JPL PUBLICATION 80-90

Synthetic Aperture Radar and Digital Processing: An Introduction

Alan Di Cenzo

February 15, 1981

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

JPL PUBLICATION 80-90

Synthetic Aperture Radar and Digital Processing: An Introduction

Alan Di Cenzo

February 15, 1981

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

The research described in this publication was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under NASA Contract No. NAS7-100.

ACKNOWLEDGEMENT

I wish to gratefully acknowledge the insight provided by Dr. Chialin Wu. Also, I would like to thank Dr. Daniel Held, Dr. Fuk Li, James Huang and Dennis Moy for their constructive comments. Finally, I would like to express my gratitude to Marie Curry and her excellent staff, as well as to my secretary Shirley Woods, for splendid assistance in preparing this report for publication.

ABSTRACT

This report is intended as a rigorous tutorial on synthetic aperture radar (SAR) with emphasis on digital data collection and processing. Background information on waveform frequency and phase notation, mixing, I, Q conversion, sampling and cross-correlation operations is included for clarity. The fate of a SAR signal from transmission to processed image is traced in detail, using the model of a single bright point target against a dark background. Finally, some of the principal problems connected with SAR processing are discussed.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

SYNTHETIC APERTURE RADAR AND DIGITAL PROCESSING: AN INTRODUCTION

I. Introduction and Motivation

The purpose of this report is to provide a basic yet rigorous introduction to synthetic aperture radar (SAR), with an emphasis on digital data collection and processing. The method will be to start with simple models, and then to refine these models so they more closely approximate the real world.

The only knowledge assumed on the part of the reader is a familiarity with elementary algebra, trigonometry, and complex arithmetic, plus a faint glimmer of differential calculus. All other concepts will be developed in this report.

Completeness is not attempted here. Instead, we trace in detail what happens from beginning to end when a SAR interrogates a point target on the ground. The principle of linear superposition can then be used to infer the result when the terrain contains distributed targets, i.e., multiple point targets. In the last section, some problems connected with SAR are sketched for the interested reader. Some knowledge of probability is assumed.

We begin in this section by describing a simple kind of imaging radar: a real aperture radar with no range compression. While this radar is not a SAR, an understanding of how it generates imagery will prove valuable in the understanding of a SAR.

Our primitive radar may be thought of as a flashlight pointing downward and outward from the side of a moving airplane. Figure 1 shows a view of the beam as seen from directly behind the aircraft. The angle β_r is called the "range" elevation beamwidth and is related to the ground swath width. Of course, the beam

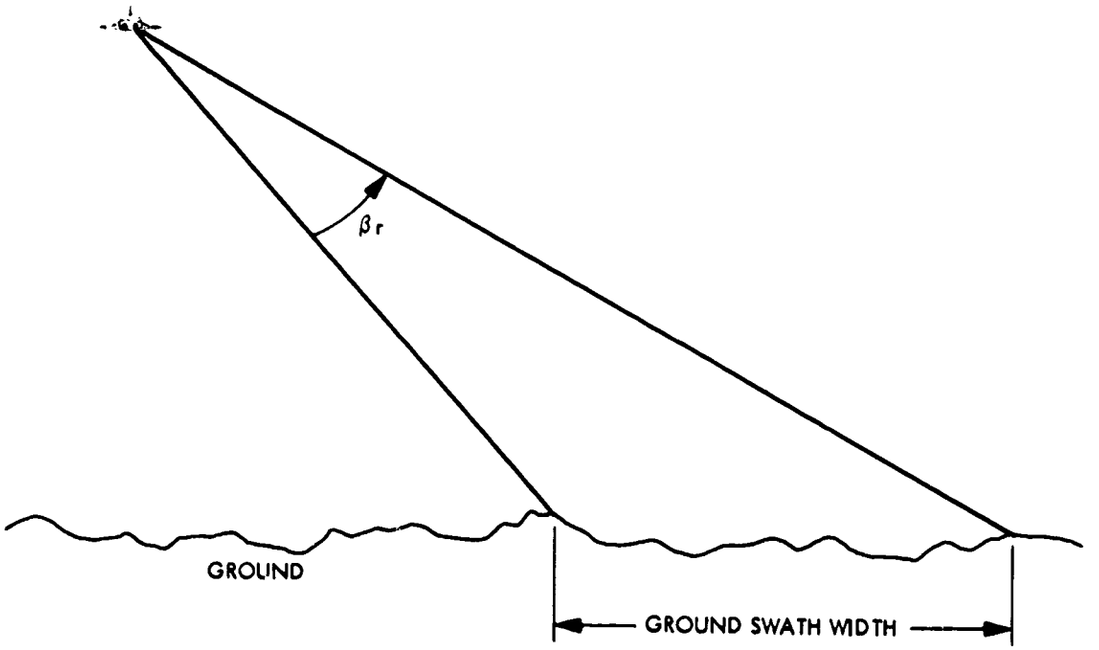


Fig. 1. Radar Beam as Viewed from Behind the Airplane

is composed not of light, but of lower frequency radio waves, and the "flashlight" is really a radar antenna. Figure 2 shows the beam as viewed from directly above the airplane. Notice that we have drawn the beam considerably narrower in Fig. 2. The small angle β_a is the azimuth beamwidth. With the proper design of a radar antenna, it is indeed possible to create a beam with distinct range and azimuth beamwidths.

Now let us suppose the radar is turned on for a very short time, and then off again. A brief burst of radio energy is emitted, which we may picture as the shaded band in Fig. 3. (Note incidentally in Fig. 3 that ground range G is related to slant range R by

$$\frac{G}{R} = \sin \gamma_e \quad)$$

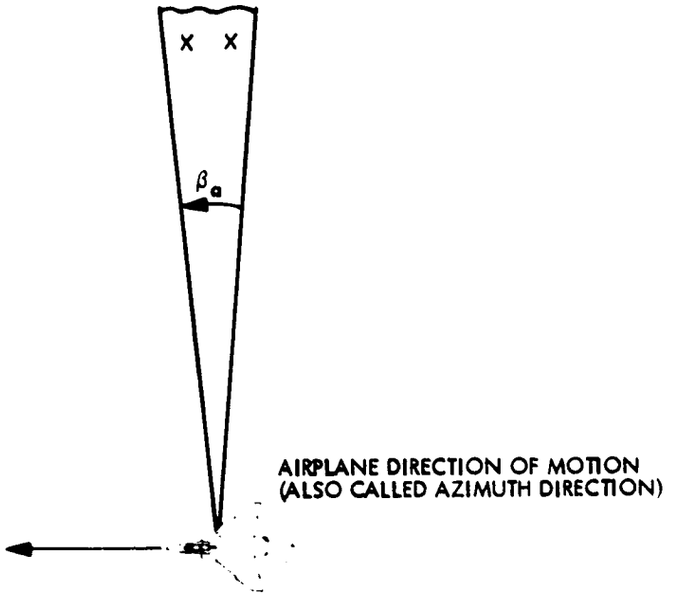


Fig. 2. Radar Beam as Viewed from Directly Above the Airplane, Looking Straight Down at the Ground

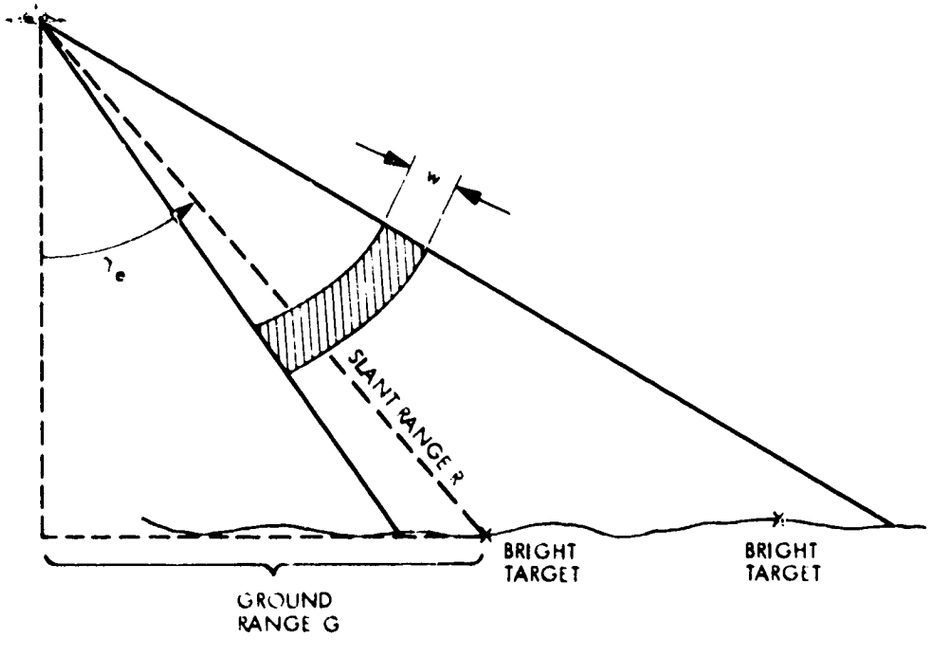


Fig. 3. Burst of Radar Illumination Viewed from Behind the Aircraft

where γ_c is the angle of elevation of the target from the nadir. After reflection from the ground, the same shaded band is shown returning to the aircraft in Fig. 4. Note that only the two designated areas (X) on the ground have enough reflectivity to return an appreciable echo, and that the two echoes will arrive back at the aircraft at different times, because they must travel different distances. Suppose we have a receiver which samples the returning echo as a function of time and records its instantaneous intensity. Then the sampled intensity will appear as in Fig. 5, or as in Fig. 6. Thus we see that by sampling the echo received from a single burst of radar energy, we have "painted" a one-dimensional strip of imagery perpendicular to the aircraft flight path. (This "strip" will henceforth be called a "range line" in this

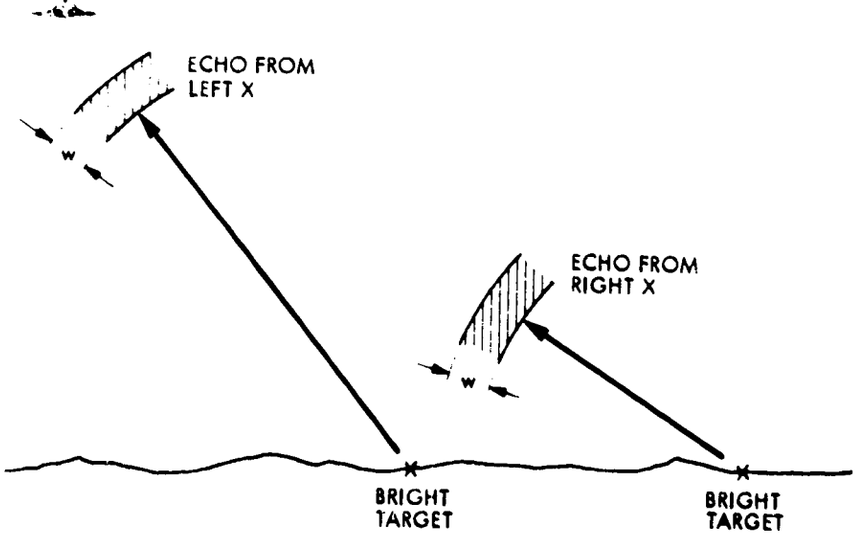


Fig. 4. Returning Echo

report, since it is oriented along the range direction, i.e., perpendicular to the flight path. The term range line will also be used later for any strip of unprocessed or processed radar data which is oriented along the range axis. The opposite notion is an azimuth line, which is a slice of data oriented in a direction parallel to the flight path.) The paintbrush moves from almost beneath

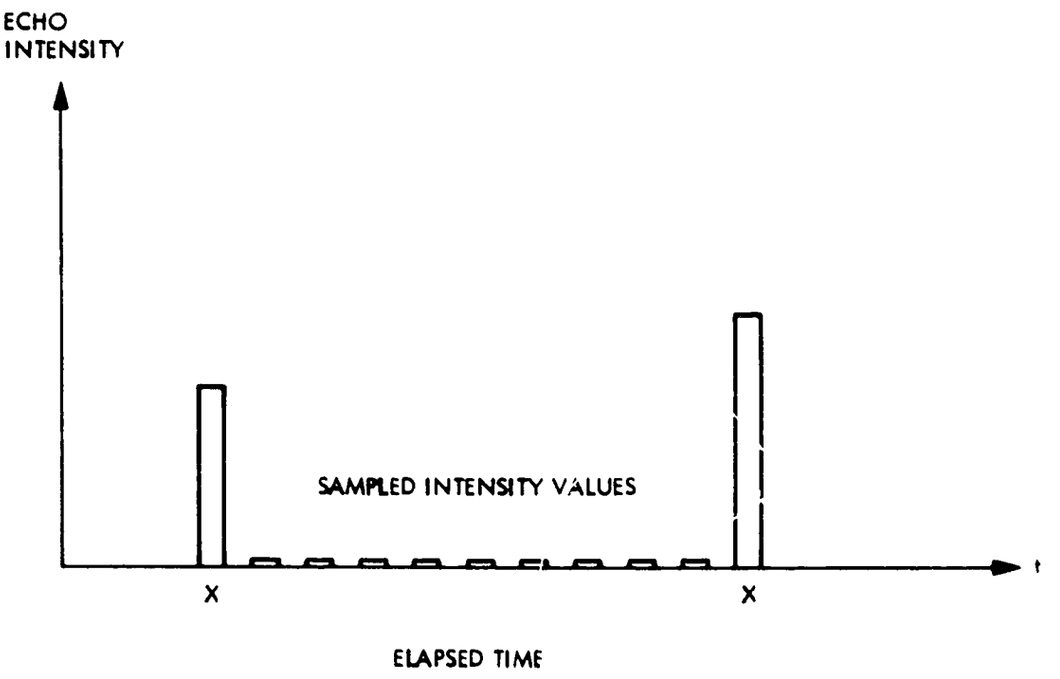


Fig. 5. Sampled Plot of Received Echo Intensity Versus Time

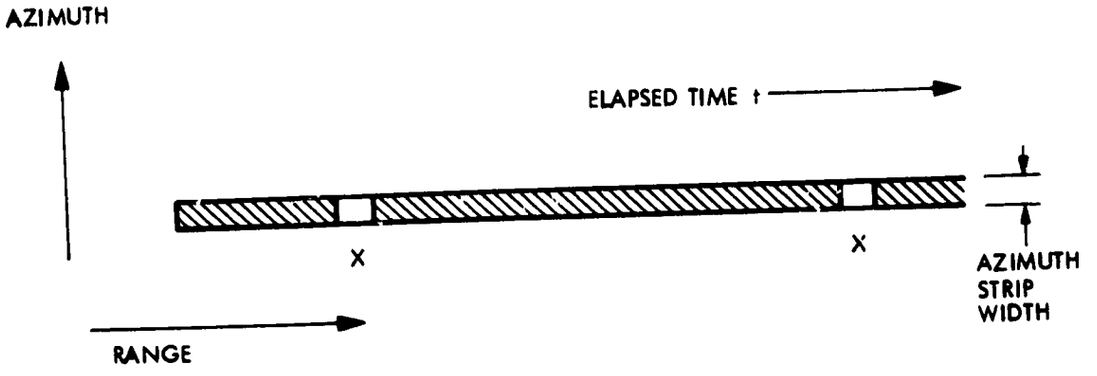


Fig. 6. Alternate Echo Display

the aircraft to far out to the side; the "movement" of the paintbrush arises from the different arrival times of the echo from different portions of the swath. Note in Fig. 6 that if the two reflectors (X) were closer together than the width w of the pulse in Figs. 3 and 4, the two targets would blur together. Thus it is imperative to transmit a very short pulse.

Now we have a strip, as in Fig. 6, but how do we produce a two-dimensional image? The answer is easy: just keep painting strips next to each other. We do this by turning on the radar again, when the aircraft has moved over, to produce the neighboring strip. Note that the "width" of our strip in azimuth (Fig. 6) is equal to the azimuth beamwidth along the ground. We cannot "see" any detail within the bright spots of Fig. 6 because the sampler can only record <u>one</u> intensity value at each instant of time. Thus the two X's in Fig. 2 cannot be separated. If we now continue, turning the radar on and off, we will "paint" a long succession of strips next to each other. The azimuth width of each strip is the ground azimuth beamwidth; and in the imagery we can only distinguish objects in one strip (i.e., one pulse) if they are further apart in range than the range pulsewidth w (actually 1/2 the pulsewidth - can you see why?). Thus our image resembles Fig. 7, if we allow four "strips."

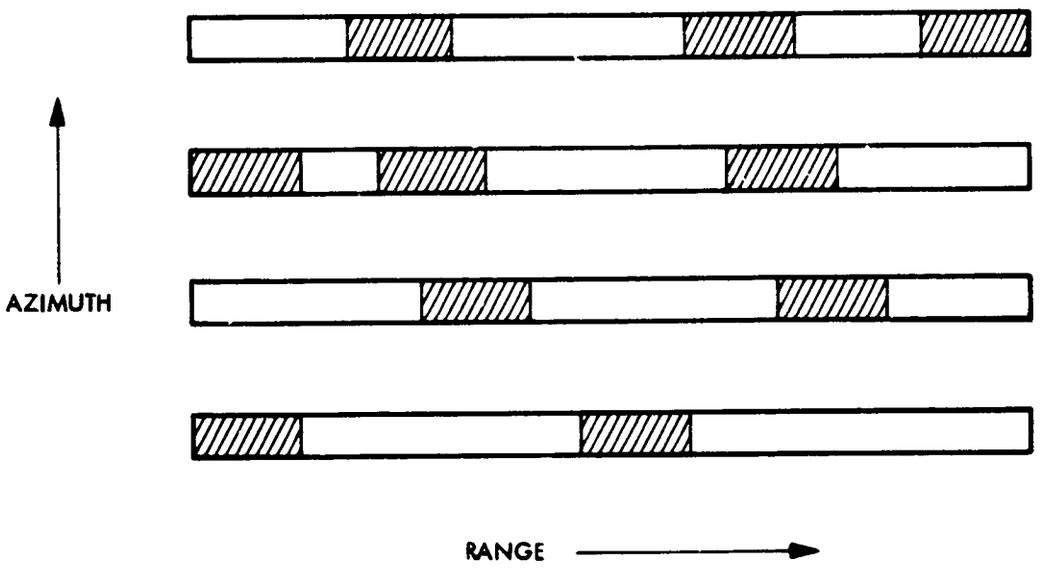


Fig. 7. The Beginnings of an Image

There are problems connected with this imaging approach. The first is that the pulse width in range (w in Fig. 3) must be kept short, as stated above. This requirement limits the energy transmitted per pulse and makes it difficult to detect weak, distant targets. The second problem is that, as already stated, we cannot distinguish the two targets (X) in Fig. 2. Although we can alleviate this problem somewhat by making β_a, the azimuth beamwidth, very narrow, we can only succeed partially, since very large antennas are required to produce narrow beams. Synthetic aperture radar (SAR) is a means of solving the second problem. In short, a SAR is simply a radar which uses the information from an entire sequence of pulses along the azimuth (and not just one) to produce a single painted strip of imagery as in Fig. 6. Thus, for example, all 4 pulses implied by Fig. 7 might be used in a SAR to provide information about the first

output strip in the figure. The result is refined resolution or ability to separate nearby targets along the azimuth.

The first problem (i.e., the problem of weak signal returns due to a short duration pulse) is solved by simply transmitting a longer pulse, often of the type known as the "linear FM chirp." The reader may well ask at this point, "Then how do we obtain the required range resolution?" The answer is that upon receiving the echo pulse we "compress" it to a shorter "pulse" by the technique of cross-correlation, to be discussed in section III-B. However, before discussing cross correlation we first require some background on electromagnetic waveforms. The reader is urged to follow the detour methodically; the results will be well worth the effort.

II. Electromagnetic Waveform Notation and Discussion

A. Phase and Frequency

Radio waves, like light waves, travel at a speed given by

$$c \sim 2.99776 \times 10^8 \text{ meters/sec.} \tag{1}$$

and are typically expressed as sinusoids:

$$s(t) = A \cos(\omega t) \tag{2}$$

Here

t = time, a real number

A = amplitude factor, greater than 0

ω = radian frequency, a real number

The right hand term in parentheses, (ωt), is the argument of the sinusoid and is called the phase. If we put, for the sinusoid of Eq. (2),

$$\phi(t) = \omega t \tag{3}$$

then we have

$$\phi'(t) = \frac{d\phi}{dt} = \omega \tag{4}$$

Thus the radian frequency is the time derivative of the phase. In other words, frequency is the rate of change of the phase.

A more general waveform is given by

$$s(t) = A(t) \cos(\phi(t)) \tag{5}$$

where both A and ϕ are real functions of t. Again, ϕ is called the phase function and $\phi'(t)$ the radian frequency. The (positive) function A(t) is the amplitude function. If we change variable by forming a new function f(t)

$$\phi'(t) = 2\pi f(t) \tag{6}$$

then f is the frequency in Hertz, or cycles per second. Thus in the special case of Eq. (2), we have

$$\phi(t) = \omega t$$

$$\phi'(t) = \omega, \text{ a constant function of time.}$$

We have then

$$f(t) = \omega/2\pi = f, \text{ a constant} \tag{7}$$

so that

$$s(t) = A \cos(2\pi i t) \tag{8}$$

From (8) we see that s(t) has period τ, where

$$\tau = 1/f \tag{9}$$

Example

The following waveform is known as a linear FM waveform. ("Linear FM" stands for linear frequency modulation.) This form of signal, also called a "chirp," is frequently used in SAR's. Define, for $-T/2 \le t \le T/2$ and real a,

$$L(t) = \cos\left(\frac{at^2}{2}\right) \tag{10}$$

Then the phase is

$$\phi(t) = \frac{at^2}{2} \tag{11}$$

and the radian frequency is

$$\phi'(t) = at \tag{12}$$

Thus the frequency in Hertz is

$$f(t) = \frac{at}{2\pi} \tag{13}$$

The term at in Eq. (12) explains the "linear" terminology, for the frequency is a linear function of time t. If $a > 0$, then a plot of frequency versus time for f(t) looks like Fig. 8.

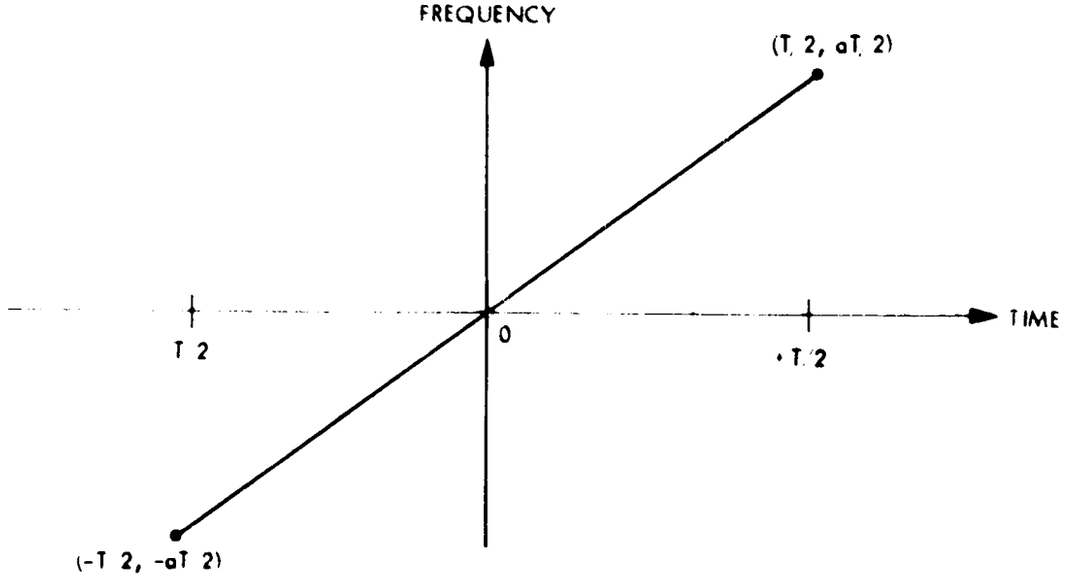


Fig. 8. Linear FM Frequency Plot

Note that at t = 0 we have

$$f(t) = \phi'(t) = 0 \tag{14}$$

while at $t = \pm\frac{T}{2}$ we have

$$\phi'(t) = \pm\frac{aT}{2} \tag{15}$$

Note also that we may differentiate the frequency in Eq. (12) and (13) to obtain the frequency rate:

$$\phi''(t) = a \text{ (radians/sec}^2\text{)} \tag{16}$$

$$\dot{f}(t) = f'(t) = \frac{a}{2\pi} \text{ (Hertz/sec)} \tag{17}$$

Eq. (16) gives the slope of the line in Fig. 8.

B. Interference and Antenna Beamwidth

Suppose we have two waveforms of unit amplitude and unit frequency in Hertz:

$$s_1(t) = \cos(2\pi t) \tag{18}$$

$$s_2(t) = \cos\left(2\pi\left(t - \frac{1}{2}\right)\right) \tag{19}$$

See Fig. 9. We see that s_1 and s_2 differ only in phase, but this is enough to cause total cancellation. Thus, if we put

$$s(t) = s_1(t) + s_2(t) \tag{20}$$

we have

$$s(t) \quad 0 \tag{21}$$

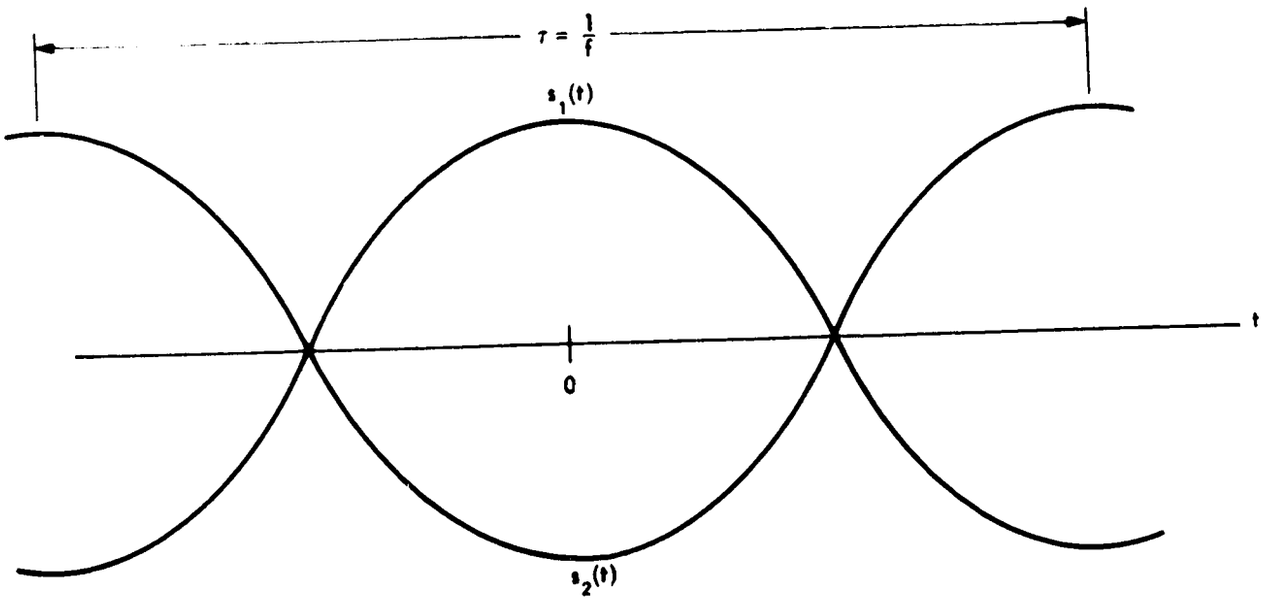


Fig. 9. Interference

We note from Eqs. (18) and (19) that Eq. (21) holds when the offset between the signals is one-half of a period. If we regard a wave as a physical structure, then we may define a wavelength as the physical distance corresponding to "the smallest period". We usually denote this quantity by λ, and observe that the "physical wavelength" λ and the "time wavelength" τ (also called the "period") are related by

$$\lambda = c\tau \tag{22}$$

where c, the velocity of light, is given by (1). Eq. (22) is only a version of the familiar relation

$$\text{distance} = (\text{rate}) \times (\text{time})$$

and if we recall that period is inversely related to frequency by

$$\tau = \frac{1}{f} \tag{9}$$

we obtain from Eq. (22)

$$\lambda = \frac{c}{f}$$

$$\lambda f = c \tag{23}$$

In words, Eq. (23) says the wavelength times the number of waves passing per second is equal to the distance light travels in a second.

If in Fig. 9 we regard s_1 and s_2 as signals possessing the same arbitrary wavelength λ, then we see that in general two signals will interfere destructively ("cancel") when they are offset by $\lambda/2$.

Antenna beamwidth is related to the above considerations. In Fig. 10 we show a point target returning a radar echo (of wavelength λ) to a radar antenna of length L. The target is offset from the perpendicular axis ("boresight") of the antenna by θ radians, and is sufficiently far away that it produces a nearly straight wavefront of parallel waves at the antenna. In the figure we suppose that the angle θ is such that the wavefront is λ units further from the right edge of the antenna than from the left, (i.e., the offset between waves (1) and (3) is λ, a wavelength). The offset between waves (1) and (2) is $\lambda/2$, and they cancel out. Looking, for example, at (1') and (2') we see that every

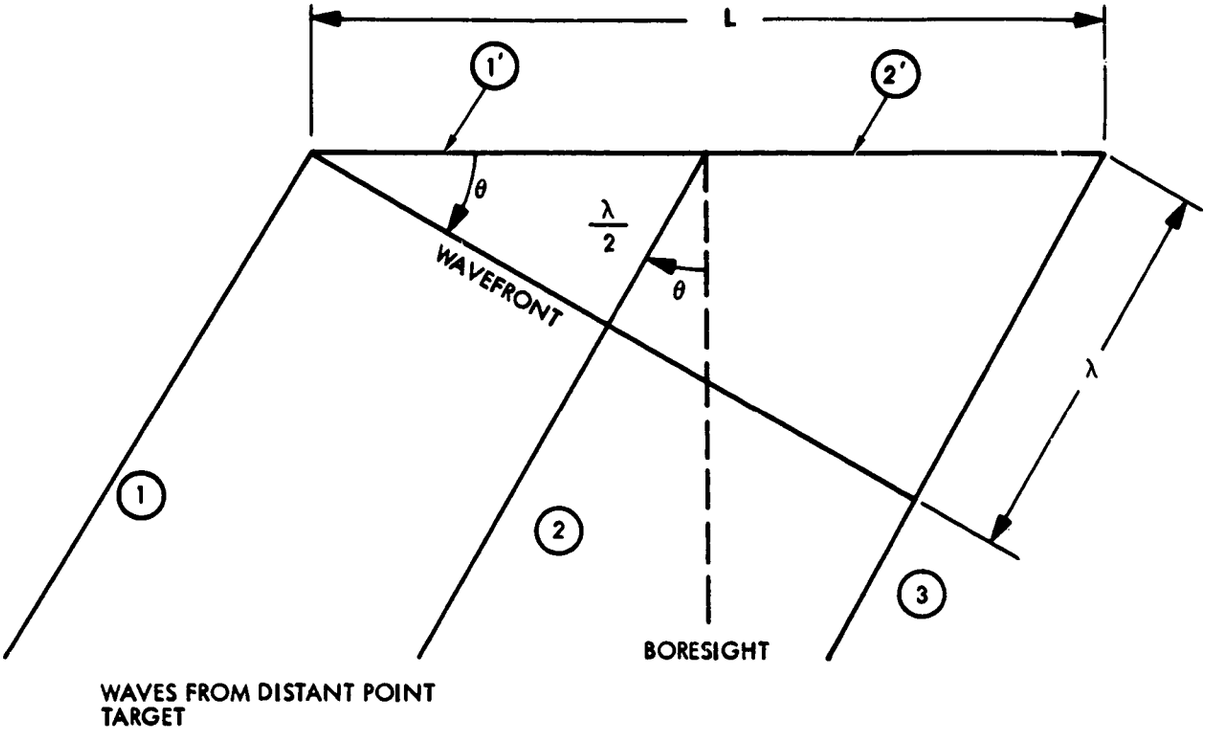


Fig. 10. Destructive Interference in a Radar Antenna

wave in the left half of the antenna has a corresponding cancelling wave in the right half, and thus we expect the net signal to be zero. This is indeed the case, and if we plot the strength of received signal from a point target as a function of offset angle, we obtain a curve resembling that in Fig. 11, where θ satisfies (see Fig. 10)

$$\sin \theta \sim \frac{\lambda}{L} \tag{24}$$

Since generally in SAR antennas, λ is much smaller than L,

$$\lambda \ll L$$

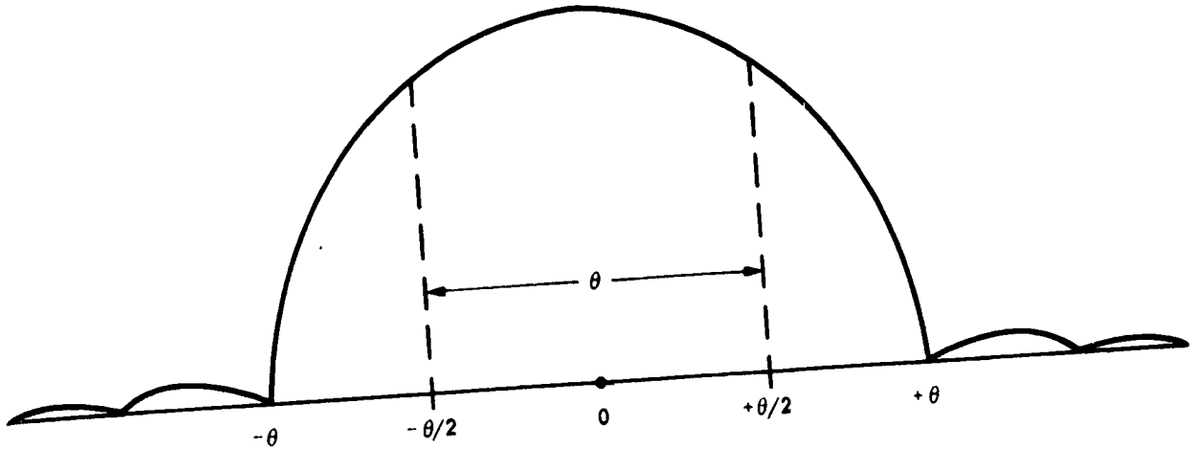


Fig. 11. Antenna Pattern

we have that θ is small, sin θ ≅ θ and hence Eq. (24) becomes

$$\theta \cong \frac{\lambda}{L} \tag{25}$$

(Generally the beamwidth θ is on the order of 1^{o} for spaceborne SAR systems, so (25) is a very good approximation.)

Thus the null-to-null beamwidth of the antenna is

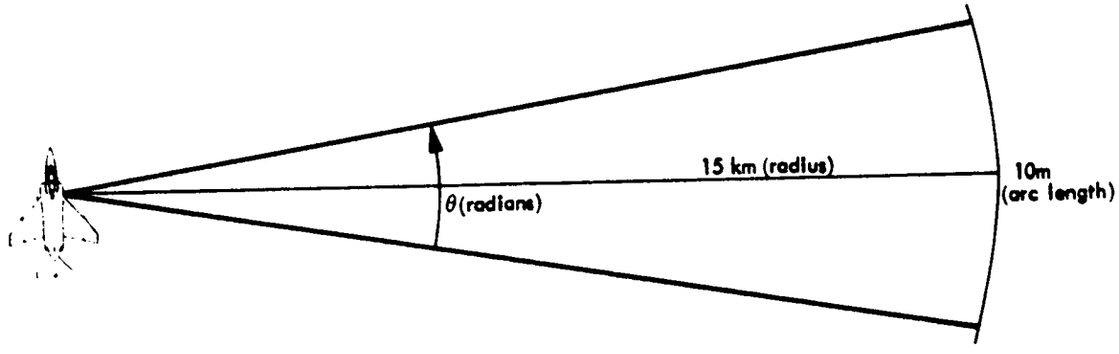
$$2\theta = \frac{2\lambda}{L}$$

However, the offset at which the return signal intensity drops to 1/2 of its maximum value is usually around ±θ/2; this offset is usually used to define "half-power" or "3 dB" beamwidth, which is thus given as 2(θ/2) = θ; hence the 3 dB beamwidth of a radar antenna is given by

$$\theta \sim \frac{\lambda}{L} \tag{26}$$

and we see that the longer the physical antenna, the smaller the beamwidth. Thus in Section I, where we desired a narrow azimuth beamwidth β_a, we see that a SAR antenna with long azimuth dimension was required.

How long? Well, if we wish to separate two targets 10 meters apart in azimuth, then our "flashlight beam" must be narrower than 10 meters on the ground along the azimuth. (See Fig. 2.) A typical radar wavelength λ might be 0.235 meters, and the distance between airplane and ground target might be 15 km. Using the relation arc length = (radians) × (radius),



we obtain

$$10\text{m} = \theta \times (15\text{ km})$$

so that

$$\theta = 0.00067 \text{ radians}$$

and from Eq. (26)

$$L = \frac{\lambda}{\theta}$$

$$L = 352.5\text{m}$$

Most airplanes are not equipped to carry an antenna this long! As we have already stated, SAR provides a solution to the problem; however, our background development must first be completed. In the meantime note that Eq. (26) implies that azimuth resolution δ_A is given by

$$\delta_A = \frac{R\lambda}{L} \tag{26'}$$

where R is the distance (slant range) between aircraft and target. (See Fig. 3.) The resolution of a SAR will be considerably better.

C. Mixing

Engineers alter the frequency of a sinusoidal signal by mixing, which consists merely of multiplying the given signal by another signal and then filtering. In this way a signal having the sum or difference frequency is produced. For our application we require the difference frequency.

As an illustrative example, consider two signals having frequencies of 100 Hz and 120 Hz. If these signals are multiplied together, we obtain a new signal consisting of two components, or "sidebands". The first component has frequency (120-100) or 20 Hz. The second component has frequency (120 + 100) or 220 Hz. A lowpass filter will isolate the 20 Hz component.

To understand how mixing works, we consider the following well known trigonometric identities:

$$\cos(a + b) = \cos a \cos b - \sin a \sin b \tag{27}$$

$$\cos(a - b) = \cos a \cos b + \sin a \sin b \tag{28}$$

$$\sin(a + b) = \sin a \cos b + \cos a \sin b \tag{29}$$

$$\sin(a - b) = \sin a \cos b - \cos a \sin b \tag{30}$$

Adding Eqs. (27) and (28) and dividing by 2 gives

$$\cos a \cos b = \frac{1}{2}\left(\cos(a + b) + \cos(a - b)\right) \tag{31}$$

Subtracting Eq. (27) from Eq. (28) and dividing by 2 gives

$$\sin a \sin b = \frac{1}{2}\left(\cos(a - b) - \cos(a + b)\right) \tag{32}$$

Adding Eqs. (29) and (30) and dividing by 2 gives

$$\sin a \cos b = \frac{1}{2}\left(\sin(a + b) + \sin(a - b)\right) \tag{33}$$

From Eqs. (31) to (33) we see that multiplication of sinusoids translates to the summing and differencing of their arguments.

To see how Eqs. (31) to (33) may be used, suppose a radar is receiving an echo of the form

$$e(t) = \cos(\omega_1 t) \tag{34}$$

where

ω_1 = constant frequency

t = time

Let us split the return signal into two channels, as in Fig. 12. We multiply the top channel signal by $\cos(\omega_1 t)$ and the bottom channel signal by $\sin(\omega_1 t)$, where this time ω_1 is a constant frequency of our own choosing.

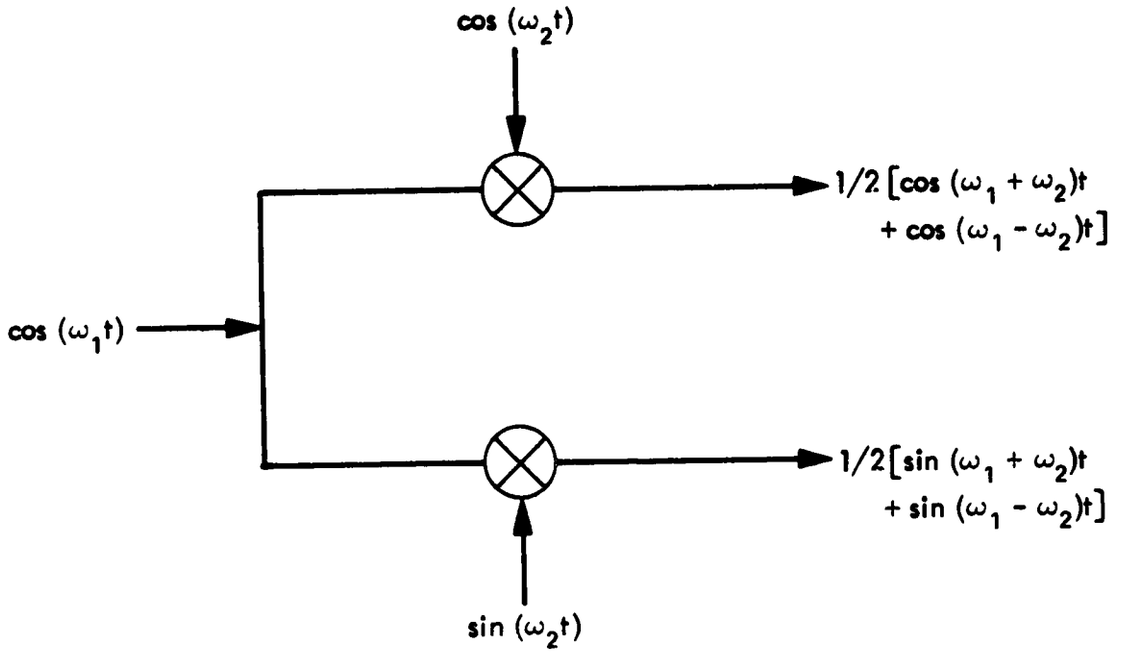


Fig. 12. The First Stage of Mixing

As is seen from Eq. (31) and Eq. (33), each channel output is a sum of sinusoids of high and low frequencies:

$$\frac{1}{2}\cos\left[(\omega_1 + \omega_2)t\right] + \frac{1}{2}\cos\left[(\omega_1 - \omega_2)t\right] \tag{35}$$

for the top channel, and

$$\frac{1}{2}\sin\left[(\omega_1 + \omega_2)t\right] + \frac{1}{2}\sin\left[(\omega_1 - \omega_2)t\right] \tag{36}$$

for the bottom channel.

Let us suppose that $(\omega_1 - \omega_2)$ is the low (positive) frequency. Then if we lowpass filter each channel we obtain the output in Fig. 13. Normally, the term mixing refers to the use of the top channel, but if we use both channels as in Figs. 12 and 13, the net result is as shown in Fig. 14, and consists of two

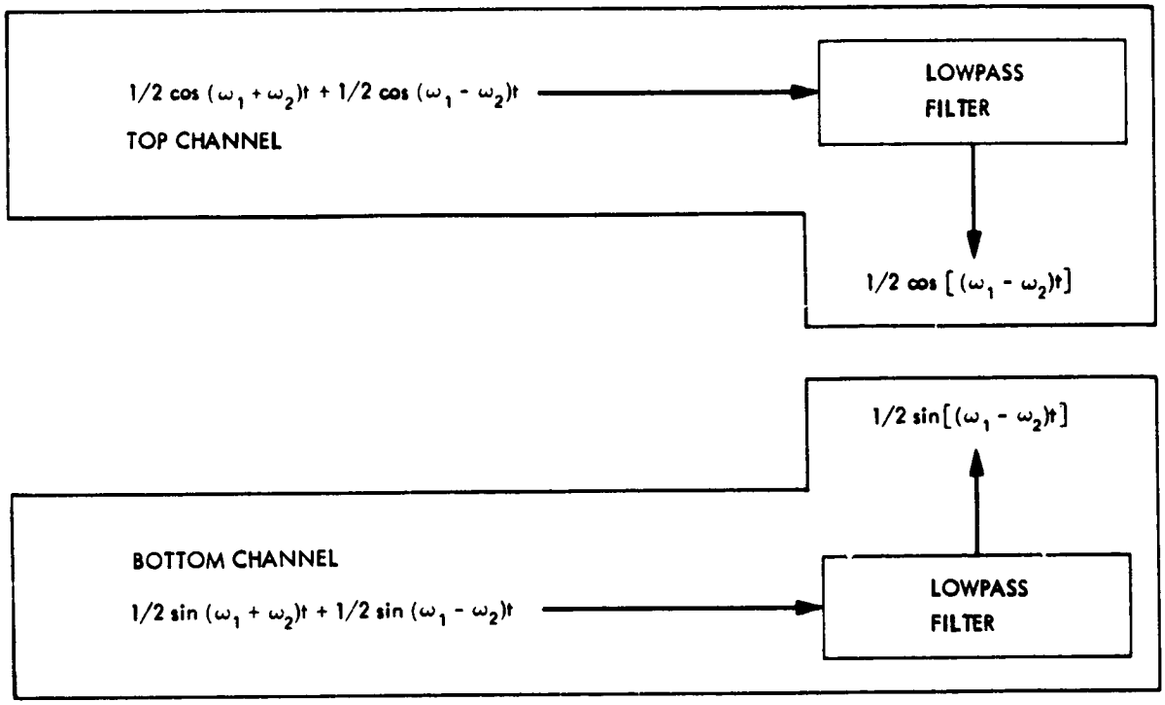


Fig. 13. The Second Stage of Mixing - Lowpass Filtering

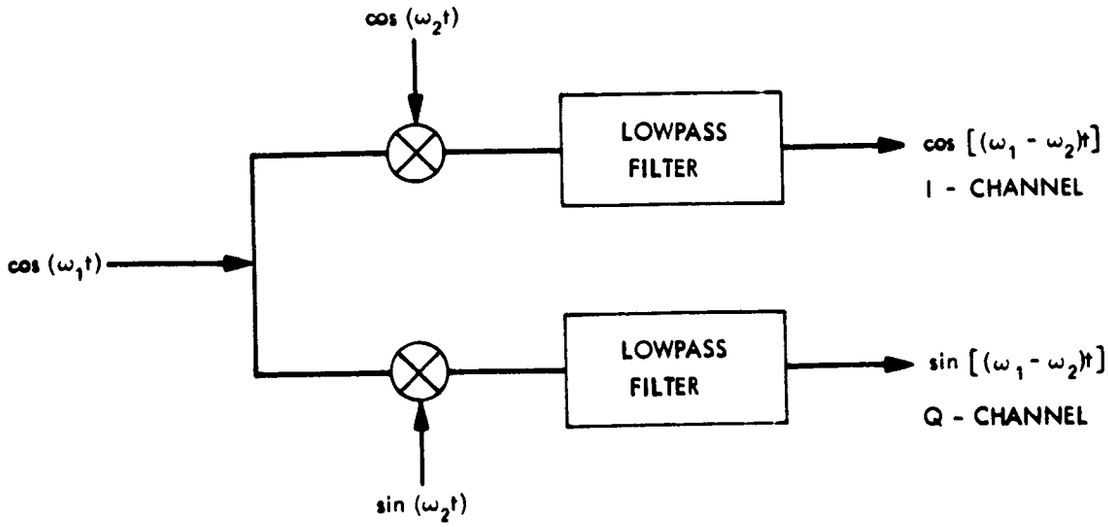


Fig. 14. Offset to I, Q Conversion, Difference Frequency

separate outputs, known as in-phase (I) and quadrature (Q). Note that in Fig. 14 the scaling (amplitude) constants have been removed.

The ordered pair of real numbers

$$\left\{\cos\left[(\omega_1 - \omega_2)t\right],\ \sin\left[(\omega_1 - \omega_2)t\right]\right\}$$

is usually abbreviated by mathematicians as[1]

$$e^{j\left[(\omega_1 - \omega_2)t\right]}$$

where

$j = \sqrt{-1}$

e = the exponential function

Thus the net result of mixing with a sine and cosine channel is shown in Fig. 15 as a complex output signal with a new frequency.

Note that it would be just as easy to use a highpass filter instead of a lowpass filter; in this case we would obtain the sum frequency instead of the difference frequency and Figs. 14 and 15 would be replaced by Figs. 16 and 17.

We are now ready to see how mixing is used to preserve coherency (i.e., azimuth phase information) in a SAR system. For the transmitted pulse we use a linear FM chirp modulating a high frequency carrier wave.

---

[1] In general, for real θ, we put

$$e^{j\theta} = \cos\theta + j\sin\theta$$

which is formally equivalent to a cartesian pair

$$e^{j\theta} = \{\cos\theta,\ \sin\theta\}$$

provided we operate on the components properly in subsequent algebraic manipulations. Note that the (+) sign in $\cos\theta + j\sin\theta$ is purely a formality; real and imaginary numbers "do not blend." We have, for example, $3 + 4 = 7$ but $3 + 4j$ stays $3 + 4j$, and is equivalent to $(3, 4)$. The equivalence is valid provided we define multiplication by $(a, b)(c, d) = (ac - bd,\ ad + bc)$.

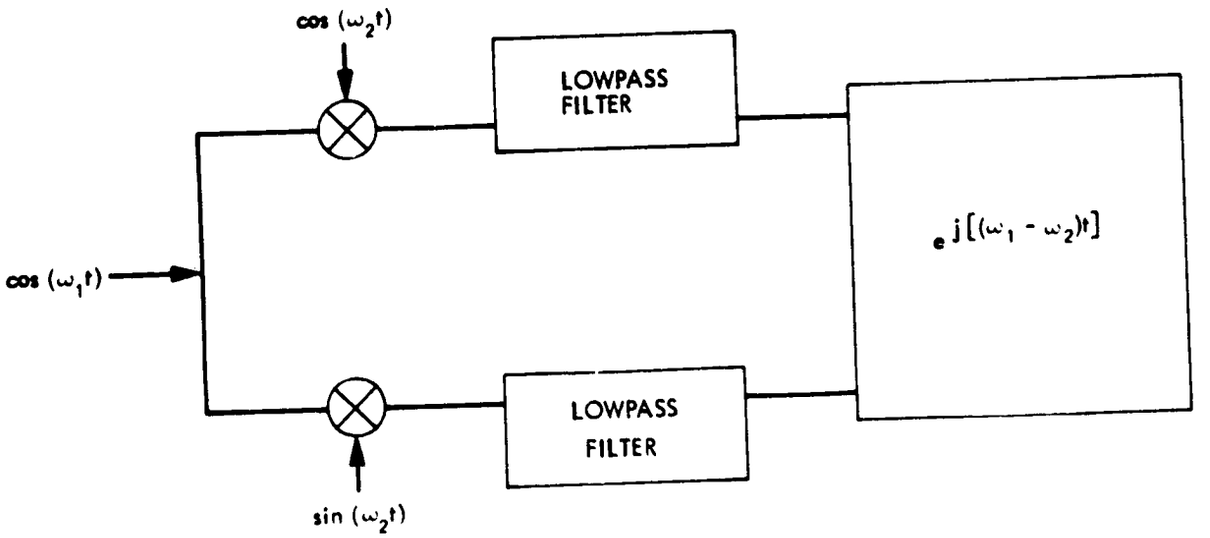


Fig. 15. Mixing Summary, Difference Frequency

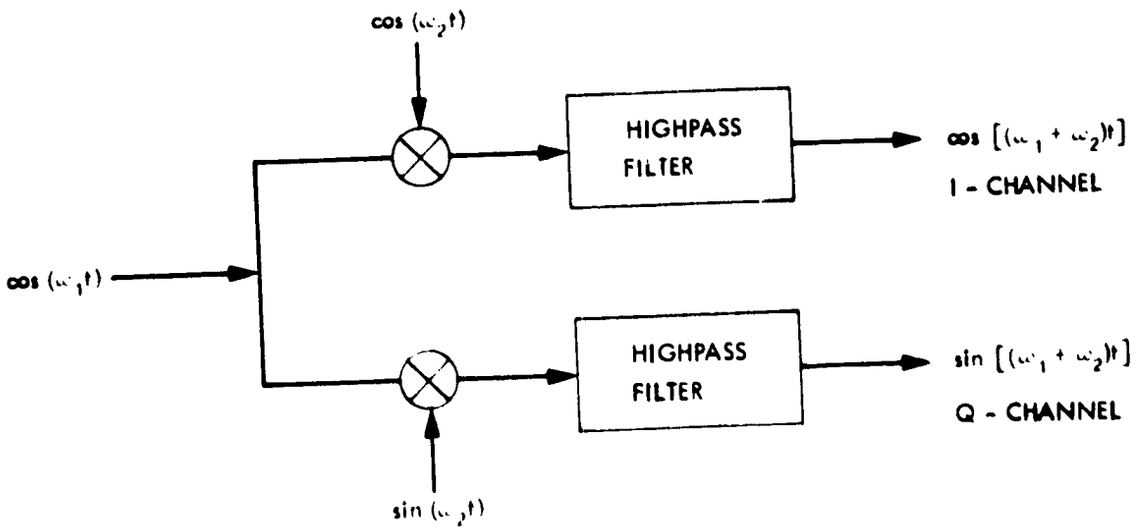


Fig. 16. Offset to I, Q Conversion Sum Frequency

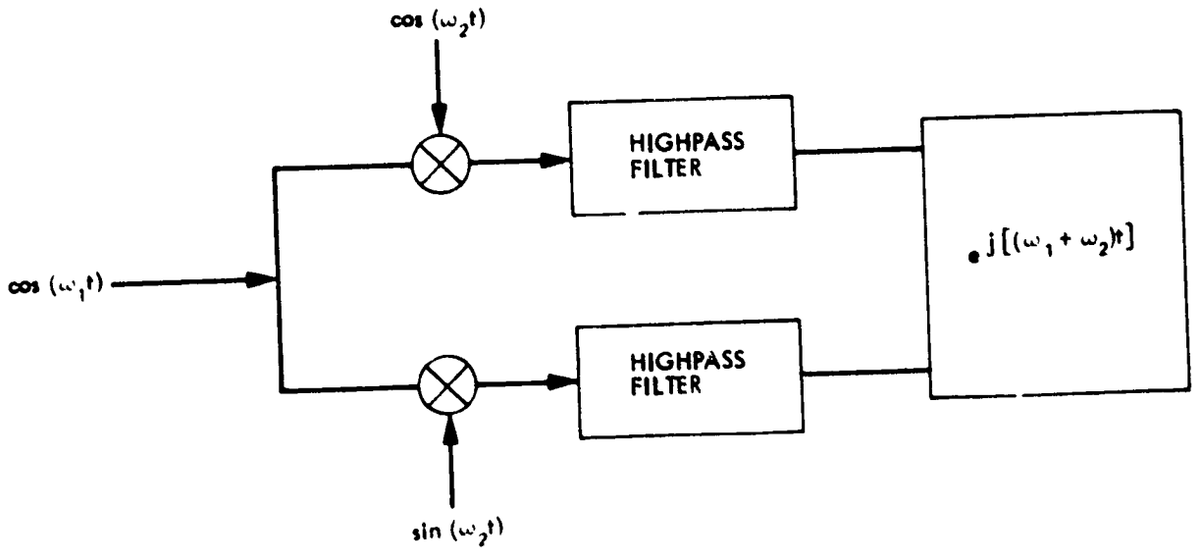


Fig. 17. Mixing Summary, Sum Frequency

Suppose that when time t satisfies $-\frac{T}{2} \leq t \leq \frac{T}{2}$, the radar transmits a pulse defined by

$$s(t) = \cos\left(\frac{at^2}{2} + \omega_0 t\right). \tag{38}$$

(Here ω_0 is considered a constant.) Outside the interval $-\frac{T}{2} \leq t \leq \frac{T}{2}$ the radar is silent, and merely listens for a return. (Note: We may assume the signal in Eq. (38) is generated either by direct modulation on a constant frequency carrier wave cos (ω_0t), or by mixing the signals $\cos\left(\frac{at^2}{2}\right)$ and cos (ω_0t); the interested reader may carry through the details of the mixing operation.) Observe also that Eq. (38) defines a linear FM waveform. A plot of its frequency versus time is shown in Fig. 18; it is like Fig. 8 but displaced vertically.

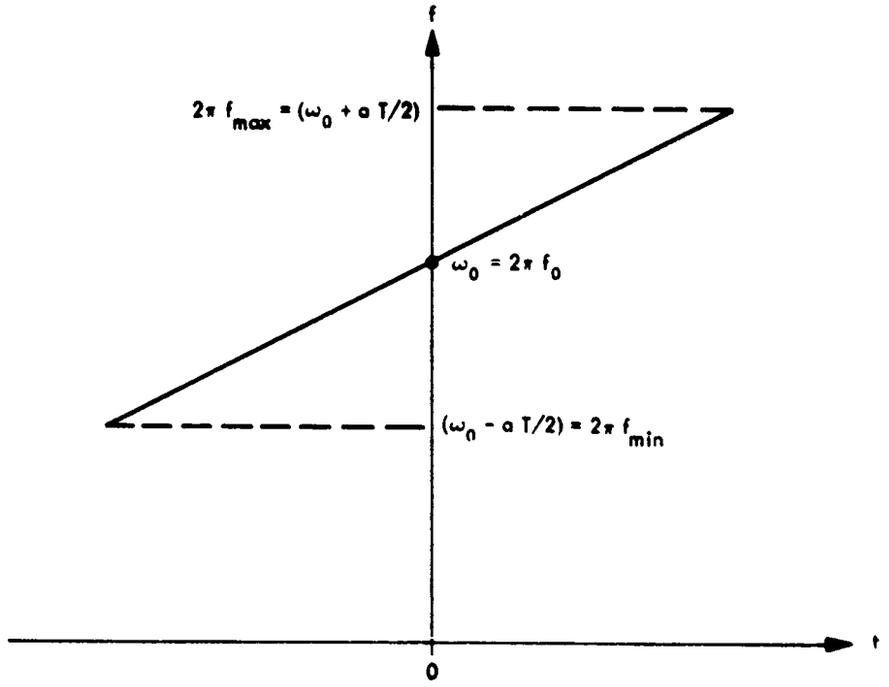


Fig. 18. Linear FM on a Carrier - The Frequency Plot

By way of example, it may be helpful to know that if we define the carrier frequency f_0 in Hz by

$$\omega_0 = 2\pi f_0$$

then a typical value of f_0 for a real radar might be 1275 MHz. A typical value of f_{max} might be 1284 MHz; f_{min} might be 1266 MHz. The signal bandwidth would then be $f_{max} - f_{min} = 1284 - 1266 = 18$ MHz.

To continue, we require that even when Eq. (38) is not being transmitted, a stable local oscillator (STALO) continuously generates a carrier waveform of frequency $\omega_0 = 2\pi f_0$. This waveform is generated in two channels, producing signals $C_I(t)$ and $C_Q(t)$, where

$$C_I(t) = \cos(\omega_0 t) \tag{39a}$$

$$C_Q(t) = \sin(\omega_0 t) \tag{39b}$$

for all time t.

In a SAR transmitter, the STALO must be extremely stable because the carrier wave is mixed with the returning radar echo. This mixing provides the critical "azimuth phase factor" Φ_a in the following way. (The use of the azimuth phase factor is given in section IV.)

Suppose the terrain is dark except for one bright point target PT on the ground at a distance R meters from the radar transmitter. (This distance is called slant range, see Fig. 3.) It takes the signal in Eq. (38) a certain time δt to traverse the round-trip distance 2R to PT and back. (We make the simplifying assumption that the aircraft does not move between transmission and subsequent echo reception; thus the radar signal travels R units in both directions.) Since, at the speed of light c, this time of traversal is given by the relation

$$c(\delta t) = 2R$$

we have

$$\delta t = \frac{2R}{c} \tag{40}$$

Thus the leading edge of the pulse, which returns δt seconds after its transmission, is received at time $-\frac{T}{2} + \delta t = -\frac{T}{2} + \frac{2R}{c}$. Similarly the trailing edge of the pulse returns at $+\frac{T}{2} + \frac{2R}{c}$, after which time the radar receives no echo. Thus in the interval

$$-\frac{T}{2} + \frac{2R}{c} < t < \frac{T}{2} + \frac{2R}{c} \tag{41}$$

the radar receives the echo e(t), which is the transmitted signal delayed by δt seconds:

$$e(t) = As\left(t - \frac{2R}{c}\right) = A\cos\left(\omega_0\left[t - \frac{2R}{c}\right] + \frac{a}{2}\left[t - \frac{2R}{c}\right]^2\right) \tag{42}$$

(The letter A denotes a positive amplitude factor proportional to the target's reflectivity, or "cross section.") (Note again that the correctness of (42) depends upon a simplifying assumption: we assume that R does not vary with t during reception of a single pulse echo. Without this simplifying assumption, the range signal mixing would be complicated by a "doppler offset frequency" which is normally neglected in SAR processing.) The returning echo is divided into two channels (I and Q) for separate mixing with the carrier components $C_I(t)$ and $C_Q(t)$. The results are indicated in Fig. 19.

As before, we represent the pair of outputs I and Q from Fig. 19 as a complex exponential $e_r(t)$; (the subscript r denotes range signal, since the

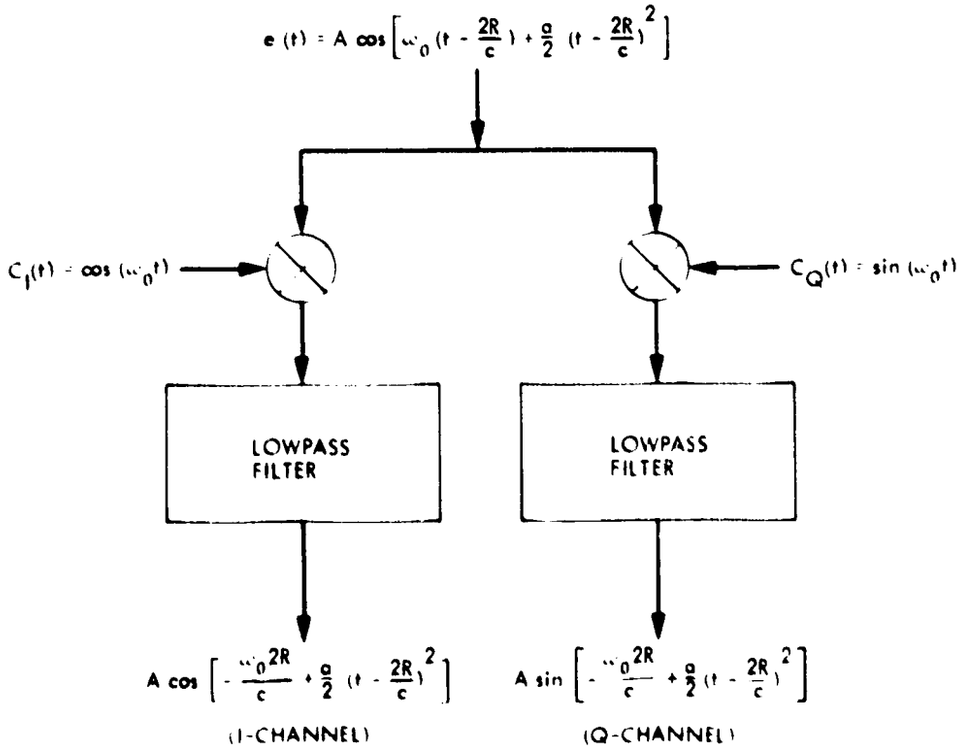


Fig. 19. Result of Mixing the Returning Radar Echo With the STALO Signals C_I and C_Q

independent variable t is "oriented" along the "range" direction. Note that e_r is the return from a single pulse).

$$e_r(t) = A \exp\left\{j\left(-\frac{\omega_0 2R}{c} + \frac{a}{2}\left(t - \frac{2R}{c}\right)^2\right)\right\} \tag{43}$$

Using the relation $e^{a+b} = e^a e^b$ we may rewrite Eq. (43) as

$$e_r(t) = A \exp\left\{j\left(-\frac{2\omega_0 R}{c}\right)\right\} \exp\left\{j\frac{a}{2}\left(t - \frac{2R}{c}\right)^2\right\} \tag{44}$$

where

$$\Phi_a = \exp\left\{j\left(-\frac{2\omega_0 R}{c}\right)\right\} \tag{45}$$

is the azimuth phase factor and

$$\phi_a = -\frac{2\omega_0 R}{c} \tag{46}$$

is its phase. We remind the reader that Eq. (44) represents the demodulated echo only on the interval

$$-\frac{T}{2} + \frac{2R}{c} \leq t \leq \frac{T}{2} + \frac{2R}{c}$$

Outside this interval $e_r(t)$ is identically zero.

We make first a few observations about Eqs. (44) to (46). First, the azimuth phase factor is independent of time t of reception. As long as the receiver is receiving signal from the point target PT, the phase factor Eq. (45) will appear as a constant term with phase given by Eq. (46). Next, the azimuth

phase in Eq. (46) may be rewritten in terms of wavelength λ. To see this, note that

$$\phi_a = -\frac{2\omega_0 R}{c} = -2\times\frac{2\pi f_0 R}{c}$$

and so

$$\phi_a = -\frac{4\pi R}{\lambda} \tag{47}$$

where we have used the relation $\lambda f_0 = c$ (cf. (23)). Finally, the factor

$$\Phi_r = \exp\left\{j\frac{a}{2}\left(t-\frac{2R}{c}\right)^2\right\} \tag{48}$$

is known as the range phase factor and we denote its phase by ϕ_r:

$$\phi_r = \frac{a}{2}\left(t-\frac{2R}{c}\right)^2 \tag{49}$$

Note that ϕ_r is of the form

$$\phi_r = \frac{a}{2}y^2 \tag{50}$$

where

$$y = t - \frac{2R}{c} = t - \beta \tag{51}$$

and

$$\beta = \frac{2R}{c} \tag{52}$$

is a function only of target separation R from the radar.

D. Sampling

Up to this point we have assumed that all signals have been analog. Actually, sampling can take place at a variety of points within the system; we choose to introduce it right here for simplicity. Thus we assume that the analog signal Eq. (44) is sampled as it is produced.

The subject of sampling is well treated in digital signal processing texts. The main issue is sampling frequently enough so that all the information in the analog signal is retained. Such ideal sampling is possible only for certain types of signals. One such type is a signal having a continuous Fourier transform of finite extent along the frequency axis; the interested reader may find the corresponding "Nyquist sampling theorem" in [1] pp. 26-30. Another such signal is one whose Fourier series has only a finite number of terms; a version of the Nyquist theorem for this second type of signal is found in [2], pp. 25-28.

The gist of these theorems is that we can keep most of the analog information of Eq. (44) if we sample it at sampling frequency f_s, where

$$f_s = \frac{aT}{2\pi} \tag{53}$$

Referring to Fig. 18 we see that f_s is the vertical extent of the frequency graph, expressed in Hertz. Thus we are saying we must sample the complex echo $e_r(t)$ at a frequency equal to its "bandwidth." For further discussion of sampling theory and alternate definitions of bandwidth, the reader is referred to [1], [2], [3].

The sampling period Δt is defined as the time interval between adjacent samples; it is the reciprocal of f_s; hence

$$\Delta t = \frac{2\pi}{aT} \tag{54}$$

(Normally we sample at a slightly higher frequency than the bandwidth, and Δt is correspondingly smaller.)

We now reexamine the situation thus far. We have followed the radar signal from the transmitter to a point target and back again, and seen how the signal is mixed down (demodulated) to baseband. ("Baseband" implies the signal is centered about zero frequency; the carrier frequency ω_0 was removed by the mixing of (44)). Next the signal was digitized (sampled) at intervals of Δt seconds. All of this manipulation is concerned with the return from a <u>single transmission</u> or <u>pulse</u>. If we knew the terrain had only one point target at range R, we could begin sampling $e_r(t)$ at $-\frac{T}{2} + \frac{2R}{c}$ and stop sampling at $+\frac{T}{2} + \frac{2R}{c}$. The sampling window would then be T seconds wide, the width of the range chirp. However, targets at a variety of slant ranges will be returning radar echoes; thus we must increase the size of our sampling window to accommodate returning pulse echoes from across the entire ground swath illuminated by the antenna in the cross-track direction. Thus we may assume we start sampling at time T_1. We take a new sample every Δt seconds until time T_2, at which time we turn off the sampler. We assume $T_1 < -\frac{T}{2} + \frac{2R}{c}$ and $+\frac{T}{2} + \frac{2R}{c} < T_2$. The time interval over which we sample is thus $\Delta T = T_2 - T_1$; the number of samples is N where

$$N = \frac{\Delta T}{\Delta t} \tag{55}$$

Time T_1 is the time of reception of the leading edge of the echo from the inmost point on the swath, and T_2 is the time of reception of the tail of the echo from the outermost point on the swath.

If PT at slant range R is the only reflecting target, then our array of N samples looks like Fig. 20.

The first 0 in Fig. 20 corresponds to $t = T_1$. The last 0 corresponds to $t = T_2$. There are N entries; hence we may denote them as a vector $v = \{v(1), v(2), \ldots, v(N)\}$. Let us suppose the kth element of v is the beginning of the echo from PT: $v(k) = e_r\left(-\frac{T}{2} + \frac{2R}{c}\right)$; we also choose ℓ so that $v(k+\ell) = e_r\left(\frac{T}{2} + \frac{2R}{c}\right)$ corresponds to the end of the echo return from PT. Then we have $v(n) = 0$ for $1 \le n < k$ and $v(n) = 0$ for $k + \ell < n \le N$, provided PT is the only reflecting target. Note that $v(k), v(k+1), \ldots, v(k+\ell)$ all have signal data from PT. How can we "compress" $v(k), \ldots, v(k+\ell)$ into one or just a few samples so as to obtain fine range resolution, as promised earlier This compression, it turns out, is the function of the cross-correlation process, and is discussed in the next section.

$$\underbrace{[0, 0, \ldots, 0}_{(k-1)\text{ ELEMENTS}}, \overbrace{e_r\left(\tfrac{-T}{2} + \tfrac{2R}{c}\right), e_r\left(\tfrac{-T}{2} + \tfrac{2R}{c} + \Delta t\right), e_r\left(\tfrac{-T}{2} + \tfrac{2R}{c} + 2\Delta t\right), \ldots, e_r\left(\tfrac{T}{2} + \tfrac{2R}{c}\right)}^{\ell+1\text{ ELEMENTS}}, \underbrace{0, 0, \ldots, 0]}_{N\text{ ELEMENTS}}$$

NOTE: IN THE ABOVE VECTOR, $e_r\left(\frac{T}{2} + \frac{2R}{c}\right) = e_r\left(\frac{-T}{2} + \frac{2R}{c} + \ell\Delta t\right)$

Fig. 20. Array of Samples from a Single Pulse Echo from a Single Point Target

III. Vectors, Matrices, Cross-Correlation and the Range Processing

In this section we develop the ideas of vectors, matrices and cross-correlation. While the reader with sufficient background may wish to skip over the explanation on vectors and matrices, the section on cross-correlation should be read carefully. Facts about cross-correlation that are especially relevant to SAR processing will be emphasized.

A. Vectors and Matrices

A vector, as described in the previous section, is simply a list of numbers in a particular order. Thus (1, 0, -3) is a vector having three components. It is distinct from the vector (-3, 0, 1). The vector v from the previous section has N complex components:

$$v = \big(v(1),\ v(2),\ \ldots,\ v(N)\big) \tag{56}$$

It is customary to denote vectors as in Eq. (56); the parentheses enclose the whole vector and separate parentheses are used to count the elements of the vector. Thus v(7) is the 7th element of v.

Vectors may be added if they have the same length. Thus if $u = \big(u(1),\ \ldots,\ u(N)\big)$ and $v = \big(v(1),\ \ldots,\ v(N)\big)$ then we define the vector $u + v$ by

$$(u + v)(k) = u(k) + v(k) \quad \text{for } k = 1, 2, \ldots, N$$

Just as naturally we may multiply a vector by a scalar (a scalar is just a number). If α is a scalar and $u = \big(u(1),\ \ldots,\ u(N)\big)$ is a vector then the vector αu is defined by

$$(\alpha u)(k) = \alpha \times u(k) \quad \text{for } k = 1, 2, \ldots, N .$$

A matrix or array is just a list of vectors of the same length. Typically the elements of a matrix are identified by affixing row and column subscripts; thus m_{ij} might be used to denote the element in the ith row and jth column of a matrix M. For example we might define

$$M = \begin{pmatrix} 1 & 3 \\ -1 & 4 \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}$$

so that $m_{21} = -1$ and $m_{12} = 3$.

The reader will note that if the digitized echo return from one radar pulse forms a vector v of length N, then the collection of returns from K pulses forms an array of N × K elements. We will (arbitrarily) assume that each pulse is represented as a (vertical) column in the matrix, so that the row index is related to the time of reception of a sample from a single pulse, while the column index identifies the particular pulse in question.

B. Cross-Correlation and the Range Processing

Let

$$u = \big(u(1),\ u(2),\ \ldots,\ u(m)\big)$$

and

$$v = \big(v(1),\ v(2),\ \ldots,\ v(m)\big)$$

be two vectors of the same length m. We define their cross-correlation vector C_{uv}, also of length m, by

$$C_{uv}(k) = \sum_{\ell=1}^{m} u(\ell)\ v(\overline{k+\ell}) \qquad \text{for } k = 1,\ 2,\ \ldots,\ m \tag{57}$$

where

$$\bar{n} \text{ means} \begin{cases} (n-1) \bmod m \text{ for } n \neq m \\ m \text{ for } n = m \end{cases}$$

so that

$$\overline{k+\ell} = \begin{cases} (k+\ell-1) \bmod m \text{ for } k+\ell-1 \neq m \\ m \text{ for } k+\ell-1 = m \end{cases}$$

For example, if

$$u = (1, -1, 0, 0, 0)$$

$$v = (0, 0, 1, -1, 0)$$

then

$$\begin{cases} C_{uv}(1) = 1 \times 0 + (-1) \times 0 + 0 \times 1 + 0 \times (-1) + 0 \times 0 = 0 \\ C_{uv}(2) = 1 \times 0 + (-1) \times 1 + 0 \times (-1) + 0 \times 0 + 0 \times 0 = -1 \\ C_{uv}(3) = 0 \times 0 + 0 \times 0 + 1 \times 1 + (-1) \times (-1) + 0 \times 0 = 2 \\ C_{uv}(4) = 1 \times (-1) + (-1) \times 0 + 0 \times 0 + 0 \times 0 + 0 \times 1 = -1 \\ C_{uv}(5) = 1 \times 0 + (-1) \times 0 + 0 \times 0 + 0 \times 1 + 0 \times (-1) = 0 \end{cases} \tag{58}$$

In other words, to cross-correlate u and v, we successively slide v "around itself" and then multiply by u and add, as in Fig. 21.

Note that C_{uv} has a maximum at $C_{uv}(3)$, indicating that the third sliding position of v gives the "best match" with u.

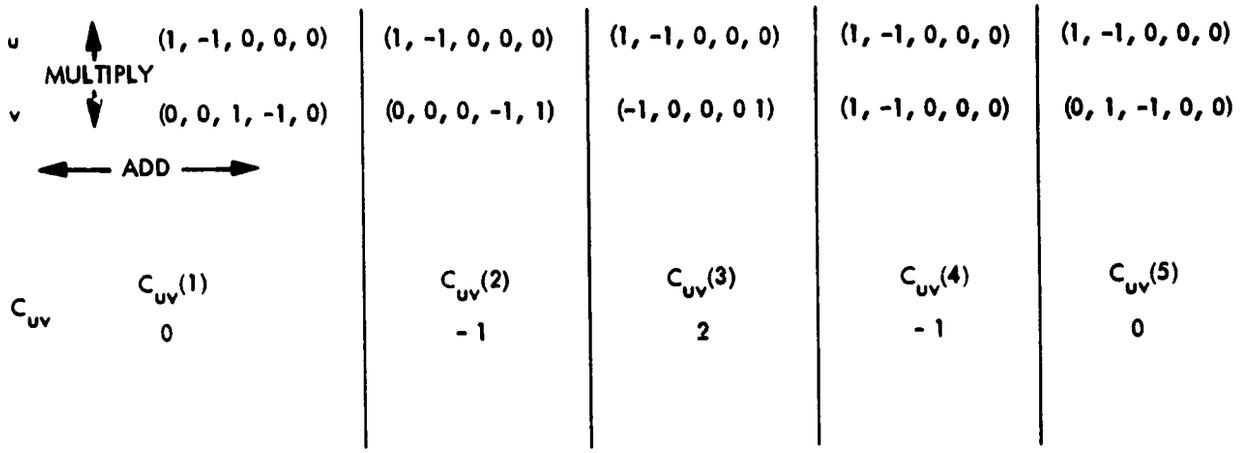


Fig. 21. Cross-Correlation of Vectors u and v

The cross-correlation operation is linear. This means that if u, v, and w are vectors and α is scalar, then

$$C_{\alpha u, v} = \alpha C_{u, v} \tag{59}$$

and

$$C_{u, v+w} = C_{u, v} + C_{u, w} \ . \tag{60}$$

To prove Eq. (59), we use Eq. (57) to write

$$C_{\alpha u, v}(k) = \sum_{\ell=1}^{m} u(\ell)\, v(\overline{k+\ell}) = \alpha \sum_{\ell=1}^{m} u(\ell)\, v(\overline{k+\ell}) = \alpha\, C_{u, v}(k)$$

Similarly, Eq. (60) is proved by writing

$$C_{u,\,v+w}(k) = \sum_{\ell=1}^{m} u(\ell)\left[(v+w)(\overline{\ell+k})\right]$$

$$= \sum_{\ell=1}^{m} u(\ell)\left[v(\overline{\ell+k}) + w(\overline{\ell+k})\right]$$

$$= \sum_{\ell=1}^{m} u(\ell)\, v(\overline{\ell+k}) + \sum_{\ell=1}^{m} u(\ell)\, w(\overline{\ell+k})$$

$$= C_{uv}(k) + C_{uw}(k)$$

As an example, the reader may verify that if u = (1, -1, 0, 0, 0), v = (0, 0, 1, -1, 0) and w = (0, 0, 0, 1, -1) then

$$C_{3u,\,v} = (0, -3, 6, -3, 0) = 3C_{u,\,v} = 3 \times (0, -1, 2, -1, 0)$$

and

$$C_{u,\,v+w} = C_{u,\,v} + C_{u,\,w}$$

$$= (0, -1, 2, -1, 0) + (0, 0, -1, 2, -1)$$

$$= (0, -1, 1, 1, -1)$$

Example

As an example we consider the linear FM waveform discussed in the previous section. Assume that a short linear FM return echo (of say 2M + 1 samples, where M is determined by the range pulsewidth T and the sample period Δt),

from a point target is embedded in a much longer vector v (of say length N) consisting of all the terrain echo samples from a single pulse. We denote the 2M + 1 nonzero linear FM samples as

$$s(-M),\ s(-M+1),\ \ldots.\ s(-1),\ s(0),\ s(1),\ \ldots,\ s(M)$$

Thus for k = 0, ±1, ..., ±M, s(k) represents a sample from the return in Eq. (44). To see how we may conveniently express (44) in terms of k, the sample index, let us translate the time axis in (44) according to eq. (51), so that time y is 0 in the middle of the echo return (44). Set

$$B = A \exp\left\{ j\left(\frac{-2\omega_0 R}{c}\right)\right\} \tag{61}$$

and

$$\sigma = -\frac{a}{2}(\Delta t)^2 \tag{62}$$

where

$$\Delta t = \frac{2\pi}{aT} \tag{63}$$

is given by (54). Then we have, for k = 0, ±1, ..., ±M,

$$s(k) = Be^{-jbk^2} \tag{64}$$

(An easy way to see the validity of (64) is to set time y = Δt in (44), corresponding to k = 1 in (64). Both b in (64) and a in (62) are commonly called the chirp

rate. For example, in (44) or (10), the rate of change of radian frequency with respect to time is

$$\phi''(t) = a$$

(cf. eqs. (12), (13)).

We wish to cross-correlate v with another vector u of similar form. More specifically, u contains the conjugate terms to Eq. (61), with B replaced by unity. Thus somewhere in u is a set of terms

$$\bar{s}(-M),\ \bar{s}(-M+1),\ \ldots,\ \bar{s}(M-1),\ \bar{s}(M)$$

where

$$\bar{s}(k) = e^{+jbk^2} \qquad \text{for } k = 0,\ \pm 1,\ \ldots,\ \pm M \tag{65}$$

All the remaining terms of u are 0. (The vector u is also of length N.)

Except when the s terms in v overlap with the $\bar{s}$ terms in u, the cross correlation vector elements are 0. When there is overlap, we may choose the index origin so that the index k runs from -M to +M:

$$C_{vu}(i+1) = \sum_{k=-M}^{M} Be^{-jbk^2} e^{+jb(k+i)^2} \tag{66}$$

At first sight, it would appear that (66) gives the correct formula for $C_{vu}(i+1)$ for all i. However, recall that $\bar{s}(k+i)$ is nonzero only for

$(k+\ell) = 0, \pm 1, \ldots, \pm M$. For $k > M - \ell$, $\bar{s}(k+\ell) = 0$. Thus the highest summation index in (66) is not M but $M-\ell$, and the correct formula for the finite duration discrete linear FM cross-correlation is

$$C_{vu}(\ell+1) = \sum_{k=-M}^{M-\ell} Be^{-jbk^2} e^{+jb(k+\ell)^2} \tag{67a}$$

for the s and $\bar{s}$ overlap region

$$-2M \le (\ell+1) \le 2M$$

and

$$C_{vu}(\ell+1) = 0 \tag{67b}$$

otherwise.

We show next that Eq. (67) may be simplified to

$$C_{vu}(\ell+1) = \frac{B \sin\left[2b\left(M + \frac{1}{2}\right)\ell - b\ell^2\right]}{\sin(b\ell)} \quad \text{for } -2M \le (\ell+1) \le 2M \tag{69}$$

The reader unconcerned with details may skip to Eq. (75).

Observe first that Eq. (67) simplifies to

$$\sum_{k=-M}^{M-\ell} Be^{j2bk\ell} e^{jb\ell^2} \tag{70}$$

Factoring out $e^{jb\ell^2}$ we have

$$C_{vu}(\ell+1) = Be^{jb\ell^2} \times \sum_{k=-M}^{M-\ell} e^{j2bk\ell} \tag{71}$$

The right hand geometric series sums to the term in brackets below, which shows

$$C_{vu}(\ell+1) = Be^{jb\ell^2}\left[\frac{e^{-j2bM\ell} - e^{j2b(M-\ell)\ell} \times e^{j2b\ell}}{1 - e^{j2b\ell}}\right]$$

$$= Be^{jb\ell^2}\left[\frac{e^{-j2bM\ell} - e^{+j(2bM\ell+2b\ell-2b\ell^2)}}{1 - e^{j2b\ell}}\right] \tag{72}$$

From both numerator and denominator we factor out $e^{jb\ell}$ to obtain

$$C_{vu}(\ell+1) = Be^{jb\ell^2}\left[\frac{e^{-j2b(M+\frac{1}{2})\ell} - e^{j2b(M+\frac{1}{2})\ell} \times e^{-j2b\ell^2}}{e^{-jb\ell} - e^{jb\ell}}\right] \tag{73}$$

Factoring out $e^{-jb\ell^2}$ from the numerator yields

$$C_{vu}(\ell+1) = Be^{jb\ell^2} e^{-jb\ell^2}\left[\frac{e^{-j2b(M+\frac{1}{2})\ell+jb\ell^2} - e^{j2b(M+\frac{1}{2})\ell-jb\ell^2}}{e^{-jb\ell} - e^{jb\ell}}\right]$$

$$= \frac{Be^{j\left[2b(M+\frac{1}{2})\ell-b\ell^2\right]} - e^{-j\left[2b(M+\frac{1}{2})\ell-b\ell^2\right]}}{e^{jb\ell} - e^{-jb\ell}} \tag{74}$$

so that finally

$$C_{vu}(\ell+1) = \frac{B \sin\left[2b\left(M+\frac{1}{2}\right)\ell - b\ell^2\right]}{\sin(b\ell)} \tag{75}$$

as promised.

Note that Eq. (75) has a peak magnitude of $|B \cdot (2M+1)|$ when $\ell = 0$. (Actually Eq. (75) is undefined at $\ell = 0$, but Eq. (67) may be easily evaluated when $\ell = 0$.) Also the first null of C_{vu} appears when ℓ satisfies

$$2b\left(M+\frac{1}{2}\right)\ell - b\ell^2 = \pm\pi$$

or

$$b\ell^2 - 2b\left(M+\frac{1}{2}\right)\ell \pm \pi = 0$$

or

$$\ell^2 - 2\left(M+\frac{1}{2}\right)\ell \pm \frac{\pi}{b} = 0$$

By completing the square we find

$$\ell = M + \frac{1}{2} - \sqrt{\left(M+\frac{1}{2}\right)^2 - \left|\pi/b\right|} \tag{76}$$

corresponds to the first positive null. Note that when $|b|$ is large, $\ell \to 0$. In fact, for fairly large $|b|$, we may write

$$\sqrt{\left(M+\frac{1}{2}\right)^2 - \left|\frac{\pi}{b}\right|} = \left(M+\frac{1}{2}\right)\sqrt{1 - \left|\frac{\pi}{b}\right| \times \frac{1}{\left(M+\frac{1}{2}\right)^2}}$$

where

$$\frac{\pi}{|b|\left(M+\frac{1}{2}\right)^2} \ll 1$$

Using a Taylor expansion (as illustrated for example in Eqs. (85), (86)) we obtain

$$\sqrt{\left(M+\frac{1}{2}\right)^2 - \left|\frac{\pi}{b}\right|} \simeq \left(M+\frac{1}{2}\right)\left(1 - \frac{\pi}{|2b|\left(M+\frac{1}{2}\right)^2}\right) \tag{77}$$

Eqs. (76) and (77) show that at the first null,

$$\ell \sim M+\frac{1}{2} - \left(M+\frac{1}{2}\right) + \frac{\pi}{|2b|\left(M+\frac{1}{2}\right)}$$

or

$$\ell \sim \frac{\pi}{|2b|\left(M+\frac{1}{2}\right)} \sim \left|\frac{\pi}{2bM}\right| \tag{78}$$

For values of ℓ greater than $\frac{\pi}{|2b|\left(M+\frac{1}{2}\right)}$ the magnitude of Eq. (75) is small compared to its magnitude for values of ℓ less than $\frac{\pi}{|2b|\left(M+\frac{1}{2}\right)}$. Thus if we plot the function $\left|C_{VU}(\ell+1)\right|^2$ versus ℓ, we obtain the response of Fig. 22.

We see from Fig. 22 that the cross-correlation of the return echo with the conjugate exponential is the key to retrieving our lost range resolution. For, while our transmitted chirp was $2M+1$ samples wide, the result of cross-correlating is only $2 \times \frac{\pi}{|2b|\left(M+\frac{1}{2}\right)}$ or $\frac{\pi}{|b|\left(M+\frac{1}{2}\right)}$ samples wide, essentially.

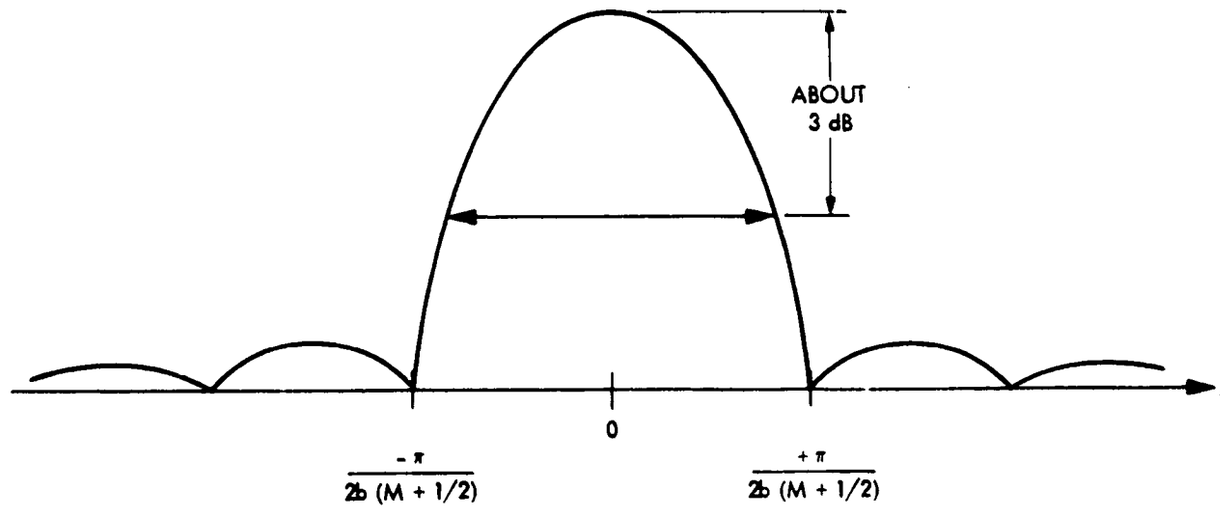


Fig. 22. Plot of Square of Absolute Value of Linear FM Cross-Correlation Function

In fact, the width of the half-amplitude (3 dB down) part of the curve in Fig. 22 (shown with the double arrow ⟷) is only $\left|\frac{\pi}{2bM}\right|$, approximately. We define this "half-width" loosely as resolution. More specifically, Fig. 22 shows the one-dimensional response to a point target. We can replace the abscissa units "ℓ" in Fig. 22 by any convenient abscissa units desired, such as time t:

$$t = \ell \Delta t$$

or slant range r:

$$r = \frac{ct}{2}$$

If we use the t-axis, the half-width in t-units is called the time-resolution. The half-width in r-units is called the slant range spatial resolution. We also say that the dimension of the half-width corresponds to one resolution cell.

When we recall that

$$b = -\frac{a}{2}(\Delta t)^2$$

from Eq. (62), then Eq. (78) assumes a special form for analog linear FM signals. For, if we let ℓ_r be the number of output samples within a range resolution cell, then

$$\left|\frac{\pi}{2bM}\right| \simeq \ell_r$$

When we realize that ℓ_r samples correspond to a time interval of $\ell_r \Delta t$ seconds, we conclude that

$$\left|\frac{\pi \Delta t}{2bM}\right| \sim \ell_r \Delta t \tag{79}$$

is an expression for the time resolution of the radar. Since

$$|b| = \frac{a}{2}(\Delta t)^2$$

we have

$$\text{Time resolution} \simeq \left|\frac{\pi \Delta t}{2bM}\right|$$

$$= \frac{\pi \Delta t}{2 \times \left(\frac{a}{2}\right)(\Delta t)^2 M}$$

$$\text{Time resolution} \sim \frac{2\pi}{2M(\Delta t)a} \tag{80}$$

Now $2M(\Delta t)$ is the approximate linear FM pulse duration and a is the radian frequency sweep rate, so that $(2M(\Delta t)a)$ is the pulse radian bandwidth. Thus $\frac{2M(\Delta t)a}{2\pi}$ is the chirp bandwidth in Hz, and the time resolution (80) is the reciprocal 1/BW of the chirp bandwidth in Hz! Now we can see why the term "compression" is used to describe the radar cross-correlation processing. After cross-correlation, the signal looks like Fig. 22 and has essentially a mainlobe time width of $\approx 1/BW$. However, the original pulse time duration was T. Dividing the original pulse duration by the "processed" pulse duration gives the "compression ratio" CR:

$$CR = T \Big/ \left(1 \Big/ BW\right) = T \times BW \tag{81}$$

Thus the compression ratio is the time-bandwidth product.

We summarize: We send out a linear FM chirp of Hertz bandwidth BW. We wait for the echo which we then correlate with a conjugate replica of the original chirp, embedded in a vector of length equal to that of the echo vector. If a point target is in the reflecting field, it will produce a lobe structure as in Fig. 22, with time resolution $\frac{1}{BW}$. The height of the mainlobe in Fig. 22 will be proportional to the reflectivity of the point target (amplitude factor B). The actual peak placement along the abscissa (labelled "0" in Fig. 22) depends directly on the slant range R to the target at the time of pulse transmission. If several point targets are found at varying locations along the image swath, linearity of the cross-correlation function shows they will produce several separate mainlobes, appearing at several different places in the output cross-correlation vector, as in Fig. 23.

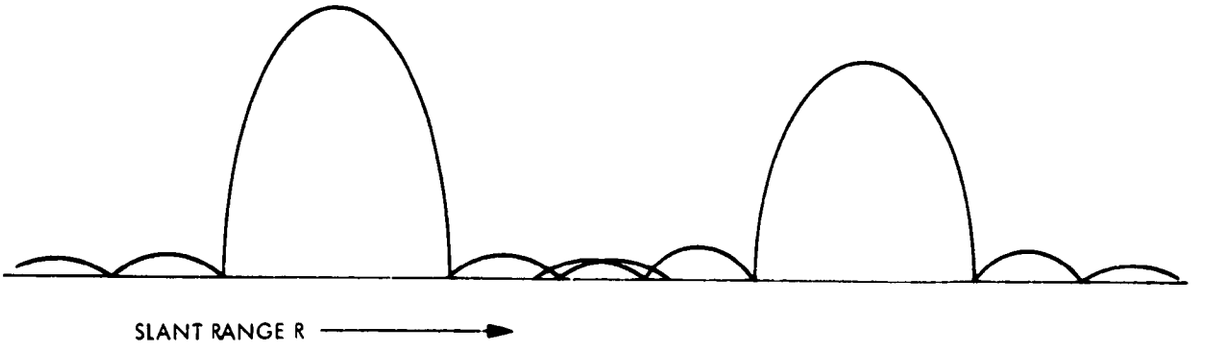


Fig. 23. Cross-Correlation Output When Two Point Targets Are Found at Different Slant Ranges

Note from Eqs. (61) and (69) that prior to taking absolute values, the curves in Fig. 23 all have azimuth phase factors of the form

$$\exp\left\{-j\frac{2\omega_0 R}{c}\right\} = \exp\left\{-j\frac{4\pi R}{\lambda}\right\}$$

where R specifies the target slant range (see Eq. (75)). And, in fact, absolute values are not taken right after range correlation. If they were, the all-important azimuth phase factor would be lost and the azimuth resolution would be limited by the antenna azimuth beamwidth, as in Fig. 2. However, the only reason we took absolute values in the above explanation was to give the reader a picture (Fig. 23) of what the range processed signal looks like.

We are almost ready to begin the discussion of the azimuth processing. However, it will be useful to first characterize the range correlated return echo a bit more fully. Suppose there is only the single point target PT in the beam when our single pulse is transmitted. Then if ℓ' is the actual index of the cross-correlated return echo, the actual form of the cross-correlation result is not Eq. (69) but is rather a translated version of Eq. (69), in which ℓ is replaced by $(\ell' - q)$. We may write

$$C_{vu}(\ell') = B \sin \frac{\left[2b\left(M + \frac{1}{2}\right)(\ell' - q) - b(\ell' - q)^2\right]}{\sin\left\{b[\ell' - q]\right\}} \qquad (82)$$

The translation index q (related to β in (51)) reflects the actual slant range to PT, and is thus a function of R. Recall also that

$$B = A \exp\left\{-j \frac{4\pi R}{\lambda}\right\}$$

is also a function of R. Thus the location of the peak in Fig. 24 is a function of R, and so is the azimuth phase $\left(-\frac{4\pi R}{\lambda}\right)$ that is attached to <u>every point</u> on the curve in Fig. 24. The azimuth phase is <u>independent</u> of index ℓ (cf. Eq. (82)).

From the last remark it may be evident that optimum azimuth processing should make use of every point on the curve of Fig. 24, <u>for every pulse echo in which PT returns a signal.</u> Such processing is currently regarded as too complex, however, and only the qth point is customarily used for the azimuth processing. We will return to this point later (Fig. 27).

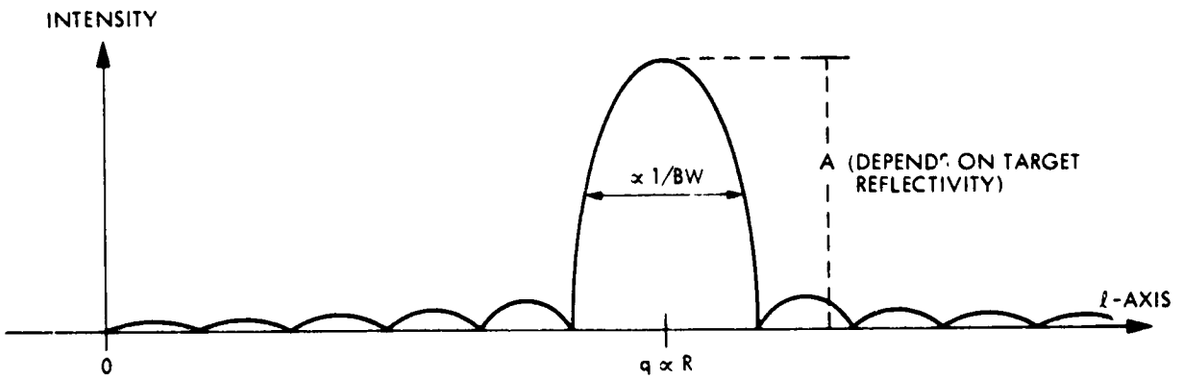


Fig. 24. Dependence of Range Correlation Peak Index on R: The Entire Curve is Multiplied by the Constant Term:

$$\exp\left\{-j\,\frac{4\pi R}{\lambda}\right\}$$

IV. The Azimuth Processing

The range processing involved operations on the information within a single pulse echo. The azimuth processing will combine the information from a number of echoes. The azimuth phase factor $\Phi_a = \exp\left\{j\left(-\frac{2\omega_0 R}{c}\right)\right\}$ of Eq. (44) is the key to this part of the processing. Recall that this factor arose from the mixing operation on the return echo. Although the azimuth processing is difficult to implement on a computer, it is conceptually simple, especially after the range processing has been understood.

The first thing we have to do is examine where the data from a point target is found within a sequence of range correlated echoes. For simplicity, we assume the radar is mounted on an aircraft traveling past a point target PT on the ground. (The more complicated case of nonlinear or orbital motion is easily understood if the aircraft scenario is understood.) In Fig. 25 the slant

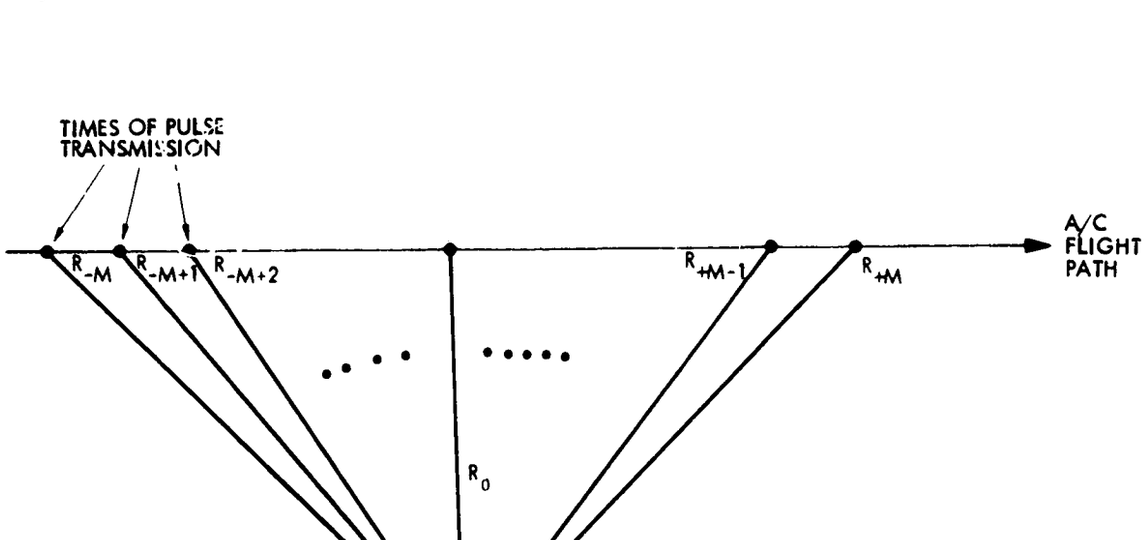


Fig. 25. Slant Ranges as a Function of Aircraft Position Relative to Target

ranges to the target are indexed as a function of the various positions successively occupied by the moving aircraft. We assume that only during pulses (-M) to (+M) does the beam illuminate PT. Thus there are (2M+1) pulses carrying information about PT. We have labeled one slant range specifically R_0 because it represents the point of closest approach. The important thing to note is that the slant range R to PT changes with each pulse. We can express this quantitatively as follows.

In Fig. 26, choose the x-axis along the flight path so that x = 0 when the aircraft is at the point of closest approach to PT. Then for any A/C position x, we have by the Pythagorean theorem that the slant range R to PT is given by

$$R = \sqrt{R_0^2 + x^2} \tag{8.3}$$

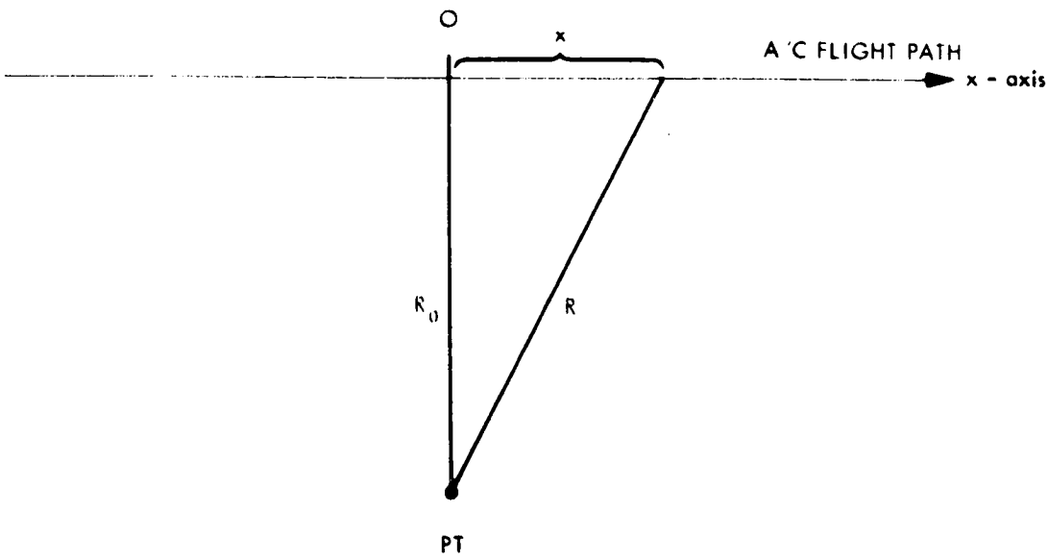


Fig. 26. Relationship of A/C Position to Slant Range

In a practical SAR, x is very small compared to R_0 because we only need to consider values of x when PT is illuminated by the narrow azimuth beam. Therefore

$$\frac{x}{R_0} \ll 1$$

and we may write Eq. (83) as

$$R = R_0 \sqrt{1 + \left(\frac{x}{R_0}\right)^2} \tag{84}$$

Eq. (84) has a radical of the form

$$F(y) = \sqrt{1 + y} \tag{85}$$

where $|y| \ll 1$. Thus we may expand $F(y) = \sqrt{1 + y}$ in a Taylor series about $y = 0$. Taking only the first two terms we have

$$F(y) \sim 1 + \frac{y}{2} \tag{86}$$

as a good approximation when $|y| \ll 1$.

Substituting $y = \left(\frac{x}{R_0}\right)^2$ in Eq. (84) we have

$$R \sim R_0 \left(1 + \left(\frac{x}{R_0}\right)^2 \Big/ 2\right)$$

or

$$R \sim R_0 + \frac{x^2}{2R_0} \tag{87}$$

which gives slant range R as a function of A C azimuth position x.

In Eq. (87), the variable x is linearly related to the pulse index m, where $-M \leq m \leq M$. The pulse index m is in turn associated with the slant range R_m in Fig. 25. When the aircraft is at position x, the slant range given by Eq. (87) will determine the location of the range correlation peak in Fig. 24. In particular, q in (82) will be determined by R which is itself a quadratic function of x. Thus if we examine the sequence of range correlated echos associated with PT, we see that the location of the peak in Fig. 24 will vary with the pulse index, as in Fig. 27 (where M = 3). Figure 27 represents the display of range correlated data over a sequence of pulses during which PT is illuminated. Each echo in Fig. 27 is a copy of the sinusoidal waveform of Eq. (75), where the location of the peak in Eq. (75) is now a function of x (i.e., the pulse index). As mentioned earlier, (e.g., p. 7 and p. 50), we wish to combine the information from all seven pulses in Fig. 27 to produce one range line of imagery. Of course, we

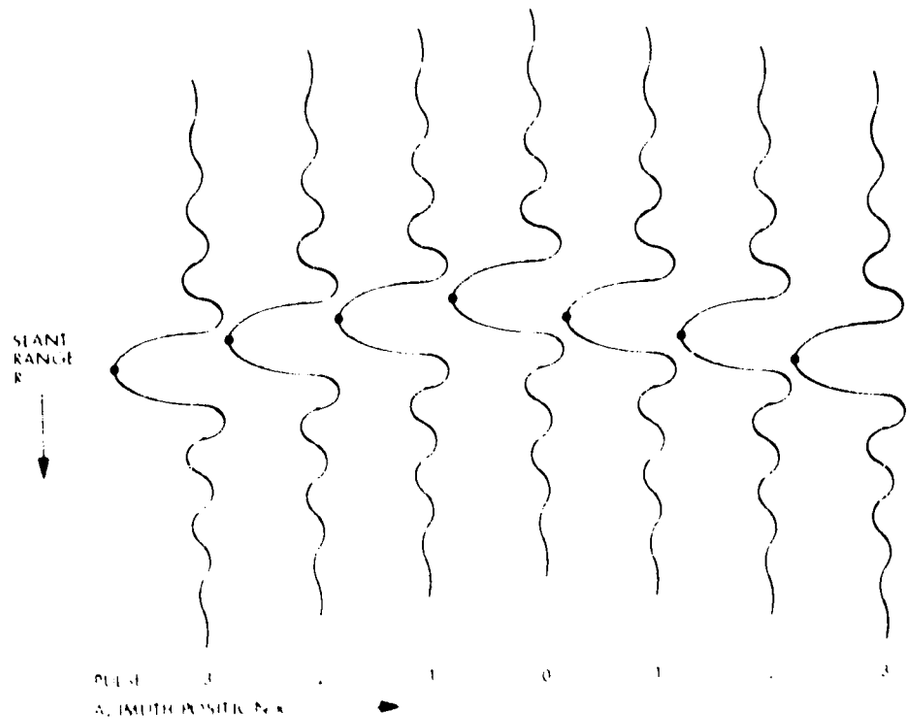


Fig. 27. Sequence of Range Correlated Echoes from PT; The Sideways Peak's Should Really Come Out of the Paper Toward the Reader.

wish to choose the strongest possible signals from PT in each pulse. To obtain the strongest signal from PT, thus, we must take the dotted samples in Fig. 27. These dotted samples follow an approximate parabola, as shown by Eq. (87). It is these dotted samples which we will extract and use for the azimuth correlation of PT. Associated with each dotted sample is an azimuth phase factor of the form:

$$B = A \exp\left(-\frac{j\,4\pi R}{\lambda}\right) \tag{88}$$

(Recall from Eq. (75) and the subsequent remarks that each dotted sample has also a magnitude factor of $(2M+1.)$

R is, remember, a function of azimuth position x. Thus, placing the dots of Fig. 27 in an "azimuth" vector

$$Az = \left(Az(1),\ Az(2),\ \ldots,\ Az(N)\right)$$

where

$$N = 2M + 1$$

we may again examine the effects of cross-correlating Az with a vector $\overline{Az}$ whose phase factor is the conjugate to Eq. (88), i.e., at position x, $\bar{a}$ has phase factor

$$\exp\left(+j\,\frac{4\pi R}{\lambda}\right)$$

(cf. Eq. (81)).

If we recall that

$$R \sim R_0 + \frac{x^2}{2R_0}$$

from Eq. (87), then we see that the correlation function takes the form

$$\exp\left\{+j\frac{4\pi}{\lambda}\left(R_0 + \frac{x^2}{2R_0}\right)\right\} \tag{89}$$

Let us throw the correlation function into a more "digital" appearing form. Since presumably aircraft position x is related to velocity v and time t by

$$x = vt \tag{90}$$

(provided t = 0 corresponds to the point of closest approach), we have for Eq. (89)

$$\exp\left\{j\frac{4\pi}{\lambda}\left(R_0 + \frac{x^2}{2R_0}\right)\right\} = \exp\left\{j\frac{4\pi}{\lambda}\left(R_0 + \frac{v^2t^2}{2R_0}\right)\right\} \tag{91}$$

Let

$$PRF = \text{pulse repetition frequency}$$

so that

$$\Delta t = \frac{1}{PRF} = \text{pulse repetition period}$$

is the length of time between pulses. (Recall that in section III, Δt denoted the range sampling period; here however we use it to denote the azimuth sampling period.) Then if $k = \pm 1, \pm 2, \pm 3, \ldots, \pm M$, the azimuth reference function of Eq. (91) is given as a function of discrete multiples of time Δt by

$$\overline{\Delta z}(k) = \exp\left\{j\frac{4\pi}{\lambda}\left(R_0 + \frac{v^2(k\Delta t)^2}{2R_0}\right)\right\} \tag{92}$$

In Eq. (92), the constant phase factor

$$\exp\left\{j\frac{4\pi}{\lambda} \times R_0\right\}$$

is usually omitted.

Now we can see that the same correlation process as described in Section III may be used. In short, the vector in Eq. (92) is embedded in a long vector, padded with zeros, called the azimuth correlation function. The conjugate (received) signal, Az(k), given by

$$Az(k) = A\exp\left\{-j\frac{4\pi}{\lambda}\left(R_0 + \frac{v^2(k\Delta t)^2}{2R_0}\right)\right\} \quad \text{for } k = 0, \pm 1, \ldots, \pm M \qquad (93)$$

is embedded in an equally long zero-padded vector. Note that the Az(k) are the dots in Fig. 27.

If we observe that Az(k) also has a constant phase factor $\exp\left\{-j\frac{4\pi R_0}{\lambda}\right\}$ which may be removed, then we see that the essential cross-correlation is between

$$Az(k) = \exp\left\{j\frac{4\pi}{\lambda} \times \frac{v^2k^2(\Delta t)^2}{2R_0}\right\} \qquad (94)$$

and

$$Az(k) = \exp\left\{-j\frac{4\pi}{\lambda} \times \frac{v^2k^2(\Delta t)^2}{2R_0}\right\} \qquad (95)$$

Eq. (95) is of the form $\exp\{jbk^2\}$ where

$$b = \frac{4\pi}{\lambda} \times \frac{v^2(\Delta t)^2}{2R_0}$$

Thus we may use Eq. (69) to conclude that the result of azimuth correlation has the form

$$I(\ell) = B \sin \frac{\left[\frac{4\pi v^2 (\Delta t)^2}{\lambda R_0} \ell \left(M + \frac{1}{2} - \frac{\ell}{2}\right)\right]}{\sin\left(\frac{2\pi v^2 (\Delta t)^2}{\lambda R_0} \ell\right)} \tag{96}$$

where now the factor B is just the amplitude factor A. From the results of Section III, we know the time resolution of Eq. (96) is the reciprocal of the azimuth bandwidth. Multiplying this time resolution by v, the aircraft velocity, gives the azimuth spatial resolution.

More specifically, we know the azimuth doppler chirp rate in Eq. (91) is the second derivative of phase. Since azimuth phase ϕ_a is

$$\phi_a = \frac{2\pi v^2 t^2}{\lambda R_0}$$

the chirp rate is

$$\ddot{\phi}_a = \frac{4\pi v^2}{\lambda R_0} \tag{97}$$

If the radar illuminates PT over a time T_a, the radian bandwidth is

$$RBW_a = \frac{4\pi v^2}{\lambda R_0} \times T_a$$

and the Hertz bandwidth is $RBW_a / 2\pi$ or

$$BW_a = \frac{2v^2}{\lambda R_0} \times T_a \tag{98}$$

The time T_a is called the coherent integration time. It is related to the physical distance L_a over which the aircraft travels while illuminating PT, by the simple formula

$$L_a = vT_a$$

(The length L_a is called the synthetic aperture length.) Therefore, Eq. (98) may be rewritten as

$$BW_a = \frac{2vL_a}{\lambda R_0} \tag{99}$$

Assuming the antenna beam pointing direction is fixed, the maximum possible length L_a over which PT is illuminated is the same as the width of the azimuth beam on the ground. This width, from Fig. 2, is given as $R_0\beta_a$. Thus for the maximum value of L_a, Eq. (99) becomes

$$BW_a = \frac{2v\ R_0\ \beta_a}{\lambda R_0}$$

Let L now denote the physical antenna azimuth dimension. Since $\beta_a = \frac{\lambda}{L}$, (cf. Eq. (26)), we have

$$BW_a = \frac{2v}{L} . \tag{100}$$

Recalling that time resolution is given as $\frac{1}{BW_a}$, we have then

$$\Delta t_a = \frac{L}{2v} \tag{101}$$

as the azimuth time resolution, and $v\Delta t_a$, or

$$\Delta x = \frac{L}{2} \tag{102}$$

as the best obtainable azimuth spatial resolution. L is the physical antenna azimuth dimension.

If, on the other hand, we use a shorter length for L_a than the maximum possible synthetic aperture length, then Eq. (99) implies a time resolution value of

$$\Delta t_a = \frac{\lambda R_o}{2vL_a} \tag{103}$$

and an azimuth spatial resolution of $v\Delta t_a$ or

$$\Delta x = \frac{\lambda R_o}{2L_a} \tag{104}$$

as a function of synthetic aperture length L_a. The reader will note from Eq. (104) that the best azimuth resolution is obtained with the longest synthetic aperture length L_a. When the longest possible synthetic aperture is used, Eq. (102) applies and the resolution is limited by the physical antenna azimuth dimension L. Therefore, in a SAR the smallest physical antenna leads to the best resolution. As an example, the SEASAT spaceborne SAR physical antenna is 10 meters along the azimuth. The synthetic aperture length, however, is more than 4500 meters.

A few remarks are necessary here to clarify the simplifying assumptions leading to the azimuth impulse response given by Eq. (96). Eq. (96) was derived assuming a point target PT is alone in the image field and produces a linear FM signal embedded in a longer vector of zeros. Since each point target along the azimuth produces a distinct nonlinear track (like the fishbones of Fig. 28) it is evident that the actual cross-correlation process must be performed with care; for example, readers familiar with FFT cross-correlation

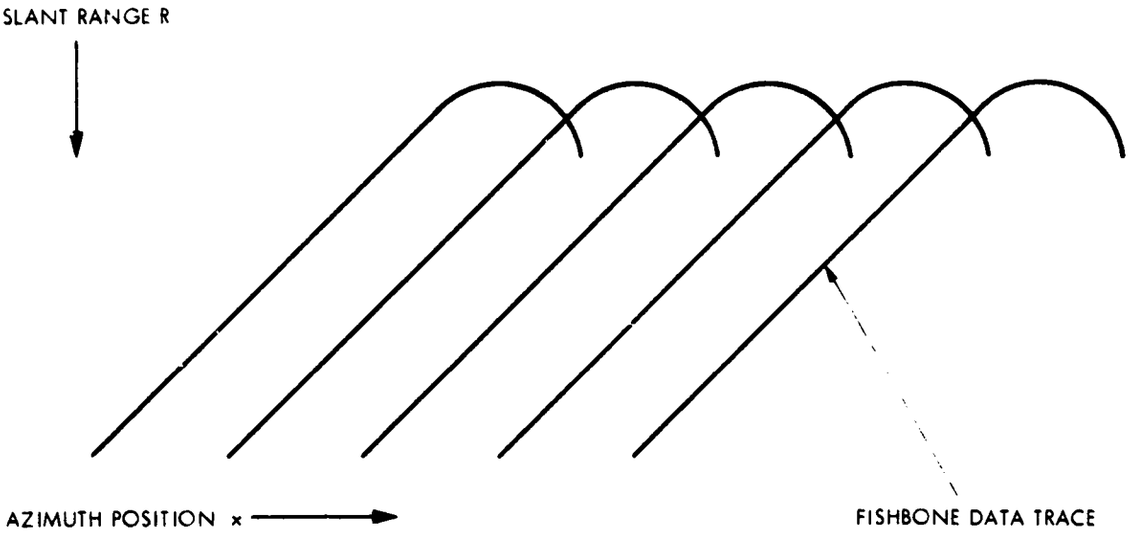


Fig. 28. Fishbone Data Traces of Successive Azimuth Point Targets at the Same Slant Range

techniques will recognize the impossibility of performing a fast correlation via an FFT for the azimuth processing because the "fishbone" data traces in Fig. 28 are not colinear. If the data samples of each "fishbone" are used separately to perform a time domain correlation, however, the absence of interfering data from the neighboring fishbones will actually result in a "better" impulse response than that depicted in Eq. (96).

We finally observe that the two-dimensional point target response ("impulse response") of a synthetic aperture radar is the product of the one-dimensional responses of (75) and (96). The index ℓ is a range index in (75); in (96) ℓ is an azimuth index. It is a useful exercise for the reader to try to visualize this impulse response.

V. Some Additional Information

In all the above analysis we assumed a single point target PT at slant range R from the radar. The linearity of cross-correlation in both the range and azimuth processing ensures that at every position in the output array, the pixel value obtained is "reasonably" proportional to the reflectivity of the corresponding location on the terrain. The illustration of this effect was already given in Fig. 23, where the two mainlobes in the figure (together with the sidelobes) add together to produce a curve which appears about the same as that in the figure.

Additional simplifying assumptions we have used may become apparent to the reader as he continues to explore the field of SAR more deeply. However, it is hoped that armed with the basic understanding provided by the above analysis, the reader will easily grasp the attendant subtleties.

For example, although an orbiting spacecraft SAR does not permit the use of the simple geometry of Fig. 25, the azimuth correlation function nevertheless still takes a form similar to (91), because the quadratic approximation is still useful. Thus the general SAR azimuth correlation function has the form

$$\overline{Az}(t) = \exp\left\{j\left(2\pi ft + \frac{2\pi \dot{f} t^2}{2}\right)\right\}$$

where f is the doppler center frequency and $\dot{f}$ is the azimuth chirp rate in Hertz/second. Some interesting problems related to SAR processing are described in the next section.

VI. Some SAR Processing Problems

A. Speckle

Because a SAR is coherent (i. e., preserves phase information), the typical SAR image is marred with irregular variations in brightness which look like a laser speckle pattern on the wall. The cause is the same for a SAR and a laser.

Each resolution cell in a SAR image (i. e., each "pixel" or correlated output point) may be thought of as corresponding to a patch on the ground of finite area. Within this patch there are many "point" targets, each contributing to the overall reflectivity of the scene, as in Fig. 29. Suppose there are N of

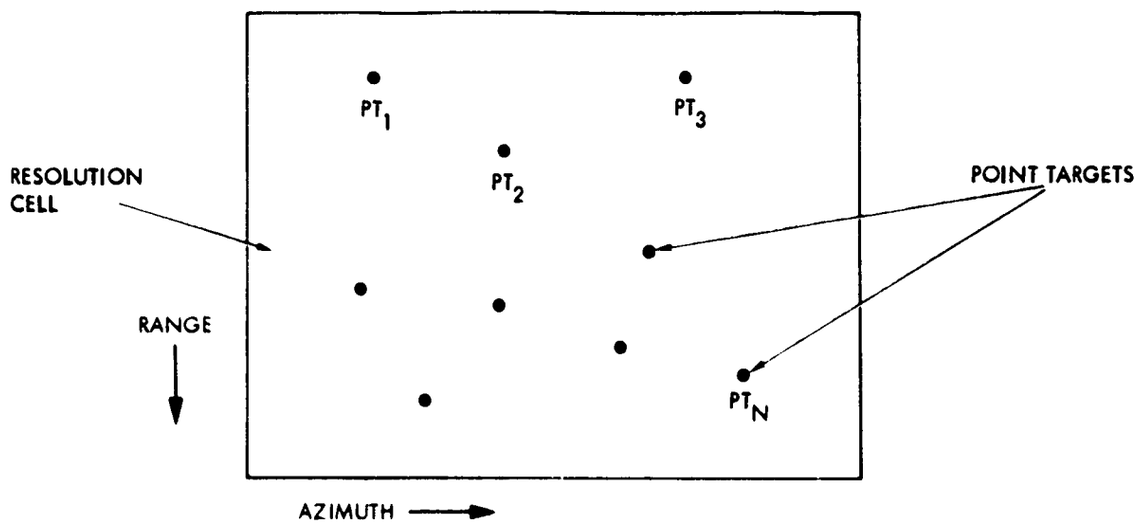


Fig. 29. Composition of a Resolution Cell

them, PT_1, PT_2, ..., PT_N where $N \gg 1$. For each k, $1 \le k \le N$, PT_k returns a signal to the radar of the form

$$s_k(t) = A_k \cos\left(\phi\left(t - \frac{2R}{c}\right) + \theta_k\right) \tag{105}$$

where

R = nominal slant range to resolution cell

$\phi(t)$ = the normal expected phase function, as in Section II. For example

$$\phi(t) = \left(\frac{at^2}{2} + \omega_0 t\right)$$

θ_k = Phase adjustment, a real number related to the actual slant range of PT_k owing to surface roughness, viewing angle, etc. (This phase adjustment θ_k also depends on dielectric and reflection-phase-inversion characteristics of PT_k). Thus θ_k is a random variable, uniform on $[0, 2\pi]$, and independent of θ_j for $k \neq j$.

$A_k = |A_k|$ = the amplitude or reflection efficiency of PT_k.

After mixing and correlation, the return from PT_k is estimated as

$$r_k = A_k \exp\{j\theta_k\} \tag{106}$$

By the linearity of cross correlation, the overall correlated return from the resolution cell is

$$S = \sum_{k=1}^{N} r_k = \sum_{k=1}^{N} A_k \exp\{j\theta_k\} \tag{107}$$

We may express each r_k in Cartesian form by

$$r_k = A_k \exp\left\{j\theta_k\right\} = I_k + j\,Q_k \tag{108}$$

where I_k and Q_k are real. Thus Eq. (107) becomes

$$S = \sum_{k=1}^{N} I_k + j \sum_{k=1}^{N} Q_k \tag{109}$$

or

$$S = I + jQ \tag{110}$$

where

$$I = \sum_{k=1}^{N} I_k \tag{111}$$

$$Q = \sum_{k=1}^{N} Q_k \tag{112}$$

From Eq. (110), we may go back to the polar representation; thus

$$S = I + jQ = A \exp\left\{j\theta\right\} \tag{113}$$

where

$$A = \sqrt{I^2 + Q^2} \tag{114a}$$

and

$$\theta = \tan^{-1}(Q/I) \tag{114b}$$

Since by Eqs. (111), (112), I and Q are sums of many random variables, we may assume by the central limit theorem (e.g., [4]) that I and Q are identically Gaussian. Thus

$$A = \sqrt{I^2 + Q^2} \tag{115}$$

is Rayleigh and

$$A^2 = I^2 + Q^2$$

is exponential, with mean equal to standard deviation. Thus the intensity A^2 of the resolution cell is a random variable depending on look angle, slant range, etc. The mean, $E(A^2)$, would presumably provide a better estimate of the "true" reflectivity of the resolution cell. However, we only have one sample value of A^2 from our correlation processing, and the signal-to-noise ratio of the exponentially distributed random variable A^2 is 0 dB. To improve the situation, we may try to obtain more samples for A^2. The interested reader is referred to the excellent paper, [5], for further details.

B. Focusing

Both the range and azimuth correlation and signal functions were of the form (Sections II, III)

$$s(t) = \exp\left\{ j \frac{at^2}{2} \right\}$$

where we choose "a" to match the return signal value. In practice, especially for the azimuth processing, it is difficult to know the true signal value of "a" because it depends on orbital information (cf. Eq. (93), for example). Mismatch between the correlation value "a" and the signal value "a" results in

degradation of the impulse response Eq. (96). There exist various feedback schemes to iteratively refine the value of "a" based on perceived image degradation. The interested reader is referred to [6] for further details on these "automatic focusing" ("autofocusing") procedures.

C. Interpolation

As in Fig. 27, it is necessary to take data values along a curve for the azimuth correlation. Rarely does a sampled grid of SAR data contain exactly the right points. The result of using samples offset from the peaks in Fig. 27 is amplitude modulation of the signal resulting in increased sidelobe levels in the impulse response (Eq. (96)) and consequent image quality degradation. The interested reader may refer to [7], [8] for more information on interpolation as a possible solution.

D. Roundoff Error

Digital processing is of course performed in a processor with finite word lengths (and all the attendant dynamic range and accuracy limitations). Further details and problem solutions are found in [1] and [9].

E. Sidelobes and Weighting

Because of the form of the impulse response in Eq. (96), image sidelobes may sometimes be mistaken for targets. More often they will limit the radar calibration accuracy. One solution is to "weight" the correlation functions. The sidelobes are thus reduced but the impulse response mainlobe is broadened. For further background, see [10], Chapter 7, and [11], [12].

REFERENCES

1. Oppenheim, A., and Schafer, R., Digital Signal Processing, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1975.
2. Ahmed, N., and Rao, K., Orthogonal Transforms for Digital Signal Processing, Springer-Verlag, New York, 1975.
3. Rabiner, L., and Gold, B., Theory and Application of Digital Signal Processing, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1975.
4. Freund, J., Mathematical Statistics, Second Edition, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1971.
5. Butman, S., and Lipes, R., "The Effect of Noise and Diversity on Synthetic Array Radar Imagery," in The Deep Space Network Progress Report 42-29, the Jet Propulsion Laboratory, Pasadena, Ca., Oct. 15, 1975, NASA Code 310-20-67-08.
6. Di Cenzo, A., "Autofocusing and Multilook Misregistration Wher the Planet is Rotating Beneath the Spacecraft," Jet Propulsion Laboratory Interoffice Memorandum 334.4-80-028 to D. Held, March 25, 1980. (JPL internal document.)
7. Korwar, V., "The Effect of Range Sampling Rate on SAR Processed Images When Range Migration is Important," Jet Propulsion Laboratory Interoffice Memorandum 334.4-095-78 to C. Wu, R. Piereson, R.M. Goldstein and J.R. Pierce, September 20, 1978. (JPL internal document.)
8. Wu, C., "Range Interpolation in the DMSP Azimuth Processing," Jet Propulsion Laboratory Interoffice Memorandum 3621-77-037 from C. Wu to R. Piereson, August 15, 1977. (JPL internal document.)

9. Di Cenzo, A., "Dynamic Range and Digital SAR Processing," Jet Propulsion Laboratory Interoffice Memorandum 334.4-242 to R. Piereson, C. Bode, November 9, 1979. (JPL internal document.)

10. Cook, C., and Bernfield, M., Radar Signals, an Introduction to Theory and Application, Academic Press, New York, 1967.

11. Di Cenzo, A., "Weighting and Digital Synthetic Aperture Radar Processing," Publication 79-94. Jet Propulsion Laboratory, Pasadena, California, November 15, 1979.

12. Di Cenzo, A., "Effect of Weighting on Time Sidelobe Suppression," NASA - New Mexico Joint SAR Technology Conference Proceedings, Las Cruces, New Mexico, March 1978.