AURORAL PHYSICS

Sydney Chapman
High Altitude Observatory, Boulder, Colorado

Records of the aurora go far back, as do the Chinese records of sunspots; for centuries the records of both were haphazard and scanty. One of the first to record the aurora systematically was the atomic chemist John Dalton (1793, 1828), who kept a “meteorological” journal for 57 years. Its first mention of the aurora was on March 24, 1787. This was well before Schwabe took up the regular observation of sunspots, in 1826. Since then sunspots have been increasingly well observed, but the auroral record has until recently remained mostly at the mercy of chance observation, and still more uncertain written record.

Scattered references to the aurora, generally fanciful or superstitious, and often obscure, are found in classical Greek and Latin literature, because the aurora is sometimes, at long intervals, visible from the Mediterranean. Likewise some medieval authors of northern Europe and Iceland mention the aurora, which there is much more often visible. The enlightened French scientist Gassendi (1658) seems to have been one of the first to regard the aurora as a natural phenomenon; he saw a great aurora from southern France on September 12, 1621, and gave it the name aurora borealis, or northern dawn. Another term commonly used for the aurora borealis is the northern lights.

Perhaps the first mention of the aurora in the periodical literature of science was by Halley (1716), who gave a description, in the Philosophical Transactions of the Royal Society, at its request, of the great aurora of March 16, 1716. This included a corona (Figure 1), lying somewhat south of the zenith at London. Halley, an ardent observer of Nature, then aged sixty, remarked that up to that time he had seen every kind of “meteor” except an aurora.

The first treatise on the aurora, by de Mairan (1733), was published by the French Academy of Sciences; it ran to a second much extended edition in 1754. It is an excellent book, still worth reading. Mairan thought that in the Southern Hemisphere lights similar to the northern lights might appear. He knew that the aurora is seen more often in higher latitudes than in France, and consequently he thought the best people to ask about the possibility would be men who had been to high southern latitudes. He consulted Frezier and Ulloa, who had rounded Cape Horn on scientific expeditions, whether they had seen such lights. Their affirmative replies do not carry conviction. The first records of indubitable southern auroras were those made in February and March 1773 by Captain Cook (1961) and others on
the two ships (then separated after fog) of his second great voyage, during their eastward journey across the southern Indian Ocean. Cook's latitude and longitude at the time of his first sight of the southern aurora were 57.9°S and 83°E. He named it the *aurora australis* (this term has at times been inappropriately used in referring to an aurora borealis seen to the south of
the observer's zenith). Such southern auroras are also called southern lights. The names aurora polaris and polar lights cover both northern and southern auroras.

Several nineteenth century writers collected the records of auroras. Up to 1872 the most extensive collection was that made by Fritz (1873). A German geographer Muncke (1833) who studied the auroral records of several arctic explorers found that although auroras are seen more often as one goes northward from mid-Europe, there is a latitude beyond which they are seen less often (and more commonly to the south than to the north). Loomis (1860) drew the first sketch of the *auroral zone* of maximum frequency of visibility of the aurora. Fritz (1874) much improved on this by his map of equilines of frequency of auroral visibility, which he called isochasms. Figure 2a is a pictorial version of his map, drawn for American readers (Gartlein 1947). The isochasms are oval curves, but are not centered on the geographical pole. Instead, as came later to be recognized, they are centered on the pole of the Earth's magnetic dipole axis, situated in north-

**Figure 2a.** Lines of equal frequency, averaged over many years, of auroral visibility (isochasms), drawn by Fritz (1873), here displayed on a map of the North American and polar regions. (Gartlein 1947, amended)
Figure 2b. Three of the outermost isochasms for the South American and polar regions; they are approximate, being antipodal to the corresponding northern isochasms.

West Greenland. The dipole axis is inclined by 11° to the geographical axis; its northern end is tilted towards America, giving the inhabitants of this region many more opportunities of seeing auroras than are available to Siberian viewers in the same latitude.

Conversely in south America, even at its southern extremity near Cape Horn (Figure 2b), auroras are much more rare than in similar latitudes near the antipodal meridian, somewhat west of Australia. Halley, on his pioneer magnetic survey voyage to the south Atlantic, 1698 to 1700, reached 52°S latitude without seeing any aurora; it could not be expected that he would have seen any, at that latitude and longitude, especially as the epoch was near sunspot minimum. Cape Horn is only about 4° further south.

Each of Fritz's isochasms bears a number, signifying how often per year, on the average of many years, the aurora might be seen if all nights were cloudless. The maximum isochasm passes over northern Canada, across Alaska, to the north of Siberia, and to the south of Iceland. A band a few degrees wide, centered on this isochasm, is called the auroral zone. The frequency of visibility falls off rapidly to the south of the zone, and less rapidly to the north. Vestine (1944) corrected Fritz's isochasms near the zone, but the outer isochasms have never been revised, though they are
Figure 3. Three maximum southern isochasms (on a map of geomagnetic dipole coordinates) inferred from observations, Nos. 2, 3, 4 being given respectively by Vestine & Snyder (1945), Bond & Jacka (1960), and Feldstein (1960); the shaded band, No. 1, shows a still earlier sketch of the southern auroral zone, drawn by White & Geddes (1939). After Hultquist (1961).

Based on rather scanty data, accumulated unsystematically over the centuries.

The land distribution in the Southern Hemisphere is less favorable for the auroral record than that in the north. Scientific records of auroras, during the precolonial period, were not to be expected from the inhabitants of the southernmost parts of America, Australia, and New Zealand, though the mention of auroras can be recognized in Maori folklore. The scanty records of the aurora australis have been used to draw tentative southern isochasms,
but none has yet been drawn for low southern latitudes, where too few auroras have been observed since the record there began. Four tentative sets of southern isochasms have so far been drawn. They are in fair agreement as regards shape and location, but Hultquist (1961) has shown that in certain longitudes the latitudes of their maximum isochasm may differ by up to 4° (Figure 3).

Auroras are high enough in the atmosphere to be visible at any time over a broad band of latitude, to north and south of where they are seen overhead. Equilines of frequency of overhead auroras are called isoaurores (Chapman 1953). Northern isoaurores were first drawn by Feldstein & Solomatina (1961). They coincide approximately with the inner isochasms, but the frequencies associated with those south of the auroral zone are less than for the isochasms. The isoaurores do not extend so far towards the equator as do the isochasms, because the auroras seen from low latitudes usually lie considerably further towards the poles, and they are very high in the atmosphere.

**Auroral Relations with Geomagnetism and Sunspots**

The slow secular variation of the Earth’s magnetic field was discovered in 1635 by Gellibrand; it proceeds from deep within the Earth. Transient changes of the geomagnetic field were discovered by Graham (1724), who found that on some days the changes are quiet and regular, and on other days they are irregular and often much larger. By correspondence between Graham in London, and Celsius (1740) and Hiorter (1747) in Upsala, it became clear that days disturbed magnetically at London were likewise disturbed at Upsala; and moreover that the aurora is associated with such magnetic disturbance. This was the first connection established between the aurora and the Earth’s magnetism. Halley, in his 1716 paper, had suggested a connection of quite another kind. He recognized that the convergence of the rays of the auroral corona (Figure 1) is only apparent, being an effect of perspective. He even suggested that they probably converge towards the observer, that is, downward, like the field lines of a magnetized sphere; he gave a diagram of these, obtained from the pattern of iron filings in the meridian plane of such a sphere. He thought that the influence of a magnet, that can act through interposed bodies like marble, is exerted by “subtle” (or penetrating) particles continually circulating through the magnet, along the field lines, in both directions. He tentatively suggested that at times certain polar regions of the atmosphere, for some reason on which he did not speculate, are sensitive to the influence of such particles, causing the aurora. His intuition that the auroral rays lie along the field lines was explicitly supported later by Wilcke (1777) at Stockholm, and confirmed recently much more accurately by Maggs & Davis (1968; vide infra.)

Halley’s idea also might suggest the inference (not drawn at the time) of conjugacy, namely that when there is an aurora in the north there is one
also in the south, at the other ends of the same field lines. Mairan's expectation that there would be auroras in the Southern Hemisphere was akin to this, but less specific. Such conjugacy has now been confirmed in large measure; it has also been observed in connection with artificial "auroras" caused by high-altitude detonation of nuclear bombs (Keys 1964). A proper test involves a knowledge of the path of the field lines above the Earth from one to the other end; the lines differ somewhat from those of a uniformly magnetized sphere or magnetic dipole, which gives only a first approximation to the field of the Earth. Only near the equinoxes are both ends of the field lines from auroral regions in darkness. Considerable technical resources have been devoted to the study of conjugacy, such as all-sky cameras and image orthicon television photography aboard two jet aircraft flying along magnetically conjugate paths over Alaska and south of New Zealand (Belon, Maggs, Davis, Mather, Glass & Hughes 1969). Hargreaves (1969) has reviewed earlier conjugacy studies of this and other kinds, which reveal some departures from conjugacy, partly related to geomagnetic disturbance. Though these deserve further attention, the substantial degree of conjugacy that prevails lessens the disadvantage of the difficulty of registering southern auroras. If the northern ones could be well observed, we should have considerable knowledge thereby of the southern ones.

Auroral observation, however, is more difficult to organize than sunspot recording; the sunspots can be seen from anywhere on the Earth during the day, when cloud is absent. The aurora is visible only over parts of the Earth, those in general the least accessible. During the International Geophysical Year (IGY: July 1, 1957 to December 31, 1958, followed by IGC, the year 1959 of International Geophysical Collaboration), auroral observation was greatly enhanced above all prior levels. The IGY auroral program and stations are described in volumes 4, 5, and 8 of the IGY Annals (1957/8); many results of the visual, all-sky, photometric and spectroscopic observations and their analysis are described in volumes 20 (Stoffregen 1962, ascaplots): 29 (McInnes 1962, visaplots); 38 (Akasofu & Kimball 1965, auroral morphology); 39 (Gartlein & Gartlein 1966, synoptic visual charts); 40 (Devlin et al 1966, spectrographic); and 45 (Akasofu, Kimball & Meng 1969, auroral substorms). Besides well-organized visual programs in many countries, there were about 120 all-sky cameras, divided between the Northern and Southern Hemispheres in the approximate ratio 3:1. Even so, never was so much as half the Arctic polar sky under observation at any one time; this was because of the land and sea distribution, and cloud.

Figure 4 (Akasofu 1963) shows the distribution of the clear fields of view, partly overlapping, of at least four auroral arcs, two of them of length 5000 km. Whether such arcs ever completely encircle the boreal pole is not known, and it seems almost impossible that it should be established by ground and airborne photography, valuable as the use of aircraft is in this connection (Akasofu 1968). Possibly polar-orbiting satellites might photo-
graph the whole polar cap when this is in winter darkness, from a height enabling this great area to be covered in one picture; however, at auroral height the dark portion of the polar cap is decidedly smaller than at the ground. The world's cloud cover is continuously photographed by satellites; but the aurora changes its form and extent far more rapidly than do the clouds. Consequently reliance cannot well be placed, in seeking knowledge of the instantaneous auroral distribution and its changes, on successive pictures of the aurora from a satellite able to cover photographically only a fraction of the polar area. The project of the photography of the whole polar region from one satellite may be technically difficult, but would be free from one obstacle that hinders observation from the ground or from aircraft, the clouds. These are all well below the aurora, at least during the season of polar darkness.

The height of the aurora remained in doubt and dispute till the beginning of this century (cf Abbe 1898). Mairan much overestimated the height, others erred in the opposite direction. There were many reports of auroras being seen below the clouds, or silhouetted against mountains. Störmer (1955) substantially settled the question, by organizing, from 1910 onwards, simultaneous photography from well-separated Norwegian stations, connected by telephone, and using the same type of camera. He found the lower border to be generally located at about 100-km height, and that the auroral light could extend upwards to heights of from 150 to 1000 km. The greatest heights he found were for sunlit auroras, seen by observers on the ground in darkness, but emitting from parts of the atmosphere still in sunlight, after twilight or before dawn. He gave the height of maximum frequency of the measured points to be 100 to 110 km; some of the frequency curves show two maxima. He found very few points at a level below 90 km. At rare intervals the aurora may extend below the level (about 80 km) of noctilucent clouds (which are summer phenomena).

Very occasionally careful observers, using Störmer's technique, have reported auroras that reach down to decidedly lower levels. The first outstanding example was an intense arc with deep crimson lower border (Harang & Bauer 1932, Harang 1951) extending down to a little above 60-km height. For some years this remained a unique observation, but subsequently Canadian observers reported further cases; Currie (1955) found three that extended down to or below 70 km, and Hill (1965) found four cases (out of 1197 pairs of all-sky auroral photographs taken in the quiet sunspot years 1964/5) that extended below 70 km (one as low as 63 km). Unfortunately no spectra of these auroras were taken—it is very desirable to obtain spectra and photometric data for such exceptionally low auroras.

Points at 100-km height in the atmosphere can be seen from the ground (or photographed by all-sky cameras) within a circle of radius about 1100 km, but the outer part of an all-sky picture is compressed, and the effective radius of good record of an all-sky camera may be estimated as 500 km, as
adopted in Figure 4. A satellite able to see the 100-km level of the atmosphere over a polar cap of angular radius 25° must be at least 770 km above the ground.

During the great geomagnetic disturbances, called magnetic storms, the aurora comes within sight of lower latitudes than usual, and it is on the rare occasions of the very greatest magnetic storms that it can be seen from places on the outermost of Fritz's isochasms, as was possible three times, from Cuba and Mexico, during the IGY. On February 4, 1872, it was seen even at Bombay; this is probably the greatest aurora on record (cf Chapman 1957). On such occasions many who see it may not recognize it as an aurora, and even when seen and recognized, too often little or no mention of it finds its way into scientific literature.

Whereas several institutions are equipped with cameras, photometers, and spectroscopes for the observation of auroras in the latitudes where they are commonly seen, there is very little provision for the observation by such means of the rare low-latitude auroras. Hence there is great need to arouse interest in such phenomena among astronomers and physicists who have applicable equipment. Where they are at all likely to have the good fortune to see low-latitude auroras, forethought should be taken to be ready if the opportunity comes. These events are great natural phenomena, with a strong astronomical connection with the Sun. Astronomical observatories, usually situated well away from city lights, are the institutions most likely to be able to make good observations of such auroras. They are also in many cases linked with warning services that send news of unexpected astronomical events, and hence can be alerted when great magnetic storms and worldwide auroras are in progress. It is desirable that the Astronomical Union or other astronomical agency should help to organize such observations. The Yerkes Observatory (vide infra) has set a fine example by the long-continued attention given by several members of its staff to the observation and recording of auroras—Barnard, Sullivan, Meinel, and Elvey (later Director of the Geophysical Institute of the University of Alaska, which under his leadership became an outstanding center for auroral research).

A low-latitude observatory may be able to contribute to auroral knowledge even after the visible aurora has faded away—by looking for the light of the NI excited atom (5200 Å). This light may continue to be emitted for an hour or more after the usual auroral emissions have ceased (Götz 1947).

The lowermost dipole latitude of great auroral displays shows a fair degree of correlation with the maximum value of the Dst (H) measure of magnetic disturbance (Akasofu & Chapman 1963); this measure corresponds to the mean strength of the encircling ring current that develops strongly around the Earth during a great magnetic storm. A low-latitude observatory that arranges to receive news of outstanding geomagnetic disturbance much increases its chance of making valuable auroral observations.
Figure 4. The distribution of auroral arcs at 8:30 and 9:30 UT on March 24, 1958, as shown by several all-sky cameras; the limits of the clear single or overlapping view of the cameras are shown, the radius of the area of good view being taken as 500 km. The auroral breakup of an auroral substorm occurred between the two epochs; in the second the arcs over Alaska are completely disrupted. After Akasofu (1963).
When Schwabe’s discovery of the 11-year sunspot cycle became widely known, in 1850, it was found within a year that the same cycle is manifested in the records of the transient magnetic variations, both on quiet and disturbed days. The relation between auroras and magnetic disturbance found in 1740 by Graham and Celsius implied that auroras should also show this influence; but for some decades this connection remained uncertain, owing to the poor quality of the auroral record. It is best manifested in the “subauroral” belts (between dipole latitudes 60° and 45°). It is clearly shown by the systematic records of the aurora made throughout three sunspot cycles at the Yerkes astronomical observatory (Meinel, Negaard & Chamberlain 1954). In auroral latitudes (above 60° dipole latitude) the aurora is so frequent that the sunspot cycle is less evident as regards frequency of visibility, though it manifests itself in other ways. In minauroral latitudes (equatorwards of 45°) the record is too scanty to give smooth frequency curves.

One of the simplest forms of auroral observation, and one of great value, in the case of the infrequent subauroral and the rare minauroral auroras, is a timed record of their appearance overhead, or of their coming within 30° of the zenith, to the north or south of it. Any such observation should be sent to the national astronomical society for its auroral section (if any), otherwise to a World Data Center.

The Auroral Spectrum

The auroral spectrum and its interpretation have been studied for a century, beginning with Ångström (1869); progress was small during its earlier half. The lines and bands, which may vary greatly in their relative strengths even during a single display, are now known and identified (Petrie & Small 1952) in great but certainly not complete detail, for the polar aurora. They are less known for the subauroral displays, and much less studied for the rarer low-altitude auroras. The spectrum was intensively observed during the IGY (Devlin et al 1966). The spectrum has been examined not only in the visible region but also, using image converters, in the infrared (Bagariazky & Fedorova 1956, Vaisberg 1959), and by rockets and satellites into the ultraviolet (Miller, Fastie & Isler 1968), which contributes a considerable share of the energy of emission. The subject is of great scope and can only be briefly summarized here. Good accounts of it have been given by Bates (1960) up to about 1959, by Chamberlain (1961) up to about 1960, and by Meinel (1966) up to about 1965; Vallance Jones (1969) has summarized more recent progress. The chief centers of auroral spectral (and other) studies include auroral or more general geophysical observatories or institutions in Alaska, Canada, Sweden, Norway, and the USSR. Important discussions of the aurora and its spectrum have been given by Krassovsky (1961, 1967, 1968) and Galperin (1963); these are references to only a few
of their more recent reviews in easily accessible English translations, and enable many of their other works to be found.

One outstanding feature of the spectrum, long a mystery, is the yellow-green atomic line 5577 Å; it was finally shown by McLennan & Shrum (1924) to result from the forbidden atomic oxygen transition 1S(0.74 sec) to 1D(110 sec). These are two metastable states whose lives are indicated in brackets—the lower level is of long duration. This emission gives the most usual color of bright auroras in polar latitudes. A transition from the level 1D to the ground-state 2P gives the red doublet 6300–6364 Å, which is mainly emitted at heights 200 to 300 km, well above the level where most of the 5577 light is produced. Polar auroras in which this red light is prominent are said to be of type A. This light is a common feature of low-latitude auroras; it has often been misinterpreted as the sky glow of a great conflagration to the north, sometimes causing fire fighters to set out to seek and suppress it—as recorded of ancient Rome, and of Victorian London; one prompted a fire department enquiry at Göttingen as recently as 1957.

Other outstanding features of the auroral spectrum are the band systems of molecular nitrogen, both neutral (first and second positive, and Vegard-Kaplan bands) and ionized (first negative and Meinel bands). Though less conspicuous visually than the green and red light of oxygen, these emissions generally involve more energy. Some polar auroras show a purplish-red lower border (these are said to be of type B), due to an enhancement of the first positive bands. Vegard, an ardent pioneer in auroral spectral studies, was the first to attribute much of the auroral light to molecular nitrogen.

Other features of the spectrum, generally minor, are contributed by O₂ (the Kaplan-Meinel bands) and O₂⁺ (first negative); there are also permitted lines of O, N, and N⁺, and forbidden lines of O⁺, N, and N⁺. The lowest level metastable states 2D of O⁺ and N are remarkable for their long lives, 3.6 and 26 hr respectively. Emissions by transitions from them to the ground states give doublets 3727 (O⁺) and 5200 (N) that are weakly present in the spectrum.

The emissions so far described come from excited neutral or ionized molecules or atoms of the chemical elements most abundant in the atmosphere. The neutral particles N₂, O₂, and O are major constituents at the main level of emission. The ionized particles and the N atoms are minor constituents. Other minor constituents at that level are H and He, the former variously in molecular, atomic, and combined form (notably as OH). The auroral spectrum includes the α, β, γ and other lines of H (Vegard 1939, Gartlein 1950, Meinel 1951), and line 10830 Å of He (Shefov 1963). The last of these is observed in sunlit auroras and in twilight without aurora.

The hydrogen lines differ from all the other features in being notably broadened and, when viewed along the direction of the field lines, displaced to shorter wavelengths. These are interpreted as Doppler effects, of the
downward and lateral motion of the emitting atoms, which are not previously particles of the polar atmosphere. Some Doppler broadening is shown, of course, by all the lines of the spectrum, and provides a means, along with the study of the intensity distribution in the bands, of inferring the kinetic, vibrational, and rotational temperature of the emitting gases; in some cases interferometers are used.

The Doppler displacement of the hydrogen lines gave the first proof that at least part of the auroral light is caused by the entry of particles into our atmosphere. As noted above, Halley attributed the aurora to particles issuing from within the Earth. The idea that the cause is particles coming to the Earth from outside arose late in the last century, Birkeland (1896, 1908, 1913) being its chief sponsor and propagator; he supported it by laboratory experiments. He thought that the particles are electrons coming from the Sun, which are deflected to the polar latitudes by the geomagnetic field. Störmer (1955) developed this idea mathematically, but only for the motion of charged particles each moving in the field without being influenced by accompanying charged particles. His calculations apply to cosmic rays, but not to a neutral stream of charged particles, half negative, half positive, which was the alternative hypothesis first suggested by Lindemann (1919).

His conception, radically differing from that adopted by Birkeland and Störmer, is now directly confirmed by satellite observations; they indicate a continual flow, explained by Parker's theoretical studies of the solar wind (1963). But the occasional intensification of the flow from specially disturbed areas on the Sun, associated with magnetic storms and bright auroras, is not yet well understood.

**Auroral Morphology**

During the IGY, asca (all-sky camera) photographs were taken at most stations at 1-min intervals during dark periods in fine weather; the exposure was 55 sec followed by 5 sec to move on the film. The program produced a great volume of record, which has been extensively studied, in particular by Akasofu and his colleagues. This has thrown much light on the more systematic features of the auroral appearances and changes, enabling the course of typical auroral substorms to be ascertained. These events are of duration of order an hour; they occur frequently with moderate intensity, and during magnetic storms, whose duration is of order a day, several polar magnetic and accompanying auroral substorms, sometimes of great intensity, may occur. These two associated substorm events are aspects of more extensive and varied magnetospheric substorms (Akasofu 1964, 1965, 1968a; Davis 1966); they involve the whole ionosphere of the Earth, auroral X-ray events, proton auroras, VLF emissions, geomagnetic micropulsations, and particle flows and redistributions, partly in response to electric fields, in the magnetosphere. This is the large region, with dimensions of order ten Earth
radii on the Earth's sunward side (and far greater on the opposite side, in what is called the tail of the magnetosphere), within which the geomagnetic field is confined by the flow of neutral ionized gas from the Sun.

"The characteristics of auroral displays depend greatly on both universal time and local time, so that if observed at a particular point on the earth it is very difficult to distinguish between those that depend on universal time and those that depend on local time" (Akasofu 1968). The description of an auroral substorm in terms of these two time variables is too complex to be detailed here. The onset is a universal time phenomenon, but the characteristics seen at different longitudes following the onset are by no means similar. Between auroral substorms there may be extensive quiet auroras; when the substorm begins (the auroral "breakup"), the arcs become rayed and thinner; they multiply and move rapidly, some to the north, others to the south; waves travel along them, with great curtains and folding in rayed forms (Figure 1). Some time after this expansion the activity decreases—the auroral arcs return to a narrower band of latitude and become quiet again; the last phase may consist of diffuse pulsating patches of auroral light. The inferences drawn from the ascastfilms have been confirmed and extended by observation of the aurora from an aircraft that in polar latitudes can remain on or near a constant local-time meridian (Akasofu 1968).

Auroral photographic techniques have steadily improved in recent years, and one post-IGY development is outstanding, the image orthicon television system. This can make successful exposures in 1/60 sec, recording at 24 or 30 per sec (Davis 1967, 1969); it has been used in jet aircraft as well as on the ground. The field of view in most cases was 12° × 16°, so that the full grand spectacle of a great aurora is not shown; but the system has been most useful in the study of conjugacy by simultaneous aircraft flights in the two hemispheres, and in revealing auroral changes too rapid to be seen even by the eye (the IGY ascastfilms were of course still less percipient). Under favorable operating conditions a linear element of low brightness and angular width 0.03° can be recorded. The spectral response resembles but is rather wider than that of the eye.

One most striking feature of the aurora on many occasions is the confinement of the visible light to thin sheets, which lie along the geomagnetic field lines (Figure 5). When a rayed arc passes across the radiant point (the magnetic "zenith"), the ray structure momentarily disappears. An ellipse containing 2/3 of 30 measures of the radiant, made at College, Alaska, by the TV technique (Maggs & Davis 1968), has a north-south minor axis of 1°, and a 2° major axis, showing how close is the alignment along the field lines (which may be slightly modified during magnetic disturbance). The thickness of various auroral sheets was measured on many occasions during 15 hr of observation time in 1966 and 1967. Errors of order 50 to 100 m may occur, partly because the picture is not taken exactly when the sheet lies
Figure 5. A close pair of auroral arcs photographed nearly edge on, by image orthicon TV camera. The lower arc is measured to have a thickness of 350 m; it is unusually regular. The forms shown in the upper part are seen sideways. After Maggs & Davis (1968).

along the radiant direction. It was found that the brighter the arc, the thinner; the median thickness among 581 measures was 230 m; the range was from the lower (instrumental) limit of 70 m up to 4440 m. Both thinner and thicker sheets certainly occur. Such edge-on photographs of the auroral sheets also often reveal a fine pleating of the sheet (Figure 6; Akasofu 1963a). Seen broadside on (or obliquely), the pleats would show as rays, owing to the triple amount of light from the pleats.

Another remarkable and very frequent feature of auroral morphology is the occurrence of multiple arcs, sometimes as many as seven or eight, which are approximately parallel (transverse to the magnetic meridian) and often move in unison, especially in quiet periods, from north to south or vice versa. Their north-to-south spacing may be 40 km. But as in Figure 5, what may appear as a single arc when seen from the side may be revealed as a close double arc when seen edge on.

Despite the incomplete coverage of the polar skies by all-sky cameras,
Feldstein (1960; see also Feldstein & Starkow 1968, Khorosheva 1962) inferred the general nature of the synoptic pattern of the auroral form and location at any instant. The arcs tend to lie in a narrow oval belt encircling the dipole pole, but not centered on it. The center is displaced by about 3° along the midnight meridian, away from the sunward direction. The radius of the oval is about 20°, so that near the midnight meridian the belt, called the auroral oval, coincides with the auroral zone; elsewhere it lies within the zone. To a first approximation the oval is fixed relative to the Sun, and the Earth rotates daily under the oval. The zone is a statistical result of the daily passage of each part of it across the midnight meridian, where it coincides with the average position of the oval, where the auroras are often most intense. As the oval is eccentric relative to the dipole pole, a station may be within the oval at midnight but outside it at noon. Auroral substorms, and polar magnetic disturbance, are best described with reference to position relative to the oval rather than to the zone. The oval encloses the area of polar-cap absorption, caused by 1–10 MeV protons, which produce a polar-glow aurora (Sandford 1962).
AURORAL PHYSICS

At times of great magnetic disturbance the radius of the oval is enlarged; this corresponds to the visibility of the aurora from subauroral or even minauroral latitudes (Akasofu & Chapman 1962).

Mishin & Popov (1969) contest the continuity of particle penetration along the Feldstein oval synoptic pattern of the aurora. Using magnetic and ionospheric as well as auroral data, they infer the occurrence of two quasicircular auroral activity zones. Akasofu (1968a), who has taken part in long-continued exploration of the auroral arctic region by jet aircraft, following the auroral bands, maintains that they confirm the oval pattern accurately and completely.

Another important (and highly variable) feature of auroral morphology is the brightness distribution or emission rate as a function of height, or of distance along the field lines. Early studies of this kind were made by Vegard (1930), who considered different spectral components of the light, and their ratio at different heights. Harang (1946, 1946a) made densitometer measurements of his auroral photographs, and inferred the variation of the scale height of the atmosphere along the rays. More recently, as during the IGY, photometers and scanning photometers, some provided with filters to isolate different spectral regions, were applied to such studies. Belon, Romick & Rees (1966) reported on the height distribution of the light of wavelength 3914(N2+), from scanning photometer records from two Alaskan stations 226 km apart, Fort Yukon and College. Romick & Belon (1967), using their data from the same two stations, went beyond the thin-sheet approximation, and by a complicated method of analysis derived contours of the volume emission rates for the 3914 and 5577 emissions (Figure 7; note that the scales of abscissae are not the same for the two sets of contours). The respective peak rates are $2.3 \times 10^4$ and $4.4 \times 10^4$ photons/cm$^3$ sec.

The unit of sky brightness used in connection with the aurora (and also for the airglow) is the Rayleigh, namely $10^6$ photons/cm$^2$ (column) sec (cf Chamberlain 1961, p. 570). Auroras are classed by visual observers according to an International Brightness Coefficient (Chamberlain 1961, p. 571).

The detailed processes of these and the other auroral optical emissions are too complex to be discussed here. Kaplan (1932) produced in the laboratory an afterglow discharge spectrum in nitrogen that he named the auroral afterglow, because of its resemblance to the auroral spectrum. This is dominated by nitrogen contributions, though they are not the most conspicuous to the eye. Even the complex nitrogen spectra obtained under controlled laboratory conditions are still under debate (Oldenberg 1969). Secondary electrons produced by the primary ones are the main agents in producing the spectrum. Green & Barth (1965) have discussed the nitrogen emission resulting from bombardment of the atmosphere by 30-keV electrons. They concluded that 40% of the energy should appear in the far-ultraviolet spec-
Figure 7. The volume distribution of emission of two components of the auroral spectrum, in the meridian plane normal to the auroral sheet; note the difference in the two scales of latitude. The isophotes show percentages of the maximum emission rate. After Romick & Belon (1967). The difference between the distributions implies that the 5577 emission is not necessarily due to the excitation of oxygen atoms by secondary electrons ejected from nitrogen molecules. The precise cause of the 5577 emission is still a difficult problem.

trum, and that another 40% goes to ionization, which later contributes a recombination spectrum.

Besides the auroral light emission, auroras at some times and places emit radio waves and possibly also Cerenkov radiations (Chamberlain 1961). Infrasonic waves from auroras have been studied by Wilson (1969, 1969a), using two microphone arrays at College and Palmer, Alaska; he associates such waves with electrojets (electric currents of limited cross section) in supersonic motion, flowing along moving auroral arcs. The ionosphere, ionized during the day by solar wave radiation, is ionized at night (as also, in certain regions, by day) by entering particles; some of these produce X rays that travel down to balloon levels, where they have been observed by balloon-borne instruments (Winckler et al 1958, 1959; Anderson 1965; Brown 1966).
THE PARTICLES THAT CAUSE THE AURORA

In 1904 Maunder interpreted the 27-day recurrence tendency shown by magnetic disturbance and auroras as indicating emission (often long continued) of limited streams of gas from active regions on the Sun. For about half a century thereafter it was generally accepted that such solar streams cause magnetic disturbance and auroras, but in 1950 it still seemed necessary to defend this idea (Chapman 1950) against the alternative proposal (Hulburt 1929) that the cause is solar ultraviolet light. Meinel’s discovery (1950), soon afterward, of the Doppler displacement of the hydrogen lines in the auroral spectrum, strengthened the particle hypothesis.

But the simple theory advocated by Birkeland and Störmer, that the aurora is caused by particles from the Sun, guided polewards by the geomagnetic field, faced a serious difficulty. Chapman & Ferraro (1930) had inferred that the impact on the Earth of a solar stream of neutral ionized gas, as proposed by Lindemann, would be prevented by the geomagnetic field, and that this field would be confined and compressed within a cavity in the solar stream (this cavity is now called the magnetosphere; Gold 1959). Chapman & Ferraro (1940) also showed that unless the solar stream is of such low density as to be unable to have any appreciable effect on the Earth, the mutual attraction of its positively and negatively charged particles makes their paths entirely different from those calculated by Störmer, who took account only of the force exerted by the magnetic field. But though concluding that the entry of particles of the stream into the cavity would be severely restricted, Chapman & Ferraro (1931) believed that during magnetic storms an electric ring current grows inside the cavity, in some way they were unable to explain. Their conception of this ring current as being of toroidal form was later corrected by Singer (1957), using ideas of Alfvén (1940, 1950) as to the paths of particles trapped within the cavity. The satellite discoveries (Van Allen & Frank 1959, Vernov et al 1960) of particles trapped within the cavity, forming the Van Allen belts, soon confirmed these conclusions.

In 1955 Van Allen and his colleagues (Meredith et al 1955) found evidence of particle entry into the auroral atmosphere, additional to and different from that of Meinel; by rockoon observations they had detected X rays (Van Allen & Kasper 1956), attributed to the close impact of energetic electrons with the nuclei of atmospheric atoms. Thus the atmosphere receives from outside both protons and electrons. Since 1955 the X rays have been extensively studied by balloon-borne recorders (Anderson 1958). The flux and energy spectra of electrons and protons have been measured at higher levels of the atmosphere by rockets, and further out by low-altitude satellites at various local times during and between magnetospheric substorms. Many of the results have been summarized by O’Brien (1967);
Akasofu (1968a) gives some later references. The details are too voluminous for inclusion here.

Even before 1955 the nature of the particles responsible for the different components of the auroral spectrum had been much discussed. It is now clear that there are three types of auroral emission, classified according to the exciting cause, which may be electrons, protons, or the heating of the atmosphere (Seaton 1955). The first two kinds, caused by electrons and protons entering from outside, may be called energetic auroras (Cole 1969), and the third kind may be called thermal. The oxygen red lines, of light from great heights, have on occasion shown Doppler broadening indicating, according to Krassovsky (1959, 1961) and Mulyarchik & Shcheglov (1963), temperatures up to 3500°K, which the electron temperature may exceed.

These types may appear singly, or two or more together. They may have different ranges of height or latitude and longitude (or local time).

The electrons are the most visually effective cause. Their energy can be inferred from the extent of their penetration of the atmosphere. The ordinary lower limit, about 100-km height, corresponds to energy of a few kiloelectron volts (keV). For the rare cases of penetration down to 80, 70, or even 60 km, the energy may range up to a few hundred keV. As the electrons descend along the field lines, they collide with atmospheric particles (mainly N₂, O, and O₂), which they may ionize, dissociate and/or excite; also the electrons are scattered, changing their pitch angles. The average energy loss to the electron, per ion-electron pair produced by the impact, is about 35 eV, so that a 10-keV electron may ionize about 300 particles. Thus the sheet of gas that emits the light of an auroral arc will be highly ionized and electrically conducting. Often the arcs move (usually, in the evening, from north to south); thus ionization of the air may be spread over a band much wider than the thin auroral sheet, though the ionization decays after the entry of particles has passed away to another region.

The thinness and movement of the auroral sheets produce notable horizontal gradients of electron density and electric conductivity; Ungstrup (1969) has discussed the influence of such gradients on the power of whistlers and very low-frequency emission coming from the magnetosphere, to penetrate to the ground. He finds that for a certain range of the ratio of the horizontal scale length of the electron density to the atmospheric scale height, auroral arcs may greatly facilitate such penetration.

The intensified nocturnal ionization associated with auroras was first observed at Tromsø during the second International Polar Year 1930/1 (Appleton et al 1937). It was extensively observed in several ways in polar (and other) regions during the IGY (the *IGY Annals*, 3, 8, 13–15, 17–19, 23 report the ionospheric program and results of the IGY). Often the D region of the ionosphere is affected when the auroral particles penetrate below 90 km; this is well recorded by riometers (Little 1957), which register the degree
of absorption of extraterrestrial radio waves (e.g. of 30 Mc/sec). Radio waves are also reflected when they impinge nearly perpendicularly on the auroral sheets (Booker 1960); such radio echoes are received from regions in sunlight by day, thus somewhat mitigating the difficulty of observing visual auroras in daylight. Doppler shift and spread of such echoes has been studied, and gives information about the motion of the electron concentrations associated with the aurora. At College, Alaska, Nichols (1957) observed such auroral echoes, using low-power continuous-wave transmissions, one directed 30°W of magnetic N, the other 30°E, and obtained Doppler shifts of opposite sign on the two antennas; his measures indicated motion along an east-west line, \( \sim 700 \) m/sec, generally eastward before midnight and westward after midnight. Such speeds cannot be due to wind, and are interpreted as indicating the motion of secondary electrons in the auroral sheet, moving along the sheet, and hence carrying an electric current in the opposite direction. Owing to the anisotropy of the electric conductivity engendered by the presence of the geomagnetic field, a westward current would be produced by a component electric field directed nearly equatorward (Akasofu 1960).

The proton precipitation occurs in general over a band broader than that of the electron precipitation; this band is not far outside the border of the auroral oval. The proton aurora is subvisual; its photons are those produced by the entering particles, in the repeated intervals during their descent in which they have temporarily acquired an electron, and thus become able to radiate.

The heating in the case of the thermal aurora has been variously attributed to atomic resistance to electric current flow in the ionosphere, to electric fields, to conduction from the magnetosphere (Cole 1969), and to absorption of hydromagnetic waves from the magnetosphere (Krassovsky 1968). Thermal auroras include the stable red arcs observed in middle latitudes (Barbier 1958, 1960; Roach & Marovich 1959).

**The Events Preceding the Entry of Auroral Particles**

According to Alfvén’s theory of the motion of trapped particles in the magnetosphere, they travel northwards and southwards between mirror points, but some have mirror points low enough for them to collide in the ionosphere and become part of it; these may be considered auroral particles. Other auroral particles not in the trapping region may enter the ionosphere along field lines that extend out into the tail of the magnetosphere. The auroral conjugacy suggests that these must come from near the median plane of the tail, where about equal numbers are moving northward and southward.

The electrons that cause the auroral arcs, as they descend to the ionosphere, must constitute thin sheets, and while the arcs are quiet these sheets
must be stable. At the auroral breakup, when the arcs become rayed and pleated and thinner, some instability is indicated, apparently of the kind studied in the laboratory by Webster (1957).

Some remaining fundamental questions concerning the events prior to the entry of the auroral particles into the atmosphere involve problems of magnetospheric and general plasma physics beyond the scope of this article. Do the auroral particles come from the Sun, as Birkeland, Störmer and others supposed, or are they particles of the Earth’s outer atmosphere, which extends far out into the magnetospheric cavity, and in some way influenced and energized by the changes in the solar wind and its magnetic field? The energy of the electrons in the solar wind is of order 25 eV (Hundhausen 1968); if any of them become auroral electrons, they must be energized in the cavity.

To what extent particles from the solar wind enter the cavity is not known. Axford & Hines (1961; see also Nishida 1966 and Brice 1967) suggested that some of them diffuse into the cavity, especially along its sides, and set up circulatory motion therein. Axford (1967; see also Dungey 1968, Piddington 1968) proposed that solar particles accumulate in the tail and inflate it increasingly until it becomes unstable, with quick release of its excess energy in the form of a magnetospheric storm or substorm. Akasofu (1964a) proposed that the solar wind includes a varying proportion of uncharged particles, whose entry into the cavity is unimpeded, bringing solar energy into it.

The selective process that restricts the entry of auroral electrons to thin sheets, often multiple, is not understood. Akasofu & Chapman (1961) proposed that these electrons come from regions closely bordering on neutral lines of the geomagnetic field; they speculated that these might lie in the ring current of the trapping region, but this is now seen to be unlikely (Parker 1962); neutral lines will, however, occur near the median plane of the tail region (Coppi, Laval & Pellat 1966).

Akasofu has estimated the energy of a substorm as $\sim 10^{22}$ ergs, and for a duration $\sim 3 \times 10^4$ sec the average rate of energy supply needed is about $3 \times 10^{18}$ ergs/sec. The kinetic energy flux in the solar wind has a variation from time to time of more than tenfold; taking a value for it of $10^{-1}$ erg/cm$^2$ sec, this, over a cross-sectional area $10^{20}$ cm$^2$, comparable with that of the cavity, gives a supply rate $10^{19}$ ergs/sec. But much of this must be diverted round and pass with the solar wind away from and beyond the Earth.

Schuster (1911), in a discussion of theories of magnetic storms, concluded that the energy must come from that of the Earth’s rotation (cf Krassovsky 1968), but it is not clear how this energy can be drawn upon.

Dungey (1968) attaches importance to interconnection of the magnetic field of the solar wind and the field within the cavity (cf Petschek 1964, Piddington 1968, Sweet 1963), proposing that this may provide a channel
for the flow of solar particles and energy into the cavity. Electric fields and magnetospheric instabilities have been considered by various authors (Fejer 1964, Kern 1962, Speiser 1968, Swift 1968, Taylor & Hones 1965), in discussing the origin and nature of magnetospheric storms.

Further observational discoveries concerning the solar wind and the magnetosphere (Ness 1968) will doubtless give guidance in the search for the true chain of events and causes that produce auroras, but it seems clear that many difficult theoretical questions will also be involved.
Alfvén, H. 1950, *Cosmical Electrodynamics* (Oxford)
Birkeland, K. 1908, 1913, *Norwegian aurora polaris expedition*, 1902–3, Parts 1 and 2 (Christiania, Oslo: Aschehoug, 801 pp.)
Cole, K. D. 1969, *IAGA Symp., Madrid*
Davis, T. N. 1967, *Aurora and Airglow*, 133 (Amsterdam: Reinhold)
(Moscow: Publ. House Acad. Sci.)
Fritz, H. 1873, *Verzeichniss beobachteter Polarlichter* (Vienna: Gerold’s Sohn)
Gartlein, C. W., Gartlein, H. E. 1966, *Northern Hemisphere Synoptic Charts ... IGy Ann.*, 39
Gassendi, P. 1658, *Opera omnia* (Lyons, France)
Gold, T. 1959, *J. Geophys. Res.*, 64, 1219
Green, A. E. S., Barth, C. A. 1965, *J. Geophys. Res.*, 70, 1083
Harang, L. 1946, *Geophys. Publ.*, 16 (13)
Hulbert, E. O. 1929, *Phys. Rev.*, 34, 344, 1167; 36, 1560
Krassovsky, V. I. 1959, *Planet. Space Sci.*, 1, 14
Lindemann, F. A. 1919, *Phil. Mag.*, 38, 669
Little, C. G. 1957, *IGy Ann.*, 3, 207
McInnes, B. 1962, *IGy Visoplots, IGy Ann.*, 29
Mishin, V. M., Popov, G. V. 1969, *IAGA Madrid Assembly*
Muncke, G. W., 1833 *Gehler’s Physikalische Wörterbuch*, 2nd ed. 1823–1845, 7(1), 113
Parker, E. N. 1962, *Space Sci. Rev.*, 1, 62
(Heiss, W. N., Ed.)
Roach, F. E., Marovich, E. 1959, *J. Res. NBS*, 63D, 297
Space Sci., 15, 1695
Sandford, B. P. 1962, J. Atmos. Terr. Phys., 24, 155
Schwabe, S. H. 1844, Astron. Nachr., 21
Seaton, M. J. 1955, The Aurora and Airglow
(Dalgarno, A., Armstrong, E. B. Eds., Pergamon)
Shchov, N. N. 1963, Planet. Space Sci., 10,
73
Spitzer, T. W. 1968, J. Geophys. Res., 73,
1112
Stoffregen, W. 1962, IGY Ascarts, IGY Ann., 20
Swift, D. W. 1968, Planet. Space Sci., 16,
329
Taylor, H. E., Hones E. W., Jr., 1965, J.
Geophys. Res., 70, 3605
Pap.
Vaisberg, O. L. 1959, Aurora and Airglow,
Rep. IYG Comm., 10, 454 (Moscow: Acad.
Sci.)
Vallance Jones, A. 1969 (See McCormac
& Omholt, p. 47)
Am. Phys. Soc., 1, 230
183, 430
Vegard, L. 1930, Z. Geophys., 6, 42
Vegard, L. 1939, Nature, 144, 1089
Vernov, S. N., Chudakov, A. E., Valukov,
P. V., Logachev, Yu.I., Nikolaev, A. G.
1960, Artificial Earth Satellites, 5(24)
(Moscow; translation in: 1961, ARS J.,
31, 967)
Elec., 49, 77
Vestine, E. H., Snyder, E. J. 1945, Terr.
Magn. Atmos. Elec., 50, 105
Webster, H. F. 1957, J. Appl. Phys., 28,
1388
White, F. G. W., Geddes, M. 1939, Terr.
Magn. Atmos. Elec., 44, 367
Handl., 273
Wilson, C. R. 1969, Planet. Space Sci., 17,
1817
Wilson, C. R. 1969a, J. Geophys. Res., 74,
(7), 1812
Winckler, J. R., Peterson, L., Arnoldy, R.,
Winckler, J. R., Peterson, L., Arnoldy, R.,
Hoffman, R. 1959, J. Geophys. Res., 64,
597