Ionospheric Propagation

The ionosphere exists between about 90 and 1000 km above the earth's surface. Radiation from the sun ionizes atoms and molecules here, liberating electrons from molecules and creating a space of free electron and ions. Subjected to an external electric field from a radio signal, these free and ions will experience a force and be pushed into motion. However, since the mass of the ions is much larger than the mass of the electrons, ionic motions are relatively small and will be ignored here.

Free electron densities on the order of 10^{10} to 10^{12} electrons per cubic metre are produced by ionization from the sun's rays. Layers of high densities of electrons are given special names called the D, E, and F layers, as shown in Figure 1. During the day the F layer splits into two layers called the F_1 and F_2 layers, while the D layer vanishes completely at night.

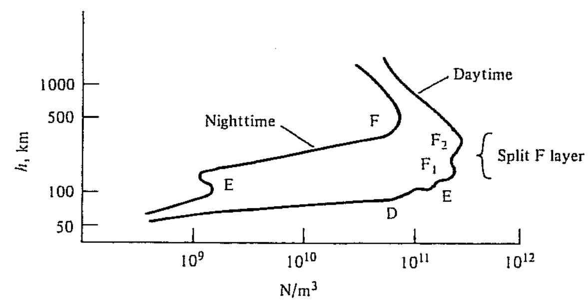

Figure 1: Electron density as a function of altitude, and various ionospheric layers

Radio waves below 40 MHz are significantly affected by the ionosphere, primarily because radio waves in this frequency range are *effectively reflected* by the ionosphere. The E and F layers are the most important for this process. For frequencies beyond 40 MHz, the wave tend to penetrate through the atmosphere versus being reflected.

The major usefulness of the ionosphere is that the reflections enable wave propagation over a much larger distance than would be possible with line-of-sight or even atmospheric refraction effects. This is shown graphically in Figure 2. The skip distance d_{max} can be very large, allowing very large communication distances. This is further enhanced by multiple reflections between the ionosphere and the ground, leading to multiple skips. This form of propagation allows shortwave and amateur radio signals to propagate worldwide. Since the D layer disappears at night, the best time for long-range communications is at night, since the skip distance is larger as the E, and F regions are at higher altitudes.

Where does the reflection come from? The reflections from the ionosphere are actually produced by refraction as the wave propagates through the ionosphere. The ionosphere is a concentrated region highly charged ions and electrons that collective form an ionized gas or plasma. This gas has a dielectric constant that is a function of various parameters, including the electron concentration and the frequency of operation. We now derive the dielectric constant of a plasma.

The electric field will produce a force on a given electron and displace it along a vector $\vec{r}$, as shown

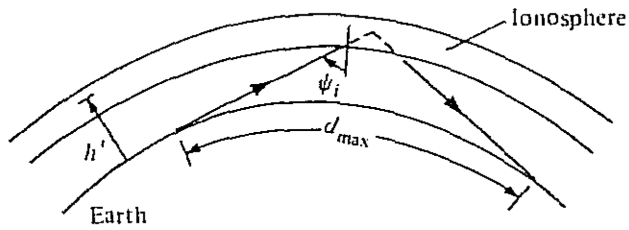

Figure 2: A single skip of a radio wave using the ionosphere

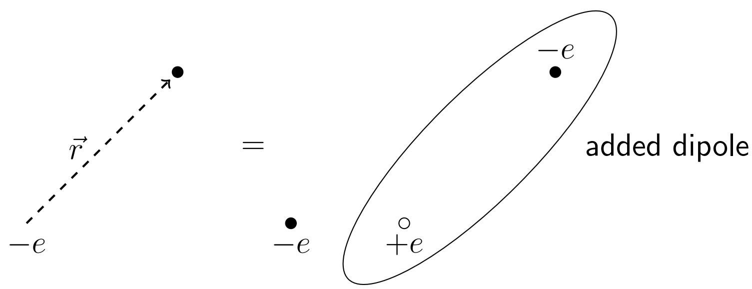

Figure 3: Representation of a moving electron as a dipole moment

on the left side of Figure 3. The displacement of an electron along this path can be modelled in an equivalent situation where the original electron remained stationary and an equivalent electric dipole is added, as shown in the right half of the figure. The dipole moment of this dipole is then equal to $-e\vec{r}$. If there are N electrons per unit volume, each displaced by $\vec{r}$ on average, then the volume polarization is

$$\vec{P} = -Ne\vec{r}. \tag{1}$$

The equation of motion for a single electron of mass $m_e = 9.109 \times 10^{-31}$ kg, charge $e = 1.6021 \times 10^{-19}$ C, with velocity $\vec{v} = d\vec{r}/dt$, and acted upon by an electric field $\vec{\mathcal{E}}$, is

$$m_e \frac{d\vec{v}}{dt} = -e\vec{\mathcal{E}} \tag{2}$$

The electron also experiences a frictional force resulting from collisions with neutral molecules. This force is added to the electric field force above, yielding

$$m_e \frac{d\vec{v}}{dt} = -e\vec{\mathcal{E}} - \nu m_e \vec{v} \tag{3}$$

where ν is the electron collision frequency. Re-writing the equation in terms of $\vec{r}$,

$$m_e \frac{d^2\vec{r}}{dt^2} = -e\vec{\mathcal{E}} - \nu m_e \frac{d\vec{r}}{dt} \tag{4}$$

We know that for sinusoidal fields, we can write equations in terms of phasors and replace d/dt with $j\omega$. Hence we can write the equation of motion on the electron in terms of phasors as

$$-\omega^2 m_e \boldsymbol{r} = -e\boldsymbol{E} - j\omega\nu m_e \boldsymbol{r} \tag{5}$$

or

$$\boldsymbol{r} = \frac{e\boldsymbol{E}}{\omega^2 m_e - j\omega\nu m_e} = \frac{e\boldsymbol{E}}{m_e\omega^2\left(1 - j\frac{\nu}{\omega}\right)}. \tag{6}$$

Substituting this into (1),

$$\boldsymbol{P} = -\frac{Ne^2\boldsymbol{E}}{m_e\omega^2\left(1 - j\frac{\nu}{\omega}\right)}. \tag{7}$$

The electric flux density can then be found as

$$\boldsymbol{D} = \epsilon_0\boldsymbol{E} + \boldsymbol{P} = \epsilon_0\boldsymbol{E} - \frac{Ne^2\boldsymbol{E}}{m_e\omega^2\left(1 - j\frac{\nu}{\omega}\right)} \equiv \epsilon_r\epsilon_0\boldsymbol{E}. \tag{8}$$

The effective relative dielectric constant of the plasma is then

$$\epsilon_r = 1 - \frac{Ne^2}{m_e\omega^2\epsilon_0\left(1 - j\frac{\nu}{\omega}\right)} \tag{9}$$

An *angular plasma frequency* can be defined such that

$$\omega_p^2 = \frac{Ne^2}{m_e\epsilon_0} \approx 3183N, \tag{10}$$

which is purely a function of electron density N. Then,

$$\epsilon_r = 1 - \frac{\omega_p^2}{\omega^2(1 - j\nu/\omega)}. \tag{11}$$

We see that in the presence of electron collisions, the dielectric constant can in general be complex. If we ignore collisions for the moment, then

$$\epsilon_r = 1 - \frac{\omega_p^2}{\omega^2} \approx 1 - \frac{81N}{f^2} \tag{12}$$

From this result we can make several important observations. Since the propagation constant of a wave travelling in a plasma is $\sqrt{\epsilon_r}k_0$,

- For frequencies $\omega > \omega_p$, the effective dielectric constant is less than unity but the propagation constant is real. Hence, the wave will be refracted by the plasma according to the variation of ϵ_r with altitude.
- For frequencies $\omega < \omega_p$, we get a negative value for the dielectric constant, which leads to an imaginary propagation constant. Hence, a plane wave in the medium will decay exponentially with distance. It is *not* absorbed (we have ignored losses / electron collisions here), but instead becomes evanescent, like a waveguide in cutoff. A wave incident on a medium with this propagation constant would be totally reflected.

- For frequencies $\omega \gg \omega_p$, the effective dielectric constant is essentially 1. Practically this happens at VHF frequencies and above. The waves simply pass through the plasma without significant refraction, but there can be other effects, especially if the plasma is magnetized by the Earth's magnetic field and the medium becomes anisotropic. Waves at these frequencies undergo *Faraday rotation* by the ionosphere, whereby there polarization vector is rotated as the wave passes through the atmosphere.

From these observations we can observe two propagation mechanisms for a wave entering the ionosphere. If it is below the plasma frequency, it is simply reflected off the ionosphere. When it is above the plasma frequency, then refraction analogous to that which occurs with atmospheric refraction occurs, though it is much more pronounced because of the large changes in refractive index. In fact, provided the angle at which the wave enters the atmosphere is not too steep, the wave will be bent quite strongly as it enters ionosphere. If the angle of incidence reaches the critical angle for total internal reflection, the wave is reflected off the ionosphere and starts to return to earth, going through the reverse process of refraction as it does so.

It is possible, in the case of refraction, that the wave enters the ionosphere at too sharp an angle (close to broadside) such that the condition for total internal reflection to occur. Then the wave will pass through the ionosphere and never return. Hence, there exists a minimum set of conditions on the electron density, frequency, and angle of incidence, for the wave to be returned to earth.

If we subdivide the ionosphere into many tiny layers as we did for atmospheric refraction, we can say that

$$n_0 \sin\theta_i = n_1 \sin\theta_1 = n_2 \sin\theta_2 = \cdots n_k \sin\theta_k \cdots$$

The condition for the wave to return to earth is to have total internal reflection, which begins when the refracted angle is $\theta = 90°$. If this happens at the kth layer,

$$n_0 \sin\theta_i = n_k \sin 90° = n_k$$

and since $n_0 = 1$,

$$\sin^2\theta_i = n_k^2 = \varepsilon_{r,k}. \tag{13}$$

From this, it follows that for a given angle of incidence θ_i and frequency f_{ob} (where ob stands for oblique incidence), the minimum electron density required to achieve total internal reflection is

$$\varepsilon_{r,k} = \sin^2\theta_i = 1 - \frac{81N_{min}}{f_{ob}^2} \tag{14}$$

If the maximium electron density present is N_{max}, we can also view the condition for the wave to be returned to Earth as follows. In the most challenging refraction case, normal incidence ($\theta_i = 0°$, $\sin\theta_i = 0$), the only possible way for the wave to be totally internally reflected is if $\varepsilon_{r,k} = 0$. This requires the frequency to be less than the critical frequency f_c, given by

$$\frac{81N_{max}}{f_c^2} = 1 \Rightarrow f_c = 9\sqrt{N_{max}} \tag{15}$$

If the electron density present is N_{max}, (14) can be rewritten in terms of the critical frequency as follows,

$$\sin^2\theta_i = 1 - \cos^2\theta_i = 1 - \frac{81N_{max}}{f_{ob}^2} \tag{16}$$

$$f_{ob} = 9\sqrt{N_{max}}\sec\theta_i = f_c\sec\theta_i \tag{17}$$

This value of f_{ob} is called the *maximum usable frequency*, and is less than 40 MHz, and can be as low as 25-30 MHz in period of low solar activity. Equation (17) is called the Secant Law.

Figure 4 shows the curved path of a refracted ray associated with frequency f_{ob}. The curve path reaches an altitude of h_1 before being returned to the Earth. If the incident and returned rays are extrapolated to a vertex, they meet at a height h', which is called the *virtual reflection height* of the ionospheric layer. The virtual height depends on conditions, the time of day, and the layer being considered, as shown in Table 1.

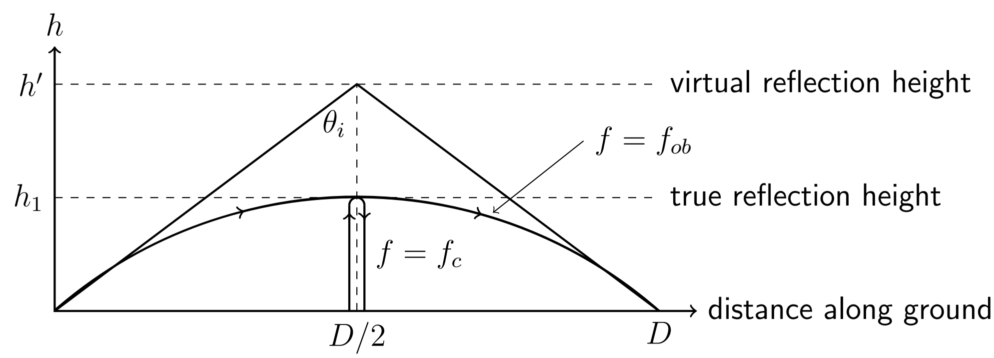

Figure 4: Virtual height

Layer	Daytime virtual refl. height	Nighttime virtual ref. height
F_2	250-400 km	—
F_1	200-250 km	—
F	—	300 km
E	110 km	110 km

Table 1: Virtual heights of various ionospheric layers

From the geometry of Figure 4, the distance traversed by the curved path can be determined from

$$\sec\theta_i = \sqrt{\left(\frac{D}{2h'}\right)^2 + 1} \tag{18}$$

The maximum skip distance $D = d_{max}$ can be achieved by aiming the radiation from the antenna so that the radiation leave the antenna parallel to the Earth's surface. This situation is shown in Figure 2. Factoring in atmospheric refraction, the maximum skip distance is given by

$$d_{max} = 2\sqrt{2KR_e h'} \tag{19}$$

which shows that very large propagation distances are possible, especially when the upper ionospheric layers are used. For example, if the F layer is used, using an effective Earth radius of $KR_e = 8497$ km gives a skip distance of 4516 km. Multiple skips are possible by using reflections of the earth to establish a multi-reflection process. This equation explains why signals can propagate so much farther at night when the D layer disappears, since it has the lowest virtual height.

In general, the broadcaster can limit the range of the transmitter by controlling the power and by manipulating the angle of radiation. Pointing the antenna skyward reduces the angle of incidence, resulting in a shorter skip distance.