are superconducting at more convenient temperatures, there already being considerable awareness of the potential benefits of having such materials. Similarly, whether the development of the laser is more properly called pure or applied is perhaps a non-issue. There is no doubt, however, about how widespread the applications of laser technology have become. The maser, the microwave equivalent of the laser, i.e. light amplification by the stimulated emission of radiation, was developed before the laser. It is perhaps of interest that the maser mechanism has been invoked to explain the unexpectedly intense radiation, at a wavelength of 18 cm, due to the OH radicals observed in certain galactic molecular clouds.

It is thought that almost immediately after the big bang the newly created Universe underwent a very short, but very intense, period of inflationary expansion. In economics, inflation occurs when too much money ‘chases’ too few goods. Research is, by contrast, deflationary: there are too few resources to meet the research demand, especially when the research is, as for astronomy, inherently expensive. It is a truism that applications that depend on basic understanding cannot take place without that basic understanding having been gained. A difficulty is that, as for research in nuclear fusion over some thirty years, achieving success in research cannot be guaranteed. Difficult decisions may well have to be made about supporting particular pieces of research, however desirable that research may appear to be. It is noteworthy, nevertheless, that the expensive rescue of the Hubble telescope in what would otherwise have been a seriously wasted investment appears to be paying dividends in the superb quality of the observations now being made and reported.