

V. 10

AD-A259 812



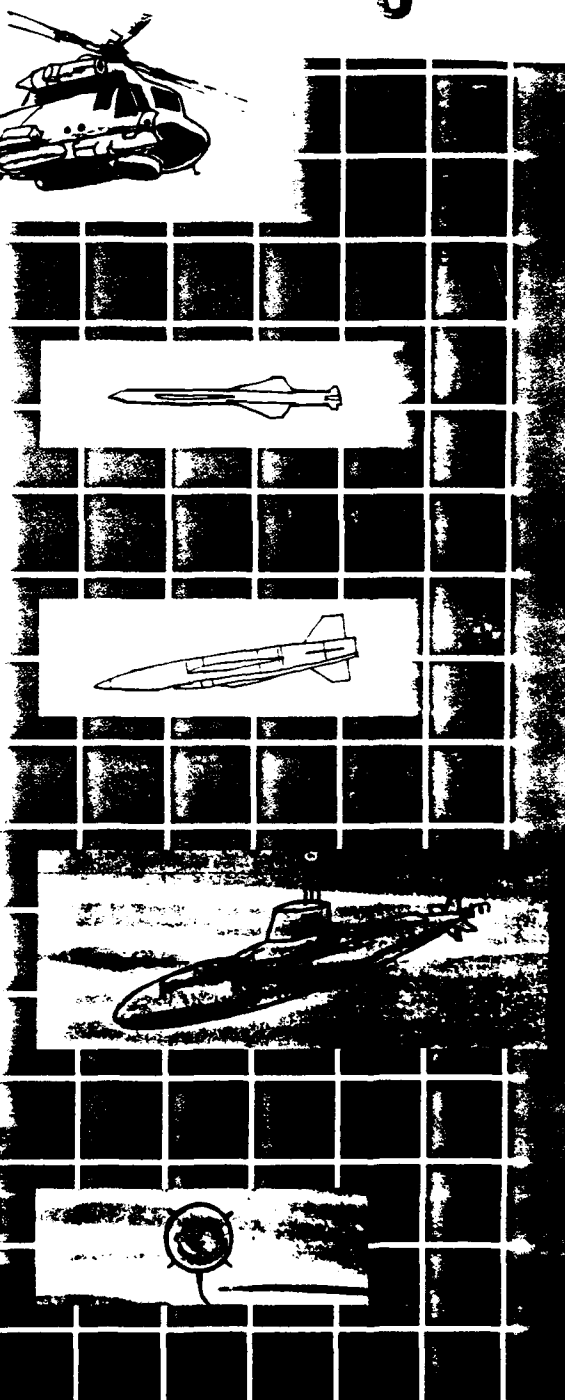
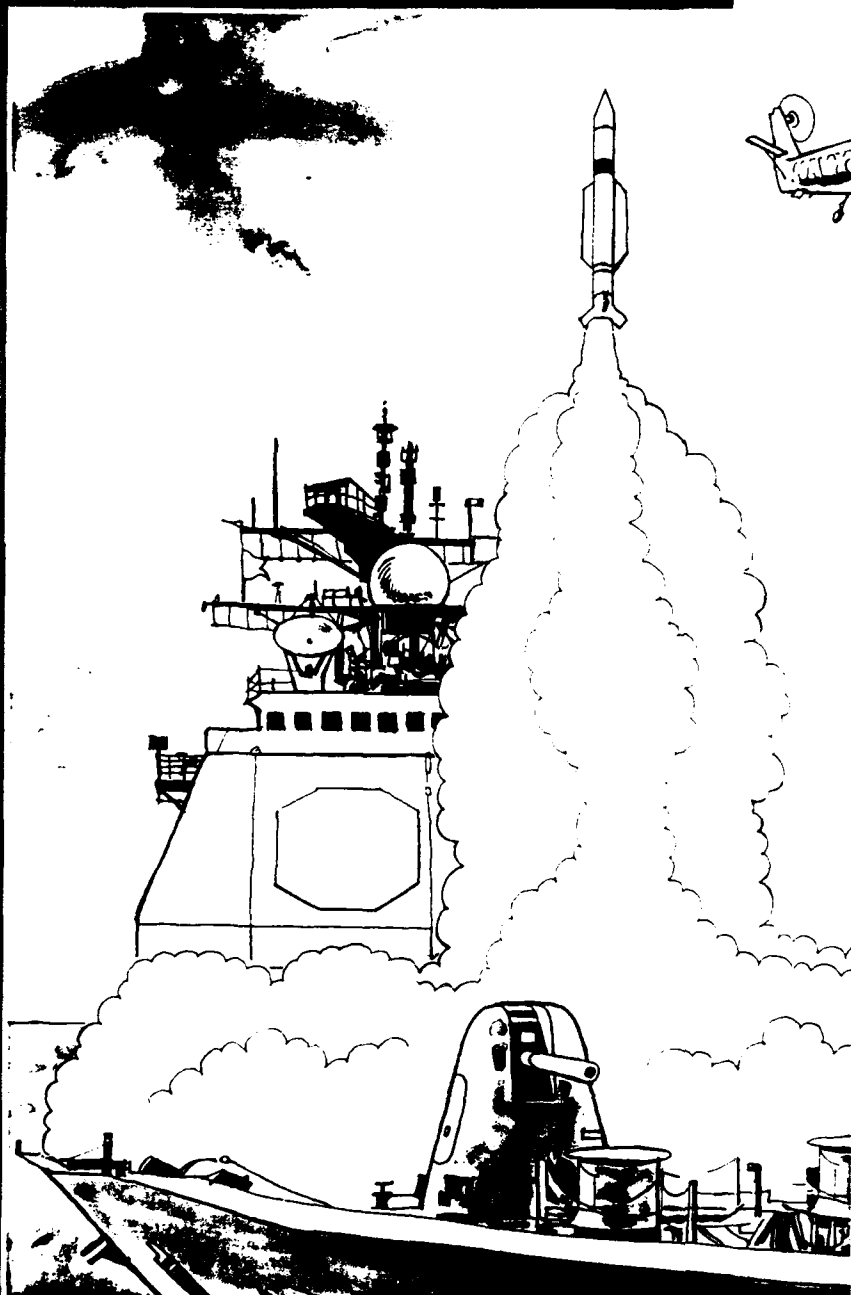
1

DTIC

ELECTE

JAN 4 1993

C



NOT STRATEGIC

NOT FOR OFFICIAL USE
EXCLUDED FROM AUTOMATIC
DECLASSIFICATION

Editorial Board

Dr. Jacques E. Goeller, *Chairman*
Dr. Horst G. Adolph
Dr. Robert S. Allgaier
Mr. Kenneth C. Baile
Mr. Sidney H. Hankerson, Jr.
Mr. Richard I. Rossbacher
Dr. Jon J. Yagla

Editorial Staff

Ms. Jean D. Sellers, *Managing Editor*
Mrs. Mary E. Montgomery, *Associate Editor*
Mr. Robert R. Coleman, *Graphic Designer*
Mrs. Pamela O. Lama, *Photographic Coordinator*

The *Naval Surface Warfare Center Dahlgren Division Technical Digest* presents unclassified articles, contributed primarily by Division scientists and engineers, on selected research and development programs. The Dahlgren Division, under the leadership of the Naval Surface Warfare Center, provides research, development, test and evaluation, engineering and fleet support for surface warfare systems, surface ship combat systems, ordnance, mines, amphibious warfare systems, mine countermeasures, special warfare systems, and strategic systems. Please address any correspondence concerning the *Technical Digest* to: Dahlgren Division, Naval Surface Warfare Center, Technical Digest (Code E262), 10901 New Hampshire Ave., Silver Spring, MD 20903-5000. Telephone (301) 394-4422 or (301) 394-4404.

About the cover: USS Ticonderoga (CG-47) class cruisers, such as the one shown here during a test launch of a Standard SM-2 anti-air missile, were the first to employ AEGIS Weapon System Mk 7. The AEGIS AN/SPY-1 radar is designed to simultaneously detect and track multiple threats similar to those represented on the right. Offering U.S. Navy ships unparalleled ability to operate in multithreat environments, AEGIS is extending the combat systems engineering concept to the entire battle force. The NSWC Dahlgren Division is Lead Laboratory and Lifetime Engineering Agent for the AEGIS combat system.

DATA QUALITY INSPECTED 1

Naval Surface Warfare Center
Dahlgren Division

Technical Digest

September 1992

Accession For	
NTIS GRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

Editorial	J. D. Sellers	3
Guest Editor's Introduction	J. R. Pollard	4

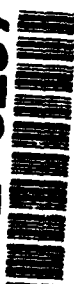
Systems Analysis

Overcoming the Reduced Radar Cross Section ASM Threat	J. F. Cavanagh	8
Refractivity Measurements Using a Vertical Array of Temperature and Humidity Sensors	J. K. Stapleton	16
Top-Level Models for Air Defense and Ship Survivability	R. L. Humphrey	26

MSI and Data Fusion

Multisensor Integration and Data Fusion in the Surface Navy	R. E. Helmick, W. D. Blair, C. F. Fennemore, and T. R. Rice	36
A Multisensor Concept for Detection of Low-Flying Targets	K. H. Krueger and R. A. Stapleton	50
ESTAR: Expert System for Track Ambiguity Resolution	M. J. Kuchinski, R. L. Taft, and H. C. Leung	56
Response of the Transfer Function of an Alpha-Beta Filter to Various Measurement Models	J. E. Gray and W. J. Murray	64

92-32974



Signal and Data Processing

A Digital Signal Processor for Improved EMS Bearing Accuracy	W. A. Masi	72
Automated Threat Library Support System	T. E. Hudson	84
The Doppler Spectrum for Accelerating Objects	J. E. Gray and G. A. Watson	88
Directional-Energy Detector Sonobuoy	T. B. Ballard	96
Wigner-Based Cycle-Rate Detection...Testing on MFSK Acoustic Data	J. Wilbur	104
Transient Resonance Scattering for Active Acoustic Classification	G. C. Gaunard	114

Component Technology

Hydrogen Spark Gap for High-Repetition-Rate Compact Accelerators	S. L. Moran	126
Next Generation High-Power Multiple Microwave Sources	H. C. Chen	140

Editorial

THE NSWC DAHLGREN DIVISION MISSION.

Provide research, development, test and evaluation, engineering and fleet support for surface warfare systems, surface ship combat systems, ordnance, mines, amphibious warfare systems, mine countermeasures, special warfare systems, and strategic systems.

Before the *Naval Surface Warfare Center Technical Digest* of September 1991 went to press, we knew there would never be another with precisely the same title. Beginning with this second issue, our *Digest* will be known as the *Naval Surface Warfare Center Dahlgren Division Technical Digest*—a small but significant change. While the first issue was in preparation, the Navy was in the process of folding a number of its research and development centers, laboratories, and various other activities into new and larger mission-oriented centers, popularly known as “megacenters.” Our organizational title, from which the *Digest* took its name, was soon to evolve to a higher level, emerging as the official name of the megacenter of which we are now part. Meanwhile, press date arrived before the effective date of the new organization.

Then, on 2 January 1992 the new Naval Surface Warfare Center (NSWC) with headquarters in Washington, D.C., was established under the aegis of the Naval Sea Systems Command. Simultaneously, we became the Dahlgren Division of that newly established Center. The Division includes headquarters in Dahlgren, Virginia, and the White Oak Detachment in Silver Spring, Maryland, with the addition of the Coastal Systems Station (formerly the Naval Coastal Systems Center), Panama City, Florida.

Thus it is that this issue of the *Naval Surface Warfare Center Dahlgren Division Technical Digest* bears a different—albeit not very different—name. We will continue publishing the same kinds of technical articles as in the past, with the addition of the mine countermeasure, amphibious, and special warfare work encompassed by our expanded mission. In the *Digest* we refer to our organization as the NSWC Dahlgren Division, the Dahlgren Division, or perhaps on second reference, simply “the Division.”

Jean D. Sellers
Managing Editor

Guest Editor's Introduction

The previous issue of the Technical Digest emphasized systems engineering principles as key to developing complex surface warfare systems. The basic functions of the modern combat system we defined as Detect, Control, and Engage. This issue, with the theme "Detection Systems and Technology," treats the first of those functions.

Detection Systems

Shipboard combat systems traditionally include an array of individual weapon systems, each consisting of a sensor for detection, a weapon for engagement, and a means of controlling both. In operation, a sensor gathers information about the ship's environment, including objects or targets of interest, and provides this information to the weapon system's operator for evaluation and, if necessary, weapon assignment. Multiple weapon systems within a given warfare area, e.g., Anti-Air Warfare (AAW), are made to work as a team by a warfare coordinator. The sensors provide information that forms a picture of the tactical situation, allowing the coordinator to assess that situation and make weapon employment decisions. Sensors can also provide information to each other to help a given sensor do its job better. Thus, the sensors found in modern combat systems enable these systems to function by providing vital information to command (for warfighting decision making), to weapons (for engagement), and to other sensors (for improved sensor system performance).

Evolving Requirements

The end to the cold war redirected the focus of the surface Navy toward regional conflict. Naval operations are now more likely to occur in shallow waters close to shore than in the open ocean. The major threats to surface ships have shifted from massive cruise missile attacks by bombers and nuclear submarines to fighter aircraft with antiship missiles, small surface ships or patrol craft operating close aboard, underwater mines, and diesel submarines armed with torpedoes. The shift in mission and threat presents major changes in detection system needs as well as new challenges in sensor technology.

One major change affecting detection is that it is becoming increasingly harder for detection systems to "see" modern targets. In some cases this is due to the small physical size of the target encountered in regional conflict, and in others to reduction of the target's radar cross section or other observables. Environments typically found in Third World regions also contribute to making targets more difficult to detect. The propagation characteristics of shallow water, for example, exacerbate an already difficult "quiet submarine" detection problem; land clutter complicates detection of air targets; and underwater mines lying on the bottom, or buried there, are extremely difficult to detect.

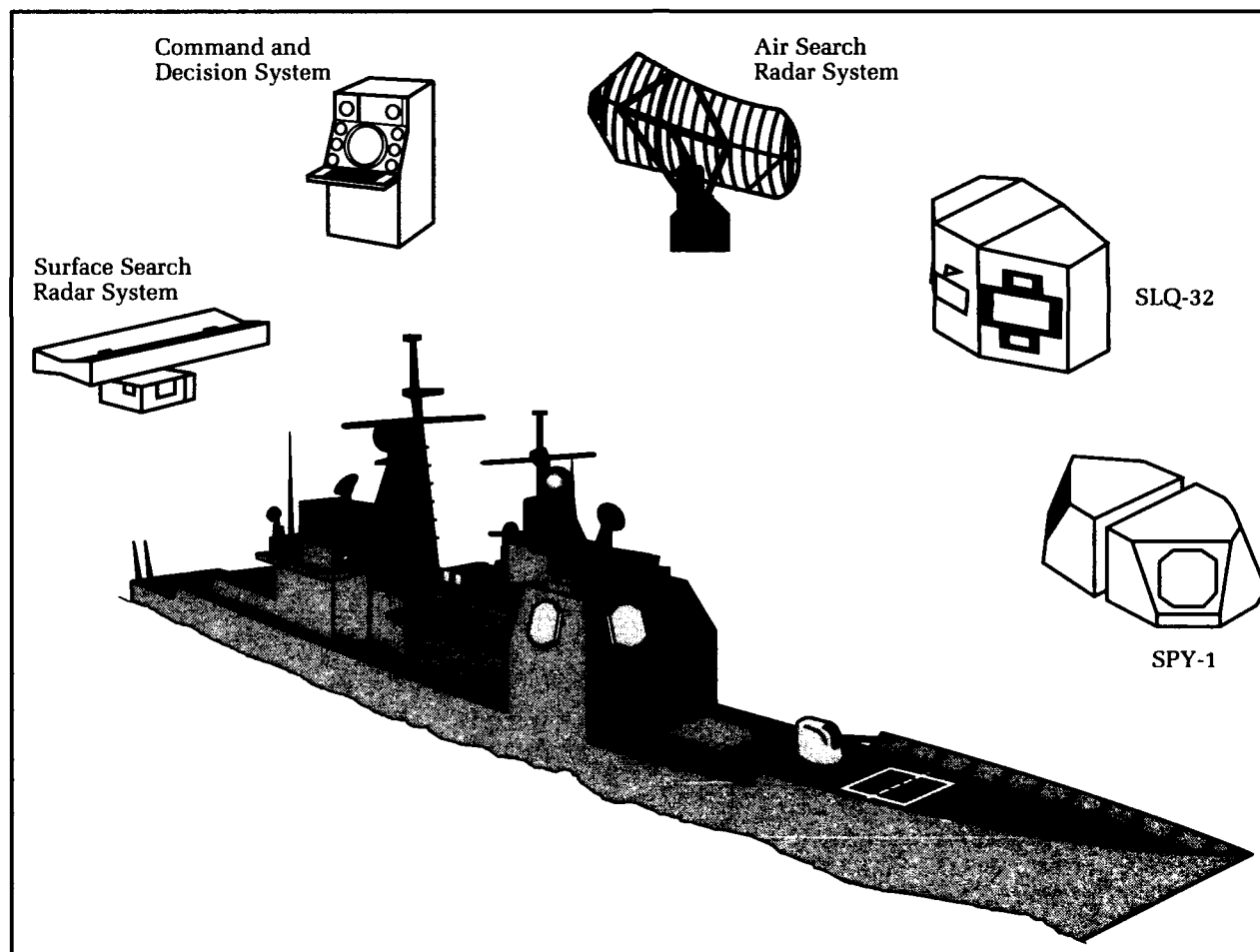
Hostile air or surface craft operating in areas of heavy commercial traffic present tracking problems, not only because of volume, but also in distinguishing one from the other. This problem is becoming increasingly significant since the rules of engagement for crisis response and regional conflicts often dictate positive target identification. In addition, under such circumstances threatening actions or other indications of hostile intent are usually required before engagement decisions can be made. Thus, sensors often must gather more than the traditional target track information to support engagement decisions. Providing future combat systems with information in a timely manner, and in the expected operating environments, poses great challenges to sensor and sensor information system designers and technologists.

Even in times of peace, surface ships operating either together or independently could be subject to surprise attack. From a detection point of

view, the antiship missile threat is particularly serious. Future antiship missiles are expected to employ techniques that will make them less observable to shipboard sensors, thus timely detection and tracking will become increasingly difficult. This and other advances in ship-defense penetration aids have prompted the Navy to position AAW self-defense as one of the top research and development priorities.

Radar has served as the primary means of detecting threat aircraft and missiles ever since it was installed aboard the destroyer *USS Leary* in 1937. Radar development progressed rapidly during World War II, evolving into the SPS-48, the mainstay of many of the Navy's non-AEGIS ships. Today, the multiwarfare capability of surface ships to deal with the AAW threat is built around the AEGIS weapon system Mark 7.

Featuring state-of-the-art technology, particularly in digital computers and radar signal processing, the AEGIS system was designed as a



AEGIS Weapon System Mk 7, deployed on Ticonderoga-class (CG-47) cruisers and Arleigh Burke-class (DDG-51) guided-missile destroyers, has a proven record as the "eyes and ears" of the fleet. The U.S. Navy relies on the system's superior detection, tracking, and engagement capability for countering the emerging AAW threat.

total weapon system from Detect through Engage. The heart of AEGIS is the advanced, automatic detect and track, multifunction, phased-array radar, the AN/SPY-1. AEGIS is designed to defeat a wide spectrum of targets, including supersonic antiship cruise missiles, manned aircraft, and surface ships. Integrated with the AEGIS system is the electronic warfare system AN/SLQ-32 with ESM (electronic support measures) sensor capability, which contributes to the detection and identification of hostile targets. The Dahlgren Division has played a continuing role in AEGIS system development as Lead Laboratory and Lifetime Engineering Agent. The detection system technologies and engineering thrusts discussed in this issue of the *Technical Digest* describe current efforts to ensure that the AEGIS combat system will keep pace with evolving threats well into the 21st century.

A Tour Through the Digest

In this issue we provide a glimpse of the NSWC Dahlgren Division's research and development in the area of shipboard detection systems. This work addresses the Navy needs and requirements described above. Our primary focus is on sensors and sensor information processing for AAW self-defense, an area that continues to elicit keen interest within the surface Navy. In addition, other categories such as underwater sensors and nonorganic surface sensors are also represented.

Systems Analysis

In the first article, J. Cavanagh addresses the overall problem facing sensor engineers in dealing with the low-flying, low-observable antiship cruise missile. Improvements in pulse doppler radars to cope with low-return signals from these low cross section missiles, and integration of infrared or other sensors with the radar, appear to be promising avenues of approach. Even so, propagation in a clutter environment continues to be a problem for the radar engineer. With the advent of more accurate propagation analysis software, accurate refractivity measurements are becoming increasingly important. J. Stapleton presents a new technique using a vertical array of temperature and humidity sensors for measurement of a radar's propagation environment. Variations in temperature and humidity near the water's surface can cause the propagation path of a radar to bend around the earth, introducing clutter and target returns from beyond the horizon. Environmental measurements are essential if we are to understand radar performance data and

design signal processors for reducing clutter and enhancing targets. In the final article in this section, R. Humphrey discusses AAW self-defense in terms of the analysis carried out in trading off and optimizing various self-defense system design alternatives.

Multisensor Integration and Data Fusion

This section expands on the concepts of multisensor integration (MSI) and data fusion as potential solutions to the AAW threat. This work addresses not only the ability of sensors and weapon systems to cope with the emerging low-observable threat, but also the creation of a consistent tactical picture from multiple-sensor data for AAW battle management. The article by Helmick et al. introduces the basic concepts of data fusion as well as definitions of terms, advantages, and issues. The authors discuss applications such as target tracking with multiple sensors, target identification, and kill assessment. C. Krueger and R. Stapleton present a more specific example of MSI, describing an integrated horizon search radar and infrared system concept to detect and track low-flying and low-observable antiship missiles. The next article, by M. Kuchinski et al., addresses the general problem of fusing track contacts to create a consistent targeting data and tactical picture from multiple sensor sources. The specific system here addresses over-the-horizon targeting for the Harpoon and Tomahawk antiship missile systems by providing a postprocessor to correlate ambiguous surface ship target track contacts from multiple nonorganic sensors. The last article in this section, by J. Gray and W. Murray, describes an improved computational technique for calculating the response of the transfer function of an alpha-beta filter to various measurement models. The technique accurately estimates future target positions and velocities from tracking radar data.

Signal Processing

Advances in signal processing are becoming increasingly critical to the detection of stealthy targets in a noisy environment. B. Masi presents a signal processing scheme to increase the signal detection sensitivity, direction-finding, and tracking accuracy of the AN/SLQ-32's ESM receiver, thus providing additional sensor information with which to solve the detection problem. Next, T. Hudson's article on an automated threat library describes an automated process that permits the rapid creation of tailored emitter libraries and subsequent deployment via satellite to ships afloat. This library enables the AN/SLQ-32 to properly identify antiship mis-

siles, attack aircraft, fire control radars, or other emitting systems. J. Gray and G. Watson discuss computation of the radar-doppler spectrum from a point-target model undergoing various accelerations. This analytical work is applicable to target identification, kill assessment, and maneuver detection, all of which are important to the ship's command in threat assessment during a tactical operation.

Next we shift from the AAW problem to the underwater detection and classification problem posed by diesel electric submarines and underwater mines. T. Ballard presents a method for processing the acoustic signals received by a sonobuoy in order to achieve automatic detection of submarines. Processing of acoustic data by the sonobuoy before transmission to an air platform allows processing equipment aboard the aircraft to handle effectively an increased number of sonobuoys, and thereby to increase search areas and missions. J. Wilbur discusses a novel signal processing procedure for cycle-rate detection of low signal-to-noise frequency and phase-modulated signals. The efficient detection and estimation of phase-modulated acoustic sonar transmissions has application to underwater telemetry, low-probability-of-intercept sonar, and other sonar transmissions obscured by noise. G. Gaunard presents a scheme for processing active sonar returns for target identification. This approach, based on the Resonance Scattering Theory, is particularly important for classification of submarines and distinguishing underwater mines from mine-like objects. The threat to surface ships and amphibious ships, as evidenced in Desert Storm, has catapulted mine countermeasures to a high Navy priority.

Component Technology

The final section presents some component technologies applicable to AAW sensors. The first article, by S. Moran, addresses research into high-power, high-repetition-rate switches for

application to charged particle beam weapons. Such switches allow stored electrical energy to be transferred to a device that in turn converts the energy into a charged particle beam. Sensors such as laser and microwave radars also utilize high-rep switches to transfer stored energy to conversion devices, e.g., a magnetron in a radar. H. Chen, in the final article, describes research in phase-controlling large numbers of microwave magnetrons connected in parallel. Such a novel approach has potential for detecting medium- and high-altitude stealth aircraft and missiles by greatly increasing the power transmitted over similar conventional radars. Good phase control offers further potential for improved performance by enabling conventional signal processing techniques to be employed.

The Guest Editor



JAMES R. POLLARD, Principal Systems Engineer in the Dahlgren Division's Combat Systems Department, received the B.S. degree in physics from Roanoke College; M.S. in electrical engineering, George Washington University; and Ph.D. in systems engineering, University of Virginia. After joining the Center (now the Dahlgren Division) in 1962, he performed research in antennas and electromagnetic coupling.

He then managed the Special Effects Weapon Research Program, which focused on high-power electromagnetic devices. After serving as head of the Weapon Systems Division and, subsequently, the Search and Track Division, he headed the Strategic Planning Group. In 1985 he received the Bernard Smith Award for creative development and implementation of a Center strategic planning process. From 1985 to 1987, he served as Warfare Systems Architect in the Space and Naval Warfare Systems Command. Currently, Dr. Pollard is conducting studies in future shipboard combat systems design.

Overcoming the Reduced Radar Cross Section ASM Threat

John Cavanagh

Future antiship missile (ASM) threats may be expected to employ stealth technology, making radar detection of low-observable targets difficult even for modern radar systems. Detection can be accomplished only at a much smaller detection range, resulting in decreased reaction time, increased target-discrimination assets, and heightened vulnerability to electronic countermeasures. When a low-observable target is also a low flyer, the detection problem becomes even more formidable. For example, characteristics of the radar, such as transmitter phase noise that could be ignored in the past, now become critical. This as well as other technical issues relative to detection of these threats are discussed. The Dahlgren Division has explored several technology trends and has concluded that the use of multiple sensors is a means of successfully defeating the projected cruise missile threat.

Introduction

Perhaps the most severe threat facing the surface Navy is the low-observable, low-flying antiship cruise missile. This article contains some background material on the nature of the difficulties encountered when trying to design a radar system to detect, track, and engage this type of threat. It will be shown that a radar alone may be insufficient for the task. For this reason, the Dahlgren Division has established a technical thrust in the area of multisensor detection and multisensor integration as a means to defend against this future threat.

Radar Cross Section

For the radar designer, the term low observable means small radar cross section. The concept of radar cross section is used to describe the fraction of the electromagnetic energy in a radar pulse that is scattered from a target back in the direction of the radar. This simple concept allows the radar engineer to replace a complex target by a fictitious isotropic radiator, the equivalent transmitter power of which is the product of the incident power density (watts/square meter) and the radar cross section (square meters). The radar cross section of an object depends on many variables, the most important being the relative orientation of the object, the radar frequency, and the polarization of the radar signal.

Figure 1 shows the radar cross section of a relatively simple shape, a cone terminated in a spherical cap. Minimum cross section is presented by this shape when the axis of the cone points in the direction of the radar. If the cone sphere is viewed from the rear, it presents a relatively large cross section. A

cross section that varies rapidly with orientation is observed for broadside presentations.

Table 1 gives the average cross section of several objects. Cross sections below about 0.01 square meter fall into the low-observable, or "stealthy," realm. The Air Force has used a butterfly analogy in connection with the B-2 bomber. This would suggest that the B-2 has a head-on radar cross section of less than 0.01 square meter.

The radar cross section of an object has only a tenuous relationship to its geometric cross section. Big objects have larger cross sections than similarly shaped smaller objects. The relationship need not be proportional. The radar cross section of a large, flat metal plate, for example, varies as the square of its geometrical area and inversely, with the square of the radar frequency. A flat plate with dimensions of 1m by 1m has a radar cross section of 14,000 square meters

Table 1. Typical Radar Cross Sections in Square Meters

Jumbo Jet	100
Fighter	1
Cruise Missile	0.1
Bird	0.01

at a frequency of 10 GHz if the plate is oriented orthogonal to the radar beam. A notable exception is a metal sphere. As long as the diameter of the sphere, d , is larger than about 5 radar wavelengths, its radar cross section is equal to its geometric cross section, $\pi d^2/4$. In addition, the radar cross section of a sphere is independent of frequency and orientation. For this reason, polished aluminum spheres are used as calibration targets for radars.

Reducing Radar Cross Section

The aircraft or missile designers have two basic tools at their disposal with which to reduce the radar cross section. These are shaping and the judicious use of radar-absorbing materials. Of these, optimum shaping has the higher payoff, inasmuch as there are no magic radar-absorbing materials. Materials with good radar-absorbing properties tend to have low physical strength and must be applied in thickness on the order of a wavelength, which is usually in the range of 3 to 30 centimeters. These materials work well on the walls of anechoic chambers, but easily suffer structural failure on the surface of high-speed airframes.

There are a few general rules for designing a reduced-radar-cross-section vehicle. Fortunately, the rules are not, for the most part, incompatible

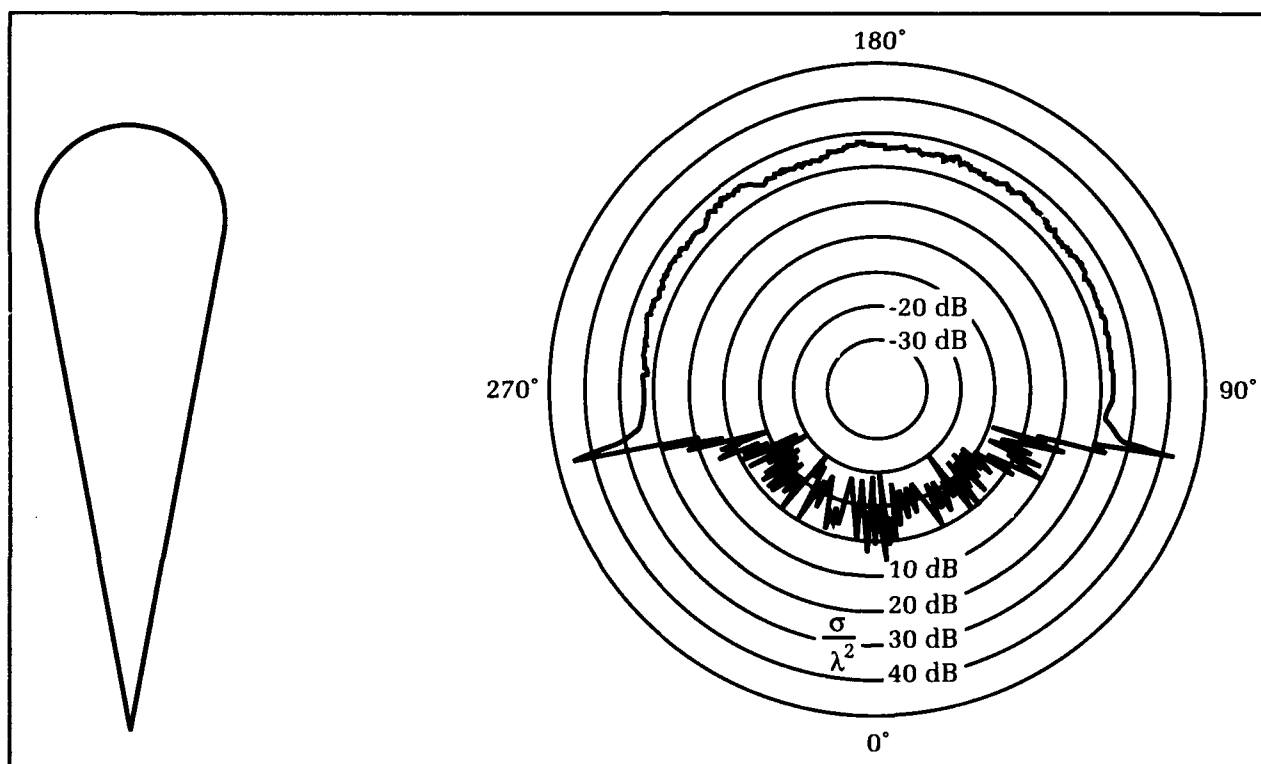


Figure 1. Radar cross section of a cone sphere.

with those for aerodynamic shaping. The following general principles apply:

1. Use shapes that reflect radar energy away from the radar.
2. Use convex, or at least flat, surfaces.
3. Avoid concave surfaces.
4. Avoid sharp discontinuities in the surfaces.
5. If any of the above rules cannot be followed, use radar-absorbing materials to reduce the back-scattered energy.

Energy is returned from an object by the general phenomenon of electromagnetic scattering. The principal forms of scattering are reflection and diffraction. The first three of the low-radar-cross-section design rules control reflection so that a significant portion of an incident radar signal is not retroreflected to the radar. Instead, the energy is reflected in some other direction. The flat angular plate structure of the F-117 stealth aircraft conforms to these rules.

Once the main reflections have been suppressed, the designer can concentrate on reduction of back scatter due to diffraction. This is accomplished by eliminating discontinuities in shape. For example, the transition from wing to fuselage should be a smooth curve instead of a sharp joint. The lower limit for the minimum cross section of the cone sphere in Figure 1 at nose-on aspect depends ultimately on the radius of curvature of the tip and the smoothness of the cone-to-sphere transition. In real objects, the ultimate cross section will be determined by skin panel joints, rivets, and holes in the skin.

Rule number 3 for designing a low-radar-cross-section vehicle is perhaps the most difficult to follow. Concave shapes make excellent wide-angle retroreflectors. Aircraft require cockpits, cruise missiles and aircraft require engine air intakes, and both need some form of sensor window. Some relief can be obtained with the use of conductive screens over the engine intakes. As long as the openings in the screen are less than about one-tenth of a radar wavelength, the screens behave much like a solid plate. The screen covering an engine intake can reflect energy away from a radar with proper orientation. There will be some residual leakage into the inlet duct, a notorious retroreflector. The use of radar-absorbing material to line the inlet will reduce the residual effects. Cockpit windows can be coated with a metal film much like the aviator's sun glasses or helmet visor. Absorbing material can be used to line the cockpit for further reduction of the residual cross section. These techniques have a price. Screens reduce airflow, and thus engine performance. Window coatings reduce visibility.

Sensors, especially radars, present a special problem. One technique applicable to cruise missiles is to stow the seeker antenna in a non-retroreflecting position until the final phase of the attack. Another approach is to use multiple-layer dielectric windows to limit signal penetration into the retroreflecting seeker housing to a narrow band of frequencies. The open literature has little to offer by way of techniques in the area of stealth technology.

Impact of Reduced Radar Cross Section

For the radar designer, low radar cross section, by itself, does not present an insurmountable challenge. In a free-space environment without jamming, it is difficult for the offense or defense to make a big impact on the detection range of a radar. The classical radar models predict a free-space detection range with an equation of the form:

$$R = [\text{Function of all variables}]^{1/4}. \quad (1)$$

It is the exponent of $1/4$ in this expression that controls the impact of radar cross section, or any other variable on free-space radar performance. In this case it takes a ten-thousandfold reduction in cross section to reduce the detection range by a factor of ten. A tenfold reduction in cross section reduces the free-space detection range to just 87 percent of the original value. A typical search radar can detect a 1-square-meter target in free space at about 200 nautical miles at some specified probability of detection and false alarm rate. Reducing the cross section to 0.0001 square meter shrinks the detection range to 20 miles.

In tactical situations, when electromagnetic countermeasures (ECM) are used, the exponent in Equation (1) changes. For a self-screening jammer, the exponent becomes $1/2$ if the jammer has enough power to exceed radar receiver noise. For a radar whose performance is limited by sea-clutter returns, the exponent approaches a value of 1. Under these circumstances, the impact of reduced radar cross section is greater than in the idealized free-space case.

Stealthy targets are not invisible. They are just more difficult to detect. The radar designer can counter reduced radar cross section with more sophisticated signal-processing techniques. This counter-countermeasure, like all others, has its price. It may increase the time required to search the surveillance volume of the radar, or it may make the radar more vulnerable to subtle forms of ECM. As the radar gains the ability to detect smaller and smaller targets, new clutter targets such as birds and insects have to be dealt with. Thus, target classification takes on a new level of complexity.

There are other advantages to low radar cross section. For the attacker, reduced radar cross section has a higher payoff in a proportional increase in jammer effectiveness than in reduced detection range. For every decibel in cross-section reduction, there is a corresponding reduction in the decibel of required jammer power for the same level of effectiveness. Alternatively, the attacker could maintain the same level of jamming effectiveness with reduced cross section and spread the jamming signal over a wider bandwidth. This is a counter to any frequency agility that might be employed by the defender.

Another payoff of reduced cross section is decreased reaction time on the part of the defender. At some range, any target, even when buried in a noise-jamming barrage, becomes detectable. If this range is too close, the defender does not have sufficient time to track the target, obtain a fire-control solution, and launch a weapon.

Radar cross section is really the only variable in the radar range equation that can be controlled by the attacker. Thus, it is prudent for the designers of naval anti-air defense systems to anticipate that future threats will have reduced radar cross sections.

The Low-Flyer Problem

The airborne attacker has another option for reducing reaction time and detection range for

the defender. This option is simply to fly an attack profile at an altitude close to the surface of the earth, as shown in Figure 2. Conventional radars have little or no performance at ranges beyond the radio horizon. For significant distances within the radio horizon, propagation losses arising from diffraction of the radar signal by the curvature of the earth limit low-elevation radar performance. The order of magnitude of these effects is equivalent to a ten-thousandfold reduction in radar cross section. The combination of a low-flying, low cross-section threat presents a formidable problem for the defender.

Figure 3 illustrates the propagation problem near the radar horizon at two common radar frequencies, S-band (3 GHz) and X-band (10 GHz).

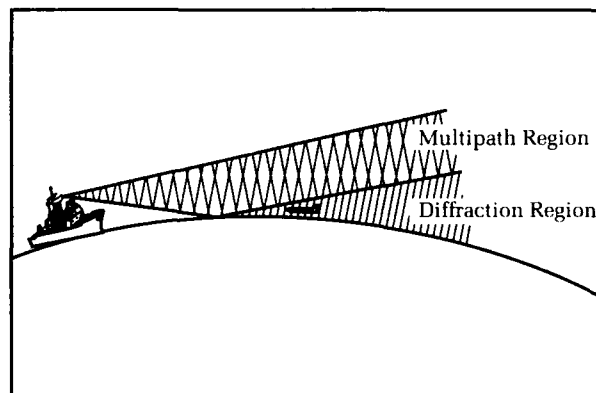


Figure 2. Low-flying ASM threat.

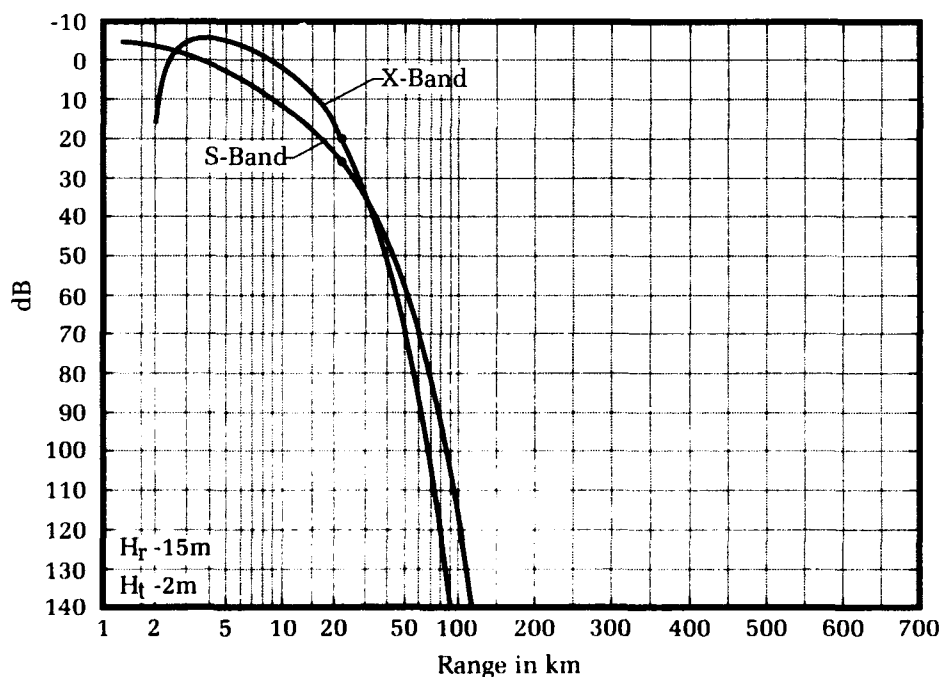
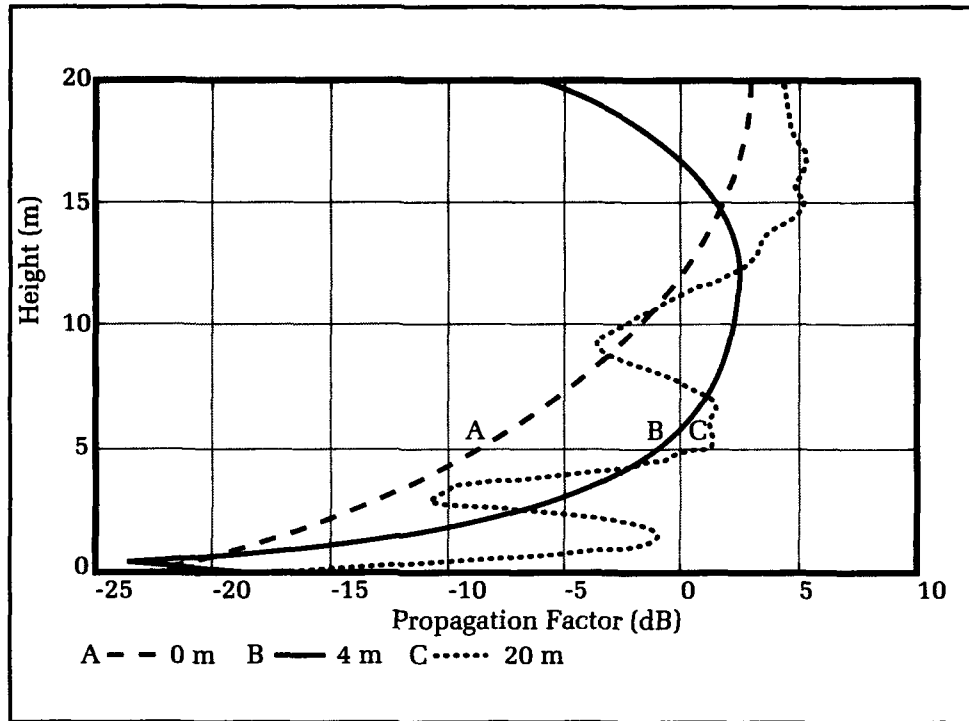


Figure 3. Propagation loss relative to free space.

Figure 4. Propagation factor in evaporation ducts.



For both frequencies, a radar height of 15 meters was used. A target height of 2 meters was assumed. The radar horizon is indicated by the solid dot on the curves at 22 km (about 12 nautical miles). Standard atmospheric refractive conditions were used in making the calculation of propagation loss with respect to free space. Note the rapid increase in relative propagation loss beyond the radio horizon, with S-band falling off more slowly. At the horizon itself, the higher frequency X-band signal suffers less propagation loss. However, this is still a significant loss. The X-band value of 20 dB at the horizon must be doubled for two-way radar propagation. This 40-dB propagation loss is equivalent to the ten-thousandfold cross-section reduction mentioned previously. Even at X-band, near-free-space condition would not be obtained until nearly 10 km. If detection cannot be achieved until this range, the defender has less than 30 seconds to track and engage a Mach 1 target.

12 In a real maritime environment, standard propagation conditions are the exception at frequencies above about 2 GHz. Over the last decade, it has been recognized that the maritime evaporation duct determines radar propagation factors at low elevation angles most of the time. In the last few years powerful computer codes have become available to calculate propagation factors in the evaporation duct. The results for such calculations are shown in Figure 4. The calculation was done for standard atmospheric refractive conditions as a reference. The other two curves were calculated for evaporation duct

heights of 4 meters and 20 meters. On a world-wide average distribution of duct heights, it is found that a height of 4 meters is exceeded 90 percent of the time and 20 meters is exceeded only 10 percent of the time. Note that propagation losses within an evaporation duct are about the same as free space or less.

While nature appears to be unusually cooperative with the designer of shipboard self-defense systems by providing an evaporation duct to counter diffraction, the evaporation duct itself presents new problems. Classical sea-clutter theory with a standard atmosphere predicts that sea clutter should disappear as the grazing angle approaches zero at the radio horizon. In the evaporation duct, however, there is a minimum non-zero grazing angle. This, in turn, gives rise to a non-zero minimum clutter level. Thus, while a reduced cross-section target may be visible in an evaporation duct, the signal return is immersed in a large sea-clutter background.

Extracting targets from clutter is amenable to two standard techniques: moving target indicators (MTI) and doppler filtering. These are actually the same technique. MTI processing is carried out in the signal time domain; doppler filtering is carried out in the signal frequency domain. Both techniques have fundamental limits on the ability to extract fast-moving targets from stationary or slow-moving targets. When these targets have low radar cross sections, the limit is determined by the phase noise spectrum of the transmitter. This particular characteristic of radar transmitters could generally be ignored in the past. Today, transmitter phase noise is of prime importance.

Radar Design Problems

No practical transmitter emits a pure signal. In fact, it can be argued that an oscillator works by means of frequency-selective amplification of phase noise arising from natural random processes within electronic devices. The spectrum of a frequency source close to the desired frequency (carrier) is shown in Figure 5. Unwanted energy appears at frequencies removed from the main carrier frequencies in the form of phase-noise sidebands.

Conceptually, it is easier to view radar performance limitations imposed by transmitter phase noise by considering a doppler radar. Echoes from a stationary radar target are returned at the same frequency as the incident signal. For a moving target, the echo is shifted in frequency relative to the incident signal by an amount proportional to the product of transmitted frequency and the target's relative radial velocity. For a Mach 1 target inbound toward an X-band radar, the doppler shift is on the order of 20 kilohertz. The ability of a doppler radar to extract moving targets from those that are stationary or slow-moving depends on its ability to separate echoes by their doppler shift and filter out the clutter signals. Filters are not perfect, and some residual clutter signals will pass through the doppler filters. Under ideal conditions, the residual clutter signals will be less than the natural noise floor of the radar receiver.

For low-radar-cross-section targets, however, the real problem arises from another source,

which generates a return signal from clutter targets with the same doppler offset frequency as the desired target. The transmitted signal contains weak phase-noise sideband at the target's doppler offset frequency. The weak signal is scattered back to the radar by a large target. This large target is the area on the surface of the sea illuminated by the radar beam. The net result is that the weak echo from the low-radar-cross-section, low-flying target must compete with the phase-noise clutter return. If the transmitter is noisy, the clutter return in the doppler filter can overwhelm or mask the target return.

Older radar tube technology represented by the magnetron tube cannot support any form of MTI processing. These tubes are too noisy. Newer, power-tube technology represented by the traveling wave tube and the cross-field amplifier will support limited MTI processing. The limit in the amount of clutter suppression that can ultimately be achieved depends on the phase-noise spectrum of the signal produced by these tubes. For the first time, the radar designer is concerned with the noise figure of the transmitter chain in addition to the traditional concern for the noise figure of the receiver chain.

Present day microwave-vacuum-power-tube technology will not support the detection and tracking of the low-observable, low-flying threat. The future does not offer much hope in this area, as there is a hiatus in power-tube development throughout the world. Radars with solid-state transmitters offer the promise of providing the needed low phase noise. However, solid-state

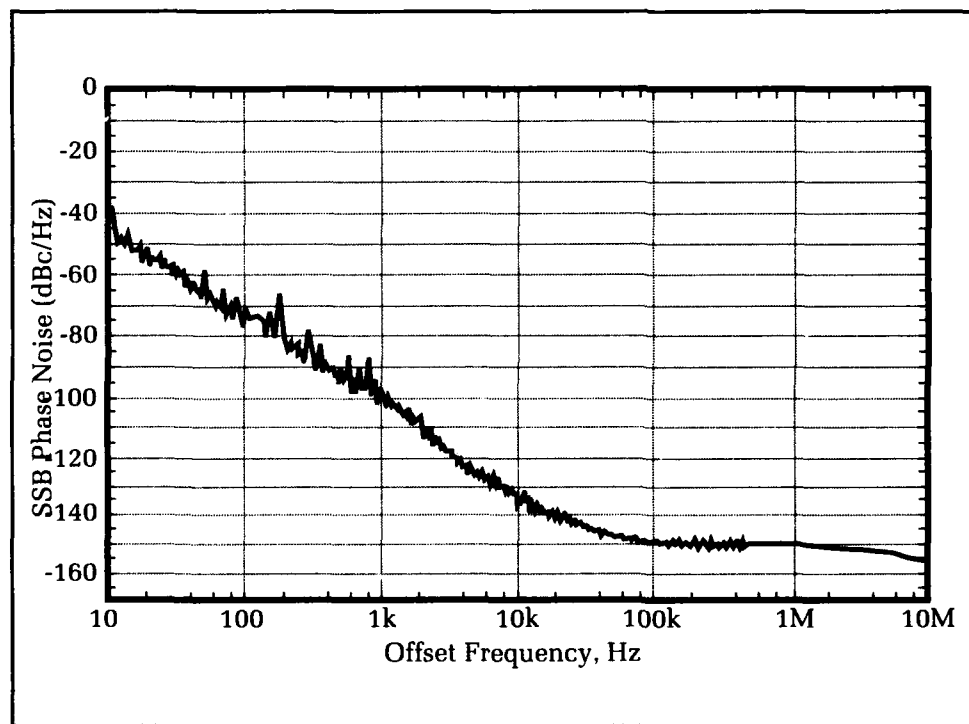


Figure 5. Typical SSB phase noise at 1 GHz.

transmitters with sufficient power to generate a low-observable target return above the receiver noise level are still in development. At the higher radar frequencies preferred for low-flying targets, solid-state amplifier transistors producing about 1 watt of power are available. Radar designers would like at least 10 watts. If this power level were available, the amplifiers would be used in an active-array antenna. With this configuration, each element of a phased array would be driven by its own amplifier. Alternative schemes in which power output of up to several thousand amplifiers would be combined to yield output powers comparable to radar-power tubes are impracticable. The best power-combining networks have unacceptable power losses. Low-power active arrays using 1-watt amplifiers are under development for the Advanced Technology Fighter Program.

Technical Trends

Solid-state radar modules are technically feasible for use in an active-array radar. This development needs to be accompanied by a parallel effort to make them affordable. The cost of solid-state transmit/receive modules needs to be brought below the \$1000 level. At the present level of development, unit costs are greater than \$10,000. The Department of Defense is supporting manufacturing technology (MANTEC) programs designed to bring the cost down to the desired level.

MANTEC uses the combined strategies of enhancing GaAs yields, employing more complex microwave integrated circuitry in the chips that go into a module, and developing automated assembly techniques. These efforts are just getting under way. However, MANTEC is an

evolutionary process with moderate risks and high payoff.

Other important technical developments supporting improved ASM defense are to be found in the area of infrared-detector arrays. Analytical estimates of the performance of infrared sensors using detector arrays have shown that they should complement radars in trying to detect the low-flying ASM. As these detector arrays become available, infrared search systems should see a revival.

In spite of all these efforts to improve radar performance, there may be circumstances in which the time and energy resources of a radar are overwhelmed. A radar alone may not be capable of doing the job of defending the ship under massive attack situations. For this reason, the concepts of multisensor detection (MSD) and multisensor integration (MSI) are being developed. Radars traditionally have a difficult time tracking low-elevation targets, but radars are the only sensors providing precision range data. Under an MSI concept, angle-track information for a low-elevation target might be derived from an infrared system while the radar continues to provide the range information. The track would be dynamically maintained by using the highest quality data available at a given instant. Under an MSD concept, a precision angle-of-arrival, electronic support measures (ESM) system might cue a radar to schedule a high-energy search dwell along the contact azimuth. This would save radar resources for the other demanding tasks going on during an attack. Conversely, the radar might request an ESM set to listen along a particular bearing to help the combat system evaluate a radar contact.

Any improved ASM defense will require even more computing capability. Future weapon

systems will provide insertion opportunities for application-specific integrated circuits, reduced-instruction set integrated circuits, very high-scale integrated circuits, and other digital-processing technologies. The speed of the latest technologies has reached a level where it is now possible to contemplate implementation of signal-processing methods that have heretofore existed only in textbooks.

Conclusion

In summary, the low-observable, low-elevation threat presents a very difficult detection and tracking problem for a radar. If target radar cross sections are projected downward, new radar technology will be required. Even with advanced technology, reliable radar detection will depend on favorable propagation conditions. When propagation conditions are not favorable, the radar is going to need some help. To this end, the Dahlgren Division is actively pursuing basic technology programs to develop that assistance. Examples include the development of auxiliary sensors such as horizon-search infrared systems and precision electronic support measures. A parallel effort is under way to develop optimum methods of interactively combining sensor data such that one sensor cues another in a synergistic total sensor suite for shipboard self-defense. These efforts are being carried out under the umbrella of the Multisensor Detection Project of the Surface Launched Weaponry Block program.

The contest between offensive and defensive systems is an age-old struggle. Only by anticipating developments in both arenas can we avoid technological surprise.

The Author



JOHN F. CAVANAGH was born 26 July 1940 in Providence, Rhode Island. He received a B.S. degree at Providence College in 1962, and M.S. and Ph.D. degrees from Purdue University in 1963 and 1969, respectively. At NSWC (now Dahlgren Division) since January 1969, he has been associated with the HERO program, EMC programs, and the EMPASS project. Since 1974, he has been a senior staff scientist in the

Search and Track Division. He is a member of the IEEE Wave Propagation Standards Committee, Chairman of United States Study Group 5 (Propagation in Non-ionized Media) of the International Radio Consultative Committee (CCIR), and International Chairman of Working Party 5B of the CCIR.

Refractivity Measurements Using a Vertical Array of Temperature and Humidity Sensors

Janet Stapleton

With the advent of more accurate propagation analysis software, refractivity measurements are seeing increased use in support of naval radar systems testing, both to enhance understanding of the widely varying signal levels observed during these tests and to guide the selection of times to run propagation-critical tests. Unfortunately, it is quite difficult to characterize completely the propagation environment over all time and space covered by such radar tests. Several different types of refractivity profiling systems have been developed, each having its own merits and shortfalls. All these refractivity measurement schemes suffer to differing degrees by being impractical for the naval operational environment. Still, during field testing, when less practical measurement systems can be used for a short duration, the profile data can be quite helpful for later analysis of recorded radar data. This article describes some of the more common refractivity measurement systems in current use and presents the results of the first test of a refractivity profiling system conducted at the Wallops Island (Virginia) Detachment during April and May 1991.

Introduction

An assessment of the RF propagation environment during radar testing has become vital to data reduction efforts. One of the most common methods of assessing RF propagation is through the use of refractivity profiles measured versus height. Several different refractivity measurement schemes have been devised and used with varying levels of success. Most involve moving a single-sensor package (consisting of a temperature sensor, a relative-humidity sensor, and a pressure sensor) vertically through the first several hundred feet of the troposphere, with a complete profile being made every 10 to 15 minutes. During recent tests at our Wallops Island (Virginia) Detachment, a vertical array of seven fixed-position temperature and humidity sensors was used to make refractivity profiles. An entire profile was measured every five seconds, allowing the temporal fluctuation of these profiles to be observed in a way not possible with most commonly used refractive profile measurement systems. The refractive profiles measured with the vertical array showed variations of 10 M (modified refractivity) units in a time period of approximately 10 minutes. Since propagation models can be sensitive to changes as small as 0.5 M unit, it is evident that the fluctuations observed could have a large impact on predicted propagation values. It is also expected that these variations occur along the

entire path covered by a radar beam, which suggests that adequately measuring the refractive environment to support the details of radar data analysis in other than the most uniform conditions is extremely difficult, if not impossible. This article discusses the results of these tests and suggests where further analysis might prove beneficial.

Background

RF refraction can be classified into several categories characterized by the refractivity gradient or the change in refractive index with height. The characteristic decrease in refractive index with height, and therefore increase in wave-front velocity with height, causes ray bending. Categories of ray bending are shown in Figure 1. The refractivity gradient considered "standard" close to the earth's surface is -39 N/km or 118 M/km , where N is refractivity and M is modified refractivity.¹ The standard refractivity gradient causes initially horizontal rays to bend downward at a rate less than the curvature of the earth. This effect extends the radar horizon beyond the calculated geometric horizon. Superrefraction and subrefraction, two other categories of RF refraction, are characterized by M gradients either less than or greater than the standard value, respectively. In cases of superrefraction, the rays are bent more than under standard conditions, which leads to an extended radar horizon. In cases of subrefraction, the rays are bent less than the standard case, resulting in a reduction in the radar horizon. A special case of superrefraction, known as trapping or ducting, occurs when the radius of curvature of the propagating rays, over some range of launch angles, is less than the radius of the earth, resulting in the rays becoming trapped in a ducting layer.

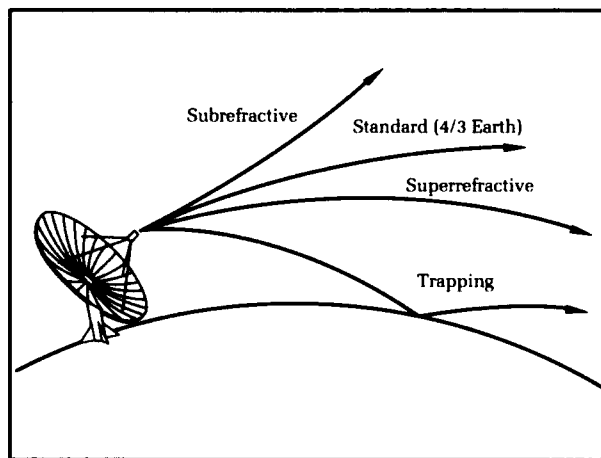


Figure 1. Ray bending for several categories of refraction.

For standard atmosphere, an effective earth-radius model is often used to calculate the propagation loss for heights and ranges of interest. In the effective earth-radius model, the earth radius is scaled an appropriate amount to account for the average degree of ray bending. (The scale factor for standard conditions is $4/3$.) After the earth radius is scaled, calculations of the radar horizon and the elevation angle to specific points in space are made using a homogeneous atmosphere and assuming the rays travel in straight lines. The $4/3$ scale factor for standard atmosphere, a worldwide median value,² would then result in an extension of the radar horizon and a reduction in the predicted elevation angle to a point in space compared to the results for the same calculations using the actual earth radius. One type of nonstandard propagation, a layer known as the evaporation duct, is always present over the ocean to some degree. The evaporation duct is characterized by a value known as duct height, which is an indication of the depth as well as the strength of the layer. The Naval Command, Control and Ocean Surveillance Center (NCCOSC) in San Diego, California, has compiled statistics on the frequency of occurrence of evaporation duct heights around the world.³ The worldwide average evaporation duct height is 13 meters, and according to the NCCOSC statistics, 90 percent of the time an evaporation duct of 4 meters or greater is present worldwide. A 23-meter or greater evaporation duct height is present 10 percent of the time. Other types of trapping layers, such as advection ducts, can dominate the evaporation duct effects.⁴ These types of ducts occur with greater frequency over water near land masses and can have dramatic effects on radar propagation. Figure 2 shows the differences in the predicted propagation for a radar at X-band for standard atmosphere, for the world

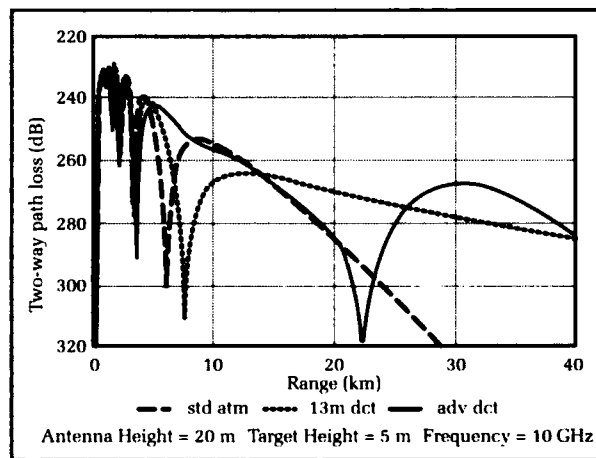


Figure 2. Two-way path loss versus range for three different refractive conditions.

average evaporative duct height of 13 meters, and for an advection duct with a -10 M change from 0 to 25 meters. The large differences in these three propagation curves show the need for assessing the refractivity environment during radar testing.

Common Refractivity Measurement Systems

Most commonly used refractivity profiling systems involve moving a single temperature, relative humidity, and pressure sensor package through the lower troposphere. The way in which the sensor package is moved and the method and rate of data transfer are the most significant differences among these systems.

Balloon refractivity profiling systems available from commercial manufacturers consist of a sensor package, a weather balloon or small aerostat, and in some cases a telemetry link. The sensor packages usually consist of a temperature sensor, a relative-humidity sensor, and a pressure sensor used to infer height above the surface. In some cases a telemetry link is used to transfer data as it is being taken to a storage unit; in other cases the sensor package is hardwired through a cable to a storage unit. The balloon is raised and lowered by an electric winch, with typical measurements of a complete profile taking 10 minutes. One of the advantages of this system is that the balloons can be used to heights near 300 meters if necessary to define the refractivity profile completely. The balloon systems are also well suited for use over water because no rigid support towers are needed; only a platform such as a boat is required for the winch and for the data logging equipment if the sensors are hardwired. One common problem with the balloon or aerostat, however, is that in winds above 10 to 12 knots, the balloon tends to become unstable. In the coastal environment, winds of this speed are common, making the use of balloons for refractivity measurements undependable. Finally, the fact that the exact height of the sensor package above the surface is derived from a pressure measurement opens up the possibility for height errors in the refractivity profiles.

An approach used in the past at the Dahlgren Division involved moving a sensor package up and down a cable attached to a support tower. The sensor package was comparable to the package described for the balloon profiling system. Early use of this system also relied on pressure to derive the height above the surface, but that system was later modified so that height was measured precisely using a height transducer. These tower profiles are limited by the support structure in two ways: the tower will most likely be located on land so that profiles over water are impossible; and the tower is usually limited to

approximately 100 feet in height, which is frequently inadequate for completely measuring the refractive profile. The time required to complete one sounding by moving the sensor package up or down the tower was on the order of three minutes, making the entire process take approximately six minutes.

Another method of measuring the refractive profile⁵ uses a model rocket to deploy a sensor package payload that takes measurements as it parachutes to the surface. Again, temperature, relative humidity, and pressure are the measured parameters. This system relies on the pressure measurement for height above the surface and on a telemetry link to transfer data. The maximum altitude for the rockets is 150 to 800 meters, with typical height resolution of the recorded data being three meters. This system is expendable, and the cost for each profile is approximately \$150.

One other approach in current use employs a sensor package attached to a helicopter flying a vertical zig-zag pattern varying in range and altitude simultaneously.⁵ The package includes a temperature, relative humidity, and pressure sensor along with a radio altimeter.

Vertical Array of Fixed Sensors

The first tests of a refractivity profiling system comprising a vertical array of fixed position

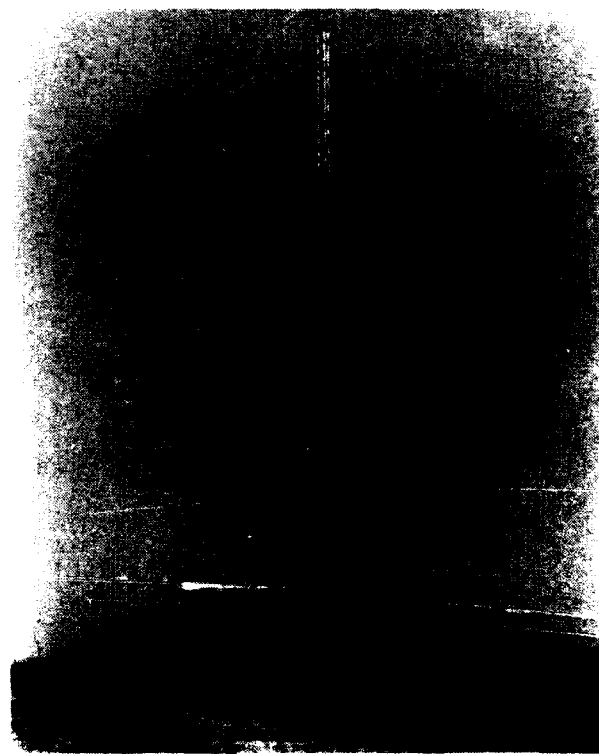


Figure 3. Full view of tower and sensors used in refractivity measurements.

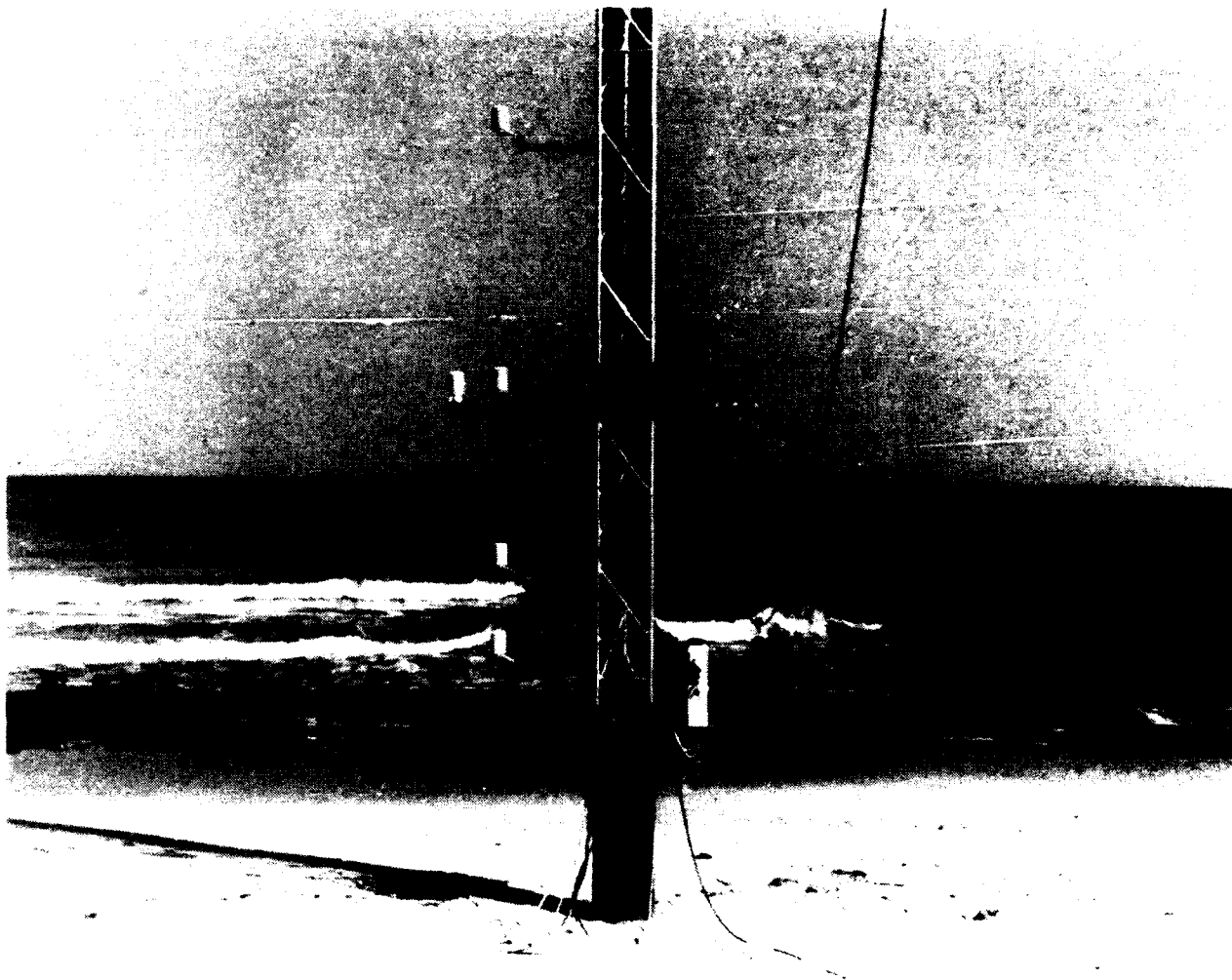


Figure 4. Front view of tower and sensors used in refractivity measurements.

temperature and relative humidity sensors were performed recently at Wallops Island. This system, shown in Figures 3 and 4, was designed to allow rapid profile measurement so that the temporal fluctuation of the refractive profiles could be studied. The other previously discussed systems were not well suited for this task. Since the sensors were fixed vertically, extra time was not required to move the sensors up or down to produce a profile, and their heights above the surface were known precisely. The vertical array was positioned near the high-water mark in order to minimize land effects.

Equipment

The main support for the equipment, a 90-foot wood pole, was sunk approximately 15 feet into the sand at the high-water mark on Wallops Island. A 75-foot aluminum tower was attached to the pole approximately 5 feet off the ground using a hinged steel plate. A "V" shaped bracket was attached about ten feet from the top of the

wood pole to secure the top of the aluminum tower. A block and tackle arrangement between the pole and the tower, along with the hinge at the tower's base, were used to raise and lower the tower so that the sensors could be serviced from the ground. The ability to raise and lower the tower was considered important for servicing the sensors, given the severity of the coastal environment and the fragile nature of temperature and humidity sensors.

Seven relative humidity and temperature sensors manufactured by Solomat Instrumentation were attached to the tower with 30-inch-long metal arms. The arms were used to place the sensors away from structures that would interfere with air flow. The fixed sensors were attached to the aluminum tower using the following spacing from the lowest sensor: 2 feet, 6 feet, 12 feet, 32.5 feet, 52.75 feet, and 73 feet. The lowest sensor was approximately 6.25 feet from the sand. The lower sensors were spaced more closely to allow greater resolution in the evaporative region.

An eighth sensor, identical to the others, was attached to a motor-driven shuttle by a slightly longer angled arm so that the movable sensor could be brought directly in line with each fixed sensor on the tower. This eighth sensor was used as the offset and drift check between the fixed sensors. The motor and gear box that powered the shuttle were attached to the wood pole at the base of the aluminum tower. The motor was controlled by a programmable drive housed in a nearby trailer along with a personal computer (PC) and data storage equipment. Since the motor drive was programmable, an automatic sequence was used to move the eighth sensor up and down the tower and stop it at each sensor position. There was also a set of manual override switches for the motor so that the program could be killed at any time and the shuttle moved to any desired location on the pole. The relative humidity and temperature signals were hard-wired to a waterproof box attached to the wood pole approximately five feet off the sand. These inputs to the waterproof box went first to a multiplexer so that all sixteen temperature and relative humidity values could be sampled by the Campbell Scientific CR10 programmable datalogger. The datalogger sampled the analog inputs and sent them via an RS485 link to the PC, which was housed in the trailer approximately 150 feet away. The RS485 signals were converted to RS232 and input to one of the serial ports on the PC.

Procedures

The PC was used to retrieve data from the datalogger, store the incoming data, calculate and display the associated refractivity profiles, and download programs to the Campbell datalogger or the motor-drive controller. The Campbell datalogger was programmed to sample all temperature and relative humidity values every five seconds. The Campbell program also applied the offset and slope corrections derived from earlier testing in an environmentally controlled chamber.

The PC's other functions were accomplished through a BASIC program which read the serial port, stored time, temperature, and humidity to a Bernoulli disk, and calculated and displayed the modified refractivity profiles. The surface pressure was either assumed to be a default value equal to 1013 millibars or a value supplied through a PC keyboard entry. The surface pressure was decreased for each successively higher fixed sensor position using a simplified form of the hydrostatic equation valid at low altitudes.⁶ This pressure value, along with the measured values of temperature and relative humidity, were used in the BASIC program as input to the Smith-

Weintraub relationship for modified refractivity.⁷ Finally, the BASIC program plotted the refractivity profiles from the above calculations.

Results

The roving sensor, used as a drift and offset check, was originally left on station approximately 30 seconds for full response, considering the manufacturer's specification of one time constant being equal to 3.5 seconds. After plotting the temperature response of the roving sensor along with its adjacent stationary sensor versus time, it became evident that much more time was required for the roving sensor to respond to the step change in temperature experienced when moving from station to station. From the data, the probe's response appeared to be on the order of six minutes. Part of the additional time necessary for the roving probe to fully respond to a change in temperature can be attributed to the 0.1-micron, filtered probe tip, which was necessary for the sea environment. Several values for the time the roving sensor spent adjacent to each stationary sensor were tested by merely altering the program used to control the shuttle motor drive. Ultimately, it was decided to leave the roving probe on station for eight minutes before moving it to the next stationary sensor's position. Eight minutes would allow for the sensor to fully respond to the overall average change in temperature and allow additional time (approximately two minutes) after full response for pertinent drift and offset data to be recorded. Figures 5 and 6 are graphs of temperature and relative humidity for the roving sensor and fixed sensor number 4. It should be noted that small perturbations in temperature could be seen on both the stationary and roving probes as the two came into agreement. Also, the lag in the response of the relative humidity sensor was usually less than for the temperature sensor. In a few cases when the roving sensor was coming from a position where the humidity was near 100 percent to a position where the humidity was as low as 70 percent, there was a lag in the response of the roving probe on the order of the temperature sensor's lag. The amount of time it took the roving sensor to respond to the average delta temperature between fixed positions was somewhat dependent on the average temperature difference between the fixed sensor sites. A larger mean delta temperature between measurement positions required a longer period of time for the roving sensor to respond. The same was true of the relative humidity readings in cases where the relative humidity delta between stations was large.

A set of data was recorded using the roving probe without its protective filter tip, but with

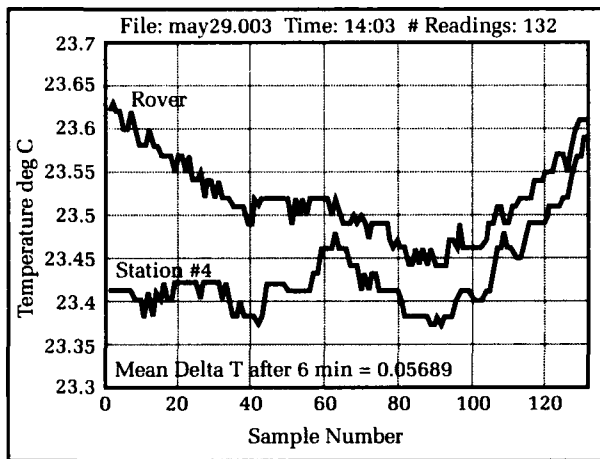


Figure 5. Temperature (rover vs. #4).

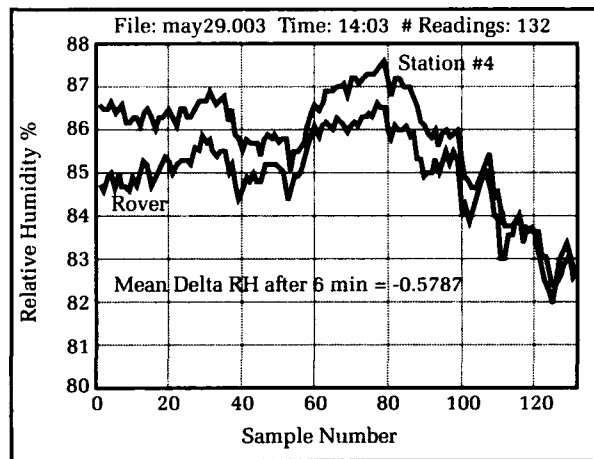


Figure 6. Relative humidity (rover vs. #4).

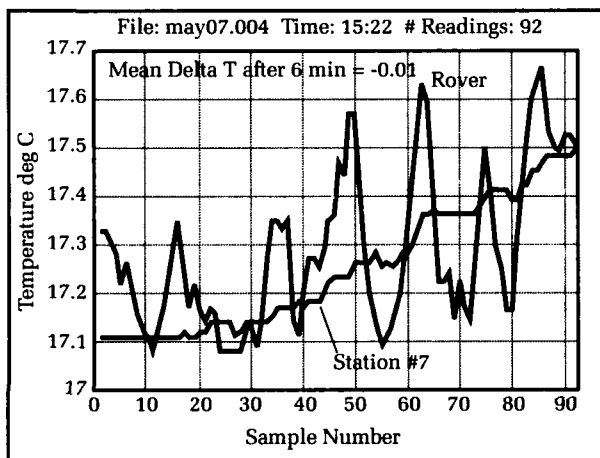


Figure 7. Temperature (rover without filter tip vs. #7 with filter tip).

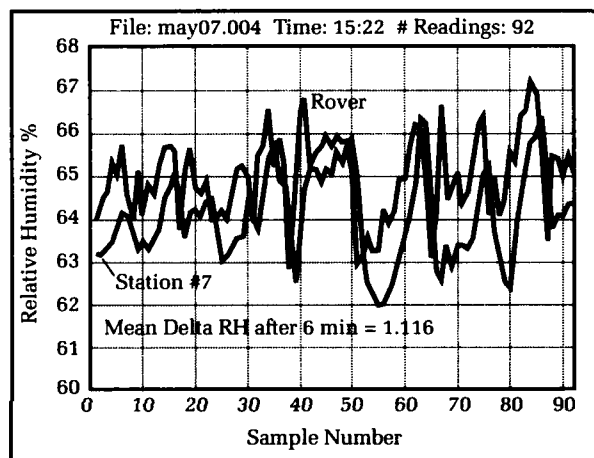


Figure 8. Relative humidity (rover without filter tip vs. #7 with filter tip).

filter tips on all fixed sensors. As described earlier, the roving sensor was brought in line with each fixed sensor and data was recorded for at least eight minutes per fixed-sensor position. Figures 7 and 8 show some results from this test. The relative humidity curves show a small delay in response, with the frequency of the fluctuations in both curves being similar. The temperature curves look as though the roving sensor's values were low-pass filtered, with the averages of both curves following nearly the same contour. The temperature curve shows a lag in response on the order of one to two minutes. The lag in the relative humidity response appears to be approximately nine seconds.

Early in the testing period, two sources of noise interfered sporadically with the temperature signals. The longer runs of data cable (the sensors higher on the pole and the roving sensor) exhibited noise problems more often than the shorter cable runs. It was determined that some of the very large spikes in the temperature data

were associated with the use of an SPS-48 radar used in high-power mode during testing at a nearby location. When the SPS-48 was used in its low-power mode, the interference was no longer present. A second source of noise resulted from the pulses sent to the stepper motor that moved the roving sensor shuttle. A large-diameter metal braid was pulled over the cable sending these pulses to the motor, and proved to be an effective shield for the motor noise. Other emitters used at Wallops Island did not prove to be a persistent problem. It should be noted for future tests, however, that careful attention should be given to proper shielding of the system.

Figures 9 through 12 show examples of refractivity profiles made at various times on two of the test days. All times given on the figures are Universal Coordinated Time (UTC), and the line marked "Std Atm" is a standard atmosphere reference. These profiles were made using the calibration offsets measured by the roving sensor. Offset values were determined by taking the

Figure 9. Delta M
unit profiles
spanning 14 min.,
April 25, 19:17 UTC.

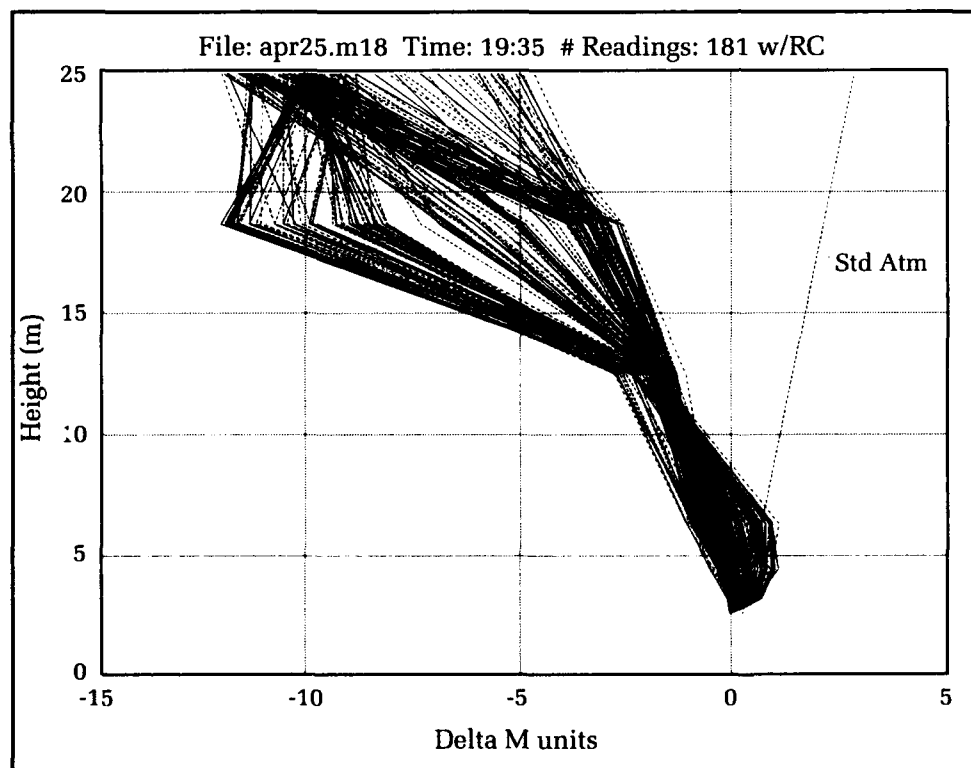
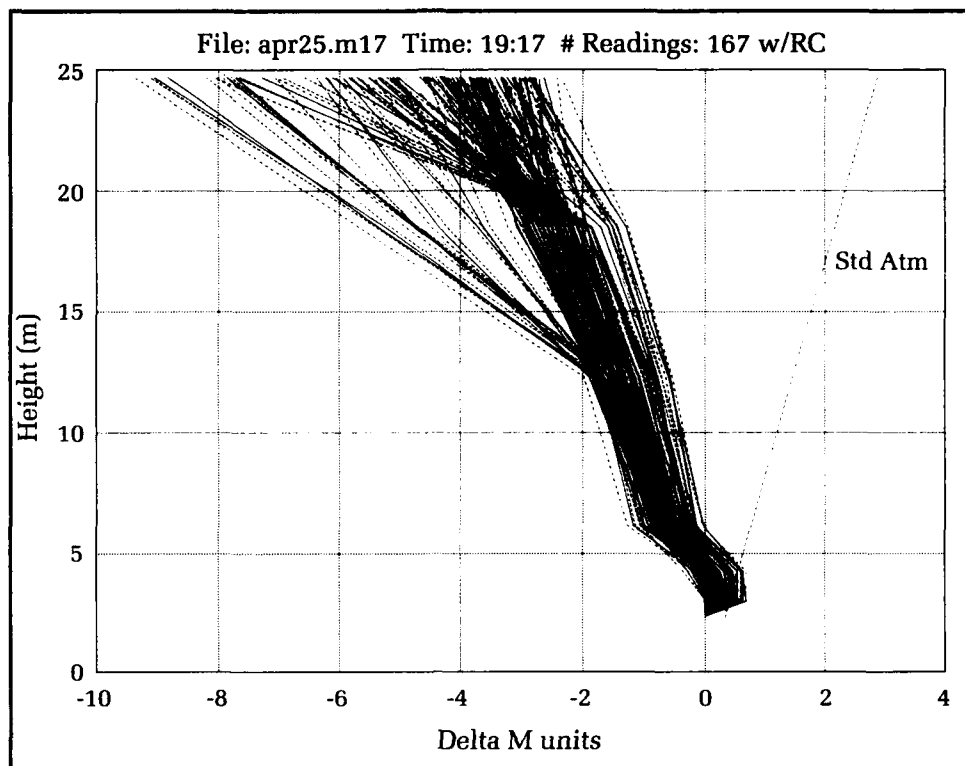


Figure 10. Delta M
unit profiles
spanning 15 min.,
April 25, 19:35 UTC.

Figure 11. Delta M unit profiles spanning 26 min., May 29, 12:48 UTC.

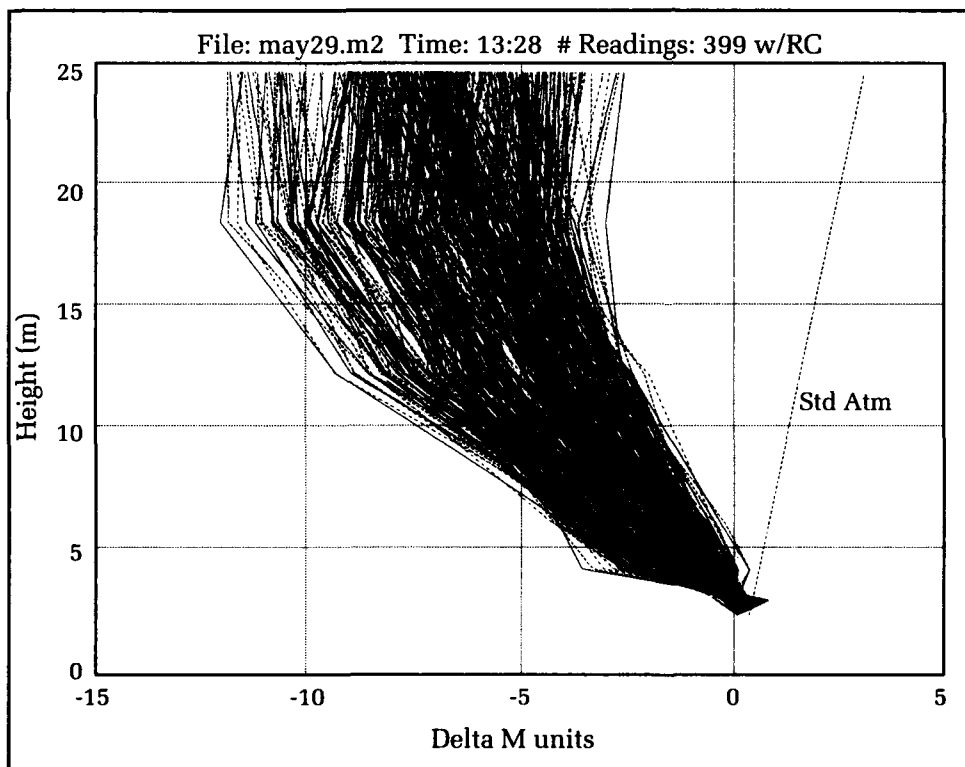
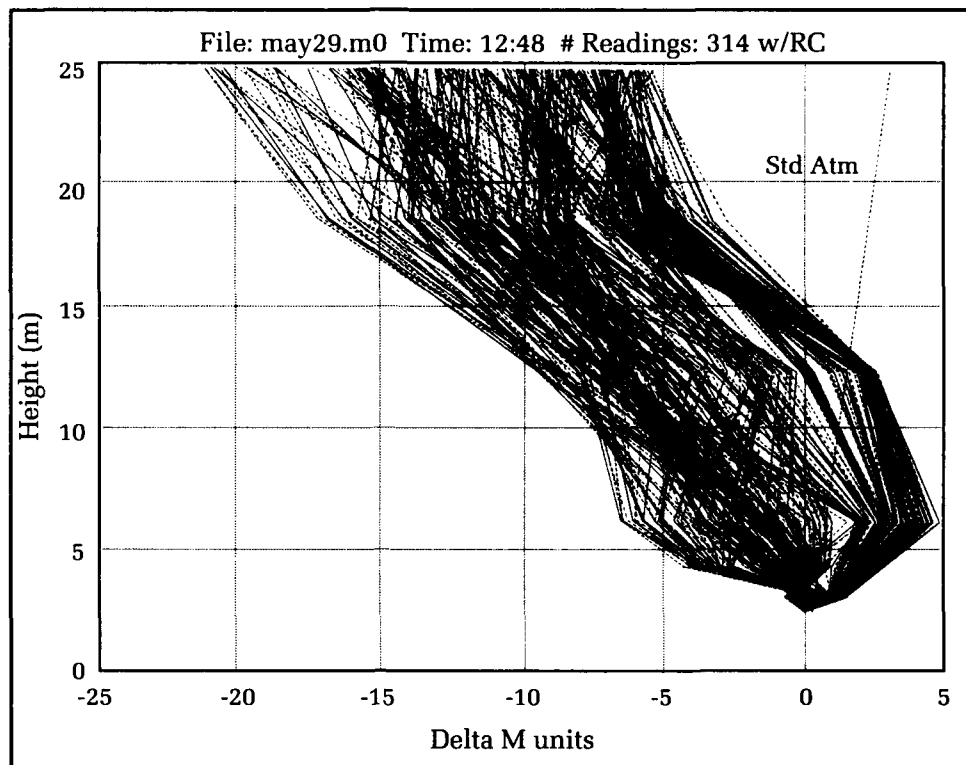


Figure 12. Delta M unit profiles spanning 33 min., May 29, 13:28 UTC.

average difference between the fixed sensor's reading of either temperature or relative humidity and the same reading made by the roving sensor. The first six minutes of data were not used in calculating the offset to allow time for the roving sensor to respond fully to changes encountered between the fixed sensor positions. This meant that approximately two minutes of data typically remained for calculation of the average offset. In some cases the period over which the offset was averaged was longer because the roving sensor was left on station for over eight minutes total. These special runs were made with the roving shuttle in manual mode. The roving sensor offsets are important in that they provide a reference measurement between all the fixed sensors. The M gradients shown in the figures are referenced to the value of M for the lowest sensor position, hence the label "Delta M units." This allows comparison of profile gradients regardless of the absolute M values. For the figures shown, most of the variability across time occurs in the top ten meters of the profiles. It is interesting to note in Figure 11 that there is a spread of 17 M units over a period of 26 minutes for the uppermost sensor. Also, the lower sensor readings in Figure 11 show over 10 M units variability for the same 26-minute period. In comparison, Figure 12 shows little variability in M for the lower sensors and approximately 9 M units of variability in the upper portion of the profile during a 33-minute period. The degree of error in the measurements of modified refractivity is approximately two M units, given the sensor accuracies of 0.2 degree Celsius for temperature and 2.0 percent relative humidity.

Another notable feature of these profiles is that the upper slope often does not become positive. This indicates that the top of the ducting layer has not been measured, and points out the maximum height limitation of this refractivity profile measurement system.

During many days of testing, especially in the afternoon when the land breezes switched to sea breezes, problems were experienced with condensation or sea-spray. The radiation shields had droplets of water around the edges of each plate. When the probes were removed for examination, droplets of water were also observed on the filtered probe tip. These water droplets obviously led to high relative-humidity readings. Salt was evident on some of the radiation shields, indicating that sea spray was also a factor. In order to try to measure the relative differences in

humidity in these conditions, the sensor values were recorded even if they indicated above 100 percent. The high-water mark on the beach was chosen so that the measurements might be able to sense evaporative ducts; however, the breaking waves and the large swings in temperature (some days as much as six-degrees Celsius) made this location difficult for measuring relative humidity. As mentioned, the relative humidity measurements for the later test days were recorded even if they indicated greater than 100 percent in order to study relative differences in humidity with height. Since the sensor used does not measure relative humidity accurately above approximately 95 percent, there is very little confidence that refractivity profiles generated from relative humidity readings in excess of 100 percent are representative of the actual profiles, and none have been included herein.

Further Analysis

Radar propagation measurements were made simultaneously with the refractive profiles discussed above. The recorded radar returns indicated a vertical propagation structure similar to that produced by using the concurrent refractivity profile as input to a parabolic, equation-based propagation model. The RF returns also showed the same fluctuations as measured in the refractive profiles. More propagation data are being reduced at this time and will be the subject of future reports.

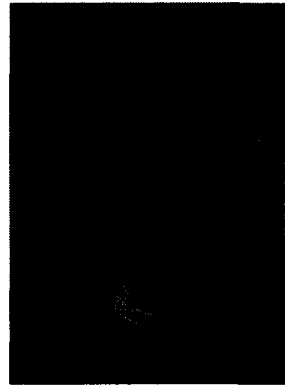
Summary

This refractivity measurement system, designed to study the temporal fluctuation of refractivity profiles, is unique in that it provides the capability to measure vertical profiles instantaneously. Most other refractivity measurement systems in current use are not suitable for this task. While the tested system proved to have limitations in its maximum height and its robustness to condensation and sea spray, it yielded a large amount of interesting data. Variations as large as 17 M units were observed at a single sensor location over a period of 26 minutes. These fluctuations in the refractive profile can significantly impact radar propagation and likely occur at all ranges in the radar's propagation path. The stochastic nature of these refractive profiles suggests that propagation modeling should be extended to include such profile variations.

References

1. Boithias, Lucien, *Radio Wave Propagation*, McGraw-Hill Book Company, 1987.
2. Hall, M. P. M., *Effects of the Troposphere on Radio Communication*, Peter Peregrinus Ltd., 1979.
3. Patterson, W. L., et al., *Engineer's Refractive Effects Prediction System (EREPS) Revision 2.0*, Technical Document 1342, Naval Ocean Systems Center, San Diego, CA, Feb 1990.
4. Rotheram, S., *Third International Conference on Antennas and Propagation ICAP 83*, part 2, IEE London, UK, 1983, pp. 9-13.
5. Rowland, John R. and Babin, Steven M., "Fine-Scale Measurements of Microwave Refractivity Profiles with Helicopter and Low-Cost Rocket Probes," *Johns Hopkins APL Technical Digest*, Vol. 8, No. 4, 1987, pp. 413-417.
6. *Smithsonian Meteorological Tables*, Sixth Revised Edition, prepared by Robert J. List, Smithsonian Institution Press, City of Washington, Fifth Reprint, Aug 1971.
7. Bean, B. R. and Dutton, E. J., *Radio Meteorology*, Dover Publications, Inc., New York, 1968.

The Author



JANET K. STAPLETON received the B.S.E.E. degree in 1985 from Tennessee Technological University. Since joining the Naval Surface Warfare Center (now Dahlgren Division) in 1985, she has performed radar system, radar pulse doppler waveform, radar sea clutter, and propagation studies. She participated in the NATO Anti-Air Warfare Program in the clutter and propagation subgroup and developed a simplified ducted sea-clutter model which has been used on several programs for analysis purposes. She has also contributed to the AN/SPY-1 Radar Element Upgrade Program in the area of waveform design and analysis. Most recently, she has performed refractivity measurements in support of radar testing, including assessment of the temporal variation of refractivity profiles.

Top-Level Models for Air Defense and Ship Survivability

Richard L. Humphrey

Tradeoffs among self-defense capability, expected threat strengths, and the ability of combatants to absorb punishment are difficult because of the large numbers of variables involved. Two top-level models are presented that give an overview of the interplay among the primary variables. The first is an air-defense model based on the premise of perfect detection and fire-control sensors. The second, based on empirical data, estimates the expected levels of damage when the ship is hit by explosive warheads. Applications include wargaming and first estimates for sizing ship concepts and defensive suites.

Introduction

Why bother with simple parametric and empirical models when we have a number of highly detailed simulation and other models for assessing Anti-Air Warfare (AAW) performance and determining the survivability of combatants after being hit by missiles, bombs, or projectiles? The answer is relatively straightforward. Simple models allow insight into the relationships among the basic parameters that may be difficult to obtain with highly detailed models. They give limiting or bounding cases for comparison with more sophisticated models. They also prove extremely useful during the initial design phases of an air-defense system and its platform.

Simple parametric models serve as reality checks and give a better understanding of the process being modeled. Realism can be added to wargaming by introducing some of the randomness of actual combat. Lastly, the models are often simple enough that results can be obtained by back-of-the-envelope methods, pocket and programmable calculators, and personal computers.

This article introduces two simple models^{1,2} that meet the above requirements, providing insight and useful tools for the study of air defense and ship damage.

The first model we will consider, for which perfect detection, C^3 (Command, Control, Communication), and fire-control capability are assumed, provides a limiting case that can be used to compare system capabilities estimated from existing models with the best-case solution. Cost, effectiveness, and survivability tradeoffs can then be made to determine the relative usefulness of obtaining additional sensor and fire-control performance. The impact of improvements in weapon performance is also easily estimated.

The second model has been shown to give a broad correlation between the estimated probability of damage and the level of loss of sensor, C^3 , and weapon

capability suffered by the ship. Again, tradeoffs can be made to match survivability, cost, and mission requirements.

Concept

The air defense of a ship or task force consists of a number of processes, including detection of the threat, evaluation and assignment of weapons, engagement, kill assessment, and re-engagement, if necessary. Threats that the combat systems, both active and passive, are unable to engage successfully would be expected to hit the targeted ship and to inflict a degree of damage dependent on the size of the ship, the size of the threat warhead, and other variables. Not all the processes involved need to be included in a simple model, as some of them are highly certain to happen.

The models are based on physical fundamentals or empirical data, and only the most influential parameters are included. The AAW model will give a best-case result based on fundamental kill probabilities and numbers of weapons that can engage the threats. This model can be characterized as an "expected value solution." The ship damage model, based on an empirical assessment of naval combatant experience during World War II, results in a "statistical" model.

Air-Defense Model

The air-defense model is built upon some parameters of AAW that, while not 100 percent true in every situation, are normally present to the extent that they can be taken as characteristic. These are: (a) the detection of the threat aircraft or missile is almost certain; (b) destruction of the threat is the primary problem; (c) weapon resources are limited to those on board the ship at the beginning of the battle or raid; and (d) these weapon resources are not replenishable during the battle or raid.

The critical remaining parameters are: (a) the number of threat aircraft or missiles attacking;

Table 1. J Derivation

$$P_k = 1 - \prod_{i=1}^n (1 - P_i)$$

$$1 - P_k = \prod_{i=1}^n (1 - P_i)$$

$$\ln(1 - P_k)^{-1} = \sum_{i=1}^n \ln(1 - P_i)^{-1}$$

$$J(P_k) \equiv \ln(1 - P_k)^{-1}$$

$$J(P_i) \equiv \ln(1 - P_i)^{-1}$$

$$J(P_k) = \sum_{i=1}^n J(P_i)$$

(b) the number of defending weapons or the number of times they can engage; and (c) the probability of kill that the weapons exhibit against the threats.

Derivation

The method used is based on the simple expected value of P_k , or total kill probability, with multiple shots when individual single-shot kill probabilities of P_i are used. Working with $1 - P_k$ as shown in Table 1, the natural log of the reciprocal of this function is taken and the working parameter, J , is defined as $J(P_k) \equiv \ln(1 - P_k)^{-1}$. A similar function $J(P_i) \equiv \ln(1 - P_i)^{-1}$ can be defined in terms of P_i . The result has been to linearize the values of P_k and P_i , allowing simple sums and multiples of J to be used to define an effectiveness parameter.

Behavior of J

The behavior of P_k and J are illustrated in Figure 1. Figure 1a shows each when P_k is plotted in a linear form. As P_k approaches one, the

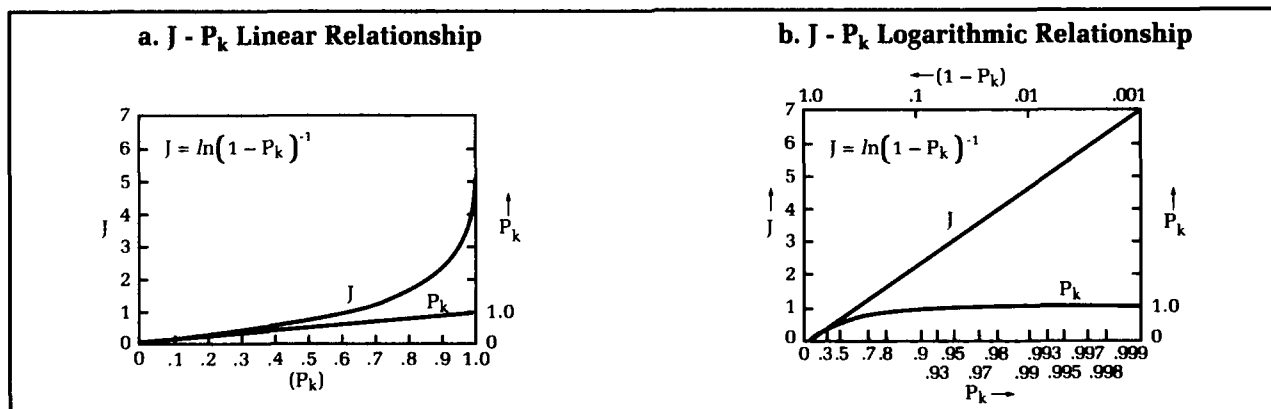


Figure 1. The behavior of J and P_k .

value of J increases rapidly. When a logarithmic form (Figure 1b) is used, J is linear while P_k asymptotically approaches unity. Note that the value of J is still relatively handy in size, even for high values of P_k . For small values of P_k , P_k and J are very nearly the same.

The linear form of J has a number of advantages. Individual single-shot J values from a number of firings or different weapons can be added together to obtain a composite J for a ship or battle group. The form is convenient, as J values can be developed as J per unit time (flux rates) for a weapon or for a magazine or an illuminator. Some of the qualities of J are amplified in Table 2 in more detail.

Surviving Threats

Returning to the basic definitions of J , it is relatively straightforward to develop a simple equation for the number of threats that will survive the AAW defenses. T is the number of threat aircraft and/or missiles, and the survivors of a particular layer of defenses are designated S . These S survivors must be handled by the next layer of defenses if a multilayered defense is used. If no inner layers exist or if the J s for all the AAW defenses have been lumped together, these survivors will impact the ship or task force. The number of threats T may have a time factor attached to it that can be matched with the corresponding J available during the time interval. The number of surviving threats is defined as:

$$S = Te^{-\Sigma J / T}$$

where J is the defensive capability available during the period of time that T threats will arrive. S will be the number of surviving threats from that raid. Details of the derivation are given in Table 3.

Table 2. Qualities of J

$$J(\Sigma) = J_1 n_1 + J_2 n_2 + J_3 n_3 + \dots$$

Where each J and n are the values associated with an individual weapon or weapon system; N is the sum of the individual N s in terms of the numbers of weapons, engagements, or salvos available.

J is linear and fully compatible with all basic arithmetic operations.

Therefore we can calculate:

J values for a given missile

J values for a ship's loadout

J flux (J per unit time) per launcher or gun

J flux per ship or battle group

J flux deliverable to a given point in the vicinity of a battle group by the elements of the battle group.

J itself is not a function of the number of targets presented.

J itself is not a function of the weapon range, but should be an effective value over the useful range of the weapon.

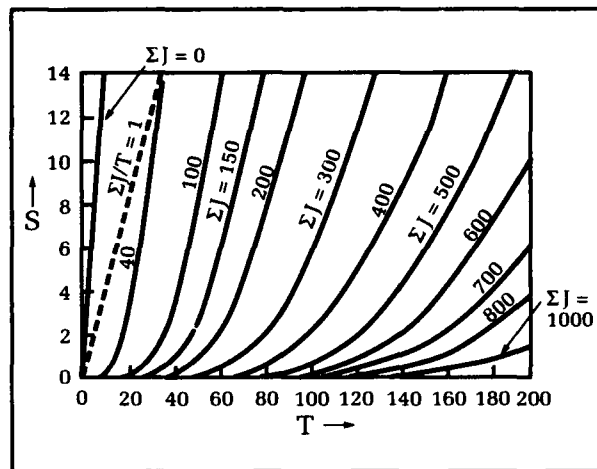


Figure 2. Typical plot of T , J , S for a large battle force.

A plot of T , J , and S for values of the three parameters that might be representative of the ranges of threats facing a large battle group with varied AAW defensive levels is given in Figure 2. Note that when J is relatively large compared to T , very few of the threats survive the defenses (bottom of the plot). As T becomes larger relative to J , that is, the number of threats begins to saturate the defenses, the number of survivors increases rapidly. The region in the area near the line labeled $J/T = 1$ is an area where very large numbers of threats will leak through the defenses since magazines will have been depleted in the region near this point.

Target-Rich Environments

Large numbers of threats pose difficult problems for AAW defenses. The fire control circuits, missile launchers, or AAW guns may become fully loaded or dedicated, thus incapable of handling additional threats. When any portion of

Table 3. Surviving Threats

$$J(P_k) = \ln(1 - P_k)^{-1}$$

$$\Sigma J / T \equiv J_1$$

$$1 - P_1 = e^{-J_1}$$

$$S \equiv T(1 - P_1) = Te^{-J_1}$$

$$S = Te^{-\Sigma J / T}$$

$$T / S = e^{\Sigma J / T}$$

$$\Sigma J = T \ln(T / S)$$

the entire detection-through-engagement chain is overloaded, the system is said to be saturated. The crucial limiting factor for this model is taken to the point at which either the fire-control channels are all occupied, or the weapon systems are all tied up or have been expended. At saturation then, N is equal to T , equal to T_s , the number of threats at the saturation point. At this point an average value of J can be defined as $J_a = J/N$. J_a represents then the discrete lumps with which weapon effectiveness can be represented. At the saturation point, one lump of J , J_a , corresponding to one missile, one salvo, or one gun burst/engagement, has been dedicated to each threat. In other words, no more J is available.

When the number of threats is small compared with the number of weapons or channels available, $T/N < 1$, the system will be effective in handling the threats. If $T/N \ll 1$, the case for light loading, a very small fraction of the threats will survive the defenses and threaten the ship. Figure 3 illustrates the proportion of surviving threats for the full range of threat loadings. As T/N becomes larger, the relative number of survivors will increase faster, with the rate depending upon the value of J_a . Low values of J_a , indicating relatively low values for average kill probabilities, show rapid increases in the rate of survival as compared to high values of J_a with corresponding high average values of kill probability per missile, salvo, or burst. The situation is analogous to the reduced efficiency of a telephone system or a time-sharing computer as the number of system users approaches the design level. The user has to wait longer for service, and is likely to encounter more delays and interruptions. Eventually the system can collapse.

At the saturation point, the maximum number of threats has been killed and any subsequent threats that arrive will get a free ride through the defenses. At the saturation point, the number of survivors is $S_s = T_s \text{EXP}(-J_a) = T_s \text{EXP}(-J/N)$. T_s is the number of threats at which the system becomes saturated. At this point all the available defensive units have been dedicated to the threats. If larger numbers of threats are then pre-

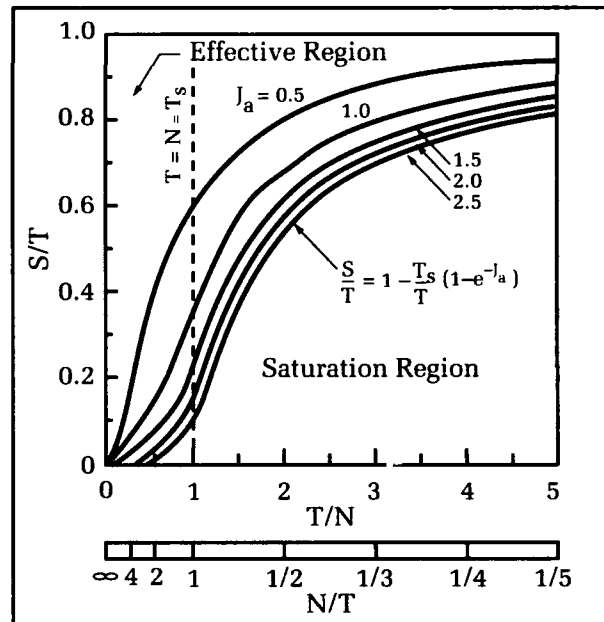


Figure 3. Behavior for large numbers of threats. T/N represents the saturation ratio; N/T is the inverse ratio.

sented, the excess numbers over T_s will get the free ride. It is now possible to estimate the fraction of surviving threats after saturation. Then, for the saturation region: $S = S_s + (T - T_s)$ and substituting for S_s , $S = T_s \text{EXP}(-J_a) + (T - T_s)$ and finally, after simplification,

$$S / T = 1 - (T_s / T)(1 - \text{EXP}(-J_a)).$$

This result is plotted in Figure 3. As the saturation ratio T/N increases, the fraction of survivors will approach unity. At large values of T/N , virtually all the threats will survive with only modest dependence on the values of J_a available before saturation occurs. This region could be called a "Target-Rich Environment," where there are many more targets than capability to handle them.

The results of such an environment are pictured in Figure 4. Col. George Custer and his regiment met a very large number of Indians at the Little Big Horn in 1876. From Custer's point of view, the battle met all the characteristics of naval AAW cited earlier: there was no problem detecting the threatening force; defending against them was the issue; Custer had limited supplies of ammunition and manpower; and he could not expect any reinforcement or replenishment during the battle. Custer and his men fought until their ammunition was exhausted. There were no survivors. The parallel in a naval environment can easily be imagined. The moral is: "Avoid having to defend in very high threat environments."

An Example

A missile launcher aboard a ship contains six missiles, each of which has a single-shot P_i of 0.6. If four threats arrive in a slow stream, what is the expected number of hits on our ship if a shoot-look-shoot firing doctrine is used? A P_i of 0.6 corresponds to a J_i of 0.916. With six missiles, J is 5.498, and J/T is then 1.374. By substituting into the survivor equation, $S/T = 0.253$, then 1.01 threats would be expected to survive the defenses and impact the ship.

Layered Defenses

Layered defenses can be readily handled by application of the model in each layer of the defenses. The survivors of one layer become the threats for the next layer. The process is illustrated in Figure 5 for the case of a three-layer defense. The first layer might be CAP (Combat Air Patrol) aircraft; the second, area missile defenses; and the third, missile or gun point defenses. The use of layering allows the number of threats killed in each layer to be estimated as well as the number of missiles expended in that layer. The number of survivors reaching the battle group or ship are the same whether calculated layer by layer, or lumped into one defensive package.

Deviations from Mean Values

This model provides mean or expected values of the number of threats killed and the number surviving the defenses. It is of considerable interest to have some estimate of the range of values that might be expected if a more complete analysis were undertaken.

The threats are always discrete whole numbers, and the defensive capabilities are always limited to the expenditure of one missile, gun round, or burst of fire. This suggests the use of a discrete distribution, such as the Poisson or binomial distributions, as an approximation for estimating the deviations that can be expected. Other appropriate discrete distributions may be used.

The following are presented as suggestions for frequent cases encountered:

If T is a number of order 1 to ~30 and P_{ka} is large (>0.15), try a Poisson distribution, where:

$$\text{Mean: } \mu = \lambda = (T - S) \text{ (Number Killed)}$$

$$\text{Std. Dev.: } \sigma = \sqrt{\lambda} = \sqrt{(T - S)}.$$

If T is large (>30) and P_{ka} is small (<0.15), try a binomial distribution where:

$$\text{Mean: } \mu = (T - S)$$

$$\text{Std. Dev.: } \sigma = \sqrt{(T - S)(1 - P_{ka})}.$$

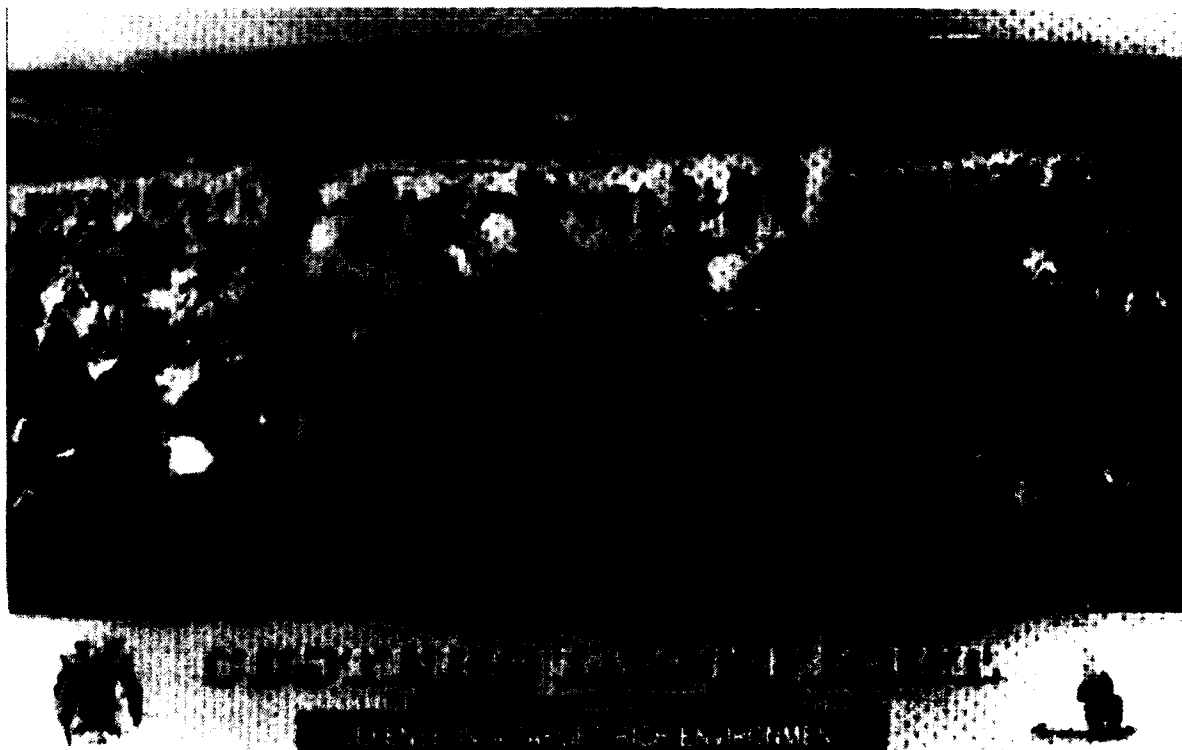


Figure 4. Defense in a target-rich environment.

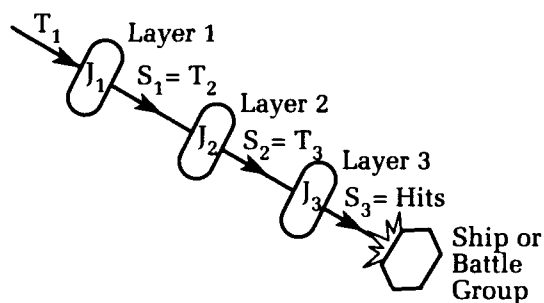


Figure 5. Schematic of a three-layer defense.

The ranges of T and average P_k (P_{ka}) are only crude guides. The user's experience with these distributions may suggest other bounds. The average P_k can be estimated by dividing the total J by the number of shots or individual weapons it is summed over. This average J can then be converted back to an average P_k . The method is relatively crude, but should be adequate for the top-level analysis we are trying to do. The example given earlier corresponds to the Poisson case where the mean number killed is $4 - 1.01 = 2.99$ and the standard deviation is $\text{SQR}(2.99) = 1.73$. Thus the number killed might be as low as $2.99 - 1.73$, or as large as 4.0 threats.

Limits and Cautions

A few things should be kept in mind when applying this method. Some time-related problems may need to be considered in selecting suitable groupings of T and J . Among these are limits on firing rates, illuminator and director handling capacities, and the bunching of the threats. Capacity limits may also be important. Among them are magazine capacities, P_i of weapons against different threats, and missile firing practices (shoot-shoot, shoot-look-shoot, etc.).

No characteristic of the threat itself such as speed, range, or maneuverability has been considered. It is assumed that the defending weapon, under the conditions of the scenario, is capable of engaging the particular threat with an expected value of P_i .

Ship Damage Model

Wide variations in the design, construction, size, and damage control capabilities of combatants make it difficult to determine in detail the amount of damage that a particular ship can sustain before being unable to achieve its mission or being sunk. Since particular characteristics of a combatant are seldom known in any detail, it is

useful if the simplest possible parameters are chosen to represent its resistance to damage. Displacement is one such parameter. The size of the warheads (with size represented by the weight of the explosive in the warhead) hitting the ship is another characteristic that may be known or estimated.

The dimensionless parameter, D/He , was chosen as the fundamental parameter. D is the full load displacement of the ship in tons, and He is the equivalent weight (pounds of TNT) of the explosive contained in the projectile, bomb, or missile warhead. The other parameter is the number of hits suffered by the ship. The further assumption is made that the warhead penetrates the structure of the ship and detonates within it. This is the most common case, but would exclude damage caused by Anti-Radiation Missiles.

Historical data from World War II was used in the correlations. Data can be drawn from several sources, including Morison,³ Korotkin,⁴ Chesneau,⁵ Campbell,⁶ and handbooks.⁷ For this particular correlation, an attempt was made to use only those examples where a single type of weapon was used against the ship. The ships ranged in size from about 1,000 tons to over 30,000 tons displacement. The data set consisted of 30 ships sunk and 49 damaged by bombs or gunfire. The data sets are plotted in Figures 6 and 7.

A form for the rule where the dimensionless ratio of displacement and explosive would be raised to some power and multiplied by a constant would be developed that would vary with the probability level and the category. The probability would correspond to the fraction of ships either damaged or sunk out of the population of ships in the category. The dimensionless ratio of D/He with a slope of one half was found to give the best fit among those tried.

The damage rule can be written as $N = C(D/He)^{1/2}$ where N is the number of hits required to damage or sink at the expected probability level in that category; P , C is a coefficient dependent on category and probability; D is the full load displacement in tons; and He is the TNT equivalent of explosive in the warhead in pounds. Table 4 gives the fits for the values of C for the damage set and for the sinking set of data.

Plots of C and the data points to which the equations were fit are shown in Figure 8. The degree of fit is quite good, with values of the regression coefficient above 0.97. It can be seen that sinking a ship, on average, requires considerably more hits than does damaging it. This is to be expected. It should also be remembered that P_d or P_s cannot be less than zero or greater than unity. Values of P_d near 0.5 can be consid-

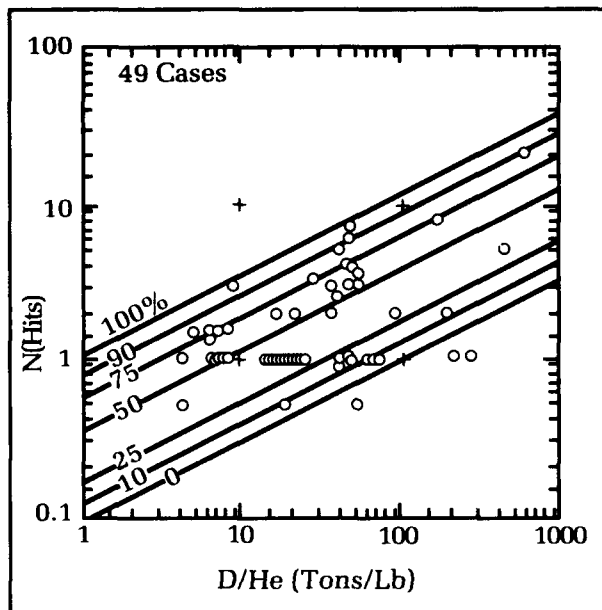


Figure 6. Damage by bombs, projectiles, and missiles.

ered to be mission kills, while a P_d near 0.9 would indicate major damage and a P_d near 0.1 would indicate relatively minor damage.

The curve fits can be used in a number of ways depending on the information needed. Figure 9 shows the number of hits required at several confidence levels to sink a 40,000-ton warship as a function of the explosive weight of the bomb or missile warhead. For a confidence level of 50 percent for sinking, Figure 10 gives the number of hits required for different size bombs and a range of warship displacements.

It can also be observed that the product of N and He , the total weight of explosive delivered on target (W), is reduced when small warheads are used, even though the number required has increased:

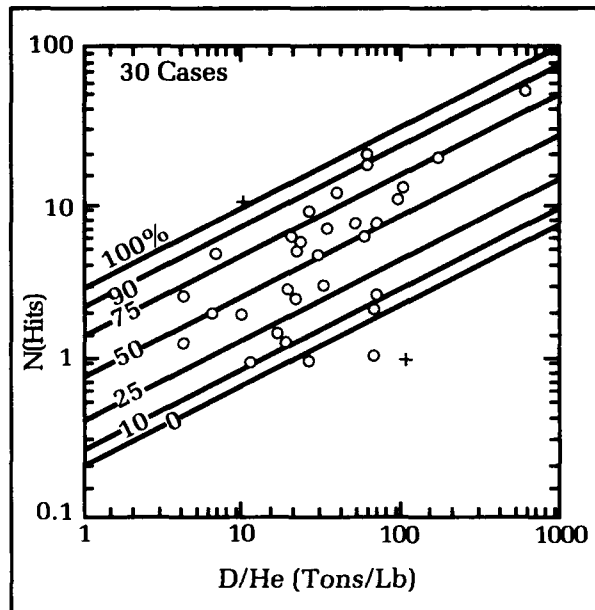


Figure 7. Sinkings by bombs, projectiles, and missiles.

$$W = N He = C [SQR(D/He)] He$$

$$W = C [SQR(D \cdot He)]$$

N increases only as the square root of the decrease in warhead size. W , however, increases as the square root of the warhead size. This result should be viewed with caution. The formula does not imply that dropping a large number of one-pound bombs is an efficient way to damage or sink a warship. Recall that penetration of the structure was assumed earlier. The example shown in Figure 11 demonstrates the behavior of W as He is changed.

The volume damaged by a warhead is often limited by venting out the sides, deck, or entry hole.⁸ Thus, a large number of smaller warheads

Table 4. Fits for the Constant C

Damaged

$$N = C(D/He)^{1/2}$$

$$C = N(D/He)^{-1/2}$$

$$C = 0.097 \times 11.06^{P_d}$$

$$P_d = [\ln(C/0.097)]/2.40$$

$$C_0 = 0.097 \quad P = 0$$

$$C_1 = 1.073 \quad P = 1$$

Sinking

$$C = 0.224 \times 14.18^{P_s}$$

$$P_s = [\ln(C/0.224)]/2.65$$

$$C_0 = 0.224$$

$$C_1 = 3.176$$

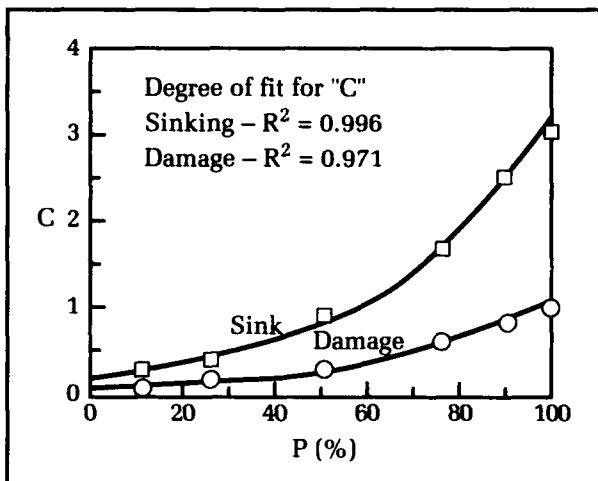


Figure 8. Coefficient curve fits for bomb, projectile, and missile.

may be made more effective by limiting venting and by damaging multiple locations in the ship with a larger total damage volume.

Damaged or Sunk?

Because the two data sets overlap, the question will arise as to how one differentiates between a ship that was damaged or one that was sunk. The probabilities may be greater than zero in both calculations. The two probabilities can be compared and a random choice made around the ratio of the two. For instance, if the probability of damage was 0.8 and that for sink-

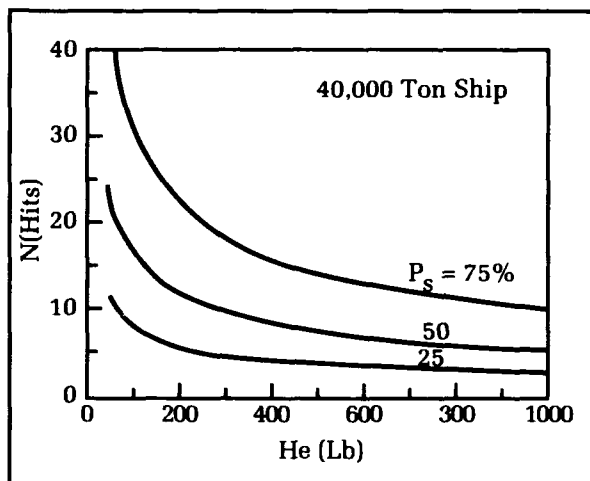


Figure 9. Number of bomb hits to sink a warship for varied warhead explosive content.

ing was 0.5, the odds would be 1.6 to 1 in favor of damage rather than sinking. The choice would be made by biasing a uniformly distributed random number generator from zero to one around the value of 0.625.

Why So Much Scatter?

The data have a great deal of scatter. There are a number of reasons for the side range of values for ships of similar size hit by similar warheads. Some of the scatter is due to shortfalls in the historical records. It is difficult in the heat of battle to be sure how many hits were

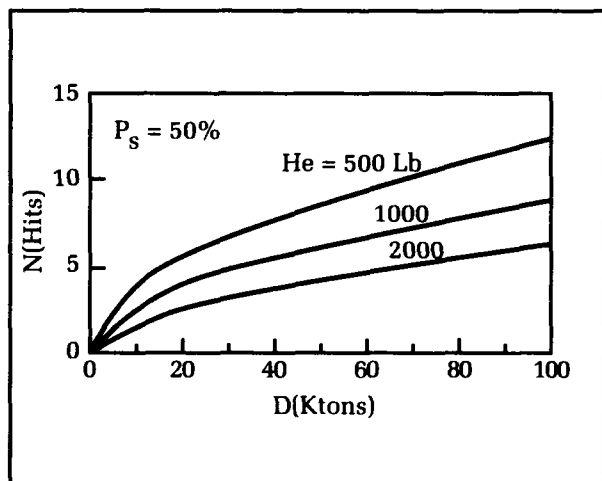


Figure 10. Number of bomb hits of three explosive weights needed to sink warships to 100 ktons.

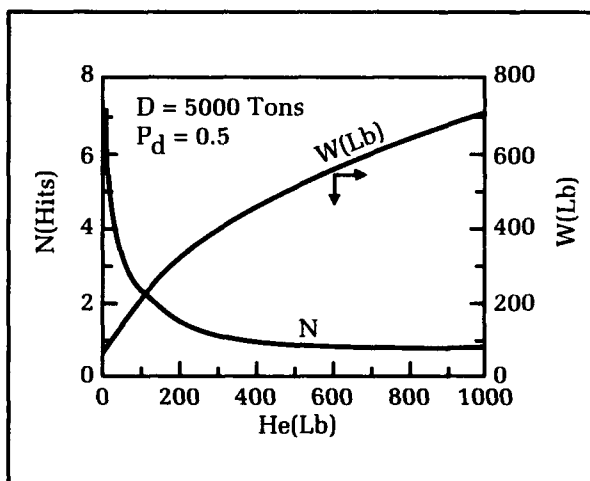


Figure 11. Effects of warhead size on number of hits and total explosive weight for a given level of damage.

taken and what the size of the bomb or missile was. Accounts on both sides will probably disagree. Observers in nearby ships may be too busy to note what happened.

The ships represented in the data set were built from about 1900 to 1945. They represent a number of generations of warship design and construction. The degree of subcompartmentation and armor can differ widely, as well as the quality of construction and maintenance.

Damage control capabilities also varied widely. In the U.S. Navy, damage control capability improved greatly during the war. Ships that would have been lost in the first year or two of the war were saved and returned to battle in later years.

It is very easy for attackers to concentrate on a wounded ship. Such a ship tends to move more slowly and is less able to fire back. Ships often take a relatively long time to sink, and the attackers may want to make sure that this one will go down. Thus, there is a tendency to use more ordnance than is absolutely necessary to put the ship out of action or to sink it.

The last factor is the "dumb luck" aspect. The British battle cruiser *Hood*, which exploded after a single hit from the German heavy cruiser *Prinz Eugen*, is such an example. A lucky hit in a vulnerable spot was enough to do the job. On the other hand, ships beyond any hope of salvage or repair often had to be sunk by additional gunfire or torpedoes.

An Example

In the earlier example, one threat penetrated the AAW defenses. Assume that this threat carried the equivalent of 500 pounds of TNT in the warhead and that the ship had a displacement of 1,500 tons. What might be the expected level of damage or the probability of sinking for this case? N is 1 and D/He is $1,500/500 = 3$. C then is $N(D/He)^{-1/2}$ or 0.577. The probability of damage, P_d , is then 0.74, indicating a level of damage greater than a mission kill and most likely of a major nature requiring extensive shipyard work. The probability of sinking, P_s , is then 0.36. From the ratio of the two probabilities, it can be seen that the ship is twice as likely to survive with heavy damage as it is to sink.

Applications

Wargaming

Realism in wargaming can be enhanced by the use of these two models. More of the uncertainty of actual combat can be simulated. The ship damage model, grounded in historical results, helps to fill in the data, lacking recent losses of warships in combat.

Initial Design Studies

The models may be useful for initial design studies of future ships as an aid in achieving balance between defensive systems and expected threats. A first look can be taken at the survivability of the combatant in battle by estimating leakers through AAW defenses and estimating the extent of damage to the ship.

Detailed Simulation Models

The AAW model is a best or limiting-case model. It can be used in conjunction with a detailed simulation to answer questions such as: How close does the proposed system come to the best case? Is it worthwhile to provide more capability to the system?

The damage model provides a historically based reference case. Does the proposed design do better or worse than the historical base? Tradeoffs among defensive systems and hull design might be indicated.

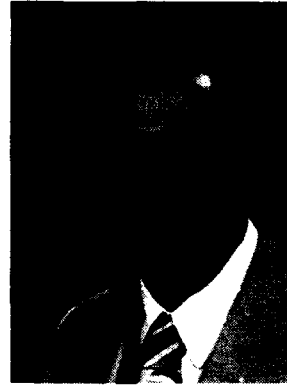
Conclusions

Each model provides insight into the critical parameters it contains. Such models provide simple first estimates that can be used in many ways to guide system design and tradeoffs. Although not as detailed as other models, they nevertheless allow the user to gain an appreciation of the limits that can be achieved in AAW system performance. They also provide a simple way to achieve a high degree of realism in wargaming of naval air and missile attacks.

References

1. Humphrey, R. L., "First Cut Model for Naval AAW Requirements," ORSA/TIMS Joint National Meeting, Philadelphia, PA, 29-31 Oct 1990.
2. Humphrey, R. L., "Warship Damage Rules for Naval Wargaming," TIMS/ORSA Joint National Meeting, Las Vegas, NV, 7-9 May 1990.
3. Morison, S. E., *History of United States Naval Operations in World War II*, Vol. I-XV, Little, Brown and Co., Boston, MA, 1962.
4. Korotkin, I. M., *Battle Damage to Surface Ships During World War II*, DTIC 437233, Feb 1964.
5. Chesneau, R., Ed., *Conway's All the World's Fighting Ships, 1922-1946*, Conway Maritime Press, London, 1980.
6. Campbell, J., *Naval Weapons of World War Two*, Conway Maritime Press, London, 1985.
7. Baumeister, T., Ed., *Mark's Standard Handbook for Mechanical Engineers*, Eighth Ed., McGraw-Hill Book Co., New York, NY, 1978, pp. 7-26 to 7-28.
8. Gates, P. J., *Surface Warships*, Vol. 3 of Brassey's Sea Power, Naval Vessels, Weapons Systems, and Technology Series, Brassey's Defence Publishers, London, 1987, pp. 166-170.

The Author



RICHARD HUMPHREY, a senior general engineer in the Performance Assessment and Analysis Branch, holds B.S. degrees in mechanical engineering (1957) and electrical engineering (1959) from the Missouri School of Mines and Metallurgy and an M.S. in mechanical engineering from the University of Tennessee. His recent work centers on methods for analysis of naval battles, analysis of combat systems for small combatants, reverse technology transfer studies, and development of measures of effectiveness for naval combat systems. Earlier, he worked in the wind tunnel area and on the Navy's High Energy Laser Program. He is a member of ASNE, ASME, IEEE, and is active in the Operations Research Society of America, where he has presented numerous papers.

Multisensor Integration and Data Fusion in the Surface Navy

R. E. Helmick, W. D. Blair, C. F. Fennemore and T. R. Rice

In recent years, interest in the synergistic use of multiple sensors to increase the capabilities of military systems has been growing. The advantages gained through the synergistic use of multisensory information can be decomposed into four fundamental aspects: redundancy, complementarity, timeliness, and cost of the information. These advantages will be discussed with reference to the surface Navy, where the use of multiple sensors can significantly impact target tracking, kill assessment, and over-the-horizon data fusion. While the interest in multisensor integration and data fusion for the surface Navy is high, the complexity of the issues involved in multisensor systems may stifle the development of combat systems that fully utilize the capabilities available through multiple sensors. Some of the issues involved in the development of such a system are sensor registration, selection of a fusion method, and selection of a data association method. These issues will be discussed relative to their impact on the surface Navy. Also, possible problems and future research directions for multisensor integration and data fusion in the surface Navy will be discussed.

Introduction

In today's surface Navy there is increasing interest in integrating stand-alone sensors into multisensor systems and fusing (or combining) data from the sensors. The reasons for this interest include the potential for improved system performance and enhanced system capabilities. Although interest is high, the complex issues involved in designing and analyzing multisensor systems may stifle the development of combat systems that fully utilize the capabilities they offer. If these complex problems can be solved, multisensor integration and data fusion will find application in many different areas in the surface Navy. Some areas currently being considered are target tracking, kill assessment, target identification, surveillance, detection, command and control, and navigation. General references for various aspects of multisensor integration and data fusion include Luo and Kay,¹ Bar-Shalom,^{2,3} Blackman,⁴ and Waltz and Llinas.⁵ The work described in this article has been supported in part by: the Dahlgren Division's AEGIS Program Office; the Naval Sea Systems Command (06DB), as part of the Division's effort as the Navy's Shipboard Gridlock/Correlation Agent; the Naval Air Systems Command-sponsored Cruise Missile Project; and the Surface-Launched Weaponry Block under the Office of Naval Technology.

Potential Advantages

A sensor is a device that provides information about features of interest in its environment. In a multisensor system, some of the sensors may be providing information about the same features in the environment while other sensors provide information about different features. In general, the advantages gained by the use of multisensor integration and data fusion can be decomposed into four fundamental aspects: redundancy, complementarity, timeliness, and cost of the information.¹

Redundant information refers to that provided by the various sensors about the same features in the environment. (Redundant information can also be provided by a single sensor acquiring information about a particular feature over a period of time.) As an example, consider two radars that are integrated into a single system to provide target tracking. Each radar provides an estimate of the position of a particular target. The feature of interest is the position of the target, and the radars are providing redundant information about that position.

Fusing redundant information will increase the accuracy (or, equivalently, reduce the uncertainty) of the information about the features observed by the sensors. This is true, in general, for most fusion schemes. An integrated system of sensors providing redundant information will have increased reliability. For a stand-alone sensor, destruction or failure of the sensor would eliminate the information coming from that sensor and could result in the loss of weapon system capability. However, in an integrated system providing redundant information, some of the remaining sensors might be able to provide information about the features of interest, and operations may proceed.

Complementary information refers to that provided by the various sensors in a multisensor system about different features in the environment. As an example, consider a radar and a thermal imaging device (TID) such as an infrared (IR) sensor that are integrated into a single system to provide short-range target search, track, and identification. The radar is used for target search and track, and the TID uses the temperature gradient of an object to produce an image that can be used for target recognition. The sensors are providing information about different features; the radar is providing kinematic information, but the TID is providing image information. Thus, the radar and the TID are providing complementary information about the environment.

Complementarity allows information to be obtained that is impossible to obtain using sensors operating in a stand-alone mode. In this example, the radar is not providing any image information for target identification, and the TID

is not providing full information about the kinematic parameters. (A TID is a passive sensor that may provide information about the azimuth and elevation of a target, but does not provide any range information.) By integrating the radar and the TID, both kinematic information (for target location) and image information (for target identification) are obtained.

A multisensor system may produce *more timely information* when compared to sensors operating in a stand-alone mode. In the previous example, the radar operating in the stand-alone mode is not providing any target recognition information. If the radar is tracking an unknown target, then identification of the target as friendly, hostile, or neutral may not occur until the target is too close to engage successfully (if it is hostile). By integrating the TID (which can operate at night as well as during the day) with the radar, it may be possible to make the identification much more quickly. Also, it may be possible to obtain more timely information if it is possible to integrate the sensors in a parallel fashion. This parallel processing will produce an increase in the speed of operation of the system.

As the demand for sensors with increased capabilities grows, the cost of design and construction of stand-alone sensors that satisfy system performance requirements will increase drastically. However, the integration of multiple sensors with nominal performance may be able to satisfy system-level performance requirements at a lower cost. That is, a multisensor system may produce *less costly information* when compared to the cost of constructing a stand-alone sensor providing the same information.

Stand-alone sensors are limited in the space they can search and the accuracy of the data they obtain. Due to redundancy and complementarity, integrating individual sensors into a multisensor system can produce better quality coverage by increasing the search space, eliminating any coverage gaps that may exist for individual sensors, and producing more accurate information. Similarly, an integrated system may be less vulnerable to countermeasures. The loss of information coming from a sensor that is being countered could be minimized by information from sensors in the system that are not affected by the countermeasures.

Development of Technologies

Many of the required technology areas for multisensor integration and data fusion are not fully developed, and this may hamper progress in designing multisensor systems. For example, integrating multiple sensors into a single system will require large amounts of data to be communicated among the various sensors in the system,

and it will also produce large amounts of data that must be processed. Advanced communication systems, computer architectures, and signal processing algorithms will be required to deal with this problem. Also, sophisticated algorithms will be required to control and schedule the various sensors in a multisensor system. The problem of integrating and fusing data from different types of sensors must be solved (e.g., active and passive sensors; electro-optical sensors and radars). Accurate alignment of the sensors will also be required. Failure to accurately align the sensors usually leads to overall system performance that is worse than the performance of the individual sensors in the system. Thus, complex system designs will be required to deal with the overall operation of a multisensor system. Differences in the security level of the information provided by the various sensors can also complicate the design of multisensor systems. These issues must be considered in the design of any multisensor system.

Definitions

Several terms occur frequently in any discussion of the multisensor integration and data fusion process as illustrated in Figure 1. Unfortunately, these terms are often used ambiguously and lead to confusion. This section defines precisely some of the more commonly used terms. These definitions are due, in large part, to Luo and Kay.¹

Multisensor integration refers to the synergistic use of multiple sensors to assist in the accomplishment of a task by a system. As an example, one may desire to search a large volume of airspace for targets and accurately track any detected targets (i.e., surveillance and tracking). This task is sometimes accomplished by using both a surveillance radar and a tracking radar. The surveillance radar can search a large volume of airspace in a short amount of time and detect tar-

gets, but it cannot track targets very accurately. Tracking radars can confirm targets (detected by the surveillance radar) and provide accurate estimates of the positions of the targets, but cannot search a large volume of space in a short amount of time. By integrating these two radars, the surveillance and tracking tasks could be accomplished that each radar could not accomplish effectively if operating in a stand-alone mode.

Integrating sensors into a single system requires that all the multisensor data be expressed in a common reference frame and be free from errors in the transformation process. Errors in the transformation process can be caused by sensor location errors, sensor offset errors, etc. **Registration** refers to the process of transforming all the multisensor data to a common reference frame such that the data is free from errors in the transformation process. **Data association** (also called data correlation) refers to the decision process for allocating sensor information with existing data sets or establishing (initializing) new data sets. Often, the data sets are tracks and/or contacts (measurements) reported by the sensors. For this reason, association is usually divided into three categories: track-to-track, contact-to-track, and contact-to-contact.

Data fusion refers to the process of combining sensor information (data) into one representational format. Data fusion can and usually does exist in stand-alone sensors. For example, a Kalman filter can be thought of as combining data that is acquired over an extended period of time from a stand-alone sensor into a single estimate of the state (the representational format). **Multisensor data fusion** refers to any stage of the multisensor integration process in which information (data) from multiple sensors is combined into one representational format. As an example, consider two tracking radars that are integrated into a single system. Each radar provides an estimate of the state vector of a particular track. The multisensor data-fusion process would combine

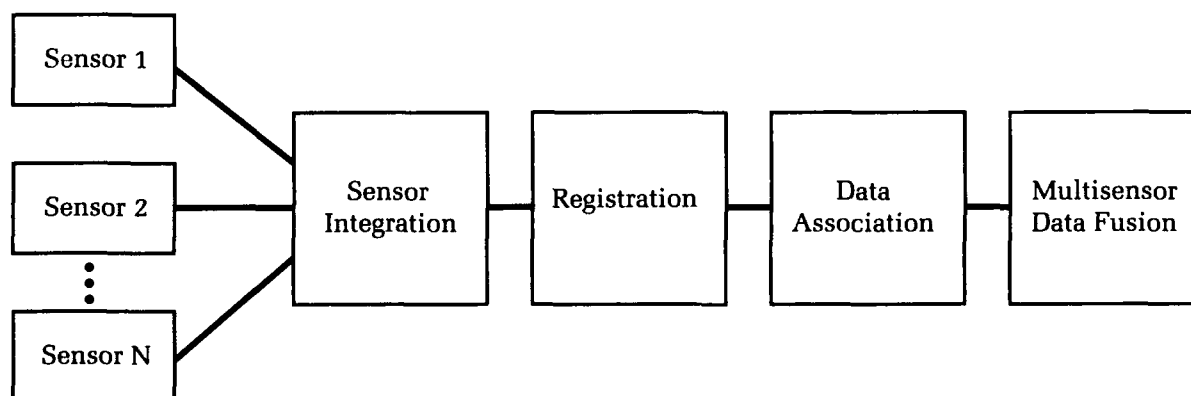


Figure 1. Illustration of the multisensor integration and data fusion process.

the separate estimates of a particular track into a single estimate of the state vector (the representational format).

The distinction between multisensor integration and data fusion is made so that the higher-level issues associated with multisensor integration can be separated from the more basic mathematical, statistical, and algorithmic issues involved with actually combining the data. Multisensor integration deals with such issues as the choice of sensors to accomplish a given task, type of system architecture to be employed (i.e., central versus distributed processors), and communication issues. Multisensor data fusion deals with such issues as obtaining mathematical models of the various sensors in the system, and the choice or design of algorithms to actually combine the sensor data. Of course, these two functions are not independent; for example, the choice of processors in the multisensor integration function would be influenced by the data fusion algorithms, and vice versa.

In the actual operation of a multisensor system, the various functions defined above will not be independent. Consider two tracking radars that are integrated into a single system. Before the separate track estimates for a particular target reported by the two radars can be fused into a single estimate, the tracks reported by each radar corresponding to this particular target must already be correlated by the data-association function. Thus, the data-fusion function depends on the data-association function to provide it with associated track pairs. Furthermore, both the data-association and data-fusion functions depend on the registration function. That is, to associate and fuse the tracks, the radars must be registered so that the track data is expressed in a common reference frame and the data is free from errors in the transformation process.

Sensor Registration

Integrating sensors into a single system requires all the sensor data to be expressed in a common reference frame. Thus, the data from each sensor must be transformed to a common reference frame. Errors in this transformation such as sensor offset errors and sensor location errors are called registration or alignment errors, and will be discussed below. As an example, consider a multisensor system consisting of two tracking radars. Track positional information from the two radars is transformed to a common reference frame. Alignment errors in each radar will introduce errors into the track positions reported by each radar. For a particular target, the track positions reported by each radar will not be the same because of these errors. The difference in position will depend on the magni-

tude of the alignment errors. If it is not known that these track reports correspond to the same target, then the data-association function may erroneously decide that they are different targets. (In this context, association refers to the process of deciding whether tracks reported by each radar correspond to the same target or different targets, i.e., track-to-track association. Association decisions are based on many track parameters, and one of these parameters is usually the difference in positions of the tracks.) Thus, the presence of alignment errors may result in the holding of separate tracks for the same target. In general, the presence of alignment errors leads to degraded system performance, the amount of degradation depending on the magnitude of the alignment errors.

To deal effectively with the alignment problem, the sensors must be registered. The registration or alignment process requires the transformation of multisensor data to a common reference frame, the estimation of the values of the alignment errors, and the use of these estimates to correct the multisensor data. The advantages of integrating sensors into a single system and fusing their data can be realized only if the sensors are registered.

The major sources of registration errors include calibration errors (i.e., offsets) in the sensors, sensor tilt, and errors in the locations of the sensors. Sensors located on ships usually undergo an initial calibration procedure. However, the calibration may deteriorate over time, or it may change if there is some change in the state of the ship (e.g., change in the amount of loading on a ship). One of the major sources of tilt errors is flexure in the platform on which the sensor is located.⁶ In shipboard applications, the action of waves on a ship and the motion of a ship through water can cause the ship's structure to flex. Flexure will cause the sensor's reference frame to tilt with respect to the ship's stabilized frame. The locations of the sensors must be known for transformation of the sensor data to a common reference frame. Errors in the locations of the sensors with respect to the origin of the common reference frame will introduce errors in the track data when it is transformed to the common reference frame.

Data Association

Data association is the decision process for correlating sensor information with existing data sets or establishing (initializing) new data sets. A data set usually contains information about a particular target or event. Thus, the data-association process will correlate sensor information to a particular target or event, or establish a data file on a new target or event. For example, a track is a state

trajectory that is estimated from a set of measurements that have been correlated to the same target.² In this case, the association process must correlate measurements to other measurements (contact-to-contact association) in the formation of new tracks, and it must also correlate measurements with existing tracks (contact-to-track association) in the maintenance and update of existing tracks. Thus, data association can exist in both stand-alone sensors and multisensor systems. In a stand-alone sensor, all the measurements and tracks are provided by a single sensor (e.g., a radar), but in a multisensor system, the measurements and/or tracks may be from multiple sensors. Of course, the multisensor case gives rise to new association problems; for example, the correlation of existing tracks reported by the various sensors (track-to-track association).

Gating is the first part of most data-association algorithms. Gating is a coarse validation test that eliminates measurements that are unlikely to correlate with a given target. For example, in target-tracking applications, a gate is constructed around a target's predicted position, and only measurements that fall within this gate are considered as candidates for correlation with the target. The gates are usually constructed based on the statistics describing the accuracy of the sensor's measurements and the accuracy of the target's predicted position. The statistical distance between a measurement and the target's predicted position is calculated, and the probability that this measurement correlates with the target is calculated. If this probability is greater than some threshold, the measurement is said to satisfy the gating criterion and is a candidate for correlation with the target. The simplest gating technique is to construct a rectangular region around the target's estimated position, and only consider measurements that fall within this region. The rectangular boundary of this region can be related to the probability threshold mentioned above. Any measurement falling within the gate will have a probability of being correlated with the target that is greater than the threshold. Another popular gating technique uses an ellipsoidal gate, where an ellipsoidal region is constructed around the target's estimated position. The ellipsoidal gate is more precise than the rectangular gate because it accounts for the statistical correlation between the components of the estimated position vector, which the rectangular gate ignores. More information on gating techniques can be found in Blackman⁴ and Bar-Shalom.²

If a single measurement falls within the gate, and it does not fall in any other target's gate, then this measurement is correlated with the target. It can then be used to update the target's position. If the measurement does not fall within

any target's gate, then this measurement becomes a candidate for the formation of a new target track. The difficulty occurs when more than one measurement falls within the gate, or a measurement falls within the gate of more than one target. This is an example of a measurement of uncertain origin. In such cases, further correlation logic is required. For the multitarget and/or multisensor problem, the major challenge is to perform data association with measurements of uncertain origin, which is further complicated by detection probabilities that are less than unity. The uncertainty in the origin of the measurements can be caused by: high density of targets or closely spaced targets; high false-alarm rates in the sensors; clutter; multipath; and countermeasures.

The data association process must relate information provided by a sensor to a number of possible data sets, each representing a hypothesis to explain the source of the information (e.g., the information is related to a particular target). If there are n existing data sets, association is the decision process that determines which of the $n + 1$ possible hypotheses best describes the source of the information. The additional hypothesis exists because the information may not be related to any of the n existing data sets. Data association methods are based on two fundamentally different models. The first model is the deterministic model, where it is assumed that the association decision is correct, and it does not account for the possibility that the decision may be incorrect. The other model is the probabilistic model that uses a Bayesian approach to account for uncertainty in the association process.

In the deterministic model, statistical methods such as classical hypothesis testing or maximum likelihood estimation are used to rank all the possible hypotheses using a statistical distance metric. The association process chooses the hypothesis with the smallest statistical distance (i.e., nearest-neighbor). This is a *hard decision* that chooses a single hypothesis, and it ignores the fact that it may not be correct. This approach can lead to severe degradation of system performance because it does not account for ambiguities when several closely ranked choices are present. These ambiguities can occur because of high density of targets, closely spaced targets, high false-alarm rate, clutter, and countermeasures (i.e., measurements of uncertain origin).

In the probabilistic model, all the hypotheses are assigned probabilities of being correct. This is an all-neighbors approach because each hypothesis has a probability of being correct, and no hard decision is made. Two methods are generally used to handle the all-neighbors problem: multiple-hypothesis tracking (MHT) or probabilistic data association (PDA).

In the multiple-hypothesis approach, all the hypotheses are retained until some can be eliminated or confirmed because of their relative probabilities. That is, decisions are deferred until additional information is gathered so that it can be used to eliminate (prune), combine or confirm existing hypotheses, or initiate new hypotheses. A drawback to this approach is the possibility of maintaining a large number of hypotheses. In the PDA approach, each candidate measurement falling within the gate is used to update the target's state estimate, where each measurement is weighted by a factor related to the probability that the measurement is correlated with the target. In this approach, all candidate hypotheses contribute to the target's state estimate, and deferred decision making is not required.

In general, data association with measurements of uncertain origin is a very difficult problem. Measurements of uncertain origin often arise in systems containing IR sensors because of the high false-alarm rate usually produced by IR sensors. This problem also occurs in radar tracking systems when there are closely spaced targets, crossing targets, splitting targets, a high target density environment, clutter, or multipath. Another difficult problem in data association is the correlation of data from passive sensors with data from active and/or other passive sensors. While the theoretical development of algorithms such as MHT and PDA has been significant, the application of these algorithms in actual systems has been limited.

Data Fusion

Data fusion refers to the process of combining sensor data (from one or more sensors) into one representational format. When selecting a data fusion method for a specific application, the type and format of the sensory information and the dynamics of the underlying process must be considered. In the single-sensor case, the data is of the same format, but in the multisensor case, the data may be of different formats. For the single-sensor case, the data is fused over time, but in the multisensor case, the data can be fused synchronously or over time. Since the fusion of redundant information is concerned with producing data that is more accurate than the data from the individual sensors, most fusion methods utilize the variances or probability distributions to characterize the redundant information. Other methods, such as logical inference, can also incorporate complementary information in the fusion process. Bayesian inference is an example of a logical inference method. Also, the dynamics of the underlying process must be considered in the fusion process. If the process is time-varying (i.e., dynamic), the weighted

average and Kalman filter methods are appropriate, while the logical-inference methods are usually restricted to static processes.

Below, some of the major types of fusion methods will be discussed. In particular, the weighted average, the Kalman filter, and the Bayesian inference method for multisensor data fusion will be discussed. The weighted average and the Kalman filter are used to fuse redundant information, where the information is either raw sensor measurements or described by probability distributions, or both. The Bayesian inference method can usually handle both redundant and complementary information, and the information is usually in the form of propositions.

Weighted Average

The weighted average method is the most intuitive or direct technique for data fusion. The weighted average method includes both the direct method, where normalized weights are used for summing the individual observations, and the least-squares method. In the direct method, parameters representing the accuracy or reliability of each of the data are utilized to compute weights for fusing the data into one representational format. For example, if N observations of a parameter vector X are available for fusion, the fused vector $\hat{X}$ is given by

$$\hat{X} = \frac{1}{C} \sum_{i=1}^N w_i X_i, \quad (1)$$

where

$$C = \sum_{i=1}^N w_i, \quad (2)$$

and X_i is the i^{th} observation of the parameter vector. The weights w_i are based on the accuracy and/or reliability of the sensor providing the i^{th} observation.

The least-squares method for data fusion is based on the minimization of the sum of the squared errors between the measurements and the projection of a parameter vector which describes the underlying process. The information within the measurements is fused into a common parameter vector through the fusion. Consider measurements given by

$$y_i = h_i X + v_i, \quad (3)$$

where y_i is the i^{th} measurement, v_i is the measurement error, and h_i relates the parameter vector X to the measurement y_i . The weighted least-squares estimate of X is the vector that minimizes the cost function

$$J(X) = \sum_{i=1}^N (y_i - h_i X)^2 r_i, \quad (4)$$

where r_i is the weight for measurement y_i . By forming the individual measurements into a vector, Equation (3) can be written as

$$Y = HX + V, \quad (5)$$

where

$$Y = [y_1 \ y_2 \ \dots \ y_N]^T, \quad (6)$$

$$H = [h_1 \ h_2 \ \dots \ h_N]^T, \quad (7)$$

$$V = [v_1 \ v_2 \ \dots \ v_N]^T. \quad (8)$$

Then the sum of squared errors in Equation (4) can be written as

$$J(X) = (Y - HX)^T R(Y - HX), \quad (9)$$

where

$$R = \text{diag}(r_1, r_2, \dots, r_N). \quad (10)$$

The weighted least-squares estimate of X is given by⁷

$$\hat{X}^{LS} = (H^T R H)^{-1} H^T R Y. \quad (11)$$

If $E[V] = 0$ and $R^{-1} = E[V^T V]$, the Gauss-Markov Theorem states that the weighted least-squares estimate is also the minimum variance estimate of X .⁸ The least-squares techniques are also used for data fusion when the measurements are a nonlinear function of the parameter vector X . The nonlinear measurement equation is first linearized, and then an iterative procedure is employed to compute the least-squares estimate.

Kalman Filter

When an underlying process is time-varying and stochastic, a Kalman filter is often utilized to fuse the data or measurements into a common representational format. Models are developed for the evolution of the underlying process with time and the measurements of that process. These models are often denoted by

$$X_{k+1} = F_k X_k + G_k w_k, \quad (12)$$

$$Z_k = H_k X_k + v_k, \quad (13)$$

where X_k is the state of the underlying process at time k , $w_k \sim N(0, Q_k)$ is the process noise, Z_k is the measurements of the process, and $v_k \sim N(0, R_k)$ is the measurement error. The matrix F_k defines the evolution of the process between time k and $k+1$, with w_k representing the possible error in F_k . In target tracking, the state process X_k may contain the position, velocity, and/or acceleration of the target, while the measurements include the position and possibly radial velocity of the target.

A Kalman filter is often employed to estimate the state of the process. The Kalman filtering update of the state estimate with a new measurement can be viewed as a weighted least-square estimation problem.⁹ For the system in Equations (12) and (13), the objective function to be minimized in the least-squares sense with respect to X_k is

$$J(X_k) = \frac{1}{2} \left[(X_k - X_{k|k-1})^T P_{k|k-1}^{-1} (X_k - X_{k|k-1}) + (Z_k - H_k X_k)^T R_k^{-1} (Z_k - H_k X_k) \right], \quad (14)$$

where $X_{k|k-1}$ is the predicted state estimate for time k based on measurements through time $k-1$, and $P_{k|k-1}$ is the covariance of the error $X_k - X_{k|k-1}$. The estimate of X_k that minimizes $J(X_k)$ is denoted as $X_{k|k}$.

The Kalman filtering equations associated with the state model of Equation (12) and the measurement model of Equation (13) are given by the following equations.

Time Update:

$$X_{k+1|k} = F_k X_{k|k}, \quad (15)$$

$$P_{k+1|k} = F_k P_{k|k} F_k^T + G_k Q_k G_k^T. \quad (16)$$

Measurement Update:

$$X_{k|k} = X_{k|k-1} + K_k [Z_k - H_k X_{k|k-1}], \quad (17)$$

$$P_{k|k} = [I - K_k H_k] P_{k|k-1}, \quad (18)$$

$$K_k = P_{k|k-1} H_k^T [H_k P_{k|k-1} H_k^T + R_k]^{-1}, \quad (19)$$

where $X_k \sim N(X_{k|k}, P_{k|k})$, with $X_{k|k}$ and $P_{k|k}$ denoting the mean and error covariance of the state estimate, respectively. The subscript notation $(k|j)$ denotes the state estimate for time k

when given measurements through time j , and K_k denotes the Kalman gain at time k .

When the state process and measurement models are linear and the process and measurement errors are Gaussian, the Kalman filter provides the minimum mean-square error estimate of the state. When the state process and measurement models are linear, but the error processes are not Gaussian, the Kalman filter is the best linear estimator in the minimum mean-square error sense. If the state process model and/or measurement model are nonlinear with respect to X_k , the models can be linearized about the most recent state estimate, and the Kalman filtering equations implemented with the linearized models. A filter using linearized models is referred to as the extended Kalman filter. In target tracking, the state of the target is often modeled in Cartesian coordinates, while the measurements of the target state are in spherical coordinates. Since these spherical coordinates are nonlinear functions of the target state, an extended Kalman filter is implemented for estimating the target state.

Bayesian Inference

States and measurements can contain information that is not kinematic. This other type of information is usually of a symbolic nature, and it may be presented in the form of a proposition. This information is usually attribute information, where an attribute is a characteristic that distinguishes between a type or class of target. For example, it may include such things as the type of emitting radar on a platform, Identification Friend or Foe response, type of hull for a ship, type of engine on an aircraft, the shape of a target's image; and the output of a sensor may be the proposition, "the target has radar type R1." Multisensor systems can provide a variety of attribute information from different sources. The problem is to fuse the attribute information from the various sensors; the ultimate goal of the fusion process is to infer the target's identity (i.e., usually target type or class).

The two techniques that are considered most often in the fusion of attribute data from multiple sensors are Bayesian inference and Dempster-Shafer evidential reasoning. Bayesian inference, based on Bayes' rule, requires *a priori* information and conditional probabilities. Let $\{x_1, \dots, x_N\}$ be a set of attribute measurements of a target, where the set of possible target types is given by $\{A_1, \dots, A_M\}$. To employ Bayesian inference in determining the target type, the following conditional probabilities must be specified:

$P(x_j | A_k)$ = probability of receiving the target's attribute measurement x_j given that the target type is A_k ,

and the prior probabilities $P(A_k)$ of each target type must also be specified. The probability that actually needs to be determined is $P(A_k | x_j)$, where

$P(A_k | x_j)$ = probability that the target is type A_k given that the attribute measurement x_j was received.

This can be computed by Bayes' rule:

$$P(A_k | x_j) = \frac{P(x_j | A_k) P(A_k)}{P(x_j)}, \quad (20)$$

where

$$P(x_j) = \sum_k P(x_j | A_k) P(A_k). \quad (21)$$

One of the attractive features of the Bayesian approach is that it can be implemented recursively to update the probabilities when new measurements are received. This recursive process can be explained as follows. A measurement x_1 is received, and Bayes' rule is used to calculate the probabilities $P(A_k | x_1)$ for $k = 1 \dots, M$. Then, $P(A_k | x_1)$ becomes the prior probability $P(A_k)$ when a new measurement x_2 is received, and Bayes' rule is then used to update the conditional probability. This process continues as new measurements are received.

Dempster-Shafer evidential reasoning is a generalization of Bayesian inference, and it was developed to overcome some of the perceived weaknesses in the Bayesian approach. For example, the Bayesian approach must employ the principle of indifference when there is no *a priori* information. The principle of indifference states that in the absence of prior information, assume that all of the prior probabilities are equally likely. Dempster-Shafer avoids this problem by allowing the assignment of a probability mass to uncertainty. Also, Dempster-Shafer can handle the problem of inconsistent information. The interested reader is referred to Blackman,⁴ where the Dempster-Shafer method is discussed in more detail. Blackman also discusses how Bayesian inference and Dempster-Shafer can be extended to include both attribute and kinematic information.

Applications

Some applications of multisensor integration and data fusion to Navy systems will be presented below. These applications include: the Shipboard Gridlock System, which provides registration and track association for radars located on different platforms and reporting over Link-11; target tracking using multiple sensors; over-the-horizon data fusion; and multisensor kill assessment and target identification.

Shipboard Gridlock System with Auto-Correlation¹⁰

The Shipboard Gridlock System with Auto-Correlation (SGS/AC) is currently used to provide a coordinated air and surface track picture for Link-11 Participating Units (PUs). One of the major goals of SGS/AC is to eliminate multiple-track reports on the same target, which can lead to data overload of Link-11 and the possibility of multiple engagements on the same target (and the consequent depletion of resources). Registration errors will introduce errors in the track data reported by the different PUs. Thus, for a particular target, the track positions reported by the various PUs may be considerably different, and it may be decided erroneously that the tracks are associated with different targets. The Shipboard Gridlock System must provide the following capabilities for each PU:

- **Registration:** The ability to transform the coordinates of all track data reported over Link-11, estimate the registration errors in the track data, and use these estimates to correct the reported track data.
- **Track Auto-Correlation:** The ability to identify common targets in the tracks held by a particular PU with tracks reported over Link-11 by other PUs. The track auto-correlation function is an example of track-to-track data association.

The registration problem is solved in SGS/AC by aligning the radars on the different PUs to a common radar (called the Gridlock Reference Unit, or GRU), where each radar holds mutual tracks with the GRU. Each PU estimates the parameters that will align it to the GRU. These alignment parameters are estimated using mutual tracks held by the PU and the GRU. In this way, each PU aligns its track picture to the GRU track picture, and the overall track picture becomes aligned to the GRU track picture. The registration errors modeled in SGS/AC are bias errors in the locations of the platforms on which the radars are located and bearing offset errors in the radars.

The track auto-correlation function uses statistical hypothesis testing, which is a deterministic model of data association, to identify any tracks held by a PU that are already reported over Link-11. If a PU identifies a track that it holds and that is also broadcast over Link-11 by another unit, then both units will not report the track over Link-11. The decisions in the track auto-correlation function are based upon five quantities: the differences in speed, heading, altitude, position along the x-axis (east), and position along the y-axis (north) between a track held by the PU and a track reported over Link-11. If these differences are large, it is less likely that

these are tracks of the same target. Conversely, if these differences are small, it is more likely that these are tracks of the same target. The actual decision to correlate a track held by the PU and a track reported over Link-11 is obtained by thresholding a likelihood function, which depends on the five quantities mentioned above. Also, the likelihood function depends on whether a pair of tracks have been previously correlated and, if so, the amount of time they have been correlated.

Target Tracking

The problem of target tracking includes track initiation, track maintenance, track-to-track association, and track-to-track fusion, where a track is the target state estimate that characterizes the trajectory at a given point in time. The track initiation or formation is the association of several detections over time and a decision that accepts the detections as having originated from a common target. When the false alarm rate is high or the probability of detection is low, the problem of track formation can be very difficult. The problem has been addressed with M out of N detection logic algorithms, probabilistic data association filters (PDAF), and MHT algorithms.³ Track maintenance is the association of measurements to existing tracks and the updating of the state estimates with the measurements. PDAF and MHT algorithms are commonly used for track maintenance for slowly maneuvering targets, while the interacting multiple model PDAF is used for highly maneuvering targets.³ Track-to-track association involves hypothesis testing for identifying tracks that correspond to a given target, while track-to-track fusion involves the fusion of tracks for a given target into a single track. Fusing the tracks reduces the computer memory required to support the targets, provides a coherent picture to the operator, and provides target-state estimates of higher quality.

The target tracking problem can be generalized to include multiple targets and/or multiple sensors. For almost all target-tracking problems in the surface Navy, one must consider the possibility of multiple targets because the simplest case of a fire-control radar tracking a single target can include a target that splits into two threats; for example, one threat and a delivery vehicle; or a threat and a decoy.¹¹ If a target splits and the case of multiple targets has not been considered, then a decoy or delivery vehicle might be engaged, leaving the threat to hit the ship. When multiple targets are present in the surveillance region, the track maintenance problem of associating the corresponding measurements and targets must be addressed. This problem of measurement-to-target association is

further complicated by false alarms and detection probabilities less than one because the actual number of targets is unknown. Two well recognized approaches to this multitarget tracking problem include the MHT algorithm⁴ and the joint probabilistic data association filter.²

The introduction of multiple sensors into the target-tracking problem is often considered to improve the data association, improve the accuracy of the target state estimates, broaden the region of coverage, and enable passive tracking. Multiple sensors can also assist in the solution of problems associated with false alarms, multipath, clutter, and jamming. In a multisensor tracking system, the association and fusion of common tracks are critical to providing to the operator a unified track picture for surveillance and weapons control, reducing the computer time and memory required to support target tracking, and providing the most accurate target-state estimates.

The PDAF or MHT algorithms are often employed for associating and processing the measurements for estimating the target state. Two standard approaches for implementing these algorithms are centralized and distributed. The centralized approach involves sending all data to a central site or computer where a common state estimate is maintained for each target. While the centralized approach provides the better solution to the target-tracking problem, processing data from multiple sensors in a central computer can significantly complicate the issues of track initiation and track maintenance. The distributed approach involves track initiation and maintenance at each sensor, with the results for each target being sent to a central site. Then in the central computer, track-to-track association and fusion are completed. While the centralized approach most often provides better solutions to track initiation and maintenance problems, the distributed approach is often implemented in practice because it requires less communication between sensor sites, a smaller central computer, and enhances survivability of the sensor system.

Over-the-Horizon Data Fusion

The multisensor data-processing problem for a naval task force (or surface action group) can be characterized in a number of different ways. One procedure is to break the tactical picture into three regions:

- **Within-the-Horizon:** In this region, the majority of the data on a contact comes from naval task force assets, and the data generated is of sufficient quality to be used to maintain the tactical picture.
- **Over-the-Horizon:** In this region, the majority of the data on a contact comes from sensors not controlled by the naval task force (i.e., over-the-horizon sensors). The information obtained from these sensors is distributed over data links and is used to maintain the tactical picture.
- **Transition Zone:** In this region, data on a contact comes from both naval task force assets and over-the-horizon sensors. In the transition zone, it is necessary to correlate (i.e., associate) contacts/tracks established using data from over-the-horizon sensors with contacts/tracks established using data from naval task force assets.

There is not a sharp separation between these regions, and the boundaries of these regions will fluctuate not only with time, but also with warfare area and emissions control condition. A brief discussion of the maintenance of the tactical picture in the over-the-horizon region is presented below.

Establishing and maintaining the over-the-horizon tactical picture has various meanings depending on what is desired from the over-the-horizon sensors. To perform the naval task-force role of sensor management and to execute long-range warfare, an effective real time over-the-horizon tactical picture is required that provides the best possible indication of the position and movement of all the platforms in the operational area.

In general, the over-the-horizon sensors are not controlled by a naval task force and cannot be tasked by the naval task force. Also, as stated above, the information obtained from over-the-horizon sensors is distributed over data links. Consequently,

- The data is time tagged but it is late (anywhere from a few minutes to several hours).
- The data received is out of time sequence.
- There may be repeated transmissions of a single sensor report (a repeated transmission may be identical to an earlier transmission; it may contain modified parameter values; it may contain additional data).
- The data reports on contacts will contain different types of data, ranging from unique ship identifiers to location/motion data (e.g., position, bearing, velocity), to attribute data (e.g., electronic intelligence, acoustic).
- The original source and prior processing of the data are, in general, unknown.
- Position and position uncertainty data are, in general, supplied, but attribute uncertainty data is usually not provided (i.e., the accuracy of the attribute data is not provided).
- There is inconsistency in data presentation (e.g., variation due to operator entry, etc.).

For the contacts detected by over-the-horizon sensors, a position estimate, expressed in terms of latitude and longitude, and position uncertainty data are provided. The position uncertainty data is represented by a bivariate normal distribution in a plane that is tangent to a spherical Earth at the position estimate. This representation of position uncertainty permits relatively easy mathematical processing and the visualization/characterization of position uncertainty as an ellipse of uncertainty.

An over-the-horizon data fusion capability can be developed using a number of approaches. Some of the principal design options are:

- Of the parameters frequently included in the over-the-horizon sensor reports, which should be used in the fusion process and what should their relative weighting be?
- How dependent should the fusion process capability be on prior knowledge (i.e., how much table look-up and specialized training should there be versus computation "on the fly," using only current data and common sense)?
- What should the data fusion decision criteria be, and how should correlation feasibility testing and correlation scoring be performed?
- Should the fusion process be a single-hypothesis technique or a multiple-hypothesis technique, and if a multiple hypothesis technique is used, how should the results of the fusion process be presented to the operator?

Application of Multisensors to Kill Assessment and Target Identification

An integral part of any combat system is a rapid and accurate kill assessment (KA) capability. There are two reasons why accuracy is important. First, unless the KA is accurate, the system will frequently indicate a no-kill when, in fact, the target is no longer a threat. This results in a re-engagement that wastes the system's resources (missiles, illumination time, and fire control channels). Second, and even more serious, an error results when the system declares the target to have been destroyed when it was not. This results in a leak. A more rapid KA is needed since more targets will be considered in a given amount of time, which is important when a combat system is experiencing a stressing target load.

Many techniques that can be applied to target identification (ID) can also be used in KA. Usually, target ID is accomplished by extracting

a pattern (or "signature") from a signal that has interacted with the target. Characteristics of this pattern are then used to identify the target. KA is done in much the same way, only one does not use the characteristics as absolutes. When doing KA, one obtains patterns before and after a missile intercept, and then looks for differences in the two patterns to determine if the target has undergone a kill.

Range and environmental conditions are a factor in the choice of sensors that can be used in KA; thus, not all sensors in the sensor suite will be able to perform KA for a given target at a given time. However, there may be several sensors that can be utilized for intercepts at various ranges and environmental conditions. The constituent parts of a sensor suite used to perform multisensor KA and their capabilities will be considered below.

Doppler Radar. Over long ranges, a high-power radar that is either continuous wave (CW) or pulse doppler (having a mode that gives a PRF of approximately 200 kHz) can be used to perform KA. Work under the Surface Launched Weaponry block has resulted in two techniques, the mean logarithm of the Fourier Transform and the cepstral process, which can detect most catastrophic kills and certain non-catastrophic kills that result from the use of a hard-kill system. Assessment of these kills is significantly faster than methods currently employed. A third technique, the velocity tracker, has been developed and can be used to detect any kill that results in a significant deceleration of an intact target.

If countermeasures such as chaff or a decoy (soft-kill weapons) are deployed, then a CW or high pulse-repetition frequency can be used to track a target in the frequency domain. Using high-resolution fast-Fourier transforms, it is possible to determine the radial velocity of the target very accurately. If a persistent change in the target's radial velocity is detected, then this can be used in conjunction with accurate angular measurements (possibly from an IR sensor) to infer whether or not the countermeasures were effective.

Broadband Radar. A broadband (or a short pulse) radar can be used to perform KA by obtaining "signatures" of the target before and after a missile intercept. Differences in these signatures will indicate whether or not the target has been neutralized. This radar can be used for target identification by comparing before-intercept signatures with known signatures.

Precision Electronic Surveillance Measures (PESM). If the target is emitting a signal, such as a missile with an active seeker, then over short to intermediate ranges, one can do KA (and target ID) using a PESM sensor.

The KA function may be performed in the following manner. First, the sensor accurately determines the angular position (elevation and bearing) of the target. Then, the angular measurements, in conjunction with a change in the radial velocity (obtained from the CW or pulsed doppler radar) can be used to detect a non-catastrophic kill by detecting changes in the velocity vector. Additionally, the cessation of the target's transmitter at the time of the missile/target intercept can be used as a piece of information to be integrated with other information to do KA.

The target ID function is performed by observing characteristics of signals emanating from the target. Some of these characteristics are: (1) frequency of the carrier signal; (2) PRF; and (3) width of the pulse. Hence, matching the measured characteristics with those of known threats can identify a target or a target class. The problem here is associating the electronic surveillance measures signal with a particular target.

Electro-Optical/Infrared (EO/IR Sensors). For reasonably short ranges, it is possible to do KA using a staring IR sensor. Since an IR sensor can supply very accurate bearing and elevation data, the angular data from the IR sensor can be fused with the PESH and radar data to obtain a very accurate track of the target, which can be used to assess non-catastrophic kills. If the IR sensor is capable of providing video (probably used in conjunction with a telescopic sight), then various techniques can be employed to do automatic KA of missile intercepts by detecting changes in the target's profile. Both of these modes (catastrophic and non-catastrophic kills) can be employed if there is an optical sight instead of an IR sensor.

Laser Radar. Finally, there is the possibility of using a laser radar in the short-range arena. A laser radar will provide the most accurate range, elevation, and bearing readings of any available sensor. This data can be fused with the data from the other sensors and used to produce a very accurate track of the target under consideration. Additionally, the laser radar can provide very accurate doppler data when used in the proper mode. Lastly, the laser radar data can supply information about the vibrational modes of an airframe, which comes from further processing the signal in the frequency domain. Being able to determine vibrational modes of the target is a possible way of doing target ID.

Among all the sensors, the CW or pulsed doppler radar has the potential for covering targets at the greatest range as well as being able to operate, even if under reduced capability, in most environmental conditions. Consequently, this should be the primary sensor used for performing KA, with the other sensors considered secondary—to be used when the right conditions

exist. For example, a pulse doppler radar and a PESH sensor can be used at intermediate ranges to observe an intercept between a weapon and a target when the missile is using its radar to search for and lock on a target. However, if the missile turns off its radar once it finds the target and uses an IR sensor to home on the target, then a missile intercept at this and later times can only be observed by the radar, since the PESH sensor will no longer have a signal to track. Likewise, using a radar and an IR sensor for performing KA is possible when the weather is clear; however, it is not possible to use the IR sensor for KA, target ID, or tracking when there is fog or rain.

Finally, the integration of the data from the various sensors must be weighted in such a way as to give the most accurate account of what happens to the target. For example, how does one weight the fact that the PESH stops receiving a transmission from the target around the time of the missile-target intercept? This is not an easy question; the answer probably depends on the other sensors that are supplying data to the KA processor. Conversely, if the PESH continues to detect an emission from the target after missile-target intercept that is identical to the emission before intercept, then the system would weight this information very strongly and probably declare a no-kill.

Conclusion

Many of the problems associated with multi-sensor integration and data fusion have been discussed, along with some of their applications in the surface Navy. However, the problems discussed in this paper were not exhaustive. For example, the problems of allocating sensor resources and multisensor architectures for communications and computation were not included. Other applications that were omitted include multisensor detection, command and control, and navigation. As concluding remarks, potential limiting factors and future research areas in multisensor integration and data fusion are discussed below.

Limiting Factors

One of the foremost problems in integrating multiple sensors is associated with registration. Poor registration can strongly hinder the data association and fusion processes. Furthermore, in the surface Navy the registration problem includes the problems associated with dynamic alignment (i.e., continuous in time), dissimilar sensors, and asynchronous data.

A second problem is the development of sensor resource allocation algorithms that enable a multisensor system to select the appropriate

sensing strategies from the sensors available to the system. This problem is twofold: the pre-selection process, and the real-time selection process. If the sensor resources cannot be managed efficiently, then the potential of a multisensor system achieving the performance of a stand-alone sensor may be limited.

A third problem is the development of a supervisory controller that manages the information and overall data fusion and integration process. The controller would manage the data flow between multiple sensors. The controller would also be required to sense system faults (e.g., loss of a sensor in the system) and maintain operation during these faults.

Future Research

Specific research problems that relate to the surface Navy include:

- Multisensor alignment
- Dynamic sensor resource allocation
- Target-state estimation
- Data association techniques
- Target identification
- Command and control
- Data management procedures
- Multisensor system design procedures
- Computer architectures for multisensor systems

While most research in the areas of multisensor integration and data fusion should be focused at solving specific problems, some should be

aimed at developing techniques that allow multisensor systems to operate in unknown and dynamic environments. Critical to this undertaking would be the development of sensor and environmental modeling.

References

1. Luo, R.C. and Kay, M.G., "Multisensor Integration and Fusion in Intelligent Systems," *IEEE Trans. Syst. Man Cybern.*, Vol. SMC-19, No. 5, 1989, pp. 901-931.
2. Bar-Shalom, Y. and Fortmann, T.E., *Tracking and Data Association*, Academic Press, Orlando, FL, 1988.
3. Bar-Shalom, Y., ed., *Multitarget-Multisensor Tracking: Advanced Applications*, Artech House, Norwood, MA, 1990.
4. Blackman, S.S., *Multiple-Target Tracking with Radar Applications*, Artech House, Dedham, MA, 1986.
5. Waltz, E. and Llinas, J., *Multisensor Data Fusion*, Artech House, Norwood, MA, 1990.
6. Carson, R.E. and Harris, R.A., *Coordinate Alignment on Flexible Ships*, Report R-1165, The Charles Stark Draper Laboratory, May 1978.
7. Gelb, A., *Applied Optimal Estimation*, M.I.T. Press, Cambridge, MA, 1974.
8. Stark, H. and Woods, J.W., *Probability, Random Processes, and Estimation Theory for Engineers*, Prentice-Hall, Englewood Cliffs, NJ, 1986.
9. Tahk, M. and Speyer, J.L., "Target Tracking Problems Subject to Kinematic Constraints," *IEEE Trans. Auto. Cont.*, Vol. AC-35, No. 3, 1990.
10. Miller, J.T., *Gridlock Analysis Report, Volume I: Concept Development*, Report FS-82-023, The Johns Hopkins University Applied Physics Lab, Feb 1982.
11. Bar-Shalom, Y., Chang, K.C., and Blom, H.A.P., "Tracking of Splitting Targets in Clutter Using an Interacting Multiple Model Joint Probabilistic Data Association Filter," *Proceedings of the 30th IEEE Conference on Decision and Control*, Dec 1991, pp. 2043-2048.

The Authors



RONALD HELMICK received the B.S. and M.S. degrees in physics and the Ph.D. degree in electrical engineering, all from West Virginia University, Morgantown, West Virginia. As a graduate research assistant, he performed research in solid-state physics and estimation and control theory. After completing the Ph.D. degree, he joined the Naval Surface Warfare Center in Dahlgren, Virginia in 1988. His work

involves multisensor alignment, data fusion, and the analysis of algorithms in the Shipboard Gridlock System. He is a member of the Institute of Electrical and Electronics Engineers (IEEE), IEEE Signal Processing Society, and the Society of Industrial and Applied Mathematics.



CHARLES F. FENMORE received the B.A. degree in mathematics from the University of Rochester in 1966 and the Ph.D. degree in mathematics from Pennsylvania State University in 1969. In 1972 and 1973, he was an assistant professor in the Mathematics Department at Clemson University. Aside from the two-year period in which he was at Clemson University, he has worked for the Naval Surface Warfare

Center (now Dahlgren Division) since 1970. He has worked on a variety of projects encompassing mine warfare (during the Vietnam conflict), DDG-963 Class anti-submarine warfare, DDG-2 Class upgrade, and design of the AEGIS destroyer, but the past ten years have been spent assisting in the design, development, implementation, and testing of the algorithms required to establish the surface tactical picture, cruise missile engagement/mission planning, and alignment of the tactical missiles.



WILLIAM DALE BLAIR received the B.S. and M.S. degrees in electrical engineering from Tennessee Technological University in 1985 and 1987, respectively. While a graduate research assistant, Mr. Blair performed robotic controls research for the Center of Excellence for Manufacturing Research and Technology Utilization at Tennessee Tech. After earning the M.S. degree, he joined the Naval System Division of FMC

Corporation in Dahlgren, Virginia as an electrical engineer. His work involved the development and evaluation of new algorithms for weapon/fire control. In 1990, Mr. Blair joined the Naval Surface Warfare Center in Dahlgren, Virginia as an electrical engineer. His work involves adaptive estimation for target tracking, target trajectory prediction, multisensor alignment, data fusion, and feedback control systems. The research has been reported in over thirty publications. Mr. Blair is a member of Institute of Electrical and Electronic Engineers (IEEE), IEEE Control Systems Society, IEEE Education Society, Tau Beta Pi, Eta Kappa Nu, and Phi Kappa Phi.



THEODORE ROGER RICE received the B.S. degree in physics in 1971, the M.S. degree in physics in 1980, and the Ph.D. degree in physics in 1983, all from North Carolina State University, Raleigh. From 1983 to 1985 he worked for the Naval Surface Warfare Center on phased-array radar modeling and signal processing. From 1985 to 1987 he worked for the Naval System Division of FMC Corporation in Dahlgren, Virginia on producing a test bed used to evaluate the performance of target

motion analysis algorithms. Since 1987, he has been working at the Dahlgren Division on various signal-processing methods related to kill assessment for naval weapons control systems and on multisensor alignment and data fusion.

A Multisensor Concept for Detection of Low-Flying Targets

K. H. Krueger and R. A. Stapleton

Low-flying missiles present severe difficulties to shipboard radar systems, which must detect and track these missiles in order to support defensive weapon systems. A system is described wherein the use of both active and passive sensors, optimized for the detection of these targets and integrated in an optimum fashion, maximizes the range at which sea-skimming cruise missiles can be detected and engaged. In the system concept, detections from a horizon-search radar and a horizon-search infrared (IR) sensor are sent to a detection processor where all the available data is used to detect targets and initiate tracks.

Introduction

The goal of this technology-base work is to define and evaluate a system concept that achieves substantial improvement in the detection and acquisition range against low-flying, low radar-cross-section (RCS) targets over that of current shipboard systems. Therefore, emphasis has been placed on developing sensor-integration techniques that operate on detection data prior to the formation of target tracks. This is in contrast to integration approaches, which rely on the sensors to report target declarations and tracks. The short time line imposed by supersonic, sea-skimming, anti-ship missiles makes rapid transition to firm track not only imperative, but suggests that interactions between the sensors be based upon detection data rather than firm tracks. The work described in this article is being performed in the Multisensor Detection (MSD) Task funded by the Surface-Launched Weaponry Technology Block.

Motivation for Integrated IR-Radar

Significant motivation exists for the integration of IR and radar sensors to improve detection of low-flying, low-RCS targets. Since clutter can be severe and long-range propagation is quite difficult, the detection of low-flying, low-RCS targets is a particularly difficult radar problem. In addition, the detection range required to support an engagement increases with target speed, making the radar low-flyer problem even worse for supersonic sea skimmers. However, IR signatures generally increase with target speed, and an optimized IR sensor will, in many cases, detect the supersonic low-flyer before the radar does. While the above-horizon performance of a radar is generally better than that at low angles, the situation is somewhat different for a shipborne IR sensor. Below about one degree elevation angle, cloud clutter is not considered to be a significant problem, primarily due to the very long path length to clouds at low angles. Clouds visible at elevation angles higher than a degree or two are normally close

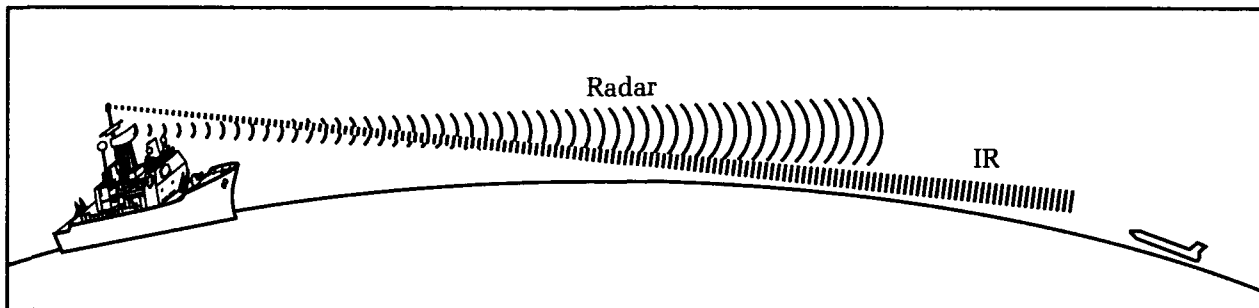


Figure 1. Horizon-search IR sensor complements radar.

enough to degrade detection performance and increase false alarm rates of the IR sensor.¹ Additionally, since approximately 50 percent of the time a ray arriving from an arbitrary direction intersects a cloud, the utility of a shipborne IR sensor to detect high-flying targets over the entire hemisphere at useful ranges is limited. Given these considerations, a vertical field of regard of about one degree appears to make sense. An additional motivation for limiting the field of regard is that sensitivity increases inversely with this parameter, so achieving the required sensitivity is made somewhat easier by minimizing the vertical field of regard. Because the low-elevation region is where radar has the most severe performance difficulties in natural environments, and coupled with the IR performance issues just discussed, an IR sensor covering the low-elevation region is certainly a natural complement to radar. This synergism is illustrated in Figure 1. Having an IR sensor that detects the high-speed, low flyer also allows some flexibility in the radar search frame time. In search-and-confirm strategies, this can result in improved detection range for slower targets and better clutter rejection due to increased time and signal-processing options. An IR sensor might also provide significant performance assistance in radar electronic counter-measure environments, although this has not been examined in any detail.

Figure 2 illustrates the rationale for specifying a high-resolution IR sensor with a relatively small vertical field of regard. The graph shows the elevation angle relative to the horizon for two constant-altitude targets as a function of range. The IR sensor is assumed to be mounted at a height of 30 meters. Notice that for even the higher altitude target, the elevation angle never exceeds 1 mRad. Consequently, a large vertical field of regard is not required for detection or tracking of these targets. It is quite clear that increasing vertical resolution is advantageous from two standpoints. The detection of vertical target motion over time becomes possible, thus adding a strong discriminant which can be used to differentiate between targets and clutter. More

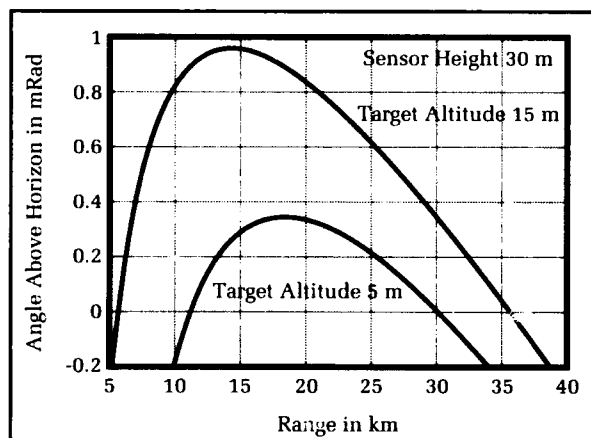


Figure 2. Elevation angle relative to horizon for constant altitude targets.

important, detections from both targets and false alarms from clutter and solar reflections on the sea surface may be spatially separated, thus allowing a much lower IR false alarm rate. A vertical resolution of about 100 μ R is a reasonable compromise between providing sufficient total vertical field of regard, adequate resolution, and keeping the number of detector elements within bounds that can reasonably be achieved with current technology.

A functional block diagram of the system concept is shown in Figure 3. Radar and IR detection data are sent to a multiple hypothesis detection processor which interfaces with a track processor and sensor resource manager. This system would report target tracks to the rest of the combat system in much the same way a single sensor currently does. However, tracks reported from this system are the result of a closely integrated pair of sensors optimized to detect and acquire low-flying targets. The detection processor uses both radar and IR data to make target declarations and request sensor actions. It should be emphasized that in this system targets are declared and tracked on the basis of radar-only contacts, IR-only contacts, and combined radar and IR contacts. This

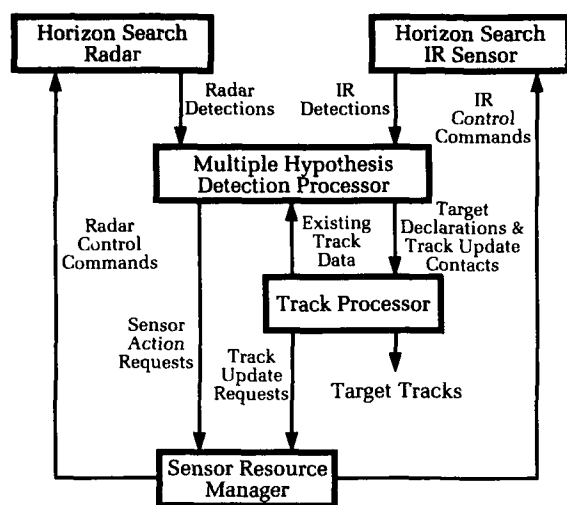


Figure 3. Multisensor detection system concept.

approach prevents the loss of performance that results from a simple "and" decision rule applied to two-sensor data.

The sensor resource manager controls the radar dwell schedule based upon information from the radar itself as well as from the IR sensor. The radar dwell schedule consists of a number of different dwell types, including horizon search, detection confirmation, track update, and cue response dwells. Priorities for each type of dwell are computed for each dwell scheduling interval, with the highest priority dwells being sent to the radar for execution during the next scheduling interval. Control comments to the IR sensor are limited to threshold control over variable-size, angular regions of interest. Controlling additional signal-processing parameters in the IR sensor through the resource manager may prove to be of benefit, but has not been examined in sufficient detail to date. An additional benefit from a centralized resource manager is that, for at least that set of targets observable by both sensors, it is possible to schedule fewer radar track update dwells since an angle track is available from the IR sensor. This can free at least a portion of the radar time and energy budget normally allocated to tracking for performing other functions. The very high-angle resolution of the IR sensor may also allow earlier detection of target maneuvers and discrimination of multiple targets flying in close proximity. Measurements of the environment are used to adjust parameters in the detection processor and resource manager to account for changing weather conditions. As an example, the relative importance assigned to IR detections would decrease with poor IR propagation conditions.

In order to gain some insight into the detection processor, it is instructive to trace the oper-

ations performed each time a new detection is reported by either sensor. In Figure 4, the detection processor maintains a time history of all radar and IR detections that have occurred over the last several seconds. A new contact, represented here by the square, is passed through an azimuth gating criterion that serves to limit the number of potential association groups subsequently formed and evaluated. The azimuth gate widens with time to accommodate uncertainties in target dynamics.

For this notional example, the contacts selected for further processing consist of the new contact labeled "A" and two existing contacts labeled "B" and "C." The four possible groupings of these three contacts are shown. It should be noted that the term association group is generalized to include groups of one, so that a score is also computed on single contacts. Once the scores for all the association groups have been computed, they are compared to two thresholds: a score exceeding the target declaration threshold results in a track initiation on that target; a score exceeding only the lower request for sensor threshold results in a radar cue, a radar detection confirmation dwell, or an IR sensor threshold adjustment. In the example shown here, the score of group (A,C) results in a request for sensor action. The particular type of sensor action is dependent upon the contacts that comprise the association group. For existing tracks, spatial gates and discrimination logic are used to send contact data to the track processor to update those tracks. Logic in the detection processor is designed to prevent those track update contacts from being used simultaneously to form the new target declarations and sensor action requests.

Sensor Considerations

In order to achieve maximum benefits in increasing the detection range against low-flying targets, the multisensor suite must be designed with the goal of integration in mind. Much higher false-alarm rates are allowed in the detection data from each sensor in the multisensor system than would be tolerated from stand-alone sensors. Real-time feedback and control of operating parameters are also important features of the sensors.

The radar that would be used in the sensor suite would nominally operate at X-band to benefit from improved low-elevation propagation relative to that at lower frequencies. The radar would need to be electronically scanned at least in azimuth to allow for fast response to cues, as well as to achieve the required variable-length dwell times. Other considerations such as weapons control may place agile beam requirements in elevation on the radar as well.

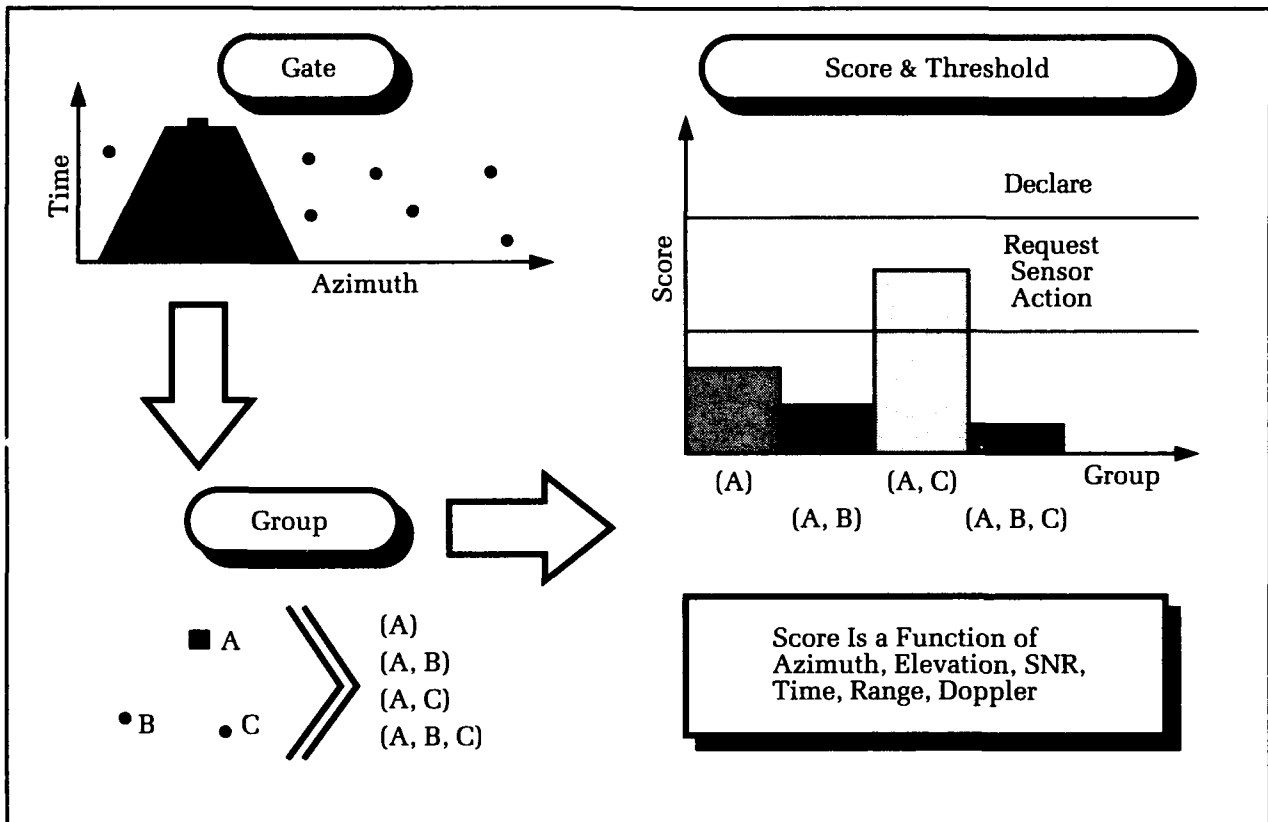


Figure 4. Detection processor conceptual operation.

The radar would need to support waveforms that have significantly enhanced detection capability used to respond to cues from the IR sensor. A cued dwell approach having an approximately 20-dB detection performance improvement relative to the normal search dwells has been developed under the MSD task in cooperation with the Naval Research Laboratory's (NRL's) Radar Division.² As can be seen from examination of Figure 5, a 20-dB increase results in about a 5-km increase in detection range in a non-ducting propagation environment. It is estimated that these cued dwells could be executed approximately once every second or two without significantly degrading the desired two-second horizon search frame time of the radar.

An IR sensor optimized for high-speed, low-flyer detection would probably be a mid-wave system operating in the 3- to 5-micron band and would have a vertical field of view on the order of one degree. A mid-wave sensor would offer improved performance over a long-wave system for this particular application due to comparatively lower extinction coefficients in the humid conditions typical of many marine operating environments. Additionally, since the spectral signature due to aerodynamic heating of high-speed, sea-skimmer missiles peaks in the 3- to 5-

micron band, the mid-wave band is probably the optimum choice. The desired sensor would ideally have a sensitivity of roughly 2×10^{-15} watts/steradian, which is about an order of magnitude greater than that of current sensors. This sensitivity level would extend horizon-limited detection performance to a wider range of

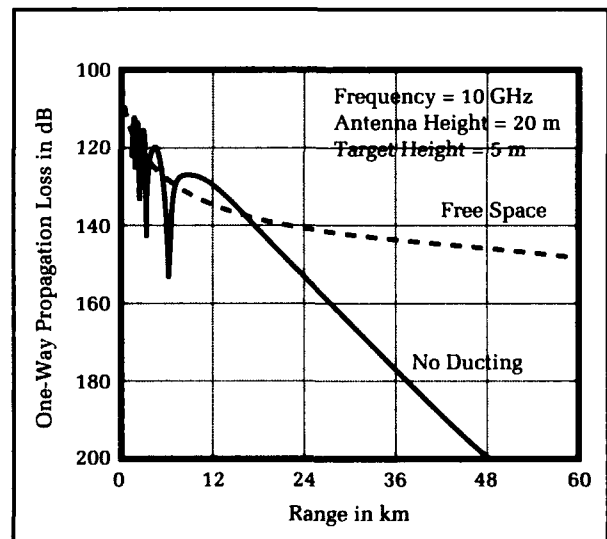


Figure 5. One-way propagation loss.

weather conditions and would provide some capability against lower signature targets such as subsonic cruise missiles. An effective horizon scan rate of 1 to 2 Hz is desired, and may be obtained with a single, rotating aperture or with multiple, fixed apertures. Multiple apertures may be necessary to achieve the required sensitivity and update rate. Vertical resolution on the order of 100 μ R is desired for reasons already discussed. Stabilization to compensate for the effects of ship motion on a high-resolution sensor is difficult while still maintaining a reasonable size and weight package, but may be achieved using stable mirror technology. The weight of the IR sensor must be kept to a minimum in order to mount it as high as possible on the ship's mast and allow extension of horizon-limited detection range against low-elevation targets. Stable mirror techniques are attractive for this application, as gyro packages mounted on mirrors internal to the sensor perform fine stabilization, thus allowing use of a lightweight, coarse mechanical stabilization of the entire sensor.

An important attribute of the IR sensor is the capability to generate target detections usable for cues on the basis of a single scan. Conventional systems built for stand-alone operation usually require multiple hits on consecutive scans in order to declare a target. This criterion, which is used to reduce the false alarm rate of the sensor, will end up significantly reducing the detection range of the multisensor system. If necessary, some type of multicolor scheme might be used in the mid-wave, horizon-optimized IR sensor in order to reduce the number of false detections on clutter and nuisance targets, while still allowing single-hit detection.

Detection performance of the multisensor suite is strongly affected by the propagation conditions seen by each sensor. Performance of the IR sensor is affected by both atmospheric extinction and atmospheric refraction. The effects of atmospheric extinction on IR sensor performance are relatively well understood and have been modeled for some time. Under certain conditions, IR refractivity caused by low-elevation temperature gradients can dominate system performance. The temperature lapse rate of the atmosphere at low elevations, due to differences in air and sea temperatures, causes abnormal ray bending along low-elevation paths and subsequent shifts in the apparent IR horizon. For the geometries involved with low-elevation targets, apparent shifts of target elevation angle on the order of 1 to 2 mR are not uncommon at large air/sea temperature differences. The areas of IR refractivity measurement and the design of models to predict the magnitude of the effects both warrant further investigation.

Nonstandard propagation conditions also affect the performance of the radar. Evaporative ducts due to abnormal changes in the refractivity of the atmosphere can severely affect detection performance of the radar against low-elevation targets. During recent MSD tests of an NRL exploratory development radar...a candidate for future multisensor experiments...return signal strength fluctuations of greater than 50 dB have been observed against a low-elevation, over-the-horizon reference target. As with IR refractivity, ducting phenomenology and the development of associated modeling tools merit further investigation.

In order to assess the impact of propagation on detection performance of the multisensor system, IR extinction coefficients and radar duct heights were computed for each of the entities in a matrix containing 400 random, worldwide weather observations. The detection performance of each sensor was then calculated, with the radar using either a standard search waveform or a cued waveform with 20 dB higher energy. This study demonstrated that the use of the IR sensor as a cueing source for the radar resulted in a 20- to 40-percent increase in detection range against the high-speed, low-elevation target.

Multisensor Detection Simulation

Multisensor Detection Simulation (MSDSIM) is a non-real-time computer simulation developed to explore and evaluate different sensor integration and control concepts. The simulation structure closely follows that of the system concept block diagram shown in Figure 3. Radar and IR models feed relatively high false alarm rate contacts to a multiple-hypothesis detection processor. A sensor resource manager schedules radar dwells and defines IR special interest zones. A scenario generator is used in the simulation to specify all pertinent radar and IR sensor parameters, thus allowing tradeoff analyses to be performed with respect to sensor characteristics. Several different detection processing and resource management schemes are currently being studied. The simulation is also structured so that joint sensor data, collection of which is planned in multisensor-detection tests, can replace the IR and radar models, allowing algorithm validation with real data.

Figure 6 shows results derived from representative MSDSIM runs for two cases using a notional sea-skimming target. The upper time line shows the sequence of events that occurs for the integrated radar IR sensor system. Because the notional target speed is approximately 1000 m/s, the numerical scale in km can also be interpreted as time-to-go in seconds. The target is first detected at 25.5 km by the IR sensor; the

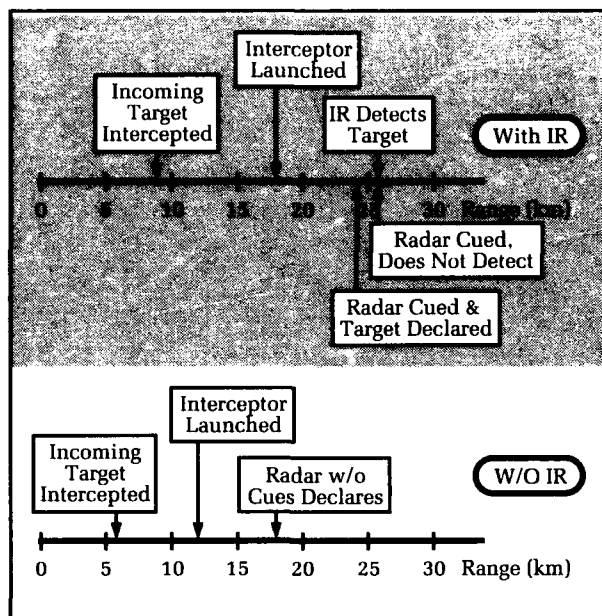


Figure 6. Time lines comparing integrated radar and IR with radar only.

radar is immediately cued but does not detect the target. A subsequent cue to the radar results in target detection, and track is initiated at a range of 24 km. The lower time line shows the performance of the radar alone without cues from the IR sensor. In this case, the radar declares the target at 18 km.

In order to provide some perspective on the benefits of the integrated sensor system, the predicted launch and intercept points of a notional weapon system are also included on both time lines. The interceptor has an average speed of 3000 fps, and a time delay of 6 seconds is assumed for threat evaluation and weapons assignment. The predicted intercept point of 8.7 km for the upper time line is at a range long enough to permit a second interceptor to be used if necessary. The 5.8-km predicted intercept range in the lower time line most likely precludes the opportunity to use a shoot-look-shoot engagement strategy in this case.

Summary

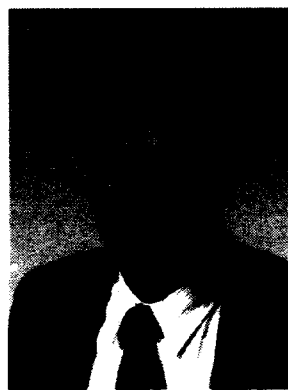
This multisensor concept is a closely integrated radar and IR system optimized for the detection and acquisition of low-flying targets. The advantage of this approach is founded on the complementary performance of radar and IR sensors against the supersonic, low-flying, low-RCS missiles. In fact, the major contribution made by the IR sensor is early detection of the supersonic sea skimmer. This IR detection information allows radar resources to be scheduled such that increased acquisition and

intercept ranges are achieved relative to a radar-only approach. The benefits of integration are fully realized only when the individual sensor characteristics are optimized based upon the performance requirements for the integrated system. Optimizing performance includes not only optimizing the integration algorithms and sensor control strategies, but also carefully specifying the requirements and characteristics of the sensors themselves.

References

1. Scribner, D. A., et al., *Advanced Staring Infrared FPA Sensor Concepts For Use In ShipBoard Search and Track Systems*, NRL Technical Report, 17 Mar 1988.
2. Trunk, G. V., et al., "A Cued Phased Array Radar System for the Detection of Low Altitude, Low Observable Targets," *Tri-Service Radar Symposium Proceedings*, 1990.

The Authors



KARL H. KRUEGER, an electronics engineer in the Electronics Systems Department since 1974, holds a B.S.E.E. degree from Youngstown State University and an M.S.E.E. from George Washington University. His major work experience includes design and development of equipment for testing electronics support measures receivers, concept and prototype hardware development for anti-radiation missile

countermeasures, design and development of a prototype transmitter for the offboard deception device system, technical concept exploration for the NATO Anti-Air Warfare program, and development and technical leadership of the Multisensor Detection program. Currently, he is working on advanced sensor concepts for AEGIS.



RON STAPLETON graduated from the Tennessee Technological University in 1985 with a bachelor's degree in electrical engineering. Since that time, he has been employed at the Dahlgren Division and has worked in the areas of ARM countermeasures, radar signal processing, and radar systems design. He is currently the project leader of the Multisensor Detection program.

ESTAR: Expert System for Track Ambiguity Resolution

M. J. Kuchinski, R. L. Taft, and H. C. Leung

*The Expert System for Track Ambiguity Resolution (ESTAR) was developed by the Combat System Technology Branch for use as an intelligent post-processor to conventional track correlation algorithms. Typically, over-the-horizon (OTH) track contacts that cannot be resolved algorithmically are placed in an ambiguity file for eventual resolution by a track console operator. ESTAR correlates ambiguous OTH track contacts in a manner similar to the decision-making process currently employed by these operators. It does so using a set of **antecedent/consequent** rules that incorporate much of a track operator's knowledge about a contact's kinematic, attribute, and electronic intelligence (ELINT) characteristics. To insure flexibility, these rules are divided into subcategories, called **contexts**, which can be ordered according to different prioritization schemes, called **scripts**. The rules are matched against information about potential correlation candidates' features, as well as those of the ambiguity itself. In addition, a data base that includes the attributes of specific ships, ship classes, and their emitter suites is also included. The inference engine that matches and selects rules and makes final recommendations is based on an extended version of **Forgy's Rete Algorithm**.*

Background

ESTAR is an expert system¹ designed to resolve OTH track ambiguities. When an OTH contact is processed by a typical track correlation system operating in the fleet, it can be categorized in one of two ways. Either it is a new track that the system has never seen before, or it is a new instance of a previous contact. An OTH contact becomes an ambiguity whenever a conventional track correlation algorithm is unable to determine in which of these categories the ambiguity belongs. Typically, an unresolved contact is placed in an ambiguity file until it can be resolved, usually by a human track operator. ESTAR is intended to serve as an intelligent assistant to that operator. As such, ESTAR functions as a post-processor to a conventional track correlation algorithm.

Normally, when a human track console operator is tasked to resolve ambiguities, the process is a manual one, often tedious and time-consuming. Several options exist for resolving an ambiguity, and each must be weighed and evaluated to determine if it is appropriate. One resolution option includes using the ambiguity to update an existing track in the track file. This is done whenever an ambiguity is determined to be a new instance of an old track. A second option is to make the ambiguity into a new track in its own right. This is done whenever the operator determines that no track in the track file matches the

ambiguity well enough to warrant an update. An operator has other options, as well. An ambiguity may be deferred, thereby postponing its resolution until a later time, or it may be deleted from the ambiguity file altogether. Figure 1 summarizes these resolution options. Regardless of how the ambiguity is resolved, however, the resolution decision is usually the operator's alone.

In making these decisions, operators frequently make use of information from outside sources, notably, various naval warfare publications as well as *Jane's Fighting Ships*. These sources provide information on the parameters of different emitters and the characteristics of various ship classes, as well as specific data on individual platforms. This information permits operators to narrow (or broaden) their decision-making options. The OTH contacts that cause ambiguities originate from a variety of sensor sources. These include high frequency direction finders, aircraft, overhead sensors (i.e., satellites), as well as passive and active radar contacts.

Overview of ESTAR

ESTAR is intended as an intelligent assistant to the track operators in the fleet. To that end, *ESTAR*'s approach to ambiguity resolution ought not be a radical departure from the operator's own. Thus, following the approach taken by most track operators, *ESTAR* sequentially examines and resolves track ambiguities. Its decisions are based on information from a varied number

of data sources. The first of these includes, quite obviously, the track and ambiguity files. In addition, a data base of specific ships, ship classes, and emitter suites is available to *ESTAR*. The purpose of this data base is to provide to *ESTAR* the same kind of information available to the human operator from the publications mentioned above.

Facts from these and other sources, along with the heuristic **antecedent/consequent** (i.e., if/then) rules that govern the utilization of these facts, form the **knowledge base** for *ESTAR*. Facts are matched with rules by means of an **inference engine** that is forward-chaining and data-driven. Currently known facts and matched rule instances are maintained in a network data structure similar to that utilized by Forgy's *Rete* Algorithm.²

ESTAR is initially provided with all tracks currently in the correlation system's track file, and subsequently with each ambiguity in turn. Like the track operators, *ESTAR* addresses the ambiguities one by one, and only after resolving (or deferring) a given ambiguity is another taken into consideration. *ESTAR* can function in one of two capacities. The first is an advisory mode, in which only recommendations on resolutions are made to the operator. The second is a fully automatic mode in which ambiguities are resolved without any intervention by the operator.

ESTAR's rule set can be loosely divided into two broad categories: those rules that result in

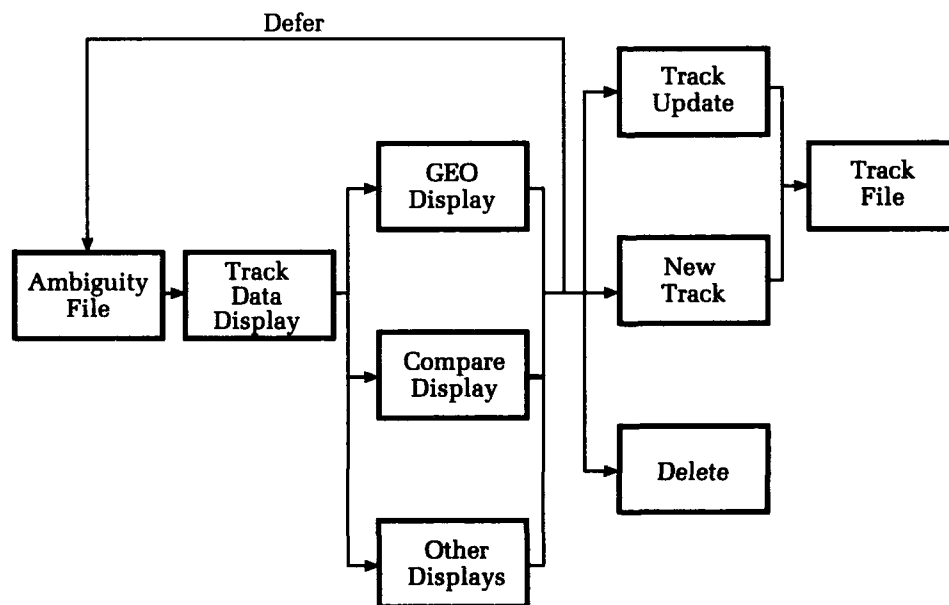


Figure 1. Summary of ambiguity resolution options.

final resolution (or deferral) of an ambiguity, and those that merely add or delete facts from the knowledge base. Concerning the latter, any change to the fact base can conceivably change which rules are instantiated. A rule **instantiation** is an instance of a rule whose antecedent matches the facts in the knowledge base. A given rule may have multiple instantiations for different combinations of facts. *ESTAR*'s decision-making process with regard to an ambiguity is not complete until all rule instantiations related to that ambiguity have had an opportunity to **fire**. A rule instantiation is said to fire when its consequent is actually carried out.

Each ambiguity that *ESTAR* receives is initially accompanied by a list of likely update candidates from among the tracks in the track file. Each candidate in this list includes numeric scores that are indicative of the candidate's goodness-of-fit compared to the ambiguity's attribute, kinematic, and ELINT data. The scores are weighted sums, with like characteristics between the ambiguity and candidate weighted positively, and unlike characteristics weighted negatively. The higher the scores, the greater the likelihood that the ambiguity is a new instance of that candidate track. These candidates, however, are used only as a departure point for *ESTAR*'s inferential reasoning process. As with a human operator, *ESTAR* is not inhibited from updating a track different from these candidate tracks if it sees fit to do so. Thus, *ESTAR* maintains its *own* candidate list, or **interest set**, for each ambiguity. This interest set is initially identical to the track correlation system's candidate list, but is modified as *ESTAR* eliminates certain candidates from consideration while entertaining the possible inclusion of others not in the original list. The final disposition of the ambiguity's interest set, once *ESTAR* has fired all instantiated rules related to the ambiguity, determines the resolution of the ambiguity. If the interest set is empty, the ambiguity is made into a new track. If the interest set contains exactly one candidate, and its correlation score is sufficiently high, the candidate is updated with the ambiguous contact. On the other hand, if the interest set contains exactly one candidate, but its correlation score is too low, the candidate is discarded and the ambiguity is made into a new track. Finally, if the interest set contains two or more candidates, the ambiguity is deferred.

The option to delete an ambiguity is used rarely, if at all, by human track operators, and *ESTAR*'s rule set reflects that conservatism. *ESTAR* will only recommend the deletion of an ambiguity if certain unusual circumstances exist. Some of these circumstances include saturation of the ambiguity file, an ambiguity report that is more than twenty-four hours old, or an

ambiguity report that adds no new information to the report base.

Knowledge Base

The information contained in the knowledge base for *ESTAR* consists largely of the reported information on the tracks and ambiguities currently in the track correlation system. In addition, known information about the characteristics of specific ships, ship classes, and emitter parameters is contained in data base form. Also, knowledge about the deductive reasoning that a track operator applies to the resolution of ambiguities was obtained from lengthy conversations with operators in the fleet. This information is contained in the antecedent/consequent rules. Knowledge from all these sources, as well as the status of the interest set, the subgroup of rules that is currently under consideration, and the number of ambiguities pending, constitute the information available to *ESTAR* in the course of rendering decisions.

It is noteworthy that at no time does *ESTAR* modify specific data fields in the correlation system's track and ambiguity records. While *ESTAR* may make decisions that alter one or more fields in a track or ambiguity, *ESTAR* retains that information only in its *own* fact base. The only time at which *ESTAR* has any impact on the track or ambiguity records as they are retained in the correlation system is when *ESTAR*, in fully automatic mode, makes a decision about a final resolution of an ambiguity. Only then are the track correlation system's track and ambiguity files modified, and then only in accordance with the prescribed algorithms for creating a new track, updating an existing track, or deleting an ambiguity. These algorithms of necessity already exist, as they must be invoked whenever a human operator renders a decision that resolves an ambiguous contact.

Report Base

The track and ambiguity records, along with the ship data base information, comprise the majority of factual information that can be matched to *ESTAR*'s rule set. The track file contains all the tracks currently in the correlation system and forms the pool from which each ambiguity's candidate list is drawn. Because *ESTAR* is intended only as a post-processor to a conventional correlation algorithm, the track and ambiguity records to which *ESTAR* has access represent contacts that have already been processed through that system. Contacts typically come into the system as OTH messages, and OTH message fields typify the kinds of informa-

tion in a given track or ambiguity record. Ambiguities are essentially identical to tracks in format and content. They differ from tracks only in that they have not been absorbed into the tactical picture, and hence have no historical information concerning their previous disposition. While *ESTAR* always has access to all tracks, *ESTAR* deals only with one ambiguity at a time.

In the course of performing the knowledge acquisition necessary to establish *ESTAR*'s system requirements, it was discovered that track operators frequently rely on outside sources of information to augment or amplify the data in a track or ambiguity's record. Those sources typically include official documents, such as Naval Warfare Publications, that detail the characteristics of friendly and hostile platforms. The information may be specific to a particular ship, related to an entire class of ships, or confined solely to an emitter and its parameters. A frequently used alternative to these official sources is *Jane's Fighting Ships*, which contains synopses of all naval combatants, worldwide.

The types of information that an operator engaged in the resolution of track ambiguities finds useful are enumerated in Figure 2. These categories of information are either already included in *ESTAR*'s ship data base or are proposed for future upgrades.

The facts in *ESTAR*'s knowledge base include other data relevant to the resolution of ambiguous contacts, but not derived directly from the track or ambiguity files or the ship data base. These facts usually relate to the tracks in *ESTAR*'s interest set, to the number of ambiguities still outstanding in the ambiguity file, or to information about which subgroup of *ESTAR*'s expert rules is currently active. Furthermore, it is the nature of *ESTAR*'s rules that they typically cause new information about a track or ambiguity to be inferred. These new facts are added to the knowledge base and can be used as the basis for further inferences.

Rule Base

The rules for *ESTAR* are broken down into subgroups, called *contexts*, which can be interleaved in any order desired. A given ordering is referred to as a script. Figure 3 illustrates a typical script. This approach ensures flexibility in the prioritization of the rules, as a script specifies the order in which the contexts will be given an opportunity to fire.

The rules themselves are each composed of an antecedent (if-component) and a consequent (then-component). The antecedents consist of **conjunctions**, **disjunctions**, and **negations**, which join together a series of **clauses**. These clauses are made up of combinations of tokens

Ship Specific:

Ship Name
Hull Number
NOSIC ID
International Call Sign

Ship Class:

Maximum Speed
Maximum Range
Maximum Threat Radius
Nationality
Propulsion Type
Ship Type (e.g., Destroyer, Cruiser, etc.)
Emitter Suite

Emitter Specific:

ELINT Numbers
Pulse Width
Scan Type
Maximum/Minimum PRF, PRI, & RF

Figure 2. Ship data base information.

- 1.) Test for Auto-Correlation
- 2.) Deletion/Saturation Criteria
- 3.) Cross-Check with Ship Data Base
- 4.) Check for Other Tracks of Interest
- 5.) Narrow by Attribute and Position Scores
- 6.) Narrow by Threat Class and Position Class
- 7.) Narrow by Emitter
- 8.) Tie-Breaker
- 9.) Interest Set Finished

Figure 3. A typical script for *ESTAR*.

that can be represented by **constants**, **variables**, or **predicate functions**. This represents an elaboration of Forgy's approach,² in which only conjunctions and negations are permitted. The inclusion of disjunctions permits explicit logical relationships and precludes the need for multiple rules with the same consequent. The use of predicate functions as part of the antecedent is another feature that extends Forgy's approach. Many decisions are made through a combination of symbolic and algorithmic reasoning, and functional clauses permit the utilization of existing algorithms where appropriate. The consequents of each rule usually consist of one or more procedures that have the side effect of modifying the fact base.

Each rule is initially specified according to a Backus-Naur-Form (BNF).³ Figure 4 exhibits this BNF, while Figure 5 exhibits a typical *ESTAR* rule. Note that the *ESTAR* rule format allows for the representation of extremely complicated logical expressions. Note also that special characters are used to specify various features. Question marks denote variables, exclamation points denote predicates, and semicolons precede variables to which returned function values are bound.

From the format illustrated in Figure 5, the rule set is transformed algorithmically into a network-like data structure. For a given rule set, this network is built only once, and is initially populated with the facts known to *ESTAR* at that time. As each ambiguity undergoes resolution, facts may be added to or deleted from the network, but the essential structure of the network remains unchanged as long as the rule set remains static.

Inference Engine

The nature of *ESTAR*'s inference engine is such that the rules, facts, and matching algorithm are all integrated into the network data structure. The rules themselves dictate the topology of the network, with each antecedent clause divided into its component tokens. Each unique token is represented by a node in the network. These include **constant-nodes**, **variable-nodes**, and **function-nodes**. Other nodes are **memory-nodes** that store current variable bindings and maintain the facts that match a given clause or logical grouping of clauses. Still other nodes, called **and-nodes** and **and-not-nodes**, maintain information about logical paths through the rules. The consequents of the rules, in the form of **terminal-nodes**, are at the bottom of the network.

Loosely speaking, any given path through the network, from top to bottom, represents a single rule, while any given path through the network, top to bottom *with a consistent set of variable bindings*, represents a single rule instantiation.

For illustrative purposes, a small subset of *ESTAR*'s rule set is specified in Figure 6, and its corresponding network is specified in Figure 7.

ESTAR's inference engine uses a forward-chaining, data-driven matching algorithm. It was designed and built in the LISP programming language in lieu of using an expert system shell. The reasons for the decision to forego an expert system shell were twofold. First, it was recognized that a production version of *ESTAR* would eventually have to run on militarized hardware. Such hardware does not typically support commercialized software packages. Second, the proprietary nature of commercial expert system shells means that the source code for their inference engine is unavailable to the user, a situation that was deemed unacceptable for a tactical application. While the inference engine was developed with the *ESTAR* application in mind, it is largely generic. Both the inference engine and the accompanying rule-specification structure could be used for other applications.

Indeed, as of this writing, a related effort is under way to develop a multiple language backend to *ESTAR*'s inference engine. When completed, it will take a user's rule set, build the accompanying inferential network as a data structure, and then reinterpret this data structure into a program structure in any of several languages, including Ada. The end result will be an inference engine that can then be embedded into the user's application program, in the user's language of choice. What's more, the reinterpretation of the data structure into "hard code" is expected to produce a significant increase in the speed at which the engine can perform its matching, a critical issue for real-time applications.

Matching Algorithm

The need for *ESTAR* to respond to an ambiguity in a timely manner mandates that the matching algorithm used in the inference engine be able to respond in near-real-time. The issues of speed and efficiency in match algorithms are ones that have arisen in the literature.^{2,4,5,6} In light of this knowledge, the inference engine was developed around Forgy's *Rete* Algorithm with certain refinements and enhancements added. The features of *Rete* that suggested it was appropriate for this application included temporal redundancy and structural similarity. Temporal redundancy assumes that the working memory of the expert system changes slowly, and exploits this by rematching only those rules affected by a new piece of data. Thus, it is most useful in those situations where information accumulates gradually. Structural similarity means that the algorithm is able to recognize the

Figure 4. Backus-Naur form for ESTAR rules.

```

<rules> ::= ( <weight> <lhs> '==>' <rhs> )

<weight> ::= 100 | <digit> | <nsdigit> <digit>
<digit> ::= 0 | <nsdigit>
<nsdigit> ::= 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9

<lhs> ::= <or_group> | <and_group>
<or_group> ::= <clause> |
              ' (or ' <and_group> <or_clause> )
<or_clause> ::= <and_group> | <and_group> <or_clause>
<and_group> ::= <clause> |
              ' (and ' <and_clause> )
<and_clause> ::= <or_group> |
                <or_group> <and_clause> |
                ' (not ' <clause> ) |
                ' (not ' <clause> ) <and_clause>

<clause> ::= [ <atoms> ] |
            [ <function> ' , ' ] |
            [ <function> ' , ' <variables> ]
<atoms> ::= <atom> | <atoms> <sp> <atoms> | ( <atoms> )
<variables> ::= <variable> | <variable> <sp> <variables>
<atom> ::= <object> | <variable> | ( <function> )
<object> ::= <char> | <char> <object>
<char> ::= <letter> | <digit> | <punctuation>
<letter> ::= A | B | C | D | E | F |
            G | H | I | J | K | L |
            M | N | O | P | Q | R |
            S | T | U | V | W | X |
            Y | Z |
<punctuation> ::= - | _
<variable> ::= ? <object>
<function> ::= <function_name> |
              <function_name> <sp> <atoms> |
              <predicate> <sp> <atom> <sp> <atom> |
              <predicate2> <sp> <atom> <sp> <atoms>
<function_name> ::= ! <object>
<predicate> ::= '!<' | '!<' | '!<' | '!>' | '!>' |
              '!<=' | '!<=' | '!<=' | '!>' | '!>'
<predicate2> ::= '!+' | '!+'

<rhs> ::= [ <function> ] | [ <function> ] <rhs>

```

```

((RULE 0602)
  90
  (AND [context check_for_other_tracks_of_interest]
    [current_ambiguity dtg_and_position_is
      ?ambig_tod ?ambig_mjd ?ambig_lat ?ambig_long]
    [?some_candidate dtg_and_position_is
      ?cand_tod ?cand_mjd ?cand_lat ?cand_long]
    [current_ambiguity major_axis_is ?ambig_maj]
    [?some_candidate major_axis_is ?cand_maj]
    [!lower_bound_on_speed
      ?ambig_mjd ?cand_mjd ?ambig_tod ?cand_tod
      ?ambig_lat ?cand_lat ?ambig_long
      ?cand_long ?ambig_mjd ?cand_mjd ;
      ?speed_bound_in_knots]
    [!<= ?speed_bound_in_knots 30 ;]
    (NOT [?some_candidate previously_in_interest_set])
    [?some_candidate class_is ?cand_class]
    (OR (AND [current_ambiguity ?i
      emitter_name_is ?ambig_emitter_i]
      [current_ambiguity ?j
      emitter_name_is ?ambig_emitter_j]
      [current_ambiguity ?k
      emitter_name_is ?ambig_emitter_k]
      [!combined_emitter_suite possible
      ?this_class ?ambig_emitter_i
      ?ambig_emitter_j ambig_emitter_k ;])
      (AND [current_ambiguity ?i
      emitter_name_is ?ambig_emitter_i]
      [current_ambiguity ?j
      emitter_name_is ?ambig_emitter_j]
      [!combined_emitter_suite possible
      ?this_class ?ambig_emitter_i
      ?ambig_emitter_j ;])
      (AND [current_ambiguity ?i
      emitter_name_is ?ambig_emitter_i]
      [!combined_emitter_suite possible
      ?this_class ?ambig_emitter_i ;])))
    ==>
    [!post_fact "(?some_candidate is_in_interest_set)"]

```

Figure 5. A typical ESTAR rule.

```
((Rule 0200) 301
  (and [context narrow]
    [current_ambiguity status_is unresolved]
    [!= (length *interest-set*) 0 ;]
    [!get_next_context ; ?next_context])
  ==>
  [!remove_fact (list "context" "narrow")]
  [!post_fact (list "context" ?next_context)]
  [!pop_script ; ])

((Rule 0260) 250
  (and [context narrow]
    [current_ambiguity status_is unresolved]
    [current_ambiguity call_sign_only]
    [?some_candidate is_in_interest_set]
    [current_ambiguity call_sign_is ?ambig_call_sign]
    (not [?some_candidate call_sign_is
          ?ambig_call_sign]))
  ==>
  [!remove_fact (list ?some_candidate
    "is_in_interest_set")]
  [!remove_from_interest_set ?some_candidate])
```

Figure 6. Subset of ESTAR rule set.

occurrence of identical features in the rules, and thus avoids making the same match multiple times. A consequence of this is that partially matched rules are maintained, and that a rule's instantiations develop incrementally.

Only one rule can fire at a time, so each is assigned a numeric priority that specifies its importance relative to other rules in its context and provides the basis for conflict resolution. The rule instantiation with the highest priority takes precedence over other rule instantiations.

User Interface

The laboratory version of *ESTAR* has a fairly elaborate user interface that can be reconfigured as the user desires. The interface includes a world map with track symbol overlays, a display for system statistics, a rule trace/justification mechanism by which the system can display and defend its decisions, and a comparison display in which an ambiguity's vital characteristics are displayed side by side with those of a potential update candidate. A display to provide ship data base information is also planned. In this form *ESTAR* could easily be used both for training and for post-interaction assessment.

Any version of *ESTAR* used in the fleet, however, will most likely have a minimal interface.

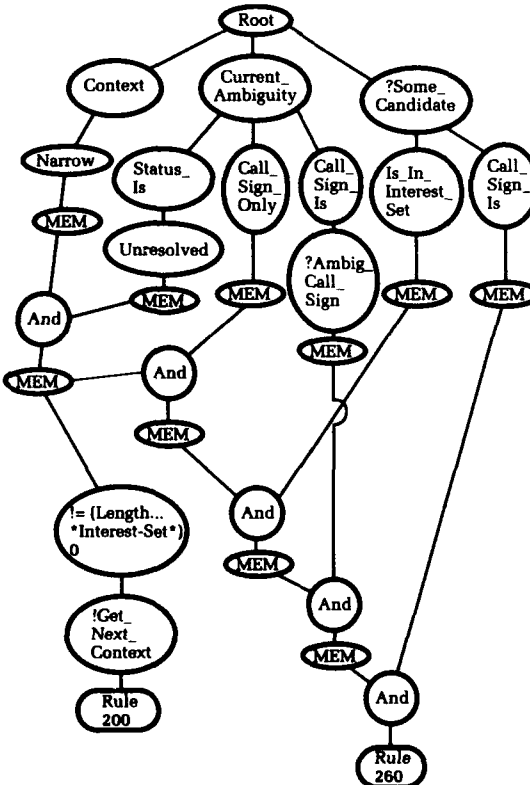


Figure 7. Portion of ESTAR inference network.

ESTAR is intended to relieve the operator of much of the burden of resolving ambiguous track contacts, and thereby decrease the operator's workload. As such it must be a nonintrusive, semiautonomous system.

Summary

ESTAR has already undergone testing in a laboratory environment. Part of that testing has involved the participation of domain experts, those same individuals whose knowledge and expertise inspired the heuristic rule set and formed the basis for the ship data base. The prototype has been generally well received by the experts and other observers, some of whom have suggested that it be used as a basis for a rule-based correlator, either to replace or to augment existing algorithmic correlators. In addition, tactical scenarios with accompanying "ground truth" data continue to be sought to serve as benchmarks for the system's performance.

The Authors



MICHAEL J. KUCHINSKI is a mathematician and task group leader of the Artificial Intelligence/Expert System group of the Engineering & Technology Branch, Combat Systems Department, and a part-time faculty member in the Computer Science Department at Mary Washington College, Fredericksburg, Virginia. He received a B.S. in mathematics from East Stroudsburg University in 1975 and an M.S. in mathematics from

West Virginia University in 1977. He has worked in applying AI to combat systems since 1983. His research interests include expert systems, genetic algorithms, neural networks and virtual environments.



HARRY C. LEUNG is a computer scientist in the Artificial Intelligence/Expert System group of the Engineering & Technology Branch, Combat Systems Department. He received a B.S. in chemistry and a B.S. in computer science from Virginia Commonwealth University in 1981 and 1983, respectively. His research interests include expert systems, neural networks, and virtual environments.

References

1. Hayes-Roth, Frederick, et al., *Building Expert Systems*, Addison-Wesley Publishing Company, Reading, MA, 1983.
2. Forgy, Charles L., *On the Efficient Implementation of Production Systems*, Ph.D. Thesis, Department of Computer Science, Carnegie-Mellon University, 1979.
3. Lewis, Philip M., et al., *The Systems Programming Series; Compiler Design Theory*, Addison-Wesley Publishing Company, Reading, MA, 1978.
4. Miranker, Daniel P., *TREAT: A Better Match Algorithm for AI Production Systems*, Technical Report AI TR87-58, Artificial Intelligence Laboratory, The University of Texas at Austin, Jul 1987.
5. Gupta, Anoop, *Parallelism in Production Systems*, Ph.D. Thesis, Department of Computer Science, Carnegie-Mellon University, 1985.
6. Oflazer, Kemal, *Partitioning in Parallel Processing of Production Systems*, Ph.D. Thesis, Department of Computer Science, Carnegie-Mellon University, 1987.



ROBERT L. TAFT is a mathematician in the Artificial Intelligence/Expert System group of the Engineering & Technology Branch, Combat Systems Department. He received a B.A. in mathematics and an M.S. in mathematics from West Virginia University in 1976 and 1978, respectively. His research interests include expert systems, neural networks, and virtual environments.

Response of the Transfer Function of an Alpha-Beta Filter to Various Measurement Models

John E. Gray and William Murray

Alpha-beta filters are commonly used in tactical naval radar tracking systems when it is necessary to track a large number of potential threats. It is important to understand the response characteristics of these filters to have an effective knowledge of system performance against different threat profiles. To gain this knowledge, details of the different response characteristics of the alpha-beta filter are used to quantify the filter's performance against different measurement models representing a target's trajectory. The transfer functions for an alpha-beta filter are used to derive closed-form (solutions) expressions for smoothed position and velocity outputs for various measurement models. The filter's response to constant velocity targets is found to be the input plus a sinusoidal transient. Constant acceleration measurement models, in addition, yield a steady-state bias that is a function of the filter parameters alpha and beta. These results are used to establish filter lag time due to target maneuvers and filter settling times. Finally, the filter's response to a sinusoidal input is determined. The response characteristics of the filter can then be used to quantify performance against a given threat's trajectory profile. This enables determination of system-level performance against various threat profiles, which is valuable for both tracking and weapon engagement.

Introduction

With a tracking radar, it is possible to measure the position of the target directly, but it is not possible to measure the velocity of the target directly.¹ A means of estimating future positions and velocities of the target is needed. Some of the earliest filters used in tracking radars are the alpha-beta filter and the alpha-beta-gamma filter. The variable-gain, alpha-beta filter combines elements of the Kalman filter² and the alpha-beta filter and has found application when large numbers of objects are being tracked. The information derived from the filter is used not only to arrive at pointing orders for radars, but also for other applications including prediction of intercept points, missile guidance orders, and computation of decelerations. It is useful to know the response of the filter to different measurement models other than the constant-velocity model. To calculate the deterministic responses of the alpha-beta filter to different measurement models, a factorization of the transfer functions is presented that leads to manageable computations of the response function of the alpha-beta filter.

The tracking equations of an alpha-beta filter for a one-coordinate target assume a model that is moving at a constant velocity. Given this assumption, the mean squared error in the filtered velocity and position estimates is minimized to get optimal estimates. This is done by choosing two coefficients that weight the differences of the predicted and measured positions and combine them with the present prediction to arrive at new estimates. The tracking equations for the alpha-beta filters are given by

$$x_p(k) = x_s(k-1) + v_s(k-1)T, \quad (1)$$

$$v_p(k) = v_s(k-1), \quad (2)$$

$$x_s(k) = x_p(k) + \alpha(x_m(k) - x_p(k)), \quad (3)$$

$$v_s(k) = v_p(k) + \left(\frac{\beta}{T}\right)(x_m(k) - x_p(k)); \quad (4)$$

where

$x_s(k) \equiv$ smoothed position at the k -th interval,

$x_p(k) \equiv$ predicted position at the k -th interval,

$x_m(k) \equiv$ measured position at the k -th interval,

$v_s(k) \equiv$ smoothed velocity at the k -th interval,

$v_p(k) \equiv$ predicted velocity at the k -th interval,

$T \equiv$ radar update interval or period.

$\alpha, \beta \equiv$ filter weighing coefficients.

Alternatively, these equations can be written as

$$\begin{bmatrix} x_s(k) \\ v_s(k) \end{bmatrix} = \begin{bmatrix} 1-\alpha & (1-\alpha)T \\ -\frac{\beta}{T} & 1-\beta \end{bmatrix} \begin{bmatrix} x_s(k-1) \\ v_s(k-1) \end{bmatrix} + \begin{bmatrix} \alpha \\ \frac{\beta}{T} \end{bmatrix} x_m(k), \quad (5)$$

$$\begin{bmatrix} x_p(k) \\ v_p(k) \end{bmatrix} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_s(k-1) \\ v_s(k-1) \end{bmatrix}. \quad (6)$$

Although these equations are for a one-coordinate model, they can be applied to a three-dimensional target by substituting successively y and z for x in Equations (1) through (4). The filter equations are usually analyzed in one dimension, and the resulting analysis is usually extended to three dimensions with the assumption that this gives similar results.

The first major advance in using the tracking equations was to find a means of optimally selecting a value for β in terms of α . The functional relationship between β and α was derived by Benedict and Bordner³ in the early sixties, using a combination of Z-transforms and variational calculus. Benedict derived an optimal relationship between the coefficients α and β , $\beta = \alpha^2/(2-\alpha)$, the so-called Benedict-Bordner relation. An optimal relationship between alpha and beta based on different assumptions, the Kalata relationship is⁴

$$\beta = 2(2-\alpha) - 4(1-\alpha)^{\frac{1}{2}}. \quad (7)$$

Numerically, the Benedict-Bordner and the Kalata relationships are identical for α less than .4. Kalata defined a variable Γ , known as the tracking index,

$$\Gamma = \frac{T^2 \sigma_v}{\sigma_w} \quad (8)$$

that is a function of the assumed target maneuverability variance σ_v^2 (deviation from modeled behavior) and the radar measurement noise variance σ_w^2 . The relationship between gamma and the filter coefficients is

$$\Gamma^2 = \frac{\beta^2}{1-\alpha}. \quad (9)$$

A convenient way to express relationships involving α and β is to introduce a damping parameter $r = (1-\alpha)^{1/2}$ that is discussed later. For the Kalata relation, the tracking index expressed in terms of r is

$$\Gamma = \frac{2(1-r)^2}{r}, \quad (10)$$

which can be readily solved for r in terms of gamma to give

$$r = \frac{(4+\Gamma) - (8\Gamma + \Gamma^2)^{\frac{1}{2}}}{4}. \quad (11)$$

Once gamma is known, the value of r can be determined using Equation (11), and the values of alpha and beta are fixed for a specified r . Tables of the values of α and β as a function of the range and the tracking index can be computed for storage in a real-time computer program.

Graphs of alpha and beta versus tracking index can be found in Kalata.

Transfer Functions

The transfer functions⁵ of the alpha-beta filter can be used to determine the response of the filter to various measurement models. The transfer function of the smoothed position is obtained by taking the Z-transform of Equation (3) and dividing the smoothed position by the measured position, which gives the result

$$H_x(z) = \frac{X_s(z)}{X_m(z)} = \frac{z[(z-1)\alpha + \beta]}{(z^2 - z(2-\alpha-\beta) + 1-\alpha)} \quad (12)$$

$$= z \left[\frac{A}{z-zp} - \frac{A^*}{z-zm} \right],$$

where,

$$A = \frac{[(zp-1)\alpha + \beta]}{(zp-zm)}. \quad (13)$$

The variables zp and zm are the poles of the transfer function, and A^* is the complex conjugate of A . The transfer function of the smoothed velocity is

$$H_v(z) = \frac{V_s(z)}{X_m(z)} = \frac{z(z-1)\left(\frac{\beta}{T}\right)}{(z^2 - z(2-\alpha-\beta) + 1-\alpha)} \quad (14)$$

$$= \left(\frac{\beta}{T}\right) z \left[\frac{C}{z-zp} - \frac{C^*}{z-zm} \right]$$

where

$$C = \frac{(zp-1)}{zp-zm}. \quad (15)$$

To determine the response to different measurement models, it is necessary to find the poles of the denominators in Equations (12) and (14). The factorization of the transfer functions is obtained by setting the denominators of the transfer functions equal to zero and solving the resulting quadratic equation, which gives two complex roots that are complex conjugates of each other. The two roots can be written in polar form as

$$zp = re^{j\theta}, \text{ and } zm = re^{-j\theta}, \quad (16)$$

where zp corresponds to one root and zm the complex conjugate of that root. Several relationships that can be derived from the above equations simplify subsequent computations:

$$r = (zpzm)^{\frac{1}{2}} = (1-\alpha)^{\frac{1}{2}}, \quad (17)$$

$$\theta = \arctan \left(\frac{\text{Im } zp}{\text{Re } zp} \right) \quad (18)$$

$$= \arctan \left[\frac{(4\beta - (\alpha + \beta)^2)^{\frac{1}{2}}}{2 - \alpha - \beta} \right],$$

$$\cos(\theta) = \frac{(zp + zm)}{2r} = \frac{2 - \alpha - \beta}{2r}, \quad (19)$$

$$\sin(\theta) = \frac{(zp - zm)}{2r} = \frac{(4\beta - (\alpha + \beta)^2)^{\frac{1}{2}}}{2r} \quad (20)$$

$$\beta = r^2(\alpha + \beta) + \alpha - 2\alpha r \cos(\theta), \quad (21)$$

$$\beta = r^2 - 2r \cos(\theta) + 1, \quad (22)$$

where Im means imaginary part and Re means real part of the complex function.

All responses of the transfer functions to specific measurement models will produce terms of a specific form as well as transient terms. The inverse Z-transform, corresponding to transients, can be written as

$$[\dots] = \frac{[\text{Im}(A zp^n)]}{(r \sin(\theta))}. \quad (23)$$

Response of Transfer Functions

The response of the transfer functions of the smoothed position and velocity to various measurement models is now considered. The measurement model for a stationary object is

$$x_m(k) = x_o U(k), \text{ while } X_m(z) = \frac{x_o z}{z-1}. \quad (24)$$

The response of the smoothed position is given by multiplying the transfer function, Equation (12), by the Z-transform in Equation (24), which results in

$$X_s(z) = H_x(z) X_m(z)$$

$$= x_o z \left[\frac{z(\alpha(z-1) + \beta)}{(z-1)(z-zp)(z-zm)} \right] \quad (25)$$

$$= x_o z \left[\frac{A}{z-zp} + \frac{B}{z-zm} + \frac{C}{z-1} \right].$$

Taking the inverse Z-transform of Equation (25), with the intermediate step

$$\text{Im}(A zp^k) = r^{k+2} [\sin(k\theta) - r \sin((k+1)\theta)], \quad (26)$$

gives the smoothed position as a function of time step k as

$$x_s(k) = x_o(1 - x1(k; \alpha)), \quad (27)$$

where

$$x1(k; \alpha) = \frac{r^{k+1}}{\sin(\theta)} [-\sin(k\theta) + r \sin((k+1)\theta)]. \quad (28)$$

The response of the smoothed position to a constant input is the constant plus a transitory term, $x1(k; \alpha)$, that decays to zero for sufficiently large k . The response of the smoothed velocity is given by multiplying the transfer function, Equation (14), by the Z-transform in Equation (24), which results in

$$V_s(z) = H_v(z) X_m(z) = x_o z \left[\frac{z\beta}{T(z-zp)(z-zm)} \right] \quad (29)$$

$$= \frac{x_o z \beta}{T} \left[\frac{A}{z-zp} + \frac{B}{z-zm} \right].$$

Taking the inverse Z-transform of Equation (29), with the intermediate step

$$\text{Im}(A z p^k) = r^{k+1} [\sin((k+1)\theta)], \quad (30)$$

gives the smoothed velocity as a function of time step k as

$$v_s(k) = \frac{x_o}{T} v1(k; \alpha), \quad (31)$$

where

$$v1(k; \alpha) = \frac{\beta r^k}{\sin(\theta)} [\sin((k+1)\theta)]. \quad (32)$$

The effect of a constant input on the smoothed velocity is a term that decays at the rate r^k .

Constant Velocity Target

The measurement model for a constant velocity target with velocity v_o is

$$x_m(k) = v_o T k U(k), \text{ while } X_m(z) = \frac{v_o T z}{(z-1)^2}. \quad (33)$$

The response of the smoothed position is given by multiplying the transfer function, Equation (12), by the Z-transform in Equation (33), taking the inverse Z-transform, which gives the smoothed position as a function of time step k

$$x_s(k) = v_o T (k - x2(k; \alpha)), \quad (34)$$

where

$$x2(k; \alpha) = \frac{r^{k+1} \sin(k\theta)}{\sin(\theta)}. \quad (35)$$

The response of the smoothed position to a linear input is the linear term plus a transitory term that decays to zero for sufficiently large k . The response of the smoothed velocity is

$$v_s(k) = v_o(1 - v2(k; \alpha)), \quad (36)$$

where

$$v2(k; \alpha) = \frac{r^k}{\sin(\theta)} [-r \sin(k\theta) + \sin((k+1)\theta)]. \quad (37)$$

The response of the smoothed velocity to a linear input is a constant plus a transitory term that decays to zero for sufficiently large k . The above transient terms may be used to determine filter settling times for a given alpha.

Linearly Accelerating Target

The measurement model for a linearly accelerating target, with acceleration $x_m(k) = a_o k^2 T^2 / 2$, is

$$x_m = \frac{a_o T^2 z(z+1)}{2(z-1)^3}. \quad (38)$$

The response of the smoothed position is

$$x_s(k) = \frac{a_o T^2}{2} (k^2 - 2 \frac{1-\alpha}{\beta} - x3(k; \alpha)), \quad (39)$$

where

$$x3(k; \alpha) = \frac{r^{k+1}}{\beta \sin(\theta)} [(r^2 - 1) \sin(k\theta) - r \sin((k+1)\theta) + r \sin((k-1)\theta)]. \quad (40)$$

The response of the smoothed position to a quadratic input is the quadratic term plus a transitory term that decays to zero for sufficiently large k plus a constant. This constant is known as a lag, which produces a bias between the filter's estimate of the position and the actual position. The response of the smoothed velocity is

$$v_s(k) = a_o T (k - \left(\frac{\alpha}{\beta} - \frac{1}{2} \right) - v3(k; \alpha)), \quad (41)$$

where

$$v3(k; \alpha) = \frac{r^k}{2\beta \sin(\theta)} [-r^2 \sin((k-1)\theta) + (2r - r^3) \sin(k\theta) + (2r^2 - 1) \sin((k+1)\theta) - r \sin((k+2)\theta)]. \quad (42)$$

The response of the smoothed velocity consists of three terms: the input, a transient term, and a lag term.

The filtering coefficients α and β can be expressed in terms of the parameter r ($1 > r \geq 0$), assuming the Kalata relationship, as

$$\alpha = 1 - r^2 \text{ and } \beta = 2(1 - r)^2. \quad (43)$$

The velocity lag coefficient (bias), L_v , in Equation (41), can be expressed as

$$L_v = \frac{\alpha}{\beta} - \frac{1}{2} = \frac{r}{1 - r}. \quad (44)$$

The magnitude of steady-state velocity bias is therefore $B_v = (a_0 T)L_v$. The position coefficient, L_x , in Equation (39), can be expressed as

$$L_x = \frac{(1 - \alpha)}{\beta} = \frac{r^2}{2(1 - r)^2}. \quad (45)$$

The magnitude of steady-state position bias is $B_x = (a_0 T^2)L_x$. Note that $L_x = 1/2 L_v^2$. These lags are graphed in Figure 1 for alpha between .1 and .8. From the definition of the steady-state bias, a steady-state velocity lag time τ_s can be defined as

$$\tau_s = \frac{B_v}{a_0} = \left(\frac{\alpha}{\beta} - \frac{1}{2} \right) T, \quad (46)$$

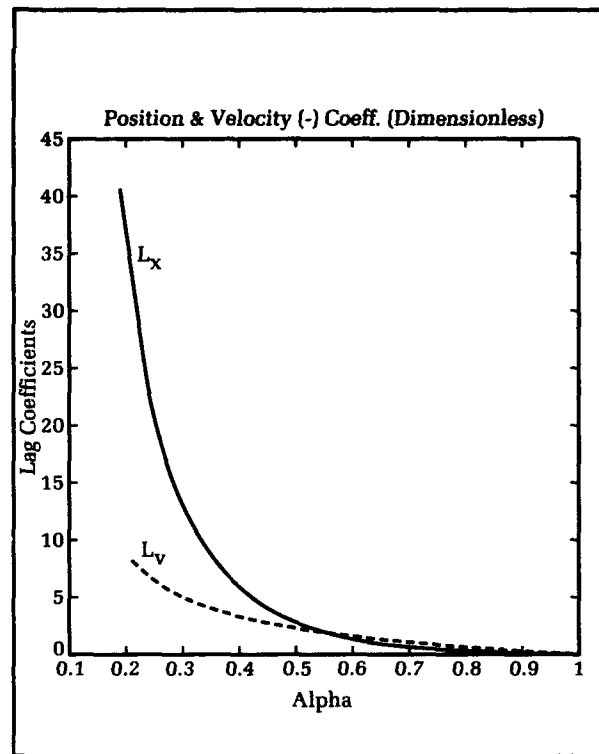


Figure 1. Steady-state position and velocity lags for linearly accelerating target.

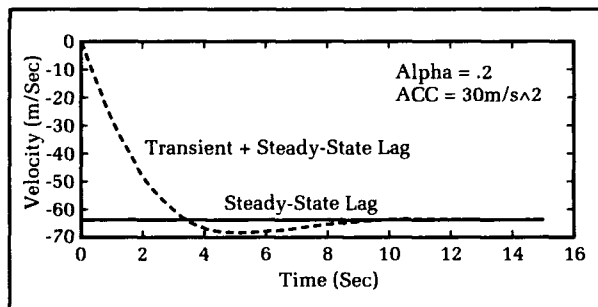


Figure 2. Transient velocity response ($\alpha = .2$) for linearly accelerating target.

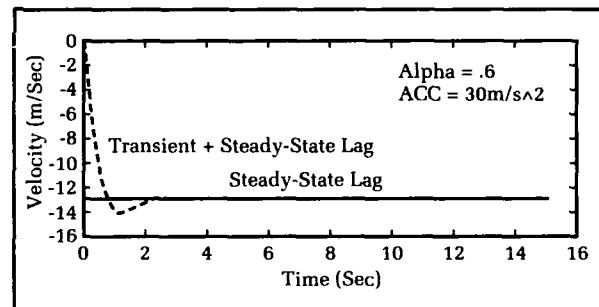


Figure 3. Transient velocity response ($\alpha = .6$) for linearly accelerating target.

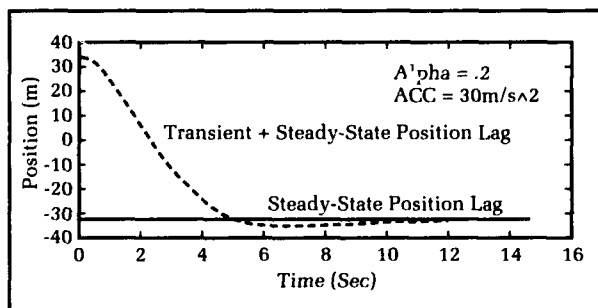


Figure 4. Transient position response ($\alpha = .2$) for linearly accelerating target.

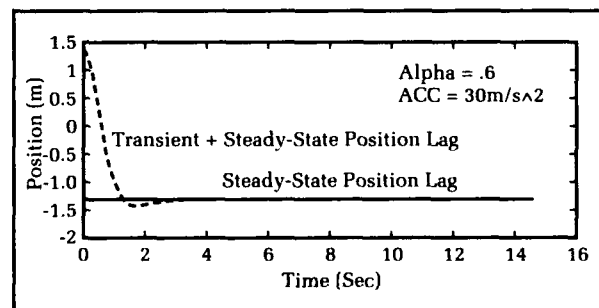


Figure 5. Transient position response ($\alpha = .6$) for linearly accelerating target.

and likewise a transient velocity lag time τ_T can be defined as

$$\tau_T = \frac{B_{Tv}}{a_0} = v_3(k; \alpha)T. \quad (47)$$

So, for an acceleration input,

$$v_s(k) = a_0 kT - a_0(\tau_s + \tau_T); \quad (48)$$

the velocity estimate lags the true velocity by $\tau_s + \tau_T$ seconds. Similar results apply to the position estimates. Note that $v_3(0, \alpha) = -L_v$. Graphs of the velocity response are shown in Figures 2 and 3 for $\alpha = .2$, and $.6$, respectively. A similar observation can be made about $x_3(k, \alpha)$, namely that $x_3(0, \alpha) = -2L_x$. Graphs of the position response are shown in Figures 4 and 5 for $\alpha = .2$, and $.6$, respectively.

Weaving Target

For a weaving measurement model

$$x_m(kT) = Q_o \sin(\omega kT) \quad \text{and} \quad (49)$$

$$X_m(z) = \frac{Q_o \sin(\omega T)}{(z - e^{j\omega T})(z - e^{-j\omega T})},$$

the response of the smoothed position consists of the sum of two terms, a steady-state term $x_{ss}(kT)$ and a transient term $x_T(kT)$, i.e., $x_s(K) = x_{ss}(kT) + x_T(kT)$. For both the transient and steady-state terms ($\alpha' = \beta - \alpha$) the term *den* is defined as

$$\begin{aligned} \text{den} = & (r^2 - 2r \cos(\theta + \omega T) + 1) \\ & (r^2 - 2r \cos(\theta - \omega T) + 1), \end{aligned} \quad (50)$$

The transient term is

$$\begin{aligned} x_T(k) = & \frac{Q_o r^k \sin(\omega T)}{\sin(\theta) \text{den}} \\ & [C_1 \sin((k-1)\theta) + C_2 \sin(k\theta) \\ & + C_3 \sin((k+1)\theta) + C_4 \sin((k+2)\theta)], \end{aligned} \quad (51)$$

where $C_1 = \alpha' r^2$, $C_2 = \alpha r^3 - 2\alpha' r \cos(\omega T)$, $C_3 = \alpha' - 2\alpha r^2 \cos(\omega T)$, and $C_4 = \alpha r$. The steady-state response is

$$\begin{aligned} x_{ss}(k) = & \frac{Q_o}{\text{den}} [D_1 \sin((k-1)\omega T) \\ & + D_2 \sin(k\omega T) \\ & + D_3 \sin((k+1)\omega T) \\ & + D_4 \sin((k+2)\omega T)], \end{aligned} \quad (52)$$

where $D_1 = \alpha'$, $D_2 = \alpha - 2\alpha' r \cos(\theta)$, $D_3 = \alpha' r^2 - 2\alpha r \cos(\theta)$, and $D_4 = \alpha r^2$. The smoothed velocity

also consists of a transient $v_T(k)$ and a steady-state term $v_{ss}(k)$, with the overall smoothed velocity being $v_s(k) = v_T(k) + v_{ss}(k)$. The transient term for the smoothed velocity is

$$\begin{aligned} v_T(k) = & \frac{Q_o r^k \beta \sin(\omega T)}{T \sin(\theta) \text{den}} [f_1 \sin((k-1)\theta) \\ & + f_2 \sin(k\theta) + f_3 \sin((k+1)\theta) \\ & + f_4 \sin((k+2)\theta)], \end{aligned} \quad (53)$$

where $f_1 = -r^2$, $f_2 = r^3 + 2r \cos(\omega T)$, $f_3 = -(1 + 2r^2 \cos(\omega T))$, and $f_4 = r$. The steady-state response of the smoothed velocity is

$$\begin{aligned} v_{ss}(k) = & \frac{Q_o \beta}{T \text{den}} [g_1 \sin((k-1)\omega T) \\ & + g_2 \sin(k\omega T) + \\ & g_3 \sin((k+1)\omega T) + g_4 \sin((k+2)\omega T)], \end{aligned} \quad (54)$$

where $g_1 = -1$, $g_2 = 1 + 2r \cos(\theta)$, $g_3 = -(r^2 + 2r \cos(\theta))$, and $g_4 = r^2$. The steady-state terms for the smoothed position and velocity are graphed in Figures 6 through 9 for $\alpha = .2$ and $.6$, $T = 1/4$ sec, and $\omega = .8$ rad./sec.

Prediction Response

Though the responses of the smoothed position and velocity transfer functions have been considered, the preceding analysis can be applied to the predicted position and velocity as well. The transfer function of the one-step-ahead predicted position is

$$H_{px}(z) = \frac{(\alpha + \beta)z - \alpha}{(z - zp)(z - zm)}. \quad (55)$$

The transfer function of the one-step-ahead predicted velocity is

$$H_{pv}(z) = \frac{\beta(z-1)}{T(z-zp)(z-zm)}. \quad (56)$$

Derivation of these transfer functions follows by taking the Z-transform of Equation (6) and noting that the predicted transfer functions are $1/z$ times the smoothed transfer functions. An example of an application of these transfer functions may be seen by determining the lags in the predicted position and velocity due to a constant acceleration model, Equation (38). The steady-state predicted bias lags are found to be

$$L_{px} = \frac{1}{\beta} = \frac{1}{2(1-r)^2}, \quad (57)$$

$$L_{pv} = \frac{(\beta + 2\alpha)}{2\beta} = \frac{1}{1-r}. \quad (58)$$

Note that $L_{px} = 1/2 L_{pv}^2$.

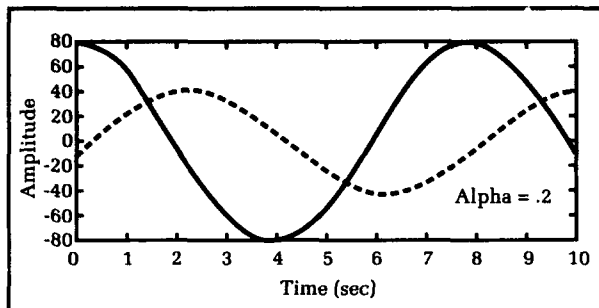


Figure 6. Velocity and filtered velocity for weaving model ($\alpha = .2$).

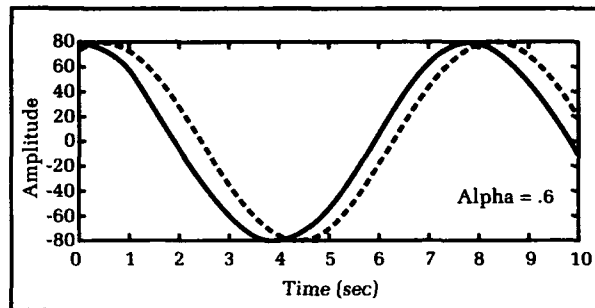


Figure 7. Velocity and filtered velocity for weaving model ($\alpha = .6$).

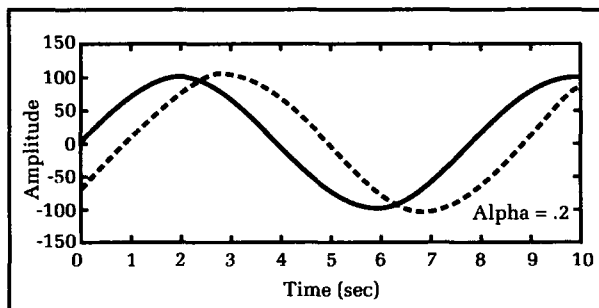


Figure 8. Position and filtered position for weaving model ($\alpha = .2$).

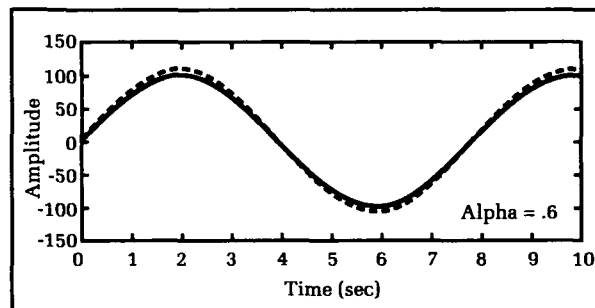


Figure 9. Position and filtered position for weaving model ($\alpha = .6$).

Conclusions

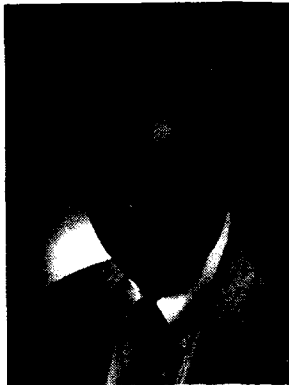
We have demonstrated the application of the factorization of the transfer functions of the alpha-beta filter to four different measurement models, which enabled us to determine exact forms for the output of the filter. The methods reported here are a simplification and unification of scattered results for different models reported in the literature. Transient and steady-state performance measures can be determined from the derived responses. These measures are used to quantify the alpha-beta filter's performance against different measurement models

representing a target's trajectory. These response characteristics can be used to determine filter parameters, to achieve desired performance, or to quantify performance against a given threat's trajectory characteristics. The response characteristics of the filter can then be used to quantify performance against a given threat's trajectory profile. This enables determination of system-level performance against various threat profiles, which has application to both tracking and weapon system engagement capabilities. This work is a unification and simplification of previously reported work⁵ that dealt with using the output of alpha-beta filters for kill assessment.

References

1. Bogler, P. L., *Radar Principles with Application to Tracking Systems*, John Wiley & Sons, 1990.
2. Gelb, A., *Applied Optimal Estimation*, Cambridge, MA, MIT Press, 1974.
3. Benedict, R. T. and Bordner, G. W., "Synthesis of an Optimal Set of Radar Track-While-Scan Smoothing Equations," *IRE Transactions on Automatic Control*, AC-7, 1962, pp. 27-32.
4. Kalata, P. R., "The Tracking Index: A Generalized Parameter for α - β and α - β - γ Target Trackers," *IEEE Transactions on Aerospace and Electronics Systems*, AES-20, Mar 1984, pp. 174-182.
5. Gray, J. E. and Murray, W. J., *Analysis of Alpha-Beta Filters for Potential Applications in Anti-Air Kill Evaluation Improvements and as Acceleration Detectors*, Naval Surface Warfare Center, NSWC TR 89-269, Feb 1990.

The Authors



JOHN E. GRAY was born on 28 August 1955 in Fort Lee, Virginia. He received B.S. degrees in physics and mathematics from the University of Mississippi in 1977 and an M.S. in physics in 1980. He also has studied physics and electrical engineering at Virginia Tech and Catholic University. He is a member of IEEE, the American Physical Society, Society for Industrial and Applied Mathematics, and Sigma Phi Sigma. He is

employed by the Naval Surface Warfare Center, Dahlgren Division, working in the areas of tracking filters, signal analysis, and electromagnetic theory applied to radars. His other areas of interest are wave phenomena, foundations of quantum mechanics, and nonlinear mathematics. He has authored over 40 technical publications.



WILLIAM MURRAY has been a member of the Weapon Control System branch at the Naval Surface Warfare Center, Dahlgren Division since 1985, after receiving B.S. and M.S. degrees in applied mathematics from the University of Lowell. Mr. Murray has been involved in system-level simulation and weapon control system upgrades to the AEGIS system. His areas of interest include track processing, track filtering, and guidance algo-

rithms for surface-to-air missiles. He is currently involved in missile developments associated with enhancing the anti-air warfare capabilities of the AEGIS system.

A Digital Signal Processor for Improved EMS Bearing Accuracy

William A. Masi

A processing scheme has been developed to increase the direction-finding accuracy and signal-detection sensitivity of the AN/SLQ-32's receiver. This procedure can yield direction-finding accuracies on the order of tenths of degrees and has the ability to process signals with a signal-to-noise ratio (SNR) of less than 0 dB at the crystal video receiver output. Other, more elaborate techniques could increase these performance parameters even further. However, the main purpose here is to show that such enhancements are possible solely through video processing without changes to the electronic hardware that processes the radio frequency (RF) signal. A simple averaging and correlation method will serve as a basis for comparison for other processing schemes. In addition, the introduction of digital signal processing will allow the implementation of algorithms to detect and characterize emitters that are presently considered low-probability-of-intercept (LPI), as well as to reduce multiple-emitter interference.

Introduction

The surface Navy's principal electronic warfare system, the AN/SLQ-32, with well over 300 units deployed on U.S. and allied combat ships, was designed in the early 1970s as a stand-alone Antiship Missile Defense System (Figure 1). The AN/SLQ-32V(1) provides early warning, direction-finding of incoming radar-guided missiles, and identification. An improved version such as the V(2) provides the additional capability of detecting those radars used to target and launch the missiles. The AN/SLQ-32V(3) adds a jamming capability to prevent or delay launch of antiship missiles.

The AN/SLQ-32 employs two multiple-beam antennas for receiving radar emissions and a lens-fed, multiple-beam array. Each antenna consists of an array of elements providing a set of high-gain beams, all existing simultaneously. The system senses RF signals at all azimuths to detect radar-guided antiship missiles and supporting radars. Identification is accomplished by comparing observed signal characteristics with parameters in a library of threat emitter characters. The electronic support measure (ESM) alerts the ship to an impending attack and controls the launch of ship countermeasures such as the MK 36 chaff rocket system.

Changes in threat scenarios and defense tactics, introduction of new U.S. weapon systems, and the increasing need to integrate shipboard sensors have placed additional demands on AN/SLQ-32. Two of these requirements involve increasing the system's detection range and providing improved bearing accuracy

on RF emitters. This article describes an approach that may well provide these improvements and establish a technological base for additional requirements such as emitter identification and interference rejection.

Approach

The main element in the direction-finding receiver is an RF lens-horn array, generally called a bootlace lens. Figure 2 shows the geometric optics ray-tracing schematic for the lens. The operation of the lens is similar to that of an optical lens.

In Figure 2 a plane wave front is seen coming from some angle of arrival (AOA). The lower part of the wave will be picked up first by the lower horn antenna and carried, through a given length of cable, to the lower feed point of the lens. The wave will then be reradiated in a circular pattern, according to Huygen's Law, through the lens. A short time later, the wave front will be picked up by the next horn in the array and subsequently reradiated from the second feed point of the lens. This process is repeated sequentially for all the array elements. The result is that the wave front is reradiated with some incremental delay from all the feed points of the lens. Now there is also a delay between the time the wave front is launched and the time it is picked up on the other side of the lens. This delay depends on what particular point on the pick-up side is chosen. However, there will be one point where the delay due to the lens transit time is complimentary to the delay due to the AOA. At this point all the parts of the wave front that were sampled come together and are focused. The amplitude here will then be a maximum. At other points the

phases will not match and the amplitude will be reduced. In the case of the AN/SLQ-32, the 35 outputs from the 35-element linear horn array are fed to the 35 inputs of lens' input arc. The output arc of the lens is sampled at 17 uniformly spaced points. These RF outputs are then fed to 17 crystal video receivers. The procedure outlined in this article uses the video output of these receivers.

A model of a bootlace lens and antenna array was first developed as a baseline configuration for the source of the array data used in the processing algorithm. The version of the bootlace lens actually used in the AN/SLQ-32 is called a Rotman lens. Although the model developed here is not that of a true Rotman lens in the sense that it is much more frequency-dependent, it can represent the energy distribution very well if the wavelength is chosen properly. Since we are interested in the energy distribution, this model will be adequate. The derivation of the model is straightforward but lengthy¹ and will not be presented here.

Digital Signal Processing Procedure to Determine Bearing Angle

Referring to Figure 2, the envelope of the electric field values at an output port, p , due to a wave with incident angle α , is equal to the sum of electric fields from all the input ports, taking into account the phase differences due to the variations in path lengths. However, there is one wave, the one with angle of incidence ($\alpha = 0$) which will, for all input ports, have equal path lengths to the center of the output feed arc, since this point is also the center of rotation for the input feed arc. Taking this length, denoted by L as the normalizing length, an expression propor-

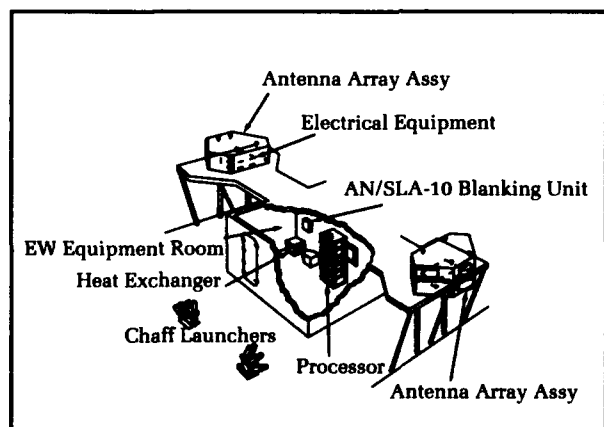


Figure 1. AN/SLQ-32V(2) shipboard installation.

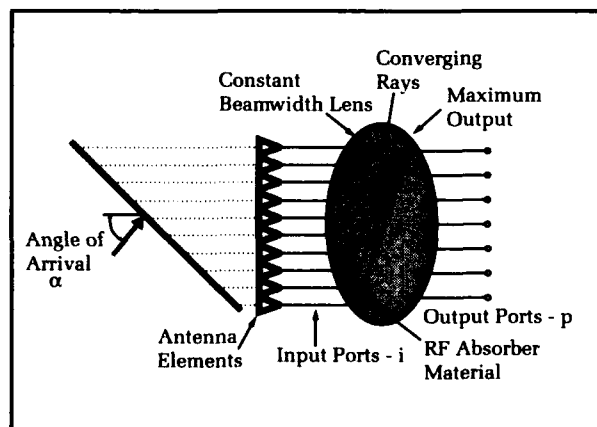


Figure 2. Bootlace lens antenna—direction-finding system.

tional for the resultant normalized electric field at port p due to a wave at incident angle can be written:

$$e(p, \alpha) = (1/N) \sum_i^N \cos[2(\pi/\lambda)(L(i, p, \alpha) - L)],$$

where N = total number of input ports.

The internal energy distribution pattern for the Rotman lens used in the AN/SLQ-32 has a beam width of ~6 degrees. To approximate this value for this model, we let the wavelength $\lambda/L = 0.3$.

If a perfect square-law detector, i.e., one whose output is proportional to its input power, is used to detect this signal, then the output would be given by

$$E(p, q) = e(p, q)^2,$$

where the constant of proportionality is set to 1.

Now let M denote the matrix of all possible energy patterns;

$$M = [E(p, q)]_m^n.$$

Then M represents an $n \times m$ matrix where the element $E(p, q)$ represents the p th output port and the q th angular resolution cell for the sector covered by the lens array. For this example, we will use a lens array that covers ± 45 degrees, has 17 output ports, and will have the incident angle resolved to angle cells of 0.9 degree. The matrix M will then be a 17×101 array. To normalize

this data set in the sense that $|E(p)| = 1$ for each q , we proceed as follows:

$$G^{<q>} = \frac{M^{<q>}}{|M^{<q>}|},$$

where $G^{<q>}$ is the q th column of the normalized matrix G .

Figure 3 shows an example of a 3-D surface plot of $E(p, q)$ for some arbitrary output port for all 101 angles of incidence; Figure 4 shows a similar plot for all 17 output ports.

Reducing Path Difference Error

Rotman and Turner² introduced a refocusing procedure to reduce the path difference error. This is done by choosing another angle of incidence, say α_1 ; then for this angle adjusting the separation of the antenna elements and the values of the delays to obtain perfect focusing at some port p as illustrated in Figure 5. This is done so as to maintain symmetry about the central axis and also equal path lengths for the normal wave, $\alpha_1 = 0$. Since symmetry is maintained, the process will also cause perfect focusing at incident angle $-\alpha_1$. This forcing of the lens to have a perfect focusing at $\alpha_1, 0, -\alpha_1$, greatly reduces its dependence on the wavelength.

Since the 17 output ports cover a 90-degree sector, their angular spacing is ~5.3 degrees.

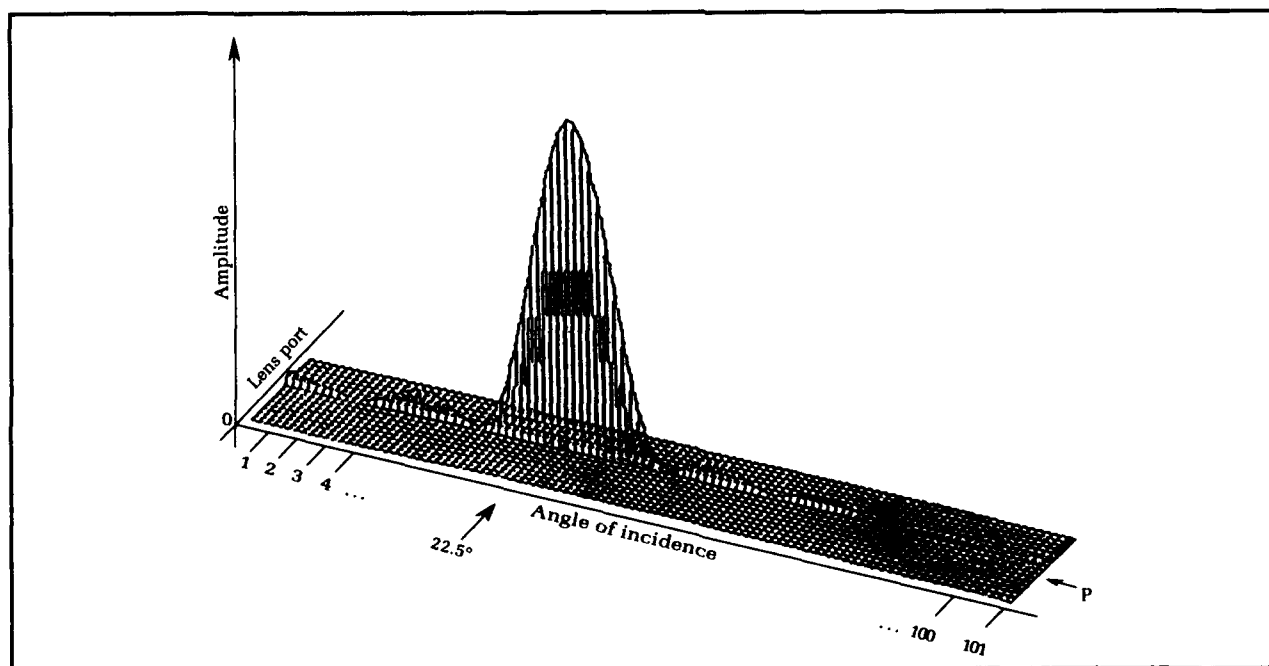


Figure 3. Energy vs. angle of incidence for p th port.

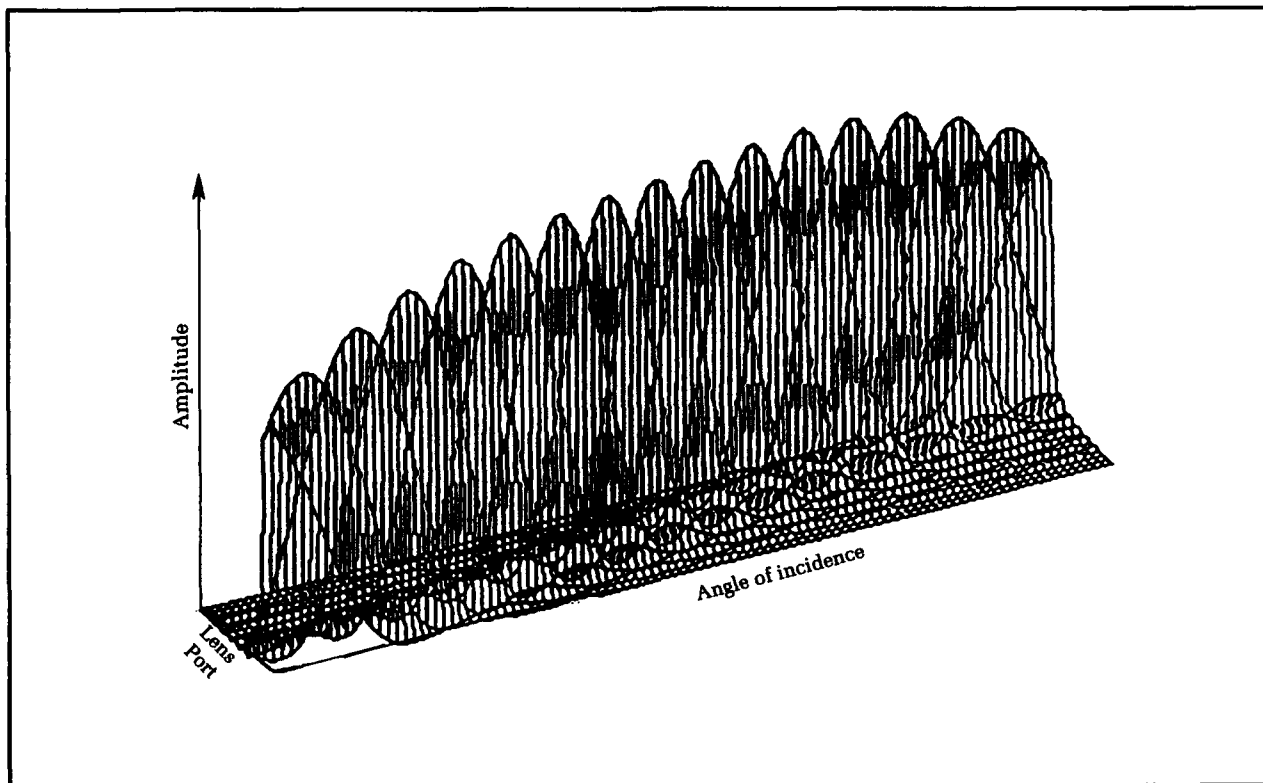


Figure 4. Energy vs. angle of incidence for all ports.

Since the wave field chosen has a granularity of 0.9 degree, the output from the ports represents the wave field sampled at 5.3-degree intervals. What we wish to do, therefore, is to interpolate the elements of the 0.9-degree wave field based on the 5.3-degree sampled data. This can be done by correlating the normalized wave field pattern matrix G with a known, normalized sampled data set. A procedure for doing this is given in the following example.

Choose some arbitrary element of the wave field, e.g., an incident wave at 22.5 degrees, as the data set. Then, denoting this set by a vector F

$$\text{where } F = [f_1, f_2, f_3, f_4 \dots f_p \dots f_N]^T,$$

$$\text{and } F = G^{<25>},$$

where f_p is the element of the set F from port p .

Next, the pattern sequence that the data should correlate to must be formed; however, this is just the wave field matrix G of all possible incident waves. Therefore, a correlation matrix, C , can be defined as

$$C = G^T F.$$

This is graphically depicted in Figure 6 using a matrix notation.

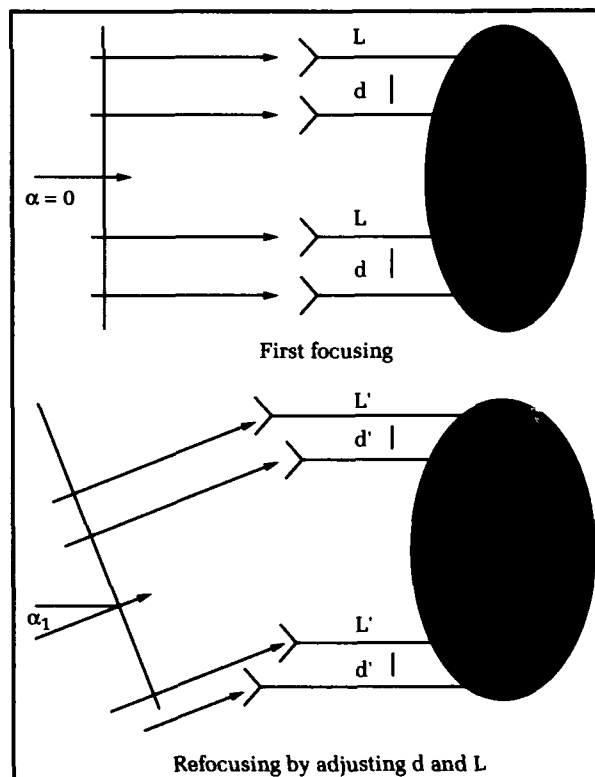


Figure 5. Rotman procedure for reducing path difference error.

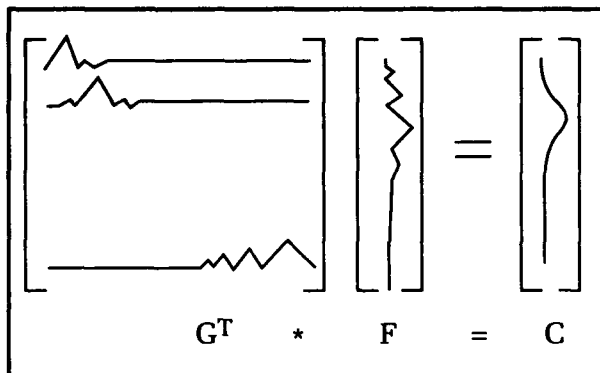


Figure 6. Graphical illustration of matrix correlation.

The data set, f_p , for the chosen incident angle and its corresponding correlation matrix element, C_q , values are plotted in Figures 7 and 8, respectively.

Let the maximum element of the matrix C be defined as $\max(C)$. Then the element index corresponding to $\max(C)$ equals the value of the index for the angle of incidence. Using the program developed by the Dahlgren Division,¹ the following results were obtained for the pertinent elements of the matrix C .

$$C_{24} = 0.98791 \quad C_{25} = 1 \quad C_{26} = 0.98953.$$

Therefore, the element index corresponding to $\max(C)$ is 25. This is the index for the wave field element with a 22.5-degree incident angle. This example illustrates the basic idea of using the correlation between a measured set of amplitudes from a bootlace lens and a reference set of amplitudes for all AOAs to determine the AOA corresponding to the measured data set. We note again that this is just one procedure, used here to illustrate the nature of digital signal processing.

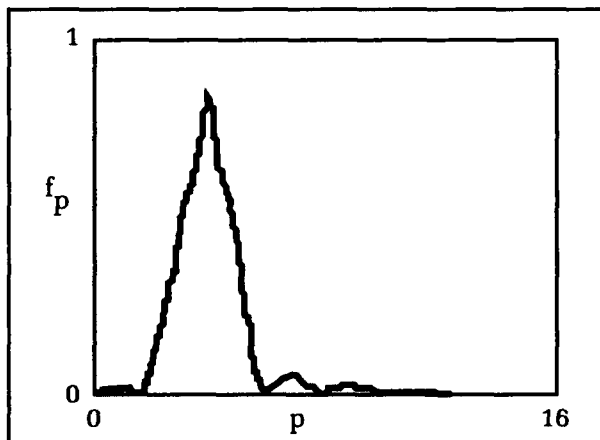


Figure 7. Data set f_p vs. port number p .

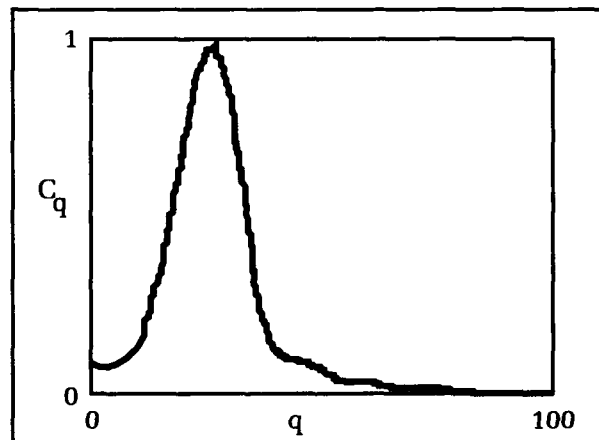


Figure 8. Correlation matrix element C_q vs. q .

Effect of Noise on Direction-Finding

The next step is to add noise to the data. This noise will contribute to the error of the measured energy distribution. If the statistical properties of the noise are known, then this error can be reduced through signal processing. However, to apply signal processing based on statistics, a set of data large enough to provide meaningful statistics must be obtained. Also, enough sampling must be done so that the signal content is high enough to affect the statistics of the collected data set. These two parameters, i.e., sample size and sample rate, are constrained by several factors. These include both hardware and software limitations. For this reason the values chosen for the following examples are very conservative, hence easily obtainable in practice; yet they nevertheless illustrate the advantages of this approach.

This example will illustrate the case when the signal contains noise. What is done here is to add noise to the signal data of the previous example. It must be noted that there is really only one source of noise in this type of system: the internal noise generated by the crystal video receiver. This is so because even though there may indeed be noisy RF sources outside the system, the wave fronts from these will impinge the array as a coherent wave, and so will be treated as a true signal which the system is designed to process. We must emphasize at this point that this is the key reason why this approach should be capable of detecting LPI emitters; i.e., detection is based on the fact that the signal appears at the receiver as a coherent wave front which a multichannel array processor can distinguish from single-channel noise. If the occurrence of other waves is considered to be a problem

because they occur when a desired wave front is also present, then the nature of the problem is defined to be one of simultaneous signal discrimination. We must note that signal discrimination is a definite requirement for any real-world application, and thus is an integral part of this project.

Because there are 17 separate crystal video receivers, all 17 channels are independent. Assuming they are also the same, we can then assume that the noise sources are independent and identically distributed. This was implemented in the software by using the random number generator routine provided with MATLAB (a mathematical software package marketed by MathWorks, Inc.). The noise was also specified to have a normal distribution. The principal processing scheme was, as in the noiseless case, correlation with reference values. In order to reduce the errors that the noise would now obviously introduce into the correlation process, a noise reduction process must also be included. Because it is easy to implement and can also be used as a standard to judge other, more complicated noise reducing methods, a moving average process was chosen to be the initial noise reducing filter.

When noise is added and averaging is used as a means to reduce its effects, the sample size becomes an important parameter, and so must be specified. Obviously, the larger the sample size, the larger the amount of information available on the signal, and so the noise reduction will be better. Therefore, the nature of the problem is such that the amount of data available will determine system performance. The principal limiting factor will then be the sample rate.

With a noiseless signal, the correlation can be carried out to arbitrarily small-angle increments and the detection sensitivity to arbitrarily low signal levels. The presence of noise will, of course, reduce direction-finding accuracy and sensitivity. In order to determine the manner and the degree to which these performance characteristics are affected, the analysis program written as an M-file for MATLAB was developed.¹ The program was run using various angular resolution cell sizes. Also, the data set size and the filter length were changed to represent various types of LPI emitters. Only one type will be presented here; the others are discussed elsewhere.¹

The total data sample size was set to 100 samples; 50 of these contained signal data. The moving average filter is a zero-phase, forward/reverse digital filter function provided with MATLAB. Its length was set to 50. When the moving average filter is the only processing done, this length sets the performance limits of the system. This is because the signal level in a data set having less than 50 samples containing signal data will be reduced by

the averaging process. On the other hand, a data set having more than 50 samples containing data will not see any improvement in SNR since the moving average window has been set to 50. The program was written to generate several families of performance measures;¹ one is plotted in Figure 9. Five groups of curves are given, with each group generated for different angular resolutions (9, 4.5, 2.25, 1.8, 0.9, 0.45, 0.225, and 0.18 degrees). Only the one for 0.9-degree resolution is given in this article. Each four-plot family of each group used different noise sets. First, a set of noise values was produced. This noise was characterized by having mean = 0 and sigma = 1. To this noise set successive incremental values of the signal were added to produce successive data sets having increasing SNR. These sets were processed to produce a single four-plot family of curves. For each family a new noise set was generated.

The first type of graph simply plots the angle error as a function of input SNR. The difference in each plot is due to the fact that different noise sets were used. This is significant since it implies that even though the noise in each case has the same statistics, i.e., Gaussian with mean

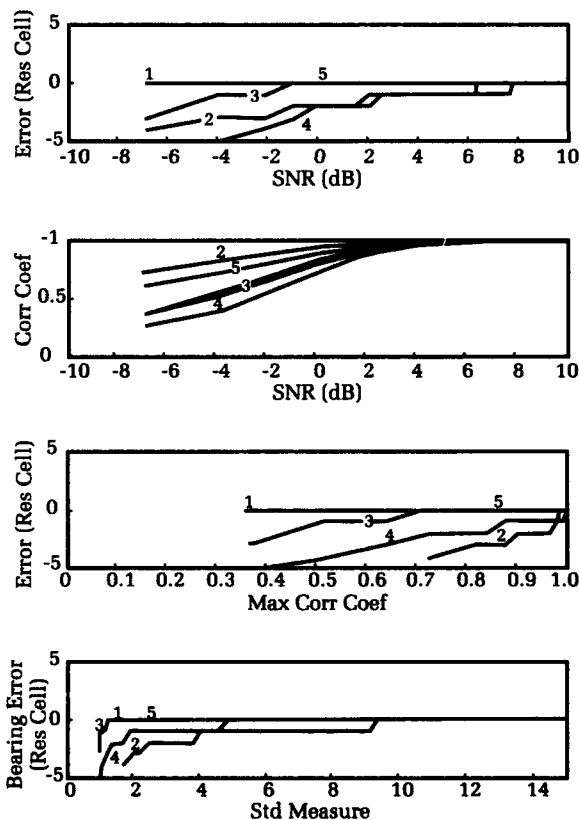


Figure 9. Bearing error plots for resolution cell = 0.9 degree.

$= 0$ and $\sigma = 1$, the actual values can differ in such a way that different data sets or equivalently different sets of measurements will give different answers for the same SNR. This is not surprising, since the parameters mean and sigma do not restrict the sample values, but only give ensemble properties. This means that a measure of confidence must also be provided to support the angle measurement. The ERROR ANALYSIS ROUTINE included in the program accomplishes this confidence measure by providing what is arbitrarily called the STANDARD MEASURE, denoted by STD MEASURE. The function STD MEASURE is plotted as one of the graphs in the family. STD MEASURE essentially gives the amount of corruption in the measured correlation vector. It is based on the correlation coefficient energy equation.³ Significantly, because it is a function of the reference values and the measured data, it does not depend upon inde-

pendent knowledge of the noise in the "signal-free" case. STD MEASURE is used as a quality indicator. Specifically, what is done is to take the envelope of the maximum STD MEASURE for the various angle resolution cells and use it as the resolution determining function. This envelope is plotted in Figure 10. If, for example, a set of data yields an angle of arrival of 10 degrees and a STD MEASURE of 4 or 5, then the bearing would be reported to a resolution cell of about 0.9 degree (± 0.45 degree). If the STD MEASURE were 12 or 14, then the measurement would be reported to a resolution cell of about .2 degree (± 0.1 degree). With good SNR, a factor of 20 improvement in bearing accuracy is anticipated. We must note that these numbers are based on a simple computer model and that the standard measure of confidence, usually expressed in terms of the standard deviation, has not been given. The reason is because, in practice, neither

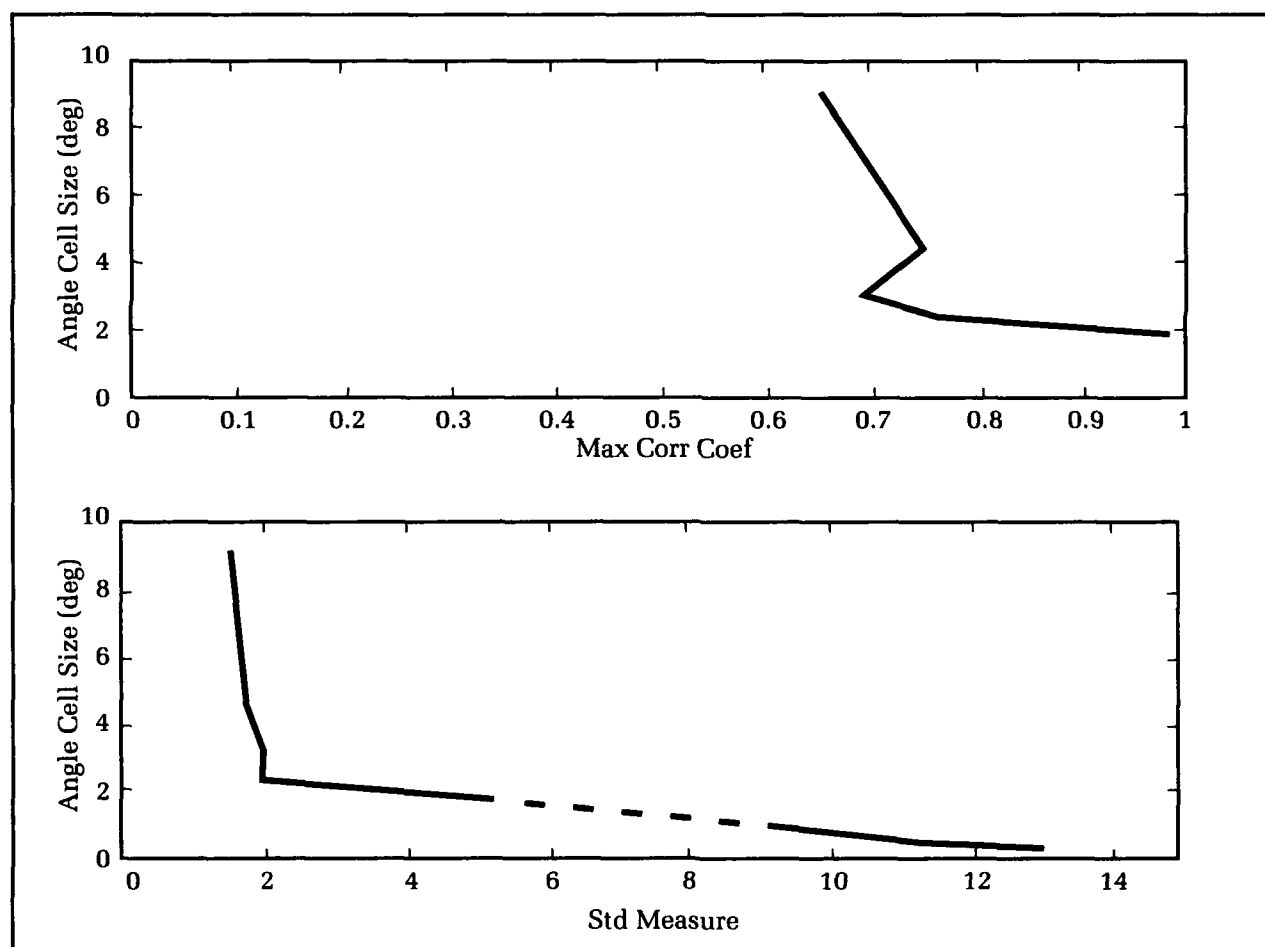


Figure 10. Angle cell size vs. maximum correlated coefficient and standard measure.

the actual distribution of the noise nor the *a priori* SNR will be known. A Monte Carlo simulation could be done, but would be of questionable value for the same reasons. A reliable confidence measure will have to wait until sufficient test data are available so that the data can be classified into the sets.

The plot MAX CORR COEF vs. SNR measures how closely the processed data values correspond to the predetermined reference values as a function of SNR. Note that the correlation coefficients themselves, at least for the high angular resolution case, are not the best confidence measure. There are two reasons for this. First, although the correlation between the measured set and the reference set may be high, the statistics of the noise may be such as to make this high correlation itself coincidental. Second, the measure of confidence is a measure of confidence to the specified resolution. When the resolution cell is small, a corresponding small change in the data set, hence also in the correlation coefficient, will put the result in or out of the correct angle cell, greatly lowering the actual reliability of the reported answer. The STD MEASURE is a better indicator, since it incorporates both the correlation coefficients and the variance of the data.

Detection of LPI Emitters

The relation between the above example and LPI emitters is as follows. We define an LPI emitter as one whose parameters fall outside those of a given receiver system. If a receiver is designed to detect the generic characteristics of an emitter, i.e., those parameters that are present due to the fact that the emitter generates an electromagnetic wave, then there would not be any LPI emitters for that receiver. Now, for the lens array system, any plane electromagnetic wave, regardless of any other distinguishing characteristics (e.g., modulation, frequency, phase coding) impinging the array will manifest itself as an increase in the coherence or correlation between the output of the lens array. Detection of this increase in coherence results in detection of the emitter. In the absence of the impinging plane wave, the outputs of the lens elements and the crystal video receivers are independent and identically distributed.

Three principal types of emitters are considered to be the leading candidates for LPI radars. These are listed in Table 1. The lens array receiver, using modern signal processing tech-

niques, should detect all of these. For the low-power, continuous waveform emitters, the radar must illuminate the target for a relatively long time for its own receiver to collect enough reflected energy to detect the target. By optimizing the detection scheme for particular classes of low-power LPIs, the system's sensitivity can be increased to approach that of a matched filter. In the above example, this would correspond to changing the sample size and the moving average filter length. Also note that for a digital system, many schemes can either be run simultaneously or the same data can be reprocessed several times in different ways.

The stochastic emitter may require an additional level of analysis using higher order statistics. A detailed explanation of the principles involved is given by Giannakis and Tsatsanis.⁴

Table 1. LPI Emitter Types

Low Power, Continuous Waveform
High Power, Single Pulse
Stochastic, Wide Band

Simultaneous Signal Discrimination

When addressing the issue of simultaneous signal discrimination, there are two cases to consider. The first occurs when the undesirable signal is known; for instance, when the electronic countermeasures (ECM) system's reflections off nearby objects is seen by the ESM receiver. The second case occurs if more than one desired intercept occurs at the same time. A distinction must be made, however, between simultaneous signals and overlapping signals. Simultaneous signals begin at precisely the same time and end at precisely the same time. Overlapping signals' start and/or end times are not the same, and only a portion of one or both signals occurs simultaneously. This is the most likely event for non-continuous waveform signals.

Signal cancellation in the time domain can be used, but is difficult to implement, especially at gigahertz frequencies, and it is hardware intensive. A better way is to transform the data to an orthogonal basis where extraneous and redundant information is eliminated and where the essential elements of the data can be classified and sorted.⁵ This may be done in a number of ways. The optimum method will depend upon the level of discrimination required, the amount of *a priori* information available, and the para-

meters of the signals (specifically, on their changeability and the degree of correlation among them). A prime candidate that seems most applicable to ECM/ESM is the Fourier transform. Consider the processing scheme depicted in Figure 11. Let the two overlapping signals be s_1 and s_2 , as shown in Figures 12 and 13, respectively. Also, let s_1 be the signal that must be removed, hence it is also the one that is known. Since s_1 is known, its Fourier transform can be determined by means of a fast Fourier transform (FFT) algorithm. This is shown in the FFT, s_1 block in Figure 11; its FFT is given in Figure 14. Since the output of the receiver is the sum of the two signals, an FFT of the output, shown in Figure 15, will be the sum of the FFTs of the two signals. Now unless the two signals

are identical, their spectra (FFTs) will be different. If the spectrum of the unwanted signal is subtracted from that of the sum, assuming they have been normalized properly, then the result (Figure 16) will be the spectrum of the desired signal. If the inverse FFT of this spectrum is taken, then the desired signal will result. Figure 17 shows that this is indeed the case.

Use of Neural Nets in Direction-Finding

We note that the Fourier transform is only one of many orthogonal bases that can be used to analyze the data sets. The main advantage is that hardware and software are available to compute the required coefficients rapidly and efficiently.

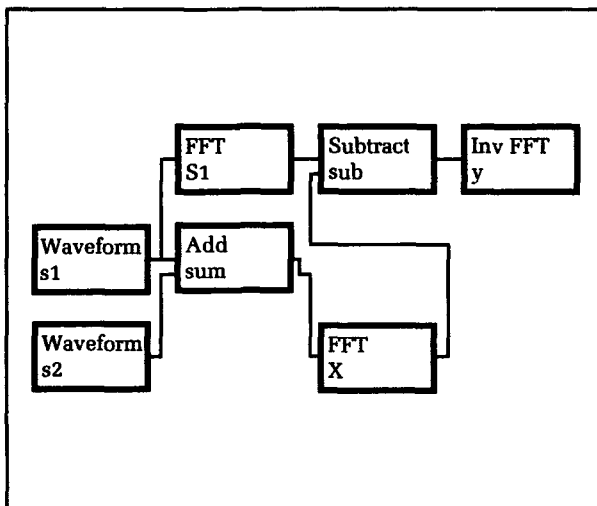


Figure 11. ECM isolation processing.

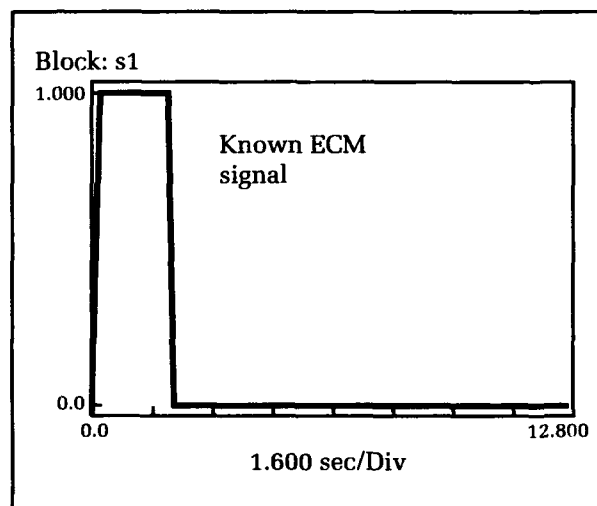


Figure 12. Signal s_1 .

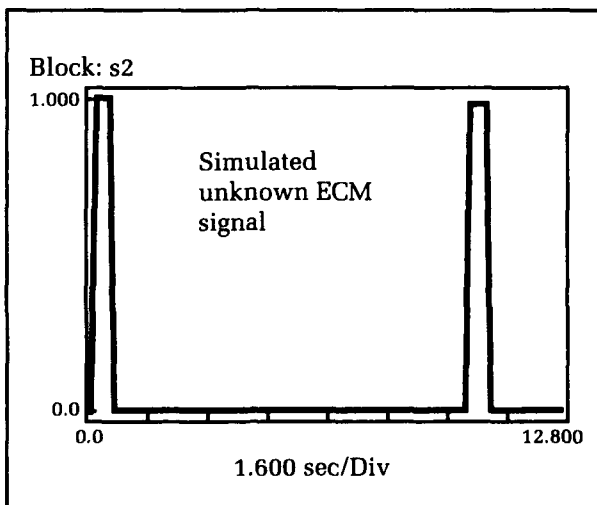


Figure 13. Signal s_2 .

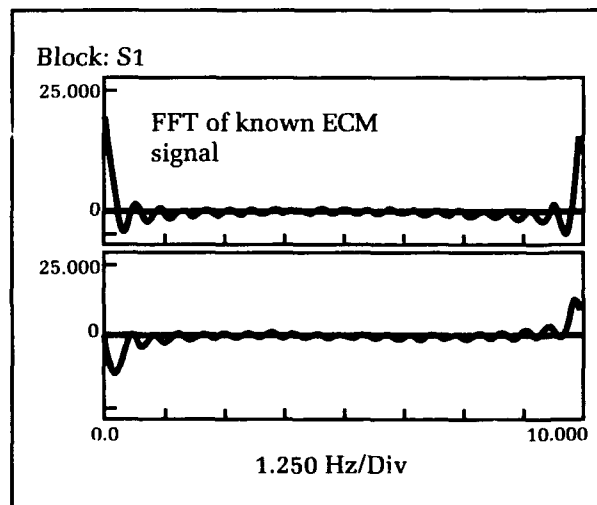


Figure 14. Spectrum of s_1 .

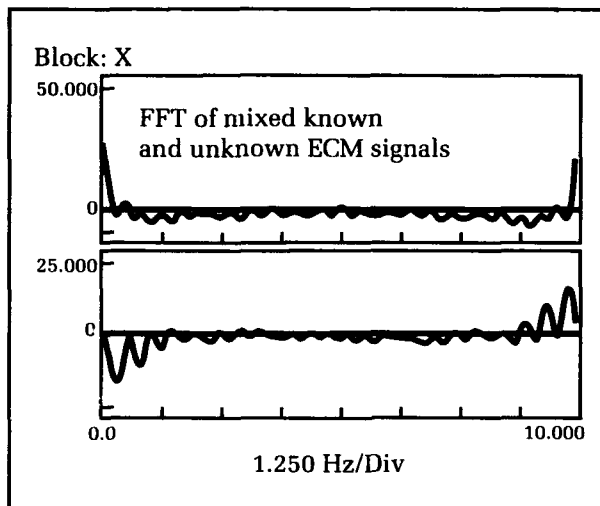


Figure 15. Spectrum of received signal.

Figure 16. Difference spectrum.

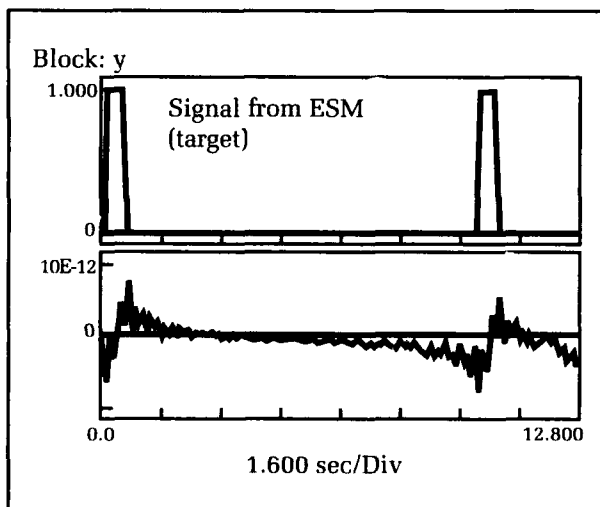
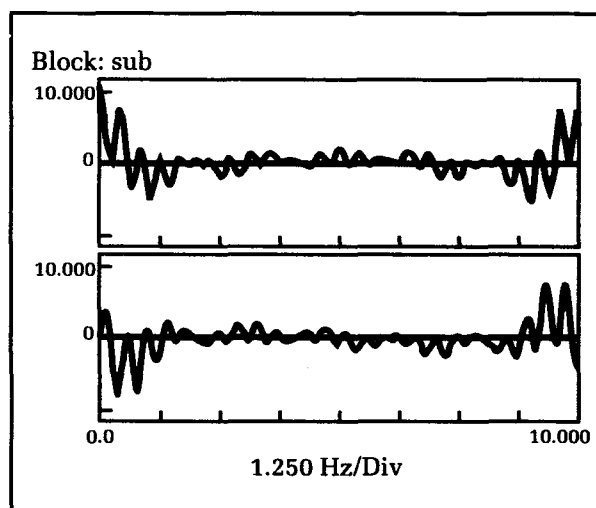


Figure 17. Inverse FFT of different spectra.

Figure 18. Net trained on 0.18-degree data.

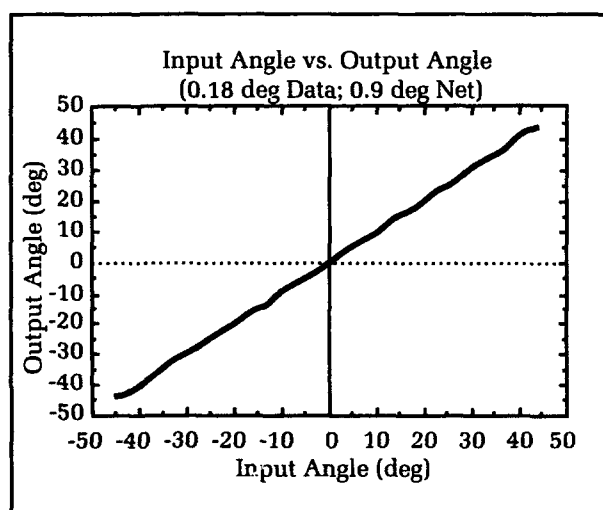
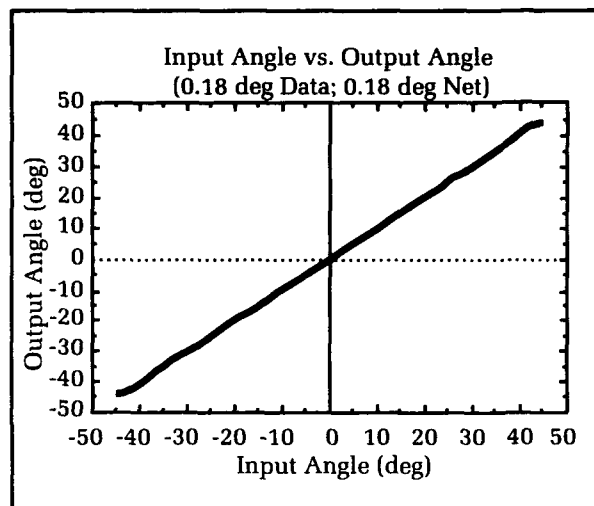
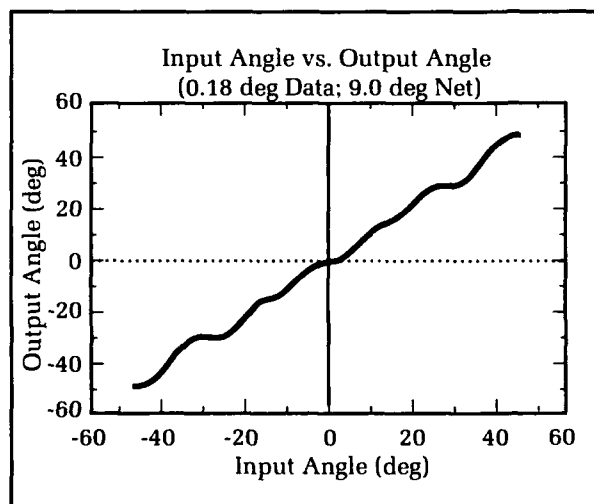


Figure 19. Net trained on 0.9-degree data.

Figure 20. Net trained on 9.0-degree data.



In addition, the coefficients themselves can be used as signal identifiers.

Since the technique used to improve direction-finding accuracy is based on pattern recognition, forms of pattern-recognizing algorithms other than cross-correlation may provide added capability. One such architecture is neural nets. A back-propagation net having 17 input elements, 34 hidden elements, and 1 output element was set up. Figure 18 shows the results of training this net, using a fast back-prop training scheme, on a wave field having a resolution of 0.18 degree, i.e., a training set of 501 patterns. When the data are fed back to the net, the results shown in Figure 18 are produced. Figure 19 shows the results of training the net on a 0.9-degree data set (101 patterns). The test set is 0.18 degree data. Figure 20 shows the result of testing with the 0.18 degree data on the net trained on only 11 evenly spaced patterns, 9.0-degree resolution. These results look promising.

Conclusions and Future Plans

The material presented here shows that it may be possible to realize significant improvements in the performance of operational electronic warfare systems by incorporating simple signal-processing functions. This approach also has the capability to deal with the electromagnetic environment in which these systems must operate. Another potential advantage is the ability to characterize the signal. This characterization can be defined as the reduction of the information content of the signal to the minimal basis consistent with the function of the system.

The next phase of this project includes tying together results from model prediction with actual experimental data. Additional future work involves incorporating filters based on higher order statistics, more extensive application of adaptive pattern-recognition networks, i.e., neural nets, and efforts to characterize signals.

References

1. Masi, W. A., *Improving the Direction Finding (DF) Accuracy and Sensitivity of the AN/SLQ-32 Using Digital Signal Processing with Application for LPI Detection and Interference Rejection*, NAVSWC TR 90-293, Jun 1990, NAVSWC, Dahlgren, VA.
2. Rotman, W. and Turner, R. F., "Wide-Angle Microwave Lens for Line Source Applications," *IEEE Transactions on Antennas and Propagation*, 1963.
3. Bendat, J. S. and Piersol, G. C., *Engineering Applications of Correlation and Spectral Analysis*, Wiley and Sons, 1980.
4. Giannakis, G. B. and Tsatsanis, M. K., "Signal Detection and Classification Using Matched Filtering and Higher-Order Statistics," *IEEE Trans. on Acoustics Speech and Signal Processing*, Jul 1990.
5. Ahmed, N. and Rao, K. R., *Orthogonal Transforms for Digital Signal Processing*, Springer-Verlag, 1975.

The Author



WILLIAM A. MASI received his B.S. degree in physics from Fordham University in 1965. He has been at NSWC, Dahlgren Division since 1966, where he has worked in the areas of electronic vulnerability and electronic warfare. He is presently developing signal-processing techniques for Navy surface electronic warfare systems.

Automated Threat Library Support System

Thomas E. Hudson

NSWC, Dahlgren Division, developed and implemented an Automated Threat Library Support System (ATLSS) for rapidly transferring and processing threat emitter data. In the current climate of unstable political alignments, international arms markets, and rapidly expanding technology, the Navy's battle group and ship commanders are faced with a constantly changing potential threat environment. Success, even survival, may depend upon the ability of tactical electronic warfare operators to recognize and respond appropriately to emitters whose characteristics or ownership may have changed. One key to this responsiveness is the timeliness and accuracy of on-board libraries of threat emitter parametric data. This article describes how ATLSS provides this data for use with the AN/SLQ-32(V) Antiship Missile Defense Electronic Warfare system installed in virtually all combatant U.S. Navy ships.

Introduction

As the Tactical Software Support Center (TSSC) for the AN/SLQ-32(V) Antimissile Defense System, the Dahlgren Division is responsible for simulation testing of geographically tailored threat libraries and distribution of the libraries to over three hundred naval combatants. We receive the threat libraries from Electronic Warfare Operational Programming Detachments (EWOPDETS) responsible for geotailoring the contents of each library.

Prior to ATLSS, we transferred threat libraries by mailing either a nine-track tape or a hard copy to the EWOPDETS, who returned information via mail. This process continued until each library was completed; normally, it took three months to complete a single geotailored library. ATLSS, on the other hand, permits rapid transfer and processing of geotailored threat emitter libraries between the Dahlgren Division and EWOPDETS located in Norfolk, Virginia, and Honolulu, Hawaii, where actual library construction from intelligence sources takes place.

We have also proved that the concept works on board ship by transferring threat data via satellite. The ATLSS secure telephone unit (STU III) transfer capability has enabled us to reduce processing time to a matter of hours.

Hardware

The hardware for ATLSS is shown in Figure 1. The personal computer (PC) is an IBM-PC/AT or compatible with 640K RAM, a 20 MB hard disk, and a 1.2 MB floppy drive. Attached to the PC is a Thorn 9903 nine-track tape drive

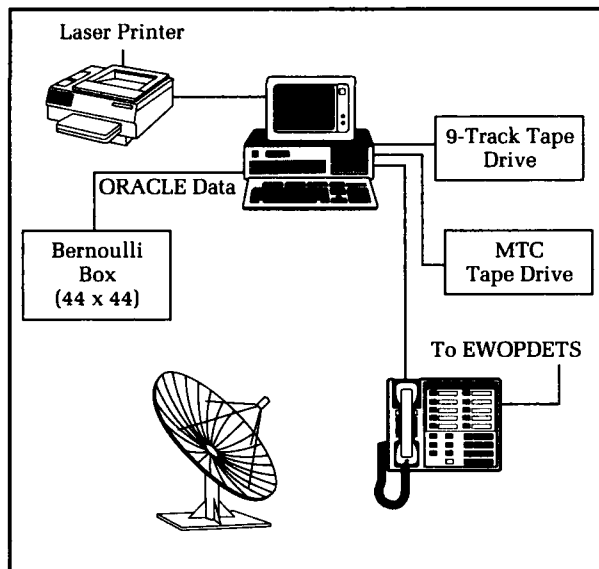


Figure 1. ALTSS hardware.

(used with either 800 or 1600 bits-per-inch tape) that loads data on Electronic Warfare Reprogrammable Library mainframes. The printer is a Hewlett Packard LaserJet. STU III interfaces to the PC via a Tempest-approved cable. This setup ensures compatibility with any configuration in the fleet.

A 44-MB Bernoulli drive maintains the threat libraries and data base on a single cartridge that can be secured when not in use. Therefore, the entire program is transportable.

The magnetic tape cartridge hardware is a tape drive unit similar to that on the AN/SLQ-32. Along with a PC, this gives a ship stand-alone capability to build a threat library for direct use on the AN/SLQ-32(V), and it gives the Dahlgren Division the capability to transfer threat libraries electronically for a "tape build" on ships.

Software

The software system, illustrated in Figure 2, consists of:

- Operating System for the host computer: either PC DOS, MS DOS version 3.0, or higher
- ORACLE Relational Data Base Management System: version 6.0
- Tape Utility: Flagstaff Engineering
- MTC Drive: Communications Package HP-IB Command Library
- Pascal Program Code
- STU III Communications: PROCOMM Plus

The ORACLE Relational Data Base Management System stores large amounts of information ready for instant use or update. The system uses two-dimensional tables to store data

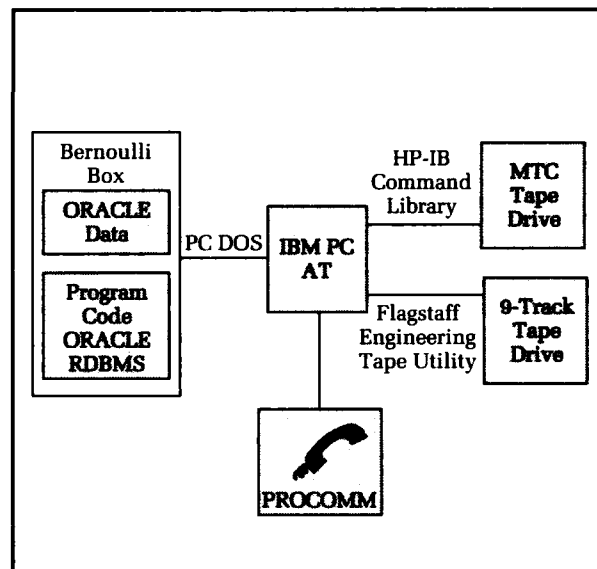


Figure 2. ALTSS software.

and allows a user to define relationships between items in a table and between items in different tables. The ORACLE data base is accessed via a high-level query language called Structured Query Language (SQL). With SQL, an operator can retrieve, insert, update and delete data; add new tables to the data base; and protect private data.

For ATLSS, ORACLE resides on one 44-mega-byte cartridge which is operated in a 44 x 44 Bernoulli Box. The system is installed so that the data files and data base system are on a single cartridge.

ATLSS Inputs

There are several ways of inputting data into the ATLSS data base: using a nine-track tape, keyboard inputting, transferring within the ATLSS data base, and transmitting via STU III from a remote site. The most efficient and cost effective method of updating and transferring threat libraries is via STU III transfer from the EWOPDETS.

The EWOPDETS have the capability of producing threat libraries on (9-track) magnetic tape from their mainframes. If electronic transfer cannot be used, the tapes can be mailed or hand-carried to the Dahlgren Division for loading into the ATLSS data base.

ATLSS Outputs

The media for ATLSS outputs is similar to the inputs with two exceptions: hard copy report of the threat library and the magnetic tape cartridge build (MTC).

Hard copies can be produced on all threat library parametric data for use as reference when desired. The MTC capability is the media used by the AN/SLQ-32(V). This capability is currently being tested for introduction to the fleet.

ATLSS Storage Functions

Data in the ATLSS data base may be stored in tables via relationships established by three corresponding fields in each table. Tables in the Baseline group contain data for all threat libraries entered in the data base. Tables in the New Release group contain data for a single threat library that was most recently added to the data base. The tables in the Difference group contain data from the New Release tables that does not correspond to data in the Baseline tables. Finally, tables in the Supplementary group are used to store data temporarily during one of the ATLSS functions or to store parametrics to which other data may be compared. The Baseline table displayed in Figure 3 is normally where a completed library is maintained. After a library has been tested and all parameters are correct, the library is entered into the Baseline.

The New Release stores data received on a 9-track tape, enabling the operator to change data or compare it to the baseline.

Remote Data is received from the EWOPDETS via the STU III link. The table has the same capabilities as the New Release.

Two-Library Comparison provides the operator with the capability to compare two different geographically tailored threat libraries. The output would result from the differences.

Baseline/New Release/Remote Comparison allows automatic comparison of newly received threat libraries to the baseline. This function will display or print the differences, which enables library builders to research, test differences, and reduce manual comparisons.

ATLSS performs an ambiguity analysis on any geographically tailored threat library. This analysis provides a listing of emitters that have all or partial overlapping of parametric values. For the AN/SLQ-32(V), the ambiguity analysis deals with Frequency, PRI and PRI Type, Scan and Scan Type. The result is a percentage of overlap for each field. Each ambiguous emitter is listed with the ambiguous mode.

Data Display/Modification Function

Baseline Data For -
Version:

No. of Emitters:

Emitter Data									
EPL No.:	Band:	Emitter:		Message:					
ELX No.:	Threat Level:	NTDS:	Modes:	Platforms:					
TRI=	Missile=	Air=	Surface=	Sub=	Land=	MLTINST=			

Mode Data									
EPL No.:	Band:	Frequency		PRI	PRI	Scan Period	Scan		
Mode	Min/Max	FA	Min/Max	MOD	Min/Max	Type	ECM	TTIA	

Definition of Terms
EPL Number: Emitter Parametric Listing, or ELINT number
Band: AN/SLQ-32V is divided into three frequency bands - 1, 2, and 3
Emitter: Emitter name is entered in this field
Message: Notes as the operator desires
ELX: Emitter Library Index - a computer-generated number
Threat Level: Threat weight on scale of 1 to 7
NTDS: Naval Tactical Data System
Frequency: The minimum and maximum frequency an emitter will operate on
FA: Frequency Agile - a bit set
PRI: Pulse Repetition Interval
PRI MOD: Modulation type
Scan: Min and max values an antenna may turn
Scan Type: Five-bit field for scan type
ECM: Electronic Countermeasures - a number in a table to designate a specific countermeasure
TTIA: Time-to impact adjustment

Figure 3. ATLSS data baseline data screen.

Data Update

Once a New Release or Remote Library has completed the research and testing, ATLSS will automatically update the Baseline library with the New Release or Remote. If a requirement exists to update a Baseline library with data from another, the capability is there. ATLSS also has the capability to update a single or all Baseline libraries with difference data.

Remote Transfer

ATLSS provides the Navy with the capability to transfer threat library data from the Dahlgren Division to any ship having satellite communications and ATLSS capability. The present method involves using STU III, accessing the satellite, and down loading to the desired ship. The Dahlgren Division developed and built a tape transport unit that is interfaced with ATLSS. With this configuration on a ship, a magnetic tape cartridge could be built for direct use on the AN/SLQ-32(V). ATLSS can be used for other EW systems by simply changing the data base to fit the needs of the library and interfacing to the media type used. For units without satellite and STU III capability, a simple solution is to have ATLSS on board and then transfer the data from units having the satellite capability.

The Dahlgren Division successfully transferred threat libraries to several aircraft carriers via ATLSS, STU III, and IMARSAT. The ship's crew built an MTC for use on the AN/SLQ-32(V).

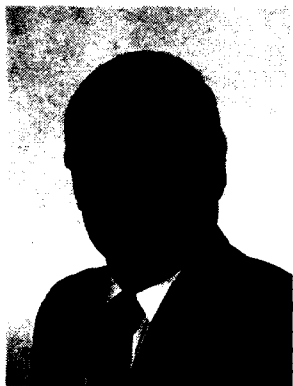
SPAWAR is presently working on a new satellite communications system that will be fully implemented in fiscal year 1994. The Chief of

Naval Operations is working to get a channel on the new system reserved for Electronic Warfare Reprogrammable Libraries. This new system would allow access to all ships without changing current equipment. Once the new system is implemented, library transfer to ships for Rapid Reprogramming or routine updates will be easily accomplished.

Conclusion

The Navy now has the capability to perform "real time" rapid reprogramming for the AN/SLQ-32(V). The process has two methods. The first method, the TSSC, is still in the system and responsible for threat engineering. This procedure would take place via satellite transfer. The second method would be within the Battle Group. With the ATLSS system installed, the Battle Group would have the capability to update libraries and write to the magnetic tape cartridge for use on the AN/SLQ-32(V). Finally, ATLSS is being modified to build libraries for other EW systems with the data being written on the appropriate media.

The Author



THOMAS E. HUDSON is an electronics technician at the Naval Surface Warfare Center, Dahlgren Division. He has attended Navy electronic and radar schools and local community colleges, and has over twenty years of experience in various aspects of electronic warfare. His most recent endeavor has been establishing and managing a viable electronic warfare reprogrammable library program for the AN/SLQ-32(V). As Electronics

Warfare Reprogrammable Library program manager, he frequently makes presentations on this topic.

The Doppler Spectrum for Accelerating Objects

John E. Gray and Gregory A. Watson

In this article, the doppler spectrum is computed for a perfectly reflecting mirror that simulates a point target model undergoing various types of accelerations. This work has application to the problems of rapid maneuver detection, target identification, and kill assessment. Accelerations and other types of noninertial motion are generally neglected effects in the doppler spectrum on the basis that these effects are second order. This assumption is dependent on the radar system being analyzed. Therefore, particular care must be taken to ensure that certain useful effects are not missed by the radar engineer. The methods presented enable determination of the doppler spectrum for any type of nonuniform motion.

Introduction

The doppler effect is a result of the relative motion of a signal source and manifests itself as a frequency shift of the wave. This frequency shift provides valuable information in some applications, such as doppler radar, while in others it is an unwanted effect to be "designed around," as in track-while-scan radar. Radar engineers generate an approximation of the exact solution by using a technique called "frame hopping," which is a double transform from the resting frame of reference to the moving frame of reference and back. This approach is usually adequate when a point target is assumed and radial velocity is the desired information. With suitable caution, this method gives acceptable results provided that the velocities are low. The difficulty is that "low" and "acceptable" are problem-dependent.

We have now developed a general method that treats noninertial motion and allows suitable approximations to be made. Derivation of the doppler spectrum makes available to the radar engineer secondary information that can provide nearly instantaneous data with applications in maneuver detection, target identification, and kill assessment.

The doppler effect^{1,2} is an important physical phenomenon that we must consider when designing radar systems. Empirical derivations of the doppler effect are discussed in radar texts.^{3,4,5} Although the data from these methods is correct and useful, it cannot easily be generalized to other than continuous wave (CW) waveforms, nor are the assumptions necessarily applicable to objects moving at other than constant velocity. To determine the effects of nonuniform motion, a re-examination of the exact derivation of the doppler effect is necessary. An exact derivation of the doppler effect starts with either a point particle or an infinite, perfectly reflecting mirror. The point particle is dynamically equivalent to the mirror.

Theory

The problem of the perfectly reflecting mirror model was first solved by Einstein, as may be seen in his paper that introduced relativity to the world.⁶ This model serves as the basis for understanding the doppler effect in radar. Generalizations of this model should serve to heighten awareness for any additional information that can be gained by examining the doppler spectrum. There are two methods for treating the noninertial movement of boundaries. Renewed interest in this area was sparked by Van Bladel^{7,8} and Mo.⁹ One method is to transform the fields in such a way that the boundary "appears" to be at rest relative to the fields. The boundary conditions are applied in the rest frame. The scattered fields are then retransformed to the original reference frame. The second method transforms the boundary conditions to a moving reference frame and solves the scattering problem with moving boundary conditions. Each method has advantages and disadvantages. The second method is easier to apply for one-dimensional fields, but computations become difficult in three dimensions because of boundary conditions.

In a previous paper,¹⁰ a general method was described for calculating the received waveform, $g(t)$, for an arbitrary transmitted waveform, $f(t)$, scattered from a perfectly reflecting surface which is undergoing a law of motion $r(t)$. This method is based upon results derived by Censor,¹¹ Cooper,^{12,13} and De Smedt,¹⁴ each of whom independently arrived at a general method for treating the arbitrary motion of mirrors that makes them amenable to examination in the Fourier domain. An exact expression for the reflected waveform, which is independent of the particular waveform $f(\bullet)$, is

$$g(\tau) = -\frac{d}{d\tau} F(2h(\tau) - \tau), \quad (1)$$

where

$$F(y) = \int^y f(u) du, \quad (2)$$

and y is the implicit variable. The function $h(\tau)$ satisfies the two relations:

$$h(\tau) + \frac{r(h(\tau))}{c} = \tau, \quad (3)$$

$$h(\tau) = t, \quad (4)$$

where c is the speed of light, t is time relative to the moving frame. For radar applications, the first-order relativistic approximation is usually all that is necessary. The proof of this approxi-

mation follows from the properties of contraction mapping.¹⁵ Assuming the first order approximation, which is the same order correction as the velocity, the received waveform is

$$g_1(\tau) = -\frac{d}{d\tau} F\left(\tau - \frac{2r(\tau)}{c}\right) \quad (5)$$

The spectral function $\hat{G}$, which is the Fourier transform of the received spectrum $g(\tau)$, is

$$-\hat{G}(\omega) = F\left(\tau - 2\frac{r(\tau)}{c}\right) e^{(-j\omega\tau)} \int_{-\infty}^{\infty} + j\omega \int_{-\infty}^{\infty} F\left(\tau - \frac{2r(\tau)}{c}\right) e^{(-j\omega\tau)} d\tau, \quad (6)$$

where ω is the frequency. It is now useful to consider a specific waveform, in particular the complex exponential, which is the standard representation of a doppler radar waveform. For this waveform, $f(t) = \exp(j\omega_0 t)$, $F(\bullet)$ is $f(\bullet)/j\omega_0$, and Equation (6) becomes

$$-\hat{G}(\omega) = \frac{e^{j\omega_0\left(\tau - \frac{2r(\tau)}{c}\right)}}{j\omega_0} e^{(-j\omega\tau)} \int_{-\infty}^{\infty} + \frac{\omega}{\omega_0} \int_{-\infty}^{\infty} e^{j\omega_0\left(\tau - \frac{2r(\tau)}{c}\right)} e^{(-j\omega\tau)} d\tau, \quad (7)$$

where ω_0 is the broadcast frequency. Note that the sinusoidal or exponential waveform is the only waveform where $f(\bullet)$ and $F(\bullet)$ have the same functional form. Therefore, the sinusoidal waveform is the only shape-preserving waveform. There can be significant differences between the received scattering information from sinusoidal and non-sinusoidal waveforms.

For practical applications, it is not appropriate to consider the frequency spectrum calculated over the interval $t \in [-\infty, \infty]$, but rather to consider it over a window of observation $t \in [(2n-1)T, (2n+1)T]$. For small windows of observation, the spectrum $\hat{G}(\omega)$ is a "snapshot" (instantaneous) of the spectrum due to the law of motion $r(t)$ in the interval $[(2n-1)T, (2n+1)T]$. A time-ordered ensemble of the snapshot spectra will be referred to as a spectrograph. For radar frequencies, ω_0 is large and the first term in Equation (7) can be neglected; the snapshot spectrum becomes

$$-\hat{G}_1(\omega) = \frac{\omega}{\omega_0} \int_{(2n-1)T}^{(2n+1)T} e^{j\omega_0\left(\tau - \frac{2r(\tau)}{c}\right)} e^{(-j\omega\tau)} d\tau. \quad (8)$$

In the radar observer's coordinates, the time t'' is represented by

$$t'' = t - \frac{2z_0(t)}{c}, \quad (9)$$

where $z_0(t)$ is the distance to the mirror at the time the acceleration begins. This would correspond to replacing Equation (3) by

$$h(\tau) + \frac{2z_0}{c} = t'. \quad (10)$$

This is nothing more than a shift of the origin, which introduces a phase shift in the Fourier spectrum. Since this shift of origin adds nothing to the physics of the problem, we can simplify the algebra by ignoring it, which has been done throughout the remainder of this article.

Examples

Constant Velocity Model

A constant velocity model is considered as the first example. For a constant velocity target, $r(t) = vt$, the received spectrum is given by substituting this model into Equation (8) ($\omega' = \omega_0(1 - 2\beta)$, $\beta = v/c$), yielding

$$\hat{G}_1(\omega) = \left(\frac{-2\omega'}{\omega_0} \right) e^{j2nT(\omega' - \omega)} \frac{\sin[(\omega' - \omega)T]}{(\omega' - \omega)}. \quad (11)$$

With $n = 0$, then taking the limit as T approaches infinity in Equation (11) yields the standard doppler frequency spectrum

$$\hat{G}_1(\omega) = \left(\frac{-2\pi\omega'}{\omega_0} \right) \delta(\omega' - \omega). \quad (12)$$

Linearly Accelerating Mirror

As a second example, the linearly accelerating mirror is examined. Let the position be represented by

$$r(t) = v_0 t - \frac{a_0 t^2}{2}. \quad (13)$$

$\hat{G}_1(\omega)$ is determined by substituting Equation (13) into Equation (8)

$$90 \quad \hat{G}_1(\omega) = \frac{-\omega}{\omega_0} \int_{(2n-1)T}^{(2n+1)T} e^{j(a\tau^2 - b\tau)} d\tau, \quad (14)$$

where $\alpha = a_0/c$, $a = \omega_0 \alpha$, and $b = \omega - \omega'$. Equation (14) can be evaluated to obtain

$$-\hat{G}_1(\omega) =$$

$$\frac{\omega}{\omega_0} \sqrt{\frac{\pi}{2a}} \left[K \left(\frac{2a(2n+1)T - b}{\sqrt{2a\pi}} \right) - K \left(\frac{2a(2n-1)T - b}{\sqrt{2a\pi}} \right) \right] e^{-\frac{jb^2}{4a}}, \quad (15)$$

where,¹⁶

$$K(x) = \int_0^x e^{\frac{j\pi y^2}{2}} dy = C(x) + jS(x). \quad (16)$$

The frequency spectrum is no longer sharply localized at the expected frequency ω' . The frequency distribution has been spread by acceleration, especially at the higher accelerations.

Constant Velocity, Oscillating Boundary

A third example is a boundary moving at a constant velocity while oscillating with an amplitude A_0 . The functional dependence is given by

$$r(\tau) = v_0 \tau - A_0 \sin(\Omega \tau), \quad (17)$$

where $\Omega (= 2\pi/T_1)$ is the frequency of oscillation. To evaluate the spectrum from Equation (17), first note that $\exp(j2\omega_0 A_0 \sin(\Omega t)/c)$ is periodic, thus it can be written as¹⁷ ($z = 2\omega_0 A_0/c$)

$$e^{(jz \sin(\Omega \tau))} = \sum_{m=-\infty}^{\infty} a_m e^{jm\Omega \tau}, \quad (18)$$

where¹⁸

$$a_m = \frac{1}{T_1} \int_{-T_1/2}^{T_1/2} e^{jz(\sin(\Omega \tau))} e^{-jm\Omega \tau} d\tau = J_m(z). \quad (19)$$

Substituting Equation (18) into Equation (8) and computing the instantaneous spectrum gives

$$\begin{aligned} -\hat{G}_1(\omega) &= \frac{\omega}{\omega_0} \sum_{m=-\infty}^{\infty} J_m(z) \int_{(2n-1)T}^{(2n+1)T} e^{j\omega' \tau} e^{jm\Omega \tau} e^{(-j\omega \tau)} d\tau \\ &= \frac{\omega}{\omega_0} \sum_{m=-\infty}^{\infty} \left(J_m(z) e^{j2nT(\omega' + m\Omega - \omega)} \right) \left(\frac{2 \sin[(\omega' - \omega + m\Omega)T]}{(\omega' - \omega + m\Omega)} \right) \end{aligned} \quad (20)$$

As the frequency of the radar increases, smaller vibration amplitudes can be detected.

Object Undergoing Air Friction Decay

The model of an object undergoing air friction decay will be considered as a fourth example. The scatterer is modeled as

$$r(t) = \frac{a_0}{\gamma^2} (1 - e^{-\gamma t}) + t \left(v_0 - \frac{a_0}{\gamma} \right). \quad (21)$$

The received frequency spectrum can be written in the form

$$G_1(\omega) = f \int_{(2n-1)T}^{(2n+1)T} e^{ja_1\tau} e^{jb_1\tau} e^{-\gamma\tau} d\tau, \quad (22)$$

where

$$a_1 = \omega_0 \left(1 + \frac{2a_0}{\gamma c} - 2\beta \right) - \omega, \quad (23)$$

$$f = \frac{-\omega}{\omega_0} e^{-jb_1}, \quad (24)$$

$$b_1 = \frac{2\omega_0 a_0}{c\gamma^2}. \quad (25)$$

A change of variables puts the received spectrum in Equation (22) in the standard form of the Pearson incomplete gamma function¹⁹

$$\hat{G}_1(\omega) = -\frac{(b_1)^{-x} f}{\gamma} [C(u_1, x) - C(u_2, x) + j(S(u_1, x) - S(u_2, x))] \quad (26)$$

where $(x = -ja_1/\tau)$

$$C(q, a) = \int_q^\infty t^{a-1} \cos(t) dt, \quad (27)$$

$$S(q, a) = \int_q^\infty t^{a-1} \sin(t) dt, \quad (28)$$

$$u_2 = b_1 e^{-\gamma(2n+1)T} \text{ and } u_1 = b_1 e^{\gamma(2n-1)T}. \quad (29)$$

This completes the discussion of examples. Other examples of interest will be considered in future papers. Some practical aspects will now be discussed, such as the question of whether nonuniform motion of the models is observable from the spectrum.

Applications

For nonuniform motion, two aspects of the information are derived from the spectrograph. The first factor to consider is the moments of a specific spectrum. If the spectrum has higher order moments ($n \geq 2$), it can be considered to be sophisticated. Also, if the real part of the spectrograph is a function of n , the spectrograph is nonstationary, and hence complicated, since several individual spectra are needed to discern behavior when noise is considered. The n -th central frequency is given by²⁰

$$\langle \omega^n \rangle = \frac{\int_{-\infty}^{\infty} \omega^n G_1(\omega) d\omega}{\int_{-\infty}^{\infty} G_1(\omega) d\omega}. \quad (30)$$

The variance in the signal is

$$\sigma^2 = \langle \omega^2 \rangle - \langle \omega \rangle^2 = \quad (31)$$

$$\frac{\beta''(0)}{(1 - \beta(0))} + j\omega_0 \beta'(0) + \frac{\beta'(0)^2}{(1 - \beta(0))^2},$$

where $\beta(t) = 2r'(t)/c$. Note that when using a non-windowed $r(t)$, the same formula would apply with $a = 0$. At radar frequencies, only the second term is important for the low-order, nonuniform motion effects being considered. Therefore, under the high frequency assumption, the magnitude of the spread is

$$|\sigma| \approx \sqrt{\omega_0 \beta'(0)}. \quad (32)$$

For the constant velocity model, the real part of the spectrograph is independent of the index n , so it is categorized as noncomplicated; for an example, see Figure 1. Also, the constant velocity scat-

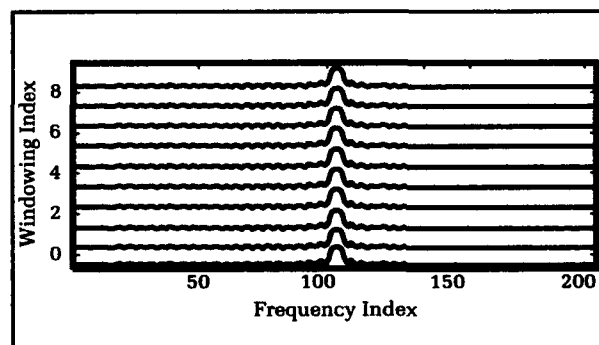


Figure 1. Spectrograph of constant velocity model. The windowing index is the number n in the interval $[(2n-1)T, (2n+1)T]$, and the frequency index is the discrete frequency.

terer is not a sophisticated signal, since $\beta'(x) = 0$. (Note that the second moment is not zero, but the spread is zero.) For the vibrating model, the spread is zero, so it is not a sophisticated signal. For the vibrating mirror, the magnitude of the sum is independent of the index n , so it is simple. The signal-to-sideband ratio, assuming the signal has been beat down to base band, is

$$SSR(m) = \frac{|\hat{G}_1(m, \omega_b)|}{|\hat{G}_1(0, \omega_b)|} \quad (33)$$

$$= \frac{|J_m(z) \text{sinc}(\omega_b + m\Omega)T|}{|\text{sinc}(\omega_b T)|},$$

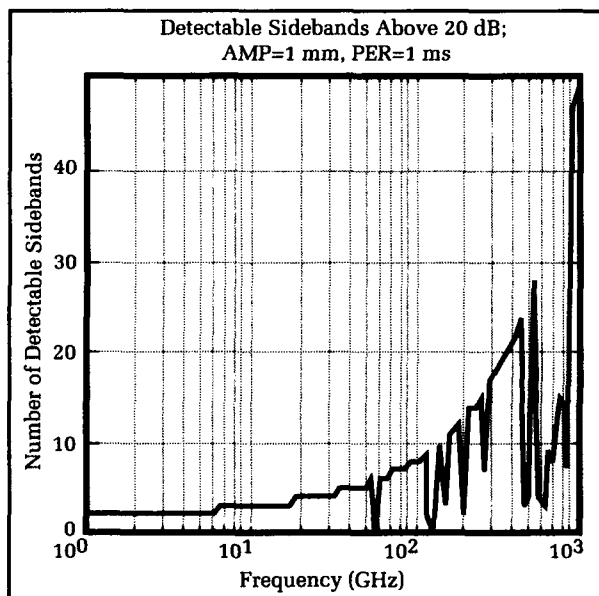


Figure 2. Number of detectable sidebands as a function of radar frequency.

with $\omega_b = 2\omega_0 v_0/c$. An example is shown in the graph in Figure 2 for a frequency range of 1-10³ GHz with an amplitude of one mm and a period of 1 msec.

The spectrum of the accelerating mirror has a sophisticated structure. The spread is

$$\sigma = \sqrt{\frac{2\omega_0 |a_0|}{c}}. \quad (34)$$

This is graphed in Figure 3 as a function of the frequency for accelerations of 1 g, 5 g, etc. The spectrograph is also complicated because it depends on the index n . Examples of spectrographs for various accelerations are shown in Figure 4. Note that the expected frequency for the accelerating boundary as a function of the index n is approximately

$$\langle \omega \rangle = \omega_0 (1 - \beta(0)). \quad (35)$$

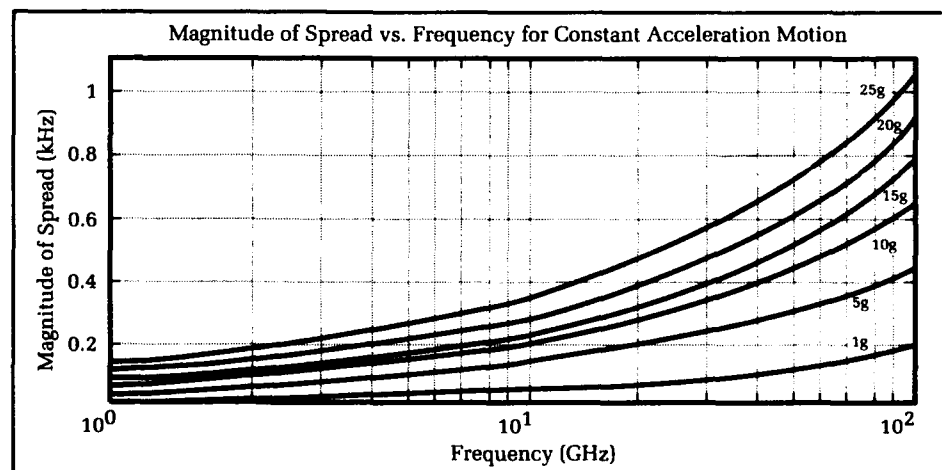
The difference between successive instantaneous spectra is

$$\langle \omega(n+1) \rangle - \langle \omega(n) \rangle = -\frac{4\omega_0 T}{c} a_0. \quad (36)$$

Moments are not discussed in radar books such as Skolnik.²¹ Instead, it is noted that difference in successive velocity estimates is $a_0 T$. This observation comes from taking the difference between successive first moments. For low accelerations, the observation that this information is just the first moment of a more complicated spectrum is irrelevant, but for higher accelerations second-moment information can be useful.

So there are two methods for determining the accelerations from spectrograph records: one is to estimate the second moments from individual instantaneous records, and the other is to estimate accelerations from the first moments of the

Figure 3. Spread in frequency due to varying accelerations.



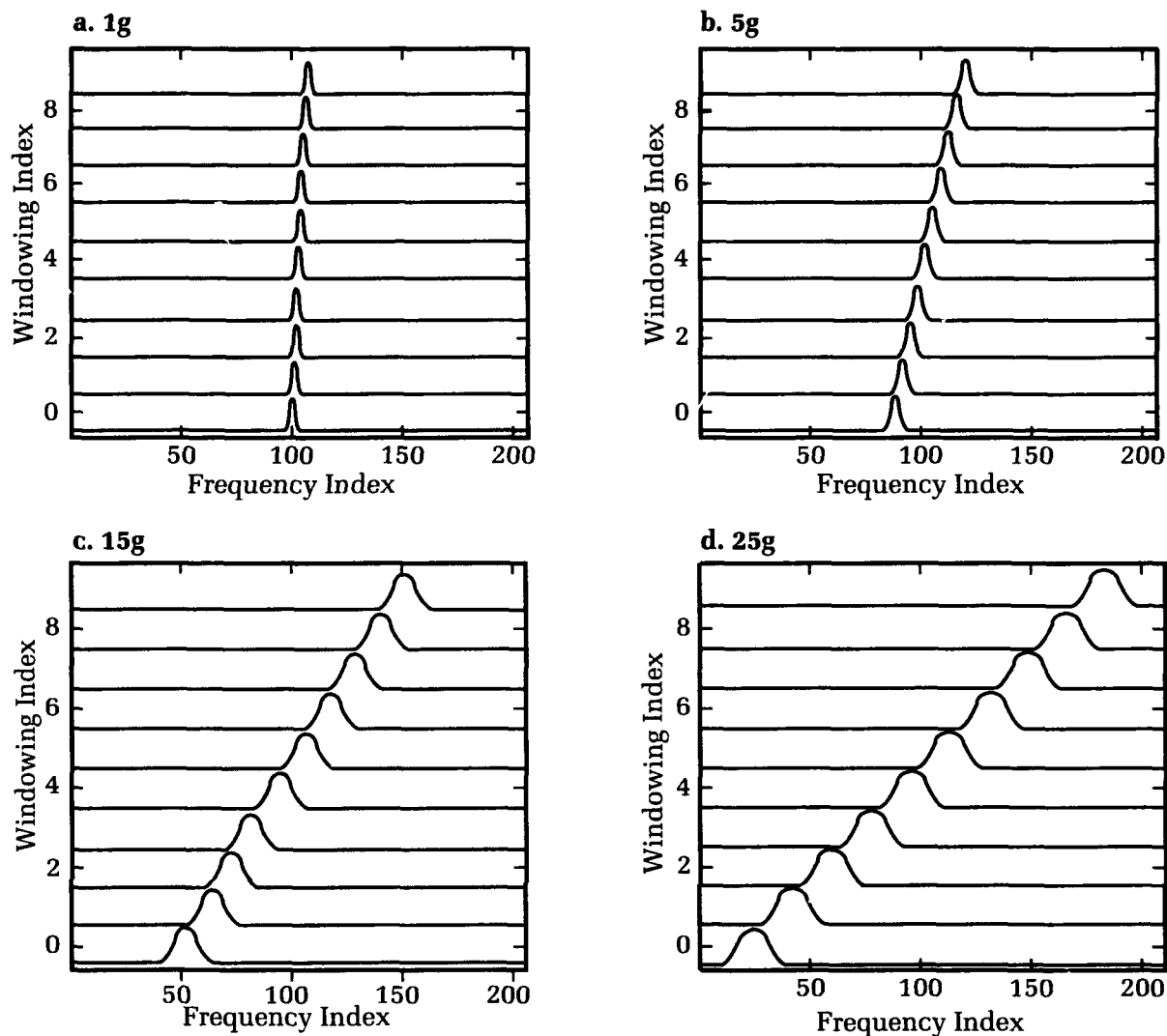


Figure 4. Spectrographs for accelerating model (1 to 25g).

spectrograph. For low noise levels, the first method will give a faster estimate, but if the noise level is high, a least-squares fit to the first moment gives an estimate of the acceleration.

Conclusions

There are exact results for the Fourier spectra that can be obtained for other than constant-velocity boundaries that can be found from the functional form of $r(\tau)$. For boundaries undergoing nonuniform motion, the resultant Fourier spectrum has been expressed in a manner such that closed-form solutions can be calculated for several different models. Exact solutions are not necessarily of great interest to the radar engineer. However, a method of approximating the correct

general solution is important in order to determine which, if any, effects of nonuniform motion are important. For some applications, such as doppler radar, the receivers are very sensitive to frequency shifts, so additional information is available in the Fourier spectrum as shown in the Applications section of this article. Additional discussions of applications that extend the work are planned. There are several areas for continuing research. The models shown to have measurable effects for CW radars need to be tested against other waveforms to determine if there are any additional useful measurable effects.

Non-sinusoidal waveforms is an area for fruitful research since the form of the received waveforms are different from those for CW waveforms. Also, the problem of reconstructing the

equation of motion from the spectrograph is an interesting application of recent developments in signal reconstruction such as wavelets and Cohen-class transforms.

Acknowledgments

We wish to express thanks to Danny Brunson of the Surface Launched Weapons Technology program for his support of this work. Also, thanks to George LeFurjah of Technology Services Corp., who insisted that closed form meant no convolution in the solution. We also wish to thank Ted Rice and J. Conte of the Weapons Direction Systems Branch for many valuable discussions and advice.

References

1. Doppler, C. J., "Über das farbige Licht der Dopplersterne und einiger anderer Gestirne des 'Himmels,'" *Abhandlungen der Königlich Böhmischen Gesellschaft der Wissenschaften*, Vol. 2, 1842, pp. 467-482.
2. Gill, T. P., *The Doppler Effect, An Introduction to the Theory of the Effect*, Logos Press Limited, 1965.
3. Levanon, Nadav, *Radar Principles*, Wiley-Interscience Publications, John Wiley & Sons, 1988, pp. 2-4.
4. Rihaczek, A. W., *Principles of High Resolution Radar*, McGraw-Hill Book Co., 1969.
5. Cook, Charles E. and Bernfeld, Marvin, *Radar Signals, An Introduction to Theory and Application*, Academic Press, New York, London, 1967, pp. 64-65.
6. Einstein, A., "Zur Elektrodynamik Bewegter Körper," *Ann. Phys. (Leipzig)*, Vol. 17, 1905, pp. 891-921.
7. Van Bladel, Jean, "Electromagnetic Fields in the Presence of Rotating Bodies," *Proc. of the IEEE*, Vol. 64, No. 3, Mar 1976, pp. 301-318.
8. Van Bladel, Jean, *Relativity and Engineering*, Springer-Verlag, Berlin, Heidelberg, New York, 1984.
9. Mo, Chin Tse, "Theory of Electrodynamics in Noninertial Frames and Applications," *J. of Math. Physics*, Vol. 11, No. 8, Aug 1970, pp. 2589-2610.
10. Gray, J. E., "The Doppler Spectrum of Accelerating Objects," *International Radar Conf., Radar-90*, May 1990.
11. Censor, Dan, "The Generalized Doppler Effect and Applications," *J. of The Franklin Institute*, Vol. 295, No. 2, Feb 1973, pp. 103-116.
12. Cooper, Jeffery, "Scattering of Electromagnetic Waves by a Moving Boundary: The One-Dimensional Case," *IEEE Transactions on Antennas and Propagation*, Vol. AP-28, No. 6, Nov 1980, pp. 791-795.
13. Cooper, Jeffery, "Scattering by Moving Bodies: The Quasi Stationary Approximation," *Math. Meth. in the Appl. Sci.*, 2, 1980, pp. 131-148.
14. Van Bladel, Jean and De Zutter, Daniel, "Reflections from Linearly Vibrating Objects: Plane Mirror at Normal Incidence," *IEEE Transactions on Antennas and Propagation*, Vol. AP-29, No. 4., Jul 1981, pp. 629-636.
15. Copson, E. T., *Metric Spaces*, Cambridge Univ. Press, 1968, pp. 112-116.
16. Papoulis, Athanasios, *Signal Analysis*, McGraw-Hill Book Co., New York, 1977.
17. Papoulis, Athanasios, *The Fourier Integral*, McGraw-Hill Book Co., New York, 1962.
18. Lebedev, N. N., *Special Functions and their Applications*, Dover Publications, Inc., New York, 1972.
19. Abramowitz, M. and Stegun, I. A., *Handbook of Mathematical Functions*, Dover Publications, Inc., New York, 1972.
20. Bracewell, Ronald N., *The Fourier Integral and Its Application*, McGraw-Hill Book Co., New York, 1987, Chap. 8.
21. Skolnik, Merrill I., *Introduction to Radar Systems*, 2nd ed., McGraw-Hill Book Co., New York, 1980.

The Authors



JOHN E. GRAY was born on 28 August 1955 in Fort Lee, Virginia. He received B.S. degrees in physics and mathematics from the University of Mississippi in 1977 and an M.S. in physics in 1980. He has also studied physics and electrical engineering at Virginia Tech and Catholic University. He is a member of IEEE, the American Physical Society, the Society for Industrial and Applied Mathematics, and Sigma Phi

Sigma. He is employed by the Naval Surface Warfare Center, Dahlgren Division working in the areas of tracking filters, signal analysis, and electromagnetic theory applied to radars. His other areas of interest are wave phenomena, foundations of quantum mechanics, and nonlinear mathematics. He has authored over 40 technical publications.



GREGORY A. WATSON received a B.S. in physics from West Chester University, West Chester, Pennsylvania, in 1988 and an M.S. in physics from the University of Connecticut at Storrs in 1990. He joined the Naval Surface Warfare Center, Dahlgren, Division as a physicist in 1990, where he has been involved in developing new and adaptive techniques for target state estimation and trajectory prediction. He has also

worked in the area of continuous wave radar waveforms to identify the relative motion effects of targets. His research interests include the use and implementation of the interactive multiple-model filter to improve tracking performance for both periodic and aperiodic data.

Directional-Energy Detector Sonobuoy

T. B. Ballard

Passive sonobuoys are acoustic listening devices deployed by air platforms to detect underwater targets. Although there are several types of sonobuoys in use, one thing common to all is that the raw data must be transmitted to the platform for processing and analysis. The processing equipment in the platform limits the number of sonobuoys that can be used effectively at any time. As targets become more quiet, the number of sonobuoys required to perform a search in a given mission area increases, thereby necessitating that some of the data processing capability be transferred from the platform to the individual sonobuoys. The NSWC Dahlgren Division has developed an automated directional-energy detection technique based on improved in-buoy signal processing incorporated into a sonobuoy by use of a simple microprocessor. Tests using the technique have demonstrated a small false-alarm rate and significantly improved detection capability compared to omnidirectional detectors. This article describes pertinent features of the technique, system sensitivity and false-alarm rate considerations, signal processing functions, and test results.

Introduction

Passive sonobuoys currently in the U.S. Navy inventory are acoustic devices used to detect submarines by listening for underwater sound information. They are generally deployed by aircraft and transmit the sound information back to the aircraft via a radio link as illustrated in Figure 1. Several types of sonobuoys differ in the number of hydrophones used and the directional characteristics of the hydrophones. However, one thing that is common to all types is that the raw data are transmitted to the aircraft for analysis. The processing equipment in the aircraft limits the number of sonobuoys that can be used effectively at any time. As targets become more quiet, the detection range of a sonobuoy is reduced; consequently, a larger number of sonobuoys must be used. This requires either an increase in the number of sonobuoys that can be processed by an aircraft or inclusion of processing capability in the sonobuoy itself. This article describes an approach to in-buoy signal processing developed by the Dahlgren Division that can result in a sonobuoy with automatic detection capability.

Sonobuoy Specifications

At the beginning of this development effort, goals were set to describe the sonobuoy performance to be achieved. The sonobuoy was to be within the mechanical framework of existing sonobuoys. It should be able to automatical-

ly detect a nearby submarine, and upon detection transmit an RF signal. The detection would be based on receiving broadband energy radiated by the submarine. Probability of false alarm was to be very small so that if a large number of these sonobuoys were deployed, the monitoring platform would not be overloaded. The sonobuoy would have an operating life somewhat longer than current sonobuoys, perhaps several days. This should be achievable because the RF transmitter would be used only upon detection, thus saving significant power. The detector was to be made as sensitive as possible within the constraint of minimum false-alarm rate. An acceptable sensitivity was to be -12 dB compared to the background noise level.

Selection of a Directional-Energy Detector

The problem of detecting broadband energy from a target was first studied by assuming an omnidirectional sensor. An analysis of this technique revealed several problems that could be solved by using a directional-energy detector.

Omnidirectional Sensor

For the omnidirectional sensor, it is assumed that the hydrophone output is passed through a bandpass filter and into a detector. With or without a target present there will be background noise, and initially let us assume that this background noise level is not time-varying and is accurately known. If the effective bandwidth of the bandpass filter is B Hertz, the noise is white Gaussian, and T seconds of data are collected; then $2BT$ independent samples will be available. Let the samples be g_i , with $i = 1, 2, \dots, 2BT$. With only background noise present, each of these samples will be normally distributed with

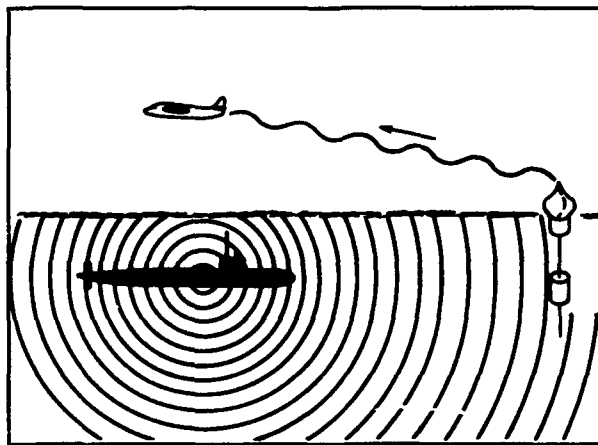


Figure 1. Passive sonobuoy transmits raw data to aircraft for analysis.

a mean of zero and a variance of σ_N^2 . With a target signal present the mean will still be zero, but the variance will be the sum of the noise power and signal power, i.e., $\sigma_N^2 + \sigma_S^2$. Let the measure of power be P

$$P = \sum_{i=1}^{2BT} g_i^2$$

Then, for the noise-only case, P has a chi-squared distribution with a mean of $2BT \sigma_N^2$ and a variance of $4BT \sigma_N^4$. If σ_N^2 is known perfectly, a detection threshold can be established for a specific probability of false alarm. The threshold will be set at

$$2BT \sigma_N^2 + k \sqrt{4BT \sigma_N^4}$$

where k depends on the probability of false alarm (PFA). For a PFA of 10^{-3} , $k = 3.09$. With the threshold set, the signal strength that yields a 50 percent probability of detection can be calculated by satisfying the following relationship:

$$2BT(\sigma_N^2 + \sigma_S^2) = 2BT \sigma_N^2 + k \sqrt{4BT \sigma_N^4}$$

This results in a signal-to-noise ratio, or SNR

$$SNR = \frac{\sigma_S^2}{\sigma_N^2} = \frac{k}{\sqrt{BT}}$$

As an example, a system can theoretically achieve a 50 percent probability of detection for an SNR of -12 dB and give 10^{-3} probability of false alarm if $BT = 2400$ and the threshold is set at $(2BT \sigma_N^2) (1.063)$. The equation for the threshold shows that a very precise knowledge of σ_N^2 is required, and that system performance will change dramatically if σ_N^2 varies a few percent. The detection threshold cannot be preset, but must be derived from a very accurate measurement of the background noise level. Once this measurement has been made and used to set the threshold, the system performance depends on this level being accurately maintained during the period of time when target detection will be attempted. It is for these reasons that an omnidirectional energy detector working at -12 dB SNR was not pursued; instead, a directional-energy detector was studied.

Directional-Energy Detector

The directional-energy detector uses a hydrophone array from an AN/SSQ-53 sonobuoy known as a Direction Finding Acoustic Receiver (DIFAR, illustrated in Figure 2). This array provides three outputs, one from an omnidirectional sensor and two having dipole responses separated by 90° . This allows a beam to be steered in any direction in the horizontal plane with an array gain of 4.7 dB. The directional-energy

detector concept uses the steerable beam to estimate received energy as a function of azimuthal angle, and then bases the detection upon greater energy arriving from one direction as opposed to all other directions. This approach makes use of array gain to improve SNR; but even more significant, a precise knowledge of the noise level is no longer required since energies from different directions are always being compared. In addition, a new dimension is added to the detection process, that of bearing to the source. This can be used to conduct tests beyond energy detec-

tion such as bearing consistency and rate of change of bearing.

To form a directional beam with the outputs from a DIFAR sonobuoy, a weighted sum is made of the hydrophone outputs. Let the hydrophone outputs be:

x_O = output from omnidirectional hydrophone

x_{N-S} = output from N-S dipole

x_{E-W} = output from E-W dipole

The assumption is made that in isotropic noise all three outputs are Gaussian and independent, and have variances of N , $\frac{N}{2}$, and $\frac{N}{2}$, respectively. To form a maximum gain beam in the direction θ , a weighted sum is made of the hydrophone outputs: $y = x_O + 2 x_{N-S} \cos \theta + 2 x_{E-W} \sin \theta$. The variable y is the output for a beam steered in direction θ . Let P_θ be the output power of the beam in direction θ . Then $P_\theta = E(y^2)$. For isotropic noise $P_\theta = 3N$. If in addition to isotropic background noise, there is a signal of power S from direction ϕ :

$$P_\theta = 3N + S[1 + 2 \cos(\theta - \phi)]^2$$

It follows that if the beam is steered directly at the signal source, then $P_\theta = 3N + 9S$, and the signal-to-noise ratio is $3 S/N$. For a single omnidirectional sensor, the signal-to-noise ratio would

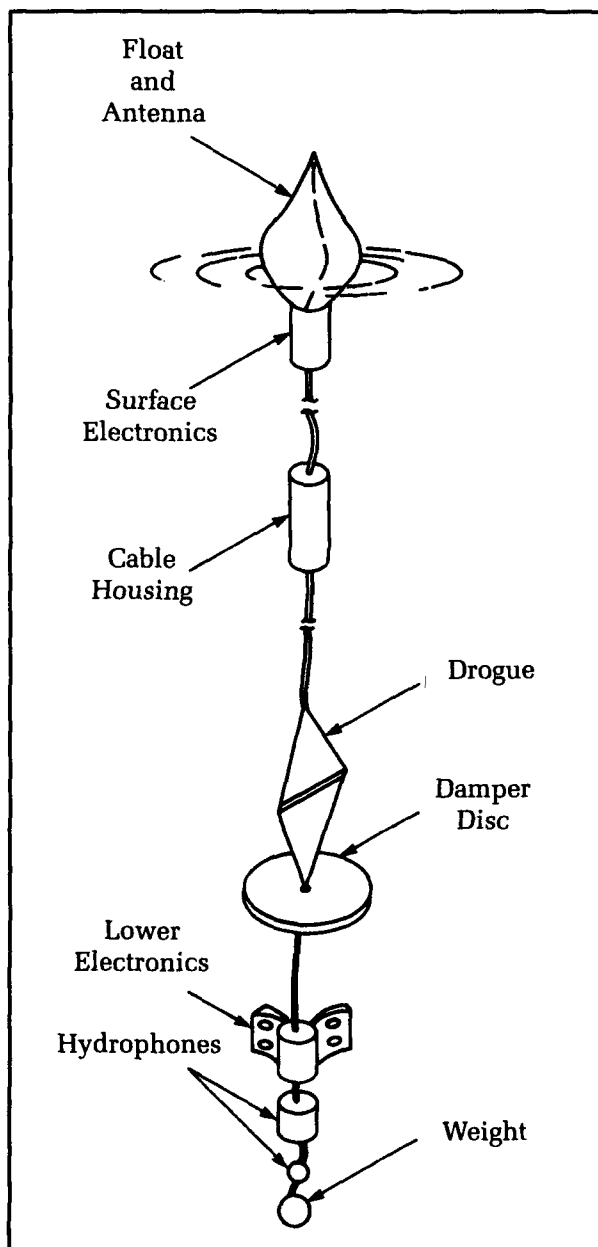


Figure 2. DIFAR sonobuoy.

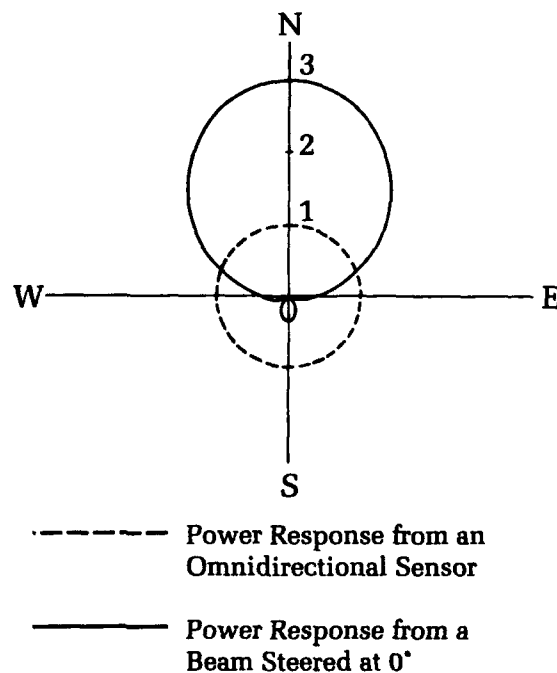


Figure 3. Response from a steered beam showing array gain.

be S/N . Thus, the use of the beam results in a gain of $10 \log 3 = 4.7$ dB.

Figure 3 is a plot of the array gain as a function of arrival angle for a beam steered to the north. Because of the width of the beam, it is possible to cover all azimuthal directions by steering eight beams spaced by 45° , i.e., at 0° , 45° , $\dots$, 315° . The maximum scalloping loss midway between two beams is .45 dB, and the average scalloping loss is .15 dB.

The forming of eight beams spaced by 45° is done by squaring the weighted sums of the hydrophone outputs.

$$\begin{bmatrix} y_0 \\ y_{45} \\ y_{90} \\ y_{135} \\ y_{180} \\ y_{225} \\ y_{270} \\ y_{315} \end{bmatrix} = \begin{bmatrix} 1 & 2 & 0 \\ 1 & \sqrt{2} & \sqrt{2} \\ 1 & 0 & 2 \\ 1 & -\sqrt{2} & \sqrt{2} \\ 1 & -2 & 0 \\ 1 & -\sqrt{2} & -\sqrt{2} \\ 1 & 0 & -2 \\ 1 & \sqrt{2} & -\sqrt{2} \end{bmatrix} \begin{bmatrix} x_O \\ x_{N-S} \\ x_{E-W} \end{bmatrix}$$

$$P_\theta = y_\theta^2$$

In the case of isotropic noise of power N and a signal of power S from an arbitrary direction:

$$\begin{aligned} \text{Average beam power} &= P_{ave} \\ &= \frac{1}{8} \sum_{i=0}^{315} P_i = 3N + 3S \end{aligned}$$

For a beam pointed directly toward the signal, which will be the strongest beam,

$$P_{peak} = 3N + 9S$$

$$\text{Let } \gamma = \frac{P_{peak}}{P_{ave}} = \frac{3N + 9S}{3N + 3S} = 1 + \frac{2S}{S + N}$$

γ has the property that, for isotropic noise and no signal, the probability distribution of γ is independent of the noise level. Therefore, for signal detection a fixed threshold can be set that is unrelated to the strength of the isotropic background noise.

To determine the correct value for a detection threshold, a computer simulation was done. For isotropic noise and values of BT from 10^2 to 10^4 , the measured statistics for γ were:

$$\text{mean} = 1 + \frac{1.15}{\sqrt{BT}}$$

$$\text{standard deviation} = \frac{.48}{\sqrt{BT}}$$

Therefore, for 50 percent probability of detection and 10^{-3} probability for false alarm, a threshold, T , can be computed.

$$T = 1 + \frac{1.15}{\sqrt{BT}} + 3.09 \times \frac{.48}{\sqrt{BT}} = 1 + \frac{2.63}{\sqrt{BT}}$$

This threshold will yield a false alarm of 10^{-3} . To compute the SNR for 50 percent probability of detection, solve

$$1 + \frac{2S}{S + N} = 1 + \frac{2.63}{\sqrt{BT}}$$

For $BT \gg 1$, then

$$\frac{S}{N} = \frac{1.32}{\sqrt{BT}}$$

This can be compared to the SNR required by an omnidetector under the same conditions of probability of detection and false alarm. For the omnidetector at 10^{-3} false alarm,

$$\frac{S}{N} = \frac{3.09}{\sqrt{BT}}$$

Therefore, the gain to be achieved by use of the directional-energy detector is

$$10 \log \left(\frac{3.09}{1.32} \right) = 3.7 \text{ dB}$$

This is based on the assumption that the signal to be detected is lined up with one of the eight beams. A loss of .15 dB is incurred from random target bearings. Therefore, the overall gain is 3.55 dB compared to an omnidetector; and more important, the system is insensitive to isotropic background noise level.

In-Buoy Signal Processing

The design of the directional-energy detector sonobuoy is most applicable to the existing DIFAR sonobuoy. Three hydrophones are used in a DIFAR sonobuoy, two directional and one omnidirectional. The hydrophone signals are multiplexed in a manner whereby the directional hydrophone signals are compass-corrected. Regardless of the sonobuoy orientation, the output signals are referenced to the magnetic north direction. The multiplexed signal is transmitted for the entire life of the sonobuoy, which can be as much as eight hours. All further signal processing is done in the receiving platform.

The directional-energy detector sonobuoy can be considered a modification of the basic DIFAR design. A block diagram is shown in Figure 4. The existing DIFAR signal flow remains unaltered except that the transmission of data is no longer continuous. The transmitter is turned on for a short time after deployment to verify the operation of the sonobuoy. After that it is turned off until a detection has been made, at which time it turns on during the period of the detec-

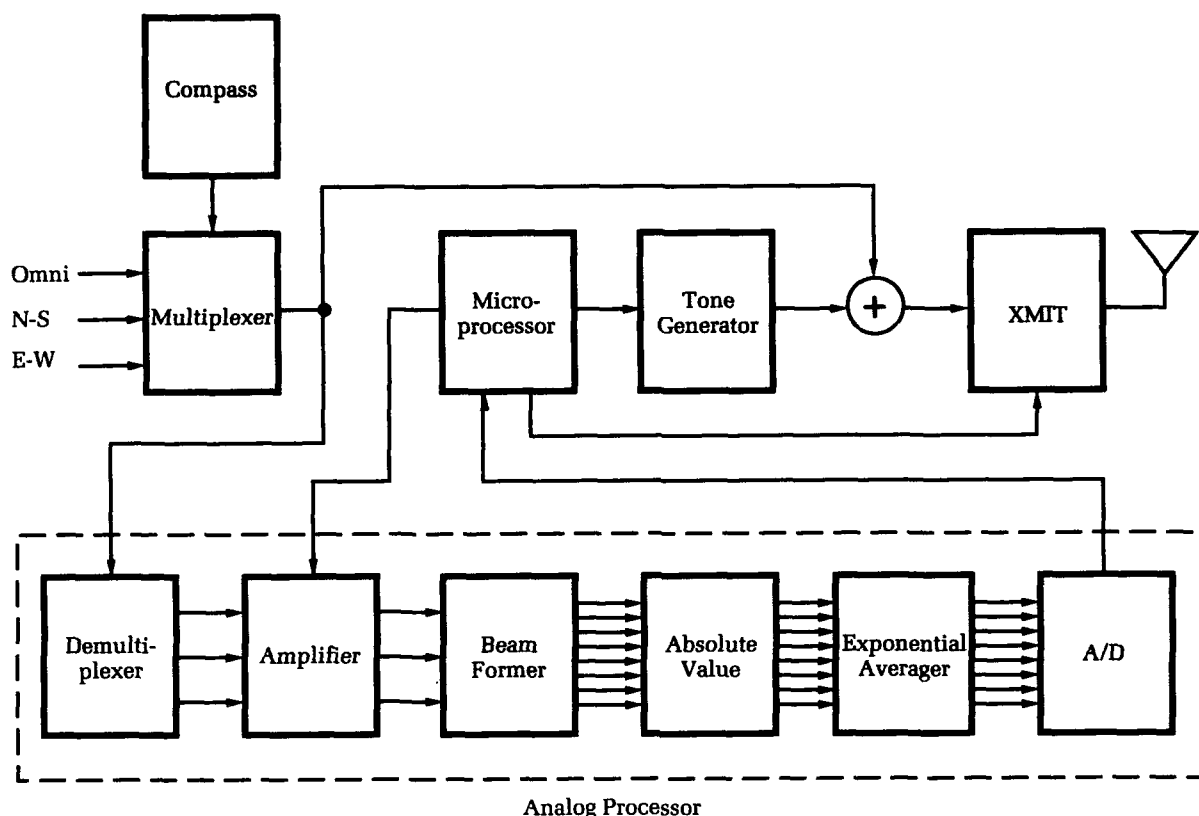


Figure 4. System block diagram, directional energy sonobuoy.

tion plus a specified period such as 15 minutes. This type of operation eliminates the large transmitter power requirement most of the time, and thus allows the sonobuoy life to be extended to several days without requiring a larger battery. Another benefit is that the number of sonobuoys deployed simultaneously is no longer limited to the total number of available radio frequency channels.

As shown in Figure 4, the directional-energy detector requires additional signal processing functions when compared to the DIFAR sonobuoy. The input to the additional functional blocks is the multiplexed DIFAR signal. It is first demultiplexed, resulting in signals similar to those that would be generated if the directional hydrophones were rigidly oriented in the North-South and East-West directions. In the amplification function, the three signals are filtered to pass only the frequency band where the target of interest radiates broadband energy, and interfering signals are minimized. An automatic gain control (AGC) is also added under microprocessor control. The gains of all three channels are identical and are controlled so that the analog-to-digital converter further down the signal flow will neither saturate nor operate at too low a level.

In operation, the beamformer follows the beam-forming equation given previously, and consists of operational amplifiers that form the required summations for eight beams spaced by 45 degrees. The absolute value function converts from amplitude to beam power. Ideally, this function would be provided by a squaring operation. However, squaring has been replaced by absolute value because of ease of implementation with little loss in performance.

The exponential averaging function is for the purpose of smoothing the eight beam powers so that data samples for the microprocessor can be at a manageable rate. With no smoothing, the input data rate would depend on the system bandwidth and could be of the order of 10,000 samples per second, a high rate for the microprocessor to handle. By choosing an exponential averager time constant of a fraction of a second, it is possible to sample each channel at a rate faster than once per second with no loss of information. In the system each of the eight channels is sampled at eight samples per second.

The analog-to-digital converter is of eight-bit resolution and operates on the eight input channels in a serial fashion with no sample and hold required. The AGC function ensures that the

input analog level is in the middle of the A/D converter range.

The microprocessor continuously receives as an input the eight smoothed beam powers. Its control functions are:

- Control of amplifier gain to provide an AGC function.
- Control of the transmitter that is turned on for a period after buoy deployment, during target detection, and for a period after target detection.
- Generation of a tone in a manner that conveys a target bearing after detection.

In terms of capability, the microprocessor can be any of a number of available eight-bit CMOS microprocessors. By use of a CMOS device the power drain is not excessive, and an eight-bit processor can handle the computational loading.

Detection Algorithm

The target-detection algorithm resides in the microprocessor. The input consists of beam powers smoothed over time in eight directions spaced by 45 degrees. The objectives are to detect a submarine that passes near the sonobuoy and to avoid false alarms caused by environmental effects or distant shipping. Theoretical analysis showed that if the detection is based on the ratio of the largest beam power to the average beam power, the system would not be affected by the background noise level in an isotropic noise environment. Furthermore, a fixed detection threshold could be used. Post-processing techniques could also be employed to ensure that the energy detections over time were from directions consistent with a target passing nearby. This section of the article describes the data flow in the microprocessor software.

Figure 5 is a block diagram that breaks the data flow into large functions and shows the data rates. The averager does averaging of the input beam powers in blocks of data, and has an

output every 30 seconds. It also performs two other functions. The first is to normalize the beam powers to make the numbers compatible with fixed-point arithmetic in the microprocessor. The second is to calculate a long-term average of each beam power and subtract it from the beam power to force the mean to become zero. This reduces the effects of hardware imbalance between the beams and helps to cancel effects of stationary directional-noise sources and distant shipping having low bearing rates.

The peak-to-average function uses as inputs the eight beam powers occurring every 30 seconds. The largest beam power is selected, and then a parabolic fit is performed using the largest beam and its two neighbors. This yields an interpolated bearing estimate, θ , and a ratio of peak power to average beam power, γ . Since under any condition one beam power will be a maximum, these two quantities can always be calculated. With no target present, γ will have a value close to 1, and θ will show no consistency if observed over time. This will also be true for a very distant target, since the averaging function will have removed the effects of stationary noise sources prior to calculating γ and θ . A target passing nearby will cause γ to increase in value and θ to show a progressive change in bearing.

The energy detector function uses a history of γ values to set a threshold for energy detection. By exponential averaging, an estimate is made of the mean and variance of the γ variable. A threshold is set at the mean plus twice the standard deviation. Each new value of γ is used to update the long-term exponential averages, and is also compared to the threshold for declaration of an energy detection. The reason for the use of an adaptive threshold rather than a fixed threshold is to model the system sensitivity to the specific acoustic environment.

The θ estimates, which occur every 30 seconds, represent directions from which peak power is received. A least-squares fit is performed on a

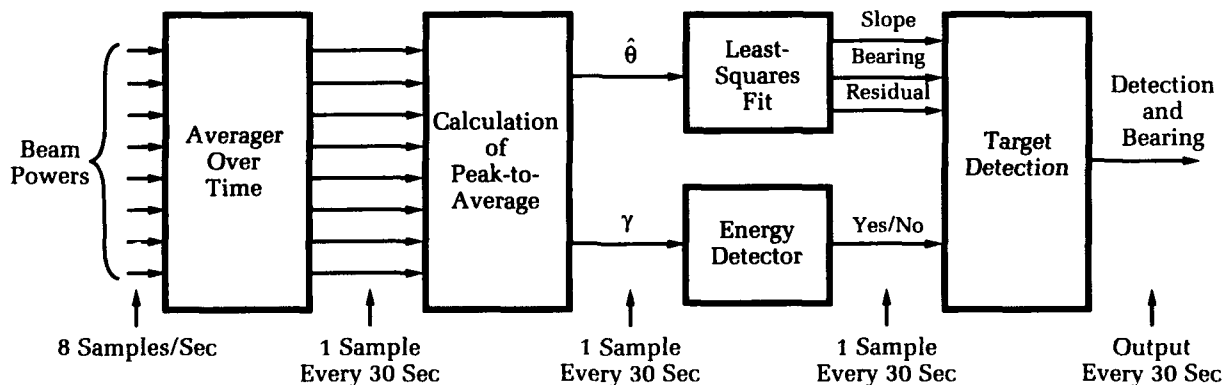


Figure 5. Functional diagram of detection algorithm.

The final function shown in Figure 5 is target detection. The inputs used in the detection decision are the bearing, slope, and residual from the least-squares fit and a record of energy detections. For a detection to be declared, all of the following criteria must be met:

-

102

The test for target detection is performed every 30 seconds. Upon target detection, the RF transmitter is turned on to transmit raw acoustic data and a message to indicate the bearing to the detected target. The transmission continues for 15 minutes after the last detection made within the sonobuoy.

Performance

The Dahlgren Division has conducted extensive tests of the directional-energy detector in the laboratory. Input data were mostly from recordings of signals from DIFAR sonobuoys used in ocean tests and from operational aircraft. These inputs were supplemented with data generated by acoustic simulators. The algorithm was first implemented on a general-purpose computer, and later on a 6805 microprocessor. In addition to being tested in the laboratory, the directional-energy detector was included as part of a prototype sonobuoy design for the Tactical Surveillance Sonobuoy project and was tested at sea.

The test program demonstrated that the design goals of the project have been met. The sensitivity of -12 dB was achieved in many acoustic environments, although the presence of noise from heavy shipping caused the sensitivity to degrade. It was difficult to establish the probability of false alarm without having a massive data base. However, the number seems to fall between 10^{-3} and 10^{-4} , indicating a false alarm every 25 hours on the average for a single sonobuoy. At this rate it would be possible for a search platform to monitor a field of 100 sonobuoys.

The feature of the directional-energy detector that leads to good performance in the underwater acoustic environment is the ability to reject signals from distant shipping, which tend to change rapidly in strength. This change in signal strength can be due to a ship's entering or leaving a convergence zone. Since this situation is common, a system having a low probability of false alarm must be insensitive to it. An omnidirectional energy detector can seldom discriminate between such a signal and one from a nearby

object. Figure 6 illustrates how the directional-energy detector reacts to this problem. The signal from the distant ship grows rapidly in strength and results in a number of energy detections. However, the estimate of bearing rate of change shows very little motion, and no detection is declared. The nearby submarine may have smaller signal strength, but the rate of change of bearing passes the threshold, resulting in a detection. Most false alarms observed in the tests conducted on this system were due to combinations of distant ships where changing signal strengths were interacting to give the impression of a high bearing rate of change.

Conclusions

The directional-energy detector may have applications in future submarine detection systems. Because the concept depends on dynamics in the input data and rejects stationary signals, it is applicable to distributed systems where many relatively insensitive sensors are used rather than a smaller number of sensors that can detect weaker signals. In applications such as shallow-water antisubmarine warfare, propagation effects may require the use of a distributed system, in which case the directional-energy detector should be considered.

The Author



THOMAS B. BALLARD received a B.S. degree from Rensselaer Polytechnic Institute, an M.S. from the California Institute of Technology, and a Ph.D. from the University of Maryland in electrical engineering. After working for ten years in the field of communications at NASA, he joined the Underwater Systems Department in 1974, where he has specialized in acoustic detection and

tracking systems as applied to surface ship and airborne antisubmarine warfare.

Wigner-Based Cycle-Rate Detection... Testing on MFSK Acoustic Data

JoEllen Wilbur

Detection and evaluation of underwater acoustic transmissions can be particularly difficult for underwater data telemetry, low-probability-of-intercept sonar, and other sonar transmissions that are obscured by noise. This article describes a novel signal-processing procedure for cycle-rate detection of low signal-to-noise ratio (SNR) frequency and phase-modulated signals. The detection procedure described herein exploits the fact that these signals maintain some form of constant-rate modulation. This constant-rate modulation yields a spectral redundancy in the data that arises at regular intervals. The cycle-rate detector performs a mapping of the data to a signal domain in which stationary noise is mapped to the origin, allowing separation of signal and noise sources that overlap in both time and frequency. Cycle tone detection is achieved when direct cycle-frequency computation is applied to frequency-hopped, pulse-train acoustic data with the signal buried 6 dB deep in noise.

Introduction

Efficient detection and estimation of pulse-modulated acoustic sonar transmissions have application to underwater data telemetry, low-probability-of-intercept sonar, and other sonar transmissions that are obscured by noise. A common denominator to signals of this type is the presence of some form of constant-rate modulation inherent in the signal. Examples of such signals include sonar signals that have undergone preprocessing in the form of multiplexing and modulation including frequency shift keyed (FSK), binary and quadrature phase shift keyed, pulse amplitude, pulse width, pulse-position modulated signals, and pulse-generated noise bursts.

Two common methods of introducing randomness to transmitted signals are phase shift keying and frequency shift keying. A random-phase/shift-keyed signal consists of a simple sine-wave signal that contains phase shifts (often 180-degree shifts) at periodic locations. For example, after every five complete cycles of the sine wave, a random decision might be made to invert or not to invert the signal. A random frequency shift-keyed signal is generated by making a random decision every millisecond to switch or not to switch between two frequencies of transmission. These examples correspond to biphase and bifrequency cases, but random selections between more than two phases or frequencies are typically used.

Such preprocessing of sonar signals has various applications. For example, low-probability-of-intercept transmitted signals are made less detectable to a

threat when given a pseudorandom character with a large time-bandwidth product. The intended recipient knows the pseudorandom pattern and is able to effectively recombine the signal in such a way as to recover it from below the noise floor, whereas the signal appears noiselike to an unintended receiver or threat receiver with no knowledge of the pseudorandom pattern.

This article describes a novel cycle-rate processing procedure for low SNR modulated transmissions developed at the Coastal Systems Station, NSWC Dahlgren Division, under an Independent Research program sponsored by the Office of Naval Research. The cycle-rate detector is designed to key on the constant-rate modulation inherent in artificially generated, noise-like signals. The constant-rate modulation in frequency- and phase-modulated transmissions causes cyclostationarity in the signal transmission. A cyclostationary signal, by definition, maintains cyclically correlated spectral spreading that can be identified, given proper transformation of the data.^{1,2,3}

The theoretical basis behind the processing procedure is reviewed. Relevant aspects of the Wigner distribution (WD) with regard to acoustic transmissions are discussed. The processing procedure is based on a mapping directly from the Wigner inner-product function to a frequency-correlation domain. The relation of the WD to the time-variant cyclic spectrum is provided. Fourier transformation of the pseudo-WD over window centers is shown to yield the cyclic correlations that correspond to the signal modulation rate. A direct cycle-frequency detection pro-

cedure, based on the theory provided herein, is described for low SNR frequency-hopped and diversity transmissions. The cycle-rate detector is then applied to frequency-hopped acoustic data collected at the White Oak Detachment's Hydroacoustic Measurements Facility, and shown to perform well in -6 dB SNR.

Theory

The WD of a signal $x(t)$ is a bilinear t to $t-\omega$ transformation of the form⁴

$$W_{xx}(t, \omega) = F_{\tau \rightarrow \omega} \left\{ x\left(t + \frac{\tau}{2}\right) x\left(t - \frac{\tau}{2}\right) \right\} \quad (1)$$

for $F_{\tau \rightarrow \omega}\{\bullet\}$ denoting Fourier transformation in τ . The WD defines a time-frequency estimator that provides a representation of signal concentration over the $t-\omega$ plane. Figures 1 and 2 give examples of the WD of respective frequency- and phase-modulated signals. Figure 1 is the surface and contour plot of a frequency shift-keyed data sequence where the frequency is randomly shifted over time between two carriers. Figure 2 gives the contour plot of a binary phase shift-keyed data sequence where the phase of a single carrier has been randomly switched between two values separated by 180 degrees to effect a large time-bandwidth product transmission. The spectral dispersion characteristic of the binary phase shift-keyed data sequence can be seen in the WD at the points in time where phase reversal occurs. The power spectral density can be obtained directly from the WD by integrating in the t direction.⁴

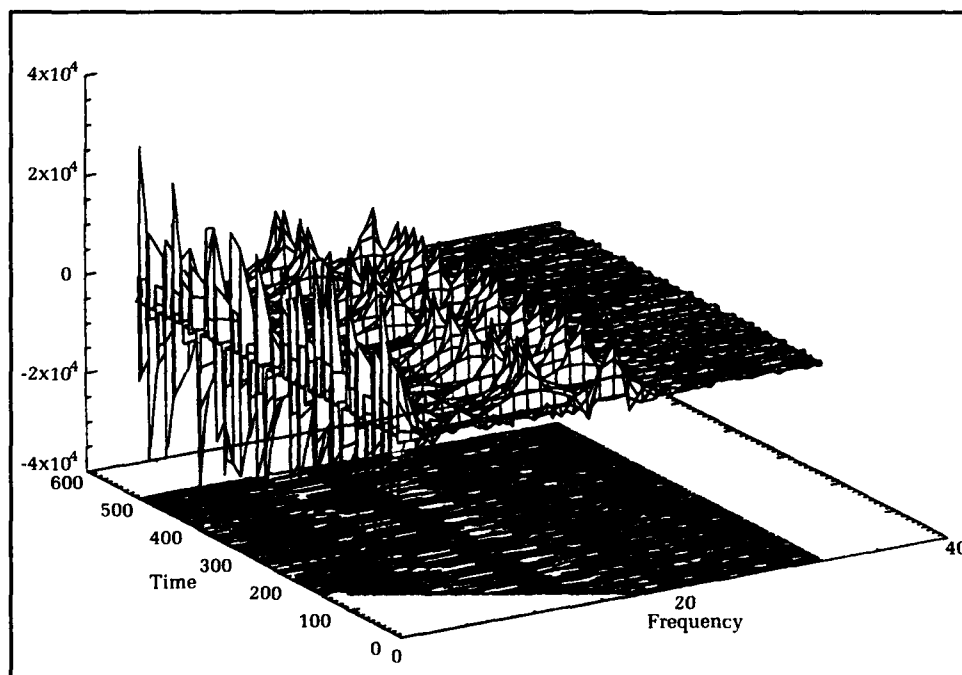


Figure 1. Wigner distribution of frequency shift-keyed data sequence where pulses are randomly hopped between two carrier frequencies.

The application of the WD to acoustic returns can be interpreted by modeling each received waveform $x(t)$ as the output of a linear time-variant random filter with input $s(t)$ such that

$$x(t) = \int_{-\infty}^{\infty} s(t-u)h(t,u)du.$$

If $h(t,u)$ is linear time-variant wss, then¹

$$\langle W_{xx}(t, \omega) \rangle = \frac{1}{2\pi}$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I_{ss}(u, \tau) R_{HH}(\tau, \sigma) e^{j(\sigma\tau - \sigma u - \omega\tau)} du d\sigma d\tau \quad (2)$$

$$I_{ss}(u, \tau) \doteq s\left(u + \frac{\tau}{2}\right) s^*\left(u - \frac{\tau}{2}\right)$$

where

$$R_{HH}(\tau, \sigma) = \langle H\left(\frac{\tau}{2}, \frac{\sigma}{2}\right) H^*\left(-\frac{\tau}{2}, \frac{\sigma}{2}\right) \rangle$$

is the autocorrelation of the linear time-variant channel transfer function. Taking the mean of the WD definition in Equation (1) yields

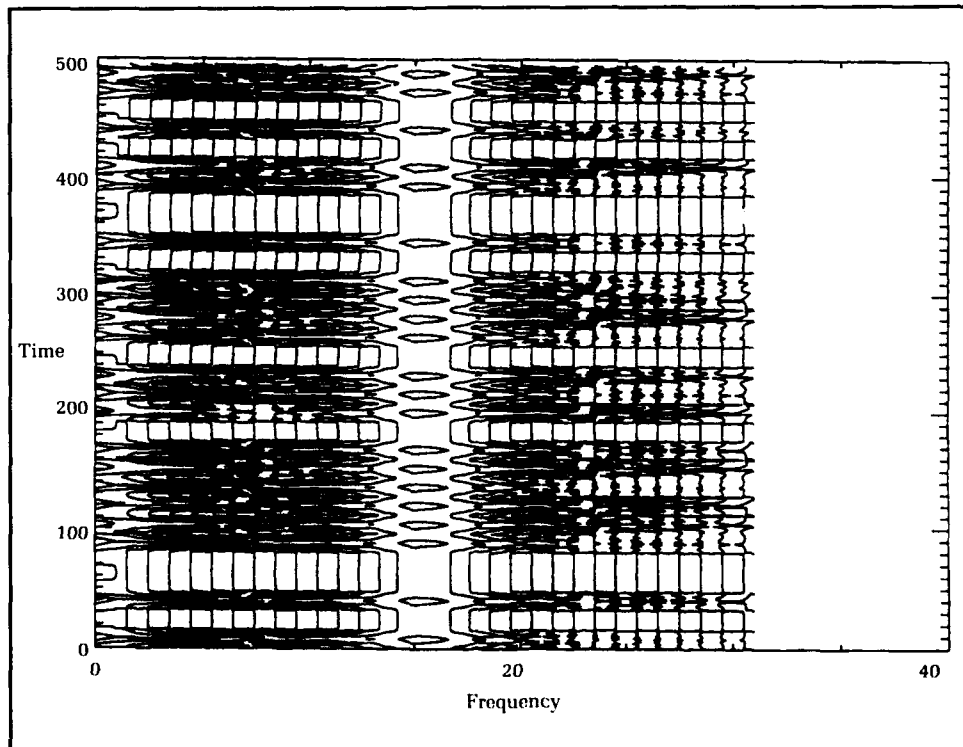
$$\langle W_{xx}(t, \omega) \rangle = F_{\tau \rightarrow \omega} \{ R_{xx}(t, \tau) \}, \text{ where}$$

$$R_{xx}(t, \tau) = \langle x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) \rangle$$

is the instantaneous autocorrelation defined for signal reception over dispersive communication channels, and whose transform $S_{xx}(t, \omega)$ is the instantaneous power spectral density.

A nonstationary signal, by definition, maintains correlated spectral spreading corresponding to the time-variant nature of the signal autocorrelation. When stationary analysis is applied to nonstationary acoustic processes, different signal sources may produce the same power-spectral density, as the power-spectral density masks the relative dependence on any correlation in the spectral spreading of the signal. For example, suppose we have a narrowband message $x(t)$ of length T_x that has been spread via a spreading function, $c(t)$, with bandwidth $B_c \gg B_x$, where B_x is the bandwidth of the original process. The resulting waveform after spectral spreading of the signal appears stationary over the signal length, indicating uncorrelated spectral spreading. In the power-spectral density, the signal appears as a wideband process where a traditional spectral coherency analysis may indicate a high correlation between the artificially spread signal and another wideband, noise-like process. If, for example, $s(t) = x(t)c(t)$, then even though $s(t)$ is effectively a wideband process, $s(t)$ exhibits high correlation among spectral components with spacing $2\omega_x$ (where ω_x denotes the center frequency of the narrowband signal $x(t)$). A signal of this form is an example of a spread-spectrum signal. A spread-spectrum signal employs a pseudorandom spreading sequence to effect a large time-bandwidth product on the

Figure 2. Contour plot of the Wigner distribution of a binary phase shift-keyed data sequence where the phase of a single carrier frequency has been randomly switched between two values separated by 180 degrees. Spectral dispersion can be seen to occur at points in time where the signal undergoes a phase reversal.



signal and is an example of the class of signal types termed cyclostationary.

The time-variant cyclic spectrum defined for cyclostationary processes^{2,5} can be shown to equate to a frequency-translated convolution between the WD time-evolutionary spectrum of the signal and a rectangular pulse $h(t)$ as below³

$$S_{xx}^{\sigma}(t, \omega) = \frac{1}{T} \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W_{xx}(u, \xi) e^{-j2\pi\sigma u} W_{rr}(t-u, \omega-\xi) du d\xi \quad (3)$$

where

$$W_{ff}(t, \omega) \doteq F_{\tau \rightarrow \omega} \left\{ f \left(t + \frac{\tau}{2} \right) f^* \left(t - \frac{\tau}{2} \right) \right\}; f = x, r. \quad (4)$$

$W_{xx}(t, \omega)$ and $W_{rr}(t, \omega)$ define the respective WD functions of the signal and a rectangular pulse of length T .

Evaluation of the time-variant cyclic spectrum along the center of the moving average estimate can be shown to reduce to³

$$C_{xx}(\sigma, \omega) = \int_{-\infty}^{\infty} e^{-j2\pi\sigma t} W_{xx}^p(t, \omega) dt \quad (5)$$

where $W_{xx}^p(t, f)$ is the pseudo-WD defined below

$$W_{xx}^p(t, \omega) = F_{\tau \rightarrow \omega} \left\{ x \left(t + \frac{\tau}{2} \right) x^* \left(t - \frac{\tau}{2} \right) h \left(\frac{\tau}{2} \right) h \left(-\frac{\tau}{2} \right) \right\}. \quad (6)$$

Transformation of the pseudo-WD over time maps the stationary noise, which has uncorrelated spectral spreading, to $\sigma \equiv 0$. Nonstationary noise is distributed randomly. When a cyclostationary signal is embedded in the noise, strong spectral peaks arise in the frequency correlation domain at integer multiples of the signal cycle rate.

A cyclostationary signal has cyclically correlated spectral spreading. The binary phase shift-keyed data sequence in Figure 2 is a common example of a direct-spreading sequence employed in spread-spectrum systems. Fourier transformation of the WD over time can be seen in Figure 3 to map the t - ω signal representation to a spectral correlation domain that identifies any correlated spectral spreading in the signal.³ If a signal is stationary, Fourier transformation will map the WD onto a line along the transformed axis that corresponds precisely to the power-spectral density.³ If, however, the signal is nonstationary, the transformed plane will yield spectral peaks where the signal exhibits high correlation between spectral components. If

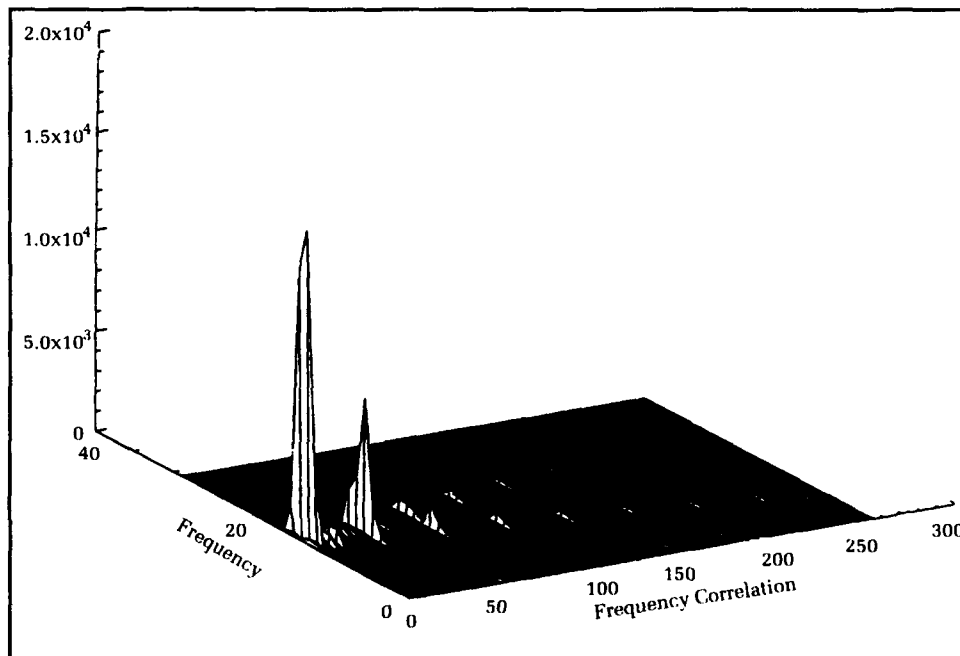
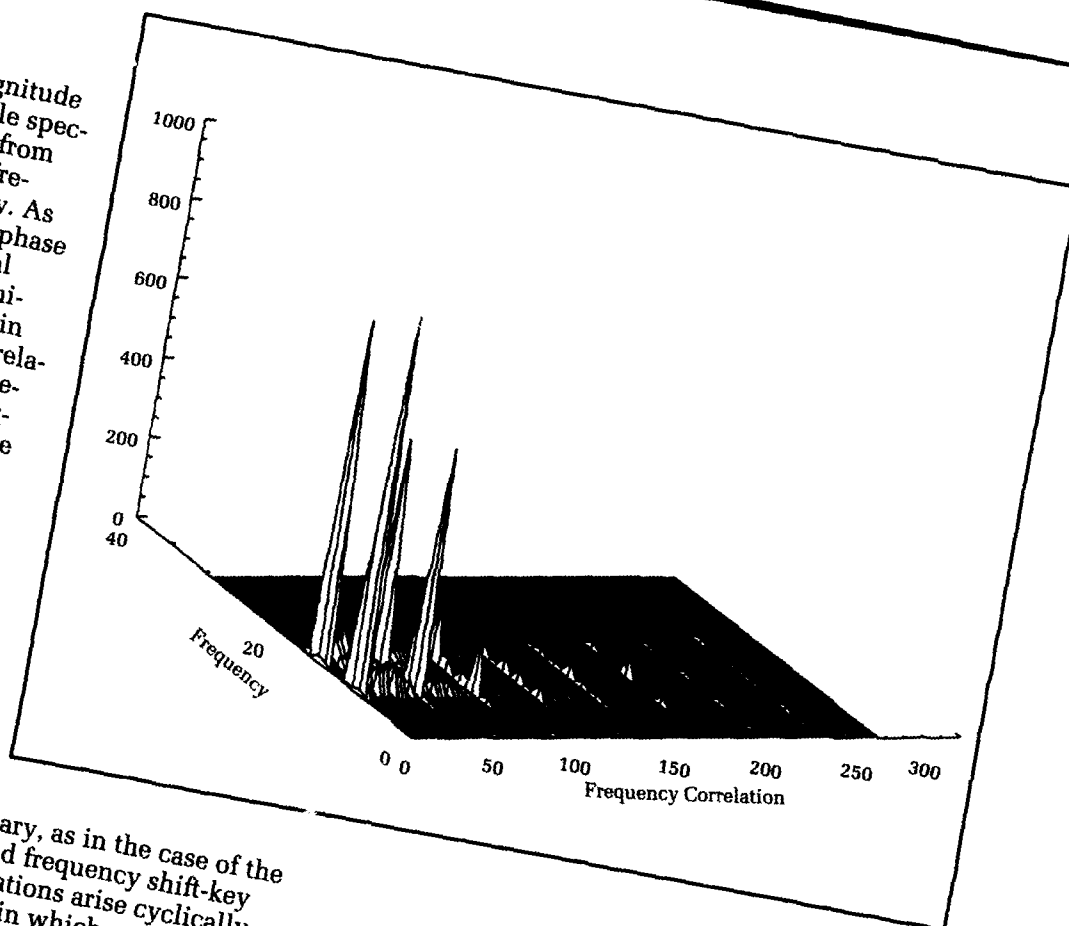


Figure 3. Magnitude plot of frequency-frequency correlation or "cycle spectrum" of binary phase shift-keyed data obtained through Fourier transformation along the time direction of the WD. Integration along frequency will yield a line spectrum or "cycle frequency" plot where the spacing between lines corresponds to the signal cycle rate.

Figure 4. Magnitude plot of the cycle spectrum obtained from the WD of the frequency shift key. As with the binary phase shift key, spectral peaks arise at equispaced distances in the frequency correlation direction corresponding to the signal cycle rate. In the transformation the power spectral density of the frequency shift key is mapped to the zero frequency correlation axis.



the signal is cyclostationary, as in the case of the binary phase shift-key and frequency shift-key examples, spectral correlations arise cyclically to yield a cycle spectrum in which spectral correlation peaks arise at integer multiples of the signal modulation rate. Figures 3 and 4 are the respective cycle spectrum plots of the binary phase shift-key and frequency shift-key examples obtained by Fourier transformation along the time direction of the pseudo-WD. Spectral peaks along the frequency correlation axis are evenly spaced a distance equivalent to the respective keying rates.

Cycle Rate Detection Processing

Direct cycle frequency estimation is computed as below,⁶

$$\begin{aligned}\tilde{S}_{xx}^i(\sigma) &= \sum_{i=1}^I \tilde{S}_{xx}^i(\sigma) \\ \tilde{S}_{xx}^i(\sigma) &= \sum_{n=0}^{N'-1} e^{-j\frac{2\pi\sigma n}{N'}} \sum_{k=K_{dc}}^{\frac{N}{2}-1} W_{xx}(n, k)\end{aligned}\quad (7)$$

where

for N' the number of window center positions, I the number of segmented blocks in the signal, K_{dc} the upper cutoff of a low frequency block and $W_{xx}(n, k)$ the generalized discrete WD which $L=1$ reduces to the pseudo-discrete WD counterpart to Equation (6). The smoothing operation on the discrete-WD inner product defined in Equation (8) provides noise immunity and cross-term reduction.

Given the Fourier-transform relation in mapping from the discrete-WD to the cycle spectrum, it becomes clear that inner product averaging corresponds to filtering of the cycle frequencies. More specifically, given $DFT_{a \rightarrow b}^P\{\cdot\}$ denotes a P

$$W_{xx}(n, k) = \sum_{m=-\frac{N}{2}+1}^{\frac{N}{2}-1} p[m]^2 i_{xx}(n, m) e^{-j\frac{2\pi km}{N}} - R_{xx}(0)$$

$$i_{xx}(n, m) = \frac{1}{2L-1} \sum_{l=-L+1}^{L-1} w(l) x(n+l+m) x^*(n+l-m)$$

$$R_{xx}(0) = \frac{1}{2L-1} \sum_{l=-L+1}^{L-1} g(l) |x(n+l)|^2 \quad (8)$$

point discrete-Fourier transform from a to b, the cycle spectrum can be rewritten as

$$S_{xx}(\sigma) = G_{2L}(-\sigma) \sum_{k=K_{dc}}^{N-1} S_{xx}^k(\sigma)$$

$$G_{2L}(-\sigma) = \frac{1}{2L-1} \sum_{l=-L+1}^{L-1} g(l) e^{j \frac{2\pi \sigma l}{N'}} \quad (9)$$

$$S_{xx}^k(\sigma) = DFT_{m \rightarrow k}^N \left\{ DFT_{n \rightarrow \sigma}^{N'} \left\{ x(n+m) x^*(n-m) \right\} \right\}$$

where $S_{xx}^k(\sigma)$ is the cycle spectrum estimate obtained via Fourier transformation over the unsmoothed discrete-WD ($L=1$).

Transformation of the discrete-WD across window centers has been interchanged with the summing operation over frequency, reducing the number of fast-Fourier transform operations from $N+N'+1$ to $N+2$. (In the case of phase-modulated signals, the cycle frequency estimate requires summation over the cycle spectrum magnitude, in which case the order of the summations cannot be exchanged.) Application of the low-frequency block is used to eliminate the DC cross-term contributions associated with the discrete-WD computation. Figure 5 illustrates the direct-cycle frequency estimate of the dry-end transmission for four records of nonrepeat-

ing 8-ary frequency-keyed pulses. The stationary part of the signal contribution, i.e., $\sigma \approx 0$, has been removed from the cycle-frequency estimate. The regular repetition of the keying pattern produces cyclic correlations that arise as dominant lobes at integer multiples of the cycle frequency or, equivalently, the reciprocal of the keying rate. The WD of the signal transmission is given in Figure 6. Figures 7 and 8 illustrate the respective time history and cycle frequency estimate of the received test data in -6 dB SNR. Here the data was transmitted between two transducers at high SNR, then added to a noise data file collected over the same passband as the 8-ary frequency shift key, allowing control over the SNR. Figure 9 gives the magnitude-squared, fast-Fourier transform commonly used to estimate the power-spectral density. The signal can be seen to lie below the noise floor.

Concluding Remarks

A signal-processing procedure for cycle-rate detection of low SNR frequency- and phase-modulated signals, which performs a mapping from the Wigner distribution inner-product function to a signal cycle frequency domain, has been described. Fourier transformation of the WD over temporal window centers equates to the time-variant cyclic spectrum evaluated along the center of a moving average estimate and serves as the basis for the cycle-rate processing procedure. Cycle-rate estimation is achieved via

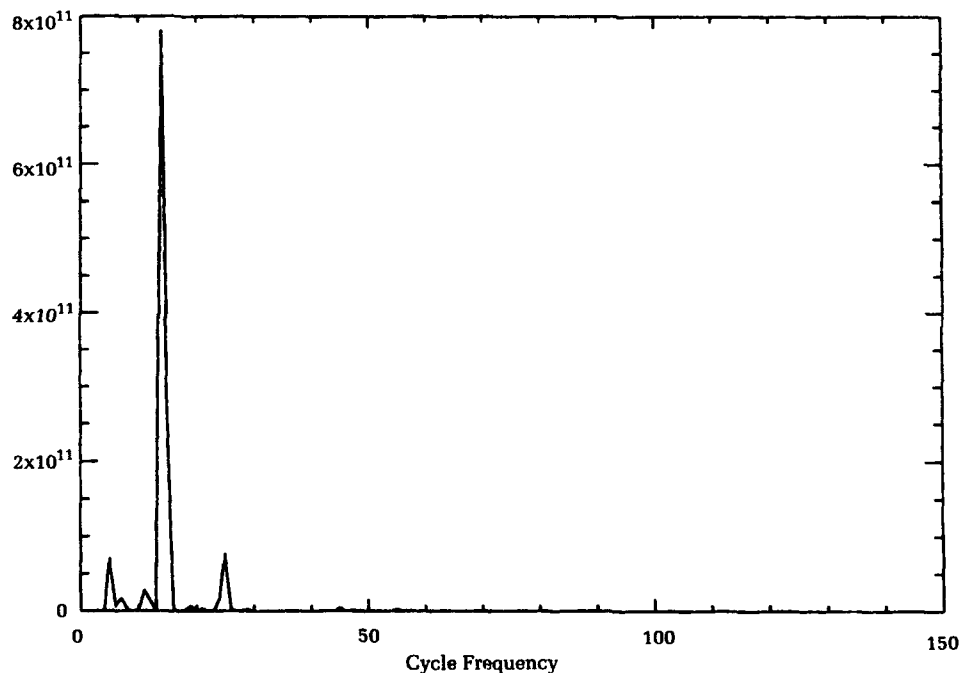


Figure 5. Direct-cycle frequency estimate of transmitted 8-ary frequency shift keyed sonar data (dry end). The stationary part of the signal, which maps to the zero correlation axis, has been removed. ($N=64$, $L=2$, $I=4$, $K_{dc}=1$.)

Figure 6. Discrete Wigner distribution of one record of 8-ary frequency shift-keyed data.

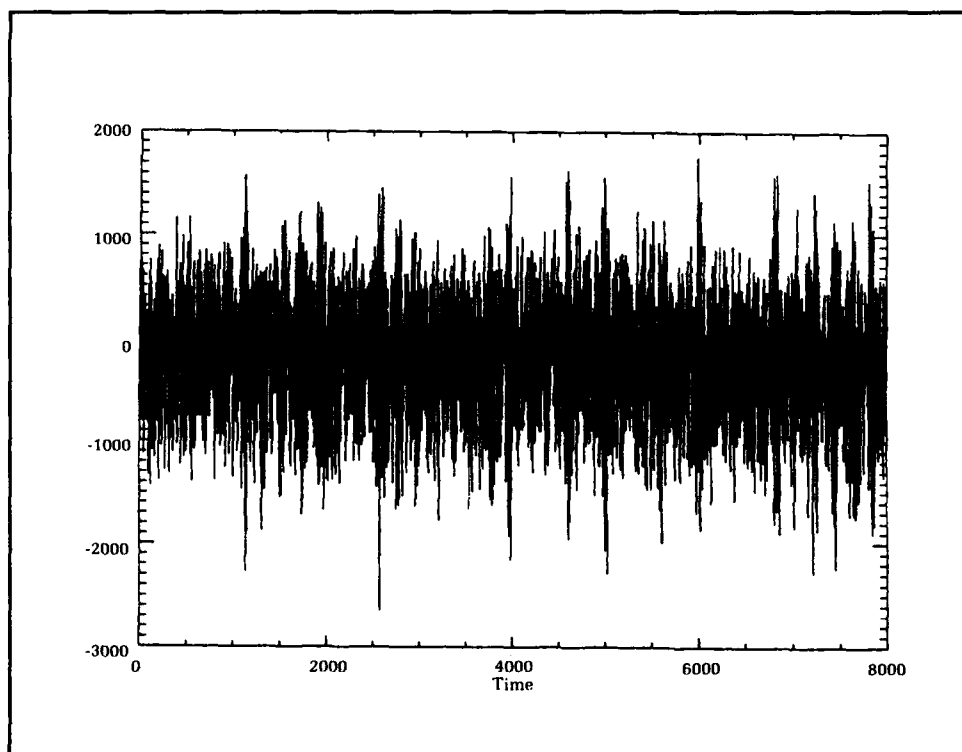
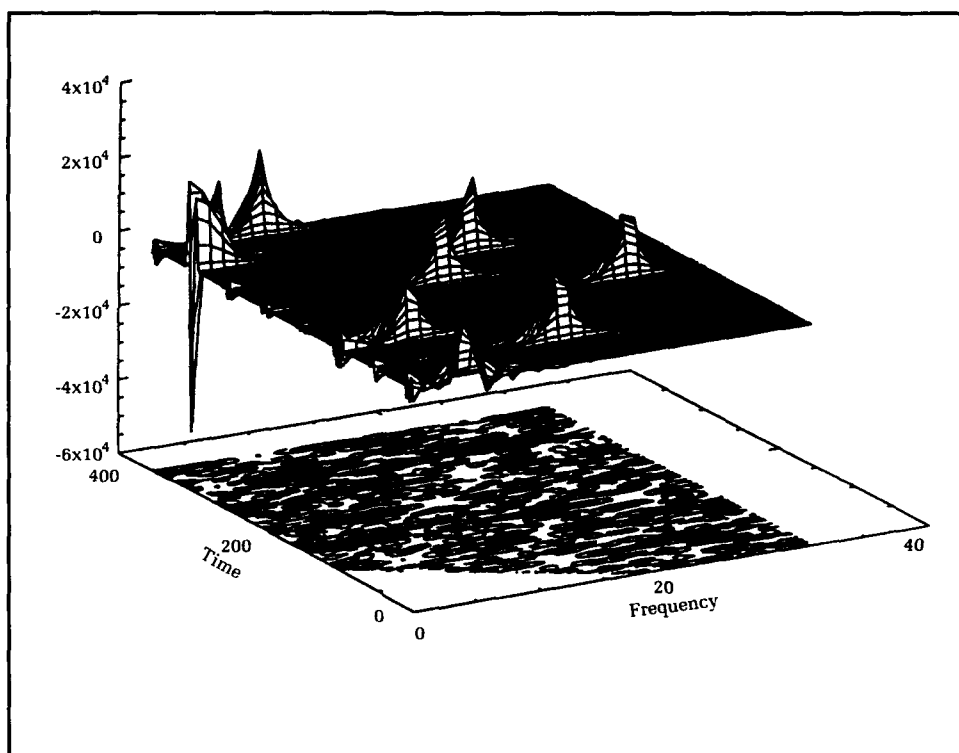


Figure 7. Time history of 8-ary frequency shift-keyed sonar data for -6 dB SNR.

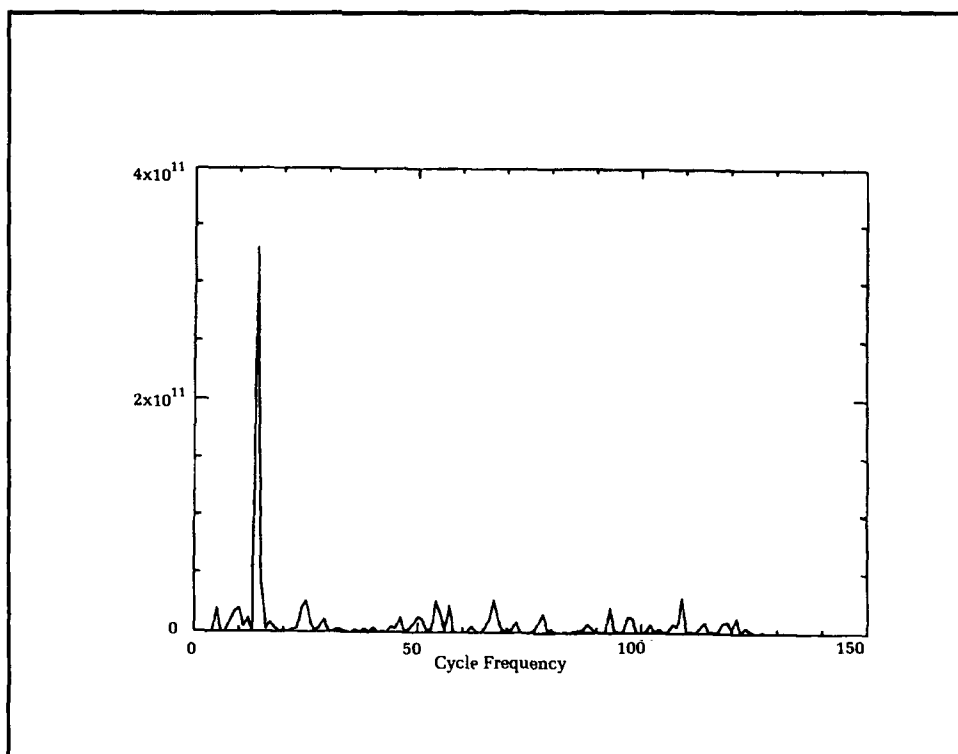
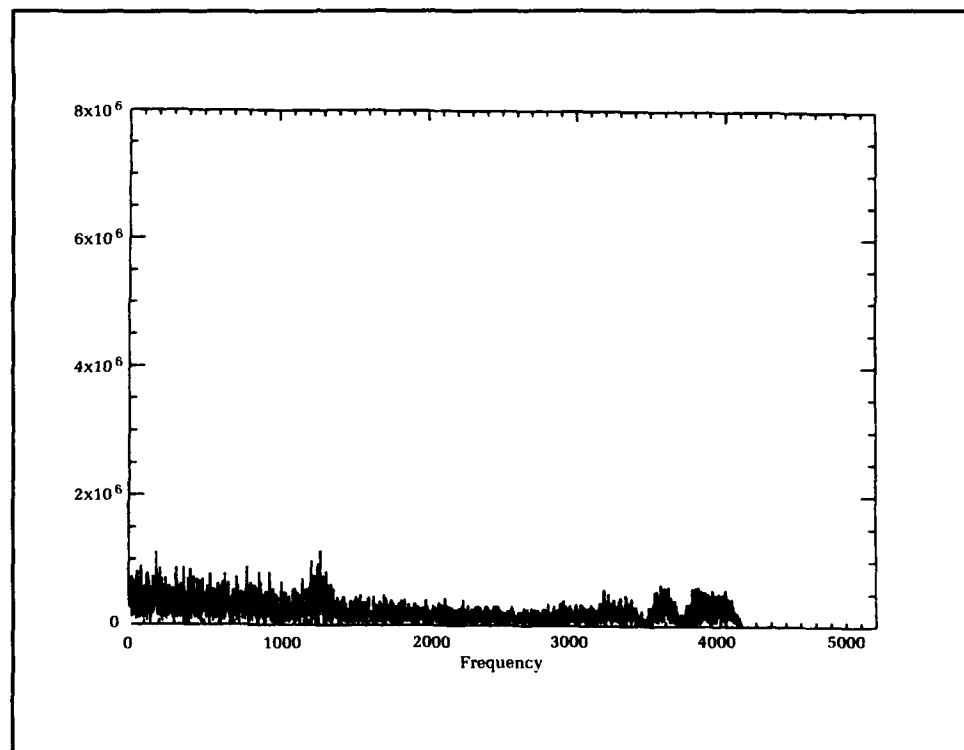


Figure 8.
Cycle-frequency
estimate of received
8-ary frequency
shift-keyed sonar
data for -6 dB SNR
[nonstationary part]
($N=64$, $L=2$, $I=4$,
 $K_{dc}=1$.)

Figure 9.
Magnitude squared
fast Fourier trans-
form of 8-ary fre-
quency shift-keyed
sonar data in -6 dB
SNR.



a weighted-average discrete-WD inner-product operation at the input to a discrete-Fourier analyzer. The weighted average operation on the discrete-WD inner-product function equates to filtering of the cycle frequencies. A low-frequency block followed by a running sum is performed at the output of the fast-Fourier transform analyzer for each windowed time-slice of the data. Fourier transformation across overlapping blocks of the summed outputs is continually averaged until cycle-tone detection occurs. The processing procedure was able to extract cycle tones when applied to frequency-hopped data in -6 dB SNR.

References

1. Wilbur, J. and Bono, J., "Wigner/Cycle Spectrum Analysis of Spread Spectrum and Diversity Transmissions," *IEEE J. Ocean Engineering* 16, Jan, 1990, pp. 98-106.
2. Gardner, W. A., "Exploitation of Spectral Redundancy in Cyclostationary Signals," *IEEE Sig. Proc. Magazine*, 8, 2, Apr 1991, pp. 14-36.
3. Wilbur, J. and McDonald, R. J., "Nonlinear Analysis of Cyclically Correlated Spectral Spreading in Modulated Signals," *J. Acoust. Soc. Am.*, Jul 1992.
4. Claasen, T. A. C. M. and Mecklenbrauker, W. F. G., "The Wigner Distribution—A Tool for Time-Frequency Signal Analysis, Parts I and III," *Philips, J. Res.*, vol. 35, 1980, pp. 217-389.
5. Gardner, W. A., *Statistical Spectral Analysis: A Nonprobabilistic Theory*, Prentice Hall, Inc., New Jersey, 1988.
6. Wilbur, J., "Cycle Frequency Estimation of Time/Frequency Diversity Transmissions in a Dense Noise Environment," *Proc. 1992 IEEE Int. Conf. Acoust., Speech, Sig. Processing*, 13-P, Mar 1992.
7. Jacobson, L. and Wechsler, H., "The Composite Pseudo Wigner Distribution (CPWD): a Computable and Versatile Approximation to the Wigner Distribution (WD)," *Proc. 1983 IEEE Int. Conf. Acoust., Speech, Sig. Processing*, 1, 1983, pp. 254-256.
8. Flandrin, P., Martin, W., and Zakharia, M., "On a Hardware Implementation of the Wigner-Ville Transform," *Dig. Sig. Processing-84*, Cappelini, V. and Constantinedes, A. G., Eds., Elsevier, 1984.

The Author



JoELLEN WILBUR received a B.S. degree in electrical engineering from the Engineering Honors Program at Virginia Tech in 1982, and M.S. and Ph.D. degrees in electrical engineering from the University of Florida in 1984 and 1987, respectively. In 1984-1985 she was a recipient of a CAD/CAM Institutional Graduate Fellowship, and in 1986-1987 she received an Army Research Office Graduate Fellowship. From

1987 to 1989, she was an assistant professor in the Department of Electrical and Computer Engineering at Clemson University. During 1988-1989, she served on the Executive Committee of the Piedmont Section, IEEE. She has been employed with the Signal, Sonar & Image Processing Branch at the Coastal Systems Station in Panama City, Florida since 1989, where she has engaged in research in underwater acoustic signal processing, with special interest in statistical spectral analysis, time-variant and cyclostationary signal processing for signal parameter estimation, LPI/LPE signal design/interception, and signal processing for mine classification. She is a member of Eta Kappa Nu, Institute for Electrical and Electronic Engineers, and American Institute of Physics. She is currently serving a three-year appointment on the Statistical Signal & Array Processing Technical Committee of the IEEE Signal Processing Society.

Transient Resonance Scattering for Active Acoustic Classification

Guillermo C. Gaunaud

The ability to classify underwater objects is important to Navy programs such as mine countermeasures and antisubmarine warfare to preclude wasting ordnance on false targets. The echoes reflected by a "ping" emerging from an active sonar when it interacts with a target in its path can be remotely sensed by a receiver. The approach presented here capitalizes on an inverse-scattering analysis that exploits certain resonance features useful for target classification in the returned echoes. Over many years, the methodology has emphasized the extraction, isolation, and labeling of the resonance features hidden within the echo. Little has been said about how these features could be used to classify a target. Although the classification of complex shapes is still a formidable task, considerable progress has been made in classifying simple shapes such as spheroidal and cylindrical shells. In the following, we will illustrate with these shapes exactly how the resonances can be linked to the physical characteristics of the target that returned the echo, thus allowing for its unambiguous characterization. The procedure, here illustrated with active acoustics (sonar), can be extended to any active return from any sensor, including radar.

Introduction

The study of target characteristics permits us to understand how a submerged elastic target behaves under acoustic interrogation. This behavior is analytically described by the object's scattering cross section (SCS) or its square root, called the "form function." The literature on techniques to predict acoustic SCS of impenetrable (i.e., rigid/soft) targets of simple shapes is quite extensive. If the scatterers are penetrable (i.e., elastic or admitting internal fields coupled to the external fields through boundary conditions on the object's surface), then the literature reduces considerably to the study of the only three shapes for which the partial differential equations governing the scattering process admit "separability," i.e., spherical, cylindrical, and flat layers. For any other shape, only numerical solutions are possible. The SCS of an elastic target in water differs substantially from that obtained by assuming the body is impenetrable (e.g., rigid). Thus, in order to predict accurately its SCS, we must account for the elastic composition of any underwater target and for its fluid-loaded condition. The structural response of an object in water is quite different from its response in vacuum or in air. Otherwise, gross errors are introduced that are unacceptable. Finally, we underline that the prediction of an object's SCS tells us only how the body scatters the known waves that are

incident upon it. It only solves the "direct-scattering problem." Active classification is considerably more difficult, since it attempts to identify the scatterer's shape and composition from the echoes it returns under various, known waveform interrogations. In short, the goal is the solution of the "inverse-scattering problem."

The solution of an inverse-scattering problem here will mean the extraction of physical information, namely composition and shape, from the active echoes returned by an underwater elastic target. This pertains to the general area of target identification. In this sense, solving an inverse-scattering problem may be relatively simple in the low- or high-frequency regions of the target's spectra because of the number of simplifying approximations that can usually be made in these two spectral regions. However, for the intermediate or resonance region of all scatterers, the above mentioned approximations usually do not hold; therefore, either exact solutions or novel types of approximations are required. This makes the inversion of the scattered waveforms extremely difficult. A certain level of success has been achieved by exploiting the presence of certain resonance features that manifest themselves in the resonance region of the scattering cross section of submerged elastic targets. Examples of how and how well one can extract identifying sets of resonance features from an echo have recently been reviewed.¹ Although the main interest over the years has been the isolation or extraction of the resonance features contained within the SCS of a penetrable target,

little has been said about what to do with these resonances once they are isolated. This resonance extraction process can be carried out theoretically, experimentally, and computationally by means of either continuous wave (CW) or pulsed interrogating waveforms. We show here some idealized examples to illustrate how the resonances, once they are isolated by either means, can be associated to the physical characteristics of the target that returned the echo, thus allowing for unambiguous target characterization. We will use an elastic shell to illustrate wave interaction and resonances, followed by a historical discussion of the Direct Scattering Problem, and finally illustrate how the physical characteristics of the target can be obtained from the returned echoes.

Elastic Shells Submerged in Water

When an acoustic beam insonifies a submerged elastic shell (Figure 1), the scattered echo returned to the sending sonar contains a number of "spiky" resonance features superimposed on a smooth "background." The echo one talks about here is the form function, $|f_{\infty}(\theta=\pi, x)|$, of the target in question. The square of this quantity is the normalized backscattering cross section (BSCS) $\sigma/\pi a^2$ of the target. The form function depends on the nondimensional frequency $x = k_1 a$, where the wave number $k_1 = \omega/c_1$, c_1 is the sound speed in the outer medium (#1), and a = shell's outer radius. For an air-

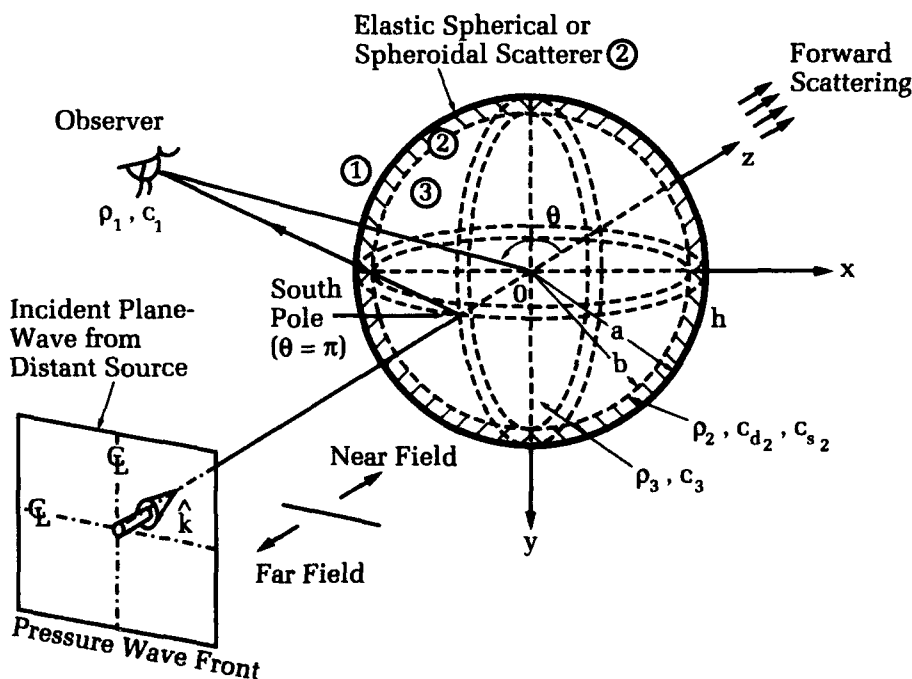


Figure 1. Shell geometry and incident (plane) wave falling on the spherical shell at its south pole.

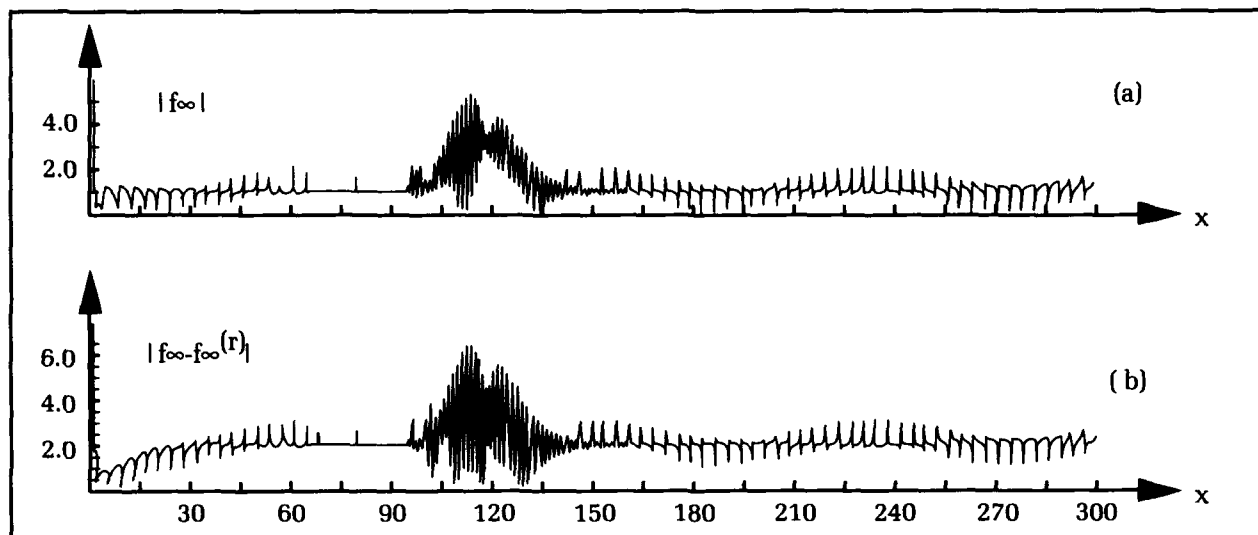


Figure 2. (a) Form function of a spherical steel shell in water of $h' = 1\%$.
(b) Residual response after subtraction of the rigid background.

filled steel spherical shell in water of inner/outer radii b/a , the plot of $|f_\infty(x)|$ is shown in Figure 2 for a thickness of $h' \equiv h/a = (a-b)/a = 1\%$, in the wide band $0 \leq x \leq 300$. The way such a plot is calculated or measured has been described in our earlier publications²⁻⁵ and will not be repeated here. Suffice it to say that this plot is exact within the context of linear elasto-dynamics, since no shell-theory approximation has been introduced to model the shell motions. To generate such plots in such wide bands requires lengthy computations because the series involved are slowly convergent and many hundreds of its terms (partial-waves) have to be added to sum it. Such wide bands are needed because some of the basic features of the response appear at large x -values. For example, there is a wide feature in the band: $95 \leq x \leq 135$ that we would have missed had we stopped the graph at, say, $x=90$. This "bump" in the BSCS is not caused by a single resonance, but by two sets of resonance families that we'll discuss below. Observation of Figure 2(a) shows large resonance features at low frequencies (i.e., at $x \sim 2$) due to shell curvature, and also in the "bump" region around $x = 120$. We said at the beginning that the resonances were superimposed on a smooth background. This is more or less evident in the summed form function of Figure 2 that has added all the constituent partial waves. This is much more evident if we look at each of the partial waves individually. It is really the partial waves that decompose naturally into smooth modal backgrounds and spiky resonance contributions. This decomposition is the basis of the Resonance Scattering Theory (RST),^{1,6-9} which postulates, and has demonstrated in many instances, that this separation is

always possible. The original idea of this decomposition was patterned after early studies of nuclear scattering cross sections,^{10,11,12} which obey an analogous principle. Considerable generalizations and extensions have taken place since their application to acoustics, including many present-day engineering applications.^{1,6,7} Resonance scattering is a subfield of Scattering Theory that studies the scattering cross sections of penetrable obstacles/targets in their resonance or intermediate spectral region. This is the frequency region hardest to analyze, because within it one can not apply a number of simplifying approximations that hold at either low or high frequencies. One must deal with either exact solutions or novel approximations in this region. In this sense, resonance scattering is the main area of concern of present-day scattering studies. Many works have shown that it is incorrect to attempt to predict the sound fields scattered by an elastic object in water by assuming it impenetrable (i.e., in this case, rigid). It is necessary to account for the object's elasticity. Rigid-body scattering is simply not a good model to analyze echoes from any underwater structure. Since a rigid body has no resonances, to assume that an elastic body is purely rigid would imply that its modal echoes are merely formed of "backgrounds" without any resonance contributions.

Separation of Impenetrable Backgrounds From Resonances

The next point that comes up is how to separate the smooth, impenetrable backgrounds ("modal" if within each partial wave, or "total"

if within the summed cross section) from the resonances superimposed on them. The majority of works pertaining to resonance scattering of underwater targets have dealt primarily with this single issue. Posed in different words, the question is: What background should one subtract from the partial waves (or from the summed form function) so that the leftover residuals clearly and accurately reproduce the target resonances for any target, and over the whole frequency spectrum? There are several answers to this question. If the target has a very high (or low) impedance relative to that of its surroundings, such as a solid metal sphere (or a gas bubble) in water, then the perfectly rigid (or soft) background does an effective job of isolating the resonances at almost any frequency. Problems arise when the target has an "intermediate" impedance, not too close to either the rigid or the soft extremes of acoustic behavior. Such a target is, for example, an air-filled, thin metal shell in water. These are, clearly, the most important targets under consideration. For such a structure, the rigid (or soft) background isolates well the resonances at the high- or the low-frequency end of the spectrum, but neither seems to work well in between. Only recently has it been possible to obtain a hybrid-type of frequency-dependent background that reduces properly to the rigid/soft cases at high/low frequencies, and that clearly isolates the resonances everywhere in between.¹³ The suppression of the rigid backgrounds from the summed form function is shown in Figure 2(b) for the same 1 percent thick steel shell in water. We now consider the partial waves and subtract the rigid modal backgrounds from each one of them (i.e., $n=1,2,3,\dots$) for the same 1 percent steel shell in water, and plot the initial resonance lobe thus

isolated, vs. x , all in the same graph, as in Figure 3. These first resonances, isolated one-at-a-time by the subtraction of the rigid backgrounds, do not look much like one would expect from standard resonances. They are quite wide, and each seems to carry a trailing tail. We can contrast this crude resonance isolation process using the rigid backgrounds with the one that results from the subtraction of the novel hybrid modal background mentioned above and displayed in Figure 4. Figure 4(a) shows the result after all the residuals are added. Figure 4(b) shows the initial resonance contained within each of the first 26 partial waves, which are associated with the S_0 - Lamb wave in the shell, all displayed in the same graph. S_0 is the

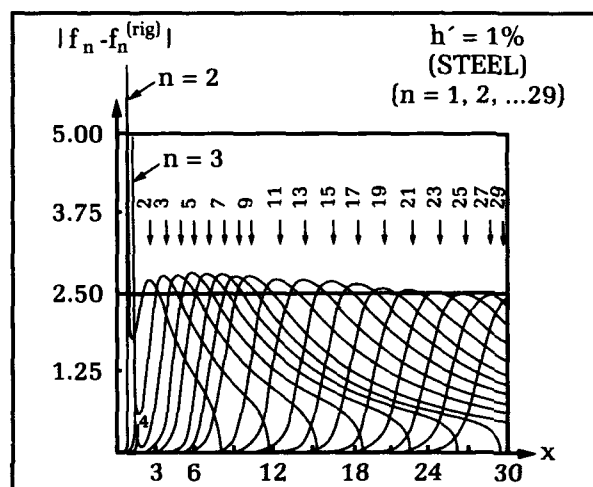


Figure 3. The leading portions of the first 29 partial-wave constituents of the "spectrogram" obtained by the subtraction of the rigid modal backgrounds, all for the same $h'=1\%$ steel spherical shell in the water. Note the trailing tails.

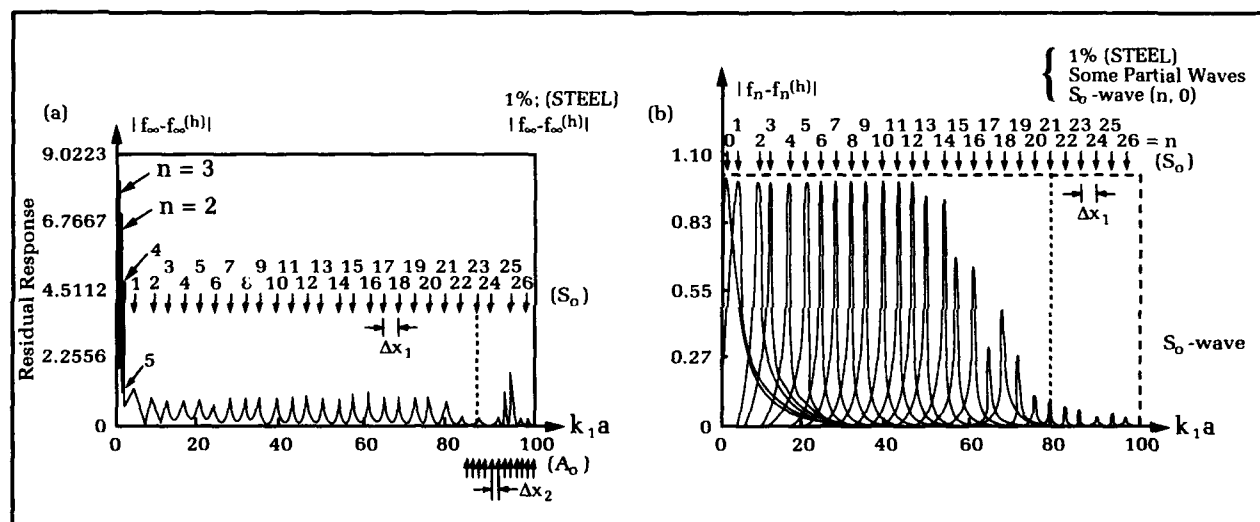


Figure 4. (a) Total residual response obtained by the novel hybrid background. (b) The leading portions of the first few modal residual responses, as obtained by the subtraction of the novel hybrid backgrounds.

generalized zeroeth-order, symmetric Lamb wave in the shell. Now the resonances are quite "clean" and lack those earlier tails, showing the appropriateness of the novel hybrid background over the whole spectrum.

Elastic Waves Causing the Resonances

We will use the elastic shell example to illustrate wave phenomena causing the resonances. Various families of resonance spikes are seen in Figure 4. Some, spaced an amount Δx_1 , exist up to $x = 85$, and others spaced an amount Δx_2 (due to the A_0 - Lamb wave), start to appear above $x = 85$. Clearly, each resonance spike thus isolated belongs to a "family." Each family is associated with, or caused by, a wave that travels either in the shell or in its external fluid. Shell waves can travel along its surfaces as surface waves, or through its interior as bulk waves. When the wavelength is close to the shell thickness, it is hard to differentiate the two types. Usually a certain bulk or surface wave in/on the shell or its surroundings causes a family of resonances, which then manifest themselves in the form function as a series of peaks. It is this ensemble of resonance spikes in the BSCS that we have learned to decipher and disentangle. It is desirable to know which of these interleaved sets of resonance features are caused by which of these bulk, surface, or external waves, what are their phase and group velocities, and most important, what information about the physical characteristics of the target (viz., shape, composition) is contained within this "code" of active spectral lines of given widths (i.e., the "spectrogram" as illustrated in a later section).

Were this totally known, we could identify the target completely, once we had extracted the spectrogram from its echoes. We know that bulk waves generate what are often called geometrical resonances, which tend to appear at high frequencies. Surface waves on an elastic target such as Rayleigh, or Whispering Gallery waves if the target is solid, or Lamb waves if it is hollow (i.e., a shell), carry most of their energy inside the target and cause what are often denoted internal resonances.^{1,6-8} Waves propagating in the fluid surrounding the target, such as the creeping waves or the Scholte and Stoneley waves, have most of their energy in that outer fluid, and lead to "resonances" that are not proper resonances of the target itself, but of the outer fluid, and are sometimes called external resonances. Resonances of this type are the only ones present for impenetrable targets, since these do not admit interior fields. We remark at this point that Rayleigh, Lamb, Scholte and Stoneley waves were all originally investigated by the authors, after whom they are

named, for flat half-spaces or infinite plates in *vacuo*. These waves have all been generalized to fluid loaded curved structures such as solid bodies or shells, but they retain their names for the flat-interface cases. The creeping waves, originally studied by Franz, owe their existence to the curvature of the target they circumnavigate. There are also some internal resonances, visible in Figures 2 through 4 at low x -values, that are due to target curvatures. The resonances associated with these waves are found as the real parts of the roots of certain characteristic equations.^{3,4} These equations correspond to the vanishing of certain determinants (of order 6×6 for a shell) with complicated frequency-dependent elements. In some cases, the roots associated with a certain type of wave remain real only within certain frequency intervals. These intervals then determine the bands over which that wave can physically propagate or exist. Usually, the speed of propagation of the external waves is close to that of sound in the surrounding fluid. There seems to be no similar restriction for the speeds of the various types of internal waves. For spherical and cylindrical targets, two integer indices, n, ℓ , label the internal resonances caused by the surface waves. For sphere problems, the first index n is such that $2\pi a = (n + 1/2)\lambda$; for cylindrical cases, it would be $2\pi a = n\lambda$. This means that the surface wave generating the given resonance occurs when the circumference of the target exactly equals a half-integer or an integer number of wavelengths λ . This condition forces the formation of a standing (surface) wave on the perimeter of the body. This standing wave then radiates, producing a scattering pattern of energy that looks like a flower with many petals. For cylinders at normal incidence, the angular pattern is a rhodonea, $|\cos n\theta|$, and for spheres, it is a Legendre polynomial $|P_n(\cos \theta)|$ of $2n$ -petals. For a body of any other shape, the term "circumference of the target," above, has to be replaced by a more complicated geodesic path, and the condition is then equivalent to a phase-matching requirement along that geodesic. The condition can then be generalized by means of the obvious integral,

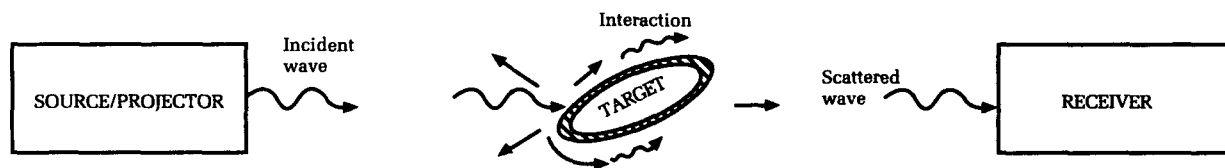
$$\int_C \frac{ds}{\lambda_\ell} = n + 1/2, \quad (1)$$

where C is the geodesic, and λ_ℓ the wavelength of the ℓ th surface wave. The above integral is for sphere-like bodies lacking azimuthal dependence. This condition, either in discrete (integer) form or in continuous (integral) form, has numerous counterparts in various areas of mechanics dealing with standing waves that go back to the 19th century.

The Direct Scattering Problem

The Direct Scattering Problem (DSP) has been treated in various cases by means of the background-suppressing method known as the RST.¹⁻⁹ The first public presentation of RST principles took place in 1977 in a Mechanics Conference that included two lectures on the subject.^{14,15} These talks were submitted for journal publication in 1977 and appeared^{16,17} the following year within a couple of months of each other. The basic idea of RST methodology stems from the realization that the backscattering cross section of a penetrable object can be decomposed into "backgrounds" and "reso-

nances," as discussed above. This decomposition takes into account the fluid-loaded condition of the target (or cavity), and the coupling that this introduces. Although the prediction of scattering cross sections of separable (i.e., spherical/cylindrical) elastic targets in water and of separable inclusions within elastic media was treated by conventional or classical means in the early fifties by Faran¹⁸ and Truell,¹⁹ respectively, the realization that the resulting cross sections could be decomposed into backgrounds and resonances did not emerge until 1977, when these splittings were formally shown to be possible¹⁴⁻¹⁷ from formulations cast in "S-Matrix" form. This formula-



Elastic Resonance-Scattering Terms

Rayleigh Waves: Surface waves originally (1885) studied by Rayleigh that propagate on the surface of flat elastic half-spaces in contact with vacuum. Here, they have been generalized to occur on the surface of curved, solid elastic bodies in contact with fluids.

Whispering Gallery (WG) Waves: For flat half-spaces in contact with vacuum, only the Rayleigh wave occurs at the interface, but for curved, solid elastic bodies in contact with fluids, not only the Rayleigh wave but many WG waves are also present.

Stoneley/Scholte Waves: Analogous to the original Rayleigh wave on the surface of flat elastic half-spaces, but now in contact with a fluid (for *all* values of the elastic parameters; Stoneley) or with another solid (for *some* values of the elastic parameters; Scholte). Here, these are generalized to occur on curved bodies in contact with fluids or solids.

Dilatation (or Compressional, or Longitudinal) Wave: Elastic body wave for which the oscillations of the elastic particles in the medium are in the direction of wave propagation.

Shear (or Transverse) Waves: The oscillations of the elastic particles occur in a plane normal to the direction of wave propagation. If they are in the vertical (or horizontal) directions within that plane, the shear wave is said to be vertically (or horizontally) polarized.

Lamb Waves: As originally (1916) studied by Lamb, these are elliptically polarized surface waves appearing on the two surfaces of an (infinite) elastic plate in contact with vacuum on both sides. They can be symmetric (S_n) or antisymmetric (A_n). Here they have been generalized to include waves on the surface of curved elastic shells, fluid loaded on both sides.

Resonance-labeling Indices ℓ , ℓ' : For solid elastic targets, within each partial wave or normal mode making up the target's backscattering cross section, there are ℓ (or ℓ') dilatational (or shear) resonance features that are labeled by these two indices.

Rigid/Soft Backgrounds: The individual (or the sum of the) partial waves that make up the backscattering cross section of a perfectly impenetrable (i.e., rigid or soft) target of identical shape to the *elastic* target under study. In the rigid (or soft) case: $(\rho c)_T \gg (\rho c)_{\text{fluid}}$ [or $(\rho c)_T \ll (\rho c)_{\text{fluid}}$].

tion was originally developed in the frequency domain (x-plane) of the BSCS, without any need to appeal to the Watson-Sommerfeld Method or to contour integrations in the mode-order, v , plane. It was later found that these two representations are connected.² Considerable development and refinement have taken place since these early days.⁶⁻⁸

All the figures we have presented so far treat the solution of the DSP for various spherical configurations, either before or after the background suppression process that results in the isolation of the body's resonances (i.e., the spectrogram) from the returned echo. An important point about this resonance isolation process is that it can also be achieved by experimental means, quickly and *in-situ*, by processing the target echoes in their "informative" region. This processing automatically suppresses the undesirable backgrounds, and has been explained many times.^{20,21} If one uses long CW pulses or CW insonifications, it consists of sampling the backscattered return after the second transient, just as the "ringing" begins. This second transient region appears after the steady-state or forced vibration region ends, just as the trailing edge of the long CW pulse ceases to interact with the body. The outcome is the capture of the spectrogram in as broad a band as desired.

Generalizing From CW To Pulsed Incidences

Transient studies using either long or short pulses are easier to carry out in practice than to model analytically. Such studies are essential, since sonar interrogation is basically a pulsed operation. It has long been known that if the

analytic expression for the steady-state form function of a target has already been obtained for CW insonification, and if the same target is insonified by a pulse, $g(\tau)$, of spectrum $G(x)$, then the backscattered pressure pulse from the target will be given²²⁻²⁵ by

$$rp_{sc}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} [f_{\infty}(x)G(x)]e^{ix\tau} dx = f_{\infty} * G, \quad (2)$$

where x and τ are a pair of nondimensional canonical variables defined by $x = ka$ and $\tau = (ct-r)/a$. This convolution result serves to predict the backscattered pressure pulse returned by any target. As an example, we consider a CW pulse (i.e., a sinusoidal wave train of finite duration) incident on a steel sphere in water. Figure 5(a) shows the form function of the sphere in the band: $0 \leq x \leq 30$. After subtraction of the suitable background, rigid in this case, we have the spectrogram shown in Figure 5(b). A resonance is seen to occur at $x \approx 8.1$, and an antiresonance at $x \approx 9.0$. A CW pulse 20 cycles in duration and carrier frequency $x_0 \approx 8.1$ is incident on the sphere. Such a pulse $g_3(x)$ has duration $\tau^* = 2\pi N/x_0 = 15.5$, as shown in Figure 6(a). The modulus of its spectrum $|G_3(x)|$ is shown in Figure 6(c). The backscattered pressure pulse as given by the above formula is shown in Figure 6(b), and its spectrum $|p_{sc}(x)|$ is displayed in Figure 6(d). Since the carrier frequency, x_0 , of the incident pulse $g_3(\tau)$ coincides by design with a target resonance, the envelope of the echo in Figure 6(b) is seen to decrease in amplitude in discrete steps.²³⁻²⁶ Such behavior is absent if x_0 does not coincide with a body resonance. Significant differences in the response appear if the incident pulse is short. If the pulse is short

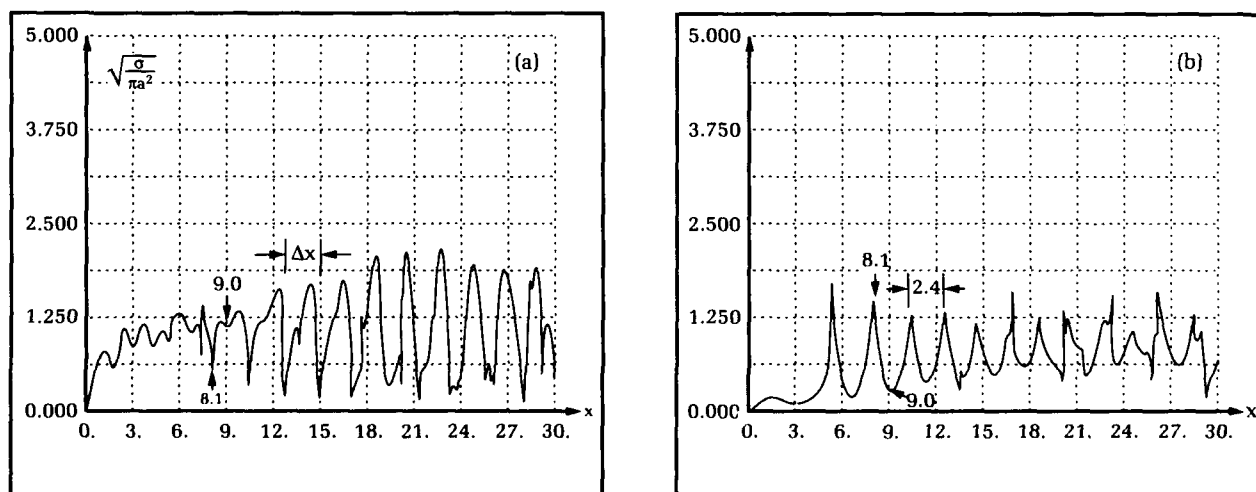


Figure 5. (a) Form function of a stainless steel sphere in water in $0 \leq x \leq 30$. (b) Residual response after (rigid) background subtraction. Resonances and antiresonances are marked by arrows at $x=8.1$ and 9.0 , respectively.

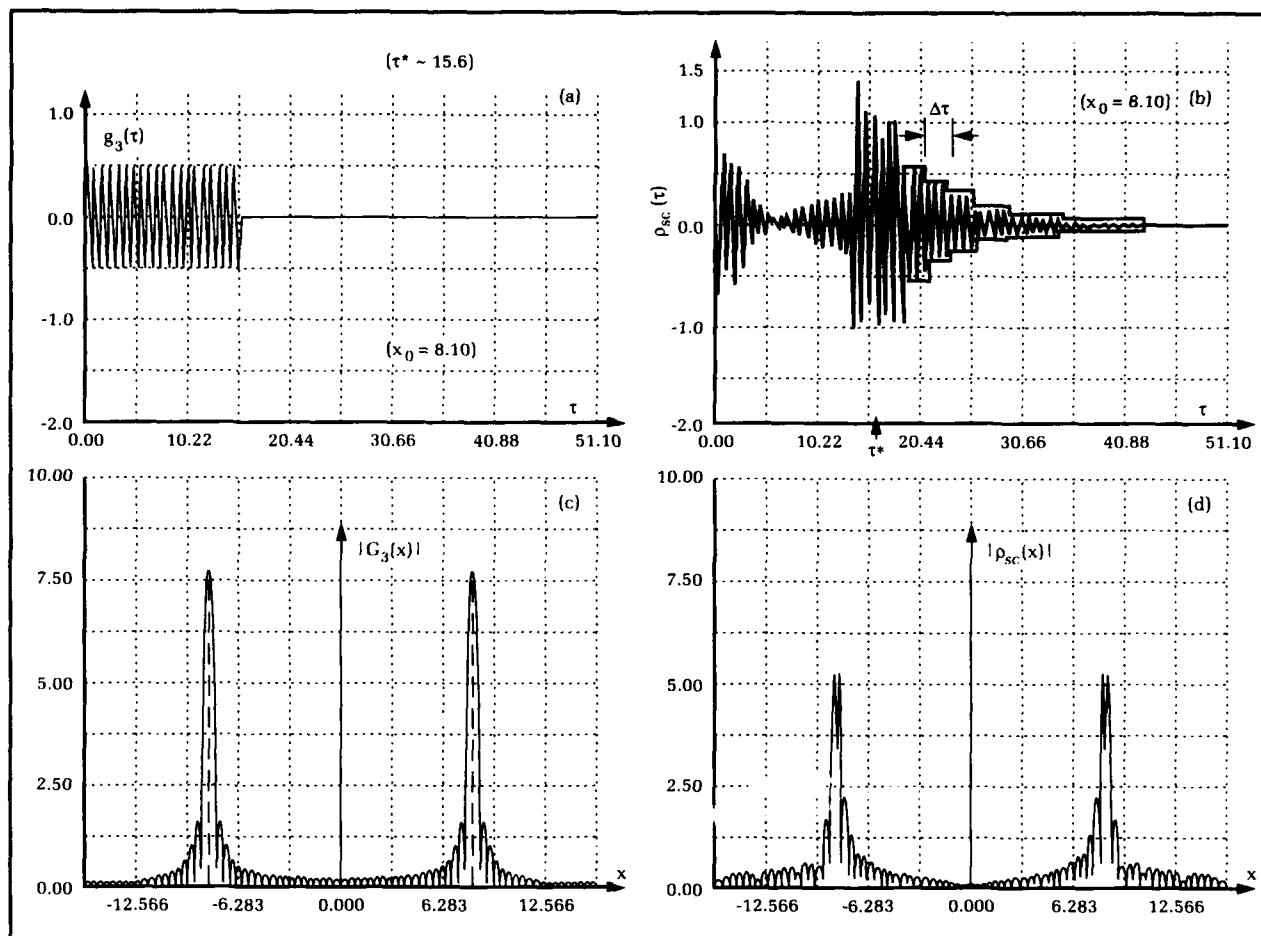


Figure 6. (a) Incident CW pulse of carrier frequency $x_0 = 8.1$ and for $N=20$ cycles. (b) Backscattered pressure pulse returned from the solid sphere. (c) Modulus of the spectrum of the incident pulse shown in (a). (d) The spectrum of the pulse in (b).

enough, the response will be a replica of the steady-state form function of the target, at least up to the carrier frequency of the incident pulse, and this is then an effective way to obtain it. Further details have been given elsewhere.²⁶ The solution of the DSP is very useful for later going backward to attempt the solution of the inverse scattering problem. At present the DSP can be solved for any type of waveform, steady-state or transient, incident on any penetrable acoustic target. Analytic solutions will be possible for separable targets. For non-separable geometries or shapes, there are a variety of very effective numerical techniques.

The Inverse-Scattering Problem

As initially stated, the Inverse-Scattering Problem (ISP) consists of the extraction of physical information about a target from its form function, which is the quantity sensed by a sonar,

and hence, the only tool one has to actively detect and classify. The spectrogram resulting after background suppression is the first item to be examined, since usually the form function is obscured by the presence of the undesirable backgrounds. Consider the spectrogram of a steel sphere in water, as displayed in Figure 5(b). The initial peaks are separated by an asymptotically uniform distance Δx , which here has value 2.4. These are the peaks caused by the Rayleigh wave circumnavigating the sphere. This separation equals the ratio of the group velocities of the induced surface waves to the sound speed in the outer fluid, viz.,

$$c_l^{(g)} / c_1 = \Delta x. \quad (3)$$

For water, $c_1 = 1.5 \times 10^5$ cm/s and since $\Delta x = 2.4$ we find $c_l^{(g)} = 3.6 \times 10^5$ cm/s, in agreement with earlier findings.²³⁻²⁵ In view of the sonar uncertainty relation, $\Delta x \Delta \tau = 2\pi$, we can estimate $\Delta \tau$ to

be: $\Delta\tau = 2\pi/2.4 \approx 2.6$. This is the value observed in Figure 6(b) for the horizontal duration of the discontinuous steps that appear in the backscattered pulse in the case of long-incident CW pulses. Examining the partial waves ($n=0, 1, 2, \dots$) contained within such spectrogram in Figure 5(b), we observe that all the partial waves contain two interleaved families of resonances.²⁷ One family has broad widths and is labeled by an index ℓ , while the other has narrow widths and is labeled by another index, ℓ' . The $n=0$ partial wave is the only one that contains only the broad set labeled by ℓ . Thus, the second index, ℓ , commonly used to label resonances, viz., (n, ℓ) , splits naturally into two subindices (here, ℓ and ℓ') associated with the types of waves (here, dilatational and shear) supported by the target, for any non-zero value of the azimuthal index n . This is in accordance with our previous findings.²⁷

Determination of Solid-Sphere Composition

Consecutive dilatational (or shear) resonances labeled by the index ℓ (or ℓ') can be shown to be asymptotically spaced a uniform value Δ (or Δ'), which is given by the expressions:²⁷

$$\begin{aligned} \Delta &\xrightarrow{\ell \gg 1} (c_{d2}/c_1)\pi \\ \text{or} \\ \Delta' &\xrightarrow{\ell' \gg 1} (c_{s2}/c_1)\pi. \end{aligned} \quad (4)$$

It follows that measuring these spacings Δ, Δ' , we can determine the c_{d2} and c_{s2} of the target. An asymptotic study of the characteristic equation of this problem²⁷ also showed that this equation uncouples and splits into two equations as $x \gg 1$. The roots of each one of these (viz., $1/X_n^{(D)} = 0$ and $X_n^{(S)} = n(n+1)/2$) give the dilatational (or shear) resonances x_ℓ (or $x_{\ell'}$), separately. For each value of n , these two types of resonances are given by

$$\begin{aligned} x_\ell &= \frac{c_{d2}}{c_1} \left(\ell - \frac{n}{2} \right) \pi + R; \\ x_{\ell'} &= \frac{c_{s2}}{c_1} \left(\ell' + \frac{n}{2} \right) \pi + R', \end{aligned} \quad (5)$$

which are families of straight lines when plotted versus ℓ or ℓ' , and where R, R' are small residuals that eventually vanish for $\ell, \ell' \gg 1$ (or $x \gg 1$). Hence, knowledge of n and ℓ (or of n and ℓ') asymptotically determines the location of any dilatational (i.e., broad) or shear (i.e., narrow) resonance. The only material parameter of the

sphere left to be determined is the density ρ_2 . It is found from the width of any dilatational resonance of any mode n (namely, $\Gamma_{n\ell}$), and the two spacings Δ, Δ' already determined, viz.,

$$\frac{\rho_1}{\rho_2} \xrightarrow{x \gg 1} \frac{\Gamma_{n\ell}}{2} - \frac{(2\Delta'/\pi)^2}{(\ell - n/2)\Delta}. \quad (6)$$

For very high x (or ℓ) values, the second term on the right side is negligible compared to the first, and then the density ratio of fluid to target is equal to the half-width of any broad/dilatational resonance, measured at half-maximum. At lower frequencies, both terms have to be taken into account since they are then comparable in size. Note that Equation (6) does not depend on the width $\Gamma_{n\ell'}$ or any shear or narrow resonance, since this is a second-order quantity. This procedure, derived for elastic targets in fluids, has been generalized to elastic targets in solid media.²⁸ The next few paragraphs will consider the most important case of submerged shells. Most algorithms for "inversion" are based on the presumed known solution for the direct problem, and in many cases, the inversions are not unique. There are, however, certain cases in which they are unique and unambiguous. Some of these results are examined below.

Complete Characterization of Shell From Its Resonances

Consider the form function of the steel shell in water of $h' = 1\%$ shown in Figure 2. For these purposes, one can consider the upper graph for the form function, since there the resonances seem to be so clearly riding on the underlying background. In the displayed band we note two quite large resonance features. One is a narrow spike at low frequencies (i.e., at $x \sim 2$), and the other is a broad "bump" centered around $x \sim 120$. Neither of these is a single resonance. Closer examination reveals that they are formed by the combined effect of families of such resonances.^{2,3,29} The low-frequency feature at $x \sim 2$ is due to the double curvature of the shell. Plates and cylinders do not exhibit such a feature. Hence, this peak tells us that the shell has double curvature like a sphere or spheroid. Were the frequency of the incident wave known, say, $f = 1$ kHz, then $x \approx 2$ fixes the shell size, viz., $a = 0.5$ m if it is spherical, and $D = 2a = 1$ m, if spheroidal.

The "coincidence" frequency x_c is defined as that frequency for which: $c_p^{ph}(x) \approx c_1$, where c_p^{ph} is the phase velocity of the corresponding (ℓ th) surface Lamb wave. The rule of thumb that has emerged from the treatment of many materials and geometries^{2,3,13} is that the above "coincidence" condition is satisfied whenever $x_c h' \approx 1$, where $h' = h/a = (a - b)/a =$ relative shell thick-

ness. To be more precise, the rule is: $Cx_c h' \approx 1$, where C is a material-dependent constant that for steel has value 0.85. The "bump" at $x \approx 120$ is always due to this coincidence effect, and we can use the relation $h' \approx 1/Cx_c$ to determine h' , which comes out to be: $h' = 1\%$, as originally assumed in order to generate Figure 2. Having the radius a and the relative shell thickness h' , we can determine the shell thickness, $h = 0.5$ cm. Finally, we can determine the composition of the shell material and its filler fluid, if any.

The shell composition emerges from a relation analogous to Equation (3), but now for shells at high frequencies, the $c_f^{(gr)}$ is to be replaced by the "plate" velocity c^{pt} , viz.,

$$\Delta x_1 = c^{pt} / c_1 = \frac{1}{c_1} \cdot \sqrt{\frac{E}{\rho(1 - \nu^2)}}. \quad (7)$$

The plate velocity is a typical property of each material. Reading $\Delta x_1 \approx 3.5$ in Figure 4(a), and since $c_1 = 1.5 \times 10^5$ cm/s, we find $c^{pt} \approx 5.25 \times 10^5$ cm/s, which can be used as a classifier for the steel in question (i.e., of $c_{d2} = 5.88 \times 10^5$ cm/s, $c_{s2} = 3.2 \times 10^5$ cm/s, and $\rho_2 = 7.8$ g/cm³). Thus, the shell is totally characterized. The composition of the filler fluid can be determined from the location of a large amplitude, quasi-thickness-resonance¹³ that appears at very high frequencies and is excited whenever half-a-wavelength of the compressional waves exactly fits the shell thickness h (namely, for $h = \lambda_d/2$). This condition is equivalent to $k_d h = m\pi$, where $m = 1, 2, 3, \dots$. Even for $m=1$, this resonance occurs at a high frequency given by $x = \pi c_{d2}/c_1 h'$. This is the cut-off or critical frequency of the S_1 -wave in the shell. For $h' = 1\%$, $c_{d2} = 5.88 \times 10^5$ cm/s and $c_1 = 1.5 \times 10^5$ cm/s, we find $x \approx 1,230$. If we determine the reflection coefficient V of the shell at high frequencies, such as $x \approx 1,230$ (far outside the band shown in Figure 2), then the impedance of the filler fluid z_3 is related to that of the outer fluid (z_1 , water, assumed known) by:

$$z_3 = z_1 \left(\frac{1 - V}{1 + V} \right), \quad (8)$$

where V is easily determined at high frequencies, because then the initial, specularly reflected portion of the return is reduced in amplitude relative to that of the incident pressure wave by an amount exactly equal to V . The z_3 of the filler can be used as a classifier when one uses a table of impedances for various materials. The filler composition can also be determined by other low-frequency approaches to be discussed elsewhere. Other techniques for discriminating between target shapes have been discussed in earlier reviews.^{1,6} The parameterization described above can be accomplished quickly, *in*

situ, and automatically, not only for sonar sensors but also for several other types of non-acoustic sensors to be found in various chapters of recent proceedings.³⁰ Among these is an application of particular importance to impulse radar.³¹

Conclusions

We have discussed the DSP and the ISP in their evolutionary context and described a methodology that exploits the presence of certain resonance features in the echoes returned by targets. The DSP predicts the spectral locations at which the resonances of a given target will make themselves noticeable in the returned echoes. The next step is the separation of the resonances from the obscuring backgrounds (unrelated to noises or noise levels) mixed with them. The extracted resonances from an echo constitute the active spectrogram of the scatterer. Spectrograms are currently obtained analytically for separable shapes, computationally for any shape if not too weird, and experimentally (for separable or non-separable shapes, weird or not). The experimental extraction of spectrograms has been previously discussed by us^{1,6,20} and it is currently achievable not only in steady-state cases for CW incidences, but also in transient situations for pulsed incidences. Finally, we considered two examples, a metal sphere and an air-filled shell, in which all the physical characteristics of these stationary targets were accurately identified from the resonance features in their SCSs. Therefore, we have illustrated the active classification capability of our methodology by completely solving these ISPs. No other way has yet been disclosed for achieving this objective.

Acknowledgments

The author appreciates the support received over the years from the Independent Research Board of the NSWC Dahlgren Division and the Office of Naval Research.

References

1. Gaunaurd, G., "Elastic and Acoustic Resonance Wave-scattering," *Applied Mech. Reviews*, Vol. 42, 1989, pp. 143-192.
2. Gaunaurd, G. and Werby, M., "Lamb and Creeping Waves Around Submerged Spherical Shells Resonantly Excited by Sound Scattering," *J. Acoust. Society of Am.*, Vol. 82, 1987, pp. 2021-2033.
3. Gaunaurd, G. and Werby, M., "Lamb and Creeping Waves Around Submerged Spherical Shells Resonantly Excited by Sound Scattering, II. Further Applications," *J. Acoust. Society of Am.*, Vol. 89, 1991, pp. 1656-1667.

4. Gaunaurd, G. and Kalnins, A., "Resonances in the Sonar Cross-sections of (Coated) Spherical Shells," *Intern. J. Solids and Structures*, Vol. 18, 1982, pp. 1083-1102.
5. Gaunaurd, G. and Brill, D., "Acoustic Spectrogram and the Complex-frequency Poles of a Resonantly Excited Elastic Tube," *J. Acoust. Soc. Am.*, Vol. 75, 1984, pp. 1680-1693.
6. Gaunaurd, G. and Werby, M., "Acoustic Resonance Scattering by Submerged Elastic Shells," *Appl. Mech. Reviews*, Vol. 43, 1990, pp. 171-207.
7. Gaunaurd, G., "Resonance Acoustic Scattering from Underwater Elastic Bodies," *Proc. 2nd IUTAM Symp. on Elastic Wave Propagation*, North Holland Publishers, 1988, pp. 335-347.
8. Brill, D. and Gaunaurd, G., "Resonance Theory of Elastic Waves Ultrasonically Scattered from an Elastic Sphere," *J. Acoust. Society of Am.*, Vol. 81, 1987, pp. 1-21.
9. Flax, L., Gaunaurd, G. and Uberall, H., "Theory of Resonance Scattering," *Physical Acoustics*, Vol. 15, Academic Press, 1981, pp. 191-294.
10. Breit, G. and Wigner, E. P., "Capture of Slow Neutrons," *Physical Review*, Vol. 49, 1936, pp. 519-531.
11. Wigner, E. P., "Resonance Reactions and Anomalous Scattering," *Physical Review*, Vol. 70, 1946, pp. 15-33.
12. Feshbach, H., et al., "On the Scattering and Absorption of Particles by Atomic Nuclei," *Physical Review*, Vol. 71, 1947, pp. 145-158.
13. Gaunaurd, G. and Werby, M., "Sound Scattering by Resonantly Excited, Fluid-loaded, Elastic Spherical Shells," *J. Acoust. Society of Am.*, Vol. 90, 1991, pp. 2536-2550.
14. Gaunaurd, G., "Resonance Theory of Elastic and Viscoelastic Wave-scattering and Its Application to the Spherical Cavity in Absorptive Media," *Proc. 1st IUTAM Symp. on Modern Problems in Elastic Wave Propagation*, J. Wiley & Sons, 1978, p. 550.
15. Uberall, H., "Modal and Surface Wave Resonances in Acoustic Wave-scattering from Cavities in Solids," *Proc. 1st IUTAM Symp. on Modern Problems in Elastic Wave Propagation*, J. Wiley & Sons, 1978, p. 239.
16. Gaunaurd, G. and Uberall, H., "Theory of Resonant Scattering from Spherical Cavities in Elastic and Viscoelastic Media," *J. Acoust. Society of Am.*, Vol. 63, June 1978, pp. 1699-1712.
17. Flax, L., Dragonette, L. and Uberall, H., "Theory of Elastic Resonance Excitation by Sound Scattering," *J. Acoust. Society of Am.*, Vol. 63, 1978, pp. 723-731.
18. Faran, J. J., Jr., "Sound Scattering by Solid Cylinders and Spheres," *J. Acoust. Society of Am.*, Vol. 23, 1951, pp. 405-418.
19. Ying, C. and Truell, R., "Scattering of a Plane Longitudinal Wave by a Spherical Obstacle in an Isotropically Elastic Solid," *J. Appl. Phys.*, Vol. 27, 1956, pp. 1086-1097.
20. Gaunaurd, G. and Tsui, C., "Transient and Steady-state Target Resonance Excitation by Sound Scattering," *Appl. Acoustics*, Vol. 23, 1988, pp. 121-140.
21. Ed. by Group d'Etudes Sur la Propagation Acoustique - GESPA, *La Diffusion Acoustique*, Cedocar Publishers, Paris, 1987.
22. Kennaugh, E. and Moffatt, D., "Transient and Impulse Response Approximations," *Proc. IEEE*, Vol. 53, 1965, pp. 893-901.

23. Brill, D., et al., "Backscattering of Sound Pulses by Elastic Bodies Under Water," *Appl. Acoustics*, Vol. 33, 1991, pp. 87-107.
24. Veksler, N., *Information Analysis in Hydroelasticity*, Acad. Sciences of the Estonian S. S. R., Valgus Publishers, Tallinn, 1982.
25. Howell, W., et al., "Complex Frequency Poles of the Acoustic Scattering Amplitude and Their Ringing," *IEEE Trans. UFFC*, Vol. 34, 1987, pp. 22-27.
26. Gaunard, G. and Wertman, W., "Transient Acoustic Scattering by Fluid-loaded Elastic Shells," *Intern. J. Solids and Structures*, Vol. 27, No. 6, 1991, pp. 699-711.
27. Ayres, V. and Gaunard, G., "Inverse Scattering from an Elastic Sphere," *J. Acoust. Society of Am.*, Vol. 82, 1987, pp. 1291-1302.
28. Gaunard, G., "Inverse Scattering Techniques for Material Characterization," *Proc Ultrasonics International '87*, Butterworth Scientific Ltd., 1987, pp. 520-525.
29. Strifors, H. and Gaunard, G., "Multipole Character of the Large-amplitude, Low-frequency Resonances in the Sonar Echoes of Submerged Spherical Shells," *Intern. J. of Solids and Structures*, Vol. 29, 1992, pp. 121-130.
30. Sahadi, F., Ed., *Proc. SPIE Automatic Target Recognition Conference*, Orlando, FL, 1991.
31. Abrahamsson, S., et al., "Target Identification by Means of Impulse Radar," *Proc. SPIE Automatic Target Recognition Conference*, Orlando, FL, 1991, pp. 130-141.

The Author



GUILLERMO C. GAUNAURD received his Ph.D. in physics/acoustics from Catholic University, Washington, D.C. in 1971. He has been a research physicist in the Research Department of the Naval Surface Warfare Center (now Dahlgren Division) since 1971, specializing in acoustics and electromagnetic radiation and scattering. Currently, he is a senior scientist and group leader in the Physics and Technology Division where he carries on a research program on the basic interaction of waves with materials. Since 1984, he has been a lecturer at the University of Maryland. He has authored over one hundred journal articles and is a member of many professional and honor societies. He is a Senior Member of the IEEE and a Fellow of the Acoustical Society of America and the Washington Academy of Sciences. His professional biography appears in Marquis' *Who's Who in America* and a dozen similar books.

Hydrogen Spark Gap For High-Repetition-Rate Compact Accelerators

Stuart L. Moran

Time-of-flight of a weapon system can limit the ability of a ship to counter fast, low, or highly maneuvering targets. The Navy is developing directed-energy weapons with near zero time-of-flight to reduce detection and sensor requirements. Recent development efforts with charged-particle-beam (CPB) weapons have concentrated on compact schemes using recirculating accelerators. Because these accelerators must handle high powers and operate at high repetition rates to achieve long ranges, high-power, high-repetition-rate switches have become a key element in CPB weapon development. Detection and identification systems may also be expected to benefit from this switch technology. The Pulsed Power Technology Branch has concentrated in-house efforts on spark-gap switches because of their high-power capabilities, simplicity, and low cost. We have recently demonstrated the use of hydrogen gas and special triggering techniques to increase the repetition rate of spark gaps. This new switch provides two orders-of-magnitude improvement over previous capability and has become the leading candidate switch technology for compact accelerators.

Introduction

New classes of sophisticated missiles have created a major threat to modern Navy ships. Highly maneuvering missiles that travel fast and low are very difficult to counter. The Navy is interested in developing technologies that allow detection of such missiles in time to respond to the threat. The amount of advance warning required by a platform to counter a missile threat is partially determined by the time-of-flight of the defending weapon system. Not only can this be a significant portion of the engagement time, but it can also limit the ability of a ship to counter a highly maneuvering target, since prediction of the target trajectory is necessary.

Directed-energy weapons, including CPB, high-energy lasers (HEL), and high-power microwaves (HPM), all have essentially zero time-of-flight—less than a millisecond. This is the major reason these weapons are considered highly desirable for ship defense. Additional reasons include a very low cost-per-shot, an almost infinite supply of ammunition, and no storage of explosives. Directed-energy weapons may be the *only* weapons capable of countering future saturated raids of highly maneuvering missiles, particularly if the raids are repeated.

The technologies developed for directed-energy weapons also have applications as sensor systems. A high-power laser could be used in a scanning mode

as a high-resolution or low-probability-of-intercept laser radar, or to detect and upset enemy optical sensors. Another application could involve impulse radars utilizing short, high-power microwave pulses to image targets for identification. However, the technology's greatest impact on detection systems may not be on the form of new techniques, but rather on the reduction of requirements for conventional sensor methods through development of speed-of-light weapons.

The Navy has had a program to develop CPB weapons for many years. CPB weapons are the most capable of the directed-energy weapons, but also the most difficult to build. CPB accelerators generate high-current electron beams, then accelerate these beams to relativistic speeds in a strong electric field. Since megajoules of energy are generally needed to damage objects physically, many kiloamps of beam current and many megavolts of acceleration are typically required. Because the beams are relativistic, they spend very little time (typically microseconds) in the accelerator, so timing of components is critical. The electrical power systems needed to drive the accelerators must operate at very high voltages, currents, and peak powers.

The beam must be propagated through the air to collide with the target. The kinetic energy of the electrons then generates enough instantaneous heat to create permanent damage to the target. The highly penetrating nature of these electrons makes it impractical to shield against this type of weapon. Propagation is difficult because of the numerous instabilities of electron beams and the interaction of the beam and the atmosphere. Electrons colliding with air molecules cause heating and rarification of the air channel. This rarification can be used to stabilize the beam and increase range, but requires that the accelerator operate in a repetitive or burst mode. This, in turn, requires that the pulsed-power systems driving the accelerator also operate in a repetitive mode. Ideally, each burst of pulses would destroy one target.

The Pulsed Power Technology Branch has been investigating the technologies required to produce the electrical power and power-conditioning systems needed to drive the electron accelerators. The effort began in 1978. Although the program has been broad-based, in-house research efforts have been concentrated on two major areas: energy storage in water-glycol mixtures,^{1,2} and high-repetition-rate spark-gap switches.³ In recent years we have achieved significant breakthroughs in both these areas.

From the beginning of the program, it was obvious that the requirements for accelerator switches would be very stringent. Theoretical studies indicated that propagation and target

interaction would require particle beams of at least hundreds of megavolts (MV) and tens of kiloamps (kA) at repetition rates of greater than a kilohertz (kHz). Since it was necessary to produce megajoules (MJ) of energy on target, 10s to 100s of megajoules would have to be stored and switched in a package suitable for a shipboard environment. When the program started, a switch that could meet the requirements for a CPB weapon did not exist. For example, high-power, spark-gap switches were typically run at less than 10 Hz. Many other types of switches could not handle the required voltages or currents. It was not clear that any switch technology would be able to meet all the requirements in a suitably sized package. Therefore, the switching research and development was broad-based, as we attempted to measure and improve the limits of known switches while, at the same time, investigating new switch concepts.⁴

This article describes the in-house spark-gap-switch research effort at NSWC Dahlgren Division. The article covers the basic design and operation of the spark gap, discusses research efforts directed toward understanding and improving switch operation, and gives specific switch requirements along with a description of ongoing experiments to demonstrate operation at full parameters. This effort has produced two orders-of-magnitude improvement in the repetition-rate capability of high-power, spark-gap switches. This technology has now become the leading switch candidate for use in a repetitive CPB accelerator.

The Basic Spark-Gap Switch

A simple spark-gap switch consists of two metal electrodes with a gas (typically pressurized air) between them. When the voltage across the gap exceeds a critical breakdown value, a narrow channel of air becomes ionized and very hot, forming a conducting channel that "closes" the switch. Temperatures reach 10,000°K inside the spark channel, allowing the gas to become a highly ionized plasma. The electrode surface becomes hot enough to maintain thermionic emission of electrons into the gas. After the switch has discharged the stored energy, the heated gas channel cools, becomes deionized and nonconducting, and the switch "opens." The process can then be repeated. In addition to simplicity, spark gaps are noted for the ability to switch extremely high voltages and peak currents. For example, the single-shot RIMFIRE spark switches at Sandia National Laboratories⁵ can switch six MV and half a mega amp (MA), providing 3,000 gigawatts (GW) of peak power.

The spark gap continues to be the most widely used switch in pulsed-power systems. Virtually

all high-power machines (nuclear simulators, fusion machines) in this country use some form of spark-gap switch. There are several reasons for this. Spark gaps have the ability to handle high voltage and high current with fast turn-on. They are simple and much easier to build than other switches, such as high-power thyratrons or vacuum tubes. As a result, they are well suited for one-of-a-kind, high-power machines. The switches can be triggered or controlled for timing purposes by distorting the electric field or by initiating a breakdown plasma with a low-energy spark or ultraviolet (UV) light.

However, the Achilles heel of the spark-gap switch is its relatively slow recovery, or return to initial conditions. Because the switched energy is conducted through a narrow channel, temperatures are high. In pressurized gas, deionization occurs relatively quickly once the energy is discharged. However, the hold-off voltage that a gas can withstand is inversely proportional to the temperature, so the gas must cool before voltage can be reapplied. Recovery to initial conditions may take several seconds.

From the beginning of the Dahlgren Division's CPB efforts, it was recognized that the spark gap is one of the few switches that could meet most of the accelerator power requirements, with the glaring exception of the repetition rate. Contracts were initiated to investigate high-speed blowing of the gas across the gap to cool the region.⁶ However, for repetition rates above a kilohertz,

the use of large blowers, or even wind tunnels, was impractical. It was obvious that improvements needed to be made on the basic recovery mechanisms of spark switches.

Spark-Gap Recovery

Based on previous experience working with high-repetition-rate spark gaps,^{7,8} in-house research was initiated to determine the dominant factors controlling spark-gap recovery.

Overvolted Experiments

The early experiments to improve spark-gap recovery centered around the obvious variables of gas species and pressure. A series of two-pulse experiments was designed to determine the recovery (hold-off) voltage as a function of time. In these experiments, two fast-rising, high-voltage pulses were applied to a spark gap with a variable time delay between them. The first pulse was used to break down the gap by "overvolting" (exceeding the breakdown limit), and the second pulse was used to determine how much voltage could be reapplied to the gap after a time delay. Plots of recovery voltage versus time were obtained for various gases and pressures.

An important result of these experiments was that the recovery curves for all gases and electrode materials tested had a plateau region at a voltage level corresponding to the static break-

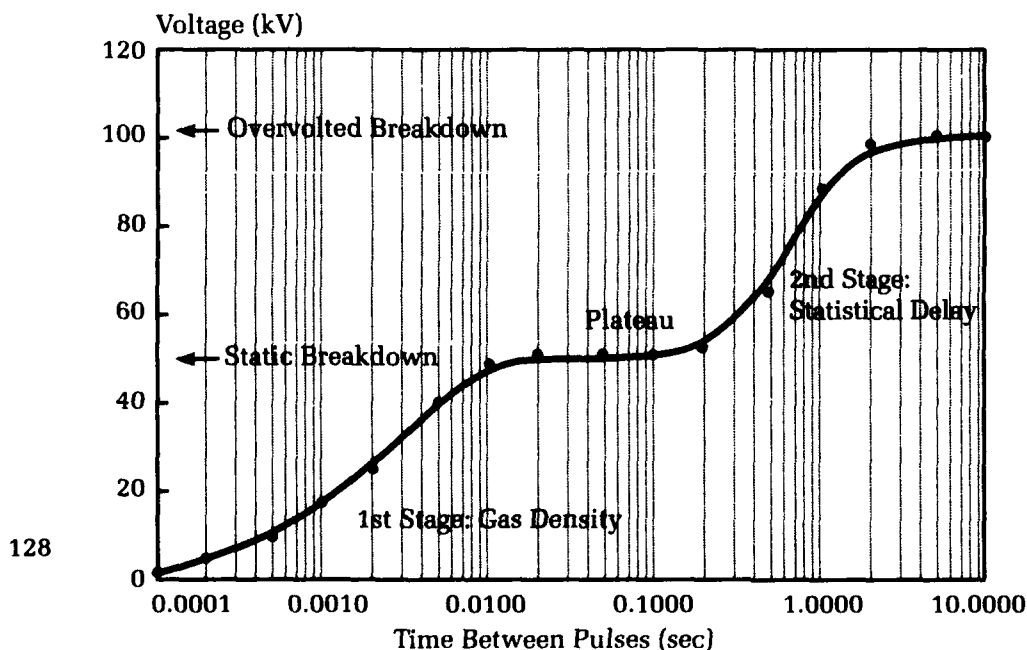


Figure 1. Two stages of overvolted spark-gap recovery. Charge-voltage rise time is 1 kV/ns in an air gap at 450 kPa (50 psig).

down voltage. It was determined that for an overvolted or untriggered spark gap, the recovery process occurs in two stages,⁹ as shown in Figure 1. The first stage was the recovery of gas density (associated with cooling), which provided the gas with its static or DC hold-off voltage.¹⁰ The second stage, which was much longer, was the recovery of statistical delay time, or the ability to be overvolted.¹¹ Overvolting occurred because of the finite time required to develop an arc channel, during which the voltage across the gap continued to rise. For a spark gap electrically triggered near static self-break, only the first stage was important for the recovery of the main gap.

Since the switches for accelerators need to be triggered for timing purposes, it became clear that recovery data on triggered switches would be necessary. Data were obtained by placing a trigger pin through a hole in the ground electrode (trigatron configuration). A low-energy voltage pulse to the trigger pin would develop ionization in the region and initiate the main breakdown at electric fields lower than self-break.

Improvements with High-Pressure Hydrogen Gas

Using triggered spark gaps, typical main-gap recovery times were shown to be about 10 milliseconds for almost all common spark-switch gases such as air, nitrogen, argon, oxygen, and SF₆. One particular gas, hydrogen, showed a marked improvement. Its high molecular speed

and thermal diffusivity allowed the recovery time to be an order-of-magnitude faster, or about 1 millisecond, as shown in Figure 2. The high thermal velocity of hydrogen allowed fast channel expansion, which reduced turn-on time and gap losses.¹² Fast recovery time was demonstrated with gap spacings from 0.1 mm to 1 cm at pressures from atmospheric up to 7 Mpa (1,000 psig). High gas pressures and small gap spacings had the advantage of a shorter arc, which reduced inductance, resistance, and allowed closer gas contact to metal surfaces. High-density gas also increased heat capacity, increased the breakdown strength, improved turbulence, and reduced statistical time. High-pressure hydrogen allowed operation of a low-loss spark-gap switch with 1 millisecond between pulses and without gas flow.

Improvements by Undervolting the Gap

For triggered gaps, the recovery-versus-time plots for all the gases tested exhibited the profile shown in Figure 2, extending typically two orders-of-magnitude in time. The amount of time needed to recover half of the self-breakdown voltage was about 10 times less than the time needed to recover the full self-breakdown voltage. Recovery of a spark-gap switch was further improved by triggering the switch well below static or self-break. Recovery time was decreased by making the operating voltage significantly less than the static (DC) breakdown voltage, thus

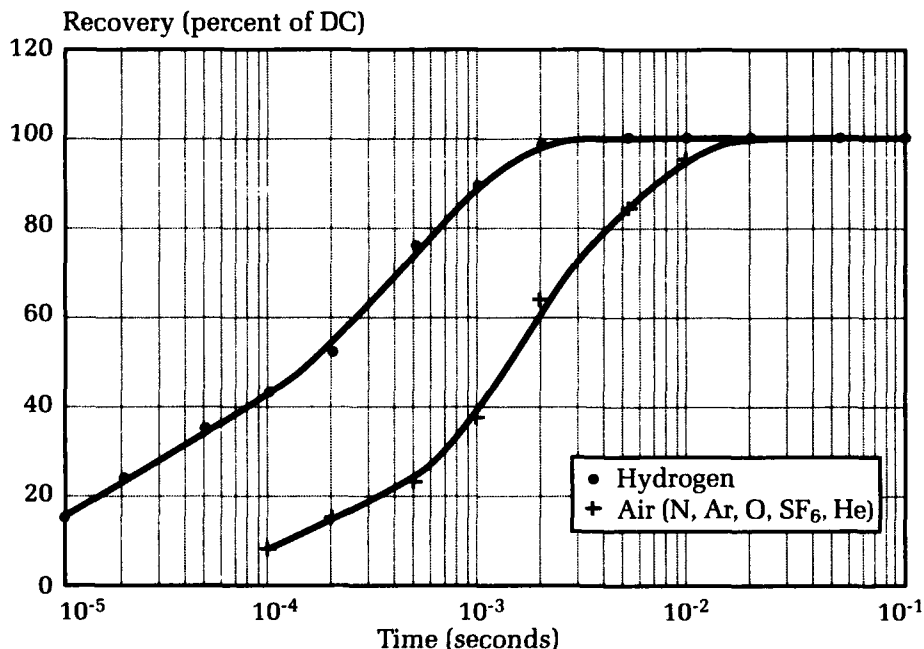


Figure 2. Recovery time for various gases in a high-pressure spark gap. Pressure is 1.4 MPa (200 psig) in a 2.5 mm gap.

allowing voltage to be reapplied before the gas returned to ambient temperature. Tests with hydrogen showed that recovery times (to the operating voltage) could be reduced an order-of-magnitude when the gap was "undervolted" by approximately 50 percent. The undervolting, in effect, allowed the gap to operate at higher temperatures. The gas did not need to cool completely in order to prevent a spark from reforming. Trigger voltages typically equaled the working voltage of the gap, although energies were much lower. Spark streamers were created from the trigger pin to both main electrodes simultaneously to allow operation at the maximum pressure.¹³

Tests were performed in 1987 at low energy (5 Joules) at 120 kV peak voltage and 200 A peak current. One-hundred-microsecond recovery (without gas flow) was demonstrated in a spark gap using high-pressure hydrogen in a trigatron configuration.¹⁴ A single-shot trigger initiated breakdown. Reapplication of the operating voltage without breakdown verified full recovery. This fast recovery time represented improvement over conventional spark gaps by a factor of 100. A patent has been obtained for these techniques and the switch design.¹⁵

Switch Description

The hydrogen switch used in the following experiments is shown in Figure 3. The pressure housing, 20 cm in diameter, is made of stainless steel to withstand gas pressure of 7 Mpa (1,000 psi). The insulators are machined from ceramic, and the electrodes and trigger pin are made from

a copper-tungsten alloy. A quartz window is provided to view the gap, and silicon O-rings are used for sealing. Trigger-gap spacings are typically equal to the main-gap spacings of about 6 mm. As a safety precaution, hydrogen gas passes through excess-flow shut-off valves and flash arresters. Isolation capacitors with a 100-kV, 100-ns rise-time trigger generator are typically used to trigger the switch.

Pulse-Forming-Line Experiments

In fiscal year 1988, fast recovery of an undervolted hydrogen switch was demonstrated at higher energies (200 Joules) using a 200-ns pulse-forming-line (PFL) as shown in Figure 4. This PFL was developed to demonstrate long-term charging of water-glycol mixtures and was

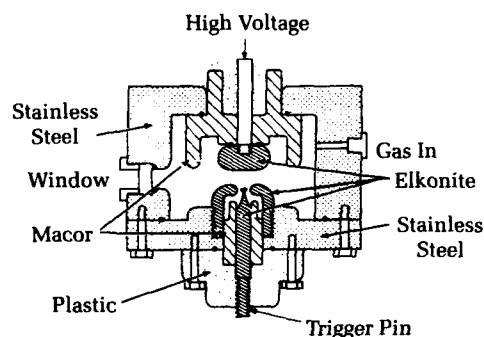


Figure 3. Cross section of high-energy switch. Diameter is 20 cm.

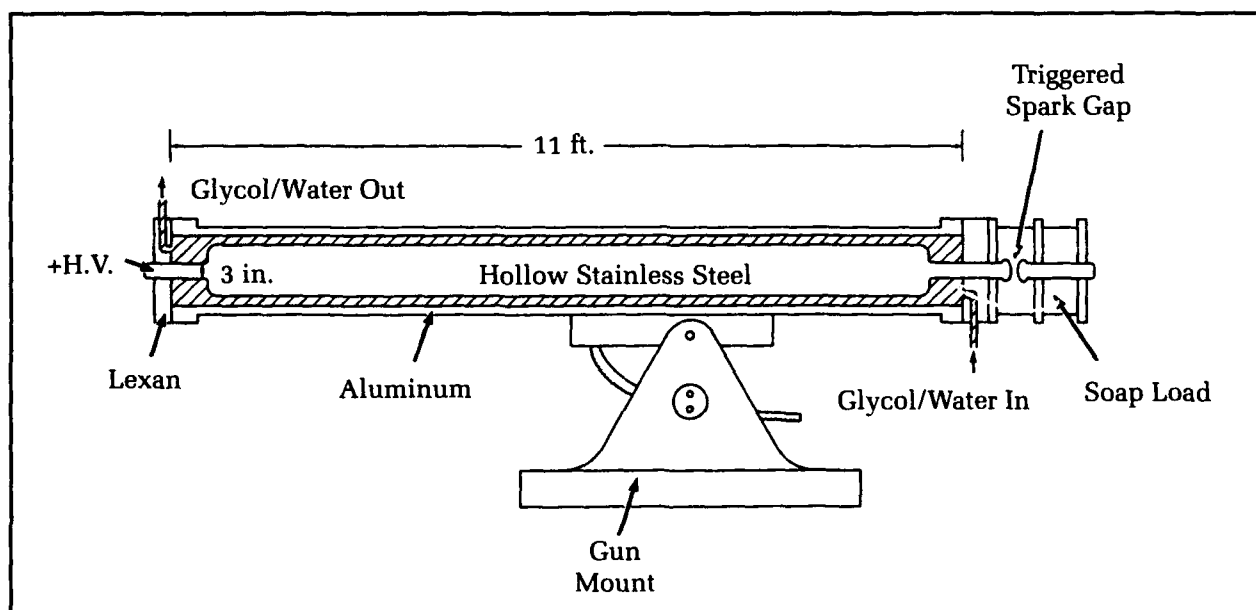
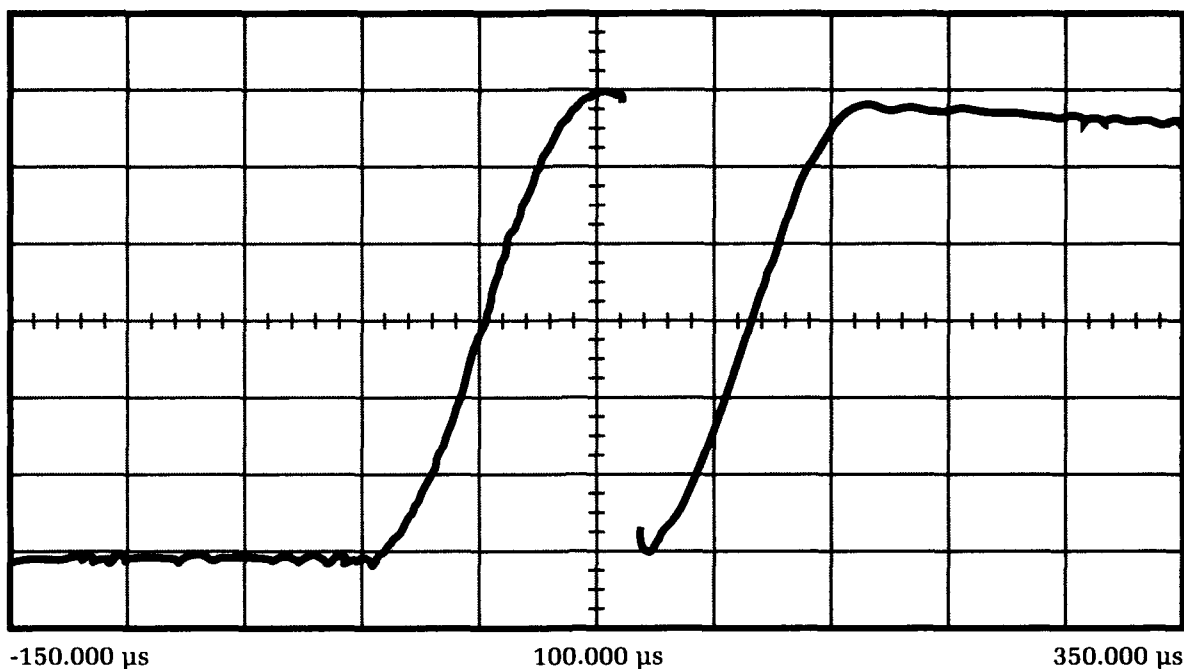


Figure 4. A 200-ns water/glycol pulse-forming-line.



Ch. 1 = 2.000 volts/div
Timebase = 50.0 μ s/div

Offset = 6.000 volts
Delay = 100.000 μ s

Figure 5. PFL voltage waveform showing full recovery at 60 kV and 35 kA. Vertical scale, 10 kV/div., horizontal scale, 50 μ s/div.

driven by the Dahlgren Division's pulsed-power facility.¹⁶ The 0.86-ohm PFL was resonantly charged using a 10-mH inductor and was terminated in a matched liquid load. The spark gap was triggered by a single-shot trigger at the peak of the first pulse. The voltage was reapplied in a 1-cos waveform (resonant charge) during the entire recovery period. The voltage waveform across the PFL is shown in Figure 5. With proper undervoltage, the spark gap did not self-fire when the second voltage pulse was applied, showing full switch recovery to 60 Kv in 100 microseconds. Current during the 200-ns pulse was 35 kA. Hydrogen gas pressure was up to 7 Mpa (1,000 psi). Multiple pulses could not be tested because a rep-rated trigger was not available. Based on these results, an undervolted hydrogen switch appeared to have potential for use in CPB accelerators, and we began looking at specific requirements.

Switch Requirements for a Compact Accelerator

CPB accelerators require extremely high voltages. Creating and controlling more than a few megavolts in a reasonable volume is very difficult. Therefore, an accelerator typically consists of a series of stages where the beam can be accel-

erated through a small portion of the field in each stage.

Basic Pulse-Power Requirements for CPB Accelerators

As an example, the pulsed-power requirements for a 50-million-electron-volt (MeV), 10-kA, 40-ns beam will be described. The acceleration process typically begins with an injector, which creates a burst of electrons from an emissive surface, accelerates them to a few MeVs, and injects them into a series of accelerating stages. Each accelerating stage produces a megavolt across a gap of a few centimeters for about 40 ns. To maintain full voltage across the gap, the current driving the accelerating stage must be at least equal to the beam current of 10 kA. After passing through about 50 gaps, the electrons will have 50 MeV of energy. The pulsed-power driver for each stage must therefore produce 10 GW of peak power (a megavolt at 10 kA) for 40 ns. The driver must also operate in a burst mode of repeated pulses to allow the beam to bore a hole through the atmosphere. The repetition rate during the burst may be on the order of 10 kHz.

To reduce the size of the accelerator, designs are under development that recirculate the beam through the same stages many times. The

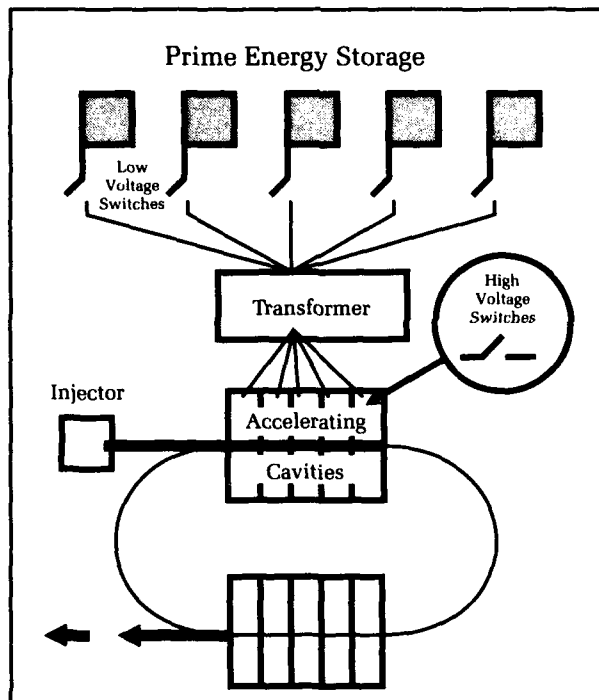


Figure 6. Pulsed-power system for a 5-pulse recirculating accelerator.

hydrogen switch technology is currently being developed for use in a recirculating accelerator being developed at Sandia National Laboratories. The accelerator will use 10 stages with multiple passes through each stage. The stages use water-dielectric cavities to allow high-energy density and low inductance. Figure 6 is a diagram of the pulsed-power train needed for a 5-pulse burst. Each prime energy store must have enough energy to charge five accelerating cavities (stages) in parallel. An identical system powers the five cavities on the other side of the accelerator. Two sets of high-rep-rate switches are needed. One set is needed at low voltage, but high energies, to charge the accelerating cavities. The other set is needed to switch the cavities, and must operate (in water) at less energy, but higher voltage. No currently available switch can meet either set of requirements. Dahlgren Division hydrogen-switch technology is being studied as a candidate for both switches.

Low-Voltage Switch Requirements

The current and energy requirements for the low-voltage switches in Figure 6 can be obtained from order-of-magnitude requirements for the accelerator described above. For a 50-MeV, 10-kA, 40-ns-long beam (10 accelerating cavities),

about 1.6 kJ of energy are required for each cavity. However, 20 to 40 percent of the energy stored in the cavity is not used (to prevent excessive drop in the voltage), so about 2.5 kJ of stored energy are required. To charge five cavities in parallel, 12 kJ of stored energy are needed, or 10 μF of capacitance at 50 kV. A 15- μs charge time would be typical for a water cavity without excessive loss. These requirements determine the energy, voltage, and peak current that must be handled by the low-voltage switches. The low-voltage switch requirements are: 170 kA peak at 50 kV for a 15- μs pulse. Each low-voltage switch needs to recover once in about 100 μs for each burst of pulses.

High-Voltage Switch Requirements

In the Sandia approach, the accelerating cavities use water as the dielectric medium, and use impedance mismatches to create reflections in the output waveform. These reflections are timed to generate multiple decaying output pulses that drive the recirculating beam. For a 50-MeV accelerator (10-MeV injector) with 10 cavities and four passes, each accelerating cavity must average about a megavolt output over four pulses. The cavity impedance mismatch provides a voltage step-up of about three, which translates to a cavity charge voltage and switch operation of about 500 kV. The switch will likely see an effective impedance of less than one ohm from the cavities. Since the rise time of the waveform must be 10 ns or less to achieve a reasonable flat-top output pulse, the switch inductance (L/R time constant) must be less than 4.5 nanohenrys (nH). The present hydrogen switches have about 100 nH of inductance, requiring 20 switches in parallel to achieve the required inductance. The current and energy in each cavity are divided among the twenty switches. Jitter requirements (deviation in switch timing) are very strict for multiple switches, since all must fire together to share the current. Nanosecond switch jitter may be required. About 20 nanofarads (nF) of capacitance are needed to store the required energy in each cavity. Therefore, each switch will handle about 1 nF. Peak current will be about 500 kA going through 20 switches, or 25 kA per switch. Total energy transferred per switch is only 125 joules. Since four passes are needed, total current conduction time is about half a microsecond. The high-voltage switch requirements are therefore 500 kV, 25 kA, 125 J, low jitter, and multiple-pulse burst operation at repetition rates up to 10 kHz.

Demonstration Experiments

Two demonstration experiments are described in this section.

Low-Voltage, High-Energy Experiments

In fiscal year 1989, successful switch testing was conducted at the parameters needed for the low-voltage switches described above. Two 10- μ F capacitor banks were connected via hydrogen switches to a common 0.1-ohm resistive-metal load that simulated the accelerator. A schematic of the circuit is shown in Figure 7. When charged to 50 kV, the capacitor banks stored 12.5 kJ each. If the first spark gap did not recover before the second gap closed, the energy oscillated between capacitor banks rather than through the load, which was very obvious from the diagnostics. Geometric inductance was about 600 nH. Separate single-shot triggers were used for each hydrogen switch. The resistive load damped the oscillations after about 1 cycle, giving a current pulse width of about 10 μ s. Peak currents were 170 kA.

Figure 8 shows successful full recovery at 100 μ s at 2.7 MPa (400 psi) at full voltage. The top trace shows the current through the load from both pulses. The time between measurable current flow is less than 80 μ s. Jitter in the switch was less than 50 ns. The system was also operat-

ed single-shot with a shorted resistive load at 260 kA peak current for an oscillating pulse 80 μ s in duration. There was no gas flow in this experiment.

This experiment showed that the recovery of an undervolted, high-pressure hydrogen switch is a surprisingly weak function of the energy discharged through the switch. In fact, the recovery time varied little from millijoules to kilojoules of transferred energy. It appears that the higher energy transfer creates a larger, brighter, more turbulent and, perhaps, slightly hotter arc, which allows excess heat to be expelled in roughly the same time period. The experiment

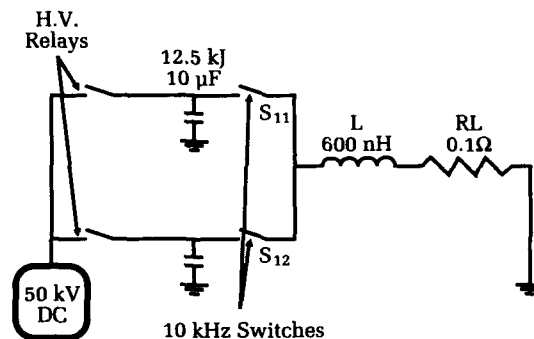


Figure 7. Simplified schematic of high-energy experiment.

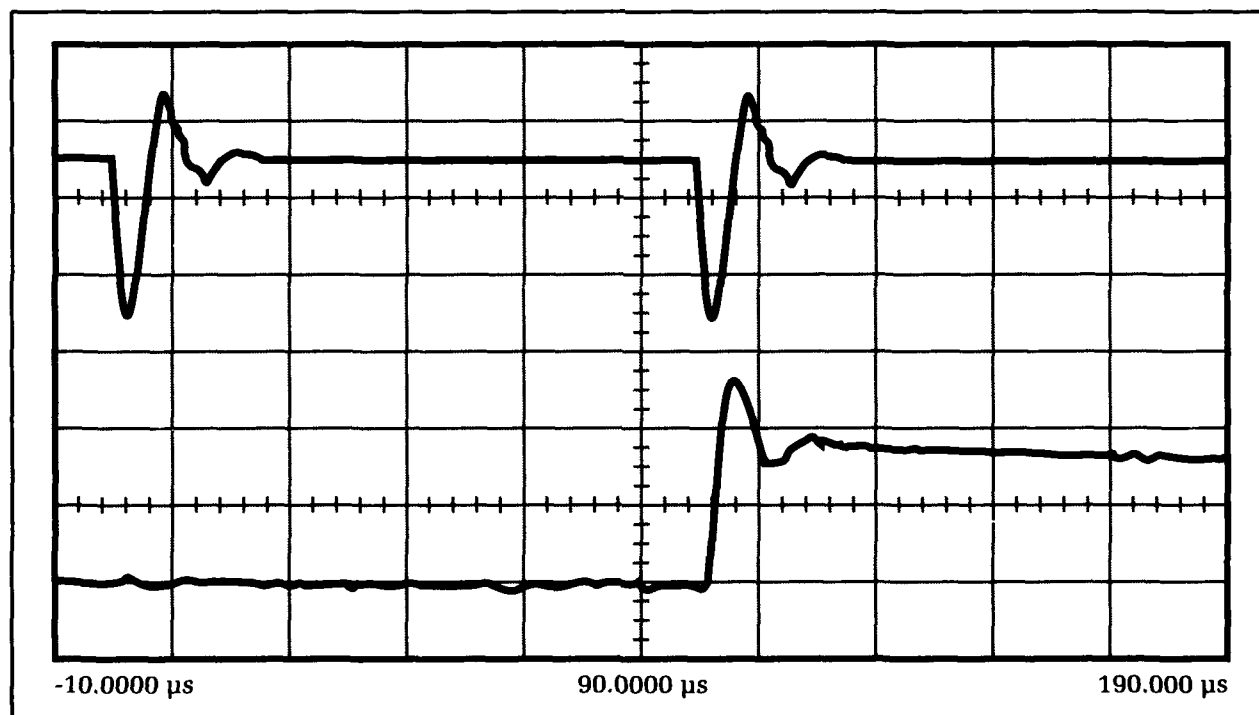


Figure 8. Current waveform showing high-energy recovery in 100 μ s at 2.7 MPa (400 psig) hydrogen after 50 kV, 12.5 kJ pulse with 170 kA peak current. Top trace (Ch-1): current through resistive load at 80 kA/div. Bottom trace (Ch-2): voltage on second capacitor bank at 25 kV/div. Time scale is 20 μ s/div.

served as a full-scale demonstration that an undervolted hydrogen switch could meet the low-voltage switch requirements needed to power a rep-rated compact accelerator.

High-Voltage, 5-Pulse Experiment

Work began in fiscal year 1990 to build a 5-pulse system that could generate the high voltages (500 kV) needed to test the hydrogen switch technology at the parameters required for their use as the high-voltage switches in rep-rated accelerating cavities. The experiment was designed to simulate the portion of an accelerating cavity that each high-voltage switch would handle. The basic thrust of the experiment is to build a 50-kV, 5-pulse system to drive a step-up pulse transformer to 500 kV. The transformer charges a high-voltage capacitor which is discharged by a single, fast-recovery hydrogen switch in a five-pulse burst. A schematic of the experiment is shown in Figure 9.

Each primary capacitor (0.7 μF) is charged through high-voltage relays to 50 kV DC, and is discharged through a fast-recovery hydrogen switch controlled by a single-shot trigger in a set-up similar to the high-energy tests described above. Since all switches are connected to a com-

mon load, each switch must recover before the next one fires. The low-voltage, 5-pulse section has been completed and successfully operated at 50 kV (875 Joules), 40 kA at a repetition rate of 5 kHz into a resistive load. The output of the low-voltage, 5-pulse section into a resistive load is shown in Figure 10. Hydrogen switches at a pressure of 225 psig were used to discharge each capacitor. The system has also operated at 10 kHz repetition rate at 45 kV charge and 275 psig.

The following year, we concentrated on constructing the high-voltage section of the 5-pulse system shown in Figure 11. The high-voltage transformer is an air-core, dual-resonant type with a 5-turn primary and an 80-turn secondary. The transformer provides 500 kV with 42 kV on the primary. Voltage rise times from the transformer are about 12 μs in order to be compatible with charging a deionized-water accelerating cell.

A 4-nF, 500-kV capacitor is needed in the high-voltage section of the experiment. The capacitor consists of nine, 0.04- μF capacitors rated at 60 kV each, stacked in series. This will create a 4.4-nF capacitor rated at 540 kV. The capacitor must be able to withstand high current and high voltage reversal. Exceeding the specified voltage and reversing the voltage will reduce lifetime, but should still be adequate for

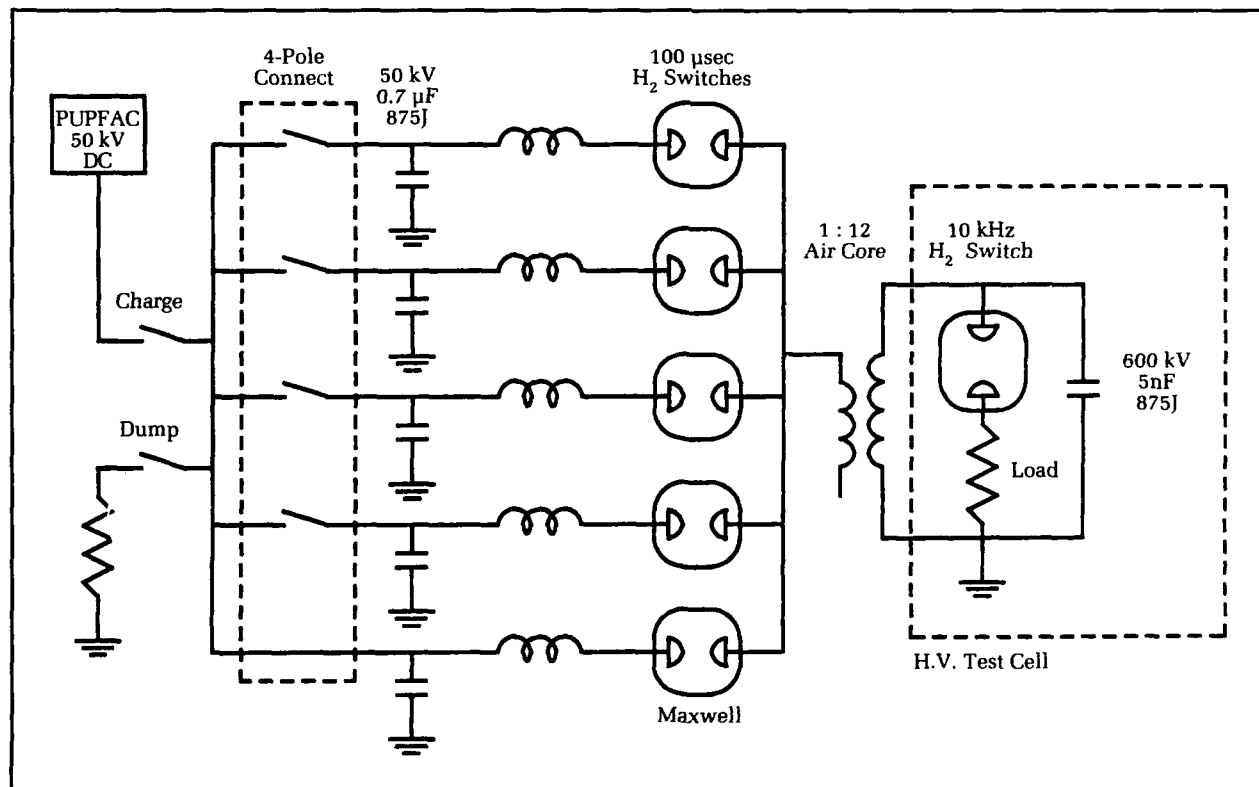
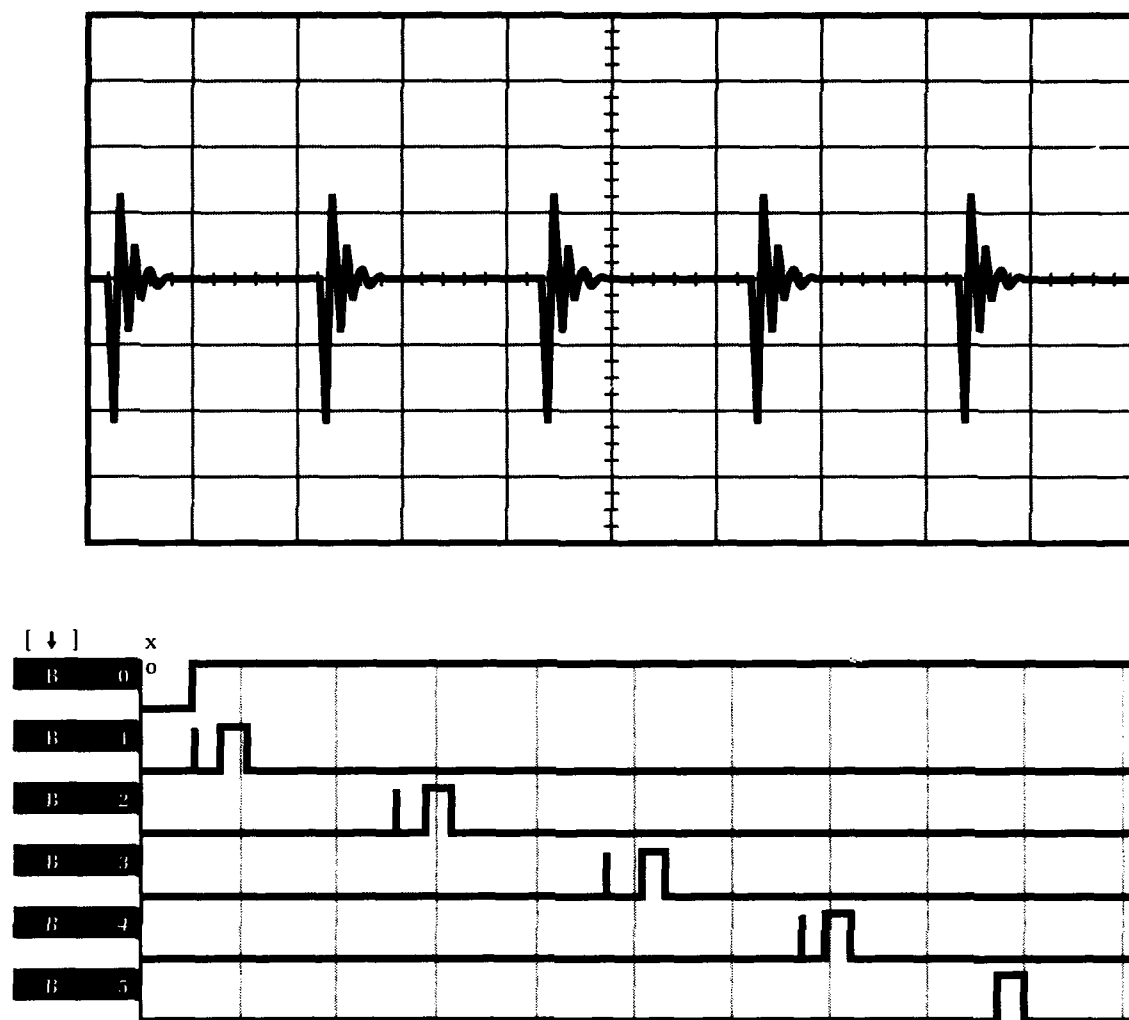


Figure 9. Schematic of the high-voltage, 5-pulse test setup.



Full recovery of a 5-pulse burst at 10 kHz. Hydrogen pressure 2 MPa (275 psig). Charge voltage 45 kV (700 Joules/pulse).

Figure 10. The output of the low-voltage section into a resistive load. Top view: Current through 0.375-ohm load 10 kA/div., 50 μ s/div. Bottom view: Light output from switches 50 μ s/div.; channels 1-4 are hydrogen switches.

our testing. The capacitor stack is submerged in water, which will resistively grade the voltages and serve as a residual discharge path to ground between pulses. Since the dielectric constant of the water is 80 and the plastic case of the capacitors is 3, much of the electric field to the wall of the grounded water tank would appear across the plastic case. The capacitors are therefore surrounded with one-inch-thick polyethylene.

The high-voltage switches are being tested under water to simulate the accelerating cavities. The deionized water is contained in an aluminum tank about 1.2 meters square and 0.75 meter tall with a divider down the center. We are using one side of the tank (in water) for the

switch, capacitor, load, diagnostics, transformer, and coil, and the other half (in air or oil) for the trigger systems. The resistivity of the deionized water is about 10 Ω M-cm. A resistive load is used to damp the oscillations.

The high-voltage capacitor is discharged by a high-voltage, pressurized hydrogen spark gap controlled by a multiple-pulse trigger. We are using a 1.3-MV switch¹⁷ from Sandia, which is an older and higher-voltage version of the switches currently used in a single-shot recirculating cavity. A drawing of the switch used in the Sandia cavities is shown in Figure 12. Modifications include reducing the main gap spacing from 4.4 cm to 2.9 cm to allow the

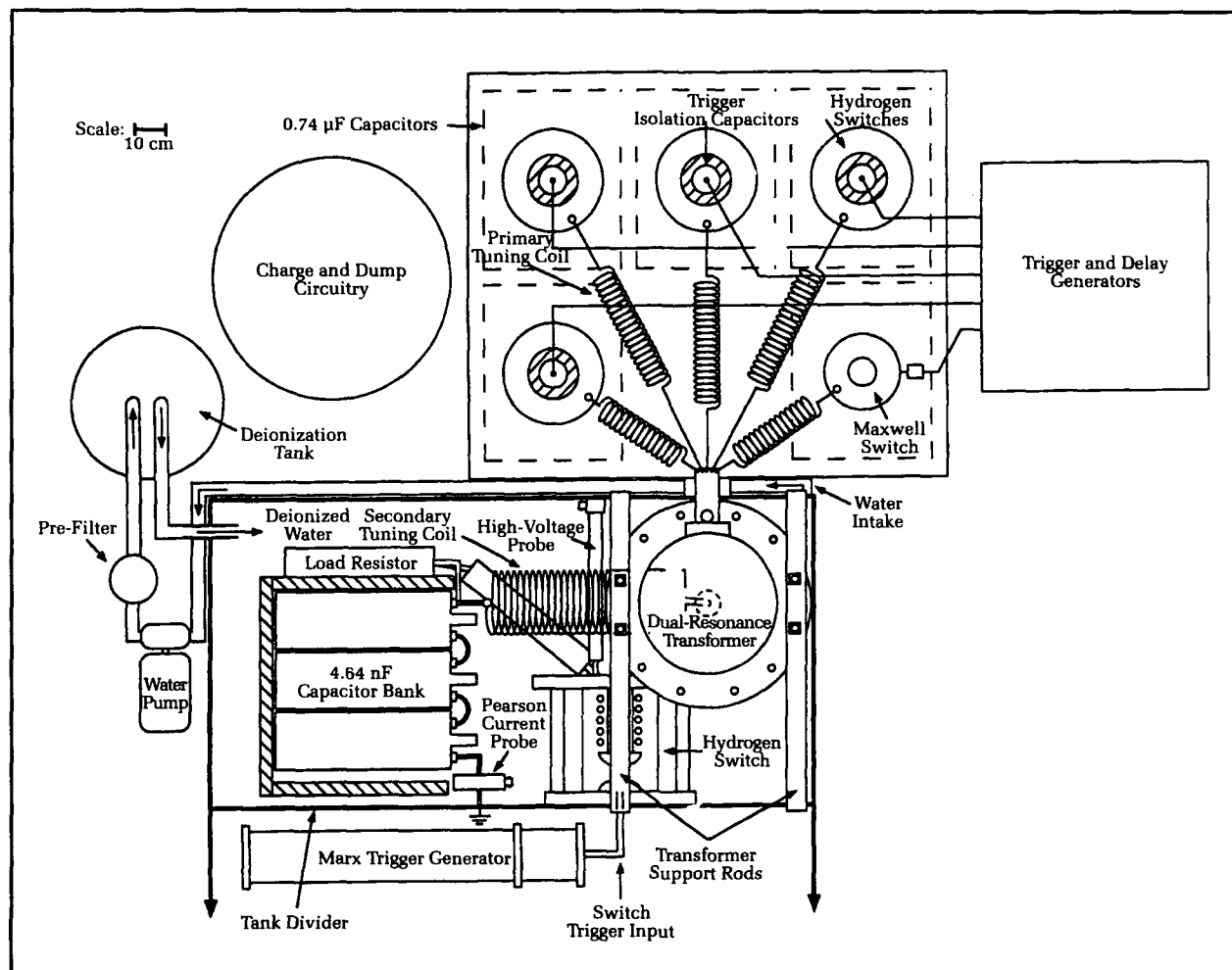


Figure 11. The 5-pulse high-voltage system.

switch to operate at lower voltages and with the lower breakdown strength of hydrogen gas. This switch has a single gap with trigger disk placed near one electrode. Based on our previous tests with hydrogen, we will not be able to undervolt the switch enough to achieve 10 kHz. However, we estimate that this switch will operate above a kilohertz in hydrogen. This will allow a very simple and inexpensive method of rep-rating Sandia's accelerating cavities to a few kHz with minimum cost for modifications.

A second switch design to be tested is the TEMPO, developed by Sandia. This is a staged switch (shown in Figure 13) in which the electric field grading is done by spiraling resistive wire around the switch. The switch is very small, an advantage for high-pressure hydrogen. It has been enclosed in a second container of SF_6 to prevent flashover on the outside of the housing. This switch should operate at a lower percent of self-break, and therefore a higher repetition rate. The major disadvantage is that the trigger requirements are more severe.

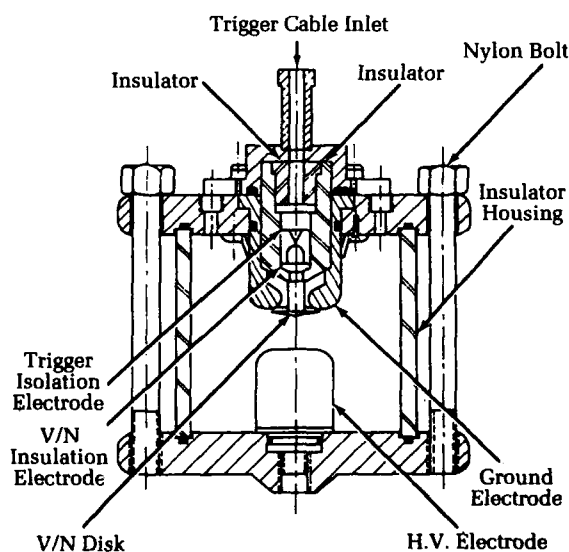


Figure 12. A single-shot "V/N" switch used in Sandia's recirculating accelerator.

One of the principal concerns is the development of a rep-rated trigger capable of a burst of five pulses at voltages of at least 100 kV. To our knowledge, such a system has never before been built. We are developing a miniature Marx generator capable of burst-mode operation. The system is centered around a MiniMarx designed at Los Alamos,¹⁸ built by Veradyne Corp., and shown in Figure 14. The Marx is designed for single-shot operation and produces a 200-kV pulse with a 2-ns rise time using a 30-kV DC input. The entire Marx is in a tube 4 inches in diameter and 2 feet long. We are attempting to operate the Marx in a burst mode by using hydrogen gas in the Marx and a Velonex high-voltage pulse generator (hard-tube pulser) as a trigger source. We have achieved 5-pulse bursts at repetition rates of 3 kHz with this system operating about 80 percent of self-break.

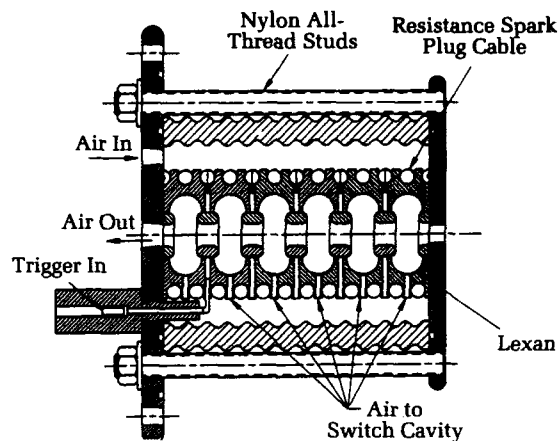


Figure 13. The TEMPO multistage switch.

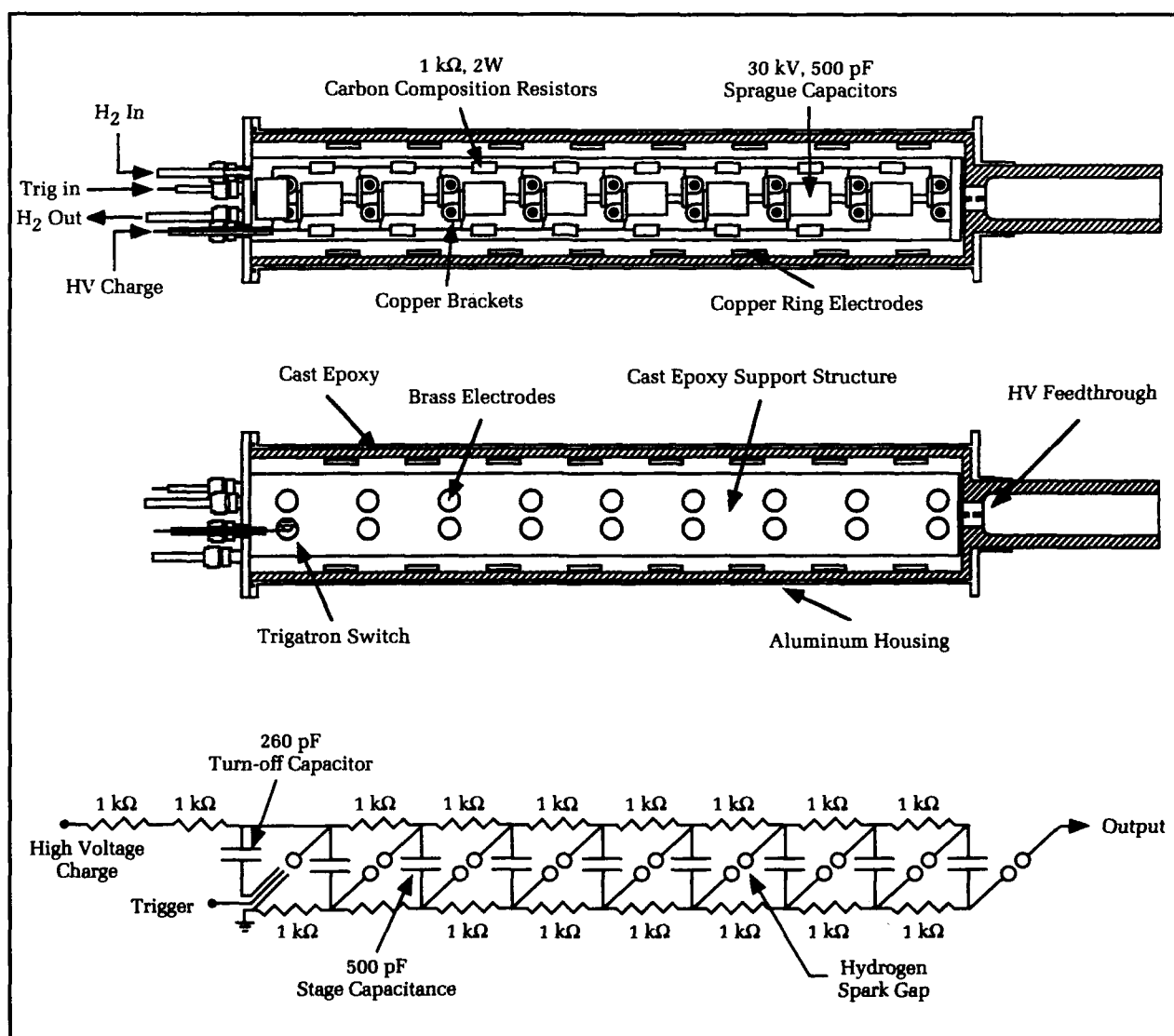


Figure 14. The MiniMarx modified for high repetition rate.

The 5-pulse system is being powered by the Dahlgren Division's Pulsed-Power Facility. Most of the diagnostics and controls are done optically. Video monitors are used to observe high-voltage switch operation. The light output from each of the six switches is monitored by a series of fiber-optic cables connected to a logic analyzer.

The high-voltage section has been assembled and filled with water, and tests are under way. We have begun testing at half voltage (250 kV) using a single primary switch. The high-voltage switch has also been successfully triggered at this voltage, and 500 μ s recovery has been demonstrated using a single-shot trigger with two high-voltage pulses. The system has successfully operated in a 5-pulse burst at 1 kHz at 250 kV. Preliminary results indicate that the switch parameters will scale to higher voltages and will work with multiple-pulse bursts. Jitter measurements also look promising. The purpose of these tests is to demonstrate that a fast-recovery hydrogen switch will operate at 500 kV in multiple-pulse bursts and will, therefore, be suitable for high-repetition-rate compact accelerator cavities.

Conclusions

A high-power spark-gap switch capable of 100- μ s recovery times has been demonstrated at high voltages and currents. Fast recovery is accomplished without gas flow using two techniques: high-pressure hydrogen gas, and triggering the main gap in a highly undervolted state. These techniques have been demonstrated at voltages up to 120 kV, peak currents of up to 170 kA, and energy transfers of up to 12.5 kJ. Tests are under way to test the concepts at 500 kV and for multiple-pulse bursts.

Significant effort and money have been expended in the Department of Energy, Department of Defense, and industry to study many different technologies in an attempt to meet the pulsed-power switch requirements for CPB weapons. After more than ten years' effort, no switch technology has been able to meet all requirements simultaneously. The Dahlgren Division's hydrogen switch comes closest to meeting the requirements and has emerged as

the leading candidate for use in high-repetition-rate recirculating accelerators.

The hydrogen switch effort has now achieved widespread recognition, including discussions at opening and closing sessions of the 1991 International Pulsed Power Conference and the 1991 DARPA/SDIO/Services Charged Particle Beam Review Conference. Papers on this effort have recently been presented at a number of conferences, workshops, and symposia.^{19,20,21}

References

1. Zahn, M., et al., "Dielectric Properties of Water and Water/Ethylene Glycol Mixtures for Use in Pulsed Power System Design," *Proc. IEEE*, Vol. 74, No. 9, Sep 1986.
2. Gehman, V. H., Berger, T. L., and Gripshover, R.J., "Water-Based Dielectrics for Energy Storage and Pulse-Forming Lines," *19th International SAMPE Technical Conf.*, Arlington, VA, Oct 1987.
3. Moran, S. L. and Hardesty, L. W., "High-Repetition-Rate Hydrogen Spark Gap," *IEEE Transactions on Electron Devices*, Vol. 38, No. 4, Apr 1991.
4. Moran, S. L. and Hardesty, L. W., "High Repetition Rate Hydrogen Spark Gap," *IEEE 1990 19th Power Modulator Symp.*, San Diego, CA, Jun 1990, p.227.
5. Humphreys, D. R., et al., "Rimfire: A Six Megavolt Laser-Triggered Gas-Filled Switch for PBFA II," *5th Pulsed Power Conf.*, Arlington, VA, Jun 1985, p.262.
6. Kuhlman, J. M. and Molen, G. M., "Performance of High-Power Gas-Flow Spark Gaps," *AIAA Paper 85-0134*, AIAA 23rd Aerospace Sciences Meeting, Reno, NV, Jan 1985.
7. Moran, S. L., *High-Repetition-Rate L-C Oscillators*, NAVSWC TR-3658, Dec 1977.
8. Moran, S. L. and Rinehart, L. F., "Voltage Recovery Time of Small Spark Gaps," *IEEE Transactions on Plasma Science*, Vol. PS-10, No. 4, Dec 1982.
9. Moran, S. and Hairfield, S., "Recovery of Overvolted Spark Gaps," *Proc. IEEE 16th Power Modulator Symp.*, Arlington, VA, Jun 1984, p.71.
10. Moran, S. and Hairfield, S., "High Pressure Spark Gap Recovery After Overvolted Breakdown," *Proc. IEEE 5th Pulsed Power Conf.*, Arlington, VA, Jun 1985, p.473.
11. Mazzola, M., Molen, G. M., and Kuhlman, J. M., "Recovery of a Gas-Blown Spark Gap with UV Preionization," *Proc. IEEE 18th Power Modulator Symp.*, Hilton Head, SC, Jun 1988, p.219.
12. Kushner, M., Kimura, W., and Byron, S., "Arc Resistance of Laser Triggered Spark Gaps," *J. of Applied Physics*, Vol. 58(5), Sep 1985, pp.1744-51.

13. Peterkin, F. and Williams, P. F., "Physical Mechanisms of Triggering in Trigatron Spark Gaps," *Applied Physics Letters*, Vol. 53, 1988, p.182.
14. Hutcherson, R., Moran, S., and Ball, E., "Triggered Spark Gap Recovery," *Proc. IEEE 6th Pulsed Power Conf.*, Arlington, VA, Jun 1987, p.221.
15. Moran, S. L., and Hutcherson, R. K., "High PRF High Current Switch," United States Patent No. 4,912,369,27, Mar 1990.
16. Moran, S., Hutcherson, R., and Ball, E., "A Facility for High-Repetition-Rate Switch Testing at Medium Power," *Proc. IEEE 6th Pulsed Power Conf.*, Arlington, VA, Jun 1987, p.252.
17. Tucker, W. K., et al., "Compact, Low Jitter, Triggered Spark Gap," *Proc. 5th IEEE Pulsed Power Conf.*, Arlington, VA, Jun 1985, p.254.
18. Platts, D. A., "10-Joule High-Voltage Trigger Micro-marx," *Proc. 3rd IEEE Pulsed Power Conf.*, Albuquerque, NM, Jun 1981, p.485.
19. Moran, S. L., *Pulsed-Power Technology Program: Accomplishments in Switching*, NAVSWC TR 89-259, Sep 1989.
20. Moran, S. L., Grothaus, M. G., and Hardesty, L. W., "High-Repetition-Rate Hydrogen Spark Switch," 1991 *DARPA/SDIO/Services Annual Charged Particle Beam Review*, Silver Spring, MD, Jun 1991.
21. Moran, S. L., Hardesty, L. W., and Grothaus, M. G., "Hydrogen Spark Gap for High Repetition Rates," *Proc., IEEE 8th Pulsed Power Conference*, San Diego, CA, Jun 1991, pp.336-339.

The Author



STUART L. MORAN received the B.S. degree in physics in 1974 from Lynchburg College, Virginia. Since 1973, he has been working at the Naval Surface Warfare Center (now Dahlgren Division), Dahlgren, Virginia, in the areas of electronic warfare, pulsed power technology, and, in particular, high-power switches. He is presently the acting head of the Pulsed Power Technology Branch. Mr. Moran has supervised and conducted research

in the areas of pulsed-power systems, high-repetition-rate switches, optical and electrical diagnostics, directed-energy warfare, and expendable jamming devices using pulsed-power sources. He also has experience in electronic countermeasures techniques, jamming tactics, expendable device design and testing, and medium-power lasers for optical countermeasures. He has presented and published papers at international and joint-service conferences concerning both pulsed power and electronic warfare.

Next Generation High-Power Microwave Sources

H. C. Chen

"Stealth" aircraft and missiles have placed new demands on high-power radars to achieve reasonable detection ranges. High-power microwave sources that have been developed intensively during recent years have application to: high-power radars; high-energy, charged-particle accelerators; thermonuclear plasma heating; and electronic warfare. The objective of the research described in this article is to generate a large number of phase-controllable, high-power signals for use in detection systems. A novel approach generates large numbers of high-power microwave sources that not only operate at the same frequencies, but also have their phases under control. This makes it possible to combine the large number of sources coherently in order to produce much higher-power levels than achievable by a single device. In addition, using these multiple signals allows for single/multiple-sensor position tracking and identification. Using magnetron arrays to produce unique microwave sources is not only feasible, but has been successfully demonstrated at the Dahlgren Division by computer particle simulation.

Introduction

A magnetron is a diode, usually cylindrical, with a magnetic field parallel to its axis. In modern usage, however, the word implies a diode that, with the aid of a magnetic field, produces short electromagnetic waves. Those magnetrons that produce radiation within the wavelength range of 1 to 30 cm are defined here as microwave magnetrons.

A magnetron (Figure 1) comprises a field-emission cathode and coupled resonators embedded in the anode block. The electrons emitted from the cathode execute $\vec{E} \times \vec{B}$ drift motion under the influence of an axial magnetic field ($\vec{B}$) and a radial electric field ($\vec{E}$) generated by the voltage between two electrodes. A certain set of wave modes can be excited in the interaction space. If the phase velocity of one of these modes synchronizes with the drifting electrons, strong wave-particle interaction occurs and beam energy is converted to electromagnetic energy. The mode selection is determined by the geometric configuration and the magnitude of magnetic field and applied voltage. For a commonly used A6 magnetron with six resonant cavities, there are two common modes. One is the so-called π mode, for which the RF electric fields in adjacent resonators are 180° out of phase with one another. The other is called 2π mode, for which the RF fields of all resonators are in phase.

In recent years, high-power microwave sources have been developed to produce a power level of a few gigawatts (GW). Extension of power to much higher

levels is highly desirable, but faces some difficulties. For example, one simple way of reaching this goal would be to increase the cavity size; however, this approach is hindered by the mode competition and limitation of the electric field sustainable in resonant cavities. Therefore, combining groups of oscillators seems to be a good approach for bringing the power up to the level unattainable by a single oscillator. In order to combine a group of oscillators effectively, one must ensure that the phase of each oscillator can be controlled. Normally, the energy in the RF field is proportional to the number of sources N ; phase-locking produces power density proportional to N^2 . The microwave sources produced can be combined coherently to extend the power level for larger detection ranges against stealth targets. Schemes of locking techniques for a large number of oscillators have been attempted experimentally by Physics International^{1,2} and others.

The magnetron has been known as one of the most efficient and rugged devices for generating microwaves at frequencies of a few GHz. Power levels of up to several GW have been achieved in relativistic magnetrons operating at voltages from several hundred kilovolts to 1 MV, drawing kiloperes of current from field emission cathodes. Contrary to experimental progress on a single magnetron, a complete analytical and theoretical description of the magnetron does not yet exist because of the complicated geometry of the anode vane structure, the highly nonlinear coupling between the fields and particle dynamics. However, we have successfully simulated self-

consistently the operation of relativistic magnetrons by using two-dimensional electromagnetic particle-in-cell codes, MASK and MAGIC.³ Transient behavior from mode excitation up to nonlinear saturation state and spoke formation has been simulated for various beam parameters and geometry configurations of magnetrons.

Experimental Studies

The phase-locking of two relativistic magnetrons has been investigated experimentally.² Two relativistic magnetrons interact directly through a short waveguide connecting one vane of a magnetron to a vane of the other. The magnetrons are powered by a split, magnetically insulated transmission line and have similar electrical and microwave operating properties. Two identical A6 magnetrons connected by a waveguide are shown in Figure 2. The magnetrons were operated in the π and 2π modes at 2.8 GHz and 3.8 GHz, respectively. Waveguide coupling bridge lengths of 5.4λ and 8.5λ were chosen in the π and 2π mode operations, respectively. The two magnetrons were turned on simultaneously and, as expected, were found to be out of phase. Unfortunately, due to the experimental setup, the bridge length was not varied to control dependence of the phase difference of the two magnetrons. In other words, this experiment failed to demonstrate either in-phase or out-of-phase operation of the two magnetrons because of even and odd multiples of the half-wavelength of the connecting bridge length.

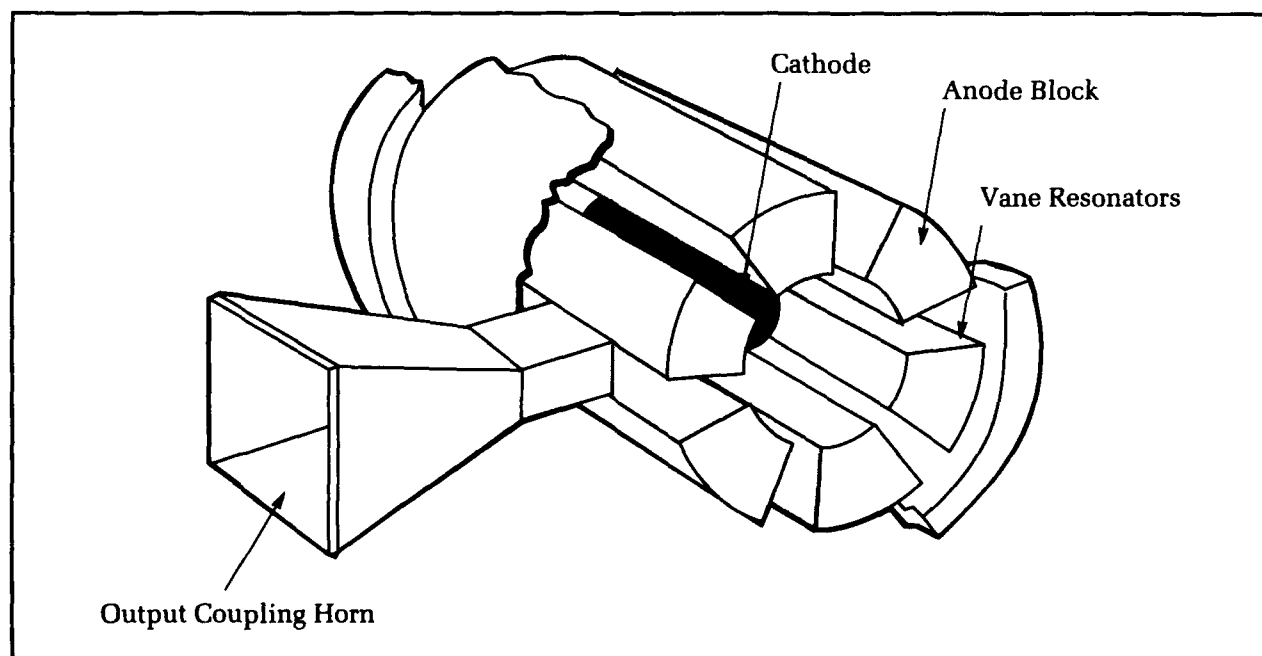


Figure 1. Schematic of a magnetron.

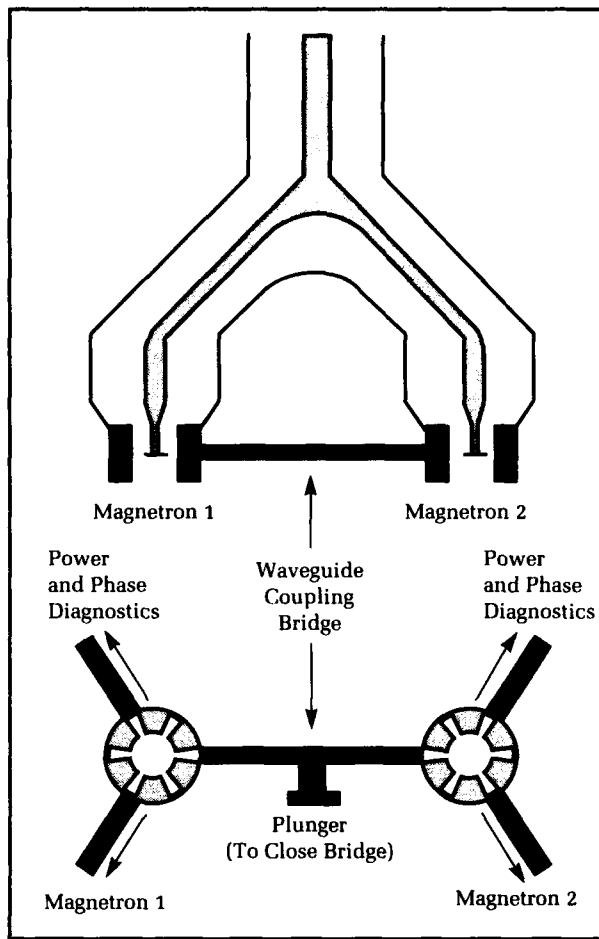


Figure 2. Identical magnetrons are connected by a waveguide.

Simulation Studies

We have studied phase-locking of dual magnetrons by means of direct particle simulation of two coupled magnetrons. Similar to the experimental setup, the phase evolution can be simulated with two magnetrons connected by a transmission line with various lengths and impedances, as illustrated schematically in Figure 3. The Physics International magnetron geometry configuration has been used in the calculations. The radii of cathode, anode, and vane are 1.27 cm, 2.1 cm, and 4.18 cm, respectively. The vane angle is 20° for the vane cavity opening. A multiple of the half-wavelength of the connecting length has been used to verify the phase-locking performance. Two magnetrons were run side-by-side and connected by a transmission line from a vane of one magnetron to a vane of the other. The MAGIC computer code has been set up to run this configuration. The phase-locking time can easily be observed from the time evolution of the phases of two magnetrons.

MAGIC is a 2-D, fully electromagnetic particle-in-cell code, developed by the Washington office of Mission Research Corporation to model self-consistently the electromagnetic fields and charged-particle dynamics of microwave and pulsed-power devices. In the simulations presented here, both the anode and the cathode are modeled as conducting surfaces. Macroparticles are created at the cathode in accordance with the space-charge-limited emission model in MAGIC. At each timestep, macroparticles of sufficient

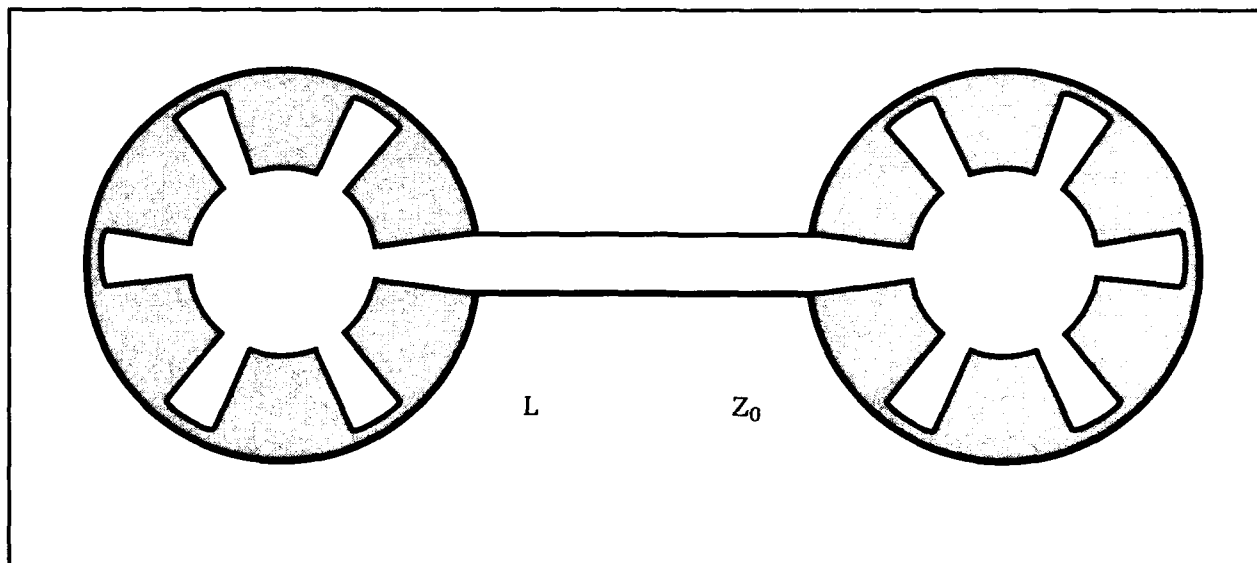


Figure 3. Two magnetrons coupled by a transmission line with length L and impedance Z_0 .

charge are introduced at the cells adjacent to the cathode to cancel any radially inward electric field ($E_R < 0$). The motion of these particles is then followed in the self-consistent electric and magnetic fields until the particles strike the anode or cathode, where they are considered lost. Unlike previous simulations of the relativistic magnetron, the simulations presented here model the full six vanes of the A6, rather than only two vanes.

Simulation Results

The analysis of phase-locking of dual magnetrons was conducted by means of direct particle simulation of two coupled magnetrons. The phase evolution can be simulated with two magnetrons connected by a transmission line with various lengths and impedances. Experimentally, the principal source of irreproducibility is the variation in the uniformity of the electron emission from the cathode surface. However, because of the unique repeatability feature of the computer run, it was necessary to vary the emission model in order to simulate properly the real irreproducibility nature of the experiment. For the simulation of two identical magnetrons connected by a transmission line, it has been found that the simulation results of phase-locking are dependent on the variation of the emission model. It is equivalent to saying that the phase-locking scheme does not work properly. The reason is very simple. In the earlier time operation of the magnetron, the emitting microwave pulse is still in the stage of random noise level. In the transient period, there is no single frequency mode domination, but only strong mode competition. Even if the magnetrons were connected by a transmission line and could talk to each other, they would fail to communicate because each lacks its own identity. Even if they could exchange information, the coupling effect leading to phase-locking would not be there. By the time each magnetron selects its own resonant mode after the transient time of a few nsec, if the phase difference between two magnetrons happens to match the phase shift across the transmission line, the transmission line will bring them to the phase-locking condition. Otherwise, two magnetrons interface with unwelcome signals and refuse to yield to each other. Therefore, it can be concluded that the two identical magnetrons cannot absolutely be coupled to phase-locking condition by a transmission line. This discrepancy between simulation and experiment needs further investigation.

Phase-Locking Conditions

The locking conditions and requirements for the phase between two mutually coupled, high-power relativistic magnetrons are summarized according to the preliminary simulation results.

Mode Purification

A single magnetron with a specific configuration can be operated at various single modes or mixed modes, depending on the characteristic values of applied magnetic fields and voltages. Pure-mode excitation after initial transients is a must in the phase-locking conditions because mode competition adversely influences the locking process. Due to the possible wave-wave interaction, it is rather obvious that phase-locking cannot be achieved if the magnetron microwave pulse contains two frequency components. For the A6 magnetron described earlier, the choice of $V_0 = 610$ kV and $B = 8$ kG ensures the π mode operation of the single magnetron.

Time-Delay Operation

In the dual-magnetron experiment, one magnetron is fired shortly after the other. In other words, two identical magnetrons are used, but the rise times of the applied voltages are different. The advantage of this operation is that the first magnetron to be turned on has enough time to establish its own mode of operation before interacting with the other magnetron. The second magnetron then turns on and receives a strong microwave pulse through the transmission line from the first magnetron. As a result, the frequency and phase of one high-power microwave source can be controlled by an injected microwave pulse from other sources through the transmission line. This is almost equivalent to the priming of the second magnetron. The length of the transmission line determines the phase difference between the two magnetrons. This works very well, especially when the transmission line is properly matched to its characteristic impedance at each terminal. In the simulation, rise times of 1 nsec and 10 nsec, respectively, were chosen for the two magnetrons. Power from the "master" magnetron with the shorter rise time selected the resonant mode and locked the phase of the "slave" magnetron via the transmission line. For a transmission line with length $L = 0.5$ wavelength and impedance $Z_0 = 0.1$ ohm, the particle distribution of two magnetrons at 12 nsec is shown in Figure 4. The oscillating RF

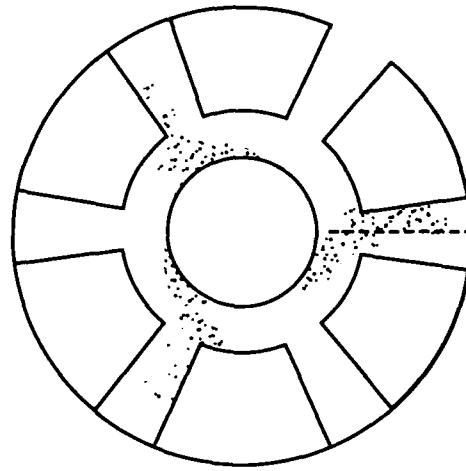
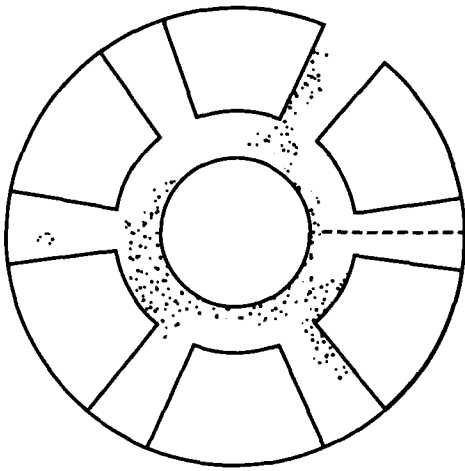


Figure 4. Magnetrons out of phase. Particle distribution at 12 nsec with transmission line length $L = 0.5$ wavelength and impedance $Z_0 = 0.1$ ohm.

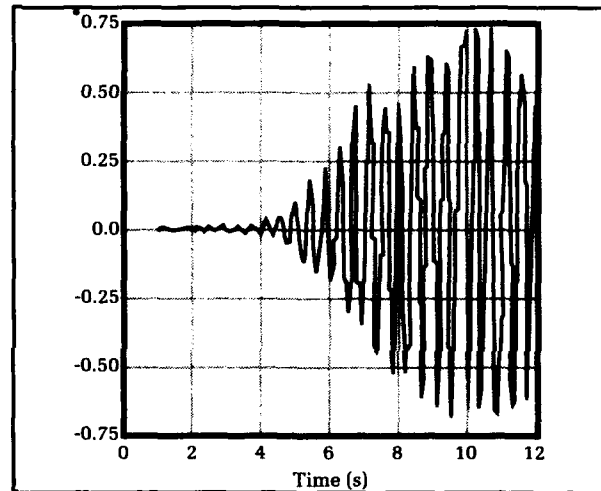
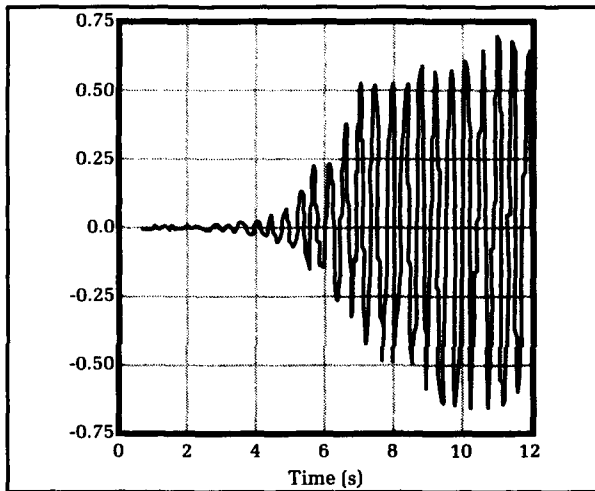


Figure 5. The oscillating RF fields of two magnetrons with transmission line length $L = 0.5$ wavelength and impedance $Z_0 = 0.1$ ohm.

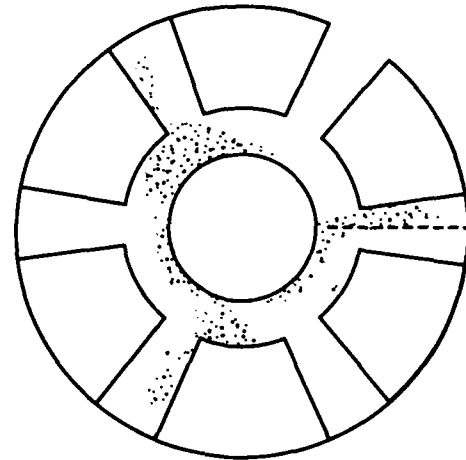
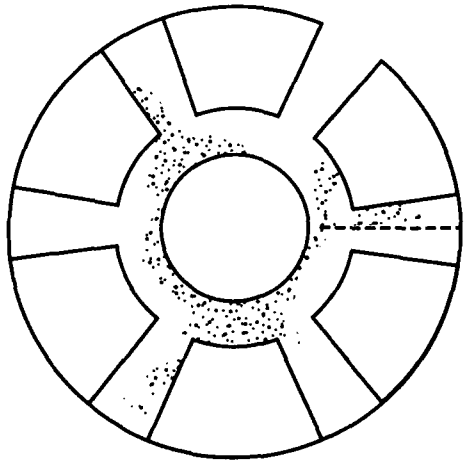


Figure 6. Magnetrons in phase. Particle distribution at 12 nsec with transmission line length $L = 1.0$ wavelength and impedance $Z_0 = 0.1$ ohm.

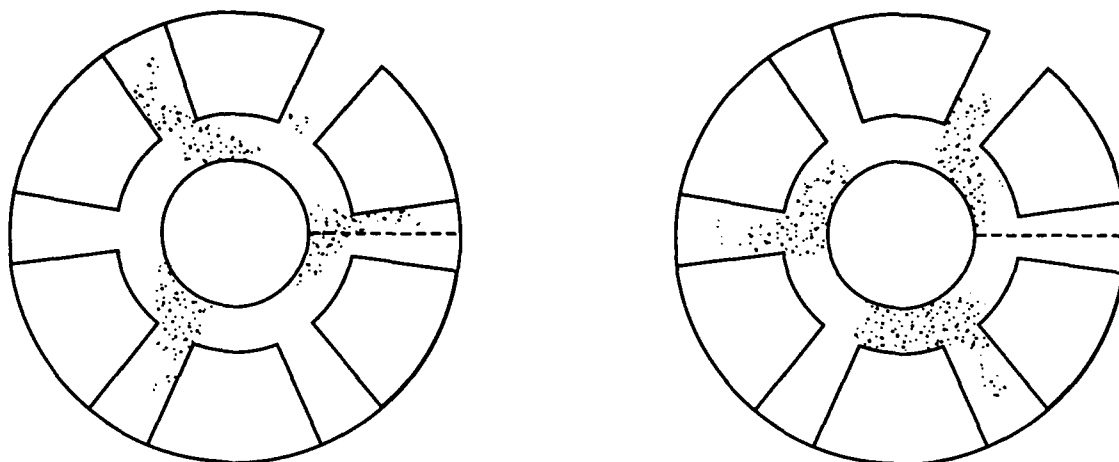


Figure 7. Magnetrons out of phase. Particle distribution at 12 nsec with transmission line length $L = 1.5$ wavelengths and impedance $Z_0 = 0.1$ ohm.

fields of two magnetrons, illustrated in Figure 5, show the π mode operation with a 180° phase difference. For further demonstration, if the length of the transmission line is increased to one wavelength, the results shown in Figure 6 indicate that the two magnetrons are in phase. Finally, if the length of the transmission line is increased to 1.5 wavelengths, the two magnetrons are out of phase again (Figure 7). It is feasible to extend the analysis to the peer-coupled case with an arbitrary number of sources. A large array of high-power, phase-locked magnetrons producing a high-power source is no longer a dream.

Conclusions

A large number of high-power microwave sources with fixed, single frequency and controllable phase has numerous potential applications. A novel approach using magnetron arrays to produce such unique microwave sources has been demonstrated numerically by means of particle simulation. Using the Physics International magnetron geometry configuration, we have demonstrated that 0.5, 1.0, and 1.5 wavelength connectors produce phase differences of 180 , 360 , and 180 degrees, respectively. Further investigation will include the influence of various vane loads, the length and impedance of the transmission line, and power extractions on phase-locking performances of two coupled magnetrons. This research has some utility in the design of future experiments relative to the phase-locking of many microwave devices. Phase-controllable, high-power multiple microwave sources hold potential value for target detection, target acquisition, target tracking, guidance, and related systems.

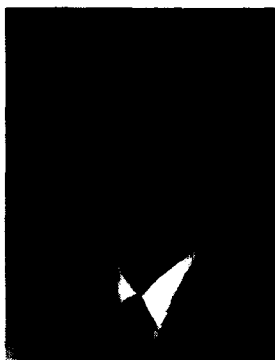
Acknowledgments

This research was supported by independent funds at the Dahlgren Division and SDIO/IST (Strategic Defense Initiative Organization, Innovative Science and Technology).

References

1. Levine, J., et al., "Strongly Coupled Relativistic Magnetrons for Phase-Locked Arrays," *Proceedings of SPIE, Microwave and Particle Beam Sources and Directed Energy Concept*, Los Angeles, CA, 16-20 Jan 1989.
2. Levine, J., Aiello, N., and Benford, J., "Design of a Compact Phase-Locked Module of Relativistic Magnetrons," *Proceedings of SPIE, Intense Microwave and Particle Beams*, Los Angeles, CA, 16-19 Jan 1990.
3. Stark, R., Chen, H., and Uhm, H., "Simulation Studies of the Relativistic Magnetron," *Proceedings of SPIE, Intense Microwave and Particle Beams*, Los Angeles, CA, 16-19 Jan 1990.

The Author



H. C. CHEN is a theoretical research physicist involved in the areas of high-power microwave generation and charged-particle beam propagation. He received a B.S. degree in electrical engineering from National Taiwan University in 1968 and M.S. and Ph.D. degrees in applied physics from the University of California, San Diego in 1971 and 1974, respectively. Before joining the Dahlgren Division, he was a research physicist at Science Application International Corporation and Systems and Applied Sciences Corporation. He has been a research associate at the Joint Institute for Laboratory Astrophysics, a postgraduate research scientist, and a research assistant at the University of California, San Diego.