Foreword and acknowledgements

This text is the first English version of the lectures on "radar theory" which I gave at the University of Oulu, Finland in 1994. As compared to the original Finnish version, the manuscript has been somewhat expanded, not so much by adding new material but by a more detailed explanation of the subject matter.

Because of my hasty timetable, I have not yet been able to give the whole manuscript to my colleagues for critical reading and therefore it is likely to contain errors, some of which may even be serious. A few people have read parts of the manuscript, found errors and made suggestions which I have corrected and taken into account. Some of these persons have been able to do this only after having seen a glimpse of the work. I am grateful to Dr. A. Aikio, Dr. A. Huuskonen and Mr. M. Markkanen for these comments. I am also grateful to Prof. E.D. Tereshchenko for making some profound suggestions which, due to the lack of time, have not yet been included in the present version but will cause changes in the next one. I thank Mr. E. Saviaro for making a draft translation from the original Finnish manuscript and drawing many of the figures.

At the present stage the manuscript contains several defects. For example, a proper reference list is missing, the whole text is still EISCAT-UHF-oriented although more weight should be put to the future ESR radar, remote measurements are almost completely neglected and sideband corrections are not explained. It would also be enlightening to introduce selected real experiments in order to demonstrate how a complete incoherent scatter measurement is constructed as well as to show ambiguity functions calculated using realistic rather than boxcar-shaped impulse responses. I hope I will be able to improve these points in the future. In one respect this manuscript is really up-to-date: new 64-bit and 128-bit alternating codes are published here for the first time.

Oulu, February 1996

Tuomo Nygrén

Contents

Chapter 1

Introduction

This text is meant to be an introduction in the theory of incoherent scatter measurements. The emphasis lies neither in the scattering process nor in the radar technology, but rather in the mathematical principles of the analogue and digital data processing in the radar receiver, in the theory of ambiguity functions and, finally, in the most important modulation methods. The purpose is to give a general picture of the working principles of the method so that those, who later want to learn about the technical aspects of the radar, will be familiar with the idea behind the applied technology and those, who want to learn about the data analysis and experiment design, will have a sufficient background on how the measurements are carried out and what affects the observed autocorrelation functions. Data analysis itself (i.e. the methods of determining the plasma parameters from the measured autocorrelation functions) as well as the error analysis are wide topics which would be subjects of a separate treatment.

In order to give some basic knowledge on the target, the theory of Thomson scattering from free electrons as well as scattering from density fluctuations in a plasma is briefly introduced following more or less the guidelines presented by Prof. T. Hagfors in various EISCAT summer schools. The theory of the plasma autocorrelation function, however, is not included but a qualitative picture based on ion and electron acoustic waves is given instead. In order show how the radio wave is generated in the transmitter, the working principle of the klystron is briefly described. The generation of narrow radar beams is explained by using the simplest theory of a parabolic dish antenna as an example.

No previous knowledge of signal theory is assumed; it is only supposed that the reader is familiar with the Fourier transform. The necessary signal theory is introduced starting from the first principles and restricted to a minimum level, e.g. the use of the concept of stochastic process is avoided as much as possible. An essential part of the course is the theory of radar ambiguity functions in the form presented by Dr. M. Lehtinen. This formalism is greatly emphasised because it presents a modern way of handling the effects of transmitter modulation and receiver filtering and it has a key role in the future standard analysis package GUISDAP (Grand Unified Incoherent Scatter Design and Analysis

Package). The ambiguity function theory is also used in explaining the ideas of the different modulation methods. Another useful tool essential in understanding the experiment design is the lag profile matrix, which is largely applied in explaining the modulation principles.

Chapter 2

Scattering process and the radar

2.1 Scattering from density fluctuations

There are always temporal and spatial variations in the refractive index n (or the permittivity $\varepsilon \propto n^2$) of the ionospheric plasma. They may be caused by thermal fluctuations, turbulence or plasma instabilities. When electromagnetic waves propagate in the plasma, these fluctuations cause partial reflection which we call scattering.

Let us consider an isotropic medium with a scalar permittivity ε. For high frequencies like in the incoherent scatter radar, this is a good approximation even in the magnetised ionospheric plasma. In this case the displacement vector $\mathbf{D}$ due to an electric field $\mathbf{E}$ is

$$\mathbf{D} = \varepsilon\mathbf{E} = \varepsilon_0\mathbf{E} + \mathbf{P}, \tag{2.1}$$

where the polarisation $\mathbf{P}$ is

$$\mathbf{P} = (\varepsilon - \varepsilon_0)\mathbf{E}. \tag{2.2}$$

If the average value of the temporally and spatially varying permittivity is $\langle\varepsilon\rangle$, we can write ε as

$$\varepsilon(\mathbf{r}, t) = \langle\varepsilon\rangle + \Delta\varepsilon(\mathbf{r}, t), \tag{2.3}$$

where $\Delta\varepsilon$ is the permittivity fluctuation. Then

$$\begin{aligned}\mathbf{P} &= (\langle\varepsilon\rangle + \Delta\varepsilon - \varepsilon_0)\mathbf{E} \\ &= (\langle\varepsilon\rangle - \varepsilon_0)\mathbf{E} + \Delta\varepsilon\mathbf{E} \\ &= \mathbf{P}_{\langle\varepsilon\rangle} + \Delta\mathbf{P}.\end{aligned} \tag{2.4}$$

Here $\Delta\mathbf{P}$ is the polarisation fluctuation

$$\Delta\mathbf{P} = \Delta\varepsilon\mathbf{E}. \tag{2.5}$$

In ideal homogeneous medium the electromagnetic wave propagates without scattering and reflection. In such a case $\varepsilon = \langle\varepsilon\rangle$ and the displacement current due to the wave is $\dot{\mathbf{D}} = \langle\varepsilon\rangle\dot{\mathbf{E}}$. This is the displacement current which is needed to make the wave propagate in the medium at a proper phase velocity. If fluctuations are present, they cause an additional displacement current density

$$\Delta\dot{\mathbf{D}} = \Delta\dot{\mathbf{P}} = \Delta\varepsilon\dot{\mathbf{E}} \tag{2.6}$$

which will emit radiation. In this way an incident electromagnetic wave will be scattered by the fluctuations of the refractive index.

Let an incident electromagnetic plane wave with an electric field

$$\mathbf{E} = \mathbf{E}_0 e^{i(\omega_0 t - \mathbf{k}\cdot\mathbf{r})} \tag{2.7}$$

enter the medium. It causes a polarisation fluctuation

$$\Delta\mathbf{P} = \Delta\varepsilon(\mathbf{r},t)\mathbf{E}_0 e^{i(\omega_0 t - \mathbf{k}\cdot\mathbf{r})}, \tag{2.8}$$

which leads to a displacement current density

$$\Delta\mathbf{j} = \Delta\dot{\mathbf{P}} = i\omega_0\Delta\varepsilon(\mathbf{r},t)\mathbf{E}_0 e^{i(\omega_0 t - \mathbf{k}\cdot\mathbf{r})}. \tag{2.9}$$

Here the time derivative of $\Delta\varepsilon(\mathbf{r},t)$ is neglected because it is much smaller than the time derivative of the electric field. A volume element d^3r' at $\mathbf{r}'$ contains a current element $\Delta\mathbf{j}\cdot d^3r'$ which causes a vector potential

$$d\mathbf{A} = \frac{\mu_0}{4\pi}\cdot\frac{\Delta\mathbf{j}\cdot d^3r'}{|\mathbf{r}-\mathbf{r}'|} \tag{2.10}$$

at $\mathbf{r}$. Since the disturbance in the electromagnetic field propagates at the speed of light, the vector potential at $\mathbf{r}$, due to a current element at $\mathbf{r}'$, is delayed by $t - |\mathbf{r}-\mathbf{r}'|/c$, *i.e.* the retarded vector potential is

$$d\mathbf{A}(\mathbf{r},t) = \frac{\mu_0}{4\pi}\cdot\frac{\Delta\mathbf{j}\left(\mathbf{r}', t-|\mathbf{r}-\mathbf{r}'|/c\right)d^3r'}{|\mathbf{r}-\mathbf{r}'|}$$

$$= \frac{i\omega_o\mu_0\mathbf{E}_0}{4\pi}\cdot\frac{\Delta\varepsilon\left(\mathbf{r}', t-|\mathbf{r}-\mathbf{r}'|/c\right)}{|\mathbf{r}-\mathbf{r}'|}\cdot e^{i\omega_o\left(t-|\mathbf{r}-\mathbf{r}'|/c\right) - i\mathbf{k}\cdot\mathbf{r}'}d^3r'. \tag{2.11}$$

In incoherent scatter radar we observe scattering from a finite volume V (Fig. 2.1). Assuming so slow fluctuations that their changes are small during the time the wave passes V and a scattering volume much smaller than its distance R to the observation point $\mathbf{r}$, the time $t-|\mathbf{r}-\mathbf{r}'|/c$ in $\Delta\varepsilon$ can be replaced by a single time $t' = t - R/c$ everywhere inside V. Then $\Delta\varepsilon\left(\mathbf{r}', t-|\mathbf{r}-\mathbf{r}'|/c\right) \approx \Delta\varepsilon(\mathbf{r}',t')$ and the term $|\mathbf{r}-\mathbf{r}'|$ in the denominator of eq. (2.11) can also be replaced by R. Furthermore, the term $i\omega_o|\mathbf{r}-\mathbf{r}'|/c$ can be written as $i\omega_o|\mathbf{r}-\mathbf{r}'|/c = ik|\mathbf{r}-\mathbf{r}'| = i\mathbf{k}_s\cdot(\mathbf{r}-\mathbf{r}')$, where $\mathbf{k}_s$ is the wave vector in the direction $\mathbf{r}-\mathbf{r}'$. With this notation the exponential function in eq. (2.11) can be developed as

$$e^{i\omega_0 t - i\mathbf{k}_s\cdot(\mathbf{r}-\mathbf{r}') - i\mathbf{k}\cdot\mathbf{r}'}$$

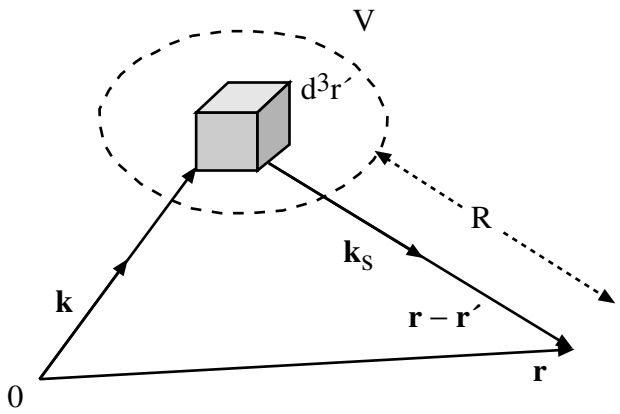

Figure 2.1: Scattering from finite volume V.

$$= e^{i\omega_0 t - i\mathbf{k}_s \cdot \mathbf{r} - i(\mathbf{k} - \mathbf{k}_s) \cdot \mathbf{r}'}$$
$$= e^{i(\omega_0 t - \mathbf{k}_s \cdot \mathbf{r})} \cdot e^{-i(\mathbf{k} - \mathbf{k}_s) \cdot \mathbf{r}'}.$$

In the above approximation the total vector potential at $\mathbf{r}$, due to the radiation scattered from V, can be written in the form

$$\mathbf{A}(\mathbf{r}, t) = \frac{i\omega_0 \mu_0 \mathbf{E}_0}{4\pi R} \cdot e^{i(\omega_0 t - \mathbf{k}_s \cdot \mathbf{r})} \int_V \Delta\varepsilon(\mathbf{r}', t') e^{-i(\mathbf{k} - \mathbf{k}_s) \cdot \mathbf{r}'} d^3 r'. \quad (2.12)$$

This can be interpreted as a plane wave propagating in the direction $\mathbf{k}_s$ with a carrier angular frequency ω_0. The amplitude of the wave is proportional to the 3-dimensional spatial Fourier transform of the permittivity fluctuation at a wave number $\mathbf{K} = \mathbf{k} - \mathbf{k}_s$.

Using the notation $\Delta\mathcal{E}$ for the Fourier transform of $\Delta\varepsilon$, the result can be rewritten as

$$\mathbf{A}(\mathbf{r}, t) = \frac{i\omega_0 \mu_0 \mathbf{E}_0}{4\pi R} \Delta\mathcal{E}(\mathbf{K}) e^{i(\omega_0 t - \mathbf{k}_s \cdot \mathbf{r})}. \quad (2.13)$$

Since the magnetic induction of the scattered wave is given by

$$\mathbf{B} = \nabla \times \mathbf{A}, \quad (2.14)$$

obviously $\mathbf{B} \propto \mathbf{E}_0 \Delta\mathcal{E}(\mathbf{K})$ and the intensity (absolute value of the Poynting vector) of the scattered radiation is

$$S_s \propto |\Delta\mathcal{E}(\mathbf{K})|^2 S, \quad (2.15)$$

where S is the intensity of the incident wave entering the scattering volume.

Starting from the refractive index of the electromagnetic wave in a plasma,

$$n = \sqrt{1 - \frac{n_e e^2}{\varepsilon_0 m_e \omega_0^2}}, \quad (2.16)$$

the plasma permittivity is

$$\varepsilon = \varepsilon_0 n^2 = \varepsilon_0 - \frac{n_e e^2}{m_e \omega_0^2}, \tag{2.17}$$

where n_e is the electron density and m_e the electron mass. Therefore the permittivity and electron density fluctuations $\Delta\varepsilon$ and Δn_e are connected by

$$\Delta\varepsilon = -\frac{e^2}{m_e \omega_0^2} \cdot \Delta n_e. \tag{2.18}$$

The corresponding relation is also valid for their Fourier transforms so that the intensity of the scattered radiation is

$$S_s \propto |\mathcal{N}_e(\mathbf{K})|^2 S, \tag{2.19}$$

where $\Delta\mathcal{N}_e(\mathbf{K})$ is the 3-dimensional spatial Fourier transform of Δn_e. This means that the amplitude of the scattered signal is proportional to the amplitude of the density fluctuation at a wave number $\mathbf{K} = \mathbf{k} - \mathbf{k}_s$.

The relation of the wave numbers $\mathbf{k}$, $\mathbf{k}_s$ and $\mathbf{K}$ is shown in Fig. 2.2. Because $|\mathbf{k}| = |\mathbf{k}_s| = 2\pi/\lambda$, $|\mathbf{K}|$ must be equal to $2k\cos\phi$ and the corresponding wavelength is

$$\Lambda = \frac{\lambda}{2\cos\phi}. \tag{2.20}$$

On the other hand, if the waves were scattered from parallel planes separated by a distance d, their path difference would be (see Fig. 2.3)

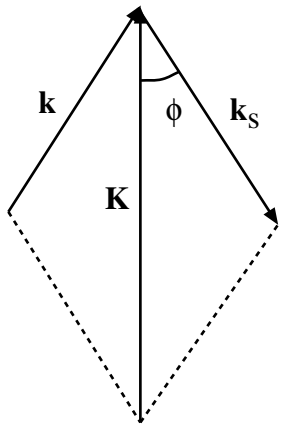

Figure 2.2: The relation of incident and scattered wave vectors and the wave vector of the permittivity fluctuation.

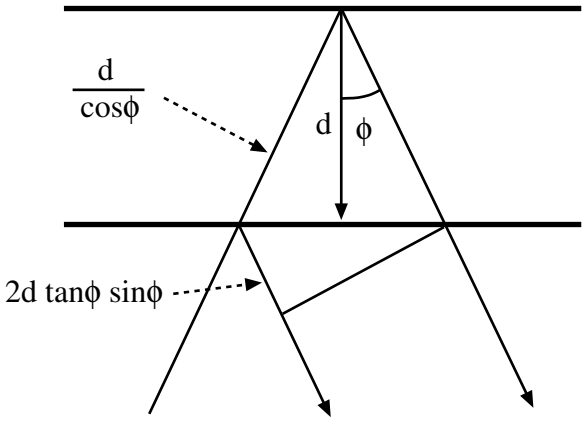

Figure 2.3: Scattering from parallel planes.

$$\Delta = \frac{2d}{\cos\phi} - 2d\tan\phi\sin\phi = 2d\cos\phi. \tag{2.21}$$

The interference is constructive, if this distance is equal to an integer number of wavelengths. When $\Delta = \lambda$, we get

$$d = \frac{\lambda}{2\cos\phi}. \tag{2.22}$$

This distance is equal to the wave length Λ in eq. (2.20). Hence we can interpret eq. (2.19) in the following way: The density fluctuation $\Delta n_e(\mathbf{r}, \mathbf{t})$ is composed of plane waves at all wave numbers, propagating in all directions as described by its spatial Fourier transform. Each wave front reflects a part of the incident radiation into the direction determined by the reflection law. When the fluctuation wave length Λ obeys eq. (2.20), reflections from adjacent wave fronts interfere constructively and incident radiation is scattered in this direction. For each direction of observation, the density fluctuation contains a Fourier component with a proper wave length and orientation to produce scattering. In other words, the scattering process chooses the Fourier component which causes constructive interference in the direction of observation. The wave vector of this component is $\mathbf{K} = \mathbf{k} - \mathbf{k}_s$ and its wave length is $\Lambda = \lambda/(2\cos\phi)$. In the case of backscattering $\phi = 0$ and $\Lambda = \lambda/2$. The mechanism is analogous to Bragg scattering in crystals.

The Fourier components of the density fluctuation are propagating waves. Since the scattering takes place at moving wave fronts, the spectrum of the scattered radiation is broadened due to the Doppler effect. The phase speed and damping of these waves depend on the physical properties of the medium, *i.e.* temperature, density *etc.* If the theory of the fluctuations and processes involved are known, the spectrum of the scattered radiation can be used to observe these physical parameters. The incoherent scatter spectrum tells us

about the electron density, plasma speed, electron and ion temperature and in some cases also on the ion-neutral collision frequency or ion concentration in a plasma.

2.2 Thomson scattering

In the previous section, scattering was considered from the macroscopic viewpoint of partial reflection from refractive index gradients without paying attention to the scattering process itself. Therefore the results are valid not only for incoherent scatter but also for other types of scattering in the ionospheric plasma or neutral atmosphere. The purpose of this chapter is to review Thomson scattering, the basic mechanism producing the faint signals observed by incoherent scatter radars.

Thomson scattering is a process by which a part of the incident energy is scattered by free electrons in the ionosphere. The electric field of the incident wave puts the electrons in oscillatory motion and they emit a small fraction of the incident energy as dipole radiation (Fig. 2.4). The same mechanism works for ions as well but, due to the great ion mass, the energy thus scattered is negligible.

The momentum equation of an electron in the oscillating electric field $\mathbf{E}_i = \mathbf{E}_0 e^{i\omega_0 t}$ of the incident wave is

$$m_e \ddot{\mathbf{r}}_e = -e\mathbf{E}_i, \tag{2.23}$$

which gives the electron velocity

$$\begin{aligned} \mathbf{v}_e = \dot{\mathbf{r}}_e &= -\frac{e}{m_e}\int \mathbf{E}_i dt \\ &= -\frac{e\mathbf{E}_0}{m_e}\int e^{i\omega_0 t} dt = i\frac{e\mathbf{E}_0}{m_e\omega_0} e^{i\omega_0 t}. \end{aligned} \tag{2.24}$$

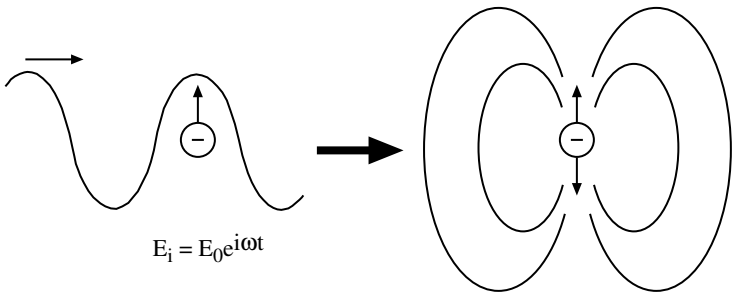

Figure 2.4: The Thomson scattering mechanism in terms of dipole radiation of an oscillating electron.

The current density due to this electron motion is

$$\begin{aligned}\mathbf{j}(\mathbf{r},t) &= \varrho_e(\mathbf{r},t)\mathbf{v}_e(t)\\ &= -e\mathbf{v}_e(t)\cdot\delta[\mathbf{r}-\mathbf{r}_e(t)]\\ &= -i\frac{e^2\mathbf{E}_0}{m_e\omega_0}e^{i\omega_0 t}\cdot\delta[\mathbf{r}-\mathbf{r}_e(t)],\end{aligned} \tag{2.25}$$

where $\rho_e(\mathbf{r},t) = -e\delta[\mathbf{r}-\mathbf{r}_e(t)]$ is the charge density of the electron and δ is the Dirac delta function. The retarded vector potential of the current density $\mathbf{j}$ is

$$\mathbf{A}_s(\mathbf{r},t) = \frac{\mu_0}{4\pi}\int\frac{\mathbf{j}(\mathbf{r}',t')}{|\mathbf{r}-\mathbf{r}'|}d^3r', \tag{2.26}$$

where

$$t' = t - \frac{|\mathbf{r}-\mathbf{r}'|}{c}. \tag{2.27}$$

This vector potential is associated with a wave scattered by the oscillating electron. By inserting (2.25) in (2.26) we obtain

$$\begin{aligned}\mathbf{A}_s(\mathbf{r},t) &= -i\frac{\mu_0 e^2\mathbf{E}_0}{4\pi m_e\omega_0}\int\frac{e^{i\omega_0 t'}}{|\mathbf{r}-\mathbf{r}'|}\cdot\delta[\mathbf{r}-\mathbf{r}_e(t')]d^3r'\\ &= -i\frac{\mu_0 e^2\mathbf{E}_0}{4\pi m_e\omega_0}\cdot\frac{e^{i\omega_0\left(t-|\mathbf{r}-\mathbf{r}_e|/c\right)}}{|\mathbf{r}-\mathbf{r}_e|}.\end{aligned} \tag{2.28}$$

If we choose our coordinate system in such a way that the electron is oscillating around the origin, $\mathbf{r}_e \approx 0$ for small oscillations and the vector potential far from the origin is

$$\mathbf{A}_s(\mathbf{r},t) = -i\frac{\mu_0 e^2\mathbf{E}_0}{4\pi m\omega_0}\cdot\frac{e^{i\omega_0\left(t-|\mathbf{r}|/c\right)}}{|\mathbf{r}|}. \tag{2.29}$$

Then $\mathbf{k}_s = \omega_0\hat{\mathbf{r}}/c$ so that $\omega_0|\mathbf{r}|/c = \mathbf{k}_s\cdot\mathbf{r}$ and

$$\mathbf{A}_s(\mathbf{r},t) = -i\frac{\mu_0 e^2\mathbf{E}_0}{4\pi m\omega_0}\cdot\frac{e^{i(\omega_0 t-\mathbf{k}_s\cdot\mathbf{r})}}{|\mathbf{r}|}. \tag{2.30}$$

This is an expression of a spherical wave with an amplitude $\propto 1/|\mathbf{r}|$. At great distances from the origin the wave fronts are nearly planes. Therefore we can use the approximation $\nabla\times \approx -i\mathbf{k}_s\times$ and Faraday's law to obtain the magnetic induction of the scattered wave as

$$\mathbf{B}_s = \nabla\times\mathbf{A}_s \approx -\frac{\mu_0 e^2(\mathbf{k}_s\times\mathbf{E}_0)}{4\pi m_e\omega_0}\cdot\frac{e^{i(\omega_0 t-\mathbf{k}_s\cdot\mathbf{r})}}{|\mathbf{r}|}. \tag{2.31}$$

The electric field of the wave is then simply $|\mathbf{E}_s| = c|\mathbf{B}_s|$ and the time average of the Poynting vector for the scattered radiation is

$$\langle|\mathbf{S}_s|\rangle = \langle|Re(\mathbf{E}_s)\times Re(\mathbf{H}_s)|\rangle = \frac{1}{2}|\mathbf{E}_s||\mathbf{H}_s|$$

$$= \frac{1}{2\mu_0}|\mathbf{E}_s||\mathbf{B}_s| = \frac{c}{2\mu_0}|\mathbf{B}_s|^2$$

$$= \frac{c}{2\mu_0}\left(\frac{\mu_0 e^2}{4\pi m_e \omega_0}\right)^2 \cdot \frac{|\mathbf{k}_s \times \mathbf{E}_0|^2}{|\mathbf{r}|^2}$$

$$= \frac{\mu_0 c}{2}\left(\frac{e^2}{4\pi m_e \omega_0}\right)^2 \cdot \frac{|\mathbf{k}_s|^2 \cdot |\mathbf{E}_0|^2 \cdot \sin^2 \mathcal{X}}{|\mathbf{r}|^2}$$

$$= \frac{\mu_0 c}{2}\left(\frac{e^2}{4\pi m_e \omega_0}\right)^2 \cdot \frac{\omega_0^2}{c^2} \cdot \frac{|\mathbf{E}_0|^2 \cdot \sin^2 \mathcal{X}}{|\mathbf{r}|^2}$$

$$= \frac{\mu_0 \varepsilon_0^2 c^3}{2}\left(\frac{e^2}{4\pi \varepsilon_0 m_e c^2}\right)^2 \cdot \frac{|\mathbf{E}_0|^2 \cdot \sin^2 \mathcal{X}}{|\mathbf{r}|^2}$$

$$= \frac{\varepsilon_0^2 c}{2} \cdot r_0^2 \cdot \frac{|\mathbf{E}_0|^2 \cdot \sin^2 \mathcal{X}}{|\mathbf{r}|^2}, \tag{2.32}$$

where

$$r_0 = \frac{e^2}{4\pi \varepsilon_0 m_e c^2} = 2.82 \cdot 10^{-15}\,\mathrm{m} \tag{2.33}$$

is so-called classical electron radius. Here $\mathcal{X}$ is the angle between the electric field of the incident wave and the scattering direction (Fig. 2.5).

Since the average Poynting vector of the incident radiation is

$$\langle|\mathbf{S}_i|\rangle = \frac{1}{2\mu_0}|\mathbf{E}_i||\mathbf{B}_i| = \frac{1}{2\mu_0 c}|\mathbf{E}_i|^2 = \frac{1}{2\mu_0 c}|\mathbf{E}_0|^2, \tag{2.34}$$

obviously

$$\langle|\mathbf{S}_s|\rangle = \frac{\varepsilon_0 c}{2} \cdot r_0^2 \cdot 2\mu_0 c\langle|\mathbf{S}_i|\rangle \cdot \frac{\sin^2 \mathcal{X}}{|\mathbf{r}|^2}$$

$$= r_0^2 \cdot \frac{\sin^2 \mathcal{X}}{|\mathbf{r}|^2}\langle|\mathbf{S}_i|\rangle. \tag{2.35}$$

Hence the power scattered to a solid angle $d\Omega$ around the direction $\mathcal{X}$ is

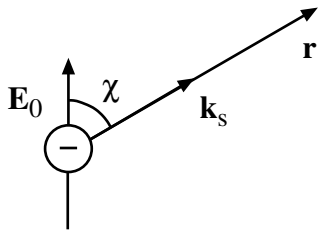

Figure 2.5: The angle between the electric field of the incident wave and the scattering direction.

$$dP(\mathcal{X}) = \langle|\mathbf{S}_i|\rangle \cdot \mathbf{r}^2 d\Omega = r_0^2 \sin^2 \mathcal{X} \langle|\mathbf{S}_i|\rangle d\Omega, \tag{2.36}$$

so that the differential cross section of Thomson scattering of a single electron is

$$\sigma(\mathcal{X}) = r_0^2 \sin^2 \mathcal{X}. \tag{2.37}$$

Note that this is different from the radar cross section defined in section 2.3. The total power scattered by a single electron is

$$P_T = \int \langle|\mathbf{S}_s|\rangle |\mathbf{r}|^2 d\Omega = 2\pi \int_0^\pi \langle|\mathbf{S}_s|\rangle r^2 \sin \mathcal{X} d\mathcal{X}$$

$$= 2\pi r_0^2 \langle|\mathbf{S}_i|\rangle \int_0^\pi \sin^3 \mathcal{X} d\mathcal{X} = \frac{8}{3}\pi r_0^2 \langle|\mathbf{S}_i|\rangle = \sigma_T \langle|\mathbf{S}_i|\rangle, \tag{2.38}$$

where the ratio of the total scattered power and the incident intensity,

$$\sigma_T = \frac{8}{3}\pi r_0^2, \tag{2.39}$$

is the total cross section of Thomson scattering, also known as the electron Thomson cross section.

2.3 Scattering from a moving electron

Unlike considered above, the electrons in the ionospheric plasma are not stationary but they may exhibit both thermal and bulk motion. Consider an electron moving with a velocity $\mathbf{v}$ in the antenna beam towards a radar transmitting at a frequency ν_0 (Fig. 2.6). The frequency of the incident wave in the rest frame of the moving electron is

$$\nu' = \nu_0 \sqrt{\frac{c+v}{c-v}}. \tag{2.40}$$

The electron emits radiation via the Thomson scattering mechanism at the same frequency in its own frame of reference. If the receiver is located at the transmitter site, the scattered radiation is observed at the frequency

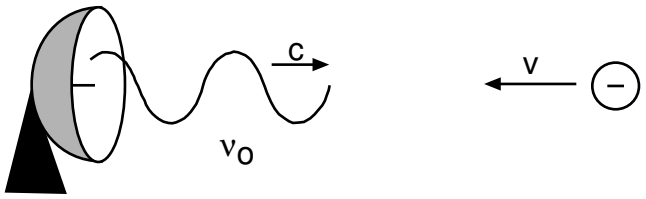

Figure 2.6: Scattering from an electron moving towards the radar.

$$\nu = \nu' \sqrt{\frac{c+v}{c-v}} = \nu_0 \frac{c+v}{c-v} \approx \nu_0 \left(1 + 2\frac{v}{c}\right). \tag{2.41}$$

Hence the observed Doppler shift is

$$\delta\nu = \nu - \nu_0 = 2\nu_0 \frac{v}{c} = 2\frac{v}{\lambda_0}, \tag{2.42}$$

where λ_0 is the radar wave length. If the electron is moving away from the transmitter, the sign of the Doppler shift is changed.

The scattering volume contains a large number of moving electrons. In the absence of bulk motion the density of electrons with a velocity component v_x along the radar beam follows the Maxwellian velocity distribution

$$\frac{dn_e}{dv_x} \propto e^{-v_x^2/v_m^2}, \tag{2.43}$$

where $v_m = \sqrt{2kT_e/m_e}$ is the most probable speed and T_e is the electron temperature. Each electron emits scattered radiation at a frequency determined by its velocity component, and therefore the spectrum of the scattered radiation is not monochromatic. The spectrum will be centred around the radar frequency and, since the scattering power of each electron is the same, the spectrum should have the same Maxwellian shape as the velocity distribution.

The velocity corresponding to the half-width of the spectrum is obtained by putting $\exp(-v_x^2/v_m^2) = 1/2$ which gives $v_x = \sqrt{\ln 2} \cdot v_m$. The spectrum half width is then

$$\Delta v = 2 \cdot 2\frac{v_x}{\lambda_0} = 4\frac{v_x}{\lambda_0} = \frac{4}{\lambda_0}\sqrt{ln2} \cdot \sqrt{2} \cdot \sqrt{\frac{kT_e}{m_e}} \approx \frac{4}{\lambda_0}\sqrt{\frac{kT_e}{m_e}}. \tag{2.44}$$

In the case of the EISCAT UHF radar, $\nu_0 = 930$ MHz and $\lambda_0 = 0.32$ m. If the electron temperature is $T_e = 1000$ K,

$$\Delta\nu = \frac{4}{0.32}\sqrt{\frac{1.38 \cdot 10^{-23} \cdot 1000}{9.11 \cdot 10^{-31}}}\,\text{Hz} \approx 1.5\,\text{MHz}.$$

In conclusion, the above consideration seems to indicate that the spectrum of the incoherent scatter signal has a Gaussian shape and the width of the spectral line obtained from a monochromatic transmission is of the order of 1 MHz.

In the 1950's the radar technology was developed to such a level that Gordon (1958) suggested an experiment of incoherent scatter from ionospheric electrons with a big antenna. The first observations were made in Long Branch, Illinois by Bowles (1958). Incoherent scatter signals were indeed observed, but the spectral width of the scattered signal was much smaller than expected. In fact, the spectrum was more in accordance with the motion of ions than electrons. The explanation was soon discovered. There are always thermal fluctuations in a plasma and, in a macroscopic sense, the scattering is caused by them as described in section 2.1. In the electron component of the plasma, the fluctuations

with wave lengths longer than the Debye length

$$\lambda_D = \sqrt{\frac{\varepsilon_0 k T_e}{n_e e^2}} \tag{2.45}$$

are mainly controlled by the motion of the massive ions since the light electrons must follow the ions in order to preserve the charge neutrality. Therefore, although the microscopic scattering mechanism is Thomson scatter by electrons, the shape of the incoherent scatter spectrum at wave lengths longer than the Debye length will be determined by the ion motion.

2.4 Incoherent scattering

As described above, the basic mechanism of incoherent scatter is Thomson scattering but the density fluctuations of the plasma have an essential role in determining the spectral shape. The expression of the incoherent scatter spectrum is quite complicated and it is not derived here. However, it is useful to give a qualitative picture on how the spectrum is formed.

The density fluctuations in the plasma can be expressed in terms of a temporal and 3-dimensional spatial Fourier transform. Each Fourier component represents a plane wave propagating in the direction of the **k**-vector of the spatial transform. These waves are not only mathematical constructions but they are real physical waves of thermal origin which must obey the dispersion relations of the medium. The two main wave modes which contribute to the thermal fluctuations are the ion- and electron-acoustic waves. The phase velocity of the ion-acoustic wave is

$$v_+ = \sqrt{\frac{k T_i}{m_i}\left(1 + \frac{T_e}{T_i}\right)}, \tag{2.46}$$

where m_i and T_i are the ion mass and temperature, respectively. This relation shows that the ion-acoustic wave is non-dispersive, *i.e.* its phase velocity does not depend on its wave length λ_+. The dispersion equation of the electron-acoustic wave is

$$v_- = f_p \sqrt{\lambda_-^2 + 12\pi^2 \lambda_D^2}, \tag{2.47}$$

where v_- and λ_- are the phase velocity and wave length, respectively, and f_p is the plasma frequency

$$f_p = \frac{1}{2\pi}\sqrt{\frac{n_e e^2}{\varepsilon_0 m_e}}. \tag{2.48}$$

The electron-acoustic wave is dispersive because its phase velocity depends on the wave length.

The Fourier component of the plasma wave causing constructive interference of the scattered radiation must fulfil eq. (2.20). In the case of backscattering the wave length of this component is $\Lambda = \lambda_0/2$. Furthermore, the component must obey the dispersion equation of one of the possible wave modes so that

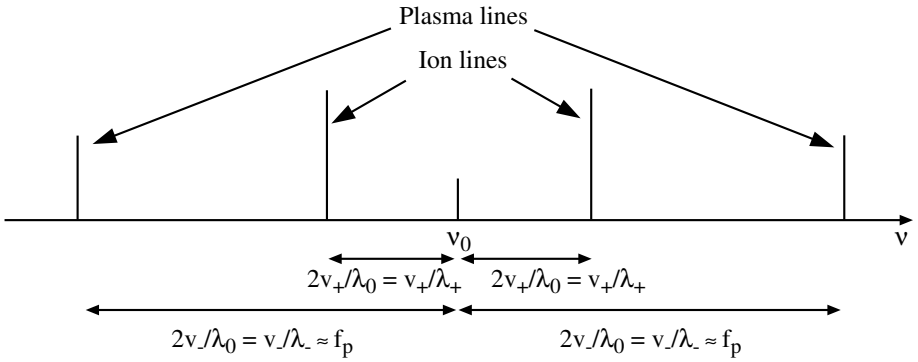

Figure 2.7: Ion and plasma lines.

it is a propagating wave and Λ is either λ_+ or λ_-. Because the phase fronts move at a phase velocity determined by the dispersion equation, the frequency of the scattered radiation is not the same as the frequency of the incident radiation but it is Doppler-shifted according to eq. (2.41). The density fluctuation contains Fourier components propagating in opposite directions, and therefore the scattering spectrum will contain both up- and downshifted lines. In the case of backscattering the Doppler shifts are $\pm 2v_+/\lambda_0 = \pm v_+/\lambda_+$ for the ion-acoustic wave and $\pm 2v_-/\lambda_0 = \pm v_-/\lambda_-$ for the electron-acoustic wave. Thus the scattering spectrum should contain four lines as indicated by Fig. 2.7.

It will be shown later that the spectrum and autocorrelation function of a signal are connected by Fourier transform. The spectrum of a monochromatic (sinusoidal) signal is a Dirac delta function and its autocorrelation function is also sinusoidal. Hence, if the scattering spectrum were composed of sharp

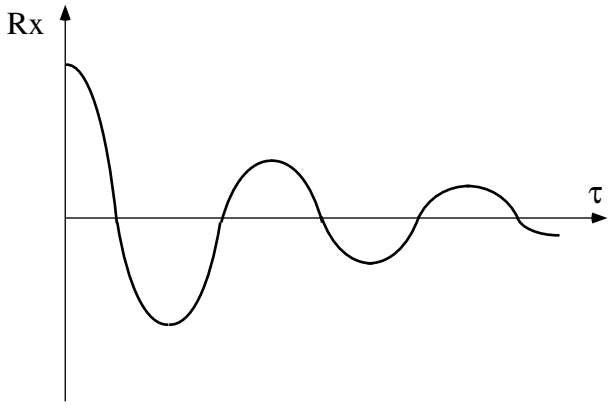

Figure 2.8: Damping of an autocorrelation function.

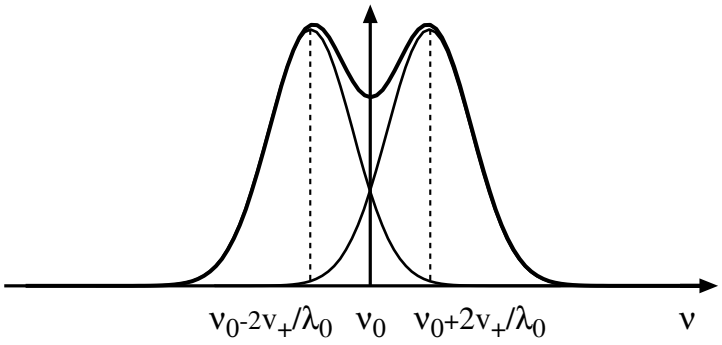

Figure 2.9: Generation of the ion line.

peaks like those shown in Fig. 2.7, the scattering signal should be composed of sinusoidal functions and each of them should be due to a sinusoidal plasma wave. In the plasma, however, new waves are thermally generated all the time and they are being attenuated by a process called Landau damping. Therefore the amplitude and phase of a certain Fourier component at a sufficiently distant time in the past have no relation to the the present amplitude and phase whatsoever. In terms of the autocorrelation function, this means that the autocorrelation function of the scattered signal is damped by a certain time constant as shown qualitatively in Fig. 2.8. Then the autocorrelation function is no more sinusoidal and the spectrum of the corresponding line is no more a delta function. The result is that the spectral lines will be broadened by Landau damping.

The broadening of the spectral lines of the ion-acoustic waves is sufficient to make them merge into a single line as shown in Fig. 2.9. This part of the spectrum is normally observed in incoherent scatter experiments and it is known as the ion line. At E and F region altitudes the ion line has a characteristic double-humped shape. The width of the ion line is roughly twice the Doppler shift of the ion-acoustic waves, *i.e.* $4v_+/\lambda_0$. If the ion and electron temperatures are nearly equal, eq. (2.46) gives $v_+ \approx \sqrt{2kT_i/m_i}$ so that the width of the ion line corresponds to the thermal speed of ions rather than electrons. This is the explanation for the narrow spectrum observed for the first time by Bowles (1958).

The remaining part of the incoherent scatter spectrum is due the electron-acoustic waves and it consists of two narrow peaks known as plasma lines (Fig. 2.7). If the radar wave length is much longer than the Debye length, the phase velocity of the electron-acoustic wave in eq. (2.47) is $v_- \approx f_p\lambda_-$ and the Doppler shifts of the plasma lines are nearly equal to the plasma frequency. Hence plasma lines can be used in determining the electron density. Unfortunately, the great ion mass makes the ion-acoustic waves dominate so that the plasma lines are only visible if electron-acoustic waves are excited *e.g.* by energetic electrons precipitating from the magnetosphere.

The term 'incoherent scatter' comes from the assumption that the signal is

scattered from evenly distributed electrons. If this were true, the phases of the elementary signals from individual electrons were completely random and the total signal would show no coherence. The coherence of a signal is measured in terms of its autocorrelation function, the longer the autocorrelation function is, the higher is the coherence level. The autocorrelation function of a truly incoherent signal would be a delta function, *i.e.* its coherence length would be zero. As seen above, the autocorrelation function of our 'incoherent scatter signal' is far from being a delta function and therefore, strictly speaking, the term is a misnomer. The name was first adopted, because the role of plasma fluctuations was not understood in the beginning and it has remained in use for historical reasons. It is our luck that 'incoherent scatter' is not incoherent. This very fact allows the determination of plasma parameters other than electron density and electron temperature.

The shape of the ion line in the incoherent scatter spectrum is determined by

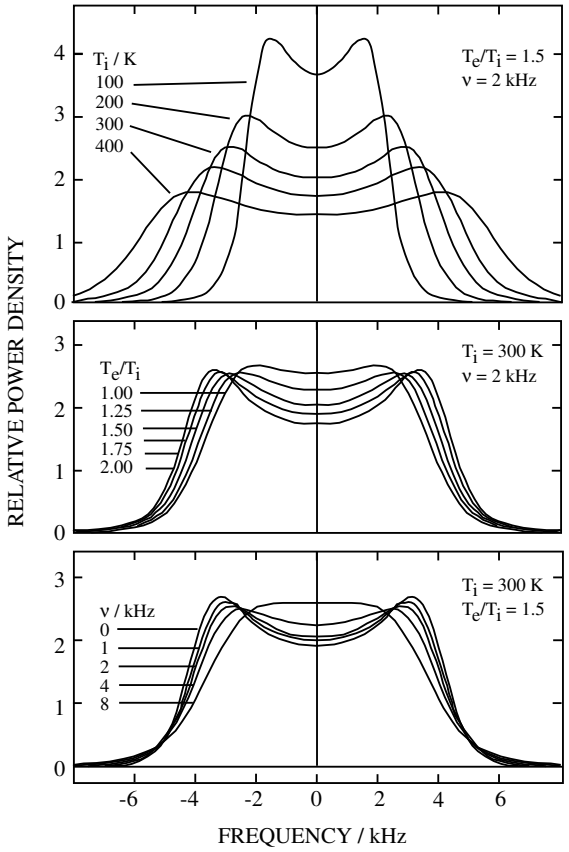

Figure 2.10: Dependence of the ion line shape on plasma parameters.

various plasma parameters. For instance, eq. (2.46) shows that the width of the ion line is determined by the ion mass and the ion and electron temperatures. A more complete theory gives an expression of the line shape which is used in determining the values of the plasma parameters by the least squares method.

In order to give a qualitative picture on the effect of various plasma parameters on the spectrum shape, spectra at various parameter combinations are plotted in Fig. 2.10. The top panel shows the broadening of the ion line with increasing ion temperature, the middle panel the deepening of the centre minimum with increasing electron temperature and the bottom panel the reduction of the steepness of the edges with increasing ion-neutral collision frequency. In principle, all these parameters can be determined from the spectrum shape, if the ion mass is known. The effects of electron temperature and collision frequency, however, are so similar that they may be difficult to separate in practice. In routine work this problem is often solved by putting the temperature ratio to unity in the lower E region and collision frequency to zero at greater heights.

The remaining parameters which can be determined from the ion line are the electron density and ion velocity. The electron density is determined from the received power (*i.e.* the integral of the spectrum) which is proportional to the number of electrons in the scattering volume. If the ionospheric plasma is in motion, the whole ion line is shifted and the bulk speed can be determined from the Doppler shift. In the case of backscattering the Doppler shift gives the line-of-sight ion velocity.

In addition to the main physical parameters which are determined from the observed ion line, a set of secondary parameters can also be calculated. These include the electric field, conductivity tensor, electric current, neutral wind velocity, Joule heating *etc.*

2.5 Generation of high transmitting power

High transmitting powers are needed in incoherent scatter radars because the scattering cross section of the electron is very small. Peak powers higher than 1 MW are common in older radars but in modern setups lower powers can be accepted. This is because higher radar duty cycles (the fraction of time the transmitter can operate at full power) can now be achieved and the modulation methods have also been greatly improved. The peak powers of the present EISCAT UHF and VHF radars are 1.7 MW and 1.5 MW, respectively (the nominal values are essentially higher), but the peak power of the EISCAT Svalbard Radar (ESR) in its first stage is only 0.5 MW. The ESR, however, is more efficient than these numbers might indicate, since its duty cycle is 25 %, whereas the duty cycles of the older EISCAT radars are only of the order of 10–15 %.

To achieve the high power needed in the incoherent scatter transmitter, a device called klystron is used. Klystron is a big electron tube shown schematically in Fig. 2.11. The tube contains an electron gun which generates a narrow electron beam in an accelerating electric field between a cathode and an anode. The electrons are directed to a tube-like radio frequency section, which contains

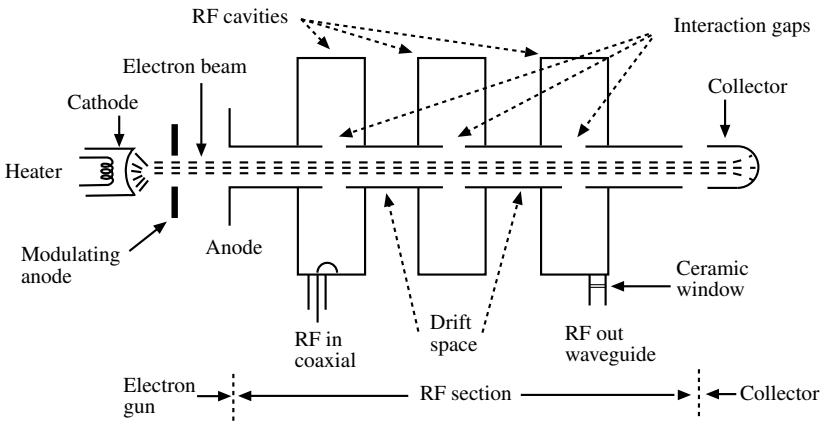

Figure 2.11: Klystron.

a few resonance cavities, and finally they hit the collector which removes them from the system.

The resonance cavities are used to produce a powerful radio frequency signal. An input signal is applied to the first cavity where its electric field accelerates and decelerates the electrons at its own frequency creating a velocity modulation in the beam. While the electrons drift in the tube, a density modulation will develop bunching the beam into sections of higher electron density. As a result, an electric field will develop between the drifting electron bunches and, if a resonance cavity is put in the place where the bunching is maximal, a strong electric field oscillating at the frequency of the input signal is observed in the cavity. A magnetic field is associated with the oscillating electric field and therefore a powerful electromagnetic signal at the frequency of the input signal is obtained. The additional cavities between the input and output cavities produce extra bunching of the beam thus increasing the gain of the klystron.

The signal is led from the output cavity to a waveguide or a coaxial cable and fed to the transmitting antenna. Waveguides are used in the EISCAT UHF and VHF radars. As any electron tube, a klystron must be evacuated and therefore a ceramic window is used between the klystron and the waveguide.

The electrons are confined to a narrow beam by an axial magnetic field generated by solenoids. It is rather essential to have a good focusing of the beam in high power klystrons since the stray electrons impinging upon the inner surface may cause overheating and destroy the tube. Only a fraction of the power in the beam is transformed to electromagnetic radiation; most of it is transformed to heat in the collector. From the collector the heat is transferred by an effective cooling system. In the EISCAT UHF and VHF radars the electrons hitting the collector cause bremsstrahlung x-rays and therefore there is radiation danger near the operating klystron. This is not the case in the ESR because of lower beam energy.

The EISCAT UHF radar has one klystron, the VHF radar has nominally two

but only one of them is in use, and the ESR has four klystrons in its first stage. The number of klystrons in the ESR can be increased to 8 or 16 and their signals are added in the waveguide. This is now possible because the modern klystrons can be synchronised with a sufficient accuracy. The EISCAT UHF and VHF klystrons are specially designed for these radars but in the ESR commercial TV tubes are used.

2.6 Radiation pattern of a parabolic antenna

The purpose of the radar antenna is to confine the transmitted power into a narrow beam and, when the antenna is used in reception, to enhance the strength of the signal from a narrow solid angle at the cost of signals from other directions. In order to give a physical idea of how an antenna reflector works, a simple derivation of the radiation pattern of a parabolic dish antenna is presented in this section. The EISCAT UHF and the ESR antennas are parabolic dishes with Cassegrain feed, *i.e.* the focused rays are reflected by a secondary mirror to the end of a feed horn at the centre of the main reflector. The EISCAT VHF antenna is a parabolic cylinder and its feed is arranged by 128 cross-dipole antennas along the focal line of the reflector.

A parabolic dish surface has the special property that it reflects all incident rays parallel to the optical axis into a same focal point. Similarly, all rays emitted from the focal point towards the surface are reflected in the direction parallel to the optical axis.

The antenna radiation pattern is the diffraction pattern of an antenna at a long distance from the reflector. From the diffraction point of view, a parabolic antenna behaves like a circular opening in a screen, illuminated by by a distant radiation source at the opposite side. Hence we only have to study Fraunhofer

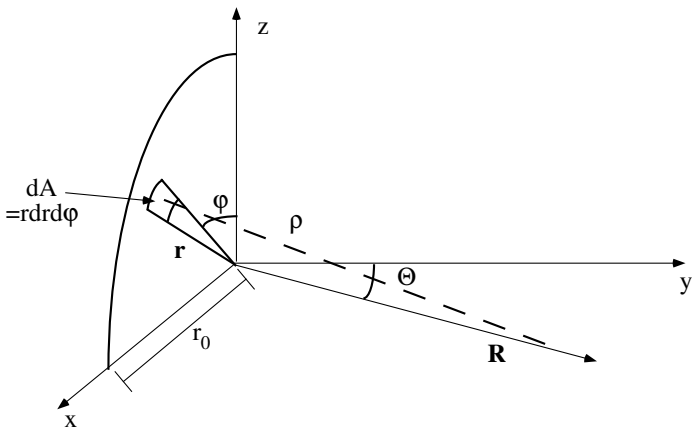

Figure 2.12: Circular opening in a screen.

diffraction in a circular slit.

In Fig. 2.12 a circular opening is located in the xz plane (only one fourth of it is shown) and the radiation enters the slit from the left. Because of the cylindrical symmetry it is sufficient to study the radiation in the xy plane. The electromagnetic field at $\mathbf{R}$ is the sum of elementary waves originating from all surface elements dA. The electric field of the elementary wave from dA is

$$dE = Ce^{i(\omega t - k\varrho)}dA, \tag{2.48}$$

where C is an amplitude coefficient, $k = \omega/c$ is the wave number and $\varrho = |\mathbf{R}-\mathbf{r}|$ is the distance from the source to the observing point. Because

$$\mathbf{r} = r(\sin\varphi\mathbf{u}_x + \cos\varphi\mathbf{u}_z) \tag{2.49}$$

and

$$\mathbf{R} = R(\sin\theta\mathbf{u}_x + \cos\theta\mathbf{u}_y), \tag{2.50}$$

the distance from the surface element to the observing point is

$$\begin{aligned}\varrho &= |(R\sin\theta - r\sin\varphi)\mathbf{u}_x + R\cos\theta\mathbf{u}_y - r\cos\varphi\mathbf{R}_z| \\ &= \sqrt{(R\sin\theta - r\sin\varphi)^2 + (R\cos\theta)^2 + (r\cos\varphi)^2} \\ &= \sqrt{R^2 - 2Rr\sin\theta\sin\varphi + r^2} \\ &\approx R\sqrt{1 - 2\frac{r}{R}\sin\theta\sin\varphi} \approx R\left(1 - \frac{r}{R}\sin\theta\sin\varphi\right) \\ &= R - r\sin\theta\sin\varphi. \end{aligned} \tag{2.51}$$

The above approximations are valid, since $R \gg r$ in Fraunhofer diffraction. The total electric field at $\mathbf{R}$ is

$$\begin{aligned}E &= \int_A Ce^{i(\omega t - k\varrho)}dA \\ &= Ce^{i\omega t}\int_A e^{ikR + ik\sin\theta\cdot r\sin\varphi}dA. \end{aligned} \tag{2.52}$$

When $R \gg r$, the wave vector of the elementary wave from dA is approximately $\mathbf{k} \approx k(\sin\theta\mathbf{u}_x + \cos\theta\mathbf{u}_y)$ and therefore an approximate expression for the electric field is

$$E \approx Ce^{i(\omega t - kR)}\int_A e^{i\mathbf{k}_\parallel\cdot\mathbf{r}}dA, \tag{2.53}$$

where $\mathbf{k}_\parallel \approx k\sin\theta\mathbf{u}_x$ is the component of the wave vector parallel to the xz plane. Note that this result is not restricted to a circular opening but it is generally valid for Fraunhofer diffraction in a slit of any shape. Eq. (2.53) shows that the diffraction pattern of the slit is obtained from a two-dimensional Fourier transform of a function which is equal to a non-zero constant inside the slit and zero elsewhere.

In a circular opening, the wave amplitude at $\mathbf{R}$ can be further developed as

$$E_0 = C \int_0^{r_0} \int_0^{2\pi} e^{ik_\parallel \cdot r \sin\varphi} r d\varphi dr$$

$$= C \int_0^{r_0} r \left[\int_0^{2\pi} e^{ik_\parallel \cdot r \sin\varphi} d\varphi \right] dr. \tag{2.54}$$

The inner integral in this expression is

$$I = \int_0^{2\pi} \cos(k_\parallel r \sin\varphi) d\varphi + i \int_0^{2\pi} \sin(k_\parallel r \sin\varphi) d\varphi. \tag{2.55}$$

Since $\sin[k_\parallel r \sin(\varphi + \pi)] = \sin(-k_\parallel r \sin\varphi) = -\sin(k_\parallel r \sin\varphi)$, the latter integral in eq. (2.55) is zero. Also, $\cos[k_\parallel r sin(\varphi + \pi)] = \cos(-k_\parallel r \sin\varphi) = \cos(k_\parallel r \sin\varphi)$ so that

$$I = 2 \int_0^{\pi} \cos(k_\parallel r \sin\varphi) d\varphi. \tag{2.56}$$

Using the zeroth order Bessel function

$$J_0(x) = \frac{1}{\pi} \int_0^{\pi} \cos(x \sin\varphi) d\varphi \tag{2.57}$$

this can be put in the form

$$I = 2\pi J_0(k_\parallel r) \tag{2.58}$$

and the wave amplitude becomes

$$E_0 = 2\pi C \int_0^{r_0} r J_0(k_\parallel r) dr$$

$$= \frac{2\pi C}{k_\parallel^2} \int_0^{k_\parallel r_0} r' J_0(r') dr'. \tag{2.59}$$

The Bessel functions of first kind obey a general recursive formula

$$\frac{d}{dx}[x^n J_n(x)] = x^n J_{n-1}(x). \tag{2.60}$$

By applying this to the Bessel functions of zeroth and first order we obtain

$$\frac{d}{dx}(r' J_1(r')] = r' J_0(r'). \tag{2.61}$$

This can be used in calculating the integral in eq. (2.59) so that the wave amplitude is

$$E_0 = \frac{2\pi C r_0}{k_\parallel} J_1(k_\parallel r_0) = \frac{2\pi C r_0}{k \sin\theta} J_1(k r_0 \sin\theta)$$

$$= C \frac{\lambda r_0}{\sin\theta} \cdot J_1\left(\frac{2\pi r_0 \sin\theta}{\lambda} \right). \tag{2.62}$$

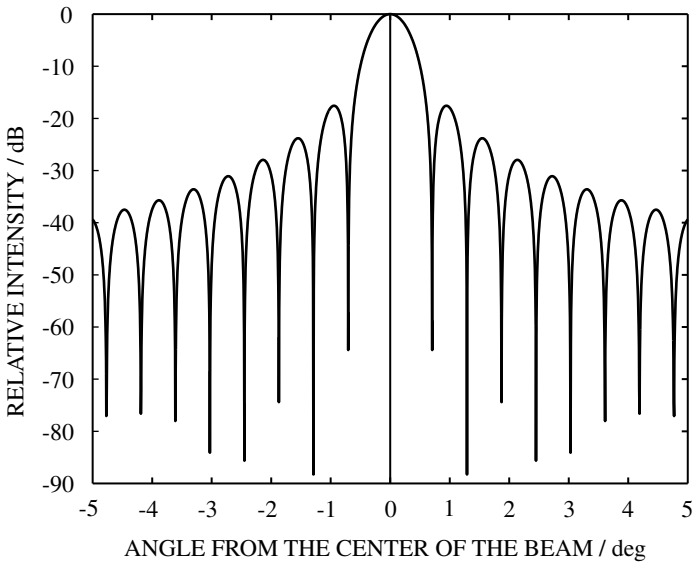

Figure 2.13: Radiation pattern of a parabolic antenna with a 32 m diameter for a frequency of 930 MHz.

The intensity is proportional to the square of the amplitude. Thus the radiation pattern of the antenna is

$$S \propto \left[\frac{\lambda r_0}{\sin\theta} \cdot J_1 \left(\frac{2\pi r_0 \sin\theta}{\lambda} \right) \right]^2 , \tag{2.63}$$

where θ is the angle between the direction of observation and the optical axis.

In the EISCAT UHF antenna $r_0 = 16$ m and $f = 930$ MHz. The antenna pattern (2.63) is plotted in Fig. 2.13 using these parameter values. We see that the pattern consists of a strong main lobe centred around the optical axis and a series of less intense side lobes surrounding it as circular structures. The intensity of the first side lobe is nearly 18 dB below that of the main lobe. The antenna beam width, defined in terms of the -3 dB points of the main lobe, is about 0.6°.

In reality there is no infinite screen around a parabolic antenna as assumed above. Therefore the antenna does not behave exactly as a circular slit but the diffraction pattern extends to angles greater than 90°. Furthermore, the secondary mirror and its supports shade the antenna, which weakens the main lobe and creates distortions in the side lobes. Conducting structures and the landscape nearby the antenna also have an effect on the radiation pattern.

2.7 Antenna gain and effective aperture

The antenna gain is a concept which is closely related to the radiation pattern discussed in Chapter 2.6. The antenna gain G in a given direction is defined as a factor by which the intensity is increased as compared to the intensity of an isotropic radiator. If the total transmitted power is P, the intensity at the point $\mathbf{r}$ would then be

$$S(\mathbf{r}) = \frac{GP}{4\pi r^2}. \tag{2.64}$$

On the other hand, the power transmitted to a unit solid angle by an isotropic radiator would be $P/4\pi$. Hence, if dP is the power transmitted by our antenna to a solid angle $d\Omega$ around the direction Ω, the antenna gain in this direction is

$$G = \frac{4\pi}{P}\left(\frac{dP}{d\Omega}\right). \tag{2.65}$$

Because the total power is

$$P = \int_{4\pi} \frac{dP}{d\Omega} d\Omega, \tag{2.66}$$

the antenna gain can be written as

$$G = \frac{4\pi\,(dP/d\Omega)}{\int (dP/d\Omega) d\Omega}. \tag{2.67}$$

Furthermore, since $dP/d\Omega \propto S$, we can also write

$$G = \frac{4\pi S(\Omega)}{\int S(\Omega) d\Omega}. \tag{2.68}$$

The radiation pattern of a directional antenna gets its maximum value in some direction and the gain corresponding to this direction is the maximum gain G_{max}. If the radiation pattern is axially symmetric like in a parabolic dish and the maximum gain is obtained in the direction $\theta = 0$, it can be calculated from

$$\frac{1}{G_{max}} = \frac{1}{4\pi S(0)} \int S(\theta) d\Omega = \frac{1}{2S(0)} \int_0^{\frac{\pi}{2}} S(\theta) \sin\theta d\theta. \tag{2.69}$$

Here it is assumed that the antenna pattern covers the half-space like in Chapter 2.6., otherwise the integration must be carried from 0 to π. In the case of a parabolic dish the intensity can be written in the form

$$S \propto \left[2\pi r_0^2 \cdot \frac{\lambda}{2\pi r_0 \sin\theta} \cdot J_1\left(\frac{2\pi r_0 \sin\theta}{\lambda}\right)\right]^2. \tag{2.70}$$

Using the notation

$$\xi = \frac{2\pi r_0 \sin\theta}{\lambda} \tag{2.71}$$

this can be put in the form

$$S \propto \left[2\pi r_0^2 \cdot \frac{J_1(\xi)}{\xi}\right]^2. \tag{2.72}$$

It can be shown that $\lim_{\xi\to 0}[J_1(\xi)/\xi] = 1/2$, so that

$$S(0) \propto \left[2\pi r_0^2 \cdot \frac{1}{2}\right]^2 = (\pi r_0^2)^2 = A^2, \tag{2.73}$$

where A is the aperture of the antenna. Then

$$\begin{aligned}\frac{1}{G_{max}} &= \frac{1}{2A^2}\int_0^{\frac{\pi}{2}} \left[\frac{\pi r_0}{\sin\theta}\cdot J_1\left(\frac{2\pi r_0 \sin\theta}{\lambda}\right)\right]^2 \sin\theta d\theta \\ &= \frac{\lambda^2 r_0^2}{2A^2}\underbrace{\int_0^{\frac{\pi}{2}} \frac{\left[J_1\left(2\pi\cdot r_0/\lambda\cdot\sin\theta\right)\right]^2}{\sin\theta} d\theta}_{I} \\ &= \frac{\lambda^2\pi r_0^2}{2\pi A^2}\cdot I = \frac{\lambda^2}{4\pi A}\cdot 2I \end{aligned} \tag{2.74}$$

and

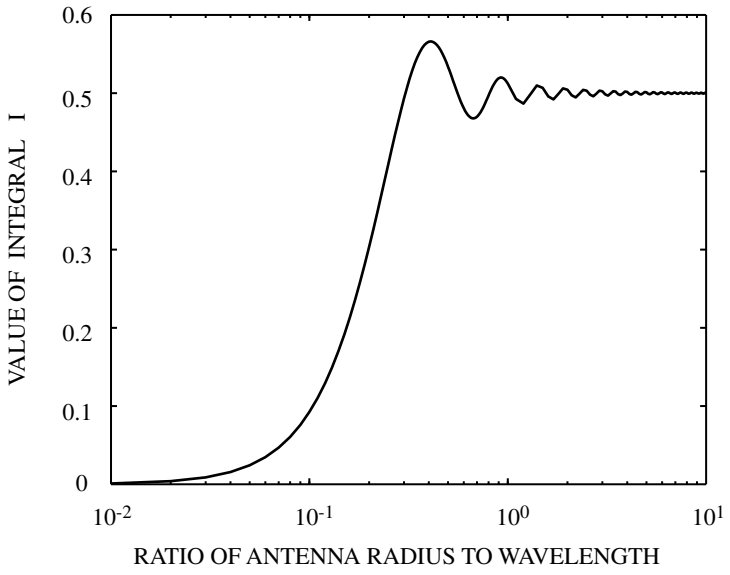

Figure 2.14: The value of the integral I in eq. (2.74) as a function of the ratio of antenna radius to wavelength.

$$G_{max} = \frac{4\pi\rho A}{\lambda^2}, \tag{2.75}$$

where $\rho = 1/2I$ is the antenna efficiency. By defining the effective aperture of the antenna as $A_e = \rho A$, the gain can be written in the form

$$G_{max} = \frac{4\pi A_e}{\lambda^2}. \tag{2.76}$$

The variation of the gain with direction is simply determined by the shape of the radiation pattern. A gain in a given direction can be expressed as

$$G(\theta) = \frac{S(\theta)}{S(0)} G_{max} = \frac{4\pi}{\lambda^2} \cdot \frac{S(\theta)}{S(0)} A_e = \frac{4\pi A_e(\theta)}{\lambda^2}, \tag{2.77}$$

where $A_e(\theta)$ is the effective aperture for the direction θ.

The value of I depends on the ratio of the antenna radius to the wavelength and it is plotted in Fig. 2.14. The oscillations in the curve are attenuated with increasing r_0/λ and the limiting value of I is obviously 0.5. At $r_0 = 10\lambda$ the value of I is already rather close to 0.5. In the case of the EISCAT UHF antenna

$$\frac{r_0}{\lambda} \approx \frac{16}{0.32} = 50,$$

so that it is a good approximation to put

$$\rho = \frac{1}{2I} = \frac{1}{2 \cdot 0.5} = 1, \tag{2.78}$$

which gives $A_e \approx A$ and

$$G_{max} = \frac{4\pi A}{\lambda^2}. \tag{2.79}$$

This result is valid for an ideal parabolic dish with $r_0/\lambda \gg 1$. In a real antenna the diffraction pattern and the antenna efficiency are affected by the factors discussed in the previous section. In addition, the antenna feed has its own efficiency which reduces the total efficiency of the antenna system.

2.8 Monostatic and multistatic radar

A monostatic radar has a single receiver which is located at the transmitter site. A multistatic radar may have several receivers and at least one of them is located far from the transmitter. In modern systems one of the receivers is always at the transmitter site but in some older radars, which used continuous transmission, reception was possible only at remote sites. The EISCAT UHF system is multistatic and both the EISCAT VHF and ESR systems are monostatic.

When the transmitter and the receiver are in different places, the scattered signal is received from the intersection of the two radar beams (Fig. 2.15). The signal is used to determine the values of plasma parameters within this

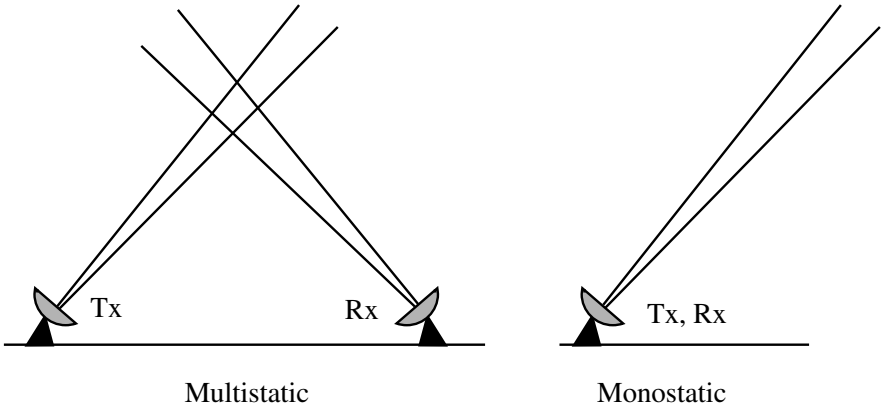

Figure 2.15: The principle of monostatic and multistatic radar.

scattering volume. Profiles of plasma parameters can in principle be determined by sweeping the receiver beam along the transmitter beam but this method is so slow that only a few scattering volumes can be obtained in practice.

In a monostatic radar, transmission is carried out with a certain modulation during a limited time interval and the scattered signal is observed during a subsequent reception period. The signal at a single instant of time comes from a scattering volume limited in the direction perpendicular to the radar beam by the beam itself and in the beam direction by the length of the transmitted pulse or pulse sequence. In the course of time the scattering volume moves upwards along the radar beam and therefore profiles of plasma parameters can be determined. The length of the scattering volume in the beam direction determines the range resolution of the incoherent scatter measurement. Different modulation methods are used for obtaining different range resolutions. Simple long pulses are applied if a resolution of some tens of kilometres is sufficient. In order to go down to a few kilometres, pulse codes or alternating codes are needed. The best possible resolutions are a few hundred metres and this is achieved by so called Barker codes.

A multistatic radar with one of the receivers at the transmitter site can also be used in monostatic mode. In fact, this is a common practice because the monostatic mode is more efficient in measuring the range profiles of the plasma parameters. In normal experiments both mono- and multistatic measurements are carried out simultaneously with different modulations. The main role of the multistatic part is the observation of the plasma velocity vector, which gives the ionospheric electric field if the scattering volume is high up in the F region.

2.9 Radar equation

Let us consider the multistatic (bistatic) radar system in Fig. 2.15. The gains of the transmitter and receiver antennas are G_1 and G_2, respectively, and the

effective aperture of the receiver antenna is A_{e2}. The position vectors from the transmitter and receiver antennas to a volume element at $\mathbf{r}$ are $\mathbf{r_1}$ and $\mathbf{r_2}$. Signal can only be observed from this volume element if it lies inside the scattering volume at the crossing of the radar beams.

If there is only a single electron inside the scattering volume, the power captured by the receiver antenna is

$$P_r(\mathbf{r}) = r_0^2 \sin^2 \mathcal{X} \cdot P_t \cdot \frac{G_1(\mathbf{r_1})}{4\pi r_1^2} \cdot \frac{A_{e2}(\mathbf{r_2})}{r_2^2}. \tag{2.80}$$

or, using the relation of effective aperture and antenna gain in eq. (2.77),

$$P_r(\mathbf{r}) = 4\pi r_0^2 \sin^2 \mathcal{X} \cdot P_t \cdot \frac{G_1(\mathbf{r_1})}{4\pi r_1^2} \cdot \frac{G_2(\mathbf{r_2})}{4\pi r_2^2} \cdot \frac{\lambda^2}{4\pi}. \tag{2.81}$$

Eq. (2.80) is also often written in the form

$$P_r(\mathbf{r}) = \sigma_0(\mathcal{X}) \cdot P_t \cdot \frac{G_1(\mathbf{r_1})}{4\pi r_1^2} \cdot \frac{A_{e2}(\mathbf{r_2})}{4\pi r_2^2}. \tag{2.82}$$

The previous three expressions are various forms of the radar equation. The quantity

$$\sigma_0 = 4\pi r_0^2 \sin^2 \mathcal{X} = 1.00 \cdot 10^{-28}\,\mathrm{m}^2 \cdot \sin^2 \mathcal{X} \tag{2.83}$$

in eq. (2.82) is called the radar cross section. In radar theory the radar cross section is commonly used instead of the differential cross section which is the equivalent standard concept in scattering theory. This is an unfortunate practice which may cause confusion but, in the case of Thomson scattering, it has the benefit that the numerical value of the radar cross section is easy to remember. The (nonphysical) idea behind the concept of radar cross section is simply that the target would scatter the incident radiation uniformly in all directions. Then the total scattered power would be the radar cross section times the incident intensity and a fraction $d\Omega/4\pi$ of it would be scattered in the solid angle $d\Omega$.

Due to the small value of the cross section, Thomson scattering is very weak. If, for example, the transmitted power is 1 MW and the cross section of the radar beam at a distance of 300 km is $10^3\,\mathrm{m} \times 10^3\,\mathrm{m} = 10^6\,\mathrm{m}^2$, the incident intensity at this distance is $1\,\mathrm{W/m}^2$. A typical value of electron density in the ionospheric F region is $n_e \approx 10^{12}\,\mathrm{m}^{-3}$ and the scattering volumes in F region incoherent scatter experiments are of the order of $V \approx 10^3\,\mathrm{m} \times 10^3\,\mathrm{m} \times 10^4\,\mathrm{m} = 10^{10}\,\mathrm{m}^3$. For a rough estimation, one can assume that the observed power is proportional to the number of electrons in the scattering volume so that the total radar cross section is about $\sigma_{tot} \approx n_e V \sigma_0 = 10^{12} \cdot 10^{10} \cdot 10^{-28}\,\mathrm{m}^2 = 10^{-6}\,\mathrm{m}^2$. The power received by an antenna with an effective aperture of $100\,\mathrm{m}^2$ at a distance of 300 km is only $P_r \approx 10^{-6}\,\mathrm{m}^2 \times 1\,\mathrm{W/m}^2 \times 100\,\mathrm{m}^2/(4\pi \cdot 300^2\,\mathrm{km}^2) \approx 10^{-16}\,\mathrm{W}$. Hence the radar receiver must be capable of detecting powers of this order even in the presence of thermal noise.

As discussed in Chapter 2.4, the electron density can be determined from the received power. The general principle of using the radar equation for this

purpose is described below. The formulas, however, only show a rough idea of the method, because the effects of transmitter modulation and receiver filtering are not taken into account.

Assume a monostatic radar so that $\mathbf{r} = \mathbf{r}_1 = \mathbf{r}_2$ and $G_1 = G_2 = G$ and consider received power from a volume element dV located at $\mathbf{r}$. The power is proportional to the number of electrons in the volume element so that

$$d^3P_r(\mathbf{r}) = n_e(\mathbf{r})\sigma P_t \cdot \frac{G^2(\mathbf{r})}{(4\pi r^2)^2} \cdot \frac{\lambda^2}{4\pi} dV, \tag{2.84}$$

where σ is the radar cross section per electron. Although eq. (2.81) gives $\sigma = 4\pi r_0^2$ in the case of backscattering, this is actually not the correct expression because the plasma effects discussed in Chapter 2.4 modify the scattering cross section. A better approximation given by the plasma theory is

$$\sigma = \frac{4\pi r_0^2}{(1 + k^2\lambda_D^2)(1 + T_e/T_i + k^2\lambda_D^2)}. \tag{2.85}$$

In spherical coordinates we can put $dV = r^2 d\Omega dr$, and therefore the power received from the height interval $(r, r + \Delta r)$ is equal to

$$\Delta P_r(r) = \frac{n_e(r)\sigma P_t \lambda^2}{(4\pi)^3 r^2} \left[\int_\Omega G^2(\Omega) d\Omega \right] \Delta r. \tag{2.86}$$

Here Δr is assumed to be small. The plasma parameters are also assumed to be constants within the cross section of the radar beam as well as within the range $(r, r + \Delta r)$. Solving from eq. (2.86) the electron density is

$$n_e(r) = C \cdot \frac{\Delta P_r(r)}{P_t} \cdot \frac{r^2}{\Delta r} \cdot \frac{1}{\sigma}, \tag{2.87}$$

where C is a constant determined by the antenna pattern and the radar wave length. When eq. (2.85) is used for the radar cross section, this gives

$$n_e(r) = C \cdot \frac{\Delta P_r(r)}{P_t} \cdot \frac{r^2}{\Delta r} \cdot \frac{(1 + k^2\lambda_D^2)(1 + T_e/T_i + k^2\lambda_D^2)}{4\pi r_0^2}. \tag{2.88}$$

Hence we see that the electron density depends not only on the received power but also on Debye length (which depends on the electron temperature and electron density!) and the ratio of electron and ion temperatures. In many cases the radar wave length is much larger than the Debye length, which simplifies the equation because the terms $k^2\lambda_D^2$ can be neglected. As a first approximation, it is also sometimes assumed that $T_e = T_i$ and then the electron density can be calculated from the formula

$$n_e(r) = C \cdot \frac{\Delta P_r(r)}{P_t} \cdot \frac{r^2}{\Delta r} \cdot \frac{2}{4\pi r_0^2}. \tag{2.88}$$

This result is sometimes called the 'raw electron density'. Note that the electron density given by eq. (2.88) is twice the value obtained using the cross section $4\pi r_0^2$.

The above equations do not explain how the power of scattered signal from a given height range can be obtained in a monostatic case. For this purpose different modulations discussed in Chapter 5 must be utilised. These modulations determine the altitude range of received power but further complications are associated with their use. Actually the determination of electron density is more complex than shown by the above equations. Instead of observing only the signal power, we measure its autocorrelation function. The concept of autocorrelation function is explained in Chapter 3. The signal autocorrelation function depends on the plasma autocorrelation function in the way described by the theory of ambiguity functions in Chapter 4. In practice the electron density is not calculated separately from eq. (2.88) but it is determined together with the other plasma parameters by fitting theoretical autocorrelation functions to the observed ones.

Chapter 3

Signal processing

3.1 Radar receiver

The signal received by an incoherent scatter radar is caused by random fluctuations in the ionosphere and therefore it consists of random variations. The theoretical concept suitable for mathematical treatment of such signals is called a stochastic process. A stochastic process is a concept which attaches a random variable to each instant of time. Every variable has its own probability distribution. The signal values are simply random values of these variables and an individual signal is called a realisation of the stochastic process. Hence, if a large set of identical ionospheres with the same values of macroscopic parameters (temperature, density, pressure, etc.) were probed by synchronised identical radars, simultaneous signal values from the radars would be different. Also, the detailed waveforms of the two signals in the same radar, received after successive transmission sequences, carry no resemblance whatsoever. In both cases, however, the signals do have the common feature that they obey the same probability distributions. This means that the statistical properties of the signals, like their spectra for instance, are identical.

We are interested in plasma parameters, which depend on the spectral shape, and therefore there is no need to know the waveform of the signal. Hence the receiver of an incoherent scatter radar can be designed for observing the received power and spectral shape only. In practice the spectrum is obtained from its Fourier transform, the autocorrelation function, which is calculated from digital samples of the signal. Therefore the receiver should contain an AD-converter. It will be seen later that a correct sampling of a signal implies a sampling frequency at least twice the highest frequency in the spectrum. The signal spectrum is band-limited around the transmission frequency so that the sampling frequency should be higher than twice the transmission frequency. In the case of the EISCAT UHF radar this would imply an unreasonably fast sampling at a frequency of the order of 1.9 GHz.

The problem can be avoided by mixing the signal to a lower frequency. This

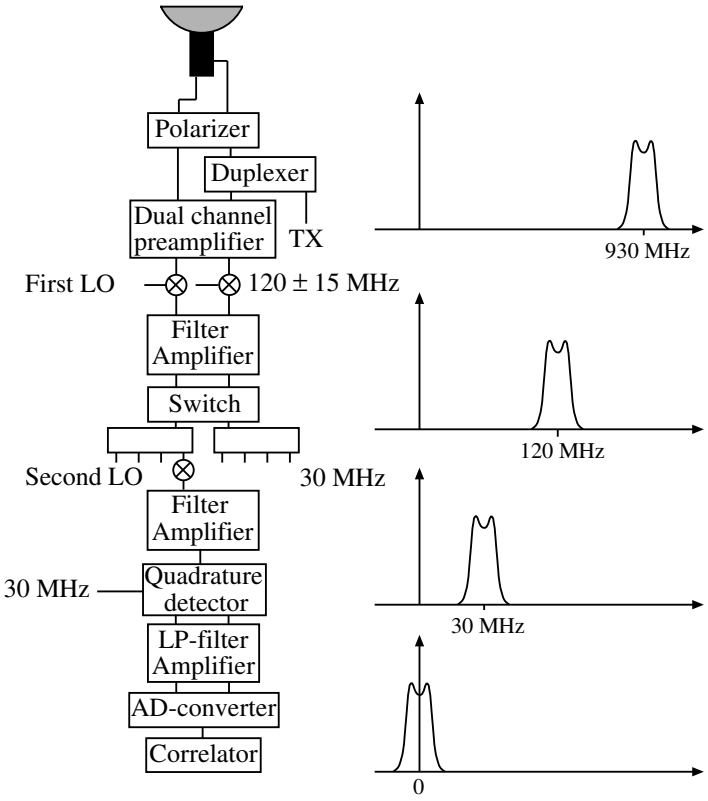

Figure 3.1: The block diagram of the EISCAT UHF receiver and the incoherent scatter spectrum at various stages of the signal path.

means creating a new signal which has the same spectral shape as the original one, but at an essentially lower frequency. The sampling frequency gets its lowest value, if the spectrum is shifted to lie around zero on the frequency axis. This can be achieved either by complex mixing or Hilbert transform.

The idea of the incoherent scatter receiver is enlightened by Fig. 3.1, which shows the block diagram of the EISCAT UHF system together with the spectra at different stages of the signal path. The radio wave captured by the antenna is directed to the feed horn by the secondary mirror and the two orthogonal polarisations are separated to different legs of the wave guide. The system is designed for processing both polarizations but, using the polarizes unit, it is possible to adjust the phase difference of the two components to give a non-zero signal in a single channel only. After the first local oscillator and subsequent filtering the system consists of several parallel channels. This arrangement is necessary because several transmitting frequencies are normally used in an incoherent scatter experiment, and these must be separated in the receiver.

When the signal arrives at the preamplifier, its spectrum lies around one

of several fixed frequencies close to 930 MHz. The mixing takes place at three stages: first to 120 MHz by the first local oscillator, then to 30 MHz by the second local oscillator and finally complex mixing to zero frequency in the quadrature detector. In addition, the signal is filtered, attenuated and amplified at various places of its path. Finally the complex signal from the quadrature detector is digitised and the data samples are fed to the correlator which calculates the autocorrelation function.

The purpose of Chapter 3 is to explain the mathematical principles of the analogue signal path in the receiver. For this purpose we have to understand the meaning of negative frequency and we also have to define the concepts of signal spectrum and autocorrelation function more specifically than in the above description.

3.2 Signal and its Fourier transform

Let us consider a real-valued signal $x(t)$. Provided it is both integrable and absolutely integrable from $-\infty$ to $+\infty$ as well as quasi-differentiable, it can be presented in terms of its Fourier transform

$$X(\nu) = \mathcal{F}\{x(t)\} = \int_{-\infty}^{\infty} x(t)e^{-i\omega t}dt \tag{3.1}$$

in the form

$$x(t) = \mathcal{F}^{-1}\{X(\nu)\} = \int_{-\infty}^{\infty} X(\nu)e^{i\omega t}d\nu. \tag{3.2}$$

Here ν is the frequency and $\omega = 2\pi\nu$ the angular frequency. We say that $x(t)$ and $X(\nu)$ make a Fourier transform pair. When $x(t)$ is real,

$$X^*(\nu) = \int_{-\infty}^{\infty} x(t)e^{i\omega t}dt = X(-\nu). \tag{3.3}$$

In terms of the modulus $|X(\nu)|$ and phase angle $\varphi(\nu)$ this can be written as

$$|X(\nu)|e^{-i\varphi(\nu)} = |X(-\nu)|e^{i\varphi(-\nu)} \tag{3.4}$$

or

$$\begin{aligned} |X(-\nu)| &= |X(\nu)| \\ \varphi(-\nu) &= -\varphi(\nu). \end{aligned} \tag{3.5}$$

Hence the modulus of the Fourier transform of a real-valued signal is an even and the phase angle an odd function of frequency. This is visualised schematically in Fig. 3.2. When considering the frequency behaviour of $|X(-\nu)|$ and $\varphi(\nu)$, these quantities are often called the amplitude and phase spectrum, respectively.

In eq. (3.2) integration is carried out from minus to plus infinity, which implies the existence of negative frequencies. Here the simple concept of frequency

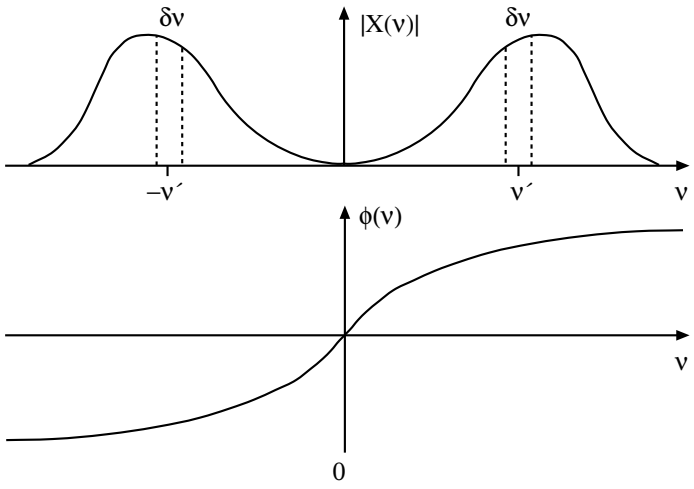

Figure 3.2: Modulus and phase angle of a real signal.

as the number of oscillations per time unit is generalised for mathematical convenience. The meaning of negative frequency can be enlightened by the following line of thought. Let us consider two frequency intervals around the frequencies ν' and $-\nu'$, both of them with a width $\delta\nu$ (Fig. 3.2). The part of $X(\nu)$ which lies within these frequency bands corresponds to a signal

$$\begin{aligned}
\delta x(t) &= X(\nu')e^{i\omega' t}\delta\nu + X(-\nu')e^{-i\omega' t}\delta\nu \\
&= \left\{|X(\nu')|e^{i[\omega' t + \varphi(\nu')]} + |X(-\nu')|e^{i[-\omega' t + \varphi(-\nu')]}\right\}\delta\nu \\
&= |X(\nu')|\left\{e^{i[\omega' t + \varphi(\nu')]} + e^{-i[\omega' t + \varphi(\nu')]}\right\}\delta\nu \\
&= 2|X(\nu')|\cos[\omega' t + \varphi(\nu')]\delta\nu. \qquad (3.6)
\end{aligned}$$

So we see that a sum of two complex signals at positive and negative frequencies ν' and $-\nu'$ creates a real signal $\delta x(t)$ at a positive frequency ν'. This is the frequency component ν' of the total signal $x(t)$. We also see that the positive and negative frequency signals together take care of the correct phase $\varphi(\nu')$ of the real signal. Obviously, the total signal $x(t)$ can be presented as a sum of narrow-band real signals $\delta x(t)$, *i.e.* by integrating $\delta x(t)/\delta\nu$ from zero frequency to infinity. This is mathematically equivalent to integrating the complex signal $X(\nu)\exp(i\omega t)$ from minus to plus infinity.

3.3 Signal power

Let us consider a real-valued physical signal $x(t)$ which differs from zero only for a limited time $[-T, T]$. The instantaneous power of the signal is $x^2(t)$ and

the mean power is

$$P = \frac{1}{2T}\int_{-\infty}^{\infty} x^2(t)dt. \tag{3.7}$$

We can now apply the Fourier transform $X(\nu)$ to obtain

$$\begin{aligned}\int_{-\infty}^{\infty} x^2(t)dt &= \int_{-\infty}^{\infty} x(t)\left[\int_{-\infty}^{\infty} X(\nu)e^{i\omega t}d\nu\right]dt \\ &= \int_{-\infty}^{\infty} X(\nu)\left[\int_{-\infty}^{\infty} x(t)e^{i\omega t}dt\right]d\nu \\ &= \int_{-\infty}^{\infty} X(\nu)X^*(\nu)d\nu \\ &= \int_{-\infty}^{\infty} |X(\nu)|^2 d\nu. \end{aligned} \tag{3.8}$$

In terms of the mean power,

$$P = \frac{1}{2T}\int_{-\infty}^{\infty} x^2(t)dt = \frac{1}{2T}\int_{-\infty}^{\infty} |X(\nu)|^2 d\nu. \tag{3.9}$$

Continuing the line of thought in Chapter 3.2 we can also calculate the mean power of the frequency component ν' in terms of the elementary signal $\delta x(t)$. If $x(t)$ is extended periodically beyond the range $[-T, T]$, it can be presented in terms of its Fourier series and the frequencies of its Fourier components are multiples of $1/2T$. Therefore it is reasonable to adopt $\delta\nu = 1/2T$. Then the average power of $\delta x(t)$ is

$$\langle \delta x^2(t)\rangle = 4|X(\nu')|^2 \langle \underbrace{\cos^2(\omega' t - \varphi(\nu'))}_{\frac{1}{2}}\rangle (\delta\nu)^2 = \frac{2|X(\nu')|^2}{2T}\delta\nu. \tag{3.10}$$

This indicates that $|X(\nu)|^2/2T$ is one half of the power density (*i.e.* power per frequency unit) at the frequency ν. Since $|X(\nu)|^2 = |X(-\nu)|^2$,

$$\langle \delta x^2(t)\rangle = \frac{1}{2T}[|X(\nu')|^2 + |X(-\nu')|^2]\delta\nu. \tag{3.11}$$

Hence one half of the power can be interpreted to come from positive and the other half from negative frequencies.

3.4 Autocorrelation function and power spectrum

As stated in Chapter 3.1, the incoherent scatter signal is a realisation of a stochastic process. Therefore the expectation value of any quantity depending on the signal — like the signal power, for instance — is determined by the

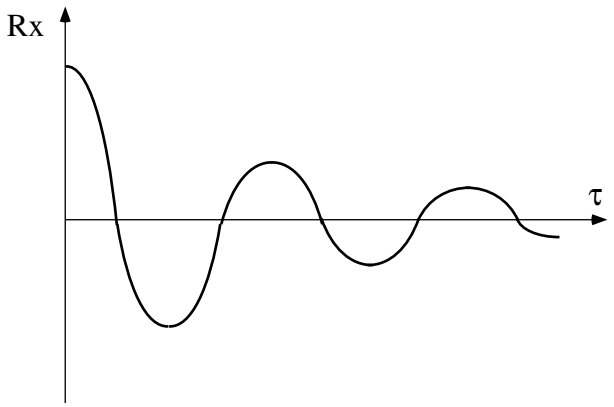

Figure 3.3: Autocorrelation function.

probability distributions associated with this stochastic process. An estimate of the expectation value can be obtained as an 'ensemble average' calculated from several realisations of the stochastic process, *i.e.* by observing the signal several times, calculating the quantity for each signal separately and taking the average of the results.

For a real-valued signal $x(t)$ we define the autocorrelation function as

$$R_x(t_1, t_2) = \langle x(t_1)x(t_2)\rangle, \tag{3.12}$$

where the angle brackets indicate the expectation value. If the signal is stationary, its statistical properties are time-independent. Then the autocorrelation function depends only on the difference of the arguments t_1 and t_2 and, instead of eq. 3.12, we can write

$$R_x(\tau) = \langle x(t)x(t-\tau)\rangle, \tag{3.13}$$

where the time difference τ is called delay or lag. In the case of so-called ergodic process (which is not defined here) the expectation value can be replaced by the time average of any realisation $x(t)$. Then

$$R_x(\tau) = \lim_{t_0\to\infty} \frac{1}{2t_0} \int_{-t_0}^{t_0} x(t)x(t-\tau)dt. \tag{3.14}$$

The autocorrelation function has its greatest value at $\tau = 0$. When τ increases, the autocorrelation function decreases towards zero, often with damped oscillations as shown schematically in Fig. 3.3. At delays longer than some value called the length of the autocorrelation function or correlation length, $R_x \approx 0$. The idea of autocorrelation function is to compare signal values at different time intervals. If R_x is positive, the signal values have often the same sign, if negative, opposite signs are usually expected. If $R_x = 0$, the relation of the signs is unpredictable. In addition to signs, the absolute values of the signals

also contribute to the autocorrelation function. All this means that, knowing the autocorrelation function, a measurement of a signal allows predictions to be made of the signal behaviour in the future. The uncertainty of this predictions grows with time distance. Nothing can be said about the signal values beyond the correlation length. In other words, signal values separated in time by more than the correlation length are practically independent of each other.

In practice, a signal can never be observed for an infinite time. It seems obvious that, by choosing a sufficiently long time interval $[-T, T]$, a good estimate for an autocorrelation function of a stationary ergodic signal $x(t)$ is given by the autocorrelation function of the signal

$$x_T(t) = \begin{cases} x(t) \text{ when } -T < t < T \\ 0 \text{ elsewhere.} \end{cases} \tag{3.15}$$

Hence

$$R_x(\tau) \approx \frac{1}{2T} \int_{-T}^{T} x_T(t) x_T(t-\tau) dt = \frac{1}{2T} \int_{-T}^{T} x(t) x(t-\tau) dt. \tag{3.16}$$

If the Fourier transform of x_T is $X(\nu)$, we see that

$$\begin{aligned} R_x(\tau) &= \frac{1}{2T} \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} X(\nu) e^{i\omega t} d\nu \right] \cdot x_T(t-\tau) dt \\ &= \frac{1}{2T} \int_{-\infty}^{\infty} X(\nu) e^{i\omega\tau} \underbrace{\left[\int_{-\infty}^{\infty} x_T(t-\tau) e^{i\omega(t-\tau)} d(t-\tau) \right]}_{X^*(\nu)} d\nu \\ &= \frac{1}{2T} \int_{-\infty}^{\infty} X(\nu) X^*(\nu) e^{i\omega\tau} d\nu \\ &= \frac{1}{2T} \int_{-\infty}^{\infty} |X(\nu)|^2 e^{i\omega\tau} d\nu \\ &= \int_{-\infty}^{\infty} S_x(\nu) e^{i\omega\tau} d\nu \\ &= \mathcal{F}^{-1}\{S_x(\nu)\}, \end{aligned} \tag{3.17}$$

where

$$S_x(\nu) = \frac{1}{2T} |X(\nu)|^2 \tag{3.18}$$

is called the power spectrum or, in more accurate terms, the power spectral density. Sometimes $|X(\nu)|^2$ is also called the power spectrum, although it is actually the energy spectral density.

Eq. (3.17) shows that the autocorrelation function and power spectrum of a signal make a Fourier transform pair

$$R_x(\tau) = \mathcal{F}^{-1}\{S_x(\nu)\}$$

$$S_x(\nu) = \mathcal{F}\{R_x(\tau)\}. \tag{3.19}$$

As explained in section 3.1, the signal received by an incoherent scatter radar is usually digitised by a AD converter. The task of the correlator is to calculate lag estimates of the autocorrelation function of the signal. The autocorrelation function contains the spectral information of the signal which is needed in determining the values of the plasma parameters in the ionosphere.

A comparison of eq. (3.18) with eq. (3.9) shows that the total mean power within the time interval $[-T, T]$ is

$$P = \int_{-\infty}^{\infty} S_x(\nu) d\nu, \tag{3.19}$$

i.e. the total power is obtained as a sum of powers from all frequencies. Because $S_x(\nu)$ is an even function of frequency, one half of the power comes from positive and the other half from negative frequencies. This result was already obtained in interpreting eq. (3.11).

3.5 Frequency mixing

The simplest way to understand frequency mixing is to consider a monochromatic signal

$$x(t) = \cos \omega t \tag{3.20}$$

If this is multiplied by $\cos \omega_0 t$, where $\omega_0 < \omega$, the result is

$$y(t) = \cos \omega t \cdot \cos \omega_0 t$$

$$= \frac{1}{2}[\cos(\omega + \omega_0)t + \cos(\omega - \omega_0)t]. \tag{3.21}$$

The amplitude spectra of these two signals are shown in Fig. 3.4. Obviously

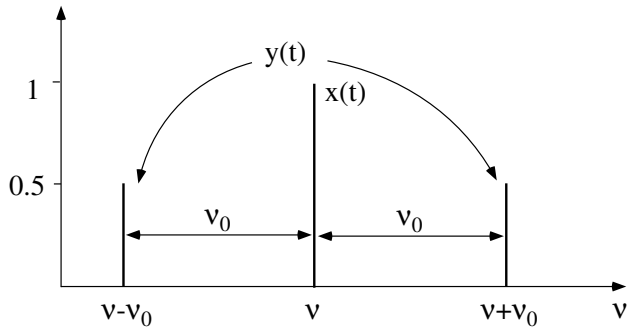

Figure 3.4: Frequency mixing of a monochromatic signal.

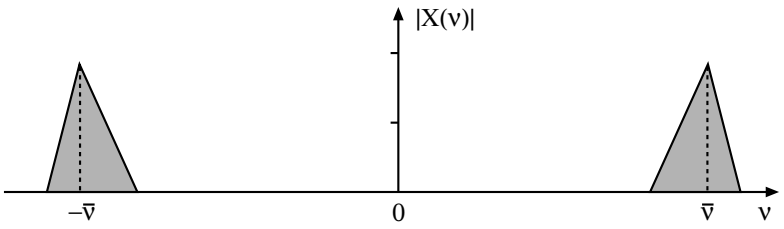

Figure 3.5: Amplitude spectrum of a narrow-band signal.

$y(t)$ consists of two monochromatic signals, one at a frequency $\nu - \nu_0$, the other at $\nu + \nu_0$ and both with amplitudes one half of the amplitude of $x(t)$. These components can be separated in the system hardware, the higher frequency by a high-pass and the lower one by a low-pass filter. The resulting signals are

$$y_u(t) = \frac{1}{2}\cos(\omega + \omega_0)t \tag{3.22}$$

and

$$y_l(t) = \frac{1}{2}\cos(\omega - \omega_0)t. \tag{3.23}$$

We say that $y_u(t)$ is mixed upwards and $y_l(t)$ downwards by the frequency ν_0. Hence the frequency of a monochromatic signal can be changed by multiplication with a different frequency and subsequent filtering.

The same principle can also be applied in a more general case. Let us consider a band-limited signal

$$x(t) = \int_{-\infty}^{\infty} X(\nu)e^{i\omega t}d\nu, \tag{3.24}$$

with its spectrum restricted to some frequency range around $\bar{\nu}$. This is illustrated by the amplitude spectrum in Fig. 3.5. When $x(t)$ is a real-valued signal, the amplitude spectrum is an even function of frequency as plotted in the figure. The phase spectrum is odd, of course.

If $x(t)$ is multiplied by $\cos\omega_0 t$, where $\omega_0 < \bar{\omega} = 2\pi\bar{\nu}$, we obtain

$$\begin{aligned} y(t) &= \cos\omega_0 t \cdot x(t) \\ &= \frac{e^{i\omega_0 t} + e^{-i\omega_0 t}}{2} \cdot \int_{-\infty}^{\infty} X(\nu)e^{i\omega t}d\nu \\ &= \frac{1}{2}\int_{-\infty}^{\infty} X(\nu)e^{i(\omega+\omega_0)t}d\nu + \frac{1}{2}\int_{-\infty}^{\infty} X(\nu)e^{i(\omega-\omega_0)t}d\nu \\ &= \frac{1}{2}\int_{-\infty}^{\infty} X(\nu-\nu_0)e^{i\omega t}d\nu + \frac{1}{2}\int_{-\infty}^{\infty} X(\nu+\nu_0)e^{i\omega t}d\nu \\ &= \int_{-\infty}^{\infty} Y(\nu)e^{i\omega t}d\nu, \end{aligned} \tag{3.25}$$

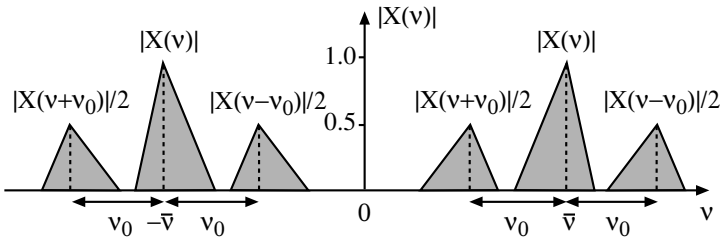

Figure 3.6: Frequency mixing of a narrow-band signal.

where

$$Y(\nu) = \mathcal{F}\{y(t)\} = \frac{1}{2}[X(\nu - \nu_0) + X(\nu + \nu_0)]. \tag{3.26}$$

Hence the spectrum of the resulting signal is composed of two parts. If the spectrum of the original signal $x(t)$ is narrow enough, it does not overlap with the spectrum of $y(t)$. Such a situation is depicted in Fig. 3.6.

In the same way as above, $y(t)$ can be divided into two components,

$$y_u(t) = \int_{-\infty}^{\infty} Y_u(\nu) e^{i\omega t} d\nu \tag{3.27}$$

and

$$y_l(t) = \int_{-\infty}^{\infty} Y_l(\nu) e^{i\omega t} d\nu, \tag{3.28}$$

where

$$Y_u(\nu) = \begin{cases} \frac{1}{2}X(\nu - \nu_0) \text{ , when } \nu > 0 \\ \frac{1}{2}X(\nu + \nu_0) \text{ , when } \nu < 0 \end{cases} \tag{3.29}$$

and

$$Y_l(\nu) = \begin{cases} \frac{1}{2}X(\nu + \nu_0) \text{ , when } \nu > 0 \\ \frac{1}{2}X(\nu - \nu_0) \text{ , when } \nu < 0 \end{cases} . \tag{3.30}$$

Via this definition the spectrum of $y_u(t)$ lies further away from the origin and the spectrum of $y_l(t)$ closer to the origin than the spectrum of $x(t)$. Then they can be separated by high- and low-pass filters, respectively.

In the EISCAT UHF radar the bandwidth of the scattered signal is of the order of 10 kHz. The applied radar modulation may broaden the spectrum to about 100 kHz. The spectrum lies around 930 MHz, so that the signal is indeed a narrow-band one. In the receiver the signal is first mixed down to about 120 MHz by the first local oscillator and then to 30 MHz by the second local oscillator. In both cases the upper sideband $Y_u(\nu)$ is removed by a low-pass filter. The frequency of the first local oscillator is fixed, but the second one is variable and its frequency is chosen to give a centre frequency of 30 MHz. A variable frequency is needed, because different transmitting frequencies around the nominal value of 930 MHz are used in the experiments.

3.6 Mixing to zero frequency

The above principle can be used in mixing the signal downwards as long as the spectrum does not extend to zero frequency. In order to mix the spectrum to zero, the method must be slightly modified. The idea is easily understood in terms of a monochromatic signal

$$x(t) = A\cos(\omega_0 t + \varphi). \tag{3.31}$$

When this is multiplied by the function

$$f(t) = e^{-i\omega_0 t} = \cos\omega_0 t - i\sin\omega_0 t \tag{3.32}$$

the result is

$$\begin{aligned} y(t) &= Ae^{-i\omega_0 t}\cdot\cos(\omega_0 t+\varphi) \\ &= Ae^{i\varphi}\cdot e^{-i(\omega_0 t+\varphi)}\cdot\ cos(\omega_0 t+\varphi) \\ &= Ae^{i\varphi}[\cos^2(\omega_0 t+\varphi) - i\sin(\omega_0 t+\varphi)\cdot\cos(\omega_0 t+\varphi)] \\ &= \frac{A}{2}e^{i\varphi}[1+\cos 2(\omega_0 t+\varphi) - i\sin 2(\omega_0 t+\varphi)]. \end{aligned} \tag{3.33}$$

Note carefully that the frequencies of $x(t)$ and $f(t)$ are the same. Thus the multiplication of $x(t)$ and $f(t)$ gives a complex function $y(t)$ which consists of an offset and an oscillation at a frequency $2\nu_0$. When this is filtered by a low-pass filter with a cut-off frequency smaller than $2\nu_0$, the output is the offset

$$z(t) = \frac{A}{2}e^{i\varphi} = \frac{A}{2}\cos\varphi + i\frac{A}{2}\sin\varphi \tag{3.34}$$

which contains the information on both the amplitude and phase of the original signal $x(t)$. They are given by

$$\begin{cases} A = 2\sqrt{z(t)z^*(t)} \\ \varphi = \arctan[\mathrm{Im}(z)/\mathrm{Re}(z)]. \end{cases} \tag{3.35}$$

In this way a monochromatic real-valued signal is mixed to zero frequency conserving the amplitude and phase information.

The same principle can be applied to a narrow-band signal. Assume that the spectrum of the signal

$$x(t) = \int_{-\infty}^{\infty} X(\nu)e^{i\omega t}d\nu \tag{3.36}$$

lies around ν_0 on the frequency axis. When $x(t)$ is multiplied by $\exp(-i\omega_0 t)$, the result is

$$\begin{aligned} y(t) = e^{-i\omega_0 t}x(t) &= e^{-i\omega_0 t}\int_{-\infty}^{\infty} X(\nu)e^{i\omega t}d\nu \\ &= \int_{-\infty}^{\infty} X(\nu)e^{-i(\omega-\omega_0)t}d\nu \end{aligned}$$

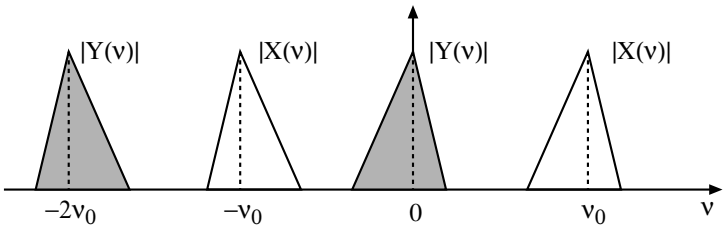

Figure 3.7: Mixing to zero frequency.

$$= \int_{-\infty}^{\infty} X(\nu + \nu_0) e^{i\omega t} d\nu$$
$$= \int_{-\infty}^{\infty} Y(\nu) e^{i\omega t} d\nu, \tag{3.37}$$

where

$$Y(\nu) = X(\nu + \nu_0). \tag{3.38}$$

This shows that complex mixing shifts the whole spectrum in negative direction by the amount ν_0. Then the upper spectral band lies around the zero frequency and the lower band around $-2\nu_0$ as indicated in Fig. 3.7. If the latter one is filtered out, all we have left is the spectrum around zero frequency. This can be put in the form

$$Z(\nu) = X_+(\nu + \nu_0), \tag{3.39}$$

where

$$X_+(\nu) = \begin{cases} X(\nu) \text{ , when } \nu > 0 \\ 0 \text{ , when } \nu < 0 \text{ .} \end{cases} \tag{3.40}$$

The corresponding signal is

$$z(t) = e^{-i\omega_0 t} \int_0^{\infty} X(\nu) e^{i\omega t} d\nu$$
$$= \int_{-\infty}^{\infty} X(\nu) e^{i(\omega - \omega_0)t} d\nu$$
$$= \int_{-\infty}^{\infty} X(\nu + \nu_0) e^{i\omega t} d\nu. \tag{3.41}$$

In the EISCAT UHF radar the spectrum is shifted to zero frequency by complex mixing in the quadrature detector, also known as the hybrid detector. After the second local oscillator the spectrum lies around 30 MHz. In the quadrature detector this signal is multiplied separately by two sinusoidal signals at 30 MHz, with a fixed phase difference of 90°. These are the real and imaginary parts of the function $\exp(-i\omega_0 t)$ in complex mixing. The output of the hybrid detector consists of two channels, one for the real part and the other for the imaginary part of the output. These together make the complex signal

$$y(t) = x(t)\cos\omega_0 t - ix(t)\sin\omega_0 t = e^{-i\omega_0 t} x(t). \tag{3.42}$$

When the real and imaginary parts are passed through low-pass filters, the higher frequencies are cut away, and the complex signal $z(t)$ is obtained. The total effect of mixing by the first and second local oscillators and complex mixing in the quadrature detector and the associated filterings is to move the narrow spectrum around 930 MHz to zero frequency. The quadrature detector and the subsequent filter (PDF or Post Detection Filter) are the last analogue components in EISCAT UHF receiver. After PFD the signal is fed to the AD-converter and post processed in digital form.

3.7 Hilbert transform

A second way of mixing a signal to zero frequency is provided by the Hilbert transform. The Hilbert transform of signal $x(t)$ is defined as

$$\hat{x}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} d\tau, \tag{3.43}$$

where it is understood that the integral means its principal value. Note that, unlike in the case of Fourier transform, the Hilbert transform of a time-domain signal remains in time domain.

The most convenient way of calculating the Fourier transform of a Hilbert-transformed signal is to use convolution which will be defined later. The result, however, can also be derived as follows. The Fourier transform of a Hilbert-transformed signal is

$$\mathcal{F}\{\hat{x}(t)\} = \int_{-\infty}^{\infty} \hat{x}(t) e^{-i\omega t} dt = \frac{1}{\pi} \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} d\tau \right] e^{-i\omega t} dt. \tag{3.44}$$

By changing the order of integration we obtain

$$\begin{aligned}
\mathcal{F}\{\hat{x}(t)\} &= \frac{1}{\pi} \int_{-\infty}^{\infty} x(\tau) \left[\int_{-\infty}^{\infty} \frac{e^{-i\omega(t-\tau)}}{t-\tau} d(t-\tau) \right] e^{-i\omega\tau} d\tau \\
&= \frac{1}{\pi} \left[\int_{-\infty}^{\infty} \frac{e^{-i\omega t'}}{t'} dt' \right] \left[\int_{-\infty}^{\infty} x(\tau) e^{-i\omega\tau} d\tau \right] \\
&= I \cdot \mathcal{F}\{x(\tau)\} = I \cdot X(\nu).
\end{aligned}$$

Here the integral I is

$$\begin{aligned}
I &= \frac{1}{\pi} \left[\int_{-\infty}^{\infty} \frac{\cos \omega t'}{t'} dt' - i \int_{-\infty}^{\infty} \frac{\sin \omega t'}{t'} dt' \right] \\
&= -\frac{2}{\pi} \cdot i \cdot \mathrm{sgn}(\nu) \int_0^{\infty} \frac{\sin \omega t'}{\omega t'} d\omega t' = -i \cdot \mathrm{sgn}(\nu),
\end{aligned}$$

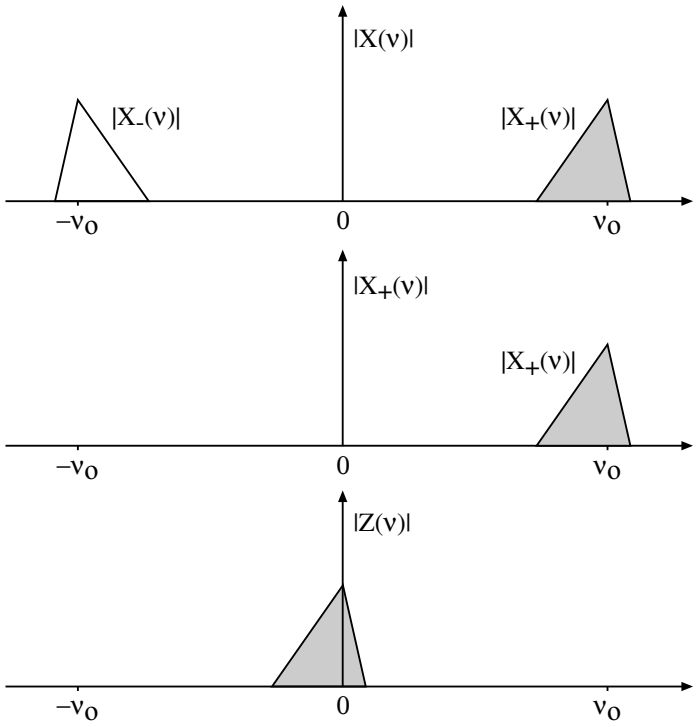

Figure 3.8: Mixing to zero frequency using Hilbert transform.

where the signum function is

$$\mathrm{sgn}(\nu) = \begin{cases} 1 \text{ , when } \nu > 0 \\ 0 \text{ , when } \nu = 0 \\ -1 \text{ , when } \nu < 0. \end{cases} \tag{3.45}$$

Combining these results gives

$$\mathcal{F}\{\hat{x}(t)\} = \hat{X}(\nu) = -i \cdot \mathrm{sgn}(\nu) \cdot X(\nu). \tag{3.46}$$

If $x(t)$ is a narrow-band signal with its spectrum around the frequency ν_0, its Fourier transform can be presented as a sum of the positive and negative frequency parts as shown by the top panel in Fig. 3.8. Thus

$$X(\nu) = X_+(\nu) + X_-(\nu) \tag{3.47}$$

and the Fourier transform of the Hilbert-transformed signal is

$$\hat{X}(\nu) = -i[X_+(\nu) - X_-(\nu)]. \tag{3.48}$$

The original signal and its Hilbert transform can be used in creating a complex signal

$$x_+(t) = [x(t) + i\hat{x}(t)]/2, \tag{3.49}$$

which has a Fourier transform

$$\mathcal{F}\{x_+(t)\} = [X(\nu) + i\hat{X}(\nu)]/2 = X_+(\nu). \tag{3.48}$$

This shows that the spectrum of $x_+(t)$ lies at positive frequencies only, as seen in the second panel of Fig. 3.8. The next step is to shift this spectrum to zero frequency. This can be done by complex mixing with $\exp(-i\omega_0 t)$. The result is

$$z(t) = x_+(t)e^{-i\omega_0 t}, \tag{3.49}$$

and the corresponding Fourier transform is

$$Z(\nu) = \int_{-\infty}^{\infty} z(t)e^{-i\omega t}dt = \int_{-\infty}^{\infty} x_+(t)e^{-i(\omega+\omega_0)t}dt = X_+(\nu+\nu_0). \tag{3.50}$$

Hence the spectrum of $z(t)$ is obtained by shifting the spectrum of $x_+(t)$ on the frequency axis in the negative direction by the amount ν_0. This is shown in the bottom panel of Fig. 3.8.

In conclusion, the above theory shows that Hilbert transform can be used in shifting the spectrum of a signal to zero frequency. The original signal and its Hilbert transform make the real and imaginary parts of a complex signal, which has a non-zero spectrum at positive frequencies only. The final shift to zero frequency is carried out by complex mixing. Note that in this procedure the negative frequency part of the spectrum is removed before mixing to zero frequency is carried out. This order is opposite to what happens in the EISCAT UHF receiver where a quadrature detector is used.

3.8 Autocorrelation function and power spectrum of a complex signal

The theory presented in the previous section makes it clear that, after the quadrature detector and post-detection filter, the incoherent scatter signal

$$z(t) = \int_{-\infty}^{\infty} Z(\omega)e^{i\omega t}d\nu \tag{3.51}$$

is complex-valued and its Fourier transform

$$Z(\omega) = \int_{-\infty}^{\infty} z(t)e^{-i\omega t}dt \tag{3.52}$$

departs from zero only within a narrow frequency range around the zero frequency. Because $z(t)$ is complex, $|Z(\omega)|$ is not necessarily even. Obviously,

$$|Z(\omega)|^2 = Z(\omega)Z^*(\omega) = \int_{-\infty}^{\infty} z(t)e^{-i\omega t}dt \cdot \int_{-\infty}^{\infty} z^*(t')e^{i\omega t'}dt'$$

$$= \int_{-\infty}^{\infty} z(t) \left[\int_{-\infty}^{\infty} z^*(t') e^{-i\omega(t-t')} dt' \right] dt.$$

Using the notation $\tau = t - t'$, so that $t' = t - \tau$ and $dt' = -d\tau$, the above equation can be developed as

$$|Z(\omega)|^2 = \int_{-\infty}^{\infty} z(t) \left[\int_{-\infty}^{\infty} z^*(t-\tau) e^{-i\omega\tau} d\tau \right] dt$$

$$= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} z(t) z^*(t-\tau) dt \right] e^{-i\omega\tau} d\tau. \tag{3.53}$$

This means that $|Z(\omega)|^2$ is a Fourier transform of a quantity closely resembling the autocorrelation function defined by eq. (3.14) for a real-valued stationary ergodic signal.

It seems reasonable to extend the definition of the autocorrelation function to include complex-valued signals. In analogy with eq. (3.16), an estimate for the autocorrelation function of a complex signal $z(t)$ is given by

$$R_z(\tau) = \langle z(t) z^*(t-\tau) \rangle \approx \frac{1}{2T} \int_{-\infty}^{\infty} z(t) z^*(t-\tau) dt, \tag{3.54}$$

when $z(t)$ is non-zero within the interval $[-T, T]$ only, and the corresponding power spectrum is

$$S_z(\nu) = \frac{1}{2T} |Z(\nu)|^2. \tag{3.55}$$

Then, according to eq. (3.53), the autocorrelation function and the power spectrum make a Fourier transform pair

$$S_z(\nu) = \mathcal{F}\{R_z(\tau)\}$$

$$R_z(\tau) = \mathcal{F}^{-1}\{S_z(\nu)\}. \tag{3.56}$$

Note that, in order to achieve this result, it was necessary to use the complex conjugate z^* in the definition of the autocorrelation function.

Because $Z(\omega) = X_+(\omega + \omega_0)$, $S_z(\nu)$ has the same shape as the power

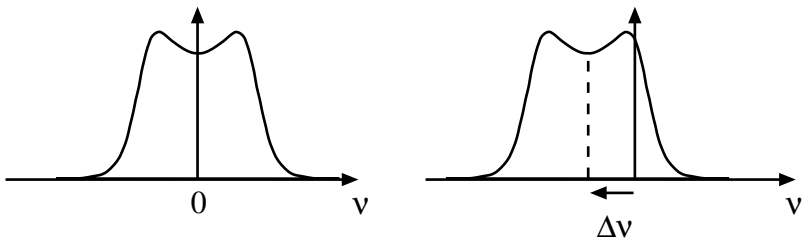

Figure 3.9: Doppler shift of the power spectrum.

spectrum of $x(t)$, but it is shifted to zero frequency. If the power spectrum of the incoherent scatter signal is symmetric around the transmission frequency, $S_z(\nu)$ is also symmetric around zero like in the left hand panel of Fig. 3.9. In that case $|Z(\nu)|$ is even and $\varphi_z(\nu)$ is odd function of frequency, so that $z(t)$ and $R_z(\tau)$ are actually real. Due to Doppler shift, however, the power spectrum of the scattering signal is usually asymmetric around the transmission frequency, and therefore $S_z(\nu)$ is also asymmetric around zero. This situation is portrayed in the right hand panel of Fig. 3.9. Then $|Z(\nu)|$ is not even and $z(t)$ and $R_z(\tau)$ are necessarily complex. Hence a non-zero imaginary part of the autocorrelation function is an indication of plasma motion with a non-zero velocity component along the radar beam.

3.9 Linear system

In the following, the word 'system' means a setup containing an input and an output, and giving an unequivocal output signal for each input signal. A real physical system, however, adds random noise in the output so that the output is not completely unequivocal.

A system is linear if it fulfils the following two conditions:

1. If the output corresponding to an input $x(t)$ is $y(t)$, then the input $a \cdot x(t)$, where a is a constant, leads to an output $a \cdot y(t)$.
2. If the outputs corresponding to inputs $x_1(t)$ and $x_2(t)$ are $y_1(t)$ and $y_2(t)$, respectively, the output corresponding to the input $x_1(t) + x_2(t)$ is $y_1(t) + y_2(t)$.

These definitions are clarified by Fig. 3.10.

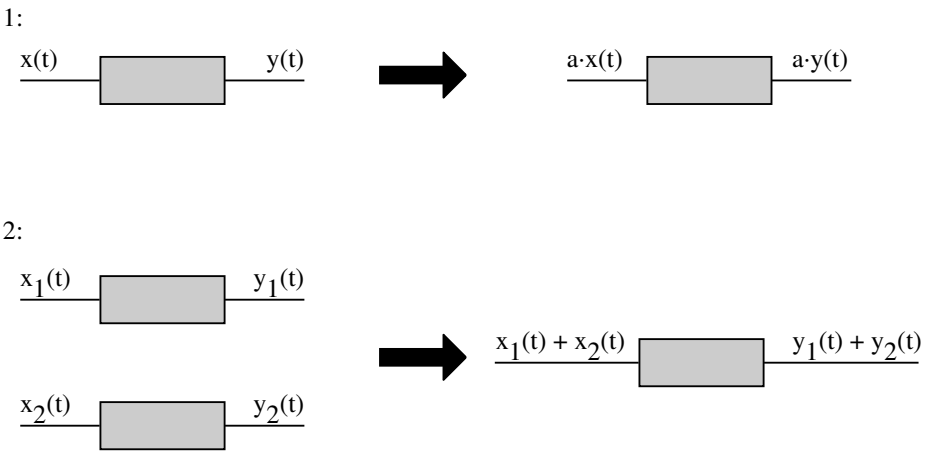

Figure 3.10: Properties of a linear system.

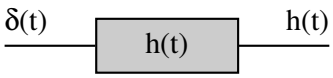

Figure 3.11: Impulse response of a linear system.

The incoherent scatter receiver is treated theoretically as a linear system. In order to fulfil this requirement, the amplifiers, attenuators and filters in different stages of the signal path should be as linear as possible. In EISCAT receivers the post-detection filter has a key role, because it affects the observed autocorrelation function and its properties must be known in the data analysis. The following text consists of the basic theory of a linear system, a knowledge which is essential in understanding the post-detection filter.

The characteristics of a linear system are ruled by its impulse response. When the impulse response is known, the output corresponding to any input can be calculated. The impulse response is defined as the output signal corresponding to an input signal equal to the Dirac delta function (Fig. 3.11).

Assume $x(t)$ is the input signal in a linear system with an impulse response $h(t)$. The task is to calculate the output signal $y(t)$. Using the definition of the delta function we can write

$$x(t) = \int_{-\infty}^{\infty} x(\tau)\delta(\tau - t)d\tau = \int_{-\infty}^{\infty} x(\tau)\delta(t - \tau)d\tau. \tag{3.57}$$

The latter equality is valid, because the delta function is even. This formula allows $x(t)$ to be interpreted as a sum of elementary signals

$$dx(t) = x(\tau)\delta(t - \tau)d\tau. \tag{3.58}$$

If the input signal is $\delta(t - \tau)$, the output is obviously $h(t - \tau)$. According to condition 2, the elementary output corresponding to an input dx is therefore

$$dy(t) = x(\tau)h(t - \tau)d\tau \tag{3.59}$$

and, according to condition 1, the total output is the sum of elementary outputs, *i.e.*

$$y(t) = \int_{-\infty}^{\infty} x(\tau)h(t - \tau)d\tau = (x * h)(t). \tag{3.60}$$

The type of integral in eq. (3.60) is known as convolution, which is often briefly marked using the notation '$*$'. Thus we have arrived at an important result: the output of a linear system is a convolution of the input and the impulse response of the system.

It is convenient to derive here some of the most important characteristics of the convolution.

1. Convolution is commutative: $(u * v)(t) = (v * u)(t)$.

   Proof:

   $$(u * v)(t) = \int_{-\infty}^{\infty} u(\tau)v(t - \tau)d\tau$$

$$= \int_{-\infty}^{\infty} u(t-\tau')v(\tau')d\tau' = (v * u)(t).$$

2. Fourier transform of convolution:
   $\mathcal{F}\{u * v\} = \mathcal{F}\{u\} \cdot \mathcal{F}\{v\}$ and $\mathcal{F}^{-1}\{U * V\} = \mathcal{F}^{-1}\{U\} \cdot \mathcal{F}\{V\}$.
   Proof:

$$\mathcal{F}\{u * v\} = \int_{-\infty}^{\infty} (u * v)e^{-i\omega t}dt = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} u(\tau)v(t-\tau)d\tau\right] e^{-i\omega t}dt$$

$$= \int_{-\infty}^{\infty} u(\tau) \left[\int_{-\infty}^{\infty} v(t-\tau)e^{-i\omega t}dt\right] d\tau$$

$$= \int_{-\infty}^{\infty} u(\tau)e^{-i\omega\tau} \left[\int_{-\infty}^{\infty} v(t-\tau)e^{-i\omega(t-\tau)}dt\right] d\tau$$

$$= \left[\int_{-\infty}^{\infty} u(\tau)e^{-i\omega\tau}d\tau\right] \left[\int_{-\infty}^{\infty} v(t')e^{-i\omega t'}dt'\right] = \mathcal{F}\{u\} \cdot \mathcal{F}\{v\}.$$

   The proof for the inverse Fourier transform is analogous.

3. Fourier transform of product:
   $\mathcal{F}\{u \cdot v\} = \mathcal{F}\{u\} * \mathcal{F}\{v\}$ and $\mathcal{F}^{-1}\{U \cdot V\} = \mathcal{F}^{-1}\{U\} * \mathcal{F}\{V\}$.
   Proof: If $U = \mathcal{F}\{u\}$ and $V = \mathcal{F}\{v\}$, obviously

$$\mathcal{F}^{-1}\{U * V\} = \mathcal{F}^{-1}\{U\} \cdot \mathcal{F}^{-1}\{V\} = u \cdot v,$$

   so that

$$\mathcal{F}\{u \cdot v\} = \mathcal{F}\{\mathcal{F}^{-1}\{U * V\}\}$$
$$= U * V = \mathcal{F}\{u\} * \mathcal{F}\{v\}.$$

   Again, the proof for the inverse transform is analogous.

The properties of the convolution integral give us a possibility to relate the Fourier transforms of the input and output signals. By applying property 2 to $y(t) = (x * h)(t)$ we immediately obtain

$$Y(\nu) = \mathcal{F}\{y\} = \mathcal{F}\{x\} \cdot \mathcal{F}\{h\} = X(\nu) \cdot H(\nu), \tag{3.61}$$

where

$$H(\nu) = \mathcal{F}\{h\} \tag{3.62}$$

is called the transfer function of the linear system. Hence the transfer function and the impulse response make a Fourier transform pair. Because

$$H(\nu) = |H(\nu)|e^{i\varphi_h(\nu)}, \tag{3.63}$$

the output signal can be put in the form

$$y(t) = \int_{-\infty}^{\infty} Y(\nu)e^{i\omega t}d\nu = \int_{-\infty}^{\infty} H(\nu)X(\nu)e^{i\omega t}d\nu$$

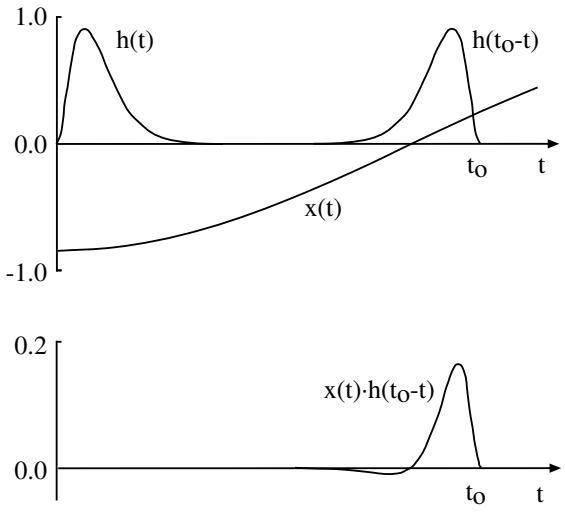

Figure 3.12: Convolution of impulse response and input signal.

$$= \int_{-\infty}^{\infty} |H(\nu)||X(\nu)| e^{i[\omega t + \varphi_x(\nu) + \varphi_h(\nu)]} d\nu. \tag{3.64}$$

This indicates how the amplitude and phase of the frequency component ν are changed by a linear system. The amplitude is multiplied by $|H(\nu)|$ and the phase is shifted by $\varphi_h(\nu)$.

The value of the convolution $(x * h)(t_0)$ is a weighted average of x from a time interval when $h(t_0 - t) \neq 0$. If t_0 indicates the present time and $h(t) = 0$, when $t < 0$, the output signal $y(t_0)$ is a weighted average of the input signal in the past. The idea is visualised in Fig. 3.12, where the top panel shows both $h(t)$, $h(t_0 - t)$ and $x(t)$, and the bottom panel the product $x(t) \cdot h(t_0 - t)$. The value of the output signal, $y(t_0)$, is the integral of the function plotted in the bottom panel. This sort of linear system remembers the past input signal from a time-span of its impulse response. The property is valid for all real physical filters.

If the impulse response is non-zero on the negative time axis, the output contains information also on the future values of the input signal. A linear system working in real time is not capable of making predictions, and therefore its impulse response is non-zero on the positive time axis only. A predicting digital filter, however, can be used for a signal which has already been observed. A simple example of such a filter is calculating a floating average.

In the EISCAT systems the complex signal from the quadrature detector is filtered by a low-pass post-detection filter. The quadrature components (real and imaginary parts) have filters of their own which have been built as identical as possible. The impulse responses of these filters deform the signal and their transfer functions deform the power spectrum. If the signal before the post-

detection filter is $z(t)$, the filter output is

$$z_h(t) = (z * h)(t) = [\mathrm{Re}(z) * h](t) + i[\mathrm{Im}(z) * h](t), \tag{3.65}$$

its Fourier transform is

$$Z_h(\nu) = \mathcal{F}\{z_h\} = \mathcal{F}\{\mathrm{Re}(z)\} \cdot H(\nu) + i\mathcal{F}\{\mathrm{Im}(z)\} \cdot H(\nu)$$

$$= H(\nu) \cdot \mathcal{F}\{\mathrm{Re}(z) + i\mathrm{Im}(z)\} = H(\nu) \cdot \mathcal{F}\{z\} = H(\nu) \cdot Z(\nu), \tag{3.66}$$

and its power spectrum is

$$S_{zh}(\nu) = \frac{1}{2T}|Z_h(\nu)|^2 = \frac{1}{2T}|H(\nu)|^2 \cdot |Z(\nu)|^2$$

$$= \frac{1}{2T}|H(\nu)|^2 \cdot |X_+(\nu + \nu_0)|^2. \tag{3.67}$$

This would seem to indicate that the spectrum of plasma fluctuations is obtained by dividing the observed power spectrum by $|H(\nu)|^2$. This is not the case, however, since the modulation of the transmitted signal also affects the signal spectrum. A unified formalism for taking into account both effects is offered by the theory of ambiguity functions, which is presented in Chapter 4. The ambiguity functions are closely associated with the altitude and lag ranges covered by the observed autocorrelation function estimates.

3.10 Transfer function of a boxcar filter. Definition of pass band.

As an example of a linear system, a filter with a boxcar-shaped impulse response is studied in this section. This is an idealised model of course, because the impulse response of a real analogue filter cannot have sharp edges. This filter, however, has the benefit that it allows easy approximate calculations in the ambiguity function theory and, for qualitative considerations, the results are not too far from those obtained using the true impulse responses. A filter with a boxcar-shaped impulse response represents one extreme case, the opposite extreme would be a filter with a boxcar-shaped transfer function, which is also sometimes called a boxcar filter.

Because the input and output signals of a linear system normally have the same unit, the dimension of the impulse response must be time^{-1}. It is practical to scale the impulse response to give

$$\int_{-\infty}^{\infty} h(t)dt = 1. \tag{3.68}$$

Since $H(\nu)$ is the Fourier transform of $h(t)$, this choice also means that $H(0) = 1$. This scaling is not useful if the amplification or damping of the system is of interest, since it pays attention only to the shape of the power spectrum,

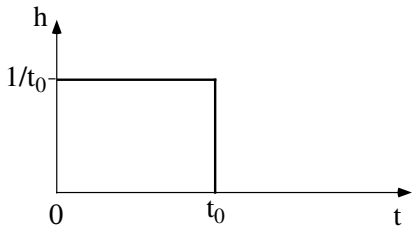

Figure 3.13: Impulse response of a boxcar filter.

not to its magnitude. In an incoherent scatter receiver, however, the signal is attenuated and amplified in various stages of its path, and a separate calibration of the output power is always necessary. Therefore the absolute scaling of the post-detection filter is not of interest to us, and it is convenient to use eq. (3.68).

With the above scaling, the impulse response of a boxcar filter is (Fig. 3.13)

$$h(t) = \begin{cases} 1/t_0 \text{ , when } 0 < t < t_0 \\ 0 \text{ , } \quad \text{when } t \leq 0 \text{ or } t \geq t_0 \text{ ,} \end{cases} \tag{3.69}$$

and the transfer function is

$$H(\nu) = \int_{-\infty}^{\infty} h(t) e^{-i\omega t} dt = \int_0^{t_0} \frac{1}{t_0} \cdot e^{-i\omega t} dt$$

$$= \frac{1}{t_0} \cdot e^{-i\omega t_0/2} \cdot \int_0^{t_0} e^{-i\omega(t - t_0/2)} dt = \frac{1}{t_0} \cdot e^{-i\omega t_0/2} \cdot \int_{-t_0/2}^{t_0/2} e^{-i\omega t} dt$$

$$= \frac{e^{-i\omega t_0/2}}{-i\omega t_0} \left(e^{-i\omega t_0/2} - e^{i\omega t_0/2} \right) = \frac{e^{-i\omega t_0/2}}{\omega t_0/2} \cdot \frac{e^{i\omega t_0/2} - e^{-i\omega t_0/2}}{2i}$$

$$= \frac{\sin \omega t_0/2}{\omega t_0/2} \cdot e^{-i\omega t_0/2} = \mathrm{sinc}(\omega t_0/2) \cdot e^{-i\omega t_0/2}. \tag{3.70}$$

It is a common practice to use the definition $\mathrm{sinc}(x) = \sin x/x$ in signal theory, although this sinc-function is nothing but the spherical Bessel function $\mathrm{j}_0(x)$.

The modulus of the transfer function is

$$|H(\nu)| = \left| \frac{\sin \omega t_0/2}{\omega t_0/2} \right| \tag{3.71}$$

and the phase

$$\varphi_h(\nu) = -\pi t_0 \nu + n\pi \mathrm{sgn}(\nu), \tag{3.72}$$

when $n/t_0 < |\nu| < (n+1)/t_0$, $n = 0, 1, 2, \ldots$. These quantities are plotted in Fig. 3.14.We see that the modulus has a strong centre peak (pass band) around zero and a set of weaker 'sidebands' separated by zero points at frequencies which are multiples of $1/t_0$. The heights of the first sidebands are about 0.22 and the other sidebands further away from the central peak are even weaker.

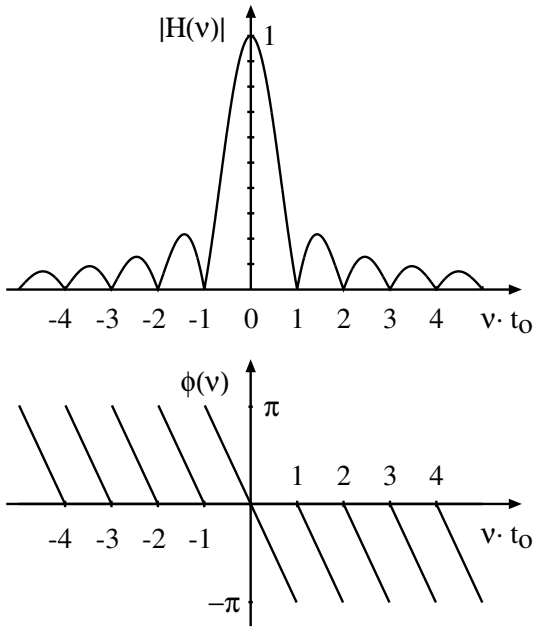

Figure 3.14: Modulus and phase of the transfer function of a boxcar filter.

This means that the system greatly attenuates the high frequencies in the signal so that it is indeed a low-pass filter, but it is not capable of completely removing all frequencies higher than some cut-off value. Hence the filter has no sharply limited pass band. The phase is discontinuous at points $\nu = \pm 1/t_0, \pm 2/t_0, \ldots$ and descends linearly with frequency within each interval $n/t_0 < |\nu| < (n+1)/t_0$. Because the impulse response is real, the modulus and phase should be even and odd functions of frequency, respectively. This condition is clearly fulfilled by the results shown in Fig. 3.14.

It is convenient to define a general concept of bandwidth which serves the practical need of comparing the pass bands of different filters. Such a definition can be given in terms of a boxcar-shaped transfer function as follows. The bandwidth of a low-pass filter is the cut-off frequency of a boxcar transfer function, which lets the same energy of white noise pass the filter as the original filter does. If ν_c is the cut-off frequency, the modulus of the boxcar transfer function is

$$|H_c(\nu)| = \begin{cases} 1 \text{ , when } -\nu_c < \nu < \nu_c \\ 0 \text{ , when } \nu < -\nu_c \text{ or } \nu > \nu_c. \end{cases} \tag{3.73}$$

Then

$$\int_{-\infty}^{\infty} |H_c(\nu)|^2 d\nu = \int_{-\nu_c}^{\nu_c} 1^2 d\nu = 2\nu_c$$
$$= \int_{-\infty}^{\infty} |H(\nu)|^2 d\nu = \int_0^{\infty} h(t)^2 dt. \tag{3.74}$$

This gives a cut-off frequency

$$\nu_c = \frac{1}{2}\int_0^\infty h(t)^2 dt. \tag{3.75}$$

In the case of a boxcar-shaped impulse response

$$\nu_c = \frac{1}{2t_0^2}\int_0^{t_0} dt = \frac{1}{2t_0}. \tag{3.76}$$

We can now calculate the ratio of output and input power at the cut-off frequency. When $\nu = \nu_c$, the value of $\omega_c t_0/2$ is $2\pi\nu_c t_0/2 = \pi/2$ and

$$|H(\nu_c)|^2 = \left|\frac{\sin \pi/2}{\pi/2}\right|^2 = \left(\frac{2}{\pi}\right)^2 \approx 0.6366^2 \approx 0.40 \approx -3.9 \text{ dB} .$$

Thus, if we assume that white noise is passed through the filter, the power at the cut-off frequency is reduced to 40 % of its value at zero frequency.

3.11 Bandwidth of an incoherent scatter signal

The post-detection filters which are used in the EISCAT receivers are electronic devices which, of course, do not have a boxcar shape. Different types of them, for instance so called Butterworth filters, are available. Regardless of the type, the above definition of the cut-off frequency gives a good idea of the filter pass band, and t_0 solved from the relation $\nu_c = 1/2t_0$ gives an estimate of the length of the impulse response. Examples of cut-off frequencies and corresponding lengths of impulse responses used in the EISCAT UHF radar are shown in Table 3.1.

Table 3.1
Typical cut-off frequencies and corresponding lengths of impulse response in the EISCAT UHF receiver.

ν_c (kHz)	t_0 (μs)
25	20
50	10
125	4

The first demand in choosing the post detection filter is that the whole ion line must be able to pass it with no severe attenuation even at the edges. The width of the ion line can be estimated as follows. According to eq. (2.50), the phase speed of an ion acoustic wave is approximately

$$v_+ = \sqrt{\frac{2kT_i}{m_i}} \tag{3.77}$$

and the corresponding Doppler shift is (see eq. 2.46)

$$\nu_+ = \frac{2v_+}{\lambda_0} = \frac{2\nu_0}{c}\sqrt{\frac{2kT_i}{m_i}}. \tag{3.78}$$

Table 3.2
Doppler shifts corresponding to the ion acoustic speeds for different ion masses and temperatures.

	ν_+ (kHz)		
m_i(u)	500	1000	2000
16	4.5	6.3	8.9
30	3.3	4.6	6.5

Table 3.2 shows the Doppler shifts calculated from this formula for ion masses 16 and 30 u (corresponding to ion species O^+ and NO^+) at different ion temperatures. Due to Landau damping, the ion line is somewhat broader than indicated by ν_+ but, nevertheless, it would look like $\nu_c = 10$ kHz would be a sufficient cut off frequency for the post detection filter. However, we also have to pay attention to the Doppler shift due to ionospheric plasma motion, which shifts the centre of the spectrum away from zero frequency. The greatest plasma velocities are encountered in the ionospheric F-layer and they are caused by electric fields. Normally the electric field E is smaller than 100 mV/m. The corresponding plasma drift speed in the geomagnetic field B is

$$v = \frac{E}{B} = \frac{100 \text{ mV/m}}{50\ \mu\text{T}} = 2000 \text{ m/s}$$

which causes a Doppler shift

$$\delta\nu = 2\frac{v}{\lambda_0} = \frac{2\nu_0 v}{c} = 12.4 \text{ kHz}.$$

Hence the width of the ion line and the Doppler shift together would give a bandwidth of about $(13 + 9)$ kHz = 22 kHz (Fig. 3.15). This is the explanation for the 25 kHz bandwidth in Table 3.1.

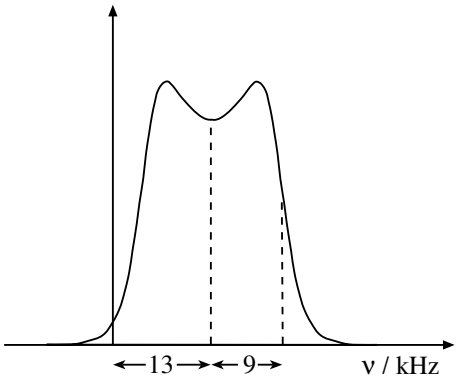

Figure 3.15: Increase of signal bandwidth due to Doppler shift.

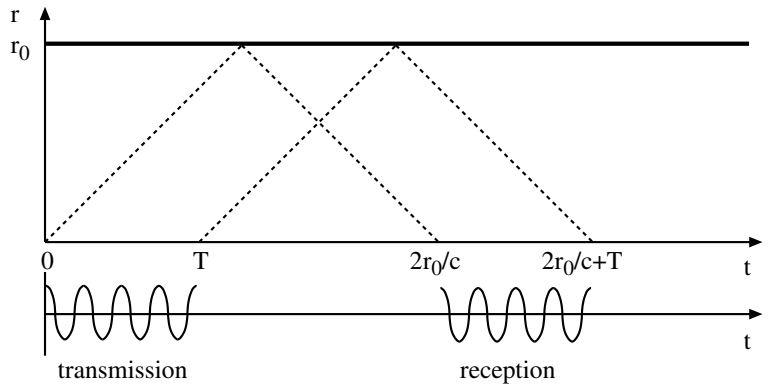

Figure 3.16: Reflection of a signal at a horizontal thin sheet.

The next task is to understand why filter bandwidths broader than 25 kHz are sometimes used. The explanation is that the modulation of the transmitted signal has a bandwidth of its own, and this modulation must also pass through the post detection filter.

Let us assume that a monostatic radar transmits a pulse with a length T at a frequency ν_0 and the pulse is reflected back by a horizontal surface at an altitude r_0 in the ionosphere (Fig. 3.16). The transmitted signal can be written as

$$x_t(t) = \text{env(t)} \cdot \cos(\omega_0 t), \tag{3.79}$$

where the modulation envelope is

$$\text{env(t)} = \begin{cases} 1 \text{ , when } 0 < t < T \\ 0 \text{ , when } t \leq 0 \text{ or } t \geq T. \end{cases} \tag{3.80}$$

This signal is shown on the left hand side in the bottom panel of Fig. 3.16, where the horizontal axis is time. In the top panel the vertical axis indicates range from the transmitter. The two tilted lines starting from the origin and $t = T$ indicate, respectively, the paths of the front and back ends of the transmitted pulse in the time-range coordinate system. The front end of the pulse hits the surface at a time $t = r_0/c$ and returns back to the transmitter at $t = 2r_0/c$. For the back end, the same incidents are delayed by T. If no phase shifts take place at the reflection point, the received signal is

$$x(t) = \text{env}(t - 2r_0/c) \cdot \cos\omega_0(t - 2r_0/c), \tag{3.81}$$

where the inevitable change in amplitude is neglected for simplicity.

If the moment when the front end of the echo reaches the receiver is chosen as the zero time, the received signal is $x(t) = \text{env}(t) \cdot \cos\omega_0 t$ and its Fourier transform is

$$X(\nu) = \int_{-\infty}^{\infty} \text{env}(t) cos(\omega_0 t) e^{-i\omega t} dt = \int_0^T \frac{e^{i\omega_0 t} + e^{-i\omega_0 t}}{2} \cdot e^{-i\omega t} dt$$

$$= \frac{1}{2}\left[\int_0^T e^{-i(\omega-\omega_0)t}dt + \int_0^T e^{-i(\omega+\omega_0)t}dt\right]. \tag{3.82}$$

Because

$$\int_0^T e^{-i\omega t}dt = e^{-i\omega T/2}\int_0^T e^{-i\omega(t-T/2)}dt = e^{-i\omega T/2}\int_{-T/2}^{T/2} e^{-i\omega t}dt$$

$$= e^{-i\omega T/2}\cdot\frac{e^{-i\omega T/2}-e^{i\omega T/2}}{-i\omega} = T\cdot\frac{e^{-i\omega T/2}}{\omega T/2}\cdot\frac{e^{i\omega T/2}-e^{-i\omega T/2}}{2i}$$

$$= T\cdot e^{-i\omega T/2}\cdot\frac{\sin(\omega T/2)}{\omega T/2},$$

we can write

$$X(\nu) = X_+(\nu) + X_-(\nu), \tag{3.83}$$

where

$$X_+(\nu) = \frac{T}{2}e^{-i(\omega-\omega_0)T/2}\cdot\frac{\sin(\omega-\omega_0)T/2}{(\omega-\omega_0)T/2}$$

$$X_-(\nu) = \frac{T}{2}e^{-i(\omega+\omega_0)T/2}\cdot\frac{\sin(\omega+\omega_0)T/2}{(\omega+\omega_0)T/2}. \tag{3.84}$$

If $T \gg 1/\nu_0$, then $X_+(\nu)$ departs from zero only in the neighbourhood of ν_0 (*i.e.* within the region $|\nu-\nu_0| \ll \nu_0$), and $X_-(\nu) \neq 0$ only in the neighbourhood of $-\nu_0$. Then, with a good accuracy, $X_+(\nu)$ represents the positive and $X_-(\nu)$ the negative frequency part of the Fourier transform. Thus the Fourier transform of the quadrature detector output is

$$Z(\nu) \approx X_+(\nu+\nu_0) = \frac{T}{2}e^{-i\omega T/2}\cdot\frac{\sin\omega T/2}{\omega T/2}. \tag{3.85}$$

This can obviously be put in the form

$$Z(\nu) = \frac{1}{2}\int_0^T e^{-i\omega t}dt = \frac{1}{2}\int_{-\infty}^{\infty}\mathrm{env}(t)e^{-i\omega t}dt = \frac{1}{2}\mathcal{F}\{\mathrm{env}(t)\} \tag{3.86}$$

so that

$$z(t) = \mathcal{F}^{-1}\{Z(\omega)\} = \frac{1}{2}\mathrm{env}(t). \tag{3.87}$$

This shows that, if the transmitted signal is modulated by a simple pulse, the output signal of the quadrature detector consists of a similar pulse. After the post detection filter the signal is

$$z_h(t) = (z*h)(t) = \frac{1}{2}(\mathrm{env}*h)(t). \tag{3.88}$$

The shape of this function for different relative magnitudes of the pulse length and impulse response length is shown in Fig. 3.17 in the case of a boxcar filter. The pulse is shown in panel *a*. In panels *b* – *d* three different impulse responses

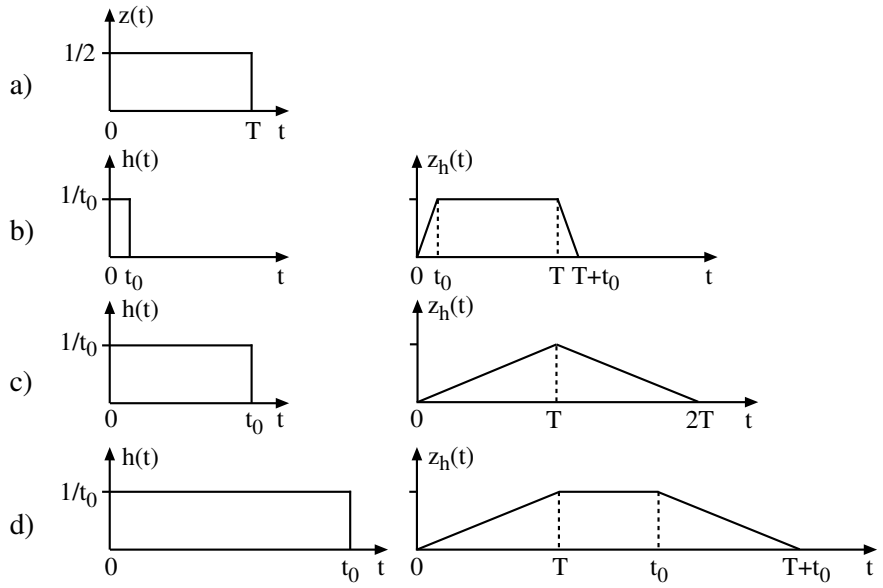

Figure 3.17: Deformation of a simple pulse by three different impulse responses.

and results of convolution are plotted. In each case the result is a pulse with sloped edges and a total length $T + t_0$, where t_0 is the length of the impulse response. If $t_0 < T$ (panel *b*), the pulse is not seriously deformed. If $t_0 = T$ (panel *c*), the output of the post detection filter is a triangular pulse. Finally, if $t_0 > T$ (panel *d*), the output signal is closer to the impulse response than the detected pulse. Because the filter bandwidth is

$$\nu_c = \frac{1}{2t_0} \tag{3.89}$$

and the bandwidth of the input signal is

$$\nu_z = \frac{1}{2T}, \tag{3.90}$$

this means in frequency domain that the signal can pass the post detection filter without too serious deformation only if ν_c is greater than ν_z.

If the reflecting surface in the ionosphere is replaced by a thin scattering layer, the scattering signal from this layer in the receiver is

$$x(t) = \mathrm{env}(t) \cdot y(t), \tag{3.91}$$

where $y(t)$ is a stochastic signal, which would be caused by a continuous illumination $\cos\omega_0 t$. Note that we discuss here only the signal due to Thomson scattering within the sheet, not the reflections at the sharp edges. This is possible since the real ionosphere can be built of a great number of such layers. Then

no sharp edges are present and only the scattering signal is observed from each sheet.

The Fourier transform of $x(t)$ is

$$X(\nu) = \mathcal{F}\{\text{env}(t)\} * \mathcal{F}\{y(t)\} = \text{ENV}(\nu) * Y(\nu), \tag{3.92}$$

where $\text{ENV}(\nu)$ is the Fourier transform of $\text{env}(t)$. If $\text{env}(t)$ is a simple pulse with a length T, the modulus of its Fourier transform is

$$|\text{ENV}(\nu)| = |\sin(\omega T/2)/(\omega T/2)|$$

and its bandwidth is $\nu_{env} = 1/2T$. From the convolution in eq. (3.92) it is obvious that the bandwidth of $x(t)$ is

$$\nu_x = \nu_{env} + \nu_y, \tag{3.93}$$

where ν_y is the bandwidth of $y(t)$.

The above consideration shows that there are two factors which affect the bandwidth of an incoherent scatter signal. One is the scattering process itself and the other the modulation of the transmitted wave. In fact, both factors have a bandwidth of their own and the total bandwidth is their sum. The post detection filter should be chosen to let the frequencies within the total bandwidth of the signal pass the filter without too much distortion but, at the same time, to reject unnecessary noise. Table 3.3. shows the approximate pulse lengths in different modulations used by the EISCAT UHF radar, together with the corresponding bandwidths of the modulation envelope. According to these estimates, the modulation bandwidth of the phase code is so wide that it alone determines the bandwidth of the post detection filter. On the other hand, the bandwidths of the longest single pulses are so narrow that they have no essential role in determining the filter band width.

Table 3.3.
Approximate pulse lengths and bandwidths of different modulations used in the EISCAT UHF radar.

T (μs)	ν (kHz)	
2 – 4	250 – 125	Phase code
10	50	Pulse code
100 – 300	5 – 1.7	Long pulse

As pointed out above, a real ionosphere can be considered to be composed of a set of adjacent thin layers. In the receiver, the scattering signal from the ith layer has the form

$$x_i(t) = \text{env}(t - 2r_i/c) \cdot y_i(t), \tag{3.94}$$

where r_i is the distance to the ith layer and $y_i(t)$ is the scattered signal from the ith layer caused by a continuous monochromatic illumination. If the transmitted

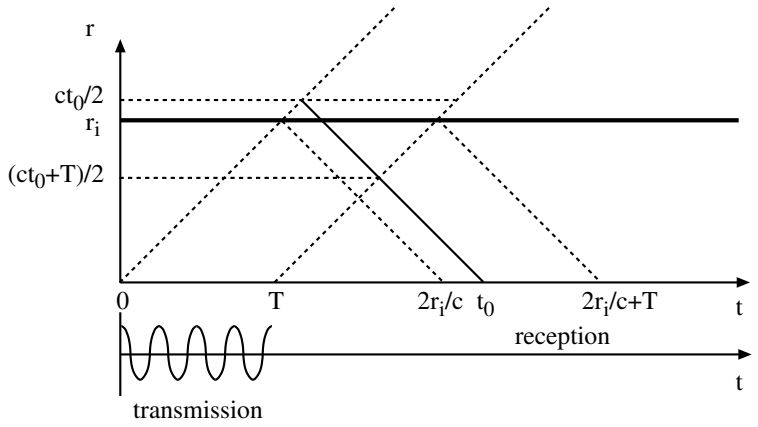

Figure 3.18: The height range of the observed signal.

pulse is modulated by a simple pulse with a length T, the scattered signal from the ith layer will be observed at the receiver of a monostatic radar during a time interval $(2r_i/c, 2r_i/c + T)$. This is shown by Fig. 3.18. On the other hand, the total signal observed at $t = t_0$ is the sum of signals scattered from all elementary layers within the height range $[c(t_0 - T)/2, ct_0/2]$. Then the above discussion on the signal bandwidth applies to each of these elementary signals. This means that the post detection filter at each channel must have a bandwidth which allows the widest possible signal from the range of interest to pass through. Even in this case the signal is deformed to some extent, which must be taken into account in the data analysis. How this is done will be explained by the theory of ambiguity functions in Chapter 4.

Chapter 4

Theory of ambiguity functions

4.1 Two-dimensional ambiguity functions

As shown previously, the incoherent scatter signal is deformed by the impulse response of the receiver, and therefore the impulse response also affects the signal autocorrelation function. A second factor which deforms the signal and its autocorrelation function is the radar modulation since it determines the scattering altitudes of the elementary signals contributing to the total signal at a given instant of time. When plasma parameters are estimated from the autocorrelation functions, both of these effects have to be taken into account. This can be done in a controlled way by means of the theory of ambiguity functions.

Let us consider a volume element d^3r at $\mathbf{r}$ with distances $r_1(\mathbf{r})$ and $r_2(\mathbf{r})$ to the transmitter and the receiver, respectively. This volume is illuminated by continuous monochromatic radiation. If there were only a single electron in the volume element, the scattered power captured by the receiver antenna would be given by eq. (2.81). The power P_z^o observed after the quadrature detector would be proportional to this power. Note that this power is inversely proportional to the squares of the distances to the two antennas and it is also controlled by the antenna gains.

When the volume element is filled with thermal plasma, a complex random voltage $dz(t, \mathbf{r})$ is observed at the output of the quadrature detector. The mean value of the square of this voltage, $\langle dz(t, \mathbf{r}) dz^*(t, \mathbf{r}) \rangle$, is proportional to the observed mean power and the respective coefficient R can be called the receiver resistance. The mean power, on the other hand, is proportional to the the number of electrons within the volume element as well as to P_z^o, the power due to a single electron. According to the definition of the autocorrelation function in eq. (3.54), $\langle dz(t, \mathbf{r}) dz^*(t, \mathbf{r}) \rangle$ is the zero lag of the signal autocorrelation function.

Then we can write the signal autocorrelation function in the form

$$\langle dz(t,\mathbf{r})dz^*(t',\mathbf{r}')\rangle = RP_z^o\sigma_e(t-t',\mathbf{r})\delta(\mathbf{r}-\mathbf{r}')d^3rd^3r', \tag{4.1}$$

where σ_e is the plasma autocorrelation function proportional to the electron density. This equation also contains the condition that signals from different volume elements do not correlate, *i.e.* $\langle dz(t,\mathbf{r})dz^*(t',\mathbf{r}')=0$ when $\mathbf{r}\neq\mathbf{r}'$. The quantity we are interested in is the plasma autocorrelation function because it is determined by the physical parameters of the plasma. One should also notice that eq. (2.84) is a special case of eq. (4.1). The factors containing the transmitted power, the radar wave length and the antenna gain and distance are included in P_z^o. The product $n_e\sigma$ is equal to $\sigma_e(0,\mathbf{r})$, the zero lag of the plasma autocorrelation function.

When the transmission is modulated with an envelope env(t), the signal received from whole the space is

$$z(t) = \int_{\mathbf{r}} \mathrm{env}(t - S(\mathbf{r})dz(t,\mathbf{r}), \tag{4.2}$$

where

$$S(\mathbf{r}) = \frac{r_1(\mathbf{r}) + r_2(\mathbf{r})}{c} \tag{4.3}$$

is the travel time of the signal from transmitter to receiver via the volume element at $\mathbf{r}$. One should notice that the strength of the elementary signals $dz(t,\mathbf{r})$ is controlled by the antenna gains so that no important contribution is obtained from regions outside the radar beams. After passing the post detection filter, the signal has the form

$$\begin{aligned} z_h(t) = (z*h)(t) = (h*z)(t) &= \int_{-\infty}^{\infty} h(t-\tau)z(\tau)d\tau \\ &= \int_{-\infty}^{\infty} h(t-\tau)\left[\int_{\mathbf{r}} \mathrm{env}(\tau - S(\mathbf{r}))dz(\tau,\mathbf{r})\right]d\tau \\ &= \int_{-\infty}^{\infty}\left[\int_{\mathbf{r}} h(t-\tau)\mathrm{env}(\tau - S(\mathbf{r}))dz(\tau,\mathbf{r})\right]d\tau. \end{aligned} \tag{4.4}$$

By defining the amplitude ambiguity function

$$W_t^A(\tau,\mathbf{r}) = h(t-\tau)\mathrm{env}(\tau - S(\mathbf{r})) \tag{4.5}$$

the signal can be written in the form

$$z_h(t) = \int_{-\infty}^{\infty}\left[\int_{\mathbf{r}} W_t^A(\tau,\mathbf{r})dz(\tau,\mathbf{r})\right]d\tau. \tag{4.6}$$

Note carefully that a different amplitude ambiguity function is attached to each time of observation. Eq. (4.6) means that, after the post detection filter, the signal is a weighted sum of elementary signals from all volume elements and all

times, and the weight in this sum is given by the amplitude ambiguity function. Of course, non-zero elementary signals are only obtained from the cross-sections of the radar beams and their amplitudes are controlled by the antenna gains. The amplitude ambiguity function introduces an additional spatial weight associated with the length of the radar modulation and a temporal weight associated with the receiver filtering.

With the help of the amplitude ambiguity function, the signal autocorrelation function can be further developed as

$$\langle z_h(t) z_h^*(t') \rangle =$$

$$\langle \int_{-\infty}^{\infty} \left[\int_{\mathbf{r}} W_t^A(\tau, \mathbf{r}) dz(\tau, \mathbf{r}) \right] d\tau \cdot \int_{-\infty}^{\infty} \left[\int_{\mathbf{r}'} W_{t'}^{A*}(\tau', \mathbf{r}) dz^*(\tau', \mathbf{r}') \right] d\tau' \rangle$$

$$= \int_{-\infty}^{\infty} d\tau \int_{-\infty}^{\infty} d\tau' \left[\int_{\mathbf{r}} \int_{\mathbf{r}'} W_t^A(\tau, \mathbf{r}) W_{t'}^{A*}(\tau', \mathbf{r}') \langle dz(\tau, \mathbf{r}) dz^*(\tau', \mathbf{r}') \rangle \right]$$

$$= R \int_{-\infty}^{\infty} d\tau \int_{-\infty}^{\infty} d\tau' \times$$

$$\left[\int_{\mathbf{r}} \int_{\mathbf{r}'} P_z^o(\mathbf{r}) W_t^A(\tau, \mathbf{r}) W_{t'}^{A*}(\tau', \mathbf{r}') \sigma_e(\tau - \tau', \mathbf{r}) \delta(\mathbf{r} - \mathbf{r}') d^3 r d^3 r' \right]$$

$$= R \int_{-\infty}^{\infty} d\tau \int_{-\infty}^{\infty} d\tau' \left[\int_{\mathbf{r}} P_z^o(\mathbf{r}) W_t^A(\tau, \mathbf{r}) W_{t'}^{A*}(\tau', \mathbf{r}) \sigma_e(\tau - \tau', \mathbf{r}) d^3 r \right] \quad (4.7)$$

By introducing a new variable $\nu = \tau - \tau'$ we obtain

$$\langle z_h(t) z_h^*(t') \rangle =$$

$$R \int_{-\infty}^{\infty} d\tau \int_{-\infty}^{\infty} d\nu \left[\int_{\mathbf{r}} P_z^o(\mathbf{r}) W_t^A(\tau, \mathbf{r}) W_{t'}^{A*}(\tau - \nu, \mathbf{r}) \sigma_e(\nu, \mathbf{r}) d^3 r \right]$$

$$= R \int_{\mathbf{r}} d^3 r P_z^o(\mathbf{r}) \int_{-\infty}^{\infty} d\nu \cdot \sigma_e(\nu, \mathbf{r}) \left[\int_{-\infty}^{\infty} W_t^A(\tau, \mathbf{r}) W_{t'}^{A*}(\tau - \nu, \mathbf{r}) d\tau \right]. \quad (4.8)$$

The integral of amplitude ambiguity functions in eq. (4.8) is known as the two-dimensional ambiguity function

$$W_{tt'}(\nu, \mathbf{r}) = \int_{-\infty}^{\infty} W_t^A(\tau, \mathbf{r}) W_{t'}^{A*}(\tau - \nu, \mathbf{r}) d\tau, \quad (4.9)$$

which is a basic concept in the analysis of incoherent scatter data. Two-dimensional ambiguity function is an unnormalised cross-correlation function (in time direction) of the amplitude ambiguity functions attached to time instants t and t'. Consequently, a different two-dimensional ambiguity function is attached to each pair (t, t') of observation times.

The autocorrelation function of the received signal can now be written as

$$\langle z_h(t) z_h^*(t') \rangle = R \int_{\mathbf{r}} P_z^o(\mathbf{r}) \left[\int_{-\infty}^{\infty} W_{tt'}(\nu, \mathbf{r}) \sigma_e(\nu, \mathbf{r}) d\nu \right] d^3 r. \quad (4.10)$$

According to this result, each lag $t - t'$ of the measured autocorrelation function is a weighted average of the plasma autocorrelation function both in space and in time. Contribution to the measurement is obtained from that part of space where both $P_z^o \neq 0$ and $W_{tt'} \neq 0$. The single electron power P_z^o includes the effects of the antenna pattern (only signal from the crossing volume of the radar beams can be non-zero), wave polarisation effects, attenuation of the signal with distance as well as receiver amplification. In monostatic case the effect of antenna pattern means that P_z^o contains spatial weighting in directions perpendicular to the radar beam. The two-dimensional ambiguity function $W_{tt'}$ includes the effects of transmitter modulation and receiver filtering. As seen from eqs. (4.5) and (4.9), the transmitter modulation also provides a spatial weighting which, in the case of a monostatic radar, controls the scattering volume in the direction of the antenna beam. The post-detection filter takes care of the weight in time direction.

4.2 Examples of two-dimensional ambiguity functions in monostatic case

In monostatic radar system the transmitter and receiver are in the same position, and the space where the scattered radiation is received from, is restricted by the radar beam. Then $W_{tt'}$ depends only on the distance r measured from the antenna or, equivalently, on $S = 2r/c$, which is the travel time of the signal from the antenna to the distance r and back. If the beam is narrow enough, σ_e does not vary within the beam cross section and therefore it also depends only on r. Although this is usually true in incoherent scatter radars with very narrow beams, the assumption may be violated when thin auroral structures are drifting across the antenna pattern.

In a monostatic case the two-dimensional ambiguity function is

$$W_{tt'}(\nu, r) = \int_{-\infty}^{\infty} W_t^A(\tau, r) W_{t'}^{A*}(\tau - \nu, r) d\tau, \tag{4.11}$$

where the amplitude ambiguity function is

$$W_t^A(\tau, r) = h(t - \tau)\mathrm{env}(\tau - S(r)), \tag{4.12}$$

and the autocorrelation function of the received signal can be written as

$$\langle z_h(t) z_h^*(t') \rangle = R \int_0^{\infty} P_{Az}^o(r) \left[\int_{-\infty}^{\infty} W_{tt'}(\nu, r) \sigma_e(\nu, r) d\nu \right] dr. \tag{4.13}$$

Here integration over the cross section of the radar beam has been carried out so that

$$P_{Az}^o(r) = \int_{A(r)} P_z^o(\mathbf{r}) dA, \tag{4.14}$$

where $A(r)$ is the area of the cross section at a distance r. A comparison with eq. (2.84) shows that eq. (4.14) involves integration of the square of the antenna

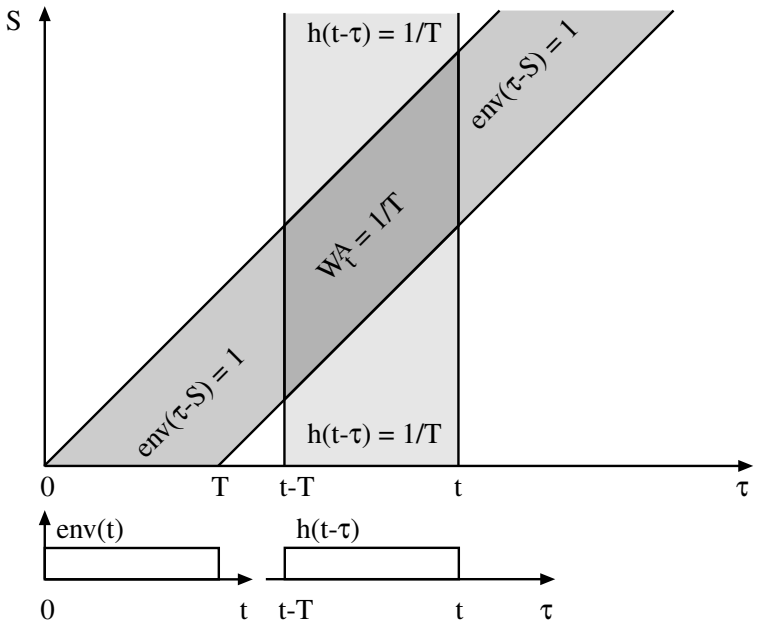

Figure 4.1: Amplitude ambiguity function of a single pulse.

gain. Instead of ν and r, we can also use ν and S as the arguments of the two-dimensional ambiguity function so that $W_{tt'} = W_{tt'}(\nu, S(r))$.

We next take two examples to demonstrate the meaning of the amplitude ambiguity function and the two-dimensional ambiguity function.The first example is a single pulse with a length T. The envelope of this modulation is

$$\text{env(t)} = \begin{cases} 1 \text{ , when } 0 < t < T \\ 0 \text{ , when } t \leq 0 \text{ or } t \geq T. \end{cases}$$

We further assume that the receiver impulse response has a boxcar shape with the same length as the envelope,

$$h(t) = \text{env(t)}/T.$$

The corresponding amplitude ambiguity function $W_t^A = h(t-\tau)\text{env}(\tau - S)$ is shown in Fig. 4.1 in τS-coordinates. The generation of this figure is understood as follows. According to the definition in eq. (4.12), the amplitude ambiguity function is a product of two terms, $h(t-\tau)$ and $\text{env}(\tau - S)$. The first one of these does not depend on S and therefore it is zero everywhere in the τS-plane except within the vertical stripe limited by the lines $\tau = t - T$ and $\tau = t$, where it has a constant value of $1/T$. The second term $\text{env}(\tau - S)$, on the other hand, is equal to unity within the region limited by lines $S = \tau$ and $S = \tau - T$ and zero elsewhere. Hence the amplitude ambiguity function is zero everywhere except within the region where the above two stripes cross. There it gets a value $1/T$.

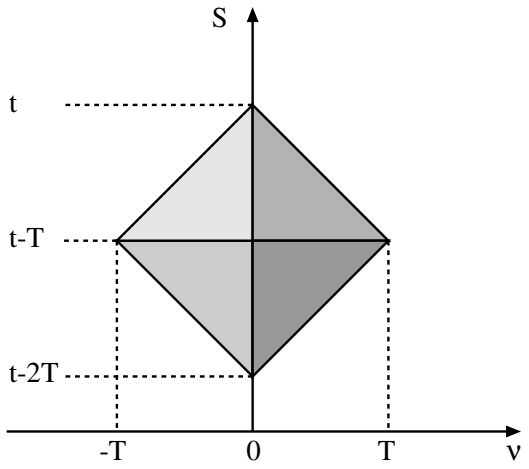

Figure 4.2: Zero lag of the two-dimensional ambiguity function of a single pulse.

This region is shown as a dark grey area in Fig. 4.1. In the course of time t, the region of non-zero amplitude ambiguity function travels upwards between the tilted lines.

The two-dimensional ambiguity functions can now be constructed using the amplitude ambiguity functions. For zero lag, *i.e.* $t' = t$, W_{tt} is simply the (unnormalised) autocorrelation function of W_t^A in lag direction. The amplitude ambiguity function has a boxcar-shaped form in τ-direction for all values $t - 2T < S < t$, otherwise it is zero. The height of the boxcar-function is $1/T$ and, when $S = t - T$, it obtains its maximum length T. The length of the boxcar-function reduces from this value linearly to zero as a function of S and the zero points are reached at $S = t$ and $S = t - 2T$. Because the autocorrelation function of a boxcar function is a triangle, $W_{tt'}$ is a pyramid with a height $1/T$. The result is shown in Fig. 4.2. With increasing t, the pyramid moves upwards along the S-axis.

The result shows that the zero lag of the measured autocorrelation is a weighted average of the plasma autocorrelation function from a lag range $(-T, T)$ and the height range corresponding to the values $t - 2T < S < t$. One must notice that, when calculating the total weight in S-direction, attenuation with distance contained in P_{Az}^o must be taken into account.

In order to study other lags with $t' - t \neq 0$, we have to use the amplitude ambiguity functions W_t^A and $W_{t'}^{A*}$ corresponding to different instants of observation time. These are investigated later in Chapter 5.x for a long pulse modulation and an impulse response shorter than the pulse length. Here it is sufficient to notice that, if $t' - t > 2T$ in the present example, the two amplitude ambiguity functions do not overlap in S-direction in the presentation of Fig. 4.1 and the corresponding two-dimensional ambiguity function will be zero. Hence lags longer than $2T$ give no information on the plasma autocorrelation function.

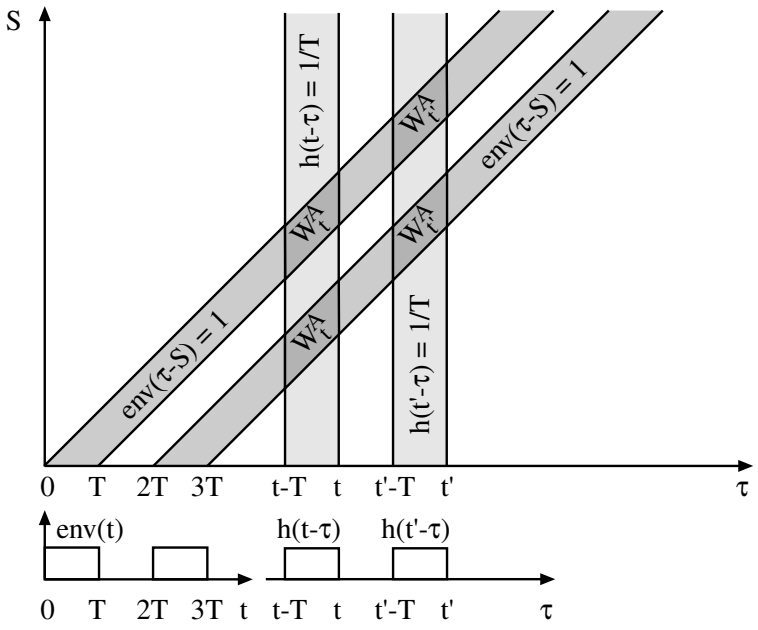

Figure 4.3: Amplitude ambiguity functions of a two-pulse code.

As a second example we take so-called two-pulse code consisting of two simple pulses of length T and a gap of the same length between them. The modulation envelope is

$$\mathrm{env}(t) = \begin{cases} 1 \text{ , when } 0 < t < T \text{ or } 2T < t < 3T \\ 0 \text{ , when } t \leq 0 \text{ or } T \leq t \leq 2T \text{ or } t \geq 3T. \end{cases}$$

As in the previous example, the impulse response has the same shape as a single pulse, *i.e.*

$$h(t) = \mathrm{env}(t)/T.$$

Amplitude ambiguity functions for two observation times t' and t with a difference $2T$ are drawn in Fig. 4.3. As before, the impulse response factor in the amplitude ambiguity function for a single observation time is non-zero within a single vertical strip in τS-plane. The difference is that the envelope factor is non-zero within two strips, and therefore the amplitude ambiguity function is non-zero within two separate regions. Note that these regions do not overlap in S-direction, but the top area starts at the same S-value as the bottom area ends. Note also that the time separation $2T$ of t' and t is chosen to make the top region of W_t^A to cover exactly the same range on the S-axis as the bottom region of $W_{t'}^A$ does.

The two-dimensional ambiguity function for zero lag is obtained using W_t^A in the same way as in Fig. 4.1. However, since the amplitude ambiguity function consists of two separate areas, the resulting two-dimensional ambiguity function

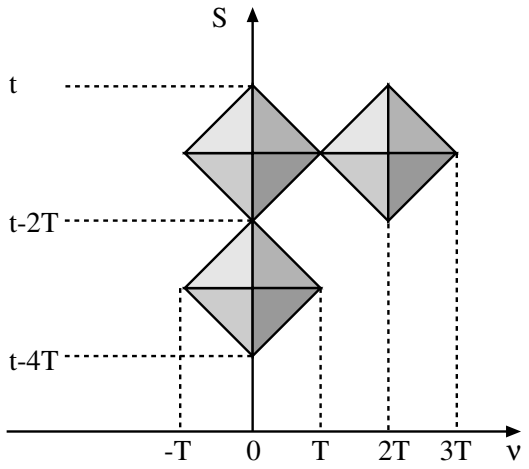

Figure 4.4: Two-dimensional ambiguity functions of a two-pulse code.

will consist of two pyramids around $\nu = 0$ at different values of S as seen in Fig. 4.4. The tops of the pyramids lie at $S = t - T$ and $S = t - 3T$ and their edges touch at $S = t - 2T$.

The two-dimensional ambiguity function for the lag $t' - t$ is obtained as a cross correlation function of the two amplitude functions W_t^A and $W_{t'}^A$. In this case the top region in W_t^A and the bottom region in $W_{t'}^A$ together form a single pyramid which is located around $\nu = 2T$, next to the top pyramid of the zero lag ambiguity function as shown in Fig. 4.4.

Altogether, we observe that the ambiguity function of the lag corresponding to the pulse separation ('first lag') produces an ambiguity function with the same shape as the zero-lag ambiguity function of a single pulse. The only difference is that the function is shifted by $2T$ in ν-direction. The zero-lag ambiguity function of the two-pulse code, on the other hand, covers a double range in altitude. We say that the zero lag has a spatial ambiguity. Spatial ambiguity is an unwanted effect because we would like to measure all lags with the same altitude resolution.

With increasing time, the two-dimensional ambiguity functions drift upwards along the S-axis. Hence, if the radar signal is sampled at intervals of $2T$ and zero lag estimates of the signal autocorrelation function are calculated from all samples as well as first lag estimates from all pairs of successive samples, altitude profiles of zero lag and first lag are obtained. The ambiguity functions of the lag profiles are obtained by shifting the ambiguity functions in Fig. 4.4 upwards in S-direction in steps of $2T$. The lag profiles give information on the altitude profiles of the respective lags of the plasma autocorrelation function. The altitude range of each zero lag observation, however, is twice the range of the first lag observation. The situation can be improved by sending a separate single pulse (perhaps at a different frequency) for obtaining the zero lag with a

better height resolution.

Two lag values are not enough for mapping a plasma autocorrelation function and therefore the modulation shown in this example has no practical value. More lags are obtained if so called multipulse codes are used. These codes are described later in Chapter 5.3. For each lag of these codes a two-dimensional ambiguity function of equal shape is formed and an altitude profile of lag estimate can be obtained. The physical parameters are obtained by least squares fitting of the observations to theoretical ambiguity functions weighted by the same ambiguity functions.

4.3 Range ambiguity function

Eq. (4.13) can be greatly simplified if the plasma autocorrelation function is approximately constant within the lag or altitude range of the two-dimensional ambiguity function. When $\sigma_e(\nu, r)$ depends only weakly on ν while $W_{tt'} \neq 0$, the signal autocorrelation functions can be presented in terms of range ambiguity functions which is one of the two types of reduced ambiguity functions. In this case we can use the approximation

$$\sigma_e(\nu, r) \approx \sigma_e(t' - t, r). \tag{4.15}$$

Then, assuming monostatic case as in Chapter 4.2,

$$\begin{aligned}\langle z_h(t) z_h^*(t') \rangle &= R \int_0^\infty P_{Az}^o(r) \left[\int_{-\infty}^\infty W_{tt'}(\nu, r) \sigma_e(t' - t, r) d\nu \right] dr \\ &= R \int_0^\infty P_{Az}^o(r) \sigma_e(t' - t, r) \left[\int_{-\infty}^\infty W_{tt'}(\nu, r) d\nu \right] dr \\ &= R \int_0^\infty P_{Az}^o(r) W_{tt'}^r(S(r)) \sigma_e(t' - t, r) dr, \end{aligned} \tag{4.16}$$

where $W_{tt'}^r$ is the range ambiguity function defined as

$$W_{tt'}^r(S) = \int_{-\infty}^\infty W_{tt'}(\nu, r(S)) d\nu, \tag{4.17}$$

i.e. the range ambiguity function is obtained by integrating the two-dimensional ambiguity function in lag direction.

The range ambiguity function can be further developed using the definition of the two-dimensional ambiguity function. In terms of the amplitude ambiguity functions,

$$W_{tt'}^r(S) = \int_{-\infty}^\infty \left[\int_{-\infty}^\infty W_t^A(\tau, r) W_{t'}^{A*}(\tau - \nu, r) d\tau \right] d\nu$$

$$= \int_{-\infty}^\infty \left[\int_{-\infty}^\infty h(t - \tau) \mathrm{env}(\tau - S) \cdot h[t' - (\tau - \nu)] \mathrm{env}^*(\tau - \nu - S) d\tau \right] d\nu$$

$$= \int_{-\infty}^{\infty} h(t-\tau)\mathrm{env}(\tau - S)d\tau \cdot \int_{-\infty}^{\infty} h(t'-\tau)\mathrm{env}^*(\nu - S)d\nu. \qquad (4.18)$$

Using a new variable $\tau' = t - \tau$, the first integral in eq. (4.18) can be written as

$$\int_{-\infty}^{\infty} h(t-\tau)\mathrm{env}(\tau - S)d\tau = \int_{-\infty}^{\infty} h(\tau')\mathrm{env}(t - S - \tau')d\tau' = (h * \mathrm{env})(t - S).$$

Because the second integral is of the same form, it must be equal to

$$(h * \mathrm{env})^*(t' - S)$$

and therefore the range ambiguity function is

$$W^r_{tt'}(S) = [(h * \mathrm{env})(t - S)] \cdot [(h * \mathrm{env})^*(t' - S)]\,. \qquad (4.19)$$

This shows that it is not necessary to first calculate the two-dimensional ambiguity function in order to obtain the range ambiguity function, but it can be directly obtained from the convolution of the impulse response and the modulation envelope.

We can now examine the range ambiguity functions in the cases of the two examples presented in Chapter 4.2. When both the modulation envelope and the impulse response are single pulses with equal lengths, $h*\mathrm{env}$ has a triangular shape as shown in the top panel of Fig. 4.5. Therefore the factor $(h*\mathrm{env})(t-S)$ in eq, (4.19) is also a triangle located between t and $t-2T$ on the S-axis. This is shown on the left hand side of the bottom panel. The range ambiguity function

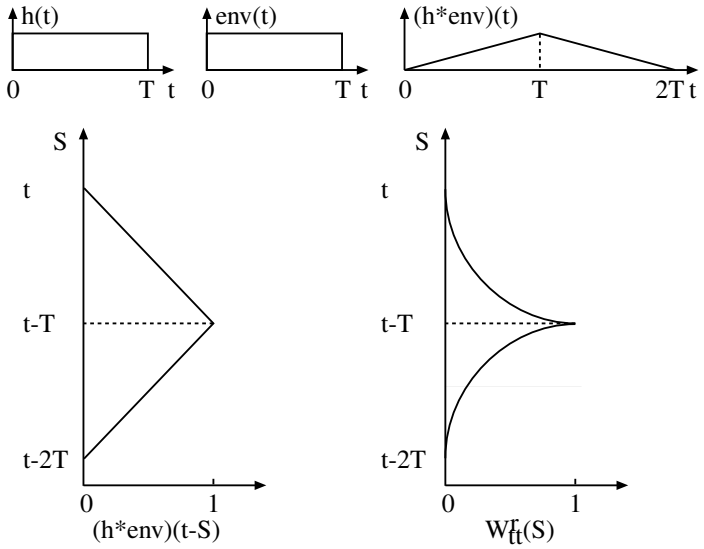

Figure 4.5: Generation of a zero lag range ambiguity function of a simple pulse.

for zero lag is the square of this function and it is plotted on the right hand side of the bottom panel. The same result is obtained, if the two-dimensional ambiguity function in Fig. 4.2 is integrated in ν-direction.The function consists of two parabolic shaped arcs on both sides of the point $S = t - T$.

The second example consists of a two-pulse code and a boxcar impulse response which is as long as the pulses in the code. The code, the impulse response and the convolution $h * \text{env}$ are shown on the top of Fig. 4.6. The convolution consists of two triangles but, unlike in the case of the envelope, there is no gap between the two elements.

The generation of the range ambiguity functions of zero lag and first lag is shown at the bottom of Fig. 4.6. The two plots at the left hand side contain the two factors $(h * \text{env})(t - S)$ and $(h * \text{env})^*(t' - S)$ of the range ambiguity function with the assumption $t' = t + 2T$. With this choice the lower triangle of $(h * \text{env})^*(t' - S)$ overlaps the upper triangle of $(h * \text{env})(t - S)$. The range ambiguity function of zero lag is obtained by taking a square of $(h * \text{env})^*(t' - S)$ and it consists of two peaks around $S = t - T$ and $S = t - 2T$. The shape of

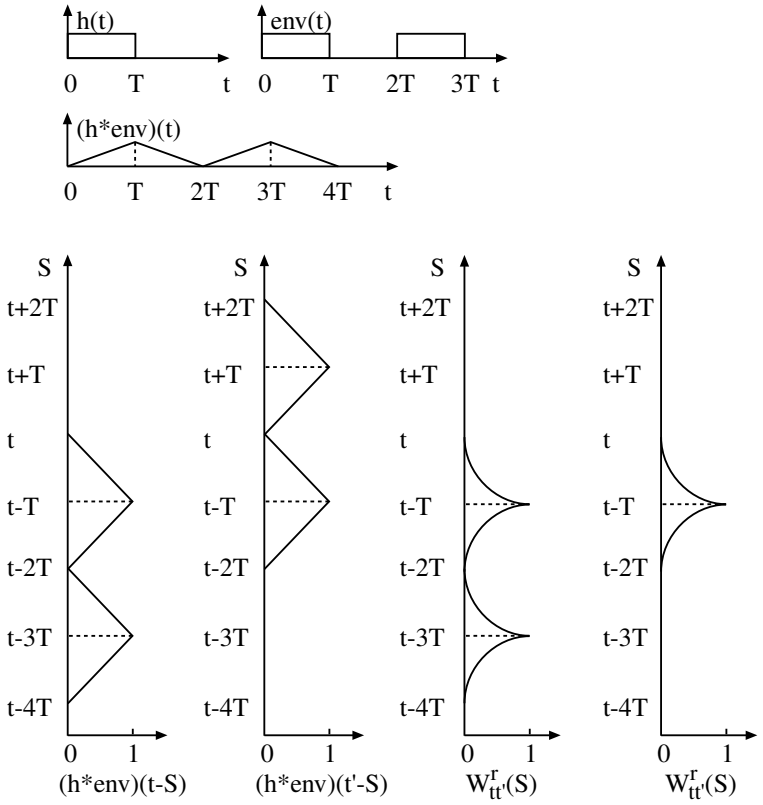

Figure 4.6: Generation of range ambiguity functions of a two-pulse code.

the peaks is the same as in Fig. 4.5. The same result is obtained by integrating the two-dimensional ambiguity function of zero lag in Fig. 4.4 in ν-direction. The presence of two peaks is an indication of the range ambiguity discussed in Chapter 4.2.

The range ambiguity function for the lag $t' - t$ (the first lag) is obtained by multiplying $(h * \mathrm{env})(t - S)$ and $(h * \mathrm{env})^*(t' - S)$. The result contains only a single peak within the region $t - 2T < S < t$. Again, this result would be obtained by integrating the two-dimensional ambiguity function of the first lag in Fig. 4.4 in ν-direction.

4.4 Lag ambiguity function

The second type of reduced ambiguity functions can be used if the plasma autocorrelation function is approximately constant within the altitude range of the two-dimensional ambiguity function. Then we have the approximation

$$\sigma_e(\nu, r) \approx \sigma_e(\nu, r_{tt'}), \tag{4.20}$$

where $r_{tt'}$ is the distance of the centre point of $W_{tt'}(\nu, r)$ from the transmitter. In this case the signal autocorrelation function in a monostatic experiment is

$$\begin{aligned}\langle z_h(t) z_h^*(t') \rangle &= R \int_0^\infty P_{Az}^o(r) \left[\int_{-\infty}^\infty W_{tt'}(\nu, r(S)) \sigma_e(\nu, r_{tt'}) d\nu \right] dr \\ &= R P_{Az}^o(r_{tt'}) \int_{-\infty}^\infty \left[\int_{-\infty}^\infty W_{tt'}(\nu, r(S)) dr \right] \sigma_e(\nu, r_{tt'}) d\nu. \end{aligned} \tag{4.21}$$

Here $W_{tt'}(\nu, r)$ is assumed to be so narrow in r-direction that P_{Az}^o does not change appreciably within the region where $W_{tt'}$ is non-zero. The autocorrelation function of the signal can now be put in the form

$$\langle z_h(t) z_h^*(t') \rangle = R P_{Az}^o(r_{tt'}) \int_{-\infty}^\infty W_{tt'}^S(\nu) \sigma_e(\nu, r_{tt'}) d\nu, \tag{4.22}$$

where $W_{tt'}^S(\nu)$ is the lag ambiguity function defined as

$$W_{tt'}^S(\nu) = \int_{-\infty}^\infty W_{tt'}(\nu, r(S)) dr. \tag{4.23}$$

The lag ambiguity function can be further developed much in the same way as the range ambiguity function in Chapter 4.3. By inserting $W_{tt'}$ in eq. (4.23) in terms of the amplitude ambiguity functions, we obtain

$$W_{tt'}^S(\nu) = \int_{-\infty}^\infty \left[\int_{-\infty}^\infty W_t^A(\tau, r(S)) W_{t'}^{A*}(\tau - \nu, r(S)) d\tau \right] dr$$

$$= \int_{-\infty}^\infty \left[\int_{-\infty}^\infty h(t - \tau) \mathrm{env}(\tau - S(r)) \cdot h[t' - (\tau - \nu)] \mathrm{env}^*(\tau - \nu - S(r)) d\tau \right] dr$$

$$= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \mathrm{env}(\tau - S(r))\mathrm{env}^*(\tau - \nu - S(r))dr \right] h(t-\tau)h(t'-\tau+\nu)d\tau. \quad (4.24)$$

Because $S = 2r/c$, the inner integral in eq. (4.24) is equal to

$$\frac{c}{2}\int_{-\infty}^{\infty} \mathrm{env}(\tau-S)\mathrm{env}^*(\tau-\nu-S)dS = \frac{c}{2}\int_{-\infty}^{\infty} \mathrm{env}(S')\mathrm{env}^*(S'-\nu)dS' = \frac{c}{2}R_{\mathrm{env}}(\nu),$$

where $R_{\mathrm{env}}(\nu)$ is the unnormalised autocorrelation function of the modulation envelope. The result does not depend on the variable τ of the outer integral. Using a new variable $\tau' = t - \tau$ we then obtain

$$W_{tt'}^S(\nu) = \frac{c}{2}R_{\mathrm{env}}(\nu)\int_{-\infty}^{\infty} h(\tau')h(\tau' + t' - t + \nu)d\tau'. \quad (4.25)$$

This can be written as

$$W_{tt'}^S(\nu) = \frac{c}{2}R_{\mathrm{env}}(\nu)R_h(t' - t + \nu), \quad (4.26)$$

where R_h is the unnormalised autocorrelation function of the impulse response. Therefore it is not necessary to calculate the two-dimensional ambiguity function in order to get the lag ambiguity function, but it can be obtained directly from the autocorrelation functions of the modulation envelope and the impulse response.

The generation of lag ambiguity functions is next demonstrated using the previous two examples. In the case of a simple transmitted pulse, the autocorrelation functions of both the modulation envelope and the impulse response have a triangular shape as shown in Fig. 4.7. Consequently, the lag ambiguity

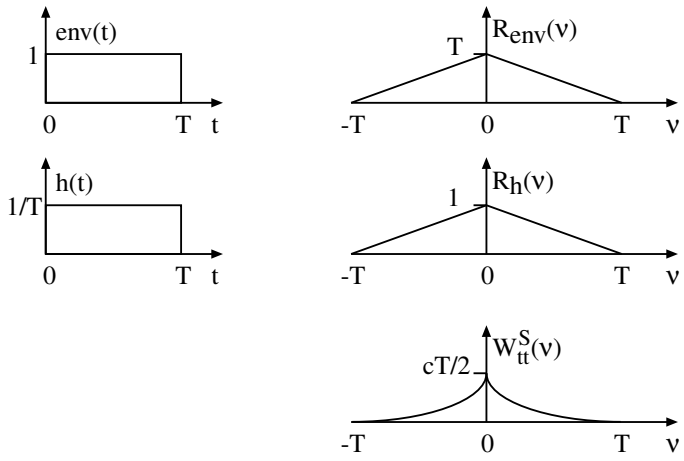

Figure 4.7: Generation of the lag ambiguity function of zero lag for a simple pulse.

function for zero lag is a single peak around $\nu = 0$, composed of two curves of parabolic shape.

The autocorrelation function of a two-pulse code consists of three triangles, with the height of the centre triangle twice the height of the other two triangles. The code and its autocorrelation function $R_{\mathrm{env}}(\nu)$ are shown in the two topmost panels in Fig. 4.8. The autocorrelation function of the boxcar impulse response $R_h(\nu)$ is shown in the next panel. As before, the length of the impulse response is equal to the pulse length. In order to obtain the lag ambiguity function for zero lag, $R_{\mathrm{env}}(\nu)$ and $R_h(\nu)$ must be multiplied. The result is a single peak with a height cT around $\nu = 0$, shown in the fourth panel.

The lag ambiguity function of the first lag corresponding to a delay $t - t' = 2T$ is obtained by multiplying $R_{\mathrm{env}}(\nu)$ and the autocorrelation function of the impulse response shifted by $2T$. The latter is shown in the fifth panel and the resulting ambiguity function in the bottom panel of Fig. 4.8. The lag ambiguity function is a peak around $\nu = 2T$ and it has the same shape as the zero lag

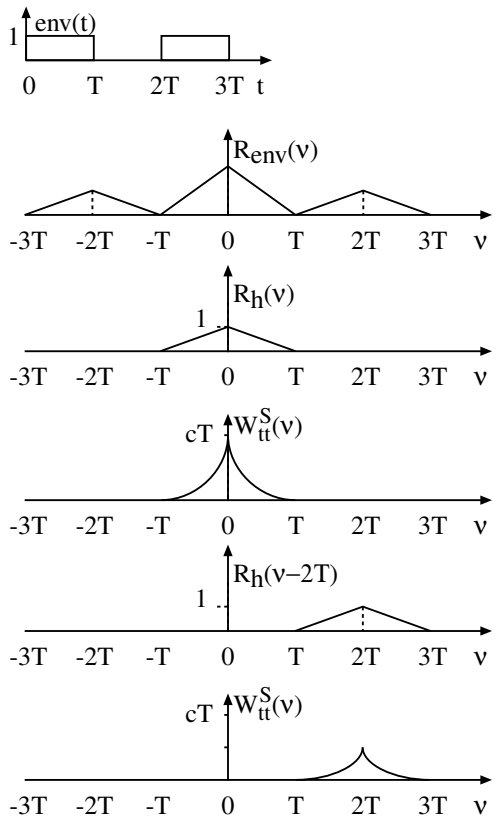

Figure 4.8: Generation of the lag ambiguity functions of zero lag and first lag for a two-pulse code.

ambiguity function, but with a peak value of only one half of the zero lag peak value. This can be easily understood with the help of Fig. 4.4 remembering that the lag ambiguity function is an integral of the two-dimensional ambiguity function in S-direction. Since the the two-dimensional ambiguity function of zero lag consists of two pyramids in νS-plane and that of the first lag contains only one pyramid, the above result is obvious.

4.5 Spectral ambiguity function

The relation between the autocorrelation function of the observed signal and the plasma autocorrelation function is completely determined by the two-dimensional autocorrelation function. The calculation of the two-dimensional ambiguity functions in an incoherent scatter experiment is a time-consuming task. It turns out that a faster analysis programme can be built if plasma scattering spectra are used instead of the plasma autocorrelation functions and the two-dimensional ambiguity functions are replaced by so called spectral ambiguity functions. Although these concepts are defined in frequency domain rather than lag domain, they can be used in connection with the signal autocorrelation function defined in lag domain, *i.e.* there is no need to calculate the signal spectrum.

Before going to the definition of the spectral ambiguity function, we give a proof of a relationship known as the Parseval formula. If $x(t)$ and $y(t)$ are complex signals which both have a Fourier transform,

$$\int_{-\infty}^{\infty} x(t)y^*(t)dt = \int_{-\infty}^{\infty} x(t)\left[\int_{-\infty}^{\infty} Y^*(\omega)e^{-i\omega t}d\nu\right]dt$$

$$= \int_{-\infty}^{\infty} Y^*(\omega)\left[\int_{-\infty}^{\infty} x(t)e^{-i\omega t}dt\right]d\nu = \int_{-\infty}^{\infty} X(\omega)Y^*(\omega)d\nu. \tag{4.27}$$

We observe that eq. (3.8), which is associated with signal power in time and frequency domains, is a special case of this formula.

The plasma scattering spectrum is defined as the Fourier transform of the plasma autocorrelation function,

$$\Sigma_e(\nu, \mathbf{r}) = \int_{-\infty}^{\infty} \sigma_e(\nu', \mathbf{r})e^{-i2\pi\nu\nu'}d\nu'. \tag{4.28}$$

Note carefully that here the notation ν is used for frequency in accordance with the notation in Chapter 3, whereas the argument ν' of the two-dimensional ambiguity function is lag. Although the plasma autocorrelation function is usually complex, its real part is always an even and imaginary part an odd function of time delay. Then, according to the properties of the Fourier transform presented in Chapter 3.2, the scattering spectrum is always real.

The spectral ambiguity function $\mathcal{W}_{tt'}$ is defined as a Fourier transform of the two-dimensional ambiguity function. Hence

$$\mathcal{W}_{tt'}(\nu, \mathbf{r}) = \int_{-\infty}^{\infty} W_{tt'}(\nu', \mathbf{r})e^{-i2\pi\nu\nu'}d\nu'. \tag{4.29}$$

According to eq. (4.10), the signal autocorrelation function depends on the integral over the lag variable of the function which is obtained as a product of the two-dimensional ambiguity function and the plasma autocorrelation function. This integral can be transformed using the Parseval formula into a form

$$\int_{-\infty}^{\infty} W_{tt'}(\nu', \mathbf{r})\sigma_e(\nu', \mathbf{r})d\nu' = \int_{-\infty}^{\infty} \mathcal{W}_{tt'}(\nu, \mathbf{r})\Sigma_e(\nu, \mathbf{r})d\nu. \tag{4.30}$$

When this is inserted in eq. (4.10), the result is

$$\langle z_h(t)z_h(t')\rangle = R\int_{\mathbf{r}} P_z^o(\mathbf{r})\left[\int_{-\infty}^{\infty} \mathcal{W}_{tt'}(\nu, \mathbf{r})\Sigma_e(\nu, \mathbf{r})d\nu\right] d^3r. \tag{4.31}$$

This equation gives a possibility to fit a theoretical scattering spectrum to the measured signal autocorrelation function. When the definition of the spectrum ambiguity function in eq. (4.29) is further developed, a new form essentially faster in numerical calculations is obtained. Using the definition of the two-dimensional ambiguity function we get

$$\begin{aligned}
\mathcal{W}_{tt'}(\nu, \mathbf{r}) &= \int_{-\infty}^{\infty} W_{tt'}(\nu', \mathbf{r})e^{-2\pi i\nu\nu'}d\nu' \\
&= \int_{-\infty}^{\infty} e^{-2\pi i\nu\nu'}\left[\int_{-\infty}^{\infty} W_t^A(\tau, \mathbf{r})W_{t'}^{A*}(\tau - \nu', \mathbf{r})d\tau\right] d\nu' \\
&= \int_{-\infty}^{\infty} W_t^A(\tau, \mathbf{r})e^{-2\pi i\nu\tau}\left[\int_{-\infty}^{\infty} W_{t'}^{A*}(\underbrace{\tau - \nu'}_{\nu''}\mathbf{r})e^{-2\pi i\nu(\overbrace{\tau - \nu'}^{\nu''})}\underbrace{d\nu'}_{-d\nu''}\right] d\tau \\
&= \left[\int_{-\infty}^{\infty} W_t^A(\tau, \mathbf{r})e^{-2\pi i\nu\tau}d\tau\right]\left[\int_{-\infty}^{\infty} W_{t'}^{A*}(\nu'', \mathbf{r})e^{-2\pi i\nu\nu''}d\nu''\right] \\
&= \left[\int_{-\infty}^{\infty} W_t^A(\tau, \mathbf{r})e^{-2\pi i\nu\tau}d\tau\right]\left[\int_{-\infty}^{\infty} W_{t'}^A(\tau', \mathbf{r})e^{-2\pi i\nu\tau'}d\tau'\right]^* \\
&= \mathcal{F}\{W_t^A(\tau, \mathbf{r})\} \cdot \mathcal{F}^*\{W_{t'}^A(\tau', \mathbf{r})\}.
\end{aligned} \tag{4.32}$$

In conclusion, the fitting of the plasma parameters to the observed signal autocorrelation function takes place as follows. The amplitude ambiguity functions are calculated from eq. (4.5) using the modulation envelope and the receiver impulse response. The spectral ambiguity functions are next obtained according to eq. (4.32). The plasma scattering spectrum is then calculated for a set of plasma parameter values and the product of the scattering spectrum and the spectral ambiguity function is integrated over the frequency domain. The result is used in calculating the right hand side of eq. (4.31). The process is repeated by varying the plasma parameters in order to make the right hand side of eq. (4.31) to fit the observed signal autocorrelation function in the least squares sense.

Chapter 5

Classical modulation methods

5.1 Lag profile matrix

In a monostatic incoherent scatter experiment a certain modulation pattern is usually transmitted at a few separate frequencies. The transmission may consist of several types of modulation. Reception of the scattered signal starts after transmission and continues during a period corresponding to the altitude range of interest in the measurement. The experiment also contains detection of the background noise level as well as power calibration via injection of known power of white noise at the front end of the receiver. After the whole cycle is completed, transmission starts again. The duration of the cycle is normally about 5–10 ms. The next transmission may use the same frequencies if the cycle is so long that no scattering signal from the pulses transmitted during the previous cycle are observed, otherwise the frequencies must be alternated between successive cycles.

In the receiver the complex signal from the quadrature detector is first low-pass filtered and then sampled by an AD converter. The samples are taken at equally spaced times $t_i = i\Delta t$, where Δt is the sampling interval and $i = 1, 2, 3, \ldots$. The result is a set of complex samples

$$x_i = z_h(t_i),\ i = 1, 2, 3, \ldots$$

where $i = 1$ corresponds the first sample taken after the transmission period. The samples are temporarily stored in a buffer memory. The correlator reads the samples, calculates lagged products $x_i \cdot x_j^*$ and adds them in a result memory, each product in its own location. The procedure is repeated for a great number of transmission periods so that sums of the products are collected in the result memory. The sums give estimates of the signal autocorrelation function, each of them corresponding to a specific lag value and altitude range. The possible lag values are multiples of the sampling interval. After a pre-set number of

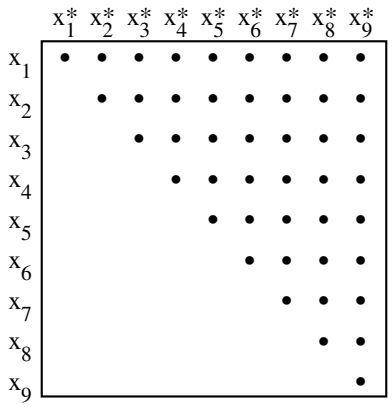

Figure 5.1: The lag profile matrix.

transmission periods the computer reads the contents of the result memory and stores it in a data file. Then the result memory is cleared and data collection can start again.

The basic temporal resolution of an incoherent scatter experiment is usually determined by the time interval between successive data transfers from the result memory to a computer file. A typical time resolution is 10 s. In such an experiment the number of lagged products added to each location in the result memory is of the order of 1000.

It is practical to organise the averaged lagged products (the contents of the result memory divided by the number of additions) into a lag profile matrix shown in Fig. 5.1. Here the data samples are indicated as a column on the left hand side of the matrix and their complex conjugates as a row above the matrix. Each matrix element is the mean product of the samples on same row at the left hand side and on the same column on the top of the matrix. In other words, the ij-element of the matrix is equal to the average value of $x_i \cdot x_j^*$. It is not necessary to fill the lower left corner of this matrix since the elements in this region would be simply complex conjugates of the elements on the opposite side of the main diagonal.

The diagonal elements of the lag profile matrix are of the form $x_i \cdot x_i^*$. This means that the estimates of the zero lag, which are proportional to the signal power, are located on the main diagonal of the lag profile matrix. The first diagonal element $x_1 \cdot x_1$ corresponds to the lowest altitude and it gives information from the height range of the range ambiguity function W_{11}^r. The range ambiguity function of the next element $x_2 \cdot x_2^*$ has a similar shape but it is shifted by an amount $c\Delta t/2$ upwards in the direction of the radar beam. Altogether, the diagonal elements give the profile of zero lags with a height separation of $c\Delta t/2$ and each element contains contribution from a height interval defined by the range ambiguity function of the zero lag.

The first side diagonal contains elements of the type $x_i \cdot x_{i+1}^*$ corresponding to

the delay Δt, which is the first lag. These estimates are arranged in the matrix to make a profile of the first lag. The element on the first row corresponds to the lowest altitude and the separation of the estimates is $c\Delta t/2$ in the beam direction. The ambiguity functions of all elements on this diagonal have the same shape. All other side diagonals follow the same principle, the second side diagonal containing the profile of the second lag with a delay $2\Delta t$, etc. The number of lag profiles in the matrix may vary from one experiment to another depending on the length of the plasma autocorrelation function and the lag resolution needed in the experiment.

5.2 Data sampling

When sampling an analogue signal it is essential to use a correct sampling frequency which is determined by the spectral width of the signal. The rule is given by the sampling theorem also known as the Nyquist theorem (not to be confused with the Nyquist noise theorem):

> A signal with a cut-off frequency ν_c can be reconstructed from data samples taken at a sampling frequency ν_s if and only if $\nu_s \geq 2\nu_c$.

This means that the sampling frequency must be at least twice the highest frequency in the signal spectrum. The lowest possible sampling frequency is called the Nyquist frequency. It is important to notice that the reconstruction of the original signal does not mean approximate interpolation between data points but an exact reproduction of the original signal value at any time instant. No exact proof of the sampling theorem is given here, but it is only pointed out that using the Nyquist frequency gives two data points per period for the highest frequency component in the signal spectrum. This is sufficient for the reproduction of this Fourier component as well as all lower frequencies in the signal.

Breaking the sampling theorem leads to aliasing. Aliasing means that any signal with a frequency $\nu_0 > 2\nu_s$ is observed as a virtual signal at a frequency

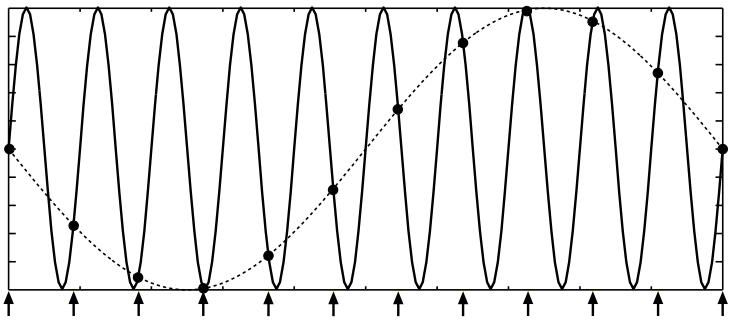

Figure 5.2: Aliasing due to undersampling of data.

$\nu_s/2 - (\nu_0 - \nu_s/2) = \nu_s - \nu_0$. Aliasing can be easily understood in terms of a simple example shown in Fig. 5.2. Here a sinusoidal signal with a frequency of 1 unit is shown by the continuous line. This curve is sampled at a frequency 1.1 units so that the sampling interval is $1/1.1 \approx 0.91$ units. The sampling times are indicated by the arrows below the horizontal axis and the data values by the filled dots on the curve. It is obvious that, if the continuous line were wiped off, the data points would seem to match the lower frequency curve indicated by the dashed line. This is an artificial signal due to aliasing and its frequency is 1.1 – 1 = 0.1 units.

In Chapter 3.11 it was noticed that, depending on the applied modulation, the bandwidth of the scattering signal in the EISCAT UHF radar is of the order of 25–250 kHz. This would imply sampling frequencies of 50–500 kHz which means sampling intervals of 20–2 μs. The sampling interval also determines the height difference $c\Delta t/2$ of the successive lagged products $x_i \cdot x_j$ and $x_{i+1} \cdot x_{j+1}$. The above sampling intervals correspond to differences of 3km – 300 m.

In addition to the height resolution, the sampling interval also affects the observed autocorrelation function and the corresponding spectrum. The lag resolution $\Delta\tau$ cannot be smaller than the sampling interval since the lag estimates are obtained only at multiple values of Δt. The spectrum, on the other hand, cannot contain frequencies higher than twice the sampling frequency. Therefore, when the signal spectrum is calculated from the sampled autocorrelation function using a discrete Fourier transform, the resulting spectral points are evenly distributed within the region $(-1/2\Delta\tau, 1/2\Delta\tau)$ on the frequency axis. This means that improving the lag resolution (diminishing $\Delta\tau$) increases the width of the calculated spectral range. The widest frequency band is obtained when $\Delta\tau = \Delta t$.

For a better understanding of the relation of a sampled autocorrelation function and the corresponding spectrum we first notice that the real and imaginary parts of the autocorrelation function are even and odd functions of lag, respectively. Hence the measured autocorrelation function can be easily expanded to negative lag values. If the longest measured lag is $\tau_n = n\Delta\tau$, the spectrum is obtained as a discrete Fourier transform of $2n + 1$ complex lag estimates. The result consists of $2n + 1$ spectral values evenly distributed within the interval $(-1/2\Delta\tau, 1/2\Delta\tau)$ so that the centre frequency is zero. Hence the frequency step in the spectrum will be $1/(2n\Delta\tau) = 1/(2\tau_n)$. This shows that the longest measured lag determines the frequency resolution and the lag resolution determines the frequency range of the calculated spectrum.

In planning an incoherent scatter experiment it is essential to know how densely an autocorrelation function should be sampled and how long the sampling should be continued; that is, what is the longest lag to be measured. This can be decided using the above results. The problem is enlightened by the following example, where the incoherent scatter spectrum is approximated by a boxcar function

$$S_z(\nu) = \begin{cases} 1/2\nu_c \text{ ,when } -\nu_c < \nu < \nu_c \\ 0 \text{ , when } \nu < -\nu_c \text{ or } \nu > \nu_c. \end{cases} \tag{5.1}$$

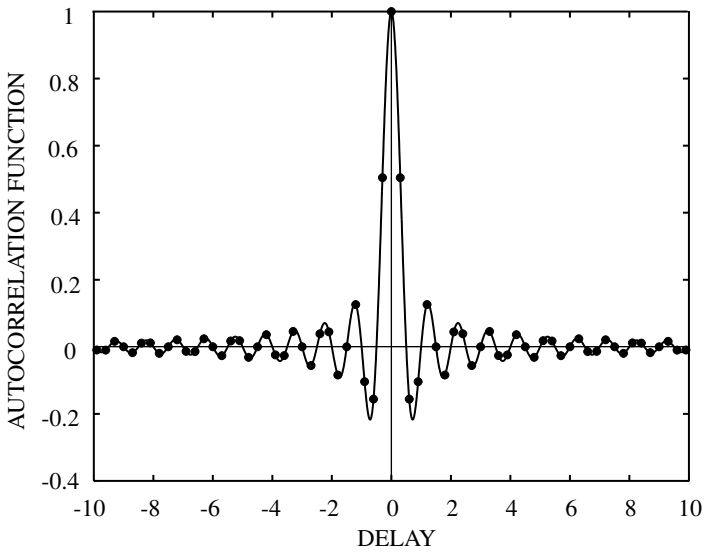

Figure 5.3: The autocorrelation function of a signal with a boxcar-shaped spectrum, sampled at intervals of 0.3.

The corresponding autocorrelation function is

$$R_z(\tau) = \mathcal{F}^{-1}\{S_z\} = \int_{-\infty}^{\infty} S_\nu e^{2\pi i\nu\tau} d\nu = \frac{1}{2\nu_c}\int_{-\nu_c}^{\nu_c} e^{2\pi i\nu\tau} d\nu$$

$$= \frac{1}{\omega_c\tau}\cdot\frac{e^{i\omega_c\tau} - e^{-i\omega_c\tau}}{2i} = \frac{\sin\omega_c\tau}{\omega_c\tau} = \mathrm{sinc}(\omega_c\tau). \qquad (5.2)$$

The autocorrelation function is shown in Fig. 5.3 within the interval (-10, 10) together with sample points taken at intervals of 0.3. Here $n = 33$ and $\Delta\tau = 0.3$ so that $\tau_n = n\Delta\tau = 9.9$. Then the spectral width of the discrete calculated spectrum is $\nu_c = 1/(2\Delta\tau) = 1.67$ and the frequency step $\Delta\nu \approx 0.05$. When the discrete data values in Fig. 5.3 are transformed using a discrete Fourier transform, the resulting spectral points are shown by filled dots in Fig. 5.4. We indeed observe that they are located within the frequency range (-1.67, 1.67) at intervals of about 0.05. We also see that the spectrum calculated from the discrete samples does not produce an exact boxcar shape, but a sawtooth pattern is generated. This is known as the Gibbs phenomenon.

Fig. 5.4 shows that the sampling interval of the autocorrelation function in Fig. 5.3 is sufficient to produce a frequency axis wide enough for the whole spectrum to be visible. The continuous line in Fig. 5.4 is calculated using the same length of autocorrelation function but with a sampling interval 0.01. The step in the frequency domain is not changed and the Gibbs phenomenon remains the same. The difference is that the highest frequency of the calculated spectrum is $\nu_c = 1/(2 \cdot 0.01) = 50$, *i.e.* the frequency axis is unnecessarily wide for the

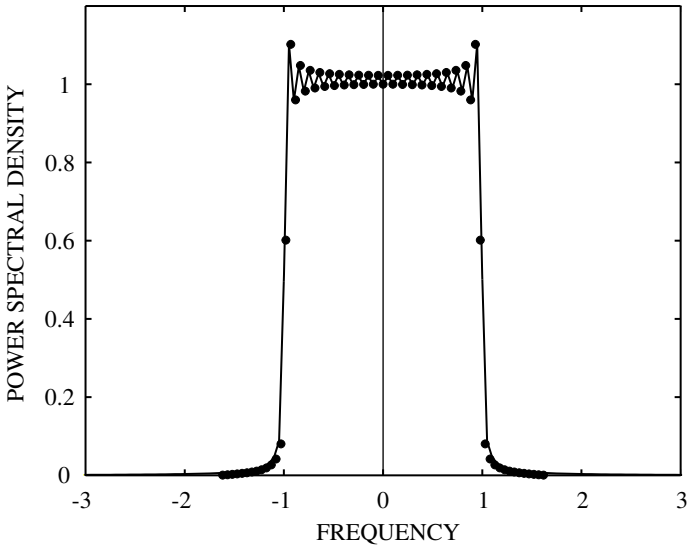

Figure 5.4: The spectrum of the autocorrelation function in Fig. 5.3 calculated from discrete samples taken at two different lag resolutions, 0.3 (dots) and 0.01 (continuous line).

presentation of the spectrum. This means that the autocorrelation function is heavily oversampled. In other words, the extra data points between those shown in Fig. 5.3 do not contain any essential new information.

In addition to a wide enough frequency axis, a sufficient frequency resolution is also needed for a proper presentation of a spectrum. The frequency step is determined by the longest lag in the autocorrelation function. In Fig. 5.4 the frequency step is short enough, as a matter of fact, even a longer step might be used. The extreme limit for observing a boxcar spectrum would be three points

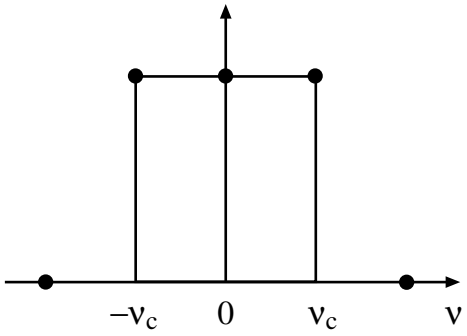

Figure 5.5: A boxcar spectrum sampled by three points.

within the box as shown in Fig. 5.5. Then the frequency step would be equal to the cut-off frequency of the spectrum,

$$\Delta\nu = 1/(2\tau_n) = \nu_c.$$

Because the ith zero point of the function $\mathrm{sinc}(\omega_c\tau)$ is at

$$\tau_i^0 = i/(2\nu_c),$$

this means that the autocorrelation function should be measured to the first zero point.

In order to obtain some sort of information on the spectrum shape, at least five points are needed within the ion line. Then the frequency step would be

$$\Delta\nu = \frac{1}{2\tau_n} = \frac{\nu_c}{2}$$

and the maximum lag

$$\tau_n = \frac{2}{2\nu_c} = \tau_2^0.$$

Hence the autocorrelation must be sampled at least up to the second, perhaps even up to the third or fourth zero point. Table 5.1 shows zero-points of the autocorrelation functions corresponding to boxcar spectra of various widths. In real incoherent scatter signals the spectral widths and the zeros of the autocorrelation functions do not follow precisely this table but, anyway, the table gives an idea of the magnitudes involved. Using the spectral widths estimated in Chapter 3.11 for the EISCAT UHF radar, one can decide that the autocorrelation functions should be measured maximally up to lags of 300–400 μs.

Table 5.1.
Lag values of zero points for autocorrelation functions corresponding to boxcar spectra of various cut-off frequencies.

ν_c (kHz)	τ_1^0 (μs)	τ_2^0 (μs)	τ_3^0 (μs)	τ_4^0 (μs)
2	250	500	750	1000
5	100	200	300	400
7	70	140	210	280
10	50	100	150	200

In conclusion, we can say that an incoherent scatter experiment must have a sufficient
- altitude resolution
- spectral width (lag resolution of the autocorrelation function)
- spectral resolution (length of the autocorrelation function).

Two factors are connected to the altitude resolution, the height span and height separation of the range ambiguity functions. The former is determined by two

factors, the modulation envelope and the receiver impulse response, and the latter by the sampling interval. In many experiments the height span and separation are the same but experiments also exist with height spans larger than height separation. This means that successive elements in lag profiles contain information partially from a same altitude region.

The height separation of the range ambiguity functions and the spectral width are closely associated since they are determined by the sampling interval. On the other hand, the height span of the range ambiguity functions and the maximum length of the signal autocorrelation function are connected to each other because they are determined by the modulation envelope.

The properties of the ionosphere are height dependent; the width of the incoherent scatter spectrum, for example, increases steeply with height in the E and lower F regions. The vertical scales of the ionospheric structures also vary with altitude. In E region, structures with a thickness of a few hundreds of metres are sometimes encountered whereas the scales in F region are usually of the order of tens of kilometres. An ideal incoherent scatter experiment should be able to fulfil both the spectral and height resolution demands at various altitudes. In addition, it should give a sufficient temporal resolution and signal-to-noise ratio. A single modulation is usually not able to meet all these demands and therefore different experiments are designed for different purposes. Experiments designed for simultaneous probing of different height ranges, E and F regions for instance, normally contain a different modulation for each region to be studied. The modulation methods most widely used in the present incoherent scatter experiments are introduced in the next chapters.

5.3 Pulse codes

Although pulse codes are not the simplest modulations applied in incoherent scatter experiments, they are probably easiest to understand. The simplest modulation consists of single long pulses, but complexities not present in pulse codes are associated with their use in measuring the plasma autocorrelation function.

A set of two pulses, a two-pulse code, was adopted in Chapter 4 to demonstrate the ambiguity functions and it was observed that this modulation allows the determination of the 'first lag', *i.e.* the autocorrelation function estimate at a delay corresponding to the separation of the two pulses. The two-dimensional ambiguity function of the first lag was found to be a pyramid located between T and $3T$ in the lag direction with its centre at $2T$, where T is the pulse length. With this method, different lags of the autocorrelation function could in principle be measured by transmitting pulse pairs with different separations. Then each separation would give a different lag but the shapes of the ambiguity functions for all lags would be identical. The zero lag would be obtained by transmitting a single pulse.

The use of pulse pairs with variable separations would be a very ineffective way of observing the autocorrelation function. A more powerful method is

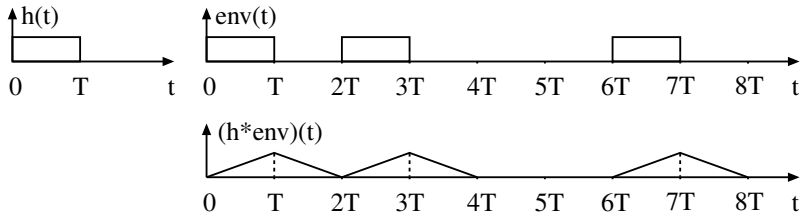

Figure 5.6: A three-pulse code and its convolution with a boxcar impulse response.

offered by multipulse codes, which are extensions of the two-pulse code. A multipulse code consists of a set of pulses with equal lengths, separated by gaps of odd multiples of the pulse length. The receiver impulse response is chosen to have the same length as the pulse. In our examples, which are designed for demonstration purposes, we use a boxcar-shaped impulse response. With a suitable choice of pulses and gaps in the modulation it is possible to measure not only a single lag but several lags of the autocorrelation function from a single transmitted pulse sequence. This greatly increases the efficiency of the modulation as compared to transmitting pulse pairs with variable separations.

The idea of a multipulse code is easily understood using a three-pulse code as an example. A three-pulse code is actually not useful in practical work since it gives only three lags of the autocorrelation function, but is serves well as an example. The most common multipulse codes contain five pulses. In a few cases, four-pulse codes have also been used.

The modulation envelope $\mathrm{env}(t)$ of a three pulse code is shown in Fig. 5.6 together with $(h * \mathrm{env})(t)$ calculated using a boxcar impulse response. We observe that $(h * \mathrm{env})(t)$ contains three triangles, the last of them being separated from the other two by a distance equal to the length of the triangle.

When samples are taken at intervals of $2T$, the sampling times are $t_i = 2iT$, $i = 0, 1, 2, 3, \ldots$ and non-zero ambiguity functions are obtained for lagged products of the form $x_i \cdot x_{i+1}^*$, $x_i \cdot x_{i+2}^*$ and $x_i \cdot x_{i+3}^*$, *i.e.* the first, second and third lag. For longer lags the ambiguity functions are zero. The generation of the range ambiguity functions for the three lags is shown in Fig. 5.7. The left hand panel shows $(h * \mathrm{env})(t_i - S)$ corresponding to the data sample x_i and the next three panels the same functions for the next three samples. The change of $(h * \mathrm{env})(t - S)$ from sample to sample is obtained by moving the set of three triangles upwards by $2T$, which is the length of a triangle. The result is that only a single triangle in $(h * \mathrm{env})(t_i - S)$ overlaps one of the triangles in $(h * \mathrm{env})(t_{i+1} - S)$, $(h * \mathrm{env})(t_{i+2} - S)$ or $(h * \mathrm{env})(t_{i+3} - S)$. Because the range ambiguity functions are obtained as products of the $(h * \mathrm{env})$-terms, the ambiguity functions for the three lags contain only a single peak as shown by the three panels on the right hand side of Fig. 5.7.

Note that, although associated with the same sample x_i, the three range ambiguity functions in Fig. 5.7 are not at the same altitude, but the ambiguity

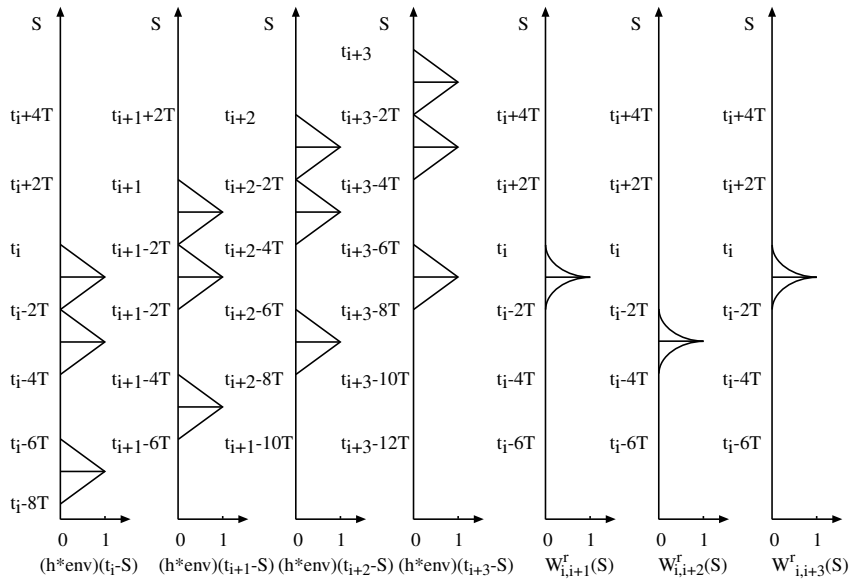

Figure 5.7: Generation of range ambiguity functions of a three-pulse code.

function of the second lag lies at a lower height than the other two. However, when the three lag profiles are calculated using each data sample as the first factor of $x_i \cdot x_j^*$, all three range ambiguity functions are obtained at each height, except at the very bottom and top of the profiles. From Fig. 5.7 one can easily see that the lagged product $x_{i+1} \cdot x_{i+3}^*$ gives a second-lag ambiguity function at the same altitude as the first and third lags plotted in the figure. Fig. 5.7 also shows why lags longer than three are not obtained for this code. This is because the bottom triangle of $h * \text{env}$ for all samples after x_{i+3} would lie at greater heights than the top triangle for x_i and, consequently, the product of the $h * \text{env}$-terms would be zero. The range ambiguity function for zero lag, of course, would consist of three peaks and therefore it has a range ambiguity much in the same way as the two-pulse code in Chapter 4.

The three-pulse code is built in such a way that only one pair of pulses overlaps simultaneously when the code is shifted with respect to itself in steps of $2T$. Longer pulse codes can be constructed using the same principle. The separations of the leading edges of the successive pulses in Fig. 5.6 are 1 and 2 in units of $2T$ and therefore a brief notation (1,2) is sometimes used of the code. The possible pulse separations obtained from the code are 1, 2 and $1 + 2 = 3$. This fulfils the condition that the same pulse separation should not appear twice in the code. If that happens, a range ambiguity appears. Since it is only the pulse separations which matters, the pulses can also be transmitted in the opposite order. In the case of the three-pulse code, the second possibility is (2,1).

The altitudes where the autocorrelation functions are determined are usually called 'gates' or 'range gates'. A gate has a centre point and a width determined by the range ambiguity functions. In the case of multipulse codes the centre points and widths of the gates are the same as those of the range ambiguity functions, but this is not necessarily true for other modulations. The idea of a gate is that lag estimates from the same gate are combined to make an autocorrelation function estimate which is used for determining the plasma parameters. Hence the height separation and the width of the gates determine the height resolution of the experiment. The gates are usually numbered in ascending order starting from the bottom of the profile.

In multipulse codes all range ambiguity functions have the same shape and therefore a single lagged product from each lag value is taken into a single gate. The lag profile matrix is a handy tool for showing how this is done. In the case of a three-pulse code we measure three side diagonals of the lag profile matrix as indicated in Fig. 5.8. Although it contains range ambiguities, the main diagonal is also often measured. In Fig. 5.7 we noticed that the lagged products $x_i \cdot x_{i+1}$, $x_{i+1} \cdot x_{i+3}$ and $x_i \cdot x_{i+3}$ give the three lags for the same altitudes. Therefore the grey-shaded elements of the lag profile matrix in Fig. 5.8 make a single range gate. The gate forms a fixed pattern which can be moved diagonally in the matrix. A single step down the diagonals means a shift of one

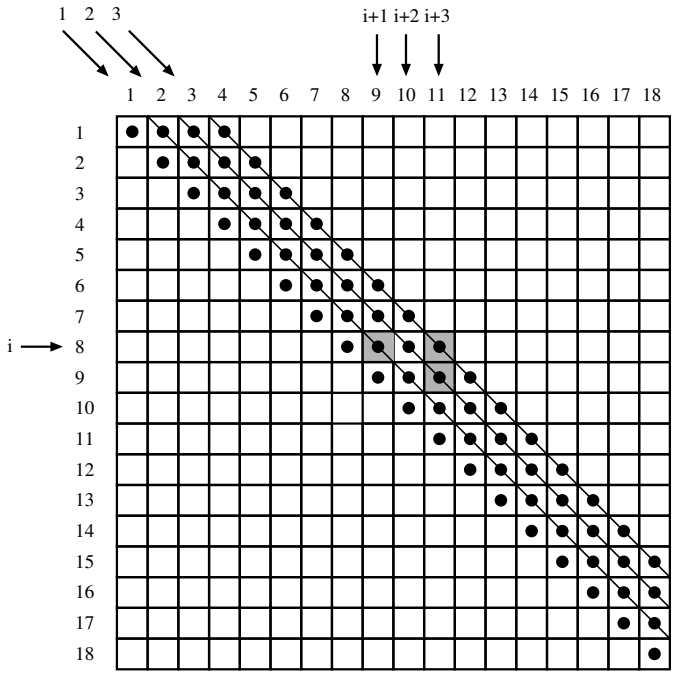

Figure 5.8: Lag profile matrix presentation of a gate in a three-pulse code.

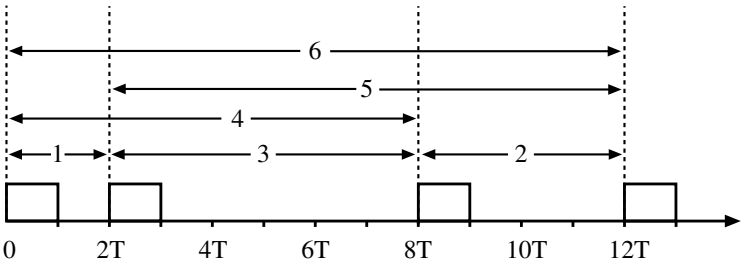

Figure 5.9: A four-pulse code in system (1,3,2) and the associated lags.

gate separation upwards along the radar beam and vice versa. Obviously, the pattern gradually glides out of the lag profile matrix at both ends so that there will be one incomplete gate at the bottom and two incomplete gates on the top of the profile.

A three-pulse code is not useful in practice, because it does not produce enough lags for fulfilling the demands discussed in Chapter 5.2. The number of lags can be effectively increased by adding the number of pulses in the code. The next possibility is a four-pulse code which can be constructed using the system (1,3,2). This code is shown in Fig. 5.9 together with the various lags it

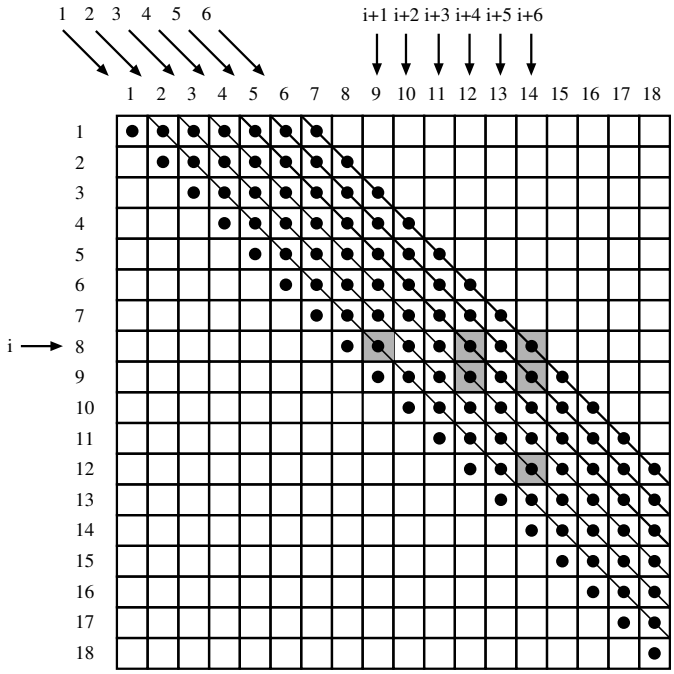

Figure 5.10: Lag profile matrix presentation of a gate in a four-pulse code.

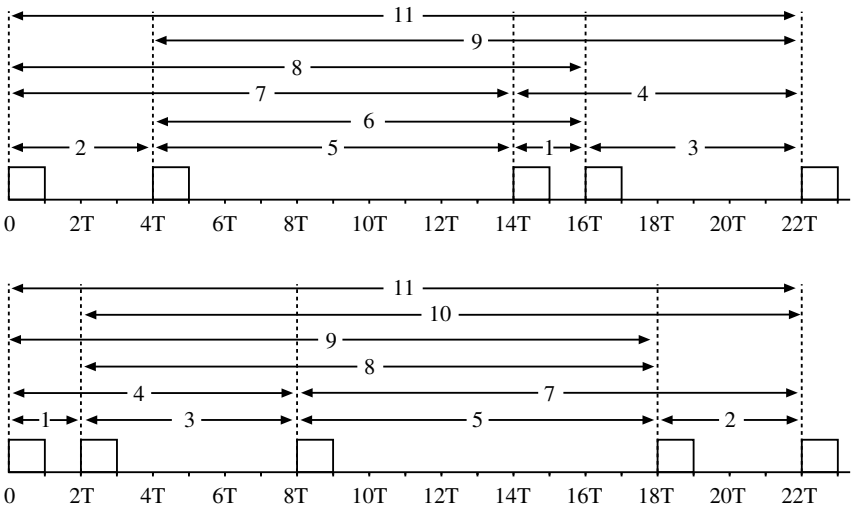

Figure 5.11: Five-pulse codes in systems (2,5,1,3) and (1,3,5,2) with the associated lags.

can produce. The lags are obtained as time differences of the leading edges of all pulse pairs and it is observed that the code is able to produce all lags from

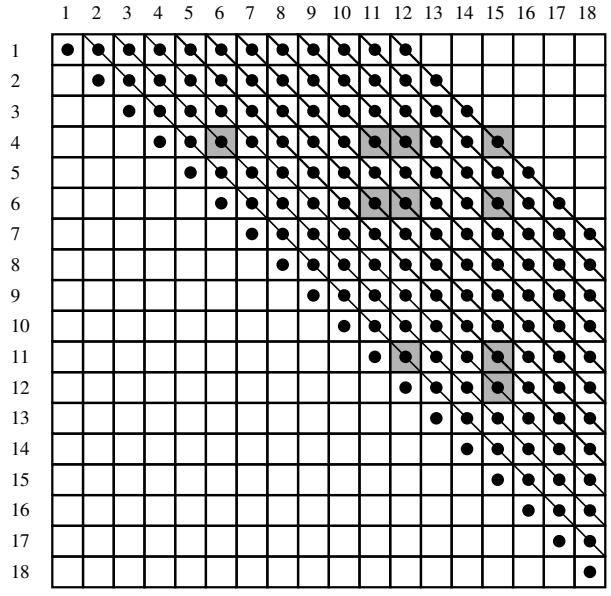

Figure 5.12: Lag profile matrix presentation of a gate in the case of the five-pulse code in system (2,5,1,3).

1 to 6. Hence, by adding the number of pulses by one, it was possible to double the number of lags. When the zero lag is measured using a separate short pulse, this number of lags is just about sufficient for fitting for instance the electron density, the ion velocity and the two temperatures.

The pattern of a single gate in the lag profile matrix is shown in Fig. 5.10. Because the first pulse takes part in lags 1, 4 and 6, the elements corresponding to these lags are found on the same row. The second pulse takes part in creating lags 3 and 5, and therefore the respective elements lie on the following row. Because the second lag is due to the third and fourth pulse and the separation of the second and third pulse is three units, the corresponding element is found on the third row below the previous two elements.

Finally, five-pulse codes using the systems (2,5,1,3) and (1,3,5,2) are shown in Fig. 5.11. The first one of these codes gives lags 1–9 and lag 11 so that lag 10 is missing and the second one lags 1–5 and 7–11 so that lag 6 is missing. Since the pulses can be sent in the opposite order, the mirror images (3,1,5,2) and (2,5,3,1) are also valid five-pulse codes. The gate pattern in the lag profile matrix is obtained in the same way as above. The pattern for the code (2,5,1,3) is shown in Fig. 5.12. Here we see that lags 2, 7, 8 and 11 are found on the same row, lags 5, 6 and 9 two rows further down, lags 1 and 4 still five rows downwards and, finally, lag 3 one row below the the previous one.

Because a pulse code is sparse, a single code does not use the duty cycle of a radar in an effective way. This fault can be avoided by interspersing pulse codes at different frequencies in such a manner that, during a gap in a single code, a second code is transmitted at another frequency. An example of this idea is presented in Fig. 5.13 where the five-pulse code (2,5,1,3) is transmitted at two frequencies and its mirror image (3,1,5,2) also at two frequencies. When the successive codes are started at intervals of T, no pulses overlap and there are only three gaps with a length $2T$ in the modulation pattern. These gaps can be filled by single pulses at other frequencies. The single pulses are useful in the determination of the zero lag.

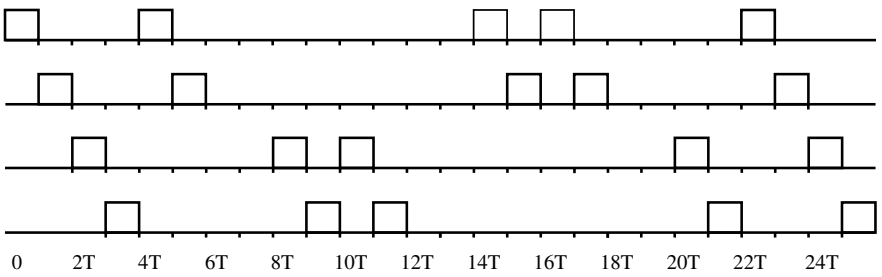

Figure 5.13: A modulation pattern containing four five-pulse codes transmitted at four frequencies.

5.4 Gating

The accuracy of a measured autocorrelation function can be improved by adding two or more neighbouring elements from a lag profile to make a single lag estimate. This means that a few elements from each lag profile are included in the same gate. The method is called gating. The effect of gating is to reduce the statistical error of the measurement while simultaneously reducing the height resolution.

Gating is demonstrated in Fig. 5.14 where the four-pulse code in Fig. 5.9 is used as an example. The difference between this and Fig. 5.10 is that the set of samples $\{x_i\}$ is divided into two subsets, $\{y_i\}$ and $\{z_i\}$ in such a way that every second term in $\{x_i\}$ belongs to $\{y_i\}$ and every second to $\{z_i\}$. Unprimed and primed numbers are used in Fig. 5.14 to separate the members of $\{y_i\}$ and $\{z_i\}$, respectively. Autocorrelation function estimates of lags 1–6 are calculated separately for the two data sets, which means that only every second side diagonal is calculated in the lag profile matrix. The gate patterns for both data sets have the same shape as in Fig. 5.10. These are indicated by shading in Fig. 5.14, where darker grey corresponds a gate for the set $\{y_i\}$ and lighter grey for $\{z_i\}$. By adding the two elements for each lag to make single lag estimates, an autocorrelation function estimate is obtained with increased statistical accu-

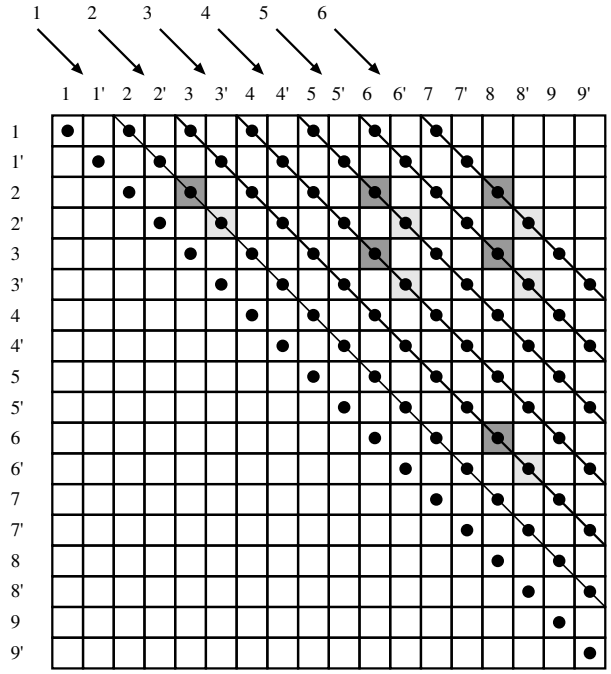

Figure 5.14: Gating of a three-pulse code with lag increment 1 and gating 1.

racy but reduced height resolution. It is, however, possible avoid the reduction of range resolution by using double sampling rate of the data.

In this example, the correlator programme only has to calculate every second side diagonal of the lag profile matrix. Then we say that the lag increment of the experiment is 2. This means that the step in lag domain is twice the sampling interval. Another property of the experiment is called gating. Gating has a numerical value which is equal to the number of lag profile elements added to a single gate minus one. In the present example gating is 1. If no gating is done, gating is zero. By varying lag increment and gating, one can construct very different experiments using the same transmitter modulation. Gating is not only associated with pulse codes but it can also be used in connection with other modulations. It is, however, a rather exceptional technique in experiment design and therefore it is no more discussed in the continuation.

5.5 Barker codes

Amplitude and phase modulation are two important modulation methods in radio transmissions. In these techniques the signal is presented in terms of variations in the amplitude or phase of a carrier wave. The modulation methods used in incoherent scatter radars can also be classified in this way. The pulse codes discussed above can be considered as amplitude modulation with only two amplitude values, zero and one. The principle of phase modulation is applied by phase codes, which alter the phase of the carrier wave according to a certain pattern.

In phase codes the transmitted pulse is divided into a number of elements and the phase of the carrier wave has a fixed value within each element. Two phase values, zero and 180^o, are used in practice, and therefore the elements are called bits. A sudden phase jump takes place at the boundary of elements with different phases. If the carrier waveform is $\cos \omega t$, the waveform within each element of zero and 180^o phase is $\cos \omega t$ and $\cos(\omega t + \pi) = -\cos \omega t$, respectively. Therefore phase 180^o can be interpreted as multiplication of the carrier by -1 and the modulation envelope of such a phase code gets values of $+1$ and -1.

The first phase codes used in incoherent scatter work were Barker codes which, however, were not originally constructed for this purpose. The Barker codes have a property that their autocorrelation functions have a triangular peak around zero, surrounded symmetrically on both sides by a set of smaller triangular peaks of equal height. If a Barker-coded signal is filtered by an impulse response which is the mirror image of the transmission envelope, the result is that both the range and lag ambiguity functions contain a sharp peak with a width corresponding to the length of a single modulation element. The values of the lag profile estimates also nearly correspond to the total energy of the transmitted pulse. Hence the modulation works approximately as if the transmitted pulse were compressed to the length of a single modulation element conserving the total energy and thus increasing the transmitting power. This is the reason why the method is sometimes called the pulse compression technique.

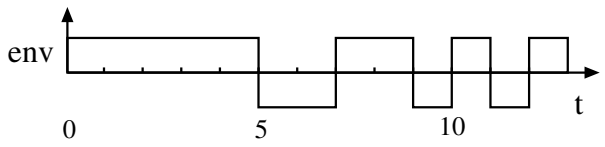

Figure 5.15: Modulation envelope of the 13-bit Barker code.

The known Barker codes up to the length of 13 bits are shown in Table 5.2. Here the phases zero and 180° are indicated by + and −, respectively. In the EISCAT radars, the 13-bit Barker code is commonly used. Its modulation envelope is plotted in Fig. 5.15.

A Barker-coded signal is filtered in the receiver with a special filter known as the Barker decoder. Since this filter is closely associated with the transmitted envelope, it is sometimes also called the matched filter. As explained later, a Barker decoder actually operates with digital data samples but the idea can be explained in terms of analogue signal processing.

Table 5.2.
Modulation envelopes of Barker codes of various lengths.

length/bits	envelope
2	+− or ++
3	++−
4	++−+ or +++−
5	+++−+
7	+++−−+−
11	+++−−−+−−+−
13	+++++−−++−+−+

If the envelope of a Barker-code is env(t), the impulse response of the Barker decoder is

$$h(t) = \mathrm{env}(-t). \tag{5.3}$$

Then

$$(h * \mathrm{env})(\tau) = [\mathrm{env}(-t) * \mathrm{env}(t)](\tau) = \int \mathrm{env}(\tau + t)\mathrm{env}(t)dt = R_{\mathrm{env}}(\tau), \tag{5.4}$$

so that the range ambiguity function is

$$W^r_{tt'}(S) = (h * \mathrm{env})(t - S) \cdot (h * \mathrm{env})(t' - S) = R_{\mathrm{env}}(t - S) \cdot R_{\mathrm{env}}(t' - S). \tag{5.5}$$

The autocorrelation function or $(h * \mathrm{env})(\tau)$ for the 13-bit Barker code is shown in Fig. 5.16. This result is easily obtained by letting a copy of the modulation envelope in Fig. 5.15 to glide on the time axis, multiplying the two

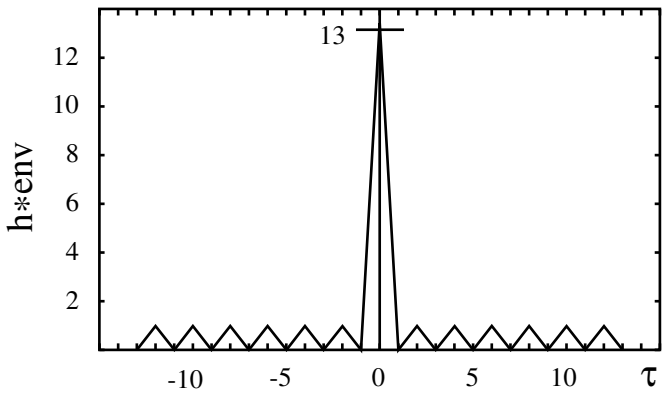

Figure 5.16: The autocorrelation function (equal to $h * \mathrm{env}$) for the modulation envelope of the 13-bit Barker code.

envelopes and observing the integral of the resulting function. When the copy completely overlaps the original envelope (zero lag), the multiplication gives the square of the envelope. This is a rectangle with a height of one unit and length of 13 units so that the value of the integral is 13. When the envelope copy is shifted to either direction, the integral decreases linearly to zero, which is reached at a time shift equal to one bit length. At this stage the multiplication produces a function which is equal to $+1$ during the first four time units, then alternately -1 and $+1$ during the next four time units and, finally, -1 during the last four time units. Thus the positive and negative areas cancel and the integral is zero. Continuing in the same manner, a sawtooth pattern is formed on both sides of the main peak with zeros and peak values of unity at odd and even multiples of the bit length, respectively. The small triangles are sometimes called side bands or side lobes of the autocorrelation function.

If the transmission consists of a single pulse modulated by a Barker-code, the range ambiguity function for zero lag is

$$W_{00}^{r}(S) = R_{\mathrm{env}}(0 - S) \cdot R_{\mathrm{env}}(0 - S) = [R_{\mathrm{env}}(S)]^2. \tag{5.6}$$

The last equality is valid because the autocorrelation function is even. Hence the range ambiguity function is given by the square of the autocorrelation function of the modulation envelope. In the multiplication all straight lines in the autocorrelation function will produce curves of parabolic shape. In the case of the 13-bit code, the central peak produces a spike with a height of 169 units and the smaller triangles create peaks with heights of unity, six of them on each side of the main peak. Also here the small peaks are sometimes called side lobes or side bands of the ambiguity function. The result is plotted in Fig. 5.17 using two different scales. The left hand panel shows the central peak and the right hand panel a magnified view revealing the structure of the side bands. The intensity ratio of the main peak and a side lobe is $13^2/1 = 169$ which is equal to $\log(13/1)^2 = 22.3$ dB.

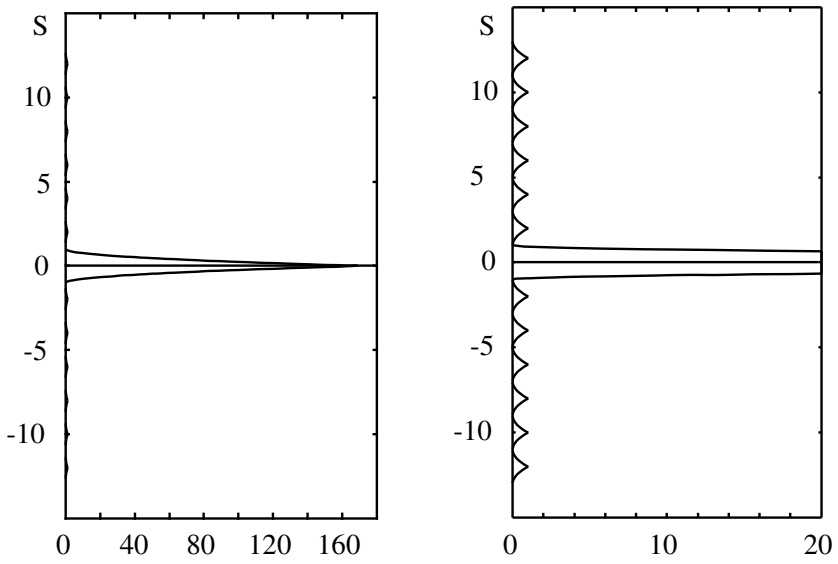

Figure 5.17: The range ambiguity function for zero lag of the 13-bit Barker code.

As indicated by eq. (4.16), the received power at a given instant of time is obtained as a sum of powers from the heights determined by the range ambiguity function of the zero lag. The power from each height is determined by both the electron density and the value of the ambiguity function. Hence, in the absence strong electron density gradients, the range ambiguity function in Fig. 5.17 indicates that almost all the received power must come from the main peak, *i.e.* from a height range corresponding to the length of a single bit in the 13-bit code. Only a small fraction comes from altitudes covered by the side bands.

Consider a fictitious experiment in which a single pulse with a length equal to that of a single bit would be transmitted with a 13-fold power and filtered with a boxcar-shaped impulse response of the same length. The range ambiguity function of this modulation would be equal to the main peak in Fig. 5.17 and the received power the same as the power from the main peak. Hence our 13-bit modulation works almost in the same way as this short pulse; only the small power from the side bands makes a small difference. Thus the Barker-coded experiment works as if a single pulse with a length of 13 bits were compressed into a length of a single bit conserving the energy and hence increasing the power to a 13-fold value. This is the explanation of the term 'pulse compression technique' discussed above.

One should notice that the above discussion is only valid if the electron density within the main peak is of the same order of magnitude as within the side bands. In some extreme situations it is possible that the electron density within several sidebands is about ten times higher than within the main peak.

Then the power emerging from the range of the side bands is not negligible and the idea of Barker-coding breaks down. The error thus created, however, can be corrected in the data analysis by a method called 'side band correction'.

As seen later, the digital filtering in the Barker decoder implies that data samples are taken at intervals of one bit length. The zero-lag range ambiguity function for each sample has the same shape as that shown in Fig. 5.17 and the separation of the main peaks for the successive samples is equal to a single bit length on the S-axis. Hence the height resolution obtained by a Barker-coded measurement corresponds to the bit length of the code. If, for example, the bit length is 4 μs, the height resolution is

$$\Delta h = \frac{3 \cdot 10^8 \,\mathrm{m/s} \cdot 4 \cdot 10^{-6} \,\mathrm{s}}{2} = 600 \,\mathrm{m}.$$

This shows that Barker-coding drastically improves the height resolution in incoherent scatter experiments.

For understanding the working principle of the Barker code, it is useful to study its amplitude ambiguity function

$$W_t^A(\tau, S) = h(t-\tau)\mathrm{env}(\tau - (S(\mathbf{r})) = \mathrm{env}(\tau - t)\mathrm{env}(\tau - (S(\mathbf{r})). \qquad (5.7)$$

This function is shown for the 13-bit Barker code in Fig. 5.18. We see that $W_t^A \neq 0$ only in a parallelogram-shaped area, which is split into $13 \times 13 = 169$ smaller parallelograms. In each of them the value of W_t^A is either $+1$ or -1; this is indicated by the pluses and minuses in the figure and the positive regions are shaded. A data sample $z_h(t)$ consists of voltages caused by elementary scattering signals $dz(\tau, \mathbf{r})$ from different heights and times, weighted by the amplitude ambiguity function as indicated by eq. (4.6). The elementary signals from a given height come from time intervals shorter than 13 bits. If this time interval is short enough, the plasma fluctuation does not change much and the elementary signal dz from each height also remains unchanged. Therefore dz can be taken as constant at a given height in eq. (4.6). Thus the impulse response calculates the total signal from this height by summing and subtracting a constant value of the elementary signal dz as indicated by the corresponding sign sequence at constant S in Fig. 5.18. When considering only the S values of the apices of the small parallelograms, we observe that the sums will be either zero or $+1$ times the elementary signal, except for a single height which gives a factor of $+13$. Hence the combined effect of the Barker modulation and the corresponding matched filter is to add and subtract signals from all heights and times in such a manner that they nearly cancel except at the height of the main peak of the range ambiguity function.

Because the idea of Barker codes is to add and subtract the amplitudes of elementary signals, we say that the Barker codes work in amplitude domain. There are also phase codes which are designed for adding and subtracting autocorrelation function estimates (powers) rather than amplitudes. We say that these codes work in power domain. A necessary condition for a phase code working in amplitude domain is that the plasma autocorrelation function is

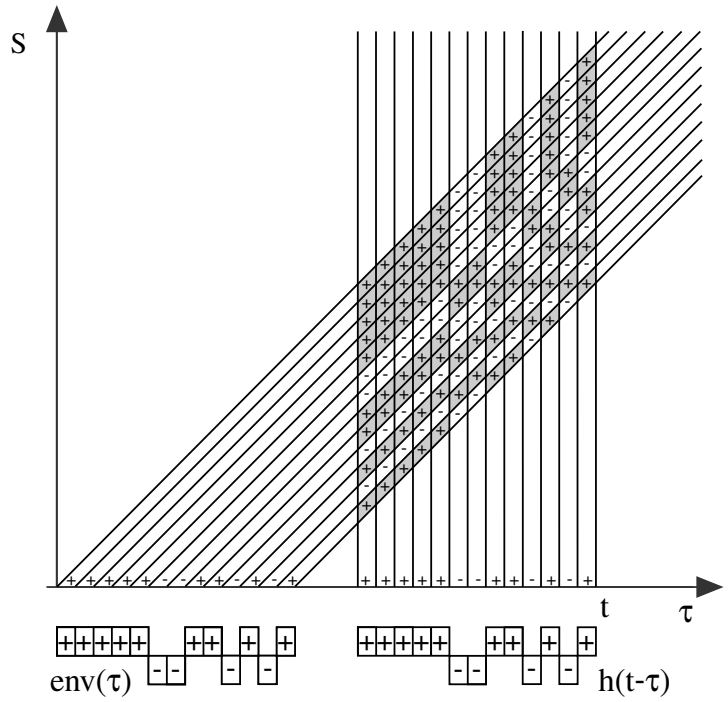

Figure 5.18: The amplitude ambiguity function of the 13-bit Barker code.

practically constant during the length of the impulse response of the matched filter. If this is not the case, the matched filter will add and subtract different values of elementary signals and the principle of Barker coding breaks down.

An incoherent scatter signal always contains random noise. Since the noise changes rapidly, its amplitudes do not cancel in the matched filter in the same way as the amplitudes of the elementary signal do. This necessarily increases the noise level of the experiment. The corresponding fictitious experiment with a compressed pulse has no elementary signals from altitudes outside the main peak of the range ambiguity function, and the associated noise is also missing.

The amplitude ambiguity function in Fig. 5.18 can be used for calculating the two-dimensional ambiguity function. The result is shown in Fig. 5.19, in terms of a three-dimensional plot in the top panel and in grey scale with contours in the bottom panel. It is enlightening to compare this with the two-dimensional ambiguity function of a simple pulse in Fig. 4.2. Obviously, Barker-coding a single pulse leads to a complicated deformation of its pyramid-shaped ambiguity function.

The above discussion only involves the measurement of power using a Barker coded modulation, although our aim is to measure the whole signal autocorrelation function with a suitable lag resolution. In pulse codes this was carried out by measuring lagged products of the form $x_i \cdot x_{i+1}^*$, $x_i \cdot x_{i+2}^*$, etc. If the

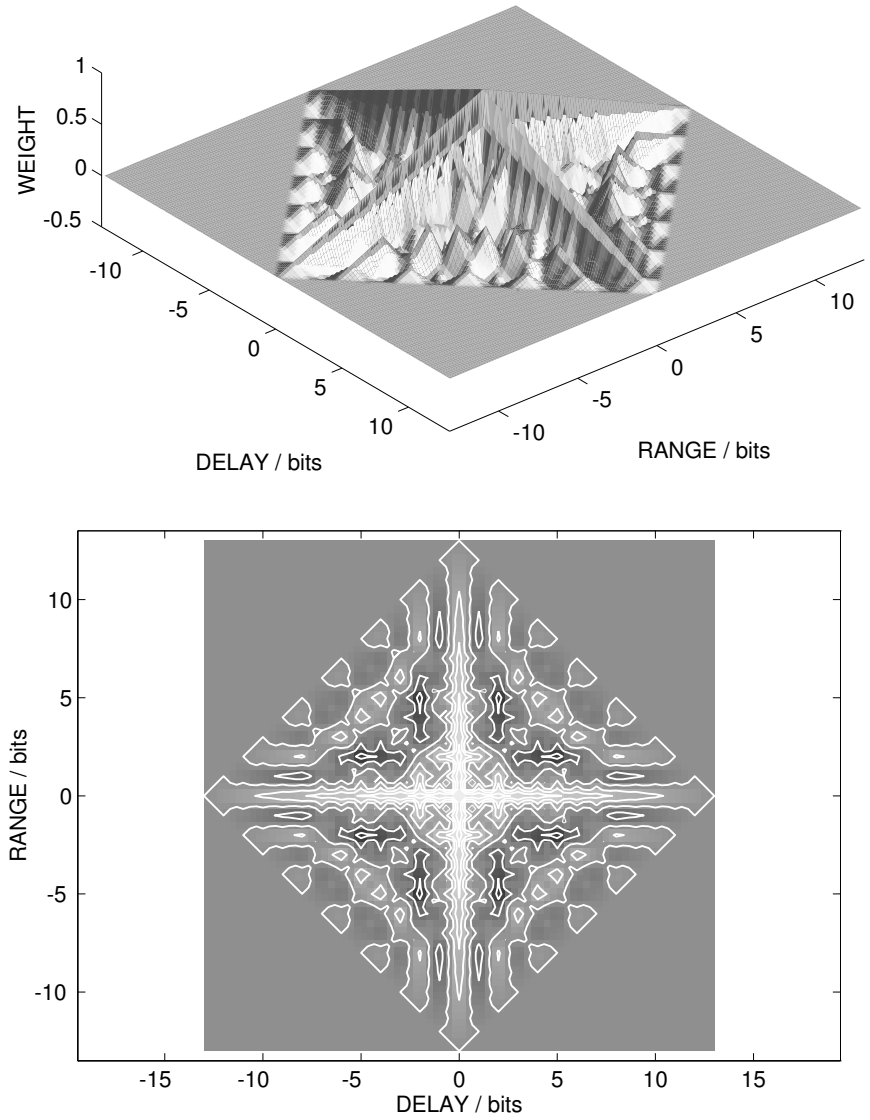

Figure 5.19: The two-dimensional ambiguity function of the 13-bit Barker code.

same is carried out for Barker-coded data samples, it will be found that the range ambiguity functions of lags longer than zero will not be composed of a single peak and a set of very small sidebands, but strong ambiguities will appear. Furthermore, the lags thus determined are not long enough. The longest lag obtained for a 13-bit code, for instance, is 12 times the bit length. If the bit length is 4 μs, which is needed for a 600 m height resolution, the longest lag obtained is only 48 μs. According to the estimates made in Chapter 5.2, this is not sufficient for a proper sampling of the signal spectrum.

5.6 Barker-coded pulse codes

The problem of meeting the demand of long lags and still maintaining the high altitude resolution offered by a Barker code can be solved by using pulse codes in which each pulse is modulated by a Barker code. In this combination of two modulations the height resolution will be determined by the Barker code and the lag resolution by the pulse code.

In the ordinary pulse codes introduced in Chapter 5.3 the time separation of the pulse front ends is equal to an even number of pulse lengths. This is necessary since, in calculating $h * \text{env}$, the pulse lengths are doubled by the effect of the impulse response of the post-detection filter (see Fig. 5.6). In the case of Barker codes the width of the main peak in $h * \text{env}$ is equal to two bit lengths rather than two pulse lengths, which leads to the fact that all pulse separations can be reduced to one half of their values in ordinary pulse codes.

Following the line of thought in Chapter 4, a two-pulse code is taken as an example for demonstrating the working principle of Barker-coded multipulse codes. The modulation envelopes of ordinary and Barker-coded two-pulse codes are shown in Fig. 5.20, where the complicated amplitude of a Barker code is briefly denoted by 'B'. In addition to phase coding, the difference in the modulations is that the Barker-coded modulation contains no gap between the two pulses; in fact it only makes sense to talk about two pulses because of the repetition of the phase code.

The convolution of the impulse response and the modulation envelope is easily obtained in the same way as for a single pulse in Chapter 5.5. The result for a 13-bit Barker code is plotted in Fig. 5.21. Obviously, each of the pulses produces a function of a shape similar to that in Fig. 5.16 and the total function is their sum. The two main peaks, however, do not overlap but their distance is

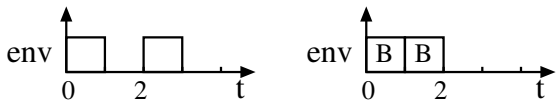

Figure 5.20: Comparison of modulation envelopes of an ordinary two-pulse code (left) and a Barker-coded two-pulse code (right). The envelope of the Barker code is denoted by 'B'.

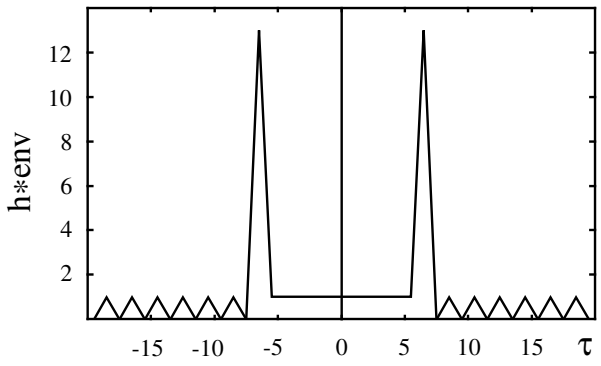

Figure 5.21: The convolution of the impulse response and modulation envelope for a two-pulse code modulated by a 13-bit Barker code.

equal to the pulse separation, *i.e.* 13 bit lengths. This separation is such that the sum of the side bands is constant between the main peaks. Since the side bands of one main peak reach only up to the other main peak, the side bands outside the double-peak structure remain untouched and consist of six triangles just like in Fig. 5.16.

The range ambiguity function of zero lag, given by the square of $h * \text{env}$, is shown in Fig. 5.22. This function contains two peaks with a height of 169 units

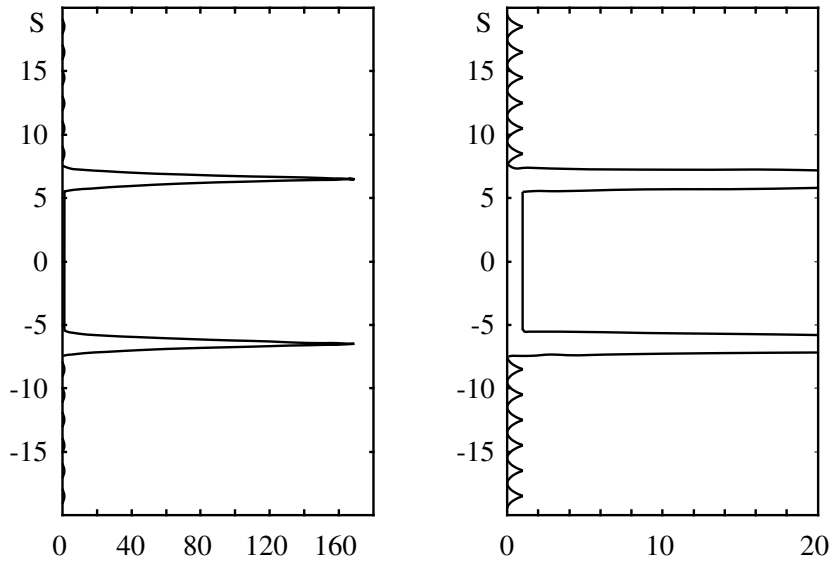

Figure 5.22: The range ambiguity function for zero lag of a two-pulse code modulated by a 13-bit Barker code.

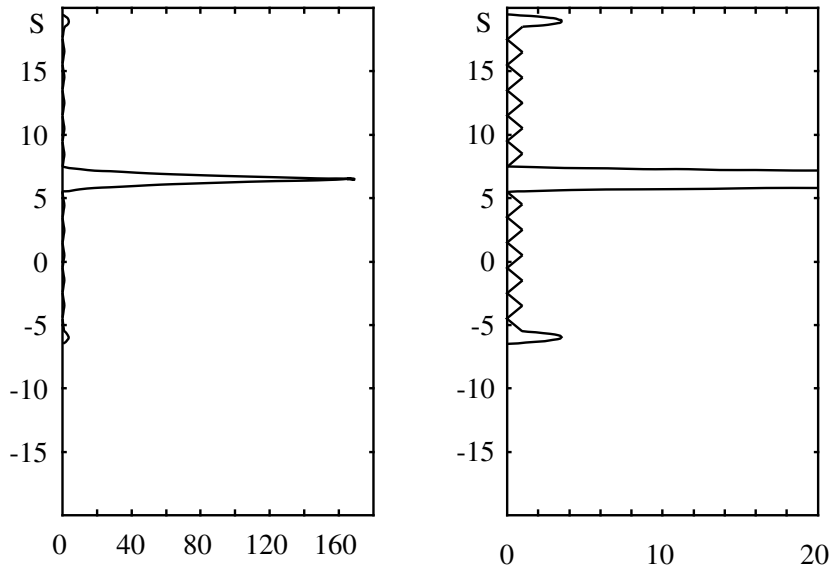

Figure 5.23: The range ambiguity function for first lag of a two-pulse code modulated by a 13-bit Barker code.

and side bands with heights of a single unit. The double-peak structure is an indication of range ambiguity of the zero lag, which is analogous to the range ambiguity in Fig. 4.6. This means that, in order to obtain a zero lag estimate with no range ambiguity, a Barker-coded single pulse must be transmitted in the same way as in ordinary multipulse experiments.

The first lag is obtained from data samples taken at intervals equal to the pulse separation. Because the sampling interval is equal to one bit length, the first lags for a 13-bit Barker code are given by lagged products of the form $x_i \cdot x_{i+13}^*$. The corresponding range ambiguity function (Fig. 5.23) is obtained by inserting two copies of $h * \text{env}$ on the S-axis with a separation of 13 units and multiplying them. Then the lower peak of one of the functions overlaps the upper peak of the other and a single main peak with a height of 169 units is formed by the product. The main peak is surrounded by side bands which consist of five triangles and higher peaks. The triangles are generated at a region where the constant values of $h * \text{env}$ overlap the triangles and the higher peaks where one of the main peaks in $h * \text{env}$ partially overlaps the last triangle in the side band structure.

Longer Barker-coded multipulse sequences can be constructed with the same principle. A Barker-coded multipulse sequence is similar to the corresponding ordinary multipulse code except that the separations of the front ends of the pulses are divided by two. This means that the same number of lags will be obtained, but the lag separations and the length of the measured autocorrelation function will be halved. The Barker-coded modulations corresponding to those

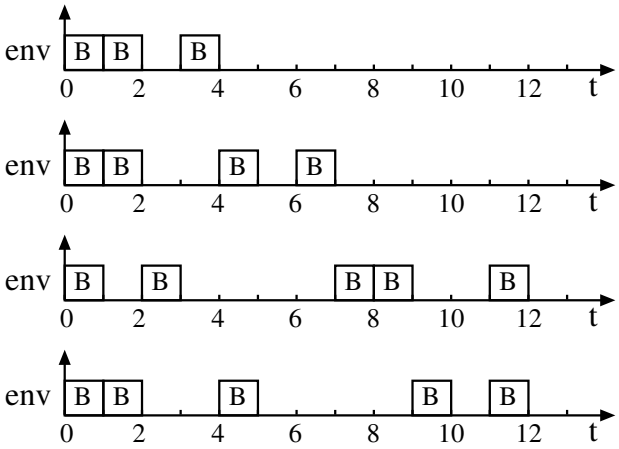

Figure 5.24: Modulation envelopes of different Barker-coded multipulse codes.

in Figures 5.6, 5.9 and 5.11 are shown in Fig. 5.24.

A practical point in designing Barker-coded experiments is that a gap of one bit length is included in each pulse. This gap is sometimes called the 'zero bit'. Hence, in effect, a single pulse in a 13-bit experiment takes 14 bits in time. The lengths of the gaps in the code are also increased to correspond to the increased 'pulse length'. The reason for this practice is that, due to the effect of filtering, the correlation length of the noise in the signal is increased, and the first lag will be contaminated by noise. In the case of a low signal-to-noise ratio the contamination may be serious. Adding the zero bit will lead to a lag increment just sufficient to give a first lag longer than the correlation length of the filtered noise.

As discussed in Chapter 3.11, the bandwidth of the post detection filter must be fitted to the bandwidth of the modulation envelope. If the bit length is 4 μs, for instance, the bandwidth of the filter should be

$$\nu_c = \nu_{\text{env}} = \frac{1}{2 \cdot 4\ \mu s} = 125 \text{ kHz}.$$

Without phase coding the bandwidth of the same pulse would only be

$$\nu_{\text{env}} = \frac{1}{2 \cdot 13 \cdot 4\ \mu\text{s}} \approx 10 \text{ kHz}.$$

This means that the use of Barker coding extends the filter pass band into a value much wider than the pass band of the incoherent scatter spectrum. The power of noise in the filtered signal, on the other hand, is proportional to the filter band width, and therefore the use of Barker coding necessarily reduces the signal-to-noise ratio of the experiment.

The altitude resolution is determined by the bit length of the Barker code and, after the bit length is fixed, the lag resolution is given by the number of bits in the code. Finally, the selected pulse code determines the number of lags and the length of the measured autocorrelation function. As shown in Chapter 5.5, a bit length of 4 μs leads to a height resolution of 600 m. A 13-bit Barker code with this bit length gives a lag increment of 14×4 μs $= 56$ μs and, if a four-pulse code is used, the longest lag is 6×56 μs $= 336$ μs. According to the estimates made in Chapter 5.2, this is a reasonable sampling of the signal autocorrelation function in the case of the EISCAT UHF radar. In fact, many experiments with 600 m height resolution have been constructed for studying the ionospheric E region using Barker-coded pulse codes.

5.7 Barker decoder

When a signal is filtered by two linear filters connected in series as shown in Fig. 5.25, the output of the first filter is

$$y(t) = (h_1 * x)(t) \tag{5.8}$$

and that of the second filter is

$$z(t) = (h_2 * y)(t) = [h_2 * (h_1 * x)](t). \tag{5.9}$$

The Fourier transforms of these equations are

$$Y(\nu) = H_1(\nu) \cdot X(\nu) \tag{5.10}$$

and

$$Z(\nu) = H_2(\nu) \cdot Y(\nu) = H_2(\nu) \cdot H_1(\nu) \cdot X(\nu) = H(\nu) \cdot X(\nu) \tag{5.11}$$

Here the total transfer function of the filter system is

$$H(\nu) = H_1(\nu) \cdot H_2(\nu). \tag{5.12}$$

The total impulse response of the filter system is obtained by an inverse Fourier transform of eq. (5.12) and the result is

$$h(t) = \mathcal{F}^{-1}\{H\} = \mathcal{F}^{-1}\{H_1 \cdot H_2\} = \mathcal{F}^{-1}\{H_1\} * \mathcal{F}^{-1}\{H_2\} = (h_1 * h_2)(t). \tag{5.13}$$

When more than two filters are connected in series, analogous equations are valid. In an incoherent scatter radar the post detection filter and the Barker

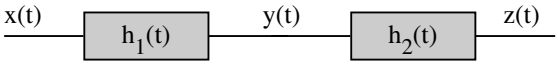

Figure 5.25: Two filters connected in series.

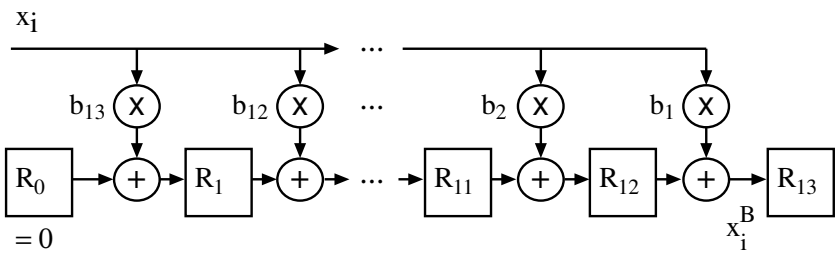

Figure 5.26: A 13-bit Barker decoder.

decoder are connected in series and the total impulse response of the resulting filter is a mirror image of the envelope of the Barker code.

Unlike the post detection filter, the Barker decoder works in digital domain. It receives the samples from AD converter and calculates their weighted averages. The number of samples included in the weighted sum is equal to the number of bits in the Barker code. In the case of a 13-bit code, the ith sample in the output of the Barker decoder is

$$x_i^{\mathrm{B}} = \sum_{k=1}^{13} b_{\mathrm{k}} x_{i-k+1}, \tag{5.14}$$

where the Barker coefficients b_{k} are

$$b_k = \begin{cases} 1 \text{ , when } k = 1, 3, 5, 6, 9 - 13 \\ -1 \text{ , when } k = 2, 4, 7, 8. \end{cases} \tag{5.15}$$

These numbers obviously give the values of the modulation envelope in the 13 bits of the Barker code in the inverse order.

A Barker decoder is constructed from multipliers, adders and shift registers. A circuit diagram for a 13-bit Barker decoder is shown in Fig. 5.26. When a digital sample x_i enters the decoder, it is multiplied separately by all Barker coefficients, each product is added to the contents of the previous register in the shift register line and the result is stored to the next register. As a result, the filtered sample x_i^{B} calculated according to eq. (5.15) will be transferred to the the last register R_{13}. It is clear that no meaningful output values are obtained before at least 13 samples have been fed to the input.

Barker decoder is a filter, which calculates weighted averages from a finite number of samples. Because the length of the impulse response of such filters is finite, they are called finite impulse response filters or FIR filters.

Let us finally show that the combined impulse response of the post detection filter and the Barker decoder is really a mirror image of the Barker code. The queue of data samples given by the AD converter can be presented using the Dirac comb in the form

$$x_i = \int_{-\infty}^{\infty} z_h(\tau)\delta(\tau - i)d\tau, \tag{5.16}$$

where z_h is the complex output signal of the post detection filter and the sampling interval is taken as the time unit. The Barker decoder calculates the weighted averages

$$x_i^{\mathrm{B}} = \sum_{k=1}^{13} b_k \cdot x_{i-k+1} = \sum_{k=1}^{13} b_k \int_{-\infty}^{\infty} z_h(\tau)\delta[\tau - (i-k+1)]d\tau$$

$$= \int_{-\infty}^{\infty} z_h(\tau) \sum_{k=1}^{13} b_k \delta[k-1-(i-\tau)]d\tau. \tag{5.17}$$

Using the notation

$$h_{\mathrm{B}}(t) = \sum_{k=1}^{13} b_k \delta(k-1-t) \tag{5.18}$$

this can be put in the form

$$x_i^{\mathrm{B}} = \int_{-\infty}^{\infty} z_h(\tau) h_{\mathrm{B}}(i-\tau)d\tau = (z_h * h_{\mathrm{B}})(i). \tag{5.19}$$

This means that our digital samples x_i^{B} can be considered as values of a fictitious signal

$$x^B(t) = \int_{-\infty}^{\infty} z_h(\tau) h_{\mathrm{B}}(t-\tau)d\tau = (z_h * h_{\mathrm{B}})(t) \tag{5.20}$$

at times i. Hence the filtered samples are the same as those obtained by digitising the signal in eq. (5.20), *i.e.* $x_i^{\mathrm{B}} = x^{\mathrm{B}}(i)$.

Obviously, $h_{\mathrm{B}}(t)$ is the impulse response of the Barker decoder. According to eq. (5.13) the total impulse response is given by the convolution of the impulse

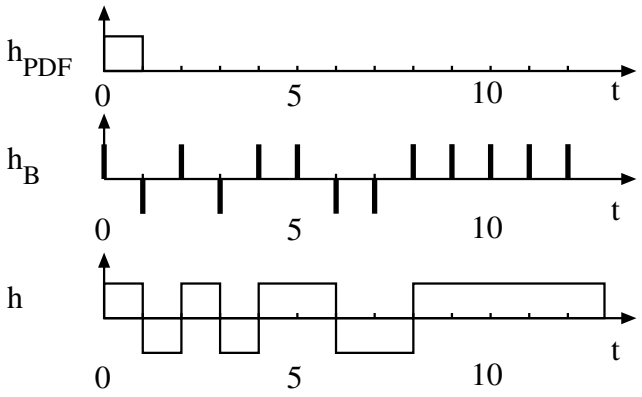

Figure 5.27: Convolution of the impulse responses of the post detection filter and the Barker decoder.

responses of the post detection filter and the Barker decoder. Thus the total impulse response is

$$h(t) = (h_{\mathrm{PDF}} * h_{\mathrm{B}})(t). \tag{5.21}$$

The calculation of this convolution is demonstrated in Fig. 5.27, where h_{PDF} is assumed to have a boxcar shape and a length equal to the bit length. The impulse response of the Barker decoder is a Dirac comb with positive and negative peaks. When the box is allowed to slide along the time axis, it overlaps only a single peak in the Dirac comb at the same time. Hence, depending on the sign of the Dirac peak, the convolution integral gives a value of $+1$ or -1 and the resulting impulse response is indeed a mirror image of the Barker envelope.

5.8 Long-pulse code

The simplest modulation which can be used in incoherent scatter measurements is a plain long pulse. Although the modulation is simple itself, some problems are associated with data sampling and use of the lag profile matrix. This leads to the fact that long-pulse experiments are in a sense more complicated than multipulse experiments are. Long-pulse experiments are mostly used in the F region, where a height resolution of 10–30 km is sufficient.

The modulation bandwidth of a long pulse is small so that the bandwidth of the post detection filter is mainly determined by the bandwidth of the incoherent scatter spectrum. Therefore the impulse response of the post detection filter is shorter than the pulse. This is a major difference as compared to the previous modulations. Another difference is that the sampling interval is shorter than the pulse length. This gives a possibility to measure several lag profiles using a single pulse.

A normal practice in a long-pulse experiment is to use a sampling interval equal to the length of the impulse response, which leads to independent data samples. The pulse length is also usually a multiple of the sampling interval. If the sampling interval is τ, the lag value of the lag profile matrix element $x_i \cdot x_j^*$ is

$$(j - i)\tau = k\tau$$

where $k = j - i$. According to eq. (4.19), the corresponding range ambiguity function is

$$W_{ij}^r(S) = (h * \mathrm{env})(i\tau - S) \cdot (h * \mathrm{env})^*(j\tau - S). \tag{5.22}$$

In a real long-pulse experiment the pulse length may be of the order of 300 μs and the length of the impulse response 20 μs (see Chapter 3.11) so that the ratio of these lengths may be about 15. In explaining the general principle of the method it is more convenient to use a shorter pulse length, and therefore a ratio of 5 or 10 is used instead in the following discussion. Plots of a boxcar-shaped impulse response, a modulation envelope and their convolution are shown in Fig. 5.28, where the pulse length is assumed to be five times the length of the impulse response. If the latter is equal to the sampling interval, the lengths of these functions are τ, 5τ and 6τ, respectively.

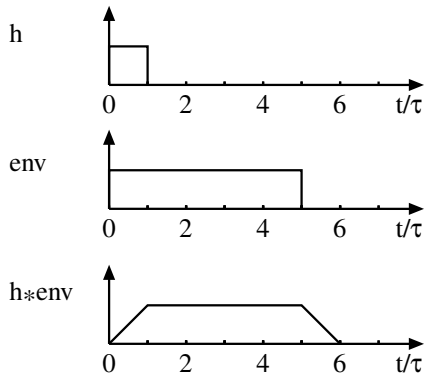

Figure 5.28: Convolution of the impulse response and modulation envelope of a simple long pulse.

The range ambiguity functions for different lag values can now be easily calculated using eq. (5.22). Their generation is presented in Fig. 5.29, where the functions $h * \mathrm{env}(t_i - S)$ for five successive sampling times are plotted in the left hand panel and, given by their products, the range ambiguity functions for lags 0–4 in the right hand panel. The range ambiguity function for the fifth lag would also be non-zero but it is so small that this lag is not measured. The ambiguity functions of all subsequent lags are zero.

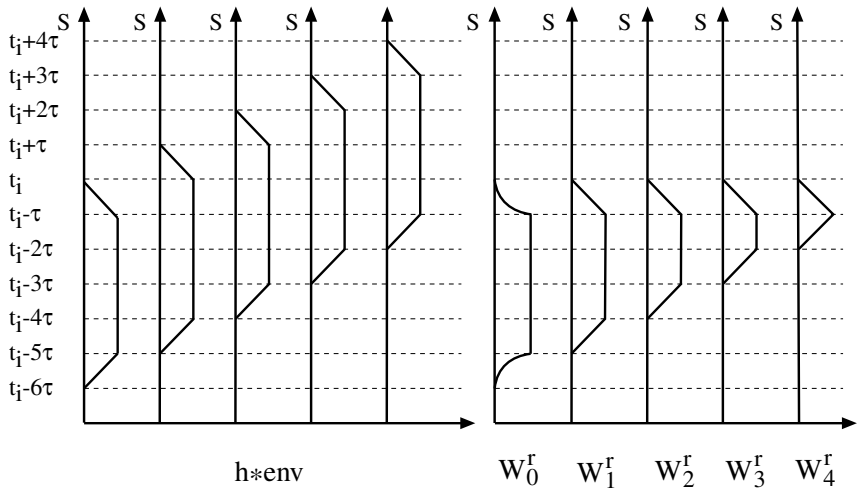

Figure 5.29: Generation of range ambiguity functions of a simple long pulse. Left panel: $h * \mathrm{env}(t_i - S)$ for five successive sampling times. Right panel: resulting range ambiguity functions for lags 0–4.

Fig. 5.29 reveals that, unlike in pulse codes, the range ambiguity functions of a simple long pulse have different lengths for different lags. The shorter lags give information from a wider altitude range so that their height resolution is worse, but on the other hand, they are more heavily weighted in the autocorrelation function than the long lags are. In this example the zero lag gives information from a height range corresponding to six sampling intervals, whereas the corresponding number for the fourth lag is two.

For general interest, the two-dimensional ambiguity functions for the long-pulse code are plotted in Fig. 5.30. We see that they are partially overlapping ridges elongated in the S-direction. The length of a two-dimensional ambiguity function is the same as that of the corresponding range ambiguity function in Fig. 5.29 and the width is equal to two time units.

The ambiguity functions shown above are properties of individual elements in the lag profile matrix. In a true long-pulse experiment more than one matrix element is usually included in the same gate, and therefore the ambiguity functions will also be different. The height resolution of a long-pulse experiment cannot be better than that of a matrix element for zero lag. If, for instance, the pulse length is 300 μs and the length of the impulse response is 20μs, the range resolution of the zero lag is

$$\frac{320\ \mu\text{s} \cdot 3 \cdot 10^8\ \text{m/s}}{2} = 48\ \text{km}.$$

One should also notice that the height separation is not the same as the height resolution. When the sampling rate is 20 μs, the separation of successive elements in a lag profile is

$$\frac{20\ \mu\text{s} \cdot 3 \cdot 10^8\ \text{m/s}}{2} = 3\ \text{km}$$

and the zero lag of each gate contains information from a 48 km height range.

In order to compare the weights of the lag profile elements $x_i \cdot x_j^*$, we approximate eq. (4.16) to give

$$\langle x_i \cdot x_j^* \rangle \approx \frac{c}{2} R P_{Az}^o \sigma_e(i-j, r_0) \int_{-\infty}^{\infty} W_{ij}^r(S) dS, \tag{5.23}$$

where r_0 is the range of the centre point of the ambiguity function W_{ij}^r. In this approximation the range ambiguity function is assumed to be so short that both the plasma autocorrelation function and the power P_{Az}^o can be taken as constants in calculating the integral in eq. (4.16). If the pulse is long, the assumption may be poorly valid especially at low altitudes.

The weight of each lag profile element relative to that of the zero lag can now be estimated using eq. (5.23). According to the principle shown in Fig. 5.29, the integral of the range ambiguity function W_{ij}^r for the lag $k\tau = (j-i)\tau$ is equal to $(n-k)\tau$, where $n\tau$ is the pulse length. This is not exactly true for the zero lag, because the ends of its ambiguity function have parabolic rather than linear shapes. Nevertheless, it is a reasonable approximation to use $n\tau$ for the

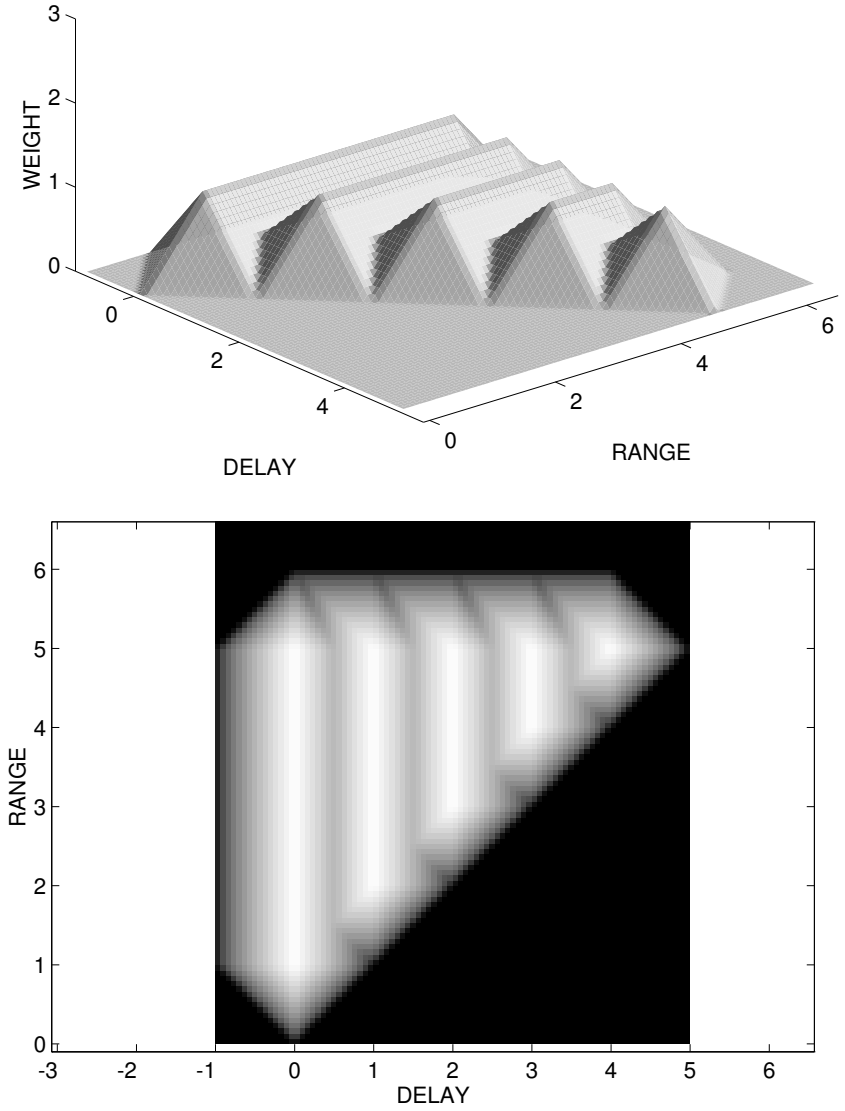

Figure 5.30: Two-dimensional ambiguity functions of a simple long pulse.

integral of the zero lag ambiguity function, especially if the pulse is long. Then the relative weight of lag $k\tau$ is

$$\mathcal{W}_k = \frac{n-k}{n}. \tag{5.24}$$

This result is used when estimating the relative weights in true long-pulse experiments where several lag profile matrix elements are summed to make a single gate.

In a classical long pulse experiment the signal autocorrelation function estimate for lag k in the ith gate is given by

$$R_z(i,k) = \sum_{j=i}^{n+i-k-1} \langle x_j \cdot x_{j+k}^* \rangle, \quad k = 0, 1, \ldots, (n-1). \tag{5.25}$$

The idea of this algorithm is to take a string of samples from a time interval equal to the pulse length (five samples in the case of the example in Figures 5.28 and 5.29), to calculate all possible lagged products and to add all those with the same lag value. The result is that the zero-lag estimate will be a sum of n elements from the main diagonal of the lag profile matrix. Similarly, $n-1$ elements from the first side diagonal will be added to make the first-lag estimate, etc. Hence a single gate will cover a triangular region in the lag profile matrix as shown in Fig. 5.31, where the pulse length is assumed to be five times the

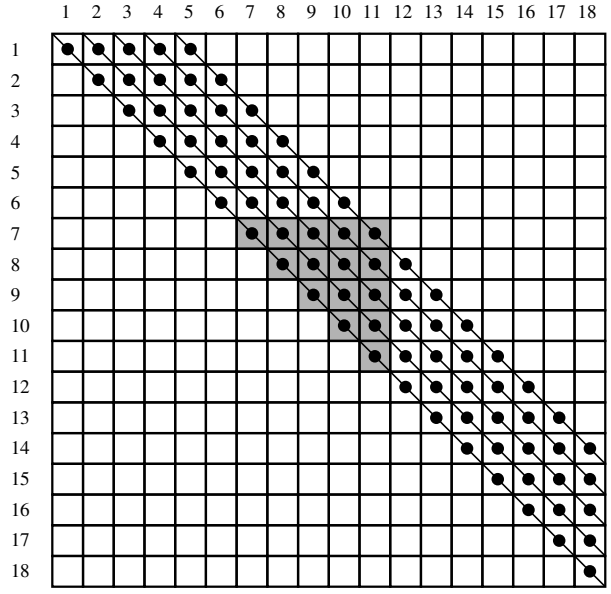

Figure 5.31: Generation of a single gate in a classical long-pulse experiment. The pulse length is assumed to be five times the sampling interval.

sampling interval. In this case five successive data samples are used for a single gate and lags 0–4 can be calculated from eq. (5.25). The next gate is obtained by shifting the triangular region diagonally by a fixed number of steps. If the shift is greater than a single step, some of the matrix elements at the longer lags are not utilised.

The classical long-pulse experiment dates back to the times when the concept of lag profile matrix was not known and the theory of ambiguity functions was not properly developed. It is a poor algorithm which is no more used. The method has two main drawbacks associated with the height ranges and relative weights of the different lags in a single gate. As shown by Fig. 5.29, the height ranges covered by the elements of the lag profile matrix are longest on the main diagonal and decrease with increasing lag number. The classical experiment increases this imbalance because it adds into a single gate more elements at short lag values than at long lags. The second point is that the classical experiment also amplifies the imbalance of the relative weights of the different lags. Therefore the long lags will contain more noise than the short lags do. This may be disastrous since the autocorrelation function is small at long lag values and it would be essential to measure it accurately there.

An obvious way of improving the experiment is to have more matrix elements in a single gate at long lag values than at short ones. In modern long-pulse experiments the autocorrelation function estimate for lag k in the ith gate is

$$R_z(i,k) = \sum_{j=i}^{i+k+v-1} \langle x_{j-k} \cdot x_j^* \rangle, \quad k = 0, 1, \ldots, (n-1). \tag{5.26}$$

According to this formula the number of matrix elements included in a single gate is equal to the lag number plus the volume index v, which determines the number of elements for zero lag. The generation of a gate is demonstrated in terms of the lag profile matrix by the shaded areas in Fig. 5.32, where two partially overlapping gates are shown. Here $n = 5$ so that lag profiles 0–4 can be calculated, and $v = 3$, which means that three elements are taken to make the zero lag. At the bottom part of the profiles a complete gate can only be constructed if $i \geq 5$ and, similarly, defective gates will also be formed at the top of the profiles. Gates are not necessarily calculated for all values of i. One possibility is to use the idea shown in Fig. 5.32, where i is incremented in steps of the volume index so that the zero lags will be independent. With this choice the longer lags of neighbouring gates will always consist partly of the same matrix elements so that these lags are not independent.

In a classical experiment the weight of lag k relative to the zero lag is

$$\mathcal{W}_k^c = \frac{(n-k)\mathcal{W}_k}{n\mathcal{W}_0} = \frac{(n-k)(n-k)/n}{n \cdot 1} = \frac{k^2}{n^2} - \frac{2}{n}k + 1. \tag{5.27}$$

In a modern experiment, on the other hand, the corresponding equation is

$$\mathcal{W}_k^m = \frac{(v+k)\mathcal{W}_k}{v\mathcal{W}_0} = \frac{(v+k)(n-k)/n}{v \cdot 1} = -\frac{1}{nv}k^2 + \frac{n-v}{nv}k + 1. \tag{5.28}$$

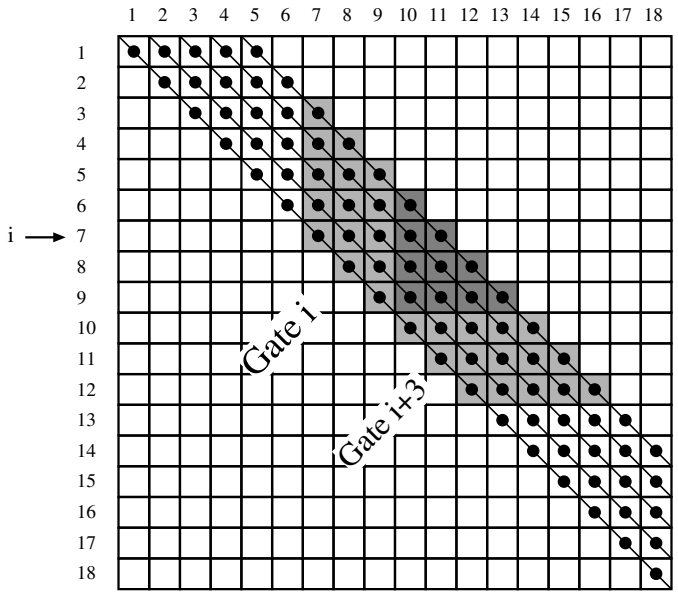

Figure 5.32: Generation of a gate in a modern long-pulse experiment. The pulse length is assumed to be five times the sampling interval and the volume index is 3. Two gates are shown, which partially overlap within the area indicated by darker shading.

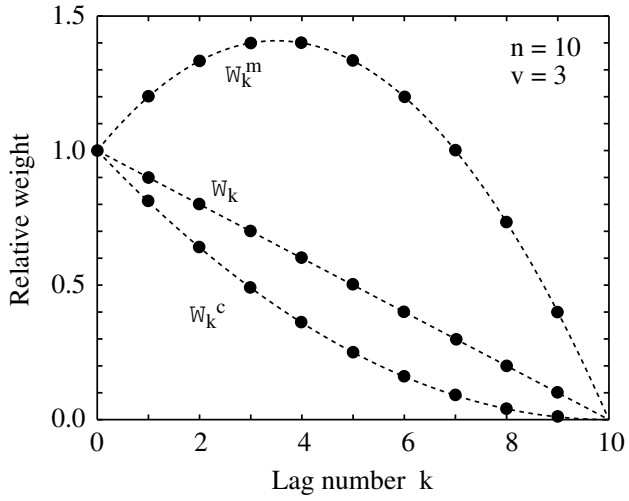

Figure 5.33: Weights of individual lag profile matrix elements and lags of classical and modern long-pulse experiments relative to the weight of the zero lag.

The relative weights are plotted in Fig. 5.33 together with the weights of the individual lag profile matrix elements. In this example $n = 10$ and $v = 3$. It is seen that the weights of the classical and modern experiment lie on upward and downward opening parabolas, respectively, and those of the matrix elements on a descending straight line. The relative weights of different lags are much more in balance in a modern than in the classical long-pulse experiment. Therefore, although the weights are not quite the same, problems similar to that in the classical experiment do not arise in the modern experiment.

The second point in the comparison of the classical and modern long-pulse experiment is the height resolution of the different lags. An individual matrix element for lag k obviously covers a range (see Fig. 5.29)

$$\Delta r_k = \frac{c}{2}(n - k + 1)\tau. \tag{5.29}$$

In the classical experiment $n - k$ elements are added in the kth lag. Since they are separated by τ in S-direction, the height resolution of the kth lag is

$$\Delta r_k^c = \frac{c}{2}[(n + 1 - k) + (n - k - 1)]\tau = c(n - k)\tau. \tag{5.30}$$

Finally, in the modern experiment $v + k$ elements are added in the kth lag so that

$$\Delta r_k^m = \frac{c}{2}[(n - k + 1) + (v + k - 1)]\tau = \frac{c}{2}(n + v)\tau. \tag{5.31}$$

These results show that the height resolutions of both the matrix elements and the the lags of the classical experiment decrease linearly with lag number, the latter one with a double rate. In the case of the modern experiment, on the other hand, the gate has the same width for all lags,which is convenient from the data analysis point of view. In this respect the modern long-pulse experiment resembles the multipulse experiments.

In conclusion, a modern long-pulse experiment is constructed to give the same gate width for all lags. This is obtained by adding lag profile matrix elements to create lag estimates in a proper way. Although the weights of these estimates are not quite identical, they are anyway close enough to give an accuracy of the same order for most of the measured lags. Only the very longest lags are measured with a smaller accuracy.

The range ambiguity functions of the measured lags are obtained as sums of the range ambiguity functions of the individual matrix elements. This is demonstrated in Fig. 5.34, where a long pulse experiment with $n = 5$ and $v = 3$ is chosen. The range ambiguity functions of the individual matrix elements are of the same form as those in Fig. 5.29. In Fig. 5.34 each ambiguity function is plotted at at $v + k$ altitudes on the S-axis, separated by τ, and their sums are formed graphically. The results are the range ambiguity functions of the long-pulse experiment. They indicate that the form of the range ambiguity function changes from gate to gate, but its length on the S-axis remains as constant, equal to $(n + v)\tau = 8\tau$.

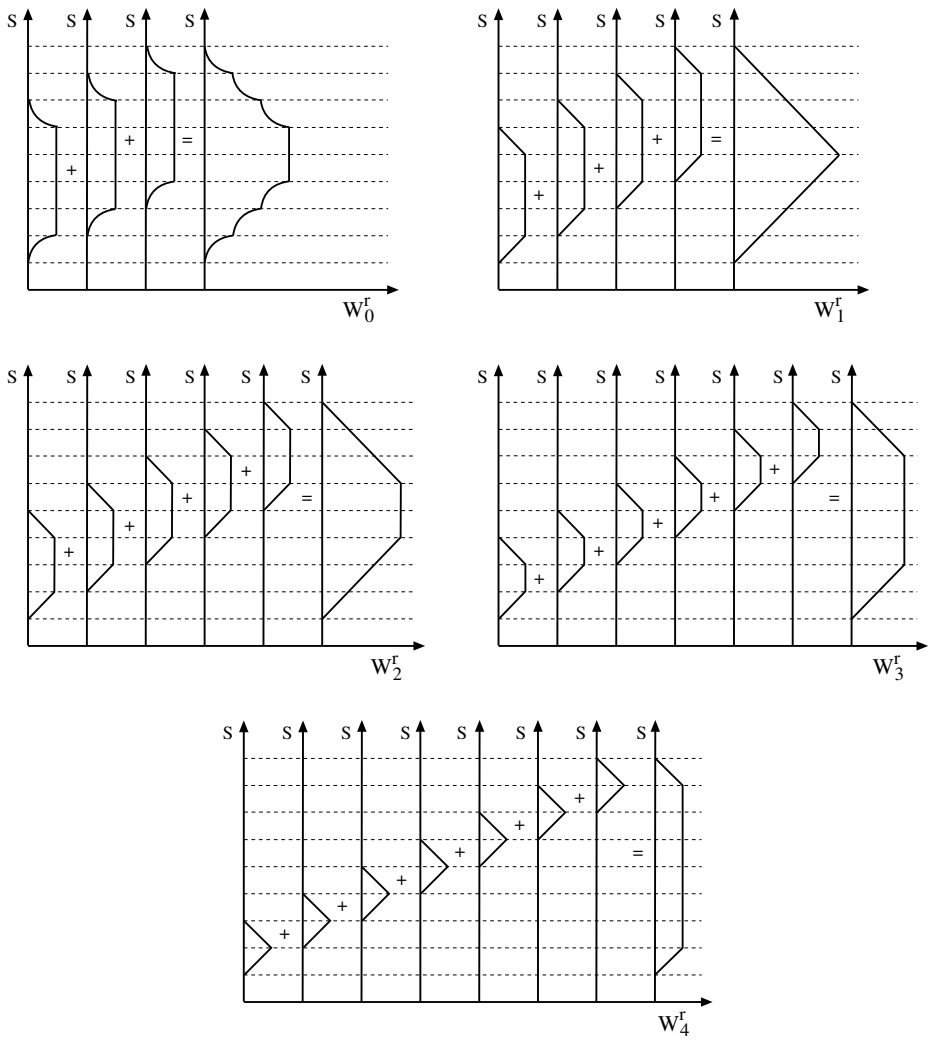

Figure 5.34: Generation of range ambiguity functions in a modern long-pulse experiment.

Chapter 6

Alternating codes

6.1 Working principle of alternating codes

The multipulse and Barker-coded multipulse modulations introduced in Chapter 5 are designed for a height resolution of a few kilometres or hundreds of metres, respectively. Although they can effectively make use of the radar duty cycle when interlaced on several frequencies, they are not optimal modulation methods. This is because lag estimates can only be calculated from pulse pairs at the same frequency channel so that samples from different frequencies cannot be correlated. A more efficient experiment would be produced if a single long pulse could be divided into a number of shorter elements and signals from all element pairs could be correlated without a disturbance from the other elements. In 1980's new modulation methods were invented which fulfil this demand. They are known as alternating codes and random codes.

Both alternating and random codes are phase codes which use two phase values, 0° and 180° just like Barker codes do. The transmission consists of long pulses which are divided into a number of bits. Reception takes place after each transmission before the next pulse is transmitted. Unlike in Barker codes, these phase patterns change from pulse to pulse and the length of the impulse response is equal to the bit length rather than pulse length, *i.e.* no decoding filter is applied (except when individual bits are further modulated by Barker codes). The codes work in power domain rather than in amplitude domain. This means that the process corresponding to decoding is not done in terms of amplitudes but in terms of lagged products; lagged products from successive pulses are added and subtracted according to certain rules so that the corresponding ambiguity functions cancel except at a single short altitude range corresponding to the bit length. In alternating codes the phase patterns are designed so that this happens exactly but in random codes the patterns are produced randomly by a code generator and the ambiguity functions cancel only in a statistical sense. The patterns change continuously in random codes but in alternating codes they make a finite set which is repeated all over again.

Two basic types of alternating codes exist, the Sulzer type and the Lehtinen-Häggström type. The benefit of the Sulzer codes is that their cycle length (the number of phase patterns in the cycle) is shorter than the number of bits in a single pattern whereas in the Lehtinen-Häggström codes the cycle length is equal to or twice the number of bits. On the other hand, the Sulzer codes have a drawback that they are known only up to the length of 12 bits, whereas much longer Lehtinen-Häggström codes are known than needed in any practical application. Also, Sulzer codes can be used only if the radar power remains constant during the transmitted pulse. There is no such limitation in the Lehtinen-Häggström codes, but they allow the transmitting power to vary in a systematic way.

In this chapter only alternating codes of the Lehtinen-Häggström type are considered. These codes are divided into two types, weak and strong. Weak codes are applicable if each bit of the code is further modulated by a Barker code, otherwise strong codes must be used. The number of bits in both codes is equal to a power of two. The cycle length of a weak code is equal to the number of bits in a single phase pattern and that of a strong code twice the number of bits.

The working principle of a four-bit weak alternating code is demonstrated in Fig. 6.1, where the matrix denoted by 'env' contains all phase patterns in the code cycle. Here the phases 0° and 180° are indicated by plus and minus, respectively. Each row in the matrix indicates a 4-bit modulation envelope in a single pulse and the four rows together make the whole cycle. The rows, *i.e.* the phase patterns of the individual pulses, are sometimes called scans and their order in the cycle is indicated by a number called the scan count.

The matrix denoted by W_1^r in Fig. 6.1 is obtained by multiplying the successive columns in the original sign matrix; the first column is the product of the first and second, the second a product of second and third, and the third a product of third and fourth. The next matrix in Fig. 6.1 is obtained from W_1^r by multiplying all its columns by the first column so that the first column will be all pluses. When the sums of the three columns are calculated, it turns out that only the first column will give a non-zero result (equal to 4), in the other two columns the pluses and minuses cancel. The last two matrices are

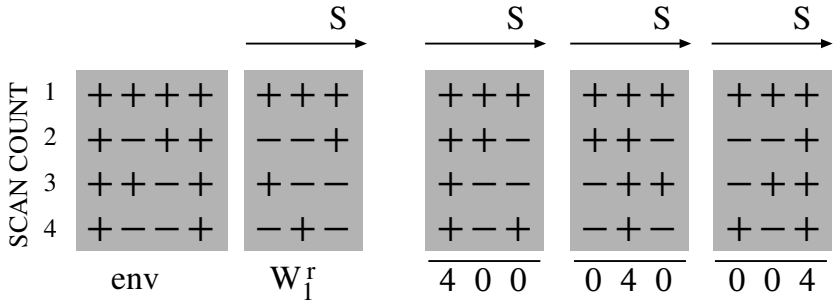

Figure 6.1: Generation of the first lag in a four-bit weak alternating code.

obtained in a similar manner using the signs of the second and third column of W_1^r. Also in these two cases the sums of only a single column (second and third) are non-zero.

For interpreting this result, assume that the signal is sampled at intervals equal to the bit length. After each transmission in the cycle, a different set of samples and lag estimates is obtained. If the impulse response of the system is equal to the Dirac δ-function, each row in the matrix W_1^r corresponds simply to multiplication of two $h*$env-terms separated by a single bit length on the S-axis. Therefore the pluses and minuses on each row indicate values of the range ambiguity functions of lag estimates $x_i \cdot x_{i+1}$ at integer values of S. Between these points the range ambiguity functions vary linearly. The sums below the last three matrices in Fig. 6.1 correspond to weighted sums of lag estimates $x_i \cdot x_{i+1}$ of the four data sets, calculated using the signs of each of the columns of W_1^r as weights. The zeros indicate that the ambiguity functions of the different scans will partly cancel in this calculation and a total range ambiguity function for the alternating code will be formed which is non-zero only within a range corresponding to a length of a single bit on the S-axis. This ambiguity function is similar to that of a simple two-pulse code. Altogether, this means that two successive data samples give first lag estimates at three different heights. These correspond to the three bit pairs which have a separation of a single bit length in a four-bit sequence.

The second and third lag can be studied in an analogous way by calculating the matrices W_2^r and W_3^r, which consist of two columns and one column, respectively. Because a four-bit sequence contains two bit pairs with a separation of two bit lengths, samples x_i and x_{i+2} will give second lag estimates at two heights. Similarly, samples x_i and x_{i+3} give only a single third-lag estimate, since only the separation of the first and fourth bit is equal to three bit lengths.

When lags 1–3 are calculated for all data samples x_i in the way described above, three lag estimates will be obtained at each height for the first lag, two for the second lag and one for the third lag. Longer weak codes work in an analogous way; if the code length is n (must be a power of two), $n - i$ lag estimates are obtained for the lag i, $i = 1, \ldots n - 1$. The zero lag cannot be determined with the same height resolution, but it contains range ambiguities much in the same way as the zero lag of a multipulse code does.

In the above discussion we assumed that the impulse response is equal to the Dirac δ-function, so that the neighbouring bits will be kept separate in calculating $h *$ env. With a realistic impulse response the successive bits will be mixed and the conclusions are no more valid. It is, however, possible to apply Barker coding to individual bits of the alternating code so that the range ambiguity functions of different scans will consist of narrow peaks and weak sidebands. Then the principle of Fig. 6.1 will apply to the main peaks which will cancel each other in the weighted sums except at a single altitude, and the alternating code will work.

In conclusion, weak alternating codes like that shown in Fig. 6.1 can be used if each bit is further modulated by a Barker code. In strong alternating codes Barker modulation is not needed. The number of scans in strong codes is twice

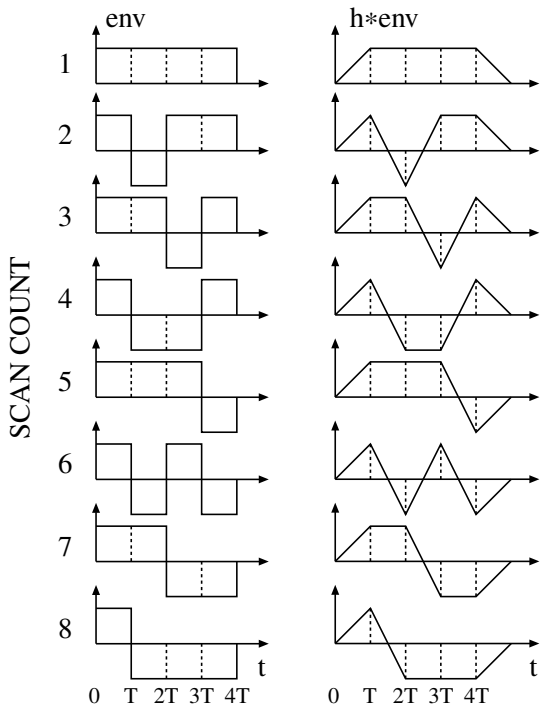

Figure 6.2: Modulation envelopes of a four-bit strong alternating code and their convolutions with a boxcar-shaped impulse response. Bit boundaries are indicated by vertical dashed lines.

the number of bits so that a four-bit strong code corresponding to the weak one in Fig. 6.1 contains eight scans. The modulation envelopes of a four-bit strong code are shown in the left hand panel of Fig. 6.2 and their convolutions with a boxcar-shaped impulse response in the right hand panel. The length of the impulse response is equal to the bit length. The code could be presented more briefly in terms of a 8×4 matrix of pulses and minuses in the same way as the weak code in Fig. 6.1. However, the same notation does not apply to the convolution $h * \text{env}$ since it does not consist of boxcar-shaped functions.

The range ambiguity functions can now be formed for all eight modulations using the convolutions $h * \text{env}$. When copies of the convolutions are moved by one bit length and multiplied by the original ones, the eight range ambiguity functions shown in the left panel in Fig. 6.3 are obtained. The ambiguity functions of the strong code are obtained in the same way as those for the weak code in Fig. 6.1. The weights are given by the values of the range ambiguity functions at $S = t_i - T$, $t_i - 2T$ and $t_i - 3T$, where T it the bit length. In these three cases the weights are $+-+-+-+-$, $+--++--+$ and $++----++$, respectively, and each choice gives a peak around the respective S-value. The peaks

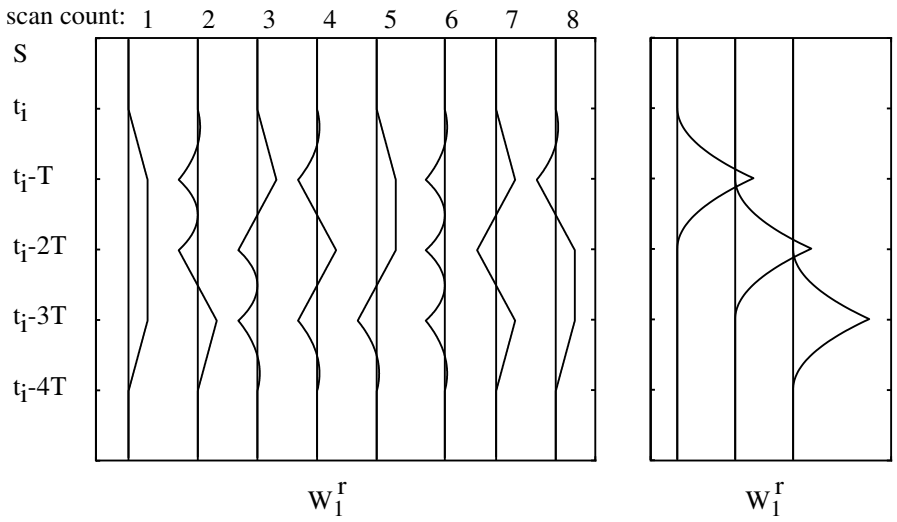

Figure 6.3: The range ambiguity functions of the first lag for the strong four-bit code in Fig. 6.2. Left panel: range ambiguity functions of the eight scans. Right panel: Range ambiguity functions of the alternating code obtained as weighted sums of the ambiguity functions of the individual scans.

are shown in the right hand panel. Hence the code gives first lag estimates at three heights separated by a distance corresponding to the bit length and each of these estimates contains information from a height range corresponding to two bit lengths.

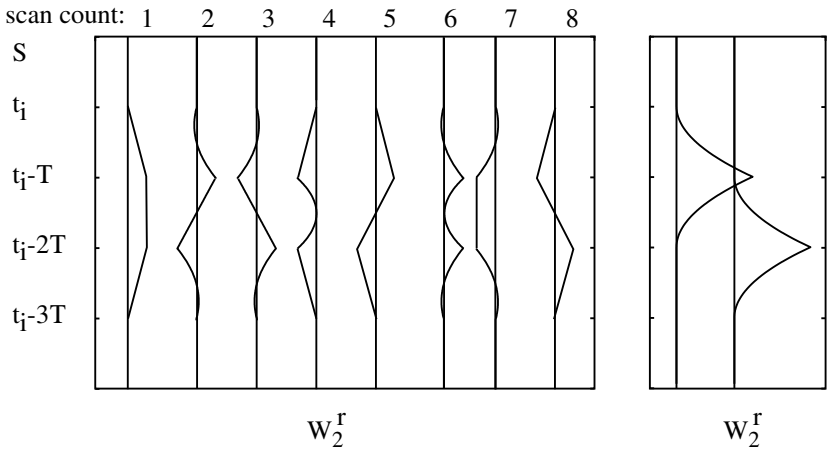

Figure 6.4: Same as Fig. 6.3 for the second lag.

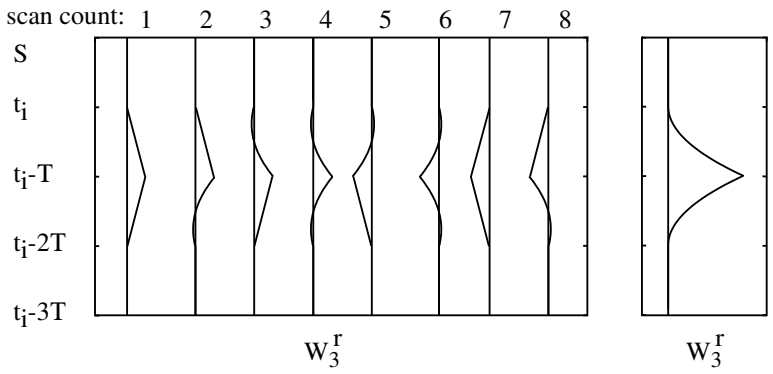

Figure 6.5: Same as Fig. 6.3 for the third lag.

The range ambiguity functions for the second and third lag are calculated in the same way, only the copies of the convolutions are shifted by two and three bit lengths, respectively. The results are plotted in Figures 6.4 and 6.5 in the same format as Fig. 6.3. It is observed that two lag estimates are obtained for the second lag (weights $++--++--$ and $+-+--+-+$) and one for the third lag (weights $+++----$). The shapes of the range ambiguity functions for all lag estimates are identical and similar to that of a simple two-pulse code, plotted in Fig. 4.6.

This example shows that a four-bit strong alternating code is capable of correlating signals from all possible bit pairs in the modulation pattern when the length of the impulse response is equal to the bit length. A weak code can do the same thing only, if the impulse response is a δ-function or an additional Barker coding is used. For this achievement a strong code needs a double number of modulation patterns as compared to the corresponding weak code.

The working principle of alternating codes was demonstrated here in terms of range ambiguity functions. One can show that two-dimensional ambiguity functions of an alternating code can be calculated in the same way and they have the same shape as the two-dimensional ambiguity function of a simple two-pulse code. As an example, all two-dimensional ambiguity functions of a strong four-bit code are plotted in Fig. 6.6. For lags 1–3 each of them is a single pyramid but the ambiguity function of the zero lag is a structured ride elongated in the S-direction. Note that the different ambiguity functions partly overlap, and only the topmost surface is visible in the figure.

The strong four-bit code works as an example but it has not much practical value. This is because it gives only three lags. Longer codes are needed for measuring signal autocorrelation functions with sufficient lag resolutions. It is fairly easy to study whether a given sign matrix makes an alternating code or not; finding the proper matrices is a much more difficult task. Going systematically through all possible sign combinations is hopeless. Consider, for example,

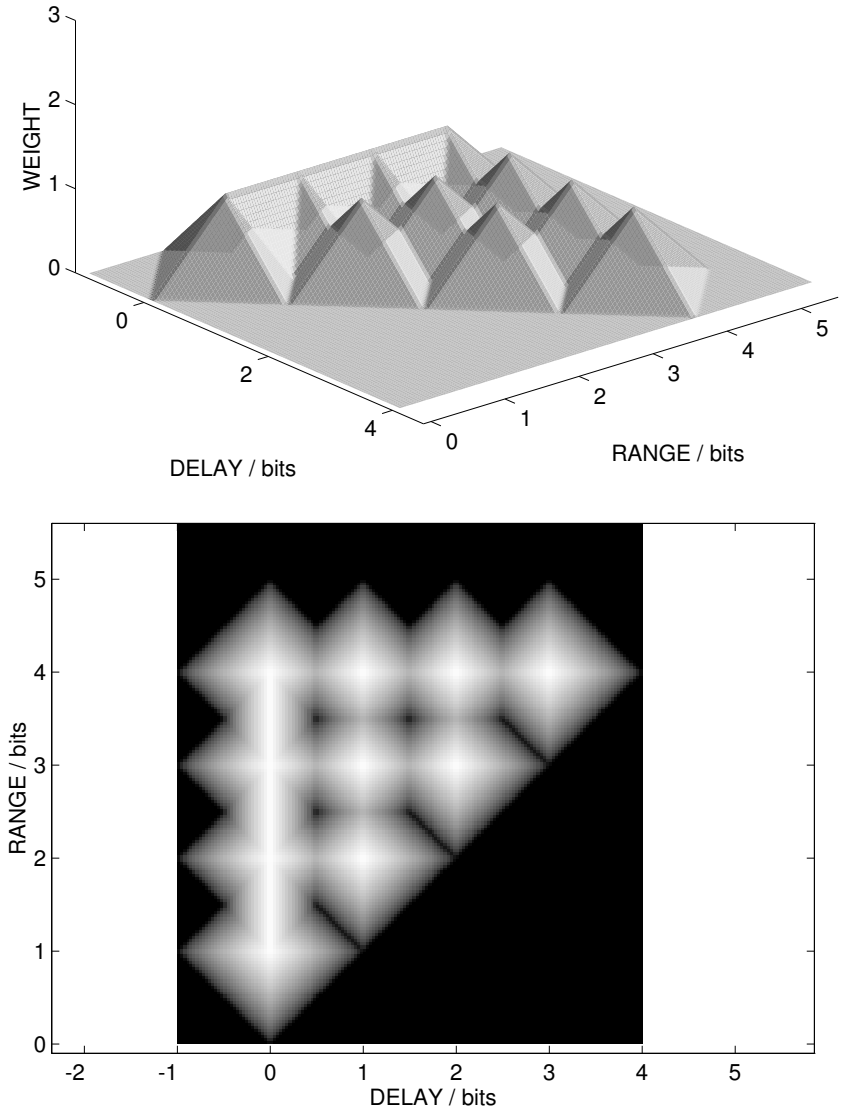

Figure 6.6: Two-dimensional ambiguity functions of a four-bit strong alternating code.

a strong four-bit code. The number of possible combinations is

$$2^{4\times 8} = 2^{32} \approx 4.3 \cdot 10^9.$$

If your computer is able to test one combination in 0.1 s, testing all possibilities would take 13.6 years. In the case of an eight-bit code the number of combinations is

$$2^{8\times 16} = 2^{128} \approx 3.4 \cdot 10^{38}$$

and the time needed would be about 10^{31} years. Considering that the age of the universe is about $15 \cdot 10^9$ years, a search for an eight-bit strong code does not look like a promising project! In spite of these pessimistic views, very long alternating codes codes have been found, and the longest strong code known at the moment has a length of 4194304 bits. This, of course, is much more than needed in any practical application.

6.2 Properties of alternating codes

As pointed out in Chapter 5.2, the signal autocorrelation function must be observed up to a certain length with a proper lag resolution. In the case of alternating codes the applied pulse length is determined by the maximum lag needed. On the other hand, the bit length determines the width of the range ambiguity function as well as the range resolution. Therefore the maximum lag also fixes the range resolution for a given code, which can only be improved by increasing the number of bits. This means that long alternating codes are needed for high resolution experiments. Consider, for instance, a typical length of an E region autocorrelation function which is of the order of 300 μs for the EISCAT UHF radar. In this case the bit length of a 32-bit code would be 10 μs and the gate separation 1.5 km. Longer codes are needed in order to obtain a better range resolution; going down to 300–400 m, for instance, implies 128-bit codes.

Because alternating codes cannot be found by a systematic testing of all sign combinations, their properties must be studied in order to find more effective search methods. The codes can be considered as strings of elementary pulses of the form

$$q_i(t) = \begin{cases} 1 \text{ , when } i < t < i+1 \\ 0 \text{ , when } t \le i \text{ or } t \ge i+1. \end{cases} \tag{6.1}$$

Here the bit length is taken as the time unit. Both positive and negative pulses may appear in the string so that a single modulation envelope of a N-bit code is

$$\mathrm{env}(t) = \sum_{i=0}^{N-1} s_i q_i(t), \tag{6.2}$$

where each s_i is either $+1$ or -1. If, for instance, $N = 4$ and $s_0 = 1$, $s_1 = -1$, $s_2 = -1$ and $s_3 = 1$, the modulation envelope is the same as the fourth scan in Fig. 6.2.

Using this notation, an N-bit alternating code with a cycle length M (M scans in the cycle) can be written in the form

$$\mathrm{env}^{(c)}(t) = \sum_{i=0}^{N-1} s_i^{(c)} q_i(t), \tag{6.3}$$

where $c = 1, 2, \ldots, M$ is the scan count. The task is to find such sign sequences

$$\{s_i^{(c)}, i = 0, 1, \ldots, N-1\}, \quad c = 1, 2, \ldots, M$$

that they generate an alternating code.

The range ambiguity function of a given scan c for the lag $t' - t$ is

$$\begin{aligned} W_{tt'}^{(c)}(S) &= (h * \mathrm{env}^{(c)})(t - S) \cdot (h * \mathrm{env}^{(c)})(t' - S) \\ &= \left[\sum_{i=0}^{N-1} s_i^{(c)} (h * q_i)(t - S)\right] \cdot \left[\sum_{j=0}^{N-1} s_j^{(c)} (h * q_j)(t' - S)\right] \\ &= \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} s_i^{(c)} s_j^{(c)} [(h * q_i)(t - S)][(h * q_j)(t' - S)]. \end{aligned} \tag{6.4}$$

Next we multiply the signs of the $(k+1)$th and $(l+1)$th elementary pulse in each transmission envelope separately and calculate the weighted sum of the range ambiguity functions of all scans using the sign products as weights. The result is

$$\begin{aligned} W_{tt'}^{kl}(S) &= \sum_{c=1}^{M} s_k^{(c)} s_l^{(c)} W_{tt'}^{(c)}(S) \\ &= \sum_{i,j=0}^{N-1} \left[\sum_{c=1}^{M} s_k^{(c)} s_l^{(c)} s_i^{(c)} s_j^{(c)}\right] [(h * q_i)(t - S)][(h * q_j)(t' - S)] \\ &= \sum_{i,j=0}^{N-1} A_{ijkl} [(h * q_i)(t - S)][(h * q_j)(t' - S)], \end{aligned} \tag{6.5}$$

where

$$A_{ijkl} = \sum_{c=1}^{M} s_i^{(c)} s_j^{(c)} s_k^{(c)} s_l^{(c)}. \tag{6.6}$$

Because $(s_i^{(c)})^2 = 1$ for all values of i, we find that

$$A_{klkl} = \sum_{c=1}^{M} (s_k^{(c)})^2 (s_l^{(c)})^2 = M \tag{6.7}$$

and eq. (6.5) can be written in the form

$$W_{tt'}^{(c)}(S) = M[(h * q_k)(t - S)][(h * q_l)(t' - S)]$$

$$+ \sum_{i \neq k, j \neq l} A_{ijkl}[(h * q_i)(t - S)][(h * q_j)(t' - S)]. \quad (6.8)$$

In this formula the first term depends only on the $(k + 1)$th and $(l + 1)$th elementary pulses, which have a separation $l - k$. For the lag $t' - t = l - k$ this term is the same as M times the range ambiguity function of a two-pulse code consisting of these two elementary pulses. If the sign sequences $\{s_i^{(c)}\}$ can be chosen in such a way that the sum term in eq. (6.8) is zero for all values t and t' with a separation $t' - t = l - k$, the whole range ambiguity function for this lag will be similar to that of a two-pulse code consisting of elementary pulses q_{k+1} and q_{l+1}. Further, if such sign sequences $\{s_i^{(c)}\}$ could be chosen that this would be true for all pairs of elementary pulses, the code would produce the same ambiguity functions as an experiment which contains all possible two-pulse codes taken from an N-pulse sequence. The difference is that, in an alternating code, these pulse pairs are contained in a single modulation envelope, whereas in a two-pulse experiment they should be transmitted separately.

The convolutions in the sum term in eq. (6.8) can be simplified when noticing that

$$q_i(t) = q_0(t - i).$$

Then

$$(h * q_i)(t - S) = (h * q_0)(t - S - i)$$

and the product of the convolutions in the sum term of eq, (6.8) can be put in the form

$$[(h * q_i)(t - S)][(h * q_j)(t' - S)] = [(h * q_0)(t - S - i)][(h * q_0)(t' - S - j)]. \quad (6.9)$$

If the impulse response is a δ-function (weak code) the convolution $h * q_0$ is non-zero only between the argument values 0 and 1. This means that all terms in the sum in eq. (6.8) are automatically zero, except those which satisfy the condition $t - S - i = t - S + l - k - j$. In the case of a strong code the length of the impulse response is equal to the bit length and the corresponding interval is between 0 and 2. Then the convolutions will put all terms in the sum to zero except those satisfying both the previous and the additional condition $t - S - i = t - S + l - k - j \pm 1$. These two relations are equivalent to

$$j - i = l - k \quad (6.10)$$

and

$$j - i = l - k \pm 1. \quad (6.11)$$

If we can choose the sign sequences $\{s_i^{(c)}\}$ in such a way that the coefficients A_{ijkl} are zeros in all cases when the convolution part in the sum term is non-zero, the sum in eq. (6.8) will disappear for all lag values. Therefore a sign sequence $\{s_i^{(c)}\}$ generates a weak alternating code, if $A_{ijlk} = 0$ for all values of

i, j, k and l satisfying eq. (6.10), when $i \neq k$ and $j \neq l$. If the same is true for indices satisfying both eq. (6.10) and eq. (6.11), the sign sequence generates a strong alternating code.

The zero lag is obtained when $k = l$. In this case eq. (6.10) gives $i = j$ and

$$A_{iikk} = \sum_{c=1}^{M} s_i^{(c)} s_i^{(c)} s_k^{(c)} s_k^{(c)} = M, \tag{6.12}$$

which is necessarily non-zero. Therefore the sum term cannot disappear for zero lag and the alternating codes cannot give zero lag estimates with a range resolution corresponding to the bit length. This was already seen in Fig. 6.6, where the zero lag ambiguity function is a single long ridge rather than a pyramid.

In conclusion, we have formulated a mathematical condition such that any sign sequence $\{s_i^{(c)}\}$ satisfying it is necessarily an alternating code. This condition can be applied in the code search.

6.3 Walsh matrix

It turns out that a search method of alternating codes can be developed using so called Walsh or Hadamard matrices. An $n \times n$ Walsh matrix can be defined by the recursive formulas

$$Wsh_1 = 1 \tag{6.13}$$

and

$$Wsh_n = \begin{pmatrix} Wsh_m & Wsh_m \\ Wsh_m & -Wsh_m \end{pmatrix}, \tag{6.14}$$

where $n = 2^{k+1}$ and $m = 2^k$, $k = 0, 1, \ldots$. Hence the dimension of a Walsh matrix is always a power of 2. All columns (and rows) of a Walsh matrix consist of an equal number of pluses and minuses, except the first one which is all pluses. This property gives a possibility to construct sign matrixes of alternating codes from columns of the Walsh matrix.

It is possible to derive alternative formulations to the Walsh matrix which are useful in later analysis. If i and j are non-negative integers, they can be put in binary form

$$i = \sum_{n=0}^{\infty} a_n 2^n \tag{6.15}$$

and

$$j = \sum_{n=0}^{\infty} b_n 2^n, \tag{6.16}$$

where each coefficients a_n or b_n is either zero or one. In terms of this presentation the Walsh matrix can be written as

$$Wsh(i, j) = (-1)^{\sum_{n=0}^{\infty} a_n b_n}. \tag{6.17}$$

This means that the (i,j)th element of the Walsh matrix is unity if, in the binary presentations of i and j, there is an even number of bits with both a_n and b_n equal to unity, otherwise the element is equal to -1. Note carefully that here the indexing of matrix elements starts from zero.

It is easily seen that the definition in eqs. (6.13) and (6.14) is equivalent to eq. (6.17). Eq. (6.17) obviously gives $Wsh(0,0) = 1$, which is equal to Wsh_1 in eq. (6.13). When the matrix is expanded in the way shown in eq. (6.14), one has to study what happens when 2^k is added to one or both of the indices of the matrix elements of Wsh_m. Doing this to one of the indices corresponds to a shift in the column or row directions and doing it to both indices corresponds to a diagonal shift. Since the indices of Wsh_m are always smaller than 2^k, the addition means that a single 'one' will appear at the position 2^k in the bit pattern of the index and all 'ones' in the original pattern will remain unaffected. Therefore, if the addition is done only to one of the indices, the sum in eq. (6.17) does not change, so that $Wsh(i, j + 2^k) = Wsh(i,j)$ and $Wsh(i + 2^k, j) = Wsh(i,j)$. If the addition is done to both indices, the sum will add by one; it will be odd if it was previously even and vice versa. Hence $Wsh(i+2^k, j+2^k) = -Wsh(i,j)$ and the two definitions indeed agree.

For later use, we now define the binary 'and' and 'exclusive or' operations as follows. If i and j are binary numbers of the form (6.15) and (6.16), their binary 'and' is

$$i \wedge j = \sum_{n=0}^{\infty} a_n b_n 2^n \tag{6.18}$$

and binary 'exclusive or' is

$$i \oplus j = \sum_{n=0}^{\infty} c_n 2^n, \tag{6.19}$$

where $c_n = 0$ when $a_n = b_n$ and $c_n = 1$ when $a_n \neq b_n$. Thus $i \wedge j$ is a binary number which has a 'one' in every position of its bit pattern where both i and j have a 'one' and $i \oplus j$ is a number which has a 'one' in every position where the digits of i and j are not the same. In addition, we define the parity of a binary number as

$$par(i) = (-1)^{\sum_{n=0}^{\infty} a_n}. \tag{6.20}$$

This means that the parity of a binary number is 1, if its bit pattern contains an even number of 'ones', otherwise it is -1.

These definitions allow us to write the Walsh matrix in two different forms. Starting from eq. (6.17) and using eqs. (6.20) and (6.18) we readily obtain

$$Wsh(i,j) = par(\sum_{n=0}^{\infty} a_n b_n 2^n) = par(i \wedge j). \tag{6.21}$$

Using the above definitions we finally show that the Wash matrix obeys the following two theorems:

Theorem 1:

$$Wsh(i,j) \cdot Wsh(i,k) = Wsh(i, j \oplus k). \tag{6.22}$$

Proof:

Using eq. (6.17) we obtain

$$Wsh(i,j) \cdot Wsh(i,k) = (-1)^{\sum_{n=0}^{\infty} a_n b_n} \cdot (-1)^{\sum_{n=0}^{\infty} a_n c_n}$$
$$= (-1)^{\sum_{n=0}^{\infty} a_n (b_n + c_n)} = (-1)^{\sum_{n=0}^{\infty} a_n d_n},$$

where $\{c_n\}$ is the bit pattern of k and $d_n = b_n + c_n$. Since, for a given index n', $b_{n'}$ and $c_{n'}$ can only be either ones or zeros, we have three alternatives:

- $b_{n'} = c_{n'} = 0 \Longrightarrow d_{n'} = 0$,
- $b_{n'} \neq c_{n'} \quad \Longrightarrow d_{n'} = 1$,
- $b_{n'} = c_{n'} = 1 \Longrightarrow d_{n'} = 2$.

In the last case

$$(-1)^{\sum_{n=0}^{\infty} a_n d_n} = (-1)^{2a_{n'}} \cdot (-1)^{\sum_{n \neq n'} a_n d_n} = (-1)^{\sum_{n \neq n'} a_n d_n},$$

so that, instead of $d_{n'} = 2$, one can put $d_{n'} = 0$. Then the numbers $\{d_n\}$ make a bit pattern of the binary number

$$l = \sum_{n=0}^{\infty} d_n 2^n,$$

such that $d_n = 0$ when $b_n = c_n$, and $d_n = 1$ when $b_n \neq c_n$. This means that

$$l = j \oplus k$$

so that

$$Wsh(i,j) \cdot Wsh(i,k) = (-1)^{\sum_{n=0}^{\infty} a_n d_n} = Wsh(i,l) = Wsh(i, j \oplus k),$$

Q.E.D.

Theorem 2:

If N is a power of two, and $0 \leq j \leq N-1$,

$$\sum_{i=0}^{N-1} Wsh(i,j) = 0 \tag{6.23}$$

for all values of $j \neq 0$.

Proof:

The statement means that the column sums of a Walsh matrix Wsh_N are always zeros, except for the first column. All elements in the first column are ones so that their sum is N. Because

$$Wsh_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

and because the Walsh matrix is built according to eq. (6.14), eq. (6.23) is necessarily valid for all values $j \neq 0$, Q.E.D.

6.4 Search of alternating codes

The original method of alternating code search is based on the properties of the Walsh matrix. In this approach only codes are investigated where the cycle length M is a power of two, *i.e.* $M = 2^m$. The sign sequences $\{s_i^{(c)}\}$, $c = 1, 2, \ldots, M$ are chosen in such a way that

$$s_i^{(c)} = Wsh(c-1, a_i), \;\; i = 0, 1, \ldots, N-1 \tag{6.24}$$

where N is the number of elementary pulses in a single scan and $0 \leq a_i \leq M-1$. This means that the signs of a single scan come from the same row of the Walsh matrix and the signs of the same elementary pulses in all scans come from the same column. Thus it is supposed that a complete alternating code can be obtained by taking N columns from the Walsh matrix Wsh_M in a proper order and arranging them into an $M \times N$ matrix. The rows of this matrix are the different scans of the alternating code.

The idea of using the Walsh matrix means that the set of sign combinations to be tested is radically limited. Because only those sign combinations which can be obtained from the Walsh matrix are studied, there is no guarantee (at least without a further mathematical study) that all possible alternating codes of a certain length are found. This is, however, quite unessential for the practical purpose of the search, since even knowing a single code of a given length is sufficient for incoherent scatter work.

It is not obvious without further proof that any sign combination calculated from eq. (6.24) will make an alternating code. Nevertheless, it is possible to show that certain conditions are valid, if such codes do exist. For the first, if a code with a cycle length M can be constructed from columns of a Walsh matrix, the columns can be chosen from the interval $0 \leq a_i \leq M-1$ and, for the second, none of the columns appears twice in the same strong code. Finally, in a weak code the number of columns to be chosen (the number of bits) is equal to $N = M$ and in a strong code $N = M/2$. Hence the cycle length of a strong code is twice the cycle length of a weak code with the same number of bits, a fact which was already seen in the example shown in Chapter 6.1.

In view of the discussion in Chapter 6.2, the effect of the choice in eq. (6.24) on the sums A_{ijkl} should be next studied. When $s_i^{(c)}$ from eq. (6.24) is inserted in eq. (6.6), the result is

$$A_{ijkl} = \sum_{c=1}^{M} Wsh(c-1, a_i) \cdot Wsh(c-1, a_j) \cdot Wsh(c-1, a_k) \cdot Wsh(c-1, a_l). \tag{6.25}$$

By applying Theorem 1 we immediately see that

$$A_{ijkl} = \sum_{c=1}^{M} Wsh(c-1, a_i \oplus a_j \oplus a_k \oplus a_l). \tag{6.26}$$

Because $0 \leq a_i \leq M$, $0 \leq a_j \leq M$, $0 \leq a_k \leq M$ and $0 \leq a_i \leq M$, also

$$0 \leq a_i \oplus a_j \oplus a_k \oplus a_l \leq M.$$

Then, according to Theorem 2, $A_{ijkl} = 0$ when the condition

$$a_i \oplus a_j \oplus a_k \oplus a_l \neq 0 \tag{6.27}$$

is valid. Thus the sign sequences $\{s_i^{(c)}\}$ obtained using eq. (6.24) make a weak alternating code, if eq. (6.27) is valid for all indices i, j, k and l satisfying the condition (6.10) when $i \neq j$ and $k \neq l$. If the indices satisfy both eq. (6.10) and eq. (6.11), the sign sequences make a strong code.

These results show that alternating codes can be searched by choosing the cycle length $M = 2^m$ and finding such a chain of $N = M$ or $N = M/2$ non-negative integers $\{a_i\}$ that condition (6.27) is valid for all combinations of four numbers in the chain within the given limitations. In this way the search has been converted to testing binary 'exclusive or' operations of small binary numbers, which is essentially faster than testing the condition $A_{ijkl} = 0$.

The search has been carried out with this method up to the length of 32 bits. The results for weak and strong codes are shown in Tables 6.1 and 6.2, respectively. The first column indicates the cycle length. The 32-bit codes are continued on two rows and the continuation lines are indicated by dots. Since every Walsh matrix index a_i in eq. (6.24) fixes a single column in the alternating code sign matrix, each code can be presented in terms of a string of Walsh indices. This is a convenient shorthand notation which is used in these tables and later on. Note that the Walsh indices are given in octal rather than decimal base.

A closer inspection of Tables 6.1 and 6.2 shows that a weak code can be obtained from a strong code by dividing the Walsh indices by two and neglecting the possible remainder. Because the search of weak codes is much faster than strong codes, this property can be made use of. Once a weak code has

Table 6.1
Walsh sequences of weak alternating codes up to the length of 32 bits. The first column indicates the cycle length and the Walsh indices are shown in octal base.

M	a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}
	...															
2	00	00	01													
4	00	00	01	02												
8	00	00	01	02	04	01	03	07								
16	00	00	01	02	04	06	02	05								
16	00	00	01	02	04	10	07	17	10	06	14	16	12	03	06	15
16	00	00	01	02	04	10	17	01	03	07	16	03	06	15	05	12
32	00	00	01	04	04	10	20	03	06	15	32	27	15	33	24	13
...	27	14	30	22	07	17	37	35	31	20	02	05	13	26	17	36
32	00	00	01	02	04	10	20	14	30	35	26	02	03	07	16	34
...	24	04	11	23	12	24	05	12	25	06	14	31	36	20	15	33

Table 6.2
Walsh sequences of strong alternating codes up to the length of 32 bits. The first column indicates the cycle length and the Walsh indices are shown in octal base.

M	a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}
	...															
4	00	01	02													
8	00	01	02	04												
16	00	01	02	04	10	03	07	16								
16	00	01	02	04	10	14	05	13								
32	00	01	02	04	10	20	17	37	21	14	31	35	24	06	15	32
32	00	01	02	04	10	20	36	03	07	16	34	06	15	32	12	25
64	00	01	02	04	10	20	40	06	15	32	64	56	33	67	50	27
...	57	30	61	44	17	37	76	72	63	41	05	13	26	54	36	75
64	00	01	02	04	10	20	40	30	61	72	55	03	07	16	34	70
...	50	11	23	46	24	51	12	25	52	14	31	62	74	41	33	67

been found, its Walsh indices are multiplied by two. Then the search of the corresponding strong codes is started from these indices and there is no need to vary them by more than one upwards. This greatly reduces the number of alternatives to be tested. The 32-bit strong codes in Table 6.2 were found in this way.

Although the above search method is effective, it is still very time consuming. The complete search for alternating codes up to the length of 32 bits took months of CPU time. It has been estimated that chances of finding 64-bit codes with the present computers are not very high. Luckily enough, a new search method has been developed which is still unpublished at the moment (February 1996). It searches directly for strong codes and it is so effective that codes up to any length needed in practice are easily found. The longest code known at the moment contains $2^{22} = 4194304$ bits. The new method was found by studying the regular behaviour of the strong 8-, 16- and 32-bit codes. Although these Walsh sequences seem to show no regularities except at the very beginning, a more careful investigation reveals a hidden structure which is repeated in codes

Table 6.3
Walsh sequence of a 64-bit strong code.

M	a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}
	...															
128	000	001	002	004	010	020	040	100	137	140	036	075	172	053	127	161
...	074	171	055	133	151	014	031	062	144	027	057	136	143	030	061	142
...	033	067	156	003	007	016	034	070	160	077	177	041	103	131	154	006
...	015	032	064	150	017	037	076	174	047	117	101	134	146	022	045	112

Table 6.4
Walsh sequence of a 128-bit strong code.

M	a_0 ...	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}
256	000	001	002	004	010	020	040	100	200	204	215	237	273	363	143	307
...	012	025	052	124	250	324	055	133	266	350	125	253	322	041	103	206
...	210	225	257	333	063	147	316	030	061	142	304	014	031	062	144	310
...	024	051	122	244	314	035	073	166	354	134	271	366	151	323	042	105
...	212	220	245	317	033	067	156	334	074	171	362	140	301	006	015	032
...	064	150	320	044	111	222	240	305	017	037	076	174	370	164	351	126
...	255	336	071	163	346	110	221	246	311	027	057	136	274	374	175	373
...	162	345	116	235	276	371	167	357	132	265	356	131	263	342	101	203

of various lengths and allows an easy construction of a strong code of any length. It is important to notice that this new method could only be found because codes up to the length of 32 bits were known. For demonstration, strong 64-bit and 128-bit codes are shown in Tables 6.3 and 6.4.

The sign matrices of alternating codes are constructed from their Walsh sequences in the way demonstrated in Fig. 6.7. The code in this example is

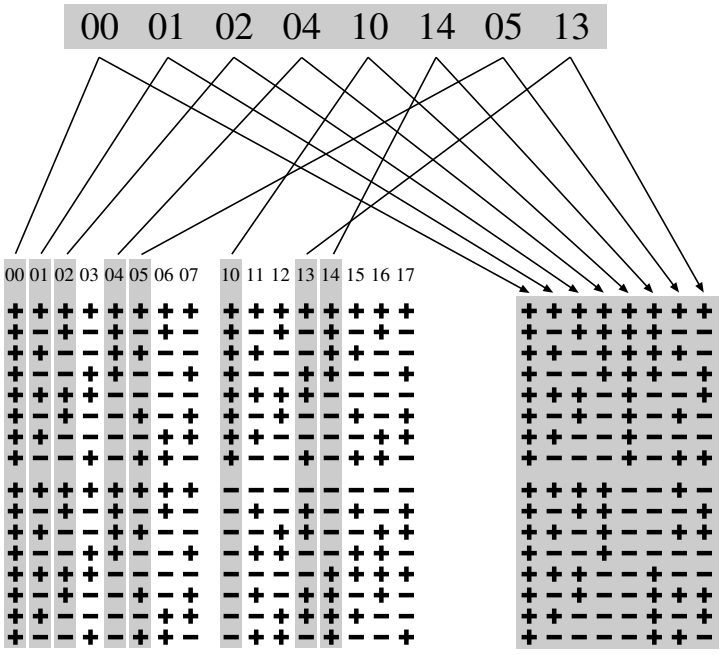

Figure 6.7: Construction of an 8-bit alternating code sign matrix from columns of an 8×8 Walsh matrix.

strong and 8-bits long with a Walsh sequence {00, 01, 02, 04, 10, 14, 05, 13} shown on the top of the figure. Hence we have to start from a 16×16 Walsh matrix, shown on the left. The columns of the matrix are numbered in octal base. The elements in the Walsh sequence are simply numbers of the columns which must be picked from the Walsh matrix and arranged in the order indicated by the sequence. The resulting sign matrix of the 8-bit code is shown on the right.

6.5 Efficiency of alternating codes

As discussed in Chapter 5.3, pulse codes can be intersperced at different frequencies in order to make use of the radar duty cycle in a maximal way. An example of such a modulation pattern is shown in Fig. 6.8, where four-pulse codes are transmitted at four frequencies and the remaining gaps are filled with two-pulse codes at two frequencies. Since no single pulses are transmitted, the zero lag cannot be measured.

The number of different lag estimates given by this modulation for a single gate is as follows:

◦ four-pulse codes, lags 1–6:	4×6	=	24
◦ two-pulse codes, lag 5:	2×1	=	2
	total		26

The total length of the pattern is 20 pulse lengths. The alternating code closest to this length is 16 bits. Hence, if we use the same range resolution, the multipulse experiment takes 25 % more of the duty cycle than the alternating code does. In an alternating code experiment the extra time available could be used for power profile measurements. Since an alternating code gives a single

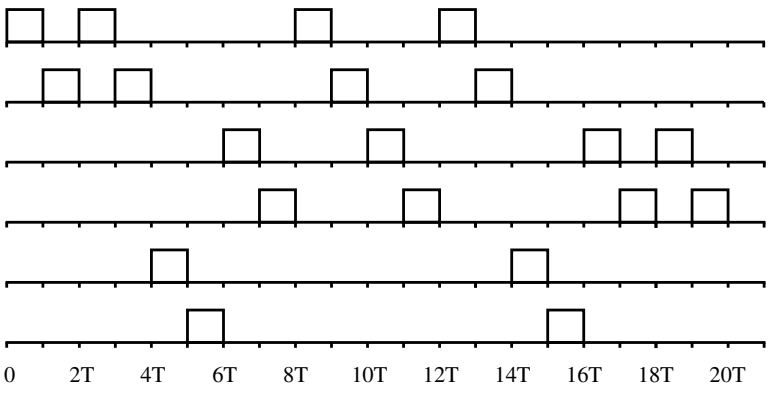

Figure 6.8: A modulation pattern consisting of four four-pulse codes and two two-pulse codes.

lag estimate for each bit pair, the total number of estimates is

$$\frac{16 \times (16-1)}{2} = 8 \times 15 = 120.$$

The above alternating code experiment is obviously more efficient than the multipulse experiment, because it gives many more lag estimates than the multipulse experiment can give in the same time. A useful number for comparing the relative efficiency of two experiments would be the ratio of observational errors when the same time is used in the measurement. A thorough analysis of such a comparison is a rather difficult task. Furthermore, the accuracy cannot often even be expressed in terms of a single number because the relative accuracy of different plasma parameters may vary from experiment to experiment. The reason for this is that the weight which a given lag has in determining the accuracy depends on the plasma parameter and experiments often do not measure the lags at the same relative accuracy, even if their lag increments and maximum lags were identical.

In general, it is well known that, if a quantity is measured several times, the standard deviation of the mean value of the measurements from the expectation value of the quantity is inversely proportional to the square root of the number of measurements made. If we neglect the fact that we actually observe different lags which have different weights in the accuracy of the plasma parameters, we can define a rough estimate to the efficiency of an experiment to be proportional to the square root of number of lag estimates it can give in unit time. Therefore the relative accuracy of the above experiments would be

$$\sqrt{\frac{120}{26}} \approx 2.1$$

This means that, if both experiments have the same temporal resolution, the

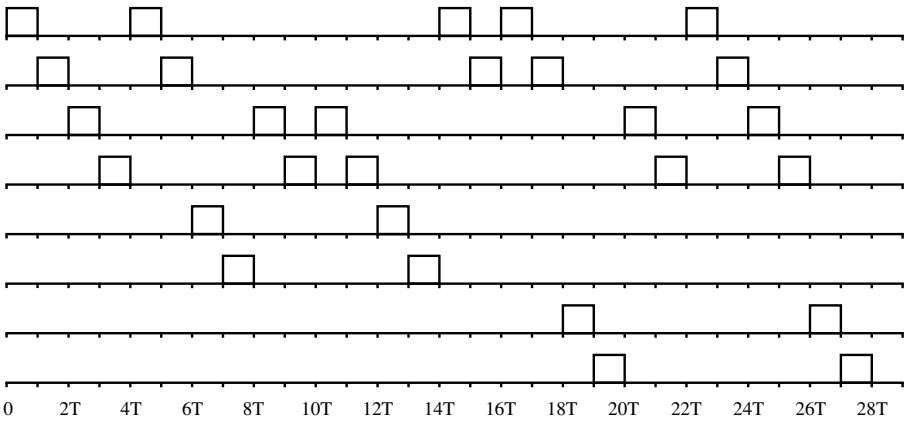

Figure 6.9: A modulation pattern consisting of four five-pulse codes and four two-pulse codes.

errors of the multipulse modulation would be more than twice the errors of the alternating code.

A modulation sequence of 28 pulse lengths can be constructed as shown in Fig. 6.9 using four five-pulse codes and four two-pulse codes. This experiment gives the following lag estimates:

◦ five-pulse codes, lags 1–9,11	4×10	=	40
◦ two-pulse codes, lag 3	2×1	=	2
◦ two pulse codes, lag 4	2×1	=	2
		total	44

The length of the closest alternating code is 32 bits and it gives

$$\frac{32 \times (32 - 1)}{2} = 16 \times 31 = 496$$

estimates. The ratio of the efficiencies of the two experiments is in this case

$$\sqrt{\frac{496}{44}} = 3.4.$$

These examples show that an alternating code experiment is 2–3 times more effective than a multipulse experiment utilising the same fraction of the radar duty cycle. The difference is caused by the fact that an alternating code can produce 5–10 times more lag estimates than a multipulse experiment can. The alternating code has this property because it allows signals from all elementary pulse pairs to be correlated, which is not possible in a multipulse experiment using several frequencies.

Chapter 7

Noise and offset

In the previous chapters interest was paid mainly to the scattering signal and only occasional references to noise were made. Like any electronic system, the incoherent scatter receiver generates noise of thermal origin which adds to the scattering signal. Cooled preamplifiers are used to reduce this effect. In addition to the thermal noise, the signal also contains radio noise of astronomical origin, which cannot be reduced by any means. The power of this noise component depends on the direction of the antenna beam. Because the incoherent scatter signal is very weak, the signal-to noise level (the ratio of singal power to noise power) is smaller than unity, often no more than about 10–20 %.

In addition to the noise, the radar signal is contaminated by a second effect. This is the offset voltage which is present at the output of the post-detection filter even in the absence of a scattering signal. The offset varies very slowly as compared to the scattering signal so that it can be considered to be constant in calculating a single autocorrelation function. The noise singal, on the other hand, is quickly varying and its power spectral density can be taken as constant before it enters the post-detection filter so that we can treat it as white noise.

Based on the above considerations, the complex output signal of the post-detection filter can be written in the form

$$z_h(t) = z_h^{(s)}(t) + z_h^{(n)}(t) + z_h^{(o)}(t), \tag{7.1}$$

where $z_h^{(s)}(t)$, $z_h^{(n)}(t)$ and $z_h^{(o)}(t)$ are the incoherent scatter signal, the noise signal and the offset voltage, respectively. The autocorrelation function of this signal for the lag $(t' - t)$ is

$$\begin{aligned}
\langle z_h(t) z_h^*(t') \rangle &= \langle [z_h^{(s)}(t) + z_h^{(n)}(t) + z_h^{(o)}(t)][z_h^{(s)}(t') + z_h^{(n)}(t') + z_h^{(o)}(t')]^* \rangle \\
&= \langle z_h^{(s)}(t) z_h^{(s)*}(t') \rangle + \langle z_h^{(n)}(t) z_h^{(n)*}(t') \rangle + \langle z_h^{(o)}(t) z_h^{(o)*}(t') \rangle \\
&+ \langle z_h^{(s)}(t) z_h^{(n)*}(t') \rangle + \langle z_h^{(s)}(t) z_h^{(o)*}(t') \rangle + \langle z_h^{(n)}(t) z_h^{(s)*}(t') \rangle \\
&+ \langle z_h^{(n)}(t) z_h^{(o)*}(t') \rangle + \langle z_h^{(o)}(t) z_h^{(s)*}(t') \rangle + \langle z_h^{(o)}(t) z_h^{(n)*}(t') \rangle.
\end{aligned} \tag{7.2}$$

Because $z_h^{(s)}(t)$, $z_h^{(n)}(t)$ and $z_h^{(o)}(t)$ are independent random variables and

$$\langle z_h^{(s)}(t)\rangle = \langle z_h^{(n)}(t)\rangle = 0, \tag{7.3}$$

the last six terms on the right hand side of eq. (7.2) are all zeros. Therefore

$$\langle z_h(t) z_h^*(t')\rangle = \langle z_h^{(s)}(t) z_h^{(s)*}(t')\rangle + \langle z_h^{(n)}(t) z_h^{(n)*}(t')\rangle + \langle z_h^{(o)}(t) z_h^{(o)*}(t')\rangle. \tag{7.4}$$

This means that the autocorrelation function of the total output signal is a sum of autocorrelation functions of the scattering signal, noise and offset. Hence, in order to obtain a lag estimate $\langle x_h^{(s)}(i) x_h^{(s)*}(j)\rangle$ of the scattering signal, one has to measure the lag estimates of the total signal, noise and offset and the result is given by

$$\langle x_h^{(s)}(i) x_h^{(s)*}(j)\rangle = \langle x_h(i) x_h^*(j)\rangle - \langle x_h^{(n)}(i) x_h^{(n)*}(j)\rangle - \langle x_h^{(o)}(i) x_h^{(o)*}(j)\rangle. \tag{7.5}$$

In principle, eq. (7.4) implies that the lag estimates of noise and offset should be measured simultaneously with the lag estimate of the total signal. The noise, however, has no height dependence as the scattering signal does, and therefore its statistical properties can equally well be measured at a different time. A normal practice is to carry out noise measurements and the calibration of the received power just before the transmission on each channel. In this way the previous transmitted pulses are already travelling well beyond the ionosphere and no significant scattering signal is received from them.

When a noise measurement is carried out, the observation also contains the offset signal. Therefore, if the total signal in a noise measurement is $z_{hn}(t)$, the autocorrelation function of the noise signal is

$$\langle z_h^{(n)}(t) z_h^{(n)*}(t')\rangle = \langle z_{hn}^{(n)}(t) z_{hn}^{(n)*}(t')\rangle - \langle z_h^{(o)}(t) z_h^{(o)*}(t')\rangle. \tag{7.6}$$

Thus we have to determine an estimate for the offset term before we can determine the noise term estimate. On the other hand, one might say that any measurement of the offset signal also contains noise and, at the first sight, we may think to find ourselves in a vicious circle. Luckily enough, the situation is better than that since a solution to the problem is offered by the different correlation lengths of the noise and offset signals.

Obviously the post-detection filter restricts the bandwidth of the noise and increases its correlation length. If the noise signal entering the post-detection filter is $z^{(n)}(t)$, the filtered noise signal is

$$z_h^{(n)}(t) = (h * z^{(n)})(t) \tag{7.7}$$

and the noise term in eq. (7.4) is

$$\begin{aligned} \langle z_h^{(n)}(t) z_h^{(n)*}(t')\rangle &= \langle \int_{-\infty}^{\infty} h(t-\tau) z^{(n)}(\tau) d\tau \cdot \int_{-\infty}^{\infty} h(t'-\tau') z^{(n)*}(\tau') d\tau' \rangle \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(t-\tau) h(t'- \ tau') \langle z^{(n)}(\tau) z^{(n)*}(\tau')\rangle d\tau d\tau'. \end{aligned} \tag{7.8}$$

Since $z^{(n)}(t)$ is white noise, its correlation length is zero and we can write

$$\langle z^{(n)}(t) z^{(n)*}(t') \rangle = R P_n \delta(t - t'), \tag{7.9}$$

where P_n is the noise power and R is the receiver resistance. Then the autocorrelation function of the filtered noise is

$$\begin{aligned} \langle z_h^{(n)}(t) z_h^{(n)*}(t') \rangle &= R P_n \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(t - \tau) h(t' - \tau') \delta(\tau - \tau') d\tau d\tau' \\ &= R P_n \int_{-\infty}^{\infty} h(t - \tau) h(t' - \tau) d\tau \\ &= R P_n \int_{-\infty}^{\infty} h(\tau'') h(\tau'' - t' + t) d\tau''. \end{aligned} \tag{7.10}$$

The integral on the bottom line of eq. (7.9) is obviously the unnormalised autocorrelation function R_h of the impulse response, which has been previously used in eq. (4.26). Hence we see that

$$\langle z_h^{(n)}(t) z_h^{(n)*}(t') \rangle = R P_n R_h(t - t'). \tag{7.11}$$

This means that the autocorrelation length of filtered noise is equal to the length of the the autocorrelation function of the filter impulse response.

Because the offset voltage varies very slowly, its autocorrelation function is much longer than that of the scattering signal or noise. Therefore the offset autocorrelation function can be taken as constant within a lag interval longer than the length of the autocorrelation function of the scattering signal. This means that an estimate of the offset term in eq. (7.5) can be obtained by measuring the autocorrelation function of the total signal at such long lags that the autocorrelation function of the scattering signal is zero. At these lag values the noise autocorrelation function is also zero and therefore no noise term is present in the calculated autocorrelation function estimate. Hence the offset correction is made by subtracting the same constant at all lag values and it is done in the same way both in calculating the noise correction term and in calculating the lag estimates of the scattering signal. This is how we find our way out of the vicious circle.

Sometimes a single offset term is measured for the whole height profile, but often the offset is measured for each gate separately. This is a good practice, because the offset can vary from gate to gate. The reason for this is that during each period of transmission the receiver experiences an electric shock. Before the beginning of reception the receiver has not necessarily recovered from the shock completely, especially if the lag profiles start at low altitudes. Therefore an offset voltage changing from gate to gate may be present in the lower part of the profile, in the upper part the offset voltage is more likely to be constant. In all cases, however, the time scale of offset variations is such that the above assumption on constant offset voltage during a length of signal autocorrelation function is valid.

In the case of zero lag eq. (7.11) gives

$$\langle z_h^{(n)}(t) z_h^{(n)*}(t) \rangle = \langle |z_h^{(n)}(t)|^2 \rangle = RP_n R_h(0), \tag{7.12}$$

so that solving for RP_n we obtain

$$RP_n = \frac{\langle |z_h^{(n)}(t)|^2 \rangle}{R_h(0)}, \tag{7.13}$$

and the noise autocorrelation function can be put in the form

$$\langle z_h^{(n)}(t) z_h^{(n)*}(t') \rangle = \frac{R_h(t'-t)}{R_h(0)} \langle |z_h^{(n)}(t)|^2 \rangle. \tag{7.14}$$

This means that, due to filtering, the correlation length of the noise increases so much that a fraction $R_h(t'-t)/R_h(0)$ of the noise power "leaks" to the lag $(t'-t)$.

An incoherent scatter experiment is usually designed in such a way that $R_h = 0$ already for the first lag. For instance in multipulse experiments (see Chapter 5.3) the length of the impulse response is equal to the pulse length T and then $R_h \neq 0$ only within the interval $(-T, T)$. The sampling interval is $2T$ so that the first lag is also obtained at $2T$. Thus the lag ambiguity function of the first lag (see Chapter 4.4) extends from T to $3T$ which is beyond the range of the noise correlation function and the noise term is zero already in the first lag. When an experiment is designed following this principle, the noise correction must be made only to the zero lag.

In the case of Barker coding the situation is a bit more complicated. A detailed analysis shows that in Barker-coded multipulse experiments a small but non-negligible fraction of noise leaks to the first lag when the length of the impulse response is equal to the bit length. The contamination can be removed by adding a "zero bit" with no transmission at the end of each sequence of 13 bits and also stretching all gaps correspondingly.

Still another point associated with noise is that, in addition to thermal and astronomical sources, the signal itself can act as a source of noise. Each signal sample contains information from a height range determined by the amplitude ambiguity function and, when two signal samples are used in calculating lag estimates other than zero lag, the corresponding height ranges only partly overlap. The correlating parts of the signals come from the common region and produce the autocorrelation function of the scattering signal. The other parts of the signals do not correlate and they show as additional noise known as self noise, which increases the statistical error of the results. Unlike ordinary noise, this is not present in the zero lag.

As a final comment it is worth pointing out that the lag estimates of the total signal, noise and offset are determined with a certain statistical accuracy. When a lag estimate of the scattering signal is calculated according to eq. (7.5), the errors of all three terms affect the error of the autocorrelation function which is used in determining the plasma parameters. This is the starting point in the error analysis of incoherent scatter measurements.