Chapter 16

Earth's Ionosphere and Upper Atmosphere

Discussion of the ionosphere requires a basic knowledge of the upper atmosphere. The reason is that the ionosphere is the partially ionized plasma region that co-exists with and is derived from the uppermost layers of a planet or moon's atmosphere. This Lecture proceeds by discussing Earth's atmosphere, describing the general physics of ionospheres, and then applying the results to Earth's ionosphere. The Lecture concludes with descriptions of ionospheric outflow of plasma, ionospheric scintillations, and waves in the ionosphere.

Aims and Expected Learning Outcomes

The **Aim** is to explore the physics of Earth's neutral atmosphere and ionosphere and the links with space weather, the magnetosphere, and solar-terrestrial activity.

Expected Learning Outcomes. You should be able to

- Describe the main regions of Earth's atmosphere and explain qualitatively the heating mechanisms and basic physics of these regions.
- Understand and describe the basic fluid equations governing the atmosphere and ionosphere.
- Give the requirements for an ionosphere to form and explain qualitatively the main sources of ionization.
- Explain qualitatively why an ionosphere's properties are expected to vary with time of day, longitude, latitude, and time of the year.
- Describe the physics of the ionospheric ambipolar electric field, what its magnitude is, and what some of its consequences are.
- Describe some of the effects on communications, and the origin, of ionospheric activity.

16.1 Earth's Atmosphere

The most basic model for a planet's neutral atmosphere involves assuming hydrostatic equilibrium. This approximation may be expected to fail in regions of the atmosphere where flows are significant; these include regions with (ordinary, water cloud) weather and/or those in which dynamically important turbulence exist, presumably associated with temperature gradients. For future reference we write the momentum equation for a neutral species, from Eq. (2.21), as

$$\frac{d(\eta_\alpha \mathbf{u}_\alpha)}{dt} = -\nabla P_\alpha + \eta_\alpha \mathbf{g} - m_\alpha \mathbf{u}_\alpha S_\alpha \ , \tag{16.1}$$

where $\mathbf{g}$ is the acceleration due to gravity and S_α represents the net source/loss of particles. (The last term is a mass-loading term.) In hydrostatic equilibrium the left hand side of (16.1) is zero, whence in the absence of sources and losses

$$\nabla P_\alpha = \eta_\alpha \mathbf{g} \ . \tag{16.2}$$

Consider a planar model with the height variable z and $\mathbf{g}$ anti-parallel to the z axis. Assuming the ideal gas law, then

$$\frac{\partial P(z)}{\partial z} = -\frac{P(z)}{H(z)} \ , \tag{16.3}$$

with $H(z) = RT(z)/g$, the scale height for the atmosphere. This equation has the exponential solution

$$P(z) = P(z_0) \exp\left(- \int_{z_0}^{z} \frac{dz'}{H(z')} \right) \ . \tag{16.4}$$

Assuming $T(z) = T_0$ and writing $P = nk_B T_0$, then

$$n(z) = n(z_0)\ e^{-(z-z_0)/H} \tag{16.5}$$

with $H = k_B T_0 / m_\alpha g$. That is, the simplest prediction for the atmosphere of a planet or moon is that the density should decrease exponentially with height. These results should be familiar to you, having been derived already in Assignment 2 for the Sun (see also Lectures.

Figure 16.1 [Abell, 1982] shows that the number density n of Earth's neutral atmosphere indeed does decrease approximately exponentially with altitude in localized regions. However, a pure exponential decrease would be a straight line in Figure 16.1 (since the abscissa axis is logarithmic in n) and it is clear that this prediction is not consistent with the observed density profile. The primary reason for the profile being only locally exponential is that the temperature and so the scale height vary with altitude. Where the temperature is reasonably constant, i.e., above 200 km altitude and below about 70 km altitude, the profile is indeed approximately exponential with an approximately straight line in Figure 16.1.

Equation (16.1) also holds individually for multiple separate neutral species. The result that $H_\alpha \propto m_\alpha^{-1}$ suggests that the atmosphere's composition will vary substantially with height, with more massive species being restricted to low altitudes and light species dominating the atmosphere at large altitudes. This explains qualitatively why planetary atmospheres are dominated by hydrogen (and associated ions) at large altitudes. However, while this idea is qualitatively correct, it turns out that Earth's atmosphere is relatively well mixed at altitudes below about 100 km (the homopause), presumably due to the effects of weather and turbulence, while the atmospheric constituents do separate out by mass at higher altitudes.

The temperature layers in Figure 16.1 are associated with absorption of solar radiation by particular molecules or atoms. Figure 16.2 describes these layers and the temperature structure in more detail.

- **Troposphere**. Lowest, densest layer of the atmosphere. Extends up to about 10 km altitude. Temperature decreases steadily upward due to radiation being absorbed by the ground and re-radiated upward. Temperature gradient drives convection and ordinary weather.

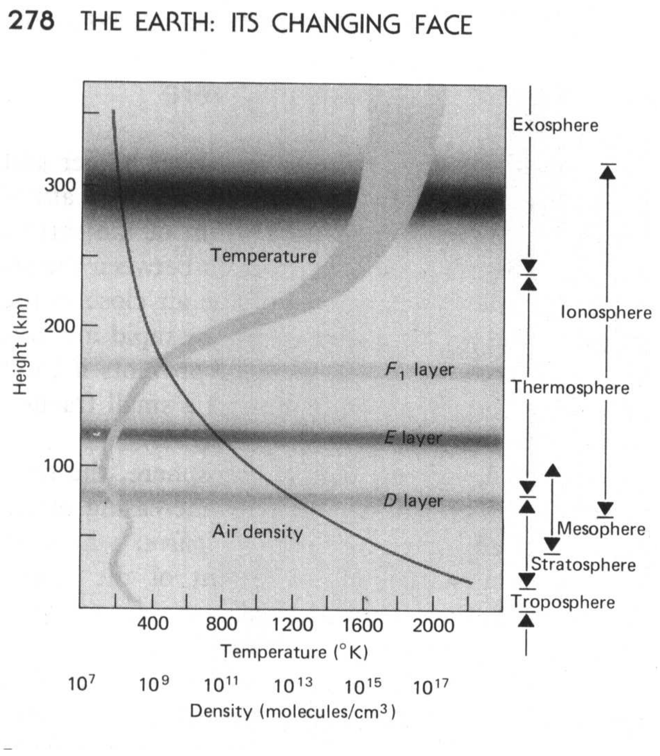

Figure 16.1: Variations in the density and temperature of Earth's neutral atmosphere with altitude [Abell, 1982].

- **Tropopause**. Boundary between troposphere and stratosphere. Location where convection ceases.
- **Stratosphere**. Cold region up to about 30 km altitude. The lower stratosphere contains the jet stream, which sometimes affects the weather in the USA and Canada.
- **Mesosphere**. Contains the ozone layer in which absorption of UV radiation leads to heating of the atmosphere. The formation of ozone involves photodissociation of oxygen molecules by UV radiation to form atomic oxygen, some of which combines with O_2 to form ozone O_3. I.E.,

$$\begin{aligned} \mathrm{O_2 + h\nu} &\rightarrow \mathrm{2O} \\ \mathrm{O_2 + O} &\rightarrow \mathrm{O_3}\ . \end{aligned} \tag{16.6}$$

  The atomic oxygen diffuses both upwards and downwards. Nitrogen molecules also undergo photodissociation, populating the upper atmosphere with atomic nitrogen and leading to species like NO. The smaller oxygen mass leads to oxygen dominating the atmospheric composition at higher altitudes, before hydrogen dominates at even higher altitudes. Note that $\mathrm{N_2}$ makes up about 78% of the atmosphere by mass, on average.
- **Mesopause** Upper boundary of the mesosphere, near about 90 km altitude. Local temperature minimum.
- **Thermosphere**. Region from 90 km to about 200 − 300 km in which the temperature increases steadily due to absorption of solar EUV radiation by

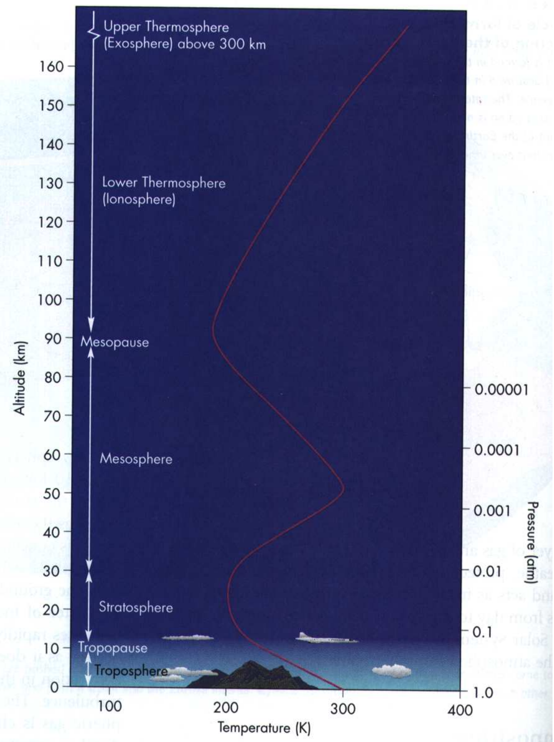

Figure 16.2: Regions of the atmosphere and associated variations in temperature [Fix, 1995].

atomic oxygen and nitrogen. The neutral atoms are often also ionized during the absorption process (photoionized), so that the ionization fraction of the thermosphere increases with height, leading to plasma behaviour.

- **Exosphere**. Region above about 200 km in which collisions become extremely rare, atoms move on ballistic trajectories, the ionization fraction is large, and the plasma behaviour becomes increasingly important.
- **Ionosphere**. This starts in the lower thermosphere near 90 km altitude and extends upwards past 300 km. Where the ionosphere ends and the plasmasphere and radiation belts etc. begin is partly a matter of definition.

Figures 16.3 and 16.4 indicate the changing nature of the atmosphere above 100 km and the start of the ionosphere. These figures also illustrate how O^+ dominates the plasma at altitudes from about 150 km to about 600 km, while H^+ dominates above about 1000 km. This difference can be important; for instance, the space shuttle encounters primarily an $O^+ - e^-$ plasma at its altitude ~ 300 km, permitting collisional charge-exchange with water outgassing from the shuttle and causing the shuttle's plasma environment to be filled with H_2O^+ pickup ions and associated plasma waves. These figures also illustrate the complicated spatial structure of the ionosphere, as partly forewarned in Figure 16.1.

16.2 Ionospheric Physics

The ionosphere is created by ionization of the neutral atoms and molecules of the atmosphere. There are two basic requirements for formation of an ionosphere: (1)

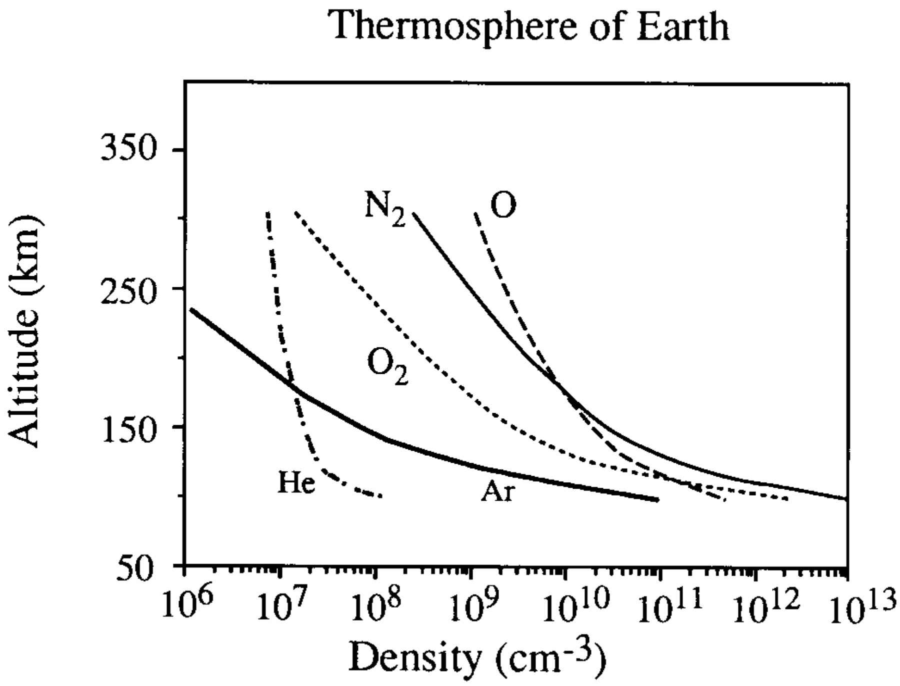

Figure 16.3: Neutral atmospheric densities for various molecular and atomic species [Cravens, 1997].

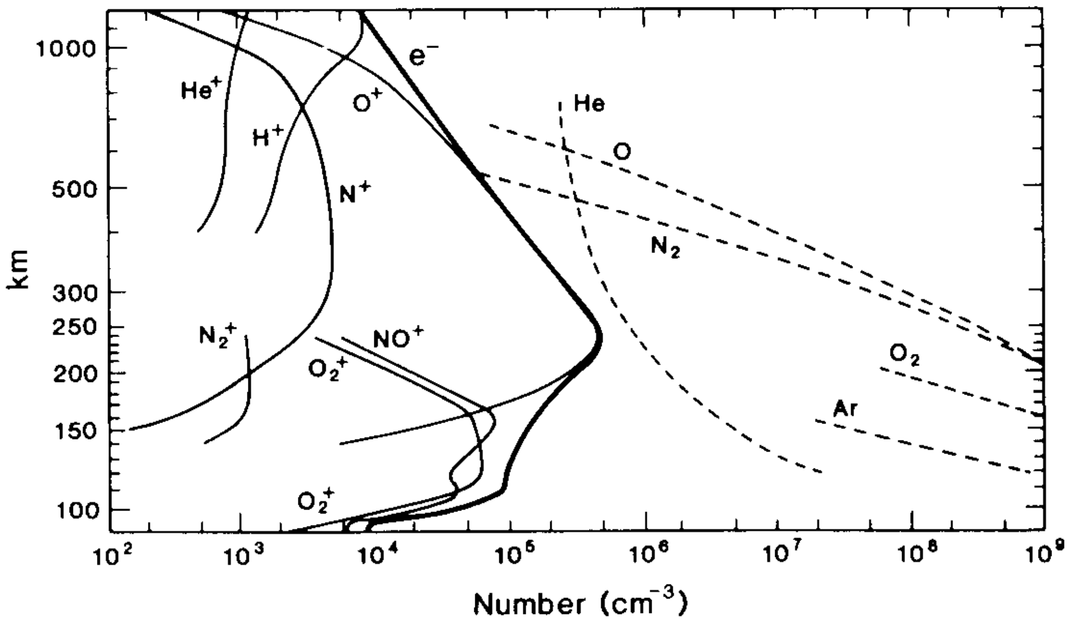

Figure 16.4: International quiet solar year daytime ionospheric and atmospheric composition based on mass spectrometer measurements [Johnson, 1969; Luhmann, 1995].

the presence of a neutral atmosphere, and (2) a source of ionization for these gases.

There are two basic sources of ionization: (1) photons and (2) energetic particles. The former process is called photoionization while the latter is called "impact ionization". The photons come primarily from the Sun (bremsstrahlung from precipitating particles is also sometimes important) while the ionizing particles can be cosmic rays, or electrons or ions from the Sun, magnetosphere, or a different region of the ionosphere.

The only requirement on the ionizing photons or particles is that their energies ($h\nu$ or $1/2mv^2$, respectively) must exceed the ionization potential or binding energy of the atom or molecule. In most cases EUV and UV solar photons with $\lambda \sim 10-10^2$ nm produce the dayside ionospheres of most planets. However, electron impact ionization is very important for Io's ionosphere [Li, 1992] and for ionization at auroral and polar cap latitudes at Earth, particularly during space weather events.

Ionization sources and losses determine the detailed structure of ionospheres. Accordingly, details of multiple chemical reactions, photodissociation, and impact ionization processes must be considered for the sources, while both recombination and chemical reactions must be considered for the losses.

16.2.1 Photoionization

Photoionization reactions can be written

$$\mathrm{M} + \mathrm{h}\nu \rightarrow \mathrm{M}^+ + \mathrm{e}^- \ , \tag{16.7}$$

where M represents an atom or molecule. The only requirement is that the photon energy hν exceed the relevant ionization potential I_M. Since typically I_M exceeds 10 – 20 eV, this means that only radiation with $\lambda < 10^2$ nm can cause photoionization. Any excess photon energy appears primarily as kinetic energy for the escaping photoelectron. Photoelectron energies can thus range from zero to hundreds of eV.

Consider a plane-parallel atmosphere which has a certain photon flux at wavelength λ incident at the top. Photoionization causes the photon flux to decrease (and the number of photoelectrons and ions produced to increase) along the path s. Figure 16.5 defines the radiation's path, the zenith direction along which the plane-parallel atmosphere is assumed to vary, and the angle χ between the line-of-sight path and the zenith direction. With $ds = dz/\cos\chi$, the flux at altitude z is given by

$$F_\lambda(z,\chi) = F_{\lambda 0} e^{-\tau_\lambda(z,\chi)} \ , \tag{16.8}$$

where $\tau_\lambda(z,\chi)$ is the optical depth given by

$$\tau_\lambda(z,\chi) = \frac{1}{\cos\chi}\int_z^\infty dz' \ n_n(z')\sigma_\lambda \ , \tag{16.9}$$

$n_n(z')$ is the neutral density, and σ_λ is the photon-absorption cross section. Assuming that $n_n(z')$ is an exponential function, as in Eq. (16.5), then the integral can be performed to yield

$$\tau_\lambda(z,\chi) = \frac{1}{\cos\chi} n_n(z) H_n \sigma_\lambda \ , \tag{16.10}$$

where H_n is the scale height for neutrals.

This equation can be used to model the source of ionization as a function of altitude, the degree of absorption of the radiation, and changes in the ionosphere with time-of-day (through the zenith angle dependence). Pursuing this last point, as χ increases, the altitude for a given τ increases, so that the total ionization rate decreases (since n_n is lower then). This implies considerable variations in the locations and amounts of ionization produced as a function of time during the

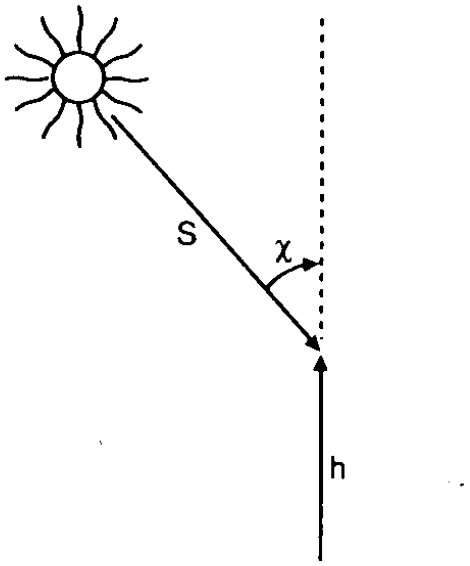

Figure 16.5: Geometry of the ray path, showing the zenith angle χ and the height z (or h) [Luhmann, 1995].

day. These daily motions and variations in the ionosphere, when considered in conjunction with recombination and other loss processes, give rise to changes in the magnetic field observed on the ground, the so-called diurnal variations that must be substracted when attempting to quantify space weather effects.

16.2.2 Impact Ionization and Losses

The ionization rate of species M due to impact ionization by energetic electrons can be written in the form

$$S_M = n_M \int dE \; \Phi_e(E) \; \sigma_M(E) \; , \tag{16.11}$$

where $\Phi_e(E)$ is the electron flux at energy E and $\sigma_M(E)$ is the corresponding cross-section [e.g., Li, 1992].

A number of electron and ion losses exist. These include radiative recombination $e + M^+ \rightarrow M + h\nu$, dissociative recombination $e + XY^+ \rightarrow X + Y$, and attachment $e + M \rightarrow M^-$. Typically these loss rates are given by the product of the number densities of electrons and the interacting species with a rate coefficient. For instance, the radiative recombination loss rate can be written

$$\frac{dn_e}{dt} = L = -\alpha n_e n_X \; . \tag{16.12}$$

These rate coefficients can be obtained experimentally.

16.2.3 Basic Theoretical Formalism

The basic approach is to use coupled fluid equations for the electrons and the multiple participating neutral and ion species, with appropriate loss terms. Number conservation, momentum conservation and energy conservation equations are used.

Below these equations are written in terms of number densities n_α rather than mass densities η_α.

$$\frac{\partial n_\alpha}{\partial t} + \nabla.(n_\alpha \mathbf{u}_\alpha) = \Sigma_i S_i - \Sigma_j L_j \ , \tag{16.13}$$

$$m_e \frac{d(n_e \mathbf{u}_e)}{dt} = -\nabla P_e - e n_e (\mathbf{E} + \mathbf{u}_e \times \mathbf{B}) - m_e n_e \mathbf{g} - m_e n_e \nu_{en} (\mathbf{u}_e - \mathbf{u}_n) - m_e n_e \nu_{ei} (\mathbf{u}_e - \mathbf{u}_i) \ , \tag{16.14}$$

and

$$m_\alpha \frac{d(n_\alpha \mathbf{u}_\alpha)}{dt} = -\nabla P_\alpha + q_\alpha n_\alpha (\mathbf{E} + \mathbf{u}_\alpha \times \mathbf{B}) - m_i n_i \mathbf{g} - m_\alpha n_\alpha \nu_{\alpha n} (\mathbf{u}_\alpha - \mathbf{u}_n) - m_\alpha n_\alpha \nu_{\alpha e} (\mathbf{u}_\alpha - \mathbf{u}_e) \ , \tag{16.15}$$

etc. The last two sets of terms in the momentum equations relate to ion-electron interactions and neutral-electron and neutral-ion interactions. These equations should then be solved simultaneously with appropriate boundary conditions. Figure 16.3 used this type of calculation.

Rather than address details here, let us focus on the important qualitative points. The first and most major point is that the electric field $\mathbf{E}$ in (16.14 – 15) cannot, in general, be ignored when gravity is retained ($\mathbf{g} \neq 0$). Instead, ionospheres tend to have significant polarization electric fields. Physically the reason is as follows: due to their much smaller masses, electrons will be able to reach much greater altitudes than ions with the same temperature, thereby causing a steady-state charge separation. This charge-separation sets up an "ambipolar" electric field which pulls ions up to higher altitudes and resists the motion of electrons to higher altitudes, setting up a situation in which approximate charge-neutrality exists. This electric field has a number of consequences that are described below.

16.2.4 The Ambipolar Electric Field

Consider just one ion species, one neutral species and electrons. Furthermore, consider only the vertical structure and vertical motions, neglect the magnetic field, and assume that the neutrals are stationary, that the ions and electrons have identical vertical velocities in the time-steady state, and that a time-steady state exists. Then all the ion-electron coupling terms cancel out below and the d/dt terms are zero. The electron and ion momentum equations are then

$$-\frac{dP_e}{dz} - n_e m_e g - e n_e E_z = n_e m_e \nu_{en} u_e \tag{16.16}$$

$$-\frac{dP_i}{dz} - n_i m_i g + e n_i E_z = n_i m_i \nu_{in} u_i \ . \tag{16.17}$$

Rewriting these equations in terms of E_z and then adding them (with the assumption that either $m_e \nu_{en} = m_i \nu_{in}$ or else the collision terms are small) yields

$$E_z = \frac{1}{2 e n_e} \left(\frac{dP_i}{dz} - \frac{dP_e}{dz} \right) + \frac{g}{2e}(m_i - m_e) \tag{16.18}$$

for $n_i = n_e$. Solving then for $u_{en} = u_{in}$ implies

$$u_e = -\frac{1}{2 n_e m_e \nu_{en}} \left(\frac{dP_e}{dz} + \frac{dP_i}{dz} + n_e g (m_i + m_e) \right) \ . \tag{16.19}$$

This ambipolar electric field has significant macroscopic consequences. Note that in the special case of an isothermal ionosphere ($P_e = P_i$), then $E = g m_i / 2e$ to a high degree of accuracy. Quantitatively, this corresponds to $E \approx 5 \times 10^{-8}$V m^{-1}. Yet this small electric field has observable macroscopic consequences! The reason is because this field acts over large distances of order 10^6 m at Earth, so that the total potential drop is of order 0.05 V which is of the same order as the atmosphere's thermal temperature ≈ 0.05 eV (to within a factor of 2). Accordingly, the field is acting to minimize plasma losses. Another important qualitative point is that plasma flows are typically present. In this one-species case, the flow corresponds to outflow of the ionospheric plasma. As discussed more below, this physics is the basis for "exospheric" models for the polar wind out of Earth's ionosphere and for the solar wind.

Before proceeding, let us just note that this system of equations resembles a diffusion equation. Assuming that T_e and T_i are constant, then the last equation above can be rewritten

$$n_e u_e = -\frac{1}{2m_e\nu_{en}}\left(k_B(T_e+T_i)\frac{dn_e}{dz} + n_e m_i g\right) \quad (16.20)$$

$$= -D\left(\frac{dn_e}{dz} + \frac{n_e}{H}\right) , \quad (16.21)$$

where H is the usual scale height and $D = k_B(T_e+T_i)/\nu_{en}$. Substituting this final form into the number conservation equation results in a diffusion equation:

$$\frac{\partial n_e}{\partial t} - D\left[\frac{\partial^2 n_e}{\partial z^2} + \frac{\partial}{\partial z}\left(\frac{n_e}{H}\right)\right] = \Sigma_i S_i - \Sigma_j L_j \ . \quad (16.22)$$

Thus, both electrons and ions can diffuse in altitude, so that the peak plasma density need not be where the peak ionization is.

16.2.5 Additional Physical Points

In the last subsection it was shown that an ambipolar electric field is set up in steady-state ionospheres so as to limit charge-separation electric fields caused by the much greater electron mobility (due to their much smaller masses). These electric fields pull ions out of the ionosphere. Schunk [1983] considered the effects of the electric field set up by the major ion species on a second, much lighter ion species (such as protons). He found that combining the number conservation equation and the momentum equation for the minor species (including the polarization field) led to an equation strongly reminiscent of Parker's gasdynamic equation for subsonic to supersonic flow. He found that this equation admitted subsonic to supersonic solutions. This model explains supersonic flows of light ions from Earth's polar ionosphere, called the polar wind.

Finally, it might be wondered why the ionospheric temperature increases with height in the thermosphere and ionosphere. The primary reason is that solar radiation is absorbed and impact ionization occurs primarily at large heights above 50 km. Another qualitative explanation based on microphysics is simply that only hotter particles can overcome the gravitational potential and reach those heights, so that the temperature should increase with height. As discussed in detail in Lecture 8, this "velocity filtration" mechanism is not, in fact, quantitatively viable for Maxwellian distributions of particles. However, for non-Maxwellian distributions this idea is quantitatively viable [e.g., Scudder, 1992] and has been proposed as an alternative model (to true heating) for the high temperature of the solar corona.

16.3 Earth's Ionosphere

Figures 16.4, 16.6 and 16.7 illustrate the density structure in Earth's ionosphere. These data can be obtained by rocket and spacecraft measurement and by radar

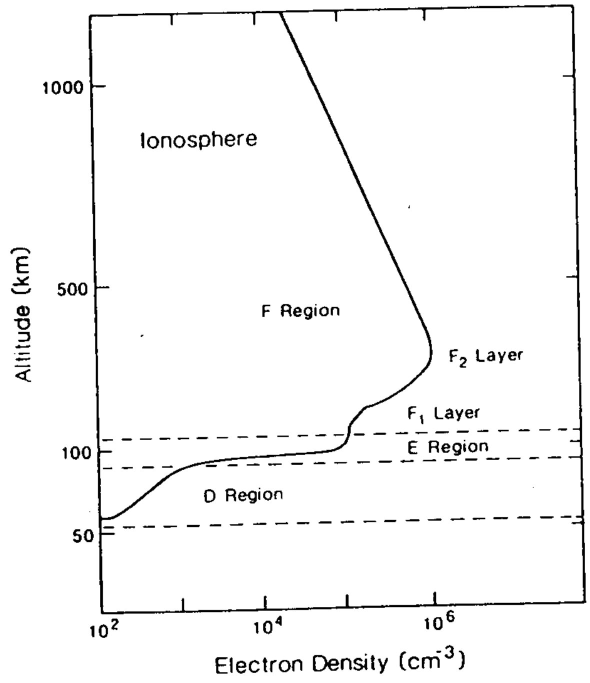

Figure 16.6: Typical plasma density profile of Earth's ionosphere, showing the D, E, and F layers, as functions of altitude [Brand, 1998].

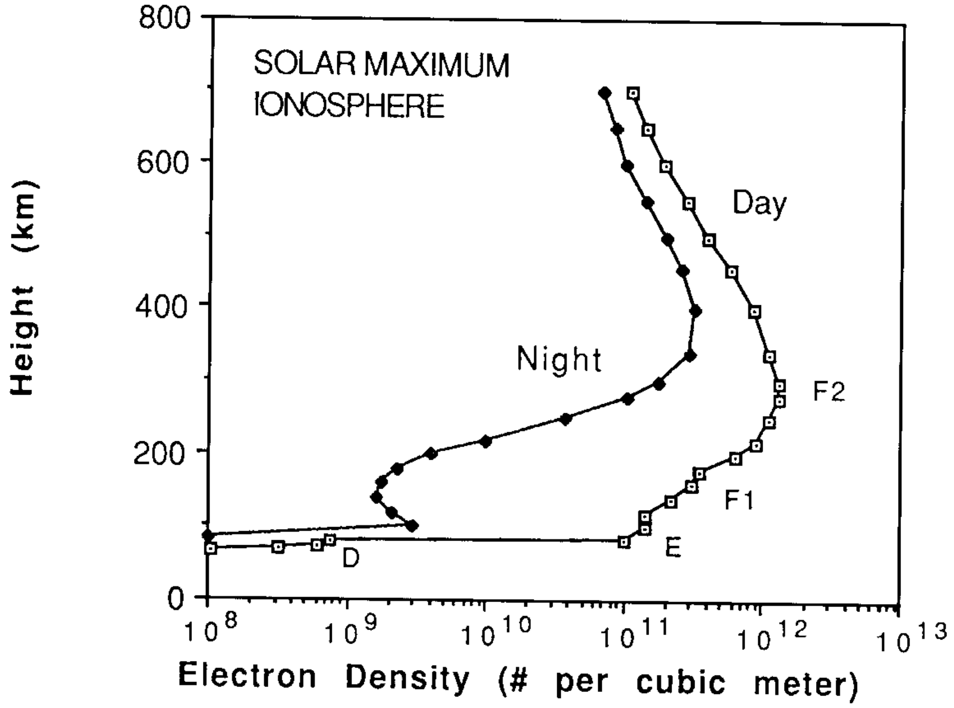

Figure 16.7: Typical ionospheric density profiles for day and night [Cravens, 1997].

sounding from the ground and space. The D region is where the ionosphere starts and becomes appreciable, covering the altitude range from about 50 to 90 km. Remembering that $f_p = 8.98 n_e^{1/2}$ kHz for n_e measured in cm^{-3}, the Figures 16.4, 16.6 and 16.7 show that the D region will reflect terrestrial signals with frequencies below about 2.5 MHz. The D layer is associated with ionization by cosmic rays and X-rays and is not well understood [Luhmann, 1995].

The E layer lies between approximately 90 and 130 km is associated primarily with ionization by UV photons and the ions O_2^+ and NO^+ [Luhmann, 1995], as can be seen in Figure 16.4. It is apparently well described by existing theory. The F layers lie above 130 km and are associated mostly with O^+ ions. The lower, F1, layer appears to be reasonably well understood, while the higher F2 layer is not. It is the E and F layers which increase the maximum plasma frequency in Earth's ionosphere to values above about 10 MHz.

Note Figure 16.7's substantial differences between the ionospheric density profiles in the day and at night. These are due primarily to the much smaller ionization rate at night. However, variations in the recombination rate with altitude and vertical transport are also important. These effects lead to substantial changes in the propagation of radio waves produced on the ground.

16.4 Space weather effects and related topics

Impact ionization of neutrals by precipitating electrons is the dominant source of ionization at auroral latitudes during geomagnetically disturbed times [Carlson and Egeland, 1995], as can be seen in the movie in the powerpoint presentation for this Lecture. This process gives rise to the enhanced conductivity in these regions, which then permits intense localized currents to flow in the "auroral electrojets" along the auroral oval. Of course, these precipitating electrons also yield their energy, via two-step processes, to produce the auroral lights. These effects are discussed in detail in the next lecture.

The large electric fields, currents, and plasma drifts relative to the neutral plasma at auroral latitudes during disturbed times lead to a variety of instabilities, particularly near the auroral electrojets at altitudes characteristic of the F region [Carlson and Egeland, 1995]. These instabilities cause the initially (relatively) homogeneous plasma to develop density turbulence, typically magnetic field-aligned, with scale sizes of order 100 m to 10 km. Radiation propagating through this density turbulence is refracted, scattered, reflected, as well as suffering phase changes and sometimes Doppler changes. These effects cause scintillation or twinkling of radio sources, analogous to the twinkling of star light by density turbulence in the atmosphere. These scintillations lead to strong amplitude fading and phase fluctuations all the way up to GHz frequencies, thereby disrupting communications up to GHz frequencies, GPS and global navigation satellite systems (GNSS), and other systems. In addition, the scattering can blind radar tracking (e.g., over-the-horizon radars), disrupt (and sometimes improve) communications. These scintillation and scattering effects can be used as a diagnostic for space weather. Figure 16.8 illustrates these scintillation effects. It should be emphasized that the density turbulence generated at auroral latitudes propagates and diffuses through the ionosphere, so that the much of the ionosphere can theoretically become disturbed and inhomogeneous as a result of auroral activity.

The enhanced UV and X-rays of solar flares cause enhanced ionization on the dayside, concentrated near local noon. This leads directly to enhanced currents and so magnetic disturbances. The enhanced ionization also changes the propagation conditions for radio and VLF waves.

Finally, just as a final topic, it is relevant that the ionospheric plasma and

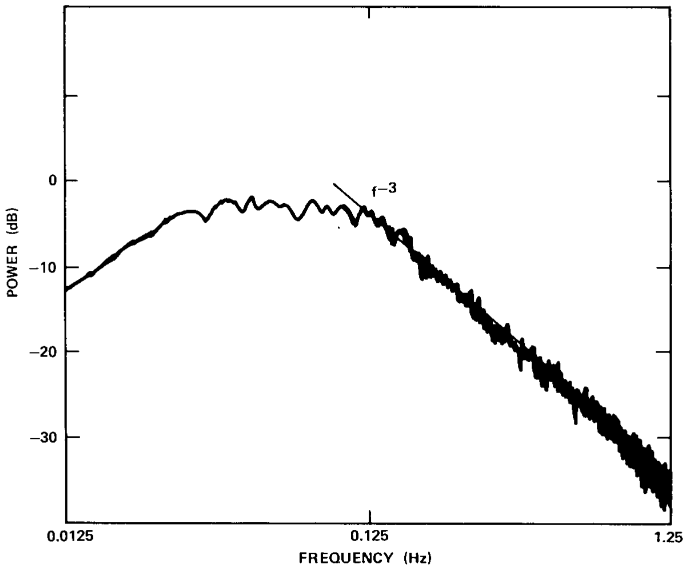

Figure 16.8: Power spectrum of 4 GHz ionospheric scintillations for an event measured in Hong Kong in 1973 [Lanzerotti, 1979] .

atmosphere contain many other waves. Some are associated with tides. Some are global disturbances, called “travelling ionospheric disturbances” (TIDs), which can travel several times around the world after a sudden event in the ionosphere and can be seen in VLBI and other astrophysical data. Some TIDs are associated with enhanced auroral activity, thereby travelling from the auroral regions towards the equator and so leading to global disturbances rather then effects solely at high magnetic latitudes.

16.5 References and Bibliography

- Abell, G.O., *Exploration of the Universe*, Fourth Edition, Saunders College Publishing, 1982.
- Brand, F., *Physics 3 & 4 Plasma Physics*, 1998.
- 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.
- Fix, J.D., *Astronomy – Journey to the Cosmic Frontier*, Mosby, 1995.
- Lanzerotti, L.J. (Ed.), Impacts of ionospheric/magnetospheric processes on terrestrial science and technology, in *Solar System Plasma Physics, Vol. III*, Eds L.J. Lanzerotti, C.F. Kennel, and E.N. Parker, North-Holland, 1979.
- Li, X., *Io's Atmosphere, Ionosphere, and its Interaction with the Io Plasma Torus*, M.Sc. Thesis, U. Iowa, USA, 1992.
- Luhmann, J.G., Ionospheres, in *Introduction to Space Physics*, Eds M.G. Kivelson and C.T. Russell, Cambridge, 1995.
- Scudder, J.D., Astrophys. J., 398, 319, 1992.