Chapter 16

Ionospheric effects in Radio Astronomy

A. P. Rao

16.1 Introduction

At the low densities encountered in the further reaches of the earth's atmosphere and in outer space, collisions between particles are very rare. Hence, unlike in a terrestrial laboratory, it is possible for gas to remain in an ionized state for long periods of time. Such plasmas are ubiquitous in astrophysics, and have been extensively studied for their own sake. In this chapter however, we focus on the effects of this plasma on radio waves propagating through them, and will find astrophysical plasmas to be largely of nuisance value.

The refractive index of a cold neutral plasma is given by

$$\mu(\nu) = \sqrt{1 - \frac{\nu_p^2}{\nu^2}}, \tag{16.1.1}$$

where ν_p the "plasma frequency is given by

$$\nu_p = \sqrt{\frac{n_e e^2}{\pi m_e}} \simeq 9\sqrt{n_e} \ \text{ kHz} \tag{16.1.2}$$

where e is the charge on the electron, m_e is the mass of the electron and n_e is the electron number density (in cm^{-3}). At frequencies below the plasma frequency ν_p the refractive index becomes imaginary, i.e. the wave is exponentially attenuated and does not propagate through the medium. The earth's ionosphere has electron densities $\sim 10^4 - 10^5$ cm^{-3}, which means that the plasma frequency is $\sim 1 - 10$ MHz. Radio waves with such low frequencies do not reach the earth's surface and can be studied only by space based telescopes. The plasma between the planets is called the Interplanetary Medium (IPM) and has electron densities ~ 1 cm^{-3} (at the earth's location); the corresponding cut off frequency is ~ 9 kHz. The typical density in the Interstellar Medium (ISM) is ~ 0.03 cm^{-3} for which the cut off frequency is ~ 1 kHz. Waves of such low frequency from extra solar system objects cannot be observed even by spacecraft since the IPM and ISM will attenuate them severely.

The dispersion relationship in a cold plasma is given by $c^2k^2 = \omega^2 - \omega_p^2$. Since this is a non linear relation there are two characteristic velocities of propagation, the phase velocity given by

$$v_p = \frac{\omega}{k} = \frac{c}{\mu} \simeq c\ (1 + \frac{1}{2}\frac{\nu_p^2}{\nu^2}) \tag{16.1.3}$$

and the group velocity which is given by

$$v_g = \frac{d\omega}{dk} = c\mu \simeq c\ (1 - \frac{1}{2}\frac{\nu_p^2}{\nu^2}). \tag{16.1.4}$$

Where for the last expression we have assumed that $\nu >> \nu_p$ (which is usually the regime of interest).

16.2 Propagation Through a Homogeneous Plasma

Even above the cutoff frequency there are various propagation effects that are important for a radio wave passing through a plasma. Let us start with the most straightforward ones. Consider a radio signal passing through a homogeneous slab of plasma of length L. The signal is delayed (with respect to the propagation time in the absence of the plasma) by the amount

$$\Delta T = \frac{L}{v_g} - \frac{L}{c} = \frac{L}{c}(1/\mu - 1) \simeq \frac{L}{c}\frac{1}{2}\frac{\nu_p^2}{\nu_2}.$$

The magnitude of the propagation delay can hence be written as

$$|\Delta T| = \frac{L}{c} \times \frac{4 \times 10^6}{\nu_{\rm Hz}^2} n_e.$$

The propagation delay can also be considered as an "excess path length" ΔL = c ΔT. Further since $(v_g/c - 1)$ and $(v_p/c - 1)$ differ only in sign[1], the magnitude of the "excess phase" (viz. $2\pi\nu(L/v_p - L/c)$) is given by $\Delta\Phi = 2\pi\nu\Delta T$. Note that since the propagation delay is a function of frequency ν, waves of different frequencies get delayed by different amounts. A pulse of radiation incident at the far end of the slab will hence get smeared out on propagation through the slab; this is called "dispersion". If the plasma also has a magnetic field running through it then it becomes birefringent – the refractive index is different for right and left circularly polarized waves. A linearly polarized wave can be considered a superposition of left and right circularly polarized waves. On propagation through a magnetized plasma the right and left circularly polarized components are phase shifted by different amounts, or equivalently the plane of polarization of the linearly polarized component is rotated. This rotation of the plane of polarization on passage through a magnetized plasma is called "Faraday rotation". The angle through which the plane of polarization is rotated is given by

$$\Theta = RM\lambda^2 = 0.81\lambda^2 \int n_e B_{||} dl.$$

and RM is called the rotation measure. For the second equality λ is in meters, n_e is in cm^{-3}, $B_{||}$ is in μG and the length is in parsecs.

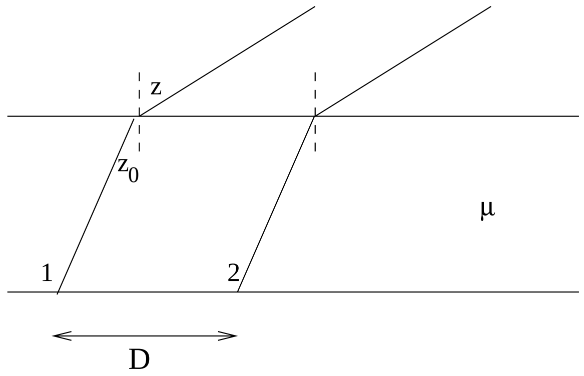

Figure 16.1: Propagation through a plane parallel ionosphere

16.3 Propagation Through a Smooth Ionosphere

For an interferometer, there are two quantities of interest (i) the delay difference between the signals reaching the two arms of the interferometer ($\delta T = \Delta T_1 - \Delta T_2$), where ΔT_1 and ΔT_2 are the propagation delays for the two arms of the interferometer, and (ii) the phase difference between the signals reaching the two arms of the interferometer ($\delta\phi = 2\pi/\lambda(\Delta L_1 - \Delta L_2)$, where ΔL_1 and ΔL_2 are the excess path lengths for the two arms of the interferometer. Generally δT is small compared to the coherence bandwidth of the signal and can be ignored to first order, however $\delta\phi$ could be substantial.

In a homogeneous plane parallel ionosphere with refractive index μ (see Figure 16.1), we have from Snell's law $\mu\sin(z_0) = \sin(z)$. The observed geometric delay is $\tau_g = \mu D\sin(z_0)/c$, since the group velocity is c/μ. From Snell's law therefore, $\tau_g = D\sin(z)/c$, the same as would have been observed in the absence of the ionosphere. A homogeneous plane parallel ionosphere hence produces no net effect on the visibilities, even though the apparent position of the source has changed. In the case where the interferometer is located outside the slab, there is neither a change in the apparent position nor a change in the phase, as is obvious from the geometry. This entire analysis holds for a stratified plane parallel ionosphere (since it is true for every individual plane parallel layer). However, in the real case of a curved ionosphere, with a radial variation of electron density, then neither the change in the apparent position nor $\delta\phi$ are zero even outside the ionosphere. Effectively, the direction of arrival of the rays from the distant source appears to be different from the true direction of arrival (as illustrated in Figure 16.2) and unlike in the plane parallel case this is not exactly canceled out by the change in the refractive index. If $\Delta\theta$ is the difference between the true direction and apparent directions of arrival, then one can compute that

$$\Delta\theta = \frac{A\sin(z_0)}{r_0}\int_0^\infty \frac{\alpha^2\mu(h)dh}{(1-\alpha^2\sin^2(z_0))}. \tag{16.3.5}$$

where z_0 is the observed zenith angle, r_0 is the radius of the earth, h is the height above the earth's surface and, $\mu(h)$ is the refractive index at height h, and A is a constant. For baseline lengths typical of the GMRT, this value is the same for both arms of the baseline. If the baseline has UV co-ordinates (u,v), then the phase difference due to the apparent change in the source position is given by

$$\Delta\phi = 2\pi(u\Delta\theta_{EW} \ + \ v\Delta\theta_{NS}).$$

[1] to first order for $\nu >> \nu_p$, as can be easily verified.

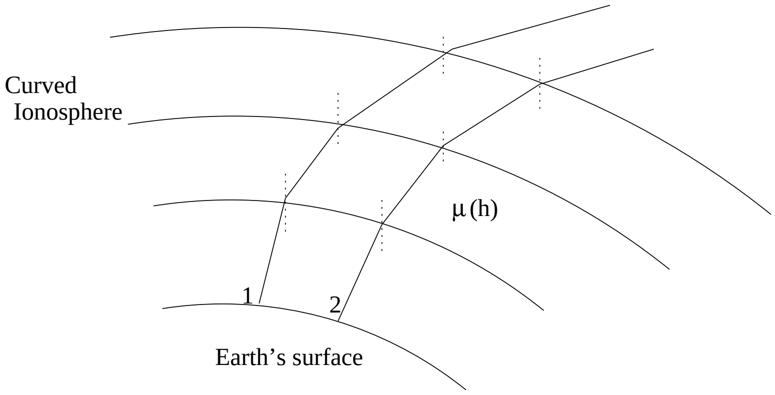

Figure 16.2: Propagation through a curved ionosphere

	Max. Val (Day)	Min Val (Night)	Freq. Dependence
TEC	5×10^{13} cm^{-2}	5×10^{12} cm^{-2}	-
Group Delay	12 μsec	1.2 μsec	ν^{-2}
Excess Path	3500 m	350 m	ν^{-2}
Phase Change	7500 rad	750 m	ν^{-2}
Phase Fluctuation	± 150 rad	± 15 rad	ν^{-2}
Mean Refraction	$6^{'}$	$0.6^{'}$	ν^{-2}
Faraday Rotation	15 cycles	1.5 cycles	ν^{-2}

Table 16.1: Typical numerical values of various ionospheric effects

Typical values for some of the ionospheric prorogation effects that we have been discussing are given in Table 16.1.

16.4 Propagation Through an Inhomogeneous Ionosphere

So far we have been dealing with an ionosphere, which, while not homogeneous, is still fairly simple in that the density fluctuations are smooth, slowly varying functions. Further, the ionospheric density was assumed to not vary with time. In reality, the earth's ionosphere shows density fluctuations on a large range of length and time scales. A density fluctuation of length scale l at a height h above the earth's surface corresponds to a fluctuation on an angular scale of l/h. For a typical length scale l of 10 km, at a height of 200 km, the corresponding angular scale is $\sim 3^o$. This means that the phase difference introduced by the ionosphere changes on an angular scale of 3^o. If this phase is to be calibrated out, then one would need to pick a calibrator that is within 3^o of the target source — for most sources it turns out that there is no suitable calibrator this close by. This problem gets increasingly worse as one goes to lower frequencies since the excess ionospheric phase increases as ν^{-2}. As discussed in Chapter 5 therefore, as long as the excess ionospheric phase is constant over the field of view, this phase can be lumped in with the electronic phase of receiver chain, and can be solved for using self-calibration.

However, for a given antenna, as one observes at lower and lower frequencies, the field

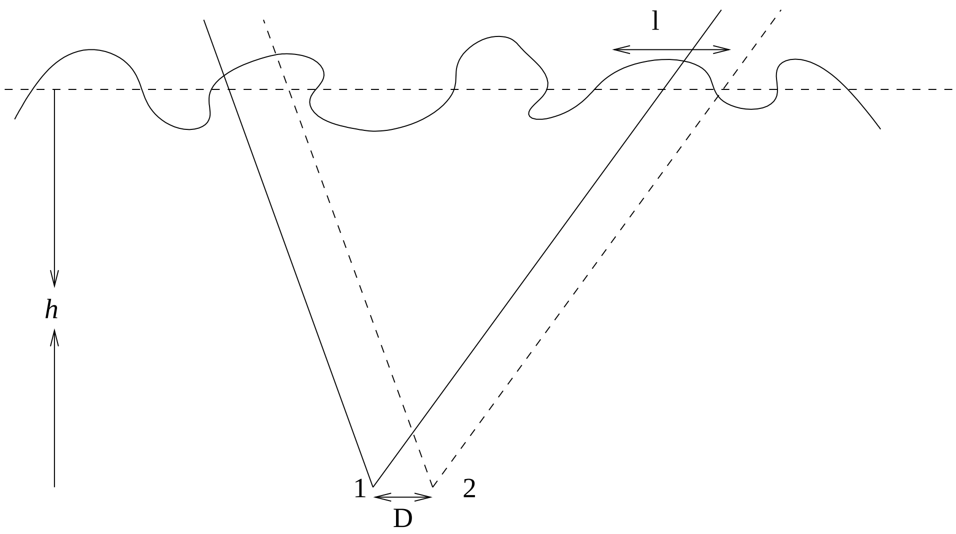

Figure 16.3: For short enough baselines, the isoplantic assumption holds even if the field of view is larger than the typical coherence length of the ionospheric irregularities. This is because both arms of the interferometer get essentially the same excess phase.

of view increases as ν^{-1}. Since the excess ionospheric phase is also increasing rapidly with decreasing frequency, one will soon hit a point where the assumption that the excess phase is constant over the field of view is a poor one. At this point the self-calibration algorithm is no longer applicable. Variations of the ionospheric phase over the field of view are referred to as "non isoplanaticity". As illustrated in Figure 16.3, when the baseline length is small compared to the typical length scale of ionospheric density fluctuations, even though the ionospheric phase is different for different sources in the field of view, the excess phase is nearly identical at both ends of the baseline. Since interferometers are sensitive only to phase differences between the two antennas, the isoplanatic assumption still holds. The non isoplanaticity problem hence arises only when the baselines as well as the field of view are sufficiently large. For the GMRT, isoplanaticity is often a poor assumption at frequencies of 325 MHz and lower.

16.5 Angular Broadening

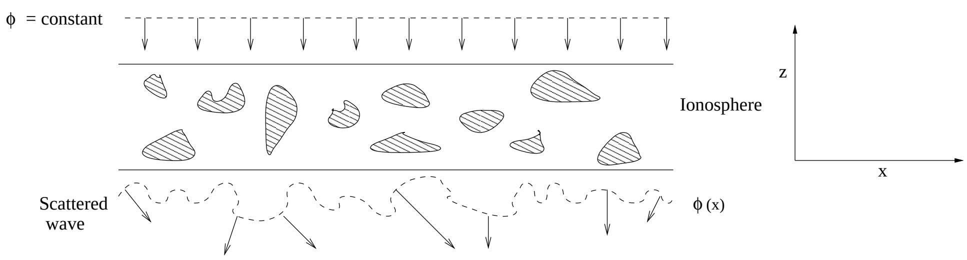

Figure 16.4: Angular broadening.

As discussed in the previous sections, the small scale fluctuations of electron density in the ionosphere lead to an excess phase for a radio wave passing through it. This excess phase is given by

$$\begin{aligned}\phi(x) &= \frac{2\pi}{\lambda}\int \Delta\mu dz,\\ \phi(x) &= C\lambda\int \Delta n(x,z)dz,\end{aligned}$$

where $\Delta\mu$ is the change in refractive index due to the electron density fluctuation, C is a constant and $\Delta n(x,z)$ is the fluctuation in electron density at the point (x,z) and the integral is over the entire path traversed by the ray (see Figure 16.4).

If we assume that $\phi(x)$ is a zero mean Gaussian random process, with auto-correlation function given by $\phi_0^2\rho(r)$, where $\rho(r) = e^{-r^2/2a_\phi^2}$, then from the relation above for $\phi(x)$ we can determine that $\phi_0^2 \propto \lambda^2\Delta n^2 L$, where L is the total path length through the ionosphere[2]. Let us assume that a plane wavefront from an extremely distant point source is incident on the top of such an ionosphere. In the absence of the ionosphere the wave reaching the surface of the earth would also be a plane wave. For a plane wave the correlation function of the electric field (i.e. the visibility) is given by $\langle E_i(x)E\_i^*(x+r)\rangle = E_i{}^2$, i.e. a constant independent of r. On passage through the ionosphere however, different parts of the wave front acquire different phases, and hence the emergent wavefront is not plane. If $E(x)$ is the electric f ield at some point on the emergent wave, then we have $E(x) = E_i e^{-i\phi(x)}$. Since E_i is just a constant, the correlation function of the emergent electric field is

$$\langle E(x)E^*(x+r)\rangle = E_i^2\langle e^{-i(\phi(x)-\phi(x+r))}\rangle.$$

From our assumptions about the statistics of $\phi(x)$ this can be evaluated to give

$$\langle E(x)E^*(x+r)\rangle = E_i^2 e^{-2\phi_0^2(1-\rho(r))}. \tag{16.5.6}$$

If ϕ_0^2 is very large, then the exponent is falls rapidly to zero as $(1-\rho(r))$ increases (or equivalently when r increases). It is therefore adequate to evaluate it for small values of r, for which $\rho(r)$ can be Taylor expanded to give $\rho(r) \simeq 1 - 1/2r^2/a_\phi^2$. and we get

$$\langle E(x)E^*(x+r)\rangle = E_i^2 e^{-\phi_0^2\frac{r^2}{a_\phi^2}}.$$

The emergent electric field hence has a finite coherence length (while the coherence length of the incident plane wave was infinite). From the van Cittert-Zernike theorem this is equivalent to saying that the original unresolved point source has got blurred out to a source of finite size. This blurring out of point sources is called “angular broadening” or “scatter broadening”. If we define $a = a_\phi/\phi_0$ then the visibilities have a Gaussian distribution given by e^{-ir^2/a^2}, meaning that the characteristic angular size θ_{scat} of the scatter broadened source is $\sim \lambda/a \propto \lambda^2\sqrt{\Delta n^2 L}$. θ_{scat} is called the “scattering angle”.

On the other hand if ϕ_0^2 is small then the exponent in eqn 16.5.6 can be Taylor expanded to give

$$\begin{aligned}\langle E(x)E^*(x+r)\rangle &= E_i^2\big[1-2\phi_0^2(1-\rho(r))\big],\\ &= E_i^2\big[(1-2\phi_0^2)+2\phi_0^2 e^{\frac{-r^2}{2a_\phi^2}}\big].\end{aligned}$$

This corresponds to the visibilities from an unresolved core (of flux density E_i^2 $(1-2\phi_0^2)$) surrounded by a weak halo.

[2]This follows from the equation for $\phi(x)$ if you also assume that $< \Delta n(x,z)\Delta n(x,z') > = \Delta n^2\delta(z,z')$.

16.6 Scintillation

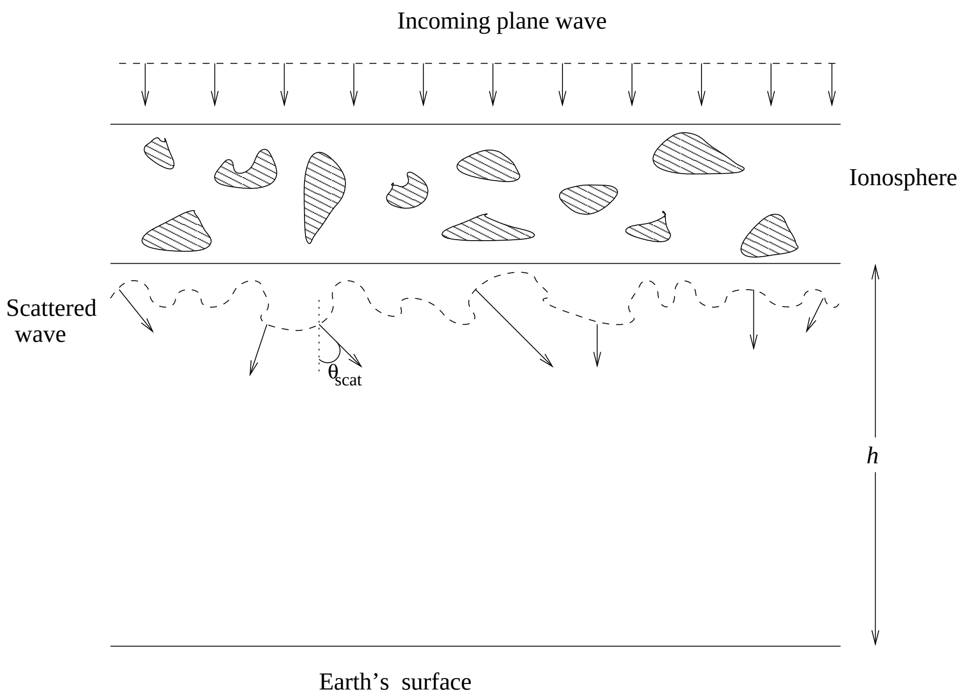

Figure 16.5: Scintillation due to the ionosphere

In the last section we dealt with an ionosphere which had random density fluctuations in it. In the model we assumed the density was assumed to vary randomly with position, but not with time. In the earth's ionosphere however, the density does vary with both position and time. Temporal variations arise both because of intrinsic variation as well as because of traveling disturbances in the ionosphere, because of which a given pattern of density fluctuations could travel across the line of sight.

This temporal variation of the density fluctuations means that the coherence function (even at some fixed separation on the surface of the earth) will vary with time. This phenomena is generically referred to as "scintillation". Depending on the typical scattering angle as well as the typical height of the scattering layer from the surface of the earth, the scintillation could be either "weak" or "strong".

As discussed in the previous section, rays on passing through an irregular ionosphere get scattered by a typical angle θ_{scat}. If the scattering occurs at a height h above the antennas, then as shown in Figure 16.5 these scattered rays have to traverse a further distance h before being detected. The transverse distance traveled by a scattered ray is $\sim h\theta_{scat}$. If this length is much less than the coherence length a, then the rays scattered by different irregularities in the scattering medium do not intersect before reaching the ground. The corresponding condition is that $h\theta_{scat} < a$, i.e. $h\theta_{scat} < \lambda/\theta_{scat}$ or $h\theta_{scat}^2 < \lambda$.

If this condition holds, then, at any instant of time, (as discussed in the previous section), what the observer sees is an undistorted image of the source, which is shifted in position due to refraction. As time passes, the density fluctuations change[3] and so

[3]but we assume that their statistics remain exactly the same, i.e. that they continue to be realization of a Gaussian random process with variance ϕ_0 and auto-correlation $\rho(r)$

the image appears to wander in the sky and in a long exposure image which averages many such wanderings, the source appears to have a scattered broadened size θ_{scat}. Provided that one can do self calibration on a time scale that is small compared to the time scale of the "image wander", this effect can be corrected for completely. On the other hand, when the $h\theta_{scat}^2 > \lambda$ the rays from different density fluctuations will intersect and interfere with one another. The observer sees more than one image, and because of the interference, the amplitude of the received signal fluctuates with time. This is called "amplitude" scintillation. Amplitude scintillation at low frequencies, particularly over the Indian subcontinent can be quite strong. The source flux could change be factors of 2 or more on very short timescales. This effect cannot be reliably modeled and removed from the data, and hence observations are effectively precluded during periods of strong amplitude scintillation.

16.7 Further Reading

1. Interferometry and Synthesis in Radio Astronomy; Thompson, A. Richard, Moran, James M., Swenson Jr., George W.; Wiley-Interscience Publication, 1986.

2. Synthesis Imaging In Radio Astronomy; Eds. Perley, Richard A., Schwab, Frederic R., and Bridle, Alan H.; ASP Conference Series, Vol 6.