1

The wind from the Sun: an introduction

> 'First accumulate a mass of Facts: and *then* construct a Theory.' *That*, I believe, is the true Scientific Method. I sat up, rubbed my eyes, and began to accumulate Facts.
>
> Lewis Carroll, *Sylvie and Bruno*

Not only does the Sun radiate the light we see – and that we do not see – but it also continually ejects into space 1 million tonnes of hydrogen per second. This wind is minute by astronomical standards; it carries a very small fraction of the solar energy output, and compared to the violent explosions pervading the universe it blows rather gently. Yet it has amazing effects on the solar surroundings. It blows a huge bubble of supersonic plasma – the heliosphere – which engulfs the planets and a host of smaller bodies, shaping their environments. It also conveys perturbations that can be seen in our daily life.

The object of this chapter is twofold. To give a concise historical account of the key ideas and observations that made our modern view of the solar wind emerge; and introduce the main properties of the Sun and of its wind, and their interpretation in terms of basic physics. The latter goal requires some tools of plasma physics, and will be developed in the rest of the book.

1.1 A brief history of ideas

The idea that planets are not moving in a vacuum is very old. In some sense our modern view of a solar wind filling interplanetary space has replaced the Aristotelian quintessence, the impalpable pneuma of the Stoic philosophers, and the swirling 'sky' introduced two thousand years later by Descartes. In some sense only, as there is a major difference: the solar wind is made of normal matter whose behaviour is – to some extent – understood, even though this matter is in a special state, a *plasma*, having unusual properties as we will see in Chapter 2.

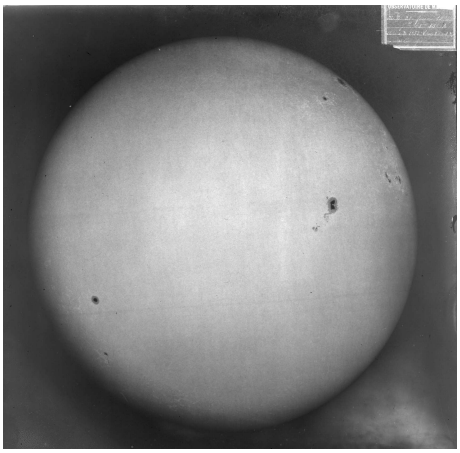

Figure 1.1 An early photograph of the Sun showing sunspots, made by Jules Janssen at the Meudon Observatory on 22 June 1894. (Observatoire de Paris, Meudon heritage collections.)

Ironically enough, the solar wind contains vortices – as did Descartes' sky and also the luminiferous ether imagined later by Maxwell – and even though those vortices have nothing in common with their ethereal ancestors, they are barely better understood. And not only does the solar wind transmit sound and light as did the ancient ether, but it also carries a host of waves that Maxwell could not have dreamt of.

Even though the idea is an ancient one, most of the solar wind story took place over little more than a century. At the end of the nineteenth century, only a couple of far-seeing scientists had imagined that a solar wind might exist. At the beginning of the twenty-first century, hordes of space probes have explored the solar wind and it is honoured with a secure place in astronomy textbooks.

Since there are eminent accounts of how this concept emerged and developed (see for example [9], [25]), it is not my intention here to trace a detailed history. I shall only give a few hints[1] as to how the ideas evolved to fit reasonably well into the logical structure of modern physics and astronomy.

1.1.1 Intermittent particle beams?

When did the story begin? Perhaps around the middle of the nineteenth century, when the British amateur astronomer Richard Carrington, who was drawing sunspots (see Fig. 1.1) from a projected image of the Sun,[2] suddenly saw two patches of peculiarly intense light appear and fade within 5 minutes in the largest sunspot group visible [7], [5]. Carrington had witnessed what we now call a

[1] Partly taken from Meyer-Vernet, N. 2005, *Bul. Inst. d'Astron. et de Géophys. Georges Lemaître* (Université Catholique de Louvain, Louvain-La-Neuve, Belgium).

[2] By far the safest way for amateurs to observe the Sun.

Figure 1.2 Auroral display seen from the ground in Lapland. (Photo courtesy of C. Molinier.) For hints on auroral physics see Section 7.3.

solar flare: a giant explosive energy release on the Sun – and a very strong one. Some time later, the magnetic field at Earth was strongly perturbed (Fig. 7.22), large disturbances appeared in telegraph systems, and intense auroras spread over much of the world (see Fig. 1.2). The connection between geomagnetic perturbations and auroras was already known, and Carrington suggested that both phenomena might be due to the special event he had seen on the Sun; but he was careful to point out that his single observation did not imply a cause-and-effect relationship. Carrington was correct; we will see later that solar flares are sometimes, although not always, followed by geomagnetic disturbances and auroras; this happens when the Sun releases a massive cloud of gas that reaches the Earth's environment and perturbs it.

Carrington was not the first to suspect the Sun of producing auroras and magnetic effects on Earth, as a correlation between the number of sunspots and geomagnetic disturbances had already been noted. The cause, however, was unclear. And the aurora was not better understood, even though its electric and magnetic nature had been identified a century before – a major improvement over the ingenious scheme based on the firing of dry gases proposed by Aristotle. To summarise the state of auroral physics around the mid 1880s (see [10]): 'The scientific theories ... are more abstruse than the popular ones, but equally fail ...', as the Norwegian scientist Sophus Tromholt put it with splendid irreverence.

All that pointed to a connection between the Sun and terrestrial magnetic disturbances, and the idea was taken seriously by some physicists near the end of the nineteenth century. Assuming 'that the Sun is powerfully electrified, and repels similarly electrified molecules with a force of some moderate number of times the gravitation of the molecules to the Sun', George Fitzgerald suggested that [12]: 'matter starting from the Sun with the explosive velocities we know possible there, and subjected to an acceleration of several times solar gravitation, could reach the Earth in a couple of days'. In other words, the Earth was bombarded by intermittent beams of charged particles coming from the Sun

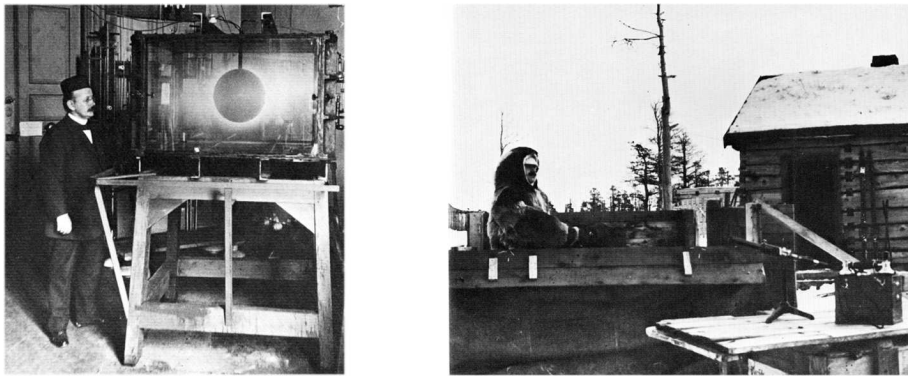

Figure 1.3 Birkeland on two of his fronts [4]. Left: with his *terella* apparatus – a magnetised sphere subjected to a beam of electrons in a vacuum chamber. Right: with some of his instruments for aurora detection.

and accelerated by an electrostatic field, just as an electrode in a giant vacuum tube.[3]

1.1.2 Permanent solar corpuscular emission?

In this context, an essential step was taken by the Norwegian physicist Kristian Birkeland, who worked in the closing years of the nineteenth century and the opening years of the twentieth (Fig. 1.3). These were fabulous times for a physicist: X-rays and radioactive decay were just being discovered, J. J. Thomson was unveiling the electron, Hendrik Lorentz was developing the electron theory and building steps on the route which led Einstein to change our vision of the Universe and Max Planck was explaining the spectrum of radiation, among other major accomplishments. Applied science was rising, too: the first aeroplane was close to being born, and Guglielmo Marconi made the first long-distance radio transmission, an engineering performance which led to the discovery of the Earth's ionosphere.

Birkeland worked on three fronts: theory, laboratory experiments with a model Earth and observation [11]. Not only did he develop the ideas put forward by Fitzgerald and others, but in order to test them he organised several polar expeditions and made the largest geomagnetic survey up to that time [4]. He also put forward a number of ingenious ideas that stand up well today, and above all, he submitted a crucial point: since auroral and geomagnetic activity was produced by solar particles and was virtually permanent, the inescapable conclusion was that the Earth environment was bombarded in permanence by 'rays of electric corpuscles emitted by the Sun'.

[3] This was written 5 years before J. J. Thomson's 1897 paper on 'cathode rays' (*Phil. Mag.* **44** 293), and showed remarkable insight. The solar wind is made of charged particles – protons and electrons – and we will see later in this book that indeed the heliospheric electric field pushes the protons outwards with a force of a few times the solar gravitational attraction.

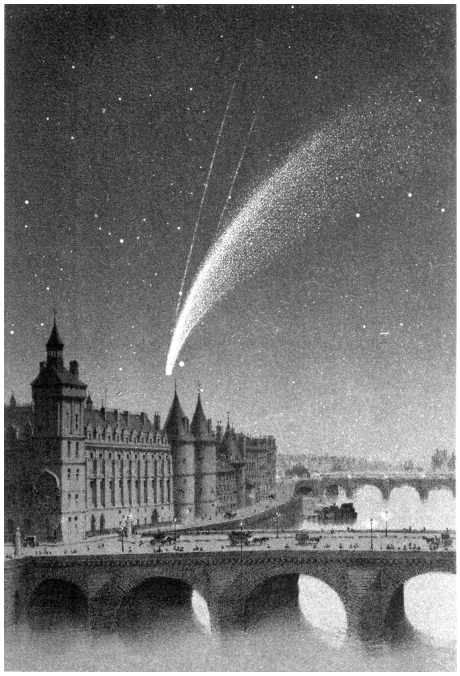

Figure 1.4 An old drawing of Donati's comet, shown over Paris on 5 October 1858. (From A. Guillemin, 1877, *Le Ciel*, Paris, Hachette.) A modern view of cometary physics may be found in Section 7.5.

Put in modern terms, Birkeland was suggesting that the Sun emits a continuous flux of charged particles filling up interplanetary space: nearly our modern solar wind. This was a great change of perspective from the picture of the Sun emitting sporadic beams separated by a vacuum. However, many of these ideas were far ahead of the time, some were incorrect, and above all, the revered Lord Kelvin submitted impressive arguments showing that the Sun could not produce geomagnetic disturbances.[4] As a result, Birkeland's work was largely ignored by the scientific community.

These ideas lay in obscurity for many years. And when the *solar corpuscular radiation* (as it was called) resurfaced – albeit on independent grounds – to explain geomagnetic activity, it was once again in the form of occasional beams emitted by the Sun by some exotic process in a (slightly dusty) vacuum.

This remained the leading view until the middle of the twentieth century, when the concept of a continuous solar emission was to re-emerge through an entirely separate line of work. Comets have two classes of tails, one class nearly straight, made of gas, the other curving away, made of dust (Fig. 1.4). The

[4]On 30 November 1892, he concludes in his Presidential Address to the Royal Society: 'It seems as if we may also be forced to conclude that the supposed connection between magnetic storms and sun-spots is unreal, and that the seeming agreement between the periods has been a mere coincidence' (Kelvin, Thompson, W. 1892, *Proc. Roy. Soc. London A* **52** 300).

curved shape of dust tails is produced by solar radiation pressure and gravity acting on the dust grains. But the gaseous tails raised an intriguing problem: they were observed to always point straight away from the Sun (with a slight aberration angle)[5] and to exhibit irregularities that appeared to be accelerated away from the Sun. What caused these properties? The current explanation in terms of solar radiation pressure acting on the cometary gas failed by several orders of magnitude.

This problem was brilliantly solved by the German physicist Ludwig Biermann in the early 1950s: solar photons could not do the job, but perhaps solar corpuscular radiation would? Biermann thus developed a model for the interaction of cometary particles with those coming from the Sun, that neatly explained the comets' gaseous tails *if* they were subjected to a permanent flux of charged particles coming from the Sun (see [3] and references therein).

Although Biermann's original arguments regarding the interaction process between particles are now known to be incorrect, this was again a crucial change of perspective from the current view. Since comets' orbits pass at all heliolatitudes, the inescapable conclusion was that the Sun was emitting particles in *all* directions at *all* times. Half a century after Birkeland's work, the concept of a continuous solar corpuscular emission was resurfacing, with stronger observational and theoretical support.

But at the same time a different conclusion was reached by the English physicist Sydney Chapman through a completely different path. The outer atmosphere of the Sun – called the *corona* (see Fig. 1.12) – was known to be very hot. Chapman, who had pioneered the calculation of the kinetic properties of gases, found that this hot ionised atmosphere conducted heat so well that it should remain hot out to very large distances. As a result, particles have such large thermal speeds even far from the Sun that they can go very far against its gravitational attraction; this makes the density decline very slowly, so slowly that the solar atmosphere should extend well beyond the Earth's orbit [6]. In other words, the Earth was to be immersed in the *static* atmosphere of the Sun.

1.1.3 The modern solar wind

How could the ubiquitous solar corpuscular flux found by Biermann coexist with this static solar atmosphere? Both are plasmas and, as we will see later, the coexistence of plasmas having such different bulk velocities has very nasty consequences. The great achievement made by Eugene Parker in 1958 was to realise that [24]: 'however unlikely it seemed, the only possibility was that Biermann and Chapman were talking about the same thing'. So Biermann's continuous flux of solar particles was just Chapman's extended solar atmosphere expanding away in space as a supersonic flow. This comes about because this atmosphere is so hot, even far away from the Sun, that neither the solar gravitational attraction nor the pressure of the tenuous interstellar medium can confine it. At

[5] With hindsight, the aberration angle of comet plasma tails, determined by the relative speeds of the radially moving solar wind plasma and the comet, yielded a correct solar wind speed of about 400 $\mathrm{km\,s^{-1}}$. For details, see Brandt, J. C. 1970, cited in Chapter 5.

Figure 1.5 Mariner 2, the spacecraft that served to identify the solar wind, and the first one to have reached a planet other than the Earth. (Image by National Aeronautics and Space Administration (NASA).)

last the modern solar wind concept was born. Parker's theory was not only an eminently elegant solution which brought together existing evidences, but it made numerous testable predictions: in particular the wind was to flow at several hundreds of kilometres per second radially away from the Sun. Irony of ironies, these ideas were so novel that the paper presenting them [23] encountered difficulty in being published in the eminent *Astrophysical Journal*, on the grounds that the author was not familiar with the subject [25]. The referees of the *Astrophysical Journal* were not the only scientists to be displeased by Parker's theory, and a hot debate followed as to whether or not the Sun was capable of emitting a supersonic wind.

Observation was needed to settle the debate. But measuring the solar wind was a heroic challenge in those years when space technology was just springing up. It took four Russian missions and seven American ones – most of which failed due to problems of launching – to get an unambiguous result. The most successful of the Russian spacecraft, Lunik II, launched in 1959 (only 2 years after the first Russian Sputnik went up and the United States launched their first satellite in reply) detected a flux of positive ions; however, this observation was not entirely conclusive because the direction of the particles' velocity was unknown. The ultimate proof came in 1962 from the American spacecraft Mariner 2 [22], which was en route for Venus after having miraculously survived an impressive series of failures (Fig. 1.5). As Marcia Neugebauer superbly puts it [20]: 'We had data! Lots of it! There was no longer any uncertainty about the existence and general properties of the solar wind.'

So ended the first age of the solar wind story. With these results Parker's ideas rose to prominence, and within a few years, in spite of some dissenting voices, the solar wind concept acquired the respected state of a physical reality. Even now, decades later, Parker's ideas serve as a reference for a large part of what is understood on the subject. However, the very elegance of this theory

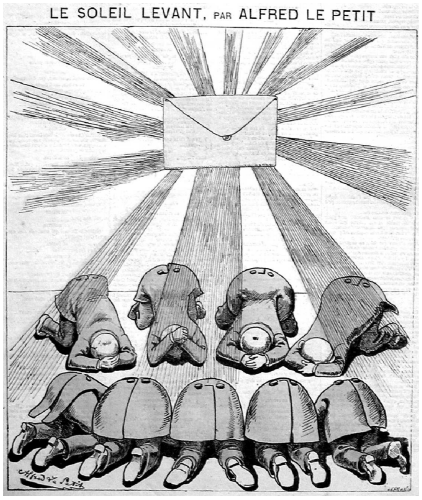

Figure 1.6 Cartoon published in the French journal *Le Grelot* (March 1873).

masked a number of fundamental difficulties that we will examine later, and of which an historical account may be found in [18]. As we will see in this book, the story has many other chapters and there is still a long way ahead.

1.2 Looking at the Sun

My Sun, the golden garden of your hair
Has begun to flame
And the fire has spread over our corn field

Already the green ears are parched
Pressed by the presence of your breath
And the last drop of their sweet is wrung for them

Strike us with the rain of your arrows,
Open to us the door of your eyes,
Oh Sun, source of beneficent light.
(Quecha poem to the Sun[6]*)*

To us the Sun is no longer a god, and we do not have to feed it with hearts and blood to keep it moving across the sky, as did the Aztecs. Modern scientists (Fig. 1.6) worship it by building beautifully instrumented observatories scattered over the Earth's surface, far underground and in space, to analyse its radiation.

[6]Ferguson, D. 2000, *Tales of the Plumed Serpent*, London, Collins & Brown.

Table 1.1 *Basic solar properties*

Mean distance to the Earth	$d_\oplus = 1.5 \times 10^{11}$ m
Mass	$M_\odot = 2. \times 10^{30}$ kg
Radius	$R_\odot = 7. \times 10^8$ m
Luminosity	$L_\odot = 3.84 \times 10^{26}$ W

The last quarter of the twentieth century saw the vertiginous growth of the scope and quality of solar data. A number of new techniques have emerged, making a revolution in the achievements of ground-based telescopes, while several beautifully instrumented spacecraft have been launched, enabling one to study wavelengths inaccessible from the ground. Among others, the satellites Yohkoh (launched in 1991), SoHO (launched in 1995), TRACE (launched in 1998) and RHESSI (launched in 2002) are currently in operation and devoted to solar observations.

We outline below some characteristics of the Sun and make a short survey of what its radiation tells us. We shall try to complete and unify this impressionistic picture later, equipped with the tools of plasma physics. A very accessible survey of solar physics is [13]; general accounts aiming at physical processes are given in [17], [14], [33].

1.2.1 Basic solar properties

Some basic properties of the Sun are listed in Table 1.1. Although these are rounded figures, they are exact to better than 1%, and may be considered as nearly constant by human standards; indeed, as we shall see later, many years will pass before changes in solar properties may oblige human beings to transfer the Earth to a more convenient star (see [2] and references therein).

The solar distance is called the astronomical unit (AU); it is used as a basic unit in the Solar System and beyond. So is the solar mass, which is negligibly altered by the solar wind ejection. Indeed, the solar wind pours out in space roughly 10^9 kg s^{-1}, which amounts to only $10^{-4} M_\odot$ over the Sun's age of a few 10^9 years. Note that the wind is not the only source of solar mass loss; the mass–energy equivalence tells us that the luminosity $L_\odot$ – the energy lost by the Sun per second via electromagnetic waves – yields a mass loss of $L_\odot/c^2 = 4.3 \times 10^9$ kg s^{-1}; this amounts to about four times the mass carried away by the solar wind, and thus barely alters the Sun's mass either; we will return later to the solar energy source.

The Sun's radius is that of the visible disc, which is almost perfectly round and whose diameter is a little more than half a degree as seen from Earth; it is sharply defined because virtually all the light we receive from the Sun originates in a thin layer a few hundred kilometres thick: the *photosphere*, where – going outwards – matter changes rapidly from completely opaque to almost completely transparent, letting radiation escape freely into space.

At the mean distance $d_\oplus$ of the Earth (but outside its atmosphere), the flux received from the Sun in the form of electromagnetic radiation by a surface perpendicular to the rays of sunlight is

$$\begin{aligned} S &= L_\odot/(4\pi d_\oplus^2) \\ &= 1.37 \times 10^3 \text{ W m}^{-2} \end{aligned} \tag{1.1}$$

which is called the *solar constant*. One square metre of the Earth's surface receives, however, much less: around 200 W in average, partly because only half the surface is sunlit and the radiation does not arrive at right angles, and partly because some of the incident energy is absorbed in the atmosphere. This energy sustains virtually all life on Earth; it may also harm it, as the Quecha poem reminds us. Despite its name, the solar constant is more variable than the other basic solar properties; it may vary by up to a few thousandths over several days and even more over long periods; it shows in fact variability at virtually all timescales, thereby raising a strong interest in climatology circles [15].

The solar luminosity enables one to derive one more basic property. If the Sun were radiating as a blackbody, that is if the Sun's disc emitted thermal radiation of temperature T_{eff}, the luminosity would be given by Stefan–Boltzmann law as

$$L_\odot = \sigma_S T_{\mathit{eff}}^4 \times 4\pi R_\odot^2 \tag{1.2}$$

where $\sigma_S = 5.67 \times 10^{-8}$ is the Stefan–Boltzmann constant in SI units. Putting the figures given in Table 1.1 into (1.2), we deduce the *effective temperature* of the Sun

$$T_{\mathit{eff}} = 5800 \text{ K}. \tag{1.3}$$

This is about 20 times hotter than the temperature at the Earth's surface, and it is this temperature difference that sustains life there. The effective temperature T_{eff} would be the actual temperature of the emitting outward layer of the Sun – the photosphere – if the radiation were thermal. We will see later that the actual photospheric temperature is close to this value.

1.2.2 The solar spectrum

The solar radiation has been studied at virtually all wavelengths, from gamma rays on the short wavelength side of the spectrum, to radio waves on the long wavelength side. How close to thermal is it?

Spectral distribution

Figure 1.7 shows the spectral distribution S_λ of the energy received from the Sun at Earth (outside the Earth's atmosphere) per unit wavelength per unit sunlit surface area perpendicular to the Sun's direction per unit time, at wavelengths ranging from 10^{-13} m to a few metres; in this range the intensity spans over

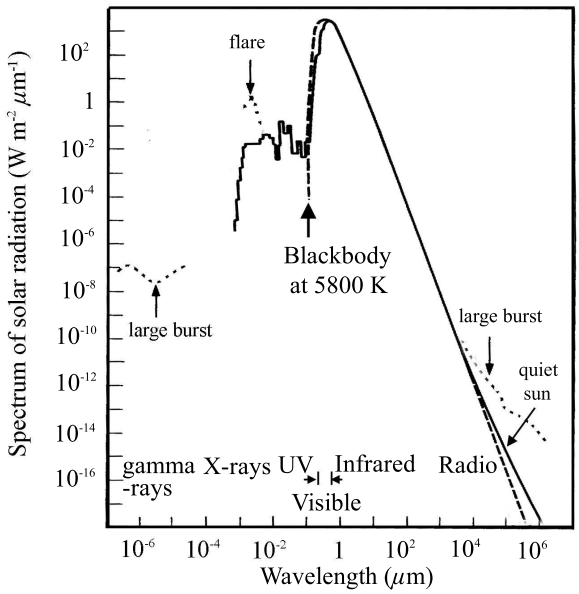

Figure 1.7 The measured spectrum of solar radiation (continuous line, with the contribution of bursts dotted), compared with the blackbody spectrum (1.4) (dashed).

23 orders of magnitude. The visible and near-infrared spectrum, which makes only a minute part of the wavelength range, takes the lion's share of the energy. The solar constant S introduced in (1.1) is the integral of S_λ over wavelengths.

On this measured spectrum we have superimposed (dashed line) the spectrum of thermal radiation:

$$S_{T\lambda} = \pi B_\lambda R_\odot^2 / d_\oplus^2 \tag{1.4}$$

where

$$B_\lambda = \frac{2hc^2}{\lambda^5} \frac{1}{e^{hc/\lambda k_B T} - 1} \tag{1.5}$$

is the Planck emissivity at wavelength λ (per unit wavelength per unit surface area perpendicular to the Sun's direction per unit solid angle per unit time), calculated with $T = T_{\mathit{eff}}$ given by (1.3).[7]

One sees that the continuous and dashed lines are rather close to each other in the near and middle ultraviolet (UV), and nearly indistinguishable in the visible and in the infrared. This means that the solar spectrum is close to thermal in that range which is the one contributing most to the total flux. This gives some sense to the value of T_{eff} we derived from Stefan–Boltzmann's law.

[7]Beware that the spectra in Fig. 1.7 are plotted in W m^{-2} μm^{-1}, i.e. not in SI units.

In contrast, one sees in Fig. 1.7 that the UV radiation below about 0.1 μm greatly exceeds that of a 5800 K blackbody; instead of falling almost exponentially the level remains of the order of 3×10^{-2} W m^{-2}μm^{-1}. This UV emission below 0.1 μm is of great importance. Integrated over wavelengths it amounts only to about 3×10^{-3} W m^{-2}, less than 10^{-5}th of the total solar flux. Yet, this radiation has a crucial property: at $\lambda = 10^{-7}$ m the frequency is $f = c/\lambda = 3 \times 10^{15}$ Hz, so that the characteristic energy hf of a photon is about 12 eV. We will see later that this is close to the binding energy of the hydrogen atom, which sets the order of magnitude of the energy needed to strip atoms of their outer electron. It is this part of the solar spectrum that is responsible for most ionised environments in the inner solar system, from planetary ionospheres (Section 7.1) to comets (Section 7.5).

At these short wavelengths, the spectral level is far from constant and becomes increasingly variable at shorter wavelengths. The largest variations can reach several orders of magnitude and occur during flares – those giant eruptions that Carrington described for the first time, and to which we will return in Section 4.5. We have sketched this variable component as a dotted line.

One sees in Fig. 1.7 that the long wavelength part of the spectrum behaves very differently. In the radio range, one can distinguish two superimposed components. The first one is a smooth spectrum – the so-called *quiet sun* radiation, which continues smoothly from the infrared with an amplitude deviating more and more from the 5800 K thermal spectrum as the wavelength increases; near $\lambda \simeq 1$ m, it would correspond to an effective temperature about 200 times hotter, that is of the order of 10^6 K. The second component is made of intermittent bursts whose amplitude may exceed the quiet level by several orders of magnitude, and whose spectrum does not even have a small resemblance to Planck's law; it is sketched as a dotted line.

Spectral lines

On the continuous spectrum shown in Fig. 1.7 are superimposed a huge number of spectral lines, which do not appear clearly on the figure because its resolution is too low; spectral lines even dominate the ultraviolet range. Millions of lines are observed in the solar spectrum, of which only a very small fraction have been analysed with the sophisticated tools of atomic and molecular physics. Since no two chemical elements have the same spectrum, spectral lines act as bar codes which enable spectroscopists to determine the chemical composition of the outer layers of the Sun, in addition to a number of physical properties. Practically all known chemical elements have been detected on the Sun. Hydrogen is by far the most abundant, about 92% in number density; the rest is mostly helium, the other elements making up only about 0.1% of the total. On the whole, and apart from a few significant exceptions, this is rather similar to the rest of the universe. We will return to this point in Section 6.5.

These lines still reveal more information. As particles are moving, the lines are Doppler-shifted, which enables one to derive the particle velocities. The random speeds produce a random frequency shift which broadens the lines,

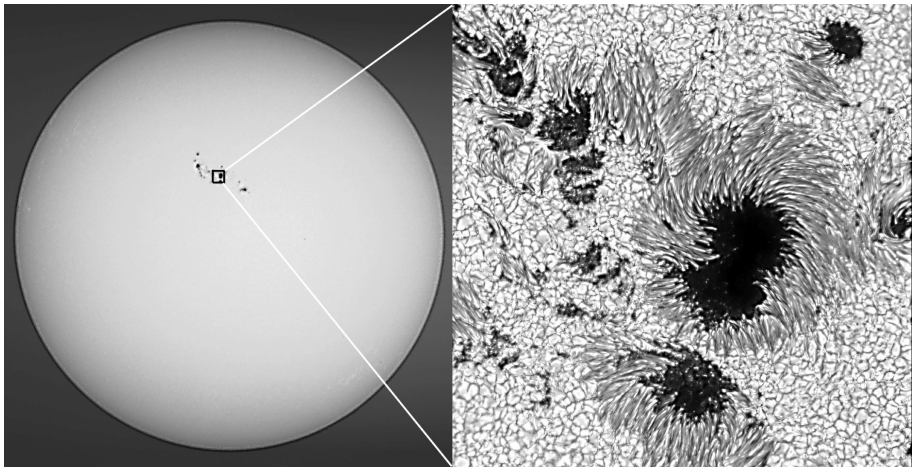

Figure 1.8 A modern photograph of the solar disc in white light, taken with the Swedish 1-metre Solar Telescope on La Palma, Canary Islands (left panel) on 15 July 2002, and a detailed view of a part of the group of sunspots visible near the centre of the disc (right panel). (Photographs courtesy of Royal Swedish Academy of Sciences.)

so that the line shapes reveal both the particle thermal speeds (that is the temperature) and the turbulent motions of the medium. A further quantity can be derived: the bulk velocity of the medium, which produces a net shift of the lines.

The spectral lines still reveal more. As the medium is magnetised, the Zeeman effect modifies the lines, which may enable one to derive the strength and direction of the magnetic field. As we shall see below, the magnetic field plays a major role in solar properties; in fact it is a key to understanding most phenomena occurring in the solar atmosphere.

1.2.3 The solar disc

Figure 1.8 is a modern image of the solar disc in white light, showing a group of sunspots (left panel) and an enlarged view of this group (right panel). Apart from sunspots, the figure shows two main features:

- the whole disc (left panel) appears darker near the edge than near the centre,
- at small scales (right panel), the space outside sunspots is covered with a granular pattern.

Both features reveal important properties of the solar surface. Let us consider in turn the disc as a whole, and its small-scale structure.

The disc as a whole

Why does the disc appear darker near the edge? To understand this *limb darkening*, we must consider where the observed light comes from and how it travels. The observed light comes from some depth within the Sun, a depth that is limited by absorption of light along its trajectory. When we look at the edge of the disc, light is travelling towards us at a small angle to the solar surface, so that it encounters more absorbing solar material than when it comes from the centre of the disc, in which case it is travelling normal to the surface. Hence we can see deeper into the Sun when we look at the centre than when we look at the edge. Therefore, the greater brightness near the centre means that deeper regions are hotter. This is not surprising; as radiation escapes into space, energy is lost so that the temperature decreases outwards.

From the variation in brightness from centre to limb, one can calculate that the temperature decreases outwards by more than 2000 K over a distance of a few hundred kilometres, where light absorption by the solar material varies from almost complete to negligible.

Why does light absorption by the solar material decrease so sharply outwards? This is because the density drops sharply; furthermore we have seen that the temperature also falls off, so that the ionisation changes. But why does the density drop so sharply? The answer lies in two simple properties of the solar material. First, it behaves as a perfect gas. We have seen that it is essentially made of hydrogen atoms, whose mass is about the proton's mass m_p; hence the pressure P and mass density ρ are related by the perfect gas law

$$P = \rho k_B T / m_p. \tag{1.6}$$

Second, the gas is close to hydrostatic equilibrium in the solar gravitational field. Thus at distance r from the Sun's centre, the outward pressure force per unit volume dP/dr balances the gravitational attraction per unit volume ρg; writing $dP/dr = \rho g$ and substituting (1.6), we deduce

$$dP/P = (m_p g / k_B T)\, dr.$$

Integrating over r in a small range of distance above the solar surface over which g and T do not vary much, we find

$$P \propto \exp(-r/H) \tag{1.7}$$

$$H = k_B T / m_p g. \tag{1.8}$$

Since the Sun's gravitational acceleration at distance $r \simeq R_\odot$ is

$$g = M_\odot G / R_\odot^2, \tag{1.9}$$

equations (1.7) and (1.8) show that the pressure and mass density decrease outwards with the scale height

$$H = \frac{k_B T R_\odot^2}{m_p M_\odot G}. \tag{1.10}$$

With the parameters in Table 1.1 and $T = T_{eff}$ (given in (1.3)), this yields $H = 1.8 \times 10^5$ m.[8] Hence over a distance of a few hundred kilometres the density falls by a large factor; this corresponds to the width of the photosphere. We shall study the photosphere in more detail in Section 3.3.

Small-scale structure

Let us now examine the solar disc at small scales (right panel of Fig. 1.8, which includes a large sunspot region). Outside sunspots, it is covered with bright irregular 'granules' of different sizes, more or less polygonal in shape, separated by darker lines. These granules have a maximum size of about 2000 km, and detailed observation shows that they seem to have a life of their own: they appear, fragment or explode, living typically for a few minutes. What is their origin? They are believed to be the uppermost signature of convective motions arising below the photosphere. As we said, the Sun is hotter in the interior, and, somewhat like water in a kettle heated from below, the solar material develops convective patterns. Spectroscopy shows that the bright central regions of granules are moving upwards, while the darker edges move downwards, with speeds of the order of 1 km s^{-1}. Granules can be thought of as fountains where hot solar material is rising near the centre, and then flows horizontally towards the edges; in doing so the material cools by radiating to outer space so that upon arriving at the edges it is cooler, and thus appears darker than the central part of the granules. Having cooled, the material is denser and sinks back into the Sun. The upward flow in the central region of the granules compensates for this disappearing cold material.

Granulation is not the only structure presumably convective in origin observed at the photosphere. It is the most conspicuous of what may be a continuum of convective flow scales. In particular, a large-scale velocity pattern – the so-called *supergranulation* – is observed, albeit with little if any intensity contrast. These large cellular structures, of scale $\sim$ (2–3) $\times 10^4$ km and living for about a day, are identified as horizontal flows moving at about 0.5 km s^{-1} towards the periphery of the cells, with much slower vertical upflows at the centre of the structures and downflows at their boundaries.

Convection plays a major role in the Sun's behaviour: on a large scale, it transports heat to the photosphere from the hotter layers below; it also drives small-scale motions of the solar material. We will return to this phenomenon in Chapter 3, and shall see that it plays an important role in the generation and structure of the magnetic field.

1.2.4 Sunspots, magnetic fields and the solar cycle

Sunspots

Let us now examine sunspots – the kind of structure Carrington was observing when he witnessed a solar flare. Sunspots can be observed with simple tools;

[8] We will see in Section 3.2 that the scale height is slightly smaller because the presence of a small quantity of atoms heavier than hydrogen increases the mean mass per particle.

in some cases they are big enough to be seen by the unaided eye.[9] Sunspots have a remarkable property: their number varies in a nearly cyclic way, with a period of about 11 years. They generally occur in groups on the solar disc. Figure 1.8 (right panel) shows a detailed view of such a group,[10] where individual spots appear as a dark *umbra* about 10^4 km across surrounded by a filamentary penumbra.

Some sunspots may last for weeks; this has enabled early observers to follow their motion and to find that the Sun appears to rotate about its axis once in about 27 days (as seen from Earth). Detailed observation shows that the Sun does not rotate as a solid body: its outer layers rotate more slowly in polar regions than at the equator. Correcting for the Earth's motion which changes the Sun's apparent rotation speed when it is observed from Earth, one finds that in an inertial frame, the Sun's outer layers rotate once in 25 days near the equator and only about once in 30 days near 60° solar latitude.

Sunspots appear dark because they are colder than the normal photosphere; since radiation varies as the fourth power of temperature (1.2), they radiate much less than their surroundings. What makes them cold? An answer was suggested in 1941 by Biermann. We have said that the normal photosphere is heated by convective motions that mix the surface and the hotter layers below. However, sunspots have a special property: they are permeated with an intense magnetic field, and as we will see later, this magnetic field holds matter so tightly that it can inhibit convective motions, thereby preventing heat from reaching them.

To summarise, sunspots exhibit three basic properties:

- they are colder than their surroundings by about 2000 K,
- their magnetic field is strong: about 0.3 T (nearly vertical at the centre); they often appear in pairs with opposite magnetic field polarities,
- their number and location on the solar disc follow regular laws, with a cycle of 11 years.

The opposite magnetic field directions observed in nearby sunspots are illustrated in Fig. 1.9; this structure is not surprising since from Maxwell equations the total magnetic flux through the solar surface must remain zero. We will return to sunspot physics in Section 3.3.

The sunspot cycle is illustrated in Fig. 1.10, which shows the number of sunspots visible on the Sun as a function of time.[11] Sunspots are not the only solar feature observed to follow such a cycle. They are the most conspicuous of a variety of structures having virtually all scales, which are governed by the magnetic field and often follow the same 11-year cycle.

[9] A warning for absent-minded readers: the Sun is by far the most dangerous astronomical object to observe; it should never be looked at, either directly or through an optical device, without an appropriate filter.

[10] Details on the structure of these sunspots may be found in Scharmer, G. B. *et al.* 2002, *Nature* **420** 151.

[11] Computed with the international sunspot number convention, till daily observations of sunspots were started at the Zurich observatory.

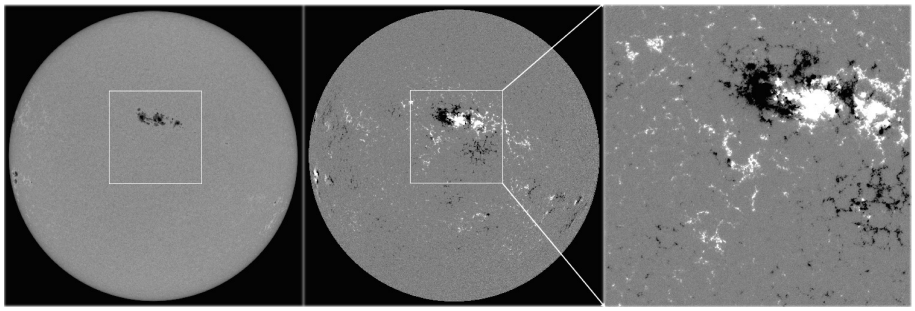

Figure 1.9 Magnetogram of the solar disc (middle panel) taken from the SoHO spacecraft on 16 July 2002; white represents upward magnetic field along the line of sight, black downward. The left panel is a white-light image of the solar surface taken a few minutes before, showing a few sunspots; the right panel is an enlarged magnetogram. (Images from SoHO/MDI, European Space Agency (ESA) and NASA, Stanford-Lockheed Institute for Space Research.)

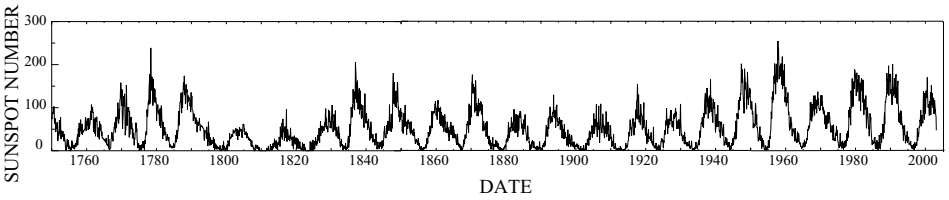

Figure 1.10 Monthly mean sunspot numbers from 1750 to 2003.

Magnetic field on the Sun

What does the photospheric magnetic field look like far away from sunspots? On a large scale and very roughly, it can be thought of as produced by a giant bar magnet whose polarity reverses every 11 years (so that the true period is 22 years). On the whole, this yields opposite average fields of the order of perhaps (1–5) $\times 10^{-4}$ T near the solar poles at the minimum of the cycle, when the solar magnetic dipole is roughly aligned with the rotation axis and the number of sunspots is minimum. This dipolar field gradually weakens and reverses its direction near the maximum of the cycle, when the number of sunspots is maximum; at that time the large-scale field is no longer dipole-like and has a complex multipolar structure.

However, observation at smaller scales reveals a more complicated pattern (Figs. 1.9 and 1.11) and shows a remarkable phenomenon: the magnetic field at the photosphere has an intermittent structure, showing concentrations as small as perhaps 100 km across,[12] where the field can be as high as about 0.15 T and is changing rapidly; since these thin structures cover only a very small fraction of

[12] Or less, since the best angular resolution achieved at present is not less than 0.1 arc–sec, and detecting smaller features requires extremely subtle techniques.

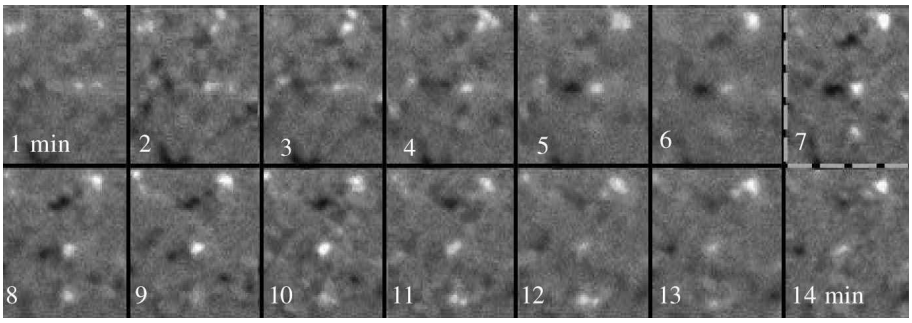

Figure 1.11 A time sequence of magnetograms of a small area on the solar surface, taken with the Swedish telescope at La Palma; the resolution is roughly 200 km on the solar surface. White represents upward vertical magnetic field, black downward. The displayed area on the Sun is approximately 7000×9000 km^2 and the magnetograms are taken at intervals of 1 min. (From [29].)

the solar surface, the average field is much smaller. Furthermore, one observes nearly everywhere a kind of magnetic carpet of fields having opposite vertical directions (Fig. 1.11), so that in 'quiet' regions the mean magnetic field may be much smaller than the average absolute value [29]. An important observational clue is that the small field concentrations generally evolve at the same timescale as convective flows, tending to accumulate at the edge of the convective cells.

What causes the solar magnetic field and its cycle? What determines the scale of its structures and their relation to one another? A hint may be obtained by comparing the energies of the magnetic field and of the particle thermal motions. The thermal energy per particle at temperature T is $3k_BT/2$; with a particle number density of about $n \simeq 10^{23}$ particles/m^3 at $T \simeq 5000$ K in the photosphere, the density of thermal energy is $3nk_BT/2 \sim 10^4$ J m^{-3}. In small magnetic structures having a strong magnetic field $B \sim 0.15$ T, the magnetic energy density $B^2/2\mu_0 \sim 10^4$ J m^{-3}, so that both types of energy are in equilibrium. In sunspots, whose magnetic field is much stronger (and temperature weaker), the magnetic field dominates and tends to hold matter, whereas in normal regions of smaller magnetic field, matter should drive the field. We shall put this rough argument on a sounder footing in Section 3.3 and will try to unveil part of the enigma of solar magnetic fields, using the tools of plasma physics. However, more than three centuries after the time of Galileo, when sunspots were first observed with a telescope, their structure is still not fully understood.

1.2.5 Around the Sun: chromosphere and corona

Chromosphere

Were the density to continue to decrease outwards at the rate estimated above, the solar atmosphere would not extend very far away ... and there would be no

solar wind. But a striking phenomenon occurs above the photosphere: instead of continuing to decrease outwards, the temperature starts to increase; over an altitude of 2000 km or so the temperature rises to 10^4 K or so. This relatively high temperature enables atoms to be excited to energy levels from which they emit spectral lines which should not be observed otherwise. In particular the red H_α line – the less energetic of the Balmer series of hydrogen – contributes to the bright purplish-red crescent seen during total solar eclipses that is at the origin of the name of this region: the *chromosphere*.

Corona

At the top of the chromosphere a still more striking phenomenon occurs: the temperature T jumps by two orders of magnitude; the density ρ decreases by roughly the same factor to prevent too large changes in the pressure, which is proportional to ρT; above this *transition region* the temperature profile flattens and remains of the order of 10^6 K over a few solar radii. This hot atmosphere is called the *corona* – the Sun's crown; not only is it very extended because the scale height increases with temperature, but it is at the origin of the solar wind.

How can the corona be hotter than the photosphere? Normally, heat should flow from the hot corona to the cold photosphere, not the opposite; how then does the Sun manage to heat the corona? This question has worried several generations of physicists who have attacked it on many fronts, but nobody has yet come out with a satisfying answer. The fundamental gaps are so important that: 'we cannot state at the present time why the Sun is obliged by the basic laws of physics to produce the heliosphere', as Eugene Parker put it not long ago [26]. We will discuss this question in Section 4.6.

We have seen that the density decreases exponentially outwards with a small scale height at the photosphere; this density decrease continues in a gentler way in the chromosphere, with a somewhat greater scale height – due to the greater temperature (cf. (1.10)); and a further density decrease by two orders of magnitude occurs in the transition region. Hence the corona is a very tenuous medium that does not radiate much.

Visible coronal radiation

The brightness of the brighter inner part of the corona in visible radiation is indeed only 10^{-6}th of the solar disc's brightness. This faint radiation is produced by scattering of sunlight in the Earth's atmosphere and barely reaches 10^{-4}th of the sky luminosity; hence, the corona cannot be seen from Earth under normal conditions. However, by a happy accident, the Sun and the Moon are seen from the Earth with virtually the same angular size, so that the Moon occults precisely the whole solar disc during total eclipses. Since this occultation occurs above the Earth's atmosphere, the sky brightness is also lowered, so that the corona can be seen on these occasions, even by the unaided eye. Total eclipses are rare, and other techniques have been developed to study the corona and the chromosphere, in particular by devising artificial

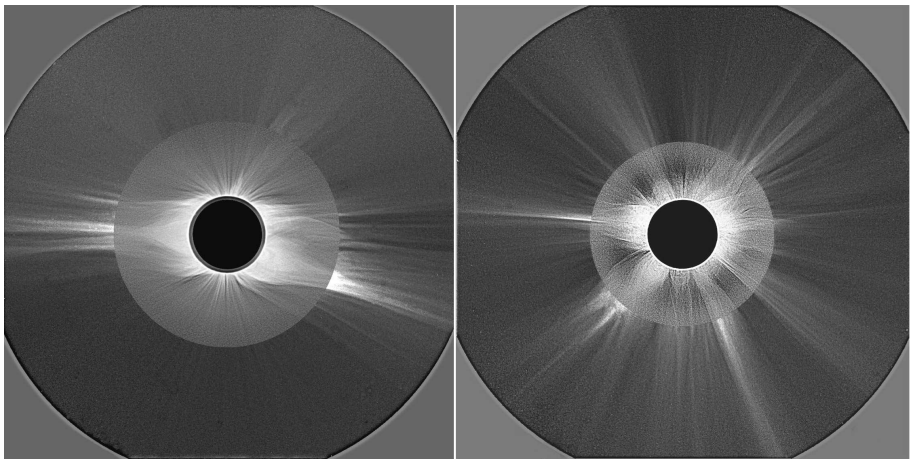

Figure 1.12 The solar corona after solar activity minimum (left panel) and near solar activity maximum (right panel). The images are composites of eclipse photographs taken respectively in Guadeloupe on 26 February 1998 (left, obs. C. Viladrich) and in Angola on 21 June 2001 (right, obs. J. Mouette), and nearly simultaneously from the LASCO-C2 coronagraph on the spacecraft SoHO (ESA and NASA). (Composites by Institut d'Astrophysique de Paris – CNRS; by courtesy of S. Koutchmy.)

eclipses inside optical instruments. Figure 1.12 is a composite obtained with both techniques, at different epochs of solar activity. The inner part shows the visible appearance of the corona observed from Earth during an eclipse; the outer part is obtained by making an artificial eclipse aboard a spacecraft on the same days.[13]

These images raise a number of questions. In the first place, what produces the observed radiation?

This question can be addressed, up to a point, in a very simple way. At $T \simeq 10^6$ K, the mean kinetic energy of a particle, $3k_BT/2$, amounts to 130 eV – roughly ten times more than the binding energy of the hydrogen atom. Hence not only do ambient electrons have enough energy for ionising hydrogen atoms by collisions and knocking out the outer electron of heavier atoms, but they can also knock out a large part of the more strongly bound inner electrons. The corona is thus a mixture of ions, including several-times ionised ones, and free electrons: a plasma. These free electrons are subjected to the solar radiation, are accelerated by the wave electric field, and in turn radiate at the same frequency; this is called *Thomson scattering*. Thomson scattering of sunlight is responsible for the visible radiation of the corona at altitudes up to a few solar radii.

From analysis of this radiation, one can deduce the density of electrons in the corona. Let us perform a simple estimate. The lower part of the corona is

[13] An opaque disc is put into the telescope to mask out the bright central emitting region.

close to hydrostatic equilibrium, hence the density is expected to fall roughly in an exponential way with a scale height given by (1.10); H is in fact twice larger because the medium is essentially made of protons and electrons, so that the mean mass per particle is $\mu \simeq (m_p + m_e)/2 \simeq m_p/2$ instead of m_p; we will return to this point later. With $T \simeq 10^6$ K, we find $H \simeq 0.1 \times R_\odot$. With such a small scale height, most of the scattered radiation comes from the electrons lying low in the corona, at distances from the Sun's centre close to $R_\odot$, so that the flux of solar radiation they receive is about $L_\odot/4\pi R_\odot^2$.

The ratio of the power radiated by one electron to the incident flux of radiation is given by Thomson's cross-section

$$\sigma_T = 8\pi r_e^2/3 \tag{1.11}$$

where the so-called classical electron radius is

$$r_e = \frac{e^2}{4\pi\epsilon_0 m_e c^2} = 2.8 \times 10^{-15} \text{ m} \tag{1.12}$$

so that each electron radiates per second:

$$\sigma_T \times L_\odot/4\pi R_\odot^2 \simeq L_\odot r_e^2/R_\odot^2 \quad \text{(radiation of one electron)} \tag{1.13}$$

from (1.11). We have seen that the coronal brightness at low altitudes is about 10^{-6}th of the solar value. Therefore the total number of scattering electrons is about

$$N \sim 10^{-6} \times R_\odot^2/r_e^2 \sim 6 \times 10^{40}.$$

A large fraction of this number lies close to the base of the corona, thus within a thin shell of surface $4\pi R_\odot^2$ and width H; we deduce the mean electron number density near the base of the corona

$$n_e \sim \frac{N}{4\pi R_\odot^2 H} \sim 10^{14} \text{ m}^{-3}.$$

We will see in Section 4.1 that the actual electron density there is indeed of this order of magnitude.

Coronal structure

The appearance of the corona in Fig. 1.12 raises another question. Near solar activity minimum, one sees bright complex structures at small and mid latitudes, with bright streamers extending more or less radially above them, whereas the polar regions appear rather uniform. In contrast, near solar activity maximum bright structures are seen all around the Sun.

What is the origin of these structures and of their variation with solar activity?

The observed emission is produced by particles, which we have seen to be electrically charged. Charged particles act as tracers of magnetic field lines,

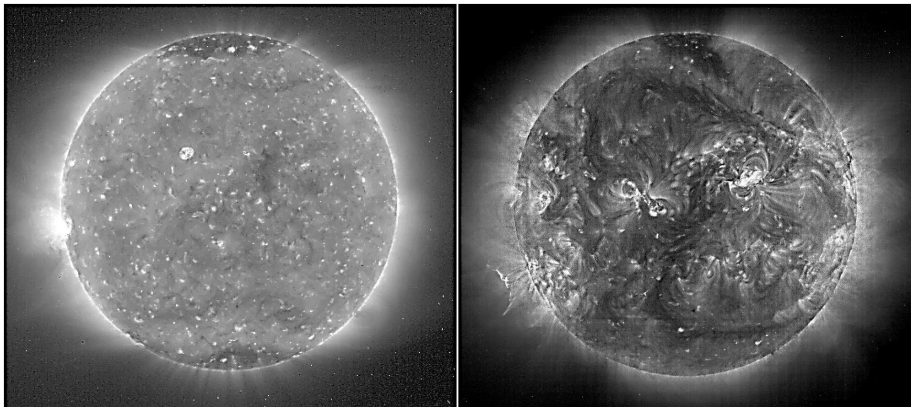

Figure 1.13 Images of the Sun from the Extreme Ultraviolet Imaging Telescope (at 195 Å) on the spacecraft SoHO on respectively 5 May 1996, when solar activity was minimum (left panel), and on 2 November 2001, when solar activity was maximum (right panel). (Images by SoHO/EIT consortium, ESA and NASA.)

somewhat as do iron filings sprinkled on a sheet of paper placed near a magnet. Indeed, as we will see in Section 2.3, the plasma and the magnetic field are intimately linked together, and one may ask whether the magnetic field might drive the plasma, somewhat as a magnet drives iron filings. A hint can be obtained by comparing the energy densities. We have seen that at the photosphere, the thermal energy of the particles is normally greater than the magnetic energy, except in the localised regions where the magnetic field is very strong. However, the fast density decrease with altitude makes the particle energy fall more rapidly than does the magnetic energy. In the upper chromosphere and the lower corona, therefore, magnetic energy tends to dominate, so that the magnetic field is expected to drive the plasma. Hence the changes in the plasma structures are driven by changes in the solar magnetic field during the sunspot cycle.

Coronal emission of ultraviolet and X-rays

We saw that the radiation responsible for the white light images of the corona shown in Fig. 1.12 is mainly produced by scattering of solar radiation by free electrons. It is interesting to compare these images to images taken in the X-ray or the extreme ultraviolet range.[14] Figure 1.13 shows such images, obtained in a narrow spectral band in the extreme ultraviolet when solar activity was respectively minimum (left panel) and maximum (right panel).

How is this emission produced? We have seen that the coronal temperature is so high that heavy elements are stripped of many of their electrons. The

[14] These images must be taken from space since radiation at X-ray and extreme ultraviolet wavelengths is absorbed by the upper atmosphere of the Earth.

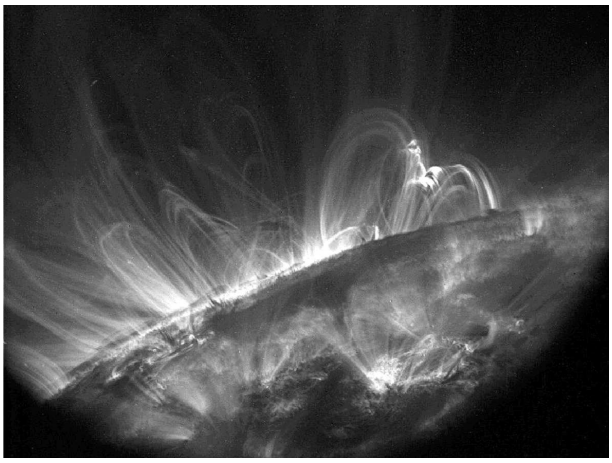

Figure 1.14 Image taken by the Transition Region and Coronal Explorer (TRACE) in April 2001 in a narrow spectral band around 171 Å, containing emission lines of eight- and nine-times ionised iron atoms that reveal structures having a temperature of about 10^6 K. (Image by Stanford-Lockheed Institute for Space Research and NASA.)

collisions with ambient electrons that ionise these elements can also knock their bound electrons into higher energy levels from which they radiate, producing spectral lines. Since the mean energy of electrons is $3k_BT/2$ and that of photons of wavelength λ is hc/λ, the plasma at $T \simeq 10^6$ K (or more) should radiate strongly at wavelengths of the order of $\lambda \simeq 2hc/3k_BT \simeq 10^{-8}$ m (or less), which corresponds to the extreme ultraviolet or the X-ray range. Furthermore, the emission is proportional to the number of emitting ions and to the number of exciting electrons, so that it increases as the density squared, thereby increasing the contrast of the structures.

Another reason for the contrast of those images taken in a narrow spectral band is that different spectral lines are emitted by different ions, which in turn are abundant in different temperature ranges. Hence a given spectral line reveals plasma at a particular temperature. For example, the spectral line in which the images of Fig. 1.13 have been obtained is emitted by iron atoms that are ionised 11 times and are abundant around 1.5×10^6 K; at lower temperatures, ambient electrons have not enough energy to produce such a high degree of ionisation, whereas at higher temperatures, they tend to produce a higher degree of ionisation.

Figure 1.13 shows two extreme types of large-scale structures: bright and dark. Bright regions appear highly structured and are especially numerous at activity maximum. As on images in the visible range, these bright structures outline the lines of force of the magnetic field. These lines emerge from the Sun and close back to its surface, forming large arches or loops, which are generally located above sunspot groups. Figure 1.14 is an image of such loops; one sees

that they have a fibril or thread-like structure. On the other hand, dark regions appear more uniform (although they do contain small-scale structures), and they radiate less essentially because they are less dense. Near solar activity minimum these dark *coronal holes* are located around the solar poles; over each coronal hole the large-scale vertical field is observed to have a uniform direction that is opposite in the north and south polar caps. Around activity maximum, when the predominant magnetic field polarity around the poles is being reversed, the polar coronal holes disappear but smaller ones appear throughout the disc.

On these basic structures are superimposed a host of time-varying phenomena having virtually all scales from the solar radius down to the smallest observable scale. What produces these features? We shall tackle this intriguing question in Chapter 4.

1.3 Observing the solar wind

2ND WITCH I'll give thee a wind.
1ST WITCH Th'art kind.
3RD WITCH And I another.
1ST WITCH I myself have all the other,
And the very ports they blow,
All the quarters that they know ...
W. Shakespeare, *Macbeth*

A solar astronomer is like a child in a toy museum, who can look but not touch (even via instruments) – until some bold and insightful space agency sends a probe to the Sun (Fig. 5.15). This involves the difficult art of determining where the radiation comes from and how it has been altered by propagation. The solar wind physicist has an easier task: he or she can make direct measurements with space probes that analyse *in situ* the solar wind particles and fields. Since the beginning of the space age, several tens of spacecraft have explored the heliosphere at virtually all latitudes up to the outskirts of the solar system, and returned a host of data (Fig. 1.15). Strangely enough, they have found that there is not a single wind but several – albeit one is more basic than the others.

1.3.1 Observing near the ecliptic

Some subtleties of space exploration

Most space probes lie close to the *ecliptic* – the plane in which the Earth and most of the planets orbit the Sun. There is a simple reason for that. A spacecraft leaving the Earth starts with a velocity vector equal to the Earth's orbital velocity plus that provided by the launcher; since the Earth's velocity is about $30\,\mathrm{km\,s^{-1}}$ and lies in the ecliptic plane, one must give to the spacecraft a velocity perpendicular to the ecliptic of at least this amount to put it into an orbit angled far from this plane; this is outside the capabilities of existing rocket technology (Fig. 1.16). As a result, there is an armada of space probes exploring the solar wind near the ecliptic.

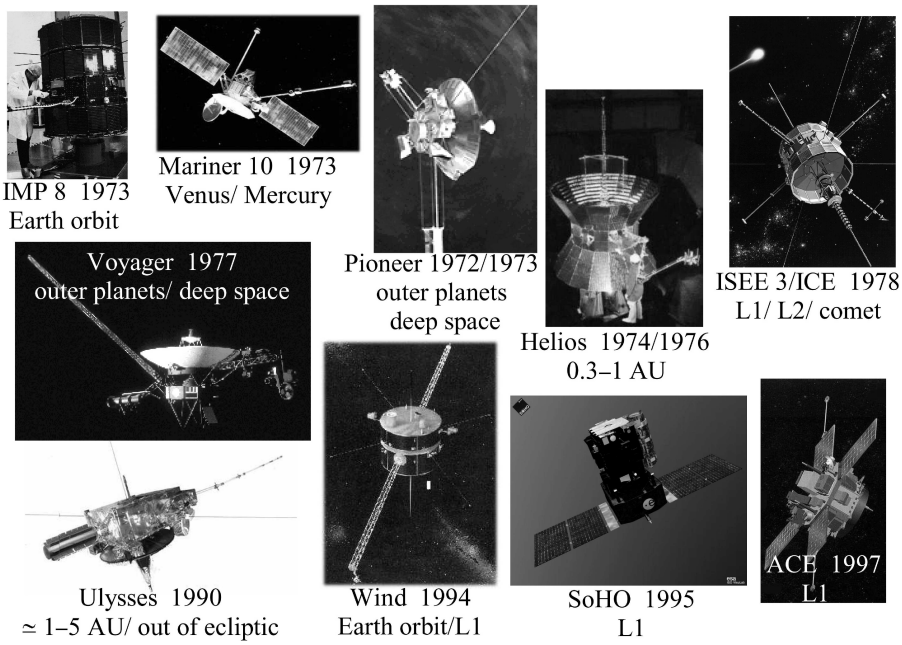

Figure 1.15 Some notable spacecraft that have explored the solar wind, with their dates of launch, and their main targets. (Images by NASA and ESA.)

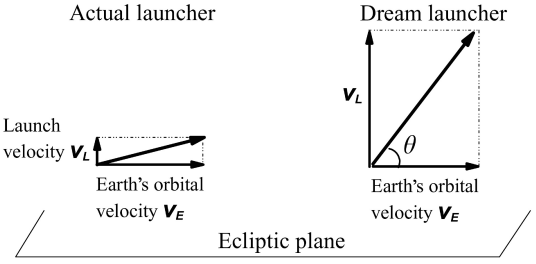

Figure 1.16 The difficulty of sending a spacecraft outside the ecliptic: because the Earth speed v_E is much greater than the launch speed v_L, the spacecraft velocity $\mathbf{v}_L + \mathbf{v}_E$ in the solar frame makes a very small angle to the ecliptic.

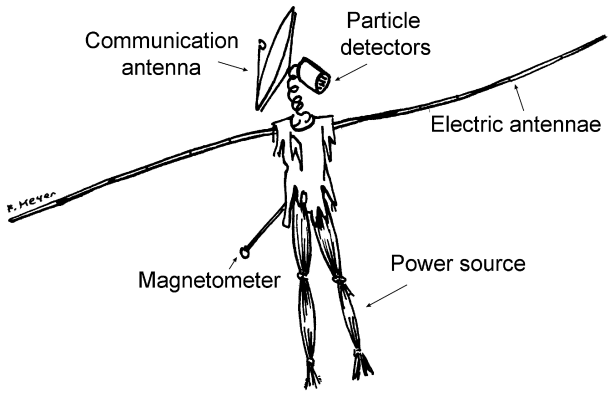

Figure 1.17 Simplified 'strawman payload' of the space agencies: the minimum required for solar wind *in situ* observation. (Drawing by F. Meyer.)

At the extremes of distances, the spacecraft Helios 1 and 2, launched respectively in 1974 and 1976, have approached the Sun up to about 0.3 AU, whereas Pioneer 10 and 11 (launched respectively in 1972 and 1973) and Voyager 1 and 2 (launched in 1977) are aiming at the outskirts of the Solar System. A horde of spacecraft are or have been watching the solar wind impinging on the Earth at 1 AU from the Sun. For doing so, they lie between the Sun and the Earth, at the right position for the Sun's and the Earth's gravitational attractions to combine to make the spacecraft orbit the Sun at the same angular speed as does the Earth.[15]

The solar wind, as the corona, is essentially made up of electrons and protons plus a small proportion of heavier ions, and it carries a magnetic field. As we shall see in Section 2.3, particles and fields are intimately coupled in plasmas, so that in order to explore them, space probes should carry at least a particle detector, a magnetometer and an electric antenna measuring waves, in addition to power and communication resources and to the necessary software (Fig. 1.17); most spacecraft generally carry additional instruments.

Several winds

Figure 1.18 shows typical measurements of key solar wind parameters: the mean velocity of protons (top panel) and the mean number density of electrons[16] (middle panel) obtained by the spacecraft WIND (operated by NASA [16]) in June 1995, close to solar activity minimum.[17]

[15]The so-called L1 Lagrange point, located at about 250 Earth's radii from the Earth towards the Sun.

[16]Because protons and electrons are the main constituents and carry opposite charges, their number densities and mean velocities must be roughly equal to keep the plasma electrically neutral.

[17]The mean velocity of protons (roughly equal to their mean radial velocity) and the density of electrons are plotted as (about) 10 minutes averages of data from respectively the ion

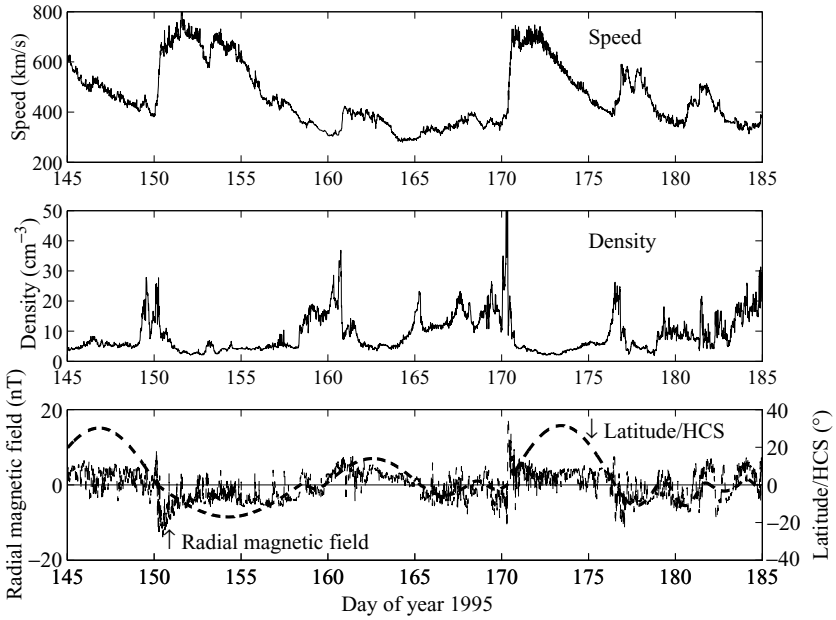

Figure 1.18 Mean proton velocity (top panel) and mean electron density (middle panel) measured on the spacecraft WIND in June 1995, at 1 AU from the Sun in the ecliptic. The bottom panel shows the radial component (Sun-centred) of the magnetic field and the latitude of WIND with respect to the heliospheric current sheet (HCS), that we will discuss in Sections 1.3.3 and 6.2. (Data courtesy of C. Salem.)

What does Fig. 1.18 tell us? In addition to small fluctuations at short timescales, the speed and the density exhibit a characteristic pattern: the speed varies by up to a factor of two or more and the density varies by up to a factor of ten, nearly simultaneously at intervals of about a week. The abrupt speed increases are followed by slow decreases; the density peaks when the speed changes abruptly, and (apart from the density peaks) the density is low (and uniform) when the speed is high and vice versa. This pattern of alternate fast and slow streams is repeated – with some modification – at the rotation period of the Sun, and is accompanied by changes in the magnetic field and in other properties. One sees in the bottom panel of Fig. 1.18 that the sign of the radial component of the magnetic field changes as a new fast stream is encountered, and remains constant within it.

---

electrostatic analyser (Lin, R. P. *et al.* 1995, *Space Sci. Rev.* **71** 125) and the thermal noise (Meyer-Vernet, N. *et al.* 1998, in *Measurement Techniques in Space Plasmas: Fields, Geophys. Monogr. Ser.* **103**, ed. R. F. Pfaff, *et al.*, Washington, DC, American Geophysical Union, p. 205) part of the WAVES receiver (Bougeret, J.-L. *et al.* 1995, *Space Sci. Rev.* **71** 231), which are the best instruments for measuring these respective properties on WIND (Salem, C. *et al.* 2003, *Ap. J.* **585** 1147).

This pattern has been observed by all spacecraft orbiting near the ecliptic at moderate distances from the Sun. It suggests the existence of several wind states: one slow/dense/structured and one fast/tenuous/uniform, in addition to transient and intermediate stages where the wind properties are still different (Section 6.3).

What is the origin of these states and of the changes between them? To understand this, we must realize that all these observations were performed near the ecliptic plane. This plane is very peculiar because the solar spin axis makes an angle of only 7.25° with the normal to the ecliptic. Furthermore near solar activity minimum, the dipolar component of the solar magnetic field is also nearly parallel to the spin axis, but not exactly so (making an angle generally larger than 7.25°). As the Sun spins, rotation of the Sun's dipolar magnetic pattern places an in-ecliptic observer alternatively in two opposite magnetic hemispheres. In contrast, a spacecraft sufficiently far from the ecliptic would remain in the same solar magnetic hemisphere as the Sun spins.

1.3.2 Exploring the third dimension with Ulysses

> O frati, dissi, che per cento milia
> Perigli siete gicenti all'occidente,
> A questa tanto picciola vigilia
> De' vostri sensi, ch'è del rimanente
> Non vogliate negar l'esperienza,
> Diretro al sol, del mondo senza gente.
> *La Divina Commedia di Dante Alighieri,*
> *Canto XXVI*[18]

So did Ulysses encourage his sailors in Dante's version of the Odyssey. Less famous is the record of a round-table discussion that gathered a lot of distinguished scientists in 1959, at a symposium on the exploration of space [28]:

> *Mr Hibbs*: I should like to ask whether there is a particular importance in performing experiments out of the plane of the ecliptic.

The fortitude of the scientists and engineers who made this dream come true in spite of the difficulties and sent the spacecraft Ulysses (Fig. 1.19) where no probe had ever flown is reminiscent of the mythical Greek warrior.

When the idea of an out-of-ecliptic mission arose, nobody knew how to realise it, and only in the 1970s did the idea appear technically feasible. The American and European space agencies then proposed a joint package of two spacecraft that were to be launched in 1983, and to sweep towards opposite sides of the ecliptic plane – using Jupiter's assist[19] – in order to pass nearly

[18] Tell me brothers, would you,/who braved a hundred thousand perils/to go ever farther to west,/now be loath to roam a realm/which reaches to the Sun/and harbours not a single living soul? (Trans. L. M. Celnikier.)

[19] Gravity assist from a planet is like playing billiards with a spacecraft, using the planetary gravitational field to deviate the spacecraft, as the edges of a billiard table deviate the ball.

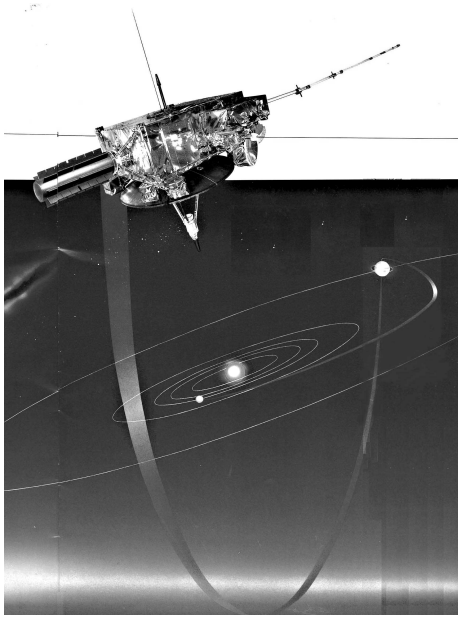

Figure 1.19 An artist's view of Ulysses, the first and only spacecraft to have reached high heliocentric latitudes (ESA and NASA).

simultaneously over opposite solar poles and to achieve a stereoscopic view of the solar wind [31]. But in the early 1980s, the National Aeronautical and Space Administration (NASA) decided to cancel the US spacecraft because of financial and technical difficulties. The project was reduced to a single spacecraft, to be built by the European Space Agency (ESA), launched by NASA with the Space Shuttle, and equipped with European and American instruments. In late 1983, however, the mission had still to wait: the spacecraft was ready but the launcher was not. And 1986 saw a catastrophic event: the Space Shuttle Challenger blew up, a few months before the planned launch of Ulysses, once more delaying the mission.

The launch took place at last in October 1990 – close to solar activity maximum. In 1991 the probe travelled in the ecliptic towards Jupiter. In February 1992 it swung around Jupiter into an elliptic orbit inclined by 80° to the ecliptic (Fig. 1.20). It then travelled into the Sun's southern hemisphere, passed over the south polar region in late 1994, crossed the ecliptic plane at 1.3 AU from the Sun, and passed over the north polar region in 1995 – near solar activity

Since the planet is moving, the spacecraft's speed with respect to the Sun changes upon reflection, even though the reflection is elastic in the frame of the planet; in this way, the spacecraft can not only change direction, but also gain or lose kinetic energy at the expense of the planet (see Gurzadyan, G. A. 2002, *Space Dynamics*, New York, Taylor & Francis). We will encounter a similar effect in Section 2.1, with the electromagnetic fields replacing gravitation, and charged particles replacing spacecraft.

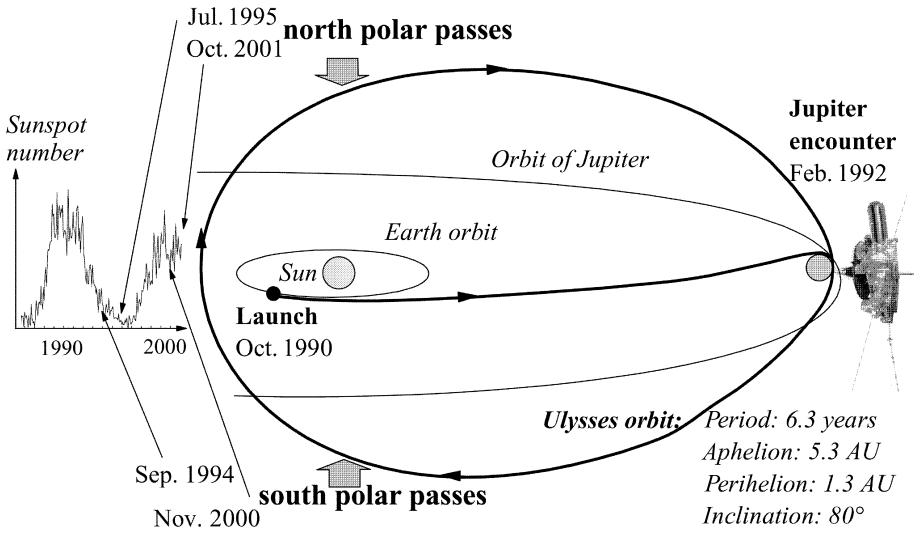

Figure 1.20 Sketch of Ulysses' trajectory. The dates of the first two passages over the solar poles are indicated on the left, along with solar activity. The orbital period is such that the fast pole-to-pole transits, covered in less than a year on the perihelion side, take place alternatively near sunspot minimum and maximum.

minimum. The second orbit took it over the polar regions once more in 2000–2001, this time near solar activity maximum. And a third orbit will take it again over the polar regions in 2007, near solar activity minimum. Ulysses carries a panel of instruments measuring the charged and neutral particles over a wide range of energies, the dust grains, the magnetic field and waves, including X- and gamma rays [32].

The orbit is especially suitable for studying the heliosphere. The orbital period is nearly half a solar activity cycle, and the pole-to-pole transit near perihelion takes less than a year – a time-span during which solar activity and distance do not change much. Hence at each passage along this part of the orbit, which take place alternately near solar activity minimum and maximum, Ulysses measures how the solar wind varies with heliocentric latitude, other parameters being roughly constant. On the other hand, the distance, latitude and solar activity vary simultaneously during the aphelion phase, when the spacecraft is moving less rapidly.

The solar wind in three dimensions

What did Ulysses find?

First of all, did Ulysses observe the recurrent pattern of two wind states observed by in-ecliptic spacecraft? Figure 1.21 shows the mean proton velocity (top panel) and the electron density (middle panel) measured by Ulysses in

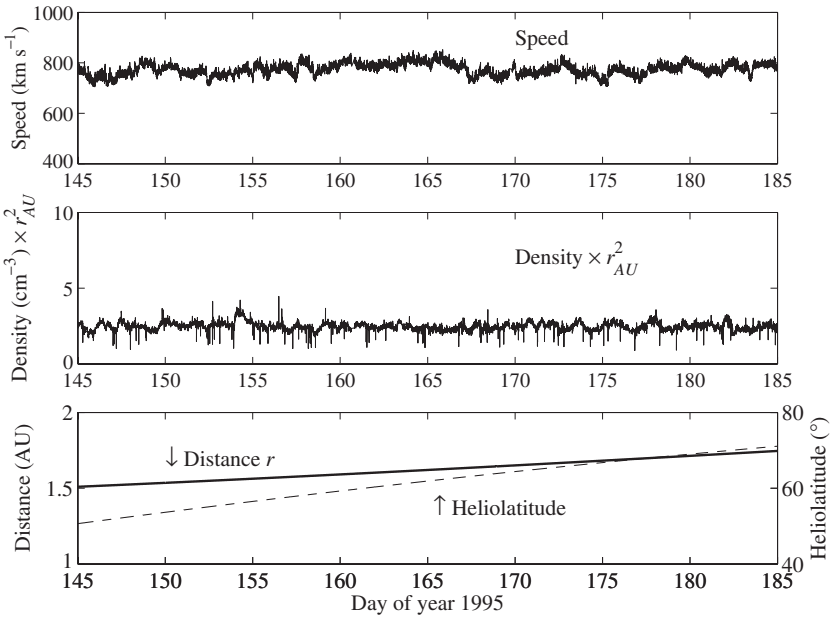

Figure 1.21 Mean proton velocity (top panel) and mean electron density (middle panel) measured on Ulysses in June 1995. The density is normalised to 1 AU by multiplying by Ulysses' distance (in AU) squared. The radial magnetic field (not shown) has a constant (positive) sign. The bottom panel shows the spacecraft heliocentric distance (thick line) and latitude (dashed line) in solar co-ordinates. (Data courtesy of K. Issautier.)

June 1995, the same month as the WIND observations plotted on Fig. 1.18 were acquired.[20] Figure 1.21 does not show any trace of the two-state wind pattern observed by in-ecliptic spacecraft. The speed remains close to 750 km s^{-1}, while the density (normalised to the distance of 1 AU) is also roughly constant at 2.5 cm^{-3}, as expected from conservation of particles expanding radially at a constant speed. Note that the density scale in Fig. 1.21 is enlarged by a factor of five with respect to that in Fig. 1.18, to emphasise the small variations. In June 1995, Ulysses was at about 1.6 AU from the Sun and 65° heliolatitude, i.e. very far from the equatorial plane of the solar magnetic dipole.

What did it observe elsewhere?

This is best summarised in Fig. 1.22, which shows how the speed changes with latitude and with solar activity. On this figure, the speed is plotted in

[20]The speed is from the SWOOPS particle electrostatic analyser (Bame, S. J. *et al.* 1992, *Astron. Astrophys. Suppl. Ser.* **92** 237), and the density from thermal noise analysis of data from the URAP instrument (Stone, R. G. *et al.* 1992, *Astron. Astrophys. Suppl. Ser.* **92** 291), that are the best ways of measuring these respective properties on Ulysses (Issautier, K. *et al.* 1999, *J. Geophys. Res.* **104** 6691).

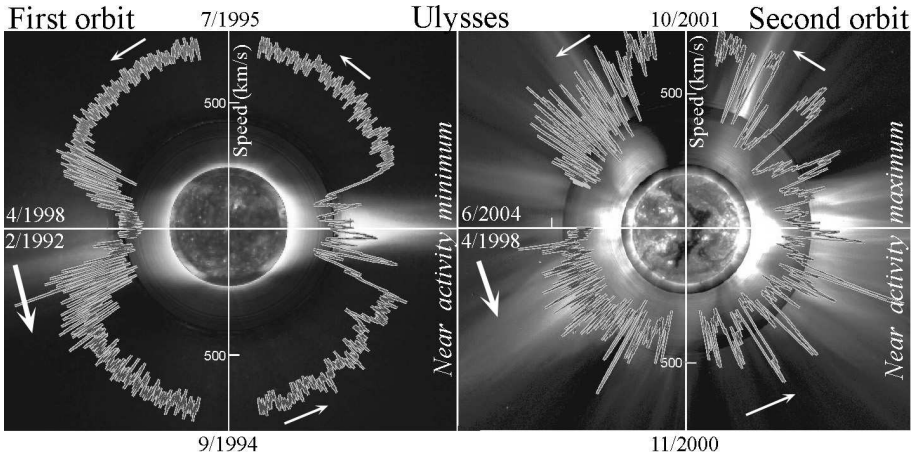

Figure 1.22 Solar wind speed as a function of heliocentric latitude, plotted in polar co-ordinates during Ulysses' first two orbits. The data are plotted over solar images obtained on 17 August 1996 (near activity minimum; left) and 7 December 2000 (near activity maximum; right), with SoHO EIT(195 Å), and LASCO instruments and the Mauna Loa K coronameter (0.7–0.95 μm). (Adapted from [19].)

polar co-ordinates, i.e. the distance from the centre is proportional to the speed at each latitude. The left and right panels show the first and second orbits, which took place respectively around solar activity minimum and maximum. The data are superimposed on images of the corona typical of these periods of solar activity. In both panels, time starts on the left-hand side (southwards) and progresses counter-clockwise along the orbit as indicated by the arrows and the dates.

Consider first the structure near solar activity minimum (left-hand panel of Fig. 1.22). The coronal image on which the data are superimposed shows the simple structure we have already seen in the left-hand panel of Fig. 1.12, with dark coronal holes on the polar caps and bright streamers extending outwards near the equatorial plane. The solar wind structure reflects this simplicity: the speed is nearly constant at all latitudes except in a narrow band of $\pm 20^\circ$ around the equator where the speed pattern resembles the two-state structure seen by near-ecliptic spacecraft. This simplicity is shared by the other properties. The sign of the radial component of the magnetic field remains constant within each hemisphere (except for some brief reversals), being outwards in the north and inwards in the south, except in a narrow equatorial band where polarities are alternating.

Around solar activity minimum, therefore, the heliosphere has an outstandingly simple structure, essentially made of a quiet, fast and tenuous wind where the radial component of the magnetic field has a constant sign (opposite in opposite hemispheres). Comparison with solar observations shows that this fast

Table 1.2 *Occurrence and variability of the fast and slow winds*

	Fast wind	Slow wind
At activity minimum	Ubiquitous outside equator	Only near equator
At activity maximum	Occurs as narrow streams	Ubiquitous outside streams
Variability	Low	Large

Table 1.3 *Basic properties of the fast and slow winds*

	Speed v ($\mathrm{m\,s^{-1}}$)	Electron density n ($\mathrm{m^{-3}}$)	Mass loss through a sphere $\left[\rho v \times 4\pi r^2\right]_{1\mathrm{AU}}$ ($\mathrm{kg\,s^{-1}}$)	Ram pressure $\left[\rho v^2\right]_{1\mathrm{AU}}$ (Pa)
Fast	7.5×10^5	2.5×10^6	10^9	2.6×10^{-9}
Slow	4×10^5	7×10^6	1.5×10^9	2.1×10^{-9}

wind arises from the inactive solar regions, especially the large coronal holes surrounding each pole, where the magnetic field has a constant polarity (opposite at opposite poles). The pattern of fast and slow winds recurring at the solar rotation period is restricted to a narrow latitude band surrounding the solar equatorial plane, that is akin to the equatorial region where bright streamers are observed in the corona. This picture is consistent with observations by near-ecliptic spacecraft, which are located at low latitudes.

Consider now the structure near activity maximum (right-hand panel of Fig. 1.22). The superimposed coronal image shows the complex picture typical of solar activity maximum, which we have already seen in the right-hand panel of Fig. 1.12, with bright streamers extending radially all around the Sun. The solar wind structure reflects this complexity, with alternating fast and slow streams of small scale – observed at all latitudes, in addition to transients. This complex structure is shared by the magnetic field, whose polarity alternates, and by other properties. Near activity maximum, therefore, the pattern of alternating fast and slow streams is observed at all latitudes, with, however, somewhat smaller speeds and scales than near activity minimum.

We have summarised in Tables 1.2 and 1.3 the basic properties of these wind states (see [21], [19]).[21]

1.3.3 A simplified three-dimensional picture

These Ulysses observations bear out – albeit with some modifications – a simple picture of the heliosphere near solar activity minimum, that had already been

[21] The values of ρv and ρv^2 are slightly greater than $nm_p v$ and $nm_p v^2$ respectively because there is a small proportion of helium.

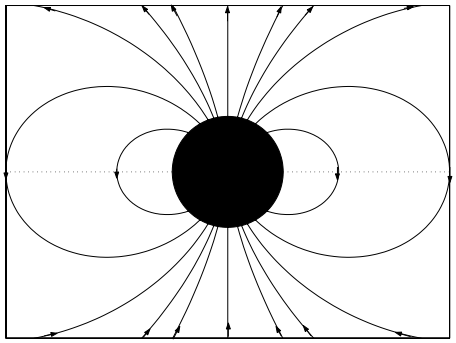

Figure 1.23 Magnetic field lines of a magnetic dipole aligned with the vertical axis. The black disc sketches the region (here the solar interior) where the currents producing magnetic fields are flowing.

hinted at from remote-sensing observations, and from data of previous spacecraft that had gone slightly outside the ecliptic.

Dipolar magnetic field

To begin with, let us assume that the Sun behaves as a huge magnetic dipole, producing a magnetic field as sketched on Fig. 1.23. The actual solar magnetic field is more complicated, but the dipole is a reasonable starting approximation because the farther out from the Sun, the smaller are the effects of the solar non-dipolar components. Indeed, outside the region containing electric currents, Maxwell's equations

$$\begin{aligned} \nabla \times \mathbf{B} &= \mu_0 \mathbf{J} \\ \nabla \cdot \mathbf{B} &= 0 \end{aligned} \tag{1.14}$$

with $\mathbf{J} = 0$, imply $\mathbf{B} = -\nabla\psi$, with the Laplace equation $\nabla^2\psi = 0$; hence the magnetic field can be developed in a multipolar series where the term of order n decreases with distance as $r^{-(n+2)}$ (e.g. [8]). The dipolar term corresponds to $n = 1$ and becomes dominant with increasing distance.

With the components of the dipolar magnetic field in the radial (B_r) and latitudinal (B_θ) directions (Appendix), the field lines are solutions of the differential equation $dr/(rd\theta) = B_r/B_\theta = -2\sin\theta/\cos\theta$, where r and θ are respectively the distance and latitude in a spherical co-ordinate system whose symmetry axis (vertical) is aligned with the magnetic dipole axis. Putting $u = \cos\theta$, this yields $dr/r = 2du/u$, whose integration gives $r \propto u^2 = \cos^2\theta$. Hence, a field line crossing the magnetic equatorial plane ($\cos\theta = 1$) at heliocentric distance L is given by $r = L\cos^2\theta$.

In this geometry, all field lines are closed; each one leaves the surface, extends outwards to a maximum distance L in the magnetic equatorial plane and then returns towards the surface on the other side (Fig. 1.23).

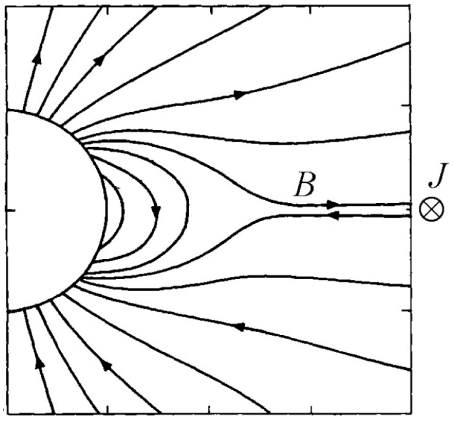

Figure 1.24 Simple model of how a dipolar magnetic field imposed at the solar surface is modified by a wind flowing along the field lines. At large distances the oppositely directed magnetic fields imply an electric current flowing in a thin sheet perpendicular to the figure, a picture that will be refined in Section 6.2. (Adapted from [27].)

A dipole plus a wind

This nicely simple picture, however, concerns a magnetic dipole in a vacuum, which the Sun is certainly not; the electric charges of the plasma around the Sun produce currents that change the magnetic field according to Maxwell's equations (1.14). Furthermore this plasma manages to flow outwards, with an energy density greater than that of the magnetic field, as we will see in Chapters 5 and 6. How does this modify the picture? We have already said (and shall explain in Section 2.3) that the plasma and the magnetic field lines are strongly tied together, so that the plasma can move along the field lines but not across them, just as beads on a necklace. Near the poles, where the magnetic lines extend nearly radially, the outgoing wind can flow unimpeded along them. In the equatorial regions, however, the situation is different because an outgoing wind would have to flow perpendicularly to the dipolar field lines (Fig. 1.23).

What happens in that case? Since the wind energy density exceeds that of the magnetic field, the wind pushes out the field, drawing the field lines outwards to the extent that they become nearly parallel to the equator and no longer return to the surface. In this way, the outgoing wind can flow along the field lines everywhere. This is sketched in Fig. 1.24, which shows a simplified solution of this problem [27]. Because the lines parallel to the equatorial plane come from opposite ends of the dipole, they represent magnetic fields having opposite directions. Hence, at large distances, the magnetic field direction changes abruptly at the equator. This implies that a thin sheet of current flows along the magnetic equatorial plane in a direction normal to the figure; in three

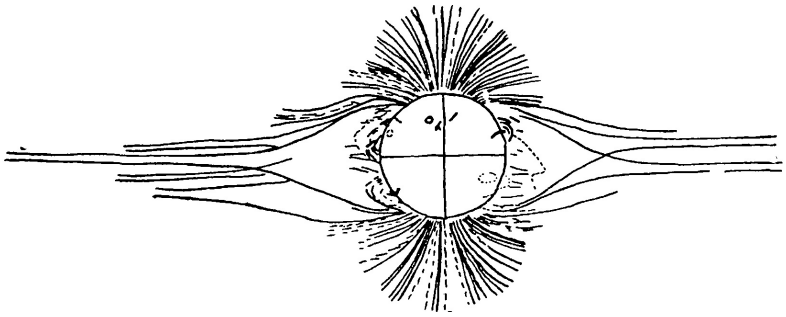

Figure 1.25 Drawing of the appearance of the corona at the 30 June 1954 eclipse, near solar activity minimum. (Adapted from [30].)

dimensions, it forms an annular current sheet that separates the magnetic fields originating at opposite poles.

Although Fig. 1.24 refers to a highly idealised situation, the real corona might sometimes resemble it, as suggested by Fig. 1.25, which is an old drawing of the visual appearance of the corona during a solar eclipse [30]. One must be cautious in interpreting such visual appearances since they result from different structures lying along the line of sight of the observer, and projection effects can be misleading. However, the figure suggests a current sheet extending into space near the solar magnetic equatorial plane, that separates regions of opposite magnetic polarities. We will see in Section 6.2 that this sheet is not a simple plane, having a complex warped shape which has been likened to the skirt of a spinning ballerina [1].

How does this picture change with solar activity? During a few years near minimum, when the solar magnetic field is not far from that of a dipole making a small angle with the rotation axis, the current sheet is a (slightly warped) plane making a small angle with the solar equator. As solar activity rises, the solar magnetic field becomes more complicated, making the sheet warp and thread its way towards polar regions, finally sometimes breaking up near activity maximum, when the large-scale solar magnetic field becomes rather disorganised. As solar activity decreases, the large-scale magnetic field reorganises itself towards a dipolar structure whose axis is again close to the spin axis, but with a direction opposite to the one in the previous cycle.

How is the wind velocity related to this magnetic structure? Naively, one expects that the flow of the wind will be unimpeded and stationary everywhere except near the current sheet where complex geometry-dependent effects should occur. In this sense, this simple picture, conceived in the mid 1970s, gives a straightforward explanation for the basic geometric structure of the heliosphere, sketched in Fig. 1.26. Returning to the WIND observations of Fig. 1.18, we can now relate them to the position of the spacecraft with respect to the heliospheric current sheet, which is shown in the bottom panel. We can see that the sign of the radial component of the magnetic field is the same as that of the latitude

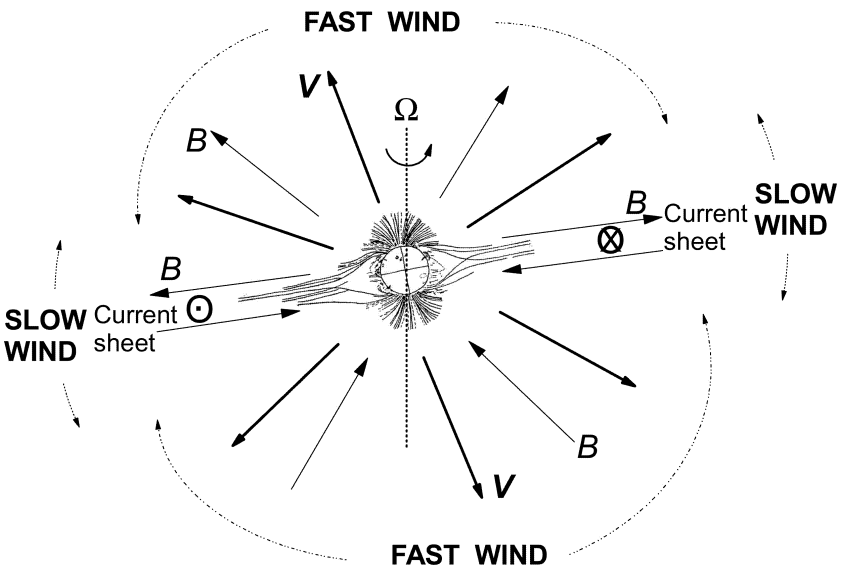

Figure 1.26 Simplified picture of the large-scale structure of the solar wind near sunspot minimum, when the solar magnetic dipole makes a small angle with the spin axis (dotted line). The velocity and field lines are sketched in bold and thin lines respectively. The magnetic polarity is the one that existed when the WIND and Ulysses observations shown in Figs. 1.18 and 1.22 were acquired; this polarity reverses every 11 years.

with respect to the current sheet, as expected from the sketch in Fig. 1.26. As the Sun spins, the WIND spacecraft in the ecliptic lies alternately above and below the heliospheric current sheet. At each crossing of the sheet, the direction of the magnetic field changes and a slow and variable wind is observed. This picture is also consistent with observations from Ulysses, which spends most of its trajectory far from the current sheet near solar minimum, and measures a fast stationary wind with a constant magnetic polarity in each solar hemisphere except at low latitudes.

We shall revisit this grossly simplified picture later. Before doing so, we have to introduce some basic plasma physics.

References

[1] Alfvén, H. 1977, Electric currents in cosmic plasmas, *Rev. Geophys. Space Phys.* **15** 271.

[2] Badescu, V. and R. B. Cathcart 2006, Use of class A and class C stellar engines to control Sun movement in the Galaxy, *Acta Astronautica* **58** 119.

[3] Biermann, L. 1953, Physical processes in comet tails and their relation to solar activity, *Mém. Soc. Roy. Sci. Liège (Ser. 4)* **13** 291.

[4] Birkeland, K. R. 1908, 1913, *The Norvegian Aurora Polaris Expedition 1902–1903,* vols. 1 and 2, Christiania, Norway, H. Aschehoug & Co.

[5] Carrington, R. C. 1860, Description of a singular appearance seen on the Sun on September 1, 1859, *Mont. Not. Roy. Soc.* **20** 13.

[6] Chapman, S. 1957, Notes on the solar corona and the terrestrial ionosphere, *Smithsonian Contr. Astrophys.* **2** 1.

[7] Clerke, A. M. 1887, *A Popular History of Astronomy during the 19th Century*, Edinburgh, Adam & Charles Black.

[8] Dennery, P. and A. Krzywicki 1967, *Mathematics for Physicists*, New York, Harper & Row.

[9] Dessler, A. J. 1967, Solar wind and interplanetary magnetic field, *Rev. Geophys.* **5** 1.

[10] Eather, R. H. 1980, *Majestic Lights*, Washington DC, American Geophysical Union.

[11] Egeland, A. and W. J. Burke 2005, *Kristian Birkeland: The First Space Scientist*, New York, Springer.

[12] Fitzgerald, G. F. 1892, Sunspots and magnetic storms, *The Electrician* **30** 48.

[13] Foukal, P. V. 2004, *Solar Astrophysics*, New York, Wiley.

[14] Giovanelli, R.G. 1984, *Secrets of the Sun*, Cambridge University Press.

[15] Haigh, J. D. *et al.* eds. 2004, *The Sun, Solar Analogs and the Climate*, New York, Springer.

[16] Harten, R. and K. Clark 1995, The design features of the GGS Wind and Polar spacecraft, *Space Sci. Rev.* **71** 23.

[17] Kippenhahn, R. 1994, *Discovering the Secrets of the Sun*, New York, Wiley.

[18] Lemaire, J. and V. Pierrard 2001, Kinetic models of solar and polar winds, *Astrophys. Space Sci.* **277** 169.

[19] McComas, D. J. *et al.* 2003, The three-dimensional solar wind around solar maximum, *Geophys. Res. Lett.* **30** 1517.

[20] Neugebauer, M. 1997, Pioneers of space physics: a career in the solar wind, *J. Geophys. Res.* **102** 26887.

[21] Neugebauer, M. 2001, in *The Heliosphere Near Solar Minimum*, ed. A. Balogh *et al.*, New York, Springer, p. 43.

[22] Neugebauer, M. and C. W. Snyder 1962, The mission of Mariner II: preliminary observations, *Science* **138** 1095.

[23] Parker, E. N. 1958, Dynamics of the interplanetary gas and magnetic fields, *Ap. J.* **128** 664.

[24] Parker, E. N. 1997, in *Cosmic Winds*, ed. J. R. Jokipii *et al.*, Tucson AZ, University of Arizona Press, p. 3.

[25] Parker, E. N. 2001, in *The Century of Space Science*, ed. J. A. M. Bleeker *et al.*, New York, Kluwer, p. 225.

[26] Parker, E. N. 2001, A critical review of sun–space physics, *Astrophys. Space Sci.* **277** 1.

[27] Pneuman, G. W. and R. A. Kopp 1971, Gas–magnetic field interactions in the solar corona, *Solar Phys.* **18** 258.

[28] Simpson, J. A. *et al.* 1959, Round-table discussion, *J. Geophys. Res.* **64** 1691.

[29] Title, A. 2000, Magnetic fields below, on and above the solar surface, *Phil. Trans. Roy. Soc. London A* **358** 657.

[30] Vsekhsvjatsky, S. K. 1963, in *The Solar Corona*, ed. J. W. Evans, London, Academic Press, p. 271.

[31] Wenzel, K.-P. 1980, The scientific objectives of the international solar polar mission, *Phil. Trans. Roy. Soc. London* A **297** 565.

[32] Wenzel, K.-P. 1992, The Ulysses mission, *Astron. Astrophys. Suppl. Ser.* **92** 207.

[33] Zirker, J. B. 2002, *Journey from the Center of the Sun*, Princeton University Press.