8

Ranging systems

8.1 Introduction

Chapters 5 to 7 considered *passive* sensors, detecting naturally occurring radiation. In this chapter and the next we shall discuss *active* sensors, which emit radiation and analyse the signal that is returned by the Earth's surface or atmosphere. We have already identified three possible classifications of remote sensing systems, distinguishing between passive and active and between imaging and non-imaging, as well as classifying them according to the wavelength of radiation employed. We can also classify active systems according to the use that is made of the returned signal. If we are principally concerned with the time delay between transmission and reception of the signal we shall call the method a *ranging technique*, whereas if we are also (or mainly) interested in the strength of the returned signal we shall call it a *scattering technique*. The distinction between the two cannot be made entirely rigorous, but it provides a useful way of thinking about active remote sensing systems. It is clear that ranging systems are simpler both to visualise and, because of their less stringent technical demands, to construct, and we shall therefore consider them first. In chapter 9 we discuss the scattering techniques.

8.2 Laser profiling

Laser profiling (or *laser altimetry*) is the simplest application of the lidar (LIght Detection And Ranging) technique.[1] Conceptually, it is extremely straightforward. A short pulse of 'light' (visible or near-infrared radiation) is emitted towards the Earth's surface by the instrument, and its 'echo' is detected some time later. By measuring the time delay and knowing the speed of propagation of the pulse, the range (distance) from the instrument to the surface can be determined. By transmitting a continuous stream of pulses, a profile of the range can be built up, and if the position of the platform as a function of time is accurately known the surface profile may then be deduced.

[1] Other applications of lidar are discussed in chapter 9.

The operation and construction of a typical laser profiler are shown schematically in figures 8.1 and 8.2. The transmitter is a semiconductor laser, usually Nd:YAG (neodymium:yttrium–aluminium–garnet) operating at 0.53 μm or 1.06 μm, or GaAs (gallium arsenide) operating at 0.9 μm. This is capable of producing a short (of the order of 1 ns), intense pulse with a small angular width. The receiver is a photodiode (see chapter 6). An interval timer with a resolution of the order of 1 ns is started by the signal that generates the transmitted pulse, and stopped by the received pulse. The travel time of the pulse, T_t, is given by

$$T_t = \frac{2H}{v_g} \tag{8.1}$$

where H is the range and v_g is the group velocity (section 3.1.3) of the pulse. As we saw in figure 3.4, the group velocity for optical and near-infrared radiation propagating in dry air differs from c, the speed of light *in vacuo*, by at most 0.03%, so the error caused by setting $v_g \approx c$ in equation (8.1) is small. Atmospheric correction of laser profile data is discussed in section 8.2.1.

The desirable features of such a system are that it should achieve a high spatial resolution at the surface (i.e. the sampled points should be close together) and a high-range resolution, and that the sensitivity should be great enough to detect signals returned from weakly reflecting surfaces.

The accuracy ΔT_t with which the travel time can be determined is normally governed by the *rise time* t_r of the received pulse, and its signal-to-noise ratio S. This can be understood from figure 8.3. If V_s is the voltage amplitude of the received pulse and V_n is the amplitude of its variation due to noise, the (voltage) signal-to-noise ratio is defined as $S = V_s/V_n$. It is evident from the figure that the greatest accuracy with which the timing of the received pulse can be determined is given by

$$\Delta T_t = \frac{t_r}{S} \tag{8.2}$$

although the precision of the system may in fact be limited by that of the interval timer.

A typical pulse transmitted by a laser profiler will have a rise time t_r of a few nanoseconds, although the received pulse may be somewhat longer if the sur-

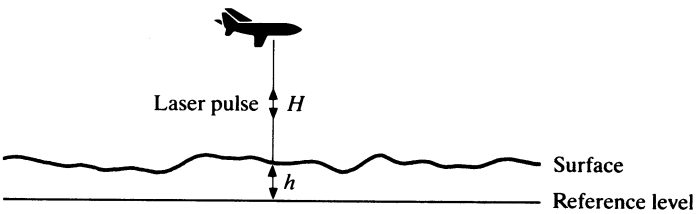

Figure 8.1. Principle of operation of a laser profiler.

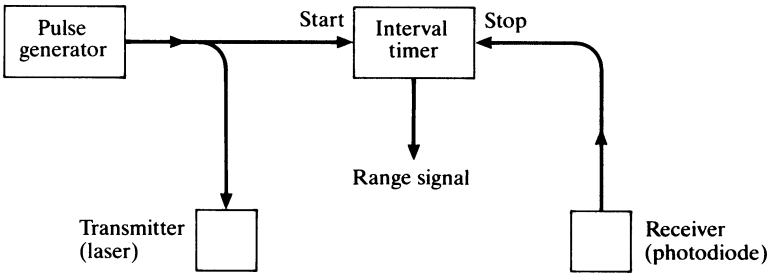

Figure 8.2. Construction of a laser profiler (schematic).

face that is being profiled is particularly rough. The signal-to-noise ratio S of the received pulse will depend on the reflectivity of the surface and the range H, as well as on system parameters such as the transmitted power, and less easily calculated influences such as the amount of incident sunlight, the weather, and atmospheric attenuation.

If pulses are transmitted at a frequency p (called the *pulse repetition frequency* or PRF) and the platform velocity is v, the linear sampling interval is v/p. If the angular beamwidth of the system is $\Delta\theta$, the linear dimension of the footprint is $H\Delta\theta$. It is obviously desirable that $H\Delta\theta$ should be no larger than some maximum value set by the nature and type of the surface under investigation. It might also be imagined that there would be no point in reducing the value of v/p below $H\Delta\theta$, but this is not so. Any decrease below this value means that, in effect, a number of independent measurements of the range is being made over a single footprint, and averaging over these measurements will therefore improve the range accuracy. Since the number of independent measurements is

$$N = \frac{H\Delta\theta}{v/p}$$

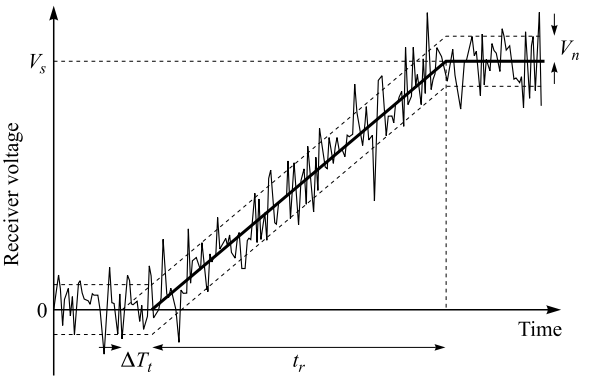

Figure 8.3. Determination of the arrival time of a noisy pulse. The accuracy ΔT_t with which the pulse can be located is $t_r V_n / V_s$.

and the improvement in range accuracy is proportional to $\sqrt{N}$, we may (provided that $N > 1$) write the range accuracy as

$$\Delta H = \frac{v_g t_r}{2S}\left(\frac{v}{pH\Delta\theta}\right)^{1/2} \tag{8.3}$$

For example, a typical airborne system might have $t_r = 5$ ns, $S = 1$, $v = 50\,\mathrm{m\,s^{-1}}$, $H = 200\,\mathrm{m}$ and $\Delta\theta = 0.001$ radian. At a PRF $p = 1000\,\mathrm{s^{-1}}$, this would give $N = 4$ and $\Delta H = 0.38\,\mathrm{m}$. Of course, a further increase in range resolution can be obtained by averaging over more than four pulses (i.e. over more than one footprint, and hence at the expense of the horizontal resolution). The system we have just described has a horizontal resolution $H\Delta\theta$ of 0.2 m. By averaging over a horizontal distance of, say, 1 m, the range resolution can be improved to 0.17 m.

Equation (8.3) shows that the measurement accuracy is increased by increasing the PRF. However, if p is increased beyond a certain point, the measured range will become ambiguous, for if the travel time T_t exceeds the interpulse period $1/p$ it will not be certain which echo belongs to which transmitted pulse. In this case, the calculated range will suffer from a *range ambiguity* of

$$H_{\mathrm{amb}} = \frac{v_g}{2p} \tag{8.4}$$

in the sense that the calculated range H may be increased or decreased by integer multiples of H_{amb} without changing the apparent travel time.[2] For this reason, it is desirable to operate the system in such a way that $H_{\mathrm{amb}} > H$, namely that

$$p < \frac{v_g}{2H} \tag{8.5}$$

For an airborne system this imposes an upper limit of tens or hundreds of thousands of pulses per second, whereas for a spaceborne system it is a few hundred pulses per second. In practice, laser profilers are normally operated well within these limits and no range ambiguities arise.

Airborne laser profiling has found applications in topographic mapping, and its high range accuracy (of the order of 0.1 m) has made it suitable for civil engineering applications. It has also proved particularly useful in reconnaissance of sea ice, where a knowledge of the freeboard (the height of the ice surface above the water level) allows the extent of the submerged portion to be estimated. Figure 8.4 illustrates typical output from a laser profiler.

The first laser profiler to be operated from space is the *Balkan-1* instrument, carried on the *Mir* space station ($H \approx 350\,\mathrm{km}$) since 1995. This operates at a PRF of $0.18\,\mathrm{s^{-1}}$, giving it an along-track sampling interval of about 45 km. The footprint width is 150 m and the range resolution 3 m. In the near future, the *GLAS* (Geoscience Laser Altimeter System) should be operated from the *EOS*-

[2] This is an example of the phenomenon of *aliasing*, discussed in greater detail in section 10.3.4.5

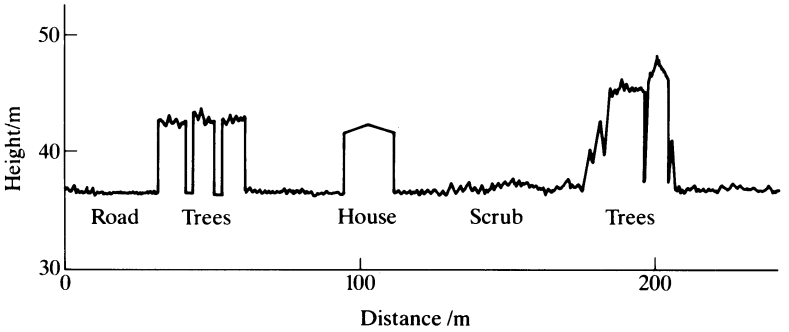

Figure 8.4. Typical output from an airborne laser profiler (modified from Jepsky, 1985).

Laser Alt satellite ($H \approx 700$ km). This instrument, which will have a PRF of $40\,s^{-1}$, will have a footprint width of 200 m and a range resolution of 0.1 m.

8.2.1 Atmospheric correction of laser profiler data

As we remarked in the previous section, the propagation speed of the pulses is given by the group velocity, and this is different from (slightly less than) the speed of light c *in vacuo*. Under normal atmospheric conditions the group velocity differs from c by at most 0.03%, so for an observation made over a range H of 1000 m the error incurred by assuming $v_g = c$ will be at most 0.3 m. For very precise measurements, the correct value of v_g should be used in equation (8.1). This depends on the wavelength, the atmospheric pressure, temperature and water vapour content.

For spaceborne or high-altitude airborne measurements, the atmospheric properties (pressure, temperature and water vapour content) are not constant along the path of the laser and it is necessary to integrate to find the travel time. Specifically, the *one-way* travel time for a path of length z is given by

$$T_t = \int_0^z \frac{1}{v_g}\, dz'$$

where v_g is the group velocity as a function of the distance z' along the path. It is often convenient to use instead the quantity P, defined by

$$P = \int_0^z \left(\frac{c}{v_g} - 1\right) dz' \tag{8.6}$$

so that

$$P = cT_t - z \tag{8.7}$$

The quantity P thus has the dimensions of a length, and is in fact equal to the one-way range error that would be incurred if we assumed that the pulse had travelled at the speed c rather than at v_g.

The quantity P is useful because, at a given wavelength, it is proportional to the integral of the number density of molecules along the path. For practical purposes, we can assume that the atmosphere consists of two components: the *dry atmosphere* (mainly nitrogen, oxygen and carbon dioxide) and the *water vapour* component. To determine the integrated number density of molecules in the dry atmosphere for a vertical path, all we need to know is the atmospheric pressure difference (specifically the difference in the partial pressures of dry air) between the top and bottom of the path. Figure 8.5 shows the value of P as a function of wavelength when this pressure difference is equal to the standard atmospheric pressure of 101 325 Pa, namely when the path traverses the entire atmosphere. For a path that does not traverse the whole atmosphere, or that makes an angle θ to the vertical, the values shown in figure 8.5 should be multiplied by

$$\frac{\Delta p}{\cos\theta}$$

where Δp is the pressure difference between the two ends of the path, expressed in atmospheres. This expression is only valid for values of θ up to about 75°.

For the water vapour component, the integrated number density of molecules for a vertical path is usually expressed as the thickness of the layer of water that would result if the water vapour were precipitated (condensed). Figure 8.6 shows the value of P as a function of wavelength for 1 m of precipitable water.[3] In table 4.1 we noted that the total mass of water vapour in the atmosphere varies typically between 6.5 and 180 kg m^{-2}. This corresponds to a range of 6.5 to 180 mm of precipitable water, so we see that a typical value of the correction due to water vapour will be of the order of 0.05 m.

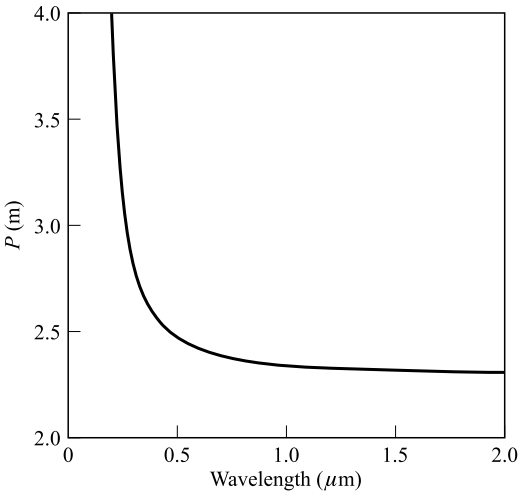

Figure 8.5. Dry-atmosphere propagation delay for one standard atmosphere.

[3] Again, for a path that is not vertical, the value of P is divided by cos θ.

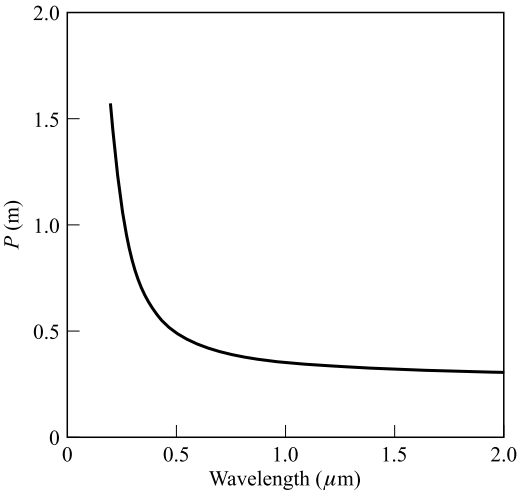

Figure 8.6. Propagation delay due to water vapour for 1 m of precipitable water.

8.3 Radar altimetry

The radar altimeter is similar in operation to the laser profiler. The basic principle, that of timing a short pulse over its round trip from the instrument to the surface and back again, is the same. Most of the differences between the kind of information obtainable from the two instruments can be ascribed to the larger beamwidth of the radar altimeter, which results from the fact that it operates at a much longer wavelength.

Figure 8.7 shows, extremely schematically, the construction of a radar altimeter. A pulse generator produces extremely short pulses at a frequency of, typically, around 10 GHz. These are fed to an antenna that radiates pulses of microwave electromagnetic radiation towards the Earth's surface. The same antenna collects the reflected pulse, and the signal is fed to a detector for subsequent analysis. The primary variable of interest is the time delay between the transmitted and received pulses, but, as we shall see, the shape of the received pulse also contains useful information.

We will first develop a very simple model of the operation of a radar altimeter, to illustrate the main features. In this model, we shall assume that the Earth's surface is flat, and that it consists of a uniform density of isotropic, incoherent, point-like scatterers. We shall also neglect the operation of the inverse square law, which we can justify if the ranges of all the scatterers

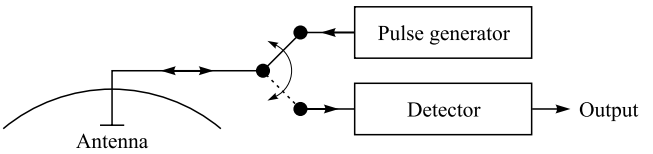

Figure 8.7. Operation of a radar altimeter (schematic).

that make a significant contribution to the received signal do not differ very much, and neglect the fall-off in sensitivity away from the axis of the antenna's main lobe. With these assumptions, the power that would be received if the antenna were to transmit *continuously* would just be proportional to the area of the Earth's surface illuminated by it. Of course, the antenna does *not* transmit continuously, and we therefore need to consider the time-structure of the received pulse.

It is clear that the return signal, if any, received at a time t after the emission of a pulse must arise from those scatterers situated at a distance $ct/2$ from the altimeter.[4] It will be convenient to describe this in terms of a 'scattering zone' that propagates away from the altimeter at a speed of $c/2$, such that any scatterer within the scattering zone contributes to the received signal at that time. This is shown schematically in figure 8.8.

If the distance from the altimeter to the surface is H, it is clear that no return signal will be received until

$$t = t_0 = \frac{2H}{c} \tag{8.8}$$

A short time Δt after this, the intersection of the scattering zone with the surface will be a circular disc of radius r, as shown in figure 8.9. Provided that $r \ll H$, this radius is given by

$$r \approx \sqrt{cH\Delta t}$$

so the area of the disc is $\pi cH\Delta t$. According to our simplifying assumptions, the received power is proportional to this area and hence to Δt. Thus, the received power will at first increase linearly with time. However, at time $t = t_0 + t_p$, where t_p is the duration of the pulse, the trailing edge of the scattering zone will just reach the surface. At times later than this the scattering zone will

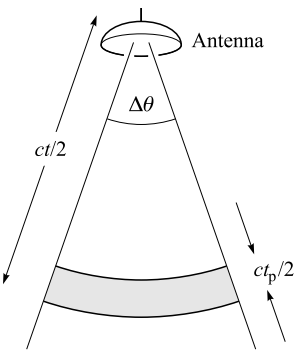

Figure 8.8. A radar altimeter emits a pulse of duration t_p beginning at time $t = 0$. $\Delta\theta$ is the beamwidth of the antenna. Any scatterer within the scattering zone (shaded) will contribute to the signal received at time t.

[4] For the time being, we will ignore the fact that the pulses propagate at the relevant group velocity and simply assume that they propagate at the speed of light c.

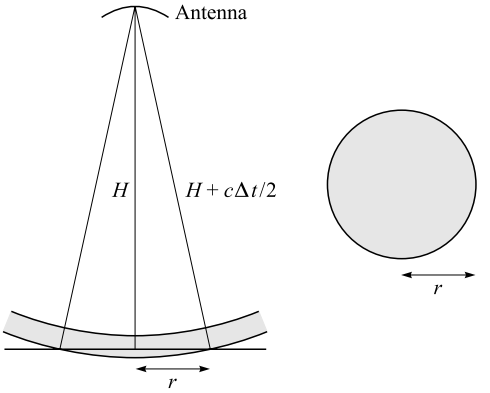

Figure 8.9. At a time $\Delta t < t_p$ after the first return signal is received, the scattering zone (shaded) intersects the surface in a disc of radius r. Left: elevation; right: plan view.

intersect the surface in an annulus, as shown in figure 8.10. The inner radius of this annulus is

$$r_1 \approx \sqrt{cH(\Delta t - t_p)}$$

and the outer radius is

$$r_2 \approx \sqrt{cH\Delta t}$$

so the area of the annulus is πcHt_p. This is independent of the time, so we can now see that the prediction of our simplified model is that the received power will increase linearly from zero at $t = t_0$ until $t = t_0 + t_p$, whereafter it will remain constant. This is shown in figure 8.11.

Inspection of figure 8.11 shows that the received power contains useful information only during the period $t_0 \leq t \leq t_0 + t_p$. It is clear that no further

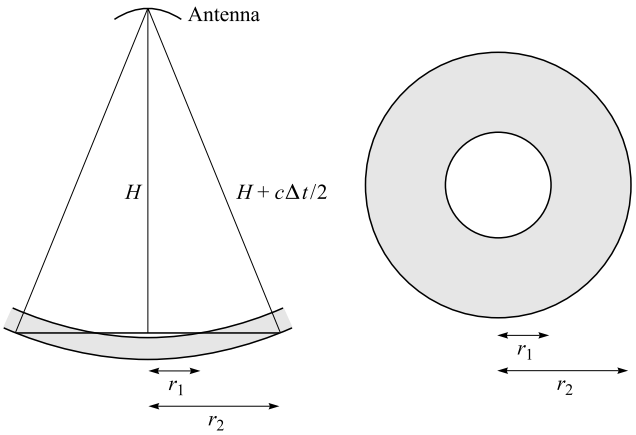

Figure 8.10. At a time $\Delta t > t_p$ after the first return signal is received, the scattering zone intersects the surface in an annulus. Left: elevation; right: plan view.

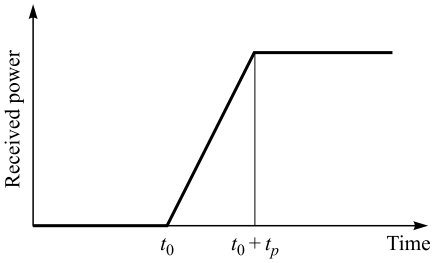

Figure 8.11. Time-dependence of the power received by a radar altimeter above a flat surface, according to the simple model derived in the text.

information is obtained at later times. This means that, in effect, the footprint of the instrument is a disc of radius r_p, where

$$r_p = \sqrt{cHt_p} \tag{8.9}$$

is the radius of the scattering disc at time $t_0 + t_p$.

In deriving this model, we have assumed that the beamwidth of the antenna is sufficiently large that variations in its response at increasingly large angles from the beam axis can be neglected. In general this will not be true, and the effect of the declining power pattern of the antenna (and also of the increasing distance of the scatterers from the antenna) as the time increases beyond t_0 will be to cause the received power to be less than predicted by our simple model, by a factor that increases with time. Thus, figure 8.11 should in fact be modified to look something like figure 8.12. We can, however, distinguish two limiting cases. The first of these is where the reduction in power as a result of the beam power pattern is negligible. The condition for this to be true is that

$$H\Delta\theta \gg 2r_p$$

in which case the altimeter is said to be *pulse-limited* and equation (8.9) correctly describes its spatial resolution. However, if

$$H\Delta\theta \ll 2r_p$$

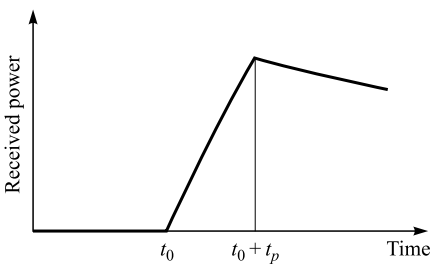

Figure 8.12. Modification of figure 8.11 to take account of the antenna power pattern and the increasing distance of scatterers from the antenna.

the altimeter is *beam-limited* and the spatial resolution is just $H\Delta\theta$. Most radar altimeters are pulse-limited, while laser profilers are beam-limited. As an example, we can consider a radar altimeter with a pulse length of 3 ns operating from a height of 800 km. Equation (8.9) shows that the radius of the beam-limited footprint will be approximately 850 m. From section 2.7 we know that the angular beamwidth (in radians) of an antenna of diameter D at wavelength λ is approximately 1.22 λ/D, so if $\lambda = 3$ cm and $D = 1$ m the radius of the beam-limited footprint will be about 30 km. In this case, clearly, the effect of the beam power pattern can be ignored and the altimeter is pulse-limited.

8.3.1 *Effect of the Earth's curvature*

Our model of the operation of a pulse-limited altimeter involved a number of simplifications, which we should now examine. One obviously incorrect assumption is that the Earth's surface is flat, and while this is entirely adequate for airborne radar altimeters it is not self-evident that it is valid for a space-borne instrument. Fortunately, the effect of the Earth's curvature can be dealt with rather simply. Figure 8.13 shows how figure 8.9 can be modified to take the Earth's curvature into account. The Earth is assumed to be a sphere of radius R, in which case simple trigonometry shows that

$$(H + c\,\Delta t/2)^2 = (H + R)^2 + R^2 - 2R(H + R)\cos\alpha$$

Assuming that $\alpha \ll 1$, this can be simplified to give

$$\alpha^2 \approx \frac{c\,\Delta t}{\frac{R^2}{H} + R}$$

and since $r = R\alpha$, we must have

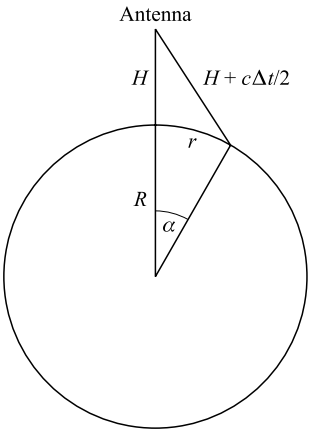

Figure 8.13. Geometry of a radar altimeter measurement when the Earth's curvature is taken into account.

$$r^2 \approx \frac{c\,\Delta t}{\dfrac{1}{H}+\dfrac{1}{R}}$$

Thus, the radius of the pulse-limited footprint can still be calculated using equation (8.9), provided that the range H is replaced by the effective height

$$H_{\text{eff}} = \left(\frac{1}{H}+\frac{1}{R}\right)^{-1} \tag{8.10}$$

Using the same example as before ($H = 800\,\text{km}$, $t_p = 3\,\text{ns}$), we see that the effective height is about 711 km, so the value of r_p is reduced from 850 m to about 800 m.

8.3.2 *Effect of coherence: range accuracy*

Our model of the response of a radar altimeter is still rather crude. We can improve on the assumption that the surface consists of a uniform distribution of isotropic scatterers by incorporating the form of the BRDF, if it is known. However, there is a more important respect in which the model fails. We have assumed that the power received by the altimeter is proportional to the number of scatterers (and hence the area) visible to it, and have thus added together the powers scattered by the various scatterers. If the radiation reaching the antenna from two scatterers is *coherent*, (i.e. if there is a definite phase relationship between the two waves), the signals are capable of *interfering* with one another. They should thus be added as vector (or more precisely phasor) quantities, with due regard to amplitude and phase.

Two points will be coherently illuminated with respect to each other if the difference between their distances from the source of illumination (the radar altimeter) is less than the *coherence length* l_c of the radiation, and their separation measured in a direction perpendicular to the propagation direction of the radiation is less than the *coherence width* w_c of the radiation. These quantities are given by

$$l_c \approx \frac{c}{\Delta f} \tag{8.11}$$

and

$$w_c \approx \frac{cH}{Df} \tag{8.12}$$

where c is the speed of light, f is the frequency of the radiation, Δf its bandwidth (which must be at least $1/t_p$, and is usually equal to it), and D is the diameter of the antenna. As before, H is the distance from the antenna to the surface. A typical spaceborne radar altimeter will have $l_c \approx 1\,\text{m}$ and $w_c \approx 10\,\text{km}$, so in practice most of the scattering zone will be coherently illuminated. The consequence of this is that the power received from a flat surface will not have the simple form shown in figure 8.11, but will instead be very

noisy, in the manner of figure 8.3 but more so, with a signal-to-noise ratio of the order of 1. However, by averaging together many pulses, something resembling figure 8.11 can be obtained. An important consequence of the fact that the signal-to-noise ratio for a single pulse is ≈ 1 is that the range accuracy of a single pulse is approximately

$$\Delta H \approx \frac{ct_p}{2} \tag{8.13}$$

consistently with equations (8.1) and (8.2).

8.3.3 *Response from a rough surface*

The last of the simplifying assumptions that we shall examine is that the surface is flat. Let us suppose instead that the surface is rough, for example an ocean surface with waves. Considering figure 8.9 again, we can see that the first return signal will now be received earlier than before, when the scattering zone just touches the tops of the waves. Similarly, the time taken for the received pulse to rise to its maximum value will be increased, because this will now correspond to the time taken for the trailing edge of the scattering zone to reach the lowest scatterers (the lowest troughs of the waves). This is illustrated in figure 8.14.

The time taken for the received signal to rise from zero to its maximum value will thus be increased beyond the value of t_p (the duration of the transmitted pulse) that we derived in section 8.3. The shape of the time-variation (the *waveform*) of the received signal will also be altered, depending on the distribution of the surface scatterers in height. In general, it can be shown that (at the same degree of approximation that we used to derive figure 8.11)

$$\frac{dP_r}{dt} \propto \int_{-\infty}^{\infty} P_t\left(t - \frac{2H}{c} + \frac{2h}{c}\right) f(h)\, dh \tag{8.14}$$

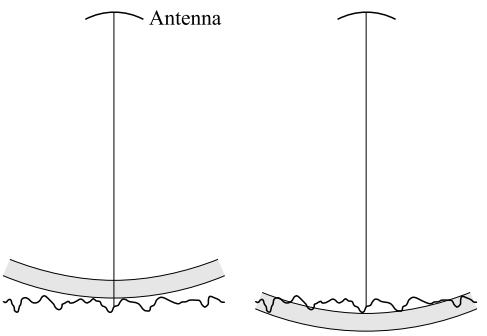

Figure 8.14. The scattering zone of a radar altimeter encounters a rough surface. Left: the instant when the first return signal is received; right: the instant when the received power reaches its maximum value.

where $P_r(t)$ and $P_t(t)$ are, respectively, the received and transmitted powers at time t, and $f(h)\,dh$ is the proportion of scatterers between heights h and $h + dh$ above the mean height of the surface.

The right-hand side of equation (8.14) is in fact a *convolution* of the surface height distribution and the shape of the transmitted pulse. Approximately, we may state that if the surface height is distributed over a range Δh, the time t_p' taken for the received power to increase from near zero to near maximum will be given by

$$t_p'^2 \approx t_p^2 + k\frac{\Delta h^2}{c^2} \tag{8.15}$$

where k is a dimensionless constant that depends on how Δh, 'near zero' and 'near maximum' are defined. Of course, equation (8.14) allows a more quantitative statement to be made if the shapes of the transmitted pulse and of $f(h)$ are known. As an example, figure 8.15 shows the waveform for a rectangular transmitted pulse (i.e. one in which the pulse is switched on abruptly, remains constant for time t_p, then is switched off again) incident on a surface for which $f(h)$ is a Gaussian distribution.

An important consequence of the broadening of the pulse received from a rough surface, from t_p to t_p', is that the radius r_p of the pulse-limited footprint is no longer given by equation (8.9). Instead, the effective resolution is coarsened to

$$r_p' = \sqrt{cHt_p'} \tag{8.16}$$

In the case of figure 8.15, for example, the rise time of the received pulse is of the order of 8 ns, so if $H_{\text{eff}} = 711$ km the radius of the pulse-limited footprint has been broadened from 800 m to about 1300 m. Pulse-broadening also implies a coarsening of the range resolution, which can be seen by substituting t_p' for t_p in equation (8.13).

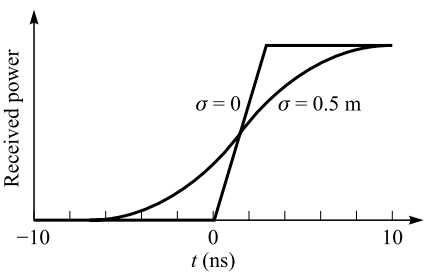

Figure 8.15. Waveform of the received power for a rectangular pulse of duration 3.0 ns incident on a flat surface and a surface with a Gaussian distribution of heights with a standard deviation of 0.50 m. The zero of the time axis is the time at which the leading edge of the pulse is received from the mean height of the surfaces.

8.3.4 Applications of radar altimetry

8.3.4.1 Sea-surface topography

Radar altimeter measurements are used extensively for characterising the topography of the ocean surface. The long-term average of the ocean surface (i.e. the average after the effects of tides, surface gravity waves, atmospheric pressure variations and wind-driven disturbances have been removed) is, with a proviso we shall discuss later, coincident with the *geoid*. This is a surface of constant gravitational potential (an equipotential surface) that is a close approximation to an *ellipsoid* of revolution with its shortest axis along the Earth's polar axis and circular symmetry about this axis. The geoid differs from this ellipsoid by distances of the order of 100 m, these variations being due to variations in the density of the Earth's mantle and lithosphere and to variations in the topography of the Earth's solid surface. Determination of the geoid is thus important for understanding the structure of the Earth and its gravity field, and also for making accurate predictions of satellite orbits (see chapter 10).

As we have already mentioned, the derivation of the mean sea surface requires the averaging of repeated measurements to reduce the magnitude of time-dependent effects. Where these phenomena vary more or less randomly in time, it is merely necessary to take the average over a time that is much longer than the correlation time of the phenomenon. However, for phenomena that are strongly periodic, such as tides, a potential problem is introduced by the fact that the sampling is also periodic, at least for a spaceborne system. It is clear that if the sampling frequency exactly matches the frequency of the tide, the tidal variation will be sampled at the same point in its cycle for each measurement, and consequently no information at all will be obtained about the temporal variability of the tide. This is another example of the phenomenon of *aliasing*, which will be discussed in greater detail in section 10.3.4.5.

Figure 8.16 is a visualisation of the global mean sea surface, using data from the radar altimeter carried on the ERS-1 satellite. It clearly shows topographic variations at a wide range of scales, but perhaps the most strikingly obvious features are those that correspond to the deep ocean-floor trenches, for example the Kermadec Trench at the lower left of the figure. The surface features corresponding to these trenches are typically 10 m deep and 200 km wide. They reflect the fact that the gravitational field strength is reduced above a trench, so that the equipotential surface moves closer to the Earth's centre in this region.

The one situation in which long-term averaging of the sea-surface topography does not result in a surface that corresponds to the geoid is where the surface has a steady motion as a result of an ocean current. In the northern hemisphere, the surface of a stream of water moving at a constant velocity v relative to the Earth's surface will be tilted such that the right-hand edge of the stream is higher than the left (in the southern hemisphere the left side is higher than the right), the angle of tilt being given by

$$\frac{2\Omega v \sin \phi}{g}$$

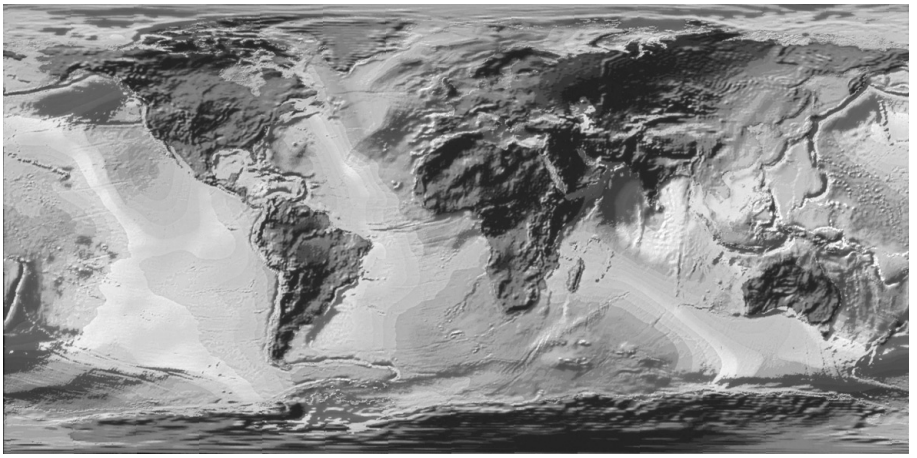

Figure 8.16. Visualisation of the global mean sea surface, using data from the ERS-1 radar altimeter. (© ESA 1994.)

where Ω is the Earth's angular velocity, ϕ is the latitude and g the gravitational field strength. This phenomenon, called *geostrophic balance*, is a consequence of the Coriolis force, and the slopes are very small. For example, the Gulf Stream has a typical velocity of $2\,\mathrm{m\,s^{-1}}$, so at a latitude of 45°N the slope is 2×10^{-5} radians, but over a typical stream width of 100 km this accounts for a difference in height of 2 m across the stream. This surface tilt can be discerned in figure 8.16.

8.3.4.2 Sea-surface roughness

The sensitivity of the waveform (the shape of the received pulse) to surface roughness, discussed in section 8.3.3, means that radar altimeter measurements can be used to determine the sea state. The simplest and most widely reported measure of sea state is the *significant wave height* $H_{1/3}$, defined as the mean height (from trough to crest) of the highest third of the waves. It is approximately related to the variance σ^2 of the surface height distribution by

$$H_{1/3} \approx 4\sigma$$

The significant wave height is related to the wind speed, and can be used to determine it, although it is also dependent on the *fetch* (the distance from land, measured in the direction in which the wind is blowing) and the *duration* (the time for which the wind has been blowing). For sufficiently large fetch and duration, the sea is said to be *fully developed*, in which case the significant wave height depends only on the wind speed, and is given by

$$\frac{H_{1/3}}{\mathrm{m}} \approx 0.02\left(\frac{v_w}{\mathrm{m\,s^{-1}}}\right)^2$$

where v_w is the wind speed 10 m above the surface. On this basis, wind speeds over oceans can be determined to an accuracy of typically $\pm 2\,\mathrm{m\,s^{-1}}$.

In fact, determination of the significant wave height is also important for measuring the sea-surface topography. The reason for this is that the height distribution $f(h)$ of scatterers, introduced in section 8.3.3, is not symmetric about the mean surface height. The consequence of this is that the mean surface height is underestimated by an amount that is typically 2–3% of the significant wave height. This effect is known as *electromagnetic bias* or *sea-state bias*.

8.3.4.3 Land and ice-sheet topography

Topographic measurements over land surfaces are considerably more difficult to make using a spaceborne radar altimeter. There are essentially two technological reasons for this, both related to the fact that land surfaces exhibit considerably larger slopes than ocean surfaces. The first of these relates to a phenomenon usually referred to as 'loss of lock' or 'loss of tracking', and it can be explained roughly as follows. If we have a spaceborne radar altimeter at an altitude of 800 km, the time interval between the emission and reception of a pulse is of the order of five million nanoseconds. If the instrument is to be capable of resolving the waveform of the returned pulse, it will have to sample the pulse at intervals of the order of 1 ns or even less. In order to keep the volume of data collected by the instrument within manageable limits, the receiving and detecting part of the instrument is therefore only activated a short time before the expected return of the pulse. This clearly requires an accurate prediction of the arrival time of the next pulse, based on the last pulse or last few pulses received. Over very smoothly varying surfaces with small slopes this is not much of a problem. However, land-surface slopes can often be so steep that the next pulse will arrive well outside the time during which the instrument is ready to accept it.

The second problem that can arise in radar altimetry of comparatively steep slopes is that of *slope-induced error*. In effect, the return pulse is derived, not

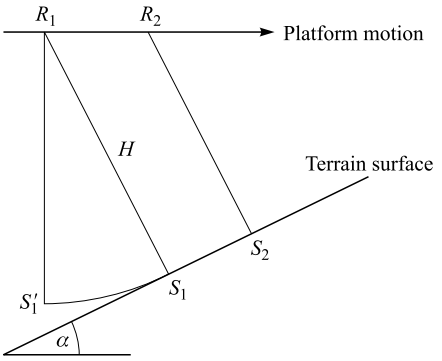

Figure 8.17. Illustration of slope-induced error in one dimension. When the radar altimeter is located at R_1, scattering actually occurs from the point S_1 but is assumed to occur at the point S_1'.

from the nadir point (the point directly beneath the instrument), but from the closest point to the altimeter. This is illustrated in figure 8.17.

When the radar altimeter is located at the position R_1, the closest point of the surface is S_1. If no correction is made for the slope-induced error, the scattering point will be assumed to be at the position S_1', where the distances R_1S_1 and R_1S_1' are both equal to the range H and S_1' is directly below R_1. For a small slope α (radians), the horizontal error is approximately $H\alpha$ and the vertical error is approximately $H\alpha^2/2$. Thus, for a spaceborne altimeter ($H \approx 800$ km) observing the ocean surface, where slopes are unlikely to exceed 10^{-4} radians, the errors are at most 80 m horizontally and 4 mm vertically, and can safely be ignored. However, over a land surface with a slope of, say, 0.03 radians, the errors are 24 km horizontally and 360 m vertically, and hence far from negligible.

Provided that the slope α does not change too rapidly, slope-induced error can be corrected. Figure 8.17 indicates how this can be achieved. If we have a second measurement from position R_2, such that the true location of the scattering point is S_2, the slope α can be deduced from the rate at which the range H changes with the along-track distance that the instrument has travelled. Once α is known, the scatterers can be assigned to their correct locations.

Two more remarks should be made about slope-induced error. The first is that the problem is in fact a two-dimensional one, so that correction requires two-dimensional coverage of the area rather than the one-dimensional transect illustrated in figure 8.17. The second is that if the angle α is large compared with the beamwidth of the antenna, little or no signal will be received from the surface unless the antenna is tilted so that its beam axis is normal to the surface.

Despite these difficulties, spaceborne radar altimetry has been applied with some success to mapping land-surface topography in comparatively flat areas, and with notable success to mapping the Antarctic and Greenland ice sheets (where the surface slopes are generally a few degrees at most).

8.3.5 *Atmospheric and ionospheric correction of radar altimeter data*

For simplicity, we have been assuming throughout our discussion of radar altimetry that the pulses propagate at the speed of light c. However, as for the laser profiler, we should really use the appropriate group velocity. The principles of correcting the measured time delay for atmospheric (principally tropospheric) delay are very similar to those for the laser profiler, discussed in section 8.2.1. There is, though, a major simplification in the case of the microwave frequencies employed for radar altimeters. The tropospheric propagation is practically non-dispersive, which means that the group velocity is equal to the phase velocity and independent of the frequency. The corresponding values of P (defined by equation (8.7)) are 2.33 metres per atmosphere for the dry tropospheric component, and 7.1 metres per metre of precipitable water. Thus, the tropospheric delay for a spaceborne observation is typically between 2.4 m

and 3.6 m depending on the amount of water vapour present in the atmosphere. Correction for this delay therefore requires that both the atmospheric pressure and water vapour distributions be known.

For a spaceborne observation we must, however, also consider the effect of the ionosphere. This was discussed in section 4.5. Rearranging equation (4.23), we find that the quantity P is given by

$$P = \frac{e^2}{8\pi^2 \varepsilon_0 m_e} \frac{N_t}{f^2} \tag{8.17}$$

where e and m_e are, respectively, the charge and mass of an electron, N_t is the total electron count, and f is the frequency. For a typical daytime total electron count of $3 \times 10^{17}\,\mathrm{m}^{-2}$ and a frequency of 10 GHz, we find that $P = 0.12$ m. If necessary (for high-precision measurements), this effect can be corrected if the total electron content is known, or, more conveniently, by using a dual-frequency radar altimeter.

8.3.6 Example: the Envisat radar altimeter

As an example of a typical spaceborne radar altimeter, we consider the RA-2 instrument that will be included as part of the payload of the Envisat satellite, orbiting at an altitude of 800 km. This is a dual-frequency altimeter, operating at both K_u-band (13.6 GHz) and S-band (3.2 GHz), so, with an antenna diameter of approximately 1 m, it has beam-limited footprints of 26 km (K_u-band) and 90 km (S-band). In its highest-resolution mode of operation, the K_u-band subsystem generates pulses of length 3.1 ns, so the diameter of the pulse-limited footprint over a smooth surface is 1.7 km. The S-band pulse length is 6.3 ns, giving a pulse-limited footprint of 2.4 km. The instrument is thus pulse-limited at both frequencies.

The range resolution for a single measurement over a flat surface is approximately $ct_p/2$, or about 0.5 m for K_u-band and 1 m for S-band. However, 100 K_u-band waveforms and 25 S-band waveforms are averaged in the instrument, improving these resolutions to approximately 5 cm and 20 cm respectively. The pulse repetition frequency for the K_u-band measurements is $1800\,\mathrm{s}^{-1}$ so each 100-waveform average is acquired in 56 milliseconds, during which time the satellite moves a distance of approximately 400 m relative to the Earth's surface. This, therefore, is the spatial sampling interval. For the S-band measurements the PRF is $450\ \mathrm{s}^{-1}$, so the spatial sampling interval is again 400 m. The principal reason for having two operating frequencies is to allow for ionospheric corrections, which are clearly necessary if range accuracies as fine as 5 cm are to be achieved.

The K_u-band subsystem can also be operated in two lower-resolution modes, with pulse lengths of 12.5 and 50 ns, respectively. These give pulse-limited footprint diameters of 3.5 and 7 km (so they are still pulse-limited), and range resolutions, after averaging, of approximately 20 and 75 cm, respectively. The purpose of these modes to allow tracking of rougher terrain.

In passing, we can make two final remarks on the technical difficulties associated with achieving accuracies of the order of 10 cm in the range measured from a spaceborne radar altimeter. The first of these is a rather obvious one. If it is desired to measure the Earth's surface topography with an absolute accuracy of 10 cm, it will be necessary to know the position of the satellite to this accuracy.

The second point concerns the production of the extremely short (nanosecond) pulses needed to obtain high range precision. It would in fact be extremely difficult or impossible to put enough energy into a pulse only a few nanoseconds long, such that the pulse would be detectable after its 1600-km round trip to the Earth's surface. Instead, *pulse compression* is used. The real pulses are, in the case of the Envisat RA-2, 20 μs long, but they are modulated in frequency. We saw in section 2.3 that a pulse of duration T must contain a range of frequencies $\Delta f \approx 1/T$. Pulse compression effectively uses the converse of this principle, which states that if we wish to construct a pulse of length t_p, we must use a range of frequencies of $\Delta f \approx 1/t_p$. The frequency modulation necessary to achieve a synthetic pulse length of 3.1 ns thus requires a bandwidth of approximately 320 MHz. In practice, this is achieved by generating a '*chirp*', which is a pulse whose frequency rises from $f_0 - \Delta f/2$ to $f_0 + \Delta f/2$ over a period of 20 μs (f_0 is the central frequency, i.e. 13.6 GHz for the K_u-band subsystem).

8.4 Other ranging systems

We have now discussed the two main types of ranging system that conform to the definition of remote sensing enunciated in chapter 1. We should also, however, briefly mention *radio echo-sounding*. This is a technique for measuring the thickness of ice sheets and glaciers, and it relies on the large attenuation length of VHF (ca. 100 MHz) radio waves in ice. Since the attenuation length in ice at these frequencies is of the order of 100 m to 1000 m (see figure 3.1), it is feasible to transmit a signal through a large body of ice and to detect the echo from the bedrock beneath it, even at a range of several thousand metres, which is typical of the Antarctic and Greenland ice sheets. This technique has been extensively and successfully employed (see e.g. Cracknell, 1981; Drewry, 1983; Rees, 1988) for mapping ice sheets and glaciers, with a range resolution approaching 1 m. However, because of the long wavelength (≈ 3 m) in free space, narrow-beam antennas are not yet technologically feasible and such remote sensing has so far been confined to observations from platforms on or relatively close to the ice surface. Satellite observations will be precluded until narrow-beam instruments can be devised and placed in orbit.

Similar techniques are used for determining the thickness of saline ice, although in this case the higher electrical conductivity and inhomogeneous structure of the medium greatly reduce the attenuation length. For this reason, high-power systems must be employed, and the distance from the platform to the ice surface must be kept small. Satellite-based remote sensing of sea-ice

thickness is an even more distant prospect than for ice sheets and glaciers, whose ice is comparatively pure and homogeneous.

Finally, we mention the use of *soil-sounding radars* in archaeological investigations. Again the technique is similar to radio echo-sounding, although the frequencies used are typical radar (microwave) frequencies rather than VHF. It has achieved a limited degree of success over dry soils in which buried artefacts produce a strong electromagnetic contrast.

PROBLEMS

1. A laser profiler is operated at a wavelength of 1 μm from an altitude of 10 000 m, and views at 45° to the nadir. Estimate the range corrections needed to account for the dry atmosphere and the water vapour component if the atmosphere contains 50 mm of preciptable water. The atmospheric pressure at 10 000 m altitude can be taken as 0.26 atmospheres.

2. (For enthusiasts.) Prove equation (8.14).

3. A radar altimeter emits a pulse whose variation of power with time is Gaussian with a length of 3.00 ns between $1/e$ points. The pulse is reflected from a surface whose height distribution is Gaussian with a range of 1.00 m between the $1/e$ points. Calculate the time for the reflected pulse to rise from 8% to 92% of its final value, neglecting coherence effects. [Note that 84% of the area under a Gaussian curve is enclosed between the $1/e$ points.]

4. (For Fourier transform enthusiasts only!) A chirp signal, in which the angular frequency rises uniformly from $\omega_0 - \Delta\omega/2$ to $\omega_0 + \Delta\omega/2$ over a time T, can be written as $\exp(i\phi(t))$, where the phase $\phi(t)$ is given by

$$\phi(t) = \omega_0 t + \frac{\Delta\omega}{2T} t^2$$

for $|t| \leq T/2$. This signal is then passed through a delay-line, whose effect on a component of angular frequency ω can be represented as multiplication by the factor

$$\exp\left(-i\omega\left(\frac{\omega_0 T}{\Delta\omega} + \frac{\omega T}{2\,\Delta\omega}\right)\right)$$

Show that the signal that emerges from the delay-line is a carrier of angular frequency ω_0, modulated by an envelope of width $2\pi/\Delta\omega$.