Chapter 17

Auroral Physics

As described in Lectures 14 to 16, Earth's auroral regions are coupled to the solar wind, magnetosphere, and ionosphere and are therefore the site of a number of space weather phenomena. These include the electromagnetic radiation of the "aurora borealis" and "aurora australis" (the "northern" and "southern lights", respectively), enhanced ionization and conductivity in the auroral ionosphere due to impact ionization processes, enhanced particle precipitation, and the rapidly varying magnetic fields associated with currents flowing in the auroral electrojet regions of the ionosphere. This Lecture contains a brief discussion of the physical processes involved in the production of the auroral lights, the different classes of auroral radiation associated with various precipitating particle populations, the existence of field-aligned currents and potential drops in the auroral regions, auroral substorms, and the auroral electrojets. The environment of the polar cap will also be discussed, including the characteristic size of the potential across the polar cap, the **E** × **B** motion of plasma across the polar cap, and the auroral electrojets.

The **Aim** of this lecture is to explore the physics of planetary auroral regions, focusing primarily on the Earth, and their connections to planetary magnetospheres and space weather.

Expected Learning Outcomes. You should be able to:

- Explain qualitatively what auroral emissions exist and why they are generated.
- Describe and explain qualitatively the different classes of auroral emissions at Earth and how they are related to magnetospheric particles and associated magnetic connections to specific regions of the magnetosphere.
- Explain qualitatively how auroral substorms evolve and how they are related to space weather events and magnetic reconnection.
- Describe the polar cap and auroral electrojets.
- Explain qualitatively how magnetic reconnection and the solar wind magnetic field determine convection patterns in the polar cap and auroral region and so the development of the auroral electrojets.
- Describe qualitatively some of the phenomena associated with currents and potential drops in the auroral zones.

17.1 The Basics of Auroral Emissions

The auroral lights appear as time-varying bands and filaments of light, often of many different colours, often moving about the sky, often showing structures oriented in

the vertical direction (qualitatively consistent with the orientation of magnetic field lines), and showing structures that are discrete (such as auroral arcs). The auroral lights do not always move, however, but sometimes are very steady and slowly varying. Green and red displays are particularly common. Examples of the auroral lights can be found in Figure 15.11, as well as in Cravens' [1997] book, the paper of Carlson and Egeland [1995], and on the web sites given to you.

Auroral emissions are produced as a result of energetic electrons and protons from the solar wind and Earth's magnetosphere colliding with the constituents of Earth's upper atmosphere and ionosphere. Typically, the emission process involves multiple steps: first, an atmospheric atom, molecule, or ion X is placed in an "excited" electronic state X^* by a collision with a precipitating electron or ion,

$$X + e \to X^* + e \tag{17.1}$$

in which the electron loses kinetic energy in the collision; second, the entity X^* either relaxes to the ground state by emitting a photon

$$X^{\star} \to X + h\nu , \qquad (17.2)$$

relaxes to a lower excited state by emitting a photon, or transfers excess energy to another atmospheric/ionospheric constituent via a collision, etc. Vibrationally excited states are also produced, leading to infrared emission. Note that these deexcitation photons have frequencies and lifetimes that are predicted very accurately by quantum mechanics. The precipitating particle can also create additional plasma particles by impact ionization processes, emit X-rays via bremsstrahlung, undergo chemical recombination reactions, heat the atmosphere and ionosphere etc.

The spectrum of Earth's auroral emissions is composed primarily of many spectral lines and bands in the infrared, the optical band, and the far ultraviolet (Figure 17.1). Bremsstrahlung X-rays are also detected. These spectral lines and bands allow the chemical composition of the atmosphere and ionosphere to be probed remotely, as well as allowing constraints to be placed on the nature, spectrum and characteristic energies of the precipitating particles. For instance, knowing the physics of (say) electron energy loss to certain chemical species and the energy levels of these chemical species, the spectrum, altitude profile, and intensity of radiation can be predicted theoretically and then compared with observations to constrain the electron energy spectrum and/or models for the atmosphere and ionosphere [e.g., Omholt, 1971; Vallance Jones, 1974; Carlson and Egeland, 1995]. These types of calculations are useful for Earth, but also for the recently discovered Jovian auroral displays [Connerney et al., 1996] where detailed comparisons with in situ data are only now becoming possible due to the Galileo spacecraft.

For definiteness, consider the following lines. The brightest visible feature of Earth's aurora is the so-called "green line" at 557.7 nm. This is due to the transition of an electron from the ¹S excited state of atomic oxygen to the ¹D state (Figure 17.2) The 630.0 nm line is a "red line", produced as the ¹D state relaxes to the ground state, that is often observed at high altitudes over the cusp and polar cap (where the atomic collision time is long compared with the de-excitation time). Note that both these lines correspond to relatively small excitation energies $\leq 4 \text{ eV}$ and so correspond to collisions by relatively slowly moving particles.

The heights at which aurorae are typically formed are between 90 and 200 km, thereby corresponding to the thermosphere and ionosphere (Figures 16.1, 16.2, 16.6 and 16.7). Characteristic intensities range from a few tenths to more than 10 kilo-Rayleighs (kR).

The energies of the precipitating electrons and protons are discussed more below. For now, however, it is enough to state that those originating in the solar wind/magnetosheath have characteristic energies of order 100 eV, those starting in



Figure 17.1: Selected portions of auroral spectra in the visible [Vallance Jones, 1974], showing the discrete nature of the auroral lines.

$$\frac{4,17}{557,7} = \frac{1}{5}, \tau = 0.8 \text{ s}$$

$$1,96 = \frac{1}{630,0} = \frac{1}{636,4} = \frac{1}{39}$$

Figure 17.2: Energy level diagram for atomic oxygen [Carlson and Egeland, 1995]. The radiative half-lives τ and wavelengths of the emitted photons are also shown.

the plasmasheet have energies of order 1 keV, while those found in the nightside auroral zone itself have energies in the range of $\sim 2 - 20$ keV.

17.2 Classes of Auroral Light and Associated Precipitation

Figure 17.3 schematically illustrates the local time and geomagnetic latitude dependences of the observed rate of precipitation of auroral particles into the upper atmosphere and the different associated classes of auroral emissions [Hartz, 1971; Carlson and Egeland, 1995]. The average flux is represented schematically by the density of the dot, triangle and star symbols, measured during multiple spacecraft transits through these regions.

The star symbols represent low energy polar cusp plasma that originates in the magnetosheath and solar wind and has characteristic energies of order 100 eV. As expected from Lecture 14, the polar cusp precipitation region is on the dayside at relatively high geomagnetic latitudes ~ 75 degrees. These precipitating particles form "dayside cusp auroras" [Carlson and Egeland, 1995]. These are typically sub-visual and relatively weak, with maximum intensities in atomic rather than molecular lines due to their generation at larger heights than other aurorae, typically in the F layer of the ionosphere. Dayside cusp aurorae apparently tend to be relatively uniform and diffuse, without arcs or localized fine structure but are relatively tightly confined in local time to near noon. Cusp aurorae usually change substantially in location and intensity in response to variations in the solar wind B_z component. For $B_z < 0$ (and so greater efficiencies for dayside magnetic reconnection and better connection of the cusp to the solar wind), cusp aurorae broaden, extend equatorward and brighten considerably. This is understandable in terms of Figure 15.5 and the decrease of closed magnetic field lines on the front side of the Earth due to magnetic reconnection for southwards B_Z . Sometimes this dayside motion equatorward coincides with the nightside aurora moving poleward, as expected for increased loading of the tail, but not always.

The dot symbols in Figure 17.3 represent where high energy (> 20 keV) auroral particles precipitate. These high energy particles precipitate on a circle with constant geomagnetic latitude ~ 65 degrees, as expected for trapped particles leaking out of the loss cone as they drift around Earth in the ring current. It turns out that these high energy particles are relatively ineffective in producing auroral emissions.

The triangle symbols in Figure 17.3 show where medium energy (0.2 - 20 keV)



Figure 17.3: Schematic illustration of the patterns of particle precipitation into Earth's upper atmosphere and the different classes of auroral emissions [Hartz, 1971; Carlson and Egeland, 1995], as functions of local time and geomagnetic latitude. The direction to the Sun is at local noon (12 hours) while the center of the tail lies at local midnight (0 hours). The radial direction corresponds to geomagnetic latitude; the geomagnetic pole is at the cross symbol. The figure is described more in the text.

particles precipitate and produce the majority of visual aurorae at Earth. The footprint made by these particles is magnetically connected to the plasmasheet and the plasmasheet boundary layers, as discussed in Lectures 14 and 15. Note that the particles precipitate onto an oval shape, not a circular shape, due to the influence of magnetic reconnection on the magnetic field lines connecting to the plasma source regions. This oval shape is the so-called auroral oval where aurorae are observed, as verified in Figure 17.4.

At least two classes of auroral activity are associated with the medium energy plasmasheet particles in Figure 17.3 and 17.4. First, there is the so-called "diffuse" aurora, which is fairly weak and smooth and is present relatively continuously around most of the auroral oval (except near noon). This emission is just due to particles precipitating from the plasmasheet and its boundary layer during their bounce motion, either due to their being in the loss cone or being scattered into the loss cone (see Lectures 1, 3 and 4, plus section 17.6 below).

The second class is due to "discrete auroral arcs" which are typically bright, localized structures that are primarily (but not always) produced during the auroral brightenings that accompany magnetospheric substorms. These structures are usually elongated in the east-west direction (along the auroral oval) but are narrow



Figure 17.4: Variation in the size and location of the auroral oval with geomagnetic activity [Feldstein and Starkov, 1967; Carlson and Egeland, 1995]. The shaded areas show the spatial regions with maximum auroral activity, averaged over many episodes, in the northern hemisphere. The coordinate system is similar to that in Figure 17.3.

in the north-south direction (normal to the oval). These occur primarily in the nightside auroral oval and typically vary rapidly with time. Spacecraft observations show that the electrons responsible for discrete arcs originate in the plasmasheet but are energised by passing through a potential drop of several to a few tens of keV, showing up as very field-aligned beams with energies of 2 - 20 keV.

Auroral displays associated with electrons and with protons are observed at Earth, having different spectral properties and typically different locations too. Typically, however, electron-driven auroras dominate at Earth. This need not be the case at other planets, or in stellar magnetospheres.

17.3 Auroral Substorm Displays

An auroral substorm may include large variations in colour, intensity and location, many substructures, and may look completely disordered to the casual observer. Looking at many auroral substorms allows one to see that a pattern exists [e.g., Akasofu, 1968; Carlson and Egeland, 1995; McPherron, 1968]. Figure 17.5 illustrates this pattern schematically. For an observer near the usual auroral oval, the auroral display usually starts with one or more arcs with low intensity (1–10 kR) elongated in the geomagnetic east-west direction (panel A). The first disturbance is a sudden brightening of the most equatorward arc in the pre-midnight sector (panel B). This represents the "onset" of the auroral substorm, as also shown in magnetometer data used for the auroral AE and AU activity indices. The brightening extends rapidly westward and poleward (panel C), rapidly forming a broad region around the midnight sector where the aurora first brightened. This region is very dynamic with arcs appearing and disappearing, brightening, folding etc. The auroral brightness may now be several hundred kR. This region continues to expand poleward (panel D) and both westwards and eastwards, corresponding to the expansion phase of the auroral (and magnetospheric) substorm. Eventually the aurora ceases to expand polewards (panel E) and starts to dim and become more homogeneous, ending the expansion phase and starting the recovery phase. After about an hour auroral activity dims at lower latitudes and quiet arcs reappear, being called the recovery phase (panel F). Typically this process takes about 1-2 hours, with about 3 hours between auroral substorms (or breakups) during geomagnetically active periods.

Figure 17.6 illustrates this pattern with a sequence obtained from the Dynamics Explorer 1 spacecraft [Frank and Craven, 1988; Carlson and Egeland, 1995]. Each picture is separated by 8 minutes.

17.4 Polar Cap and the Auroral Electrojets

The polar cap is a region in which the terrestrial field lines are magnetically connected to the solar wind. Accordingly the solar wind's convection electric field is mapped across the polar cap, causing there to be a potential induced across the polar cap. Since the solar wind electric field is typically in the dawn-dusk directions, the potential is also in the dawn-dusk directions. The size of the polar cap, the extent of the region with open field lines, and so the size of polar cap potential depends on the amount of magnetic reconnection occurring at the magnetopause. Accordingly the polar cap potential is larger when the IMF B_z component is southwards. Measured values range from about 20 kV in quiet times to about 150 kV in active times.

Due to its magnetic connection to the solar wind and magnetosheath, the polar cap is where energetic particles and mildly superthermal plasma from the solar wind



Figure 17.5: Schematic pattern of an auroral substorm [Akasofu, 1968].



Figure 17.6: An example of an auroral substorm [Frank and Craven, 1988].

and magnetosheath can be observed. This is why Earth's magnetic poles are good places to observe cosmic rays, for instance; the drawback, on the other hand, is that workers obtain higher radiation doses at higher magnetic latitudes. Precipitation of these solar wind wind and magnetosheath particles into the upper atmosphere, and associated generation of waves by beams or loss cone distributions of mirrored particles leads to additional coupling between the ionosphere and the solar wind over the polar cap.

Figure 17.7 [Hughes, 1995] shows how plasma and the frozen-in field lines move in the magnetosphere and across and around the polar cap during times of southward IMF and enhanced magnetic reconnection near the nose of the magnetopause. Note that the plasma flows anti-sunward across the polar cap due to the (reconnected) magnetic field lines being in the solar wind, so that they are dragged across the polar cap. This motion is, in one sense, opposite to that implied by simply considering the $\mathbf{E} \times \mathbf{B}$ drift for plasma and associated magnetic field lines in the tail, which must move sunwards from the nightside to the dayside. Note that the sunwards-convecting plasma also undergoes ∇B and curvature drifts around the Earth (as for the ring current). [Alternatively this could be thought of, loosely, in terms of the plasma being deflected by the high-B region close to the Earth.] This allows both the anti-sunwards flow across the polar cap and the sunwards $\mathbf{E} \times \mathbf{B}$ drift to be accomodated, because now the plasma and field lines convect from the nightside to the dayside by forming return paths on the dusk and dawn sides at



Figure 17.7: Flow of plasma and associated frozen-in field lines in the magnetosphere and across the polar cap as a result of magnetic reconnection at the magnetopause for southwards IMF [Hughes, 1995]. The return paths of the anti-sunwards plasma flow across the polar cap lead to the auroral electrojets.

lower latitudes. The flow pattern thus has two cells. These return paths give rise to the so-called "auroral (or convection) electrojets" responsible for the magnetic perturbations caused by auroral activity.

It should be questioned why these convection patterns give rise to any currents. The reason for currents arising is due to the weakly collisional nature of the ionospheric plasma: both electrons and ions experience the magnetic and electric forces, leading to $\mathbf{E} \times \mathbf{B}$ drift but the more frequent collisions of ions rather than electrons with atmospheric neutrals and the smaller electron mass lead to the electrons undergoing the $\mathbf{E} \times \mathbf{B}$ drift while the ions have a much smaller net drift, thereby leading to a current flowing in the $-\mathbf{E} \times \mathbf{B}$ drift direction. The currents are concentrated in the auroral oval due to the much higher conductivity there (due to higher plasma densities), thereby leading to currents flowing in the auroral electrojets.

Finally, note that the directions of the convection velocities and currents reverse when the IMF reverses directions. This implies that substantial variations should exist in the magnetic fields observed at auroral, polar cap, and mid latitudes during space weather events, as indeed described in Lecture 15 and associated references. Figure 17.8 illustrates these differences schematically [Cravens, 1997] but in both cases for $B_z < 0$: for $B_z > 0$ the directions in the lefthand image reverse and the righthand image is irrelevant. Note that the current in the electrojets during substorms is considerably larger and more concentrated near midnight than during quiet times, with the direction primarily imposed by the ∇B drift of energetic electrons coming from the magnetotail reconnection sites.



Figure 17.8: The different directions of plasma and current flow in the auroral electrojets during quiet times ("convection electrojets") and substorms for $B_z < 0$ [Cravens, 1997].

17.5 Auroral Zone Physics

The nightside auroral zone is full of exotic, exciting and often unexplained phenomena. The aim of this Section is to introduce you to several of these phenomena, most of which are relevant elsewhere in space physics or astrophysics or plasma physics, and not to explain them.

Large field-aligned currents are common in the auroral zone, as shown in Figure 17.9 [Ijjima and Potemra, 1978], thereby connecting the ionosphere and the magnetosphere. These currents are usually inferred from magnetometer data using Ampere's Law $\nabla \times B = \mu_0 \mathbf{J}$. Note that the currents flow down on one side of the auroral oval and then up on the other side, requiring current to flow around and across the auroral oval as discussed above in the context of the auroral electrojets. Figure 17.10 shows that the flux of downgoing electrons typically peaks in the current regions of the auroral zone and is small over the polar cap. The figure also shows that electric fields exist pointing into and out of the auroral oval. Moreover, consideration of the field between 1138 UT and 1146 UT over a spatial distance of order $1R_E$ yields a potential ~ 40 mV m⁻¹ × 6 × 10⁶ m ~ 250 kV across the polar cap. This is the solar wind convection electric field mapped across the polar cap.

An important result that is still not understood in detail is the existence of relatively steady, large (several to tens of kV), field-aligned potential drops in the downward and upgoing current regions. Why is this surprising? Because the plasma is essentially collisionless and so electron motion along the magnetic field would normally be expected to short out any potential drops , i.e., field lines are usually



Figure 17.9: Pattern of field-aligned (Birkeland) currents for relatively quiet conditions [Iijima and Potemra, 1978].

expected to be equipotentials.

Figure 17.11 shows typical results from the FAST spacecraft [Ergun et al., 1998a]. Note that the electrons are downgoing and close to mono-energetic, indicative of acceleration by a potential. Similarly, the ions are almost mono-energetic after about 49:50 UT, indicative of acceleration by a potential below the spacecraft. Note that this acceleration was by a parallel potential that endured for tens of seconds, much longer than required for normal electron motions to short out the potential [Ergun et al., 1998a].

Many ideas have been proposed to explain the development and preservation of these potential drops in regions with field-aligned currents: e.g., Alfven waves propagating from the magnetotail reconnection sites and reflecting from the ionosphere, electrostatic double layers, anomalous resistivity (which increases the effective collision frequency and impedes particle transport to short out the potential), and magnetic mirroring of electrons and ions with different characteristics. An alternative, recent suggestion is that nonlinear, solitary wave structures support most of the potential drops along auroral field lines [Ergun et al., 1998b]. Figure 17.12 shows two such structures: they have spatial scales of order several Debye lengths, carry potentials of order 100 V each, move at speeds of order 3000 km s⁻¹, and often come in long trains capable of maintaining potentials of order several kilovolts. These structures are consistent with those predicted for electron hole modes,



Figure 17.10: Observations from the S3-2 satellite during a dawn-to-dusk South Pole pass on 19 September 1976 [Harel et al., 1981; Wolf, 1995]: electron energy flux in the downswards direction, the transverse magnetic field deflection, and the electric field.

corresponding to clumps of positive charge moving at speeds of order the downgoing electrons [Ergun et al., 1998c].

17.6 Auroral Kilometric Radiation

Cyclotron maser radiation is one of the two major emission mechanisms for radio waves generated in our solar system. Examples include Auroral Kilometric Radiation (AKR) from Earth's auroral regions, Jupiter's decametric radiation, solar microwave bursts associated with flares and solar active regions, planetary radio emissions from the other outer planets. This mechanism is also believed important for various stellar radio bursts.

Cyclotron maser emission involves generation near the electron cyclotron frequency Ω_{ce} and its harmonics in regions where $\Omega_{ce} >> \omega_{pe}$. Considering the wave-particle resonance condition

$$\omega - s\Omega_{ce} - k_{\parallel}v_{\parallel} = 0 , \qquad (17.3)$$

it is clear that generation at frequencies near $s\Omega_{ce}$ requires the product $k_{\parallel}v_{\parallel}$ to be small (here the subscript \parallel means parallel to the ambient magnetic field). This condition is generally met by the radiation being produced with wavevectors almost perpendicular to the magnetic field. Symmetry then implies that the radiation is produced in a cone with large half-angle ~ 90 degrees and symmetry axis along **B**.

What particle distributions can drive cyclotron maser emission? Generalizing



Figure 17.11: High resolution FAST observations of the near-midnight auroral zone [Ergun et al., 1998a]. The dashed line separates the downward and upward current regions. (a) DC electric field perpendicular to **B** and along the satellite path, (b) Electric field in a 4 kHz bandwidth. (c) East-west component of **B**. (d)-(e) Electric field spectrograms of plasma waves. (f) Electron energy spectrogram. (g) Electron pitch-angles, with 180 degrees upgoing and 0 and 360 degrees downgoing. (h)-(i) Similar to (f) and (g) but for ions.



Figure 17.12: A fast solitary wave at 0.5 μ s resolution, moving anti-Earthward at about 2000 km s⁻¹ [Ergun et al., 1998b].

(9.6), the growth rate Γ can be written

$$\Gamma \propto -\Sigma_{s=-\infty}^{\infty} \int d^3 \mathbf{p} \left(\frac{s\Omega_e}{\gamma v_\perp} \frac{\partial}{\partial p_\perp} + k_{\parallel} \frac{\partial}{\partial p_{\parallel}} \right) f(\mathbf{p}) \,\,\delta(\omega - s\Omega_e/\gamma - k_{\parallel}v_{\parallel}) \,\,, \quad (17.4)$$

where **p** is the electron momentum, γ is the Lorentz factor, and Ω_e is the electron gyrofrequency. Both the \parallel and \perp gradients in the electron distribution function can thus drive wave growth. It turns out that beam distributions (\parallel gradient) can, but that distributions with large gradients $\partial f/\partial v_{\perp}$ are more favoured. Socalled "loss cone distributions" have large gradients $\partial f/\partial v_{\perp}$ since they are missing particles with small pitch-angles, as expected for particles bouncing back and forth in the magnetic mirrors formed by dipole magnetic fields. Accordingly, cyclotron maser radiation generated by loss cone distributions of electrons is expected to be ubiquitous from planetary auroral regions and solar active regions, as is indeed observed.

Earth's AKR is one of the brightest radio sources in our solar system. As illustrated in Figure 4.9 from 0000 to about 0600 UT, AKR typically has frequencies from about 100 - 500 kHz and occurs in broadband bursts. AKR brightens during magnetic and auroral substorms and can be used as a remote indicator of geomagnetic activity.

17.7 References and Bibliography

- Carlson, H.C., Jr., and A. Egeland, in *Introduction to Space Physics*, Eds M.G. Kivelson and C.T. Russell, Cambridge, 1995.
- Cravens, T.E., Physics of Solar System Plasmas, Cambridge, 1997.
- Ergun, R.E., C.W. Carlson, J.P. McFadden et al., FAST satellite observations of electric field structures in the auroral zone, **Geophys. Res. Lett.**, 25, 2025, 1998a.

- Ergun, R.E., C.W. Carlson, J.P. McFadden et al., FAST satellite observations of large-amplitude solitary structures, **Geophys. Res. Lett.**, **25**, 2025, 1998b.
- Ergun, R.E., C.W. Carlson, J.P. McFadden et al., Phys. Rev. Lett., 81, 826, 1998c.
- Feldstein, Y.I., and G.V. Starkov, Dynamics or auroral belt and polar geomagnetic disturbances, **Planet. Space Sci.**, **15**, 209, 1967.
- Frank, L.A., and J.D. Crave, Imaging results from Dynamic Explorer 1, Rev. Geophys. Space Phys., 26, 249, 1988.
- Harel, M., R.A. Wolf, et al., Quantitative simulation of a magnetospheric substorm. 1. Model logic and overview, J. Geophys. Res., 86, 2217, 1981
- Hartz, T.R., Particle precipitation patterns, in *The Radiating Atmosphere*, Ed. B.M. McCormac, Dordrecht:Reidel, 1971.
- Hughes, W.J., Magnetopause, magnetotail, and reconnection, in *Introduction* to Space Physics, Eds M.G. Kivelson and C.T. Russell, Cambridge, 1995.
- Iijima, T., and T.A. Potemra, Large-scale characteristics of field-aligned currents associated with substorms, J. Geophys. Res., 83, 599, 1978.
- McPherron, R.L. Magnetospheric dynamics, in *Introduction to Space Physics*, Eds M.G. Kivelson and C.T. Russell, Cambridge, 1995.
- Omholt, A., The Optical Aurora, Springer-Verlag, 1971.
- Vallance Jones, A., Aurora, Dordrecht: Reidel, 1974.
- Wolf, R.A., Magnetospheric configuration, in *Introduction to Space Physics*, Eds M.G. Kivelson and C.T. Russell, Cambridge, 1995.