1
Development of the Ideas and Instruments of Modern Solar Research

1.1
Early Telescopic Discoveries on the Sun

Scientific inquiry into the nature of the sun began around 1610 with the first telescopic observations of sunspots (called maculae or blemishes) by J. Fabricius, Galileo Galilei, C. Scheiner, and others in Western Europe. Naked-eye sightings of dark markings on the sun's disk had been reported in China and elsewhere at least 15 centuries earlier, and this evidence was almost certainly familiar to Galileo and his contemporaries. But the telescope showed that the westward motion of the spots was fastest near disk center and relatively slower near the east and west limbs. Using this observation, Galileo was able to argue from the projection effects expected in circular motion that the spots must be dark markings on the sun's surface rather than planets or other bodies in distant orbit and transiting its disk.

By 1630, the Jesuit astronomer Scheiner had used the spots' daily motion to accurately measure the sun's equatorial rotation period of 27 days as seen from the orbiting earth and to determine the 7° tilt of its equator relative to the ecliptic plane. Scheiner also detected the longer rotation period of spots at high solar latitudes and inferred that the sun's angular rotation rate decreases toward the poles. The well-known confrontation between Galileo's presentation of these early discoveries on the sun and the Church's doctrine of immaculate and stationary celestial bodies made a direct impact on the cosmology of the day that remains unique in the history of solar research.

To understand the fate of solar research in the two centuries following these promising beginnings, we should recall that the sun's nature was about as dimly perceived in Galileo's time as is the nature of quasars today. A useful estimate of its distance and size became available only after 1673, when the distance of Mars from Earth was measured using that planet's parallax observed between Paris and French Guiana. The Earth-sun distance then followed from J. Kepler's third law (of 1619) to within 10 % of the correct value. The sun's mass relative to that of the Earth (determined from the orbital periods and distances of the Earth and moon) was not published by Isaac Newton until some 50 years after Galileo's first observations. The Earth itself was properly "weighed" by H. Cavendish's experiment only in 1798. Even the order of magnitude of the sun's surface temperature remained uncertain

Solar Astrophysics. Peter V. Foukal

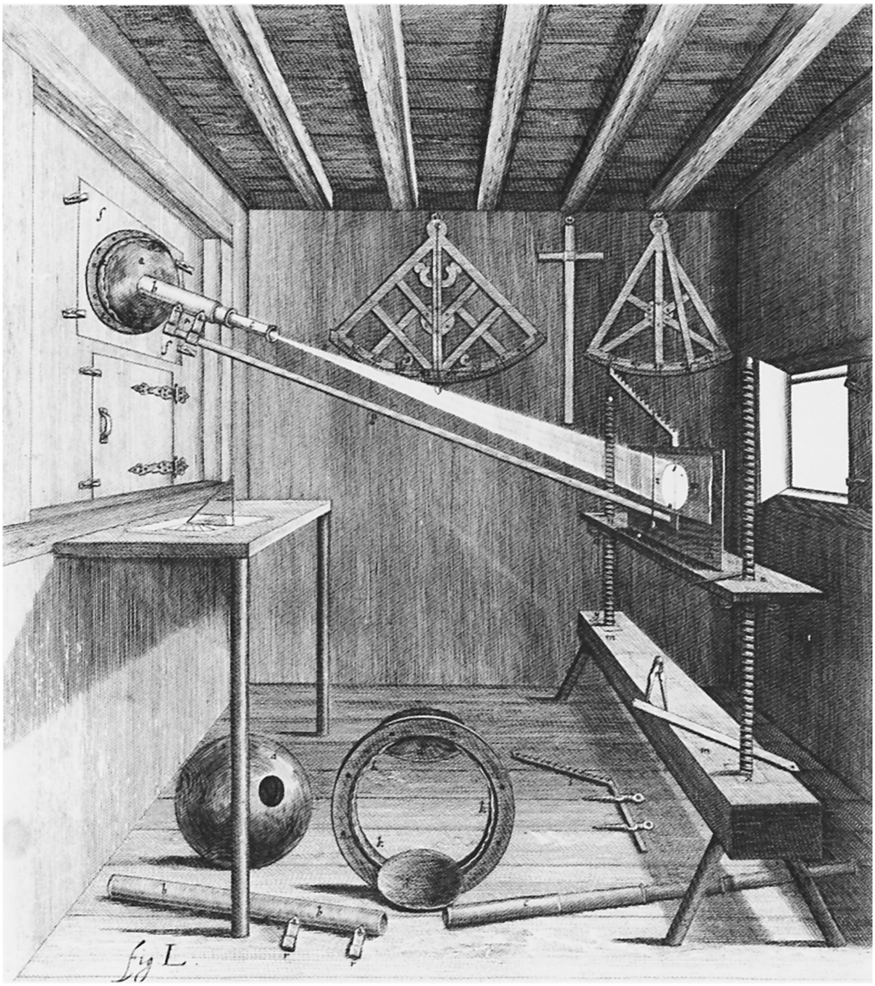

Fig. 1-1 Eyepiece projection scheme used by J. Hevelius to observe sunspots and faculae. From his *Selenografia*. By permission of the Houghton Library, Harvard University.

until after the 1880s, when the fourth-power relationship between the temperature of an opaque body and its emittance came to be accepted through the experimental work of J. Stefan and the thermodynamic arguments set forth by L. Boltzmann.

Given the slow development of the physics and chemistry required to interpret solar observations, it is not surprising that for roughly two centuries after the first telescopic discoveries solar research consisted mainly of sunspot observations. Observers in the early 17th century used non-achromatic refractors with spatial resolution of 10 to 15 arcsec, and for detailed work the solar image was usually projected onto a screen (Fig. 1-1). (Never look at the sun through any optical instrument without expert advice – serious eye damage could result). Drawings by J. Hevelius (Fig. 1-2) and Scheiner show that the darkest, roundish central umbra of the spot is often surrounded by a roughly annular lighter region, the penumbra. The relative positions and areas of individual spots within a group were observed to change from hour to hour.

These observers also noted bright irregular patches comparable in size to spots, that could often be observed in association with sunspots near the limb. Scheiner called them faculae or little torches. The contrast of faculae is much lower than that of spots, and they are not visible at all in white light near disk center. Faculae, and less often, spots, can be observed to recur on subsequent solar rotations, implying lifetimes of up to several months. Figure 1-2 shows the march of faculae and a sunspot across the disk from day to day.

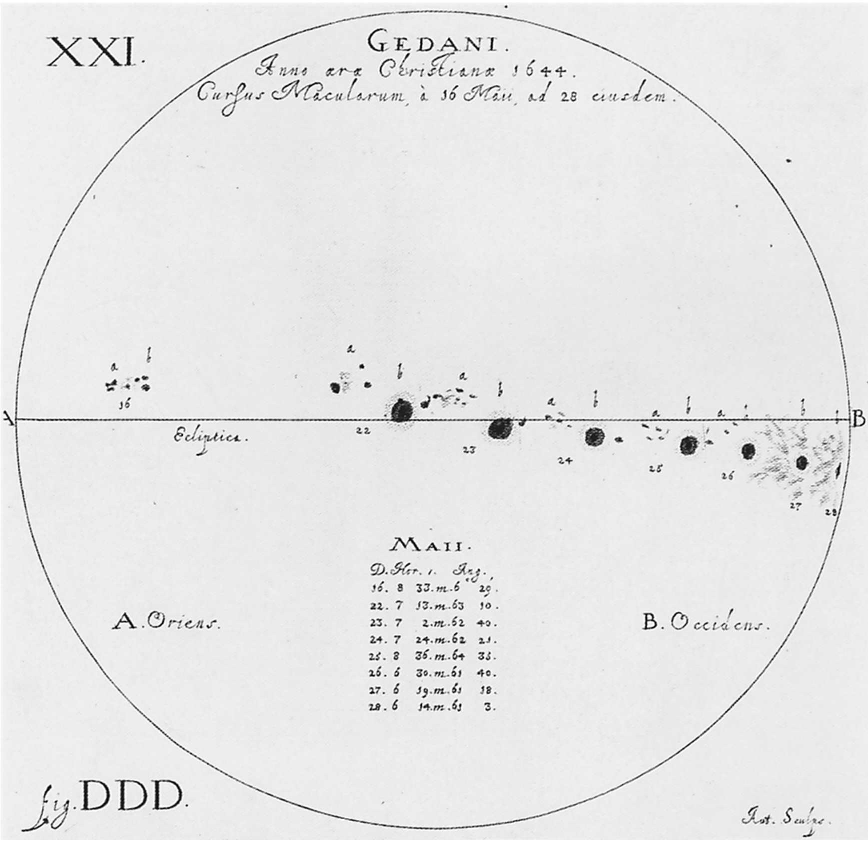

Fig. 1-2 Sunspots and faculae observed by J. Hevelius on May 16–28, 1644. The position of the large spot is shown on successive days as solar rotation moves it from east to west across the disk. Faculae (hatched) are seen only near the limb. From the *Selenografia*. By permission of the Houghton Library, Harvard University.

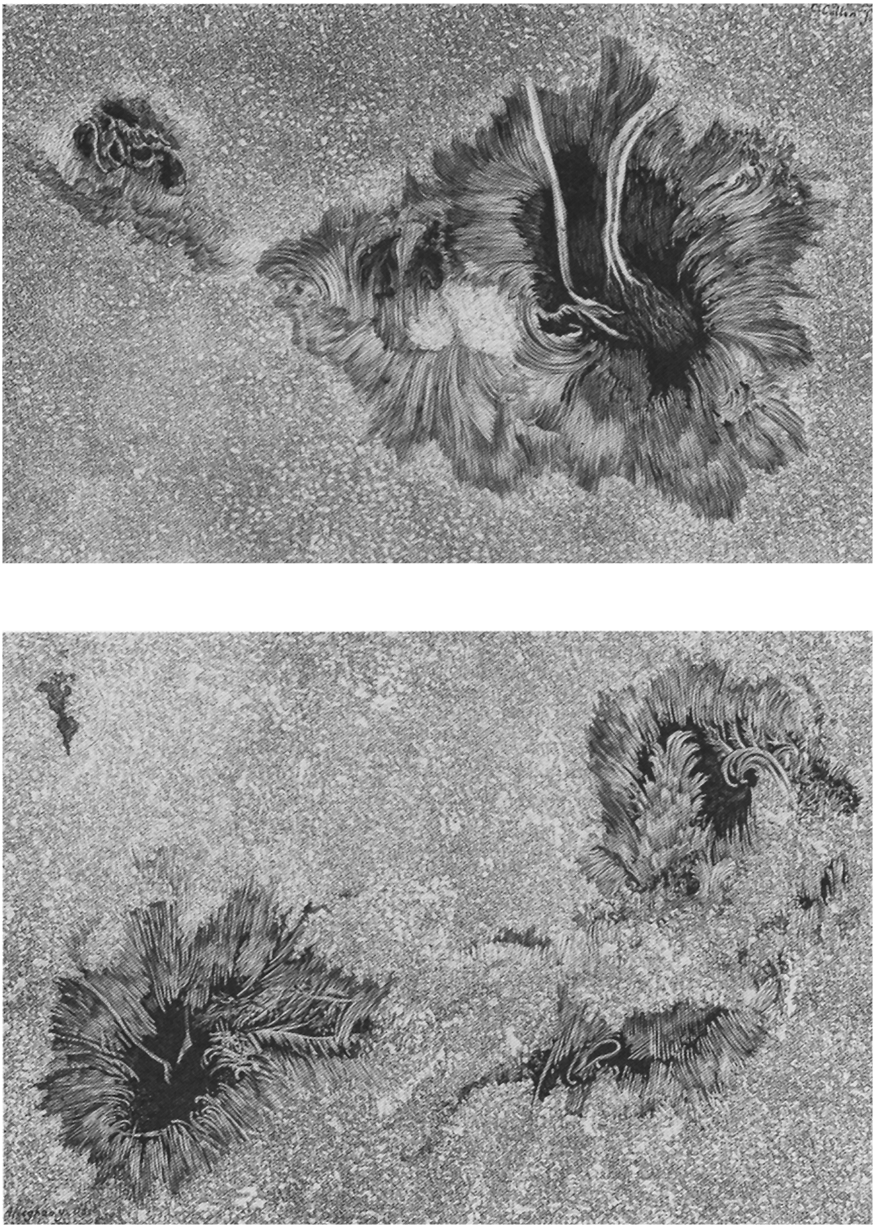

Fig. 1-3 Drawings of sunspots made by S. P. Langley on September 21, 1870 (top) and March 5, 1873 (bottom), using the 13-inch refractor at Allegheny Observatory. Note the light "bridge" across the umbra and the fine penumbral detail. Granulation can be seen in the photosphere outside the spots. From S. P. Langley, *The New Astronomy*, 1888.

The development of more compact and achromatic refractors in the 19th century eventually made possible visual observations of greatly increased resolution, such as the superb sunspot drawings by S. Langley shown in Fig. 1-3. But the records compiled by 17th-century observers continue even now to provide new insights into the sun's behavior. For instance, we find that spots became rare after about 1645 and remained so until around 1715. This broad minimum of sunspot occurrence was rediscovered by E. W. Maunder in the late 19th century and has since been named after him. Since the processes responsible for sunspot generation and for the solar differential rotation may be related, efforts have also been made to reconstruct the solar rotation rate and its latitude profile from the rotation of spots observed immediately before 1645 and after 1715, to check for a possible change during the Maunder Minimum. The data used included Hevelius' drawings, such as Fig. 1-2. More than 300 years after their discovery, the mechanisms responsible for the generation of sunspots, the relation of spots to faculae, and the differential rotation of the sun's surface continue to present problems at the forefront of solar research.

Toward the end of the 18th century, advances in stellar astronomy associated in large part with the work of William Herschel gradually led to the view that the sun is a star similar to the thousands of bright points seen in the night sky, only located much closer to the Earth. How much closer became clear in 1837 with the reliable measurement of the stellar parallax of 61 Cygni by F. Bessel. Study of the sun's physical structure and chemical composition as a valuable key to understanding the stars provided a powerful impetus to solar research in the 19th century that continues to the present day.

1.2 The Spectroscope and Photography

The spectroscope was the new instrument that enabled astronomers to carry the idea of the sun as a star to useful conclusions. After initial studies on the dispersion of solar light by Newton and others in the 17th and 18th centuries, J. Fraunhofer built the first spectroscope useful for quantitative analysis in 1814 and used it to measure the positions of over 500 of the remarkable dark lines seen in the solar spectrum. Fraunhofer's original spectrum is illustrated in Fig. 1-4. A certain amount of experimental work on the spectra of the sun, stars, and incandescent laboratory sources ensued. But it required the insight of G. Kirchhoff almost 50 years later to discover the simple laws governing the emission and absorption of light from solid and gaseous bodies.

The most important impact of Kirchhoff's work on solar research lay in his demonstration that the thousands of dark absorption lines observed in the spectrum of the solar disk have, in general, a one-to-one correspondence to bright emission lines in incandescent vapors observed spectroscopically in the laboratory. From these foundations, Kirchhoff and others showed that the sun's surface consisted of hot gases made up of elements found on Earth. This discovery surely ranks with the most far-reaching findings of natural science, since it gave valuable support to the

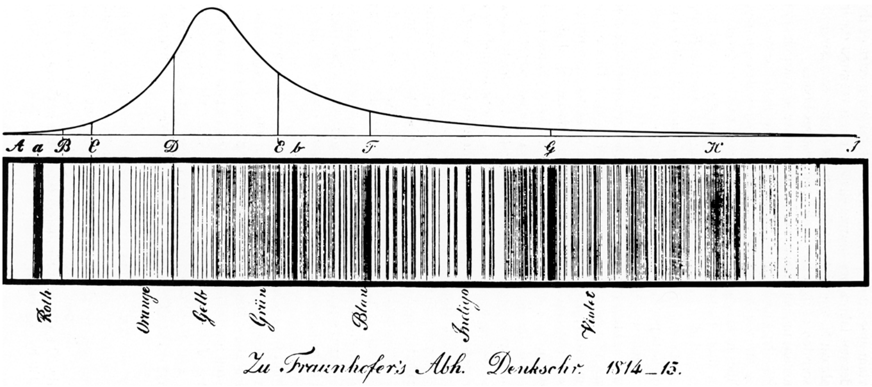

Fig. 1-4 Fraunhofer's original drawing of the solar spectrum from 1814. Red is to the left, violet to the right. The H-line of Ca II, the b-lines of Mg I, and D-line of Na I correspond to modern nomenclature. From J. N. Lockyer, *Solar Physics*, 1874.

rather daring cosmological principle that all corners of the universe obey the same laws of nature.

The development of photography in the 1840s made possible more objective recording of structure seen on the sun's disk in white light. One of the first results was to demonstrate that the darkening of the solar disk toward the limb claimed from visual observations was real. This limb darkening of the white-light-emitting layer in the sun's atmosphere (later called the photosphere) was eventually explained by K. Schwarzschild, in a classic paper in 1906. Schwarzschild used newly developed ideas of radiative transfer to show that the observed limb darkening by a factor of roughly 2.5 can be understood in terms of the decreasing temperature outward through the photosphere if energy transport through this layer is by radiation.

By the 1870s, photographs such as Fig. 1-5 were obtained of the pattern of small mottles, called granulation, that cover the photosphere. The bright roundish granules (also seen in the drawings of Fig. 1-3) are typically a few arc seconds in diameter and are separated by narrower dark intergranular lanes. It was later recognized that these granules resemble the pattern of convective cells found in H. Bénard's experiments of 1901 on fluids heated from below, so that they probably represent hot rising gas elements convecting heat to the sun's surface.

This association of granules with convection is generally accepted on the basis of modern observations and simulations, but the validity of the early evidence discussed above has been questioned. The remarkable regularity of the granulation seen in Fig. 1-5 seems to arise at least in part from a recently discovered pattern of fine cracks in the emulsion. The cells seen in Bénard's experiments are considered to have been caused by surface tension effects rather than by convection. This ironic episode shows how compensating errors in observations and interpretation can underlie a correct result.

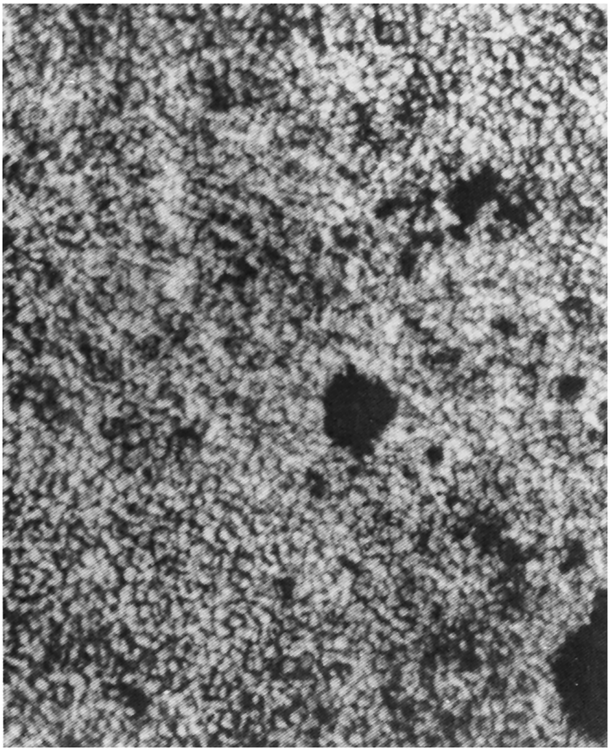

Fig. 1-5 A photograph of the photosphere, showing some spots and pores and perhaps granulation, taken in Paris by J. Janssen on July 5, 1885.

It appears that, although most of the energy is carried by radiation through the photosphere, the granulation seems to represent the "overshooting" of convective elements from the deeper, opaque layers of the sun where convection carries most of the heat flux. The structure of a deep convection zone, extending roughly 30 % of the sun's radius below the photosphere, was first worked out in 1930 by A. Unsöld.

1.3 Solar-Terrestial Research and the New Astronomy

A different line of inquiry that had great impact on 19th-century solar research, and continues to provide important motivation today, was initiated by the discovery of a significant 11-year periodic variation in the number of sunspots by H. Schwabe in 1843. Soon afterward, R. Wolf and E. Sabine connected these results to the modulation in geomagnetic storm occurrence discovered earlier by J. Lamont and by Sabine in research on magnetic disturbances in Germany and in Canada. This was the first good evidence for a variable solar influence on Earth, and the beginning of solar-terrestrial research.

More insight into the specific aspect of solar activity responsible for geomagnetic disturbances came after the first observations of great solar eruptions within spot groups, reported in white-light observations of the photosphere in 1859 by R. Carrington and R. Hodgson. The occurrence of a great geomagnetic storm within less than a day suggested that these "flares" on the sun, rather than the sunspots themselves, were directly responsible for the terrestrial disruptions. The impressive correlations found thereafter between geomagnetic storms and disk passage of large spot groups gradually led to our relatively recent acceptance of the view that emissions of charged particles travel from the sun to Earth.

Observations during total solar eclipses provided the only method to study the solar atmosphere outside the bright photospheric disk until the 1860s. The ability to predict eclipses dates back to Babylonian times, and improvements in technique introduced by the great astronomer Hipparchus of Rhodes in the second century B.C. increased accuracy to a few hours. The extended white-light corona had been noted in the early 18th century, along with naked-eye sightings of the pinkish coronal condensations that we now call prominences. Eclipse watchers since the 1706 event had also remarked on the red light of the narrow layer called the chromosphere, visible briefly just as the moon covers and uncovers the photosphere at beginning and end of totality.

During the American Revolutionary War hostilities were suspended in a part of the state of Maine for one day to permit the Reverend Prof. Williams and his collaborators from Harvard College to observe the eclipse of 1780. This expedition was reportedly the first to note the phenomenon of Bailey's beads, which was rediscovered at the 1836 eclipse by the English amateur astronomer F. Bailey, and is caused by solar light shining through the lunar valleys at the edge of the moon's disk.

But until the event of 1836, few physical observations of the sun were made, although the times of eclipse contacts were systematically recorded. At eclipses beginning in 1842, some of the foremost astronomers of the day set up experiments intended to determine whether the corona, prominences, and chromosphere were solar or whether they originated in a (then – plausible) lunar atmosphere. Good photographs obtained at two sites 250 miles apart during the 1860 eclipse established that prominences were indeed some kind of cloud or condensation in the sun's atmosphere.

Kirchhoff's work on the interpretation of spectra encouraged spectroscopy at the 1868 eclipse. Here J. Janssen observed the yellow (D_3) line λ 5876 in prominences, later found to be radiated by neutral helium, an element not discovered on Earth until 1895 by the chemist W. Ramsay. The green coronal emission line at λ 5303 was also first observed during the eclipse of 1869. This was the first of 24 coronal emissions to be discovered in the visible spectrum which could not be identified with any known element, and for many decades posed a vexing problem for spectroscopy. The explanation came in 1941, when the spectroscopist B. Edlén, proceeding from earlier evidence put forward by W. Grotrian, showed how these lines ascribed to the mystery element "coronium" actually arose from forbidden transitions between low-lying fine structure levels of highly ionized heavy atoms such as Fe and Ca. In 1871, Janssen identified greatly broadened Fraunhofer absorption lines in the coronal continuum spectrum and showed that they originated in scattering of the photospheric Fraunhofer lines by coronal particles. He thus established that the white-light corona was solar in nature. Advances in understanding such as these, achieved through the application of spectroscopy to the sun and stars, came to be recognized as the New Astronomy, or astrophysics as we now call it.

The 1868 eclipse also led to the first use of monochromatic imaging of solar features, through the invention of the "spectrohelioscope" to observe prominences. Most of the emission from prominences occurs in a few narrow spectral lines, so the principle of this instrument was to use a prism spectroscope to form an image in a narrow spectral passband including one of these lines. Most of the continuum light of the solar disk scattered in the Earth's atmosphere was thus excluded. Using this technique, J. Janssen in France, and soon thereafter, J. Lockyer in England, were the first to observe prominences and also the chromosphere, outside of eclipse. Drawings of prominences and of the chromosphere made by W. Huggins and A. Secchi using spectrohelioscopes in the 1870s revealed structural details such as the chromospheric jets (seen at the limb in Fig. 1-6) that were rediscovered after 1945 and named spicules. These visual observers also documented the intricate shapes of prominences, such as the loops illustrated in Fig. 1-6, and their often rapid motions. Around this time, comparative study of the eclipsed corona near sunspot cycle maximum and near the minimum (see Fig. 1-7) first revealed the differences in morphology characteristic of these periods of high and low activity. The intricate loops and plumes visible near maximum activity led to the first suspicion that electromagnetic forces play a role in the dynamics of the solar atmosphere.

At the eclipse of 1870, C. Young observed the "flash" spectrum of the chromosphere, which blazes out for a few seconds immediately before and after totality. In 1913, S. Mitchell showed that this spectrum could not originate in scattering of the Fraunhofer lines since it contains high excitation lines weak or absent in the photospheric spectrum. His reasoning also led to the first suggestion that the chromospheric temperature might exceed that of the photosphere. The identification in 1941 of the "coronium" lines with forbidden transitions in Fe X, Fe XIV, Ca XV, and other highly stripped ions proved that the temperature of coronal gases lying above the chromosphere was very high, over one million degrees. These high temperatures continue to stimulate research into the processes of nonthermal heating in the chromosphere and corona. Dissipation of waves or electric currents seems to be required since thermal heating of these layers by conduction, convection, or radiation from the cooler photosphere would violate the second law of thermodynamics.

The structure of the chromosphere on the disk was first revealed by observations with the spectroheliograph, developed independently in 1891 by H. Deslandres and by G. Hale. This instrument combined the narrow passband of the spectrohelioscope with the advantages of permanent and accurate recording made possible by the photographic plate. Its principle of operation (Fig. 1-8) relies on stepping the entrance slit of a spectrograph across the solar disk at exactly the same rate that the exit slit is moved across the photographic plate. A full image of the solar disk can

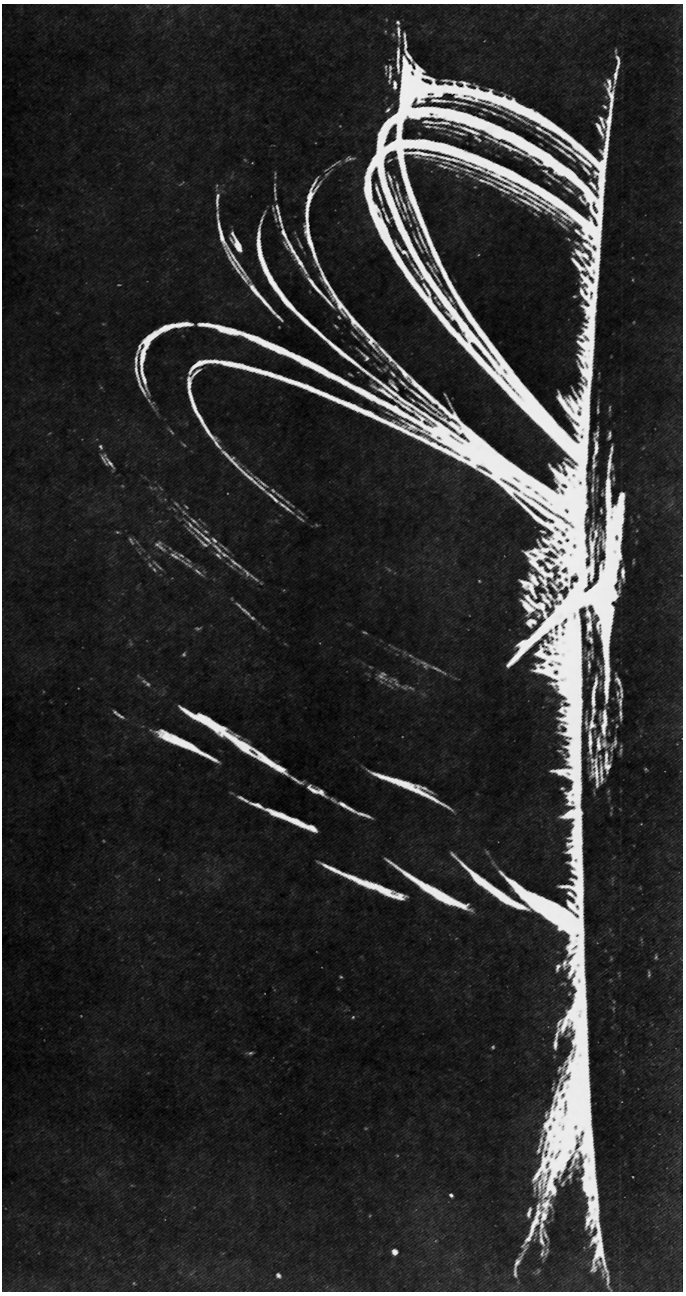

Fig. 1-6 Coronal loops drawn by A. Secchi from spectrohelioscope observations in the Balmer alpha line on October 5, 1871. The thin radial structures at the limb are chromospheric spicules. From C. Young, *The Sun*, 1895.

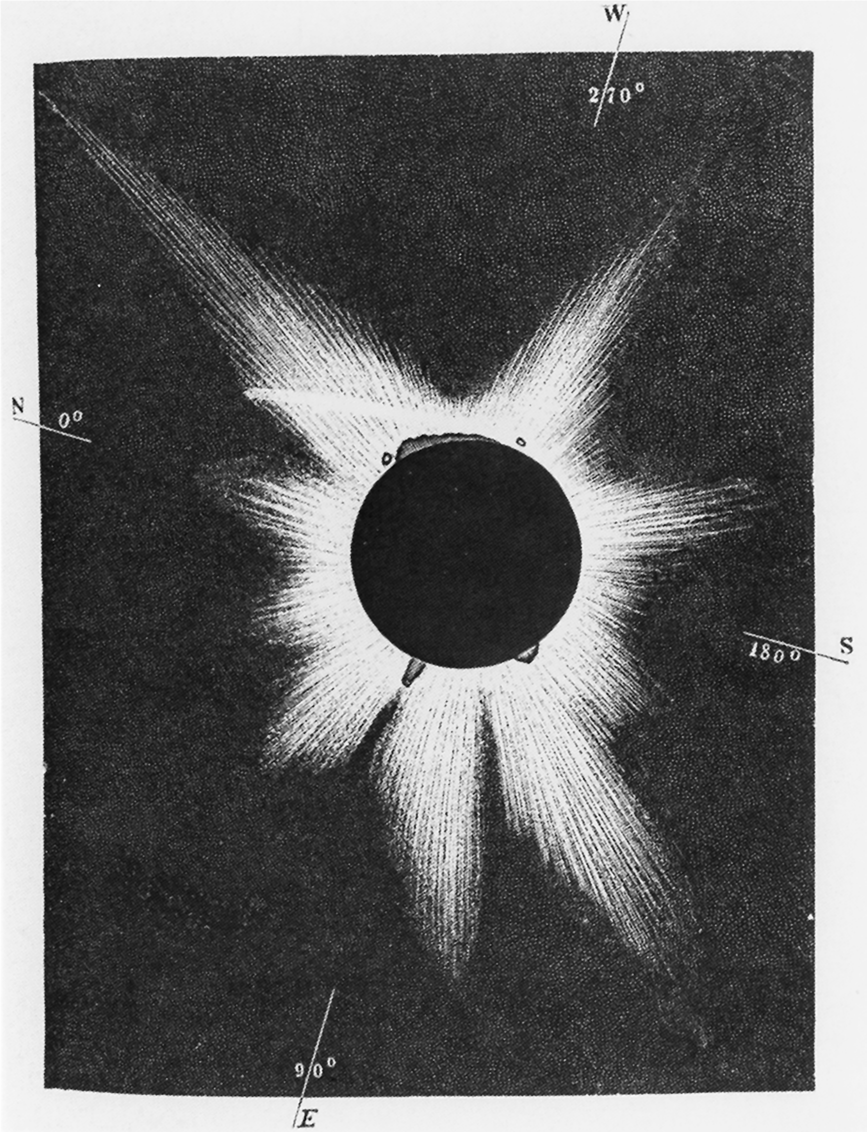

Fig. 1-7 A drawing of the corona of 1868 observed at Mantawa-lok-Kekee, Malaysia. This was the eclipse at which helium was first detected spectroscopically. The horizontal bright streak across the upper left coronal streamer is a flaw in the woodcut (not a comet). From J. N. Lockyer, *Solar Physics*, 1874.

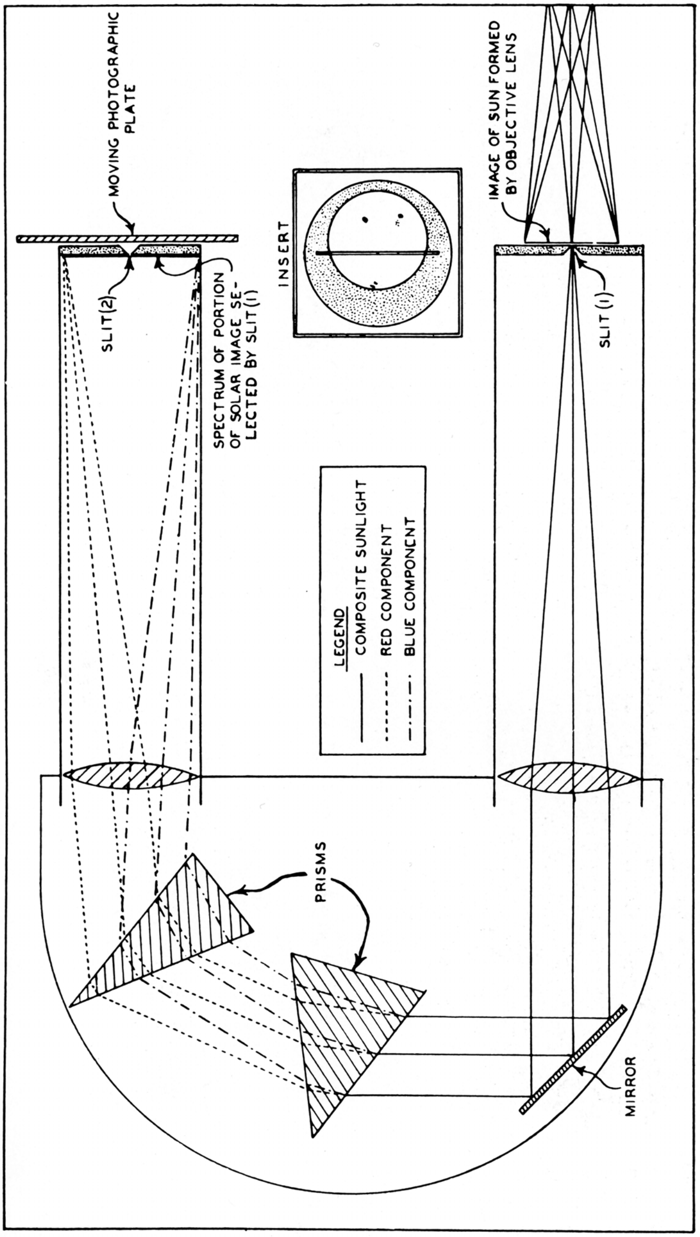

Fig. 1-8 The principle of the spectroheliograph. Slit 2 is adjusted in position and width to select a wavelength and pass-band width of interest. The photographic plate is then translated across slit 2 at exactly the same rate as the solar image is translated across slit 1. By permission of Mt. Wilson and Las Campanas Observatories, Carnegie Institution of Washington.

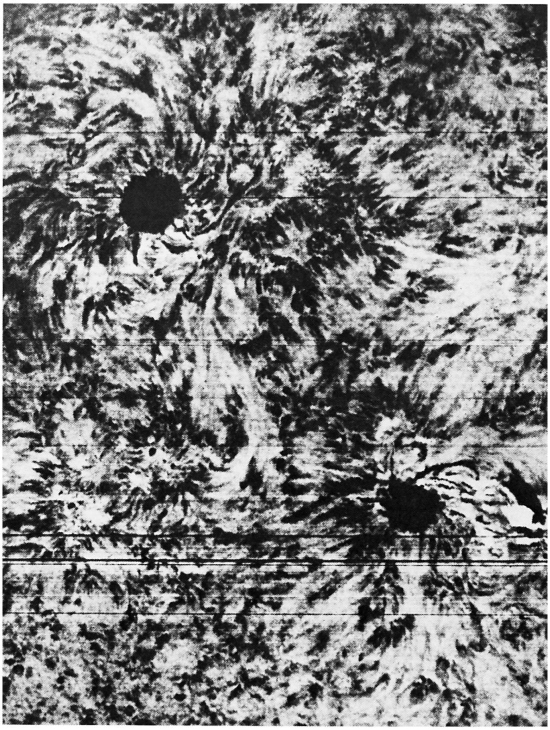

Fig. 1-9 The presence of intense magnetic fields in sunspots was suggested to Hale by vortex-like shapes in chromospheric structures similar to those shown in this Hα spectroheliogram taken on September 9, 1908. By permission of Mt. Wilson and Las Campanas Observatories, Carnegie Institution of Washington.

then be produced in any spectral line or continuum region by setting the grating angle so that the spectrograph forms an image of the entrance slit at the exit slit in the desired wavelength.

This instrument made it possible to photograph the spatial structure of the chromosphere on the disk in the resonance line of singly ionized calcium at 3934 Å, known as the Ca II K-line. Later, observations were obtained in the strongest Balmer absorption line Hα at 6563 Å when red-sensitized emulsions became available in 1908. Hale's spectroheliograms in these lines, one of which is illustrated in Fig. 1-9, revealed the intricate dark structures of chromospheric fibrils and mottles, which later work related to the spicules that are seen in emission at the limb in Fig. 1-6. Hale's spectroheliograms also showed the bright chromospheric plages seen when faculae are observed on the disk in the cores of the Hα and Ca K-lines. Hale and Deslandres were nominated jointly for the Nobel prize for their invention of this remarkable instrument.

1.4
Solar Chemical Composition and Energy Generation

Kirchhoff made the first solar identifications of absorption lines of sodium, iron, magnesium, and other heavy elements around 1860. The presence of hydrogen in the sun's composition was revealed about ten years later by A. Angstrom and others. H. Rowland's development of excellent diffraction gratings led to his publication in 1897 of a superb atlas of the solar spectrum between atmospheric cutoff around λ 2975 and the visual limit near λ 7331. From these spectroscopic data, he was able to increase the number of elements identified in the sun to 36 by the end of the century.

The major advance to quantitative analysis of the solar line spectrum came in the 1920s after the development of N. Bohr's atomic theory and of M. Saha's ionization equation. This progress in atomic and statistical physics enabled H. Russell in 1928 to use Rowland's eye-estimates of line intensities to establish the rough relative abundances of elements in the solar atmosphere. The surprising result that the sun consisted mainly of hydrogen, first noted by C. Payne, was by no means immediately accepted.

Around 1920, A. Eddington explained how hydrogen burning to helium might provide the energy to fuel the sun's luminosity. Building on earlier ideas, he suggested that the hydrogen fuel might be sufficient to account for the Earth's age implied by fossils. This age had been recognized by then to greatly exceed the $20 \cdot 10^6$ year time scale that H. van Helmholtz had estimated in 1854 for conversion of the sun's gravitational potential energy to radiation. Helmholtz's theory that a slow contraction of the sun provided the energy required to fuel its heat and light output remained the only reasonable explanation of the source of the sun's power output until about the turn of the century. The rate of steady contraction implied in the present epoch would be well below the detection limit for measurements of change in solar diameter.

Two sets of nuclear fusion reactions, called the carbon-nitrogen cycle and the proton-proton chain, were put forward in 1938 by H. Bethe, C. Critchfield, and C. von Weizsäcker to specify how the nuclear burning might proceed. The first direct test of these mechanisms in the sun was provided by the neutrino experiment of R. Davis, located deep in the Homestake gold mine in South Dakota. This experiment began operation in 1969 and detected a significantly lower neutrino flux than was predicted by standard solar models. The discrepancy is now known to result from then-unknown neutrino properties discussed in Chapter 6, and the solar models appear to be correct.

The success of nuclear fusion in explaining the luminosity of stars rested until this finding, on the good agreement between the observed and calculated distribution of stars in the Hertzsprung-Russell plot of stellar luminosity against color. Impressive regularities found in the relative solar abundances of elements heavier than helium are also well explained through the theory of stellar evolution put forward by M. Burbidge, G. Burbidge, W. Fowler, and F. Hoyle in 1957. Their explanation is based on successive fusion episodes, which build up the elements lighter than iron.

The sun's enormous power output was demonstrated by the first good calorimeter measurements of the total solar irradiance or "solar constant" obtained in 1837 by S. Pouillet in France and by J. Herschel working in South Africa. Pouillet's radiometer is shown in Fig. 1-10. Measurements of the spectral distribution of solar power were advanced by S. Langley's invention of the thermoconducting bolometer around 1881. With this sensitive detector, Langley mapped the sun's spectrum to 5.3 μm. He also showed the importance of molecules in the Earth's atmosphere, such as H_2O, CO_2, and CO, in the seasonally variable terrestrial absorption of solar infrared light.

C. Abbot, working with Langley at the Smithsonian Institution, put into use better radiometers for the solar constant measurements. Together with his collaborators, he then carried out an epic program of daily solar constant observations between 1923 and 1952 at a worldwide chain of Smithsonian mountain stations. Abbot's measurements demonstrated that the solar radiation transmitted by the Earth's atmosphere was constant to better than 1 % over several decades. But the absolute value of the solar constant remained uncertain at the 5 % level until the sun's ultraviolet flux below 3000 Å was properly measured by rockets in the 1950s.

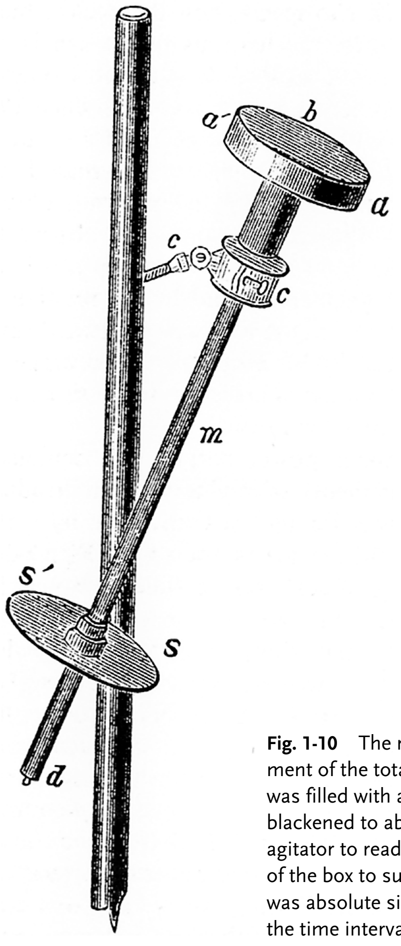

Fig. 1-10 The radiometer used by S. Pouillet in 1837 for measurement of the total solar irradiance. The hollow metal box marked *a-a′* was filled with a known volume of water. The upper surface *b* was blackened to absorb sunlight. The shaft *m* contained a thermometer-agitator to read the water temperature increase caused by exposure of the box to sunlight for a known interval of time. The measurement was absolute since the volume of the water, the blackened area, and the time interval were known.

Modern radiometry from space now provides the *absolute* value of the total solar irradiance to about 0.3% accuracy. Daily measurements since 1978 have clearly shown the decreases of the sun's output caused by dark sunspots, and the increases produced by bright faculae. These remarkably reproducible measurements, first achieved by R. Willson at JPL, and by J. Hickey at the Eppley Laboratories, also showed that the sun is almost 0.1% brighter, not darker, near peak sunspot activity. The explanation for this unexpected behavior is discussed in Chapter 13.

1.5
The Mt. Wilson Era of Large Telescopes

Using first the horizontal Snow telescope, and then the 60-ft and 150-ft tower telescopes built at Mt. Wilson near Los Angeles in 1907 and 1912 (Fig. 1-11), Hale and his collaborators made the first solar observations using high-dispersion spectro-

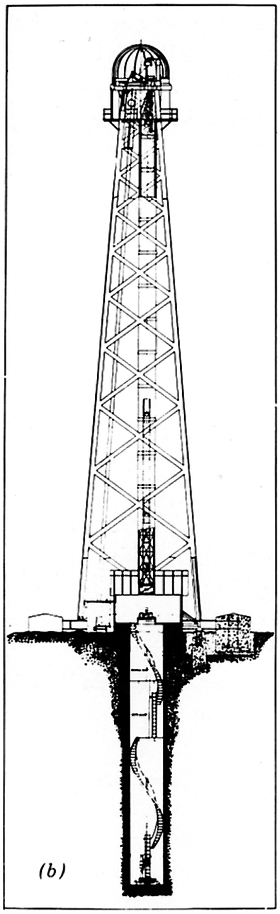

Fig. 1-11 (Left) The 150-ft tower telescope still in active use on Mt. Wilson and (right), cutaway view of the 150-ft solar tower. The deep pit below ground level contains the spectrograph, with its grating at the bottom. By permission of Mt. Wilson and Las Campanas Observatories, Carnegie Institution of Washington.

graphs and large image scale to achieve good spectral and spatial resolution on the sun. Following the example of J. Lockyer, E. Frankland, and other chemists, they compared their solar spectra with those they obtained of laboratory plasmas at various temperatures and pressures. The laboratory at Mt. Wilson, where these comparisons were carried out, is shown in Fig. 1-12. The most important results had to do with the physics of sunspots and of the solar magnetic field. Hale's work confirmed (by comparison of the umbral spectrum with arc and spark spectra in the laboratory) that the dark umbra is cooler than the brighter photosphere. This fact was less obvious than it seems now, since the theory of radiative transfer in gaseous atmospheres was only beginning to emerge from the work of A. Schuster, K. Schwarzschild, and others.

High-dispersion sunspot spectra also showed a puzzling widening and sometimes splitting of the lines in umbrae. Hale's Hα spectroheliograms in 1908 had shown vortex-like chromospheric dark filaments wound around the spots (see Fig. 1-9) resembling the configuration of iron filings around a magnet, which led him to suspect strong umbral magnetic fields. Proceeding from P. Zeeman's (1896) analysis of the magnetic field splitting of spectral lines, Hale obtained umbral spectra through a Nicol prism and found the expected zigzag pattern caused by the oppo-

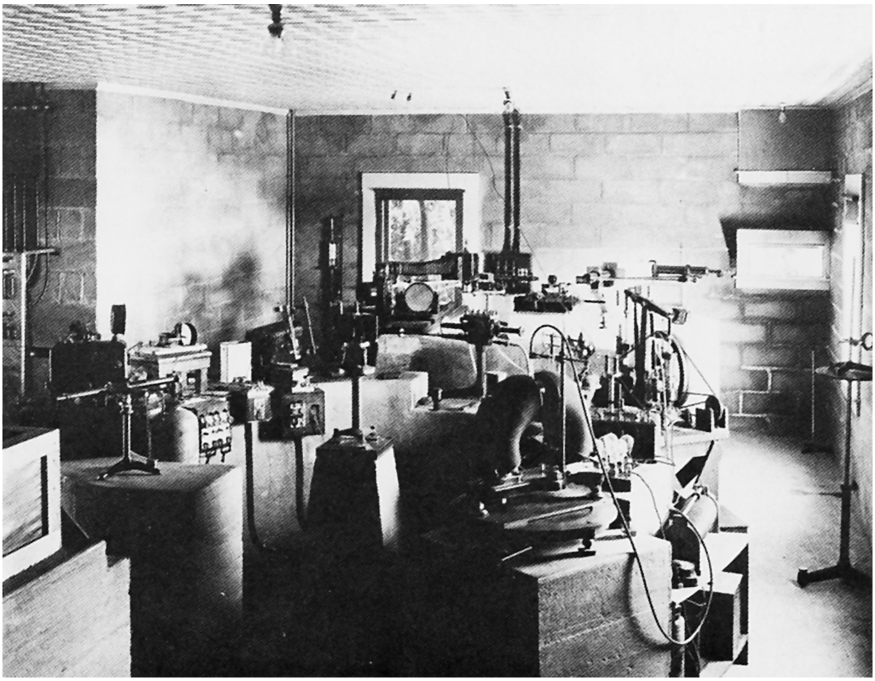

Fig. 1-12 Interior of early physical laboratory on Mt. Wilson, showing spectrograph and magnet used in studies of the Zeeman effect. By permission of Mt. Wilson and Las Campanas Observatories.

site circular polarization of the Zeeman-split line-wing components in sunspot fields of up to 3,500 G.

This first discovery of extraterrestrial magnetic fields set the stage for the systematic study by Hale and S. Nicholson of the laws of sunspot polarity and their changing behavior over the spot cycle. Their work showed that the oscillation of sunspot number recognized since 1843 constituted half of a 22-year cycle of sunspot magnetic polarities. The most obvious aspect of this 22-year cycle is the reversal of polarity of the east and west spots in active regions that accompanies the onset of a new 11-year cycle in spot number.

Hale's further evidence for a general polar magnetic field of intensity roughly 50 G with opposite sign at the north and south poles, was never fully accepted. Only after the development of the photoelectric magnetograph by H. W. Babcock in 1951 was a weak 1–5 G polar field reproducibly detected in spatial averages over the polar regions and observed to change in polarity every 11 years along with the spot fields.

The principle behind this very important instrument can be sketched as follows. Two exit slits of a high-dispersion spectrograph are placed in the red and blue wings of a magnetically sensitive photospheric absorption line, whose intensity profile is shown in Fig. 1-13. The solid and dashed $I(\lambda)$ curves denote the profiles from a magnetic region of the sun when left or right-hand circularly polarized light is alternately admitted to the spectrograph. Since the two wings of a normal Zeeman triplet broadened by a magnetic field are oppositely circularly polarized, admitting alternate senses of polarization by rotating a quarter-wave plate placed in series with a Nicol prism, moves the profile back and forth. The difference in intensity measured between the two photomultipliers placed in the wings changes from one profile to the other. This variation of the difference signal is amplified and recorded. Its amplitude is roughly proportional to the *net* magnetic flux (the effect of opposite magnetic polarities will tend to cancel) in the solar region whose light is admitted through the entrance slit of the spectrograph.

The magnetograph revealed for the first time, the distribution of photospheric magnetic fields outside of spots and faculae (Fig. 1-14). Observations with improved versions of this instrument by the 1960s showed a network pattern of magnetic fields of characteristic cell dimensions 20,000–40,000 km covering the quiet photosphere. This magnetic network was found to coincide well with the chromospheric network defined by the bright emission in Ca K spectroheliograms outside of active regions.

The suggestion of a velocity shift of spectral lines by C. Doppler in 1842 eventually led to the first measurements of the relative blue- and redshifts of Fraunhofer absorption lines at the sun's east and west limbs and thus to spectroscopic measurement of the photospheric gas rotation rate. These measurements (by N. Duner, H. Vogel, and C. Hastings between 1870 and 1890) showed a similar rotation rate and equatorial acceleration to the accurate sunspot rotation rates obtained by Carrington in 1865. But the plasma rates could be extended to latitudes ±75°, whereas spots could be used as tracers only over the active latitudes to roughly ±40°. Motions around sunspots also proved interesting. J. Evershed's 1909 observations in Kodaikanal, India, revealed a strong horizontal outflow extending several spot diameters

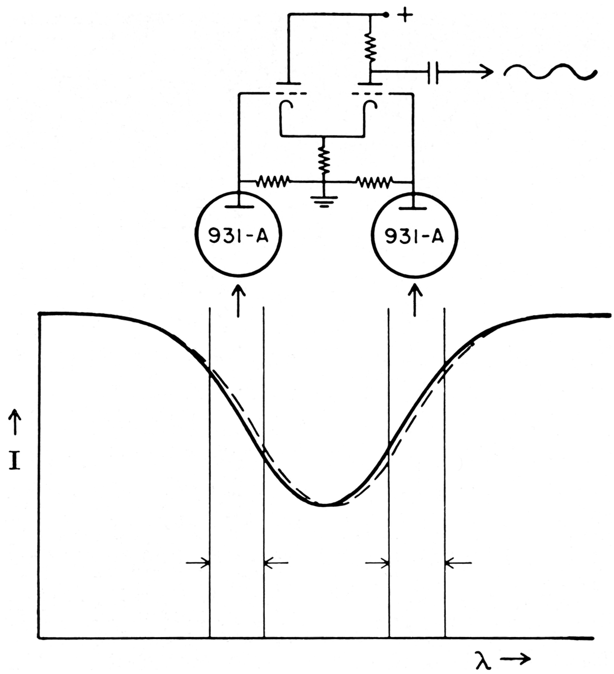

Fig. 1-13 Schematic diagram showing the principle of operation of the original Babcock magnetograph. Profiles of an absorption line for the two states of opposite circular polarization are shown solid and dashed. The symbols *I* and *λ* refer to intensity and wavelength respectively. The two detectors are type 931-A photomultipliers. By permission of the University of Chicago Press.

from the umbral edge. This Evershed effect is only now being incorporated into a satisfactory dynamical theory of spot coolness and stability.

The improved accuracy of line wavelength determinations achieved by upgrading the Rowland atlas in 1928 to use the cadmium red line interferometric wavelength standard led to increased awareness of the limb redshift of Fraunhofer lines relative to their disk center position. The nature of this shift, noted earlier by Hale and his collaborators at Mt. Wilson, generated lively debate. A small redshift relative to the laboratory wavelength, of constant magnitude across the disk, was expected from general relativity. But the conspicuous center-to-limb redshifts of up to 0.5 km s^{-1} required a different explanation. Only in the 1950s was a viable mechanism put forward in terms of the net Doppler blueshift near disk center caused by the bright upflowing granules, relative to the smaller contribution of dark downflowing material of intergranule lanes.

In the late 1950s, R. Leighton devised a technique for studying periodic velocity signals on the sun, using the spectroheliograph at the 60-ft Mt. Wilson tower. The slit was placed in the wing of a line, and the spectroheliograph was scanned across

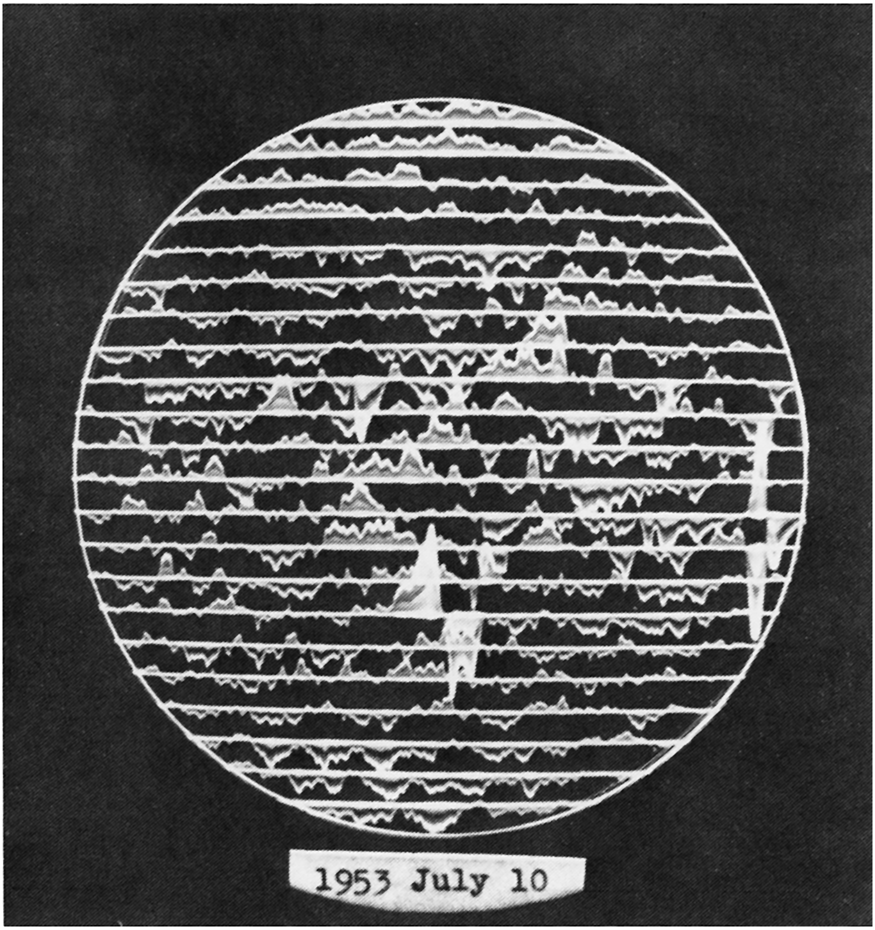

Fig. 1-14 An early photoelectric magnetogram obtained with the Babcock magnetograph. Each trace was made by a separate scan across the sun. A deflection equal to the distance between traces is produced by a field (averaged over the spectrograph entrance slit) of about 10 G. By permission of the University of Chicago Press.

the sun and back taking P minutes in each direction. A negative of the scan in one direction was then overlaid in perfect spatial registration on a transparent positive of the scan in the opposite direction. Points on the sun that happened to be Doppler shifted in the same sense on both the negative and positive were registered strongly as dark or bright on the resultant print, whereas points whose Doppler shift changed sign during the time, cancelled to a neutral gray in the Doppler summation print. The variable time delay across the summed print was ideally suited to detecting oscillations of unknown period in the solar velocity field over the range of periods shorter than the maximum time delay $2P$.

Using this technique, Leighton and his graduate students, R. Noyes and G. Simon, detected and studied the 5-min oscillation of the photosphere. The fruitful interpretation of the oscillation as standing acoustic waves trapped in resonant cavities below the photosphere was made in the early 1970s, when the observations of F. Deubner and the calculations of R. Ulrich independently showed the rich mode structure of these oscillations when their oscillatory power is plotted in the plane of spatial wavenumber k versus temporal frequency ω. The position of the modes in the k–ω plane, their width, and their splitting yield valuable information on the structure of the sun's interior, on the dynamics of wave damping by turbulent solar convection, and on the sun's internal rotation profile.

The Doppler cancellation techniques developed by Leighton also showed that the network magnetic fields coincided roughly with the edges of an outflowing horizontal velocity field of roughly 0.5 km s^{-1} centered within each of the cells. These 20,000–40,000 km diameter velocity cells were called supergranular convection. The strong magnetic fields measured at their boundaries in magnetograms were thought to be intensified from weaker fields pushed to the edges by the ram pressure of the convergent supergranule flows. More recent magnetograms obtained with higher spatial resolution confirm the indirect findings of J. Stenflo that the network fields are of order 10^3 G, thus similar in intensity to those in a spot.

1.6 Advances in Coronal Physics and in the Theory of Solar Activity

Instruments used for coronal studies did not progress much for 50 years after the introduction of the spectroheliosope at the 1868 eclipse. Only a few minutes per year were available to study the corona outside of the brightest prominences (which were visible outside of eclipse through a spectrohelioscope) using portable equipment set up at remote sites. The idea of blocking out the photospheric disk at the image plane of a telescope had occurred to many, but the difficulties of overcoming scattered light were not solved until B. Lyot implemented some ingenious precautions and built the first working coronagraph at the Pic du Midi, in France, in 1930.

The principle of the coronagraph is illustrated in Figure 1-15. The photospheric disk is imaged on a convex occulting disk at the prime focus of a refracting telescope and reflected out of the beam. One of Lyot's important advances was in placing a field lens (D) behind the occulting disk (C) to image the primary objective lens (B) onto a circular diaphragm (E). This diaphragm was made somewhat smaller than the objective lens image, so that rays originating at the edge (A) of the lens were blocked. Lyot had found that diffraction from the edge of the lens contributed a large fraction of the total scattered light.

Through this arrangement only coronal light gathered by the central part of the objective was used to form the final image of the corona. To reduce scattering in the lens glass itself to a minimum, Lyot used a single lens of bubble-free glass instead of an achromatic doublet, which causes multiple internal reflections from the glass-air surfaces (unless modern antireflection coatings unavailable in 1930 are used).

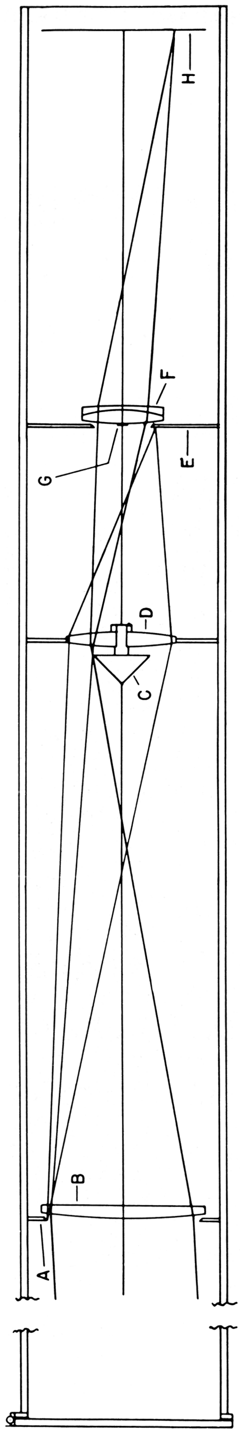

Fig. 1-15 Optical parts of a Lyot coronagraph. See text for explanation. By permission of the University of Chicago Press.

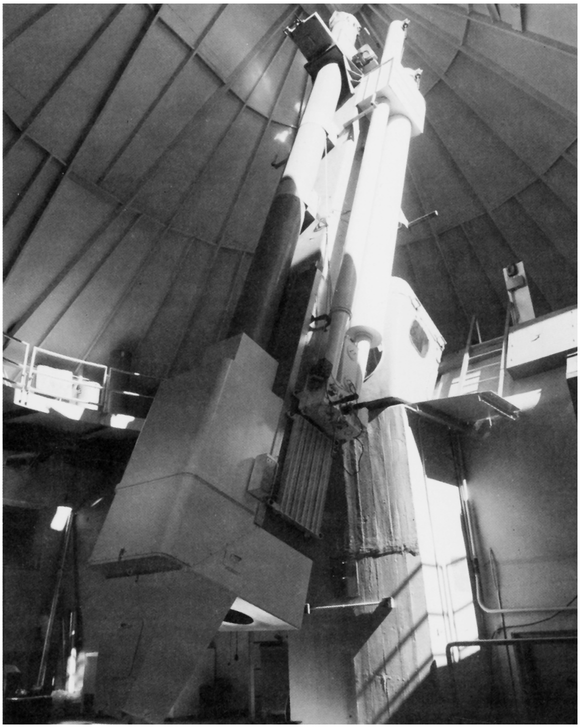

Fig. 1-16 The 16-inch aperture coronagraph at Sacramento Peak Observatory in New Mexico. This instrument is in active use for coronal studies and other observations requiring low scattered light. National Solar Observatories photograph.

Finally he is reported to have taken care to clean the objective with oil from the tip of his nose, applied with a well-laundered handkerchief. A large coronagraph is shown in Fig. 1-16.

Lyot also co-invented the birefringent filter (with Y. Öhman of Sweden), which enabled him to take monochromatic pictures of coronal and chromospheric phenomena far more rapidly than could be done with a spectroheliograph. The principle of the filter relies on the interference produced between the ordinary and extraordinary rays passing through a birefringent crystal such as calcite. Since the refractive index of the crystal differs for the two rays, they emerge with a phase difference that depends (for a given calcite thickness) on the wavelength of the light. Transmission is minimal for wavelengths at which the block of calcite yields a phase difference of π. A practical filter consists of several calcite elements, the thinner elements remove unwanted transmission sidebands at wavelengths other than that selected. The birefringent filter can be built to produce narrow passbands of 0.25 Å or less, with a few percent transmission, uniform over a field of view of roughly 1°. Most high-resolution photographs and time-lapse films of the sun in narrow passbands have been made using variants on this type of filter.

The results of these instrumental advances were rewarding; Lyot, R. McMath, and others made time-lapse movies of coronal structures at the limb in the green and red coronal lines λ 5303, λ 6374, or in Hα. These films followed the slow changes in the intricate shapes of quiescent prominences, and the rapid changes in active prominences near spots, including the predominant downflow of cool material from the corona called coronal rain. They also captured the occasional magnificent brightening of the chromosphere in an active region during the eruption of a solar flare. Sometimes the films showed the outward acceleration of material in flare-related surges (where the ascending material is later seen to return downwards along the same path) and sprays (where its return is not observed). Then the flare brightening expands upward, eventually forming intricate arches of postflare loops similar to those seen in Fig. 1-6. Very rarely, a flare is so bright as to be seen in white light, as was the case of the first flare ever reported in 1859.

The energy released in flares, and the intricate structure of prominences, represent two aspects of the control that electromagnetic forces exert over the solar plasma. Important advances in plasma electrodynamics were achieved through the work of H. Alfvén, T. Cowling, L. Biermann, T. Gold, and others in the 1940s and 1950s. Alfvén showed in 1942 that, under the conditions of reasonably good electrical conductivity and enormous scale commonly encountered in cosmic plasmas, their self-induction is exceedingly high and the motions of magnetic field lines and plasma particles can be usefully visualized as “frozen” together. That is, the plasma can move along the field lines but not perpendicular to them. This makes it easier to visualize, for example, how horizontal magnetic fields might support the cool dense material of prominences against gravity.

The energy density of the field measured in spots and plages was known to be amply sufficient to provide the energy for flares. But here the self-induction of the plasma intrudes, since it limits the rate of electric current discharge. The mechanism of energy release in flares is still not clear, but present evidence indicates that the resistive dissipation of electric currents into heat takes place in very thin layers, perhaps a kilometer or less. The self-induction is limited here by the diminished thickness, and intense compression may increase the plasma resistivity through creation of turbulence. Alfvén also showed in 1942 that plasma motions perpendicular to the field could give rise to transverse oscillations whose restoring force is the tension in the magnetic field lines. The propagation and dissipation of Alfvén waves combined in various ways with longitudinal pressure waves have been widely studied in attempts to explain how the high temperature of the chromosphere and corona are maintained.

Using these new concepts of magnetohydrodynamics, and regularities he observed in solar magnetic field behavior, H. W. Babcock put forward in 1961 a highly successful phenomenological model of the sunspot cycle. In this model the differential rotation (in both latitude and depth) of the sun’s convection zone provides the power to wind up (and thus intensify) a weak, initially poloidal field. This intensification leads to the 11-year oscillation in sunspot number and also produces the global change in direction of the field.

The physical processes behind Babcock's model were based on new ideas in dynamo theory suggested by Cowling and others. E. Parker's calculations in 1955 on magnetic buoyancy, which tends to make fields appear at the photosphere, and field helicity, which is necessary to explain the field reversal process, put these ideas on a firmer basis. Leighton added to the model in 1964 by showing how the evolution of the supergranular velocity cells he discovered could cause the magnetic fields to carry out a random walk at the photosphere, thus leading to a general dispersion of the field producing a relaxation of the oscillator.

A regenerative dynamo mechanism is not required to explain the existence of a solar magnetic field. Unlike the Earth, the sun is a large enough conducting sphere that its self-induction would prevent the ohmic decay of any primordial field that might have been present at its birth 4.5 billion years ago. But some kind of oscillator is required to understand the regular 11-year and 22-year variability of the solar field. To produce more detailed models, a better understanding of solar convection and its interaction with magnetic fields is required, and numerical simulations are beginning to provide useful insights.

1.7 Observations at Radio, Ultraviolet, and X-Ray Wavelengths

The sun's emission at radio frequencies was first identified in 1942 by J. Hey in the United Kingdom and by J. Southworth in the United States using World War II radar receivers. For reasons of military secrecy, their results were not published until after 1945. Rapid progress was then made in categorizing solar radio signals both spatially and spectrally in Australia, the United Kingdom, and the United States. The thermal continuum background at centimeter and decimeter wavelengths was found to be slowly varying with the rotation and evolution of plages, and various types of bursts were classified, together with their relation to flares and other disturbances of the active sun. Meter wave emissions from nonthermal sources were studied using rapid sweeping in frequency. Observations in meter waves could track plasma ejections from flares and eruptive prominences as they moved outward through the coronal plasma. Centimeter wave observations provided some of the first data used to estimate the thickness of the transition region, the height interval above the ten-thousand-degree chromosphere in which plasma temperatures rise to the million-degree values of the corona.

The low spatial resolution achievable with single radio dishes because of diffraction, and the relatively wide range of plasma temperatures contributing to the radio emission at a given frequency, led solar astronomers to seek other techniques in the ultraviolet and X-ray spectral regions for more detailed and easily interpreted data on the corona and transition region. Recent advances in the use of radio interferometry, and in aperture synthesis (utilizing many dishes electronically locked together), have changed the situation dramatically. The highest spatial resolution imaging (approximately 0.1 arc sec or 70 km on the sun) could, in principle, be achieved in centimeter waves using aperture synthesis at the Very Large Array (VLA). Unfortunately, it is difficult to achieve coverage of extended solar features in reasonable times at these spatial resolutions.

The sun's ultraviolet radiation down to roughly 3000 Å was photographed in 1840 by A. Becquerel. A rough idea of the sun's emissions at ultraviolet wavelengths below the atmospheric cut-off imposed by ozone (O_3) and molecular oxygen in the Earth's atmosphere was inferred from the results of ionospheric studies in the 1920s and 1930s. Some early detections of solar UV light were reported using balloons flown to roughly 20 km by K. Kiepenheuer and his colleagues in the 1930s. But the first solar UV spectra were obtained by workers at the Naval Research Laboratory (NRL) starting in 1946 using spectrometers flown on captured German V2 rockets. These spectra down to λ 2100 showed that the solar UV flux over this spectral range was significantly lower than would be expected from black-body extrapolation of the visible-light radiation curve to wavelengths below 3000 Å.

Subsequent rocket and satellite-borne spectrometers showed that at even lower UV wavelengths the flux observed far exceeded that expected from such a black-body curve and was increasingly variable over all time scales at wavelengths below approximately 2000 Å. It is now well established that most of this variation is caused by bright ultraviolet plage regions (also bright in the visible light of Ca K and Hα) moving across the disk as the sun rotates. Recent satellite-borne measurements also clearly reveal the 11-year ultraviolet flux variations at these wavelengths.

The ultraviolet region below 2000 Å contains the strongest lines of most of the abundant ions formed in the temperature transition between the chromosphere and corona. Since these lines are each formed in a restricted temperature regime within this 10^4–10^6 K temperature interval, they offer a unique opportunity to "unwrap" the physical structure of this interesting layer of the solar atmosphere. Images in strong ultraviolet emission lines at $\lambda < 1500$ Å show the intricate magnetic structures of the chromosphere, transition regions, and corona on the solar disk since the continuum emission of the photosphere is very weak at these wavelengths. The diagnostic techniques of ultraviolet spectroscopy were applied to studies of the plasma densities and temperatures using spectrometers flown on the *OSO* satellites, on *Skylab*, and most recently on the SOHO, TRACE and SDO spacecraft.

The first rockets bearing X-ray detectors were flown by NRL in 1949, and solar radiations were detected in the spectral region below 100 Å. Below roughly 300 Å optical imaging using normal incidence mirrors fails because of excessive absorption from the metal surfaces. X-ray telescopes using, instead, grazing incidence optics (Fig. 1-17) were first flown on rockets by a team at American Science and Engineering in the late 1960s. As in the ultraviolet, the photosphere and chromosphere are too cool to emit a significant X-ray flux, so such telescopes image the million-degree structures of the corona on the disk. This advance made possible study of the hottest coronal and flare emissions of $T > 5 \cdot 10^6$ K on the disk.

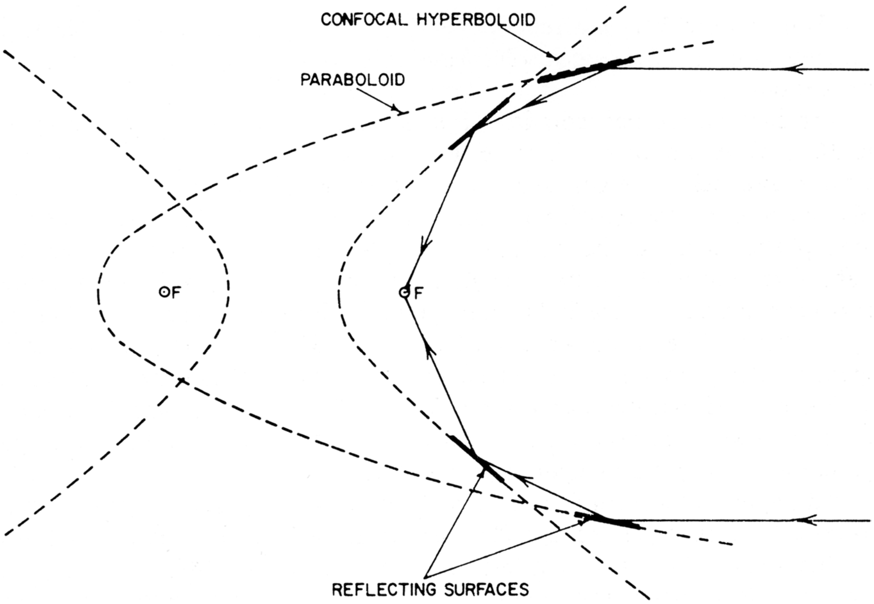

Fig. 1-17 The principle of a soft X-ray imaging telescope. Light entering from the right is reflected at grazing incidence from surfaces of confocal hyperboloid and paraboloid mirrors and is brought to the focus at F. Film is usually used as a detector. By permission of L. Golub.

1.8
The Solar Wind and Heliosphere

The X-ray and ultraviolet techniques for observing the solar corona and transition region on the disk led to the discovery, in the late 1960s, of extended, sharply demarcated areas of very low coronal emission, called coronal holes. For decades, correlations between geomagnetic disturbances and passage of active regions across the disk had indicated that seemingly quiet areas of the sun outside of active regions seemed to be responsible for many 27-day recurrent geomagnetic storms. These mysterious quiet areas were called M-regions. The good correlation found between disk passage of long-lived coronal holes and increased solar particle fluxes near Earth showed that these areas of low coronal plasma density are the long-sought M-regions responsible for a substantial fraction of the sun's output of particles and fields into the interplanetary medium.

The idea of intermittent particle flows from the sun was widely accepted by the turn of the 19th century from the evidence of a several-day time lag between solar flares and their geomagnetic disturbances. In the 1950s, the work of L. Biermann also showed that deflections of comet plasma tails could not be explained by solar radiation pressure, but required charged particles. The concept of a steady, rather than intermittent, outflow of coronal plasma was put forward already in the 1930s by J. Bartels and others, but it was predicted from a hydrodynamical model by Parker in 1958 on the theoretical grounds that solar gravity could not entirely contain the gases of the million-degree corona. This solar wind of roughly 400 km s^{-1} speed and 1–10 particles cm^{-3} density, whose existence was highly controversial for several years, was confirmed by *in situ* measurements from the *Mariner 2* space probe to Venus in 1962.

Sporadic (not recurrent) geomagnetic storms can often be linked to flares or to the eruption of filaments from the chromosphere. The disturbance travels as a shock wave at $\sim$ 500 km s^{-1} and produces particle and field enhancements at 1 AU within 2–3 days. The solar eruption is much more promptly heralded by ultraviolet and X-ray enhancements that affect the ionosphere within about 8 min of the flare occurrence, and within roughly an hour by the arrival (for the largest flares) of relativistic particles in the energy range 0.1–1 GeV. The production of these relativistic particles in some flares is not well understood. Recent observations of flare-produced γ-rays, and also primary neutrons, indicate nuclear processes occurring within the acceleration region. The first such observations were made in 1972 from the *OSO 7* spacecraft, using a sodium iodide scintillation counter.

This new evidence also sheds light on the variable isotope composition of helium nuclei detected at Earth from flares. Important information on how flare- or filament-produced disturbances move through the corona and interplanetary medium is being gathered from white-light coronagraphs flown on spacecraft. In these coronagraphs an additional external occulting disc is mounted well in front of the objective. Scattered light levels below 10^{-7} of disc brightness can be achieved in the absence of atmospheric scattering, using such an arrangement, although the external occultation limits the field of view to the outer corona.

Further from the sun than a few solar radii, the progress of interplanetary disturbances has been monitored by spacecraft such as *Helios* moving between about 1.2 and 1/3 AU (1 AU = 1 Astronomical Unit, the mean sun-Earth distance), from *Ulysses* passing over the solar poles, and from the Earth's orbit by means of very low frequency radio waves emitted by plasma oscillations in these disturbances. These radio waves, in the kilometer wavelength range, can only be observed by antennas on spacecraft, since their frequencies lie below the cut-off imposed by the Earth's ionosphere.

The solar wind punches out a volume in the interstellar material through which it passes, called the heliosphere. Its dimensions are still quite uncertain. The so- called termination shock marking the transition between the supersonic expansion of the wind, and the interstellar material, was passed by the Voyager I probe at a distance of 94 AU from the sun; Voyager 2 detected it about 10 AU closer. Both probes are now in interstellar space, several times Pluto's distance from the sun.

1.9 Modern Solar Instrumentation

The sun has now been observed in emissions ranging from neutrinos and high energy charged particles through the full spectrum of photon energies from γ-rays to kilometric radio waves. So it can no longer be said that that entirely untouched wavelength frontiers exist for future solar observations. Still, many opportunities exist for application of basic techniques to the less well probed parts of the spectrum. For instance, neutrino imaging and spectrometry are possible, and remain to be exploited. Or, broad band thermal detectors now enable imaging in the sun's total, integrated light, instead of just restricted spectral ranges accessed by various photon detectors like CCDs.

Without special precautions, even a small solar telescope would act as a solar furnace. Most modern solar telescopes, therefore, share a common basic design intended to reduce heat load. Traditionally, such a design is described as a photoheliograph. Here, a large objective (either a lens or now more commonly a mirror) forms a primary image on a heat shield of temperature-resistant and often actively cooled, material. Only a small part of the primary image is allowed to pass through a field limiting aperture; most of the image is reflected back out through the front of the telescope or diverted sideways into a cooled heat sink. Behind the field limiting aperture the light passes through re-imaging optics that greatly enlarge the image and feed it to various focal plane instruments like cameras, spectrometers, or magnetographs. These various optical paths are usually accessed with beam splitters, or by moving the smaller focal plane instruments around on motorized stages on an optical bench.

Other ways of reducing heat load can be useful in specialized applications. Using bare glass mirrors on the primary and secondary can produce a quite satisfactory image that is only $(0.04)^2$ times, or about 0.01%, as bright as if the mirrors were aluminized. If limitation of the spectral range is no disadvantage, then dielectric coating of the objective can turn it into a band pass filter transmitting over only tens or hundreds of Angstroms around a chosen wavelength.

The achievement of higher angular resolution has, in the past, provided one of the main avenues to progress. The highest resolution is now being achieved with the 1-meter aperture Swedish Solar Telescope (SST) on La Palma in the Canary Islands, and with the recently completed New Solar Telescope (NST) at Big Bear Lake in California. These telescopes have demonstrated resolution approaching 0.1″ or about 70 km on the sun. This is somewhat higher than the resolution achieved by the Solar Optical telescope, SOT, on the Hinode spacecraft. But SOT's resolution is achieved continuously over a relatively large field of view, whereas the ground based telescopes rely on shorter periods of excellent seeing, and/or limitation to the smaller field covered by an adaptive optics system.

Such an adaptive optics servo system, which senses and removes deformations of the wavefront entering the telescope caused by Earth's atmospheric seeing, will be an important feature of the next-generation Advanced Technology Solar Telescope (ATST). This giant instrument, with a 4-m aperture, is planned for construction on the peak of Haleakala in Hawaii. Its design is illustrated in Fig. 1-18. Its large aperture, besides yielding even higher angular resolution, is required to gather enough light to perform vector magnetic studies of the smallest-diameter photospheric flux tubes.

The ultimate in spatial resolution will be achieved by a spacecraft called the Solar Probe Plus (Fig. 1-19) now being prepared for launch into an eccentric orbit that will take it to within about 8 solar radii of the photosphere. That is into the outer corona where it is thought that the solar wind originates. Exposure to solar irradiance about 500 times higher than measured at 1 AU will produce temperatures of over 2000 °C at the carbon composite heat shield that will protect the instruments. A range of in situ measurements of particles, and of both magnetic and electric fields, are planned. Imaging of the corona will also be carried out. Unfortunately, the time spent near perihelion will be limited, but the findings should still be of great interest. The Solar Orbiter mission to be launched in a few years by the European Space Agency will make both in situ plasma measurements and image over a wide range from hard X-rays to visible light. Its orbit will take it to a perihelion at 0.28 AU.

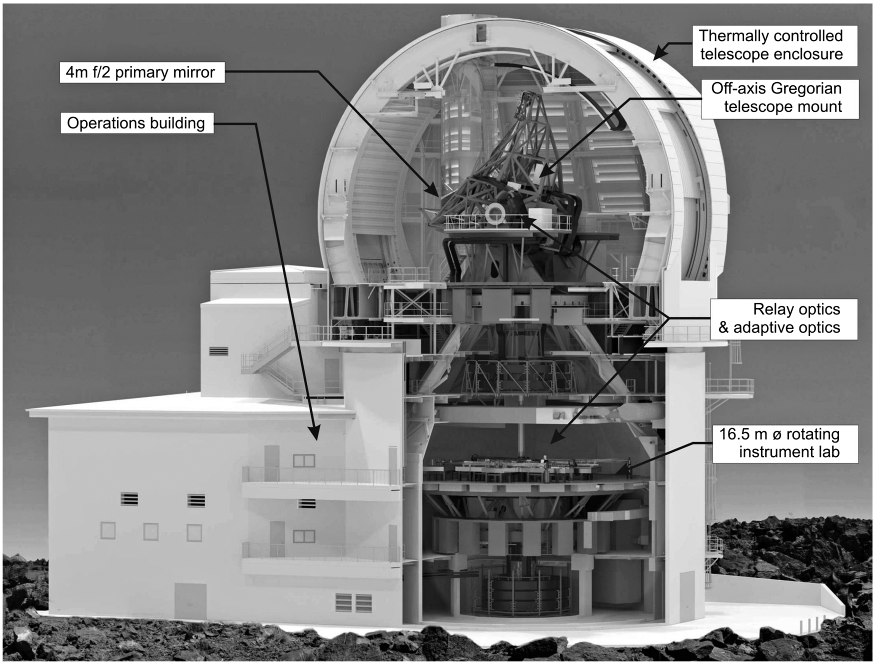

Fig. 1-18 A design for the proposed 4-meter class Advanced Technology Solar Telescope, making use of adaptive optics to achieve angular resolution better than 50 milli-arc secs. The ATST is optimized for wavelength coverage extending beyond 20 microns in the infra-red, and low instrumental polarization. Figure kindly povided by the U.S. National Solar Observatory.

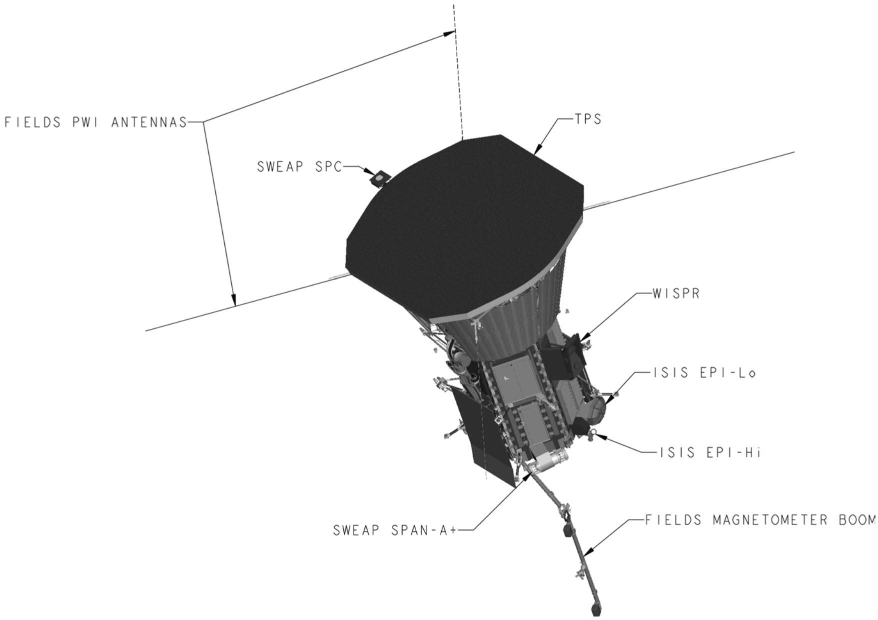

Fig. 1-19 Artist's impression of the Solar Probe Plus showing the thermal protection system (TPS), the imager (WISPR), the solar particle experiments (SWEAP, SPC, SPA), and the fields detection antennas and boom. The imaging instrument will operate in the penumbra of the heat shield, where moderate temperatures are expected. By courtesy of NASA.

Coronagraphs, such as illustrated in Fig. 1-16, continue to produce daily patrol observations of magnetic active regions near the solar limb; their coronal radiations are often visible well before they rotate onto the disk, so they provide information useful for activity predictions. In a modification of coronagraph design that has proved very fruitful, a second occulting disk is placed in *front* of the coronagraph objective, shielding it from direct solar illumination. Only rays from the outer regions of the solar corona are able to pass around this external occulting disk and are imaged. In this way, scattered light levels as low as 10^{-7} can be reached in space-borne imaging above the scattering caused by the Earth's atmosphere. Such an externally occulted coronagraph, flown on the Skylab Space Station by the High Altitude Observatory, enabled the discovery of coronal mass ejections (CMEs) whose understanding has revolutionized space weather research (see Chapter 14). The LASCO coronagraphs of this design built by the U.S. Naval Research Laboratory and located between Earth and the sun on the SOHO mission contribute greatly to the warnings of CMEs headed toward Earth.

Coronal imaging on the disk is carried out using grazing incidence reflecting optics as described in Section 1.7, but the images obtained from spacecraft are now captured on charge coupled detectors (CCDs) instead of the film that was carried above the atmosphere and then returned, in earlier X-ray imaging from rockets and from the Skylab Space Station. The most up-to-date X-ray telescope is the approximately 30-cm aperture XRT being flown by the Smithsonian Astrophysical Observatory (SAO) on the Japanese Hinode mission. It covers the wavelength range between roughly 2–200 Å. Perhaps the most spectacular images of the corona on the disk have been obtained by EUV telescopes operating at wavelengths between about 150–1000 Å. The lines at these wavelengths tend to be emitted by plasmas that are somewhat cooler and illuminate more sharply defined loops than those emitted in the soft X-ray range. The most sophisticated implementation of these normal incidence reflecting telescopes, developed mainly at the Lockheed Corporation and at SAO, is the Atmospheric Imaging Assembly (AIA) recently launched on the Solar Dynamics Observatory (SDO).

Modern magnetographs are based on the same principles used in the Babcock magnetograph described in Section 5.1, but the measurement now includes the transverse component of the field, not only the longitudinal (i.e. line-of-sight) component measured in Babcock's instrument. This is accomplished by analyzing not only the circular polarization of the line wings, but also its relation to the linear polarization of the line core. Linear polarization measurements are more difficult because optical elements can themselves generate spurious linear polarization. However, several vector magnetographs have now been in routine operation for decades and, with careful precautions, are able to measure photospheric vector fields reproducibly. Such measurements are often made by raster scanning a solar image across the entrance slit of a spectra-polarimeter, to build up an image of the vector field over the field of view. However, for space-borne applications with size and weight limitations such an arrangement is less practical than the use of narrow-band filters as the monochromator, together with ccd detectors.

More generally, imaging of solar brightness, magnetic fields and line of sight motions is increasingly carried out using filters of the birefringent designs described in Section 1.6, or based on the Fabry Perot and Michelson interferometers described in standard optics texts. The Helioseismic and Magnetic Imager (HMI) recently launched on the SDO mission is an example of the last type. Its layout is illustrated in Fig. 1-20. All of these filters exhibit a change in wavelength for off- axis rays, so they must be used in reasonably well collimated light if a narrow passband is to be achieved. The birefringent filters tend to give a more uniform passband across the field of view. However, their transmission is low – typically only a few percent. The FP design is best suited when the narrowest passbands, even below 100 milli-Angstroms, are required. The all-reflecting Michelson design can operate well into the infrared.

In the most recent implementations of all three of these designs, scanning in wavelength over a wide range is carried out using liquid crystal (LC) elements whose retardance can be changed, thus enabling wavelengths to be accessed in random order in milliseconds, with no image shift or moving parts to generate vibration. The first successful LC filters, of the birefringent type, were developed in the early 1990s at Cambridge Research and Instrumentation, Inc.

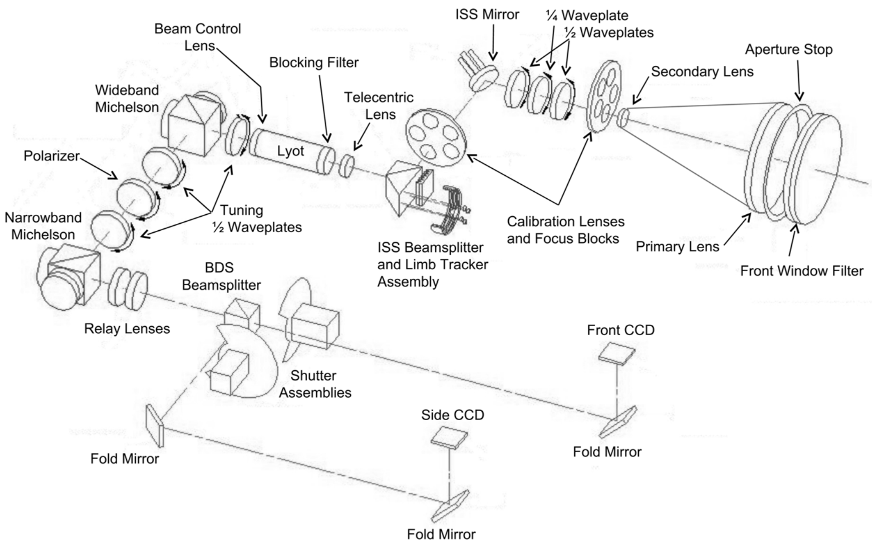

Fig. 1-20 Optical layout of the Helioseismic Magnetic Imager (HMI) showing the imaging optics, polarization analyzers, Lyot and Michelson monochromators, and detectors. By permission of J. Schou, Stanford University.

In the radio frequency range, improvements at the Very Large Array in Socorro, New Mexico, provide better frequency coverage for solar burst observations, and for active region observations from about 1 to 50 GHz. The availability of the VLA for solar work will, however, probably remain a challenge. Upgrades at the Owens Valley Solar Array in California will achieve similar angular resolution to that generally accessible at the VLA, enabling comparable resolution in flare microwaves as achieved in flare hard X-rays by spacecraft such as RHESSI. The new ALMA millimeter wave interferometer array located in Chile's Atacama Desert (Fig. 1-21) has been designed to include studies of the chromosphere, in principle at angular resolutions down to a few milli-arc secs. This would be far better than achievable at visible or ultraviolet wavelengths used in the past for chromospheric studies. In the decimeter band, important advances are expected from the Chinese Spectral Radioheliograph which will enable imaging in a frequency range where analyses have relied so far mainly on unresolved flux data.

Total solar irradiance (TSI) variation has been monitored since 1978 by a series of space-borne pyrheliometers whose basic principle is similar to that described in Section 1.4, except that the solar heating is compared to joule heating of an electrical coil wound onto the receiver. The receiver itself is also now a cavity that traps solar radiation better than the flat plate used by Pouillet in 1837. Despite sophisticated design improvements, though, the successive measurement sets obtained from various spacecraft have exhibited large calibration offsets of up to 0.5%, although the reproducibility of each set in space was an order of magnitude better. The reason for

Fig. 1-21 The Atacama Long Baseline Millimeter Array (ALMA) in its present form, with 39 of the 64 12-meter aperture dishes in place. The site in the Chilean desert is located at 5,000 meters altitude. By permission of the National Science Foundation

these perplexing offsets now seems to have been identified with unanticipated scattering from the entrance apertures used in some of the pyrheliometer designs. This advance was achieved at the University of Colorado; the flight pyrheliometer was characterized using a stabilized laser light source and a more accurate cryogenic radiometer. This technique, widely employed in metrology laboratories since the 1990's, will be used to calibrate future flight pyrheliometers, reducing the need to overlap successive missions to achieve reproducibility over many decades.

Variability of the solar spectral irradiance in the ultraviolet at wavelengths below about 250 nm has been measured with various degrees of success by a series of space-borne spectro radiometers flown since the 1970s. Such measurements require maintenance of spectroradiometer calibration in the face of degradation of both optical surfaces and detectors in the harsh space environment. Ingenious calibration schemes including flying on-board standards lamps, or observing UV-emitting stars, have been used to surmount these challenges, particularly at the U.S. Naval Research Laboratory and at the University of Colorado. The sun's brightening around the peak of the 11 year cycle, and the origin of this brightening in plage magnetic structures imaged in the upper photosphere and chromospheres, has now been well established. Measurement of the much smaller (<1%) SSI variation at longer wavelengths has been more difficult, although findings reported recently from the SORCE mission indicate that the behavior of SSI has now been measured with useful precision over the decline of spot cycle 23, well into the near infrared.

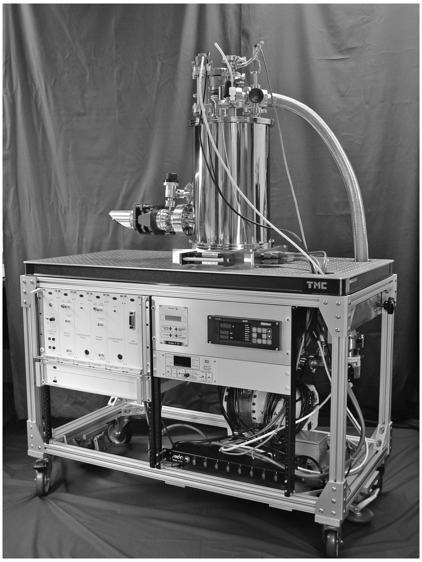

Fig. 1-22 A cryogenic radiometer as used for characterization of flight pyrheliometers and spectra-radiometers, constructed by L-1 Standards, Inc., under licence from Cambridge Research and Instrumentation, Inc. Intensity-stabilized, polarized, laser light is introduced into the cryo-radiometer cavity through a Brewster-angled window in the horizontal snout. The same laser beam is used to illuminate the flight radiometer. By permission of Steven Lorenz, L-1 Standards, Inc.

Stable and continuous monitoring of certain key metrics of solar behavior plays an often less – heralded but very important role in solar research. Maintenance of the basic indices of solar activity such as the sunspot number can be challenging as the institutions and individuals involved change over decades and centuries. Workshops dedicated to the understanding of scale changes in the spot number since the 17^{th} century, are very important to accurate understanding of solar activity in the past and future. The recently opened SOLIS facility (Fig. 1-23) at the U.S. National Solar Observatory (NSO) is an example of resources allocated to providing high quality synoptic solar magnetograms and images on a daily basis.

The time scale for planning, funding and bringing new observational facilities on line continues to increase. Once, instruments could be built or modified in a few days or weeks. Now years to decades are required. To some extent this is unavoidable as we build ever larger versions of basically similar instruments, to achieve higher angular and spectral resolution. Fortunately, there are still relatively unexplored areas that require mainly some imagination and perseverance to investigate. The equipment required is often modest. One example is the exploration of plasma electric fields, which are accessible through the Stark effect using an approach not much different than used to study magnetic fields with the Zeeman effect. Although

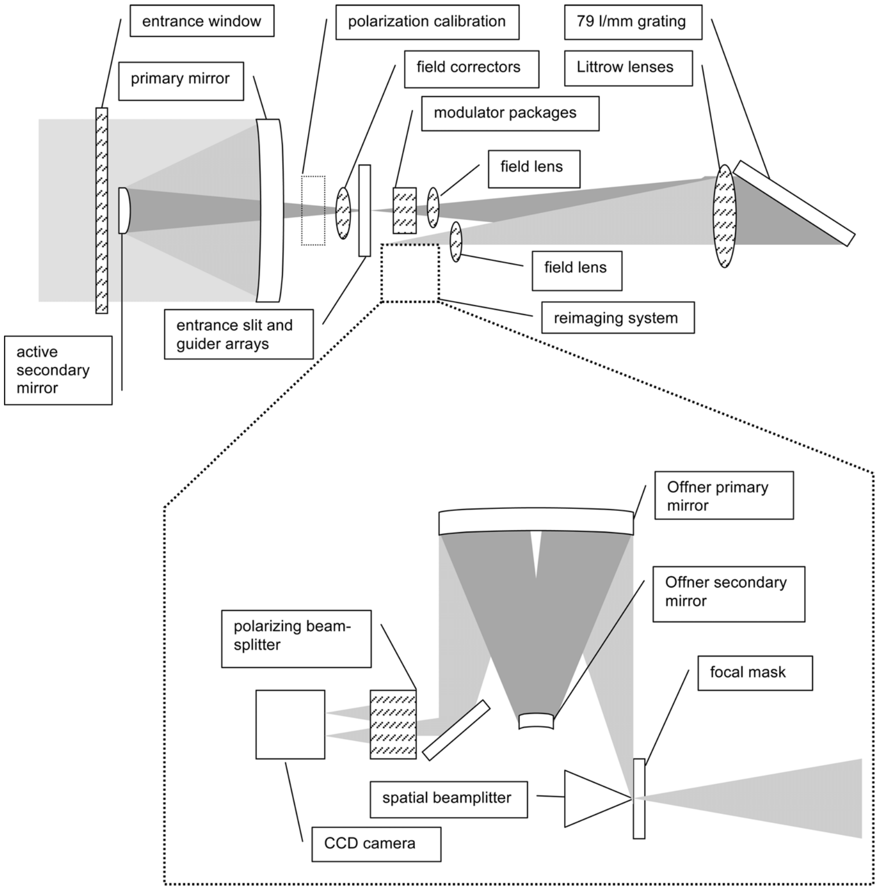

Fig. 1-23 Optical layout of the SOLIS synoptic telescope at the U.S. National Solar Observatory, Kitt Peak, Arizona. Kindly provided by L. Bertello.

they play a parallel role to magnetic fields in a wide range of solar plasma phenomena, their study has received nowhere near the same level of attention. Opportunities like this to perhaps achieve major breakthroughs await next-generation solar observers willing to think "outside the box".

Additional Reading

Historical Development of Solar Research

A. Berry, *A Short History of Astronomy*, Dover, (1961).

L. Goldberg, "Introduction", in *The Sun*. G. Kuiper (Ed.) University of Chicago Press, p. 1, (1953).

A. Meadows, *Early Solar Physics*, Pergamon, (1970).

Solar Instrumentation

E. Hecht, "Optics" 4th ed. Addison-Wesley (2002).

M. Kundu, "Techniques of Solar Observations", in *Solar Radio Astronomy*, Wiley-Interscience (1965).

N. Ness, "Magnetometers for Space Research" *Sp. Sci. Revs*. **11**, 459, (1970).

K. Ogilivie, et al. "SWE, A Comprehensive Plasma Instrument", *Sp. Sci. Revs*. **71**, 55, (1995).

R. Dunn, "High Resolution Solar Telescopes", *Solar Phys*., **100**, 1 (1985).

S. Keil, (Ed.),"Innovative Telescopes and Instrumentation for Solar Astrophysics" SPIE Proceedings, **4853**, (2003).

W. Pesnell et al., eds., "The Solar Dynamics Observatory", *Solar Phys*., **275** (2012).

General Reading

J. Eddy, "The Sun, the Earth, and Near-Earth Space", NASA/U.S. Government Printing Office, (2009).

K. Lang, "The Sun from Space", Springer Verlag, (2000).

L. Golub and J. Pasachoff, "Nearest Star", Harvard Univ. Press, (2001).

Exercises

1. Explain how Galileo was able to distinguish between sunspots and planets transiting the solar disc using his observations of their foreshortening, and of their decreasing rate of motion, near the limb.
2. Describe the triangulation method used by Aristarchus to estimate the sun's distance (A. Berry, *A Short History of Astronomy*, Dover, 1961, p. 34). Compare the precision of his technique to that used in the 17th and 18th centuries, based on the transits of Venus (op. cit., p. 284), and to modern methods based on radar echoes from Venus, Mercury, and Mars and particularly from the asteroid Eros [see, e.g., D. Muhleman, *Monthly Notices Roy. Astron. Soc*., **144**, 151 (1969)].
3. Describe the principle behind determinations of the sun's mass (see, e.g., K. Lang, *Astrophysical Formulae*, Springer-Verlag, 1980, p. 544) and give the

main sources of error and the accuracy of modern values [see, e.g., L. Goldberg, in *The Sun*, G. Kuiper (Ed.), University of Chicago Press, 1953, p. 17].

4. Compare the principle of operation of Pouillet's calorimeter (Fig. 1-10) and of modern pyrheliometers [R. Willson, *Appl. Opt.*, **18**, 179 (1979)]. Explain why such relatively insensitive thermal detectors rather than more common astronomical detectors, such as photomultipliers or photodiodes, must be used to measure the total solar irradiance.
5. Explain the basic operating principles of a) a spectroheliograph, b) a magnetograph, c) a coronagraph, using a diagram in each case. State the main advance in solar observations achieved with each instrument.
6. The solar image formed by even a small-aperture mirror or positive lens can be very hot. But if the ratio of the focal length to the aperture diameter is increased (i.e. the focal ratio is large) the heating decreases (as the inverse square of this ratio). Calculate the f-ratio for which the irradiance at the focal plane is equal to that of natural sunlight.