CHAPTER 5:   WAVE SENSING SYSTEMS

INTRODUCTION

This chapter addresses both passive and active wave sensing systems, starting with interferometry and progressing to radar and synthetic aperture radar (SAR).

5.1 PASSIVE APERTURE SYNTHESIS

Wave sensing systems often employ multiple apertures intercepting signals that are then combined to yield images representing the angular distribution and correlation properties of those arriving signals. Single aperture systems were discussed in Chapter 3. To understand how signals intercepted at various positions might be combined to yield images, it is useful to understand the statistical relationships among these signals as a function of the arriving wave distribution.

Consider the aperture in the x-y plane illustrated in Figure 5.1-1. Assume multiple plane waves are arriving from different directions φ_x, φ_y. We might imagine each such arriving plane wave to have been radiated by a point source at infinity located in that direction; there might be a continuum of such point sources radiating statistically independent signals. Throughout this discussion such statistical independence will be assumed.

The total observed complex electric field $\overline{\underline{E}}(x, y)$ can be represented as an integral over 4π steradians of the contributions associated with each of these separate plane waves. In this case the total electric field can be written as

$$\overline{\underline{E}}(x, y) \cong \int_{4\pi} \overline{\underline{E}}\left(\varphi_x, \varphi_y\right) e^{+\frac{2\pi}{\lambda}\left(x \sin \varphi_x + y \sin \varphi_y\right)} d\Omega \tag{5.1.1}$$

where $\overline{\underline{E}}(\varphi_x, \varphi_y)$ is the uniform plane-wave expansion characterizing the arriving signals. The phase factor is a simple consequence of geometry.
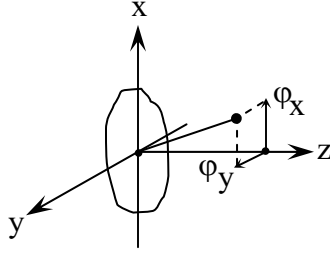
Figure 5.1-1 Aperture intercepting arriving plane-wave distribution

In practice we are usually interested in angles φ_x, φ_y that are small compared to unity, in which case the plane wave expansion associated with a given aperture distribution is approximately:

$$\overline{\underline{E}}(\varphi_x, \varphi_y) \cong \frac{1}{\lambda^2} \int_A \overline{\underline{E}}(x, y) e^{-j\frac{2\pi}{\lambda}(x\varphi_x + y\varphi_y)} dxdy \qquad (5.1.2)$$

Again the phase factor has a simple geometric explanation.

These first two equations can be simplified further if we define the new variables $x_\lambda \triangleq x/\lambda$, and $y_\lambda \triangleq y/\lambda$:

$$\overline{\underline{E}}(x_\lambda, y_\lambda) \cong \int_{4\pi} \overline{\underline{E}}(\varphi_x, \varphi_y) e^{j2\pi(x_\lambda \varphi_x + y_\lambda \varphi_y)} d\Omega \qquad (5.1.3)$$

$$\overline{\underline{E}}(\varphi_x, \varphi_y) \cong \iint_A \overline{\underline{E}}(x, y) e^{-j2\pi(x_\lambda \varphi_x + y_\lambda \varphi_y)} d\Omega \qquad (5.1.4)$$

This Fourier relationship, (5.1.3) and (5.1.4), can be related to an associated Fourier transform between the autocorrelation function of the electric field in the aperture and the angular intensity distribution associated with the transmitted or arriving plane waves:

$$\begin{array}{ccc}
\overline{\underline{E}}(x_\lambda, y_\lambda) & \overset{\sim}{\leftrightarrow} & \overline{\underline{E}}(\varphi_x, \varphi_y) \\
\downarrow & & \downarrow \\
\underline{R}_{\underline{E}}(\overline{\tau}_\lambda) & \overset{\sim}{\leftrightarrow} & \left|\overline{\underline{E}}(\varphi_x, \varphi_y)\right|^2 \propto G(\overline{\varphi}) \text{(transmitting)}
\end{array} \qquad (5.1.5)$$

where the autocorrelation function of the complex aperture field is defined as:

$$\underline{R}_{\underline{E}}\left(\overline{\tau}_\lambda\right) \triangleq \iint\limits_{-\infty}^{\infty} \overline{\underline{E}}\left(\overline{r}_\lambda\right)\overline{\underline{E}}^*\left(\overline{r}_\lambda - \overline{\tau}_\lambda\right)dx_\lambda dy_\lambda \tag{5.1.6}$$

Note that the aperture electric field $\overline{\underline{E}}$ is not stochastic; stochastic signals are considered later. Note that the transmitted intensity distribution $\left|\overline{\underline{E}}\left(\varphi_x,\varphi_y\right)\right|^2$ is proportional to the gain $G\left(\overline{\varphi}\right)$ that the aperture would have if it were a transmitting antenna.

Equation (5.1.5) quickly leads to a useful limit on the maximum angular resolution obtainable from a finite aperture if perfect image reconstruction is desired. For example, assume the aperture is uniformly illuminated, which can be represented by a cylindrical form for $\overline{\underline{E}}\left(x_\lambda,y_\lambda\right)$, as illustrated in Figure 5.1-2. Its autocorrelation function $R_{\overline{\underline{E}}}\left(\overline{\tau}_\lambda\right)$ therefore also has a finite region of support, which is a circle of radius D/λ.
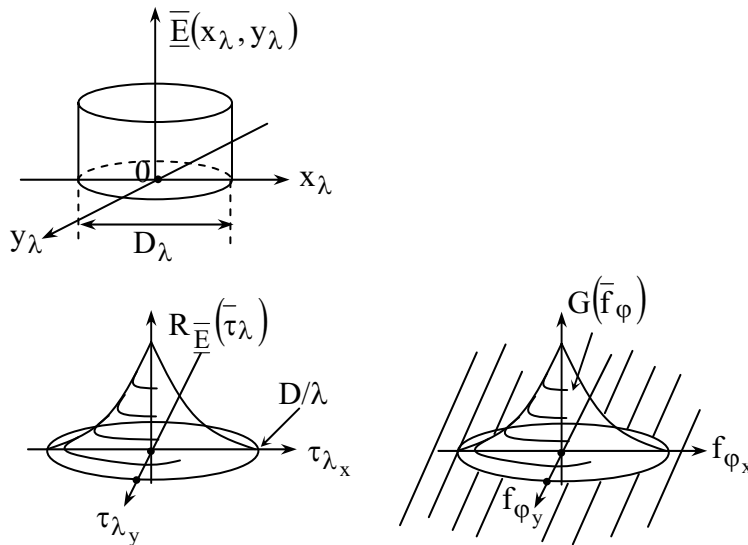

Figure 5.1-2   Aperture excitation function and corresponding aperture autocorrelation function and gain angular-frequency response.

Equation (5.1.5) then suggests that the gain $G\left(\overline{\varphi}\right)$ of this transmitting aperture is the Fourier transform of this pointed conical autocorrelation function. The Fourier transform of the gain function $G\left(\overline{\varphi}\right)$ can alternatively be interpreted as the angular spectral response of the aperture $G\left(\overline{f}_\varphi\right)$, where f_φ is defined as cycles per view-angle radian. That is, an antenna pattern with much fine structure (good angular resolution) would have significant amplitude at higher angular frequencies f_φ. An important result follows:

$$R_{\overline{\underline{E}}}\left(\overline{\tau_\lambda}\right) \leftrightarrow G\left(\overline{\varphi}\right) \leftrightarrow G\left(\overline{f}_\varphi\right)$$

$$\text{so } R_{\overline{\underline{E}}}(\tau) \cong G\left(f_\varphi\right)$$

$$(5.1.7)$$

The form of the aperture autocorrelation function $R_{\overline{\underline{E}}}\left(\overline{\tau_\lambda}\right)$ is seen to be the same as the aperture angular response function $G(f_\varphi)$, which has no response beyond a maximum angular frequency $f_{max} = D/\lambda$ cycles/radian. That is, a finite aperture of diameter D is blind to Fourier components in the angular intensity distribution at frequencies above D/λ cycles/radian. Angular frequency components above this D/λ limit can be estimated only by using *a priori* information.

This relationship between the power received by an antenna $T_A\left(\overline{\varphi}\right)$ as a function of angle $\overline{\varphi}$ and the angular-frequency response (or gain) of the antenna $G\left(\overline{f}_\varphi\right)$ can be represented by:

$$
\begin{array}{cccc}
T_A\left(\overline{\varphi}\right)= & G\left(\overline{\varphi}\right) & * & T_B\left(\overline{\varphi}\right) \\
\updownarrow & \updownarrow & \updownarrow & \updownarrow \\
T_A\left(\overline{f}_\varphi\right)= & G\left(\overline{f}_\varphi\right) & \bullet & T_B\left(\overline{f}_\varphi\right)
\end{array}
$$

$$(5.1.8)$$

This expression makes clear that the sensed angular-frequency distribution of received power (or antenna temperature) $T_A\left(\overline{f}_\varphi\right)$ is zero where $G\left(\overline{f}_\varphi\right)$ is zero, i.e., beyond D/λ cycles/radian. Reactive antennas, not discussed here, can sense somewhat higher angular frequencies; this option is seldom used except for antennas smaller than ~λ/2.

The foregoing can now be generalized to the case of uncorrelated stochastic signals arriving from different directions. Exceptions to this assumption of zero correlation are rare and most commonly occur when the arriving signals are coherently reflected by obstacles within view of the receiving antenna. An example of such problems is the fading associated with broadcast television "ghosts."

We represent narrowband stochastic signals by the slowly varying random complex amplitude of the otherwise sinusoidal signal. That is, we let:

$$\overline{E}\left(x_\lambda, y_\lambda, t\right)\left(vm^{-1} \ ster^{-1}\right) = Re\left\{\overline{\underline{E}}\left(t, x_\lambda, y_\lambda\right)e^{j\omega t}\right\}$$

$$(5.1.9)$$

Where $\overline{\underline{E}}\left(t, x_\lambda, y_\lambda\right)$ is a slowly varying random signal.

Substituting (5.1.9) into (5.1.5), we obtain

$$\overline{\underline{E}}(x_\lambda,y_\lambda,t) \quad \leftrightarrow \quad \overline{\underline{E}}(\varphi_x,\varphi_y,t)$$
$$\Downarrow \qquad\qquad\qquad \Downarrow$$
$$E\left[R_{\underline{\overline{E}}}(\overline{\tau}_\lambda)\right] \qquad\qquad\qquad\qquad (5.1.10)$$
$$\|$$
$$\phi_{\overline{\underline{E}}}\left(\tau_{x_\lambda},\tau_{y_\lambda}\right) \quad \leftrightarrow \quad E\left[\left|\overline{\underline{E}}(\varphi_x,\varphi_y,t)\right|^2\right]$$

where $\overline{\underline{E}}(\varphi_x,\varphi_y,t)$ is the slowly varying plane wave expansion corresponding to the slowly varying complex aperture distribution $\overline{\underline{E}}(x_\lambda,y_\lambda,t)$. The double arrows imply irreversibility for two reasons: the autocorrelation operator $R_{\overline{\underline{E}}}(\overline{\tau}_\lambda)$ is irreversible, and so is the expectation operator $E\left[R_{\underline{E}}\right]$. We can replace the expected value of the aperture autocorrelation function by its equivalent notation used previously for stochastic signals, $\varphi_{\overline{\underline{E}}}\left(\tau_{x_\lambda},\tau_{y_\lambda}\right)$, which is the Fourier transform of $E\left[\left|\overline{\underline{E}}(\varphi_x,\varphi_y,t)\right|^2\right]$. If the units for $\overline{\underline{E}}(x_\lambda,y_\lambda,t)$ are volts/meter, then the units for $\varphi_{\overline{\underline{E}}}\left(\tau_{x_\lambda},\tau_{y_\lambda}\right)$ are $(\text{volts/meter})^2$, and the units for $E\left[\left|\overline{\underline{E}}(\varphi_x,\varphi_y,t)\right|^2\right]$ are $(\text{volts/meter radian})^2$.

If the plane wave expansion corresponds to arriving waves associated with an angular intensity distribution $I(\varphi_x,\varphi_y,f)\left(\text{Wm}^{-2}\text{Hz}^{-1}\text{ster}^{-1}\right)$, then this intensity distribution can be readily found via:

$$\frac{\phi_{\overline{\underline{E}}}\left(\tau_{x_\lambda},\tau_{y_\lambda},f\right)}{2\eta_o B} \leftrightarrow \frac{E\left[\left|\overline{\underline{E}}(\varphi_x,\varphi_y,f)\right|^2\right]}{2\eta_o B} = I(\varphi_x,\varphi_y,f) \qquad (5.1.11)$$

where the units for these three terms are: $(\text{vm}^{-1})^2 \text{ ohm}^{-1} \text{ Hz}^{-1}$, $(\text{vm}^{-1}\text{rad}^{-1})^2 \text{ ohm}^{-1} \text{ Hz}^{-1}$, and $\text{Wm}^{-2} \text{Hz}^{-1} \text{ster}^{-1}$, respectively, where radians and steradians are not physical units, but only serve a mnemonic purpose.

As an example of how this formalism might be utilized, consider a square thermal source of $T_B(^\circ K)$ located at $\varphi_x = \varphi_y = 0$, as illustrated in Figure 5.1-3.
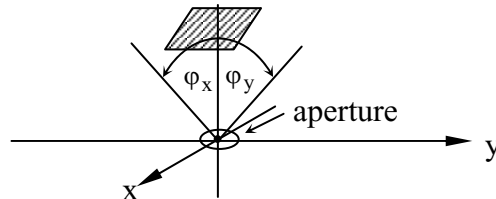

Figure 5.1-3 Aperture response to a square thermal source

Assume the source is 0.01 radians on a side, so that the solid angle $\Omega_s \cong 10^{-4}$ steradians. In this case, for waves arriving from ϕ_x, φ_y within Ω_s.

$$E\left[\left|\overline{E}\left(\varphi_x,\varphi_y\right)\right|^2\right] = \left(kT_B/2\lambda^2\right)2\eta_o B \tag{5.1.12}$$

where B is the bandwidth (Hz) over which the electric field is to be measured, η_o is the impedance of free space, and T_B (°K) is the brightness temperature of the square source. The corresponding intensity arriving at the aperture from angles within the source is

$$I_x\left(\varphi_x,\varphi_y,f\right) = kT_B/2\lambda^2 \ \left[\text{Wm}^{-2}\ \text{Hz}^{-1}\text{ster}^{-1}\right] \tag{5.1.13}$$

The resulting complex electric field autocorrelation function within the aperture can be found from the square intensity distribution via a Fourier transform of the form (5.1.3), as suggested in Figure 5.1-4, where we have used (5.1.11) to introduce the factor $2\eta_o B$.
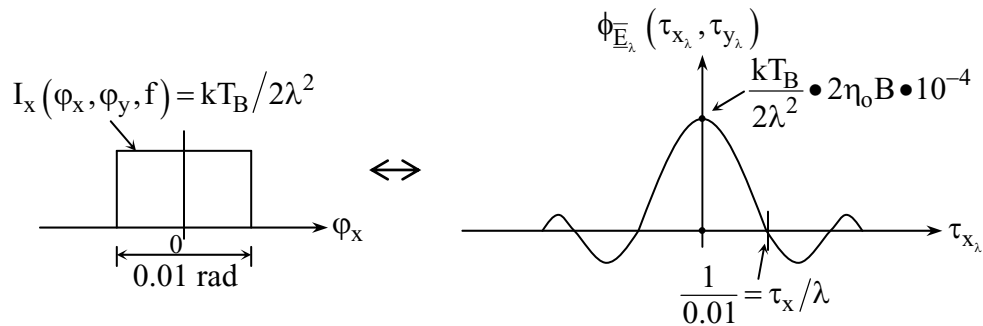
Figure 5.1-4 Fourier relationship between angular intensity function and complex field autocorrelation function in an aperture.

Note that

$$\phi_{\underline{E}_x}(0,0)/2\eta_o = E\left\{\left|\overline{E}(x,y,t)\right|^2\right\}/2\eta_o = \frac{kT_B}{2\lambda^2}\Omega_s B = S_o \tag{5.1.13}$$

Where S_o is the source flux (watts/m^2). Thus we can observe the complex field autocorrelation function $\phi_{\underline{E}_x}\left(\tau_{\lambda_x}, \tau_{\lambda_y}\right)$, and from its Fourier transform learn the shape of the intensity distribution. Then, using Figure 5.1-4 or (5.1.13), the value of T_B can readily be obtained.

Previously we have considered autocorrelation functions in time and in space, and so we may also consider autocorrelations of the complex field as a function of simultaneous offsets in space and time,. For example, for a given polarization, we can take the Fourier transform between time and frequency to obtain:

$$\phi_{\overline{E}_x}\left(\tau_{\lambda_x},\tau_{\lambda_y},\tau\right)/2\eta_o\left[Wm^{-2}\right] \overset{\substack{time/\\frequency}}{\longleftrightarrow} \phi_{\overline{E}_x}\left(\tau_{\lambda_x,}\tau_{\lambda_y},f\right)/2\eta_o\left[Wm^{-2}Hz^{-1}\right]$$

$$\underset{angle}{space}\diagup \updownarrow\left(\overline{\tau\lambda}\leftrightarrow\overline{\phi}\right) \qquad\qquad (5.1.14)$$

$$I_x\left(\overline{\phi},f\right)=\frac{kT_B\left(\overline{\phi},f\right)}{2\lambda^2}\left[Wm^{-2}Hz^{-1}ster^{-1}\right]$$

In (5.1.14) the Fourier transform between time and frequency can be followed by the Fourier transform between spatial offset $\left(\overline{\tau_\lambda}\right)$ and angle $\left(\overline{\phi}\right)$. The separate time and space Fourier transforms in combination yield the transform between $\varphi_{\overline{E}_x}\left(\tau_{\lambda_x},\tau_{\lambda_y},\tau\right)$ and $I_x\left(\overline{\phi},f\right)$.

These Fourier relationships may now be used to characterize aperture synthesis systems which operate on signals that are intercepted by two or more small antennas and then processed so as to yield the angular response characteristics of much larger antennas. This process is called *aperture synthesis*.

Assume two small antennas can be moved within a rectangular field measuring τ_{λ_x} by τ_{λ_y} wavelengths. These dimensions define the maximum values of $\overline{\tau}$ for which the field autocorrelation function $\varphi_{\overline{E}}(\overline{\tau})$ can be measured. Interchanging the outputs of the two small antennas converts $\overline{\tau}$ to $-\overline{\tau}$, so that

$$\phi_{\overline{E}}\left(\overline{\tau}\right)=\phi_{\overline{E}}^{*}\left(-\overline{\tau}\right) \qquad\qquad (5.1.15)$$

As a result, the area covered in $\overline{\tau}$ space has four times the area of the available field, as suggested in Figure 5.1-5. Within a field of dimensions $\overline{\tau}_{\lambda_{max}}$ it is necessary only to measure $\varphi_{\overline{E}}(\overline{\tau})$ for values of $\overline{\tau}_\lambda$ in quadrants A and B because quadrants A* and B* contain the complex conjugates of observations made in quadrants A and B.
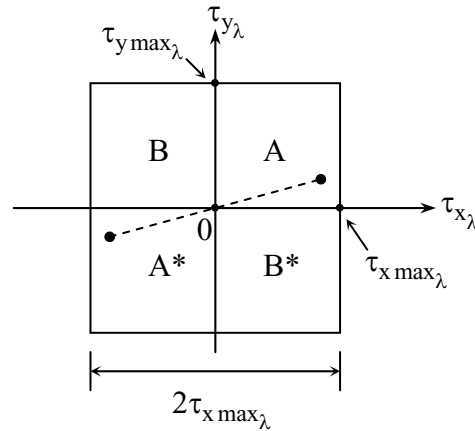
Figure 5.1-5 Observable range of $\bar{\tau}$ within a field of dimensions $\bar{\tau\lambda}_{max}$.

The dimensions of the observed space $\bar{\tau}_\lambda$ is $2\tau_{\lambda_{max}}$, or $2L_x/\lambda$ by $2L_y/\lambda$. These are the dimensions of the weighting function $W(\bar{\tau}_\lambda)$ which masks the true autocorrelation function $\varphi_{\underline{E}}(\bar{\tau}_\lambda)$, and therefore determines the angular resolution and effective beamwidth of the synthesized aperture:

$$W(\bar{\tau}_\lambda) \bullet \phi_{\underline{E}}(\bar{\tau}_\lambda)$$
$$\updownarrow \qquad\qquad\qquad (5.1.16)$$
$$W(\bar{\varphi}) * |\underline{E}(\bar{\varphi})|^2$$

Figure 5.1-6 illustrates this simple relationship graphically. The Fourier transform of a rectangular boxcar function is a two-dimensional sinc function with nulls at angles that are multiples of $\lambda/2L_x$ and $\lambda/2L_y$. Note that the synthesized antenna pattern actually has negative response values at some angles, unlike single-aperture antenna patterns for which the antenna gain is always positive. The 3-dB antenna beamwidths are slightly larger than the positions of the first nulls at $\lambda/2L$. If the weighting function $W(\bar{\tau}_\lambda)$ is tapered for larger values of τ_λ , the amplitudes of these sidelobes can be diminished at the expense of a small increase in antenna beamwidth.

244

$$W(\overline{\tau_\lambda}) \leftrightarrow W(\overline{\varphi}) =$$

$$2\tau_{x\,max_\lambda} = 2L_x/\lambda$$

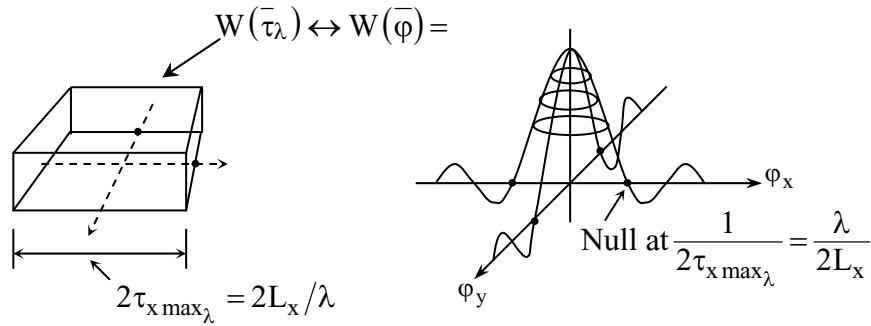Null at $\dfrac{1}{2\tau_{x\,max_\lambda}} = \dfrac{\lambda}{2L_x}$

Figure 5.1-6 Maximum-resolution antenna pattern synthesized using a finite rectangular field

Note that the beamwidth of the synthesized pattern available with a rectangular field of size L_x and L_y is half that achieved with a uniformly illuminated aperture having the same dimensions as that field; $\lambda/L_x, \lambda/L_y$. See Figure 5.1-2 for an example of the limits to the angular resolution of a finite aperture. Aperture synthesis permits substantially better angular resolution because each of the separate responses to pairs of antennas can be amplified independently before these point-pair responses are superimposed to yield the final synthesized system response.
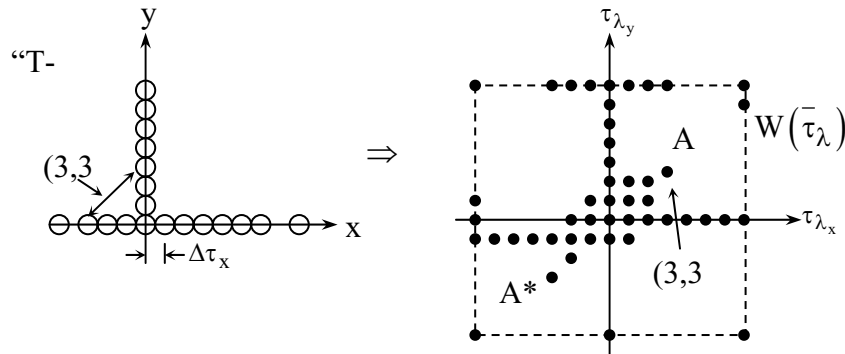
Figure 5.1-7 Observed sampled autocorrelation function for a discrete T-array of antennas

It is not necessary to position pairs of small antennas in all portions of the field. For example, the simple "T-array" in Figure 5.1-7 suffices to fill the rectangular field $\overline{\tau}_\lambda$ with a sampled grid of observations corresponding to all possible pairs of antennas in the T-array.

More efficient than the T-array in its use of a finite field is the "U-array", for which three sides of the field are lined with small aperture antennas.

245

Because the observations of $\underline{\varphi}\left(\bar{\tau}_\lambda\right)$ are generally comprised of discrete samples arranged on a rectangular grid, the weighting function $W\left(\bar{\tau}_\lambda\right)$ is both finite and sampled, with readily understood consequences. Since:

$$\frac{\phi\left(\tau_{x\lambda},\tau_{y\lambda},f\right)}{2\eta_o B} \leftrightarrow \frac{\left|\overline{E}\left(\bar{\varphi},f\right)\right|^2}{2\eta_o B} = I\left(\bar{\varphi},f\right)\left[\mathrm{Wm}^{-2}\mathrm{ster}^{-1}\mathrm{Hz}^{-1}\right] \tag{5.1.17}$$

it follows that multiplication by the sampled weighting function $W\left(\bar{\tau}_\lambda\right)$ produces the result:

$$\left[\phi_{\underline{\overline{E}}}\left(\bar{\tau},f\right)/2\eta_o B\right]\bullet W\left(\bar{\tau}_\lambda\right) \leftrightarrow W\left(\bar{\varphi}\right)*I\left(\bar{\varphi},f\right) \tag{5.1.18}$$

which can be represented graphically, as illustrated in Figure 5.1-8.
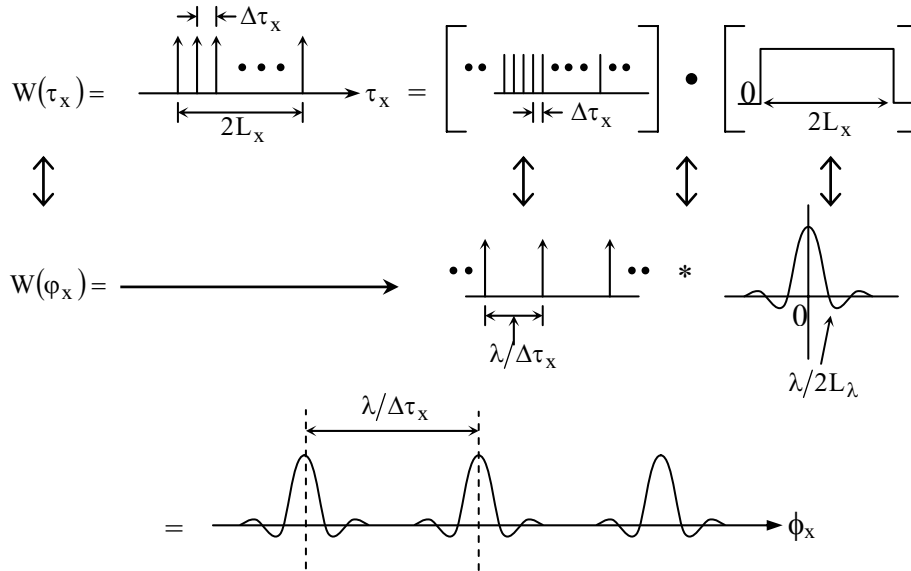
Figure 5.1-8   Aliasing of synthesized images due to discrete sampling of the complex field correlation function $\varphi_{\underline{\overline{E}}}\left(\bar{\tau}_\lambda\right)$

Note that the finite extent of the $W\left(\bar{\tau}\right)$ produces sinc functions of infinite extent, although they have generally negligible values for angles beyond $\lambda/2L$, and therefore synthesized images always risk aliasing, even for very small values of antenna displacement $\Delta\tau$. Fortunately, in many applications the source of interest is alone in the field of view of the individual antenna elements, and the rest of space is relatively empty. Examples of this situation occur when strong astronomical sources are imaged against the blank background of surrounding astronomical space, or when strong radiating objects are observed against an extended passive background.
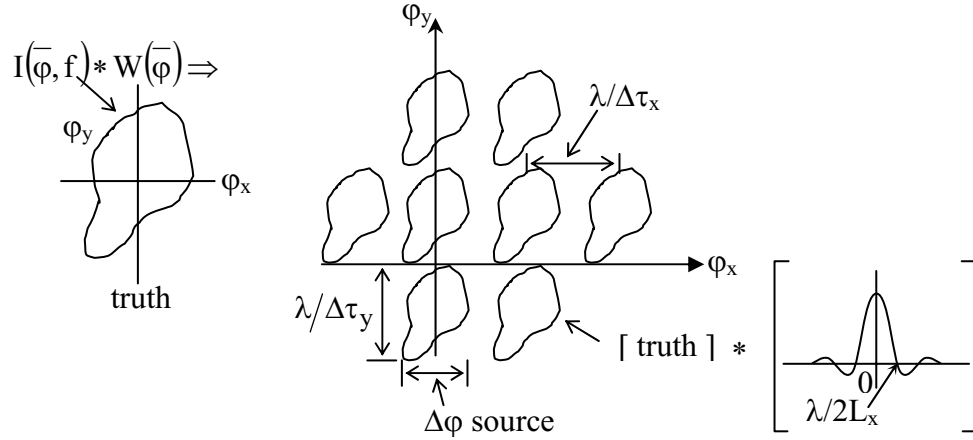
Figure 5.1-9   Image aliasing in synthesized images due to discrete antenna placement

For example, consider the case where the true source distribution in the sky is the pear-shaped object illustrated in Figure 5.1-9.   In this case the discrete set of observations $\left(\phi_{\underline{E}}(\overline{\tau}\lambda,f)/2\eta_o B\right)\bullet W(\overline{\tau}\lambda)$, when Fourier transformed, yields a set of duplicate aliased images with spaces between their centers of $\lambda/\Delta\tau_x$ and $\lambda/\Delta\tau_y$.   Each of these replicated images has been blurred by a sinc function having its first nulls at $\lambda/2L_x$ and $\lambda/2L_y$.   If the source does not have a small angular extent, then the resulting images may be aliased and overlapped, producing confusion which is difficult to remove without additional information.   The estimated retrieved source intensity distribution illustrated in Figure 5.1-9 is characterized by:

$$\hat{I}(\overline{\phi},f) = \left[I(\overline{\phi},f) * W(\overline{\phi})\right] \bullet G_{AB}(\overline{\phi}) \tag{5.1.19}$$

To minimize the effects of image aliasing it is therefore advisable to use spacings between adjacent antenna positions $\Delta\tau_\lambda$ which are less than the reciprocal of the source width in radians, and preferably a small fraction of that.

Often it is possible to obtain multiple antenna spacings with a smaller fixed number of antenna positions because the viewing geometry evolves in time producing effective antenna movement.   For example, astronomical aperture synthesis systems often use a small number of antennas which can be moved on a railroad track to a small number of discrete positions where, for each such position, the rotation of the earth sweeps the effective $\overline{\tau}$ across a significant portion of the space $\overline{\tau}\lambda$, as suggested in Figure 5.1-10.
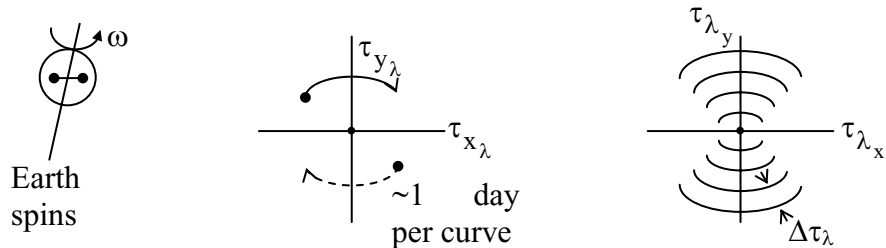
247

Figure 5.1-10 Typical $\underline{\varphi}(\bar{\tau}_\lambda)$ observed by two antennas over a single day and over four days.

As the earth spins a single pair of antennas observes a projected baseline $\bar{\tau}_\lambda$ that changes as suggested in the figure, sweeping out a curve in $\bar{\tau}_\lambda$ space, together with its complex conjugate mirror image. Repositioning the antennas daily produces an array of such curves, also as illustrated. To avoid aliasing it is important that the spacings between these curves, e.g. $\Delta\bar{\tau}_\lambda$, are small. Furthermore, if these curves do not cover the space $\bar{\tau}_\lambda$ completely, the resulting gaps can lead to excessive sidelobes in the synthesized image. The radioastronomy community typically replaces the variables τ_{λ_x} and τ_{λ_y} by their equivalents: u,v. An important advantage often enjoyed by the astronomical community is that the source features of interest are often simple, such as the source diameter, or the separation of two point images. In these simple cases very sparse coverage of the u,v plane suffices to yield the desired image parameters.

5.2 INTERFEROMETERS

Aperture synthesis involves mathematically combining the signals from many pairs of antennas, or a moveable pair, to form a synthesized image. In other cases, a single pair of antennas can provide the desired information. Interferometry was first demonstrated for astronomical purposes by Michaelson, who placed two small mirrors in the 100-inch aperture of the Mount Wilson telescope and caused the reflected beams to interfere, forming a fringe pattern where the observability of the fringes depended on the diameter of the star being observed. Similar interferometers are used for aperture synthesis and source localization. Figure 5.2-1 illustrates how signals from a pair of radio antennas might be generated before being combined.
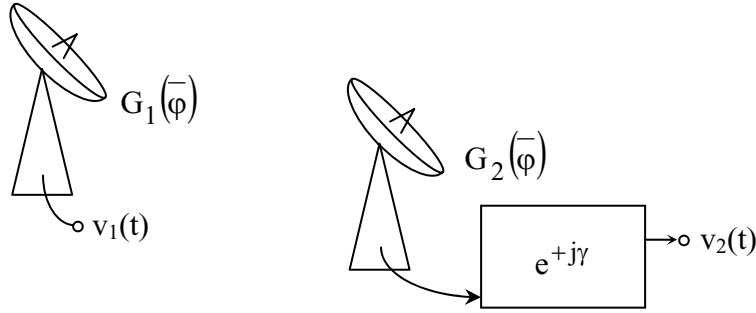
Figure 5.2-1   Characterization of a pair of antennas used for interferometry.

The output voltages from each antenna $v_i(t)$ are each proportional to the square roots of the corresponding antenna gain $G_i(\overline{\varphi})$. One of the antennas has an additional phase adjustment $e^{j\gamma}$. In the narrowband case where we can represent these signals as having slowly varying complex envelopes $\underline{E}(x, y, t)$, the voltages $v_i(t)$ may be represented as:

$$v_1(t) \cong k_1 R_e \left\{ \sqrt{G_1(\overline{\varphi})} \bullet \underline{E}(x, y, t)e^{j\omega t} \right\}$$

$$v_2(t) \cong k_2 R_e \left\{ \sqrt{G_2} \bullet \underline{E}(x - \tau_x, y - \tau_y, t) \, e^{j\gamma}e^{j\omega t} \right\}$$

(5.2.1)

where the constants k_i reflect the fact that each receiver may have amplifiers with different gains.

For many interferometers the computed output is the expected value of the product of the two output voltages:

$$E[v_1(t)v_2(t)] = \frac{k_1 k_2}{2} Re \left\{ \sqrt{G_1 G_2} \bullet E\left[\overline{\underline{E}}(x, y, t) \bullet \overline{\underline{E}}^*\left(x - \tau_x, y - \tau_y, t\right) \right] e^{-j\gamma} \right\}$$

(5.2.2)

where $G_1 G_2$ is often replaced by the cross gain G_{12}, and where we recall:

$$E[a(t)b(t)] = \frac{1}{2} R_e \left\{ \underline{A}\underline{B}^* \right\}$$

(5.2.3)

Equation (5.2.2) must be modified if the bandwidth of these signals exceed some limit which depends on the separation L between the antennas, and on the source angle φ_x projected on the x-axis connecting the two antennas, which are configured as suggested in Figure 5.2-2.
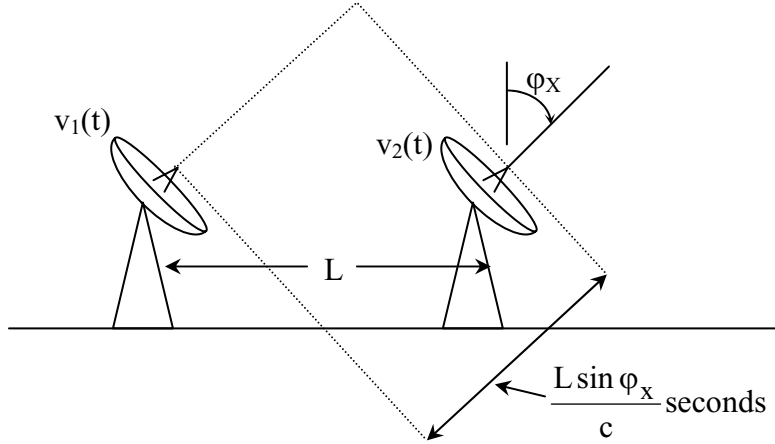
Figure 5.2-2 Envelope delay in a two-element interferometer

Bandwidth enters into (5.2.2) when the slowly varying envelope $\underline{E}(x, y, t)$ varies sufficiently rapidly that the delay in arrival time at the two antennas, $L \sin \varphi_x / c$, approaches the envelop correlation time $\tau \cong 1/2B$. In this case (5.2.2) may be written as:

$$E[v_1(t)v_2(t)] = \frac{k_1 k_2}{2} G_{12}(\overline{\varphi}_s) E\left[\left\{R_e \underline{E}\left(x, y, t + \frac{L \sin \varphi_x}{2c}\right)\underline{E}^*\right.\right.$$
$$\left.\left.\left(x - \tau_x, y - \tau_y, t - \frac{L \sin \varphi_x}{2c}\right) \bullet e^{j\omega L \sin \varphi_x / 2c} \ e^{-j\gamma}\right\}\right]$$

$$(5.2.4)$$

The expected value of the product of the two envelopes can be simplified, yielding:

$$E[v_1(t)v_2(t)] = \frac{k_1 k_2}{2} G_{12}(\overline{\varphi}_s) R_e\left\{e^{-j\gamma} e^{j2\pi L \sin \varphi_x / \lambda} \bullet \phi_{\underline{E}(t)}\left(\frac{L \sin \varphi_x}{c}\right)\right\} \qquad (5.2.5)$$

where the factor $e^{j\gamma} e^{j2\pi L \sin \varphi_x / \lambda}$ characterizes the monochromatic fringe, and $\phi_{\underline{E}(t)}(L \sin \varphi_x / c)$ characterizes the envelope that modulates the fringes. This fringe modulation envelope is a sinc function corresponding to the rectangular bandpass filter illustrated in Figure 5.2-3, and yields a null when $\sin \varphi_x = c/2BL$ seconds.
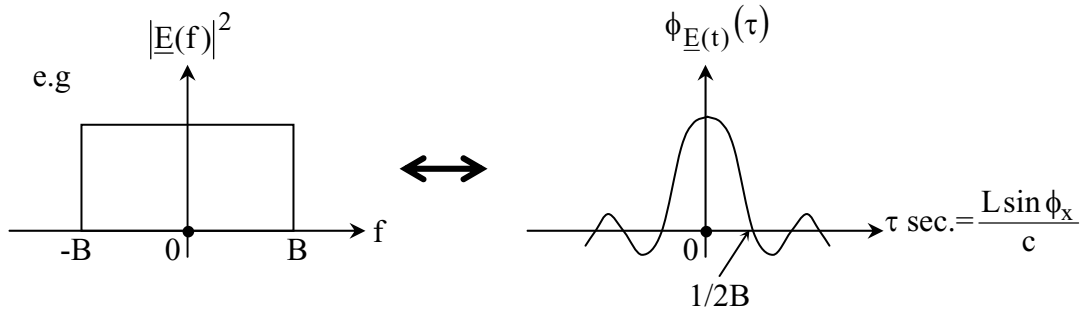
Figure 5.2-3   Fringe modulation envelope resulting from finite bandwidth in a two-element interferometer

The resulting combined fringe pattern is illustrated in Figure 5.2-4, which shows the influence of the cross gain G_{12}, the monochromatic fringe, and the fringe envelope.
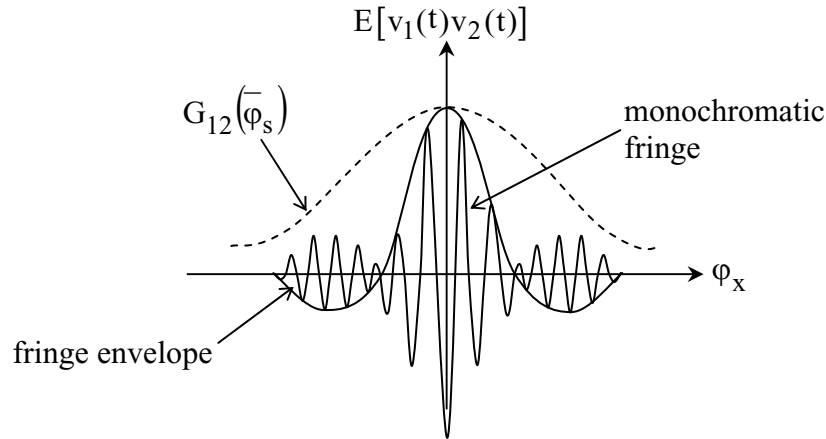

Figure 5.2-4   Bandwidth-limited angular response of polychromatic interferometer

The polychromatic fringe pattern illustrated in Figure 5.2-4 represents the superposition of a continuum of monochromatic fringe patterns at each frequency within the bandwidth B. Because all colors contribute equally to the central fringe, it is called the *white fringe*. For visible wavelengths, as we move toward the first null in the polychromatic fringe pattern the shorter-wavelength blue band approaches its null first, so the residual fringes are predominantly red. As the red polychromatic band later approaches its first null, the blue band has re-emerged and the fringes appear bluish. Further out in the fringe pattern the nulls of the different wavelengths are so out of phase that the colors become washed out and, with respect to the human visual system, again appear white.

The fringe envelope defined in (5.2.5), $\phi_{\underline{E}(t)}\left(L\sin\varphi_x/c\right)$, reaches its first null as suggested in Figure 5.2-3 when $\left(L\sin\varphi_x\right)/c = 1/B$ seconds, so that:

$$\varphi_{x_{null}} = \sin^{-1}(c/LB) \approx c/LB \text{ for } \varphi_x \ll 1 \tag{5.2.6}$$

Thus the angular resolution of a polychromatic system viewing a source which is white over the bandwidth B can have extremely high angular resolution when LB is large compared to c.

For example, consider an interferometer with two radio or optical sensors L = 100 meters apart that produce signals which are coherently combined to form a fringe pattern within a bandwidth B of 1 GHz. In this case (5.2.6) suggests the polychromatic first null will occur at an angle of $c/LB = 3 \times 10^8 / 100 \times 10^9 = 3 \text{ mrad} \cong 10 \text{ arc min}$. If the source being observed is approximately monochromatic $(B \lesssim c/\lambda)$, then the first null is governed instead by the monochromatic fringe pattern.

Equation (5.2.5) governs interferometers where a pair of signals is multiplied. In practice other interferometer circuits are often used to synthesize source angular distributions, as suggested in (5.1.10). Many interferometers simply add the two received signals a and b coherently before squaring them, as suggested in Figure 5.2-5 and equation (5.2.7). The overbars in (5.2.7) and Figure 5.2-6 signify the time average of the associated variable.

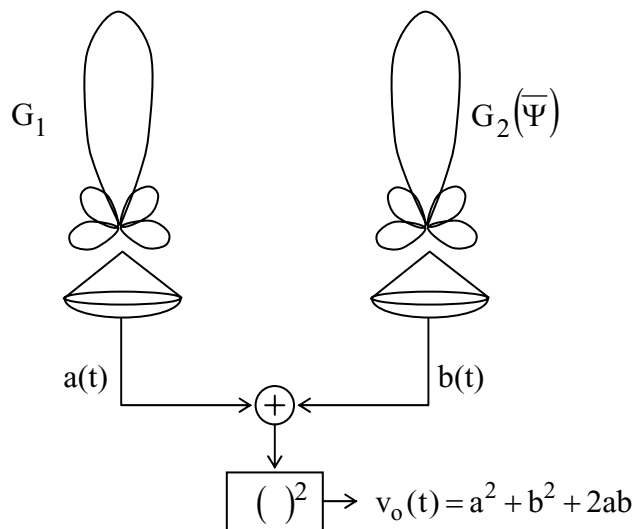$$\overline{v_o(\psi)} = \overline{a^2} + \overline{b^2} + 2\overline{ab} \tag{5.2.7}$$

Figure 5.2-5: Simple adding interferometer

The interferometer output $\overline{v_o(\psi)}$ given by (5.2.7) is illustrated in Figure 5.2-6, and takes the form of a fringe pattern with inter-null spacings of λ/L and having amplitudes modulated by the single-antenna gain pattern. We have assumed in Figure 5.2-6 that $G_1(\overline{\psi}) = G_2(\overline{\psi})$.
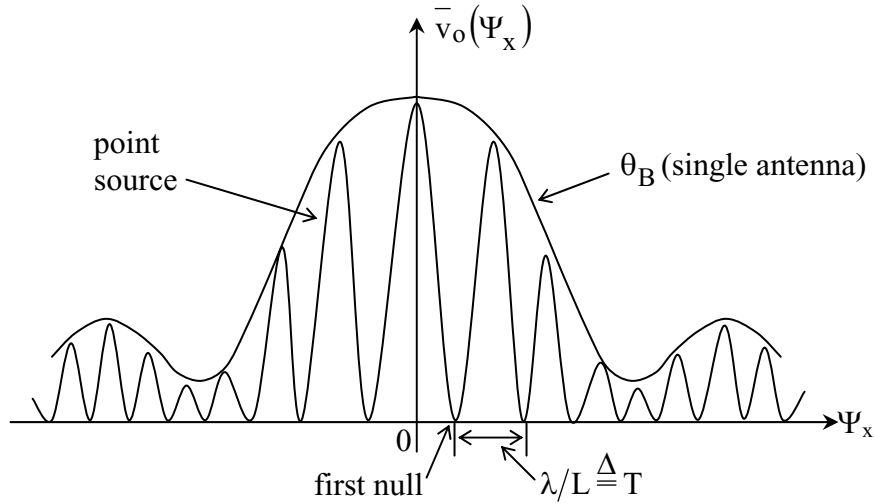
Figure 5.2-6: Point-source response of a symmetric two-antenna interferometer.

Nulls occur only when the two signals a and b have the same amplitude and are 180° out of phase, as suggested in Figure 5.2-7. If the signals are in phase when $\psi = 0$, then the first null occurs when:

$$\psi_{null} = \sin^{-1}(\lambda/2L) \cong \lambda/2L \text{ if } \lambda \ll L \tag{5.2.8}$$

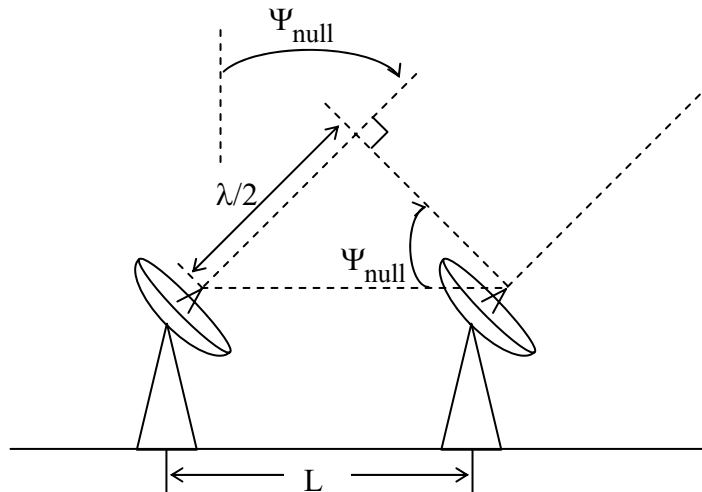Therefore the spacing between nulls is twice $\lambda/2L$, or λ/L.

Figure 5.2-7: Angular positions of nulls in a simple adding interferometer

If two point sources are present and radiating incoherently in the same frequency band, the response function of the interferometer becomes the linear superposition of the point source responses illustrated in Figure 5.2-6. If the two sets of nulls spaced at λ/L do not coincide perfectly, then the interferometer response to a spread source becomes that illustrated in Figure 5.2-8, where the fringe pattern now has a DC term and the fringes are bounded by envelopes having amplitudes v_{max} and v_{min} at the origin where $\psi_x = 0$. Both figures 5.2-6 and 5.2-8 illustrate the DC term $\overline{a^2 + b^2}$, and the AC term $\overline{2ab}$ given by (5.2.7).
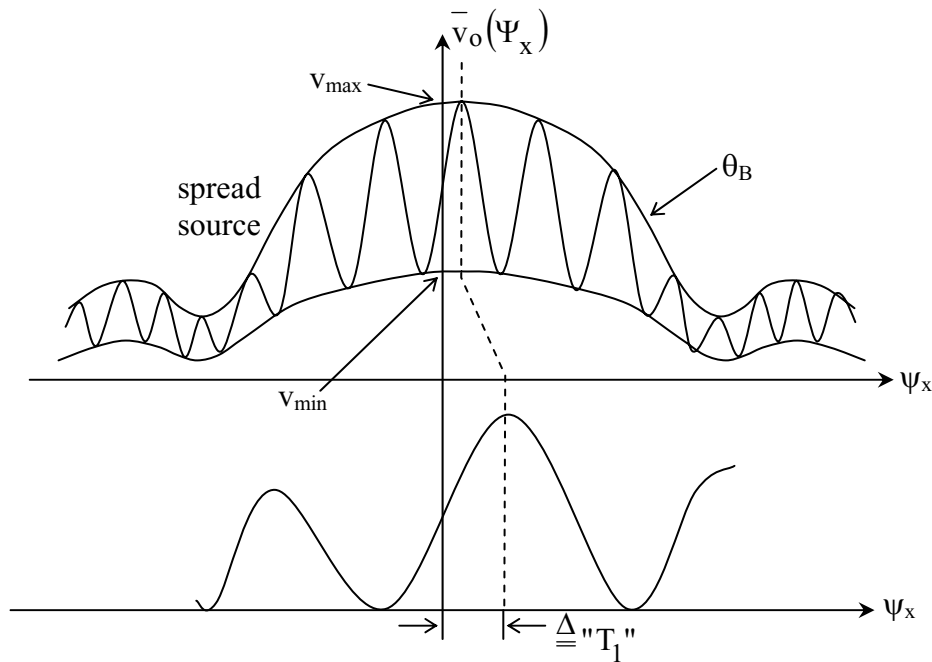
Figure 5.2-8: Response of a simple adding interferometer to a spread source

Simple interferometers yield four parameters of interest when observing a spread source seen against a blank background; these are v_{max}, v_{min}, T_1, and T, where T is defined as the distance between nulls λ/L. The *complex fringe visibility* $\underline{V}$ can be deduced from the interferometer output:

$$\underline{V} \triangleq \frac{v_{max} - v_{min}}{v_{max} + v_{min}} e^{2\pi T_1/T} \propto \phi_{\underline{E}}\left(\overline{\tau\lambda}\right) \leftrightarrow T_A\left(\overline{\psi}\right) \tag{5.2.9}$$

It can be shown that the complex fringe visibility $\underline{V}\left(\overline{\tau\lambda}\right)$ is directly proportional to $\phi_{\underline{E}}\left(\overline{\tau\lambda}\right)$, which is the approximate Fourier transform of the angular antenna temperature distribution $T_A\left(\overline{\psi}\right)$.

The Fourier relationship between the aperture autocorrelation function and the source angular distribution, or between the fringe visibility function and the source angular distribution, can be understood in terms of the sinusoidal response of the adding interferometer to a monochromatic point source. This relationship can be seen mathematically by considering (5.2.11), which states once again that the antenna temperature, as a function of antenna pointing angle $\overline{\psi}$, is the convolution of the antenna gain $G(\overline{\psi})$ with the sky brightness distribution $T_B(\overline{\psi})$. The antenna temperatures could, of course, characterize non-thermal sources such as communication signals provided that the signals arriving from different directions are uncorrelated. If we Fourier transform each term and operator in (5.2.11), we obtain (5.2.10), which states that the observed autocorrelation function of the aperture field distribution $\phi_A(\overline{\tau\lambda})$ is the true field autocorrelation function $\underline{\phi}_E(\overline{\tau\lambda})$ times the windowing function introduced by the finite nature of the observing aperture $R_{\underline{E}}(\overline{\tau\lambda})$, where $R_{\underline{E}}(\overline{\tau\lambda})$ is suggested by Figure 5.2-9 for a uniformly illuminated square aperture of dimensions D_λ.
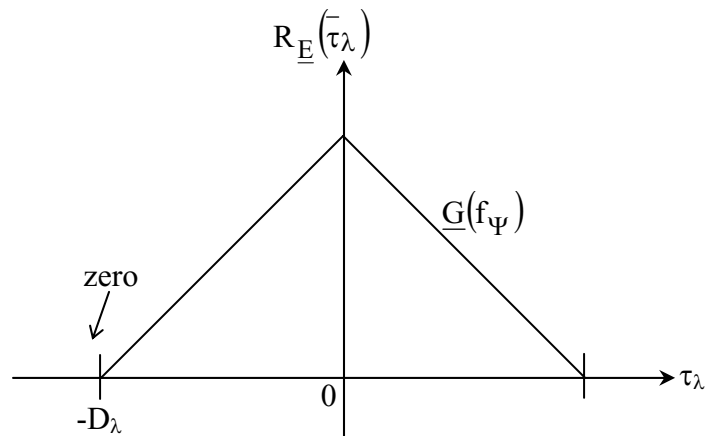

Figure 5.2-9: Field autocorrelation function characterizing a uniformly illuminated square aperture

We may also compute the Fourier transforms of each of the variables and operators in (5.2.11) to yield (5.2.12), which states that the angular Fourier transform of the antenna temperature $\underline{T}_A(f_\psi)$ is the product of the antenna gain and sky brightness Fourier representations

$$\phi_A(\overline{\tau\lambda}) = R_{\underline{E}}(\overline{\tau\lambda}) \quad \bullet \quad \phi_E(\overline{\tau\lambda}) \tag{5.2.10}$$

$$\updownarrow \qquad \sim\!\!\updownarrow \quad \sim\!\!\updownarrow \quad \sim\!\!\updownarrow$$

$$T_A(\overline{\psi}) = \underline{G}(\overline{\psi}) \quad * \quad T_B(\overline{\psi}) \tag{5.2.11}$$

$$\updownarrow \qquad \sim\!\!\updownarrow \quad \sim\!\!\updownarrow \quad \sim\!\!\updownarrow$$

$$\underline{T}_A(f_\psi) \quad \underline{G}(f_\psi) \quad \bullet \quad T_{\underline{B}}(f_\psi) \tag{5.2.12}$$

As noted earlier, the angular spectral response of the antenna gain $\underline{G}(f_\psi)$ is the same as the autocorrelation function of the aperture field distribution, and therefore is zero for a broad range of frequencies when the aperture dimensions D_λ are finite.

In the case of the ideal two-point interferometer in Figure 5.2-1, the aperture field autocorrelation function $R_{\underline{E}}(\overline{\tau\lambda})$ consists of three impulses spaced D_λ apart, where the central impulse has twice the amplitude of the others. This trio of impulses also characterizes the angular frequency response for that interferometer spacing, $\underline{G}(f_\psi)$, which therefore consists of a cosine wave plus a DC term when $\gamma = 0$, and a sine wave plus a DC term when $\gamma = \pi/2$. Together they represent the real and imaginary parts of $\underline{G}$. These cosine and sine responses of the two element adding interferometer are illustrated in Figure 5.2-10.
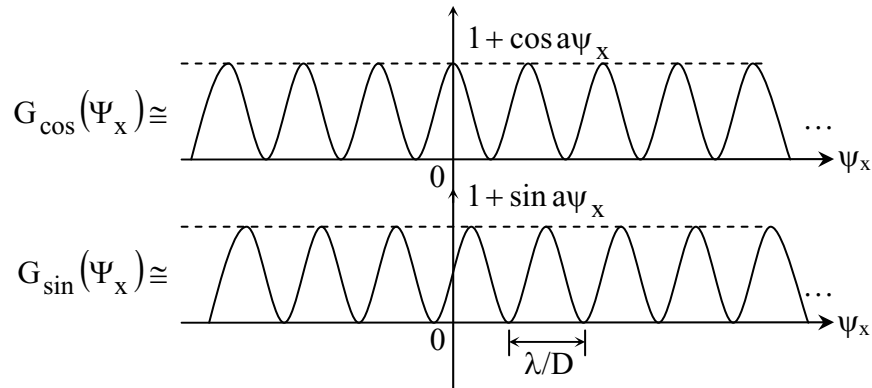

Figure 5.2-10: Cosine and sine responses for a two-element adding interferometer viewing a point source at $\psi = 0$

The antenna temperature $T_A(\overline{\psi}=0)$ is the interferometer gain function $G(\overline{\psi}')$ times the source brightness temperature $T_B(\overline{\psi}')$, integrated over all angles $\overline{\psi}'$. If the two path lengths into the adder and square law device illustrated in Figure 5.2-5 are equal, then the antenna gain $G(\overline{\psi})$ is the cosine response of Figure 5.2-10. If a phase delay γ of 90° is added to one arm of this interferometer, the response peaks at a slightly displaced angle ψ_x, as suggested by the sine term of Figure 5.2-10. Therefore, once the DC terms are omitted, the remainder are simply the

sine and cosine terms of the source distribution at the angular frequency λ/D, where D is the separation between the two interferometer apertures. If we observe enough frequency components in two dimensions, the result can be Fourier transformed to yield the source angular distribution, as shown previously.

The ideal monochromatic point response of a two-aperture interferometer is a cosine wave, as illustrated in Figure 5.2-10, having a period of λ/D. If the source and receiver are broadband, then the fringe spacing λ/D varies over this band so that the interferometer output is actually the superposition of the continuum of fringe patterns corresponding to all the frequencies which have been added in phase, as suggested in Figure 5.2-11.
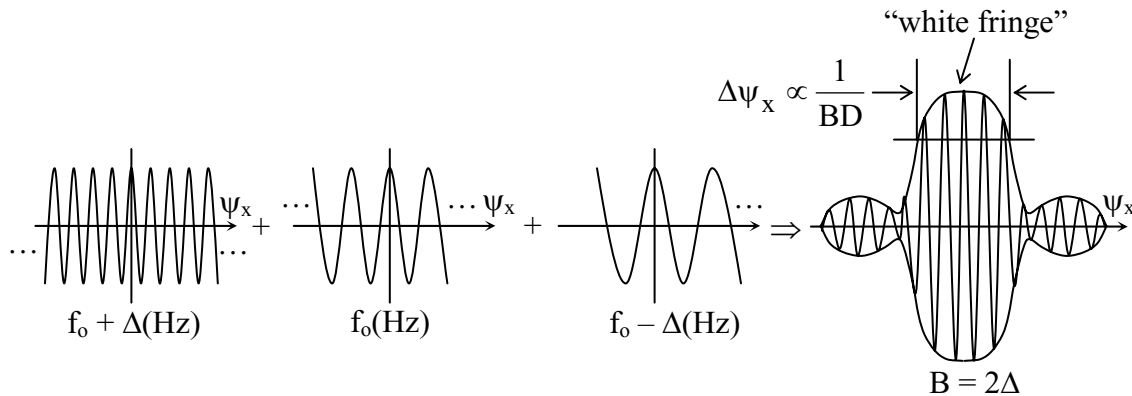
Figure 5.2-11: Creation of white fringe in a broadband two-element interferometer

As noted before, the central fringe of an optical interferometer operating in the visible region appears to be white because all colors add equally at the origin. At angles away from this "equal-delay" central fringe, the fringes become increasingly colored and then gray. If a delay line is added to one interferometer, the position of the white fringe shifts correspondingly in angle. Very narrow effective antenna beams can be created when observing broadband sources by using this technique with a variable delay line that can sweep the antenna pattern across the source.

The DC terms appearing in the simple adding interferometer are routinely eliminated by using Dicke-switch or multiplying interferometers, as illustrated in Figure 5.2-12.
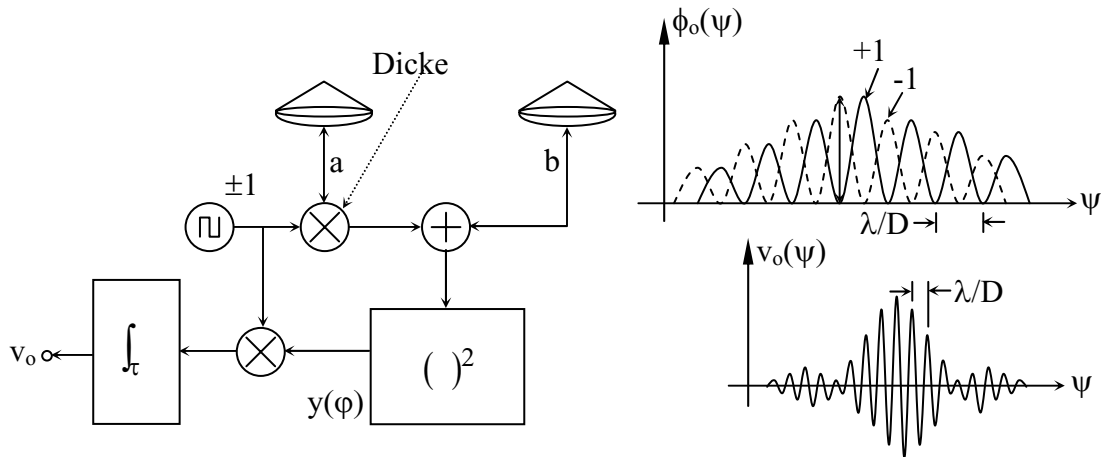
Figure 5.2-12: Two element Dicke-switched adding interferometer

Since the interferometer response when the Dicke switch is in the +1 position is 180° out of phase with the response for the −1 switch position, and since the interferometer subtracts these two responses continuously at high speed, the radiometer output $\phi_o(\psi)$ preserves only the product term $< 2ab >$ of (5.2.7) as the source traverses the beam.

It is generally more convenient to have an interferometer that directly produces the real and imaginary parts of the field autocorrelation function $\underline{\phi}(\overline{\tau_\lambda})$, even when the source is stationary in the sky. Two such circuits are suggested in Figures 5.2-13 and 5.2-14. The first figure is a Dicke-switched multiplying interferometer where, if the switch were omitted, the interferometer would produce the same output, but would be more subject to the effects of gain fluctuations. The circuit of Figure 5.2-14 is a *lobe-scanning interferometer*, where the local oscillators driving the two receivers are at slightly different frequencies, causing the fringe pattern to sweep at a known rate across the source, permitting both the sine and cosine terms to be measured.
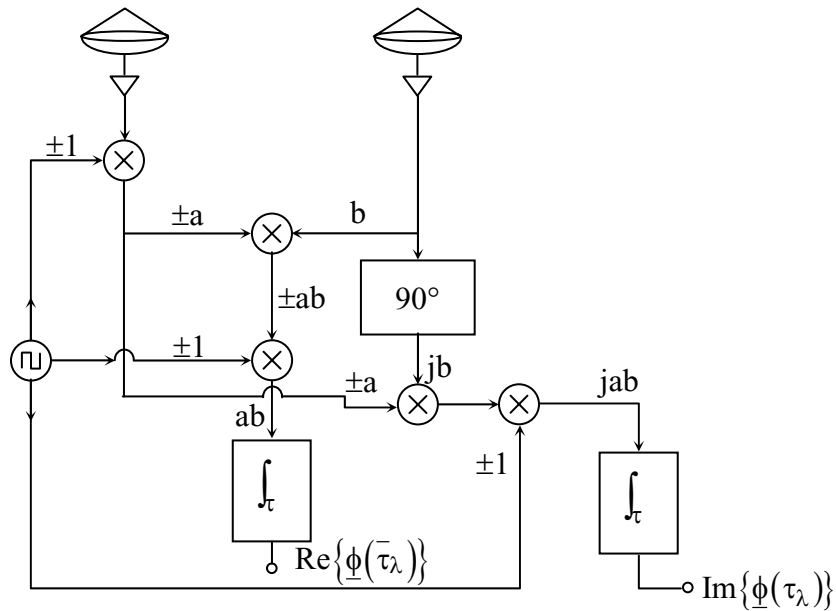
Figure 5.2-13: Multiplying Dicke-switched interferometer

In the lobe-scanning interferometer stable outputs are obtained for stationary sources by demodulating the lobe-scanning response at the lobe modulation frequency ω_m by means of the two multipliers immediately preceding the final integrators. The signal combination path that involves no delay produces the real part of the field autocorrelation function, whereas the path delayed 90° yields the imaginary part. Because these multipliers produce two sidebands, the undesired one must be filtered out. The upper-sideband (USB) filter performs this function at r.f., and the two integrators perform the same function at baseband.
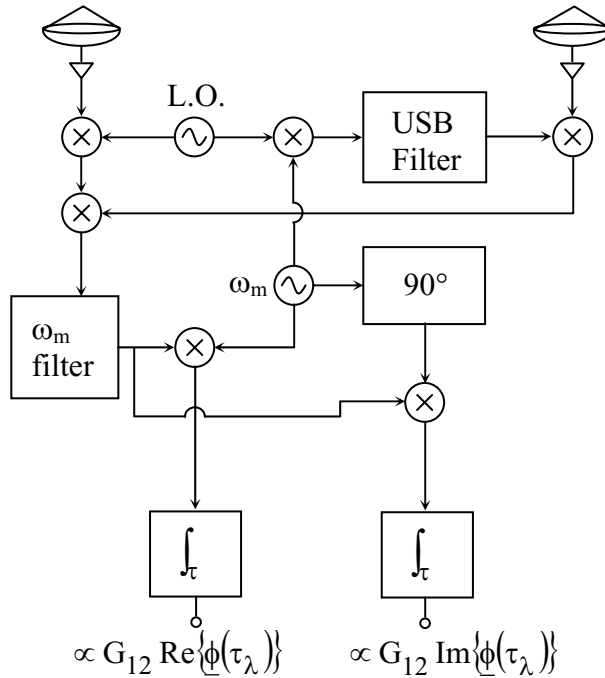
Figure 5.2-14: Lobe-scanning interferometer

One use of such interferometers is to detect unresolved point sources, while ignoring all extended ones. For example, it is possible to detect radio-emitting point sources like stars with such interferometers by integrating hours or days. Since almost all astronomical radio sources are spatially extended, this technique can be quite successful, but does require any integration drift to be calibrated. This can be done in Dicke-radiometer fashion by periodically offsetting the demodulation frequency ω_m from its true value to force the integrator to zero.

Traditional analog systems are increasingly being replaced by digital equivalents. One popular example is the *cross-correlation interferometer spectrometer*, such as the one illustrated in Figure 5.2-15
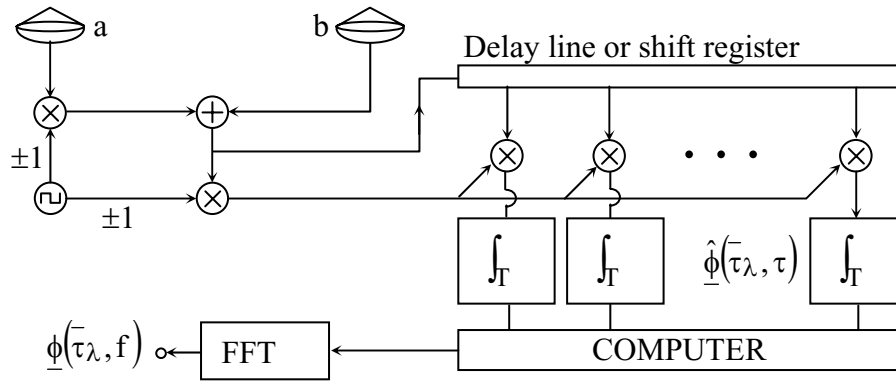
Figure 5.2-15: Cross-correlation Dicke interferometer spectrometer

In this case the two signals a and b are added, digitized, and fed to a delay line or shift register and also to multipliers operating on each of the outputs of that delay line. These products are each integrated for a period of time T, at the end of which the computer produces the Fourier transform, yielding the desired field autocorrelation function $\underline{\phi}\left(\overline{\tau\lambda}, f\right)$. In this case the Fourier transform is performed only in delay space τ(seconds), not in spatial-offset space $\overline{\tau\lambda}$. As before, the Dicke switch permits use of an adder instead of an ideal multiplier, because it cancels out the DC terms leaving only the multiplication terms in the final output. The multipliers operating on the delay line outputs are typically digital, where the A/D converters can precede or follow the adder combining the two signals a and b. Note that when a = b, the output $\underline{\phi}\left(\overline{\tau\lambda}, f\right)$ simply becomes $\underline{\phi}(0, f) = S(f)$; that is, the device operates purely as a spectrometer.

A more common variation of this circuit is suggested by Figure 5.2-16, where the Dicke switches are omitted and the two antenna signals are multiplied following the delay line. Note that Figure 5.2-16 has omitted the necessary A/D converters and band limiting filters.

If the antennas are a considerable distance apart, separate local oscillators may be used, but they must somehow be synchronized, for example, by using the circuit illustrated in Figure 5.2-17.
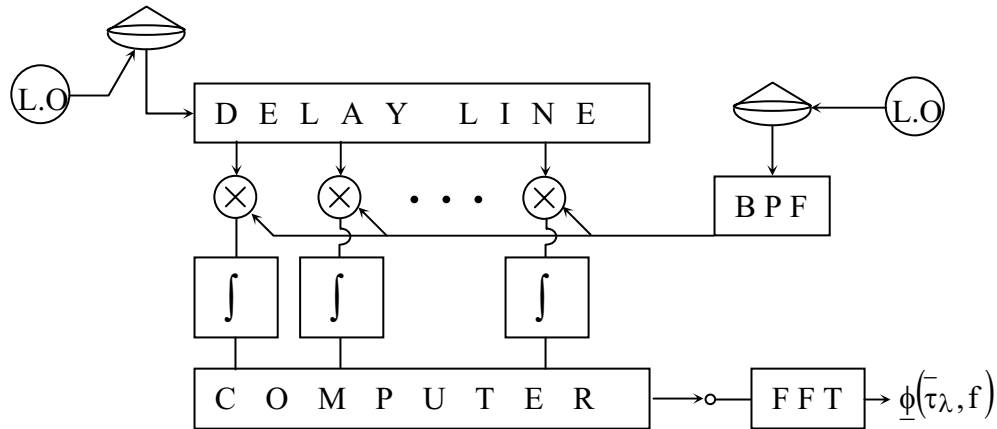
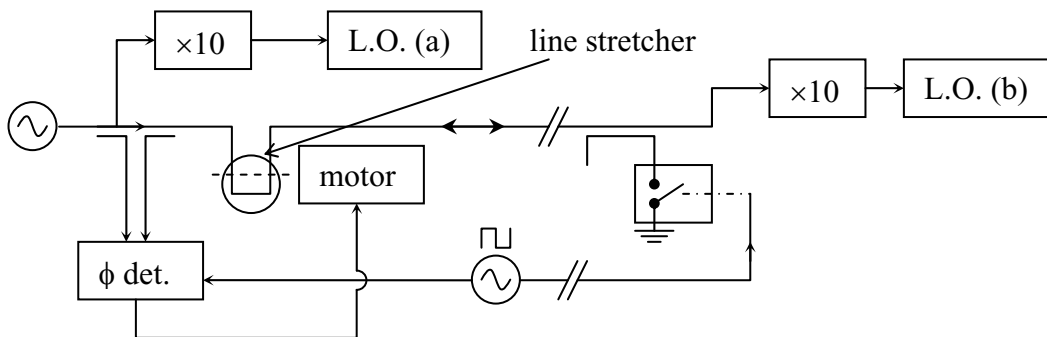Figure 5.2-16: Cross-correlation multiplying interferometer spectrometer

Figure 5.2-17: Local oscillator phase sychronization circuit

The fundamental oscillator on the left side of the figure is typically multiplied by some integer to produce the first local oscillator (a). It is also similarly multiplied to produce the second local oscillator (b), but after passing through a long transmission line connecting the two sites and a variable-length "line stretcher." The line length is continuously and rapidly varied by a servo system observing the relative phase between the original fundamental oscillator signal and a reflected version which has traveled twice through the unknown length of the transmission line connecting the two sites. In order to distinguish the desire reflected signal from reflections which could occur anywhere else in the system, the signal reaching the second site is modulated by a switch which is controlled by a Dicke oscillator, as illustrated. The feedback control system is operated so that the phase difference, as observed by the phase detector, remains zero. This technique is used also in communications and telemetry systems.

   If the distance between two sites is too great to permit even this technique to work, then highly stable independent clocks, one at each site, can be used. This technique is widely used in *very-long-baseline interferometry* (VLBI). In this case the data at each site is mixed to baseband using the accurate clock, converted to digital form, and stored either for manual delivery or for

asynchronous delayed transmission. These baseband digital signals a′ and b′ from the sites are later brought together with time synchronization controlled by a coarse system such as that provided by the satellite based global positioning system (GPS), which readily permits 10^{-8} second accuracy, or the more primative LORAN-C system, which permits one microsecond accuracy. Although temperature stabilization can improve the intrinsic stability of crystal oscillators, which ranges from $\sim 10^{-5}$ to values of $\sim 10^{-8}$, cesium-beam clocks of $\sim 10^{-12}$ accuracy or hydrogen masers approaching $10^{-14} - 10^{-15}$ accuracy are generally preferred.

If the clocks are perfect, then the two signals a′ and b′ can be cross-correlated to find the time offset, permitting its correction. Although there is an unknown fixed phase offset that results, this can be established by definition or by observing a reference source in the sky. Since the clocks drift slightly, however, the time over which phase coherence is obtained is finite. The signals are then averaged over these finite intervals to determine the magnitude of the correlation coefficient, and to track the phase shift. This phase shift can be unwrapped to extend the time intervals over which the magnitude can be averaged. Unwrapping involves removal of the 2π phase ambiguities that result whenever the true phase crosses the $0/2\pi$ boundary.

For example, if a cesium clock accurate to 10^{-12} drifts π radians in approximated one minute, then $\hat{\underline{\phi}}(\tau_\lambda, f)$ might be computed for blocks lasting several seconds before the results are averaged; in this case only $\left|\hat{\underline{\phi}}\right|$ can be determined. If within a diffuse source there is a strong monochromatic point source, then the phase offset $\phi_o(t)$ can be determined by analyzing its contribution and that offset can then be used to interpret the remaining signal. If such a point source resides within view of all antennas, then its use is straight forward. Alternatively, each site can have two antennas, one of which is pointed to the common point source located elsewhere in the sky.

Another technique for resolving fringe ambiguities in very long baseline interferometers involves the use of very broad bandwidths. Consider a source region that has hundreds or millions of fringes across its width, and assume we wish to determine on which fringe F a particular source sits. Figure 5.2-18 illustrates the problem and the technique.
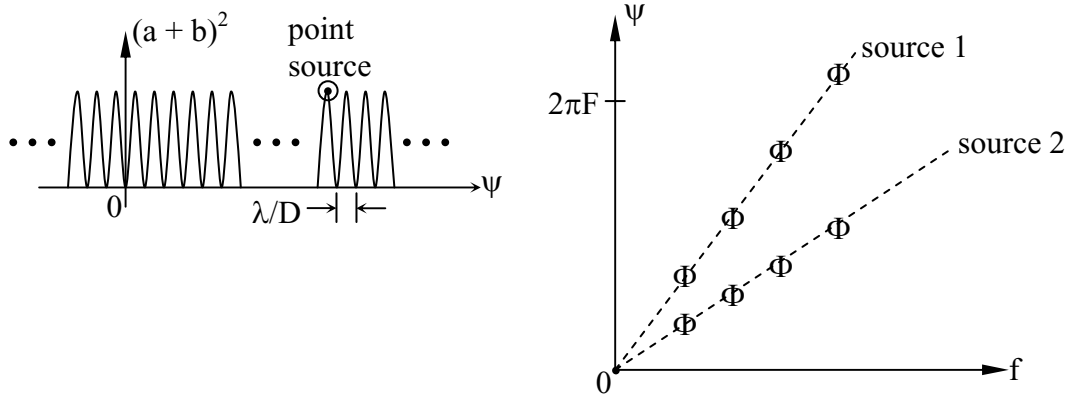
Figure 5.2-18: Fringe identification using broadband observations.

The fringe pattern of Figure 5.2-18 has a period of λ/D, where D is the interferometer baseline projected perpendicular to the line of sight to the source. Therefore for each source the fringe phase relative to the phase origin is $\psi = 2\pi F$, where $F(\lambda)$ is the wavelength-dependent fringe number. Since we can determine phase only modulo 2π at a single frequency, we need more observations. For example, if we were to observe at all frequencies, we could readily determine the fringe position for each source of interest by extrapolating its observed phase to zero frequency. In practice it suffices to observe the fringe spacing over single or multiple disconnected modest bandwidths, where these modest bandwidths are sufficient to determine the slope of phase as a function of frequency with accuracy sufficient to extrapolate ψ without ambiguity to zero frequency or to the next band in which phase is measured. Accurate phase determinations at two bands simplify the task of extrapolation to a third band, and so on. A modest number of bands spaced exponentially usually suffices to remove fringe ambiguity so the separations between two sources can be accurately determined, even if the absolute position may be somewhat uncertain because of relative drifts between the two local oscillators. Clearly one must switch the interferometer from one frequency band to another across all necessary bands within the coherence time of the clock pair.

In those cases where only the magnitude of the aperture field autocorrelation function is desired, phase information and clock synchronization requirements can be discarded entirely, provided the source signal strength is strong. Perhaps the first instrument employing this technique was that of Hanbury-Brown and Twiss in their visible-band optical interferometer at Narrabri, Australia. This instrument used two large parabolic mirrors on moveable platforms which could be separated at distances of hundreds of meters along a circular track. At the focus of each light bucket was a phototube with a bandwidth of ~1 GHz.
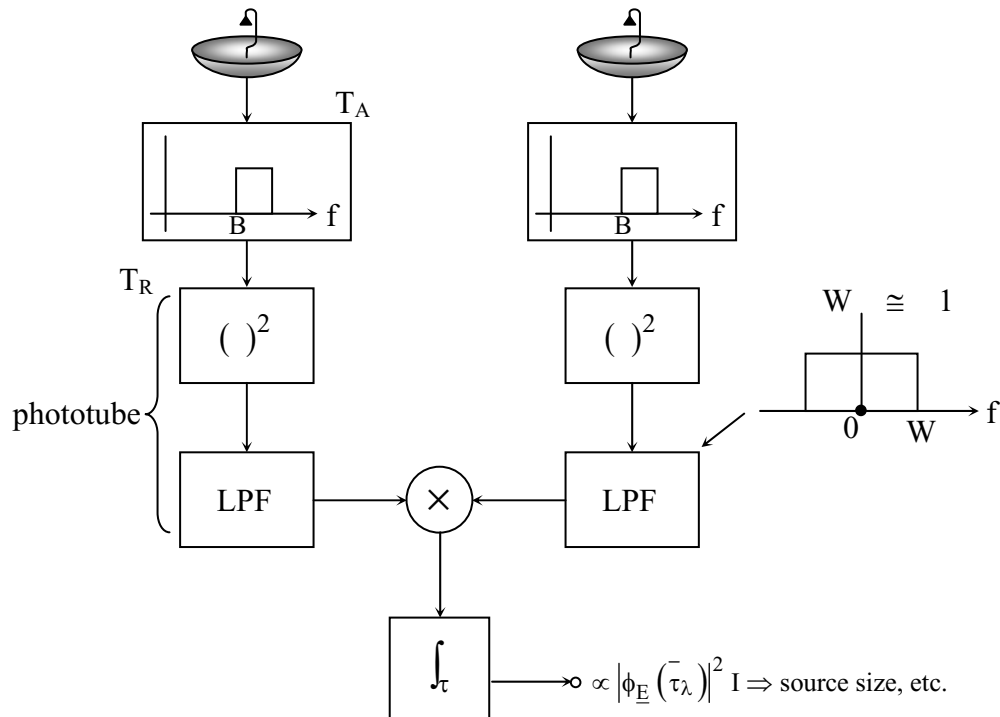
Figure 5.2-19: Hanbury-Brown Twiss Interferometer

In this system, often called a *Hanbury-Brown Twiss interferometer*, the detected 1-GHz bandwidth signals from the two phototubes are multiplied and integrated to yield an output which is proportional to the square of the field autocorrelation function $\left|\phi_{\underline{E}}\left(\bar{\tau}_\lambda\right)\right|^2$. This relationship is easily derived by recalling the formula for the expected value of the product of four jointly gaussian random variables. In this case, the two optical intputs, x and y, are first squared in the phototube because only power in the arriving waves is measured, and then low-pass filtered by the finite response time of the phototube to produce the 1-GHz bandwidth signals which are then multiplied. Averaging produces the expected value

$$E[xxyy] = \overline{x^2}\ \overline{y^2} + 2\overline{xy}^2 \tag{5.2.13}$$

where $\overline{xy}$ is $\phi_{\underline{E}}\left(\bar{\tau}_\lambda\right)$ here. That is, the output of the Hanbury-Brown Twiss interferometer is the square of the field autocorrelation function $\phi_{\underline{E}}\left(\bar{\tau}_\lambda\right)$. Although we have lost the phase information in $\phi_{\underline{E}}$, if the source is known to be an even function of position, then $\phi_{\underline{E}}$ is real and the phase is known *a priori* to be zero. In this case the source intensity distribution can be produced exactly, even though all phase information has been lost. The fact that a system

lacking phase reference can nonetheless yield angular resolution of $\sim\lambda/D$ may be surprising. Several simple ways of interpreting this situation are explained below.

First, however, we should note the poor sensitivity of this configuration. Because only the envelopes of the two signals are being correlated in the Hanbury-Twiss interferometer, there is a penalty in system sensitivity for the weak signal case where $T_R \gg T_A$. In this situation the ratio of the rms fluctuations in the output signal $\phi_{o_{rms}}$ to the average value of the output voltage $\langle \phi_o \rangle$ is proportional to $(T_R/T_A)^2$:

$$\frac{v_{o\,rms}}{\langle v_o \rangle} \cong \frac{T_R^2}{T_A^2 \sqrt{2W\tau}} \tag{5.2.14}$$

where W[Hz] is the bandwidth of the detector output. In this case it is important to concentrate on the brightest sources and to use a fast cooled photomultiplier tube, a very large aperture, and a very long integration time τ.

The output of the Hanbury-Brown Twiss interferometer $\left| \underline{\phi}_{\underline{E}}\left(\overline{\tau\lambda}\right) \right|^2$ is related to the source angular brightness distribution $\left| \underline{E}\left(\overline{\psi}\right) \right|^2$ as suggested by equation (5.2.15).

$$
\begin{array}{ccc}
\underline{E}(x,y) & \leftrightarrow & \underline{E}\left(\overline{\psi}\right) \\
\downarrow & & \downarrow \\
\underline{\phi}_{\underline{E}}\left(\overline{\tau\lambda}\right) & \leftrightarrow & \left| \underline{E}\left(\overline{\psi}\right) \right|^2 \Rightarrow I\left(\overline{\psi}\right) \\
\downarrow & & \downarrow \\
\left| \underline{\phi}_{\underline{E}}\left(\overline{\tau\lambda}\right) \right|^2 & \leftrightarrow & R_{\left| \underline{E}\left(\overline{\psi}\right) \right|^2}\left(\Delta\overline{\psi}\right)
\end{array}
\tag{5.2.15}
$$

The top four variables in (5.2.15) correspond to the aperture field distribution $\underline{E}(x,y)$ and its transformed representations in displacement space $\overline{\tau\lambda}$, and source-angle space $\overline{\psi}$. The same transformations that link the first Fourier-pair of variables to the pair on the second line of (5.2.15) can be applied a second time to the second row so as to yield the third row. That is, the squared magnitude of any complex quantity is the Fourier transform of the autocorrelation function of the Fourier transform of the original. More particularly, the output of the Hanbury-Brown Twiss interferometer $\left| \underline{\phi}_{\underline{E}}\left(\overline{\tau\lambda}\right) \right|^2$ is the Fourier transform of the autocorrelation function $R_{\left| \underline{E}\left(\overline{\psi}\right) \right|^2}\left(\Delta\overline{\psi}\right)$ of the angular plane wave expansion of the source, squared; that is, the ouput is the Fourier transform of the autocorrelation function of the source angular intensity distribution

$I(\overline{\psi})$. Thus, to the extent $\underline{\phi}_E(\overline{\tau}\lambda)$ is a real function, which is equivalent to the source angular intensity distribution $I(\overline{\psi})$ being an even function of position, $I(\overline{\psi})$ can be recovered exactly.

Figure 5.2-20 and 5.2-21 suggest two ways to understand physically why the thermal radiation becomes increasingly decorrelated in directions separated by angles θ larger than $\sim \lambda/D$, where D is the diameter of the blackbody source. We can consider the radiation traveling towards the antennas as originating from a combination of many point sources, such as g and h in Figure 5.2-20. The radiation from g and h superimpose at antenna a with an offset in time which differs by roughly 2π from that associated with the superposition of the same two rays at antenna b, at which point the signals a and b become decorrelated. The separation angle θ that produces decorrelation and a differential phase lag 2π at the central frequency is λ/D.
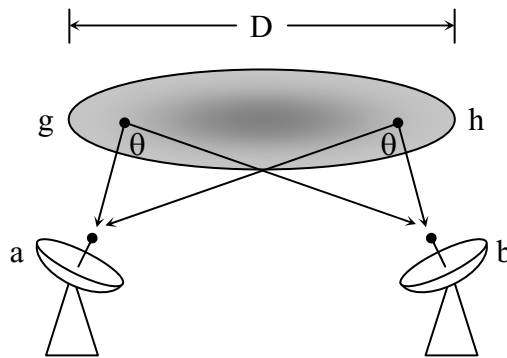
Figure 5.2-20: Decorrelation of source radiation propagating in different directions.

We can also arrive at a similar conclusion using a planewave expansion, as suggested in Figure 5.2-21. If we imagine the source to be a hole in the sky behind which there is a thermal blackbody, then the shapes of the diffracted beam patterns propagating toward the two antennas a and b will be identical, but the signals associated with these patterns, one per planewave, will be completely uncorrelated. Therefore the only correlation between signals arriving at a and b arise because of the finite beamwidth θ_B of the beams diffracted by the source aperture. This beamwidth is $\sim \lambda/D$.
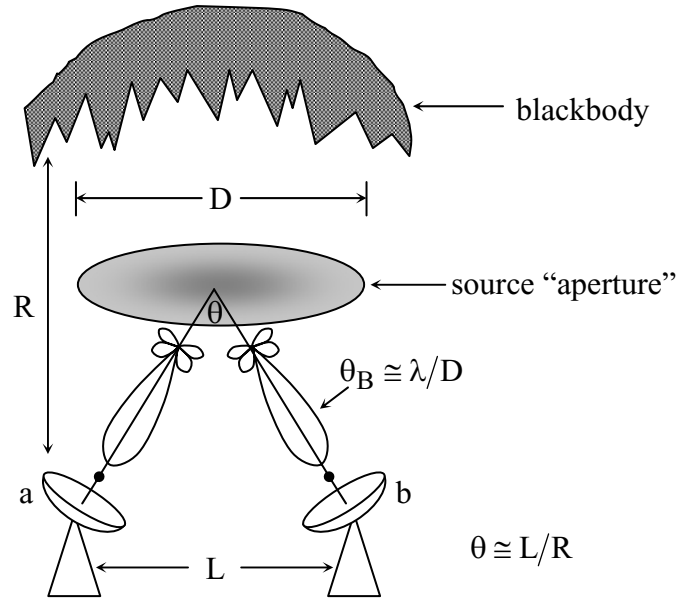
Figure 5.2-21: Decorrelation of source-diffracted radiation propagating in different directions.

The sources observed by Hanbury-Brown and Twiss were stars of diameter $\sim 10^{10}$ meters and therefore θ_B was on the order of a milliarcsecond. This also equals the angle L/R at the point of decorrelation, where R is the distance to the star, and L is the separation distance between the two antennas a and b at the point of decorrelation. Thus decorrelation occurs when:

$$L/R \cong \lambda/D \qquad (5.2.16)$$

Manipulating (5.2.16) yields:

$$\lambda/L \cong D/R \cong \theta_s \qquad (5.2.17)$$

where θ_s is the angular dimension of the source (radians) as seen by the observer.

5.3 RADAR

Radar systems deduce the range to objects, and perhaps their radial velocity and shape. *Monostatic radar* is the most common, which means that the same antenna is used for transmission and reception. If the two antennas are located at different positions, it is called *bistatic radar*. If the image is constructed by coherently combining different radar echoes obtained while the radar antenna is moving relative to the source, or visa versa, it is called *synthetic aperture radar* (SAR). That variant of SAR where the radar antenna is regarded as stationary and only the target moves is sometimes called *inverse synthetic aperture radar*

(ISAR). Since SAR images can preserve phase, it is possible to determine phase differences between consecutive images using *Interferometric synthetic aperture radar* (InSAR) Discussion of SAR and ISAR is postponed to Section 5.4.

A simple characterization of the radar problem is provided by Figure 5.3-1.
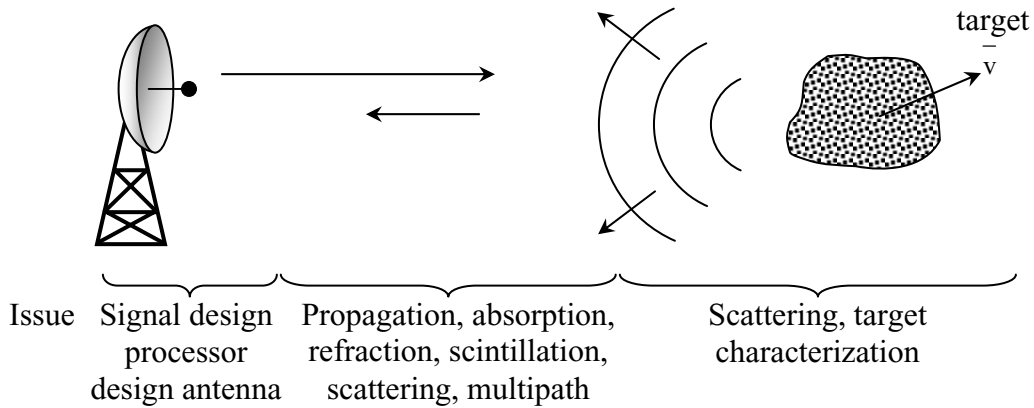

| Issue | Signal design processor design antenna | Propagation, absorption, refraction, scintillation, scattering, multipath | Scattering, target characterization |

Figure 5.3-1: Radar issues

The target generally reflects the incident signals in directional patterns that may be complex, depending on the shape and electrical properties of the target. If the target is a rigid body, then it can be characterized as having a position $\bar{r}$, a velocity $\bar{v}$, and an angular velocity vector $\bar{\omega}$. In a bistatic configuration, it is important to know how the incident energy is scattered over all 4π steradians for any particular position of the target. In the monostatic case, only the scattering cross-section σ corresponding to retro-reflected radiation is of interest.

The basic *radar equation* which quantifies the power received by a radar P_r in terms of the power transmitted P_t is given by (5.3.1), where G_t and A_t are the gain and effective area of the transmitting antenna, respectively, R is the distance to the target, and σ is the target scattering cross-section (m^2).

$$P_r = \frac{P_t}{4\pi R^2} \bullet G_t \bullet \frac{\sigma}{4\pi R^2} \bullet A_t = P_t \left(\frac{G\lambda}{4\pi R^2} \right)^2 \frac{\sigma}{4\pi} \qquad (5.3.1)$$

The product of the first two factors $P_tG_t/4\pi R^2$ corresponds to the flux density (Wm^{-2}) at the target, while the product of the first three factors corresponds to the flux density reflected by the target and measured at the transmitting antenna. As suggested in (5.3.1), the equation can be further simplified using the relation $A = G\lambda^2/4\pi$. Note that the factor $\sigma/4\pi R^2$ in (5.3.1) effectively defines the *scattering cross-section* σ as the equivalent capture cross-section for a target that scatters the incident radiation isotropically.
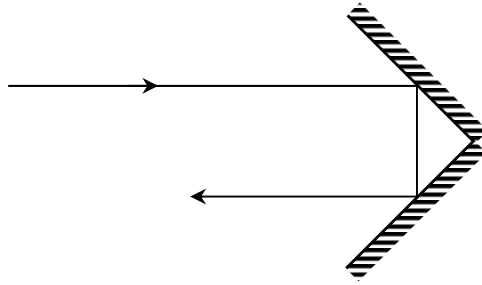
Figure 5.3-2: Corner reflector

It is very important to note that targets can have scattering cross-sections substantially larger than their physical size. For example, a corner reflector composed of two metal plates at right angles and perpendicular to the plane of incidence reflect the power incident upon them in a narrow retro-reflected beam having a width of $\sim \lambda/D$ radians, where D is the diameter of the retro-reflector. Corner reflectors can be effective over nearly one steradian if they are composed of three flat sheets at right angles, forming a three-dimensional corner. It is easy to show that the scattering cross-section σ of a good retro-reflector can be approximately $4\pi D^2/\lambda^2$ times larger than its physical cross-section.

It is also interesting to note that power received is proportional to the square of the gain $G(\theta)$, and that a radar exhibits a narrower beamwidth than would the same antenna used in a passive mode. Most important is the fact that the returned echo for an unresolved target is proportional to R^{-4}, which severely limits the practical range of radars having modest power and antenna gain.
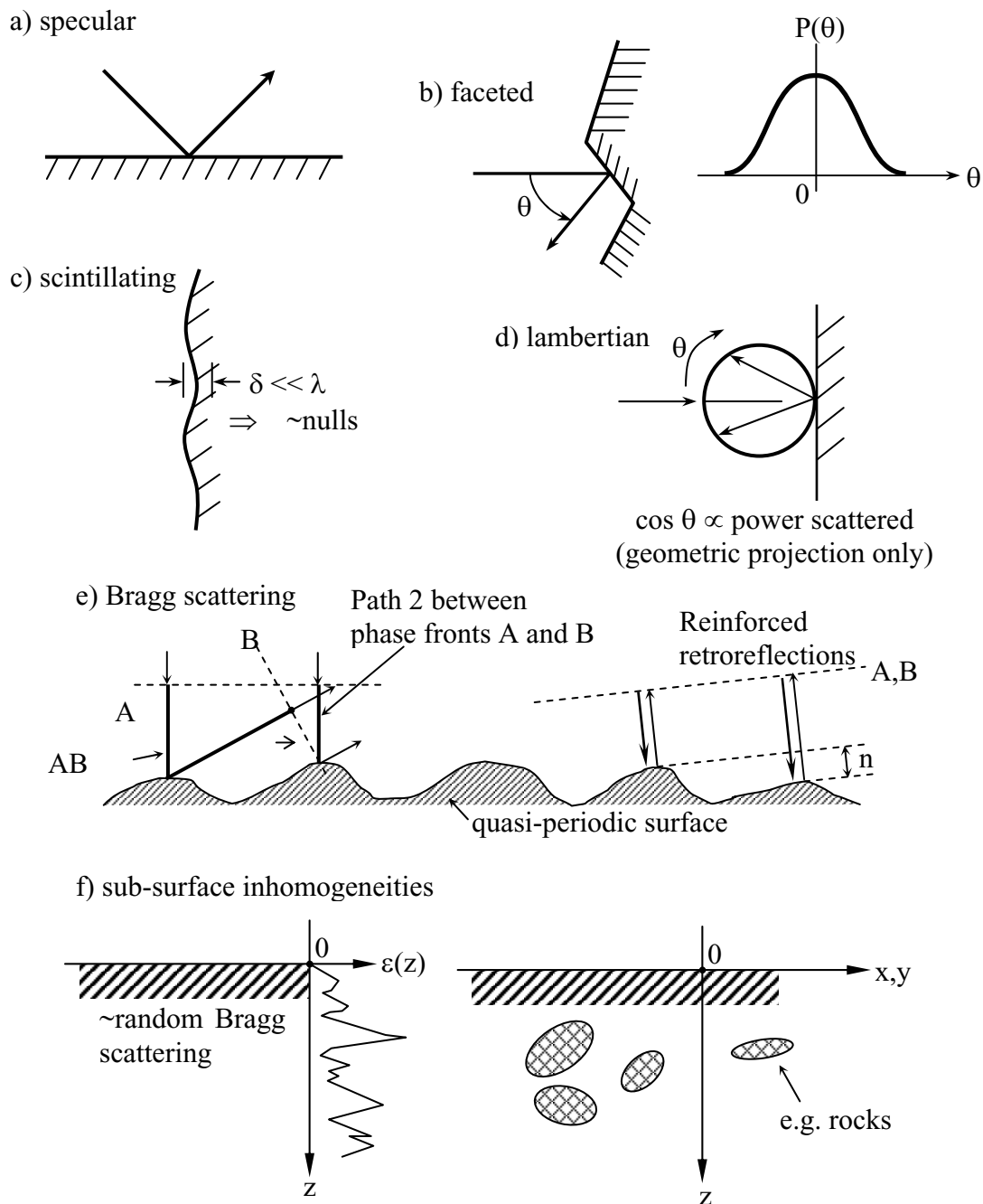
a) specular

b) faceted

P(θ)

θ

c) scintillating

$\delta \ll \lambda$

$\Rightarrow$ ~nulls

d) lambertian

θ

$\cos \theta \propto$ power scattered
(geometric projection only)

e) Bragg scattering

Path 2 between
phase fronts A and B

B

A

AB

Reinforced
retroreflections

A,B

n

quasi-periodic surface

f) sub-surface inhomogeneities

0

$\varepsilon(z)$

~random Bragg
scattering

z

0

x,y

e.g. rocks

z

Figure 5.3-3: Target scattering laws

The scattering cross-section of a target depends on the character of its surface, as suggested in Figure 5.3-3. Six important types of surfaces are frequently encountered. *Specular surfaces* are flat and smooth on a scale of the wavelength, and preserve the plane-wave character of the incident wave upon reflection, where the angle of reflection equals the angle of incidence. *Scintillating surfaces* are similar but with surface deviations $\delta \ll \lambda$, where these indentations and protrusions have spatial extents much larger than a wavelength. The resulting scattered

wave is generally planar, but with randomly positioned nulls. *Faceted surfaces* are composed of specular surfaces linked to form a zigzag surface, where the facets are generally large compared to a wavelength and the angles of deflection are generally less than a radian. *Lambertian surfaces* have a horizontal spatial scale much smaller than a wavelength and as a result the scattered radiation is more nearly isotropic and therefore in proportion to the projection angle from which it is viewed; that is, the reflected intensity is proportional to $\cos \theta$, where θ is the angle between the surface normal and the viewer. *Bragg scattering* results when the target surface is generally planar but modulated periodically so that the reflected signal tends to reinforce in some directions and cancel in others, as suggested in Figure 5.3-3e.

Sub-surface inhomogeneities can take various forms. For example, the dielectric constant $\epsilon(z)$ can be a function of depth z, so that multiple reflected waves are produced which interfere constructively in some frequency bands and destructively in others, and thus assumes some spectral character. Alternatively they may be introduced by rocks of one dielectric constant imbedded in a medium of another. To the extent the rocks are small compared to a wavelength, the sub-surface scattering will take on a Lambertian character, and to the extent the rocks are large compared to a wavelength, they may assume a more faceted character.
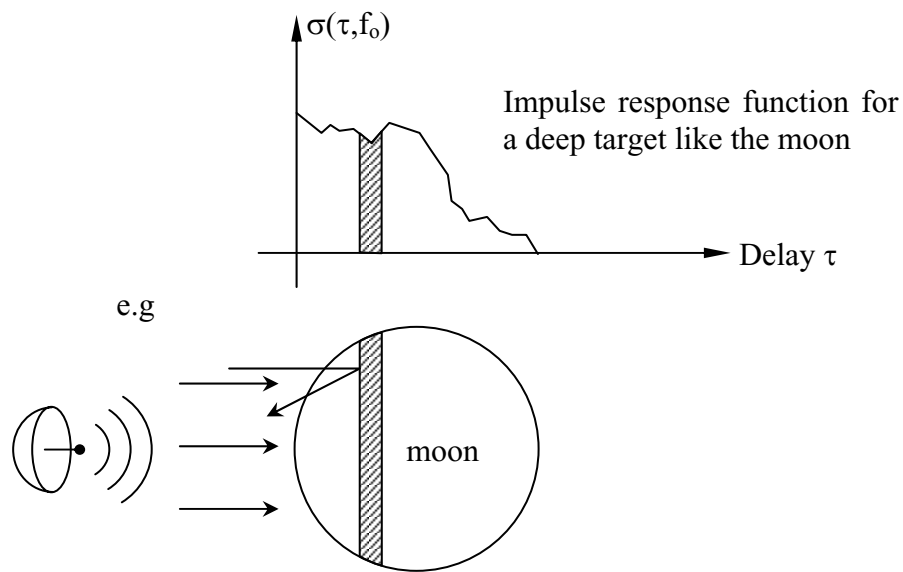
Figure 5.3-4: Lunar scattering cross-section $\sigma(\tau,f_o)$ as a function of pulse delay τ, as observed at r.f. frequency f_o.

Some targets exhibit combinations of these scattering processes. For example, consider the target response of the moon when its entire surface is illuminated by an earth-based radar emitting an impulse of duration T(sec) and length cT meters, which is small compared to the lunar radius. Figure 5.3-4 illustrates how the signal received at any instant comes from a narrow circular ring on the moon of depth cT/2, where the factor of two results because of the round-trip

path taken by the echo. If the thickness of this ring is small compared to the depth of the target we say we have a deep target.

Because the moon is large, the pulse duration T can be made sufficiently short that useful range resolution is obtained, while still providing enough pulse width to permit Fourier analysis to detect doppler shifts introduced by lunar rotation, as suggested in Figure 5.3-5.
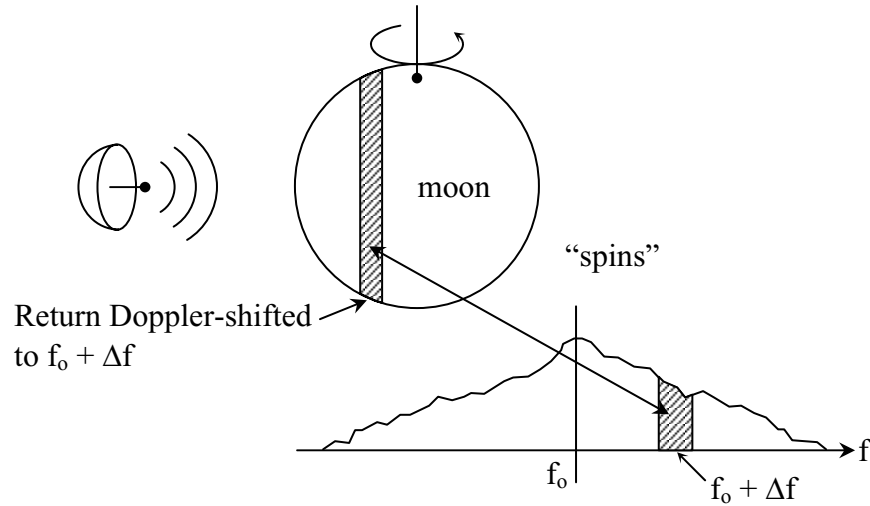
Figure 5.3-5: Doppler-shift zones observed by lunar radar.

Because the moon is effectively rotating slightly with respect to observers on earth because of a small wobble, a small doppler shift Δf can be introduced in the echo, as illustrated.

A map of the lunar radar reflectivity can be obtained, as suggested in Figure 5.3-6, by dividing each returned radar pulse into different range bins corresponding to different delays, and then Fourier transforming each such sub-echo, thus dividing it into different doppler shifts.
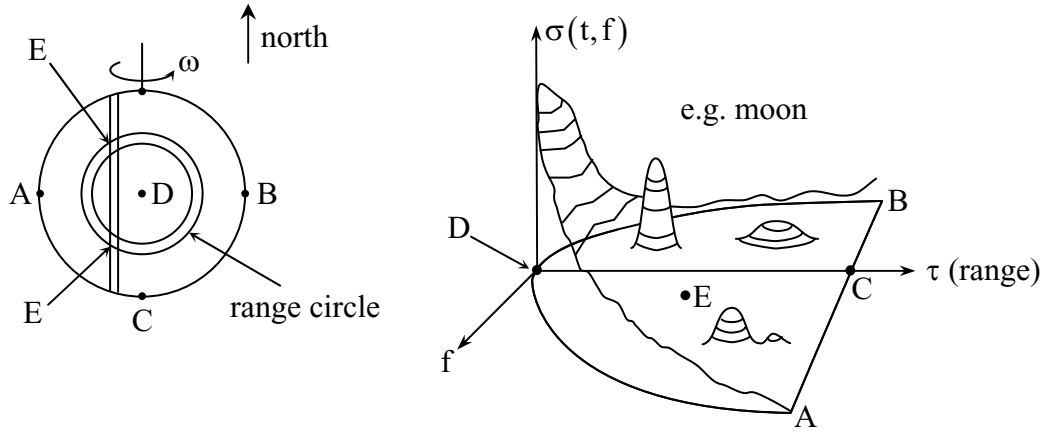
Figure 5.3-6: Range-doppler response for a pulse radar illuminating the moon.

The first return in any pulse is associated with the lunar point D closest to the radar, and subsequent parts of the echo are associated with a range circle of steadily increasing radius, until the the limb of the moon is reached at A,C,B, beyond which no further echo is received. For any particular range circle, doppler analysis, as suggested in Figure 5.3-5, can isolate the total power scattered by the points E illustrated in the figure which are associated with the given doppler band. There is, unfortunately, a north-south ambiguity between echoes from the points E in the northern and southern hemispheres; their echo strengths are superimposed and can be disentangled only if the radar antenna beam illuminates one hemisphere in preference to the other. The observed echo strength as a function of delay τ and doppler offset f thus resembles a map of the moon, albeit folded over on itself at the equator.

The figure suggests how the scattering cross-section $\sigma(\tau,f)$ corresponding to the point D is larger because this reflection typically has a large specular component. By observing the angular width of this specular component, information about the slope distribution can be obtained for those specular facets with widths several times larger than the wavelength. Craters or other features which strongly retro-reflect show up within the space (τ,f) as regions of enhanced reflectivity, as suggested in the figure. The semicircular shape of this space $\sigma(\tau,f)$ where σ is non-zero results because no delays τ can be greater than those corresponding to the points A, B, and C, and for any particular range τ, the doppler width will be limited. At the point D the maximum possible doppler offset f is zero. The echo corresponding to the points E is also illustrated in the figure.

Simple considerations show that the lateral resolution of a range-doppler radar observing the moon is limited. Consider the simple radar pulse illustrated in Figure 5.3-7.
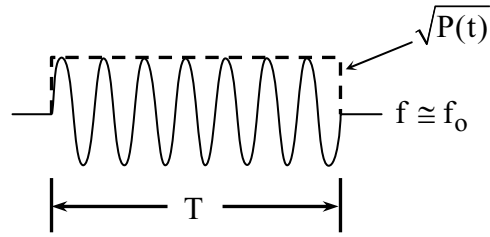
274

Figure 5.3-7: CW radar pulse

The pulse consists of a sinusoid at frequency f_o Hz truncated by a boxcar window of length T seconds; the amplitude of the sinusoid is proportional to $\sqrt{P(t)}$. The resulting received signal r(t) will be a slowly modulated sinusoid as suggested in (5.3.2).

$$r(t)(\text{volts}) = k\sqrt{\sigma}\sqrt{P(t)}\cos\omega_o t + m(t) \qquad (5.3.2)$$

The received signal r(t) is proportional to the constant k, which represents the effects of antenna gain and other factors in the radar equation, plus receiver amplifier gains, etc. The additive term m(t) corresponds to receiver noise, which is usually gaussian and white over the band of interest. This expression for a point target must be convolved with the distribution of the scattering cross-section σ with range τ and/or doppler offset Δf.
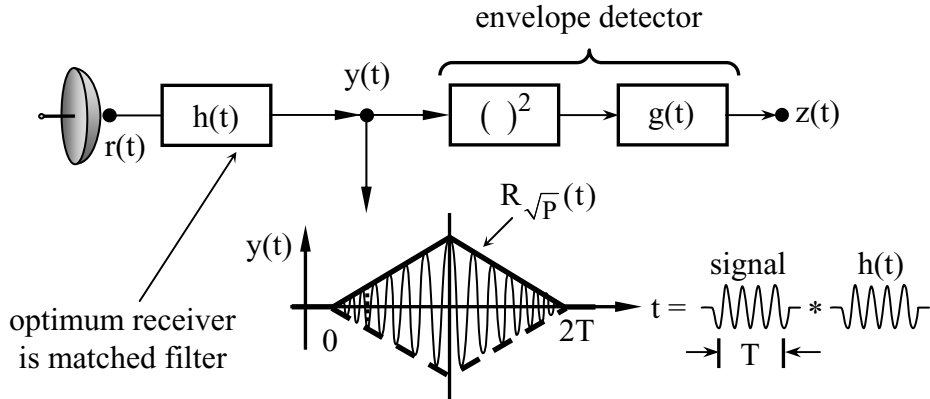
Figure 5.3-8: Matched filter receiver for CW pulsed radar

Estimating the scattering cross-section σ of a point target and the delay τ becomes a communications receiver design problem, where a matched filter receiver can yield optimum performance by testing every possible delay. However, this bank of filters can be replaced by a single filter matched to the transmitted waveform. This filter output can be scanned by the receiver for its maximum value, the amplitude and timing of which provides the desired cross-section and range information. Figure 5.3-8 illustrates the form of such an optimum receiver.

275

Consider the output y(t) from an optimum matched filter h(t), where h(t) is the time-inverted transmitted pulse. The output y(t) is the convolution of the input signal with the filter response function h(t), and assumes the triangular form suggested in Figure 5.3-8. The envelope of that response $R_{\sqrt{P}}(t)$ is defined as:

$$R_{\sqrt{P}}(t) \triangleq \int_{-\infty}^{\infty} \sqrt{P(\tau)} \frac{\sqrt{P(t-\tau)}}{\sqrt{P(0)}} d\tau \qquad (5.3.3)$$

where $\sqrt{P(\tau)}$ is the voltage envelope of the transmitted pulse, as suggested in Figure 5.3-7 and again in Figure 5.3-8. More particularly, using (5.3.2) we obtain the filter output y(t):

$$y(t) \triangleq \left[\sqrt{\sigma}\sqrt{P(t)} \cos \omega_0 t + m(t) \right] * \left[\frac{\sqrt{P(-t)} \cos \omega_0 t}{\sqrt{P(0)}} \right] \qquad (5.3.4)$$

where the ideal filter response h(t) is the matched filter response normalized by the transmitted signal voltage $\sqrt{P(0)}$.

One way to detect the envelope of the filter output is to square the output and then low-pass filter it using the filter g(t), where this low pass filter has a time constant equal to several periods f_0^{-1} of the transmitted pulse. The matched filter response to a point source has total duration 2T, where T is the duration of the transmitted pulse. The output of the envelope detector z(t) is:

$$z(t) \cong k^2 \sigma_0 R_{\sqrt{P}}^2(t) + n(t) \qquad (5.3.5)$$

where σ_0 is the scattering cross-section of the point source, R(t) is the envelope of the ideal point-source matched-filter response, and n(t) is the effective additive noise at the output. For an ideal point target the response z(t) is given for this receiver in Figure 5.3-9, where this response $R_{\sqrt{P}}^2(t)$ is the *ambiguity function* in delay.
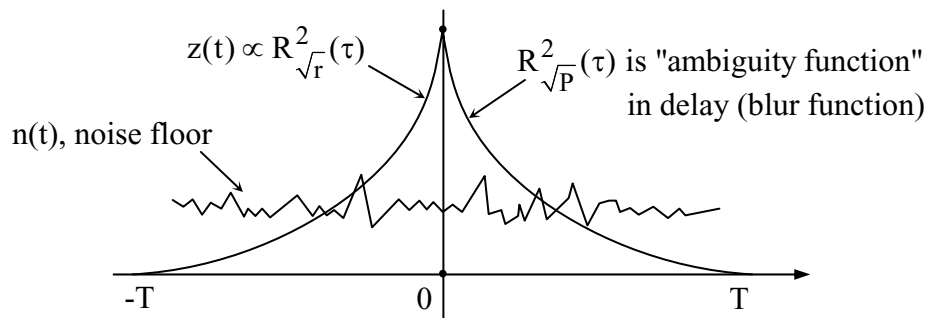

Figure 5.3-9: Ambiguity function in delay for a pulsed CW radar

276

In this case the ambiguity function is the square of a triangle wave, and the output in the presence of multiple targets may or may not be the superposition of their corresponding ambiguity functions. For example, if the targets are all stationary then their echoes will be coherent and interfere constructively or destructively, preventing superposition. If the targets have different doppler shifts, then the echoes will add incoherently and the system response z(t) will be the superposition of the corresponding ambiguity functions, each with its own delay and amplitude corresponding to the target cross-section and range. In this case where the target echoes add incoherently the expected value of the output is:

$$E[z(t)] = k^2 \sigma(\tau) * R^2_{\sqrt{P}}(\tau) + n(t) \tag{5.3.6}$$

This integral equation (5.3.6) can generally be solved for the scattering cross-section σ as a function of range τ, but the range accuracy is limited by the width of the ambiguity function and the noise level. If the detector is linear instead of quadratic, a slightly different ambiguity function results, having the triangular form of $R_{\sqrt{P}}(t)$ for boxcar CW pulse transmissions.

Optimum receiver design for radar signals is similar to that for communication signals. In each case an ensemble of possible received waveforms might result, each corresponding to a different target, and our task is to identify the one with the maximum *a posteriori* probability. A variation of the receiver architecture illustrated in Figure 5.3-8 can be constructed for analyzing signals where both doppler and range are of interest, as illustrated in Figure 5.3-10. The received signal r(t) is amplified and then passed through several filter banks in parallel. Their outputs, y(t), produce responses resembling the ambiguity function and feed a comparative peak detector that identifies the highest response in range/frequency as the best estimate for delay D and doppler shift Δf.
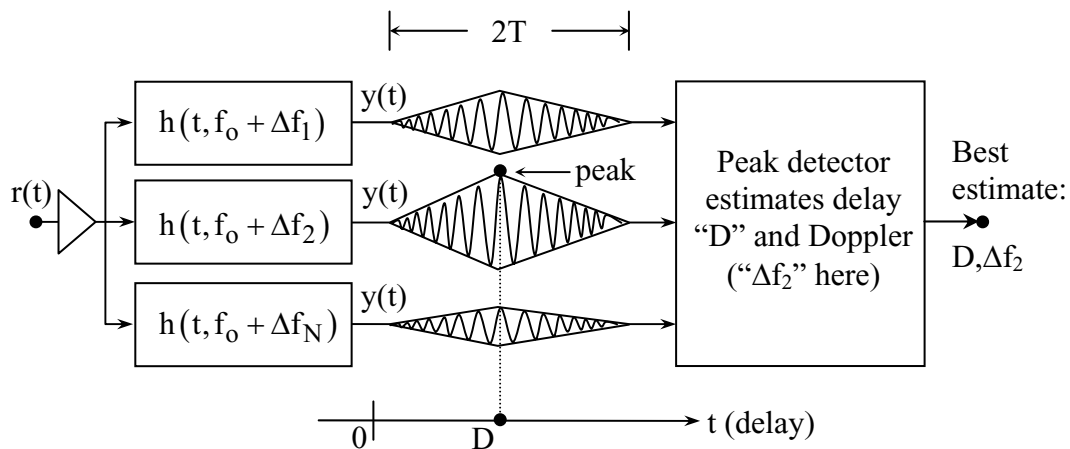

Figure 5.3-10: Matched filter radar receiver for delay and Doppler estimation

An alternative form of the same receiver would insert square-law detectors operating on the output of each filter, together with low pass filters which produce ambiguity functions resembling those of Figure 5.3-9. The matched filters in Figure 5.3-10 would take the form:

$$h(t, f_o + \Delta f_1) = P(-t)\cos(-2\pi(f_o + \Delta f_1)t)$$ (5.3.7)

where P(-t) is the time inverted waveform of the transmitted pulse, as illustrated in (5.3.7).

In the case of joint detection of range and doppler it is useful to construct the range-doppler ambiguity function. Consider the original transmitted CW pulse and its doppler-shifted counterpart illustrated in Figure 5.3-11.
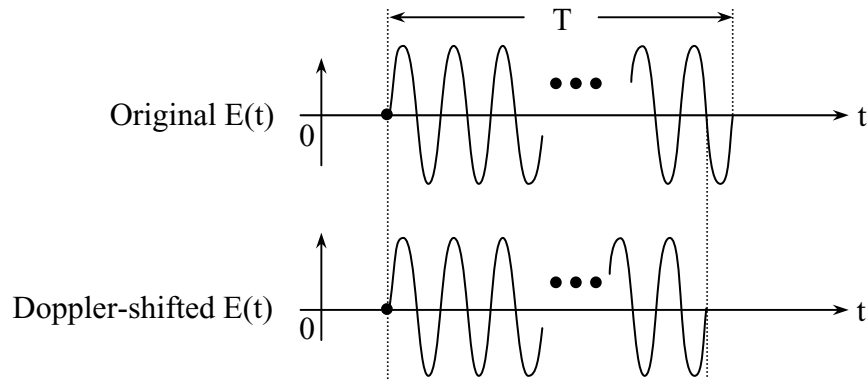
Figure 5.3-11: Original and Doppler-shifted CW pulse

The matched filter for the transmitted pulse for Figure 5.3-11 would yield a null response for the illustrated doppler-shifted waveform because it has been compressed one-half cycle over the pulse duration. That is, by shifting $\Delta f/f = (1/2$ cycle$)/(fT$ cycles$)$ we have arrived at the first null in the matched filter response $y(\tau, \Delta f)$. Therefore this first doppler null in matched filter response occurs when $\Delta f = 1/2T$ Hz. The complete doppler ambiguity function for y(t) is illustrated in Figure 5.3-12, and suggests a doppler resolution of 1/2T Hz. If we are certain the target is a point source, and the receiver noise level is low, we can determine the location of the peak and the doppler shift substantially more accurately than the widths of the range-doppler ambiguity function.
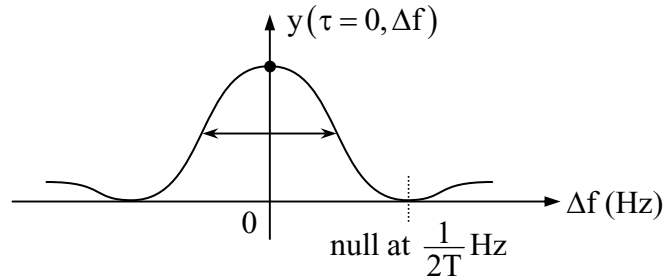
Figure 5.3-12: Doppler ambiguity function

Figure 5.3-13 suggests the form of the combined range-doppler ambiguity function for a linear detector.
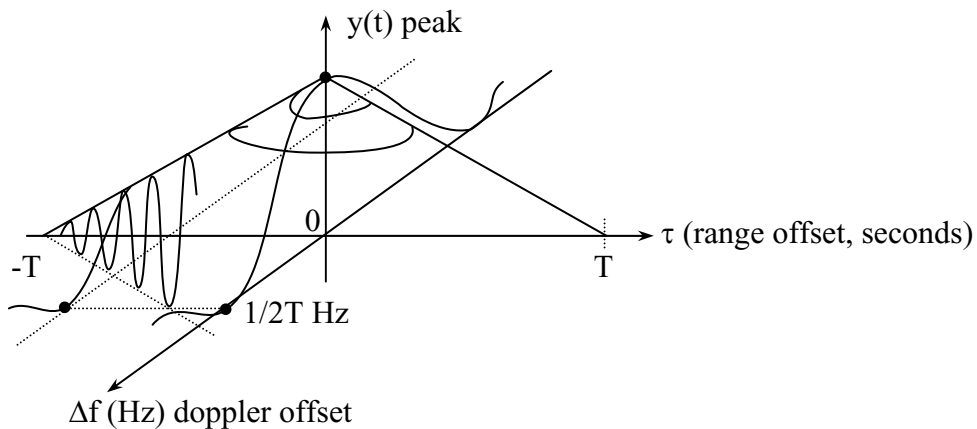

Figure 5.3-13: Range-doppler ambiguity function for a CW radar pulse
of duration T

For this pulse the half-power width in range is approximately T seconds, and the doppler width is approximately 1/2T Hz, so the time-frequency uncertainty for a CW pulse is $\Delta\tau\Delta f \cong 1/2$, roughly consistent with the Heisenburg uncertainty principle. It is also useful to talk about the *time-bandwidth product* of this CW radar pulse, which is $BT \cong 1$, where $B \cong 1/T$ Hz.

The output $z(\tau,\Delta f)$ of the matched-filter range-doppler receiver of Figure 5.3-10 can generally be expressed as the convolution of the range-doppler ambiguity function $R(\tau,\Delta f)$ with the range-doppler target response function $\sigma_s(\tau,\Delta f)$. In this case our resolution is limited largely by T in delay, and 1/2T in doppler resolution. If we are certain we have a point source, then our delay and doppler resolution might approach ~0.01T and 1/100T, or better, provided the SNR is sufficiently high. Using iterative interpretation techniques similarly high resolution is possible if we are certain we have only two point sources, provided they are not too close

279

together. In practice, we are often not sufficiently certain of the source structure to achieve these levels of performance.

Modern radar systems often transmit pulses exhibiting time-bandwidth products that are much larger than one, for example, by using the sort of waveform illustrated in Figure 5.3-14.
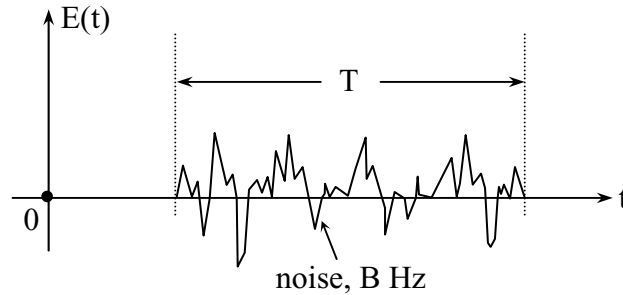

Figure 5.3-14: Pseudo-random white noise radar pulse having a large time-bandwidth product

Unlike the CW signal considered previously, such a pseudo-random noise signal can have almost arbitrarily large bandwidth B. *Pseudo-random noise* (PRN) is any noise-like waveform that can be reproduced exactly, and only appears to be random. It typically is implemented as a random binary sequence, such as that illustrated in Figure 5.3-15.
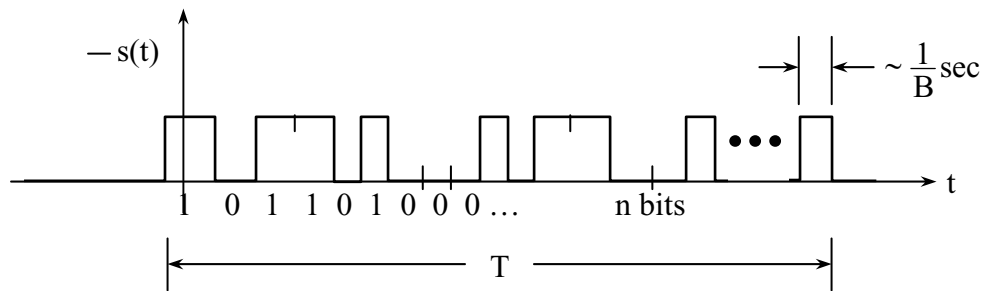

Figure 5.3-15: Binary pseudo-random noise sequence

The bandwidth of the binary noise signal is roughly the reciprocal of a single pulse width, and since the pulse length T can be made arbitrarily large, so can the time-bandwidth product BT >> 1.

The ambiguity function $z(\tau, \Delta f)$ for BT >> 1 typically takes the form illustrated in Figure 5.3-16.
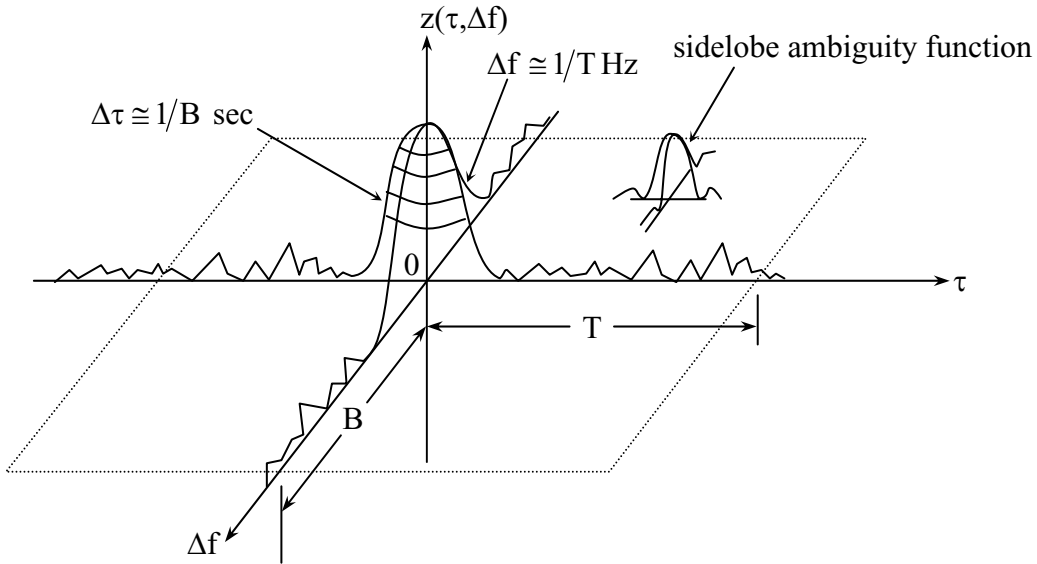
Figure 5.3-16: Range-doppler ambiguity function for a large time-bandwidth product.

In general we try to utilize PRN sequences for which the ambiguity function in range is sharply peaked with $\Delta\tau \cong 1/B$ seconds, and the doppler ambiguity $\Delta f \cong 1/T$ Hz. Unfortunately, such ambiguity functions have random fluctuations which are non-zero outside the narrow peak, and extend to $\pm B$ and $\pm T$ before becoming zero. For such a PRN pulse, the time bandwidth product is:

$$\Delta\tau\Delta f \cong 1/BT << 1 \tag{5.3.8}$$

In general, such large time-bandwidth functions are designed by trial and error rather than by any analytic means. The criteria used to select PRN sequences may involve minimization of the rms response outside the central peak, or may simply limit the maximum sidelobe response. Very often such waveforms take the simple form of a chirp, which is a sine wave that systematically drifts up or down in frequency over the pulse duration. This simplicity often offers certain implementation advantages; the matched filter can be implemented using analog circuits.

In general, such large time-bandwidth functions are designed by trial and error rather than by any analytic means. The criteria used to select PRN sequences may involve minimization of the rms response outside the central peak, or may simply limit the maximum sidelobe response. Very often such waveforms take the simple form of a chirp, which is a sine wave that systematically drifts up or down in frequency over the pulse duration. This simplicity often offers certain implementation advantages; the matched filter can be implemented using analog circuits.

5.4 SYNTHETIC APERTURE RADAR

5.4.1 Unfocused synthetic aperture radar

Synthetic aperture radar (SAR) exploits the relative movement between a radar antenna and a target field, and the coherent nature of the illuminating radar pulse, to synthesize apertures that can be substantially larger than the instantaneous physical aperture. That is, because the transmitted radar pulses are coherent with one another, the signals received by the radar while in several different sequential positions can be coherently combined to achieve the same resolution as would a conventional multi-antenna interferometric array, such as those discussed in Section 5.1.

Consider the typical fixed array of antennas illustrated in Figure 5.4-1, which can achieve a synthesized antenna beamwidth $\theta_B \cong \lambda/L$, where L is the overall length of the illustrated linear array. A synthetic aperture radar system mounted on a moving airplane moving at some velocity $\overline{v}$ can achieve comparable resolution by summing the signals received from multiple pulses as the airplane traverses a distance L in a line sufficiently straight that phase errors due to random aircraft position are not deleterious. Before summing, each pulse is appropriately delayed to produce simultaneity and phase coherence.
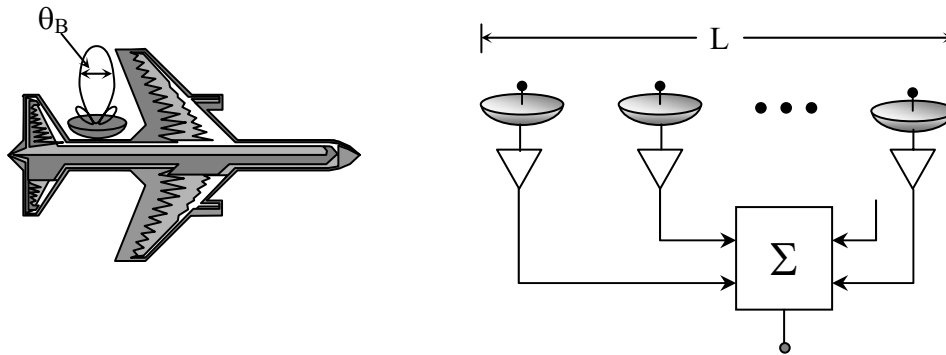

Figure 5.4-1: Synthetic aperture equivalents of a fixed antenna array and a single moving antenna

Let's consider the system response to an ideal point reflector as a moving SAR platform passes by while emitting a series of brief CW pulses. This configuration, the transmitted waveform, and the phase and amplitude of the received echo are sketched in Figure 5.4-2.
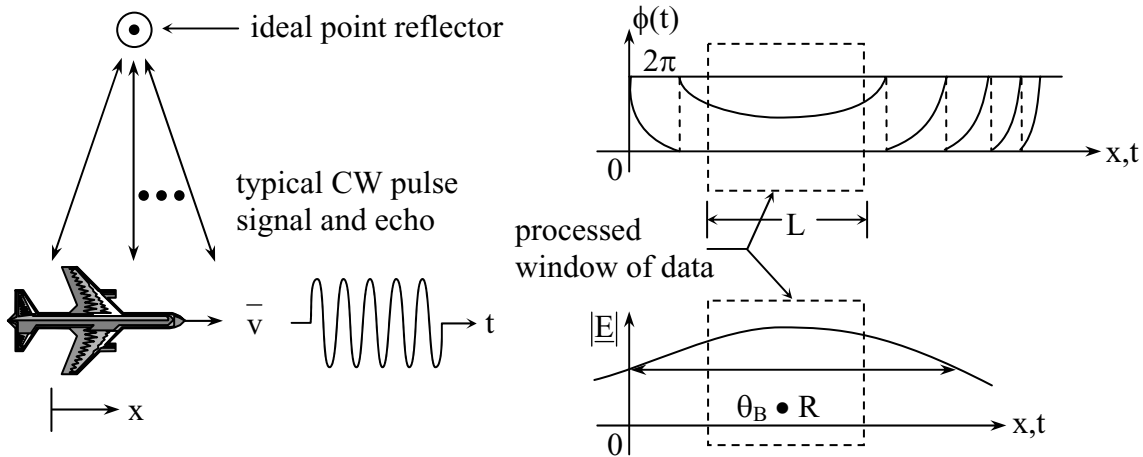
Figure 5.4-2: Phase and magnitude response for an ideal point reflector
observed by SAR

If the antenna on the moving platform is transmitting at right angles to the platform velocity vector, then the phase $\phi(t)$ of each pulse returned as the platform is abreast of the target will change very slowly, as suggested by the values of $\phi(t)$ in Figure 5.4-2. The amplitude of the received signal $|\underline{E}|$ decays as $1/R^2$ where R is the target range; the received power decays as $1/R^4$. If we are in the far field of the synthesized antenna, so that $R >> 2\,L/\lambda^2$, then the received pulse power and phase variation $\phi(t)$ over the path L will be small compared to 2π. In an unfocused SAR the window of data processed at any one time is of sufficiently short duration that this far-field approximation applies and the phase $\phi(t)$ associated with a single ideal point reflector is essentially constant. If the scattering cross-section of the target is a function of angle, then the phase and magnitude response for the received train of impulses may deviate from those illustrated in Figure 5.4-2. The resulting sampled aperture field $\underline{E}(x)$ can be Fourier transformed to yield an image as shown later in (5.4.5).

The objective of SAR is to produce a 2-D or 3-D characterization of the source. In the 3-D case we seek the source scattering cross-section σ_s as a function of azimuth, elevation, and range. In general, the resolution in range is governed by the pulse time-bandwidth product, the resolution in azimuth is governed by the length L of the synthesized aperture, and the range in elevation (orthogonal to the radar velocity vector) is governed by the diameter of the radar antenna in elevation.

In the case illustrated in Figure 5.4-2 the window width L for which the data is processed would be limited by the phase errors over the aperture, because the point source in this case is too close to the transmitter. Normally the antenna diameter D is chosen so that its antenna beamwidth $\lambda/D \cong \theta_B$ is sufficiently small that the magnitude of the response $|\underline{E}|$ decays before the phase $\phi(x)$ drifts more than ~1 radian. In this case the antenna diameter D controls the

effective aperture $L \cong \theta_B R \cong \lambda R/D$ and the processing window used to derive the target cross-section σ at R and x_o. The geometry is suggested in Figure 5.4-3.
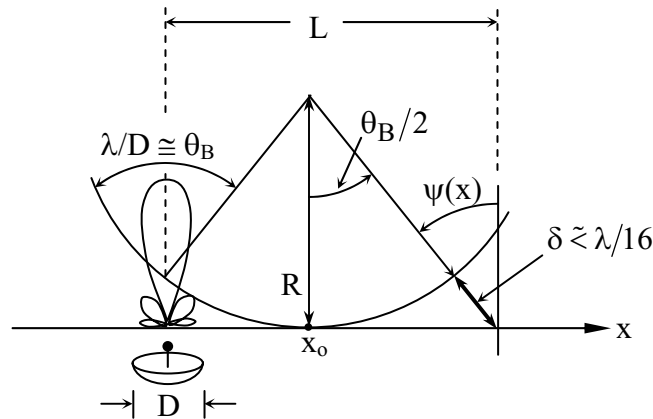

Figure 5.4-3: Geometry controlling angular resolution of unfocused SAR's

This geometry leads to the important result that the maximum lateral resolution at the range of the target is approximately D, the diameter of the moving aperture antenna. That is, the lateral (width) resolution w of the unfocused SAR at the range R of the target is approximately:

$$w \cong R\lambda/L = R\lambda/(\lambda R/D) = D \tag{5.4.1}$$

The temptation to seek high spatial resolution by shrinking D is balanced by the resulting requirement for more transmitter power to maintain a given SNR and by the need to maintain phase coherence over longer flight distances L.

Figure 5.4-3 also presents the geometry that leads to the requirement that $R \geq 2L^2/\lambda$ in order for the target to be in the far field of the synthesized aperture where the SAR is unfocused. This minimum range R of an unfocused SAR can be found by evaluating the phase discrepancy δ at the edge of the synthesized aperture. Referring to the figure, the square of the maximum range is:

$$R^2 + (L/2)^2 = (R+\delta)^2 \cong R^2 + 2\delta R \tag{5.4.2}$$

Solving (5.4.2) for δ yields:

$$\delta \cong L^2/8R \leq \lambda/16 \tag{5.4.3}$$

To set $\delta = \lambda/16$ implies a roundtrip phase error of $\lambda/8$, or 45°. Solving (5.4.3) for R yields the desired minimum range for an unfocused SAR:

$$R \geq 2L^2/\lambda \tag{5.4.4}$$

The nature of the reconstructed image can be seen from:

$$
\begin{array}{ccccc}
w(x) & \bullet & \underline{E}(x) & \leftrightarrow & \hat{\underline{E}}(\psi_x) \\
\updownarrow & & \updownarrow & & \updownarrow \\
W(\psi) & * & \underline{E}(\psi) & = & \hat{\underline{E}}(\psi)
\end{array}
\tag{5.4.5}
$$

where $w(x)$ is the sampled boxcar weighting function corresponding to the length L of the synthesized aperture, and $\underline{E}(x)$ is the aperture field distribution; $W(\Psi)$ and $\underline{E}(\psi)$ are their respective Fourier transforms. For the case where the aperture field distribution $\underline{E}(x)$ is a constant and we ignore the fact that $w(x)$ is sampled (because the sampling rate is very high), Figure 5.4-4 suggests the point target response function is a sinc function of angle ψ_x.
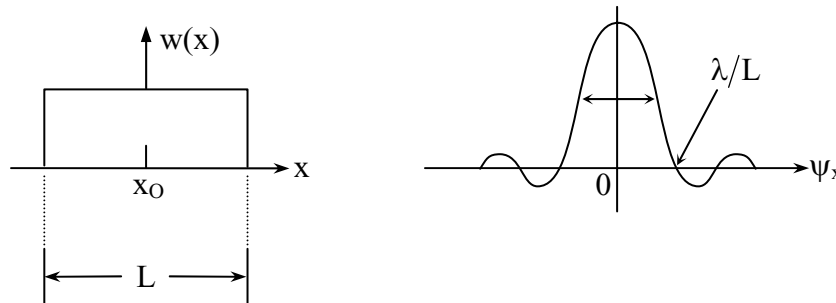

Figure 5.4-4: Point target response function for an unfocused SAR

Because we wish to avoid multiple point source responses (aliasing) in the synthesized image we must have a sufficiently high *pulse-repetition frequency (PRF)*, where $PRF = v/L'$. L' is the distance the aircraft moves between pulses and v is the aircraft velocity. Aliased images will appear at angles that are multiples of λ/L' and we wish these to lie beyond the first null of the antenna at λ/D, or $2\lambda/D$ to be conservative. That is, we would like $\lambda/L' > 2\lambda/D$, or $L' \stackrel{<}{\sim} D/2$. Thus we want $PRF \stackrel{>}{\sim} 2v/D$, but not so high as to introduce confusion, as explained below.

If we process the unfocused SAR data separately for each range box ΔR by Fourier transforming the corresponding data, we could obtain the synthesized radar image suggested in Figure 5.4-5. The contour plots in the figure correspond to the inferred scattering cross-section at that range and longitudinal position x. The minimum range is determined by ignoring radar echoes received earlier, and is chosen in order to meet the requirement $R \stackrel{>}{\sim} 2L^2/\lambda$. The maximum range is established by ignoring echoes arriving too late, and this boundary is usually

set by the fact that echoes beyond a certain range are usually too weak to be heard. In this situation the strong range dependence $P_r \propto R^{-4}$ is beneficial. The maximum range can also be limited by the antenna pattern if the radar is flying sufficiently high that it is looking down upon the target. This geometry is elaborated further in Figure 5.4-6.
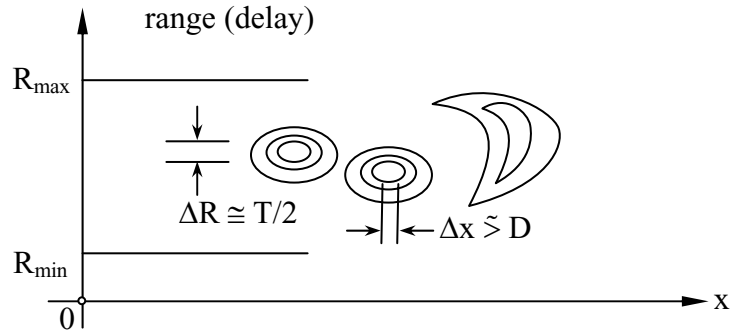

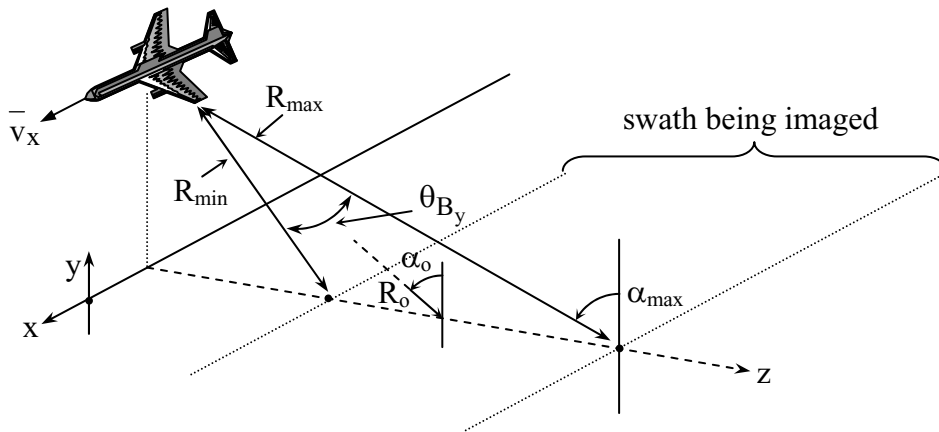Figure 5.4-5: Representative synthesized SAR image

Figure 5.4-6: Determinants of swath width geometry for SAR

Because we want to receive only one radar echo at a time, we choose our pulse repetition frequency such that:

$$2(R_{max} - R_{min})/c < (1/PRF) = L'/v_x \qquad (5.4.6)$$

where L' is the distance the platform moves between successive pulses, and v_x is the velocity of the radar platform. Unless R_{max} is limited, tardy echoes could arrive so late that they overlap those from the following pulse. We may approximate the range difference $R_{max} - R_{min}$ as $\theta_{B_y} \bullet R_o \bullet \tan\alpha_o$, where R_o and α_o are the average values over the imaged swath, as

suggested in Figure 5.4-6. By substituting this relation into (5.4.6) under the additional assumptions that $L' < D/2$ and the beam is circular so that $\theta_{B_y} \cong \theta_{B_x} \cong \lambda/D$, it follows from (5.4.6) that:

$$2\theta_B R_o \sin\alpha_o/c < L'/v < D/2v \qquad (5.4.7)$$

Therefore the distance the platform travels L' between pulses must be less than $\sim D/2$ in order to avoid aliasing in the reconstructed image.

Since the area A mapped per second is $\sim vR_o \sin\alpha$, it follows from (5.4.7) that:

$$D \gtrsim 4A/c \qquad (5.4.8)$$

Therefore to achieve good spatial resolution with an unfocused SAR, we would like to employ small values of D, but this makes it difficult to cover large swath areas rapidly. Coverage rates greater than $\sim 1000 \ km^2/sec$ require $D \gtrsim 1m$, and perhaps more complex systems
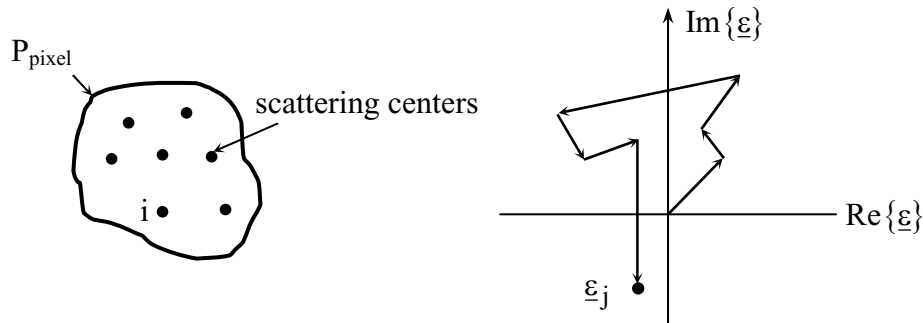

Figure 5.4-7: Complex received phasor for one pixel and one SAR pulse, when the pixel contains multiple scattering centers

The raw synthesized image produced by an unfocused SAR typically exhibits considerable *speckle*, or random amplitude fluctuations from pixel to pixel. This is because the inferred scattering cross-section at any particular point is deduced from the squared magnitude of a sum of phasors, one phasor per pulse. Consider the case illustrated in Figure 5.4-7, where a single resolved SAR pixel contains many scattering centers, such as it might if ordinary terrain were being imaged. The phasors associated with these scattering centers and a single pulse might add to produce a total phasor $\underline{\varepsilon}$, as illustrated for a single pulse.

In the synthesized image many such random phasors are superimposed, one per pulse j, to produce the final synthesized image, so that the total power associated with a single pixel in the final synthesized image can be represented as:

287

$$P_{pixel} = \left| \sum_j \varepsilon_j \right|^2 = \left| \sum_{i,j} a_{ij} e^{j(\omega t + \phi_{ij})} \right|^2 = r^2 \tag{5.4.9}$$

where the subscript i refers to the scattering center within the pixel, and the index j refers to the pulse number in the set being superimposed. Since the phase angles ϕ_{ij} are typically distributed uniformly over 2π radians, it follows that the sum of the real parts of the total phasor and the sum of the imaginary parts of the same phasor are individually gaussian random variables with zero mean and variance σ^2. The probability distribution of the magnitude squared of this sum of phasors is Laplacian, and is given by:

$$p\{r^2\} = p\left\{ \left| \sum_j \varepsilon_j \right|^2 \right\} = \frac{1}{4\sigma_p^2} e^{-r^2/2\sigma_p^2} \tag{5.4.10}$$

where σ_p^2 is the variance of r^2. The distribution for $p\{r\}$ is suggested in Figure 5.4-8.

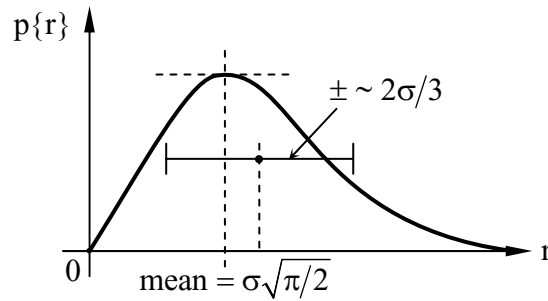
Figure 5.4-8: Probability distribution for the SAR phasor magnitude

For this probability distribution it follows that

$$E[r] = \sigma\sqrt{\pi/2} \tag{5.4.11}$$

$$E[r^2] = \sigma^2(2 - \pi/2) \tag{5.4.12}$$

From these expressions we can show

$$\sqrt{E[r - \bar{r}]^2} \cong 2\sigma/3 = \sigma_p \tag{5.4.13}$$

Note that the form of the probability distribution is independent of the number of phasors being added. As a result, in raw SAR images having full spatial resolution the ratio of the standard deviation σ_p of the phasor magnitude squared to the mean phasor intensity $\bar{p}$ is approximately

$$\frac{\sigma_p}{\bar{p}} \cong \frac{2\sigma/3}{\sigma\sqrt{\pi/2}} = 0.53 \qquad (5.4.14)$$

This important result says that any full-resolution SAR image will be extremely grainy because the standard deviation of pixel echo strength is comparable to its mean value. The standard solution to reducing this noise is simply to blur the image spatially. The point is that when one sums larger number of phasors, the standard deviation does not decline with their number N, whereas the standard deviation relative to the mean does decline inversely with the square root of the number of real numbers averaged. That is, if we average M^2 pixels the ratio of (5.4.14) becomes 0.53/M.

Another way to average is to run multiple unfocused SAR's in parallel. One SAR might look slightly forward, one might look to the side, and one might look to the rear, as suggested by Figure 5.4-9. This generally requires that the transmitting antenna be a multibeam phased array, although a mechanically steered solid aperture could be used. Yet another approach is to have only one antenna system, but to operate in multiple bands, where each band yields an independent image. Of course the same total bandwidth could alternatively be used to yield more range resolution and pixels, which then could be averaged spatially to achieve a comparable result.
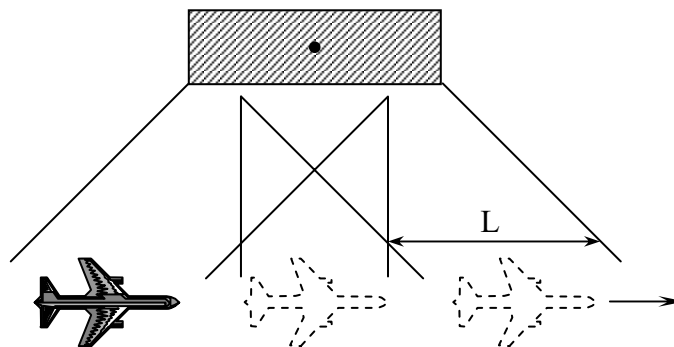

Figure 5.4-9: Multi-look SAR configuration

It is clear from the foregoing analysis that the same SAR performance results whether the antenna is moving linearly with respect to the source, as might happen when imaging geological formations from an aircraft, or whether the antenna is stationary and the source flies by in a straight line, as might a satellite in orbit about the earth. Less obvious is the fact that the target might follow an arbitrary known trajectory and rotate.

5.4.2 Focused SAR or Phase-focused SAR

If the strip map imaged by SAR is within the near-field or Fresnel region of the synthesized aperture of length L, then a range-dependent phase correction must be applied before reconstructing the image at that range. These corrections are unique to each point in the reconstructed image.

For example, consider a single point target that yields the phase history illustrated in Figure 5.4-10.
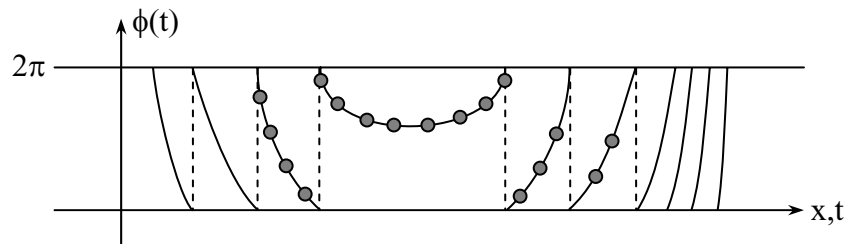

Figure 5.4-10: SAR point-target phase history; each dot corresponds the phase of a single received CW pulse

When the moving platform is abreast of the target near the point of closest approach the phase of the received CW pulse $\phi(t)$ changes very slowly; otherwise the phase can move through 2π very rapidly. If the physical aperture D is sufficiently large, pulses outside the constant-phase region are in the sidelobes of the physical antenna and therefore have negligible impact. For example, it is easy to show that if the roundtrip phase error were $\lambda/2$ at the extreme edge of the synthesized aperture where x = $\pm$L/2, then the minimum range R_o must be larger than $2D^2/\lambda$ A more conservative metric for aperture phase error, $\delta < \lambda/16$, yields $R_o > 16D^2/\lambda$. Since the phase errors associated with targets at each distance are known as a function of down-track aperture position x, the phase corrections necessary to produce constant phase prior to reconstruction of the image are known and can be employed.

It is important to note that by focusing the phase for each imaged point, the length L of the synthetic aperture can be increased substantially, and the spatial resolution in x can be less than D, provided that the finite aperture D is steered so as to observe the target as it passes by. This can be accomplished by using a mechanically rotated antenna, or by using a phased array with either a single steered beam or multiple beams observing several directions simultaneously. The limiting down-track resolution approaches λ as the synthesized aperture L approaches the target range R. The resolution in range remains approximately c/B, where c is the velocity of light and B is the pulse bandwidth.

To approach these limiting resolutions for very long synthesized apertures, it may also be necessary to delay the pulse envelope as well as the phase. That is, envelope delay corrections should be employ if the extra roundtrip pulse delay experienced at the ends of the aperture where $x = \pm \, L/2$ is a significant fraction of the pulse length (meters) of c/B.