1 · The Solar System

The Solar System is made up of one star (the Sun), the eight planets with their satellites (Table 1.1) and various minor members such as asteroids, comets and meteoroids, plus a vast amount of thinly spread interplanetary matter. The Sun contains 99.86% of the total mass of the System, while Jupiter and Saturn account for 90% of what is left. Jupiter is the largest member of the planetary family, and is in fact more massive than all the other planets combined. Mainly because of Jupiter, the centre of gravity of the Solar System lies just outside the surface of the Sun.

The Solar System is divided into two parts. There are four comparatively small, rocky planets (Mercury, Venus, the Earth and Mars), beyond which comes the zone of the Main-Belt asteroids, of which only one (Ceres) is over 900 km in diameter. Next come the four giants (Jupiter, Saturn, Uranus and Neptune), plus a swarm of trans-Neptunian objects, of which the largest known are Eris and Pluto. For many years after its discovery, in 1930, Pluto was regarded as a true planet, but in August 2006 the International Astronomical Union, the controlling body of world astronomy, introduced a new scheme of classification, as follows:

A **planet** is any body in orbit round the Sun which is massive enough to assume a spherical shape, and has cleared its immediate neighbourhood of all smaller objects. All these criteria are met by the eight familiar planets, from Mercury to Neptune.

A **dwarf planet** is spherical, but has not cleared its neighbourhood. Three were listed: Eris, Pluto and Ceres.

Small solar system bodies (SSSBs) are other bodies orbiting the Sun.[1]

Natural satellites are objects in orbit round planets, dwarf planets or SSSBs rather than directly round the Sun itself.

Distances from the Sun are conventionally given in **astronomical units** (a.u.). The a.u. is defined as the mean distance between the Earth and the Sun: in round numbers 149 600 000 km (93 000 000 miles in Imperial measure). Jupiter is approximately 5.2 a.u. from the Sun; one light-year, used for interstellar distances, is equal to 63 240 a.u.

It now seems that the distinctions between the various classes of bodies in the Solar System are much less clear-cut than used to be thought. For example, it may well be that some 'near-Earth' asteroids, which swing inward away from the main swarm, are ex-comets which have lost all their volatiles, and many of the small planetary satellites are certainly captured SSSBs.

All planets, dwarf planets and SSSBs move round the Sun in the same sense, and (with one exception) so do the larger satellites orbiting their primary planets, though many of the tiny 'asteroidal' satellites move in the opposite (retrograde) sense. The orbits of the planets are not strongly eccentric, and are not greatly inclined to that of the Earth, so that to draw a plan of the main Solar System on a flat piece of paper is not grossly inaccurate. However, dwarf planets and SSSBs may have paths which are more eccentric and inclined, and comets come into a different category altogether. Those with periods of a few years or a few tens of years have direct motion, but brilliant comets come from the depths of space, and often travel in a retrograde sense. Their periods may amount to centuries, or to thousands or even millions of years.

It is also notable that six of the planets rotate in the same sense as the Earth, though the axial periods are different – over 58 Earth days for Mercury, less than 10 hours for Jupiter. The exceptions are Venus, which has retrograde rotation, and Uranus, where the rotational axis is tilted to the orbital plane by 98 degrees, more than a right angle. The cause of these anomalies is unclear.

ORIGIN AND EVOLUTION OF THE SOLAR SYSTEM

In investigating the past history of the Solar System, we do at least have one important piece of information: the age of the Earth is 4.6 thousand million years, and the Sun, in some form or other, must be older than this. We are entitled to be confident about the Earth's age, because there are several reliable methods of research, and all give the same value. There are no modern dissentients, apart of course from the Biblical Fundamentalists.

Many theories have been proposed. Of particular note is the 'Nebular Hypothesis', usually associated with the name of the eighteenth-century French astronomer Pierre Simon de Laplace, though he was not actually the first to describe it; the original idea was put forward in 1734 by Emanuel Swedenborg, of Sweden, who carried out useful scientific work but who is best remembered today for his later somewhat eccentric theories (he was on excellent terms with a number of angels, and gave graphic accounts of life on all the planets!). Swedenborg's suggestion was elaborated by Thomas Wright in England and Immanuel Kant in Germany, but the Nebular Hypothesis in its final form was due to Laplace, in 1796.

Laplace started with a vast hydrogen gas-cloud, disc-shaped and in slow rotation; it shrank steadily and threw off rings, each of which produced a planet, while the central part of the cloud – the so-called solar nebula – heated up as the atoms within it began to collide with increasing frequency. Eventually, when the temperature had risen sufficiently, the Sun had been born, and the planets were in orbits which were more or less in the same plane. All seemed well – until mathematical analysis showed that

Table 1.1 *Basic data for the planetary system*

Name	Mean distance from Sun (km)	Orbital period	Orbital eccentricity	Orbital inclination	Equatorial diameter (km)	Equatorial rotation period	Number of satellites
Mercury	57 900 000	87.97 d	0.206	7° 0′ 15″ .5	4878	58.6 d	0
Venus	108 200 000	224.7 d	0.007	178°	12 104	243.2 d	0
Earth	149 598 000	365.25	0.017	0	12 756	23h 56m 4s	1
Mars	227 940 000	687.0 d	0.093	1° 51′	6794	24h 37m 23s	2
Jupiter	778 340 000	11.86 y	0.048	1° 18′ 16″	143 884	9h 50m 30s	63
Saturn	1427 000 000	29.5 y	0.056	2° 29′ 21″	120 536	10h 14m	61
Uranus	2869 600 000	84.0 y	0.047	0° 46′ 23″	51 118	17h 14m	27
Neptune	4496 700 000	164.8 y	0.009	1° 34′ 20″	50 538	16h 6m	13

a thrown-off ring would not condense into a planet at all; it would merely disperse. There were other difficulties, too. Most of the angular momentum of the system would reside in the Sun, which would be in rapid rotation; actually, most of the angular momentum is due to the planets, and the Sun is a slow spinner (its axial rotation period amounts to several Earth weeks). In its original form, the Nebular Hypothesis had to be given up.

In 1901, T. C. Chamberlin and F. R. Moulton proposed an entirely different theory, according to which the planets were pulled off the Sun by a passing star. The visitor's gravitational pull would tear out a cigar-shaped tongue of material, and this would break up into planets, with the largest planets (Jupiter and Saturn) in the middle part of the system, where the thickest part of the 'cigar' would have been. Again there were fatal mathematical objections, and a modification of the idea by A.W. Bickerton (New Zealand), involving 'partial impact,' was no better. However, the theory in its original form remained in favour for some time, particularly as it was supported by Sir James Jeans, a leading British astronomer who was also the author of popular books on astronomy which were widely read (and in fact still are). Had it been valid, planetary systems would have been very rare in the Galaxy, because close encounters between stars seldom occur. As we now know, this is very far from being the truth. Another modification was proposed later by G. P. Kuiper, who believed that the Sun somehow acquired enough material to produce a binary companion, but that this material never formed into a true star; the planets could be regarded as stellar débris. This idea never met with much support.

In many ways our current theories are not too unlike the old Nebular Hypothesis. We do indeed begin with a gas-and-dust cloud, which began to collapse, and also to rotate, possibly because of the gravitational pull of a distant supernova. The core turned into what we call a proto-star, and the solar nebula was forced into the form of a flattened disc. As the temperature rose, the proto-star became a true star – the Sun – and for a while went through what is called the T Tauri stage, sending out a strong 'stellar wind' into the cloud and driving out the lightest gases, hydrogen and helium. (The name has been given because the phenomenon was first found with a distant variable star, catalogued as T Tauri.) The planets

built up by accretion. The inner, rocky planets lacked the gas which had been forced out by the stellar wind but, further away from the Sun, where the temperature was much lower, the giant planets were able to form and accumulate huge hydrogen-rich atmospheres. Jupiter and Saturn accreted first; Uranus and Neptune built up later, when much of the hydrogen had been dispersed. This is why they contain less hydrogen and more icy materials than their predecessors. It is fair to say that Jupiter and Saturn are true gas-giants, while Uranus and Neptune are better described as ice-giants.

In the early history of the Solar System there was a great deal of 'left-over' material. Jupiter's powerful pull prevented a planet from being formed in the zone now occupied by the Main-Belt asteroids; further out there were other asteroid-sized bodies which make up the Kuiper Belt. All the planets were subjected to heavy bombardment, and this is very evident; all the rocky planets are thickly cratered, and so are the satellites – including our Moon, where the bombardment went on for several hundreds of millions of years. (Earth was not immune, but by now most of the terrestrial impact craters have been eroded away or subducted.) It is widely believed that the gas-giants, particularly Jupiter, have acted as shields, protecting the inner planets from even more devastating bombardment. See Table 1.2 for planetary and satellite feature names.

In other ways, too, the young Solar System was very different from that of today. The Sun was much less luminous, so that, for example, Venus may well have been no more than pleasantly tropical. It is also likely that there was an extra planet in the inner part of the System, which collided with the proto-Earth and produced the Moon (though there are differing views about this). The outer planets at least may not have been in their present orbits, and interactions with each other and with general débris is thought to have caused 'planetary migration'; it has even been suggested that at one stage Uranus, not Neptune, was the outermost giant. We cannot pretend that we know all the details about the evolution of the Solar System, but at least we can be confident that we are on the right track.

How far does the Solar System extend? It is difficult to give a precise answer. The main System ends at the orbit of Neptune (unless there is a still more remote giant, which is unlikely though

Table 1.2 *Planetary and satellite feature names*

Arcus (arcus)	Arc-shaped feature
Catena (catenæ)	Chain of craters
Cavus (cavi)	Hollows; irregular steep-sided depressions
Chaos	Irregular area of broken terrain
Chasma (chasmata)	Deep, elongated, steep-sided depression
Colles	Small hills
Corona (coronæ)	Ovoid-shaped feature
Dorsum (dorsa)	Ridge
Facula (faculæ)	Bright spot
Farrum (farra)	Pancake-shaped structure
Flexus (flexûs)	Low curvilinear ridge
Fluctus (fluctûs)	Flow terrain
Flumen (flumina)	Channel that might carry liquid
Fossa (fossæ)	'Ditch'; long, narrow depression
Insula (insulæ)	Island
Labes (labes)	Landslide
Labyrinthus (labyrinthi)	Complex of intersecting ridges or valleys
Lacus	Lake
Lenticula (lenticulæ)	Small dark spot
Linea (lineæ)	Dark or bright elongated marking, either curved or straight
Macula (maculæ)	Dark spot or patch
Mare (maria)	'Sea'; large, comparatively smooth plain
Mensa (mensæ)	Flat-topped prominence with cliff-like edges
Mons (montes)	Mountain
Oceanus	'Ocean'; very large dark plain
Palus (paludes)	'Marsh'; small, often irregular plain
Patera (pateræ)	Irregular crater-like structure with scalloped edges
Planitia (planitiæ)	Low-lying plain
Planum (plana)	Plateau, or high plain
Promontorium (promontoria)	'Cape' (promontory)
Regio (regiones)	Large area, clearly different from adjacent areas
Rill (rills)	Crack-like feature (also spelled 'rille')
Rima (rimæ)	Fissure
Scopulus (scopuli)	Lobate or irregular scarp
Sinus (sinus)	'Bay'
Sulcus	Groove or trench
Tessera (tesseræ)	'Parquet' (tile-like, polygonal terrain)
Tholus (tholi)	Small, dome-like hill
Undæ	Dunes
Vallis (valles)	Valley
Vastitas	Extensive plain
Virga (virgæ)	Coloured streak

not impossible), but comets and many trans-Neptunians recede to much greater distances, and the Oort Cloud lies well over a light year away. The nearest stars beyond the Sun, those of the α Centauri group, are just over four light years away. Therefore, it seems fair to say that the effective border of the Solar System is of the order of two light years from us.

At present the Solar System is essentially stable, but this state of affairs cannot last for ever. The Sun is becoming steadily more luminous, and in no more than four thousand million years will have swelled out to become a red giant star, far more powerful than it is today. Mercury and Venus will be destroyed; Earth may survive, because the Sun's loss of mass will weaken its gravitational pull, and the planets will spiral outward to a limited degree. Yet even if our world does survive, it will be in the form of a red-hot, seething mass. Next, the Sun will collapse to become a tiny, feeble, super-dense white dwarf star, and scorching heat will be replaced by numbing cold. In the end the Sun will lose the last of its power, and will become a dead black dwarf, perhaps still attended by the ghosts of its remaining planets. It is even possible that following the merger between our Galaxy and the Andromeda Spiral, the Solar System, or rather, what is left of it – may end up in the outer part of the Milky Way, or in the depths of intergalactic space.

However, for us, all these crises lie so far ahead that we cannot predict them really accurately. We know that the Solar System has a limited lifetime, but as yet it is no more than middle-aged.

ENDNOTE

1 I was a member of that IAU Commission for many years, but felt bound to retire in 2001 as I was no longer able to travel to meetings. Had I been present at the 2006 meeting I would have put forward some alternative proposals, because the Resolution, as passed, seems to be unclear. Of the two largest Main-Belt asteroids, why should Ceres be a dwarf planet and Pallas an SSSB? Of the trans-Neptunians, a good many, such as Quaoar and Varuna, are considerably larger than Ceres. I would have retained the main planets and their satellites, and lumped the rest together as 'planetoids'. Moreover, one can hardly regard a comet as a 'small' body; the coma of Holmes' Comet of 2007 was larger than the Sun, though admittedly its mass was negligible. I suspect that the 2006 Resolution will be revised before long, but meanwhile it must be accepted.

2 · The Sun

The Sun, the controlling body of the Solar System, is the only star close enough to be studied in detail. It is 270 000 times closer than the nearest stars beyond the Solar System, those of the α Centauri group. Data are given in Table 2.1.

DISTANCE

The first known estimate of the distance of the Sun was made by the Greek philosopher Anaxagoras (500–428 BC). He assumed the Earth to be flat, and gave the Sun's distance as 6500 km (using modern units), with a diameter of over 50 km. A much better estimate was made by Aristarchus of Samos, around 270 BC. His value, derived from observations of the angle between the Sun and the exact half Moon, was approximately 4 800 000 km; his method was perfectly sound in theory, but the necessary measurements could not be made with sufficient accuracy. (Aristarchus also held the belief that the Sun, not the Earth, is the centre of the planetary system.) Ptolemy (*c*. AD 150) increased the distance to 8 000 000 km, but in his book published in AD 1543 Copernicus reverted to only 3 200 000 km. Kepler, in 1618 gave a value of 22 500 000 km.

The first reasonably accurate estimate of the Earth–Sun distance (the astronomical unit) was made in 1672 by Giovanni Cassini, from observations of the parallax of Mars. Some later determinations are given in Table 2.2.

One early method involved transits of Venus across the face of the Sun, as suggested by J. Gregory in 1663 and extended by Edmond Halley in 1678; Halley rightly concluded that transits of Mercury could not give accurate results because of the smallness of the planet's disc. In fact, the transit of Venus method was affected by the 'Black Drop' – the apparent effect of Venus drawing a strip of blackness after it during ingress on to the solar disc, thus making precise timings difficult. (Captain Cook's famous voyage, during which he discovered Australia, was made in order to take the astronomer C. Green to a suitable site (Tahiti) in order to observe the transit of 1769.)

Results from the transits of Venus in 1874 and 1882 were still unsatisfactory, and better estimates came from the parallax measurements of planets and (particularly) asteroids. However, Spencer Jones' value as derived from the close approach of the asteroid Eros in 1931 was too high. The modern method – radar to Venus – was introduced in the early 1960s by astronomers in the United States. The present accepted value of the astronomical unit is accurate to a tiny fraction of 1%.

THE SUN IN THE GALAXY

The Sun lies close to the inner ring of the Milky Way Galaxy's Orion Arm. It is contained within the Local Bubble, an area of rarefied high-temperature gas (caused by a supernova outburst?). The distance between our local arm and the next one out, the Perseus Arm, is about 6500 light-years. The Sun's orbit is somewhat elliptical, and passes through the galactic plane about 2.7 times per orbit. The Sun has so far completed from 20 to 25 orbits (20 to 25 'cosmic years').

ROTATION

The first comments about the Sun's rotation were made by Galileo, following his observations of sunspots from 1610. He gave a value of rather less than one month.

The discovery that the Sun shows differential rotation – i.e. that it does not rotate as a solid body would do – was made by the English amateur Richard Carrington in 1863; the rotational period at the equator is much shorter than that at the poles. Synodic rotation periods for features at various heliographic latitudes are given in Table 2.3. Spots are never seen either at the poles or exactly on the equator, but from 1871 H. C. Vogel introduced the method of measuring the solar rotation by observing the Doppler shifts at opposite limbs of the Sun.

THE SOLAR CONSTANT

The solar constant may be defined as being the amount of energy in the form of solar radiation per second which is vertically incident per unit area at the top of the Earth's atmosphere: it is roughly equal to the amount of energy reaching ground level on a clear day. The first measurements were made by Sir John Herschel in 1837–8, using an actinometer (basically a bowl of water; the estimate was made by the rate at which the bowl was heated). He gave a value which is about half the actual figure. The modern value is 1.95 cal cm^{-2} min^{-1} (1368 Wm^{-2}).

SOLAR PHOTOGRAPHY

The first photograph of the Sun – a Daguerreotype – seems to have been taken by Lerebours, in France, in 1842. However, the first good Daguerreotype was taken by Fizeau and Foucault, also in France, on 2 April 1845, at the request of F. Arago. In 1854 B. Reade used a dry collodion plate to show mottling on the disc.

Table 2.1 *The Sun: data*

Distance from Earth:
    mean   149 597 893 km (1 astronomical unit (a.u.))
    max.   152 103 000 km
    min.   147 104 000 km
Mean parallax: $8'' .794$
Distance from centre of the Galaxy: ~26 000 light-years
Velocity round centre of Galaxy: ~250 km s^{-1}
Period of revolution round centre of Galaxy: ~225 000 000 years
(1 'cosmic year')
Velocity toward solar apex: 19.5 km s^{-1}
Apparent diameter: mean $32'$ $01''$
                           max. $32'$ $25''$
                           min. $31'$ $31''$
Equatorial diameter: 1 391 980 km
Density, water = 1: mean 1.409
Volume, Earth = 1: 1 303 600
Mass, Earth = 1: 332 946
Mass: 2×10^{27} tonnes (>99% of the mass of the entire Solar System)
Surface gravity, Earth = 1: 27.90
Escape velocity: 617.7 km s^{-1}
Luminosity: 3.85×10^{23} kW
Solar constant (solar radiation per second vertically incident at unit area at 1 a.u. from the Sun); 1368 W m^{-2}
Mean apparent visual magnitude: −26.78 (600 000 times as bright as the full Moon)
Absolute magnitude: +4.82
Spectrum: G2
Temperature: surface 5500 °C
                      core ~15 000 000 °C
Rotation period: sidereal, mean: 25.380 days
                       synodic, mean: 27.275 days
Time taken for light to reach the Earth, at mean distance: 499.012 s
(8.3 min)
Age: ~4.6 thousand million years

The first systematic series of solar photographs was taken from Kew (outer London) from 1858 to 1872, using equipment designed by the English amateur Warren de la Rue. Nowadays the Sun is photographed daily from observatories all over the world, and there are many solar telescopes designed specially for this work. Many solar telescopes are of the 'tower' type, but the largest solar telescope now in operation, the McMath Telescope at Kitt Peak in Arizona, looks like a large, white inclined tunnel. At the top is the upper mirror (the heliostat), 203 cm in diameter; it can be rotated, and sends the sunlight down the tunnel in a fixed direction. At the bottom of the 183 m tunnel is a 152 cm mirror, which reflects the rays back up the tunnel on to the halfway stage, where a flat mirror sends the rays down through a hole into the solar laboratory where the analyses are carried out. This means that the heavy equipment in the solar laboratory does not have to be moved at all.

SUNSPOTS

The bright surface of the Sun is known as the photosphere, composed mainly of hydrogen and helium and it is here that we see the dark patches which are always called sunspots. Really large spot-groups may be visible with the naked eye, and a precisely dated Chinese record from as far back as 28 BC describes a patch which was 'a black vapour as large as a coin'. There is a Chinese record of an 'obscuration' in the Sun, which may well have been a spot, as early as 800 BC.

The first observer to publish telescope drawings of sunspots was J. Fabricius, from Holland, in 1611, and although his drawings are undated he probably saw the spots toward the end of 1610. C. Scheiner, at Ingoldstädt, recorded spots in March 1611, with his pupil J. Cysat. Scheiner wrote a tract which came to the notice of Galileo, who claimed to have been observing sunspots since November 1610. No doubt all these observers recorded spots telescopically at about the same time (the date was close to solar maximum when spot groups should have been frequent) but their interpretations differed. Galileo's explanation was basically correct. Scheiner regarded the spots as dark bodies moving round the Sun close to the solar surface; Cassini, later, regarded them as mountains protruding through the bright surface. Today we know that they are due to the effects of bipolar magnetic field lines below the visible surface.

Direct telescopic observation of the Sun through any telescope is highly dangerous, unless special filters or special equipment is used. The first observer to describe the projection method of studying sunspots may have been Galileo's pupil B. Castelli. Galileo himself certainly used the method, and said (correctly) that it is 'the method that any sensible person will use'. This seems to dispose of the legend that he ruined his eyesight by looking straight at the Sun through one of his primitive telescopes.

A major spot consists of a darker central portion (umbra) surrounded by a lighter portion (penumbra); with a complex spot there may be many umbræ contained in one penumbral mass. Some 'spots' at least are depressions, as can be seen from what is termed the Wilson effect, announced in 1774 by A. Wilson of Glasgow. He found that with a regular spot, the penumbra toward the limbward side is broadened, compared with the opposite side, as the spot is carried toward the solar limb by virtue of the Sun's rotation. From these observations, dating from 1769, Wilson deduced that the spots must be hollows. The Wilson effect can be striking, although not all spots and spot-groups show it.

Some spot-groups may grow to immense size. The largest group on record is that of April 1947; it covered an area of 18 130 000 000 km^2, reaching its maximum on 8 April. To be visible with the naked eye, a spot-group must cover 500 millionths of the visible hemisphere. (One millionth of the hemisphere is equal to 3 000 000 km^2.)

A large spot-group may persist for several rotations. The present record for longevity is held by a group which lasted for 200 days, between June and December 1943. On the other hand, very small spots, known as pores, may have lifetimes of less than an hour. A pore is usually regarded as a feature no more than 2500 km in diameter.

The darkest parts of spots – the umbræ – have temperatures of around 4000 °C, while the surrounding photosphere is at well over

Table 2.2 *Selected estimates of the length of the astronomical unit*

Year	Authority	Method	Parallax (arcsec)	Distance (km)
1672	G. D. Cassini	Parallax of Mars	9.5	138 370 000
1672	J. Flamsteed	Parallax of Mars	10	130 000 000
1770	L. Euler	1769 transit of Venus	8.82	151 225 000
1771	J. de Lalande	1769 transit of Venus	8.5	154 198 000
1814	J. Delambre	1769 transit of Venus	8.6	153 841 000
1823	J. F. Encke	1761 and 1769 transits of Venus	8.5776	153 375 000
1867	S. Newcomb	Parallax of Mars	8.855	145 570 000
1877	G. Airy	1874 transit of Venus	8.754	150 280 000
1877	E. T. Stone	1874 transit of Venus	8.884	148 080 000
1878	J. Galle	Parallax of asteroids Phocæa and Flora	8.87	148 290 000
1884	M. Houzeau	1882 transit of Venus	8.907	147 700 000
1896	D. Gill	Parallax of asteroid Victoria	8.801	149 480 000
1911	J. Hinks	Parallax of asteroid Eros	8.807	149 380 000
1925	H. Spencer Jones	Parallax of Mars	8.809	149 350 000
1939	H. Spencer Jones	Parallax of asteroid Eros	8.790	149 670 000
1950	E. Rabe	Motion of asteroid Eros	8.798	149 526 000
1962	G. Pettengill	Radar to Venus	8.794 0976	149 598 728
1992	Various	Radar to Venus	8.794 148	149 597 871

Table 2.3 *Synodic rotation period for features at various heliographic latitudes*

Latitude (°)	Period (days)
0	24.6
10	24.9
20	25.2
30	25.8
40	27.5
50	29.2
60	30.9
70	32.4
80	33.7
90	34.0

Table 2.4 *Zürich sunspot classification*

A	Small single unipolar spot, or a very small group of spots without penumbræ.
B	Bipolar sunspot group with no penumbræ.
C	Elongated bipolar sunspot group. One spot must have penumbræ.
D	Elongated bipolar sunspot group with penumbræ on both ends of the group.
E	Elongated bipolar sunspot group with penumbræ on both ends. Longitudinal extent of penumbræ exceeds 10° but not 15°.
F	Elongated bipolar sunspot group with penumbra on both ends. Longitudinal extent of penumbræ exceeds 15°.
H	Unipolar sunspot group with penumbræ.

5000 °C. This means that a spot is by no means black, and if it could be seen shining on its own the surface brightness would be greater than that of an arc-lamp. The accepted Zürich classification of sunspots is given in Table 2.4.

Sunspots are essentially magnetic phenomena, and are linked with the solar cycle. Every 11 years or so the Sun is at its most active, with many spot-groups and associated phenomena; activity then dies down to a protracted minimum, after which activity builds up once more toward the next maximum. A typical group has two main spots, a leader and a follower, which are of opposite magnetic polarity.

The magnetic fields associated with sunspots were discovered by G. E. Hale, from the United States, in 1908. This resulted from the Zeeman effect (discovered in 1896 by the Dutch physicist P. Zeeman), according to which the spectral lines of a light source are split into two or three components if the source is associated with a magnetic field. It was Hale who found that the leader and the follower of a two-spot group are of opposite polarity – and that the conditions are the same over a complete hemisphere of the Sun, although reversed in the opposite hemisphere. At the end of each cycle the whole situation is reversed, so that it is fair to say that the true cycle (the 'Hale cycle') is 22 years in length rather than 11.

The magnetic fields of spots are very strong, and may exceed 4000 G. With one group, seen in 1967, the field reached 5000 G. The preceding and following spots of a two-spot group are joined by loops of magnetic field lines which rise high into the solar atmosphere above. The highly magnetised area in, around and above a bipolar sunspot group is known as an *active region*.

The modern theory of sunspots is based upon pioneer work carried out by H. Babcock in 1961. The spots are produced by bipolar

magnetic regions (i.e. adjacent areas of opposite polarity) formed where a bunch of concentrated field lines (a 'flux tube') emerges through the photosphere to form a region of outward-directed or positive field; the flux tube then curves round in a loop, and re-enters to form a region of inward-directed or negative field. This, of course, explains why the leader and the follower are of opposite polarity.

Babcock's original model assumed that the solar magnetic lines of force run from one magnetic pole to the other below the bright surface. An initial polar magnetic field is located just below the photosphere in the convective zone. The Sun's differential rotation means that the field is 'stretched' more at the equator than at the poles. After many rotations, the field has become concentrated as toroids to either side of the equator, and spot-groups are produced. At the end of the cycle, the toroid fields have diffused poleward and formed a polar field with reversed polarity, and this explains the Hale 22-year cycle.

Each spot-group has its own characteristics, but in general the average two-spot group begins as two tiny specks at the limit of visibility. These develop into proper spots, growing and also separating in longitude at a rate of around 0.5 km s^{-1}. Within two weeks the group has reached its maximum length, with a fairly regular leader together with a less regular follower. There are also various minor spots and clusters; the axis of the main pair has rotated until it is roughly parallel with the solar equator. After the group has reached its peak, a decline sets in; the leader is usually the last survivor. Around 75% of groups fit into this pattern, but others do not conform, and single spots are also common.

ASSOCIATED PHENOMENA

Plages are bright, active regions in the Sun's atmosphere, usually seen around sunspot groups. The brightest features of this type seen in integrated light are the faculæ.

The discovery of faculæ was made by C. Scheiner, probably about 1611. Faculæ (Latin, 'torches') are clouds of incandescent gases lying above the brilliant surface; they are composed largely of hydrogen, and are best seen near the limb, where the photosphere is less bright than at the centre of the disc (in fact, the limb has only two-thirds the brilliance of the centre, because at the centre we are looking down more directly into the hotter material). Faculæ may last for over two months, although their average lifetime is about 15 days. They often appear in areas where a spot-group is about to appear, and persist after the group has disappeared.

Polar faculæ are different from those of the more central regions, and are much less easy to observe from Earth; they are most common near the minimum of the sunspot cycle, and have latitudes higher than 65° north or south, with lifetimes ranging from a few days to no more than 12 min. They may well be associated with coronal plumes.

Even in non-spot zones, the solar surface is not calm. The photosphere is covered with granules, which are bright, irregular polygonal structures; each is around 1000 km across, and may last from 3 to 10 min (8 min is about the average). They are vast convective cells of hot gases, rising and falling at average speeds

Table 2.5 *Classification of solar flares*

Area (square degrees)	Classification
Over 24.7	4
12.5–24.7	3
5.2–12.4	2
2.0–5.1	1
Less than 2	s

F = faint, N = normal, B = bright.
Thus the most important flares are classified as 4B.

of about 0.5 km s^{-1}; the gases rise at the centre of the granule and descend at the edges, so that the general situation has been likened to a boiling liquid, although the photosphere is of course entirely gaseous. They cover the whole photosphere, except at sunspots, and it has been estimated that at any one moment the whole surface contains about 4 000 000 granules. At the centre of the disc the average distance between granules is of the order of 1400 km. The granular structure is easy to observe; the first really good pictures of it were obtained from a balloon, Stratoscope II, in 1957.

Supergranulation involves large organised cells, usually polygonal, measuring around 30 000 km across; each contains several hundreds of individual granules. They last from 20 h to several days, and extend up into the chromosphere (the layer of the Sun's atmosphere immediately above the photosphere). Material wells up at the centre of the cell, spreading out to the edges before sinking again.

Spicules are needle-shaped structures rising from the photosphere, generally along the borders of the supergranules, at speeds of from 10 to 30 km s^{-1}. About half of them fade out at peak altitude, while the remainder fall back into the photosphere. Their origin is not yet completely understood.

Flares are violent, short-lived outbursts, usually occurring above active spot-groups. They emit charged particles as well as radiations ranging from very short gamma-rays up to long-wavelength radio waves; they are most energetic in the X-ray and EUV (extreme ultraviolet) regions of the electromagnetic spectrum. They produce shock waves in the corona and chromosphere, and may last for around 20 min, although some have persisted for 2 h and one, on 16 August 1989, persisted for 13 h. They are most common between 1 and 2 years after the peak of a sunspot cycle. They are seldom seen in visible light. The first flare to be seen in 'white' light was observed by R. Carrington on 1 September 1859, but generally flares have to be studied with spectroscopic equipment or the equivalent. Observed in hydrogen light, they are classified according to area. The classification is given in Table 2.5.

It seems that flares are explosive releases of energy stored in complex magnetic fields above active areas. They are powered by magnetic reconnection events, when oppositely directed magnetic fields meet up and reconnect to form new magnetic structures. As the field lines snap into their new shapes, the temperature rises to tens of millions of degrees in a few minutes, and this can result in clouds of plasma being sent outward through the solar atmosphere into space; the situation has been likened to the sudden snapping of

a tightly wound elastic band. These huge 'bubbles' of plasma, containing thousands of millions of tonnes of material, are known as Coronal Mass Ejections (CMEs). The particles emitted by the CME travel at a slower speed than flare's radiations and reach Earth a day or two later, striking the ionosphere and causing 'magnetic storms' – one of which, on 13 March 1989, caused power blackouts for over nine hours in Quebec, while on 20 January 2005 a flare caused widespread disruption of communications, and scrambled detectors on space-craft. And, in October 2007, material from a CME stripped the tail off Encke's periodical comet, which was then at about the distance of the orbit of Mercury.

Research carried out by A.G. Kosovichev and V. V. Zharkova has shown that flares produce seismic waves in the Sun's interior. They also cause shock waves in the solar chromosphere, known as Moreton waves, which propagate outwards at speeds of from 500 to 1000 km s^{-1}, and have been likened to solar tsunamis! They were first described by the American astronomer G. Moreton in 1960, though there had in fact been earlier observations of them by Japanese solar observers. However, nothing matches the brilliance and the effects of Carrington's flare of 1859. This was exceptional in every way.

A major CME is very likely to produce brilliant displays of auroræ. Cosmic rays and energetic particles sent out by CMEs are dangerous to astronauts moving above the protective screen of the Earth's atmosphere and, to a much lesser extent, passengers in very high-flying aircraft.

Flares are, in fact, amazingly powerful and a major outburst may release as much energy as 10 000 million one-megaton nuclear bombs. Some of the ejected particles are accelerated to almost half the velocity of light.

THE SOLAR CYCLE

The first suggestion of a solar cycle seems to have come from the Danish astronomer P. N. Horrebow in 1775–1776, but his work was not published until 1859, by which time the cycle had been definitely identified. In fact the 11-year cycle was discovered by H. Schwabe, a Dessau pharmacist, who began observing the Sun regularly in 1826 – mainly to see whether he could observe the transit of an intra-Mercurian planet. In 1851 his findings were popularised by W. Humboldt. A connection between solar activity and terrestrial phenomena was found by E. Sabine in 1852, and in 1870 E. Loomis, at Yale, established the link between the solar cycle and the frequency of auroræ.

The cycle is by no means perfectly regular. The mean value of its length since 1715 has been 11.04 years, but there are marked fluctuations; the longest interval between successive maxima has been 17.1 years (1788 to 1805) and the shortest has been 7.3 years (1829.9 to 1837). Since 1715, when reasonably accurate records began, the most energetic maximum has been that of 1957.9; the least energetic maximum was that of 1816. (See Table 2.6.) The numbered solar cycles are given in Table 2.7.

There are, moreover, spells when the cycle seems to be suspended, and there are few or no spots. Four of these spells have been identified with fair certainty: the Oort Minimum (1010–1050), the Wolf Minimum (1280–1340), the Spörer

Table 2.6 *Sunspot maxima and minima, 1718–2000*

Maxima	Minima
1718.2	1723.5
1727.5	1734.0
1738.7	1745.0
1750.5	1755.2
1761.5	1766.5
1769.7	1777.5
1778.4	1784.7
1805.2	1798.3
1816.4	1810.6
1829.9	1823.3
1837.2	1833.9
1848.1	1843.5
1860.1	1856.0
1870.6	1867.2
1883.9	1878.9
1894.1	1899.6
1907.0	1901.7
1917.6	1913.6
1928.4	1923.6
1937.4	1933.8
1947.5	1944.2
1957.8	1954.3
1968.9	1964.7
1979.9	1976.5
1990.8	1986.8
2000.1	1996.8
	2008.9

Minimum (1420–1530) and the Maunder Minimum (1645–1715). Of these the best authenticated is the last. Attention was drawn to it in 1894 by the British astronomer E. W. Maunder, based on earlier work by F. G. W. Spörer in Germany.

Maunder found, from examining old records, that between 1645 and 1715 there were virtually no spots at all. It is significant that this coincided with a very cold spell in Europe; during the 1680s, for example, the Thames froze every winter, and frost fairs were held on it. Auroræ too were lacking; Edmond Halley recorded that he saw his first aurora only in 1716, after forty years of watching.

Since then there has been a cool period, with low solar activity; it lasted between about 1790 and 1820, and is known as the Dalton Minimum.

Records of the earlier prolonged minima are fragmentary, but some evidence comes from the science of tree rings, dendro-chronology, founded by an astronomer, A. E. Douglass. High-energy cosmic rays which pervade the Galaxy transmute a small amount of atmospheric nitrogen to an isotope of carbon, carbon-14, which is radioactive. When trees assimilate carbon dioxide, each growth ring contains a small percentage of carbon-14, which decays exponentially with a half-life of 5730 years. At sunspot maximum, the magnetic field ejected by the Sun deflects some of the cosmic rays away from the Earth, and reduces the level of carbon-14 in the

Table 2.7 *Numbered solar cycles*

Cycle	Began	Ended	Duration, years	No. of spotless days (throughout cycle)
1	Mar 1755	June 1766	11.3	
2	June 1766	June 1775	9.0	
3	June 1775	Sept 1284	9.3	
4	Sept 1784	May 1798	13.7	
5	May 1798	Dec 1810	12.6	
6	Dec 1810	May 1823	12.4	
7	May 1823	Nov 1833	10.5	
8	Nov 1833	July 1843	9.8	
9	July 1843	Dec 1855	12.4	
10	Dec 1855	Mar 1867	11.3	~654
11	Mar 1867	Dec 1878	11.8	~406
12	Dec 1878	Mar 1890	11.3	~736
13	Mar 1890	Feb 1902	11.9	~938
14	Feb 1902	Aug 1913	11.5	~1019
15	Aug 1913	Aug 1923	10.0	534
16	Aug 1923	Sept 1933	10.1	568
17	Sept 1933	Feb 1944	10.4	269
18	Feb 1944	Apr 1954	10.2	446
19	Apr 1954	Oct 1964	10.5	227
20	Oct 1964	June 1976	11.7	272
21	June 1976	Sept.1986	10.3	273
22	Sept 1986	May 1996	9.7	309
23	May 1996	Dec 2008	12.6	>730
24	Dec 2008			

Solar minimum of ~2009 (Cycle 24). This was the deepest for many years. In 2008 there were no spots on 266 days (73%), and 2009 was even lower. This recalls 1913 (311 spotless days).

atmosphere, so that the tree rings formed at sunspot maximum have a lower amount of the carbon–14 isotope. Careful studies were carried out by F. Vercelli, who examined a tree which lived between 275 BC and AD 1914. Then, in 1976, J. Eddy compared the carbon–14 record of solar activity with records of sunspots, auroræ and climatic data, and confirmed Maunder's suggestion of a dearth of spots between 1645 and 1715. Yet strangely, although there were virtually no records of telescopic sunspots during this period, naked-eye spots were recorded in China in 1647, 1650, 1655, 1656, 1665 and 1694; whether or not these observations are reliable must be a matter for debate. There is strong evidence for a longer cycle superimposed on the 11–year one.

The law relating to the latitudes of sunspots (Spörer's law) was discovered by the German amateur Spörer in 1861. At the start of a new cycle after minimum, the first spots appear at latitudes between 30° and 45° north or south. As the cycle progresses, spots appear closer to the equator, until at maximum the average latitude of the groups is only about 15° north or south. The spots of the old cycle then die out (before reaching the equator), but even before they have completely disappeared the first spots of the new cycle are seen at the higher latitudes. This was demonstrated by the famous 'Butterfly Diagram', first drawn by Maunder in 1904.

The Wolf or Zürich sunspot number for any given day, indicating the state of the Sun at that time, was worked out by R. Wolf of Zürich in 1852. The formula is $R = k(10g + f)$, where R is the Zürich number, g is the number of groups seen, f is the total number of individual spots seen and k is a constant depending on the equipment and site of the observer (k is usually not far from unity). The Zürich number may range from zero for a clear disc up to over 200. A spot less than about 2500 km in diameter is officially classed as a pore.

Rather surprisingly, the Sun is actually brightest at spot maximum. The greater numbers of sunspots do not compensate for the greater numbers of brilliant plages.

SPECTRUM AND COMPOSITION OF THE SUN

The first intentional solar spectrum was obtained by Isaac Newton in 1666, but he never took these investigations much further, although he did of course demonstrate the complex nature of sunlight. The sunlight entered the prism by way of a hole in the screen, rather than a slit.

In 1802 W. H. Wollaston, in England, used a slit to obtain a spectrum and discovered the dark lines, but he merely took them to be the boundaries between different colours of the rainbow spectrum. The first really systematic studies of the dark lines were carried out in Germany by J. von Fraunhofer, from 1814. Fraunhofer realised that the lines were permanent; he recorded 5740 of them and mapped 324. They are still often referred to as the Fraunhofer lines.

The explanation was found by G. Kirchhoff, in 1859 (initially working with R. Bunsen). Kirchhoff found that the photosphere yields a rainbow or continuous spectrum; the overlying gases produce a line spectrum, but since these lines are seen against the rainbow background they are reversed, and appear dark instead of bright. Since their positions and intensities are not affected, each line may be tracked down to a particular element or group of elements. In 1861–1862 Kirchhoff produced the first detailed map of the solar spectrum. (His eyesight was affected, and the work was actually finished by his assistant, K. Hofmann.) In 1869 Anders Ångström, of Sweden, studied the solar spectrum by using a grating instead of a prism, and in 1889 H. Rowland produced a detailed photographic map of the solar spectrum. The most prominent Fraunhofer lines in the visible spectrum are given in Table 2.8.

By now many of the known chemical elements have been identified in the Sun. The list of elements which have now been identified is given in Table 2.9. The fact that the remaining elements have not been detected does not necessarily mean that they are completely absent; they may be present, although no doubt in very small amounts.

So far as relative mass is concerned, the most abundant element by far is hydrogen (71%). Next comes helium (27%). All the other elements combined make up only 2%. The numbers of atoms in the Sun relative to one million atoms of hydrogen are given in Table 2.10.

Table 2.8 *The most prominent Fraunhofer lines in the visible spectrum of the Sun*

Letter	Wavelength (Å)	Identification	Letter	Wavelength (Å)	Identification
A	7593	O_2			
a	7183	H_2O			
B	6867	O_2			
(These three are telluric lines – due to the Earth's intervening atmosphere.)					
C(Hα)	6563	H	b4	5167	Mg
D1	5896		F(Hβ)	4861	H
D2	5890	Na	f(Hγ)	4340	H
E	5270	Ca, Fe	G	4308	Fe, Ti
	5269	Fe	g	4227	Ca
b1	5183	Mg	h(Hδ)	4102	H
b2	5173	Mg	H	3968	Ca1[1]
b3	5169	Fe	K	3933	

Note: one Ångström (Å) is equal to one hundred-millionth part of a centimetre; it is named in honour of Anders Ångström. The diameter of a human hair is roughly 500 000 Å. To convert Ångströms into nanometres, divide all wavelengths by 10, so that, for instance, Hα becomes 656.3 nm.

Helium was identified in the Sun (by Norman Lockyer, in 1868) before being found on Earth. Lockyer named it after the Greek ηλιος, the Sun. It was detected on Earth in 1894 by Sir William Ramsay, as a gas occluded in cleveite.

For a time it was believed that the corona contained another element unknown on Earth, and it was even given a name – coronium – but the lines, described initially by Harkness and Young at the eclipse of 1869, proved to be due to elements already known. In 1940 B. Edlén, of Sweden, showed that the coronium lines were produced by highly ionised iron and calcium.

SOLAR ENERGY

Most of the radiation emitted by the Sun comes from the photo-sphere, which is no more than about 500 km deep. It is easy to see that the disc is at its brightest near the centre; there is appreciable limb darkening – because when we look at the centre of the disc we are seeing into deeper and hotter layers. It is rather curious to recall that there were once suggestions that the interior of the Sun might be cool. This was the view of Sir William Herschel, who believed that below the bright surface there was a temperature region which might well be inhabited – and he never changed his view (he died in 1822). Few of his contemporaries agreed with him, but at least his reputation ensured that the idea of a habitable Sun would be taken seriously. And as recently as 1869 William Herschel's son, Sir John, was still maintaining that a sunspot was produced when the luminous clouds rolled back, bringing the dark, solid body of the Sun itself into view[1].

Spectroscopic work eventually put paid to theories of this kind. The spectroheliograph, enabling the Sun to be photographed in the light of one element only, was invented by G. E. Hale in 1892; its visual equivalent, the spectrohelioscope, was invented in 1923, also by Hale. In 1933 B. Lyot, in France, developed the Lyot filter, which is less versatile but more convenient, and also allows the Sun to be studied in the light of one element only.

But how did the Sun produce its energy? One theory, proposed by J. Waterson and, in 1848, by J. R. Mayer, involved meteoritic infall. Mayer found that a globe of hot gas the size of the Sun would cool down in 5000 years or so if there were no other energy source, while a Sun made up of coal, and burning furiously enough to produce as much heat as the real Sun actually does, would be turned into ashes after a mere 4600 years. Mayer therefore assumed that the energy was produced by meteorites striking the Sun's surface.

Rather better was the contraction theory, proposed in 1854 by H. von Helmholtz. He calculated that if the Sun contracted by 60 m per year, the energy produced would suffice to maintain the output for 15 000 000 years. This theory was supported later by the great British physicist Lord Kelvin. However, it had to be abandoned when it was shown that the Earth itself is around 4600 million years old – and the Sun could hardly be younger than that. In 1920 Sir Arthur Eddington stated that atomic energy was necessary, adding 'Only the inertia of tradition keeps the contraction hypothesis alive – or, rather, not alive, but an unburied corpse.'

The nuclear transformation theory was worked out by H. Bethe in 1938, during a train journey from Washington to Cornell University. Hydrogen is being converted into helium, so that energy is released and mass is lost; the decrease in mass amounts to 4 000 000 tonnes per second. Bethe assumed that carbon and nitrogen were used as catalysts, but C. Critchfield, also in America, subsequently showed that in solar-type stars the proton–proton reaction is dominant.

Slight variations in output occur, and it is often claimed that it is these minor changes which have led to the ice ages which have affected the Earth now and then throughout its history, but for the moment at least the Sun is a stable, well-behaved Main Sequence star.

The core temperature is believed to be around 15 000 000 °C, and the density about 10 times as dense as solid lead. The core