

# Terrestrial Ionospheres

Stanley C. Solomon

High Altitude Observatory  
National Center for Atmospheric Research  
Boulder, Colorado  
stans@ucar.edu

## Introduction

Early investigation of the ionosphere through its effect on radio waves resulted in description by means of layers, principally the *D*, *E*, and *F* layers, the latter subdivided into  $F_1$  and  $F_2$ . This terminology continues to influence our current concept of the nature of energy deposition in atmospheres, although the misleading “layers” has given way to “regions”. It arose from the observation of systematic variation in the height at which the critical frequency of reflection occurs in ionospheric radio sounding; this method cannot detect ionization above the peak of a region, which explains the appearance of layers. Radar and spacecraft measurements now give a more complete picture of peaks and valleys and reveal the complex morphology of the ionosphere. *Chamberlain and Hunten* [1987] provide a referenced discussion of the historical literature. An overview of the altitude dependence and variability of Earth’s ionosphere is given in Figure 1, showing the diurnal and solar-cycle changes and the locations of the named regions. Space and planetary exploration has also found that the Earth’s ionosphere is unique, just as its atmosphere is unique, and for some of the same reasons.

An additional historical artifact in terminology is the ionosphere itself. Because it was discovered before the neutral thermosphere in which it is contained, anything above the stratosphere is often referred to as the ionosphere, resulting in a common misconception that this region of the atmosphere is mostly ionized. In fact, it is mostly neutral, ranging from less than a part in a million ionized during the day at 100 km altitude to about 1% ionized at the exobase (~600 km, depending on solar activity). Even at 1000 km, there is only of the order of 10% ionization. At several thousand km, where ions (mostly protons) finally become dominant, the region is defined as the plasmasphere. Figure 2 shows the results of two standard semi-empirical models of the thermosphere and ionosphere, the Mass Spectrometer Incoherent Scatter (MSIS) model [*Hedin*, 1987] and the International Reference Ionosphere (IRI) [*Bilitza*, 1990], for typical daytime conditions at moderate solar activity. Charge neutrality is assumed, so the electron density profile is the sum of all the ion density profiles. Only the major ions are shown. It is clear from this plot that  $O^+$  is the most important ion, particularly in the extensive  $F_2$  region above ~200. The  $F_1$  region from ~150 to ~200 km appears as a mere plateau in the profile, but is distinguished by a transition to molecular ions, particularly  $NO^+$ . The low levels of  $N_2^+$ , given the dominance of  $N_2$  at these altitudes, is noteworthy. The *E* region from ~100 to ~150 km exhibits a small peak, dominated by  $O_2^+$  and  $NO^+$ . The weak ~70 to ~90 km *D* region is not shown on this plot.

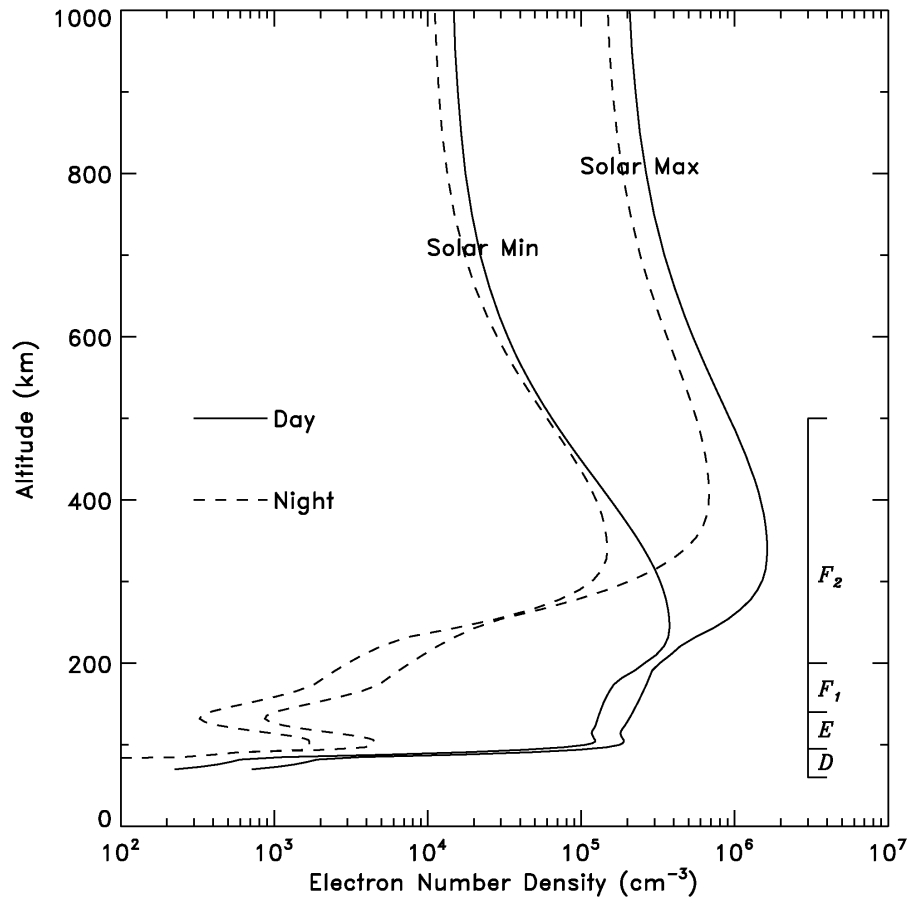


Figure 1. Overview of the altitude distribution of Earth's ionosphere for daytime and nighttime conditions, at high and low solar activity. From the International Reference Ionosphere (IRI) empirical model [Bilitza, 1990].

At night, the ionosphere below 200 km almost disappears. Small residuals remain, and sporadic layers of metallic ions in the  $E$  region sometimes appear, but basically the night ionosphere is confined to the  $F_2$  region. The peak electron density erodes and its altitude varies, but the  $F_2$  ionosphere persists through the night. At auroral latitudes, a new  $E$  region can occur, co-located with auroral emission features. Sometimes the  $F_2$  region is enhanced in the presence of aurora; at other times it is greatly attenuated.

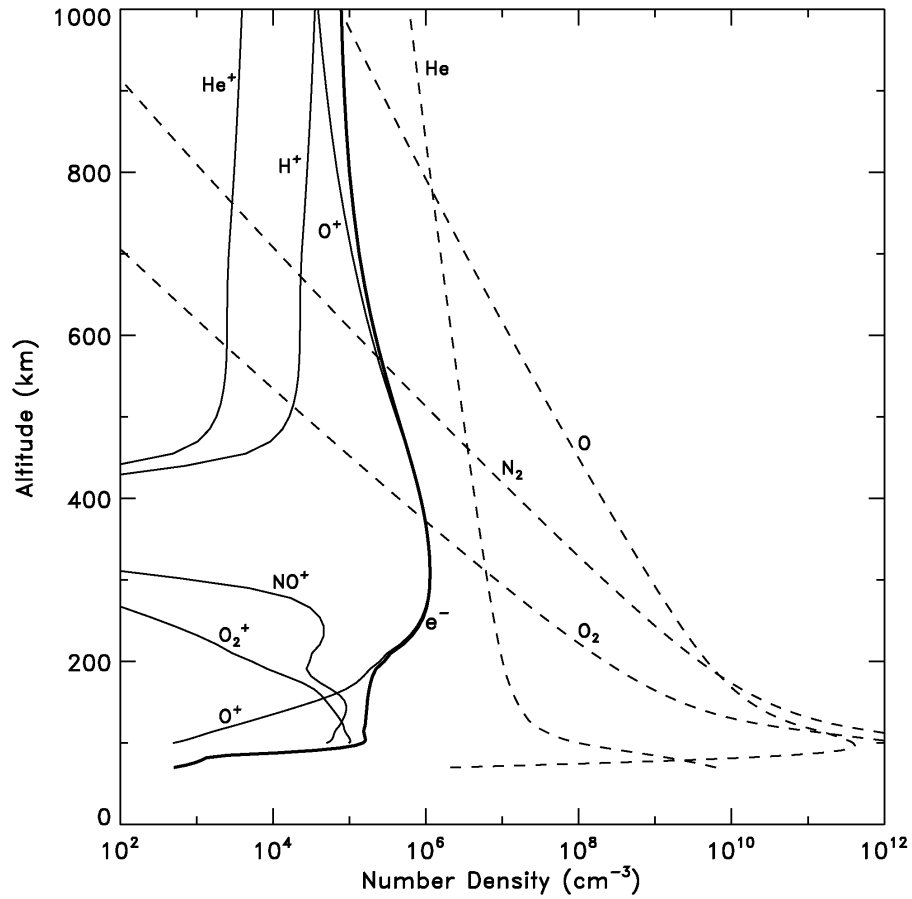


Figure 2. Semi-empirical models of the thermosphere/ionosphere: the Mass Spectrometer Incoherent Scatter (MSIS) model [Hedin, 1987] and International Reference Ionosphere (IRI) [Bilitza, 1990].

### Ionization

The ionosphere is created by ionizing radiation, including extreme ultraviolet (EUV) and X-ray photons from the sun, and corpuscular radiation that is mostly energetic electrons and mostly occurs at high magnetic latitude as auroral “precipitation.” The solar photon output at these wavelengths, from  $\sim 1$  nm to the H I Lyman- $\alpha$  line at 121.6 nm, varies by factors ranging from  $\sim 2$  to  $>100$  over the 11-year solar activity cycle, and is additionally variable on shorter time scales, including especially the 27-day solar rotation period. This causes dramatic variation in the temperature and density of the thermosphere and ionosphere. Changes in the solar wind and interplanetary magnetic field also affect the thermosphere/ionosphere through geomagnetic perturbations that result in transfer of energy from the magnetosphere, both in the form of auroral particle ionization and in the form of heat from the resulting currents imposed in the polar regions. An additional form of energy transfer is the generation of energetic electrons released in the ionization process. These electrons, referred to as photoelectrons in the case of photoionization and secondary electrons in the case of particle impact ionization, have enough energy to excite, dis-

sociate, and further ionize the neutral atmosphere as well as heat the ambient plasma. Solar ionization and its by-products provide most of the ionization and heating of the thermosphere, and account for most of its 11-year and 27-day variation, but geomagnetic activity accounts for much of the shorter term variation on time scales from hours to days.

The details of ionospheric formation can be explained through examination of the photoionization and photoabsorption cross sections of thermospheric constituents. The ionization continua of  $N_2$ , O, and  $O_2$  all peak in the vicinity of 60 nm at tens of Megabarns ( $10^{-18} \text{ cm}^2$ ). This causes their energy to be deposited largely in the  $F_1$  region. This can be calculated for any particular wavelength by applying Beer's law to an exponential atmosphere, which was done by *Chapman* [1931]. For a single-constituent, planar, isothermal atmosphere with density  $n_0$  at some reference altitude  $z_0$  and scale height  $H$ , a simplified form of the Chapman production function is:

$$q_z = I_\infty \sigma_i n_0 \exp \left[ -\frac{z - z_0}{H} - \tau_z \right]$$

where  $q_z$  is the ionization rate as a function of altitude,  $I_\infty$  is the unattenuated ionization frequency,  $\sigma_i$  is the ionization cross section, and  $\tau_z$  is the optical depth. This function peaks at the altitude of unit optical depth, and reduces rapidly below that altitude. The smaller the cross section, the larger the column density at unit optical depth, and the lower the peak. Shortward of 60 nm, cross sections decrease and the radiation penetrates to lower altitude. The intense solar He II emission at 30.4 nm deposits most of its energy near 150 km and 1-10 nm soft X-rays can penetrate to 100 km. Most of the  $E$  region is produced by longer wavelength radiation, however, particularly the C III line at 97.7 nm and the H I Lyman- $\beta$  line at 102.7 nm. These do not have enough energy to ionize  $N_2$  and O but penetrate through gaps in the  $N_2$  absorption spectrum to ionize  $O_2$  to  $O_2^+$ . Longward of 103 nm, only the important minor species NO has a low enough ionization potential to be ionized by solar radiation. H I Lyman- $\alpha$  happens to fall at a low point in the  $O_2$  absorption spectrum and so penetrates below 90 km, where ionization of NO to  $NO^+$  and subsequent products create the  $D$  region. Thus, while the Chapman production function is approximately correct for any species at each wavelength, ionized regions are created by the superposition of many such functions for different species and wavelengths. In general, the more polychromatic the absorbed irradiance, the broader the resulting region of energy deposition.

## Recombination

Positive ions and electrons have generally fast collision rates, so one would suppose that ionospheric production would be balanced by recombination and that the ions would be short-lived after sunset. However, atomic ions colliding with electrons have the problem common to all two-body reactions that a single molecule is unlikely to result, because there is nothing to carry away surplus kinetic energy. Photon emission following collision of an atomic ion with an electron can stabilize the resulting atom; this *radiative recombination* is quite slow, with rate coefficients of the order of  $10^{-12} \text{ cm}^3 \text{ s}^{-1}$ . Although radiative recombination occurs and is important in the highest reaches of the ionosphere, it is insufficient as a loss mechanism for ions and electrons given their observed  $F$  region densities. Since the solar ionization frequency is  $\sim 10^{-6} \text{ s}^{-1}$ , ion densities would be several orders of magnitude larger than observed if radiative recombination were the only loss mechanism. The recognition by *Bates and Massey* [1947] that atomic ions must yield their charge to molecular ions in order to undergo rapid *dissociative recombination* is regarded as seminal. Dissociative recombination, schematically  $XY^+ + e^- \rightarrow X + Y$ , has

rate coefficients of the order of  $10^{-7} \text{ cm}^3 \text{ s}^{-1}$  and is the fundamental loss mechanism for ions in planetary ionospheres. In order for an atomic ion to become a molecular ion, *atom-ion interchange*, schematically  $X^+ + YZ \rightarrow XY + Z^+$ , or *charge exchange*, schematically  $X^+ + YZ \rightarrow X + YZ^+$ , must occur. Charge exchange reactions are typically fast if energetically possible, but atom-ion interchange rates depend on the nature of the reacting molecule, since a bond must be broken.

In regions of the atmosphere where molecules dominate, recombination chemistry is simplified since it is essentially a balance between ionization and dissociative recombination. Again following Chapman's work, a common approximation is use of an effective recombination rate coefficient  $\alpha_{\text{eff}}$ , the density-weighted average of the ion recombination rates. In photochemical equilibrium, the production rate  $q = \alpha_{\text{eff}} [M^+] [e^-]$  where  $M^+$  is the sum of the ions. Assuming charge neutrality, this yields  $[e^-] = (q/\alpha_{\text{eff}})^{1/2}$ . Applying the Chapman production function to obtain  $q$  results in a Chapman "layer", considering as above the caveats associated with use of that term. Thus, in molecular atmospheres, electron density varies approximately as the square root of the ionization rate profile. This is a particularly useful form for auroral ionization, where electrons (and sometimes protons or heavier ions) penetrate to  $\sim 100 \text{ km}$  or lower.

Although the  $F_2$  region has some of the morphological appearance of this type of layer, it is at the wrong altitude, and in the atom-dominated region. It is not a Chapman layer at all, but a result of diffusive processes.  $O^+$  has an increasingly long lifetime as altitude increases and the molecular fraction of the thermosphere decreases. Above  $200 \text{ km}$  it becomes subject to diffusion, but is still chemically controlled up to the peak of the  $F_2$  layer near  $300 \text{ km}$ . Above this altitude, ambipolar diffusion takes over, where "ambipolar" refers to the effect of electrical attraction between the ions and nearly massless electrons, resulting in a scale height for  $O^+$  about twice that of  $O$ . The  $F_2$  region varies in response to thermospheric winds and electric fields, so the mid-latitude and equatorial ionosphere can be greatly influenced by auroral processes at high latitudes through their effect on thermospheric dynamics.

Figure 3 is provided as a guide to understanding the ion-neutral chemical processes described above. It is a greatly simplified schematic, but contains the essential species and reactions necessary to describe ionospheric photochemistry from  $100\sim 600 \text{ km}$ . The vertical direction implies approximate ionization potential. Ionization occurs primarily on the three major species  $N_2$ ,  $O$ , and  $O_2$  by photon, photoelectron, and auroral particle impact.  $N_2^+$  quickly loses its charge through dissociative recombination, atom-ion interchange with  $O$  to make  $NO^+$ , and, at lower altitude, charge exchange with  $O_2$  to make  $O_2^+$ . Thus it is always low in density and negligible in the absence of production.  $O^+$  loses its charge by atom-ion interchange with  $N_2$  and  $O_2$ . Reaction with  $N_2$  is slow,  $\sim 10^{-12} \text{ cm}^3 \text{ s}^{-1}$ , because of the high strength of the  $N\equiv N$  bond. Reaction with  $O_2$  is faster,  $\sim 10^{-11} \text{ cm}^3 \text{ s}^{-1}$ , but there is far less  $O_2$  available. This is why  $O^+$  is long-lived in the Earth's ionosphere, and why the  $F_2$  region exists.  $O_2^+$  loses its charge through dissociative recombination or through reaction with the odd nitrogen species  $NO$  and  $N$ , which control the balance between  $O_2^+$  and  $NO^+$  in the  $E$  region.  $NO^+$ , daughter of all the above and the "terminal ion", is subject only to dissociative recombination. Figure 3 thus describes a machine for dissociating molecular gases. Ionization goes in the top, and dissociation comes out the bottom, because dissociative recombination is the only significant way out.

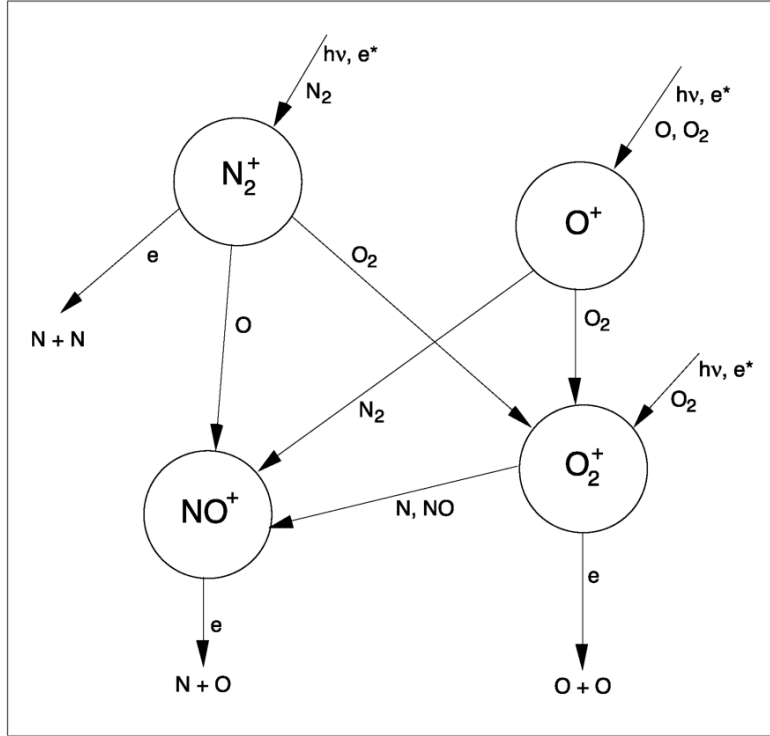


Figure 3. Simplified diagram of ionospheric chemistry in the Earth's upper atmosphere.

Figure 3 does not describe the ionosphere at very high altitude because it neglects hydrogen, which reacts by charge exchange with  $O^+$  to make  $H^+$ .  $N^+$  is also a significant minor ion, created by photodissociative ionization of  $N_2$ , that is neglected here. Doubly-ionized species are also ignored in this simplification. Although ground state  $O^+(^4S)$  does not have enough energy to make  $N_2^+$ , metastable  $O^+(^2D)$  and  $O^+(^2P)$  are created by photon and electron impact ionization, and these can charge exchange with  $N_2$  to form  $N_2^+$ . It is possible that vibrational excitation of thermospheric  $N_2$  can also accelerate the reaction of  $O^+$  with  $N_2$ .

In the *E* region, there is a complex interplay between that  $O_2^+$  and  $NO^+$ , due to the involvement of odd nitrogen with the ion chemistry, since  $O_2^+$  is converted to  $NO^+$  by reaction with NO and N. NO in particular is highly variable with solar activity, geomagnetic activity, and location, so this is a complicated problem. Older empirical and theoretical models which assumed that  $O_2^+$  is the dominant *E* region ion, due to its production by solar H I Lyman-*B* radiation, have been superseded by evidence that  $NO^+$  is generally observed to be the dominant *E* region ion, and considerable recent observational and modeling advances in understanding the high levels of NO and its importance to radiational cooling as well as ion chemistry have occurred.

In the *D* region, ion chemistry is entirely different due to the higher neutral density which allows three-body attachment, particularly  $2O_2 + e^- \rightarrow O_2 + O_2^-$ . This sets in motion a complicated negative-ion chain involving carbon, nitrogen, and hydrogen compounds, including water, that finally results in *mutual neutralization* of negative and positive ions, schematically  $M^+ + M^- \rightarrow M + M$ .  $NO^+$  created by H I Lyman- $\alpha$  ionization also initiates an involved positive-ion sequence, again involving hydration processes.

For the  $F$  regions, Figure 3 is a relatively sound synopsis nonetheless. There have been many recent advances, but most of the fundamental chemistry was worked out during the Atmosphere Explorer program, as reviewed by *Torr and Torr* [1978].

### Venus and Mars

The terrestrial planets, Venus, Earth, and Mars, are so named due to their fundamental similarity, and are presumed to have had common elemental origins. However, their subsequent evolution varied, due to their differing distance from the Sun, the smaller size of Mars, and the lack of rotation of Venus. Thus, their atmospheres are entirely different, and so are their upper atmospheres and ionospheres. Early exploration of Venus and Mars found that instead of persistent, high-altitude,  $F_2$ -type ionospheres, these planets had less dense, lower-altitude ionospheres that more resembled Chapman “layers”, were greatly attenuated at night, and consisted mostly of  $O_2^+$  and other molecular ions [e.g., *Schunk and Nagy*, 1980]. The presence of  $O_2^+$  seems especially perplexing, since Earth is the planet we generally associate with the unusual and quite reactive oxygen molecule. At higher altitude,  $O^+$  becomes an important species in the ionospheres of Venus and Mars, as on Earth, but at significantly lower density and without the same degree of persistence throughout the night.  $CO_2^+$  is a minor ion on both planets. There is a basic similarity in their ionospheres, despite the vastly different density of their lower atmospheres. Figures 4 and 5 provide an overview of the typical altitude structure.

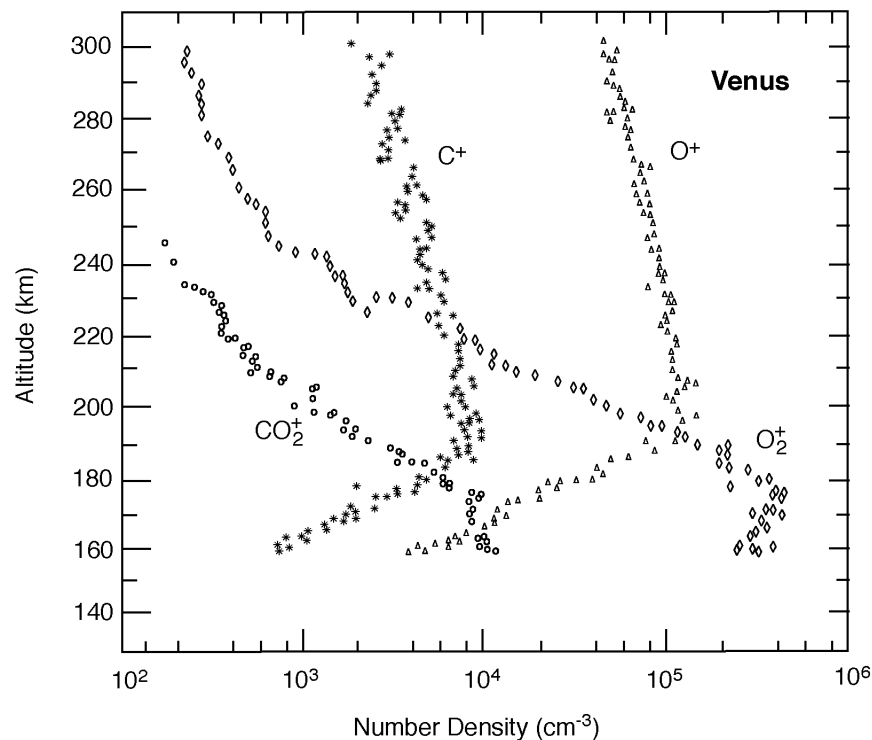


Figure 4. Measurements of the ionosphere of Venus from the ion mass spectrometer on the Pioneer Venus spacecraft [Taylor et al., 1980].

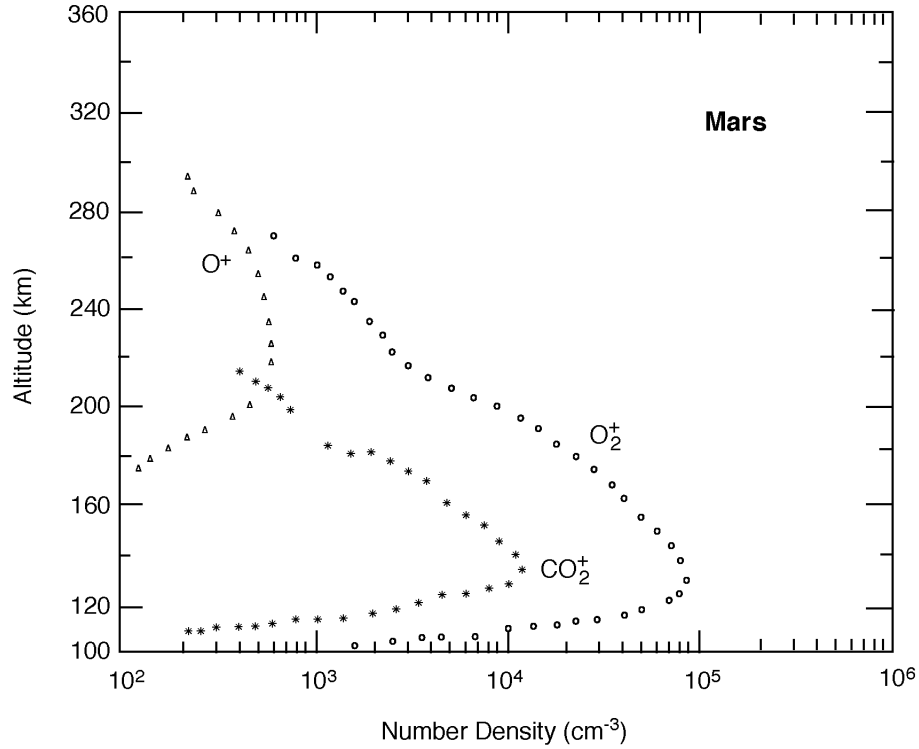


Figure 5. Measurements of the ionosphere of Mars from the ion mass spectrometer on the Viking-1 spacecraft [Chen et al., 1978].

The reason that the ionospheres of Venus and Mars are different from that of Earth is that the molecular compositions of their atmospheres are different, and therefore the compositions of their thermospheres are different. Table 1 gives a simple overview of the abundance of the primary atmospheric gases in the three terrestrial planets.

Aside from the large differences in surface pressure, the atmospheres of Venus and Mars are similar in composition, and  $N_2$  is an important species on all three planets.  $N_2$  requires more energy to dissociate than the oxygen compounds, however, so at thermospheric altitudes, atomic oxygen becomes important on all three planets. The Venus and Mars thermospheres are distinguished by high levels of  $CO_2$  (and also  $CO$ ) due to the underlying atmospheric composition, as shown in figures 6 and 7.



<i>Planet</i>	<i>Molecule</i>	<i>Abundance (bars)</i>	<i>Fraction of Total</i>
Venus	CO <sub>2</sub>	86.4	0.96
	N <sub>2</sub>	3.2	0.035
	Ar	0.006	0.00007
	H <sub>2</sub> O	0.009	0.00001
Earth	N <sub>2</sub>	0.77	.77
	O <sub>2</sub>	0.21	0.21
	H <sub>2</sub> O	0.01	0.01
	Ar	0.0093	0.0093
	CO <sub>2</sub>	0.0004	0.0004
Mars	CO <sub>2</sub>	0.0062	0.95
	N <sub>2</sub>	0.00018	0.027
	Ar	0.00010	0.016
	H <sub>2</sub> O	4 x 10 <sup>-7</sup>	0.00006

Table 1. Atmospheric composition of the terrestrial planets [Goody and Walker, 1972].

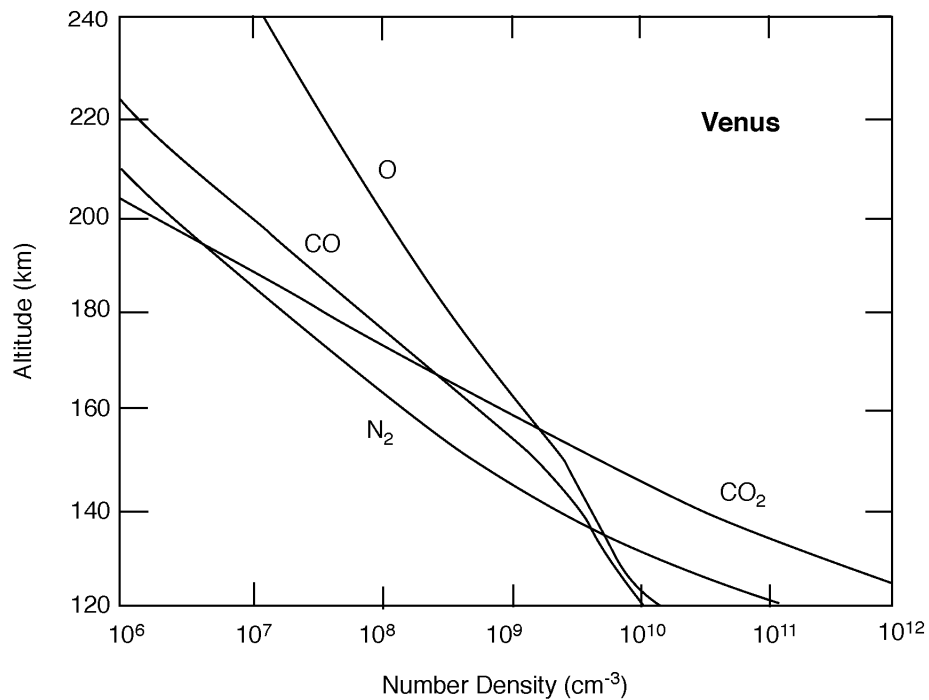


Figure 6. Typical composition of the Venus thermosphere, Fox and Bougher [1991], after Hedin [1983].

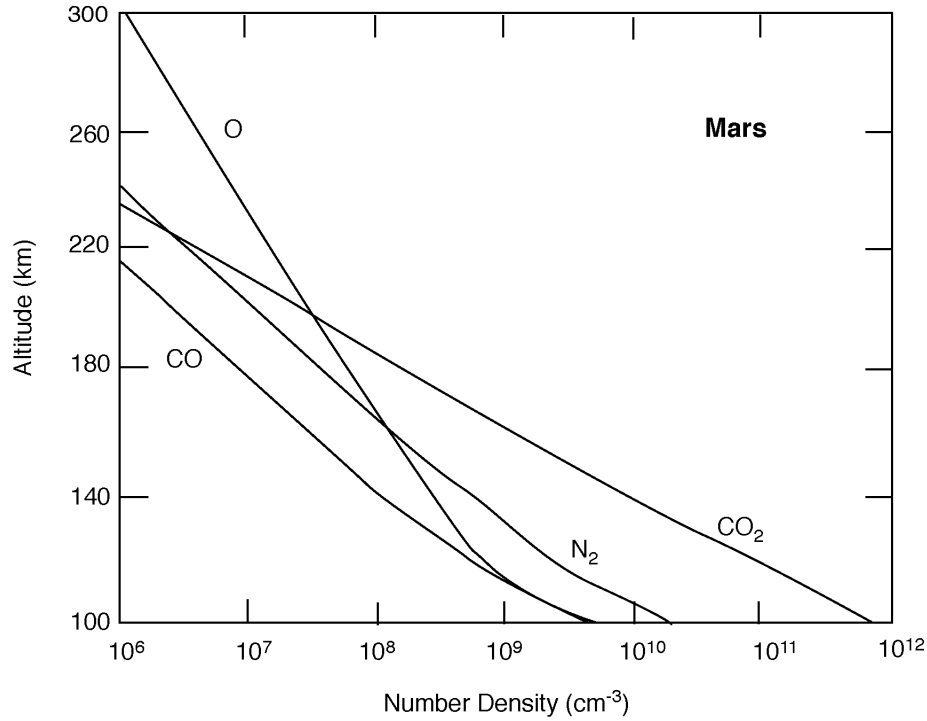


Figure 7. Typical composition of the Mars thermosphere, Chen et al. [1978], after Nier and McElroy [1977].

On Earth,  $O^+$  is a long-lived species in the high ionosphere because the  $O^+ + N_2$  reaction is so slow and there are few other molecules to react with to make a short-lived molecular ion. On Venus and Mars, the reaction  $O^+ + CO_2$  is quite fast,  $\sim 10^{-9} \text{ cm}^3 \text{ s}^{-1}$ , because  $CO_2$  is much less strongly bound than  $N_2$ . (The  $N \equiv N$  bond is the strongest known chemical bond in nature.) The reaction of  $O^+ + CO_2$  produces  $O_2^+$ , which is also produced by the reaction  $CO_2^+ + O$ . Ionization of the major thermospheric gases on Venus and Mars,  $O$  and  $CO_2$ , is thus quickly converted into  $O_2^+$ , which dissociatively recombines, resulting in the observed ionospheric morphology, lacking a significant  $F_2$  region. A simplified schematic of these processes is shown in figure 8. Thus, ironically, although life-supporting Earth is the planet associated with  $O_2$ , Venus and Mars are the planets with  $O_2^+$  ionospheres.

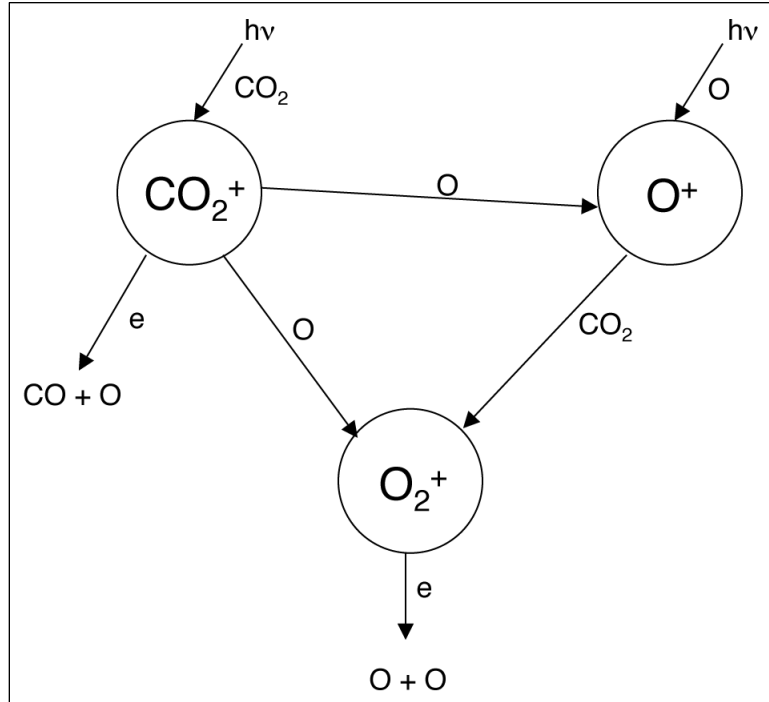


Figure 8. Simplified schematic of ion chemistry on Venus and Mars.

The  $F_2$  ionosphere is unique to Earth among the known planets. This is due to its peculiar atmosphere, lacking in  $\text{CO}_2$ , dominated by  $\text{N}_2$ , and carrying its oxygen in unusual and reactive states. Venus and Mars have nitrogen as well, but carbon and oxygen dominate their upper atmospheres, so it has little effect. Earth has a significant carbon budget, and once had much higher levels of  $\text{CO}_2$  in its atmosphere, but most of its carbon is currently locked up in the crust in the form of carbonate rocks. Thus, the  $F_2$  ionosphere may be a recent event in the history of Earth, an artifact of geology and biology.

Ion density at high altitude is relatively easy to detect and measure in planetary exploration due to its effect on radio signals. Thus, it is intriguing to speculate what an  $F_2$ -type ionosphere would signify if observed on some other planet. A full exposition of the processes which removed carbon from the Earth's atmosphere is beyond the scope of this discussion, but there are significant inorganic processes that extract dissolved carbon from the ocean and deposit it in rocks, which would continue to dominate even if the well-known biological pathways for absorption of  $\text{CO}_2$  from the atmosphere did not exist. Therefore, although it cannot be claimed that the  $F_2$  ionosphere is an indicator of biological processes, it may be related to the presence of an ocean, which is at least a precursor, if not a prerequisite, for the development of replicating or even aerobic chemical compounds. At the current state of Earth's evolution, with a small but ongoing effort underway to restore our planet's carbon balance to its natural state, it may be pertinent to consider just how unusual our atmosphere has become, and whether such a configuration of neutral and ionized gases can long endure.

## References

- Bates, D. R., and H. S. W. Massey, The basic reactions of the upper atmosphere II, The theory of recombination in the ionized layers, *Proc. Roy. Soc (London)*, *A192*, 1, 1947.
- Bilitza, D., ed., *International Reference Ionosphere 1990*, NASA publication NSSDC 90-22, Greenbelt, Maryland, 1990.
- Chamberlain, J. W., and D. M. Hunten, *Theory of Planetary Atmospheres*, Academic Press, 1987.
- Chapman, S., The absorption and dissociation or ionizing effect of monochromatic radiation in an atmosphere on a rotation Earth, *Proc. Phys. Soc.*, *43*, 26, 1931.
- Chen, R. H., T. E. Cravens, and A. F. Nagy, The Martian Ionosphere in Light of the Viking Observations, *J. Geophys. Res.*, *83*, 3871, 1978.
- Fox, J. L., and S. W. Bougher, Structure, luminosity, and dynamics of the Venus thermosphere, *Space Science Reviews*, *55*, 357, 1991.
- Goody, R. M., and J. C. G. Walker, *Atmospheres*, Prentice-Hall, Englewood Cliffs, N.J., 1972.
- Hedin, A. E., H. B. Niemann, W. T. Kasprzak, and A. Seif, Global empirical model of the Venus thermosphere, *J. Geophys. Res.*, *88*, 73, 1983.
- Hedin, A. E., MSIS-86 thermospheric model, *J. Geophys. Res.*, *92*, 4649, 1987.
- Schunk, R. W., and A. F. Nagy, Ionospheres of the Terrestrial Planets, *Rev. Geophys. Space Phys.*, *18*, 813, 1980.
- Nier, A. O., and M. B. McElroy, Composition and structure of Mars upper atmosphere: Results from the neutral mass spectrometers on Viking 1 and 2, *J. Geophys. Res.*, *82*, 4341, 1977.
- Taylor, H. A., Jr., H. C. Brinton, S. J. Bauer, R. E. Hartle, P. A. Cloutier, and R. E. Baniell, Global observations of the composition and dynamics of the ionosphere of Venus: Implications for the solar wind interaction, *J. Geophys. Res.*, *85*, 7765, 1980.
- Torr D. G., and M. R. Torr, Review of rate coefficients of ionic reactions determined from measurements made by the Atmosphere Explorer satellites, *Rev. Geophys. Space Phys.*, *16*, 327, 1978.