NEAR EAST UNIVERSITY

Faculty of Engineering

Department of Electrical and Electronic Engineering

PHASED ARRAY RADAR

Graduation Project
EE- 400

Student: Nidal Aabed (992186)

Supervisor: Assoc. Prof. Dr. Sameer Ikhdair

Nicosia - 2002

ACKNOWLEDGEMENT

First of all, my supervisor Prof. Dr. Sameer Ikhdair how a great person you are, you were always helping us and answer our questions even the trivial ones. I would like to thank my teachers in Electrical and electronic Department especially Mr.Özgür Özerdem, Dr.Kadri Bürüncük, Mr.Mehrdad Khaledi.

Secondly, I would like to thank my parents for their moral and finical support as well as encouragement. Without you I would never ever reach this point.

I dedicate this work for the martyrs in my country Palestine wishing that we will reach independence and satisfaction.

Many thanks to my friends who I will never forget them especially wesam abu rajab,ehab and wael ashour,ahmed wadi, sameh mashaqi and ahmed qatanani.

Finally thanks for my University Near East University for everything.

TABLE OF CONTENTS

CHAPTER ONE

INTRODUCTION

Radar is an instrument that operates by transmitting an electromagnetic radio wave toward some object. This wave is reflected by the object and some of the reflected wave is received back by the antenna of the radar. Receiving this wave means making detection, or in other words, it means detecting that some object stands in the direction of propagation of the wave. Radar system or technique for detecting the position, movement, and nature of a remote object by means of radio waves reflected from its surface.

Radar involves the transmission of pulses of electromagnetic waves by means of a directional antenna; some of the pulses are reflected by objects that intercept them. The reflections are picked up by a receiver, processed electronically, and converted into visible form by means of a cathode-ray tube. The range of the object is determined by measuring the time it takes for the radar signal to reach the object and return. The object's location with respect to the radar unit is determined from the direction in which the pulse was received. The information secured by radar includes the position and velocity of the object with respect to the radar unit. In some advanced systems the shape of the object may also be determined.

Phased arrays get around the beam width limit by using several small apertures to achieve the same result as one large aperture. radio telescopes can operate to detect signals from the sky or to broadcast terrestrial messages. While the mechanics of the actual transmitters and receivers may differ, the science here is the same. While a linear array produces a narrow *fan beam* of radiation, a plane array can be created to emit a thin *pencil beam*.

In practice, arrays are often placed in more complicated arrangements. Arrays may be multi-dimensional. Some may feature irregular element separation. The mathematics for arbitrary arrangements becomes complicated quickly and won't be dealt with here. Suffice it to say that with careful numerical modeling, arrays can be designed to fulfill a wide variety of purposes. Upcoming sections will focus on some applications of acoustic arrays. Optical telescopes yield for one basic reason; the wavelength of light they deal with is tiny compared to the diameter of the aperture it must pass through (the main lens/mirror of the device).

Since many of the specific characteristic properties are not comparable to those of standard ultrasonic equipments and -probes and are in addition not common within the medical diagnostic field one has to introduce some procedures specific for phased array systems in NDT .The reliability and reproduceability of inspections with such complex systems are much more depending on the correct functioning and the correct calibration of the system then any other ultrasonic inspection equipment. Since the calibration procedures are in general not sensitive enough for the detection of all parameter deviations from the their standard values, it is nescessary to consider special steps and tools for the measurement of the most important probe-and equipment characteristics

Maximum cable length between the probe and the equipment. Number of transmitter and receiver channels, number of probe channels. Sample rate, resolution in bits, maximal and minimal dynamic range, maximal pulse repetition rate of the the A-scan conversion and data storage. amplitude transfer behavior (linearity), dynamic range in dependency on the amplification calibration and adjustment, frequency transfer behavior (bandwidth).Crosstalk between different element channels at the probes and at the transmitter/receiver channels within the equipment. Signal to noise ratio for a given target and at a given distance. Software depending parameters like angles, angle beam divergence, focusing. Sensitivity variations due to changes of the skewing- or incidence angles and of the focusing. The potential for increased target handling capacity available in Track While Scan radars is limited by the requirement to position the radar antenna mechanically.

Existing mechanical scanning methods are inherently slow and require large amounts of power in order to respond rapidly enough to deal with large numbers of high speed maneuvering targets. With mechanically scanned systems, antenna inertia and inflexibility prevent employment of optimum radar beam positioning patterns that can reduce reaction times and increase target capacity. With electronic scanning, the radar beams are positioned almost instantaneously and completely without the inertia, time lags, and vibration of mechanical systems. In an era in which the numerical superiority of adversaries is expected to remain large, electronic scanning can offset that advantage.

The fundamental principles underlying the concept of electronic beam steering are derived from electromagnetic radiation theory employing constructive and destructive interference. The electromagnetic energy received at a point in space from two or more closely spaced radiating elements is a maximum when the energy from each radiating element arrives at the point in phase.

All elements are radiating in phase, and the resultant wave front is perpendicular to the axis of the element array. And subsequent diagrams show only a limited number of radiating elements. In actual radar antenna design several thousand elements could be used to obtain a high-gain antenna with a beam width of less than two degrees. The wave fronts remain perpendicular to the bore sight axis and are considered to arrive at a point target in space at the same time. The path lengths from the elements to point *P* equalize as *P* approaches infinity. Thus, in situations where the target range is very large compared to the distance between elements, the paths from the elements to point *P* are almost parallel. Under these conditions, energy will arrive at point *P* with the same phase relationship that existed at the array. To achieve beam positioning off the bore sight axis, it is necessary to radiate the antenna elements out of phase with one another.

In a phase-scanned radar system, the radiating elements are fed from a radar transmitter through phase-shifting networks or "phasers." The aim of the system is again to position the beam at any arbitrary angle, at any time. In this case the means of accomplishing the phase shift at each element is simply to shift the phase of the incoming energy to each element.

These phasers are adjustable over the range 0 to + 2 radians. No matter which of the three possible methods of phase scanning is used in a phased array system, the objective is a relative phase shift of the energy being radiated by each element in the array.

In practice there are limits to the useful angular displacement of an electronically scanned radar beam. One limit is caused by the element pattern. The antenna pattern of an array is the product of the array pattern and the element pattern. In the simple examples given in this section we have assumed that the element pattern was omnidirectional. A practical array element pattern is not omnidirectional, so the elements limit the scan angle. Another limit is caused by the element spacing. A large scan angle requires a close element spacing. If the scan angle exceeds that which can be accommodated by the element spacing, grating lobes will be formed in the other direction.

In my project, I will discuss the main purpose of phase array radar and the most common components which are included As well as the applications of phase array radar. And also the principle operation of radar and phase array.

CHAPTER TWO

PRINCIPLES OF RADAR

2.1 History of Radar

Radar and penicillin actually have something in common- They were both discovered by accident. Radar, which stands for Radio Detection and Ranging, was developed for military purposes during World War II. The British and US Military used Radar to locate ships and airplanes. However, annoying blips were consistently appearing on the radar screen. It turned out; these annoying radar returns were raindrops.

Well someone saw this hindrance as a wonderful opportunity. In 1957, the US Government created the WSR-57 (weather surveillance, 1957) which became the primary radar for the weather service for nearly 40 years. Advances in technology helped usher in the WSR-88D, D for Doppler. Also known as NexRad for next Generation Radar, this is what the meteorologists currently use to help save lives and predict your weather. The NEXRAD has 750,000 Watts of power and a 460 Km range. Most notably, improved radar has allowed meteorologists to see wind fields, determine precipitation rates and most importantly identify potential tornadic cells.

2.2 Radar System

Radar (Radio Detection and Ranging) is an instrument that operates by transmitting an electromagnetic radio wave toward some object. This wave is reflected by the object and some of the reflected wave is received back by the antenna of the radar. Receiving this wave means making detection, or in other words, it means detecting that some object stands in the direction of propagation of the wave. Radar system or technique for detecting the position, movement, and nature of a remote object by means of radio waves reflected from its surface. Although most radar units use microwave frequencies, the principle of radar is not confined to any particular frequency range. There are some radar units that operate on frequencies well below 100 megahertz (megacycles) and others that operate in the infrared range and above.

The term *radar,* an acronym for *ra*dio *d*etection *a*nd *r*anging, is also used to denote the apparatus for implementing the technique. Radar, was originally used as a detection device in military aircraft during World War II. It was then modified after the war for meteorological use in the weather community. The physical structure of snow makes it more difficult to track on conventional radar than other types of precipitation. Snow flakes have lower moisture and higher air content than other types of precipitation. This quality is physically apparent in the holes found in a snowflake's structure. Radar can detect many properties of the target object: first of all, its distance, which is proportional to the delay time of the reception of the reflected signal.

Furthermore, radar can detect many other properties of the target object by comparing the characteristics of the emitted and received signals. In fact, the beam initially emitted by radar is somewhat like the beam of a laser light: it has a single frequency and it is *coherent* (meaning that the phase of the beam is the same across the entire wavefront). One of the greatest advantages is that the beam's characteristics (such as its frequency, its duration and its strength) are perfectly known and can be compared to the characteristics of the reflected signal. If, for example the frequency of the two signals is different, this means that the detected object is moving in the direction of the observer (this is due to a Doppler effect).

Radar involves the transmission of pulses of electromagnetic waves by means of a directional antenna; some of the pulses are reflected by objects that intercept them. The reflections are picked up by a receiver, processed electronically, and converted into visible form by means of a cathode-ray tube. The range of the object is determined by measuring the time it takes for the radar signal to reach the object and return. The object's location with respect to the radar unit is determined from the direction in which the pulse was received. In most radar units the beam of pulses is continuously rotated at a constant speed, or it is scanned (swung back and forth) over a sector, also at a constant rate. The velocity of the object is measured by applying the Doppler principle: if the object is approaching the radar unit, the frequency of the returned signal is greater than the frequency of the transmitted signal; if the object is receding from the radar unit, the returned frequency is less; and if the object is not moving relative to the radar unit, the return signal will have the same frequency as the transmitted signal.

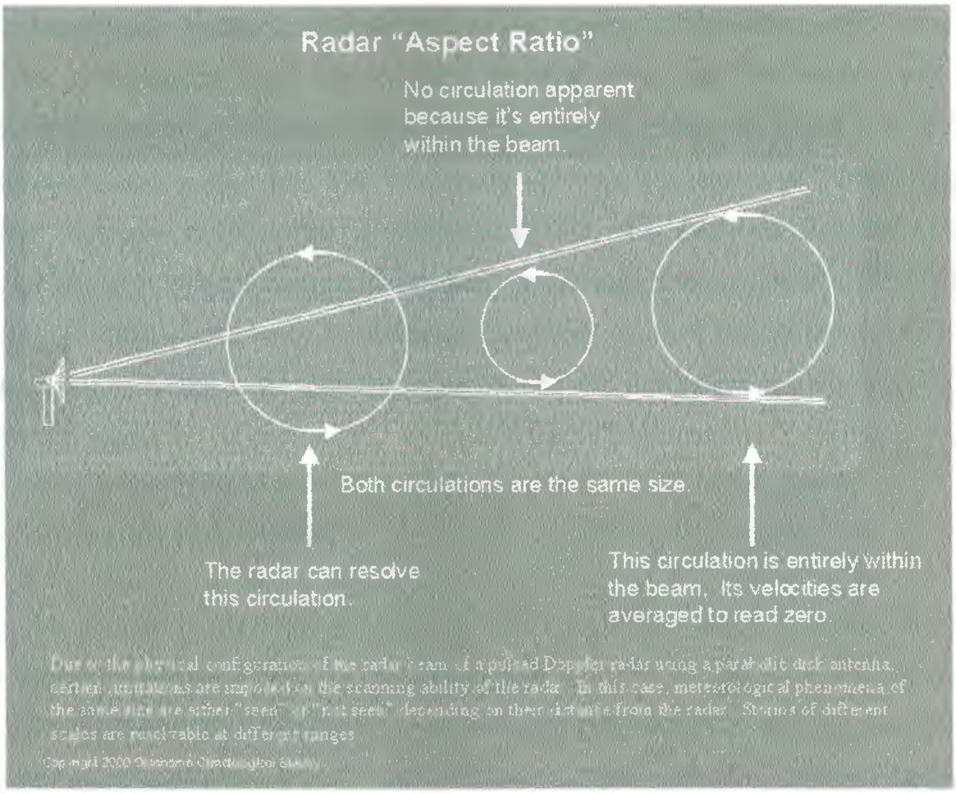

Figure 2.1 Circulation of radar

System or technique for detecting the position, movement, and nature of a remote object by means of radio waves reflected from its surface. Although most radar units use microwave frequencies, the principle of radar is not confined to any particular frequency range. There are some radar units that operate on frequencies well below 100 megahertz (megacycles) and others that operate in the infrared range and above. The term *radar,* an acronym for *ra*dio *d*etection *a*nd *r*anging, is also used to denote the apparatus for implementing the technique.

2.3 Principles of Radar

Radar involves the transmission of pulses of electromagnetic waves by means of a directional antenna; some of the pulses are reflected by objects that intercept them. The reflections are picked up by a receiver, processed electronically, and converted into visible form by means of a cathode-ray tube. The range of the object is determined by measuring the time it takes for the radar signal to reach the object and return.

The object's location with respect to the radar unit is determined from the direction in which the pulse was received. In most radar units the beam of pulses is continuously rotated at a constant speed, or it is scanned (swung back and forth) over a sector, also at a constant rate. The velocity of the object is measured by applying the Doppler principle: if the object is approaching the radar unit, the frequency of the returned signal is greater than the frequency of the transmitted signal; if the object is receding from the radar unit, the returned frequency is less; and if the object is not moving relative to the radar unit, the return signal will have the same frequency as the transmitted signal.

2.4 Applications of Radar

The information secured by radar includes the position and velocity of the object with respect to the radar unit. In some advanced systems the shape of the object may also be determined. Commercial airliners are equipped with radar devices that warn of obstacles in or approaching their path and give accurate altitude readings. Planes can land in fog at airports equipped with radar-assisted ground-controlled approach (GCA) systems, in which the plane's flight is observed on radar screens while operators radio landing directions to the pilot. A ground-based radar system for guiding and landing aircraft by remote control was developed in 1960. Radar is also used to measure distances and map geographical areas (shoran) and to navigate and fix positions at sea. Meteorologists use radar to monitor precipitation; it has become the primary tool for short-term weather forecasting and is also used to watch for severe weather such as thunderstorms and tornados. Radar can be used to study the planets and the solar ionosphere and to trace solar flares and other moving particles in outer space. Various radar tracking and surveillance systems are used for scientific study and for defense.

For the defense of North America the U.S. government developed (c.1959–63) a radar network known as the Ballistic Missile Early Warning System (BMEWS), with radar installations in Thule, Greenland; Clear, Alaska; and Yorkshire, England. A radar system known as Space Detention and Tracking System (SPADATS), operated collaboratively by the Canada and the U.S., is used to track earth-orbiting artificial satellites

2.5 Stealth Technology

Designs and materials engineered for the military purpose of avoiding detection by radar or any other electronic system. Stealth, or antidetection, technology is applied to vehicles (e.g., tanks), missiles, ships, and aircraft with the goal of making the object more difficult to detect at closer and closer ranges. Since radar is the most difficult form of detection to elude, avoidance is generally accomplished by reducing the radar cross section (RCS) of the object to within the level of background noise; for example, the reported goal of U.S. military designers is to make a fighter plane with an RCS the size of a bird. The RCS is the area of an imaginary perfect reflector that would reflect the same amount of energy back to the receiving radar antenna as does the actual target, which may be much larger or even smaller than the RCS. A pickup truck, for example, with its flat surfaces and sharp edges has an RCS of approximately 200 sq m, but a smooth-edged fighter jet has an RCS of only 2 to 4 sq m. The RCS of any given object, however, differs at various angles and radar frequencies. Much about stealth technology remains classified, but among the antidetection techniques used in the U.S. Air Force F-117 Stealth fighter plane (which probably has an RCS of 1 sq m or less) are a low profile with no flat surfaces to reflect radar directly back, the intensive substition of radar opaque composites in place of metals, and an overall coating of radar absorbing material. The implementation of stealth technology usually requires such compromises as reduced payload capacity, aerodynamic instability, and high design, production, and maintenance expenses

2.6 Development of Radar

Radar was developed (c.1935–40) independently in several countries as a military instrument for detecting aircraft and ships. One of the earliest practical radar systems was devised (1934–35) by Sir Robert Watson-Watt, a Scots physicist. Although the technology evolved rapidly during World War II, radar improved immensely following the war, the principal advances being higher power outputs, greater receiver sensitivity, and improved timing and signal-processing circuits. In 1946 radar beams from the earth were reflected back from the moon. Radar contact was established with Venus in 1958 and with the sun in 1959, thereby opening a new field of astronomy—radar astronomy.

2.7 Radar basics

To understand how radar detectors work, you first have to know what they're detecting. The concept of measuring vehicle speed with radar is very simple. A basic speed gun is just a radio transmitter and receiver combined into one unit. A radio transmitter is a device that oscillates an electrical current so the voltage goes up and down at a certain frequency. This electricity generates electromagnetic energy, and when the current is oscillated, the energy travels through the air as an electromagnetic wave. A transmitter also has an amplifier that increases the intensity of the electromagnetic energy and an antenna that broadcasts it into the air. A radio receiver is just the reverse of the transmitter: It picks up electromagnetic waves with an antenna and converts them back into an electrical current. At its heart, this is all radio is -- the transmission of electromagnetic waves through space.

Radar is the use of radio waves to detect and monitor various objects. The simplest function of radar is to tell you how far away an object is. To do this, the radar device emits a concentrated radio wave and listens for any echo. If there is an object in the path of the radio wave, it will reflect some of the electromagnetic energy, and the radio wave will bounce back to the radar device. Radio waves move through the air at a constant speed (the speed of light), so the radar device can calculate how far away the object is based on how long it takes the radio signal to return. Radar can also be used to measure the speed of an object, due to a phenomenon called Doppler shift. Like sound waves, radio waves have a certain frequency, the number of oscillations per unit of time. When the radar gun and the car are both standing still, the echo will have the same wave frequency as the original signal. Each part of the signal is reflected when it reaches the car, mirroring the original signal exactly. But when the car is moving, each part of the radio signal is reflected at a different point in space, which changes the wave pattern. When the car is moving away from the radar gun, the second segment of the signal has to travel a greater distance to reach the car than the first segment of the signal. As you can see in the diagram below, this has the effect of "stretching out" the wave, or lowering its frequency. If the car is moving toward the radar gun, the second segment of the wave travels a shorter distance than the first segment before being reflected. As a result, the peaks and valleys of the wave get squeezed together: The frequency increases.

Based on how much the frequency changes, a radar gun can calculate how quickly a car is moving toward it or away from it. If the radar gun is used inside a moving police car, its own movement must also be factored in. For example, if the police car is going 50 miles per hour and the gun detects that the target is moving away at 20 miles per hour, the target must be driving at 70 miles per hour. If the radar gun determines that the target is not moving toward or away from the police car, than the target is driving at exactly 50 miles per hour. Police officers have been catching speeders this way for more than 50 years. Recently, many police departments have added a new sort of speed detector, one that uses light instead of radio waves. In the next section, we'll see how this cutting edge devices work.

The lidar gun clocks the time it takes a burst of infrared light to reach a car, bounce off and return back to the starting point. By multiplying this time by the speed of light, the lidar system determines how far away the object is. Unlike traditional police radar, lidar does not measure change in wave frequency. Instead, it sends out many infrared laser bursts in a short period of time to collect multiple distances. By comparing these different distance samples, the system can calculate how fast the car is moving. These guns may take several hundred samples in less than half a second, so they are extremely accurate.

2.8 Smile for the Camera

Police may use handheld lidar systems, just like conventional radar guns, but in many areas, the lidar system is completely automated. The gun shines the laser beam at an angle across the road and registers the speed of any car that passes by (the system makes a mathematical adjustment to account for the angle of view). When a speeding car is detected, the system triggers a small camera, which takes a picture of the car's license plate and the driver's face. Since the automated system has collected all of the evidence the police need, the central office simply issues a ticket and sends it to the speeder in the mail. In the next sections, we'll see how detector devices help speeders evade radar and lidar speed traps. We'll also find out what the police can do to figure out who's using a radar detector.

2.9 Picking Up Signals

In the previous sections, we saw how police use traditional radar as well as new laser technology to catch drivers speeding. As it turns out, conventional radar is relatively easy to detect. The simplest radar detector is just a basic radio receiver, something like the one you use to pick up FM and AM radio stations. The air is full of radio signals -- they're used for everything from television broadcasts to garage door openers -- so for a receiver to be at all useful, it must pick up only signals in a certain range. The receiver in a radio is designed to pick up signals in the AM and FM frequency spectrum, whereas the receiver in a radar detector is tuned to the frequency range used by police radar guns. Periodically, the frequency range used by the police is expanded, and speedster's everywhere a basic radar detector won't do you much good if the police officer drives up behind you and turns on the radar gun. The detector will alert you, but by that time, the officer already has all the information he or she needs. In many cases, however, detectors pick up the signal before the speeding car can be tracked. Police often leave their radar guns turned on for a long period of time, instead of activating them after sneaking up behind a car.

Radar guns have a cone- or dish-shaped antenna that concentrates the radio signal, but the electromagnetic wave quickly spreads out over a wide area. The radar gun is configured so that it only monitors the speed of a particular target, not everything in the vicinity, so chances are a detector will pick up the radio signal well before the radar gun recognizes the car. Of course, with this sort of detector, you're relying mostly on the luck of the draw -- if the police officer decides to target you before any other car, you're caught. Modern detectors offer much more extensive protection for speeders.

2.10 Jamming Signals

In the last section, we looked at conventional radar detector, which pick up police radar with a simple radio receiver. This sort of detector is a completely passive device: It simply recognizes the presence of radar. More sophisticated detectors actually take an active role in eluding the police.

In addition to the basic receiver, these devices have their own radio transmitter, which emits a jamming signal. Essentially, the signal replicates the original signal from the police radar gun, but mixes it with additional radio noise. With this information added, the radar receiver gets a confusing echo signal, and the police can't make an accurate speed reading.

Modern detectors may also include a light-sensitive panel that detects the beams from lidar guns. These devices are more difficult to evade than traditional radar because the beam is much more focused and it doesn't carry well over long distances. By the time a detector recognizes the presence of the laser beam, the car is most likely in the beam's sights already. Some speeders try to get around these systems by reducing the reflectivity of their car. A black surface reduces reflectivity because it absorbs more light. Drivers can also get special plastic covers that reduce the reflectivity of license plates. These measures reduce the effective range of the lidar system, but not the range of the driver's detector. With this extra time, a speeder might be able to slow down before the lidar gun can get a read on his or her speed. Speeders may also use a laser jammer. This works basically the same way as a radar jammer. In addition to a light-sensitive panel, the detector has its own built-in light-emitting diodes (LEDs) that produce a light beam of their own. When this beam shines on the lidar system, the receiver can't recognize any reflected light and so can't get a clear speed reading.

It's important to note that none of these systems are 100-percent effective; even with a top-of-the-line detection and jamming system, the police still might catch you speeding. Also, since police periodically introduce new speed-monitoring technology, a detector might suddenly become outdated. Whenever this happens, the fully equipped speeder has to dump everything and pick up all new equipment. Of course, there is always one surefire way you can avoid speeding tickets, no matter what technology the police come up with.

2.11 Synthetic Aperture Radar

Environmental monitoring, earth-resource mapping, and military systems require broad-area imaging at high resolutions. Many times the imagery must be acquired in inclement weather or during night as well as day. Synthetic Aperture Radar (SAR) provides such a capability. SAR systems take advantage of the long-range propagation characteristics of radar signals and the complex information processing capability of modern digital electronics to provide high resolution imagery. Synthetic aperture radar complements photographic and other optical imaging capabilities because of the minimum constraints on time-of-day and atmospheric conditions and because of the unique responses of terrain and cultural targets to radar frequencies.

Synthetic aperture radar technology has provided terrain structural information to geologists for mineral exploration, oil spill boundaries on water to environmentalists, sea state and ice hazard maps to navigators, and reconnaissance and targeting information to military operations. There are many other applications or potential applications. Some of these, particularly civilian, have not yet been adequately explored because lower cost electronics are just beginning to make SAR technology economical for smaller scale uses. Sandia has a long history in the development of the components and technologies applicable to Synthetic Aperture Radar -- 40 years in radar, antenna, and miniature electronics development; 30 years in microelectronics; and 25 years in precision navigation, guidance, and digital-signal processing. Over the last decade, we have applied these technologies to imaging radars to meet the needs of advanced weapon systems; verification and nonproliferation programs; and environmental applications.

Sandia's expertise in electromagnetics, microwave electronics, high-speed signal processing, and high performance computing and navigation, guidance and control have established us as world leaders in real-time imaging, miniaturization, processing algorithms, and innovative applications for SAR. A detailed description of the theory of operation of SAR is complex and beyond the scope of this document. Instead, this page is intended to give the reader an intuitive feel for how synthetic aperture radar works. Consider an airborne SAR imaging perpendicular to the aircraft velocity as shown in the figure below. Typically, SARs produce a two-dimensional (2-D) image.

One dimension in the image is called range (or cross track) and is a measure of the "line-of-sight" distance from the radar to the target. Range measurement and resolution are achieved in synthetic aperture radar in the same manner as most other radars: Range is determined by precisely measuring the time from transmission of a pulse to receiving the echo from a target and, in the simplest SAR, range resolution is determined by the transmitted pulse width, i.e. narrow pulses yield fine range resolution.

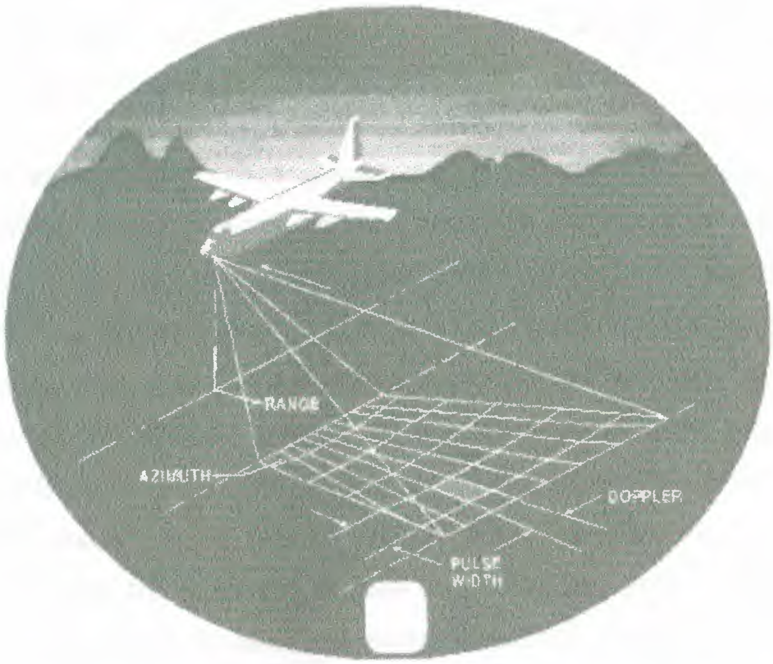

Figure 2.2 Synthetic Aperture Radar Imaging Concepts

The other dimension is called azimuth (or along track) and is perpendicular to range. It is the ability of SAR to produce relatively fine azimuth resolution that differentiates it from other radars. To obtain fine azimuth resolution, a physically large antenna is needed to focus the transmitted and received energy into a sharp beam. The sharpness of the beam defines the azimuth resolution. Similarly, optical systems, such as telescopes, require large apertures (mirrors or lenses which are analogous to the radar antenna) to obtain fine imaging resolution. Since SARs are much lower in frequency than optical systems, even moderate SAR resolutions require an antenna physically larger than can be practically carried by an airborne platform: antenna lengths several hundred meters long are often required. However, airborne radar could collect data while flying this distance and then process the data as if it came from a physically long antenna.

The distance the aircraft flies in synthesizing the antenna is known as the synthetic aperture. A narrow synthetic beam width results from the relatively long synthetic aperture, which yields finer resolution than is possible from a smaller physical antenna. Achieving fine azimuth resolution may also be described from a doppler processing viewpoint. A target's position along the flight path determines the doppler frequency of its echoes: Targets ahead of the aircraft produce a positive doppler offset; targets behind the aircraft produce a negative offset. As the aircraft flies a distance (the synthetic aperture), echoes are resolved into a number of doppler frequencies. The target's doppler frequency determines its azimuth position.

While this section attempts to provide an intuitive understanding, SARs are not as simple as described above. Transmitting short pulses to provide range resolution is generally not practical. Typically, longer pulses with wide-bandwidth modulation are transmitted which complicates the range processing but decreases the peak power requirements on the transmitter. For even moderate azimuth resolutions, a target's range to each location on the synthetic aperture changes along the synthetic aperture. The energy reflected from the target must be "mathematically focused" to compensate for the range dependence across the aperture prior to image formation. Additionally, for fine-resolution systems, the range and azimuth processing are coupled (dependent on each other) which also greatly increases the computational processing.

CHAPTER THREE

APPLICATIONS OF PHASED ARRAY TECHNOLOGY

3.1. Overview

The angular resolution (beam width) of a detector is proportional to the wavelength divided by the aperture diameter. Hence, for longer wavelengths, a different technology must be used for higher resolution. Phased arrays synthesize larger apertures from an array of elements. The basic theory is explained with 2-element arrays and then expanded with an N-element linear array. Beam-steering and nulling are briefly discussed. This technology is widely used in radar and radio-astronomy, however, in this paper, I present a number of acoustic applications of phased arrays including side-scan sonar, passive sea floor detectors, and towed arrays. Keywords: Phased Arrays, Sonar, Beam forming, Acoustic Technology.

3.2. Aperture Theory

3.2.1. The high School Approach:

Optical telescopes yield for one basic reason; the wavelength of light they deal with is tiny compared to the diameter of the aperture it must pass through (the main lens/mirror of the device). However, when longer wavelengths are used, resolution becomes increasingly worse. The angular resolution θ is written as

$$\theta=1.22\lambda/D \tag{3.1}$$

where D is the diameter of the aperture and λ is the wavelength of light being considered. The 1.22 factor arises from the circular aperture and will be explained shortly.

3.2.2. The Fourier approach

If we consider a point source represented by a Dirac delta function δ(x,y) and view it through an circular aperture of radius a, f(x,y), the observed pattern will be the square of the Hankel transform of the convolution of the source and aperture functions or

$$
\begin{aligned}
f(x,y) &= (\pi a^2)^{-1} \ \mathrm{rect}(r/2a) \\
h(r) &= (\pi a^2)^{-1} \ \mathrm{rect}(r/2a) * \delta(r) \qquad (3.2) \\
H(q) &= FT(h(r)) \\
&= FT((\pi a^2)^{-1} \ \mathrm{rect}(r/2a)).FT(\delta(r)) \\
&= \frac{4}{\pi} jinc(2aq) = \frac{4}{\pi} \frac{J_1(2\pi aq)}{aq} \qquad (3.3)
\end{aligned}
$$

The observed intensity at the detector is then $(H(q))^2$ as represented in figure 1. The location of the first minimum is $\pm$0.61 from the axis, half the 1.22 seen in fig.2.1.

This is the *beam width* so important for radio astronomy, radar, sonar, and other applications. Notice that the strength of the *side lobes* decreases with angle.

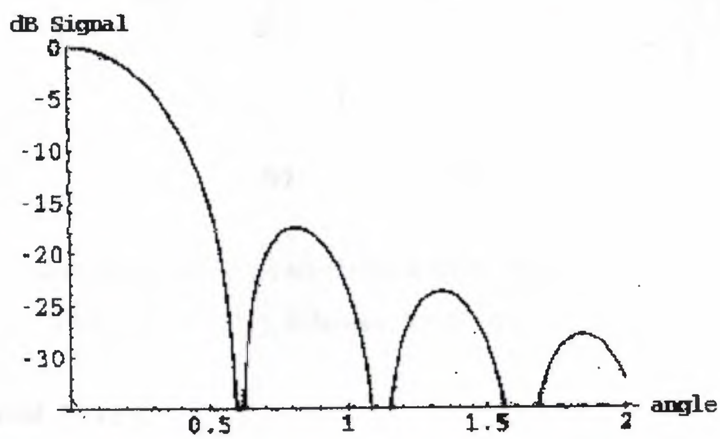

Figure 3.1 Aperture function from a uniform pinhole.

3.2.3. Obtaining Better Resolutions

From fig.2.1, we see that optical images with one arc-second resolution/beam width are possible using a 10cm perfect telescope. However, for the same resolution with microwaves, would require a telescope 250 meters across, something very difficult to engineer. There is a better way to improve resolution.

3.3. Phased Array

Phased arrays get around the beam width limit by using several small apertures to achieve the same result as one large aperture. The most dramatic example of this may be the Very Long Baseline Interferometer, a series of 25m radio telescope spaced all over the earth operating at radio wavelengths. By themselves they sport mediocre beam widths insufficient for scientific studies. But when linked together as an *array* the size of earth, they have a beam width of milliarcseconds, far better than any current optical instrument. Before we explore the workings of a phased array, it should be noted that the equations and science here makes no distinction between receiving and transmitting devices. For instance, radio telescopes can operate to detect signals from the sky or to broadcast terrestrial messages. While the mechanics of the actual transmitters and receivers may differ, the science here is the same.

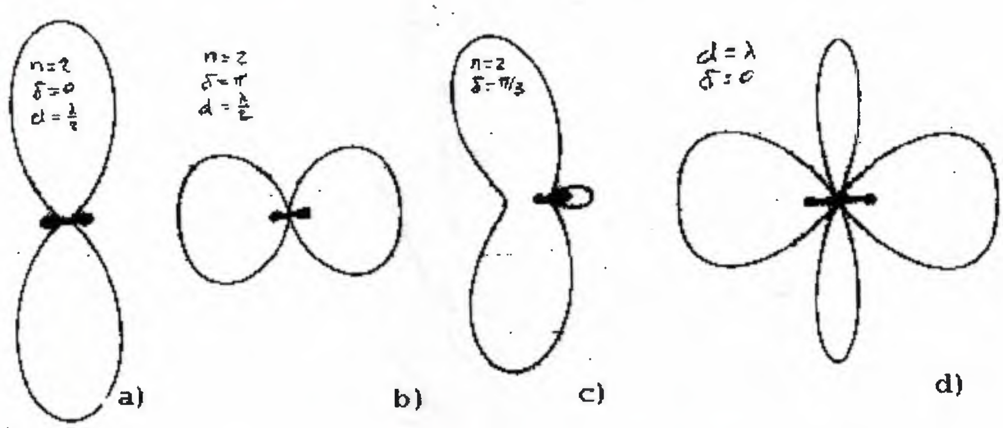

Figure 3.2 Power patterns from a two-element array. Broadside (a), endfire (b), arbitrary phase (c), full-wavelength spacing (d).

3.3.1. 2-element Array:

Let us examine the simplest case of a 2-element array. Assume we are dealing with devices much smaller than the wavelength of interest and that each broadcasts isotropically.

Let us also assume that all measurements are made in the far-field. The power radiated at a given angular position is the square of the sum of the fields put out by the two elements

$$E(\theta) = \exp(i\pi \frac{d}{\lambda}\cos\theta + \delta) + \exp(-i\pi \frac{d}{\lambda}\cos\theta + \delta) \tag{3.4}$$

$$\mathrm{I}(\theta) = \mathrm{E}(\theta)^2 = \cos(\pi \frac{d}{\lambda}\cos\theta + \delta)^2 \tag{3.5}$$

where θ first, is the viewing angle with respect to the array axis, d is the separation between the array elements, and δsecond, is the relative phase of the two emitters.Equation.4. generates an intensity pattern with two lobes, the *main lobe* and *back lobe*. By varying the element separation and phase, a number of interesting patterns can be produced (figure 3.2). The case where the main lobe projects perpendicular to the line between the elements is known as *broadside* whereas the on-axis case is referred to as *end fire*. It should be noted that the beam produced by a linear array (of which two elements is the simplest case) is a conical one. The actual radiation pattern in 3-dimensional space is the solid of rotation.

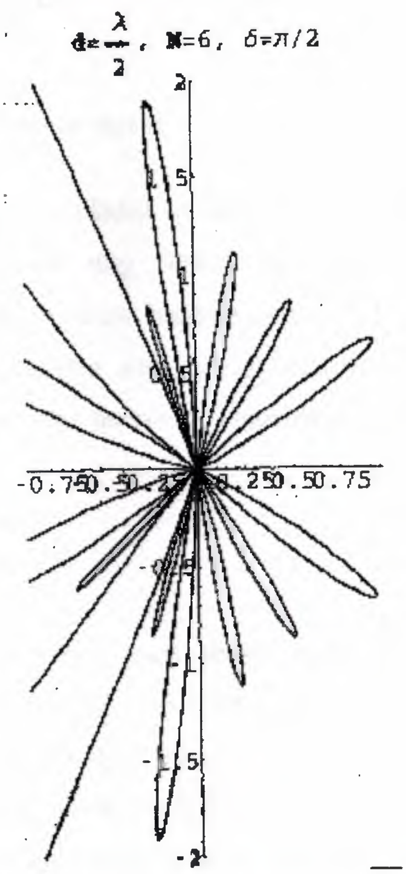

Figure 3.3 Power pattern from a 7-element evenly spaced, unshaded array set at phase.

3.2.4. N-element Array:

Now let's generalize to a linear array of N elements all spaced d/λ apart. (Figure 3.3) One way to quantify this array is as a series of 2-element arrays .... (1,2)+(1,3)+...+(1,*N*) and sum the output (Kraus 1988).

$$E(\theta) = \sum_{n=0}^{N-1} \exp(in\frac{2\pi d}{\lambda}\cos\theta + in\delta) \qquad (3.6)$$

With N=2, Eq.6. reduces to fig3..4. The more elements used, the sharper the main beam and back lobe will be. The usual games may be played with the array spacing.

3.2.5. Two-Dimensional Arrays

Extending the above analysis to multi-dimensional arrays is simple in concept. While a linear array produces a narrow *fan beam* of radiation, a plane array can be created to emit a thin *pencil beam.*

3.2.5.1. Arrays of Arbitrary Arrangement

In practice, arrays are often placed in more complicated arrangements. Arrays may be multi-dimensional. Some may feature irregular element separation. The mathematics for arbitrary arrangements becomes complicated quickly and won't be dealt with here. Suffice it to say that with careful numerical modeling, arrays can be designed to fulfill a wide variety of purposes. Upcoming sections will focus on some applications of acoustic arrays.

3.2.6. Aperture Shading:

Recalling the Fourier approach used above, recall that we used a ``square" aperture with equal transmission at all points with some circle for our calculations. If instead the transmission varied over the surface of the aperture, the Fourier transform, and hence the shape of the beam would be different. Several outcomes can be obtained by skillful aperture shading. First let's examine an aperture where the transmission is``shaded" by a Gaussian function .

$$f(x,y) = e^{-\pi r^2} * rect(r)$$
$$h(r) = e^{-\pi r^2} * rect(r)$$
$$H(q) = e^{-\pi q^2} jinc(q) \tag{3.7}$$

This pattern has about the same beam width as that seen in Fig.2. The one striking diffe an aperture shaded such that it becomes a narrow annulus. rence is that the side lobes are roughly 16 decibels lower. Secondly, let's examine .

$$f(x,y) = e^{-\pi r^2} * rect(r)$$
$$h(r) = e^{-\pi r^2} * rect(r)$$
$$H(q) = e^{-\pi q^2} jinc(q) \tag{3.7}$$

This main beam is only 2/3 the width of the unshaded aperture but the main lobe is a full 10 decibels lower while the side lobes remain more or less unaffected. (Figure 4).

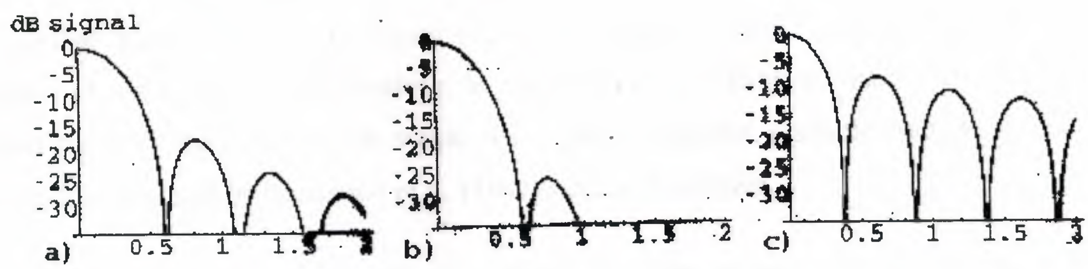

Figure 3.4 Power patterns for unshaded aperture (a), Gaussian shaded aperture (b), annulus shaded aperture (c).

Applying this technique to phased arrays involves simply giving more or less weight to individual elements. In applications where there is plenty of signal and better resolution or lower side lobes are required, shading may be a good option.

3.2.6.1. Beam Steering

Notice that by changing δ in Eq.6., the main lobe and backlobe both shift one direction or the other (figures 3.2,3.3). This represents one of the most significant advances over fixed detectors: fast beam steering. With a single dish antenna, to change the direction, motors and mechanical devices must physically turn the hardware.

With arrays, the same effect can be gotten by changing the relative phases of the different elements. This results in a lighter structure with nearly response time limited by electronic rather than mechanical factors.

The downside is that adjusting the phases is nearly always more difficult than rotating the detector. The methods of phase adjustment vary depending on the medium and the frequency being used. Optical devices can be phase-delayed by using mechanical or electronic delay lines or by passing the beams through regions of adjustable index of refraction. Acoustic devices and some lower frequency electromagnetic devices operate at low enough frequency that each element can be individually driven and the phases adjusted by software.

3.2.6.2. Noise Cancellation With Nulls

In all previous examples, all array elements have had a consistant phase relationship with their neighbors. However, just as shading offers many options in beamforming, *null steering* allows sources of noise to be filtered out. The principle of null steering is that the beam is shaped such that a source of noise coincides with a direction of very low power/sensitivity in the beam. The full-blown mechanics of detailed beamforming is beyond the scope of this paper, however, a simple example of null steering presented by Lombardo et.al. (1993) may be illustrative.

In their experiment, Lombardo et.al. (1993) used an acoustic twin-line towed sonar array. It was composed of two 5km strings of *hydrophones* separated by a distance comparable to the wavelength of the energy being studied (the low frequency sounds of submarines, in this case). With the large number of elements in each string, the mainlobe was quite sharp and could be pointed by either steering the towing ship or adjusting the phases. Unfortunately, a source located in the left beam was indistinguishable from that in the right beam.

To solve this *left-right ambiguity*, the line separation distance was adjusted such that.$\mathbf{d=\lambda/4}$ Signals arriving from the right would reach the left line a time $\mathbf{+\pi/4}$ out of phase with the right line. Signals arriving from the left would reach the left line $\mathbf{-\pi/4}$ out of phase with those from the right. When the $\mathbf{-\pi/4}$ phase is added to the data from the left line, the mainlobe will be brought into phase, while the backlobe data will sum to zero. There is slight degredation in the mainlobe strength, but the backlobe drops to -100dB (Figure 3.5).

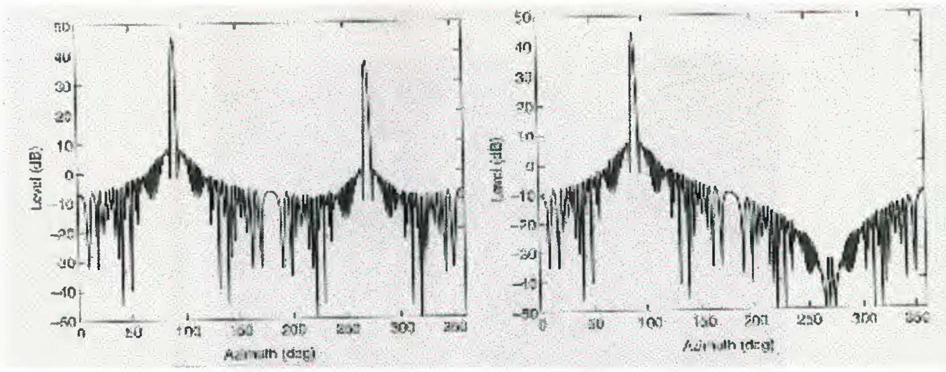

Figure 3.5 An example of nulling. Before: main lobe and back lobe are of comparable strength. After: back lobe drops by 100dB. (Lombardo et. al. 1993)

3.2.7. Acoustic Arrays

The basic sonar system is fairly straight forward. A projector is used to emit a pulse of sound energy. The accoustic waves travel outward and rebound to varying degrees from objects in the field. The reflected wave is then picked up by a hydrophone. Modern sonar systems use piezo-electric or magnetostrictive components both to produce the *ping* and measure the very small pressure variations associated with the *return.* They can be manufactured in many ways, but most are made from relatively cheap ceramic materials (Urick, 1983). This is refered to as *active sonar*. *Passive sonar* is simply the receiving half of the active system. Figure 3.6 shows a modern acoustic array.

Even from a very simple one-component *echo sounder* system such as this, a number of data can be found. From the time delay between transmission and detection, the distance to the reflecting object (henceforth the *target*) can be found given the velocity of sound in the medium (more on this later). The frequency shift of the returned signal can be used to determine the radial velocity of the target. Thirdly, the strength of the returned signal can help guess the properties of the target. Larger cross-sections will yeild larger returns, however, different materials will reflect differently. Orientation of target surfaces also matter.

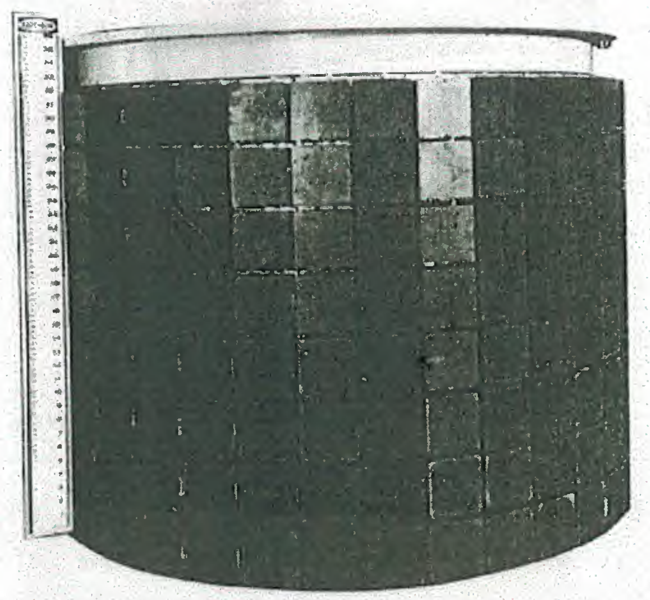

Figure 3.6 A cylindrical, 192 element echo ranging array. Circa 1980's. The unit is 0.9 meters tall. (Urick 1983).

3.2.7.1. Speed of Sound

One ``feature" of unique to sonar applications is the high variability of sound speed in a medium. In water, the primary arena for sonar, the sound propagation speed is a function of temperature, pressure and chemical composition (in this case, salinity) (Coates, 1989). Thus in a real body of water, the speed profile varies with seasonal changes and currents. A typical deep-water profile is shown in figure 3.7a.

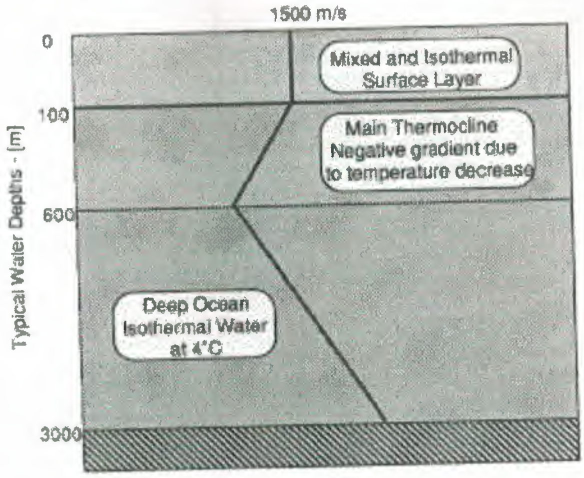

Figure 3.7a: Sound speed in ocean water (Coates 89).

A consequence of this feature is that sound rays (the path locally normal to the wave fronts) will refract and bend with depth. In the limit where the water depth is small compared to the horizontal distance, grazing rays will reflect and refract. This phenomenon can be exploited in several ways. Rays from a source located in the surface layer (down to perhaps 100 meters) will refract downward into the deep water whereupon they will be refocussed upward. These *convergence zones* form concentric rings around the horizontal location of the source (Lombardo et. al. 1993) (Figure 7b). Hydrophones placed on the surface at these regions will hear a much stronger signal than those placed in the ``dead zones" between.

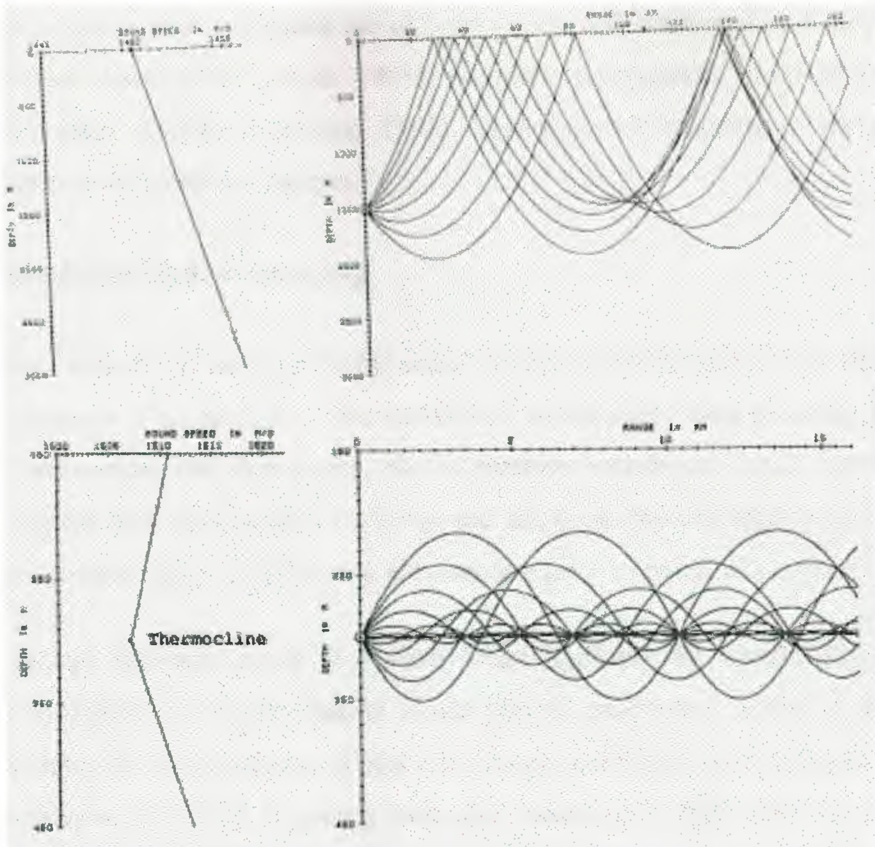

Figure 3.7b,c Ray traces of sound above and below the thermocline (Coates 1989).

For sources at greater depth, upward-heading rays may refract downward and never actually reach the surface (figure 3.7c). Military submarines utilize this phenomenon by cruising below the *thermocline*, the depth where the sound speed reaches a minimum, thus hiding their acoustic trace from surface ships. Sub-hunting surface craft counter this by deploying deep-running passive arrays (Lombardo et.al. 1993) or by using data from sea-bed arrays (Davis et.al. 1997).

3.2.8. Applications of Acoustic Arrays

Sonar technology and phased array technology can and have been quite successfully merged with very impressive results (Lombardo et.al. 1993; Boyles & Biondo 1993; etc). A few of the multitude of sonar array applications include passive ``listenning" arrays for subarine detection. Marine biologists use similar arrays in tracking animal life such as whales. Geologists are able to detect submarine movement of magma and seabed materials.

Oceanographers make extensive use of both vertical and horizontal seafloor arrays to study surface waves (Davis et.al. 1997), accoustic propagation speed in various shallow-water areas (Boyles & Biondo 1993), high-resolution mapping of the ocean floor, and seasonal temperature changes.

3.2.8.1 Side-Scan Sonar Imaging

Certainly one of the more profitable areas of underwater acoustics, it is also one with huge numbers of applications. This technology is frequently used by archeologist, geologists, prospectors and developers, not to mention search-and-rescue teams and practically anyone else who wants to explore the sea-floor. Indeed, high quality side-scan units are commercially available now for a modest price to practically anyone.

The typical side-scan sonar is a short array (typically 50 wavelengths long) encased in a *towfish*, a torpedo-shaped object towed underwater behind a ship. It typically operates in the hundreds of kilohertz range providing good azimuthal and range measurements. The high frequency limits it to operating in fairly shallow water (a hundred meters). (Figure3.8)

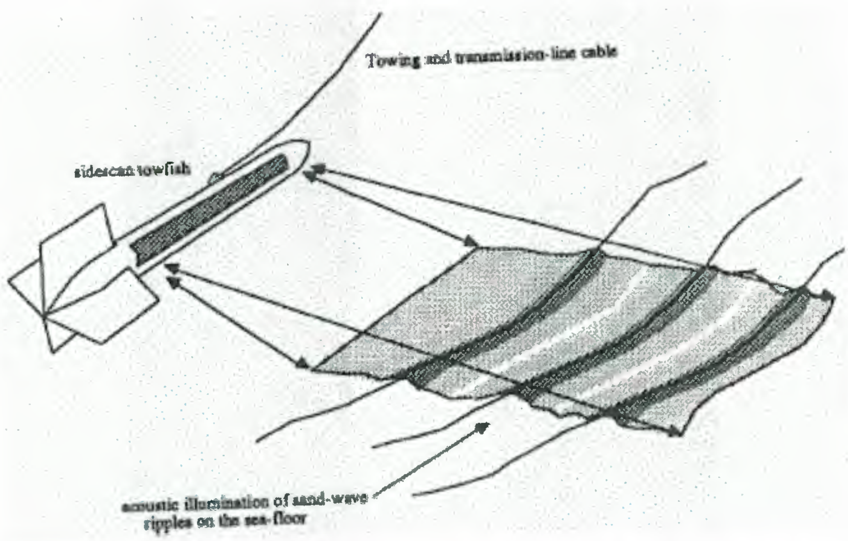

Figure 3.8 Side-scan sonar system. The towfish is typically a meter or so long (Coates 1989).

As the towfish is pulled through the water, it emits a sonar pulse in a fan-pattern covering a line of seabed. The strength of the return plotted against the delay can then be interpretted as a the illumination of a source as a function of distance.

Subsequent pulses give the next lines in what soon becomes a 3-dimensional image. While this data seems very arcane, it is a representation the human brain is extremely good at interpretting. Strong signals belie something strongly ``illuminated" while areas of little or no signal are ``shadows". Figures2.9 and 10 show particularly stunning samples of these images.

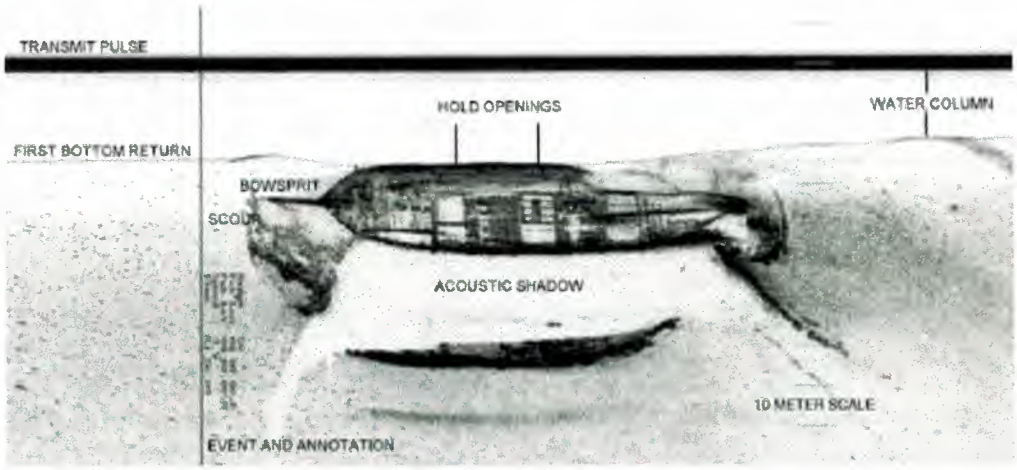

Figure 2.9 Side-scan sonar image of a wreck

Figure 2.10 Side-scan sonar image of seafloor geology .

3.2.8.2. Passive listening Arrays

The world's naval industries are constantly engaged in a race to make their submarines stealthier and develop better methods of detecting enemy vessels.

Davis et.al. (1997) advanced the cause by using a passive, two-dimensional array layed on the ocean floor connected by fiberoptics to a central telemetry station. This device successfully measured ambient sound in the frequency range 0.01 to 6000 Hz and detected passing ships. Similar arrays feature the ability to monitor surface waves height (the height variation of the water causes pressure fluctuations in the depths) and to monitor local marine life.

A different approach to the same challenge was developed by Lombardo et.al. (1993) with their twin-line towed array. this 5km twin-line array was towed behind a ship in the deep waters south of Hawaii. Using careful beamsteering and nulling (figure 3.5), they were able to scan the acoustic environment and resolve individual surface ships many hundreds of kilometers distant.

3.2.8.3. Air-based Acoustic Arrays

It is no surprise that practically all acoustic arrays operate underwater. There are several reasons for this. Gasses are not good conductors of acoustic energy unlike denser media. The temperature variations in the air are more extreme and faster varying than those in water leading to less well understood refraction systems. Finally, other sensing systems, such as radar and simple visual observation, can operate in air while they are hampered in marine environments.

An interesting non-marine application of passive accoustic arrays is found in a series of military listening posts scattered across the American Southwest (Hoffman 1996). Originally intended to detect the extremely low frequency accoustic (infrasound) signatures of atomic weapon tests, it has proven useful in tracking large meteorites entering Earth's atmosphere.

3.2.8.3. Astronomical Requirements for the Millimeter Array Correlator

Taking full advantage of the sensitivity and flexibility of the Millimeter Array (MMA) will require an impressive correlator. The signals from 40 telescopes (possibly as many as 128, if major foreign collaborations materialize) must be correlated over bandwidths of at least 2 GHz, and preferably 8 GHz, per polarization, producing 4-32HTz of (bandwidth times polarizations). This should be split amongst at least 4, and preferably 8, independently-tunable baseband pairs. There should be 500-1000 channels (with two polarization products each) over the full 8 GHz, and one should be able to trade bandwidth for channels in a fairly flexible way. Standard sustainable integration times of 0.1 second are required, with sustainable integration times of $40(\frac{\delta \text{ m}}{\text{D}})^2$ mscc being highly desirable. This gives a required sustainable data rate of at least 250 million visibilities per second. The spectral dynamic range, measured either as the accuracy of continuum subtraction or the ratio of the peak to the spectral sidelobes of a narrow signal, must be $10^5 - 10^6 : 1$. The prospect of a collaborative project with international partners leaves several of the most important correlator requirements uncertain.

This memorandum aims to set forth the astronomical requirements for the correlator of the proposed Millimeter Array (MMA). The current working design for that correlator is described in *MMA Memorandum 166* (Escoffier 1997) and *MMA Memorandum 194* (Rupen & Escoffier 1998), the latter of which also discusses the limits on the expansion of that design. Here we step back to ask what astronomers would like to be able to do with the instrument, and what requirements those desires set for the correlator. The emphasis is naturally on the more challenging projects, those which push the correlator to its limits; at the same time we try to specify a reasonable minimal as well as the ``dream" correlator. The underlying philosophy is that the correlator should not rule out plausible experiments which would otherwise be allowed by the design. Obviously not all of these experiments will be doable when the MMA first opens; but if the VLA experience is any guide, the original correlator will remain in use for many decades following first light.

To avoid confusing what is intended primarily as a scientific discussion, and to restrict the length of an already-lengthy document, we defer consideration of cost equations and other practical matters of correlator design to a later memorandum. Similarly, we do not here compare the requested to the current design specifications.

3.2.9. Some Notes on Nomenclature

We use the same terminology as *MMA Memo 194*. The numbers given here are for reference only, to make a somewhat confusing discussion more concrete, and are based on the current notional design of the MMA system.

- **Channel:** the resolution element of frequency; also referred to as a *spectral point*. A channel may be have associated single, dual (XX & YY), or full (all Stokes) polarization correlator products. We have tried to be careful to state specifically which is required in each instance: *e.g.*, 4000 channels dual polarization, which means 4000 frequency elements with two polarization products each. Other authors sometimes refer to this as ``8000 channels"; here 8000 channels always means 8000 chunks which do not overlap in frequency. In this notation, the total number of complex correlator products is the number of channels times the number of Stokes parameters required for each channel (4000 channels dual polarization corresponds to 8000 complex correlator products).
- **IF:** abbreviation for ``intermediate frequency"; a terribly confusing term we try to avoid, as it is used to mean so many different things. Insofar as it must be used, an IF is a wire coming out of a receiver package. In the current design for the MMA, each wire carries an 8 GHz bandwidth. There is provision for switching halves of the IFs (that is, a 4 GHz bandwidth) independently, and each half is sometimes called an IF. The current plan has a total of 16 GHz being sent down from each antenna to the correlator, for a total of 4 independent (4 GHz) IFs.
- **Baseband (BB):** the signal presented to a sampler. In the MMA case, each BB has a bandwidth of 2 GHz or less, and can be flexibly positioned within an IF band, or switched between IFs.

The canonical design has 4 pairs of basebands, with the two basebands of each pair having the same frequency but different (linear) polarizations.

We often refer to `independently-tunable' basebands or baseband pairs. This is intended to mean, ``capable of being tuned to different frequencies, with reasonable flexibility." There could be some restrictions on which exact frequencies are possible, and the LO chains for the BBs need not be totally independent (e.g. they might share a first LO or some such). The MMA will require much more flexible tuning than the VLA; for the correlator this doesn't matter, and discussion of the LO system, tuning capabilities, and the like, is outside the scope of this document. In this parlance the VLA currently has two 50 MHz BB pairs.

- **Total bandwidth:** in this memo, the total bandwidth is the total frequency coverage, *NOT* the total frequency coverage times the number of correlator products. Again the required polarization products will usually be stated explicitly; for instance, an observation covering 230-232 GHz with all four Stokes parameters recorded simultaneously will be referred to as ``2 GHz full polarization."
- **Correlator:** one source of confusion is the dividing line between the correlator and the other electronics systems. Here we do not maintain a rigorous distinction, but have generally drawn the line at the samplers. The correlators takes as input the sampled signal from each antenna, and gives as output spectra (either frequency or lag) for each baseline and polarization product. There have been many discussions about how much processing is done within the correlator itself, with the tendency being to do as little as possible; so for instance resampling or smoothing the spectra, either in frequency or in time, will probably be done after the correlator.

We assume throughout that the MMA will eventually observe through all atmospheric windows between ~ 30 GHz and ~ 850 GHz, although it will probably not be equipped with all of the requisite receivers initially. For concreteness we also assume that linear polarizations (X and Y) are recorded.

3.2.9.1. Correlator Requirements

The lower limit to the number of antennas is set by the desire for excellent `snapshot' uv-coverage. Snapshot observations are of more importance to the MMA than to other interferometers for several reasons.

First, the atmosphere at millimeter and submillimeter wavelengths is highly variable both in opacity and in phase coherence, and one wishes both to take advantage of brief periods of exceptionally fine weather, and to image large regions on the sky rapidly to minimize systematic effects across the resulting maps. This becomes progressively more important at higher frequencies. Second, the MMA's excellent sensitivity makes very short observations attractive, if the instantaneous uv-coverage is good enough to ensure accurate images. Finally, mosaics are expected to become rather common with the MMA, given the small primary beams at these frequencies; the quality of those mosaics will be set in part by the consistency and completeness of the uv-coverage of each pointing. The second and third of these points were quantified in Cornwell, Holdaway, and Uson (1993) and further elaborated in a series of MMA memoranda by Holdaway and others (Holdaway 1990, 1992; Holdaway & Foster 1994; Wright 1997). All three *desiderata* considered together led to the design goal of 40 antennas for the MMA (cf. the original MMA proposal to the NSF [1990]).

Another, less strong argument for a large number of antennas is the desire for multiple subarrays, each with very good mapping capabilities. This is discussed further below (§3.7).

More significant are the recent discussions between NRAO, ESO, and Japan's NAO concerning possible partnership(s), which would result in a much larger array. The most recent European proposal (Downes *et al.* 1997b) suggests 64*12m dishes, set by the desire to maximize collecting area while minimizing the number of antennas, subject to the constraint that the antennas maintain excellent performance for mosaicing and for high frequency work. The complexity of the correlator itself, and the resulting data rates, are the main arguments for minimizing the number of dishes. On the other hand, it is clearly easier to build superb small dishes than superb big ones, and alternative suggestions range from ~128*8m to 90*10m dishes. Since we will probably not know until at least the end of 1999 whether ESO and/or NAO will actually fund the proposed joint project, the number of telescopes could be a factor two lower than suggested here. So the bottom line is that the MMA correlator should be able to handle at least 40 antennas, and perhaps as many as 80-128 if these major foreign partnerships materialize.

Unfortunately the number of antennas is *not* something that can readily be changed in the correlator at some later date; see *MMA Memo 194*. If we are to have e.g. 80 telescopes eventually it is far preferable to design the correlator to handle them from the start. Of course a correlator which can handle more telescopes than are actually built would allow the addition of further dishes after the completion of the original project (cf. Downes *et al.* 1997b).

3.2.9.2. Maximum Total Interferometric Bandwidth

Many observations will benefit from the widest possible bandwidths. For continuum experiments this is important mainly for sensitivity, and so will directly affect virtually all such observations. Projects for which sensitivity is paramount, even for the ``maximal'' combined US and European array, cover every area of millimeter astronomy: the detection of proto-Jupiters in other solar systems; high-resolution dust mapping around young stellar objects (YSOs); the observation of pulsar emission at wavelengths which minimize the effects of dispersion; the search for dust emission in high-redshift galaxies and proto-galaxies; and so on and on. Sensitivity becomes progressively more important at higher frequencies, as the atmospheric contribution leads to higher and higher system temperatures; and at higher resolutions, since the same total flux is spread over a larger number of synthesized beams. Some of the most interesting science expected to come from the MMA depends on high-resolution, low-noise images. The importance of wide bandwidths in achieving the desired sensitivity cannot be stressed enough.

Wide-band continuum observations will also provide accurate single-band ``colours'' (spectral indices).

While a few narrow basebands spread across the band might suffice for bright sources (the Sun, the planets, Sgr A, M87), wider bandwidth (e.g., 1 GHz) `chunks' spread over a receiver's frequency window would allow similar analysis of more standard sources, in particular the measurement of thermal dust temperatures in Galactic YSOs and extragalactic disks.

The above argues for a wide continuum bandwidth using dual polarization for sensitivity. Full polarization imaging is also important, primarily for stellar emission, dust polarization mapping, Faraday rotation observations, and planetary work (where the polarization fraction is of order 1%).

The signals are expected to be quite faint, but the results will be well worth the effort, particularly in mapping the magnetic fields in molecular clouds and accretion disks. Anumber of spectral line experiments also require or benefit from wide bandwidths.

- *Pressure-broadened planetary lines* can be up to 3 GHz across. In general one would like to measure a line-free continuum at the same time as well.
- *Solar radio recombination lines* are expected to be broad and shallow, requiring total bandwidths of 1 to 2 GHz.
- *YSO outflows* could cover up to 600 km/s (full-width at zero intensity, FWZI), including the extremely-high-velocity CO `bullets' in sources like HH111. This corresponds to ~ 1.7 GHz at 850 GHz, the highest frequency proposed for the MMA. Adding a reasonable area for continuum subtraction brings this to ~2 GHz total bandwidth.
- *Radio recombination lines in ultracompact H II regions* may be up to 250 km/s broad, for the most energetic ionized outflows. Observing H, He, and C recombination lines at once adds another 150 km/s; adding a bit for baseline determination brings the total to ~550km/s or 1.6 GHz at 850 GHz.
- *Large-scale surveys of Galactic gas* need to cover ~600km/s total bandwidth, allowing for the full range of Galactic rotation.

However, it seems unlikely that these will be carried out at the highest frequencies, so the total required bandwidth is more modest, 0.7 GHz at 345 GHz.

• *Radio recombination lines arising from narrow line regions of active galactic nuclei* (including Sgr A*) require a total velocity coverage of at least 1500 km/s, or ~ 5GHz at 850 GHz. These high-frequency lines, observed at high resolution with high sensitivity, are expected to prove very interesting in constraining models of the dense gas surrounding AGNs (Anantharamaiah 1997, priv. comm.).

• *Surveying a galaxy cluster* covering ~10,000km/s in CO J=2-1 at 230 GHz, expected to be the most sensitive transition for such searches, requires 8 GHz total bandwidth. Higher frequency transitions would demand even wider bandwidths. Of course the full bandwidth need not be observed simultaneously.

• *Spectral line surveys* in general, either local (Galactic) or at high redshifts, benefit from the widest possible bandwidths. For example, 8 GHz at 45 GHz covers CO J=1-0 emission from redshifts of 1.5 to 2.1. Again such surveys could be carried out by time-sharing among several frequency settings.

Most of these experiments would involve dual polarization for sensitivity. The case for full polarization wide-band spectral line measurements is much less clear, confined to Zeeman splitting observations of masers covering a wide range of velocities; however this could easily be done in a series of frequency settings, and will probably be a rare enough project that the additional time required will not be an issue.

In sum, a bandwidth of 2 GHz, producing two correlator polarization products (XX & YY), is necessary for a wide variety of spectral line experiments. Observations of pressure-broadened planetary lines and radio recombination lines associated with the Sun or AGNs require double that bandwidth (4 GHz with two polarization products) to fit the line into a single frequency setting, while the very broadest of such lines at the highest frequencies might require up to 5 GHz. For continuum observations sensitivity demands the widest bandwidths possible.

A large number of experiments would also benefit from the ability to sacrifice polarization products for bandwidth; most experiments not involving linear polarization measurements would prefer to cover, e.g., 16 GHz producing only the parallel-hand (XX, YY) polarization products, to always getting full Stokes information but only over 8 GHz. In any event the correlator should match the rest of the instrument in allowing the maximum bandwidth permitted through the receivers, backends, etc; with the current systems design, with 2*8GHz sent to the correlator from each antenna, this would imply all four polarization products for an 8 GHz bandwidth.

3.3. Spectral Dynamic Range

The term ``spectral dynamic range" (SDR) is used to mean at least three different things:

1. The ratio of the peak continuum signal to the root-mean-squared (rms) noise in a continuum-subtracted image. A high SDR in this sense corresponds to having a very flat frequency response.

2. The ratio of the peak continuum signal to the accuracy with which one can measure a very deep absorption line. A high SDR in this sense corresponds to correctly measuring a very wide range in correlation coefficient.

3. The ratio of the peak of a narrow (in frequency) signal to its spectral sidelobes. A high SDR in this sense corresponds to very little cross-talk between frequency channels.

The MMA, especially in the larger versions suggested for the collaboration with the Europeans, will be a very sensitive instrument, with noise levels at 230 GHz in one minute as low as 30μJy in the continuum and 7mJy (15mk, for the smallest configuration) in a 0.2 km/s channel. One will often be looking for weak signals in the presence of strong confusing sources, either continuum or line (e.g., masers), requiring a high SDR potentially in all three senses.

1. *Flatness:* the ratio of the continuum brightness to the noise level in a channel map will often need to be as high as $10^5:1$ and for many experiments $10^6:1$ or higher in observations of faint lines on top of very bright continuum sources. Examples include rarefied species near bright YSOs, H II regions, and the like; radio recombination lines in ionized outflows; and searches for faint absorption lines against very bright AGNs.

2. *Absorption:* This is probably the least important type of SDR to an astronomer, as it limits the accuracy of a measurement rather than the possibility of a detection. The cases where this would matter a great deal are probably limited to searches for faint substructure in high-opacity lines, for instance searching for very weak emission superposed on very strong absorption. Probably one would like to be able to believe 1% variations in highly opaque lines, but it seems unlikely that even higher accuracy would regularly be required.

3. *Spectral sidelobes:* the desired limit on the leakage of a strong signal in one channel into other channels in the same baseband, is set primarily by deep observations of emission in line wings of sources with very bright emission at the line center.

> Examples include many of the most interesting objects in the sky: planetary absorption lines (e.g. Gurwell 1996); molecular outflows (e.g. Yu & Chernin 1997 [VLA 1623/CO], Cernicharo & Reipurth 1996 [HH111/CO]); YSOs (e.g., Hogerheijde *et al.* 1997 [T Tauri/HCO+], Olmi *et al.* 1996 [G10.47+0.03/CH_3 CN]); star forming regions (e.g. Wink *et al.* 1994 [W3(OH)], Shepherd, Churchwell, & Wilner 1997 [ON2]); stars (e.g., Dayal & Bieging 1995 [IRC+10216]); and even external galaxies (e.g. Shen & Lo 1995 [M82], Sofue & Irwin 1992 [NGC 3079]). In all these cases the line centers peak at some 10s of Kelvin, while the rms noise in one minute on source might be a few milliKelvin (for the most compact configuration). This neglects even more difficult cases, such as looking for thermal emission around masers, searching for faint (rarefied?) species in the same BB as stronger lines, planetary radar experiments (where the strong zero-velocity return signal creates a problem similar to the maser case), and observations (at the lower frequencies at least) in the presence of strong radio frequency interference. This suggests that $10^5:1$ will be desired fairly often, and $10^6:1$ may not be that unusual a requirement.

Of course, achieving these levels is not purely a correlator problem; at the VLA we almost always Hanning smooth the raw spectra to beat down the spectral sidelobes, and (so long as one has enough bits) one can do this or more sophisticated apodisation after the correlator. The discussion here refers to the SDR after such apodisation.

3.3.1. High-quality Imaging

Although it is not clear how to relate image quality to the correlator design, for completeness it may be worth mentioning the dynamic range (peak to off-source rms noise level) MMA images will achieve. With noise levels as in the last section and taking 10 hours as a reasonable long integration, one expects dynamic ranges of

- $10^7:1$ for a few special sources - planets, masers, strong AGNs;
- a few to 10 times $10^5:1$ for many sources - including continuum and strong line emission from young stellar objects, AGB stars, H II regions, the Galactic center, and less impressive AGNs;
- $10^4:1$ fairly routinely, in many cases in only a few minutes of integration.

The above dynamic ranges assume the joint US+European project with a collecting area of ~ 7,000 m^2; for the MMA alone (2,000 m^2) they should be reduced by a factor ~ 3.5 Such excellent images will be useful scientifically in many contexts, ranging from searching for extragalactic counterjets at frequencies where the core is not so bright as to limit the imaging, to looking for proto-Jupiters around main sequence stars, to mapping faint, extended gas around young stellar objects. Most of these projects require not simply high dynamic range, but high on-source accuracy as well, with peak on-source errors of 1% being a good target (cf. Cornwell, Holdaway, and Uson 1993). Nothing in the correlator should prohibit making such high-quality images.

Another kind of dynamic range relates to the total range of data within a single uv-data set - the ratio of the peak short-spacing flux to the rms noise on the longest baseline.

This is in some ways a more difficult number to derive, since it involves comparing the integrated flux density of a source to the rms noise in an integration period.

Probably the brightest sources for which one might achieve thermal noise on the longest baselines are the planets, with brightness temperatures of 200-300 K (for Jupiter and Venus). The flux density measured on the shortest baseline would then be of order $15{,}000(\frac{T_B}{250K})(\frac{\delta m}{D})^2 JY$ where D is the telescope diameter. With a noise level between 5 and 20 mJy on a single baseline in 1 second for frequencies up to about 230 GHz, the peak ratio of flux density on the shortest baseline to rms noise on the longest baseline would be of order $10^6 : 1$.

3.3.2. Radio Frequency Interference

Although radio frequency interference (RFI) has not in the past been much of a problem for millimeter interferometers, RFI has been increasing even at the high frequencies used by the MMA, and will certainly be an issue for at least the lower observing bands by the time it is built. Currently the main frequency allocations above 30 GHz have been to satellites, with a few areas in Q band going to stratospheric balloons and automobile radar systems. Little use has yet been made of the frequency space already allocated, but some important features are already clear. First, most services proposing to operate at these high frequencies do so because they need fairly wide bandwidths. This implies that typical RFI signals will be at least a few MHz wide. Second, at millimeter wavelengths it is more difficult to generate high transmitter power, while high gain beams require only small antennas. Thus one would expect most transmissions to be highly beamed at specific areas. Most importantly, the sidelobe levels of radio astronomy antennas are likely to have similar gain to those at centimeter wavelengths, i.e. of order 10 dB for those near the main beam and order 0.1 dB at angles greater than about 50 degrees from the main beam. Since the collecting area for a given gain is proportional to wavelength squared, the sidelobe sensitivity to interference decreases with increasing frequency. This leads to the third point, that the bulk of the worrisome interference will come from satellite downlinks.

Since those satellites are likely to surround the globe, we will not escape their transmissions however remote the observatory site. If the satellite downlinks use time multiplexing like IRIDIUM, they will transmit in brief bursts (IRIDIUM uses 4.5 msec packets) which we may be able to flag if the signal can be recognized and discarded on ~ msec time scales.

Such time-sharing may not be very common however, since most satellite allocations are designated specifically as space-to-Earth or Earth-to-space, whereas IRIDIUM takes advantage of an unusual secondary uplink allocation within a primary downlink band. While the RFI situation is currently fairly benign, we cannot afford to be complacent. At the VLA the bulk of the interference above ~ 15GHz is internally generated, due mostly to the LO systems (the 100-200 MHz `birdies'). This should be avoided if at all possible at the MMA; it will do us little good to have clear skies if we bring with us our own headlights. Further, although no allocations have yet been made above 300 GHz, those are to be discussed in the 1999 World Radiocommunications Conference.

3.3.3. Miscellaneous Constraints

In addition to the major requirements discussed above, the correlator must allow for a number of more specific constraints:

- *Fast telescope motions:* Slew speeds of order 1^0/sec, and the correspondingly fast dump times, will require rapid updates to the correlator model.In the (common) fast switching case there may be savings because one is switching between only two positions, but OTF (or even `point and shoot') maps will be more challenging.

Unlike the VLA, the MMA will often be pointing at significantly different positions during adjacent integration periods (dump times).

- *Rapid frequency switching:* several experiments would benefit from rapid frequency switching within a single band, though the availability of more than the minimal 4 BB pairs would take care of most cases.

More important and more difficult is the problem of phase-referencing one band to another, e.g. to take advantage of a strong maser transition. In this case one might want to switch between frequency bands in a few seconds (for the more extreme examples). Whether this mode should be considered vital, and how often it might be used, will be discussed in an upcoming memo on phase calibration (Holdaway *et al.*, in prep.).

• *Maximum baseline:* Many of the most interesting MMA observations require high resolution, between several 10s of milliarcseconds and an arcsecond. Examples are legion, including e.g. searches for proto-Jupiters (or tell-tale gaps in disks) around YSOs, imaging of stellar photospheres, and resolving AGN accretion disks out to Virgo and beyond. The desired resolution corresponds to baselines ranging from a few to 10s of kilometers, with 10km being the maximum currently discussed. The correlator should therefore allow for baselines up to 10km long. Although it is unlikely that the initial array will include such long baselines, the existence of prospective international partners specifically interested in such high-resolution experiments is a strong argument for allowing for this capability from the start. Even longer baselines have occasionally been discussed in the context of major international partners.

• *Phased-array output:* For some types of experiments, most notably gated pulsar and VLBI observations, one wishes to phase up the array in real time and write out the summed data. For VLBI one wants the raw summed data, not averaged over time; for pulsars one can average down (see *Gating*, below).

• *Gating:* Pulsars are very weak (but detectable; Kramer *et al.* 1997) at MMA frequencies; writing out the (phased array) correlator output in time bins sufficient to phase resolve the pulse profile would therefore be useful in increasing the signal-to-noise of their emission. This sort of gating in time would also be useful for on-line RFI excision, particularly of periodic interference. Various flux calibration schemes (measurements of ON/OFF loads and the like) might benefit as well from the availability of multiple data streams. The required time resolution for pulsars would be a tenth of a millisecond or better, and one might want to employ 10 or 20 different time bins across the ~ msec pulse, to sample the pulse profile.

RFI may require time samples as fine as ~ 1 msec (IRIDIUM for instance pulses with a 45 msec period, and its individual data packets are a tenth as long).

- *Burst mode:* Some objects (e.g. solar flares) vary so rapidly that it might be advantageous to write out partial data at a much higher rate than normally allowed.

One might also consider trading baselines for time resolution, for instance for large scale surveys limited by total observing time rather than sensitivity, at least during the early days of the instrument. A special burst mode is probably not necessary if dump times of some 10s of milliseconds are considered `standard' (see 3.6).

The Sun is an example of a source which [2]The Sun and Other Bright Sources: dominates the system temperature; during bursts, that system temperature will vary dramatically. Similarly other sources contribute significantly to the system temperature (e.g., planets), and there too T_{sys}

- will vary somewhat as the source wobbles around in the primary beam of each antenna. That variation must be monitored and the visibility data corrected. This affects the correlator because its power inputs may not be held at constant and optimum levels. The consequent less-than-optimum performance is probably not a big issue, as this should affect only a small number of sources.

3.4. Summary

From the above discussion, the main requirements for the MMA correlator are as follows:

- *Number of antennas:* ≥40: at least 40 for good imaging, up to 80-128 if we can afford more collecting area but don't want huge dishes.
- *Maximum bandwidth:* There are several obvious break points for this.
  - ≥2GHz with two polarization products, corresponding to ≥4GHz of (bandwidth times polarization products), for a wide variety of spectral line experiments.

○ ≥4GHz with two polarization products for pressure-broadened planetary lines and wide RRLs.

○ 8GHz with full (4) polarization products to match the maximum bandwidth sent down from each antenna in the current systems design.

○ As wide as possible to maximum continuum sensitivity and the efficiency of spectral line surveys.

- *Frequency resolution:* 10 Hz - 100 MHz, with the ability to select in factors of two or something similarly flexible. A few to 10 Hz is needed (but only over a few hundred channels) for bistatic radar; other experiments require only 2 kHz resolution, for sampling thermal/dynamical linewidths of low-mass systems (comets, YSOs, dense cores, etc.), resolving molecular absorption lines, and Zeeman splitting. SETI searches on the other hand would benefit from 0.1 - 1 Hz resolution, over as wide a band as possible, but this should not be permitted to drive the correlator design. The lowest interesting resolution is that corresponding to features in the bandpass, due to poor delay settings or to the RF/IF systems themselves. One wants this low resolution for continuum experiments, to keep the data rate down as much as possible.

- *Number of basebands:* 4-8 independently-tunable baseband pairs. Desirable for flexibility in placement of continuum bandwidth (for spectral index studies, and to avoid atmospheric emission and RFI) and choice of multiple line transitions. This last is crucial to the MMA, not so much because its lack would disallow vital experiments, but because observing several lines at once is a far more efficient use of the instrument.

- *Number of channels:* 500-1000 (with dual polarization products) spread over 8 GHz, desirable for a wide variety of line experiments, ranging from searching for molecular clouds in elliptical galaxies, to imaging the entire primary beam with negligible sensitivity losses due to bandwidth smearing. Spectral line surveys and unbiased line searches could benefit from as many as 1000 channels per GHz over the entire band, but these can be carried out by time-multiplexing amongst several frequency settings; it is however interesting to note that there is no obvious experiment which would benefit greatly from *more* than this number of channels. These numbers assume that one can trade bandwidth for frequency in a fairly flexible way, i.e. obtain more channels across a narrower bandwidth.

- *Dump times:*
    - o *Autocorrelation spectra:* $3.0(\frac{\delta\ m}{\mathrm{D}})m\sec$, to allow good atmospheric subtraction during on-the-fly measurements at 350 GHz
    - o *Phased-array data:* 10 μsecfor pulsar work
    - o *Interferometric data:* $40(\frac{\delta\ m}{\mathrm{D}})^2 m\sec$ to allow surveys of large regions on the sky. This requirement, if it must be relaxed, should at least be kept below 100 msec, to avoid significant decorrelation due to phase variation.
    - o *Max. dump time:* 10 seconds or longer, if one can hold the phase steady during that time. Without active phase correction the maximum dump time would probably be about 1 second.
- *Subarrays:* at least 4; prefer 6-8. Desirable for many reasons, most importantly for flux/phase calibration, to speed up mapping large areas, to lower the data rate, and to allow the use of part of the array for single-dish measurements. More important if there are more telescopes and for larger total collecting areas.
- *Total power measurements:* must be fully supported for taking data similar to those from the interferometer. Vital for flux calibration and mosaicing.
- *Ancillary data:* must record a variety of weather, telescope, and related information with the visibilities. Important primarily for calibration (phase, flux, and pointing) and flagging.
- *Spectral dynamic range:* $10^5 - 10^6 : 1$, for many projects (weak lines near strong thermal sources (stars, YSOs, H II regions), faint line wings, etc.). Conveniently this numerical requirement is roughly the same for both the flatness of the frequency response and the limit on spectral sidelobes.
- *RFI excision:* allow for very strong, transient signals; possibly employ on-line flagging and masking, particularly for the longer integration times. May require flexible, programmable on-line flagging.
- *Maximum baseline:* 10 km, to allow eventual resolutions of a few milliarcseconds. Might be even longer if a major foreign collaboration materializes.

3.5. Remaining Uncertainties

As the astute reader will have noticed, several areas of these correlator specifications would benefit from more careful study. While the need for ancillary data (figure 3.9) is obvious, whether it would be useful to apply any calibration derived therefrom on-line is not. Although doing this in the correlator would significantly complicate the design, the prospect of averaging down the data before writing them out is very attractive. Perhaps one could employ an intermediate, real-time processor directly after correlation to do some simple calibration and flagging before the data are written to disk. Similarly the maximum channel width is important in limiting the output rate, but will be determined by the accuracy of the delay settings and the frequency characteristics of the LO and related systems.

By far the largest source of uncertainty however is the possibility of significant foreign partners. If either the European LSA or the Japanese LMSA does in fact merge with the MMA, the budget will grow considerably, and a rather different instrument will result. This uncertainty is reflected in several areas of the correlator specifications. The most obvious is the number of antennas, which might go from 40 to a hundred or more. This together with the larger collecting area leads to a desire for more subarrays and more stringent dynamic range limits. Larger antennas have smaller primary beams, which make one want shorter integrations to allow mapping a given sky area in the same total time. Some of the joint proposals, particularly the Japanese, also push for longer baselines.

CHAPTER FOUR

CHARACTERISTIC OF ULTRASONIC PHASED ARRAY

4.1. Overview

The inspection possibilities with ultrasonic phased array technology are closely linked to the reliable operation of the ultrasonic of the electronic equipment and the soundfield properties of the array probes.

Since many of the specific characteristic properties are not comparable to those of standard ultrasonic equipments and -probes and are in addition not common within the medical diagnostic field one has to introduce some procedures specific for phased array systems in NDT .The reliability and reproduceability of inspections with such complex systems are much more depending on the correct functioning and the correct calibration of the system then any other ultrasonic inspection equipment. Since the calibration procedures are in general not sensitive enough for the detection of all parameter deviations from the their standard values, it is nescessary to consider special steps and tools for the measurement of the most important probe-and equipment characteristics

Figure3.1 shows a schematic presentation of an ultrasonic phased array equipment with probes and single elements, cable and cable adapter, the transmitter, the receiver, the A-scan converter and the data acquisition computer. At the different parts of this block diagram are indicated the figure numbers, which contain more detailed information for this specific equipment partition. All the parameters which will be discussed in the following have to discriminate two different purposes: There are data sheet parameters which should be guaranteed by the equipment manufacturer and there are operational parameters which can be adjusted, calibrated and modified by the user of such a system . Although this basic difference is very common in standards for the characterization of ultrasonic equipments, in the case of phased array equipments we have to reconsider this seperation in view of the different technical concepts of the used equipments.

Data sheet parameters and design criterias are closely linked. E.g. for the probe design (element size, element spacing, distribution of differently dimensioned element sizes, sensitivity apodizing etc.), it is important to know the relationship between grating and main lobes at the minimum and the maximum skewing -or incidence angles. This can be optimized by a suitable modelling of the probe soundfield structure [1,2]. The software may influence the grating lobes by an amplification apodizing at the border elements of an array or by randomly chosen small variations of the delay times. It is therefore recommanded to document the grating lobes for the different wave modes and to regard them also as an important operational parameter .

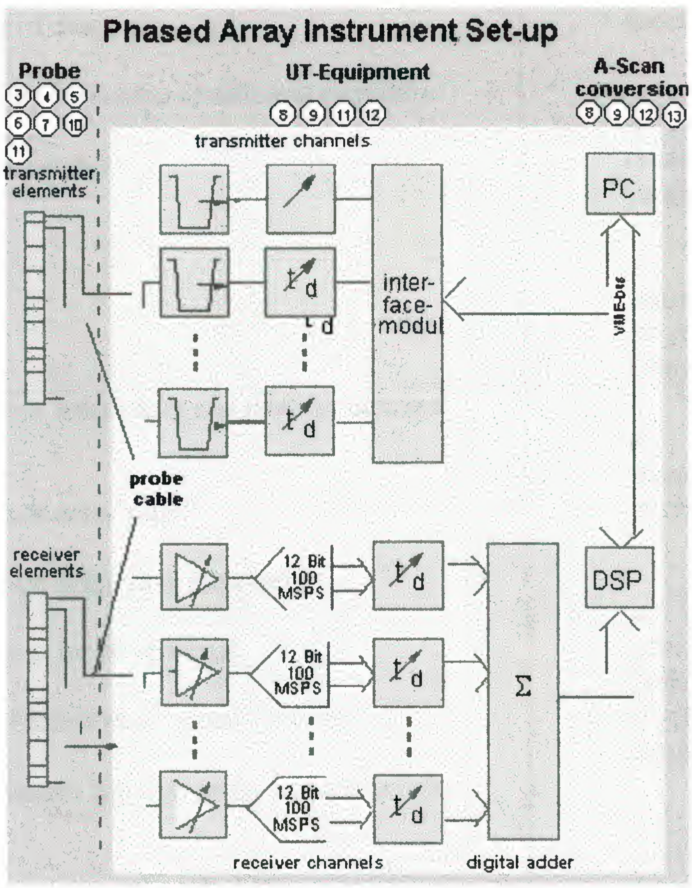

Figure 4.1 shows a schematic presentation of an ultrasonic phased array equipment

Data Sheet Parameters	Operational Parameters
o Proto type, frequency and wave mode o size of the effective transducer dimensions o range of the skewing or the incidence angle o delay path length o wedge angle o number of elements o sensitivity variations of different elements o frequency range	o Crosstalk between probe element channels o sensitivity variations due to: o skewing - or incidence angles and focussing o software controled parameters o (angles, beam divergency, focussing,grating lobes)
o number of transmitter and receiver element channels o maximum delay time o maximum delay time increment o number of probe channels o (T/R - Mode/Pulse - Echo - Mode) o Cable length between probe and equipment	o Crosstalk at transmitter/receiver channels o amplitude transfer behaviour(linearity, noise, saturation) o frequency transfer behaviour(bandwidth) o signal to noise ratio delay offset o gain

	compensation of DAC curves (TGC)
o Sample rate o resolution in bits o real time storage rate of A-Scans	o dynamic range and amplification(effective number of bits, digital offset, undersampling of echopulses)

Table 4.1 gives a list of parameters for both groups. Data sheet **parameters are e.g.:**

For the probe: probe type,-frequency and wave mode, number of elements, size of the elements, sensitivity variations of different elements, size of the effective transducer dimensions, delay path length, range of variation of the skewing and the incidence angle. Angle of incidence for a zero delay-time distribution or wedge angle.

For the equipment: maximum cable length between the probe and the equipment., number of transmitter and receiver channels, number of probe channels(in T/R-Mode and in Pulse-Echo-Mode), frequency range, maximum delay time, minimal delay time increment.

For the A-scan converter: sample rate, resolution in bits, maximal and minimal dynamic range, maximal pulse repetition rate of the the A-scan conversion and data storage. Due to the software based variability of some of those parameters, many of them have to be regarded as well as operational parameters.

Some typical operational parameters are: amplitude transfer behaviour (linearity), dynamic range in dependency on the amplification calibration and adjustment, frequency transfer behaviour (bandwidth).Crosstalk between different element channels at the probes and at the transmitter/receiver channels within the equipment.Signal to noise ratio for a given target and at a given distance. Software depending parameters like angles, angle beam divergency, focussing. Sensitivity variations due to changes of the skewing- or incidence angles and of the focussing, sensitivity corrections depending on the element channel and depending on the angle and the focussing, sensitivity corrections related to the DAC curves (TGC).

In the following specific problems with some of the listed parameters are described in detail.

4.2. Probe Related Parameters

Beside the classical soundfield and spectral characterization of conventional probes there are three typical sets of parameters which must be recorded and measured at a phased array probe:

- element sensitivity
- element separation (acoustic and electric insulation)
- directivity pattern and grating lobe suppression

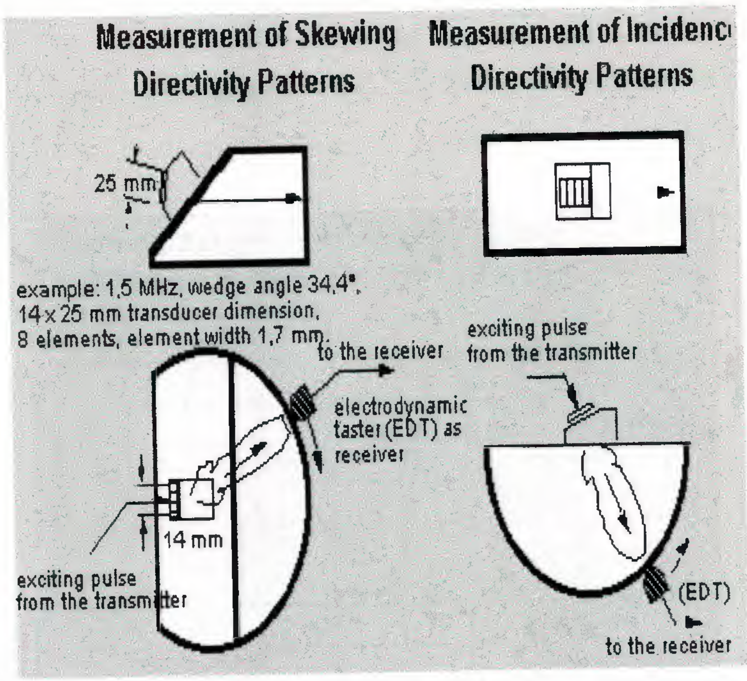

Figure 4.2 presents two directivity pattern measurement arrangements

Figure4.2 presents two directivity pattern measurement arrangements with electrodynamic sensors as contactless microphons on test blocks with a cylindrical scanning surface. Recommanded is a fine grain ferritic or austenitic steel block, similar to that used for the IIW block (ISO 2400).

With the help of such a block the individual maximum amplitudes of the elements can be measured as well as their pulse shape and spectrum. The measurement equipment of figure 3 should use as a standard pulse e.g. a needle pulse at a 50 ohm resistor with 5 nsec rise time and 250 V exitation voltage.

If squarewave pulses are used the pulse duration should be in the range of 0.8 times half the period of the center frequency.

The quality of the directivity measurement with the arrangement of fig.4.2 can be characterized by the comparison between a theoretical calculation based on the model of [1] and [2] and the measurement as shown in Figure4. 4 Since the element sensitivity may deviate from each other it is nescessary to measure it e.g. with semicylindrical test blocks like those shown in fig.4.3 for probes with a skewing or an incidence angle variations.

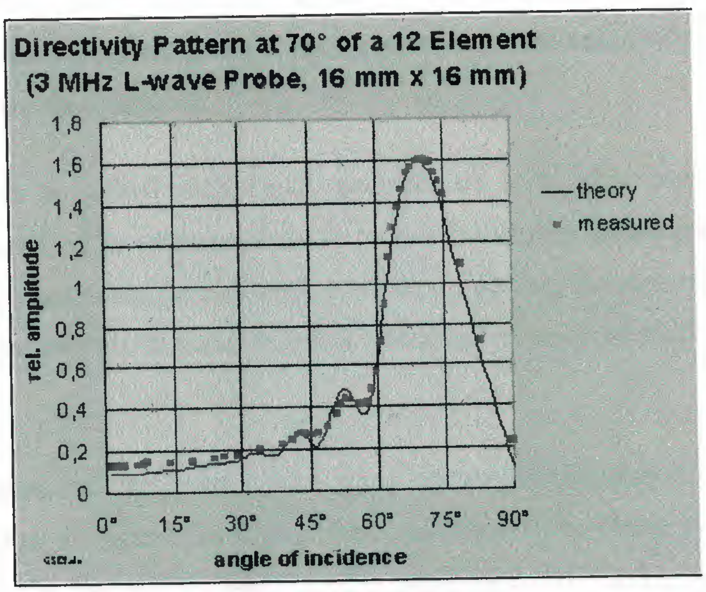

Figure 4.3 probes with a skewing or an incidence angle variations

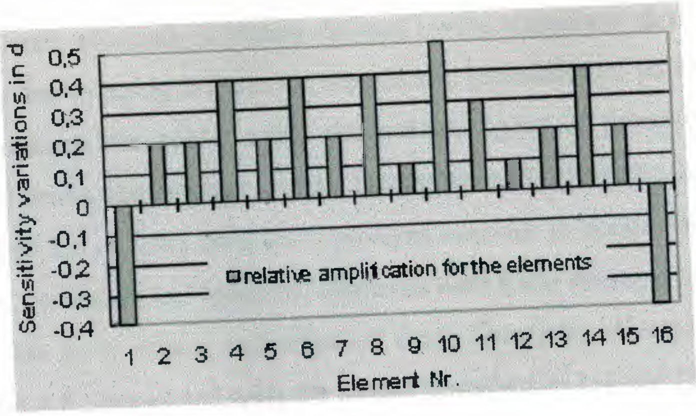

Figure 4.4 an example for the variation of the element sensitivities

Figure4.4 shows an example for the variation of the element sensitivities, which can be compensated by a corresponding software based correction curve individually applied to the different element receiver channels. The variability of the different elements is strongly depending on the production of the probe, especially the cutting between the elements. For the modern composite piezoelectric materials, there are different procedures to cut a transducer surface into different elements resulting in a more or less strong sensitivity variation. A strong variation of the sensitivity reduces the dynamic range of the whole system.

Fig.4.5 shows a typical pulse and spectrum of a 2 MHz array element. The possible presence of secondary vibration modes is strongly depending from the design of the probe concerning the element thickness, the cutting, the element width and the kind of the piezoelectric material (standard ceramic material or piezo composite material).

The pulse shape and the spectrum are a good indicator for those modes. Therefore it is recommanded to document pulse shape and spectrum of the elements in order to check for possible changes due to aging or other phenomena. Deviating spectrum and pulse shapes of the individual elements are also giving first hints on irregularities e. g. due to a bad element separation. The element separation or the crosstalk between different channels is the second important influence being typical for phased arrays. The separation can be reduced by acoustic or electric cross talk.

Whereas the acoustic crosstalk is mainly defined by the transducer design and is not changing to much due to aging influences, the electric crosstalk can be influenced by many other parameters like cable, cable connectors, aging of some electronic components within the probe, the receiver and the transmitter and others. At the probe it is interesting to observe the element directivity patterns in order to document possible crosstalk deficiencies. As shown in Figure4.6, elements with a too strong crosstalk can easely be detected due to a strong reduction of their directivity divergency. The theoretical curve in fig.4.6 , calculated with the model described in [1] and [2], defines the ideal separation status between the elements. A whide divergency is needed in order to guarantee a large range of steerable angles for a phased array probe.

If the directivity pattern is too strongly reduced by crosstalk influences, it is nescessary to compensate the reduced steerability by sensitivity corrections which is resulting in a decreased dynamic range of the whole system. The grating lobes of a probe system can also be calculated according to [1] and [2] and be compared with the practical measurements. It is recommanded to store all the probe parameters into a data bank for the purpose of fast comparison and the early detection of aging phenomena.

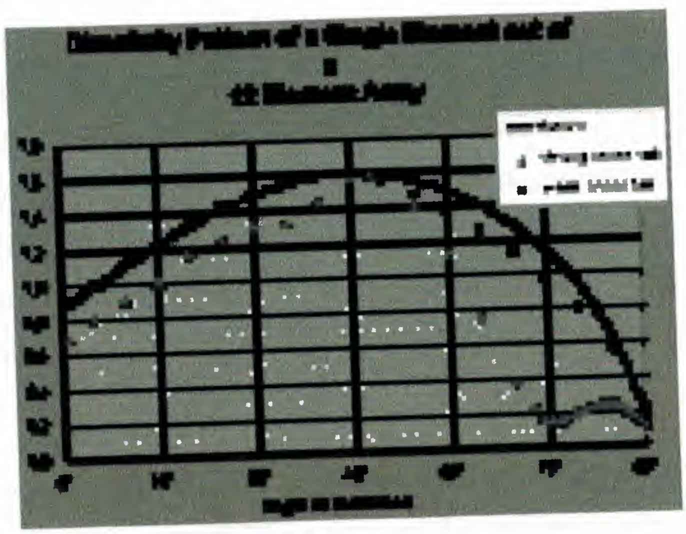

Figure 4.5 a typical pulse and spectrum

Echopulse and Spectrum
of an Element

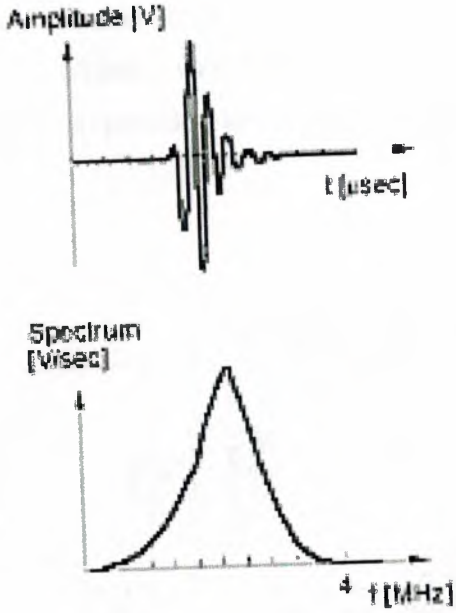

figure 4.6 document possible crosstalk deficiencies

4.3. Equipment Related Parameters

The cable and the equipment have to be checked for transmitter, receiver and A-scan conversion parameters. At the transmitter channel the constancy of the pulse exitation voltage and the pulse shape as well as the independency from the different delay times are essential parameters which should be documented. This can be measured qualitatively by a suitable control software or more quantitatively with a special hardware support. The receiver channel can be characterized by two transfer functions (Fig4.7): the amplitude transfer function to check the linearity of the amplifier behaviour , the frequency transfer function to check possible limitations of the frequency band width.

From the amplitude transfer behaviour one can derive the maximum dynamic range between noise and saturation level. The noise level depends on the amplification settings and for digital equipments also on the handling of the lower bits after the A /D conversion. It is recommanded to check the amplitude transfer behaviour for the whole equipment and for the single channels at different amplification setting.

It must be mentioned that especially during the sensitivity calibration procedure using large reference reflectors like a back wall or a larger side drilled hole it may happen that one element channel delivers a signal amplitude above the saturation level of that individual receiver channel. This will not be detected observing only the total receiver output signal which adds up all individual element channels. It is therefore an essential for phased array equipments to provide a kind of monitoring for the saturation at individual element channels.

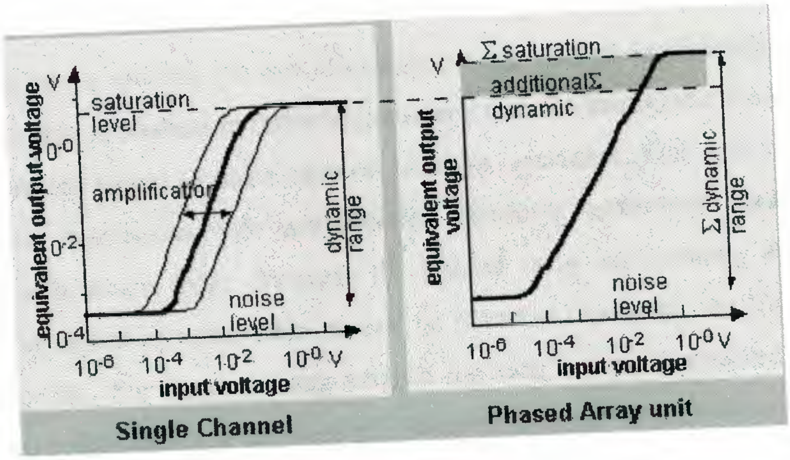

Figure 4.7 Amplitude Transfer Function

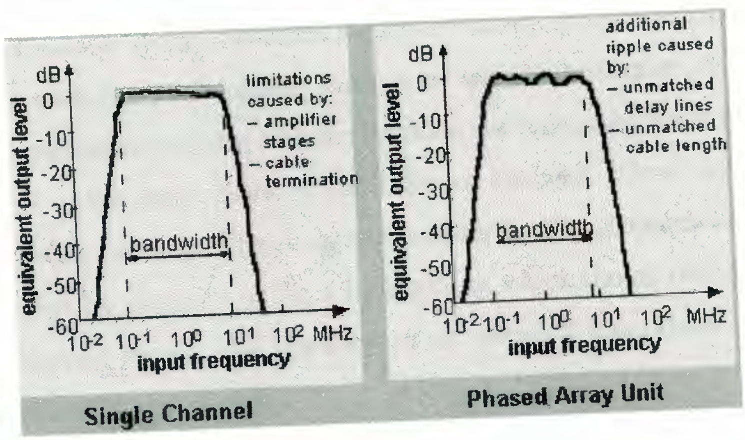

figure 4.8 Frequency Transfer Function

Fig. 4.9 shows the dynamic range problem at the different stages of a receiver channel. It is of course depending on the kind of amplification settings in the preamplifyer stages and on the kind of digitalization used. The different corrections to be applied at the element channel amplification - e.g. to compensate for the element sensitivity variations, the angular sensitivity dependency or the DAC curves with the TGC - are complicating the problem and are often requiring an individual optimization. It is interesting to remark that the relationship between the sampling frequency of the digitization in relationship to the carrier frequency of an ultrasonic pulse plays an important role for the maximum transferable dynamic range. Depending on the kind of sensitivity needed for a specific job one has to choose an optimal amplification setting during the calibration procedure in order to guarantee that the inspection is not limited by problems with the signal to noise ratio or with the saturation. This problem exists also at automatic inspections with conventional ultrasonic equipments using linear amplifiers but it is much more dramatic for phased array equipments, where the limitations of the signal to noise ratio cannot be detected observing the total output signal at the receiver. For the frequency transfer function it is recommanded to use a continuous wave exitation at the input of all element channels in order to avoid a too strong influence from slightly deviating delay times. The use of a sinus burst pulse results in a more or less strong dependency of the measured transfer behaviour on the burst length. Fig.4.8 (bottom) shows examples for the frequency transfer function of an ultrasonic equipment with the total transfer behaviour and the behaviour of a single element. The typical oscillations of that transfer function are indicating some problems with the homogeneity of the delay times of the different channels. Those oscillations should not be greater than +-0.5 dB. The A-scan conversion can be characterized by the transfer of the stepped pulse sequence of Fig.4.12 (bottom), which should produce at the output an equally stepped signal. The linearity or an intended logarithmic or other behaviour can be checked by that curves.

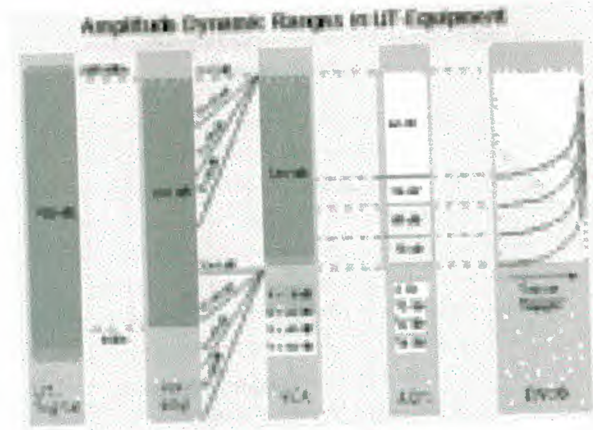

Figure 4.9 shows the dynamic range problem at the different stages of a receiver channel

4.4. Specialized Methods for the Probe Parameters

Most important to characterize the probes are directivity patterns at the different angles and focussing parameters. (Fig.4.10). With the arrangement presented in Fig.4.3 (right) an example of the patterns for a 0° - L-wave probe, covering several angles of incidence has been measured. The directivity patterns in Fig.3.10 have grating lobes, which are as more important as the main lobe is decreasing. Those grating lobes appear especially at the limiting edges of the usable ranges for the incidence or the skewing angles.

From the dependency of the maximum amplitude of the main lobe from the angle of incidence or the angle of skewing one may derive sensitivity correction curves in order to guarantee for each angle of incidence a constant defect detectability. The angles choosen by the controling software have to be compared with the ones determined by the measurements at a block according to fig.4.3. The curves in fig.4.11 are giving an example for the resulting skewing (Ø=phi) and incidence (α=alpha) angles and for the sensitivity changes at the different angles for the case of a skewing angle beam probe.

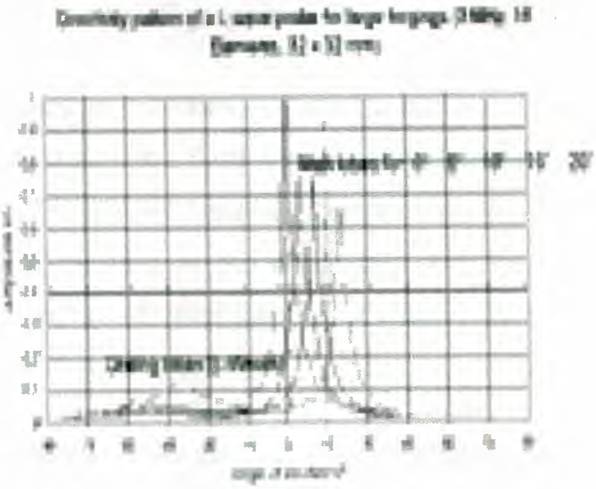

Figure 4.10 The directivity patterns have grating lobes

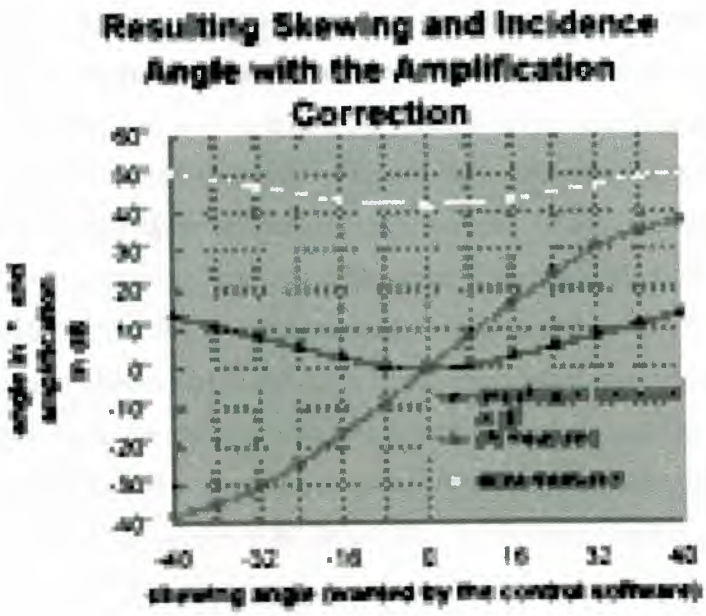

Figure 4.11 an example for the resulting skewing (Ø=phi) and incidence

4.5. Specialized Methods for the Equipment Parameters

There are several different possibilities to check the equipment parameters, either at the laboratory or at on site conditions. Some checks can be based on a special control software without the use of additional hardware components. By this the following parameters can be checked: crosstalk at the transmitter- and the receiver channels (switching some channels on or off) functionality of probe-, cable- and equipment channels (e.g. using internal probe echos).

Most of the software based checks are part of an integral equipment surveillance and give a good first impression on the functionality of all single channels. But they in general are not able to indicate quantitatively degraded parameters like the crosstalk behaviour or the sensitivity variations at the different channels. The digital A-scan converter and the data aquisition system can also be checked by a suitable control software, e.g. testing the linearity of the A-scan converter by a stepwise amplification variation using a reference echo from a testblock, which can easily be compared with a stored reference curve. If a more quantitative figure about the different parameters is required e.g. in order to trace aging influences it is nescessary to use a special hardware support. The Fig.4.12 top shows a hardware adaptation box, which has been conceived at BAM. This box acts as a probe replacement and produces with the aid of a function generator different pulse shapes and amplitudes. It allows to measure quantitatively the different transmitter-pulses, the receiver channel transfer functions and the delay time influence. Fig. 4.12 (middle and bottom) are showing the transmitter pulses produced by that box and an artificial echo sequence of sinus bursts at different amplitudes with a 6 dB step in between. The amplitude accuracy of this test hardware can be better than 0.1 dB allowing its use as a reference source for a quantitative equipment evaluation. Most of the specialized methods for the equipment parameters can be carried out as a quick surveillance, that means that they can be a build-in part of an inspection procedure. Even the hardware supported checks can be carried out very easily because all of the different checks with the parameter box are software based and do not need more than 5 or 10 minutes. Of course most available equipment are provided with some build-in self-check procedures, e.g. carried out during the switch on procedure of an equipment.

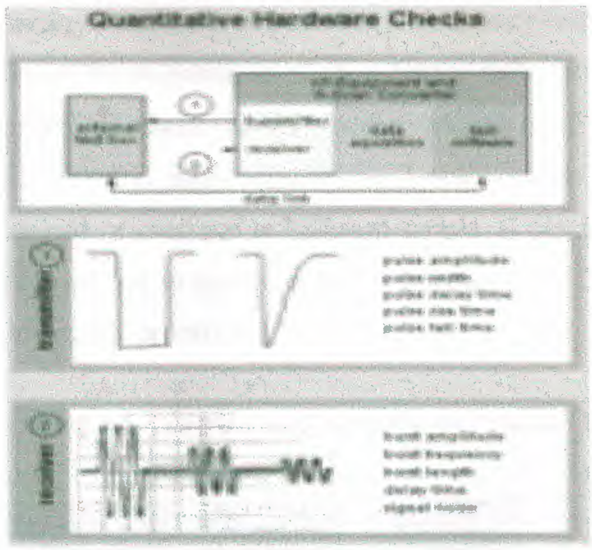

Figure 4.12 a hardware adaptation box, which has been conceived at BAM

4.6. Surveillance of the Control Software

A very important factor for the reliable inspection by ultrasonic phased array equipment is the quality assurance of the used software. A first basic check of a control software is of course the verification of the intended skewing or inclination angles and of the intended focussing and also of the relationship between grating lobes and main lobe. This can be done either by the directivity pattern measurement device in fig. 3.3 or with specialized test blocks containing suitable side drilled holes. An interesting check could also be based on an auto-focussing procedure searching automatically the optimal angle and focussing at side drilled holes in order to compare it with the values calculated by the controling software. Another problem for the software check is the surveillance of the administration purposes of a control software. The control software of a phased array equipment has not only to guarantee at the different probe positions the required angles and focussing values, but also the amplification, the individual TGC curves, the sensitivity correction values among others have to be selected from look-up tables and the data of the A-scan converter have to be stored on files allowing the off line evaluation. The correctness of this administration task of the control software can only be verified with measurements at suitably tailored test blocks together with an automatic scanning.

Fig.4.13 gives an example of TD-scans measured at a block with side drilled holes at three different angles. The different positions of the reflector indications at the TD-Scans are allowing an immediate check of the correct operation of the software structure. In order to detect possible influences of the A-scan conversion and data storage rate, the scanning should be executed at different speeds. A carefully designed software quality assurance system has to guarantee that a software once checked is not changed during the whole inspection procedure.

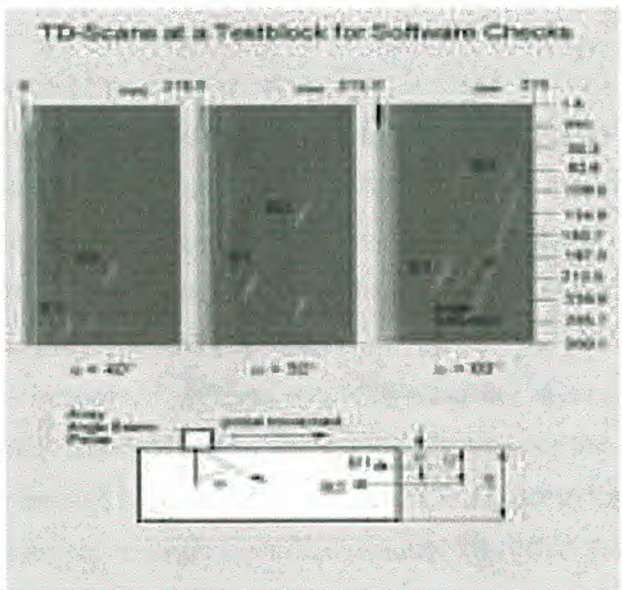

Figure 4.13 an example of TD-scans measured at a block with side drilled holes at three different angles

CHAPTER FIVE

ELECTRONIC SCANNING AND THE PHASED ARRAY

5.1. Electronic Scanning and the Phased Array

5.1.1. Introduction

The potential for increased target handling capacity available in Track While Scan radars is limited by the requirement to position the radar antenna mechanically.

Existing mechanical scanning methods are inherently slow and require large amounts of power in order to respond rapidly enough to deal with large numbers of high speed maneuvering targets. With mechanically scanned systems, antenna inertia and inflexibility prevent employment of optimum radar beam positioning patterns that can reduce reaction times and increase target capacity. With electronic scanning, the radar beams are positioned almost instantaneously and completely without the inertia, time lags, and vibration of mechanical systems. In an era in which the numerical superiority of adversaries is expected to remain large, electronic scanning can offset that advantage. The specific benefits of electronic scanning include:

(1) increased data rates (reduction of system reaction time),

(2) virtually instantaneous positioning of the radar beam anywhere within a set sector (beam position can be changed in a matter of micro-seconds),

(3) elimination of mechanical errors and failures associated with mechanically scanned antennas,

(4) vastly increased flexibility of the radar facilitating multi-mode operation, automatic multi-target tracking, highly directional transmission of missile guidance and control orders, interceptor and general air traffic control from one radar at virtually the same time.

5.1.2. Principles of Operation

The fundamental principles underlying the concept of electronic beam steering are derived from electromagnetic radiation theory employing constructive and destructive interference. These principles can be stated as follows:

The electromagnetic energy received at a point in space from two or more closely spaced radiating elements is a maximum when the energy from each radiating element arrives at the point in phase.

All elements are radiating in phase, and the resultant wave front is perpendicular to the axis of the element array. and subsequent diagrams show only a limited number of radiating elements. In actual radar antenna design several thousand elements could be used to obtain a high-gain antenna with a beam width of less than two degrees. The wave fronts remain perpendicular to the boresight axis and are considered to arrive at a point target in space at the same time. the path lengths from the elements to point *P* equalize as *P* approaches infinity. Thus, in situations where the target range is very large compared to the distance between elements, the paths from the elements to point *P* are almost parallel. Under these conditions, energy will arrive at point *P* with the same phase relationship that existed at the array. To achieve beam positioning off the boresight axis, it is necessary to radiated the antenna elements out of phase with one another. depicts the phase shift necessary to create constructive interference along a line joining an arbitrary point *P* with the center of the array. In order to achieve this constructive interference at Point *P*, the energy arriving from all radiating sources must be in phase and arrive at the same time. The energy from element *e1* must travel a path length *r1*, while the energy from element *e2* must travel the longer path length *r2*. The electric field magnitudes from the two elements at point *P* are given by equation (1-5) as follows:

$$E1 = E0 \sin \underline{2}(r1 - ct) + 1 \qquad (5\text{-}1)$$

$$E2 = E0 \sin \underline{2}(r2 - ct) + 2 \qquad (5\text{-}2)$$

Since these two electric fields must be in phase at point *P* for constructive interference to occur, the arguments of the sine functions must be equal.

$$2(r1 - ct) + 1 = \underline{2}(r2 - ct) + 2 \qquad (5\text{-}3)$$

$$1 - 2 = \underline{2}(r2 - ct) - \underline{2}(r1 - ct) \qquad (5\text{-}4)$$

thus

$$1 - 2 = = \underline{2}(r2 - r1) \qquad (5\text{-}5)$$

The path length difference, *r2 - r1,* approaches the value *d* sin as the distance *R* in increases. where point *P* has been moved an infinite distance from the source, the paths from the two sources have become parallel and the path length difference is exactly *d* sin . For points at a distance *R* less than infinity, *d* sin is still a good approximation for the path length difference, *r2 - r1,* as long as *R* is large compared to the element spacing, *d.* Thus, applying this distance approximation yields

$$r2 - r1 = dsin\ (R>d) \tag{5-6}$$

For practical radar applications, *d*(the distance between elements) is on the order of a few centimeters while *R* is on the order of kilometers, a difference of several orders of magnitude; therefore, the distance approximation is valid for all radar applications in the far radar field (ranges greater than 1 km).

Applying the distance approximation to the expression already obtained for the required phase difference between two radiating elements yield equation (5-6).

$$= \underline{2\ d}\,\sin \tag{5-7}$$

where:

= the phase shift between adjacent elements expressed in radians.

= the free space wavelength in meters.

d = the linear distance between radiating elements in meters.

= the desired angular offset in degrees.

5.1.3. Methods of Beam Steering

The previous discussion addressed the theory required to compute the relative phase shift between adjacent radiating elements in order to position the beam of an array-type antenna to a specific angle off of the antenna boresight axis. In practice there are three methods of accomplishing this phase difference.

5.1.4. Time Delay Scanning

The employment of time delay as a means of achieving the desired phase relationships between elements allows greater flexibility in frequency utilization than other methods.

However, in practice the use of coaxial delay lines or other means of timing at high power levels is impractical due to increased cost, complexity, and weight. To accomplish time delay scanning, variable delay networks are inserted in front of each radiating element. By proper choice of these time delays, the required effective phase shift can be applied to each element. The time delay between adjacent elements required to scan the beam to an angle, , is given by:

$$t = \underline{d}\ sin \qquad (5\text{-}8)$$

5.1.5. Frequency Scanning

One of the simpler methods of phased-array radar implementation is frequency scanning. This method is also relatively inexpensive. The length of the serpentine wavelength line (*l*) is chosen such that for some center frequency, *f0*, the length of signal travel between elements is an integral number of wavelengths, or

= *n* (*n* = any integer greater than zero)

where

0 = wavelength in the serpentine line at frequency *f0*.

Thus, when the excitation frequency is *f0,* the serpentine line will cause all elements to radiate in phase, and the beam will be formed along the boresight axis. If the excitation frequency is increased or decreased from the center frequency, *f0*, the line length, *l*, will no longer represent an integer number of wavelengths. As the excitation energy travels along the serpentine line, it will reach each successive radiating element with a uniformly increasing positive or negative phase shift. This results in the beam being deflected by an angle from the boresight axis. Thus, by varying the radar transmitter frequency about some base frequency, the beam can be positioned in one axis. the frequency scanned array has been simplified to a two- element system with the boresight axis normal to the plane of the elements. The feed is folded into a serpentine form to allow close element spacing while maintaining the required line length (*l*) between elements.

RF energy at 5,000 MHz is fed at the top of the array, and the elements are separated by distance *d* equal to (.03 meters). At time t = 0, the energy enters the serpentine feed line, and antenna *A1* radiates immediately starting at zero phase.

Since the period of the wave form is $T = 1/f$, then T = 1/5,000 MHz or 200 sec. Therefore, it takes 200 sec for one wavelength to propagate from *A1*. If the distance *l* traveled in the serpentine feed between *A1* and *A2* equals one wavelength or any integer number of wavelength (*L = n* where *n = 1, 2, 3, . . .),* then

$$t = \underline{l} = \underline{n} = n\,\underline{c} = \underline{n} = nT\ c\ c\,f\,f\,c \qquad (5\text{-}9)$$

where

t = elapsed time, and

T = the period of the wave form

Therefore, the energy from *A2* will always be in phase with *A1*. The beam formed by the array will be on the boresight axis. Note that this represents a broadside array (the elements transmit in phase). If the frequency is changed to 5,500 MHz, the period T becomes 181.81 sec; however, the wave form still takes $t = l/c$ or 200 sec longer to reach *A2* than to reach *A1* when fed into the serpentine line as depicted in figure 7-8. Note that the wave form from *A2* is no longer in phase with *A1*. Energy from *A2* lags *A1* in phase by 200 sec - 181.81 sec or 18.19 sec. The amount of phase shift can be determined by:

$\underline{t}$ = T 2 (radians) <u>18.19 sec</u> = 181.81 sec 2 radians = (.1)(2) = 0.6286 radians or 36o

Since there is a phase difference, , then the beam axis can be located as follows:

= <u>2 *d*</u> sin 0.6286 = <u>2 (.03m) sin</u> (.0545m) sin = 0.1817 = 10.47o

In this illustration the distance from *A1* to a point on space (*R1*) is greater than from *A2* to that point (*R1*). The wave forms will arrive in phase at the point because *A2* lags *A1* when the energy is transmitted. Thus, as frequency is varied, the beam axis will change, and scanning can be accomplished in one axis (either elevation or azimuth). The principles are employed in the AN/SPS-48 and AN/SPS-52 series radars as well as in the older AN/SPS-39. Variation in frequency tends to make these radars more resistant to jamming than they would be if operated at a fixed frequency, and it also provides a solution to the blind speed problem in MTI systems.

Frequency scanning does impose some limitations in that a large portion of the available frequency band is used for scanning rather than to optimize resolution of targets.

Additionally, this imposes the requirement that the receiver bandwidth be extremely wide or that the receiver be capable of shifting the center of a narrower bandwidth with the transmitted frequency.

Equation (5-3) gives the relationship between the percentage variation in frequency (bandwidth) and the scan angle, which is referred to as the Wrap-up Ratio.

The wrap-up ratio is the ratio of the sine of the maximum scan angle to the percentage change in frequency required to scan.

5.1.6. Phase Scanning

In a phase-scanned radar system, the radiating elements are fed from a radar transmitter through phase-shifting networks or "phasers.". The aim of the system is again to position the beam at any arbitrary angle, , at any time. In this case the means of accomplishing the phase shift at each element is simply to shift the phase of the incoming energy to each element. These phasers are adjustable over the range 0 to + 2 radians. The task of the system is to compute the phase shift required for each element, and set each phaser to the proper value to accomplish the desired beam offset. While phase scanning is more expensive than frequency scanning, it is much less expensive (in dollars, weight, and power losses) than time delay steering. To a first approximation, the bandwidth of a phase-scanning antenna in % of f0 is equal to the normal beamwidth in degrees. Thus a 1o beamwidth antenna working at 10 GHz can radiate over a 100 MHz bandwidth (i.e., +50 MHz) without distortion.

Energy reception. To receive energy transmitted in a steered beam by any of the three scanning methods, the applied frequency, time, or phase relationships are maintained at each element, which has the effect of making the radar sensitive to energy from the direction of transmission. Thus, each of the scanning methods is completely reversible and works equally well in reception of energy as in transmission. Unfortunately, some beam positions do not have the same phase shift in the reverse direction. When non-reciprocal phase shifters are used, it is necessary to change the phase-shifter setting between transmit and receive to maintain the same phase shift on receive as was used on transmission.

5.1.7. Computation of Required Phase Relationships

No matter which of the three possible methods of phase scanning is used in a phased array system, the objective is a relative phase shift of the energy being radiated by each element in the array. The incremental phase shift required between two adjacent elements is given by equation (5-1). When using this equation is will be assumed for consistency that represents the phase lead given to each element with respect to its adjacent element in the direction of the chosen reference element. Thus, when positioning the beam to the same side of the boresight axis as the reference element, each array element must lead the next element closer to the reference by the same amount, >0. When positioning the beam on the opposite side of the boresight axis from the reference by the same amount, <0 (negative or effective phase lag). This convention can be extended to include a sign convention for the angle . Choose the reference element as the top most element and the farthest right element when looking from the antenna along the boresight axis. Also choose elevation angles as positive above the boresight axis and negative below, and choose azimuth angles as positive in a clockwise direction (scanning to the right) and negative in a counterclockwise direction. (Do not confuse the above definitions of lead/lag in the spatial domain with the electrical definitions of lead/lag in the radian domain.). In order to position the beam above the boresight axis, the angle will be positive and thus sin will be positive also. This yields a positive between elements. To determine the phase applied to each element , simply use the relationship

e= *e* , *e* = 0,1,2, . . .

Similarly, to position the beam below the boresight axis requires the use of a negative angle . This yields a negative and equation (5-4) again yields the applied phase of each element.

If the reference element has a phase of zero, compute the phase applied to element five when the beam is scanned 40o above the boresight axis.

Find : = c = .06m

Find : = 2 *d* sin = 2 (.03) sin 40o .06 = 2.02 radians

Find : = *e* = 5(2.02 radians) = 10.10 radians

Note that this result is greater than 2 radians. In practice, the phasers can only shift the phase of the energy going to an element by an amount between -2 and 2 radians.

Expressed mathematically

-2 < < 2

Thus, the phase shift applied to element five must be

= 10.10 radians - 2 radians = 3.82 radians

In other words, element five must lead the reference element (element zero) by an amount of 3.82 radians.

Example 2. A Hypothetical Three-Dimensional Search Radar

In this example, azimuth information is obtained by rotating the antenna mechanically in a continuous full-circle search pattern. Range information is obtained in the standard way by timing the pulse travel to and from the target. The phase shifters (0), (1), (2), (3) control the elevation position of the beam. It is desired to control the elevation of the beam in 0.0872 rad (5o) steps from +1.047 rad (60o) to -0.2617 rad (15o) with respect to the antenna boresight axis, which is +0.2617 rad (15o) displaced from the horizontal. The system operational parameters are as follows:

Antenna rotational speed---10 rpm

Pulse repetition rate (PRR)---400 pps No. of elevation beam positions---16 No. of pulse per beam position---2

For each beam position the amount of phase shift is calculated for each radiating element. The resultant phase shifts are applied, and then two pulses are transmitted and received. This example is, of course, hypothetical. Some operational 3-D radars, such as the Marine AN/TPS-59 function similarly. It is important to note that the concept of controlling beam position by varying the relative phase of radiating elements is common to frequency-scanned, phase-scanned, and time-delay scanned arrays. The difference is in the methods employed to achieve the proper phase relationships between the radiating elements.

Example 3. Full Three-Dimensional Phased Array-Radar

The logical extension os the simple system of example 2 is the realization of a fully operational three-dimensional phased-array radar system to compute direct radar beams in elevation and bearing. Such a system is used in the Ticonderoga-class Aegis cruisers, in the AN/AWG-9 radar for the F-14 aircraft, in the Air Force COBRA DANE surveillance system, and in the Army Patriot missile acquisition and guidance radar.

The task now is to position the beam in both elevation and azimuth by electronic means. As in the two-dimensional case, the phase shift for each element in the array must be computed and applied prior to transmitting a pulse of radar energy. The array is made up of many independent elements. A unique element can be designated by the subscripts (*e,a*) for the *eth* row and *ath* column. The equations governing the positioning of the beam are presented below. Elevation Above or Below the Boresight Axis (Elevation scan)

= *e 2d*(sin EL)/ (5-5) Azimuth to the Right or Left of the Boresight Axis (Azimuth scan)

= *a 2d*(sin AZ)/ (5-6)

The phase shift for each unique element is simply an additive combination of the above equations.

= e + a = *e2 d*(sin EL) + *a2 d*(sin AZ) = 2 *d* [*e*(sin EL) + *a*(sin AZ)]

The example has been patterned after radars that use phase scanning in both directions (elevation and azimuth). Other systems have been designed using a combination of array scanning systems. The AN/SPS-33 radar, which is no longer in use, was frequency scanned in elevation and phase scanned in azimuth.

5.1.8. Synthetic Aperature Radar

The synthetic aperture radar (SAR) is discussed here because of its similarities with conventional linear array antennas. SAR permits the attainment of the high resolution associated with arrays by using the motion of the vehicle to generate the antenna aperature sequentially rather than simultaneously as conventional arrays. As an example of SAR. The eight elements in the figure will now represent points in space where the platform is located when the radar radiates energy as it travels from point seven to point zero. At each point along the path, data is gathered from the echos received, and this information is stored. Upon collecting the data at position zero, all the stored data from positions one through seven are combined with the data from position zero and processed as the data would be from an eight-element linear array with simultaneous inputs from all elements. The effect will be similar to a linear-array antenna whose length is the distance traveled during the transmission of the eight pulses.

The "element" spacing of the synthesized antenna is equal to the distance traveled by the vehicle between pulse transmissions. SAR, commonly used in aircraft and more recently in ships, is sometimes called side-looking radar or SLR. In SLR, the platform travels in a straight path with a constant velocity. Its radar antenna is mounted so as to radiate in the direction perpendicular to the direction of motion. SARs in this configuration can be used to gain imaging data of the earth's surface in order to provide a maplike display for military reconnaissance, measurement of seastate conditions, and other high resolution applications.

5.1.9. Other Considerations

This has been a highly simplified treatment of electronic scanning and the phased-array radar. In addition to positioning the main lobe of energy, other considerations are:

(1) The suppression of side lobe interference.

(2) Array element excitation amplitudes and geometry to achieve various beam radiation patterns.

(3) Combining the phased array with a track-while-scan function.

(4) Modulating the radiated energy to transmit information (for missile guidance commands, etc.)

In practice there are limits to the useful angular displacement of an electronically scanned radar beam. One limit is caused by the element pattern. The antenna pattern of an array is the product of the array pattern and the element pattern. In the simple examples given in this section we have assumed that the element pattern was omnidirectional. A practical array element pattern is not omnidirectional, so the elements limit the scan angle. Another limit is caused by the element spacing. A large scan angle requires a close element spacing. If the scan angle exceeds that which can be accommodated by the element spacing, grating lobes will be formed in the other direction.

5.2. Principles of Underwater Sound

5.2.1. Principles of underwater sound

The effectiveness of the present-day submarine depends upon its ability to remain undetected for long periods of time while it searches, tracks, or attacks from beneath the sea surface. This medium of concealment, however, is advantageous to the submarine only so long as it is not detected or deprived of its ability to detect. Before a submarine can be attacked, it must be detected and its subsequent positions determined within the requirements of the available weapons system. Detection and position fixing can take place in two ways. There may either be some radiation or reflection of energy from the submarine to the searcher, or else the submarine may disturb one of the natural, static, spatial fields, such as the earth's magnetic field, thereby betraying its presence.

The choice of energy to be used for underwater detection is determined by three factors:

1. Range of penetration in the medium.
2. Ability to differentiate between various objects in the medium.
3. Speed of propagation.

Of all the known physical phenomena, light has excellent differentiation ability and high speed of transmission, but its range in water is very limited, on the order of tens of meters, thereby restricting its operational usefulness. This is not to say that light will never be used in ASW, for continuing experimentation and perfection of laser light detectors might yet add this method to the arsenal. Radio frequency waves also are propagated with extreme rapidity and to great distances through certain mediums, but sea water is essentially impervious to them for most frequencies. VLF signals will penetrate only about 10 meters, whereas higher frequency penetration depths can be measured in millimeters. Magnetic and gravitational field distortions are detectable only at very short ranges because the anomaly diminishes proportionally with the inverse of the range cubed. While their detection range is greater than either light or radio frequency, it is only of the magnitude of several hundred meters and therefore is insufficient for normal surveillance.

Acoustic energy, while lacking the propagation speed of electromagnetic waves, is capable of being transmitted through the sea to distances that are operationally significant. Because of this, sound is the physical phenomenon used for antisubmarine warfare, underwater communications, and underwater navigation. It must not be inferred, however, that sound is a panacea. It too has significant limitations to its effective employment, all of which must be thoroughly understood by the operators of underwater sound equipment. The optimum use of sound requires a thorough understanding of its limitations so that these effects can be minimized. For example, sea water is not uniform in pressure, temperature, or salinity, and all these characteristics have important effects on sound propagation through the sea. The requirement for predicting these effects on sonar performance has become a necessity, and a difficult one at that.

5.2.2. Fundamental Concepts

All sound, whether produced by a cowbell or a complicated electronic device, behaves in much the same manner. Sound originates as a wave motion by a vibrating source and requires for its transmission an elastic medium such as air or water. For example, consider a piston suspended in one of these mediums. As the piston is forced to move forward and backward, the medium is compressed on the forward stroke and decompressed or rarefied on the return stroke. Thus, a wave motion or series of compressions and rarefactions is caused to move from the source out through the medium. In the fluid medium the molecular motion is back and forth, parallel to the direction of the piston's movement. Because the fluid is compressible, this motion results in a series of detectable pressure changes. This series of compressions and rarefactions, such as is produced by the piston, constitutes a compressional wave train. Another way of explaining the phenomenon of acoustic wave propagation is to consider the medium of transmission as a loosely packed collection of mass elements connected by springy bumpers. A disturbance of the elements at some point (e.g., piston motion) moves along in the fluid by the successive extension and compression of the springs as the elements swing back and forth, each communicating its motion to its neighbor through the connecting bumpers. In this way, the agitation of a cluster of elements is propagated through the medium even though the individual elements do no more than move about their equilibrium positions without actually migrating.

The sound wave propagates parallel to the source resulting in a longitudinal wave. Recall from radar principles that the electromagnetic wave propagated perpendicular to the source, resulting in a transverse wave.

5.2.2.1. The Sound Wave

As the wave passes a given point, the fluid elements are compressed and extended in the manner depicted by the sine wave's oscillations above and below the static pressure. The compressions and rarefactions are so labeled on the curve. There are two important things to note on this curve. The first is the maximum amplitude of the sine wave, labeled P, which represents the maximum pressure excursion above and below the static or hydrostatic pressure that exists in the fluid at the location of the wave train. The second thing to note is that the wave train would be passing, or propagating at the speed of sound **c** in the medium.

One final point should be made about the sine wave representation Though somewhat difficult to imagine and more difficult to describe pictorially, the displaced parallel line a represent the motion of the elements within the field as the wave train passes. As the elements are compressed and extended, their motion can also be mathematically described by a sine wave; however, the elements would be oscillating to and fro about their static position. The maximum amplitude would then be the maximum displacement from the static position. To provide an example of the order of magnitude of these displacements, consider that the faintest 1,000 Hz tone that can just be heard in air has pressure variations of only 2/10,000,000,000 of one atmosphere of pressure. The corresponding particle displacement is about 10-9 cm. By the way of comparison, the diameter of an atom is about 10-8 cm.

As this pressure disturbance propagates through the medium, the pressure at any point in the medium can be expressed as a function of the distance, r, from the source and time, t, since a particular wave departed the source.

$$P(r,t) = P(r)\sin | \underline{2} (r - ct) | (5\text{-}2) | | \qquad (5\text{-}10)$$

5.2.2.2. Intensity

A propagating sound wave carries mechanical energy with it in the form of kinetic energy of the particles in motion plus the potential energy of the stresses set up in the elastic medium. Because the wave is propagating, a certain amount of energy per second, or power, is crossing a unit area and this power per unit area, or power density is called the intensity, I, of the wave. The intensity is proportional to the square of the acoustic pressure. Before giving the defining relationship for intensity, however, two variables must be explained. The value of peak pressure, P, as shown in figure 8-1 is not the "effective" or root-mean-square pressure. An analogy exists between acoustic pressure and the voltages measured in AC circuits. Most voltmeters read the rms voltage. The rms value of a sinusoidal voltage is simply the peak voltage divided by the square root of two.

5.2.2.3. Characteristic Impedance

The second variable that must be explained is the proportionality factor that equates intensity to effective pressure squared. It consists of two terms multiplied together - fluid density, and the propagation speed of the wave, c. The quantity, c, is called the characteristic impedance; it is that property of a sound medium that is analogous to resistance or impedance in electrical circuit theory, where power equals voltage squared divided by resistance. Additionally, it can be illustrated by a simple example: When two dissimilar mediums, such as air and water, are adjacent to each other, the boundary between the two is called a discontinuity. When sound energy is traveling through one medium and encounters a discontinuity, part of the energy will be transferred across the boundary and part will be reflected back into the original medium. The greater the difference between the characteristic impedances, the greater will be the percentage of energy reflected. (The difference between the c values for air and water in SI units is approximately 1.5 x 106.) Thus, when sound is traveling through water and it reaches the surface, only a small amount is transmitted into the air. Most of the energy is reflected by the air/ocean boundary back into the water. Obviously, it is important to maintain a consistent set of units when comparing characteristic impedances, and care must be exercised when dealing with different sources of acoustic information.

5.2.2.3. Measurement of Acoustic Parameters

A convenient system is needed in order to measure and discuss acoustic parameters. Pressure is defined as a force per unit area. Although many people are familiar with the British units of pounds per square inch (psi), it has long been the convention in acoustics to use metric units, namely newtons per square meter (N/m2), or dynes per square centimeter (dynes/cm2). Of the two metric units, the dynes/cm2 has been the most commonly used. It has an alternate name, microbar (bar), and is equivalent to approximately 1/1,000,000 of a standard atmosphere. For underwater sounds, a reference pressure of 1 bar was established from which all others were measured. The corresponding reference pressure for airborne sounds was 0.0002 bar, because this was the approximate intensity of a 1,000-Hz tone that was barely audible to human ears. The previously less commonly used N/m2 also has an alternate name, a Pascal (Pa), and the reference standard derived from this was the micropascal (Pa), which is equivalent to 10-6N/m2.

With such a profusion of reference standards and measurement systems, there were ample opportunities for misunderstandings as an operator or planner consulted different sources of acoustic information. In 1971 the Naval Sea Systems Command directed that thereafter all sound pressure levels should be expressed in the Systeme Internationale (SI) units of micropascals. Although all new publications contain the updated standards, older references will not until they are revised.

5.2.2.4. Sound Pressure Level

In theoretical investigations of acoustic phenomena, it is often convenient to express sound pressures in newtons/m2 and sound intensities in watts/m2. However, in practical engineering work it is customary to describe these same quantities through the use of logarithmic scales known as sound pressure levels. The reason is related, in part, to the subjective response of the ear. The human ear can hear sounds having pressure disturbances as great as 100,000,000 micropascals and as small as 10 micropascals. A problem is encountered when discussing pressures that vary over so great a range, in that the minimum audible disturbance is one ten-millionth that of the maximum. In underwater acoustics, useful pressures having even greater variations in magnitude are commonly encountered.

In order to make the numbers more manageable, both in magnitude and for actual manipulation, logarithms are used rather than the numbers themselves. Suppose two acoustic signals are to be compared, one having a Pe, of 100,000,000 Pa and the other a Pe of 10 Pa. Their ratio would be

$\underline{P1} = \underline{100{,}000{,}000\ Pa} = 10{,}000{,}000 = 107\ P2\ 10\ Pa$

In underwater acoustics, however, the attribute of primary interest is sound intensity, or power, rather than pressure. As with pressure, acoustic intensities are referenced to some standard intensity, designated Io, and the logarithm of the ratio taken. Intensity level is therefore defined as

$$IL = 10 \log(I/Io) \qquad (5\text{-}11)$$

where IL is measured in dB. However, as there is only one intensity reference (10-12 watt/m2 in air) and many pressure references, IL must be able to be expressed in terms of pressure. By inserting equation (5-11) into equation (5-12), a new expression of IL can be obtained, which is based on pressure rather than intensity per sec.

$$Pe2\ IL = 10 \log pc \qquad (5\text{-}12)$$

or

$$IL = 10 \log \underline{Pe2} = 20 \log \underline{Pe}\ Po2\ Po \qquad (5\text{-}13)$$

Under the assumption that the reference intensity and the reference pressure are measured in the same acoustic wave, then a new sound level can be defined called sound pressure level.

$$SPL = 20\log \underline{Pe} \qquad (5\text{-}14)$$

Since the voltage outputs of the microphones and hydrophones commonly used in acoustic measurements are proportional to pressure, acoustic pressure is the most readily measured variable in a sound field. For this reason, sound pressure level is more widely used in specifying sound levels, and this is also why only pressure references are used in underwater acoustics. Note that IL and SPL are numerically equivalent.

$$IL = 10\log\underline{I} = 20\log\underline{Pe} = SPL\ \ Io\ Po \qquad (5\text{-}15)$$

where

$Io = P2\ \underline{o}\ pc$

5.2.2.5. Decibels

SPL has the dimensionless units of decibels (dB). The decibel system was selected by acousticians for a number of logical reasons. First, it is a logarithmic system, which is convenient for dealing with large changes in quantities. It also simplifies computations since multiplication and division are reduced to addition or subtraction, respectively. Second, human senses have an approximate logarithmic response to stimuli such as light, sound, and heat. For example, the human ear perceives about the same change in loudness between 1 and 10 units of pressure as it perceives between 10 and 100 units of pressure. And finally, in the area of underwater acoustics, the primary interest is in ratios of power levels and signal levels rather than absolute numerical values. In the decibel system, the bel is the fundamental division of a logarithmic scale for expressing the ratio of two amounts of power. The number of bels to express such a ratio is the logarithm to the base 10 of the ratio. Acousticians decided the bel was a unit too large for application in their field, and subsequently adopted the decibel (1/10 bel) as their basic logarithmic unit. The conversion factors in table 8-1 can in themselves be cumbersome to use, but when expressed in dB, only addition or subtraction is required. When converting from a pressure referenced to 1 bar to one referenced to 1 Pa, simply add 100dB. When converting from 0.0002 bar to 1 Pa, simply add 26dB. If converting from 1 Pa to the others, merely subtract the appropriate values. Note that the new micropascal reference standard is small enough that negative values of decibels are rarely encountered. As an aid in interpreting and understanding the decibel scale and its relation to intensity and pressure, it is useful to remember that

a factor 2 in intensity is + 3 dB
a factor of 0.5 in intensity is -3 dB
a factor of 10 in intensity is + 10 dB
a factor of 0.1 in intensity is -10 dB
a factor of 2 in pressure is + 6 dB
a factor of 0.5 in pressure is -6 dB
a factor of 10 in pressure is + 20 dB
a factor of 0.1 in pressure is -20dB

In understanding intensity levels and sound pressure levels, it is important to note that the decibel scale is a ratio of power or energy, no matter what quantities are being ratioed. A problem commonly arising in acoustic calculations is that of obtaining the overall intensity level after the individual intensities within the applicable bandwidths have been calculated. Such a situation is encountered in calculating a term in the sonar equations (to be discussed later) called noise level, which is actually a combination of ambient noise and self-noise. Because we are dealing with decibels, it is not possible to merely add similar intensity levels together and work with their sum. For example, two 30 dB signals combine to give a total intensity level of 33 dB, not 60 dB as might be expected. The reason for this is that, as shown above, doubling the intensity is represented in decibels by a + 3 dB change. The process is more complicated when dealing with levels of unequal value. This process can be expanded to include any number of intensity levels. However, when dealing with more than two intensities, it is often easier to use anti-logs, to convert each IL back to its intensity units, add all the intensities together, then reconvert to dB levels. Either method may be used and should result in the same numerical value of noise level.

Although standardization has been reached for measuring intensities, such is not the case for other quantities. Ranges are expressed in yards, kilometers, and nautical miles. Depths are given in feet, meters, or fathoms. Sound speed is stated in feet per second or meters per second and ship speed in knots. Temperatures are commonly specified in degrees Fahrenheit or degrees Celsius. These diverse units should warn the user to exercise due caution when discussing the various facets of underwater sound to ensure that misunderstandings do not occur. In this text, the SI units of measure will be used wherever possible.

5.2.2.6. Speed of Sound in the Sea

From physics it will be remembered that when gas is the transmitting medium, the denser the gas, the slower the speed of sound, and yet the speed of sound in water is about four times greater than that in air. Although this seems contradictory, it is not, because there is another more important factor that influences the speed of sound. In truth, the speed of sound is determined primarily by the elasticity of the medium and only secondarily by the density.

5.2.2.7. Bulk Modulus

Elasticity is defined as that property of a body that causes it to resist deformation and to recover its original shape and size when the deforming forces are removed. Of specific concern is volume elasticity or bulk modulus - that is, the ratio of force per unit area (stress) to the change in volume per unit volume (strain).

Thus, Bulk Modulus = Stress Strain

In order to bring about a change in the volume of a liquid, it is necessary to exert a force of much greater magnitude than is required to bring about an equivalent change in the same volume of air. Therefore, the value of bulk modulus is much greater for a liquid than for a gas. This bit of information, however, is meaningless until it is applied in the formula for the speed of sound.

The speed of sound, c, in a fluid is equal to the square root of the ratio of bulk modulus to density. Thus,

c = Bulk Modulus Density

Although seawater is almost a thousand times denser than air, the enormous bulk modulus of water is the more important factor determining sound speed. Of concern, however, are not the differences of the two mediums but the conditions in water that cause changes in sound speed. Contrary to the assumptions made up to this point, the ocean is not a homogeneous medium, and the speed of sound varies from point to point in the ocean. This variation in sound speed is one of the most important characteristics affecting the transmission of sound. The three main environmental factors affecting the speed of sound in the ocean are salinity, pressure, and temperature.

5.2.2.8. Salinity

Salinity, which on the average ranges from 32 to 38 parts per thousand (ppt), is fairly constant in the open ocean. A change of salinity will cause a small corresponding change in density with a resulting change in bulk modulus, causing variation of sound speed. The greatest variation in salinity in the open ocean exists in the vicinity of "oceanic fronts," which are narrow zones separating water masses of different physical characteristics, usually exhibiting very large horizontal gradients of temperature and salinity .

Even greater variation in salinity can be expected around the mouths of rivers, heavy ice, and in areas of extraordinary rainfall (e.g., the monsoon) where a layer of fresh water overrides a layer of salt water. A change in salinity of one part per thousand will result in a change in sound speed of approximately 1.3 meters per second.

5.2.2.9. Pressure

Pressure in most circumstances is more important than salinity, but in the sea its change is constant and thus predictable. It also causes a change in bulk modulus and density, and the result is an increase in sound speed of 0.017 m/sec for every meter of depth increase. This slight change, which is important when temperature remains constant, causes a sound beam to bend upward at great depths as will be discussed later.

5.2.2.10. Temperature

Temperature, the foremost factor affecting sound speed, usually decreases with depth, and this leads to an accompanying decrease in sound speed at the rate of approximately 3 m/sec per degree Celsius. Below a depth of about 1,000 m, however, temperature is fairly constant, and the predominant factor affecting sound speed becomes pressure. At first glance it would seem that a temperature decrease would increase sound speed due to the increased water density, but not so. As the temperature of a medium decreases, bulk modulus decreases while density increases. Considering these effects in terms of the sound speed formula in equation (5-10), it is clear that a decrease in temperature brings an attendant decrease in sound speed. It also should be noted that temperature differs bulk modulus and density at a variable rate. A change in temperature at one point on the scale, therefore, affects sound speed differently than an equal change at another point on the scale. It should be noted that the effect of temperature is relatively large compared to the other factors.

It takes a depth change of about 165 meters to cause the same change in sound speed as a one-degree temperature change. As will be discussed, temperature is therefore the only factor normally measured and evaluated under operational conditions.

5.2.2.11. Sound Speed Equation

Dealing with these three factors to arrive at values for bulk modulus and density, and thence sound speed, is very cumbersome. To overcome this, numerous empirical relationships have been developed for converting the three factors directly to sound speed. A simplified version of such sound speed equations developed by Wilson in 1960 is present below.

$$c = 1449 + 4.6T + 0.055T2 + 0.003T3 + (1.39 - 0.012T)(S - 35) + 0.017d \quad (8\text{-}11)$$

where

T = temperature in degrees Celsius

S = salinity in parts per thousand

d = depth in meters

Given accurate temperature, salinity, and depth data, this equation is accurate within 0.6 meters/sec, 96 percent of the time. By way of contrast, the equation for the speed of sound in air is approximately

$$c = 331.6 + 0.6T$$

In making calculations involving the transmission of sound through the sea, it frequently is adequate to use a standard speed rather than the more accurate value given by equation 5-11. Although in seawater c can vary from a low of about 1,420 m/s to over 1,560 m/s depending on environmental conditions, a standard speed of 1,500 m/s may be assumed for computation purposes unless otherwise noted.

5.2.3. Field Observations of Sound Speed

Knowledge of sound velocity is important to the ASW tactician and physical oceanographer because of the effect that variations in sound velocity have upon acoustic absorption and refraction. Two different devices are in use today for finding the speed of sound in the sea.

The first device is called a bathy-thermograph. As previously stated, temperature is the predominant ocean variable affecting sound speed.

Not only is it relatively easy to measure, but when applied to empirical relationships such as equation (5-11), sound speed can be computed.

Older BT systems employed a mechanical device that was lowered on a cable and the temperature was scribed on a smoked piece of glass. This had a number of inherent disadvantages that have been overcome through the development of the expendable bathythermograph (XBT), which does not require retrieval of the sensing unit..It consists of a thermistor probe that is ejected from the launching platform and sinks at a known non-linear rate. The XBT is connected to a special recorder on board the launching platform by a fine wire. As it sinks, ther thermistor changes its electrical resistance with changing temperature, and as a result a temperature vs. depth trace is obtained. Because the wire uncoils from both the probe and its launcher, there is no tension on the wire until the probe has sunk to its full depth. At this point, the wire snaps and the recording stops. Variants of the basic XBT have been developed for use aboard submarines and from aircraft through incorporation into sonobuoys.

When the XBT temperature vs. depth trace is converted to sound speed vs. depth, it produces a sound speed profile very similar to that obtainable from a sound velocimeter, and is of sufficient accuracy for operational requirements.

5.2.3.1. Sound Velocimeter

The second and most accurate method is the sound velocimeter. Its principle advantage is that it can measure sound speed directly, without need for conversions, by transmitting a pulse of sound over a very short path on the order of 1/2 meter or less. When the pulse arrives at the receiver, another pulse is then triggered from the transmitter; this is known as the "sing-around" or "howler" principle. The faster the speed of sound in the water in which the velocimeter is submerged, the faster the pulse will travel and the sooner it will arrive at the receiver to trigger the succeeding pulse. Since nearly all the time delay between pulses occurs as acoustic delay in the water, the PRF of the pulses is determined by the local sound speed and is directly proportional to it. Thus, knowing the path length and observing the PRF can lead directly to computation of sound speed. Until recently sound velocimeters were expensive and awkward to use, thus eliminating their use tactical-ly. The recent development of the expendable sound velocimeter (XSV) has made it possible to reduce sound velocity measurement er-rors to less than .25 meters per second at reasonable expense with-out reduction of the mobility of combatant units.

Today's sophis-ticated sonars and acoustic navigation systems can provide improved information in many oceanic regions when actual sound-velocity pro-files are used rather than extrapolated sound velocity values based on temperature profiles and assumed salinity data. Based on the variability in the sea with time, a policy of regular periodic mea-surement of the velocity profile is required during an ASW opera-tion. Normally, one or two ships in the force are assigned the bathythermograph guard duty. These ships periodically measure the temperature or velocity profile, and promulgate it to all ASW units in company. These sound velocity profiles are essential in deter-mining the sound propagation paths available.

5.2.4. Typical Sound Speed Profiles

It is important to remember that while temperature is the dominant factor, the sound-speed profile is really a composite of the pressure, salinity, and temperature profiles as shown in the area of ocean fronts, where salinity may vary up to 3 ppt. from assumed values, the use of temperature data alone may result in an error of up to 4.2 meters per second in the calculation of sound speed.

The profile may be divided into four major layers each having different thermal characteristics. Just below the sea surface is the surface layer, in which the speed of sound is susceptible to daily and local changes of heating, cooling, and wind action. The surface layer may contain isothermal water that is formed due to mixing by the action of wind as it blows across the water. Below the surface layer lies the seasonal thermocline - the word "thermocline" denoting a layer in which the temperature changes rapidly with depth. The seasonal thermocline is characterized by a negative sound-speed gradient that varies with the seasons. During the summer and fall, when the near-surface waters of the sea are warm, the seasonal thermocline is strong and well defined; during the winter and spring, and in the Arctic, it tends to merge with, and be indistinguishable from, the surface layer. Underlying the seasonal thermocline is the permanent thermocline, which is affected only slightly by seasonal changes. Below the permanent thermocline and extending to the sea bottom is the deep isothermal layer, having a nearly constant temperature of about 4oC, in which the speed of sound has a

5.2.4.1. Positive Gradient Because of the Effect of Pressure on Sound Speed.

Between the negative speed gradient of the permanent thermocline and the positive gradient of the deep isothermal layer, there is a speed minimum toward which sound traveling at great depths tends to be bent or focused by refraction. This is the deep sound channel and will be discussed later. The refraction of sound, however, is much more complex than this simple four-layer ocean model would indicate. There are ocean eddies, fronts, interfaces between currents, underwater mountains and ridges. For instance, a submarine detected in the Labrador Current but crossing into the Gulf Stream has been compared to a person going out of an open field and disappearing into the nearby woods.

5.2.5. Ray Theory

The propagation of sound in an elastic medium can be described mathematically by solutions of the wave equation using the ap-propriate boundary and medium conditions for a particular problem. The wave equation is a partial differential equation relating the acoustic pressure P to the coordinate x, y, z, and the time t, and may be written as

$$\underline{2P} = c2\ (\underline{2P} + \underline{2P} + \underline{2P})\ (8\text{-}12)\ t2\ (x2\ y2\ z2) \qquad (5\text{-}15)$$

5.2.5.1. Normal-Mode Theory

There are two theoretical approaches to a solution of the wave equation. One is called normal-mode theory, in which the prop-agation is described in terms of characteristic functions called normal modes, each of which is a solution of the equation. The normal modes are combined additively to satisfy the boundary and source conditions of interest. The result is a complicated math-ematical function which, though adequate for computations on a digital computer, gives little insight, compared to ray theory, on the distribution of the energy of the source in space and time. Normal-mode theory is well suited for a description of sound prop-agation in shallow water, but will not be discussed in this text.

5.2.5.2. Ray Acoustics

The other form of solution of the wave equation is ray theory, and the body of results and conclusions therefrom is called ray acous-tics. The essence of ray theory is (1) the postulate of wave fronts, along which the phase or time function of the solution is constant, and (2) the existence of rays that describe where in space the sound emanating from the source is being sent. Like its analog in optics, ray acoustics has considerable intuitive appeal and presents a picture of the propagation in the form of the ray diagram.

For almost all operational problems, the sound-speed gradient, with respect to horizontal changes of location, can be assumed to be zero. The major gradient of interest is the vertical gradient, dc/dz, where dz is the amount of depth change. If a source of sound at the surface of the sea radiates omnidirectionally, a wave front expanding from this source in all directions transfers energy from one particle in the water to another, and by this means the wave is propagated. If some point on this wave front is selected, and from it a line is drawn in the direction of energy propagation, then connecting these points as the wave expands in space will result in a line called a ray. A sound wave, or ray, which enters another medium or layer of the same medium having a different characteristic impedance, will undergo an abrupt change in direction and speed. Depending upon the angle of incidence and the abruptness of change in c, a portion of the impinging acoustic energy will be reflected off the medium boundary, and a portion will be refracted or bent passing through the boundary. A sound ray will always bend toward the region of slower sound speed.

One of the most important practical results of ray theory is Snell's Law, which describes the refraction of waves in mediums of variable speeds. Snell's Law states that the angle of incidence, 1, at a boundary is related to the angle of refraction 2, by the following expression:

$$\underline{\sin 1} = \underline{c1}\ (5\text{-}13)\ \sin 2\ c2 \qquad (5\text{-}16)$$

where

c1 = sound speed in medium 1

c2 = sound speed in medium 2

If the wave is considered to be passing through three horizontal layers or strata, in each of which the sound speed is considered to be constant, then Snell's Law can be rewritten as

$$\underline{c1} = \underline{c2} = \underline{c3} = \underline{cn}\,(5\text{-}14)\ \cos 1\ \cos 2\ \cos 3\ \cos n \qquad (5\text{-}17)$$

where

cn = speed of sound at any point in the medium

n = angle made with horizontal at that point

Note that the angle in equation 5-14 is the complement of the angle usually expressed in Snell's basic law. It is commonly referred to as the grazing angle or angle of inclination. This expression is the basis of ray computation used by most computers, since it enables a particular ray to be "traced out" by following it through the successive layers into which the speed profile may have been divided. In a layered medium having layers of constant speed, the rays consist of a series of straight-line segments joined together, in effect, by Snell's Law.

In practice, however, temperature does not change abruptly, but rather the gradient will normally decrease or increase at a measurable rate. For such a situation, the sound speed at any depth z would be given by

c(z) = co + gz (5-15) where co = speed at the surface or transducer depth

g = speed gradient dc/dz between the surface and depth z

The net result is that, in reality, ray traces appear as curves rather than straight lines. By combining equations (5-14) and (5-15), an expression can be developed for the radius of curvature R of any ray at any point along the ray path, as shown by equation (5-16)

$$R = \underline{-co} = \underline{c}\ \ g\ g\cos \qquad (5\text{-}18)$$

Under operational conditions, values of R are very large, approaching several tens of kilometers.

5.3. Propagation Paths

5.3.1. Thermal Structure

The thermal structure of the ocean governs the refractive con-ditions for a given water mass. Despite infinite vertical temperature variations in the ocean, the temperature structure normally can be related to three basic types: (1) isothermal,

(2) negative temperature gradient, and (3) positive temperature gradient. In discussing sound propagation, it is customary to use the temperature profile as an indicator of sound speed conditions at various depths, because it has the greatest effect. It must be remembered, however, that changes in sound-beam direction result from changes in the sound-speed profile, which is influenced not only by temperature but pressure and salinity as well.

5.3.2. Direct Path

In an isothermal condition, the water's temperature is almost constant. If there is a slight decrease in temperature, and it is just balanced out by the pressure increase, the result is an iso-sound-speed condition. This causes a straight-line ray, leaving the source in lines that continue with little or no change in angle. Long ranges are possible when this type of structure is present.

When there is a negative temperature gradient, sound speed decreases with depth, and sound rays bend sharply downward. This condition is common near the surface of the sea. At some horizon-tal distance from the sound source, beyond where the rays bend downward, is a region in which sound intensity is negligible ; it is called a shadow zone. The magnitude of the tem-perature gradient determines the amount of bending of the sound beam and thus the range of the shadow zone. For example, if the decrease in temperature to a depth of 10 meters totals 2oC or more, the shadow zone would begin beyond a horizontal range of 1,000 met-ers due to the sharp curvature of the sound beam. When the temperature of the water has a positive gradient, sound speed increases with depth, and sound rays are refracted upward. Longer ranges are attained with this temperature structure than with a negative gradient because the rays are refracted upward and then reflect off the surface.

Unless the surface of the sea is very rough, most of the rays are repeatedly reflected at the surface to longer ranges. Circumstances usually produce conditions where combinations of temperatures occur. One of these combinations includes a layer of isothermal water over water with a negative gradient. Approximately 90 per cent of all the bathythermograph records from all over the world show this type of thermal structure. One ray, labeled "the critical ray," becomes horizontal at the boundary or division between the isothermal layer and the negative gradient. The speed of sound is a maximum at this boundary point.

Consequently, we define the layer depth (z) as that depth of greatest sound speed (c) above the seasonal thermocline.One half of the critical beam bends toward the upper region at a reduced speed, and the other half bends toward the lower region at a reduced speed. The angle that the critical ray makes with the horizontal at the point of projection is called the critical angle. All rays in the sound beam directed at an angle less than the critical angle will follow paths entirely within the isothermal layer and will be bent upward to the surface. All rays directed at an angle greater than the critical angle follow paths that penetrate the boundary and are subsequently refracted downward. No rays enter the region bounded by the two branches of the split critical ray, and for this reason it is also called a shadow zone. Sharp shadow zones are not fully developed because of diffraction and other effects, though the sound intensity in this area is quite low. Submarine commanders deliberately use this phenomenon, when it exists, to attempt to escape detection when approaching a target. The optimum depth for close approach to a target with minimum probability of counter-detection is approximately Best depth = 17 Z where z is the layer depth in meters. This is accurate down to a layer depth of 60 meters. Below that, the best depth for approach is a constant 60 meters below layer depth.

5.3.3. Convergence Zone

In the deep ocean, temperature usually decreases with depth to approximately 1,000 meters. Deeper than this, temperature is a constant 4oC and sound speed increases as a result of pressure. A negative speed gradient overlays a positive speed gradient, allowing equal speeds at two different depths with slower speed conditions in between. Sound originating in the thermocline, traveling nearly parallel to the surface initially, thus bends toward greater depths.

But as the detected sound penetrates into deep water, it passes the region of minimum sound speed and enters the deep isothermal layer. Now the gradient in sound speed operates in the other direction; the sound path bends upward rather than downward, and the sound returns to the surface. This produces a convergence zone, where the sound waves concentrate as if they had been focused. It typically lies at a distance of about fifty kilometers from the source. Beyond this convergence zone is a second zone of silence, where again the acoustic waves diffract downward; then another convergence zone, fifty kilometers out, and so forth. The mapping of these zones is a routine part of submarine operations; by measuring deep-water temperatures witha bathythermograph, one obtains data that readily allow a computer to calculate the appropriate sound-wave paths. When water with a negative speed gradient overlays a positive speed gradient, a sound channel is produced. Under these circumstances, any sound signal traveling in this area is refracted back and forth so that it becomes horizontally channeled. Sound rays originating with an initial upward inclination are refracted upward. Rays from a sound source in this layer that make a small angle with the horizontal are roughly sinusoidal, crossing and recrossing the layer of minimum speed. This reinforcement of rays within the sound channel can continue until the sound is absorbed, scattered, or intercepted by some obstacle. Sounds traveling in this manner sometimes are received at extremely great distances from the source. These long ranges occur primarily as a result of two factors: absorption is small for low-frequency sound and most of the sound energy from a sound source at the axis is confined to the channel.

Under certain circumstances, a sound channel can exist near the surface of the sea. In a surface layer with a strong positive temperature gradient the upward bending of sound rays combined with reflections from the surface will form such a channel. Sonar ranges many times greater than normal have been observed where sound channels exist. However, the conditions that produce such sound channels near the surface are rare and not very stable. The region of minimum sound velocity, at a depth exceeding a kilometer, is the deep sound channel. It acts as an acoustic waveguide; sound propagating either upward or downward encounters sound-velocity gradients and bends back into this channel. Within the deep sound channel, sound undergoes only a cylindrical spreading loss, in which intensity drops off only as the first power of the distance. If sound spread uniformly in all directions, known as spherical spreading, this falloff would follow the square of the distance.

The concept of cylindrical and spherical spreading and their importance to sonar is presented in the next section. Other rays may follow propagation paths that are reflected at the sea surface and bottom (ray C), or refracted and reflected from either the sea bottom or the sea surface (rays D and E). Ray D is commonly called the refracted surface reflected (RSR) path, while ray E is called the refracted bottom reflected (RBR) path.

5.3.4. Bottom Bounce

In addition to being refracted by varying conditions in the medium, sound can be reflected in the manner of a light beam striking a mirrorlike surface and lose little of its intensity. The two surfaces that can produce this type of reflection are the surface of the water and the bottom of the ocean. Rarely if ever are these surfaces smooth enough to give a mirror reflection, but in many instances the majority of the sound is reflected as a beam. Some sonars make use of this phenomenon which is called bottom bounce. A beam is directed against the bottom from which it is reflected to the surface of the water. From the surface it is reflected back to the bottom again. Thus, the ray bounces from one to the other until its energy is dissipated or until it strikes a target and returns to the sonar. As with reflected light, the angle of reflection is equal to the angle of incidence. Obviously, ranging over a fairly flat bottom is more effective than ranging over a rough or sloping bottom. This path is highly dependent upon depth and absorption of sound by the ocean bottom. For simplicity in the figure, the bottom-bounce path is depicted as a single beam, whereas in the real world, a wider beam corresponding to a multidirectional sound source would be the normal case. The convergence zone path is shown as having depth based on the concept of a wide sound beam or "bundle" of sound rays emanating from a source. By definition, the convergence zone is formed when the upper ray of the bundle becomes horizontal.

This is the depth at which the sound velocity is equal to the high of the velocities at either the surface or at the bottom of the surface layer. The difference in depth between the ocean bottom and the depth at which the upper ray becomes horizontal is called the depth excess. The depth excess available defines the depth of the bundle of rays or "amount of sound" that forms the convergence zone path. As the sound energy travels upward and approaches the surface, the path narrows, tending to focus the sound energy resulting in convergence gains.

A maximum of 1200 ft depth excess is required to produce an operational useful convergence zone with reliable convergence gains. (As velocity increases with depth, the corresponding amount by which the velocity of the bottom ray exceeds that of the top ray in the convergence zone path bundle is called velocity excess.). Although many paths are available, only by observing the environment carefully and paying close attention to his equipment will the operator be able to use them to best advantage and not find them a liability.

5.3.5. Sound Propagation Through the Sea

The sea, together with its boundaries, forms a remarkably complex medium for the propagation of sound. It possesses an internal structure and a peculiar upper and lower surface that create many diverse effects upon the sound emitted from an underwater source. In traveling through the sea, an underwater sound signal becomes delayed, distorted, and weakened. The transmission loss term in the sonar equations expresses the magnitude of these effects.

5.3.5.1. Transmission Loss

Consider a source of sound located in the sea. The intensity of the sound can be measured at any point in the sea, near to or far from the source. For purposes of measuring intensity at the source, the intensity measurement is generally taken at one unit distance from the source and labeled Io. The intensity can then be measured at any distant point where a hydrophone is located and denoted I. It is operationally significant to compare the two values.

One way to do this is to form the ration Io/I. Note that if the ratio, denoted n, is greater than unity, the intensity at the source is greater than at the receiver, as would be expected. If n = Io/I, then

10 log n = 10 log Io - 10 log I = sound intensity level at the source minus sound intensity level at the receiver.

The value 10 log n is called the transmission loss and, of course, is measured in decibels. Most of the factors that influence transmission loss have been accounted for by scientific research, and can be grouped into two major categories: spreading and attenuation.

5.3.5.2. Spreading Loss

To understand spreading loss, it is convenient to imagine a theoretical ocean that has no boundaries and in which every point has the same physical properties as any other point - i.e., an infinite, homogeneous medium. In such a medium, sound energy would propagate from a point source in all directions along straight paths and would have a spherical wave front.

5.3.5.3. Attenuation Loss

Attenuation of sound energy in seawater arises principally through the action of two independent factors, absorption and scattering, with an additional contribution from bottom loss. Absorption. The primary causes of absorption have been attributed to several processes, including viscosity, thermal con-ductivity, and chemical reactions involving ions in the seawater. (1) The viscosity of the medium causes sound energy to be convert-ed into heat by internal friction. (2) Some sound energy is con-verted into heat because sound waves alternately raise and lower the temperatures. (3) Suspended particles are set to oscillating by the sound waves and in this process some of the sound energy is dissipated in the form of heat. This is especially the case if the particles are air bubbles. While each of these factors offers its own unique contribution to the total absorption loss, all of them are caused by the repeated pressure fluctuations in the medium as the sound waves are propagated. They involve a process of conver-sion of acoustic energy into heat and thereby represent a true loss of acoustic energy to the environment.

Experimentation has produced a factor , called the absorption coefficient, which when multiplied by the range gives the total loss in dB due to absorption. Water temperature and the amount of magnesium sulphate ($MgSO4$) are important factors influencing the magnitude of , because the colder the average water temperature and the greater the amount of $MgSO4$ present, the greater will be the losses due to absorption. However, it is the frequency of the sound wave that causes the most significant variation in the absorption coefficient.

While the formula will change slightly with temperature and geographical location, an equation for the value of in decibels per meter for seawater at 5oC is = 0.036f2 + 3.2 x 10-7f2 (8-21) f2 +3600

where

f = frequency in kHz

While this formula is rather cumbersome, the important thing to observe is that increases roughly as the square of the frequency. This relationship is of major importance to the naval tactician. It tells him that if higher frequencies are chosen for sonar operation in order to achieve greater target definition, the price he must pay is greater attenuation. The higher the frequency, the greater the attenuation and the less the range of detection. For this reason, where long-range operation of sonar equipment is desired, the lower the frequency used the better. Figure 8-23 depicts typical values of the absorption coefficient of seawater at 5oC for varying frequencies.

To obtain the transmission loss due to absorption, is merely multiplied by the range in meters. Thus,

TL = r (5-18) Scattering

Another form of attenuation is scattering, which results when sound strikes foreign bodies in the water, and the sound energy is re-flected. Some reflectors are boundaries (surface, bottom, and shores), bubbles, suspended solid and organic particles, marine life, and minor inhomogeneities in the thermal structure of the ocean. The amount of energy scattered is a function of the size, density, and concentration of foreign bodies present in the sound path, as well as the frequency of the sound wave. The larger the area of the reflector compared to the sound wavelength, the more effective it is as a scatterer. Part of the reflected sound is re-turned to the source as an echo, i.e, is backscattered, and the re-mainder is reflected off in another direction and is lost energy. Back-scattered energy is known as reverberation and is divided into three types: volume, surface and bottom.

Volume reverberation is caused by various reflectors, but fish and other marine organisms are the major contributors. Additional causes are suspended solids, bubbles, and water masses of markedly different temperatures. Volume reverberation is always present during active sonar operations, but is not normally a serious factor in masking target echoes.

The one exception involves the deep scattering layer (DSL), which is a relatively dense layer of marine life present in most areas of the ocean. During daylight hours, the layer is generally located at depths of about 600 meters and does not pose a serious problem. At night, however, the layer migrates toward the surface and becomes a major source of reverberation. It is rarely opaque to sound when detected with a sonar looking down on it from directly above, as with a fathometer, but this is not the case with a search sonar transmitting in a more or less horizontal direction. By pinging horizontally, the sound waves encounter many more organisms, and the effect can vary from partial transmission of sound to total reflection and scattering, thereby hiding a submarine.

Surface reverberation is generated when transmitted sound rays strike the surface of the ocean, i.e., the underside of the waves. It is always a factor in active sonar operations, and is directly related to wind speed because it controls wave size and the angle of incidence. Bottom reverberation occurs whenever a sound pulse strikes the ocean bottom. In deep water this condition normally does not cause serious problems, but in shallow water, bottom reverberation can dominate the background and completely mask a close target. The amount of energy lost through scattering will vary with the roughness of the bottom and the frequency of the incident sound.

Sound reflected from the ocean floor usually suffers a significant loss in intensity. Part of this loss is caused by the scattering effects just described, but most of it results from the fact that a portion of sound energy will enter the bottom and travel within it as a new wave. The net result is that the strength of the reflected wave is greatly reduced. The amount of energy lost into the bottom varies with the bottom composition, sound frequency, and the striking angle of the sound wave. The total of these losses can vary from as low as 2 dB/bounce to greater than 30 dB/bounce. In general, bottom loss will tend to increase with frequency and with the angle of incidence. Soft bottoms such as mud are usually associated with high bottom losses (10 to 30 dB/bounce); hard bottoms such as smooth rock or sand produce lower losses.

While it is possible to derive equations that will compute precise values of TL, associated with each of these additional scattering and bottom loss factors, the ocean characteristics are so variable that there is little utility in doing so.

It is customary, therefore, in operational situations, to make an educated guess as to their values and lump them together into one term "A," known as the transmission loss anomaly, which is included in the transmission loss equation.

5.3.6. Total Propagation Loss

It would be useful to have a simple mathematical relationship that would describe all the effects of the various factors influencing transmission loss as they occur in the ocean. But the state of the physical conditions encountered in the ocean are very complex and not at all easy to represent. A few mathematical models do exist that provide close approximations for some sets of conditions, but at present, no single model accounts for all the conditions encountered.

A simplified model used to obtain approximate values of transmission loss for the spherical spreading case is

$$TL = 20 \log r + r + A \qquad (5\text{-}19)$$

and for the cylindrical spreading case

$$TL = 10 \log r + r + A \qquad (5\text{-}20)$$

It is important to realize that sound transmission in the ocean is three-dimensional and that transmission loss versus horizontal range alone is not sufficient information for most operational situations. Areas of no sonar coverage occur at various intervals of range because of refraction, reflection, and interference between waves traveling via different paths. Therefore, while the TL equations are interesting and somewhat useful, they are not always totally accurate.

5.4. Sound Sources and Noise

Background noise, like reverberation, interferes with the reception of desired echoes. Unlike reverberation, however, it does not result from unwanted echoes of the transmitted pulse but from active noise-makers located in the ship or in the water. Noise produced by these sources is classified as self-noise and ambient noise. Self-noise is associated with the electronic and mechanical operation of the sonar and the ship. Ambient noise encompasses all of the noises in the sea.

5.5. Sample Problem Using Wenz Curves

A SOSUS station is searching for an enemy submarine known to be producing sound at a frequency of 300 Hz. It is suspected that the sub is patrolling in shipping lanes that currently have 6 ft. seas. Use Wenz curves to determine an approximate value of ambient noise.

Using the nomogram, combine the signals of 65 dB and 66 dB IL1 –

IL2 = 66-65 = 1

From the nomogram add 2.4 dB to 66 dB ... 68.4 dB.

5.6. The Sonar Equations

The key to success in antisubmarine warfare is initial detection. For this work the major sensor in use today is sonar, both active and passive, and the present state of knowledge of the physical world suggests no change in this situation for many years. An understanding of sonar can only be achieved through a comprehension of the sonar equations and the concept called figure of merit. Many of the phenomena and effects associated with underwater sound may conveniently and logically be related in a quantitative manner by the sonar equations. For many problems in ASW, the sonar equa-tions are the working relationships that tie together the effects of the medium, the target, and the equipment, so that the operator can effectively use and understand the information received and provide prediction tools for additional information. Therefore, the purpose of this section is to spell out the sonar equations and figure of merit, to state the specifics of their usefulness, and to indicate how the various parameters in the sonar equations, in-cluding the figure of merit, can be measured.

5.6.1. Signal to Noise Ratio.

The sonar equations are based on a relationship or ratio that must exist between the desired and undesired portion of the received energy when some function of the sonar set, such as detection or classification, is performed. These functions all involve the re-ception of acoustic energy occurring in a natural acoustic back-ground. Of the total acoustic energy at the receiver, a portion is from the target and is called signal. The remainder is from the environment and is called noise.

The oceans are filled with noise sources, such as breaking waves, marine organisms, surf, and dis-tant shipping, which combine to produce what is known as ambient noise. Self-noise, in contrast, is produced by machinery within the receiving platform and by motion of the receiving platform through the water. Further, in active systems, scatterers such as fish, bubbles, and the sea surface and bottom produce an unwanted return called reverberation, which contributes to the masking of the desired signal.

The function of the design engineer is to optimize the signal-to-noise (S/N) ratio for all conditions as detailed in the original design specifications of the sonar set. The operator, using his knowledge of the design specifications, his known ability in certain circumstances, the predicted conditions extrapolated from previously determined measurements, and actual on-board measurements, can then predict the detection probability. In order to predict performance, the operator's interaction with the sonar set must be defined or quantified in a manner that provides a measure of predictability for varying signal and noise levels. This quantity, known as Detection Threshold (DT), attempts to describe in a single number everything that happens once the signal and its accompanying noise are received at the sonar. Detection threshold is defined as the signal minus noise level required inboard of the hydrophone array in order that an operator can detect a target. Actually, the business of detecting a sonar signal is a chance process for several reasons, one of which is that a human being is involved. The decision to call a target may be either right or wrong if a target is really present, then there is a detection; if a target is not present, then there is a false alarm. Hence, the definition of DT is normally qualified by adding the requirement that an operator "can detect a target on 50 percent of those occasions for which a target presents itself." Thus, if the average value of provided signal-to-noise equals the average of required signal-to-noise, a detection occurs in 50 percent of the times that a detection could occur. To summarize:

If average provided = average required, then detection probability is 50%
If average provided > average required, then detection probability is 50% to 100%
If average provided < average required, then detection probability is 50% to 0%

Note that the instantaneous value of the provided or required signal-to-noise can vary over a wide range due to the variability of operators, and an individual operator's moods, as well as time fluctuations in propagation loss, target radiated signal, and own ship noise. Hence, while the average value of provided signal-to-noise may be less than the average value of required signal-to-noise, at times the instantaneous value may be greater than the required value, and a detection may occur. Thus, a probability of detection greater than zero exists. Putting this all together, it can be seen that if detection is to occur with a specified degree of probability, then the signal, expressed in decibels, minus the noise, expressed in decibels, must be equal to or greater than a number, the Detection Threshold, which also is expressed in decibels.

$$S - N \geq DT \qquad (5\text{-}21)$$

This equation is the foundation upon which all the versions of the sonar equations are based, and is simply a specialized statement of the law of conservation of energy.

The next step is to expand the basic sonar equation in terms of the sonar parameters determined by the equipment, the environment, and the target.

Two pairs of the parameters are given the same symbol (Own Sonar Source Level/Target Source Level, and Self-noise Level/Ambient-noise Level) because they are accounted for similarly in the sonar equations. This set of parameters is not unique, nor is the symbolism the same in all publications, but they are the ones conventionally used in technical literature. It should be noted as the discussion progresses how each of these parameters will fit into the mold of the basic sonar equation.

Parameter Active/Passive Symbol Determined by Own sonar source level Active SL Equipment Self noise level Active/Passive NL Equipment Receiving Directivity Active (Noise Limited Index Passive DI Equipment Detection Threshold Active/Passive DT Equipment Transmission Loss Active (2TL) TL Environment Passive TL Reverberation level Active RL Environment Ambient Noise Level Active NL Environment Passive Target Strength Active TS Target Target Source Level Passive SL Target. The transmission loss is frequency dependent .

When calculating the acoustic signal resulting from a signal of interest superimposed on the ambient background noise of the ocean (signal to noise ratio), this calculation must be made using signals of the same frequency to be valid. Since sonar systems are designed to operate in specific frequency bands, calculations for these systems must relate to the design frequency band of the specific sonar to be valid.

5.6.2. Passive Sonar Equation

A passive sonar depends on receiving a signal that is radiated by a target. The target signal can be caused by operating machinery, propeller noise, hull flow noise, etc., but the same fundamental signal-to-noise ratio requirement must be satisfied. At the receiver, the passive equation begins as $S - N \geq DT$. If the target radiates an acoustic signal of SL (Target Source Level), the sound intensity is diminished while en route to the receiver because of any one or more of the following: spreading, ray path bending, absorption, reflection, and scattering. The decrease in intensity level due to this is called Transmission Loss (TL) and is also measured in decibels. Hence the intensity level of the signal arriving at the ship is

$$S = SL - TL \tag{5-22}$$

Noise, N, acts to mask the signal and is not wanted. Therefore, the receiver is composed of many elements, sensitive primarily in the direction of the target so that it can discrimminate against noise coming from other directions. This discrimination against noise can be referred to as a spatial processing gain and is called the Receiving Directivity Index, DI. DI gives the reduction in noise level obtained by the directional properties of the transducer array. Therefore, in the basic equation, noise is now reduced and becomes

$$N = NL - DI \tag{5-23}$$

There are two things to note in this simple equation: the first is that DI is always a positive quantity, so that NL - DI is always less than or equal to NL; the second is that the parameter NL represents both Self-noise Level and Ambient-noise Level, for by its definition it is the noise at the hydrophone location and can come from any, or all, sources.

$$SL - TL - NL + DI \geq DT \qquad (5\text{-}24)$$

which is the simplest form of the passive sonar equation. In words, equation 5-24 says that the source level of the target minus the loss due to propagation through the medium, minus the sum of all interfering noises plus improvement by the spatial processing gain of the receiver, must be equal to or greater than the detection threshold for a target to be detected with the specified probability of detection. However, the greater-than or equal-to condition is normally written as an equality. It is then understood to mean that if the left-hand side's algebraic sum is greater than DT, detection is possible with a greater probability than that specified by DT. If the sum is less than DT, detection probability decreases. Generally speaking, these two conditions imply that either detection is highly probable or seldom occurs. As a further aid to understanding the passive sonar equation.

5.6.3. Active Sonar Equation

In an active sonar, acoustic energy is transmitted, and the received signal is the echo from the target. Two different, but related, equations are needed to describe the active sonar - one for an ambient-noise-limited situation and the other for the reverberation-limited situation. As developed previously, sonar performance is governed by the requirement that signal minus noise must be equal to or greater than detection threshold. The differ-ence in the two active sonar equations that satisfy this require-ment depends upon the characteristics of the noise that is actually present at the receiver when the signal is detected. The ambient noise may be described as either isotropic-i.e., as much noise power arrives from one direction as from any other - or as reverb-eration, in which noise returns primarily from the direction in which the sonar has transmitted. Before developing the active sonar equations, the two types of noise should be briefly explained. Ambient noise consists of those noises present even when no sound is being radiated by a sonar. These include such noises as sea animals, machinery, propulsion noises generated by the echo-ranging platform, and the turbulence generated in the vicinity of the sonar. This type is the same as the noise level term discussed in the passive sonar equation. The second type, reverberation, consists of a multiplicity of echoes returned from small scatterers located in the sound beam and near the target when they reflect the transmitted energy.

The combined effect of echoes from all of these scatterers produces a level of noise at the receiver that will tend to mask the returning echo from any wanted target. The development of the active sonar equation is similar to that for the passive equation. In other words, the formal sonar parameters will be fitted to the signal and noise. If a sonar transmits an acoustic pulse with an initial source level of SL dB, the transmitted pulse will suffer a transmission loss in traveling to the target. The target will scatter acoustic energy, some of which will return to the sonar. The back-scattered intensity is called target strength and is related to the scattering cross section of the target. The returning echo will again undergo a propagation loss, and thus the signal at the sonar will be

$$S = SL - 2TL + TS \qquad (5\text{-}25)$$

As long as the source of the radiated energy and the receiver for the echo are located together, the transmission loss experienced is equal to twice the one-way transmission loss. When the echo returns, under some conditions the reverberation background due to the initial transmission will have disappeared, and only ambient noise will be present. This noise will be identical to that described in the passive sonar equation, modified by the receiving directivity index. The fundamental relationship can then be expressed as

$$SL - 2TL + TS - NL + DI \geq DT \qquad (5\text{-}26)$$

which is the basic active sonar equation used when the sonar is operating in a noise-limited situation. Limited. If, on the other hand, the echo returns when the reverberation background has not decayed to a level below the ambient noise level, the background noise is given by RL. In this case, the parameter DI, defined in terms of an isotropic background, is inappropriate, inasmuch as reverberation is by no means isotropic. For a reverberation background the terms NL - DI are replaced by an equivalent reverberation level observed at the hydrophone terminals, and the sonar equation takes the form.

$$SL - 2TL + TS - RL \geq DT \qquad (5\text{-}27)$$

which is known as the reverberation-limited active sonar equation. Detailed quantification for the new term, RL, is difficult at best, for it is a time-varying function resulting from the inhomogeneties in the medium. One thing to note is that in the normal conversion from the basic equation to the active equations, the inequality again becomes an equality. As discussed under the passive sonar equation, it is understood that when the terms on the left-hand side exceed the detection threshold by a significant amount, detection is highly probable, and when it is significantly less than the detection threshold, detection seldom occurs. Of special interest in the active sonar equations is the term TS, and the fact that it usually is on the order of 15 to 25 dB. The variability of the value of TS is a function of the target aspect presented to the incoming signal. A beam target presents a greater reflective area then the bow, hence more energy is reflected from a beam target then from a bow target. Because TS is 10 times the logarithm of the reflected intensity divided by the inbound intensity, this statement apparently says that more energy is reflected than is incident, a condition clearly not possible.

The key lies in the definition of terms:

I reflected is the intensity, I, of the reflected signal measured one meter from the target, assuming the target is a point source.

I inbound is the intensity, I, of the signal inbound from transmitting ship to the target measured at a point on the target. Intensity is actually power per unit area striking the target at some point, and thus the total sound power striking the target is I inbound times an effective area. If one assumes that the major portion of this power is reflected not from the original effective area (which is almost the same as the profile area of the target), but instead from a point source, it necessarily follows that the reflected energy computed in this way must be greater because of the reduced area from which the energy emanates. Thus I reflected is greater than I inbound if both are defined as indicated above.

In this case, there is no such wave as the one that is defined as originating from the point source. This construct is merely a convenient way of duplicating the actual measured value of I re-flected when the wave is 1,000 meters or more away from the point source enroute back to the transmitting ship. Thus, if one were to measure I reflected and I inbound both at 1,000 meters from the target, then I inbound would definitely be greater than I reflect-ed.

Therefore, I reflected would have been computed to suffer a greater attenuation in traveling 1,000 meters from the constructed point source than I inbound will suffer in going 1,000 meters to the target. The explanation for this is the rapid attenuation due to spreading from the point source as compared to that undergone by the inbound wave, which is very near a plane wave when it is within 1,000 meters of the target.

CONCLUSION

Radar is all about using radio waves to detect the presence of objects and to find their positions. Modern radar classifies or identifies targets and even produce images of objects.A transmitter sends out a radio signals, which will scatter of any thing that it encounters an a small amount of energy scattered back to a radio receiver. The radar considered as very important invention in out life that is used in many applications such as military and weather forecasts.....etc

Furthermore, radar can detect many other properties of the target object by comparing the characteristics of the emitted and received signals. In fact, the beam initially emitted by radar is somewhat like the beam of a laser light: it has a single frequency and it is *coherent* (meaning that the phase of the beam is the same across the entire wavefront). One of the greatest advantages is that the beam's characteristics (such as its frequency, its duration and its strength) are perfectly known and can be compared to the characteristics of the reflected signal. Radar involves the transmission of pulses of electromagnetic waves by means of a directional antenna; some of the pulses are reflected by objects that intercept them. The reflections are picked up by a receiver, processed electronically, and converted into visible form by means of a cathode-ray tube. The range of the object is determined by measuring the time it takes for the radar signal to reach the object and return. The object's location with respect to the radar unit is determined from the direction in which the pulse was received.

Phased arrays get around the beam width limit by using several small apertures to achieve the same result as one large aperture. The most dramatic example of this may be the Very Long Baseline Interferometer, a series of 25m radio telescope spaced all over the earth operating at radio wavelengths. By themselves they sport mediocre beam widths insufficient for scientific studies. But when linked together as an *array* the size of earth, they have a beam width of milliarcseconds, far better than any current optical instrument.

Before we explore the workings of a phased array, it should be noted that the equations and science here makes no distinction between receiving and transmitting devices. For instance, radio telescopes can operate to detect signals from the sky or to broadcast terrestrial messages. While the mechanics of the actual transmitters and receivers may differ, the science here is the same.

In practice, arrays are often placed in more complicated arrangements. Arrays may be multi-dimensional. Some may feature irregular element separation. The mathematics for arbitrary arrangements becomes complicated quickly and won't be dealt with here. Suffice it to say that with careful numerical modeling, arrays can be designed to fulfill a wide variety of purposes. Upcoming sections will focus on some applications of acoustic arrays.

Phased-array output: For some types of experiments, most notably gated pulsar and VLBI observations, one wishes to phase up the array in real time and write out the summed data. For VLBI one wants the raw summed data, not averaged over time; for pulsars one can average down.

The cable and the equipment have to be checked for transmitter, receiver and A-scan conversion parameters. At the transmitter channel the constancy of the pulse exitation voltage and the pulse shape as well as the independency from the different delay times are essential parameters which should be documented. This can be measured qualitatively by a suitable control software or more quantitatively with a special hardware support. The receiver channel can be characterized by two transfer functions the amplitude transfer function to check the linearity of the amplifier behaviour , the frequency transfer function to check possible limitations of the frequency band width.

There are several different possibilities to check the equipment parameters, either at the laboratory or at on site conditions. Some checks can be based on a special control software without the use of additional hardware components. By this the following parameters can be checked: crosstalk at the transmitter- and the receiver channels (switching some channels on or off) functionality of probe-, cable- and equipment channels (e.g. using internal probe echos).

The fundamental principles underlying the concept of electronic beam steering are derived from electromagnetic radiation theory employing constructive and destructive interference.

These principles can be stated as follows:

The electromagnetic energy received at a point in space from two or more closely spaced radiating elements is a maximum when the energy from each radiating element arrives at the point in phase.

All elements are radiating in phase, and the resultant wave front is perpendicular to the axis of the element array. and subsequent diagrams show only a limited number of radiating elements. In actual radar antenna design several thousand elements could be used to obtain a high-gain antenna with a beam width of less than two degrees. The wave fronts remain perpendicular to the boresight axis and are considered to arrive at a point target in space at the same time. the path lengths from the elements to point *P* equalize as *P* approaches infinity. Thus, in situations where the target range is very large compared to the distance between elements, the paths from the elements to point *P* are almost parallel. Under these conditions, energy will arrive at point *P* with the same phase relationship that existed at the array.

Additionally, this imposes the requirement that the receiver bandwidth be extremely wide or that the receiver be capable of shifting the center of a narrower bandwidth with the transmitted frequency. However, in practice the use of coaxial delay lines or other means of timing at high power levels is impractical due to increased cost, complexity, and weight. To accomplish time delay scanning, variable delay networks are inserted in front of each radiating element. By proper choice of these time delays, the required effective phase shift can be applied to each element.

REFERENCES

[1] See G. J. Wheeler, *Radar Fundamentals* (1967); W. S. Burdic, *Radar Signal Analysis* (1968); H. Cole, *Understanding Radar* (1985); M. Skolnik, *Radar Handbook* (1989).

[2] D'Addario, L.R. 1989, *MMA Memo No. 55: Millimeter Array Correlator Cost Equation.*

[3] D'Addario, L.R. 1989, *MMA Memo No. 56: Millimeter Array Correlator: Further Design Details.*

[4] *Characteristics of a Millimetre Array, with and without a single Large Element and an LE pointing correction algorithm.*

[5] Holdaway, M.A. 1990, *MMA Memo No. 61: Imaging Characteristics of a Homogeneous Millimeter Array.*

[6] Thompson, A.R. *et al.* 1997, *MMA Memo No. 165: System Design Considerations for the Atacama Array.*

[7] www.yahoo.com/prenciples+radars

[8] www.ecyclopedia.com

[9] www.google.com/phased+array+radar