

Northern Lights by Fridtjof Nansen (1910-1911) - Cover illustration from his book *Nord I Takeheimen* (In Northern Mists) The image of the northern lights is from a woodcut created from a crayon sketch he made in 1893.

Chapter 7 The Ionosphere

A Introduction

The upper atmosphere, beginning at about 50 km altitude, is partially ionized by ultraviolet and x-ray radiation from the sun. This region of partially ionized gas extends upwards to about 1000 km altitude. The region is termed the ionosphere. The ionosphere is important as a source of plasma for the magnetosphere, and as a medium which reflects radio waves at frequencies from a few Hz up to several Megahertz. Like the rest of the earth-sun system we have explored it is a dynamic region with an amazing variety of features. The most striking is the aurora, which reflects the influence of the hot magnetospheric plasmas on the high latitude atmosphere. Figure 7.1 shows an auroral image acquired by DMSP, in a snapshot of the auroral activity found in visible wavelengths. (DMSP orbits at an altitude of about 500 km, looking down on the atmosphere. The aurora occurs at altitudes below 100 km.) These bright glows are visible from the surface of the earth, and the northern lights (aurora borealis) and southern lights are a topic of perpetual interest for space physicists.

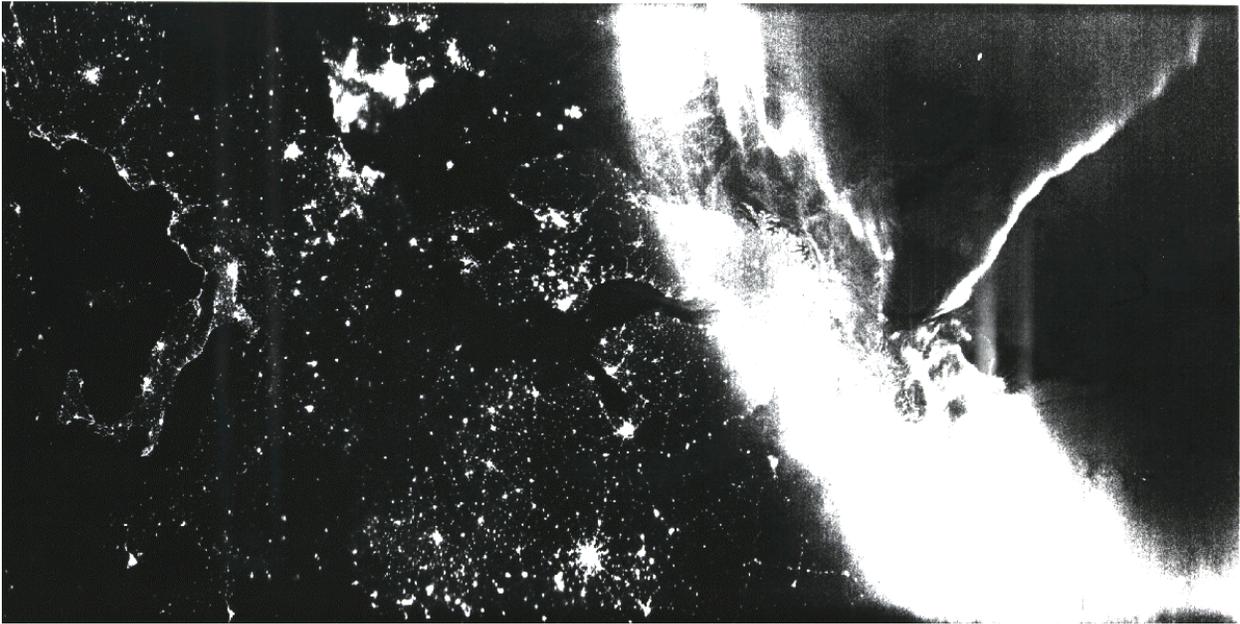


Figure 7.1 Auroral image from the Defense Meteorological Satellite Program (DMSP) Operational Linescan System. The OLS is the visible imager, with 2.7 km resolution (low resolution mode). This image of the aurora was obtained over Europe on 11 December 1988, at ~2120 UT. The satellite traverses a polar orbit from 76.8 N, 45.8 E (far right) to 39.6 N, 9.7 E (far left) over a 10 minute interval. The 'boot' of Italy is on the left, the bright spot on the center, lower edge of the image is Moscow. The auroral oval cuts across the northern edge of Scandinavia in this image, extending 8-10 degrees in latitude at a given longitude. There is a small 'polar' arc extending northward from the main auroral glow. Image courtesy of Dr. Charles Pike, USAF (Geophysics Lab, Hanscom AFB) See also <http://web.ngdc.noaa.gov/dmsp/source/aurora.html>.

Embedded in the upper atmosphere there are several fairly distinct layers where positive ions and electrons are present in sufficient numbers to affect radiowave propagation. The variations in

electron density with altitude led to the subdivision of the ionosphere into the D, E and F regions as shown in Fig. 7.2. The electron density in all of these regions varies with time of day, season and solar activity. On the average, the electron density is greater in the daytime than at night, greater in summer than in winter, and greater during the maximum phase of the sunspot cycle than at sunspot minimum. However, there are noteworthy exceptions. For instance, the peak electron density of the F₂-layer is greater in winter than in summer. These exceptions will be described in section 7.

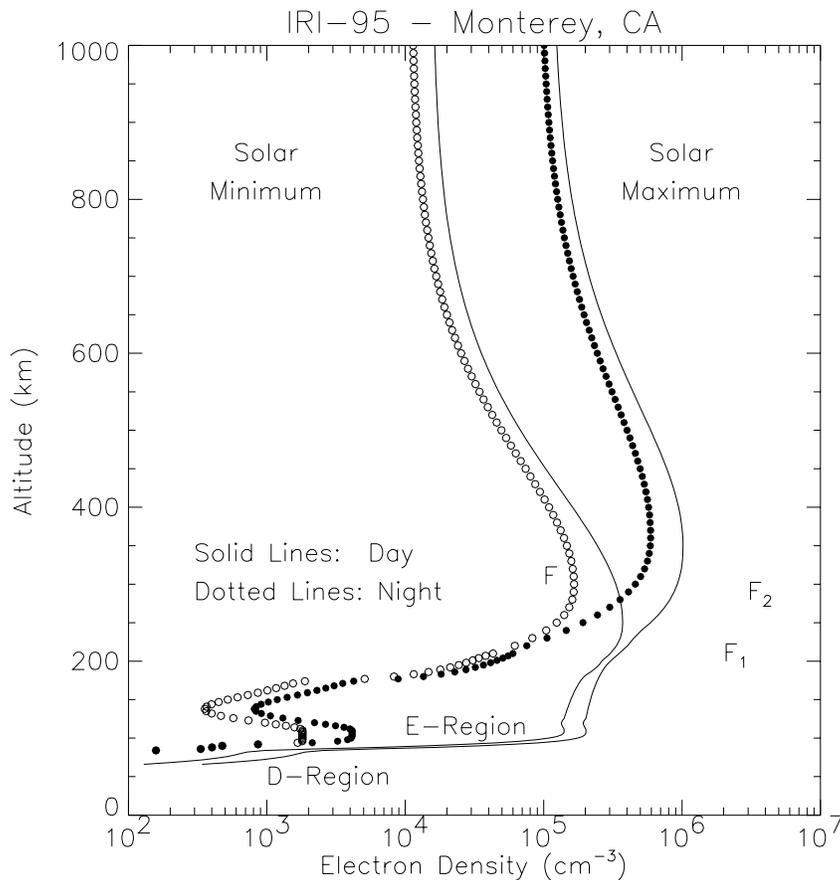


Figure 7.2 Electron density as a function of altitude for the mid-latitude ionosphere. The International Reference Ionosphere 1995 (IRI-95) model was run for the location of Monterey, CA (geographic latitude = 36.5°, geographic longitude = 238°), for July 4, 1989 (solar maximum) and July 4, 1995 (solar minimum). Calculations were done at 0LT (local midnight), and 12 LT (local noon). Note that for the more traditional metric unit for density, m⁻³, you multiply by 10⁶. [<http://nssdc.gsfc.nasa.gov/space/models/iri.html>]

The ionosphere serves as a high altitude reflector for short-wave broadcasting and long-range communication. High frequency (3-30 MHz) radio waves are reflected by the F-region or the daytime E-region, while very-low-frequency (less than 30 kHz) radio waves are reflected by the D-region. When a radio wave travels through the ionosphere, the electrons are set into oscillation at the frequency of the wave. The energy of oscillation is obtained from the wave. The electrons lose some of this energy by colliding with the neutral molecules of the upper atmosphere. In these collisions, much of the energy of oscillation that the electrons acquired from the radio wave is transferred to the neutral molecules, and appears as random kinetic energy. Consequently, the radio wave is attenuated. The higher the rate of collision of electrons with neutral molecules the

greater is the attenuation. Hence, most of the attenuation generally occurs in the D-region, where the density of the neutral molecules and the collisional frequency of the electrons is greater than in the upper region.

Furthermore, it turns out that the attenuation is inversely proportional to the square of the frequency, so that waves of frequency near the lower end of the HF band are more severely attenuated than those near the upper end of the band. In practice, therefore, the optimum frequency used by a communication circuit is the highest frequency that will be reflected by the ionosphere.

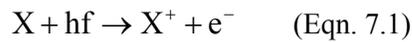
During the occurrence of a solar flare, the enhanced x-ray radiation from the sun causes the electron density in the D-region to increase by a large factor. This results in the complete absorption of high-frequency radio waves, and the disruption of short-wave communication over the earth's sunlit hemisphere. Beginning about 26 hours after a solar flare, the global F₂-region electron density undergoes substantial variations owing to the interaction of the disturbed solar wind with the ionosphere. These variations could last for several days after the flare.

At high latitudes, a significant fraction of the ionization is produced by charged particles that have been dumped from the radiation belts at times of magnetic storms, and also by charged particles from the sun that have been diverted to these latitudes by the geomagnetic field.

B Formation of an Ionospheric Layer (Chapman Theory)

The ionospheric plasma is electrically neutral i.e. there are an equal number of positive and negative charges in a reasonable size volume. But due to their much larger mass (typically a factor of 30,000 or more) the ions are much less mobile and their motions can be ignored for many purposes such as the interaction with radiowaves. The ionosphere is not entirely or even predominantly composed of charged particles. Neutral particles far outnumber the ions and electrons even in the layers of highest charge concentration. For example in the F₂ region the charged particle density is only about 10^{-3} that of the neutral particle density.

The primary process by which charged particles are produced in the upper atmosphere is the ionization of the neutral gases by solar ultraviolet and x-radiation, particularly at low and mid latitudes. The formation of layers in the ionosphere can be qualitatively understood by considering the ionization of a typical atom or molecule in the atmosphere



The rate of this reaction, $q(z)$, at a given altitude, z , will be proportional to the product of the concentration of the molecules, $n(z)$, and the intensity of the solar radiation, $I(z)$, at that altitude. Thus,

$$q(z) \propto n(z) I(z) \quad (\text{Eqn. 7.2})$$

At very high altitudes, there are lots of photons but few molecules. At very low altitudes there are lots of molecules but very few photons to cause ionization. Thus, the rate of electron production $q(z)$ peaks at some intermediate level as illustrated in Figure 7.3

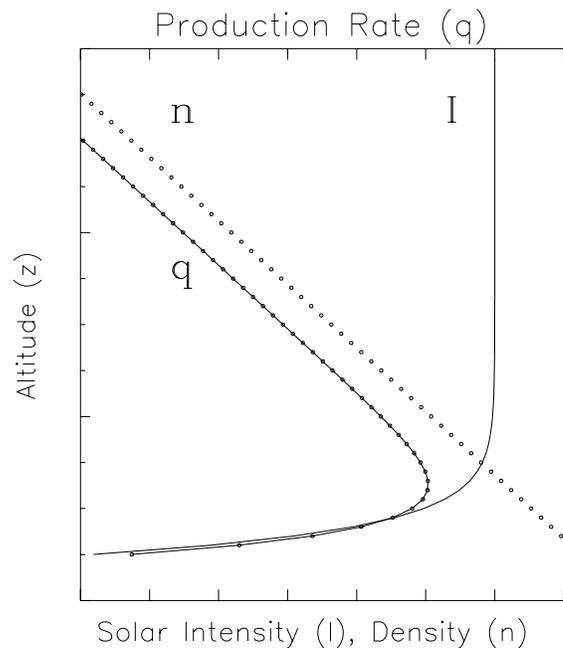


Figure 7.3 Schematic illustration of Chapman's model

In 1931, Sidney Chapman worked out a simplified theoretical model of this process by assuming that:

the atmosphere consists of only one gas, and is horizontally stratified so that the concentration, n , of atoms (or molecules) depends only on the altitude, z ;

the atmospheric gas density decreases exponentially with altitude, as $n(z) = n_0 e^{-z/H}$, and;

(3) the radiation from the sun is monochromatic.

Despite the rather drastic simplifications the theory yields remarkably good results for each atmospheric constituent. The conclusions of Chapman's theory may be summarized as follows.

1 Solar Intensity

First, the intensity of the solar radiation at the altitude z is given by:

$$I(z) = I_{\infty} e^{-\sigma n(z) H \sec \chi} \quad \text{photons/m}^2 \text{ s} \quad (\text{Eqn. 7.3})$$

where I_{∞} is the unattenuated intensity, i.e. the intensity at the “top” of the atmosphere, is the **absorption** cross-section of the gas for the particular wavelength of interest ($\sigma \sim 10^{-22} - 10^{-27} \text{ m}^2$ for the earth’s atmosphere at typical wavelengths of solar radiation. $n(z)$ is the number density of the atmospheric gas molecules at the altitude z ,

H is the scale height of the atmosphere,

$\left(\frac{kT}{mg}\right)$ and χ is the angle between the direction of the solar radiation and the vertical; χ is often called the solar zenith angle.

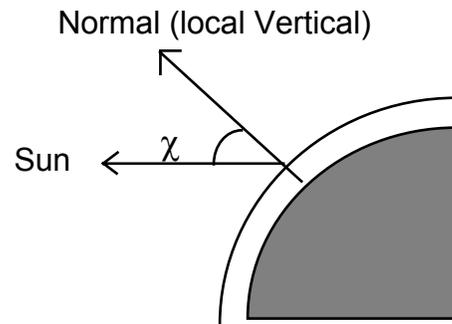


Fig. 7.4 Definition of solar zenith angle

Note that (a) $n(z)$ increases and consequently $I(z)$ decreases as the radiation propagates downward, and (b) $I(z)$ is greater at noon, when χ is small than at sunrise or sunset when χ is large.

2 Peak Altitude

Assuming that each photon produces one electron, the rate of electron production at the altitude z is:

$$q(z) = \sigma n(z) I(z) \quad (\text{Eqn. 7.4})$$

(In reality, the primary electron produced by a highly energetic photon will itself produce several additional electrons by cascaded collisions, in which case the above expression should be modified by the effective number of electrons produced by each photon. This modifying factor can range from 10-20 in the UV range. Note that due to an unfortunate shortage of Greek symbols, the same symbol ‘ σ ’ is used here for the **ionization** cross section. It will be similar to, but not identical, to the absorption coefficient.)

Now, the number density n_m at the altitude where $q(z)$ is maximum is given by :

$$\sigma n_m H \sec \chi = 1 \quad (\text{Eqn. 7.5})$$

Hence, the altitude at which $q(z)$ is a maximum is

$$z_m = H \ln (n_0 \sigma H \sec \chi) \quad (\text{Eqn. 7.6})$$

At this altitude, the intensity of the solar radiation is $I_{\infty} e^{-1} = I_{\infty}/e$. (Eqn. 7.7)

Note that z_m is independent of the intensity of the solar radiation. It depends only on the properties of the atmosphere and the solar zenith angle. When χ increases, so does z_m . Thus the altitude of peak electron production is higher at sunrise and sunset than at noon.

example: using $n_0 = 2 \times 10^{25} \text{ m}^{-3}$, $H = 8.71 \times 10^3 \text{ m}$

$$\begin{array}{llll} \chi = 0^\circ & \sigma = 10^{-18} \text{ cm}^2 & (10^{-22} \text{ m}^2) : & h_m = 145 \text{ km} \\ & \sigma = 10^{-20} \text{ cm}^2 & (10^{-24} \text{ m}^2) : & h_m = 105 \text{ km} \end{array}$$

$$\begin{array}{llll} \chi = 60^\circ & \sigma = 10^{-18} \text{ cm}^2 & : & h_m = 151 \text{ km} \\ & \sigma = 10^{-20} \text{ cm}^2 & : & h_m = 111 \text{ km} \end{array}$$

3 Peak Production Rate

The rate of electron production at the peak is given by:

$$q_m = \sigma n_m \frac{I_\infty}{e} = \sigma \left(\frac{1}{\sigma H \sec \chi} \right) \frac{I_\infty}{e} = \left(\frac{I_\infty}{e H} \right) \cos \chi \quad (\text{Eqn. 7.8})$$

[Note: here as above, 'e' is the 'exponential', not the charge of the electron]

Note that q_m is proportional to I_∞ and $\cos \chi$. As χ increases q_m decreases. Thus the peak electron production is greater at noon than at sunrise and sunset.

Example (hypothetical)

$$I_\infty = 5 \times 10^{11} \text{ photons/m}^2 \text{ sec}, \quad H = 8.71 \times 10^3 \text{ meters}, \quad \chi = 10^\circ$$

$$\text{then } q_m = \frac{5 \times 10^{11}}{(2.72)(8.71 \times 10^3)} \cos 10^\circ = 2.08 \times 10^7 \text{ electrons/m}^3 \text{ sec}$$

4 Production Rate

Now, define q_0 to be the peak rate of electron production when $\chi = 0$, and z_0 to be the altitude for the peak production when $\chi = 0$.

$$\text{Then, } q_0 = \frac{I_\infty}{eH} \quad \text{and} \quad q_m = q_0 \cos \chi \quad (\text{Eqn. 7.9})$$

$$\text{Also, } z_0 = H \ln(n_0 \sigma H) \quad (\text{Eqn. 7.10})$$

A careful rearrangement of all the terms eventually leads to:

$$q(h) = q_0 \exp [1 - h - \sec \chi \exp(-h)] \quad (\text{Eqn. 7.11})$$

$$\text{where: } h = \frac{z - z_0}{H}$$

This is called the Chapman (ion-pair) production function. Plots of $q(h)$ versus h for various values of χ are shown in Figure 7.5.

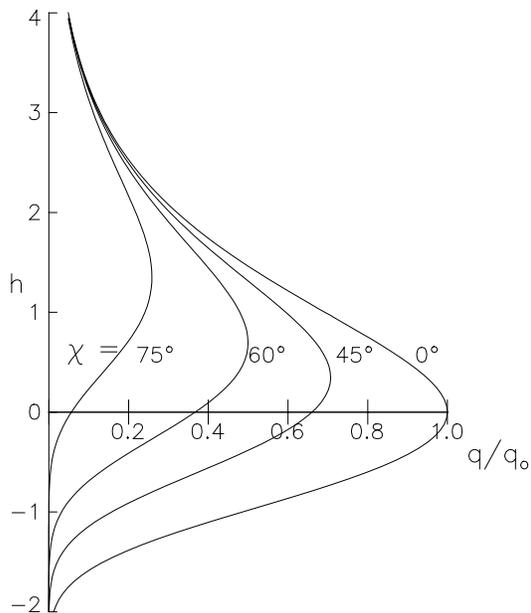


Figure 7.5 : Normalized Chapman production rate versus reduced height, z , parametric in solar zenith angle χ , at the equator

6 Loss Processes

The above discussion of how the ionization process occurs provides the major fraction of the knowledge need to understand why the Chapman layers exist. The ions which are generated in this fashion can decay quite quickly, however, because direct collisions between the ions and neutrals lead to the elimination of the plasma. A primary loss mechanism is dissociative recombination, where electrons combine with various molecular (positive) ions.

C Ionosphere morphology - Altitude Structure

Different atmospheric gases are ionized by various components of the solar radiation to produce the different ionospheric regions.

1 D-region: 60 km to 90 km.

Formation

This is the lowest region of the ionosphere and is thus produced by the penetrating component of the incident radiation, namely short wavelength ultraviolet (Lyman α with $\lambda = 121.6$ nm) and x-rays. The D region is formed primarily by the ionization of the trace atmospheric constituent NO ($[\text{NO}] \approx 10^7 \text{ cm}^{-3}$ as compared with $[\text{N}_2] \approx 10^{14} \text{ cm}^{-3}$ at 85 km) by the relatively intense Lyman- α radiation ($I_\infty \approx 3.3 \times 10^{11} \text{ photons/cm}^2 \text{ sec}$) from the sun. Ionization of N_2 and O_2 by solar x-rays is a secondary process. The contribution of this latter process is small except during a solar flare. The dominant loss process is the dissociative recombination of electrons with various molecular positive ions.

Electron Density

The electron density increases from about 100 electrons/cm³ at 60 km to about 10⁴ electrons/cm³ at 90 km, around noon. It is greater in the summer than in the winter, and greater at sunspot maximum. At night, when there is no incident radiation, the electrons quickly recombine with the molecular positive ions, so that the D-region disappears, except at latitudes greater than about 65°, where particle bombardment sustains the ionization.

Anomalies

Sudden Ionospheric Disturbance (SID)

During a solar flare, the electron density in the D-region increases by a large factor as a result of the considerable increase in the solar 'hard' x-rays of wavelength less than 10 Å. This increase is a factor of 100 to 1000 depending on the severity of the flare. Since the increased electron density occurs at altitudes where the electron collision frequency is high, radio waves propagating in the ionosphere are almost completely absorbed, and high-frequency (HF) communications are disrupted over the sun-lit hemisphere. This is sometimes called a radio blackout. A related phenomenon is the PCA event.

Polar Cap Absorption (PCA)

Polar Cap Absorption is the name given to the severe attenuation suffered by a HF radio wave propagating in the ionosphere at a high latitude near the polar caps, in the daytime or at night, soon after a solar flare. It is caused by the high flux of solar protons emitted during a large flare, and deflected to the polar regions by the geomagnetic field.

2 E layer: 90km-140 km:

This is the best understood region of the ionosphere, and the first 'layer' identified in ionospheric research (It was the 'electric' layer - hence the e-layer)

Formation

The E-region is formed primarily by the ionization of O_2 . The solar radiations primarily responsible for the ionization are Lyman- β of wavelength 1025.7 \AA ($I_\infty \approx 3.6 \times 10^9 \text{ photons/cm}^2 \text{ sec}$), and the C_{III} line of wavelength 977 \AA ($I_\infty \approx 4.4 \times 10^9 \text{ photons/cm}^2 \text{ sec}$). An additional production process is the ionization of N_2 (and O_2) by X-rays of wavelength in the range $10\text{-}100 \text{ \AA}$. The N_2^+ ions are converted to O_2^+ and NO^+ ions by rapid charge exchange. The net charge production rate is about $4000 \text{ electrons/cm}^3 \text{ sec}$ at 105 km for $\chi = 10^\circ$. At high latitudes, particle radiation makes a significant contribution to the ionization at all hours.

The dominant ions in the E-region are O_2^+ and NO^+ , so that the dominant loss process is the dissociative recombination of the electrons with these ions.

Electron Density

The peak electron density around noon, at equinox at the equator (i.e., $\chi = 0$) is $\cong 2 \times 10^5$ el/cm³. The diurnal, seasonal and latitudinal variation is in approximate agreement with the Chapman theory. The height of the peak varies with χ in agreement with the theory, with $h_0 \cong 105$ km and $H \cong 8$ km. Chapman theory shows that the production rate, defined by q_{\max} , is linearly proportional to $\cos \chi$. If the dominant loss process is dissociative recombination, i.e., $q = \alpha N_e^2$, then $N_{e \max}$ should vary as $\cos^{0.5} \chi$. The maximum electron density is found by experiment to vary with χ as $\cos^{0.6} \chi$. The slight difference in the exponent (0.6 versus 0.5) can be accounted for by the height variation of the scale height and of the recombination coefficient. Note that the functional dependence implied by χ can be either time of day or latitude.

Based on a large number of measurements, the solar-cycle variation of the electron density may be expressed by:

$$(N_e)_{\max} \cong a (1 + 0.004R_z) \quad (\text{Eqn. 7.12})$$

where R_z is the sunspot number and $a \cong 1.3 \times 10^5$ el/cm³ at $\chi = 0^\circ$ ('a' varies slightly from month to month)

The E-region persists even during the night, with electron densities in the range 500-10,000 electrons/cm³. The nighttime E-region is thought to be maintained by solar extreme ultraviolet (EUV) radiation, primarily Lyman- α and Lyman- β which has been scattered from the exospheric hydrogen - e.g. the geocoronal glow.

Disturbances (Anomalies)

Within the E-region, local enhancements in electron density are frequently observed. These are known as sporadic-E, or E_s . Ground-based observations (global network of ionosondes) show that E_s is more prevalent in summer than in winter, at mid-latitudes. At the geomagnetic equator (actually the magnetic dip equator), E_s is observed mainly in the daytime, throughout the year. Rocket-borne experiments have shown that, at mid-latitudes, E_s is a thin layer, of thickness in the order of a few hundred meters, with N_e greater than $(N_e)_{\max}$ of the E-layer. The processes involved in the production of E_s are rather complicated.

3 F-region: 140km-1000km

This is the region that is primarily responsible for the reflection of radio waves in high-frequency communication, broadcasting, and OTHR (over-the-horizon radar) - hence the most important of the ionospheric regions.

Formation

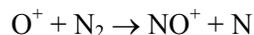
The primary production process in the F-region is the ionization of atomic oxygen, O, by solar radiation of $\lambda < 911 \text{ \AA}$. The spectral bands responsible for the ionization are the Lyman continuum: $800 - 910 \text{ \AA}$ ($I_{\infty} \cong 1.0 \times 10^{10} \text{ ph/cm}^2 \text{ sec}$); the wavelength range $200-350 \text{ \AA}$, including the strong He II line at 304 \AA ($I_{\infty} \cong 1.5 \times 10^{10} \text{ ph/cm}^2 \text{ sec}$); and the wavelength range $500-700 \text{ \AA}$. A secondary production process is the ionization of molecular nitrogen and oxygen by solar radiation of $\lambda < 796 \text{ \AA}$. Peak electron production occurs in the height range of $\cong 160-180 \text{ km}$, but the peak N_e occurs at a greater height (see below).

The primary positive ions produced by the above radiations are O^+ , N_2^+ and O_2^+ . Various chemical process then convert these ions into different ions, so that the observed dominant ions are:

O^+ , NO^+ and O_2^+ in the height range 140-200 km (F_1), and
 O^+ in the height range 200-400 km (F_2).

The F-region divides into two layers, called F_1 and F_2 , particularly in the summer in the daytime. The F_1 -layer, forms a ledge in the electron density profile at the bottom of the F_2 -layer.

The main loss process in the F_1 -layer ($\cong 140-200 \text{ km}$) is the dissociative recombination of the electrons with the molecular positive ions. In the F_2 -layer ($\cong 200 \text{ km}$ to $\cong 400 \text{ km}$) the main chemical loss process is a two-stage process in which ion-molecule reactions first convert O^+ to the molecular ions NO^+ and O_2^+ by the reactions



and



and the electrons then recombine dissociatively with these ions. This loss process obeys a linear law, i.e., $L = \beta(h) N_e$, with the loss coefficient $\beta(h)$ decreasing with height faster than the production $q(h)$ decreases with height, so that the electron density increases with height above the level of peak production (as noted earlier). The F_2 -peak is formed by the combined action of this linear loss process and the transport of electrons by plasma diffusion. In fact, the transport of electrons by various processes plays an important role in the morphology of the F_2 -layer.

Above 500 km, significant amounts of the hydrogen ion H^+ are observed and, at heights greater than about 1200 km, H^+ is the dominant ion, especially at night.

Electron Density

The maximum electron density of the F₁-layer is about 3×10^5 el/cm³, around noon. The layer disappears at night. The peak of the F₂-layer is about an order of magnitude higher in density. The diurnal, seasonal and latitudinal variation of the electron density and height of the F₂ - layer are NOT in accordance with the χ -variation. For instance, $(N_e)_{\max}$ is:

(a) greater in winter (2.5×10^6 at $R_Z = 200$) than in summer (7×10^5 at $R_Z = 200$). This is called the 'winter anomaly',

(b) greater on either side of the equator than at the equator - the Appleton (equatorial) anomaly, and

(c) greater before noon, or afternoon, than at noon in the summer.

Points a and c are illustrated in Figure 7.6, which shows how winter densities are higher than summer, and shows the increase in electron density with increasing solar activity.

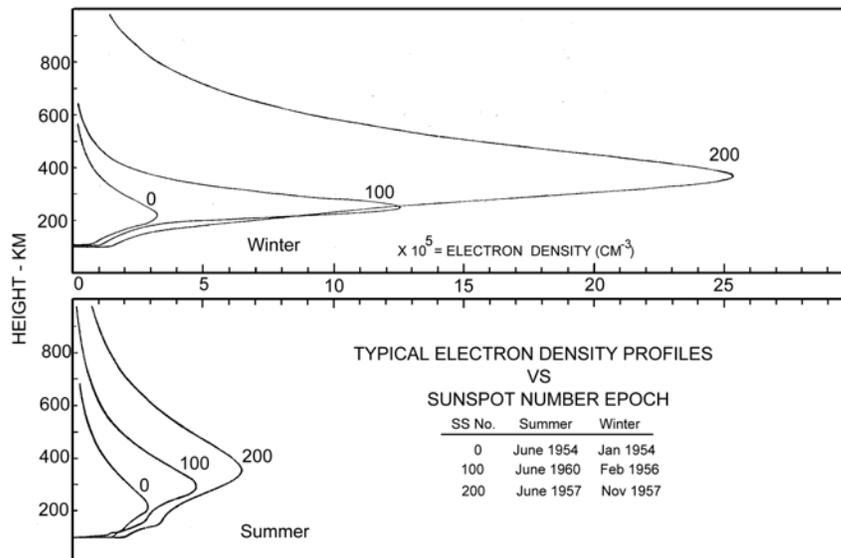


Figure 7.6 Typical electron density profiles at three (Zurich) sunspot epochs for winter and summer noon conditions at Washington DC (Belvoir). Magnetically quiet days were chosen for the above profiles. J. W. Wright, Dependence of the Ionospheric F Region on the Solar Cycle, *Nature*, page 461, May 5, 1962.

Disturbances (Anomalies)

The departures from a simple solar control of the electron density in the F₂-region are referred to as anomalies. These anomalies are the result of several factors:

- a) marked seasonal and solar-cycle variation of the temperature and consequently the scale height H,
- b) seasonal variation of the O/O₂ and O/N₂ ratios,
- c) transport of electrons by diffusion, winds (global atmospheric circulation), electromagnetic forces ('fountain effect') etc.
- d) transport of electrons between geomagnetic conjugate points.

The F₂-region is also disturbed by solar flares. About 24-36 hrs after a solar flare, particle radiation from the sun causes $(N_e)_{\max}$ in the F₂-layer to increase for a few hours at high latitudes. This is followed by a decrease which lasts for several hours. Such ionospheric disturbances could seriously affect high-frequency communication and broadcasting, as well as OTHR.

D Ionosphere morphology - Latitude Structure

The ionospheric structure described above is descriptive of the mid-latitude ionosphere. Near the magnetic equator, and at high latitudes, there are substantial deviations from the simple forms described by Chapman theory. These are largely due to the influence of the magnetosphere on the earth's upper atmosphere.

The most prominent feature of the ionosphere is the aurora, illustrated at the beginning of the chapter in Figure 7-1. The aurora occur at high latitudes - typically near 60-65° magnetic latitude. The bright auroral display is most obvious to the human eye (and satellites such as DMSP) near local midnight. In fact, the aurora occurs in an oval, or ring shaped distribution centered on the magnetic poles, and extends to all longitudes. Figure 7-7 shows an example of the Dynamics Explorer 1 imagery which shows the auroral oval extending around the northern magnetic pole. When the DE-1 imagery emerged in late 1981, it confirmed concepts such as the auroral oval with remarkable scenes such as the one shown here, and demonstrated for the first time how the aurora extended into the dayside polar region.

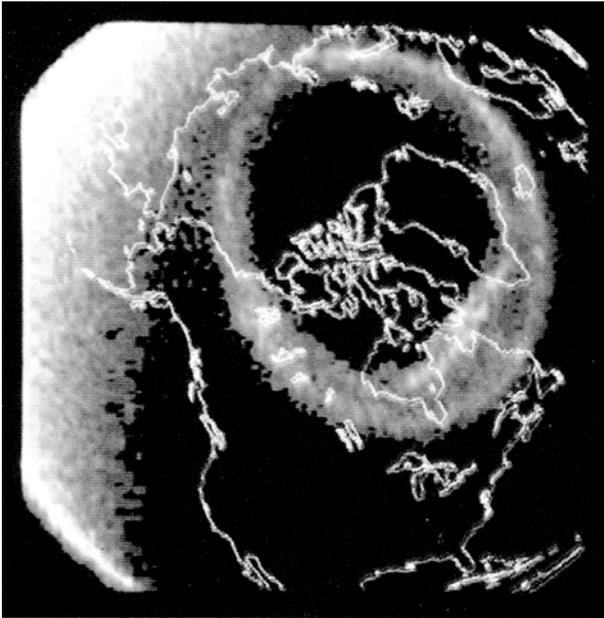


Figure 7.7 Auroral imagery from DE-1. This image was taken on 8 November 1981 by the Dynamics Explorer 1 Spin-Scan Auroral Imager. The filter used here passed the O_I wavelengths of 130.4 and 135.6 nm (far UV). A coastline map was superimposed on the image, which was taken at an altitude of several earth-radii above the northern polar cap. The illuminated hemisphere is to the left - the polar region is largely in darkness. These images were the first to show an uninterrupted look at the entire auroral oval.

Figure is from Plate 1 of, *Images of the Earth's Aurora and Geocorona from the Dynamics Explorer Mission*, L. A. Frank, J. D. Craven, and R. L. Rairden, *Advances in Space Research*, vol. 5, No. 4, pp. 53-68, 1985.

These auroral displays are produced by electrons from the plasma sheet, precipitating along magnetic field lines which extend in the night sector to $L = 5$ to 10. Some of the most intense displays map to geosynchronous orbit. The auroral electrons typically have energies of 1-10 keV, and often have a mono-energetic spectrum which in recent years has been interpreted as due to acceleration by electric fields pointed along the auroral magnetic field lines. (McIlwain first inferred this from sounding rocket measurements in 1960.) Figure 7-8 shows the topology described here.

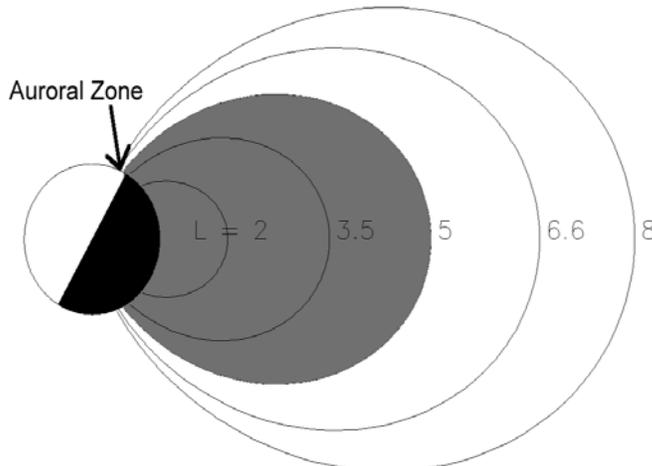


Figure 7.8. Diagram of auroral field lines. The sun is to the left. The gray shaded region is meant to illustrate the plasmasphere. Outside this region ($L > 5$ typically), is the plasma sheet. Plasma sheet electrons with energies of 1-20 keV move along the magnetic field lines. Those which fall into the (magnetic) loss cone precipitate into the atmosphere, producing a diffuse auroral glow. At the high latitude edge of this region, more intense aurora can be seen (visible to the human eye). This 'discrete' aurora is thought to result when plasma sheet electrons are accelerated to higher energies by electric fields aligned along the magnetic fields within a few thousand kilometers of the earth's surface. Magnetic fields poleward of this region map to the tail lobes, and perhaps out into the solar wind.

There is also a substantial structure at lower latitudes. Figure 7.9 shows unique imagery from early in the space age - the UV camera carried to the moon on Apollo provided a unique look at the earth's atmosphere. Thin bands of light are seen to extend from the daylight hemisphere into the dark hemisphere. These bands reflect the equatorial structure associated with the equatorial fountain - aka the Appleton Anomaly.

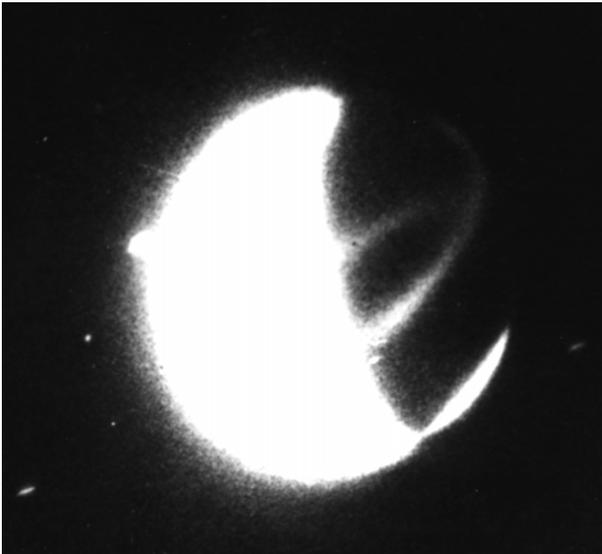


Figure 7.9. UV image of Earth, April 21, 1972. This image was recorded with the far UV camera carried on the Apollo 16 spacecraft. The image was acquired with a pass-band of 1250-1600 Å. The exposure time was 10 minutes. The bright regions are from atomic oxygen and molecular nitrogen emissions in the upper atmosphere. (Note the hydrogen geocorona is not visible in this image.) The sun is to the lower left, producing the bright airglow in the sunlit hemisphere. The southern auroral zone is the bright cap to the lower right; there is a less obvious brightening in the partially sunlit northern hemisphere auroral zone. The equatorial airglow bands extend into the night, curving from $\sim \pm 20^\circ$ latitude to near the equator at the right side.

George R. Carruthers and Thornton Page, Apollo 16 Far-Ultraviolet Camera/Spectrograph; Earth Observations, *Science*, 1 September 1972, volume 177, pp 788-791.

This remarkable image reflects an important latitudinal structure in electron density. Figure 7.10 shows this structure, and the substantial change in electron density observed near the equator. This structure is due to the solar driven upwelling of the neutral atmosphere at the equator, which in turn forces an upward flow of the plasma there. This has been termed the equatorial fountain. The plasma drifts back downward 5-10 degrees away from the equator. These density variations, and the turbulence associated with the equatorial fountain, make the low latitudes a region in which it is difficult to utilize equipment such as OTHR.

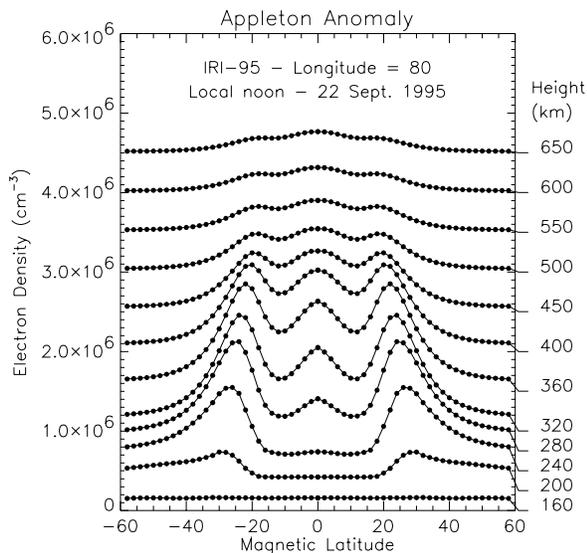


Figure 7.10 Equatorial density profiles. The profiles shown here were obtained from the IRI-95 model, run for 22 September 1995. The resulting latitude profiles were plotted beginning at the bottom with a profile at 160 km altitude. Subsequent plots were offset upward by 0.25×10^6 (up to 320 km) and then by 0.5×10^6 . The zero points are indicated by the short horizontal lines leading from the plot to the altitude labels

E Radiowave propagation in the Ionosphere

When an electromagnetic wave is incident on the ionosphere the oscillating electric field of the wave causes the electrons to oscillate with respect to the much heavier and therefore relatively stationary positive ions. This process results in the refraction or reflection of radiowaves (depending on their frequency) from the ionosphere as we shall see in more detail below. In general the influence of the ionosphere on wave propagation decreases with increasing frequency and waves of frequency higher than the so called "plasma" frequency will pass through the ionosphere and reach a satellite. Because of its reflective properties at sufficiently low radio frequencies the ionosphere behaves like a "mirror in the sky" and is used for short wave communications, OTHR (over the horizon radar) and radio navigational systems.

High frequency (3 - 30 MHz) radio waves are reflected from the dayside E region or the F region (200-300 km altitude) while very low frequencies (less than 30 kHz) are reflected from the D region (70-100 km altitude). Radio waves are also partially absorbed by the ionosphere as a result of electron collisions with air molecules. This is particularly true in the D region where the density of neutral air molecules is largest. The increased D region ionization caused by an intense solar flare can result in the complete absorption of high frequency waves and disruption of short wave communication, particularly over the sunlit hemisphere of the earth.

To understand how a radiowave is reflected from the ionosphere we need to develop a few concepts:

(1) The refractive index of a medium n is defined as: $n = \frac{c}{v_p}$

where :

c = the velocity of an electromagnetic wave in a vacuum

= 3×10^8 m/s, and

v_p = the phase velocity of the wave in the medium

In most substances such as glass or water $n > 1$ and n varies with frequency. It is greater for violet light (high f) and smaller for red light (lower f).

In a plasma medium the index of refraction (neglecting the geomagnetic field) is:

$$n = \sqrt{1 - \frac{Ne^2}{\epsilon_0 m \omega^2}} = \sqrt{1 - \frac{80.6 N}{f^2}} \quad (\text{Eqn. 7.13})$$

where N = electron density in (electrons/ m³) and f is the frequency of the wave in Hz. (In this section of the text, we must deviate from the previous practice of using the lower case "n" for electron density, since "n" is used for the index of refraction.)

This formula exhibits a characteristic frequency which determines the boundary between a real and an imaginary index of refraction. This characteristic frequency, known as the plasma frequency, is defined by the equation

$$f_p = \sqrt{80.6 N} \quad \text{or} \quad \omega_p = \sqrt{\frac{Ne^2}{\epsilon_0 m}} \quad (\text{Eqn. 7.14}).$$

This represents the natural frequency of oscillation of the electrons around the stationary ions. If, in contrast with the behavior of light in dielectric materials such as glass. If , the index of refraction is imaginary. This means that in the plasma medium, electromagnetic waves cannot propagate below a critical frequency for which the medium is conducting.

(2) Snell's Law tells us how a wave is refracted (bent) at the interface between two media.

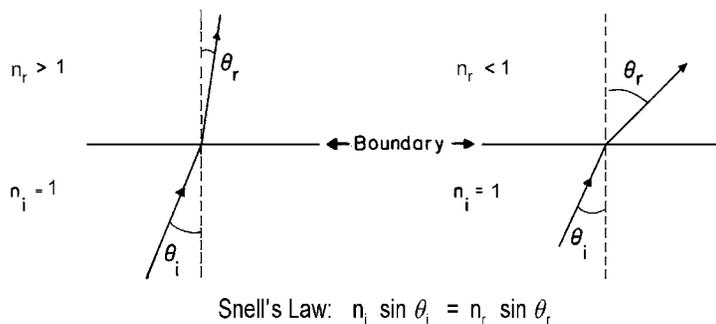


Figure 7.8 Snell's Law - Single Boundary

Consider now a layered medium in which each layer has a lower index than the one below it as shown.

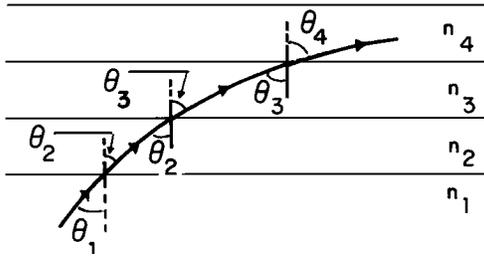


Figure 7.9 Snell's Law - Continuously varying index of refraction

Applying Snell's Law at each boundary we obtain

$$\frac{n_2}{n_1} = \frac{\sin \theta_1}{\sin \theta_2} \quad \frac{n_3}{n_2} = \frac{\sin \theta_2}{\sin \theta_3} \quad \frac{n_4}{n_3} = \frac{\sin \theta_3}{\sin \theta_4} \text{ etc.}$$

This can also be written in the form

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 = n_3 \sin \theta_3 = \dots \text{ or in general}$$

$$n \sin \theta = \text{constant} \quad (\text{Eqn. 7.15})$$

Let us now apply this to the ionosphere as shown in figure 7.10

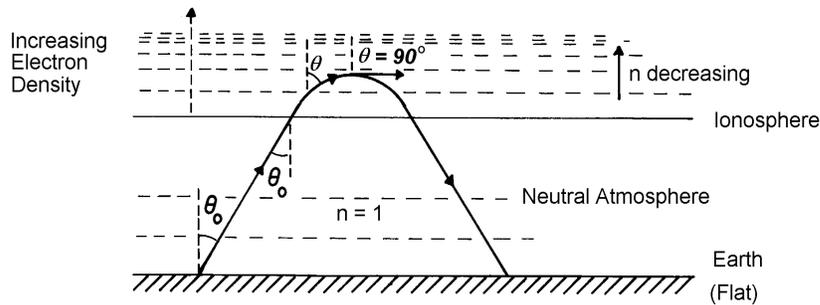


Figure 7.10 Snell's Law in the ionosphere

Eqn. 7.15 must hold for all points of the ray path. Before the wave enters the ionosphere $\theta = \theta_0$ and $n = 1$. Thus the angle of incidence at the base of the ionosphere is also θ_0 as can be seen from the sketch.

As we enter the ionosphere the electron density increases and the index of refraction begins to decrease below 1. To satisfy Snell's Law the angle θ must increase and the ray bend further and further away from the vertical. In order to have reflection take place θ must go to 90° at the top of the trajectory. We can calculate what the electron density must be for this to take place

$$n_m \sin 90^\circ = (1) \sin \theta_0 \quad \text{or} \quad n_m = \sin \theta_0 \quad (\text{Eqn. 7.16})$$

But we had above in eqn.s 7.13 and 7.14:

$$n = \sqrt{1 - \frac{Ne^2}{\epsilon_0 m \omega^2}} = \sqrt{1 - \frac{80.6 N}{f^2}} = \sqrt{1 - \frac{\omega_p^2}{\omega^2}} = \sqrt{1 - \frac{f_p^2}{f^2}} \quad (\text{Eqn. 7.17})$$

Thus the minimum electron density N_m needed to reflect this ray which left the ground at an angle θ_0 is given by

$$\sin \theta_0 = \sqrt{1 - \frac{80.6 N}{f^2}} = \sqrt{1 - \frac{f_p^2}{f^2}} \quad (\text{Eqn. 7.18})$$

where f = frequency of wave (Hz)

Solving for N_m we obtain (using $1 - \sin^2 \theta_0 = \cos^2 \theta_0$)

$$N_m = \frac{f^2}{80.6} \cos^2 \theta_0 \quad (\text{electrons/m}^3) \quad (\text{Eqn. 7.19})$$

If we now consider the special case of vertical incidence we have that $\theta_0 = 0$ which gives for the minimum required electron density N_0

$$\sqrt{1 - \frac{80.6 N_0}{f^2}} = \sqrt{1 - \frac{f_p^2}{f^2}} = 0 \quad (\text{Eqn. 7.20})$$

Thus we see that for vertical incidence $n = 0$ when the incident wave frequency equals the plasma (or critical) frequency.

To summarize: For vertical incidence reflection will occur at a level in the ionosphere where

- (a) $n = 0$ Index of refraction vanishes
- (b) $f = f_p$ radio frequency = plasma frequency
- (c) $v_p = \infty$ Phase velocity becomes infinite
- (d) $v_g = 0$ Group velocity = velocity of energy transport becomes 0.

If we now compare the minimum electron density required for reflection for the vertical and oblique cases we see that

$$\text{Oblique:} \quad N_m = \frac{f^2}{80.6} \cos^2 \theta_0 \quad (\text{Eqn. 7.21 a})$$

$$\text{Vertical:} \quad N_m = \frac{f^2}{80.6} \quad (\text{Eqn. 7.22 b})$$

A smaller electron density is required for oblique incidence ($\cos \theta$ is always ≤ 1)

If we look at the same problem from the point of view of the frequencies which will be reflected for given electron density at a certain height we find that the frequency f_θ of the oblique wave is related to the reflected frequency f_p at vertical incidence by

$$f_\theta = \frac{f_p}{\cos \theta} = f_p \sec \theta \quad (\text{Eqn. 7.23})$$

which shows that in general $f_\theta > f_p$.

If N_{\max} represents the maximum (peak) electron density in an ionospheric layer then the highest frequency that will be reflected at vertical incidence is called the "critical" frequency of the layer and is given by

$$f_o = \sqrt{80.6 N_{\max}} \approx 9\sqrt{N_{\max}} \text{ with } N_{\max} \text{ (electrons/m}^3\text{)} \quad (\text{Eqn. 7.24})$$

and following our earlier discussion we see that for oblique incidence the critical frequency will be

$$(f_\theta)_c = f_o \sec \theta \quad (\text{flat earth}) \quad (\text{Eqn. 7.25})$$

For large values of θ (approaching 90°) this leads to a substantial increase in the available frequencies for communications. Note that this formula does not take into account the earth's curvature, and further modifications are necessary for large θ .

The applicability of this result is illustrated in figure 7.11, an ionospheric sounding utilizing radar at frequencies of 1-10 MHz. The radar is swept in frequency (horizontal axis), and the range-gated returns (vertical axis) give the altitude at which the signal is reflected.

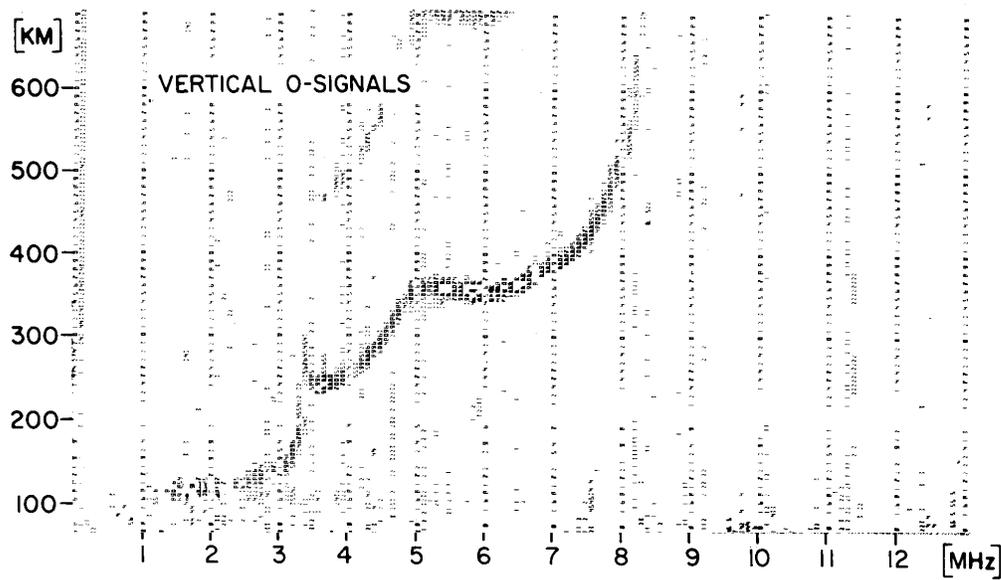


Figure 7.11 Data from a digital ionosonde at the AFGL Goose Bay Ionospheric observatory. These data were acquired on 16 June 1980 at 1720 AST. (local noon?) There is a standard set of 128 range bins at each frequency. Height resolution is 5 km, each frequency step is 100 kHz. Handbook of Geophysics, page 10-6. See also Bibl, K. and B. W. Reinisch, The universal digital ionosondes, Radio Science, vol. 13, pages 519-530, 1970.

F References

Friedman, Sun and Earth, Scientific American Press, New York, 1986

Kelley, Michael C, The Earth's Ionosphere, Academic Press, San Diego, 1989.

McIlwain, C. E., Direct measurement of particles producing visible auroras, Journal of Geophysical Research, 65, page 2727, 1960.

Ratcliffe, J. A., An Introduction to the ionosphere and magnetosphere, Cambridge University Press, 1972.

Handbook of Geophysics and the Space Environment, edited by Adolf S. Jursa, Air Force Geophysics Laboratory, 1985; Chapter 10, The Ionosphere, Section 10.1 Measuring Techniques, S. Basu, J. Buchau, F. J. Rich, and E. J. Weber

G Problems

1. Determine the production curve resulting from the photoionization of an atmospheric gas by monochromatic solar radiation incident vertically on the atmosphere. The rate of electron production at the altitude h is given by:

$$q(h) = I(h) \sigma n(h) \text{ electrons/m}^3 \text{ s}$$

where

$I(h)$ = intensity of the solar radiation (in photons/m²s) at the altitude h
 σ = ionization cross-section (in m²) of the gas molecules for the particular solar radiation. and
 $n(h)$ = number density (in m⁻³) of the gas molecules at the altitude (h)

(a) For n , use the barometric relation with a scale height of 30 km and $n = 5 \times 10^{16}$ molecules/m³ at 150 km altitude

(b) For the solar radiation, use $I(h) = I_0 e^{-(1000-h)/h_0}$, h_0 is the scale height for the decrease in solar intensity with altitude. Take $I_0 = 1.0 \times 10^{14}$, $h_0(\text{km}) = 10^{17}/n$. note that the last formula should be read: $h_0(\text{km}) = \frac{10^{17}}{\text{density}}$

(c) For the ionization cross section, take $\sigma = 2 \times 10^{-21}$

Plot the neutral density, $n(h)$, solar intensity, $I(h)$, and the production rate, $q(h)$ as a function of altitude for $h = 250$ to 800 km altitude.

2. Figure 7.12 shows an electron density profile. Estimate the scale height for plasma above 700 km altitude (Go from 700-1500 km). Use this scale height to estimate the temperature. Note that the answer depends upon which mass you use in the scale height formula $\left(H = \frac{kT}{mg}\right)$. Try the electron mass, proton mass, and the mass of atomic oxygen (16 AMU). Which is most reasonable?

3. From Figure 7.12 estimate the electron density at 200 km and 400 km. What are the radio frequencies that will be reflected at these altitudes assuming vertically traveling waves ?

4. Compare the neutral gas density and plasma (electron) density as a function of altitude. Plot the total neutral density (number/volume) as a function of altitude (e.g. figure 6.5b). On the same curve, plot the electron density.

5. Using the data found in Figure 7.11, estimate the electron density profile. ($\theta = 0$)

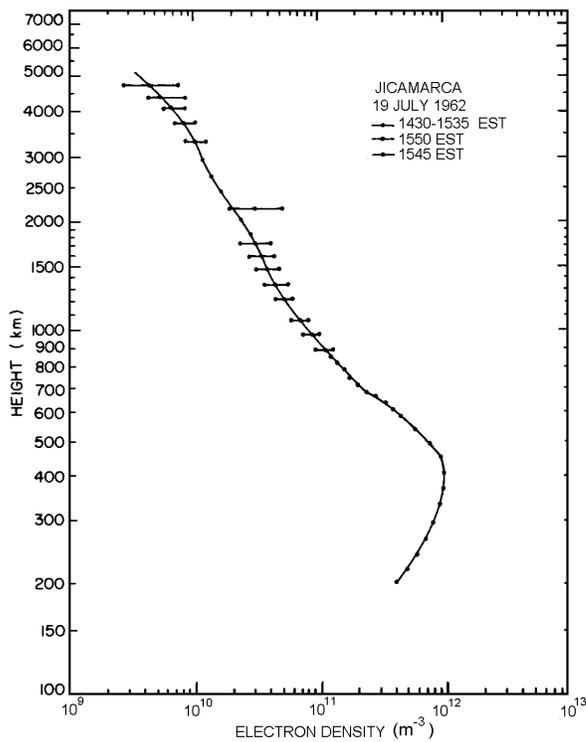


Figure 7.12 An electron density profile obtained at Jicamarca that extends to almost one earth radius. [Bowles, K. L. and Staff of Jicamarca Radio Observatory Institute, Profiles of Electron Density Over the Magnetic Equator Obtained Using the Incoherent Scatter Technique, *Tech Report Note 169*, National Bureau of Standards, Boulder, CO, 1963]. [Found in Handbook of Geophysics, page 10-16]

6. The atmosphere of a hypothetical planet is composed of nitrogen, N_2 , and oxygen, O_2 . Above 150 km, the atmosphere is isothermal at a temperature of 1500 K, and its composition varies with altitude as a result of diffusive separation.

The number densities of N_2 and O_2 are $5 \times 10^8 \text{ cm}^{-3}$ and $5 \times 10^7 \text{ cm}^{-3}$ respectively at 200 km; and their **absorption** cross-sections for solar radiation of wavelength 75 nanometers are $3.1 \times 10^{-17} \text{ cm}^{-2}$ and $2.9 \times 10^{-17} \text{ cm}^{-2}$ respectively.

The **ionization** cross-section of N_2 for the same radiation is $2.3 \times 10^{-17} \text{ cm}^{-2}$.

The acceleration due to gravity is 9.81 m/s^2 , and may be assumed to be independent of height.

If the solar irradiance at the top of the atmosphere is $2.5 \times 10^9 \text{ photons/cm}^2 \text{ s}$ at 75 nm, what would be the irradiance at 200 km when the radiation is vertically incident on the atmosphere?

Calculate the ionization rate of N_2 at 200 km due to the 75 nm radiation.

Assuming that the only loss process for N_2^+ is charge transfer to O_2 , and that this reaction rate is $1.5 \times 10^{-11} \text{ cm}^3 \text{ s}^{-1}$ calculate the PCE density of N_2^+

Take $\gamma = 1$ (one electron per incident photon).

Blank page