11 Aurorae

M. Blanc and J. Mäkinen

11.1 The physical phenomenon

11.1.1 The nature of the aurorae

Anyone who has had the opportunity of travelling to high northern or southern latitudes during the long winter night, and has discovered the polar aurorae for themselves has undoubtedly been left with an indelible impression. Because they are usually too faint to be seen in the presence of city lights, you have to go to a really dark, remote place, far from any urban activity, before you are able to see their true beauty. They then appear in an incredible variety of forms, from a very faint, diffuse, greenish patch to a spectacular show of coloured curtains dancing in the sky, apparently set in motion by an invisible wind (Fig. 11.1). If you are lucky enough to see the aurora (which just takes a little bit of patience and effort), you can expect a wonderful experience: the beauty of the aurora is enhanced by the darkness of the night, the silence of the polar wilderness and, very often, the cold that you have to withstand to watch the show until it comes to a (temporary) halt.

Although the true beauty of the aurorae can perhaps be appreciated only from high latitudes, it is certainly wrong to suppose – as the vast majority of people imagine – that it cannot be seen elsewhere. It is certainly visible (although more rarely) at lower geographical latitudes. The aurora is associated with the Earth's magnetic field, and has even been seen at the geomagnetic equator, so essentially anyone might see an aurora, wherever they are on Earth, although it could be a once-in-a-lifetime experience. Events such as the great auroral storms of 1938 January and 1989 March covered such a vast area and were so spectacular that they generated considerable interest among the public and the news media. In any case, amateur astronomers, who naturally seek out dark skies for their other observing programmes, are ideally placed to observe any aurorae that do occur. But what is the physical nature of the aurora?

The aurora has been known since ancient times, and has attracted the curiosity of some of the great names in the history of science and astronomy, such as Aristotle and Seneca (who both personally witnessed the appearance of an aurora), and later Gassendi, Galileo Galilei and Edmond Halley. The first use of the term 'aurora borealis' (northern dawn) is commonly attributed to the French astronomer Gassendi, although recent research suggests that his famous 17th-century contemporary Galileo may have been the first to use that form of words. It is also possible that Gregory of Tours used the description aurora borealis as early as the 6th century.

The first scientist to have an approximately correct idea of the origin of the aurora was the French scientist Jean-Jacques Dortous de Mairan (1678–1771), whose paper *Traité Physique et Historique de l'Aurore Boréale*, published in 1733 by the French

Fig. 11.1. *A typical auroral display occurring during the polar night. (Photo: Heikki Ketola, Ursa Astronomical Association)*

Académie des Sciences, was the very first monograph on the aurora. De Mairan interpreted the aurora as the result of the interaction between the Earth's upper atmosphere and the extended atmosphere of the Sun (what we now call the corona), along the surface in space where the gravitational forces of the Sun and the Earth are in balance. If a 'bubble' of solar atmosphere were pushed over to the Earth's side of this boundary, it would fall into the Earth's atmosphere, and its reaction with the constituents of the terrestrial atmosphere would produce the fire that was seen as the aurora.

De Mairan was obviously wrong about the physical nature of the interaction between the solar and terrestrial atmospheres, but this was inevitable in his day, because electromagnetism had yet to be discovered. He was right, however, in claiming that this interaction was the primary cause of the aurora.

How do we understand and describe this interaction nowadays? Taken as a local phenomenon, the aurora is rather simple (Fig. 11.2). Energetic electrons originating in the Sun's atmosphere move along the geomagnetic field lines. Their energies range from a few hundred eV to hundreds of keV, with most of the population having energies between approximately 500 eV and 10 keV. At high altitudes, say above about 400 km, because of the magnetic forces acting on charged particles, these electrons move in helical trajectories around the Earth's magnetic lines of force. When these electrons reach lower altitudes, they move through a progressively denser and denser atmosphere, because the density of the Earth's atmosphere, like that of any planet, increases exponentially with decreasing altitude. The approximate density at any altitude is given in Fig. 11.3, which shows, for example, that at 400 km there are between 10^6 and 10^8 atoms or molecules per cubic centimetre. The precise value depends on the exospheric temperature, i.e., the temperature of the upper atmosphere above 200 km, which varies with solar activity. At 100 km, however, this density has increased to over 10^{12} atoms or molecules per cubic centimetre. The probability of a collision between an electron that is moving down a magnetic field line and the ambient atoms and molecules therefore increases exponentially with decreasing altitude. At a certain height, a collision becomes inevitable; lower down, collisions become more and more frequent, until the incident electron finally loses all its kinetic energy.

Where does that energy go? A small fraction is transformed into kinetic energy of the atoms and molecules with which the electron has collided. Most of the energy, however, goes to other electrons, namely the electrons in the electronic shells of those atoms and molecules, because in collisions the maximum amount of energy is exchanged between particles of similar masses. Some of the electrons in the shells receive enough energy to escape from their atom or molecule. They become free electrons, leaving behind newly created positive ions.

The X-ray and ultraviolet components of ordinary sunlight carry enough energy to eject electrons from the outer shells of atmospheric oxygen and nitrogen, giving rise to populations of ions and electrons at altitudes greater that 80 km, and thus forming the ionosphere. Radio communications over long distances exploit the reflective properties of ionospheric layers. During auroral activity, localized clouds of increased ion and electron concentration may disrupt, or sometimes enhance, the propagation of radio signals.

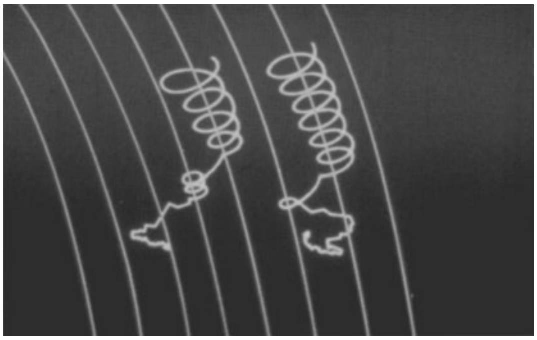

Fig. 11.2. *The basic mechanism generating the aurora. An energetic electron originating in the Sun precipitates into the Earth's upper atmosphere after spiralling round the lines of the Earth's magnetic field, which are nearly vertical over the polar regions. Collisions with atoms and molecules modify its trajectory, slow it down, and transfer energy into excitation of those atoms and molecules, which later release this excitation energy by radiating a photon. This causes the auroral emission.*

Most of the energy from the incoming electrons is transferred to electrons that do not completely escape from their atoms or molecules, but simply move to a different shell with a higher energy level. This produces an excited atom or molecule, a physical system that is always unstable, and which tends to return to a state of lower energy at some later time. When this happens, the excited electron drops to a lower energy level and the excess energy is carried away by an emitted photon. The photon's energy corresponds to one of the characteristic, spectroscopic emission lines of the emitting atom or molecule.

The aurora is thus a set of emission lines of the components of the Earth's upper atmosphere excited by incoming electrons of solar origin. (A somewhat similar process, involving excitation by sunlight, and lower particle energies, produces the much fainter phenomenon known as airglow.) Examination of a typical auroral spectrum (Fig. 11.4) shows that it consists of a few major lines and bands produced by transitions of molecular nitrogen, atomic nitrogen, and atomic oxygen, which are the main components of the Earth's upper atmosphere. The group of blue and green lines and bands below 500 nm is mainly caused by permitted transitions of molecular nitrogen N_2. They are the main component of auroral emission below 100 km, in the lower part of the aurora. But the most intense auroral emission lines are primarily the yellow-green lines at 557.7 nm and then the red line at 630.0 nm, both caused by an excited state of oxygen, O I. There is a special story about these two lines. During early studies of the auroral spectrum, considerable efforts were made to identify the components of the spectrum. If the aurora were just atmospheric airglow, the auroral lines obviously had to lie at the known wavelengths for emission from the main components in the atmosphere, oxygen and nitrogen.

The mystery arose because the lines at 557.7 and 630.0 nm just do not exist in the atmospheric spectrum as seen at the surface of the Earth, or measured in a

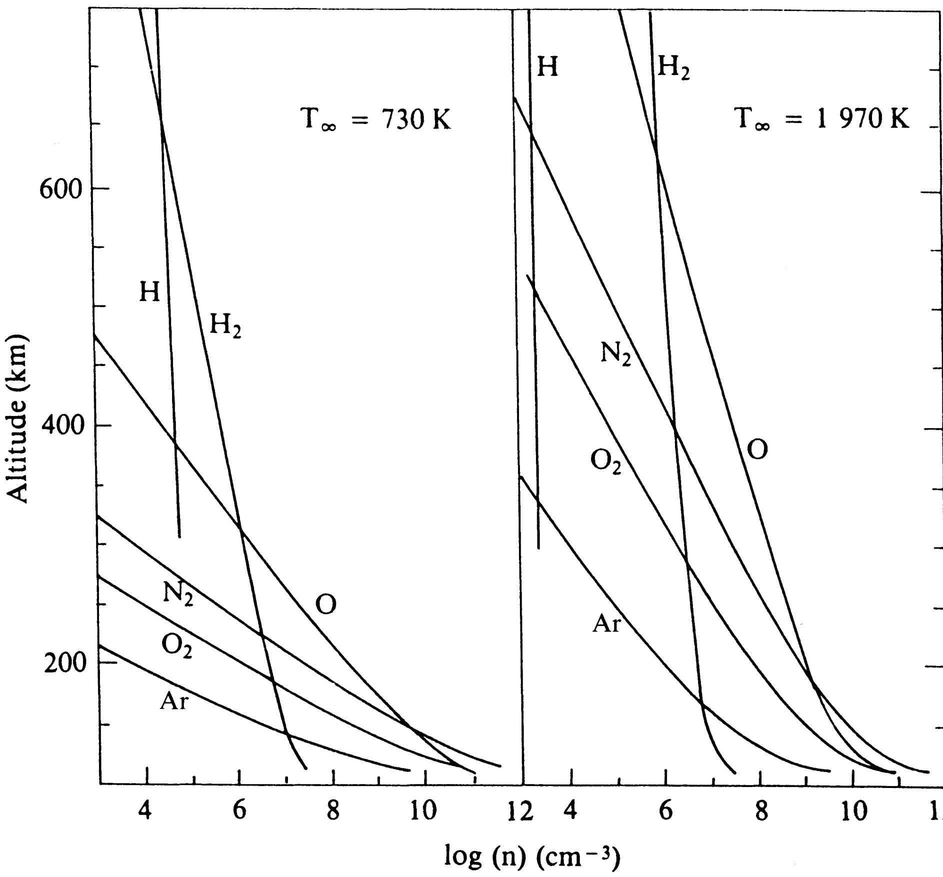

Fig. 11.3. *Typical vertical variation of the main components of the Earth's upper atmosphere. (After Giraud & Petit, 1978)**

laboratory! The reason is that these lines are so-called 'forbidden' lines. They are produced by a special type of excited state of oxygen, known as a 'metastable state', which has a very long lifetime before it is de-excited by the emission of a photon. These long lifetimes – 0.75 s for the green line and 110 s for the red line – may be compared with the typical lifetime of about one millionth of a second for the other emission lines. Because of their long lifetimes, these metastable states have absolutely no chance of radiating a photon at ground level, because the excess energy is lost through collisions with surrounding molecules in the dense atmosphere long before it can be radiated away. It is only in a very rarefied medium, where the time separating successive collisions with neighbouring molecules is longer than the lifetime of these metastable species, that they have a chance of radiating a photon.

At auroral heights, the atmospheric density is sufficiently low for the forbidden emissions to be produced. The 577.7 nm oxygen emission occurs above 95 km

* Notes references and bibliography to volume 2 commence on p. 1141

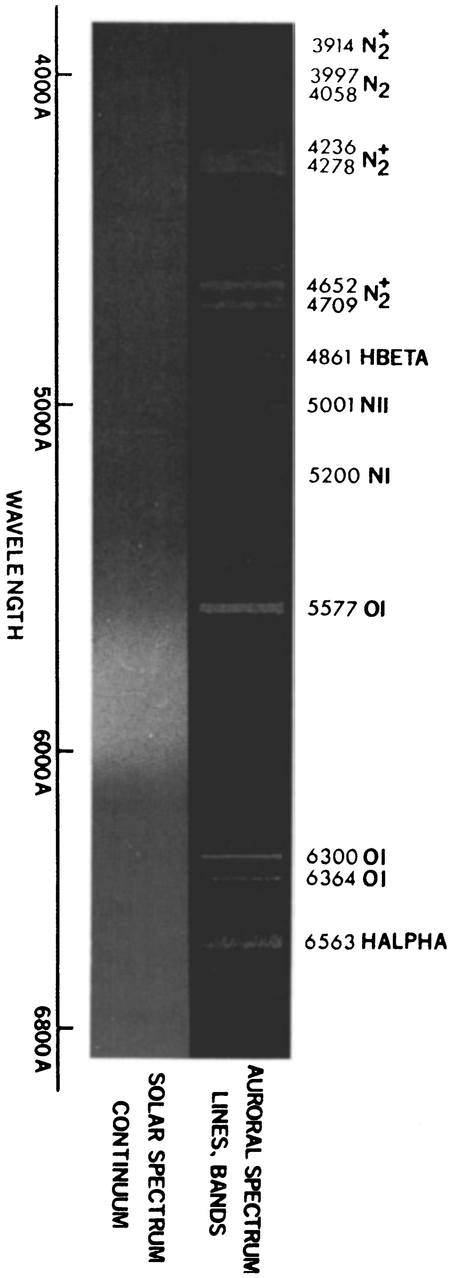

Fig. 11.4. *The spectrum of the aurora in the visible region showing the locations of the main emission lines and bands, and indicating the atom or molecule responsible for each emission. (Photo: A. Vallance Jones, in Eather,* Majestic Lights, *1979.)*

Table 11.1. *Typical altitudes reached by auroral electrons*

Electron energy (eV)	Minimum altitude (approx.) (km)	Emission	
100	300	Red	(630.0 nm)
1000	130	Green	(557.7 nm)
20 000	95	Blue & purple	

altitude. Red oxygen emission at 630.0 nm occurs higher in aurorae, at altitudes in excess of 250 km, but is suppressed (quenched) at lower altitudes. In bright aurorae, the rays may therefore show colour differences along their length. It was only in 1923 that this phenomenon was understood, and therefore that the main emission lines of the aurora were thus definitely identified as produced by atmospheric oxygen.

As may be seen from its spectrum, the aurora may display a fantastic variety of colours, with its emission lines including violet, blue, green, yellow and red. But these different colours are produced at different altitudes as we have just seen. (Although auroral colour is largely controlled by the energy of the incoming electrons, with the more energetic electrons penetrating deeper before they are stopped by collisions, the condition of the atmosphere also plays a part, such as whether the region is actually in sunlight. For example, purple N_2^+ rays occur in sunlight.) A few typical electron energies, and the altitudes at which they produce auroral emissions are given in Table 11.1. Conversely, when we look at an aurora, we can guess from its colours the average energy of the electrons producing it. Because the typical energy of auroral electrons is a few keV, the most common colour is yellow-green. The altitude of maximum emission, which has been studied for several decades by triangulation techniques, averages 107 km for yellow-green aurorae.

11.1.2 The global phenomenon: aurorae and the magnetosphere

Despite centuries of observations from the ground, a global view of the polar aurora was not achieved until the advent of the space age, when satellites were able to carry cameras and take pictures from space. NASA's Dynamics Explorer-1 satellite, launched in 1981, had an eccentric polar orbit that lay several Earth radii above the poles. This allowed it, for the first time, to observe the global distribution of the aurora over an entire hemisphere. Figure 11.5 shows one of the pictures taken by Dynamics Explorer's UV camera. The ultraviolet emission seen in the photograph comes entirely from the Earth's upper atmosphere, because the lower atmosphere is not transparent to UV light. The upper left-hand portion of the picture shows the crescent of the sunlit side of the Earth, extending from the limb at the top to the day/night terminator at the bottom. But the most spectacular feature is the oval ring of light visible on the dark side of the Earth in the polar region. This luminous ring simply corresponds to the overall distribution of the aurora, as seen from space. Examining its geometry in detail, it is found that the centre of the oval

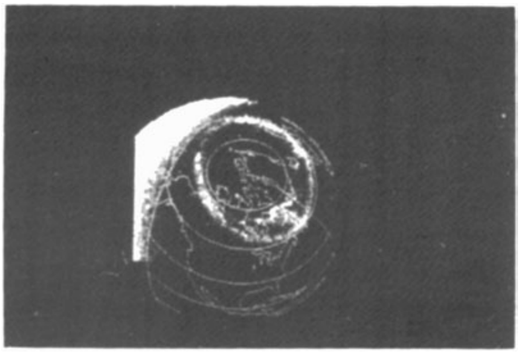

Fig. 11.5. *One of the overall pictures of the auroral oval, taken by the ultraviolet camera on board NASA's* DYNAMICS EXPLORER *satellite. The sunlit side of the Earth is to the left. (Image: Dr L.A. Franck, University of Iowa, Iowa City.)*

is located close to the magnetic pole, but displaced about 500 km towards the night side. Consequently, the aurora lies at lower magnetic latitudes on the night side than on the day side. It is also noticeable that the extent of auroral emission is broader in latitude towards the night side than it is towards the day side.

One of the major findings from DYNAMICS EXPLORER studies is the permanence of the auroral oval. An oval is indeed present at any time around each of our magnetic poles, but the intensity of the emission, and the radius and width of the oval vary with time. We shall see shortly the consequences of this for ground-based observers. But first we have to understand the origin of the auroral oval.

As we have seen, the aurora is produced by the interaction of energetic electrons with the upper atmosphere. It may be likened to the visible image produced by a giant cathode-ray tube, where the screen consists of our atmosphere. Behind – or in our case, above – the two-dimensional reality of the screen there is the three-dimensional electron accelerator that generates and guides the electron beams. In the example of a TV monitor, it consists of electron guns and electrostatic optics. In the case of the aurora, the electron accelerator is the Earth's magnetosphere, and its energy source is to be found in the interplanetary medium and the solar wind. Let us briefly describe how it works.

The magnetosphere is the region of the Earth's environment in which the dynamics of particles are dominated by the Earth's magnetic field. The degree to which the terrestrial magnetic field extends into space is limited by its interaction with the solar wind, which is a continuous outflow of electrons and protons from the Sun, moving at an average speed of 400 km/s, with typical densities of a few particles per cubic centimetre. When the charged, solar-wind particles meet the Earth's magnetic field, they are reflected by it, so the terrestrial magnetic field forms a real obstacle, which compresses and deflects the flow of the solar wind. The geometry of this interaction is represented in Fig. 11.6. This figure shows a cross-section of the magnetosphere in a plane containing the axis of the Earth's magnetic dipole, and the Earth–Sun line. The Sun is to the left, so the solar wind flows from left to right towards the obstacle

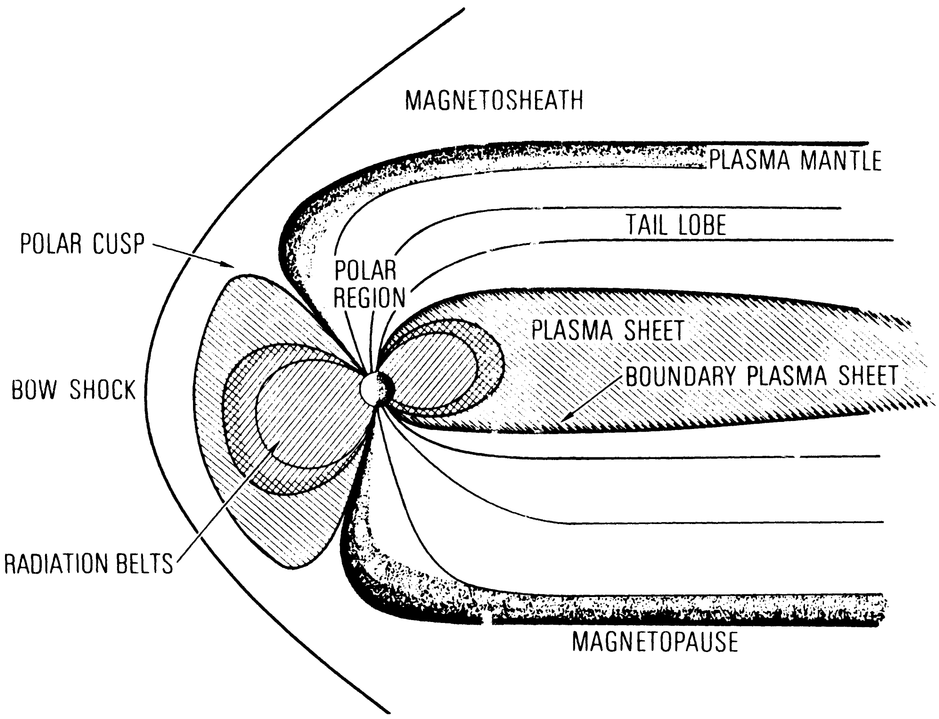

Fig. 11.6. *A cross-section of the Earth's magnetosphere in the plane containing the Sun–Earth line and the axis of the Earth's magnetic dipole. The diagram shows the Earth's bow shock in the solar wind, the magnetic field lines, and the main regions and particle populations of the magnetosphere.*

presented by the Earth's magnetic field. The line farthest to the left represents the 'bow shock'. Because the solar wind is flowing at supersonic speeds with respect to the obstacle, a steady shock forms upstream of the Earth – just like the bow shock in front of supersonic aircraft. Here the solar wind is decelerated to a subsonic value, and at the same time it is compressed. In the region between the bow shock and the magnetopause (which is the external boundary of the magnetosphere), the compressed solar wind flows around the obstacle, before it again accelerates to supersonic speeds. Because of this interaction, the magnetic field of the Earth is confined inside a comet-like cavity, which forms what we call the magnetosphere. A few typical magnetospheric field lines are shown in the figure. In the polar regions, the Earth's field lines emerge from the northern and southern polar caps, before running back in two lobes in the elongated magnetotail. It is well-known that the Earth's magnetotail, like those of all planetary magnetospheres, is very elongated. It extends well beyond the orbit of the Moon, and is probably at least 1 AU long.

Except for the set of magnetic field lines that are connected to the tail lobes, all the other geomagnetic field lines are organised in closed loops which connect approximately symmetrical points on the Earth's surface in the two magnetic hemispheres.

Energetic electrons and ions may become trapped on these closed field lines, and some may be precipitated into the upper atmosphere to produce the aurora. Where do they come from? They come from the solar wind. As may be seen from the figure, there are two special points on the day side of the magnetopause where the magnetic-field intensity may become very small, and thus allow solar-wind particles to penetrate into the Earth's magnetosphere. These regions in the northern and southern hemispheres are called the polar cusps. Solar-wind electrons that drift into the upper atmosphere via the polar cusps have relatively low energies, and so give rise to only diffuse, weak aurorae on the Earth's day side.

Electrons and protons that do enter at the polar cusps subsequently drift towards the night side of the magnetosphere in the lobes, forming a domain in the magnetosphere plasma called the plasma mantle. They then converge towards the central region of the tail, where they accumulate to form a giant reservoir of particles called the plasma sheet. Finally, plasma-sheet particles move back towards the Sun. As they do so they also drift around the Earth, extending the plasma sheet to all longitudes and local times and forming a circular belt that completely encircles the Earth. During this transport, solar-wind particles gain a significant amount of energy. They have an energy of 10–100 eV in the solar wind, and reach several keV in the regions of the plasma sheet that are nearest to the Earth. It is the plasma-sheet particles that form the immediate source of the aurora.

To really understand the connection of these magnetospheric particle domains with the aurora, it is necessary, however, to visualize them in three dimensions, and to see how they project down (along the field lines) onto the upper atmosphere. This is shown in Fig.11.7. The left-hand side shows a slightly different view of the magnetosphere to that in Fig. 11.6, but with the same particle domains. On the right-hand side, each of these particle domains is projected onto a polar map of one terrestrial hemisphere. It will be seen that the projection of the plasma sheet forms an oval, nearly centred on the geomagnetic pole, but slightly displaced towards the night side. It is on this oval that, at any given time, solar-wind electrons may cascade into the atmosphere and excite an aurora. This oval obviously corresponds to the auroral oval shown in the DYNAMICS EXPLORER ultraviolet images.

11.1.3 Spatial and temporal variations in the aurora

Now that we know how the aurora is distributed over the Earth, it is easier to understand when and how it may be seen by an observer on the ground. The general pattern shown by the auroral ovals is essentially steady in a reference frame fixed relative to the Sun. They may expand towards lower latitudes, or contract radially towards the pole, depending on the changing conditions in the solar wind, but they do not rotate with the Earth. Under quiet geomagnetic conditions, the ovals may be regarded as fixed in space above the rotating, solid body of the Earth. Terrestrial observers therefore rotate beneath the auroral ovals. This is illustrated in Fig. 11.8, which, by way of example, shows the motion of Scandinavia under the auroral oval in the course of one rotation of the Earth.

This diagram also shows the variation in the type of aurora that may be seen with local time. When Scandinavia is on the day side of the Earth, the auroral