

AD-A245 236

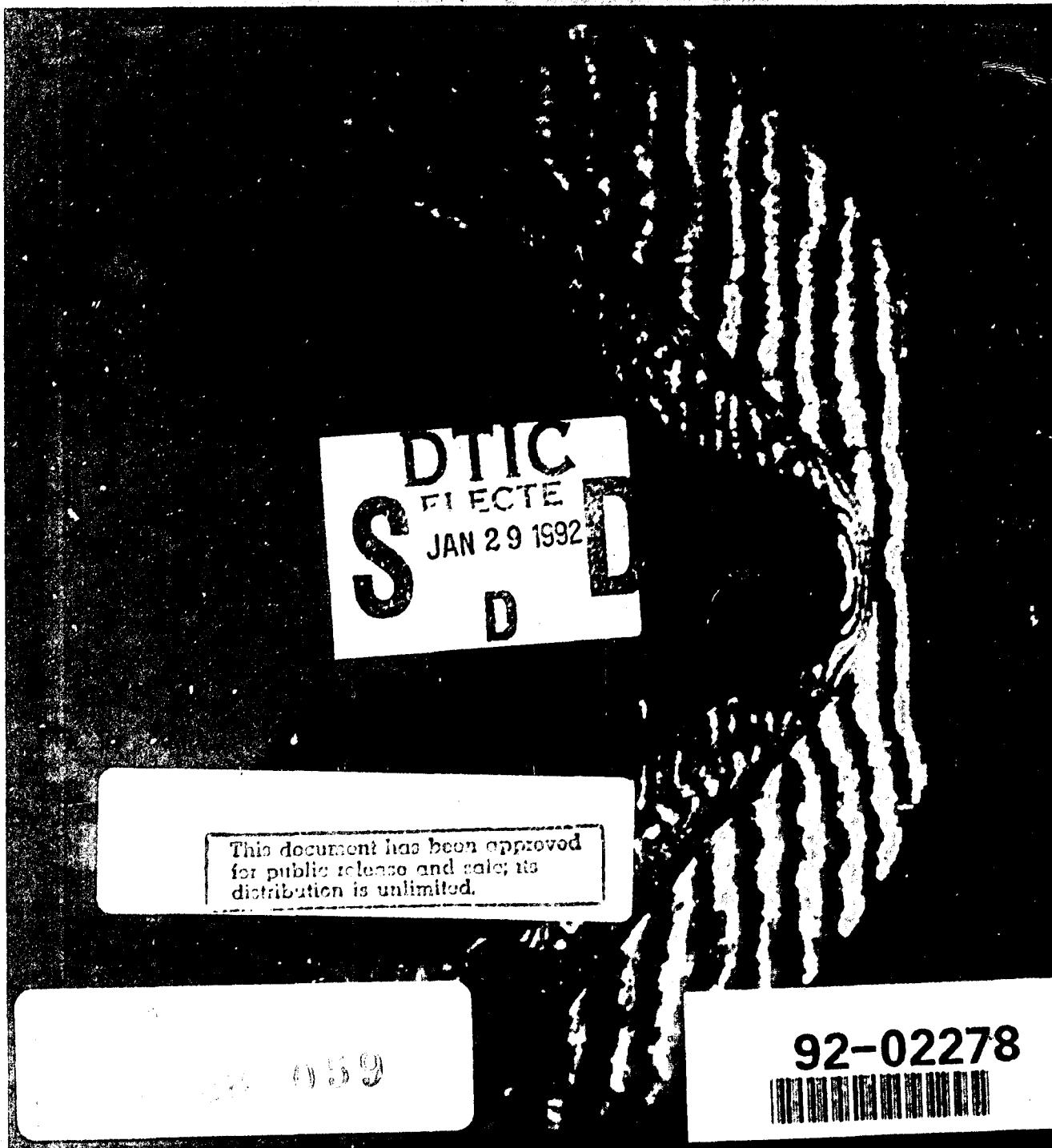


Naval Surface Warfare Center

6

Technical Digest

September 1991



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 Technical Digest presents unclassified articles, contributed primarily by Center scientists and engineers, on selected research and development programs. Under the leadership of the Space and Naval Warfare Systems Command, the Naval Surface Warfare Center is the Navy's principal research, development, test and evaluation center for surface ship combat systems, ordnance, mines, and strategic systems support. Please address any correspondence concerning the Technical Digest to: Naval Surface Warfare Center, Technical Digest (Code E221), 10901 New Hampshire Ave., Silver Spring, MD 20903-5000. Telephone (301) 394-4422 or (301) 394-4404.

About the cover: A computer-colorized interferogram shows the flow field around a model undergoing tests in NAVSWC's Hypersonic Wind Tunnel #2 at a Mach number of 5. The model is a simulation of the indented shape that may occur on a re-entry vehicle nose tip because of ablation due to atmospheric heating. The holographic interferometry technique used here gives aerodynamic engineers a non-intrusive means of measuring air densities in complicated flow fields.

Naval Surface Warfare Center

TECHNICAL DIGEST

September 1991

A Message from the Technical Director	T. A. Clare	3
Guest Editors' Introduction		4

Combat Systems

A Vision of Naval Surface Force Structure in 2030	V. A. Meyer	8
System Challenges of Technology Transition	T. C. Henderson	16
Objectives, Principles, and Attributes: A Structured Approach to Systems Engineering	D. K. Kreider and R. E. Nance	22
Advanced Distributed Processing Technology and ADMRALS	J. Blanton	32
A Chemical Warfare Naval Simulation Model for Surface Ships	T. J. Yencha	46
A Standardized Approach for Implementing Fiber Optics in Navy Surface Warfare Systems	D. R. Knudsen, G. D. Brown, and J. P. Ingold	54

Surface-Launched Weapon Systems

An Integrated Artificial Neural System for Target Identification	T. Holland, T. Tarr, and A. Farsaie	64
Target State Estimation and Prediction for Tactical Weapons Fire Control	W. D. Blair	72
Innovative Missile Airframe Concepts	L. Schindel and S. R. Hardy	84
The Use of Holographic Interferometry for Flow Diagnostics	W. C. Spring, III, W. J. Yanta, and K. U. Gross	92

Celsian-Based Ceramics for Advanced Radomes	I. Talmy, D. Haught, E. Wuchina, and J. Zaykoski	104
An Introduction to Directed Energy Technology	L. H. Luessen	112

Underwater Systems

An Electromagnetic Noise Collection and Processing System, "Arctic Research Buoy"	J. F. Scarzello, D. S. Lenko, and LCDR G. G. Durante	126
On the Validity of Normal Mode Theory for Modeling Undersea Acoustic Propagation	J. I. Arvelo, Jr.	134
A Neural Network for Passive Acoustic Discrimination between Surface and Submarine Targets	R. H. Baran	142

Space and Strategic Systems

Platform Attitude Determination by the Use of Global Positioning System	A. G. Evans, B. R. Hermann, and B. L. Miller	158
---	--	-----

A Message from the Technical Director

It gives me a great deal of pleasure to welcome all of our new readers to the premier issue of our Technical Digest. This and future issues will introduce you to the scientists and engineers at the Naval Surface Warfare Center, and to their many and varied technical contributions in support of the Navy.

The United States Navy has a long and proud tradition of service to our nation, dating from the founding of the Republic. Throughout its history, the Navy has both supported and utilized the applications of emerging technology to improve its fighting capabilities—to move forward from sail to steam to nuclear propulsion; from long glass and signal flags to sophisticated electronic surveillance and communications systems; from round shot and boarding pikes to missiles capable of reaching unseen targets at ranges of hundreds, even thousands, of miles.

Scientists and sailors, working together, made this happen. The history of the naval service and the history of science and engineering in the nineteenth and twentieth centuries are intertwined, as seen by the work of such outstanding individuals as Fulton-Colt-Dahlgren-Taylor-Michelson-Edison-Millikan-Goddard-Norden-Atanasoff—all of whom, along with many more, contributed their special talents to help meet the technical needs of the Navy.

We at the Naval Surface Warfare Center, as well as our colleagues at the other Navy research and development laboratories, are proud to be a continuing part of this tradition. This Center's advances in science, technology, and engineering will help assure that future naval forces will continue to be able to go in harm's way, whenever and wherever called upon. We are pleased to be able to share some of the results of our work with you in this forum.

As with any new venture, the introduction of the Technical Digest was neither simple nor straightforward. I would particularly like to thank the members of the Editorial Board for their many hours of unselfish dedication in helping to launch this publication, and our Editorial Staff in the Center's Technical Information Division for their production of the finished document. Both groups demonstrated true Navy "can-do" spirit in transforming the concept of this Center's technical publication into reality.



THOMAS A. CLARE

Accession For	
NTIS CR#&I DTIC TAB Unannounced Justification	
By _____ Distib. to /	
Availability Control	
Dist	Availabilty Special
A-1	

Guest Editor's Introduction

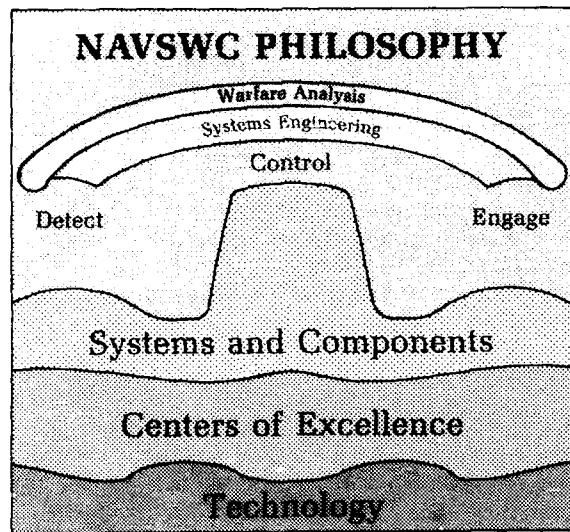
"Research and Technology—Shaping Future Naval Systems." For the initial issue of the Naval Surface Warfare Center's *Technical Digest*, our Board of Editors selected this rather broad theme as one reflective of the Center's mission and programs. Further, we sought a theme to encompass technical articles representative of our overall mission as the Navy's principal research, development, test and evaluation center for surface ship combat systems, ordnance, mines, and strategic systems support.

The future of naval warfare is entering a critical era which, on the one hand, offers hitherto unknown technological possibilities, and on the other, limits those possibilities within the scope of affordable costs. To make wise choices, we must strengthen the bridge between engineering and technology. Concurrently, we seek methods of reducing the time it takes to bring new developments to fruition. The question then becomes, How do we manage research and technology in order to shape those naval systems that will best protect the national interests in the years ahead? How do we assure that our naval forces will be able to perform future missions and defeat future threats—threats projected to differ substantially from those of the recent past in both nature and sophistication? Several of the articles in this issue address these concerns, some in rather general terms, and others from the standpoint of specific developing technologies or research and development programs.

The Center's philosophy for developing complex surface warfare systems is founded on systems engineering principles. Only a comprehensive, disciplined engineering approach to development of combat systems can achieve the affordable warfighting capability necessary for the future. The accompanying figure illustrates the basic functional elements of the combat system, namely: "Detect" the target, "Control" the weapon, and "Engage" the target. We believe the Center must continue to maintain full-spectrum capability in these critical functions. A sound technology base provides not only the foundation, but also a wealth of opportunities to meet the needs of the Navy well into the twenty-first century. The vision to anticipate these future needs relies upon strong "Warfare Analysis" and "Systems Engineering." Accordingly, we begin the *Technical Digest* with articles describing the Center's vision of the force structure for the 2030 time frame and some of the research and technology that may well be "shaping future naval systems." The articles represent a wide variety of subject matter, presented in four major categories: Combat Systems, Surface-Warfare Systems, Underwater Systems, and Space and Strategic Systems. In future issues, we will tend to focus on more specialized subject areas. The next issue, for example, will concentrate on Detection Systems, the first element in our combat system model.

Combat Systems

The first two articles in the Combat Systems grouping discuss the evolution of naval surface warfare in a time of significant change. Our naval force structure is likely to be altered dramatically by fundamental shifts in the geopolitical arena, advances in technology, and stringent cuts in the defense budget. Meyer summarizes the impact of these changes and projects long-range changes in force structure, surface-ship design, and warfighting capabilities. These, in turn, suggest many technology needs, including reduced signature, improved sensor performance, and revolutionary new weapon systems. Henderson approaches evolutionary change



from the perspective of the AEGIS combat systems baseline. He discusses a systematic approach to performance upgrades for the 2030 time frame, giving several examples of technology innovation. Among them is an open combat systems architecture—a distributed, bus-connected processing architecture that allows insertion of new technology more readily.

The complexity of modern systems, coupled with accelerating technological advances, requires a disciplined systems engineering approach to the development of naval systems. Kreider and Nance postulate a concept for engineering large-scale systems, based on an "objectives, principles, attributes" structure. This methodology is an extension of a previous effort to apply the same structure to software engineering.

The next two articles provide good examples of the utility of simulation and modeling, first at the force level and then for a specific ship-level problem. Blanton addresses complex, multiwarfare analysis problems involving fleet, point and area defense, outer-air battle, space-based surveillance and target assets, and some limited antisubmarine and electronic warfare. He describes a model based on Advanced Distributed Processing Technology originated at NAVSWC several years ago, which has since matured and been applied to several large-scale warfare analysis problems. Yencha demonstrates the utility of a simulation model for estimating the degradation of a ship's effectiveness when exposed to a chemical attack. Most important, the simulation provides a useful tool to design future ships for survivability against chemical attacks which, based

on Operation Desert Storm experience, could pose a serious threat. In the last article of this section, Knudsen, Brown, and Ingold describe an ongoing program for implementing fiber-optic data transfer networks in surface Navy ships.

Surface-Launched Weapon Systems

Surface-launched weapon systems such as Tomahawk and the 16-inch naval gun were used quite successfully in the recent Persian Gulf conflict. Considering the success enjoyed by the U.S. and its coalition allies during the Gulf War, one might well ask, Why be concerned about incorporating new technology into the weapon systems? However, as naval targets become faster, more stealthy, more coordinated and maneuverable, surface-launched weapon systems must likewise keep pace in order to maintain superiority. Six Digest articles highlight promising technology for tomorrow's fleet. In the first, Holland, Tarr, and Farsaie describe how Artificial Neural Nets (ANN) can be used to determine the type of target a remote infrared (IR) sensor sees, based on actual IR data. The ANN was trained on targets typical of those encountered by the Marines, but has potential value for other applications as well.

Blair demonstrates how a relatively new type of filter, Interactive-Multiple-Model, can improve the target-state estimate of a high-speed target over existing filters. Of course, the better one can predict the target state, the better the fire control solution and the more likely a successful engagement by a friendly weapon or decoy.

The next two articles are oriented toward innovation in the airframe area. Schindel and Hardy address improvements in airframe design for engaging either a very long-range target or a high-speed, low-altitude cruise missile. In the second article, Spring, Yanta, and Gross describe an innovative way to determine flow-field information around a high-speed weapon configuration by means of holographic interferometry (HI). This diagnostic tool, readily adaptable to many wind tunnels, is a nonintrusive (outside the tunnel test section) technique based on changes in light waves as they pass through regions of the flow field. NAVSWC maintains a full range of wind tunnels for aerodynamic testing of weapons. Holographic interferometry is one of the alternatives pioneered at NAVSWC for determining flow-field information.

Talmy, Haught, Wuchina, and Zaykoski, in an article on celsian-based ceramics for radomes, address the materials problems encountered with all-weather operation of high-speed, termi-

nally guided missiles. The goal here is to find a material that will allow the missile to fly at Mach numbers of 7 or higher without resorting to radome covers or coatings to withstand operation in a rain-erosive environment.

In the final article in the Surface-Launched Weapons section, Luessen describes the potential value to the Navy of directed-energy technology, particularly in self-defense of surface ships.

Underwater Systems

The Underwater Systems program at NAVSWC encompasses antisubmarine warfare (ASW) and antisurface warfare (ASUW), involving numerous weapon concepts and missions. Examples include warhead concepts for torpedoes and mines, design and development of new underwater weapon systems, signal processing, and related analyses of new weapon concepts for underwater ocean environments. Environments of concern vary from the harsh and unfamiliar world of the arctic and antarctic, to the shallow water envisioned for many third world engagement scenarios. Articles in the Underwater Systems section address the environment, the potential benefits from improvements in shallow-water acoustic modeling, and new techniques in signal processing, such as adaptive neural nets.

The first article focuses on electromagnetic noise associated with the arctic aurora, in particular, the potential for this noise to disturb many types of electrical systems, including communication and weapon systems. Scarzello, Lenko, and Durante describe an Arctic Research Buoy that collects electromagnetic noise data in the polar and aurora zones, and then transmits the data to the continental United States via satellite for use in designing naval systems.

The recent focus on regional conflicts emphasizes the need for improvements in shallow-water anti-submarine warfare. Of particular interest is the active detection of underwater targets in coastal regions with adverse acoustic environments. Arvelo discusses the development of a normal mode sound propagation model which considers sound-speed variation, sediment and sub-bottom elasticity, surface and bottom roughness, and range dependence. Good agreement between the predictions and test measurements is presented to validate the model.

Underwater mines have historically played an important role in undersea warfare. Being entirely autonomous, they generally rely on passive sensors for target detection. The advent of embedded computers and low-power processors has enabled mines to process more data with the potential of classifying targets to improve effectiveness and countermeasure resistance. Baran's article presents a neural network that discriminates between surface and submarine targets—a first step toward solving the classification problem.

Space and Strategic Systems

The Center has a long history of supporting the fleet in space and strategic systems and related technologies, particularly in fire-control software and re-entry body technology. Future issues of the Digest will address this important subject area. We include in this issue one article on space systems, in which Evans, Hermann, and Miller discuss work of the past eight years, demonstrating that the Global Positioning System can provide users with real-time platform attitude information that significantly enhances the operation of the Inertial Navigation System, especially in the azimuth (heading) component.

The Guest Editors



HARRY E. CRISP has been employed at the Center since 1971. He holds a B.S. degree in electrical engineering from Clemson University and an M.S. and Ph.D. in electrical engineering from Auburn University. His graduate research was in automatic control theory. During his first few years at the Center, he did technology work in the application of digital controllers to weapon control systems. Subsequently, he led

the Modular Fire Control Program and other technology projects for the Center. He performed a temporary assignment at the Naval Sea Systems Command as project engineer for the Mk 86 Fire Control Program. After managing the Mk 86 research and development program within the Center, he was named head of the Weapons Control Technology Branch. In addition, he has served as head of the Information and Control Technologies Branch and was Center coordinator for the Independent Exploratory Development program from May, 1985 through March, 1988. Currently, he is head of the Center's Technology Base Program Office.



FRANKIE G. MOORE, Principal Aerodynamicist in NAVSWC's Weapons System Department, holds B.S., M.S. and Ph.D. degrees in aerospace engineering from Virginia Polytechnic Institute and State University. Beginning at NAVSWC as a Co-op Student from 1963 to 1967, he subsequently performed research into new projectile design concepts; developed approximate aerodynamic prediction codes, and served

as aerodynamics and structures technology principal. He has headed the Center's Mathematical Analysis and Aeromechanics Branches, and, until recently, was the NAVSWC Independent Exploratory Development coordinator. Current assignments include computational aerodynamics tasks and Weapons Systems Department focus for internal technology base resources.



WILLIAM T. MESSICK, a senior program manager in the Center's Strategic Systems Department, received a B.S. degree in mechanical engineering from Drexel University in 1968, and M.S. and Ph.D. degrees in mechanical engineering from the University of Maryland. At NAVSWC since 1968, his experience includes research in static and dynamic response of structures and materials and development

of high-temperature structural ceramic antenna windows. He has managed the Surface-Launched Missile Materials Technology Program, served as acting head of the Research and Technology Department's Materials Division and, since 1985, has been manager of the Weapons and Spacecraft Materials Technology Block Program. Dr. Messick has been chairman of the Joint Directors of Laboratories (JDL) Electromagnetic Windows Working Group. He has served on the National Materials Advisory Board Study of High-Temperature Synthetic Fibers; the Tri-Service Materials Development Coordinating Committee, and the Aerospace Industries Association Composites Materials Roadmap on Ceramics. He is an Associate Fellow of the American Institute of Aeronautics and Astronautics.



CHARLES F. McCLURE was born in Madison, Wisconsin in 1940. He received a B.A. in physics from the The Johns Hopkins University in 1962 and a Ph.D. in theoretical physics from the University of Maryland in 1971. Dr. McClure joined what was then the Naval Ordnance Laboratory as a summer student in 1961 and for many years was involved with the test and evaluation of major weapon systems such as

SUBROC, the Mk 48 torpedo, the CAPTOR mine, and other mine systems. Currently Dr. McClure is a group head in the Systems Analysis Branch of the Underwater Department and is working on problems related to the effectiveness and lethality of underwater warheads against Navy targets.

A Vision of Naval Surface Force Structure in 2030

Victor A. Meyer

This Naval Surface Warfare Vision is a strategy-framed, top-down description of surface warfare, circa 2030. Surface warfare is naval warfare conducted from surface platforms against surface, subsurface, air, and space forces. It also includes organic off-board assets and support from nonorganic assets and joint/allied forces.

The focus of this 2030 Vision is on warfighting. A surface warfare concept of operations is postulated and evaluated in several different naval campaign models. The campaign models are varied, depending on future world geopolitical trends. A naval force structure is postulated and analyzed in terms of the surface ships and the associated combat systems required to implement the concept of operations. Ship and combat system concepts are generated for notional ship types needed to fulfill the postulated force structure. This rebalancing of force structure would provide a more flexible and adaptable surface Navy needed to defend the national interests of the U.S. across the full spectrum of conflict from naval presence and deterrence to low-intensity conflict or global war.

Introduction

The United States Navy is destined to undergo significant change in the coming decade. Driven by fundamental changes in the geopolitical arena, advances in technology, and stringent cuts in the defense budget, our naval force structure is likely to be altered substantially as we plan for the 21st century. It is vital that we understand these changes and their implications so that we can manage this transition successfully.

A survey of geopolitical events in 1991 finds extraordinary change. The dramatic decline of the Soviet Union and the dissolution of the Warsaw Pact have significantly reduced the threat of a war in central Europe. Super-power polarization no longer dominates, having been pushed aside by the pragmatism of economic survival. The twin engines of social and economic change are overwhelming the old order in scenes marked by turbulence and instability. Nationalism founded in ethnic, racial, and religious mores is reshaping governments on both sides of the old Iron Curtain. Ideological battles have given way to demands for individual freedoms and improved living standards. The struggle to control access to the world's resources on the one hand, and the need to protect the world's environment on the other, will be played out on center stage in the years ahead.

The threat to U.S. naval forces in the future will no longer be limited to the latest Soviet missile or submarine. While we cannot abandon completely a strategy that has held in check for 45 years an often adventurous and sometimes hostile Soviet Union, we must balance that strategy with naval forces designed to counter the more likely occurrence of regional and low-intensity conflict. No less violent in nature than larger scale engagements, regional and low-intensity conflicts present unique warfare requirements that arise from the diversity of the forces involved, the rules of engagement that govern such conflict, and the usually common characteristics of being in close proximity to a land mass. These conditions will demand that future naval forces be more flexible and adaptable to mission requirements and a diverse threat spectrum, while at the same time operating in a more limited

battle space. This will ultimately affect the type of naval forces needed to perform warfare tasks.

The technology outlook and its potential for naval applications hold great promise in the years ahead, tempered by the high price of technology development and the long-lead time needed to put advanced systems into operational use. High technology systems, such as directed energy, signature reduction, advanced electronics/photonics, and space systems, need to be pursued in a highly interactive manner, however, if technology is to be used effectively and affordably in naval applications. Strategy-framed concepts of operations must be defined to serve as constructs upon which to evolve innovative, advanced systems concepts. By encouraging a healthy balance between the technology "push" and the warfare requirements "pull" process, we can develop advanced naval systems that will meet the threat within the constraints of cost.

Because of fundamental problems in the U.S. economy related to large trade and budget deficits, record personal and corporate indebtedness, and interest on the three-trillion-dollar national debt, it is very likely that the defense budget will decrease substantially in the foreseeable future. This is further compounded by political expectations of a "peace dividend" as we see our traditional foe enmeshed in sweeping economic and political reform. Additionally, manpower for the all-volunteer force will be more scarce and costly, at least through the mid-90s, when the demographic curve reverses the current decline of military-age recruits.

Logically, the Navy and Marine Corps might expect a less substantial cut than the other services because of the forward-deployed, flexible nature of their forces. Peacetime politics are likely to prevail, however, and the sea services may expect to be cut by their "fair share."

Force Structure Trends

The implications of these events in terms of a future naval force structure can be analyzed¹ to provide some interesting insights.

- In the anticipated future "unstable peace" we are likely to need naval forces in more places, but have fewer forces to respond. New balances will have to be struck to match peacetime deployments with shifting national interests, and ultimately to reshape force structure to accommodate more diverse tasking.
- Better global intelligence and multispectral surveillance (e.g., electromagnetic, acoustic, optical) will be needed so that naval forces can be alerted early and updated frequently

as they respond to time-sensitive situations. Maldeployments must be minimized.

- Future naval force structure must be sufficiently flexible and adaptable to operate across the complete spectrum of conflict from peacetime presence and crisis response to limited, regional, or global conventional war.
- Interoperable maritime forces that have common interfaces with joint service and allied forces are true force multipliers. The operational value and political benefits will be worth the effort to develop force interoperability.
- Increased communications, data fusion, and information management will be available to manage a multi-sensor, all-source data base distributed across the geographically expanded battle space.
- Increased instances of low-intensity conflict will require that naval forces be available early to keep the conflict from escalating and to maintain a sustained effort. These forces should be compatible with friendly, indigenous naval forces to help train them and to foster cooperation.
- There will be an increased requirement to perform mine countermeasures, shallow-water antisubmarine warfare and coastal warfare, both as a part of low-intensity conflict operations and for defense of our own coasts against submarine-launched missiles, covert operations, and drug smuggling.
- Changing demographics, educational skills, and system automation will require improvements in training, personnel assignments and rotations, and possible adjustments in the active/reserve force mix.
- Reduced target detectability and response time will shrink the battle space and place emphasis on countermeasures and point-defense systems against both missiles and torpedoes; this will lead to emphasis on prevention of targeting via stealth and deception.
- High technology systems will become increasingly costly to develop and maintain over their life cycle. Great care will be needed to introduce new technology where it will have maximum leverage.
- Improved battle damage assessment will be needed against over-the-horizon targets to avoid wasting a limited inventory of high technology weapons.

Roles and Missions

Having noted some of the more significant trends that will affect our Navy in the future, it is important that we examine the roles and

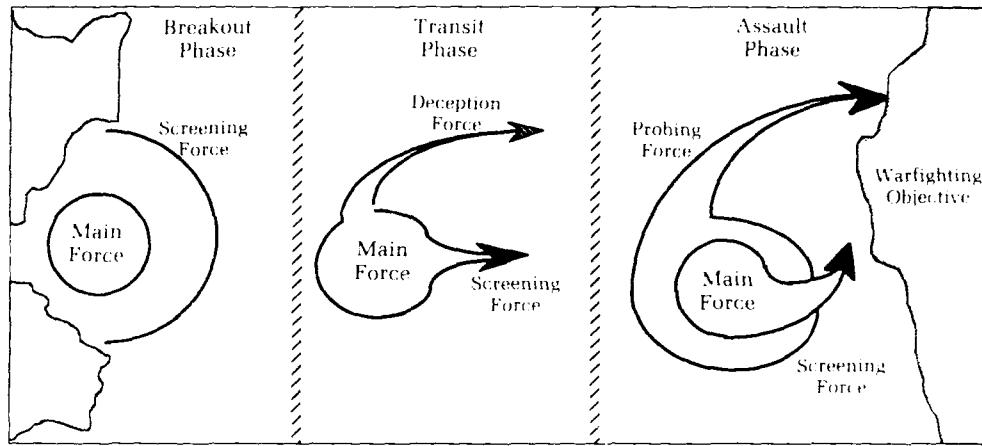


Figure 1. Surface force concept of operations.

missions the Navy will be required to perform in the years ahead. The Navy's mission "to conduct prompt and sustained combat operations at sea in support of U.S. national interests" will remain clear. So will the fundamental elements of our maritime strategy — a forward, global, and flexible force which, in conjunction with our allies, can deter or defeat any threat to our national security. This policy will continue to be functionally implemented by sea control and power projection forces that include the elements of strategic deterrence and strategic sealift.

It is at this point, however, that choices concerning the future naval force structure must be made. What balance between the various force structure elements is needed to carry out the Navy's mission and strategy in the next decade and into the 21st century? How does one provide a flexible, adaptable force in the numbers needed to deter any potential Soviet threat, yet still operate effectively and efficiently in low-intensity conflict, all within the limits of a severely reduced naval budget? How should the future force be employed and deployed? The answer to the last question can be instructive in formulating the answers to the first two.

Concept of Operations

If we postulate a future naval operating force as having two elements, a main force and a screening force, as depicted in the concept of operations shown in Figure 1, we can analyze how they interact during a typical deployment scenario.² During the breakout phase, the screening force may perform mine countermeasures and shallow-water ASW for the main

force that will provide the mobility, sustainability, and command and control functions of the force. In the open-ocean transit phase, the main force provides deep-water antisubmarine warfare (ASW) and wide-area antiair warfare (AAW) protection to the total force, while the screening force may be used to augment the ASW screen or act as a deception force. Near the warfighting objectives, the screening force may be used to probe and stretch enemy defenses in an attempt to discover weaknesses, deceive and imbalance defenses, and cause maldeployments that will expose critical enemy vulnerabilities to strikes by the main force. Using the maneuver warfare attributes of surprise, speed, boldness, and initiative to create a high tempo of operations, enemy weak points are rapidly and repeatedly struck to achieve the maximum physical as well as psychological effect. The aim is to overwhelm the enemy and destroy his morale, hence, his willingness to resist. The warfare functions of this concept of operations may be assigned to the main force and the screening force as follows:

- | Main Force | Screening Force |
|--|--|
| <ul style="list-style-type: none"> ● Force C³I & Surveillance ● Air Superiority ● Strike Warfare ● Force Defense ● Amphibious Warfare ● Force Logistics | <ul style="list-style-type: none"> ● Local Area Reconnaissance ● Screening and Probing ● Deception, Diversion, Countermeasures ● Dispersed AAW, ASW, ASUW ● Mine Laying and Countermeasures ● Special Operations ● Combat Search and Rescue |

To derive a future force structure requires assigning these force functions to ship platforms, an exercise that can begin with the existing force structure and evolve into future concepts. Within the main force, we can assign the role of force surveillance, air superiority, strike warfare, and force defense to the aircraft carrier and its embarked air wing. The need for high performance, manned aircraft to carry out these missions will very likely continue well into the next century with a CVN-type ship, probably improved in signature reduction and survivability, as the combat-proven platform. Assisting in the strike warfare and force defense tasks is the battle force combatant, a multimission, aviation-capable ship that complements the carrier by providing dispersed strike and defense-in-depth capabilities to the main force. A conceptual view of the future battle force combatant is shown in Figure 2. Its salient features are smooth topsides, a multifunctional conformal antenna array, an open-ocean ASW suite, an integrated hard-kill/soft-kill self-defense system, and increased ship survivability features.

A new ship type assigned to perform the main force functions of amphibious warfare and force logistics, and to assist in force defense and strike warfare, is the sea control ship. The sea control ship can be configured alternatively as an amphibious assault aviation ship, a sea control ship for antisubmarine warfare, a command and fire support ship, or, with the well deck replaced by bunkers, a combat logistics ship. As depicted in Figure 3, the sea control ship is characterized by an angled jump flight deck, an amphibious well deck, and a multimission combat suite. Its primary characteristic is the capability to support STOVL (short takeoff, vertical landing) aircraft for amphibious operations, ASW sea control, or combat logistics support. It can also carry a wide ranging payload of smaller scout-fighter vehicles, such as mine countermeasures craft, coastal patrol boats, helicopters, or mini-submarines that can operate from either the flight deck or the well deck. In addition, different weapon loads can be accommodated by varying the loadout of the vertical launch system magazine.

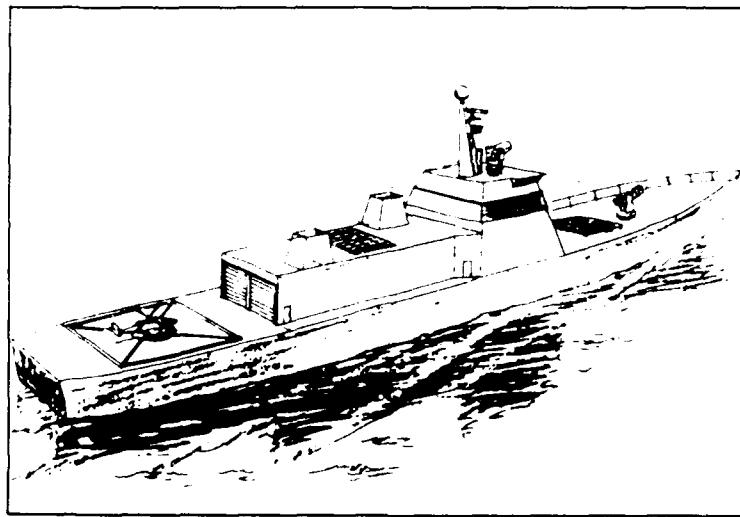


Figure 2. Battle force combatant.

- Multimission battle force combatant, 10,000 T
 - Dispersed Strike
 - Force AAW, ASW, ASUW
 - Independent Ops, Special Ops
- Aviation capable
 - OTH surveillance, classification, localization, attack
 - Helo, UAV, AIC
 - Air Intercept Control

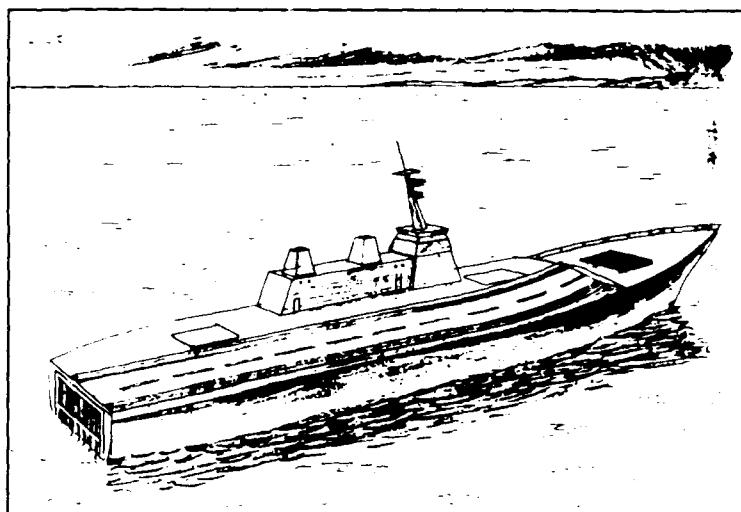


Figure 3. Sea control ship.

- Functional replacement for LHD, LHA, BB, LCC, AOE, CVS; 40,000 T
- Modular payload, ski-jump flight deck, well deck
 - STOVL, helos, UAVs
 - LCAC, SES, MCM, UUV, Deception vehicles
- Multimission
 - AMW, ASW, STK, C&C, LOG

The operational concept for the screening force requires ship types designed for single-purpose missions in areas of high risk and variable environmental conditions. This difficult problem has traditionally been solved by using small, special purpose ships and craft. Minesweepers, coastal patrol craft, and landing craft are examples. The limitation on these small craft has always been availability when and where needed. Their short-legged endurance and inherent seakeeping limitations have usually limited their use by the U.S. in overseas operations.

The mother ship concept solves this problem and expands the types of vehicles which can be carried. It embarks the scout-fighters in a well deck or hangar where they are safely stowed

12 until near the theater of operations. During open ocean transits, the mother ship remains under the defense of the main force. While transiting, it may use its payload as a deception force or to assist in antisubmarine warfare.

It is in the operational area, however, that the mother ship and its scout-fighters really begin

to function. There, the scout-fighters are deployed to expand the screen to gather intelligence, probe for gaps and vulnerabilities in enemy defenses, disrupt his timelines, and deceive him into making maldeployments. This is maneuver warfare doctrine designed to outpace the enemy, stretch out and imbalance his forces, thereby creating opportunities for decisive strikes by the main force. Operationally, scout-fighters may also be used to lay mines, insert and extract special warfare forces, and perform shallow-water antisubmarine warfare. The mother ship is limited only by the type of scout-fighter vehicle available as the modular payload. Figure 4 shows a conceptual rendering of the mother ship.

The scout-fighter vehicles are key to the screening force concept. They are single-mission platforms which rely on their small size, stealthy features, and flexibility in operations (e.g., shallow draft, high speed, maneuverability), both to avoid contact with the enemy and to countermeasure his weapons.

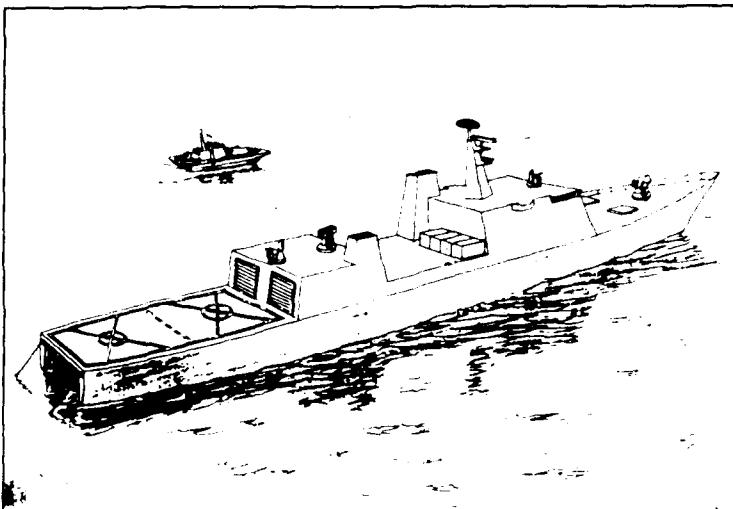


Figure 4. Conceptual mother ship.

- Functional replacement for LPD, LSD, LST, LKA, AS, AD, AR; 20,000 T
- Modular payload for flight deck, well deck
 - Manned or autonomous air, surface, sub-surface vehicles: 5-500 T
 - Vans, plug-in modules/PCs, RO-RO
- Tailored mission
 - Force deception, MCM, shallow-water ASW, NSW, NFS, maintenance and repair

They are heavily oriented toward offensive missions and surveillance, having only a minimum self-defense weapon suite. They are constructed as simply and inexpensively as possible so as to be affordable in the numbers needed to be effective and to minimize the political risk should one be lost.

A scout-fighter may take many forms, as illustrated in Figure 5—a helicopter, surface craft, mini-submarine, or autonomous vehicle. It can be as large as an ASW corvette, in which case it would operate from alongside the mother ship and under its control, or as small as a remotely piloted vehicle. It should be modular enough to operate from the mother ship, but not so constrained by its own support systems that it cannot be replaced *in toto* as new technology models become available. Ideally, it will have the flexibility of a helicopter and the maintenance requirements of a patrol boat. The mother ship-scout fighter concept also offers a way to transition to autonomous vehicles in future air, surface, and subsurface applications.

To be practical, the future force structure outlined above must be affordable and it must be realizable from today's Navy. It must also provide a balanced and effective deployment strategy for peacetime and wartime contingencies. The deployment scheme shown in Figure 6 satisfies these requirements and can be achieved with 10 to 12 carrier battle groups and 6 to 8 sea control groups. This scheme supports a global, three-ocean strategy with four three-carrier battle forces available in the Atlantic/Mediterranean, the Pacific/Indian Ocean, and the option of deploying a carrier or sea control group to the southern hemisphere. In peacetime, a normal one-in-three deployment schedule can be maintained in these areas to provide naval presence and crisis response, while allowing for meaningful training in the operational theater. The sea control ships could be deployed to areas under U.S. or allied air control such as the Greenland-Iceland-Norway gap, off Japan and the Philippine Islands, and in the Mediterranean, to train in antisubmarine operations. They could also be used in certain

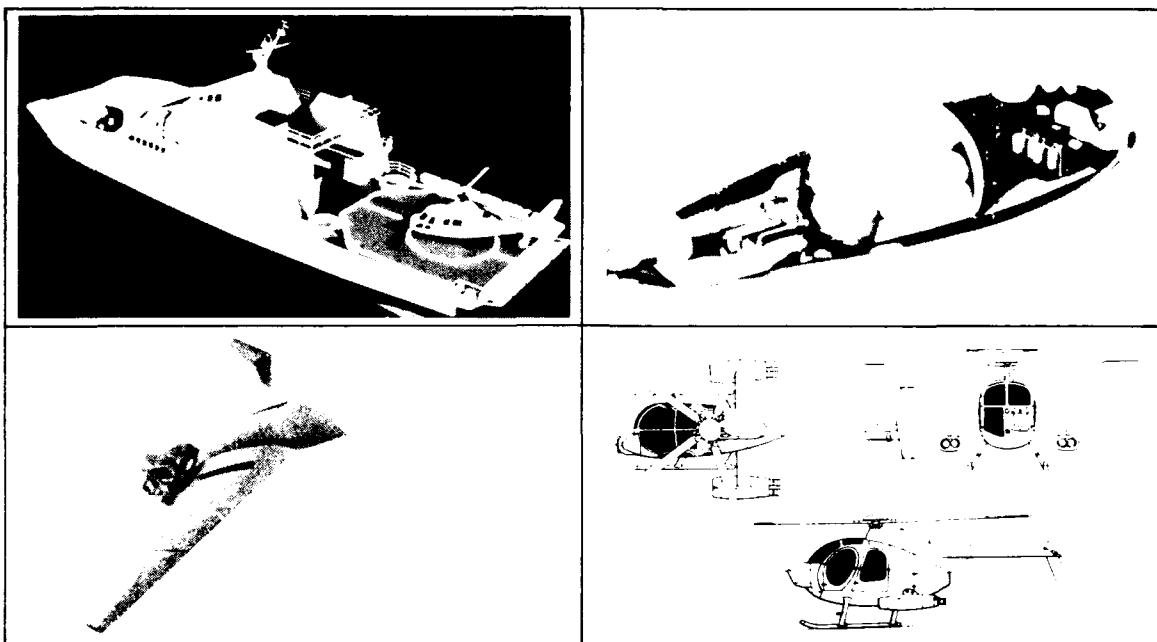
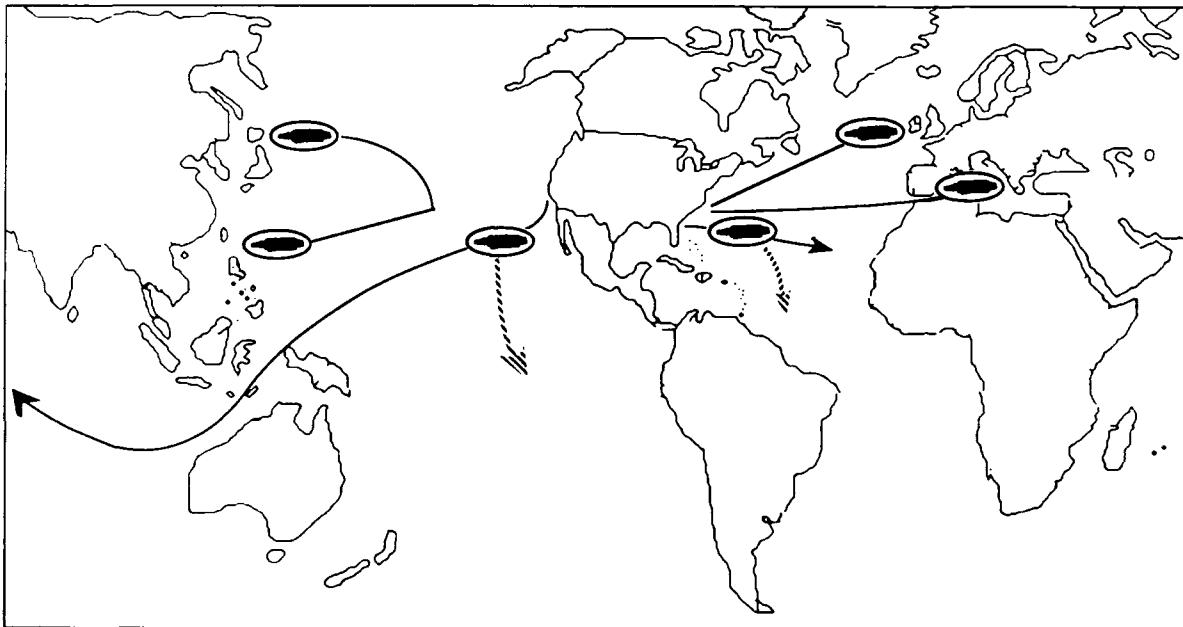


Figure 5. Examples of scout-fighters. Clockwise from the top left: surface craft, mini-submarine, helicopter, and autonomous air vehicle.

14



- CVN and BFC II for N. Atlantic, NW Pacific, Indian, Med, etc.
- Sea control ship for N. Atlantic, Kuriles, Med, and West Pac
- Mother ship for PG, Caribbean, Sea of Japan, Aegean, etc.

Figure 6. Peacetime/wartime deployments.

regional and low-intensity conflict roles currently filled by aircraft carriers and battleships which will likely be retired as we transition to a smaller Navy.

As one addresses the affordability of these changes in force structure, sea control ships are envisioned as modern, flexible ships that would be manned with much smaller crews and require fewer escort ships. They could be produced at about half the cost of an aircraft carrier; therefore, they represent a net reduction in both outlay and operating expenditures compared to the ships they would replace.

Likewise, if each of the 10 to 12 carrier battle groups were complemented by a mother ship and its embarked scout-fighters, a fleet-wide screening capability would be available at less cost and greater flexibility than an equivalent number of surface combatants. This would bring the screening force capability to battle force operations where the tactics can be developed to make this concept fully effective.

Summary

In summary, we are witnessing profound change in the world on virtually every front. In this uncertain time, political instability could easily spill over into armed conflict. The U.S. Navy will be relied on increasingly as an instrument of foreign policy to carry out its peacetime presence mission and to help control crisis situations. This role demands flexible and adaptable naval forces, available worldwide on short notice, to join with our allies in preserving the peace.

National priorities and fiscal realities will force reduction in the U.S. defense budget in the years ahead. We should therefore make a conscious effort to invest a portion of our

acquisition dollars in a naval force structure that will specifically address the lower end of the spectrum of conflict. Low-intensity conflict, support of friendly Third World governments, drug interdiction, evacuation of nationals, and similar events can be effectively and efficiently handled by ships such as the sea-control ship and the mother ship-scout fighter combination. These ships offer great flexibility at lower risk and cost and would restore needed balance to our naval force structure.

References

1. Dept. of the Navy, *Sea Control 2020 Vision*, Space and Warfare Systems Command, Washington, DC, 1990, p.10.
2. Meyer, V. A., *Surface Warfare Vision 2030*, NAVSWC TR 90-619, Dec 1990.

The Author



VICTOR A. MEYER is a senior systems engineer and naval analyst with the Surface Warfare Analysis Office. A 1962 graduate of the U.S. Naval Academy, he holds a masters degree in mechanical engineering from the University of Notre Dame and graduated with honors from the Naval War College in 1988. At NAVSWC, he has managed numerous combat system and ship programs ranging from battleships to fast patrol boats. He also has had underwater experience, having spent nine years in the submarine service, and he currently commands the Naval Reserve Undersea Warfare Unit (OP-02 Det 106) in the Pentagon. Mr. Meyer leads the Surface Warfare Vision 2030 project at NAVSWC.

System Challenges of Technology Transition

Thurman C. Henderson

Keeping fleet warfighting capabilities ahead of threat technology in a rapidly changing world is a significant challenge. For this reason, the Center is involved in several long-range planning efforts to determine needs and formulate concepts for future naval surface combatant combat systems. The Center is exploring long-range technology and weapon trends, critical technology clusters, and system development approaches which can position the Navy to meet the challenges of the future. One such initiative is the AEGIS Combat System Round Table. This effort is using AEGIS in-house and industry team expertise, systems engineering approach, and lessons learned to project needs and formulate system concepts for the future.

Introduction

Rapid technology advances, coupled with the proliferation of modern weapon systems throughout developing and third world countries, will greatly impact the nature and level of capabilities required in future naval systems. Stealth, stressful target kinematics, extremes in attack profiles, sophisticated multimode guidance, and "smart" countermeasures are but a few of the weapon features which technology advances are already fostering and will continue to accelerate. A better understanding of technologies and their potential impacts on future systems and military operations is required to position the Navy to meet its future needs. The AEGIS Combat System Round Table effort enables us to better understand and project those needs and the required characteristics of future naval surface combatants and forces.

The David Taylor Research Center (DTRC) initiated an effort almost three years ago to identify critical hull, mechanical and electrical (HM&E) technology clusters, advanced warship concepts, and future force operational concepts.¹ The AEGIS Combat System Round Table, which was initiated by the Technical Director, AEGIS Shipbuilding Program, in the spring of 1990, is patterned after the DTRC process. The Round Table draws from the experience of the AEGIS in-house and industry engineering team, supported by participants from the research and development community, to identify weapon technology trends, critical technology clusters, and combat system concepts to complement the DTRC studies and establish a direction for future Navy surface combatants. An increasingly far-term perspective is being forced upon those who are involved in weapon system research and development activity because of the very rapid rate and breadth of technology advances and the long development lead time and operational life span of major naval systems.

The Basic Problem

The basic problem is twofold. First, the challenge is to look into an uncertain future, project the types and characteristics of forces which may be required, and then to identify the actions that must be taken to position the Navy to meet the envisioned future needs. Some progress has been made in this area already by the DTRC and AEGIS Round Table, and further effort

is planned. A second aspect, which is a lingering problem in all military systems, is to find better ways to keep forces combat-capable, at an affordable cost, when threat-driven performance requirements change much more frequently than the replacement interval for systems.

Significant threat changes can occur rather quickly within a 10 to 15 year period or less, even though the life span of major surface combatants and systems is much longer. Development of a new system typically takes 12 to 15 years, followed by an additional 30 to 40 years of service life (with perhaps a mid-life upgrade). The challenge continues to be to find affordable technological and programmatic means to keep forces combat-capable during the development, procurement, and service life of individual units and systems.

Building on Past Achievements

Several steps have been taken in the AEGIS Program to stay ahead of the threat and changing operational needs. One significant step has been to rely heavily on key performance factors to assess potential changes and focus the developments—both to drive new combat system baseline designs and to judge potential backfit upgrades to operational baselines. The objective is to enhance and preserve system performance as embodied in the quantitative performance factors. A qualitative statement of the performance factors may be expressed as:

- Reaction Time—target detection to weapon motion
- Firepower –number of weapons per unit time
- Electronic Countermeasures and Environmental Resistance - effective multi-warfare capability in countermeasures and severe natural environments
- Continuous Availability—sustainability, graceful degradation; automatic error detection and recovery, automatic fault detection and isolation
- Coverage—effective detection and engagement volume (out, up, down), depth of fire.

Another step taken by the AEGIS Program involves engineering forward-fit ship and combat system changes for new construction ships in planned periodic upgrades. Advances in ship HM&E over the years have been incorporated in several specific ship hull configurations, and the combat system has been engineered and built in combat system baselines. This approach allows new technologies and performance capabilities to be incorporated into forward-fit

ships in a timely manner, while also allowing strict system engineering control to be maintained. Examples of new capabilities which have been incorporated include ship and combat system shock modifications, introduction of the Vertical Launching System (VLS), and introduction of new computers.

This approach to date has resulted in multiple hull configurations and six AEGIS combat system baselines. Four of the baselines are operational, the fifth is in development, and a sixth is under consideration. Figure 1 shows the first AEGIS ship, the CG-47 Ticonderoga-class cruiser, and major systems and elements which constitute the Baseline 1 Combat System of the ship. The multiwarfare-capable combat system is built around the AEGIS Weapon System Mark 7. Figure 2 shows some of the major warfighting enhancements already implemented or under way in the forward-fit baselines. Changes have been made in all warfare areas in order to keep the combat system capable against threat growth and other changes in operational need over the long duration of the program.

As shown in Figure 2, when a capability is introduced in a new construction baseline, that capability is then incorporated in subsequent baselines. For example, the Mk 26 rail missile launcher of Baseline 1 was replaced by the VLS in Baseline 2 and in all subsequent baselines. Similarly, the SPY-1A radar was upgraded to the SPY-1B in Baseline 3 and to the SPY-1B(V) in Baseline 4 cruisers. A variant of the SPY-1B(V), the SPY-1D, was developed for Baseline 4 and follow-on destroyers to accommodate the single deckhouse of the destroyer class. The SPY-1 Engineering Development Model (EDM-4B) capability is under development for Baseline 6. Similarly, major advances in antisubmarine warfare and other areas of the combat system have been incorporated into the baselines over time.

In making forward-fit upgrades to keep new construction ships combat-capable, the AEGIS Program has taken steps to ease potential future backfit upgrades in older AEGIS ships. The fact that each combat system baseline is built upon the preceding baseline helps; but beyond that, the AEGIS Program has also taken other steps in the forward-fit developments to preserve backward compatibility with earlier baselines. The approach which has been used allows substantial commonality in system computer programs and eases potential backfit of selected performance features from more recent baselines into earlier baselines. The system design and computer program capture is not an inconsequential benefit in a highly automated,

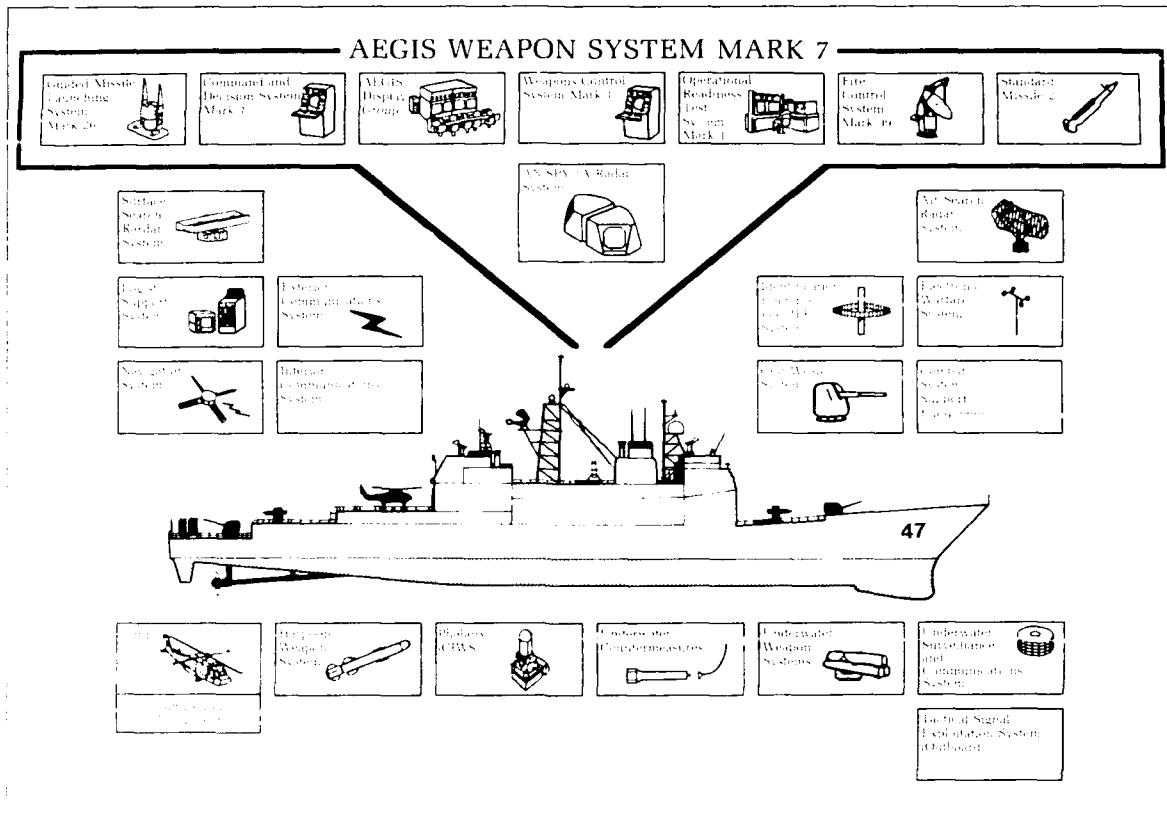


Figure 1. AEGIS ship combat system.

complex, computer-intensive system like AEGIS. Figure 3 shows the size of the Baseline 1 tactical computer programs (excluding several million additional lines of code for the Operational Readiness Test System and AEGIS Display System data bases) and the incremental increases in tactical computer code which have been added to accommodate subsequent baseline upgrades through Baseline 4. As previously indicated, many of the major changes in the baselines are shown in Figure 2, but many other smaller changes are not reflected here.

A rather modest increase in new computer code was required to implement Baselines 1 through 3. More substantial performance enhancements were added between Baselines 3 and 4, since the transition to the more capable UYK-43 and UYK-44 computers could accommodate the level of change required. The common "superset" design for Baseline 4 cruisers and destroyers positioned both ship classes to potentially capture upgrades in follow-on baseline developments. For example,

the planned SM-2 Block IV capability of Baseline 5 can readily be backfitted into all Baseline 4 ships. Similarly, Baseline 2 and 3 ships can capture much of the design, computer programs, and combat system performance from Baseline 4 (including SM-2 Block IV capability) if selected equipment such as UYK-43 computers and UYQ-21 display consoles are backfitted into the Baseline 2 and 3 ships. The future backfit plans for AEGIS ships are contained in the Ticonderoga (CG-47) Class Warfighting Improvement Plan.² The extent to which this plan is executed will depend on budget considerations and other factors.

Future Opportunity

The Round Table studies to date suggest that U.S. technology advances will allow systems to stay ahead of the projected threat, and further, that systems can be developed to defeat the technological threat out to the present visible horizon (about 2010). No revolutionary depar-

tures in sensor and weapon types and force operations appear to be mandatory. Radars, missiles, force organic air capability, sonars, and the like are expected to continue to be core-required capabilities even out to the 2030 time frame, but the operational characteristics and performance capabilities of the advanced systems will be substantially different from present systems. New capabilities such as directed-energy, self-defense weapons may also begin to appear in the surface forces by about 2010, but these systems are expected to have very limited applicability for many years after their initial introduction.

Building technologically advanced systems to defeat the threat and satisfy other operational objectives appears to be feasible. Transitioning that capability in quantity into the fleet in a cost-effective manner remains a major engineering and program management challenge. Lessons learned need to be applied, and innovative means need to be found. The challenge will not soon disappear in this era of

rapid technology advances. The problem is well recognized, however, and a number of ideas are being assessed and formulated in ongoing studies like the DTRC and AEGIS Round Table efforts. The ideas include moving toward an open combat system architecture to more readily accommodate change in all dimensions, including command and control, weapon systems, and communication systems. Substantial technology advances are being made in distributed high-capacity processors, bus interconnection systems, and smart display workstations which could support such an architecture. One of many notional concepts at the pictorial level in this area is shown in Figure 4.³

Benefits could also be realized by placing greater reliance on long-range forecasting and greater focus on designing to technologically feasible, rather than projected near-term (5 to 10 year future threats). Because of the long development and operational life span of systems, some level of performance margin—

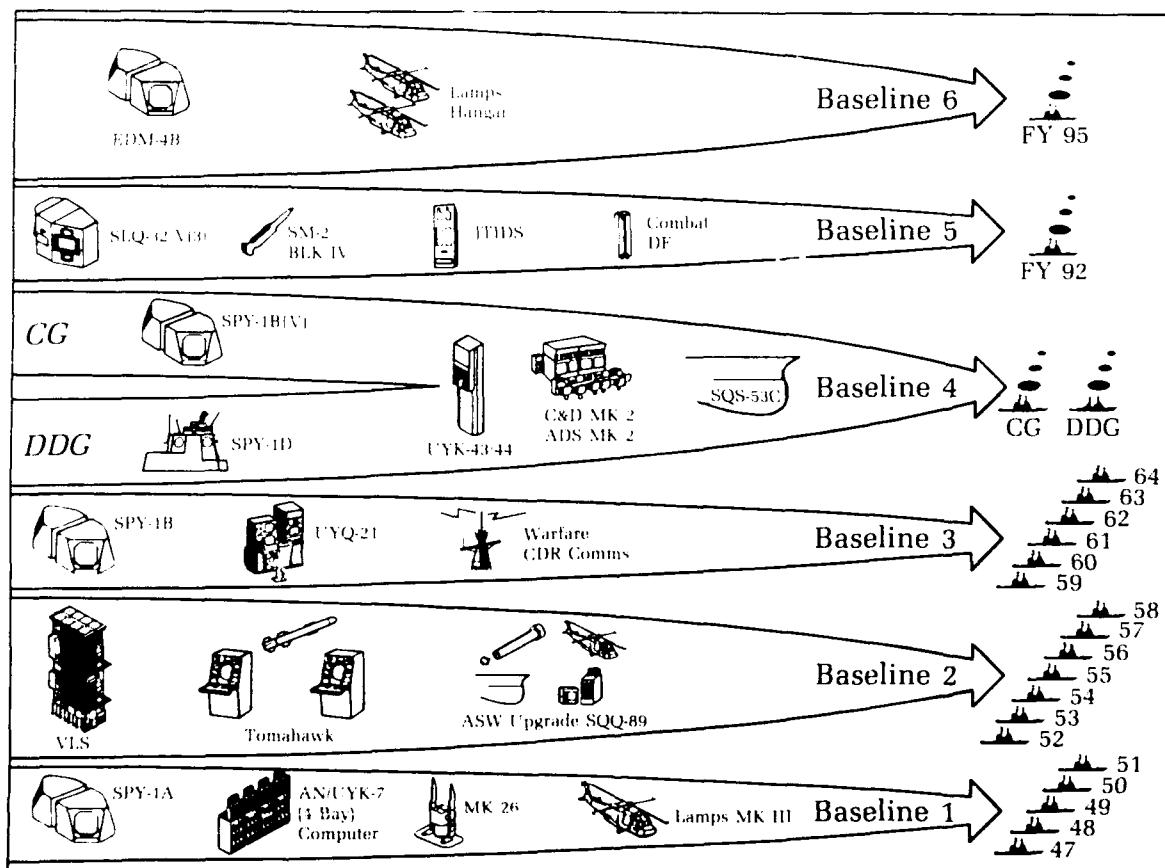


Figure 2. AEGIS combat system baselines.

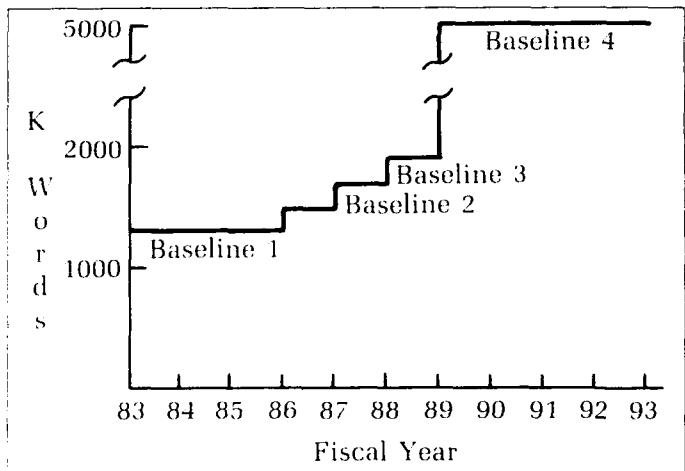


Figure 3. Computer program code active in AEGIS.

balanced by affordability—needs to be incorporated into initial future-system designs.

Fewer surface combatants of higher quality with enhanced organic aircraft capability may

offer the most attractive approach in a time of reduced budgets. The quality is required to survive in the expected difficult multiwarfare threat environments, and fewer units might

20

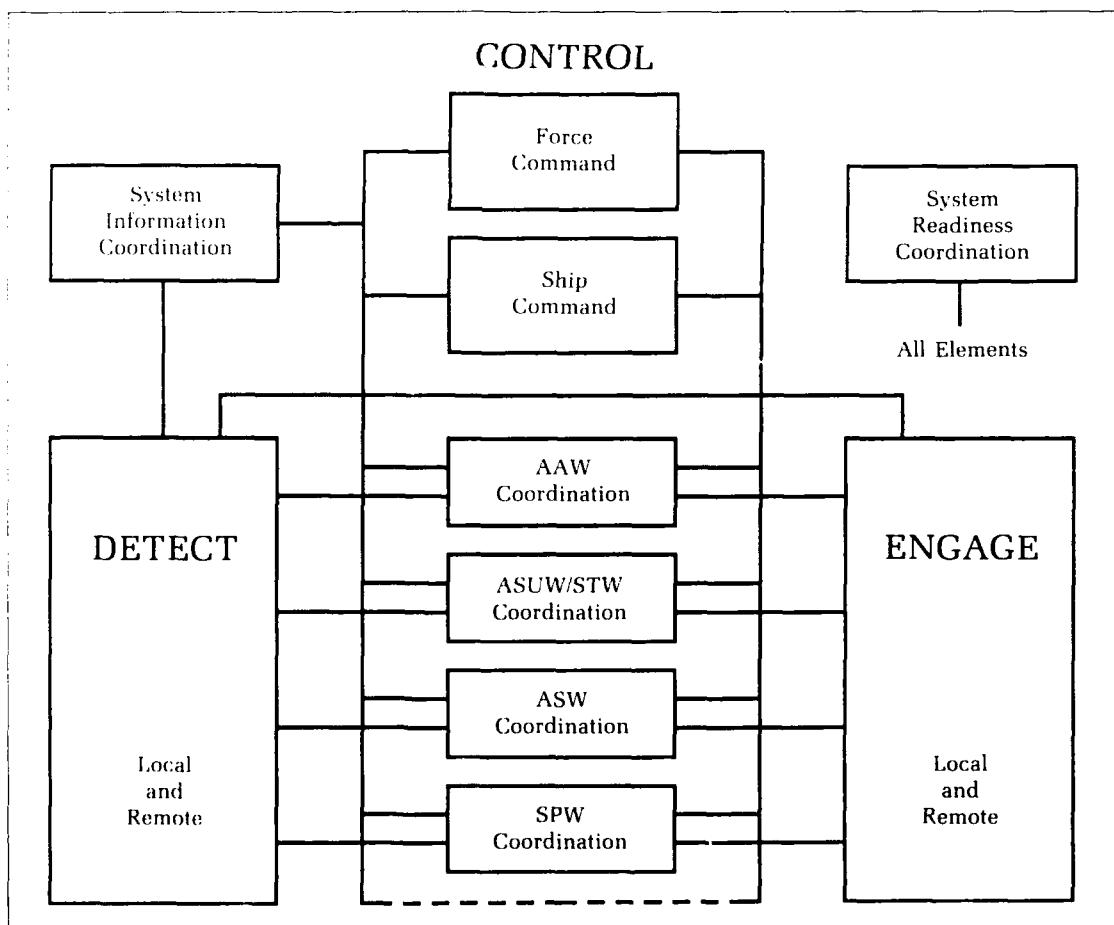


Figure 4. Notional combat system structure.

allow more substantial periodic upgrades to keep systems more combat-capable. There are numerous other benefits including ability to place significant ordnance ashore, to operate without replenishment for long periods of time, and to tolerate war damage.

Netting units together with high-capacity, jam-resistant data links so that older units can more fully benefit from more effective sensor and weapon capabilities in newer force units also has appeal. The Cooperative Engagement Capability has demonstrated potential in this area, and additional possibilities exist as progress continues to be made in airborne and remote underwater sensor and weapon systems.

Conclusions

Technology transition into fleet systems in a manner which will keep them combat-capable in an era of rapidly evolving operational needs remains a critical challenge. Applying lessons learned and focusing on long-term trends, needs, and concepts are mandatory if successful strategies and approaches are to achieve this challenging and complex goal.

References

1. Graham, Captain C. and Bosworth, CDR M., "Designing the Future U.S. Naval Surface Fleet for Effectiveness and Productivity." *SHAME 1989 Ship Production Symposium*, Arlington, VA, 13-15 Sep 1988.
2. Chief of Naval Operations, *Ticonderoga (CG 47) Class Warfare Improvement Plan*, Washington, D.C., Department of the Navy, 2 Jun 1988.
3. Cullen, Richard P., *A Concept for Future Force Combat Systems: An Enduring Structure*, NAVSWC TR 90-121, Feb 1990.

The Author



THURMAN C. HENDERSON, a senior physicist in the Combat Systems Department, has served as AEGIS Principal Combat System Engineer since 1985. He was instrumental in establishing the AEGIS Systems Engineering Branch and served as its head from 1981-85. He received a B.S. degree in physics from King College, Bristol, Tennessee, in 1965 and accepted a position at NAVSWC in 1966. He completed more

than 40 hours of graduate courses under the NAVSWC continuing education program and received a Masters degree from the University of Maryland in 1974. Mr. Henderson performed T&E and computer modeling in HERO, EMC and nuclear effects; served as analyst and engineer in electronic warfare, wide-area surveillance, and intelligence systems; directed a team in the development of radar and passive detection computer models for aircraft, ship, submarine, and space-based systems; and directed R&D in advanced radar systems. He has served as U.S. representative to NATO studies in technology and future combat system concepts and chaired U.S. Navy studies in DDG-51 AAW, Fleet Outer Air Battle, Fleet Command and Control, and Combat System Concepts for Future Navy Surface Combatants. He is lead evaluator for General Electric's Aerospace and Defense IR&D Program. Mr. Henderson received the Bernard Smith Award in 1985 for technical ability and superior leadership in expanding AEGIS combat system engineering at NAVSWC.

Objectives, Principles, and Attributes: A Structured Approach to Systems Engineering

David K. Kreider and Richard E. Nance

The complexity of modern systems and the acceleration of technological advancements with their insertion into engineering applications are a few of a multitude of factors requiring a disciplined systems engineering approach to naval system development. A "total" systems engineering perspective on naval systems development is the challenge of the future. This perspective requires the engineering community to evolve from the prevalent "commodity viewpoint" to a more comprehensive system engineering structure. The objective, principle, and attribute (OPA) concept provides a formalized approach to structuring systems engineering efforts, and to developing and evaluating tools and methodologies to support the evolution to this systems engineering structure.

The classical systems engineering approach appropriately focuses on defining needs (operational requirements), understanding the interrelationships of these needs, and developing a system to satisfy these needs. The OPA concept broadens this focus by providing an infrastructure for a more structured approach to development and evaluation of tools and methodologies. The OPA approach is based on the concept of relating objectives (high-level requirements), principles (use of fundamentals, e.g., functional decomposition) and attributes (desired technical performance characteristics). In addition, it provides a means of relating systems engineering efforts and promoting increased understanding of systems engineering and its relationship to other disciplines. A direct consequence of this increased understanding is an improvement in the application of the systems engineering process.

Introduction

Rapid technology advancement, accompanied by expectations of more capable and complex systems, are forcing a "total system" perspective on naval engineers. The autonomous development of combat system elements can no longer be accepted; each element must be created as a system functioning within a hierarchy of systems. This mandatory refocus from a predominantly "commodity viewpoint" to that of a more comprehensive systems engineering viewpoint represents a major challenge.

Such a challenge can be met successfully with systems engineering methodologies, tools, and training programs integrated into a structured environment facilitating the application of the systems engineering process. Providing the guiding framework of the systems engineering process, supported only by independently developed, nonintegrated tools and methods to aid the engineering community, is inadequate. A promising concept to support this future evolution is the objectives, principles, and attributes framework applied within the systems engineering process.

The primary objective of this article is to propose the OPA framework as a means for structuring, understanding, and guiding systems engineering activities within the naval system development process. A secondary

objective is to define how concepts advanced in the OPA framework can facilitate the systems engineering process by providing a means of developing, evaluating, and selecting tools to support the process.

What Is Systems Engineering?

While no definition of systems engineering is universally accepted, the scope and goals of systems engineering are easily understood. Within this article, systems engineering is considered to be "a logical sequence of activities and decisions transforming an operational need into a description of system performance parameters and a preferred system configuration."¹ A more comprehensive definition of systems engineering is "the application of scientific and engineering efforts to (a) transform an operational need into a description of system performance parameters and a system configuration through the use of an iterative process of definition, synthesis, analysis, design, test, and evaluation; (b) integrate related technical parameters and ensure compatibility of all physical, functional, and program interfaces in a manner that optimizes the total system definition and design, and (c) integrate reliability, maintainability, safety, survivability, human factors and other such factors into the total engineering effort to meet cost, schedule, and technical performance objectives."^{2,3}

Several key concepts of systems engineering, reflected not only in these definitions but throughout systems engineering literature, warrant further emphasis. These concepts include:

- Systems engineering emphasizes global optimization over local subsystem or component optimization.
- The systems engineering process transforms a user's need into requirements, specifications, and a preferred system configuration to satisfy that need.
- Systems engineering adopts a total life cycle perspective.
- Systems engineering is an iterative process.
- Systems engineering is interdisciplinary in nature.

The systems engineering process is often described in terms of phases, activities, and areas of knowledge, as shown in Table 1. In addition, the process is often graphically represented through waterfall charts or axis diagrams such as Hall's morphology. While differences in terminology and categorization of the phases and activities exist, in general, each representation addresses the same phase and activity areas.⁴

In application of the systems engineering process, the early phases define the high-level objectives, requirements, and functions for the system being developed. The systems engineer coordinates and integrates the interdisciplinary efforts to transform these high-level requirements into more detailed requirements and specifications. Continuing in an iterative process, the requirements and specifications are eventually transformed into a system definition, design, and a preferred system configuration to meet the user's needs.

Naval Systems Engineering

Recognizing systems engineering as an appropriate approach for meeting the naval system development challenges is not a recent idea. Systems engineering has been prevalent in the military since the 1940s.^{3,5,6} However, primarily due to technological advances, systems have continually advanced in size and complexity. Other factors also influence the advancement of and need for systems engineering, for example: increasing development and implementation time; increasing development, training, and operational costs; decreasing development and acquisition funding; the need for preventing or avoiding errors or failures of earlier development efforts, and forestalling rapid obsolescence due to technological advances. All these factors, along with others not listed, are extremely prevalent today.

In the 1950s and 1960s, systems engineering was conducted primarily at an equipment level (radars, missiles, receivers, etc.). Within the surface Navy, the nature of system development at an equipment level has reflected a "commodity perspective." This implies that system elements are generally developed independently, then integrated into the fleet in a piecemeal fashion. Such development often fails to take into account the overall impact of programmatic decisions on the system, and interoperability aspects are often overlooked altogether. These potential negatives can result in system characteristics such as incompatibility, electromagnetic interference, undesirable performance characteristics, and poor reliability and maintainability when operated in conjunction with or in the presence of other systems.

In the 1970s and early 1980s, systems engineering efforts became associated with weapon systems and combat systems development for such programs as AEGIS. These programs recognized that integrating individual system elements into larger systems could potentially alleviate shortfalls or improve performance. However, this type of systems development within the surface Navy often focuses on "how" to physically integrate

scenario During the breakout phase, the screening force may perform mine countermeasures and shallow-water ASW for the main

- Countermeasures
- Special Operations
- Combat Search and Rescue

Table 1. The Systems Engineering Process⁴

LIFE CYCLE PHASE	ACTIVITIES
Identify Need	Identify the Need & Define Problem Define the System & Objectives Define within Hierarchy of Systems
Planning	Program Organization Logistic & Program Planning (schedule, milestones, etc.)
System Studies	Exploratory Studies Scientific Research Feasibility & Trade-off Studies Data and Information Collection Analysis (functional, cost, risk, effectiveness, modeling...)
Requirements & Specifications	Define Systems Constraints Requirements Definition Requirements Synthesis Specification Generation
Design	Architectural Design Functional and/or Hierarchical Decomposition Preliminary & Detailed Design Reliability & Maintainability System Selection System Synthesis & Development Prototype Construction
Test & Evaluation	Demonstration and Validation Requirements Verification Testing and Evaluation Data Collection
Production	Construction & Site Preparation Quality Control Distribution Industrial Engineering
Operation	Activation and Logistic Support Monitor Performance Training Maintenance & Operation System Upgrades or Modifications Phase-Out, Recycle, or Disposal

existing systems versus the need for overall requirements to be met by the integrated systems. Systems development through integration that focuses on the physical connections continues to promote development from a commodity viewpoint. As a result, effective integration cannot overcome the shortfalls inherent in development of individual system elements unless focus is maintained on the overall system requirements.

24 This is not to say that systems integration is not desirable. Due to development cost and acquisition time constraints, naval development will continue to be initiated with the integra-

tion of current fleet assets as the basis. However, systems engineering is more comprehensive than system integration. In applying the systems engineering process, the Navy needs first to identify fleet needs and develop overall requirements, then determine the best mix of systems integration, new systems development, and technology insertion to meet these requirements. Existing elements (or subsystems) must not be permitted to impose binding constraints on the total system capability.

The Space and Naval Warfare Systems Command (SPAWAR), established in 1985, initiated force systems engineering to identify

Navy needs and requirements for the future fleet. SPAWAR is tasked with the development and acquisition of command, control, and communication systems, space systems, and naval warfare systems. This mission includes force-level warfare systems architecture and engineering (WSA&E) to transform top-level warfare requirements and architectures into high-level system specifications.⁷

As the application of the systems engineering process at a force level continues to evolve through SPAWAR WSA&E and other efforts of similar nature, progress at these levels has been slowed by uncertainties and risks inherent with efforts in previously unexplored areas, or attendant to the insertion of new technology. Mahan, Beach, and Betts provide numerous examples of how traditions, experience, and perceptions have inhibited the willingness to attempt new alternatives to warfare, weapons, or processes within the naval community.^{8,9,10} In order to support the evolution of systems engineering at a force, combat, weapon or other level, a means to structure or relate efforts and to facilitate the systems engineering process is needed.

The Need: A More Formalized Approach

The problem of how to improve the application of the systems engineering process does not stem from the lack of design methodologies and tools, or the inadequacy of the process itself.¹¹ Rather, the problem can be framed by several fundamental questions, among them:

- How to provide more structure to the systems engineering process and more discipline to the activities within the process?
- How to develop, select, and evaluate tools and/or methodologies for use in applying the systems engineering process?
- How to improve the understanding of, support of, and communication within the systems engineering process?

These questions must be asked at all levels of naval systems engineering, and during all phases and activities of the systems engineering process.

Origin of the OPA Concept

Challenged to evaluate comparatively two software development methodologies, Arthur and Nance¹² proposed the Objectives, Principles, and Attributes framework in the mid-1980s. This approach, initiated with the distinction of "methodology" from "method" or "tool," emanates from recognition of basic premises:

Software objectives are desirable project level properties. Software principles govern the development process so that desirable properties (objectives) can be achieved (Figure 1). Software attributes, as a consequence of the application of designated principles, are induced in the products (code and documentation).

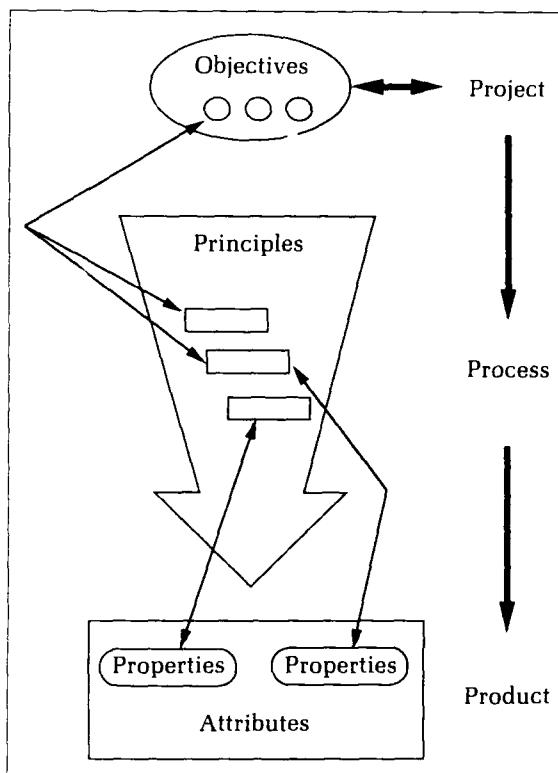


Figure 1. The objectives, principles, and attribute framework for software development.

A more recent study investigates the utility of applying the OPA concept to the discipline of systems engineering, and concludes that the concept appears beneficially applicable.⁴

Recognition of the importance of objectives and principles is not new to systems engineering or naval development approaches. For example, combat system architecture (CSA) efforts in the late 1970s stipulated the identification of high-level requirements as the precursor to specifying naval architectures.¹³ Such requirements correspond to objectives within the OPA concept. The requirements are obtained by implementing specific design principles. However, the CSA definition of "principles" in terms of design characteristics falls short of providing the necessary process definition afforded by the OPA framework.

The Objectives, Principles, and Attributes Concept

Central to the OPA concept is the premise that tools embody methodology principles, and the application of the tools implements the principles. Principles correspond to systems engineering fundamentals or techniques such as hierarchical and functional decomposition. The application of a principle governs the development process in a manner to promote the attainment of a desired objective. Within naval system development, objectives (e.g., reliability and maintainability) correspond to high-level system goals or requirements from which specifications (A-Specs) can be defined. The presence of a tool in a development environment is based on its support of one or more principles; but more important, this methodology guides in the partitioning of principles among tools to achieve an integrated support environment.

Another OPA premise is that use of a particular principle induces specific attributes in the final product or system. Attributes are desirable characteristics corresponding to technical performance characteristics in naval system development. The presence or absence of attributes is examined to measure success in attaining desired objectives.

One might contend that objectives, principles, and attributes are applied currently in systems engineering practice. If true, the current implementation is poorly defined and understood, inadequately documented, and primarily employed at the equipment level. For example, techniques are available for designing and measuring reliability at an equipment level. The objective (reliability) and an attribute (mean-time-between-failure) are readily apparent. The techniques (principles), while often not as apparent, do exist. What is missing are the well-defined relationships. As a result, the relationships may not be adequately comprehended or exploited to realize the full potential benefit of the OPA framework.

While the OPA framework has the potential to provide structure and discipline, the application raises several practical concerns:¹²

- While a set of objectives can be identified, this set might not be complete, and additions should be allowed.
- Objectives may be interdependent (for example, maintainability may have an adverse effect on reliability).
- Objectives can differ in importance from one application to another.
- Attributes of a large, complex system might be evident in one component, yet missing in others.

26

The following sections provide examples that illustrate and clarify the extension of the OPA framework to systems engineering.

Systems Engineering Objectives

The systems engineering process can be applied at varying levels, depending on how the boundaries of the target systems are drawn. Each level of application of systems engineering, existing system, or system being developed, has objectives. Objectives are viewed as invariant with the level of application. While one set of objectives may not apply to all systems, a small sub-set of objectives may be recognizable in the majority of systems. This set may depend on application domain factors (military, communication, medical, etc.). For example, naval combat system objectives might include:

- Adaptability - the ability of the system to accommodate change. Change may involve adjusting to changes within the environment, modifications to meet new operational requirements, or enhancements to improve performance.
- Feasibility - operational, technical, and economic viability.
- Maintainability - the ability of the system or its subsystems (components) to be maintained.
- Capability - the functional and operational capability of the system and its components under specified conditions.
- Reliability - the ability of the system to perform in a satisfactory manner for a specified period of time when operated under specified conditions.
- Compatibility - the ability of the system to be integrated with or operated in the presence of other systems without mutual interference.
- Verifiability - the ability to verify that the system supports the naval mission and meets its system requirements.
- Operability - the ability to perform as intended in an effective and efficient manner for a specified period of time.

This list of objectives is not to be considered all-inclusive or applicable to all systems.

Systems Engineering Principles

Several engineering fundamentals or techniques provide the counterparts of principles within the OPA framework: hierarchical and functional decomposition, information hiding, life-cycle verification, expansionism, and concurrent documentation are ready examples.

Hierarchical decomposition refers to the hierarchical breakdown of the system into subordinate parts (subsystems and components). Functional decomposition first identifies the primary functions of the system and then progressively breaks these down into subfunctions.¹⁴ Hierarchical and functional decomposition help assure that all elements of the system are fully recognized and defined, and all facets of system development, operation, and support are adequately covered.²

Information hiding is a concept in which the components of the system "know" only the necessary information about the other components comprising the system. System details likely to change are "hidden" inside individual system component design.¹⁵ This does not imply that the engineers developing the individual components are unaware of the "global" system objectives and the functionality of the other components, but that the system components need not necessarily be aware of the details of how other components provide their functionality. This principle is applicable to the promotion of modular design and adaptability within systems since modifications or enhancements to one module or subsystem have less impact upon the entire system or other subsystems.

Life-cycle verification confirms adherence to system requirements throughout the design, development, and maintenance of the system life cycle.¹⁶ It also involves verifying that principles, tools, and methodologies are used properly throughout the systems engineering process.

Expansionism involves, and may be defined as, the hierarchy of systems within the realm of systems. As such, it is closely related to or may be considered part of hierarchical decomposition. Just as consideration of the hierarchy of subsystems or functions within a system of interest is important, recognition and consideration of a system's relationships and position within the hierarchy of systems is also important. This principle aids the system engineer in defining the objectives and requirements of the system of interest in relation to other systems.

The principle of concurrent documentation promotes requirements and decision traceability. This may reduce redundancy in past and present efforts, reduce the probability of perpetuating previous design errors, promote understanding of system development objectives, and ultimately reduce development time and cost.

Systems Engineering Attributes

Attributes are desirable characteristics of the system product. An attribute may or may not be visible or quantitatively measurable. For example, the technical performance characteristic mean-time-between-failure is generally measurable. However, an attribute such as simplicity may be found to be less formally defined and more subject to human judgment. Examples of attributes and their relation to specific objectives are provided in Table 2.

Table 2. Systems Engineering Attribute Examples

OBJECTIVE	PRINCIPLE	ATTRIBUTE
Maintainability	Functional Decomposition	Accessibility Simplicity Mean-Time-between-Maintenance Availability
Reliability	Functional Decomposition	Redundancy Mean-Time-between-Failure Availability
Verifiability	Documentation	Traceability Documentation Adequacy

Benefits of the OPA Approach

Applying the OPA framework within the discipline of systems engineering has three major potential benefits. First, it provides a means or structure for relating the different levels of systems engineering (Figure 2). As efforts move from an equipment level toward the force level, complexity increases. Requirements are to be developed in a manner such that efforts at lower levels support attaining higher level objectives (top-down requirements definition).

design and development. That is, objectives and requirements are defined from the top down and design/implementation occurs from the bottom up.

The second major benefit of the OPA framework is a disciplined procedure for developing, evaluating, and selecting tools and methodologies to support the application of systems engineering. The linkage between objectives and principles serves as the fundamental basis for this benefit. Principles provide the basis for tool and methodology development, and attributes provide the basis

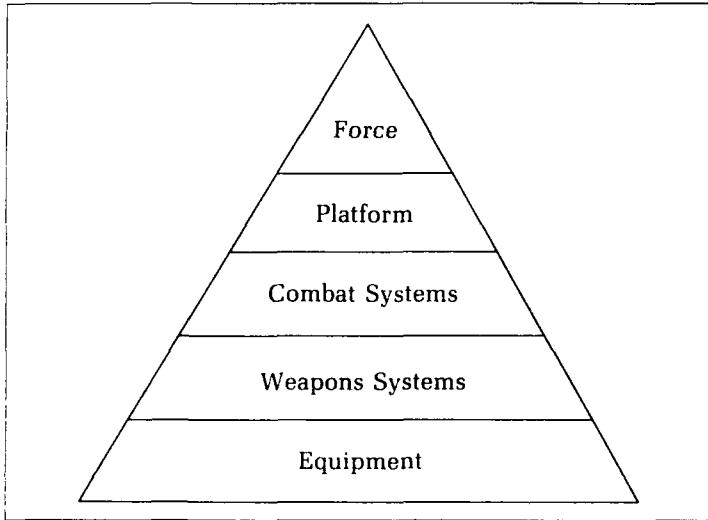


Figure 2. Pyramidal illustration of the levels of systems engineering.

Each level of systems engineering has objectives, and emphasis on particular objectives may differ between levels. For example, it is possible for greater emphasis to be placed upon reliability at the weapon system and equipment levels than at the force level. At the former levels, reliability could support the attainment of objectives receiving greater emphasis at the force level such as maintainability, supportability, or performance. As a result, a hierarchy of objective levels is derived. The relationships among objectives at different levels within the hierarchy provide the means for relating the corresponding levels of application of the systems engineering process (Figure 3). A crucial assumption is that the achievement of objectives defined for the encapsulating system (highest level) is based on synthesis of objectives realized in the embedded systems (lower levels).

In addition, the demonstration of linkage in levels of application of the systems engineering process aids in understanding the "total" systems engineering effort. This illustration depicts a top-down/bottom-up approach to

for evaluation. For example, existing tools and methodologies can be evaluated to determine the principles employed. This evaluation can serve as an indication of the objectives the application seeks to promote. Similarly, development of new tools or methodologies would be based on the principles promoting the desired objective. The success or "goodness" of the new tools and methodologies is then evaluated by measuring or evaluating the attributes present in the product. By determining which objective(s) are attained by employing a specific tool or methodology, the system engineer is provided with a basis for selection of tools that best support the efficient and effective development of the system of interest. Thus, the selection of the tool or methodology is a function of the objective desired and the level of systems engineering being addressed.

A third benefit is that the achievement of objectives by use of specific principles, and the measurement of attributes and principles to indicate the attainment of these objectives, is easily understood by engineers and technicians

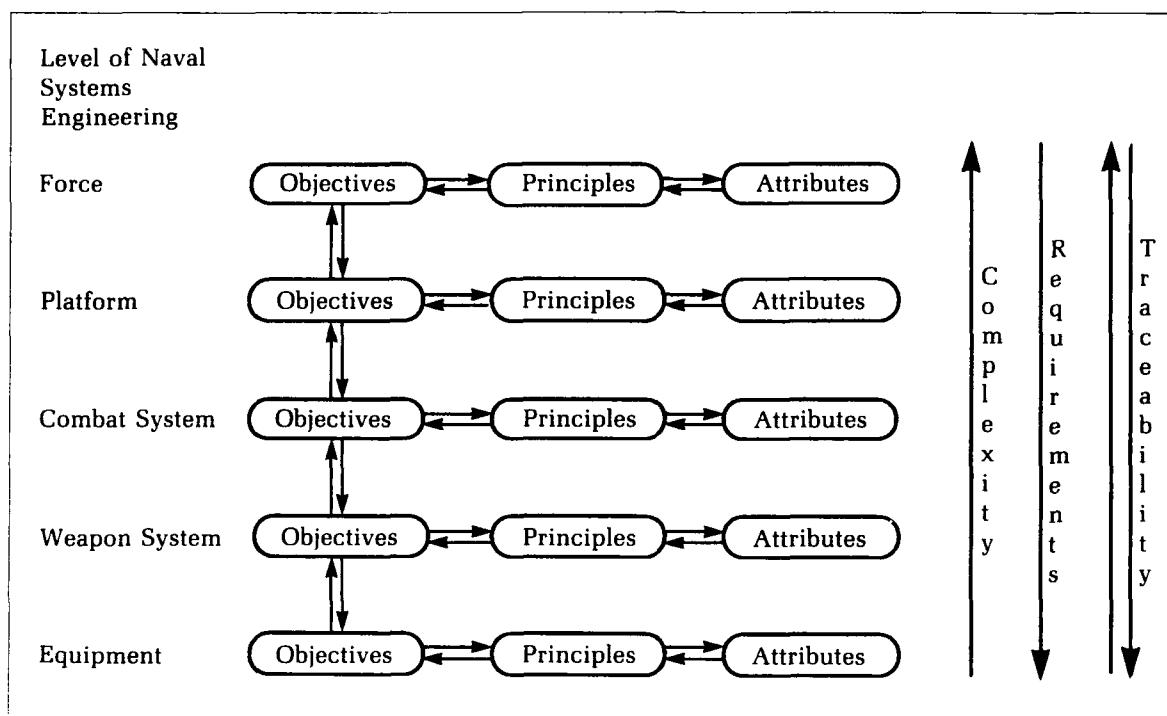


Figure 3. OPA hierarchy.

regardless of background or experience level. The natural appeal of this interplay among objectives, principles, and attributes facilitates communication and promotes understanding of the methodology and the intended application, strengths, and limitations of tools implementing the methodology. This aids in achieving valid, credible results. A direct and important consequence of this increased understanding and credibility is improvement in the application of the systems engineering process.

In summary, the relationship between levels of objectives provides structure for relating levels of systems engineering. The linkage among objectives, principles, and attributes provides a disciplined approach to tool development, evaluation, and selection. Furthermore, if in developing tools and methodologies the hierarchy of objectives is adhered to, the ability to integrate tools or their results across the total systems engineering spectrum should be enhanced. This ability to integrate directly supports the goal of establishing an integrated tool environment supporting the entire systems engineering process.

Conclusion

To date, the discipline of systems engineering has primarily provided engineers with conceptual support that increases the probability of developing an effective system to solve the right problem. It is time to target more ambitious goals. Better means of implementing, understanding, and relating efforts within the systems engineering process are needed. The OPA framework provides an infrastructure to facilitate the implementation of the systems engineering process and supports future evolution of the systems engineering discipline. By providing a structured approach for relating systems engineering efforts and a discipline for integrated tool development to support the system engineer, the systems engineering process may become more formalized and less subjective. In addition, the OPA framework is adaptable, flexible, and promotes improved communication and understanding by the use of consistent terminology.

Systems engineering objectives, principles, and attributes exist now only in part, they are

only partially defined and inadequately understood, and require further investigation and development. To implement the OPA framework in support of naval or other community systems engineering efforts requires:

- a more definitive understanding of the systems engineering process,
- establishment and definition of an objective hierarchy (if multiple levels of systems engineering exist),
- definition of principles and their relationship to the objectives,
- definition of desirable systems attributes related to the objectives,
- development of methodologies based on the concept of objectives, principles, and attributes, from which the specification of tool requirements can be derived,
- development of tools based on requirements driven by principles, and
- integration of the requisite tools to produce a system engineering environment supporting system development.

References

1. Military Standard Engineering Management, MIL-STD-499A(USAF), Department of Defense, 1 May 1974.
2. Blanchard, B. S. and Fabrycky, W. J., Systems Engineering and Analysis, Englewood Cliffs: Prentice-Hall, Inc., 1981.
3. Systems Engineering Management Guide, Defense Systems Management College, Fort Belvoir, VA, 1986.
4. Kreider, D. K., Objective, Principles, and Attributes: An Approach to Methodology Development and Evaluation for Systems Engineering, M.S. Dissertation, Virginia Polytechnic Institute and State University, 1990.
5. Checkland, P., "Rethinking A Systems Approach," *Journal of Applied Systems Analysis*, Vol. 8, 1981, pp. 3-13.
6. M'Pherson, P. K., "Systems Engineering: an Approach to Whole-System Design," *The Radio and Electronic Engineer*, Vol. 50, No. 11/12, Nov/Dec 1980, pp. 545-558.
7. Lehman, J., Memorandum, Space and Naval Warfare Systems Command (SPAWAR), concerning SPAWAR organization and operational concept, Washington, D.C., 6 Nov 1985.
8. Mahan, A. T., *The Influence of Sea Power Upon History, 1660-1783*, 12th Edition, Boston: Little, Brown, and Co., 1918.
9. Beach, E. L., *The United States Navy 200 Years*, New York: Henry Holt and Co., 1986.
10. Betts, R. K., Editor, *Cruise Missiles Technology, Strategy, Politics*, Washington, D.C.: The Brookings Institution, 1981.
11. Warfield, J. N., "Selecting Participation Methodologies For Systems Design," 1983 *Proceedings of the International Conference on Systems, Man, and Cybernetics*, Vol. 2, 1983, pp. 762-764.
12. Arthur, J. D., Nance, R. E. and Henry, S. M., *A Procedural Approach to Evaluating Software Development Methodologies: The Foundation*, Technical Report SRC-86-008, Systems Research Center, Virginia Polytechnic Institute and State University, Blacksburg, VA, Sep 1986.
13. Cullen, R. P. and Hill, D. L., "Combat System Architecture: Description and Use in Surface Combatants," NAVSWC, Dahlgren, VA, 1981.
14. Eisner, H., *Computer-Aided Systems Engineering*, Englewood Cliffs: Prentice-Hall, Inc., 1988.
15. Opferman, D. C. and Yacobellis, R. H., "A Design Methodology for System Quality," *AT&T Technical Journal*, Vol. 65 May-Jun 1986, pp. 60-72.
16. Dandekar, A. V., *A Procedural Approach to the Evaluation of Software Development Methodologies*, VPI&SU, 1987.

The Authors



DAVID K. KREIDER received a B.S. in Engineering Science and Mechanics from the University of Tennessee in 1984 and an M.S. in Systems Engineering from Virginia Polytechnic Institute and State University in 1990. He is a systems engineer in the Warfare Systems Architecture and Engineering Program Office of the NAVSWC Combat Systems Department. Currently, he is applying the systems engineering process

within WSA&E efforts to define and evaluate current and future naval force architectures. Prior to this, he was the test and evaluation engineer for the RAIDS Program. In addition, Mr. Kreider developed requirements for and initiated the establishment of land-based test site interoperability among the Naval Ocean Surveillance System Center, NAVSWC, and the Wallops Island Naval facilities, and defined battle group critical experiments to be conducted utilizing this capability.



RICHARD E. NANCE is the Rear Admiral John Adolphus Dahlgren Professor of Computer Science and Director of the Systems Research Center at Virginia Polytechnic Institute and State University. He received B.S. and M.S. degrees from North Carolina State University in 1962 and 1966, and the Ph.D. degree from Purdue University in 1968. He has served on the faculties of Southern Methodist University and

Virginia Tech, where he was Department Head of Computer Science, 1973-1979. Dr. Nance has held research appointments at NAVSWC and at the Imperial College of Science and Technology (UK). Within the Association for Computing Machinery (ACM), he has chaired two special interest groups: Information Retrieval (SIGIR), 1970-71 and Simulation (SIGSIM), 1983-85. He has served as Chair of the External Activities Board, the Outstanding Service Awards Subcommittee, the ad hoc Conference Procedures Committee and the ad hoc Film Committee that produced *Computers in Your Life*. The author of over 80 papers on discrete event simulation, performance modeling and evaluation, computer networks, and software engineering, Dr. Nance has served on the editorial panel of *Communications ACM* for research contributions in simulation and statistical computing, 1985-89, as area editor for computational structures and techniques of *Operations Research*, 1978-82, and as department editor for simulation, automation, and information systems of *IIE Transactions*, 1976-81. He is the founding editor-in-chief of the *ACM Transactions on Modeling and Computer Simulation* and a member of the Advisory Board of the *ORSA Journal on Computing*. Since 1983, Dr. Nance has served as principal or co-principal investigator on 14 externally sponsored projects with total funding exceeding \$1.5 million. Dr. Nance received an Exceptional Service Award from the TIMS College on Simulation in 1987. He is a member of Alpha Pi Mu, Upsilon Pi Epsilon, ACM, IIE, ORSA, and TIMS.

Advanced Distributed Processing Technology and ADMRALS

J. Blanton

A generalized methodology for the analysis and simulation of large complex problems on an internetworked distributed computer system has been developed. This methodology, the distributed processing system, and the associated network communication control system are collectively referred to as ADPT (Advanced Distributed Processing Technology). This approach, which includes a highly modular functional representation of the components of a generic complex system, has been applied to the development of a naval battle force, multiwarfare simulation environment called ADMRALS (Attack and Defense of Maritime Resources in Adverse Locales Simulation). ADMRALS represents a collection of newly developed and previously existing models which operate concurrently and interactively on the distributed system. At present, fleet point and area defense, outer air battle, space-based surveillance and targeting assets, and some limited antisubmarine and electronic warfare models are integrated into the system to provide a battle force systems analysis capability not previously available. This highly flexible, synergistic simulation system can be used to conduct weapon system performance and effectiveness analyses, provide value-added studies of proposed new systems, and support the fleet architecture assessment process.

This article describes the methodology used to develop this system and describes models currently included. This work has been sponsored primarily by the Naval Surface Warfare Center (NAVSWC) through the Space and Warfare Systems Command (SPAWAR) Warfare System Architecture and Engineering (WSA&E) Program.

Introduction

Attack and Defense of Maritime Resources in Adverse Locales is a force level, multi-warfare analysis system that is being developed in the Space and Surface Systems Division, NAVSWC. It represents a specific application of a broader methodology called Advanced Distributed Processing Technology that is also evolving at NAVSWC.

This article describes the basic, underlying concepts associated with ADPT and the computer science technology that has emerged as a result of this work. A brief overview of the ADMRALS application and current capabilities is then presented.

Underlying ADPT Methodology

Figure 1 is a notional representation of a large class of systems. A system in this class consists of many objects, entities, or components of varying degrees of complexity. These entities typically interact with one another through a mutually shared environment. These interactions may themselves be complex as well as dynamic. As the system evolves, interactions may result in the creation of new objects or the elimination of existing objects.

Analyzing or modeling such a system is difficult for several reasons. True experts in detailed nuances generally are associated only with specific subsystem components. Therefore, the expertise of many component experts

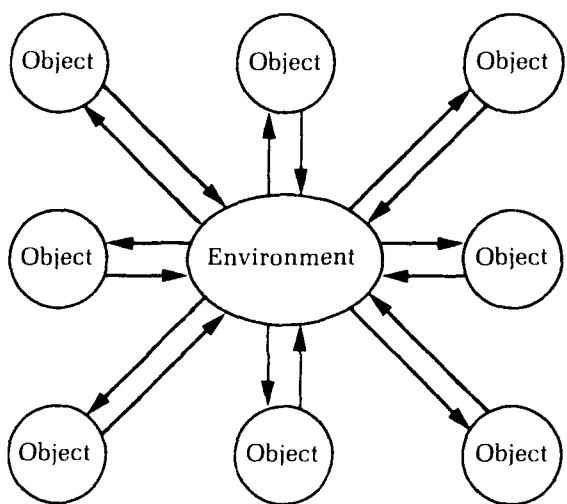


Figure 1. Typical system representation.

must somehow be coordinated and integrated in any attempt to understand the whole system.

Differing analysis needs also affect modeling requirements. Some analysts may require detailed, high fidelity models at the component level. These detailed models may be totally inappropriate for broader, high-level systems analysis requirements. This implied need for flexibility and interchangeability of models and the lack of global consensus on the acceptability of models pose still more difficulties for the system analyst.

Difficulties in analyzing large, complex systems lead to software development problems. Computer simulations of such systems have traditionally been implemented as in-line code on large mainframe computers. As the system definition changes or analysis requirements change, the code is modified. These modifications can be difficult and often have unexpected side effects. Serious software maintenance problems can occur even if good software design standards have been practiced.

Large mainframe computers are powerful and have widespread utility for many applications. They are also expensive in initial cost, in maintenance, and for the support infrastructure generally associated with them. A mainframe computer also represents an essentially closed architecture; expansion of capabilities is limited. It is possible for a complex systems analysis application to outgrow the computer on which it is hosted.

The various aspects of ADPT help ameliorate system analysis problems, software development problems, and computer architecture constraints. ADPT is based on a simple underlying

precept: Large complex problems can be made tractable by decomposing them into a finite set of generic functions and associated control maps. This philosophy implies that many, if not all, systems and subsystem components or objects inherently consist of a finite number of functional attributes and interconnectivities among these attributes. If this is true, and if those functions and interconnectivities (control maps) can be generically defined, a template can be developed that supports systems analysis and model development.

As a first step toward this template definition, consider one of the objects in the notional system of Figure 1. Each object has a two-way interaction with its environment. In general, it is affected by its environment, and in turn exerts some influence or effect on that environment. This interaction is shown in Figure 2, where the beginning of the development of a usable set of generic functions can be seen. At any instant, environmental influences are perceived by the object through generic sensory functions. Based on its perception of the environment, the object then exercises decision-making functions. Finally, as a result of its perception and decisions, the object executes some action which affects the environment. The simple, high-level object decomposition shown in Figure 2 is not very useful for the practical development of models for complex systems. It does, however, provide some insight into the thought processes that led to the development of the more detailed system of generic functions and interconnectivities shown in Figure 3. This function template has been used in all current ADPT applications to develop models of systems components or objects. For most of these applications these objects have tended to be platforms such as

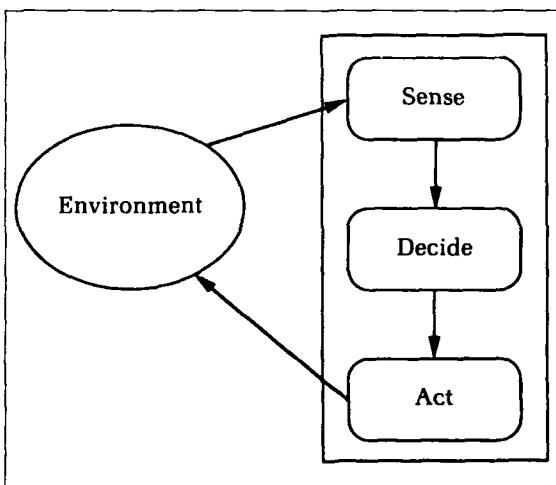


Figure 2. High-level functionalization.

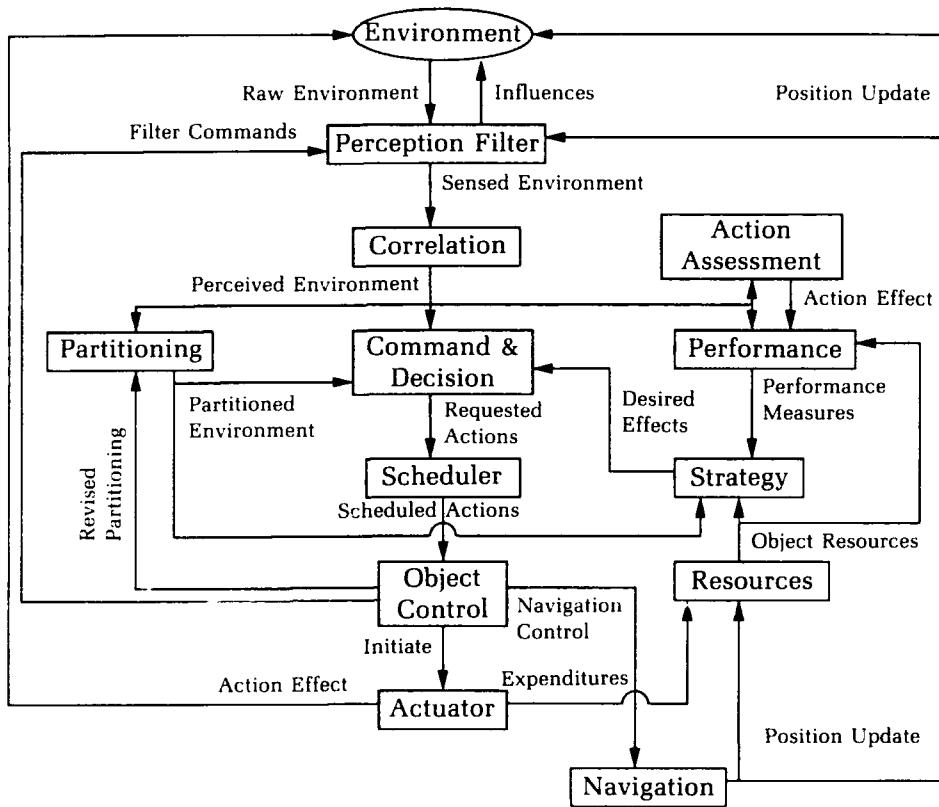


Figure 3. ADPT generic functions.

ships, aircraft, or missiles. All functions are not necessarily required in every model or for all possible applications or analysis needs. To date, no additional function has been required. This functionalization is not necessarily considered to be unique. For some problems, it should perhaps be modified. However, experience has shown that it supports a broad class of complex system analysis applications, especially those which include human or other intelligent decision-making processes. Any process inherent in such a system can be represented by one of these generic functions, and any object in such a system can be modeled through the combination of all or a subset of these functions.

From a software development perspective, the generic functions may be thought of as libraries of algorithms or software modules of similar functional utility. Perception Filter includes models of things which interface with the environment to obtain information. Correlation models integrate information from various perception filters. Action assessment models provide evaluation of action previously taken. Partitioning algorithms partition or segment the perceived environment. Performance,

Command and Decision, Strategy and Scheduler are all models of intelligent or decision-making processes. Resources are models that count assets or resources. Object Control initiates actions, Actuator exerts influences on the environment, and Navigation consists of models which move objects in time and space.

The lines connecting the functional blocks indicate the flow of information which must be shared among functions. If the functional software modules are designed to include the logical hooks necessary to accommodate inter-function information flow, considerable advantages can be realized. By maintaining the logical interfaces, a module can be changed or replaced with little or no impact on other parts of the code. This modular design with carefully defined interface control allows a great deal of flexibility in modifying and interchanging modules for varying analysis needs. The effects of changes are localized. System models can be developed in a "building block" fashion and varying degrees of model fidelity can be supported.

Functions near the top and bottom of Figure 3 tend to represent hardware or models of

hardware. The functions in the center of Figure 3 represent the various aspects of decision-making processes. In most traditional system models, some or all of these decision-making processes are only implicitly reflected in the code. The functionalization shown in Figure 3 includes explicit representation of these decision-making processes. Consequently, a strategy module, for example, can be changed in a larger model as easily as a particular sensor or perception filter algorithm can be changed.

Another aspect of the functional decomposition of Figure 3 is the existence of an implied hierarchy of model development. Each generic function receives information from other parts of the system (its environment), performs some action based on that information, and exerts some influence on other parts of the system. In this sense, each function block resembles the generic object of Figure 2. Consequently, each function can itself be broken down further by the same functional decomposition already discussed. The hierarchical nature of this functional decomposition allows an analyst to conduct, in principle, top-down system analyses and model development by successively applying the function decomposition until any desired level of detail is reached.

The greatest utility of this functional decomposition is realized when it is used from the top down for system analysis as well as for the entire software design and development process. Even when this methodology is not used in the initial design process, certain benefits can often be obtained. If some or all of the functions and interconnectivities in Figure 3 can be identified in an existing program or model, the code or software corresponding to those functions can be extracted, placed in the appropriate generic software libraries, and reused for other applications. In some cases, if the proper object/environment interactions are present, models of objects can be reused in their entirety. This may be done even if the internal representation of the object model does not reflect the functional decomposition of Figure 3. Generally, however, it is more appropriate to extract and reuse parts of existing programs. Functions at the top and bottom of Figure 3, which often correspond to physical objects, are usually more readily identifiable and reusable than are functions corresponding to decision-making process.

ADPT Computer Technology

The previous discussion described a new systems analysis and software engineering methodology. Model development based on the generic system/object functions of Figure 3 is

independent of target host computer architecture. Software development, modification, and maintenance advantages will be obtained with ADPT whether the host computer system is a large mainframe, a desktop, or a distributed processing network. Because of the highly modular nature of this approach, models developed in this manner are particularly well suited to implementation on a distributed processing system. For some classes of problems, certain additional advantages can be realized through such an implementation.

A distributed processing system is an open architecture. The network may start with a small number of inexpensive machines. The computer suite can then be expanded as appropriate for the problem being addressed. New computer technology can be added incrementally. Very large problems may be addressed and execution- or through-put time reduced through the use of concurrent processing. The major technological developments of ADPT support the utilization of distributed systems and concurrent processing.

Concurrent processing can be defined as a number of synchronized sequential programs, communicating with one another and executing concurrently on multiple processors. Since concurrent processing consists of the execution of a number of sequential programs, all considerations for good sequential program development also hold for the individual programs included in a concurrent processing system. There are, however, additional important factors to be considered. These factors include inter-program communication across the distributed network, synchronization among the various programs or processes, and the assignment of processes to the various computing nodes on the network. These issues will now be addressed.

The current ADPT processing suite located in NAVSWC's Space and Surface Systems Division is shown in Figure 4. This system was recently separated into two independent networks to accommodate both classified and unclassified operations. Each network consists of a number of desktop workstations, a computer configurer¹ as a network file server, and a mini-super-computer. Also included in each network is a graphics workstation. Not shown in Figure 4, but available to the networks, are a number of detachable hard disk drives and an overhead projection system. The computer hardware suite, which uses 4.3BSD UNIX and System V UNIX operating systems, is interconnected on a local area network (Ethernet IEEE 802.3). Message transfers are accomplished using Transmission Control Protocol/Internet Protocol (TCP/IP) and a specially designed TCP/IP application interface and network

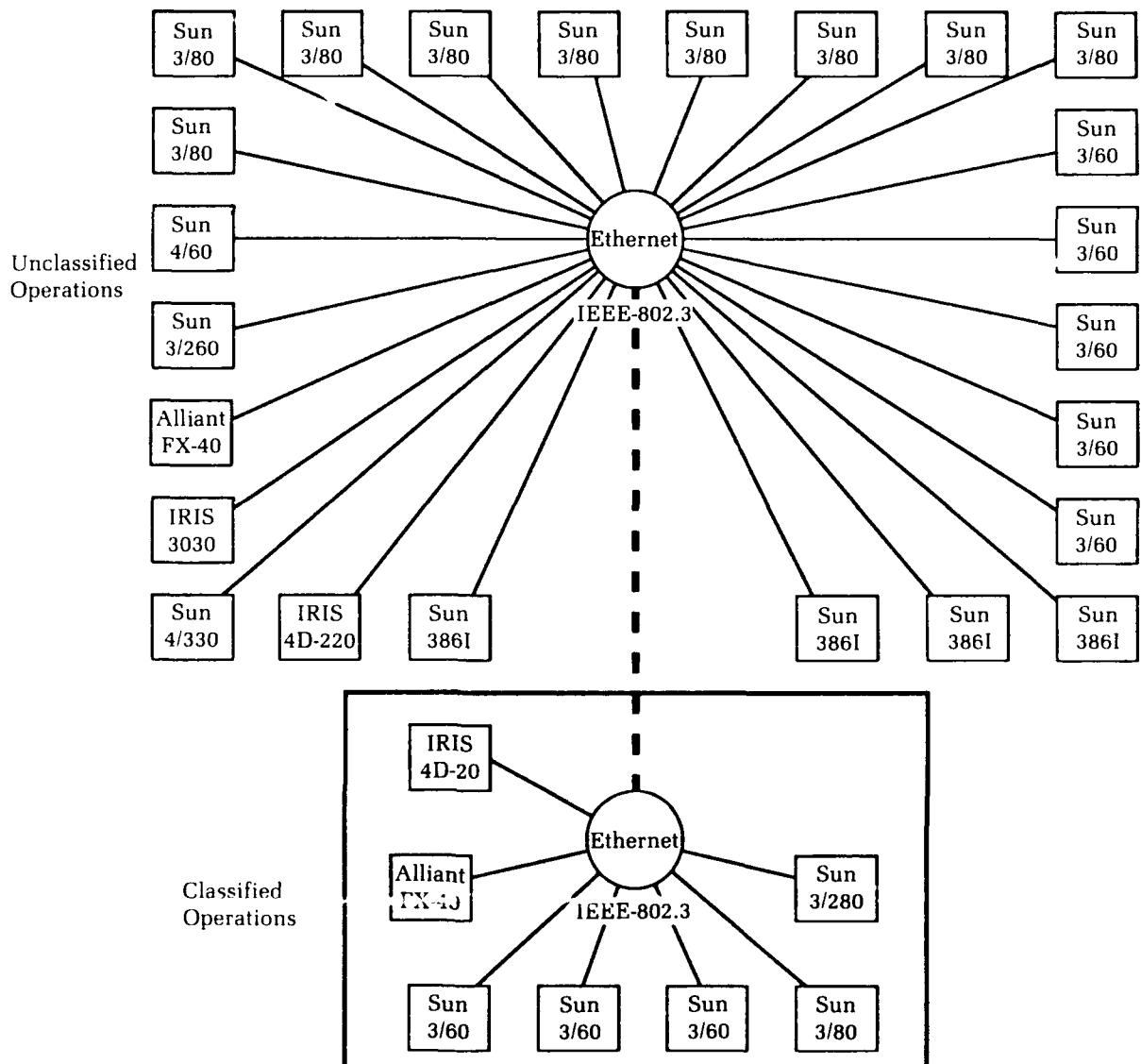


Figure 4. Current ADPT network.

communication controller program called COMTROL.

This program, written in C programming language, was developed solely in the Space and Surface Systems Division. COMTROL is itself distributed. It may be implemented in two layers consisting of a Master COMTROL in the top layer and multiple Slave COMTROL in the second layer. See Figure 5. The two-layered structure was originally developed to circumvent a UNIX limitation that allowed only 64 files to be opened at one time. This meant that only 64 processes could be executed concurrently on the distributed system. With the

36

current configuration, and depending on the size of the application, up to 64 Slave COMTROL can be implemented, and as many as 4096 processes can be executed concurrently.

Each Slave COMTROL maintains a list of all client applications or processes that it serves. The Master COMTROL maintains a list of all connections on the network. These lists are used to support communications or message transfers from any client process to any other client process connected to the system. The various COMTROL and client applications can be hosted on any processor on the network.

If a Slave COMTROL and a process that it serves are hosted on a common processor, communications between them occur within the UNIX domain. If a set of processes and their intercommunications are static, that is, if the processes are always present and their communications requirements do not change over time, COMTROL does not need to be used to transfer data among them. Those message transfers occur directly in the UNIX domain among processes on common machines and point-to-point using TCP/IP for processes on different machines.

The greatest advantage of COMTROL is its ability to manage communications in a dynamic application environment. It was designed to support analyses of a system such as the one in Figure 1, where objects (processes) may be created or removed in an unpredictable manner as the system evolves. When a new process is created, it provides the system with a unique identification when it logs on to the system through COMTROL. It is provided with a unique system address and is then able to communicate with any other process on the system. Similarly, when a process ceases to exist, COMTROL automatically supports its removal from the system.

COMTROL provides a highly capable but simple interface for network communications. Its operation is totally transparent to the system user or programmer. A more detailed description of the ADPT network communication system is given in NSWC TR 91-5.¹

In addition to supporting communication among concurrent processes on a distributed system, it is necessary to synchronize execution of the various processes. This is done in ADPT through the use of blocking communication. Individual processes must be designed so that execution is "blocked" or suspended until a logically needed message is received. A simple

example of this is shown in Figure 6. Both processes in Figure 6 are initialized concurrently. Process 1 begins execution to produce data needed by process 2, which waits to receive that data. When the appropriate data is received, process 2 continues execution, and process 1 suspends execution at its "Receive" until the proper message is passed to it over the network. Clearly, this method of synchronization is application-dependent. Placement in the code of the blocking receive or send instructions depends on the problem being solved. The need for careful top-down software design is evident. Again, the generic functions and interfaces in Figure 3 are helpful in this design process. While the proper use of blocking communications guarantees the logical synchronization of concurrent processes, it is also clear that it can lead to underutilization of some computing resources on the network. Mismatching of process computational intensity and computer power will lead to excessive computer idle time as the processes on some machines wait for needed messages. A general, automatic methodology for optimally assigning processes to computers does not presently exist. Ongoing research in this area includes the investigation of noninvasive system load-monitoring techniques and the development of a first-generation, automatic load-balancing methodology.

To date, the ADPT functions approach and the distributed processing methodology discussed above have been applied to several different problem areas. The first of these was ADMRALS, which will be described in more detail later. The second application includes the integration of generic, space-based sensor assets into the ADMRALS force-level simulation system. NAVSWC has used this system in first-cut performance and value-added analyses for several candidate space-based Wide Area

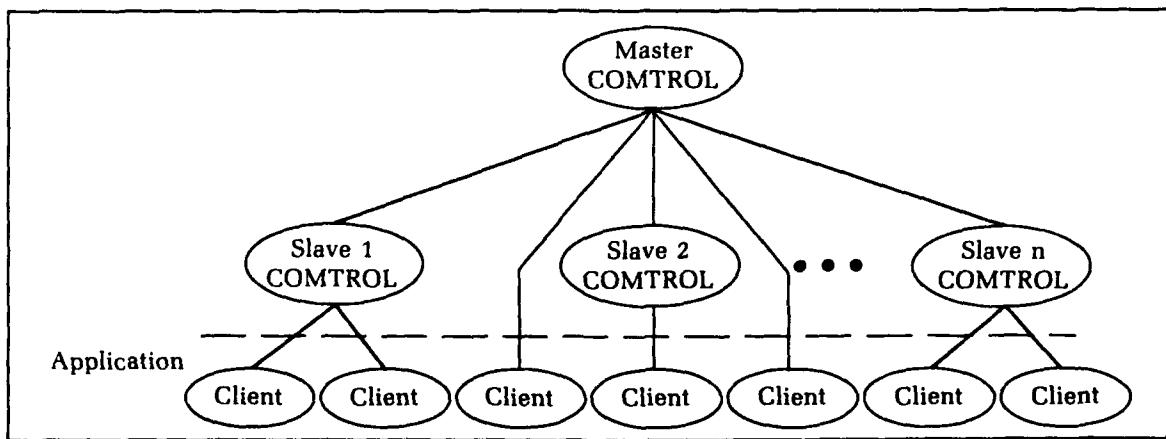
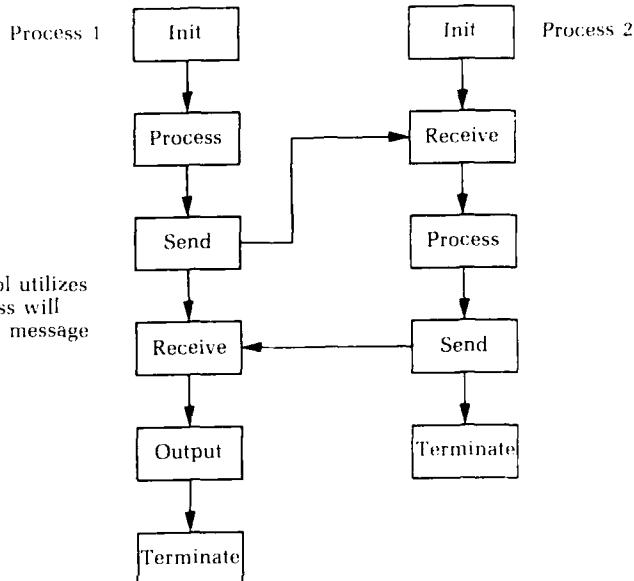


Figure 5. Communication control.

Figure 6. ADPT simulation control utilizes blocking communications. A process will suspend execution until a specified message is received from another process.



Surveillance Tracking and Targeting (WASTT) systems.

A simulation program called AGEM (Anti-Satellite Global Engagement Model) was developed using ADPT. This model was used for extensive system effectiveness, parametric trade-off, and force-on-force analyses in support of the Navy antisatellite development effort of FY89/90. AGEM is currently being modified to support system analysis studies of proposed Navy anti-tactical ballistic missile systems.

ADPT is also being used to design and develop an Advanced AEGIS Weapon System Simulation (AAWSS). The primary focus of this work is the development of a detailed simulation of the AEGIS Weapon System. At the ship level in this application, part of the functional decomposition of the weapon system corresponds to combat system elements that include SPY, Command and Decision, Weapons Control System, and Standard Missile. The interconnectivities among the various elements correspond to a subset of the actual combat system data transfers. Hooks are being provided so that the AAWSS can, in principle, be integrated into the force-level, multi-warfare simulation system, ADMRALS.

38

To date, all the applications of ADPT have fallen in the modeling, simulation, and systems analysis arena. Other possible applications include human-in-the-loop training systems, real-time distributed systems, and the solution of numerically intensive computational mechanics problems.

ADMRALS Overview and Background

In FY86, SPAWAR began its WSA&E initiative. This effort spanned the entire Navy laboratory community. The stated objectives of the WSA&E program included conducting battle force systems engineering and architecture analyses, translating requirements into specifications, and providing systems engineering at the multi-warfare level. Because of the great costs and long-lead times associated with developing, testing, and introducing new weapon systems into the fleet, the need for a force-level, multi-warfare system analysis tool was evident. There did not exist, however, any single group or organization in the Navy with all the knowledge, expertise, and resources necessary to system engineer the fleet at the force level. This was the task assumed by the WSA&E community.

Early in FY87, ADPT was seen as an environment in which models and expertise of various warfare area experts could be integrated into a synergistic, multi-warfare analysis system to support the WSA&E process. The ADPT application called ADMRALS is the result of that analysis environment development effort. ADMRALS is hosted on the distributed processing system shown in Figure 4. Communication and synchronization are maintained through the use of COMTROL and the blocking communication technique (called TIMER for the ADMRALS application) previously discussed. The generic ADPT functions of Figure 3, specialized for applications to warfare systems or objects, are shown in Figure 7. All object models designed specifically for integration into ADMRALS follow the template of Figure 7. Several existing models developed by various agencies for other applications have also been integrated into the ADMRALS environment. In some cases, models have been integrated intact. In other cases, the existing models have been somewhat disassembled along functional lines and the appropriate parts have then been integrated. Models developed specifically for ADMRALS are written in the C-programming language so that they may readily interface with COMTROL, which is also coded

in C. The previously existing models have been written in either PASCAL or FORTRAN. In these cases, C-wrapper programs have been developed to interface the various models with the network.

In its current state, ADMRALS includes relatively sophisticated anti-air warfare (AAW) capabilities (point defense, area defense, and outer-air battle). It includes generic WASTT space assets. It has some antisubmarine warfare (ASW) models and includes limited electronic warfare (EW) capabilities. Table 1 shows the major models included in ADMRALS, their origins, and the programming language in which they are written. Figure 8 shows Blue Force platforms currently modeled. Current ADMRALS models and capabilities will now be discussed.

Point and Area Defense

The first existing model integrated into the ADMRALS distributed environment was the Fleet Requirements Analysis Model (FRAM). FRAM is a time-step-driven, battle group point and area defense model written in FORTRAN. It was originally developed as a sequential program to be executed on a mainframe computer. It originated in and has been used extensively

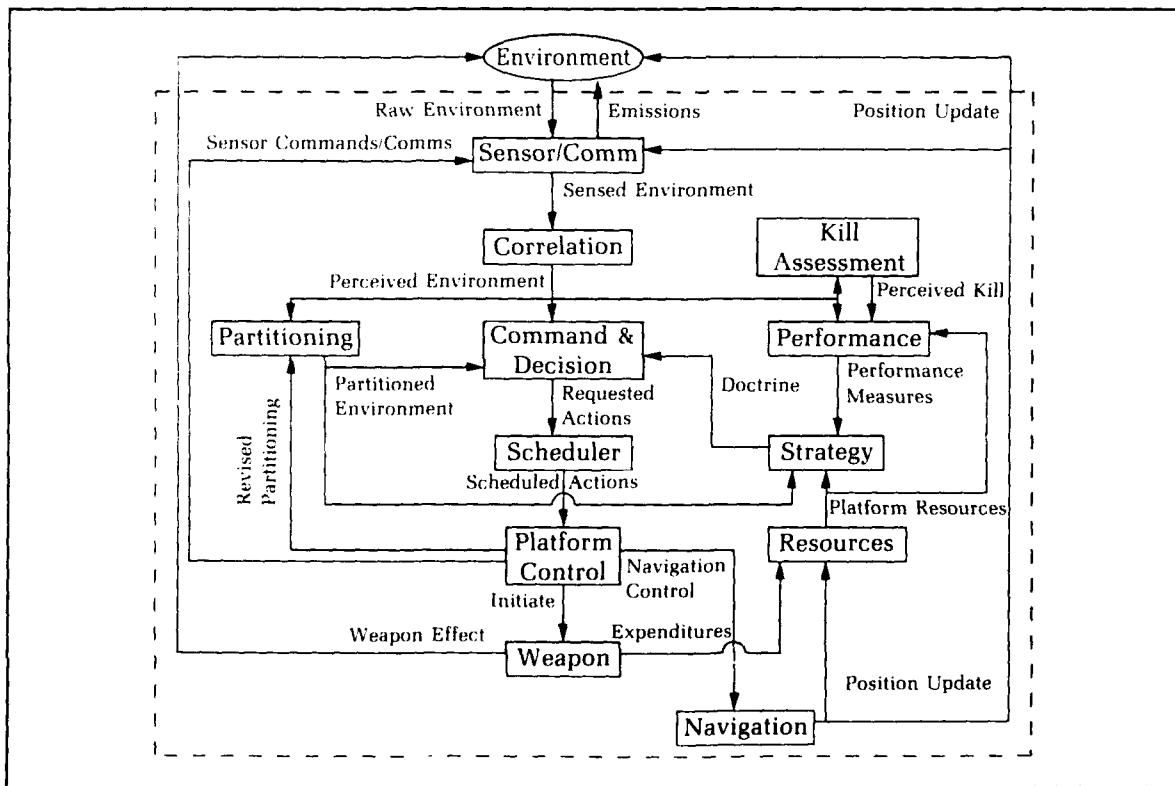


Figure 7. ADMRALS multi-warfare functions.

Table 1. Current Models in ADMRALS

Warfare Area	Model	Origin	Language
AAW(IAB)	FRAM TSMM	NAVSWC Contractor	FORTRAN PASCAL
AAW(OAB)	Fighter E-2 AAWC CV	NAVSWC NAVSWC NAVSWC NAVSWC	C C C C
ASW	RADS HDM SUBSUB ASWC CG/DD	Contractor Contractor Contractor NAVSWC NAVSWC	PASCAL FORTRAN FORTRAN C C
Space	Satellite	NAVSWC	FORTRAN
Red Threat		NAVSWC	C

by the Combat Systems Engineering and Assessment Division at NAVSWC. In early FY87, FRAM was obtained by the ADMRALS/ADPT development team. It was decomposed as much as possible along the functional areas of Figure 7, interfaced with COMTROL and TIMER, and rehosted on a prototype distributed-processing network. A relational data base back end for post-processing analyses and an animated graphics package were also added. FRAM remains in ADMRALS to model point and area defense, but the nature of its implementation on the distributed system has changed somewhat as other warfare area models have been added to the ADMRALS system. For example, the original Red Threat model in FRAM has been removed, and a more robust model has been developed to support point and area defense as well as other warfare areas for which the FRAM threat was never intended.

FRAM provides the ability to model virtually any type and number of Blue Force surface platforms. Platforms are created by entering data specifying their physical descriptions. These data describe platform locations relative to fleet center, types and numbers of sensors, launchers, weapons, illuminators, communication links, the performance characteristics of various platform subsystems, etc. Varying levels of area defense coordination can be specified. Blue Force ship damage and degradation can be modeled. Blue Force decoys and, to some extent, Blue Force EW weapons, can also be included.

40

The Outer Air Battle

In late FY87 and FY88, outer air battle (OAB) models were developed in the Space and Surface Systems Division to expand the AAW capabilities provided by FRAM. These models include an anti-air warfare commander (AAWC), a carrier (CV), an airborne early warning platform (E2), and fighters (F-14 and F/A-18). Each of these models was designed and developed from the top down using the highly modular functional/object-based approach associated with ADPT. A high-level flow chart for the F-14 is given in Figure 9. The correspondence of this F-14 specific model to the generic functions of Figure 7 is evident. The decision-making process, or "pilot," of the F-14 was modeled using both traditional rule bases and a neural network.²

The AAWC model is primarily a decision-making process. Its input from the environment consists of communications from the E-2 models, from space assets, and from an ES-3 model. These communications describe the current tactical situation. Based on this tactical picture, the rule base is exercised to make appropriate decisions. Output from the AAWC to its environment includes orders to the E-2s and the CVs. Actions initiated by the AAWC include establishing combat air patrol patterns for fighters and E-2s. The characteristics of these patterns can be defined by simulation user input. The AAWC also requests launches from the CVs, maintains an inventory of air resources, establishes fighter loiter and

refueling points (called "sponges") and initiates, through the E-2s, fighter attacks on Red air assets.

The CV model provides the launching platform for Blue Force aircraft. User inputs to the CV model include specification of numbers and types of aircraft on board, recycle times for each aircraft type, aircraft weapon types, and loadout and total initial CV consumables loadout. In response to commands from the AAWC, the carrier moves aircraft among various alert states (cold, 30 minutes, 15 minutes, and 5 minutes) and launches aircraft. The model decrements resources as aircraft are loaded and launched. It also includes catapult restrictions for certain types of aircraft (e.g., propeller-driven aircraft cannot be launched from waist catapults).

When the carrier model "launches" an aircraft, a new program or process corresponding to that aircraft is initiated. The new program is automatically assigned to a computer on the distributed network. COMTROL automatically establishes the appropriate communication connectivities with other processes, and the aircraft

operates interactively with other objects in the system. If the aircraft is destroyed during the simulation or when it returns to the carrier, it logs off and is removed from the system. Simulations have been run with four CVs operating concurrently and with as many as 100 aircraft processes created during the simulation.

The E-2 model provides long-range radar detection and acts as a local coordinator for fighter aircraft. The E-2 correlates its own radar contacts with those received from the fighters under its control to obtain a local tactical picture. Given this tactical picture, the E-2 issues vectoring commands to the fighters. The E-2 utilizes a neural network to cluster the targets in order to determine the vectoring scheme. If more fighters are required to meet the threat than are under the E-2's control, they are requested from the AAWC. The E-2 notifies the AAWC when it must return for fuel and when fighters have located enemy aircraft. When an E-2 is being replaced it transfers the control of its fighters to another E-2.

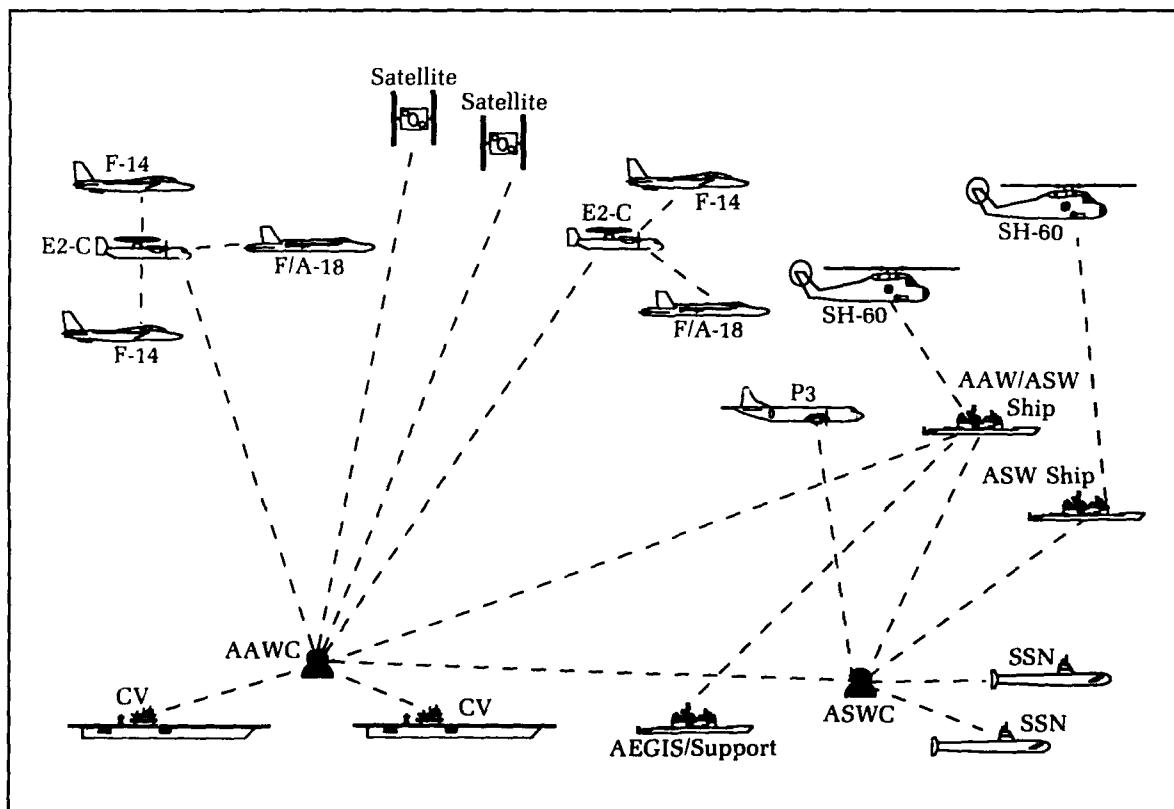


Figure 8. Current ADMRALS simulation capabilities.

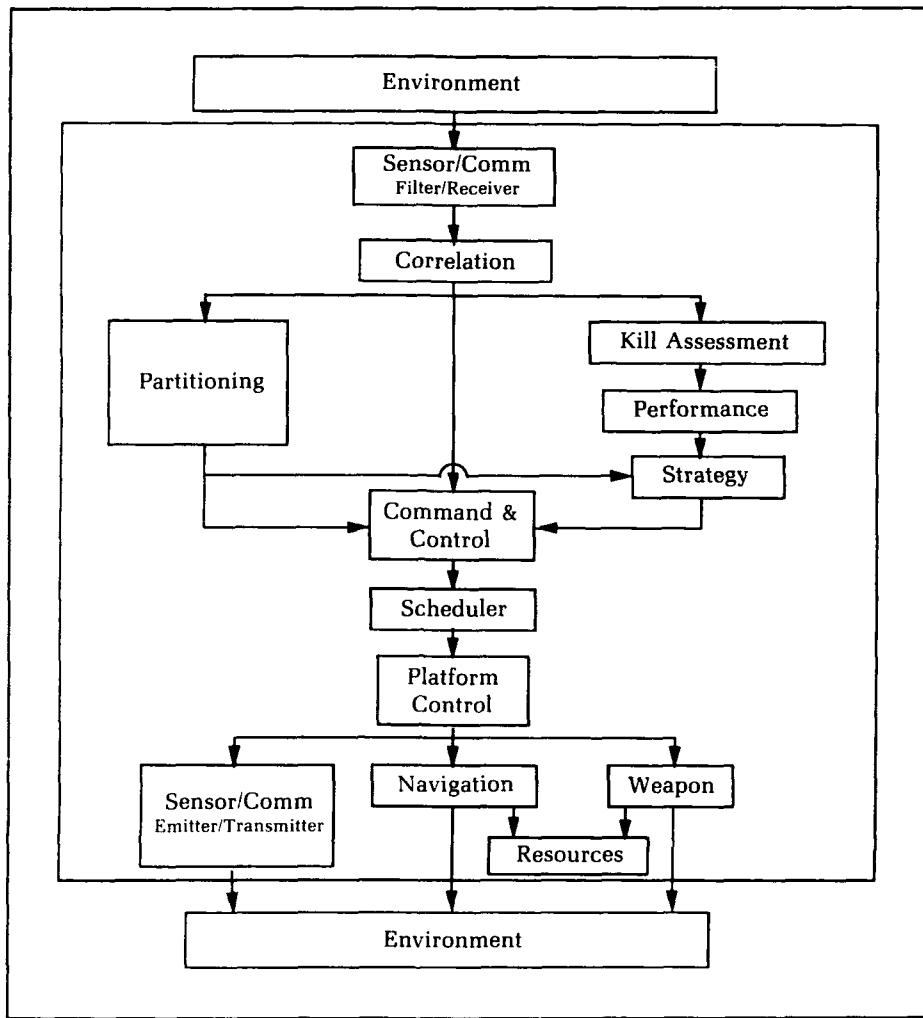


Figure 9. Fighter flow diagram.

The fighter model is used to simulate the F-14s and F/A-18s used in the OAB. The fighter observes the environment with a user-specified combination of its radar, infrared, camera, and visual sensors. If anything is detected, the fighter notifies the E-2. If the fighter has not received a vectoring command from the E-2 and has one or more target tracks, it will fire a weapon at each of the tracks. Otherwise, it will follow the targets assigned by the E-2 and attack when within range. The fighter contains a rule base for target prioritization. For example, a fighter will attack an enemy bomber before a fighter (if both are tracked). The fighter is capable of carrying a variety of weapons, depending on the user's preference. The fighter reports to the E-2 when it must return to the CV for fuel or weapons. The

fighter contains a missile flyout model for Blue Force air-to-air missile engagements. The fighter also contains a rule base to determine chase logic based on engagement geometry.

Space Assets

Space-based WASTT assets can be included in the ADMRALS system at the user's option. Arbitrary satellite constellations (numbers of satellites, spatial distribution, and orbital elements) can be defined. Orbits are propagated using Keplerian motion. Moderately high-fidelity infrared sensor capabilities are modeled for the satellites. Sensor bandwidth, spot size, sensitivity, and search pattern can be specified by the user. Atmospheric attenuation effects are included, but require a prior off-line execu-

tion of a separate program. As currently implemented, when a space-based infrared sensor detects and/or tracks a Red Force threat, it communicates directly to the AAWC, which includes that information in its decision-making process. Communication delays to simulate ground processing of satellite sensor data can be implemented easily.

Anti-Submarine Warfare

The ASW models implemented in ADMRALS were all obtained through the Center for Naval Analyses. These models include HDM (Helicopter Dipping Model), RADS (Rapid Acoustic Detection Simulation), and SUBSUB (Submarine on Submarine). Each of these models was originally developed to simulate a single one-on-one engagement. Consequently, significant additional code had to be developed to navigate the Blue Force platforms and otherwise interface the ASW models with the rest of the ADMRALS system.

HDM is a statistical model that determines the probability of detection of an enemy submarine by a helicopter equipped with a dipping sonar when both platforms are within a given area. It was necessary to create models for the helicopter navigation and torpedo launching in order to include HDM in ADMRALS. It was also necessary to create a cruiser/destroyer (CG/DD) ASW interface model in order to launch/recover/reload the helicopter as well as to give it initial detection bearings on the Red Force submarines.

RADS, a P-3 model similar to HDM, determines the probability of detection of an enemy submarine within a given area. It was necessary to create navigation and torpedo launching models to facilitate the use of RADS.

SUBSUB is a probabilistic submarine-versus-submarine engagement model. It utilizes environmental information and acoustic data from each of the platforms to determine the probabilities of detection and kill for each submarine. It was necessary to create a navigation function for the Blue Force submarines. (Red submarines are navigated by the Red Force Threat Generator.) An ASWC model was created to tie together the different aspects of ASW within ADMRALS. The ASWC communicates with the P-3s and CG/DDs and allocates new ASW assets.

These models give the ADMRALS system limited ASW capabilities. The current degree of sophistication of ASW within ADMRALS, however, is much lower than the current AAW capabilities.

Red Threat

All Blue platforms in ADMRALS are fully reactive. Each Blue Force object utilizes sensor models to collect information about its environment. It uses rule bases or neural nets to make decisions. Then it reacts, according to its rules, on its current perception of the environment. Currently, Red objects are not generally reactive. Red platforms proceed along scripted tracks and initiate actions such as jamming or weapon launch at *a priori* designated times or positions. The Red Threat model offers a great deal of flexibility to the ADMRALS user. It can be used to generate attack profiles for any number of Red Force surface, subsurface, or airborne platforms. These attack profiles can include any number of waves, course changes, speed changes and, in the case of aircraft, altitude changes. As currently implemented, Red on Blue air-to-air engagements are modeled in Blue aircraft since Red is not reactive. When either a Blue or Red platform is destroyed, it is automatically removed from the simulation.

Electronic Warfare

EW modeling in ADMRALS is somewhat limited because the Red Threat is scripted as currently implemented. Future plans include elimination of these limitations through the development of a more reactive Red Threat model.

In the current implementation, Red jamming is modeled by reducing Blue radar detection ranges accordingly. Similarly, Blue jamming is modeled by decreasing Red weapon release ranges. A model has been developed for a generic airborne Blue Electronic Support Measures platform. This platform passively collects track information on attacking Red aircraft and transfers that information to the AAWC.

ADMRALS User Interfaces

Simulating a multi-wave Red attack on a large Battle Force (four or more carriers) presents both simulation set-up and data analysis problems. Several features are included in the ADMRALS System Design (Figure 10) to help manage these problems. A relational data base system is used to provide the user interface to part (FRAM and Red Threat) of the simulation set-up. Other user input for OAB and ASW models is accomplished through edit files. The relational data base also supports the analysis of simulation results. During a typical scenario execution, all data produced can be

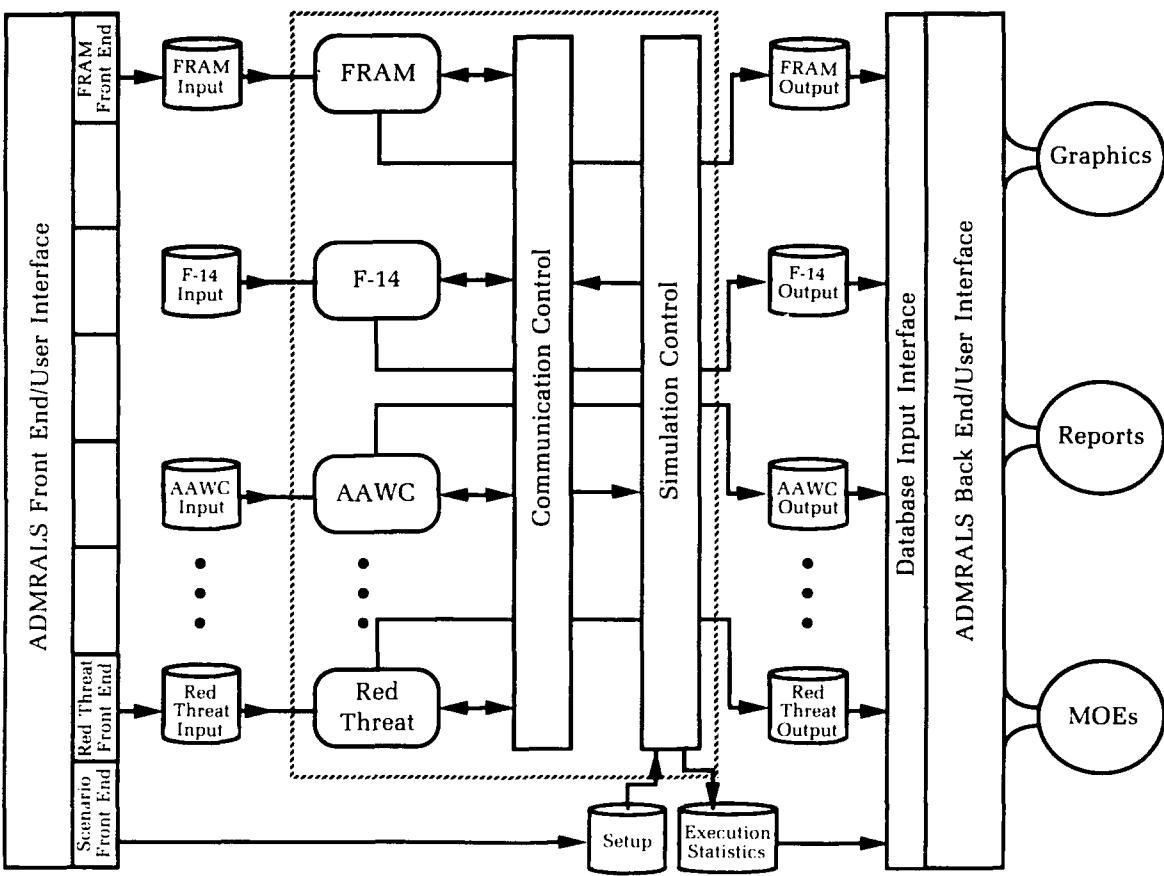


Figure 10. ADMRLS system design.

saved in the various memory locations of the processing network. After the execution, all or any part of this data can be loaded into a relational data base. Various standard post-processing reports or Measure of Effectiveness (MOE) tables can then be produced automatically, or an analyst may use relational data base query language to produce non-standard tables and reports.

Part of the data produced in a simulation includes track files (position versus time) of all Red and Blue platforms and missile trajectories. These track files are input into an animated graphics system that includes color graphics workstations and an overhead, large-screen projection system. This graphics capability has proven to be an invaluable analysis tool. Analysts can observe the simulation and detect nuances for closer examination in a way that would not be possible if limited to hard-copy output.

44

Summary and Future Plans

ADMRLS and ADPT originally began as a single project. As that project matured, it became evident that the highly modular, functional systems engineering and software development methodology and the distributed network communication and control technology had broader utility than the WSA&E application. Considerable effort has gone into formalizing this general methodology. This effort is continuing. The approach and the distributed network have now been used for several other applications. More research is needed in various aspects of distributed processing. Network performance in terms of execution time has been less than anticipated. The main reason for this appears to be non-optimal assignment of processes to computer resources (load balancing). Several potential improvements in this area have been identified, and

efforts are under way to implement them. Additionally, future research is needed in nonintrusive system monitoring and state-of-the-art methods for distributed network process synchronization and communication.

The greatest strengths of the ADMRALS application are AAW and space asset modeling capabilities. Future plans include development of more sophisticated EW capabilities and the addition of other warfare areas. Ongoing efforts include the development of a more reactive threat model to support enhanced EW and Anti-Surface Warfare. A model interface requirements document is being written so that other organizations may develop object or platform models that will readily interface with the ADMRALS system.

Acknowledgments

Many individuals have contributed to the development of ADPT and ADMRALS. The original concept of generic functional decomposition of complex systems was due to Dr. A. Dan Parks of NAVSWC's Space Sciences Branch. Albert P. Shuler, also from this branch, assisted in defining the specific function areas and was instrumental in the early distributed processing research. The COMTROL network communication package was developed by Dr. L. Talsma of the Physical Sciences Software Branch. Ron J. Guidry and Emanuel E. Skamangas made major contributions to efforts formalizing ADPT methodology, and they have been largely responsible for the implementation of new and existing models in the ADMRALS environment. Mr. Guidry and Mr. Skamangas also made significant contributions to the preparation of this paper.

References

1. Talsma, Don L., TARDIS—A Message-Based Networking Package for ADPT, NSWC TR 91-05, Dahlgren, VA, Jan 1991.
2. Rogers, G., Solka, J. L., and Steffen, D., Neural Network Implementation of an F-14 Battle Management Fusion Algorithm Rule Base, NSWC TR 90-169, Jun 1990.

The Author



JEFFREY N. BLANTON is currently the Senior Technical Advisor for the Space and Surface Systems Division, where he also serves as Program Manager for the various ADPT/ADMRALS-related projects. He received his professional education as an aerospace engineer at the University of Virginia. He earned his Bachelor's degree in 1968, his Master's in 1971, and Ph.D. in 1976. Dr. Blanton has been at the Naval Surface

Warfare Center since August, 1976. He served as Head of the Space Sciences Branch from 1985 to 1990. His major interests include estimation theory, space mechanics, and satellite attitude dynamics. While at NAVSWC, Dr. Blanton has received numerous performance awards and letters of appreciation. In 1981 he received the Defense Mapping Agency Research and Development Award. He is a member of AIAA, the Virginia Academy of Sciences, Theta Tau, Tau Beta Pi and Sigma Xi.

A Chemical Warfare Naval Simulation Model for Surface Ships

Thomas James Yoncha

The Naval Surface Warfare Center (NAVSWC) has developed a computer software system, the Chemical Warfare Naval Simulation (CWNAVSIM), to simulate the impact of a chemical agent attack upon a surface ship. This system will simulate a point or line source weapon at any point in space and track the liquid, primary vapor, and secondary vapor to a target (ship). It will then deposit liquid, chemical warfare (CW) agent on the surfaces of the ship and subsequently evaporate this liquid. The vapor concentrations at the ship ventilation inlets are calculated as a function of time and transported throughout the ship's internal compartments. For each compartment, a time-history of concentration and dosage is calculated. Computer graphics are output depicting the liquid cloud enveloping the ship and depositing agent upon the ship. A typical chemical attack simulation clearly shows how a ship's fighting effectiveness can be seriously degraded from dosages to compartments, causing casualties, replacement of casualties with less trained personnel, and shutdown of vital systems within the ship.

Introduction

Recent experience in the Persian Gulf clearly demonstrates the possibility of U.S. naval ships coming under chemical attack during regional conflicts. Design of future ships for survivability and improved survivability of current ships will become a necessity. In this regard, CWNAVSIM is a valuable tool to the U.S. Navy in that it can predict the effect of changes in ship design to enhance survivability at very low cost. Physical testing of similar design changes would be very costly or impossible to accomplish. This simulation can determine the best location to install CW agent detectors, ship entry/exits (air locks), and the filter capacity needed on shipboard collective protective systems to survive CW attacks. It can determine, in an operational sense, if (and how much of) a ship must be decontaminated after an attack. As a case in point, CWNAVSIM was used to predict the results of a CW attack on a naval ship in an elevated temperature environment during the Persian Gulf War. It was also used to determine the effect of a biological attack in the same scenario.

One scenario of concern was that one of our ships on patrol might be fired upon from a shore battery. In a typical incident of this type, a shell containing a CW agent might explode about 20 meters in front of the destroyer and 50 meters up in the air. The CW agent, which initially is in liquid form, is dispersed into a cloud of liquid droplets by the explosion. At this point several processes begin simultaneously. First, the liquid droplets begin to fall like rain, with the heavier droplets falling fastest. Second, since the CW agent is volatile, the droplets also begin to vaporize, forming what is called the primary vapor cloud, which begins to flow with the local air currents. Some fraction of the liquid droplets falls into the ocean and ceases to be part of the problem. The remaining fraction of the droplets falls as a liquid onto the ship's decks and other exposed surfaces. These droplets spread out on the ship's surfaces and also begin to evaporate, forming what is termed the secondary vapor cloud. The primary and

secondary vapor clouds both move with the prevailing local winds and, depending upon ship's speed and heading, some fraction of the vapor in these clouds enters the ship's interior through various ventilation air intakes. Once inside the ship's air handling systems, the CW agent begins to affect the crew members. As crew members become incapacitated, the ship's performance begins to degrade in predictable modes. This entire process is modeled in CWNAVSIM and discussed in detail in this article.

Overview of the Simulation

CWNAVSIM is a series of computer programs used in simulating the transportation, evaporation, liquid deposition, weathering effects, and mission impact of a chemical agent attack on a naval ship. The simulation actually consists of three major submodels: Deposition and Weathering of a Chemical Attack on a Naval Vessel (DAWN), the Ventilation Model (VENM), and the Naval Unit Resiliency Model (NURA). Each of these comprises a group of programs. The entire simulation consists of modules that can be replaced or modified as needed for: (1) data base generation; (2) cloud tracking; (3) agent deposition; (4) agent evaporation (weathering); (5) vapor generation; (6) vapor tracking; (7) vent history; (8) graphics display; (9) internal ventilation transportation; and (10) mission degradation. Figure 1 is a high-level program flow chart. Figure 2 is a companion flow chart that provides more detail. Some of the abbreviations in Figure 2 are: OPSIT for Operational Situation, OTE for

Operational Threat Environment, HVAC for Heating, Ventilation and Air Conditioning system, and Ship OPS for Shipboard Operations.

The CWNAVSIM is currently running on a Tektronix XD88/30 computer. This computer is a RISC architecture system which performs about 2.2 single-precision MFLOPS. The operating system is a Tektronix version of UNIX called UTEK, V3.2.E. The programs were compiled with XD88 series C and F77 compilers, both by Green Hills Software. Graphic programs use the Tektronix OnRamp 3D Graphics Library.

Preparation of Ship Data Bases

Referring to Figures 3 and 4, it can be seen that ships, as seen by the computer and portrayed in graphics, are made up of a large number of quadrilaterals referred to here as quads. The data bases contain the real-world coordinates, normals, etc., of these quads. Presently, data bases for the FFG, DD-963, LST-1179, LHA, and CG-26 class ships are available for use.

Quantification of Flow Around the Ship

The quad data base is used in all of the subprograms in CWNAVSIM, including a potential flow model¹ that determines the characteristics of the airflow about the ship's surfaces. The characteristics of the flow about these surfaces are maintained in a data file for use in estimating evaporation from the surfaces. Given an arbitrary initial flow about a vessel, the relative wind speed at the surface of each quad

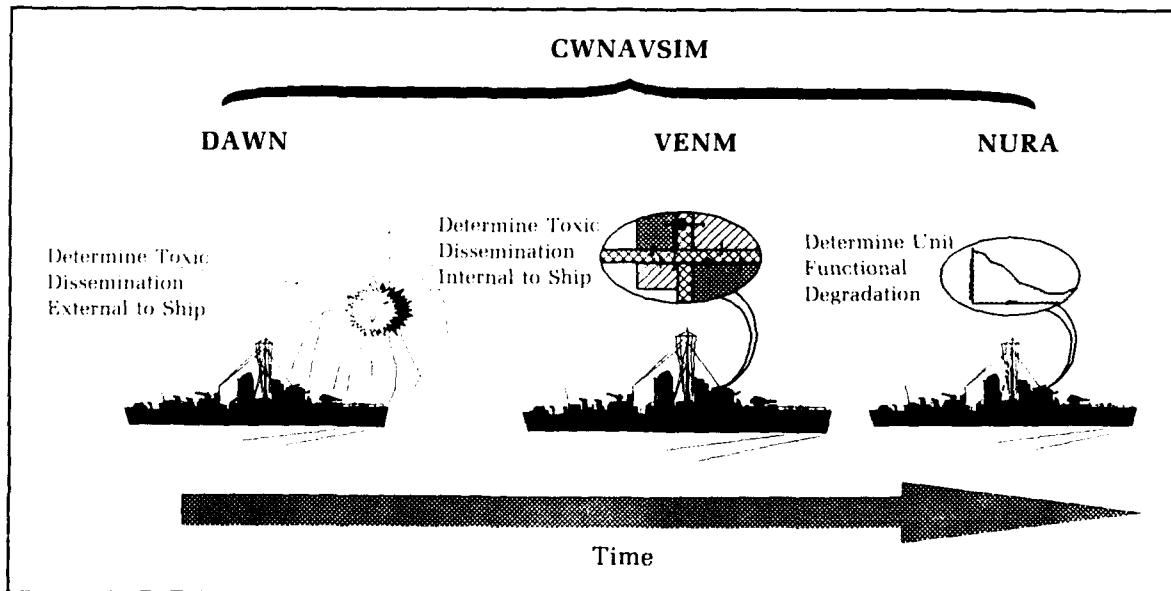


Figure 1. Program flow.

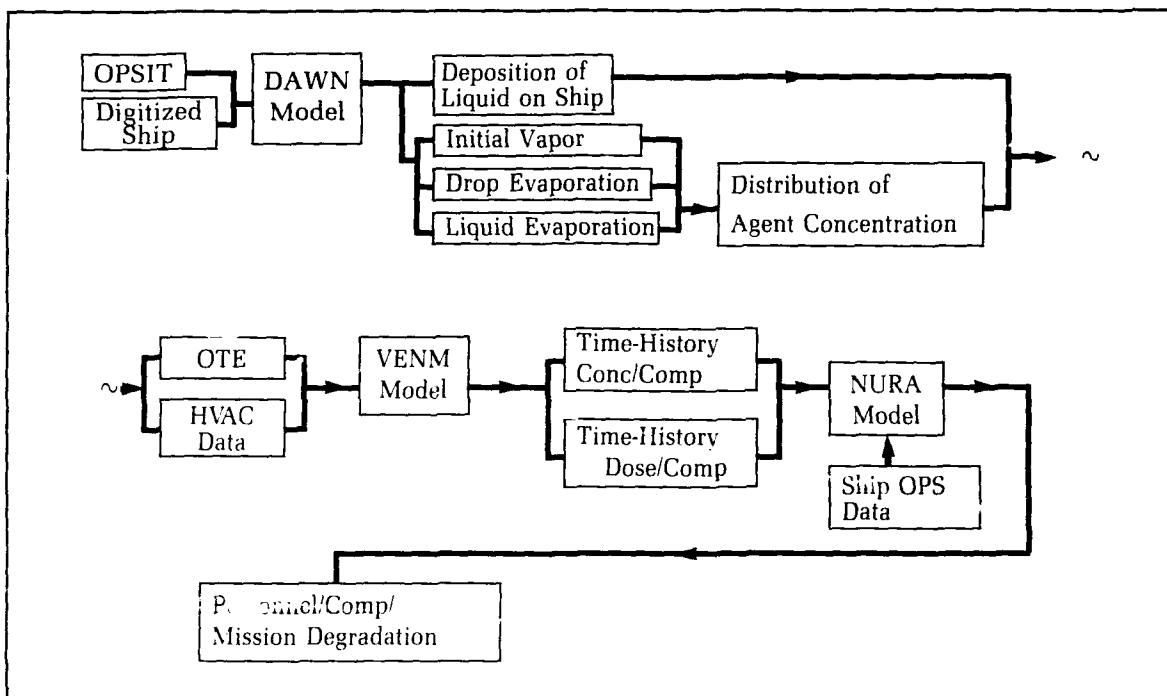


Figure 2. Description of program flow.

is determined. The absolute wind flow at these points is used to determine the evaporation rate of a CW agent from the ship's surfaces.

Since the velocity components of an onset flow may be different with each cloud track, the actual surface velocities may be calculated quickly, given the onset flow components and the potential flow coefficients at the center of each surface quad.

Description of the Agent Burst

The simulation maintains a data file which is used by a program module to track the movement of the agent cloud. This includes:

- agent type GB (nerve agent), GD (nerve agent), HD (distilled mustard), VX (nerve agent), AC (hydrogen cyanide)
- agent temperature (celsius)
- agent mass (as vapor)
- mass medium diameter of drops (log normal distribution)
- Litchfield slope of droplet diameter distribution (standard geometric deviation of the radii of particles in the distribution. When this value is 1.0, all the particles are the same size.)
- line source length in meters (for line source only)
- fall angle from horizontal (for line source only)

- line angle from wind (degrees) (for line source only)
- release height (meters)
- wind speed and direction
- atmospheric pressure (mm Hg)
- atmospheric temperature (Celsius)
- temperature profile slope value (deg C/m)
- horizontal cloud sigma at 100 meters
- vertical cloud sigma at 100 meters from release point
- horizontal cloud sigma at release point
- vertical cloud sigma at release point

An agent data file is also maintained for input to the weathering program. These parameters include: surface type (epoxy, alkyd, urethane), surface temperature (celsius), droplet spread factor (the size of the circle that a droplet spreads out to upon impact with a surface), target emission rate (mg/sq m/min) (the rate of emission of vapor agent from the entire target), and desorption rate (mg/sq m/min) (the time rate of desorption of a liquid agent from a desorbing surface).

Tracking of Initial Agent Clouds

A cloud-tracking module has been developed that will track and quantitatively describe the atmospheric transport and diffusion of a chemical contaminant through a navy ship.^{2,3,4}

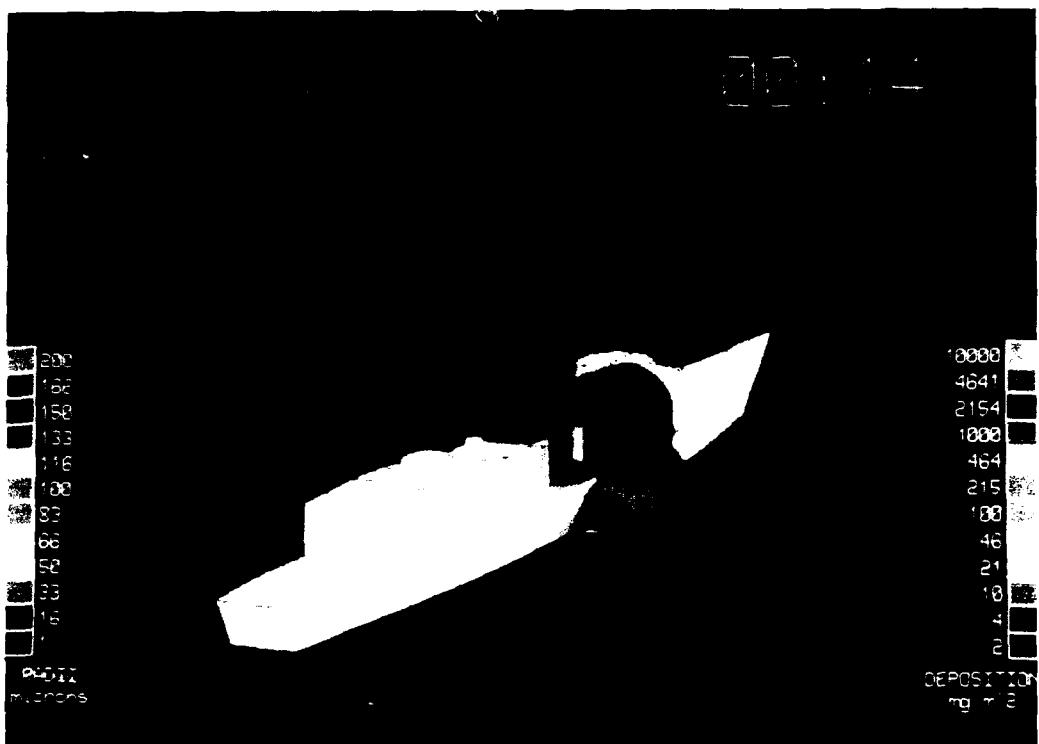


Figure 3. Typical small weapon size chemical warfare attack.



Figure 4. Typical large weapon size chemical warfare attack.

The initial aerosol cloud is regarded as a collection of 50 envelopes (cloudlets), each of which contains droplets spanning a small range in size. By making use of the droplet size distribution, each range of droplet size is replaced by an average size, suitably defined. For each time step, the radius of the droplets is calculated along with the speed, vertical position and velocity, and size of the associated envelope. When all the envelopes have been processed, the vapor distribution of the aerosol cloud is calculated. The droplet temperature is calculated at the end of each time step, along with all temperature-dependent parameters. A time-tagged data record for each cloud envelope (cloudlet) is included in the standard output. The cycle is then repeated until either the time of interest has been reached or the entire droplet mass is no longer airborne. A more complete description of the method and basic flow used in the program may be found in the references.

The program is also used to calculate the final cloud track. The final cloud track file contains the quantitative cloudlet information and the three-dimensional positions of the cloudlets as they interact with the airflow about the ship.

Agent release parameters, which include wind speed and direction, ship speed, and agent release point, are input by the user. A preliminary track for each cloudlet is calculated. Once a cloudlet enters the flow boundary around the ship, the preliminary cloudlet information, including quantitative droplet data and a final cloud track, are produced. The XYZ positions of the chemical cloud envelopes are computed at select times as the envelopes interact with the airflow about the ship. The XYZ potential flow model is used to compute positions on streamlines that do not touch the ship's surfaces. In order to speed processing, several features were added to the program.

- The flow boundary around the ship body is defined as an area extending outward (as a function of cloud radius) from all ship semi-axis limits. All positions within this area are tracked along flow streamlines. All positions outside this area are tracked using initial flow and fall rate. This method eliminates the need to compute streamline velocities at points where the interaction of the ship with the airflow is negligible.
- All velocity calculations include the fall rate of each cloudlet.
- Data on cloudlet position is used to determine whether it is inside or outside the flow boundary, or if it has been removed from the flow.

- Subclouds are removed from the flow when a subcloud is deposited upon the ocean surface or when it enters the ship's body.

Determination of Agent Deposition on the Ship

The portion of the liquid cloud that is eventually deposited on the ship's surfaces is determined next. Each cloud containing liquid is subdivided into a large number of cubes. The motion of center point of each cube is assumed to represent the motion of the entire cube. The dimensions of the cube are known, so the motion of the volume enclosed by the cube is known as well. The XYZ potential flow model discussed above gives the motion of the cubes relative to the quads that describe the ship's surfaces. The amount of agent deposited is determined by finding the fractional volume of the cube that is swept out as it moves past each surface quad.

Determination of Absorption, Evaporation, and Desorption

The next step is to predict the persistence of and the vapor emission from chemical agent droplets on a moving naval vessel considering surface temperatures and the wind flow about the vessel. The evaporation or weathering module is an adaptation of the model to predict droplet persistence on a moving vehicle.³ The evaporation/absorption model selected for use in CWNAVSIM is the constant radius of curvature (CRC) model for monodispersed droplets. The CRC Model considers the evaporation/absorption history of a single droplet and applies a constant factor to account for the total deposition within the area. In order to use this model without major modification, the deposition is assumed to be uniform over each area considered. Evaporation and absorption quantities for the total vehicle are determined by summing across the different areas.

The desorption process is assumed to commence after all liquid on the surface is either evaporated or absorbed. Desorption data is a user input. This input is in the form of desorption rate per unit area as a function of elapsed time for surface dryness. In the computer model, desorption of each area is considered separately, and quantities for the total vehicle are determined by summing across the different areas.

The CRC Model also includes routines to determine the decay and the removal by pick-up of liquid deposited on the surface. As neither process is important for this naval

application, these routines were omitted when the CRC model was adopted for use in the weathering model in CWNAVSIM. Several additional simplifying assumptions were made: (a) liquid droplets are deposited without overlap on a uniform smooth plane; (b) the relative wind velocity vectors remain constant throughout the weathering process and are the same as the tracking velocity vectors; (c) relative wind velocities over vehicle surfaces are determined by using the flow coefficients at the center of each area to calculate the absolute velocity over each area; (d) the temperature remains constant during the entire weathering process in each quad; (e) liquid droplet temperature is the same as that of the surface to which the droplet is attached; (f) the evaporation, absorption, and desorption processes in each area are assumed to be independent of these processes in adjacent areas; (g) all surfaces are assumed dry before liquid is deposited; (h) all surfaces are assumed to be smooth and of the same composition (e.g., alkyd paint on metal, urethane on metal, etc.). Droplet behavior in cracks, crevices, and greased areas is not considered.

For each area, the program outputs the initial deposition and a time history of drop height, evaporation rate, absorption rate, and desorption rate.

Tracking the Primary Vapor

Knowing the final liquid cloud track, the positions, dimensions, concentration, and generation times of a series of primary vapor clouds are determined. As time progresses, the droplets defining a liquid cloud evaporate. These droplets leave behind a primary vapor trail. The amount of vapor in the clouds is estimated by calculating the change in mass of a cloudlet (all the droplets) over a period of time. This mass difference is the total mass contained within the vapor cloud. The volume of the vapor cloud during this time period is equal to the volume swept out by the moving cloudlet. This volume is divided into discrete units, and the generation time and position of each unit are calculated. These discrete units are termed the primary vapor cloudlets. The XYZ positions of the cloudlets are tracked by the potential flow model to determine how they interact with the ship.

Tracking the Secondary Vapor

Given the agent mass change rate per unit time of a liquid agent on a defined surface, secondary vapor clouds are created over time to approximate the evaporation process. Given that the surfaces upon which an agent is

deposited may be quite large (>100 square meters), the simulation breaks each contaminated surface into a grid, where all grid areas are equal. The center of each of these plane grid sections is considered the starting point of an associated secondary vapor cloud. Once the surface position of each vapor cloud is known, the height above the surface through which the cloud is distributed is computed. Using the turbulent boundary layer theory, the boundary layer thickness is computed, and the height of the vapor cloud is assigned this value.

In much the same method used with the primary vapor, the change in mass due to evaporation of the liquid from a surface is found per unit time. This mass difference becomes the mass of the vapor confined in the secondary vapor cloud. The cloud volume is determined by the area of the grid section and the height of the boundary layer. Using the potential flow model, the XYZ positions of the secondary vapor clouds are then computed at select times as the envelopes interact with the airflow about the ship.

Final Output Files

Given the positions of all primary and secondary vapor clouds for all times and all vent positions on a ship, the time at which any vapor cloud intersects a vent is calculated. Since the vapor positions are along a streamline of a single vapor cloud, and the input is a series of these streamlines or tracks, an equation of the line joining two points on a track is found. This track line segment (representing the motion of the center of the vapor cloud) is checked against all vent positions. If this line is within a vapor cloud's radial distance from a vent, then an intersection is said to have occurred. Using geometric analysis, the time that the vapor cloud first intersects the vent and the time that it ceases to intersect the vent are found. From this we may determine the time duration of exposure of the vent to the vapor cloud. The data from the primary and secondary cloud interactions with the vents are merged and the concentrations are added whenever more than one cloud is involved. The data is sorted by vent identification and by time (first to last) to be input to the VENM model.

Graphics Production

An executive program which is mouse driven allows a user to enter a graphic label, determine the identification of various quads on the target, and show the track of a liquid cloud. A program called Depshow.c generates a three-dimensional display of the agent as it is

deposited on the vessel. A program called Evapshow.c generates a three-dimensional display of the agent as it evaporates off the vessel surface. Finally, a program called Cloudshow.c generates a three-dimensional display of the agent cloud track as it interacts with the vessel while showing the deposition and evaporation of the agent. Liquid clouds are displayed as shaded images colored to represent agent particle diameter. Ship areas are colored to represent the amount of liquid agent deposited upon surfaces. Figure 3 is an example of the graphical output. A legend on the left describes liquid cloud aerosol radii, while the right legend describes liquid agent deposition upon the ship. This example shows a small CW weapon attack. The sea surface grid is 100 x 100 feet. Figure 4 is an example of a large weapon CW attack. The graphics in these figures visually depict the results from CWNAVSIM. They are extremely useful in determining ship design parameters and operational doctrine as previously discussed. These graphics amplify the effect of the tabular data output from CWNAVSIM and allow an analyst to see the effect of CW attacks quickly.

Determination of Internal Agent Concentrations

The VENM⁶ was developed by combining appropriate portions of an existing NAVSWC model (SHPPIN) and a McLean Research Center model (kinetic flow model). The VENM module has three operating modes; DAWN interface, homogeneous cloud traverse (HCT), and penetrating hit (PH). All modes output a file to transfer data to the NURA module.

In the DAWN interface mode, VENM is driven by input from the DAWN module, i.e., the target is under attack by an external point or line source. In the HCT mode, VENM is operated in a stand-alone manner and simulates the target ship traversing a homogeneous vapor agent cloud in which all ventilation inlets see the same concentration for a specific time. In the PH mode, VENM is also operated as a stand-alone module, but now simulates a weapon penetrating the ship before functioning. Operating in the PH mode requires the user to estimate physical damage caused by a penetrating hit, since this is not calculated by VENM. Agent can enter the ship through two routes: through ventilation inlets (both from external vapor and vapor exhausted from a ship's ventilation exhaust system), and through penetrating hit-breach areas. Agent concentration in breached areas is considered to be the sum of vapor created by the initial functioning of the weapon and vapor created by evaporation, with time, of liquid released by the

52 external vapor and vapor exhausted from a ship's ventilation exhaust system), and through penetrating hit-breach areas. Agent concentration in breached areas is considered to be the sum of vapor created by the initial functioning of the weapon and vapor created by evaporation, with time, of liquid released by the

weapon. Two types of agent vapor plumes are calculated: a plume flowing out the breached area and plumes flowing out ventilation exhaust vents. All plume geometry is considered Gaussian and follows laminar flow on the external surface of the ship. No agent loss to sea spray or condensation (AC cooling devices) is calculated. Ventilation system material conditions are used during transportation calculations.

Determination of Mission Degradation

NURA is an adaptation of the Army unit resiliency model (AURA) developed by the U.S. Army's Ballistic Research Laboratory. The AURA detailed description is left to the original documents.⁷ The difference between AURA and NURA is in the data bases. The AURA model assumes a two-dimensional battlefield with assets distributed as an army unit. The NURA module assumes an array of ship compartments with assets distributed as a naval unit.

Typical Results

The CWNAVSIM yields both graphical and tabular output. Previous sections detailed the graphical output, which dealt with liquid aerosol clouds and liquid deposition; data is also available in a tabular format. Many megabytes of tabular data are output, including agent vapor and liquid data, dosage data outside and inside the ship, concentrations outside and inside the ship, dosage by compartment, casualty data, and ship mission degradation data. Figure 5 illustrates the degradation to the

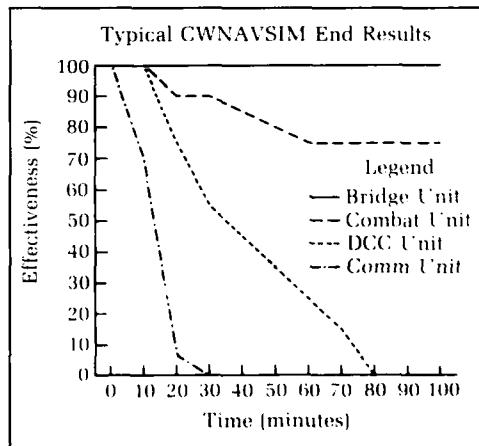


Figure 5. Degradation to ship's functional departments caused by CW attack.

ship's functional departments caused by the CW attack. This degradation includes the effect of CW casualties, the degradation caused by replacement of experienced casualties with less experienced personnel, and shutting down various ship systems due to manpower attrition. It can be seen that the attack can have serious, degrading effects with time, particularly in the communication unit. This is probably due to casualties to personnel, particularly experienced people needed to carry out vital command functions. While not as significant, the ship's ability to conduct combat operations is still seriously degraded. It is clear that current and future ships must be designed for survivability against chemical and biological attacks.

Summary

The CWNAVSIM will track liquid and vapor CW agents from the point of functioning to a naval target, deposit and evaporate liquid agent, and assess the degradation of ship's effectiveness. Benefits to the Navy from using the CWNAVSIM include the ability to quickly and cheaply determine the effects of CW attacks. It allows design engineers to determine the best location for CW detectors and airlocks without laborious and costly physical ship modifications and testing. An engineer can also size the filter needed for collective protective systems. In an operational sense, whether to decontaminate the ship at all, or parts of it, can be determined by shipboard personnel if they can run this simulation on the ship. Shipboard personnel could also determine when, and which, personnel should don what kind of protective equipment. CWNAVSIM may also be used as a CW training aid, or, in non-CW applications, it can track friendly missile and gunfire exhaust, smoke plumes, and toxic spill plumes.

References

1. Dawson, C., *The XYZ Potential Flow Program*, Report 3892, Naval Ship Research and Development Center, Bethesda, MD, Jun 1972.
2. Saucier, R., *A Mathematical Model of the Atmospheric Transport and Diffusion of a Chemical Contaminant*, Technical Report ARCSI-TR-81071, U.S. Army Armament Research and Development Command, Chemical Systems Laboratory, Aberdeen Proving Ground, MD, Nov 1981.
3. Saucier, R., *A Mathematical Model of the Vapor and Mass Distribution from a Falling, Evaporating Aerosol Cloud*, Technical Report ARCSL-TR-81007, U.S. Army Armament Research and Development Command, Chemical Systems Laboratory, Aberdeen Proving Ground, MD, Mar 1981.
4. Saucier, R., *NUSSE3, Model Description*, Technical Report CRDEC-TR-87046, U.S. Army Armament Munitions Chemical Command, Chemical Research, Development & Engineering Center, Aberdeen Proving Ground, MD, May 1987.
5. Winkler, R., Imhoff, S., *Computer Program for the Prediction of Droplet Persistence on a Moving Vehicle-Weathering Model (VEHW)*, Contractor Report ARCSL-CR-83013, U.S. Army Armament Research and Development Command, Chemical Systems Laboratory, Aberdeen Proving Ground, MD, Jan 1983.
6. Blacksten, H. R., Yencha, T. J., *VENM Maintenance Manual*, DPG Document No. DPG/TA-90/003B, Technical Analysis and Information Office, U.S. Army Dugway Proving Ground, Dugway, UT, Nov 1989.
7. Klopčić, J., Roach, L., *An Introduction to the Use of the Army Unit Resiliency Analysis (AURA) Methodology: Volume I*, Memorandum Report BRL-MR-3384, U.S. Army Ballistic Research Laboratory, Aberdeen Proving Ground, MD, Sep 1984.

The Author



THOMAS JAMES YENCHA is a project leader and senior physicist in the Chemical Systems Branch of the Naval Surface Warfare Center. He received a B.S. degree in physics in 1972 and a M.S. degree in physics in 1974 from Wilkes University in Wilkes-Barre, Pennsylvania. The author joined NAVSWC in July of 1974 and, after a job rotation period, became a member of the Chemical Systems Branch. Over the years, he has performed as a naval environmental pollution programmer, a source emission test team leader, chemical warfare detector test engineer, and a chemical warfare simulation software development engineer. Mr. Yencha has received several NAVSWC Sustained Superior Performance Awards. He is an active member of the Military Operations Research Society (MORS) and has presented papers at MORS Symposia at the U.S. Naval Academy in 1989, at the U.S. Air Force Academy in 1985, and at Kirtland AFB in 1982. He also presented technical papers at CRDEC Modeling Conferences in 1984 and 1985.

A Standardized Approach for Implementing Fiber Optics in Navy Surface Warfare Systems

D. R. Knudsen, G. D. Brown and J. P. Ingold

In response to increasingly stringent requirements, technology advances, and direction from the Chief of Naval Operations, the Navy is beginning to use fiber-optic technology in surface warfare systems. This article addresses the particular fiber-optic technology that the Navy, through the Naval Sea Systems Command's Fiber Optic Program Office, is standardizing for use. The article stresses the importance of using standardized components and design guidelines and describes the Naval Surface Warfare Center's (NAVSWC's) contribution to that effort. One of the key benefits of fiber optics is a reduction in component types. It is important, therefore, that all personnel involved in the development of Navy surface warfare systems be aware of these standards so this benefit can be realized.

Introduction

Fiber-optic technology in conjunction with other new technologies is beginning to be incorporated into communications and control in Navy surface warfare systems. Increasingly stringent requirements provide a continuous need for quick reaction to multiple threats in severe environments. These requirements dictate highly reliable and survivable data communication systems. Fiber-optic technology can meet the data communication needs demanded by these requirements. The Chief of Naval Operations has therefore directed the use of fiber optics in new systems where benefits can be realized. The technology has been proven and is ready to be engineered into shipboard systems.

Fiber optics offer many benefits for naval surface combatants.^{1,2} The use of fiber optics allows higher data transmission rates, reduces electromagnetic interference (EMI), reduces cable weight and space requirements, allows systems designers greater flexibility in the design of systems, and significantly reduces the numbers and types of components needed for logistics support of installed systems.

To introduce this technology efficiently to surface warfare systems and achieve the maximum benefits, standardized components and design guidelines must be used. If different components are used on different systems, the component and logistic costs will increase and subsequently reliability, maintainability, and availability requirements may not be met.

Data Communications Requirements

Surface warfare systems have requirements for widely varying signal types, including voice, video, and data. Each of these can be either analog or digital. These signals can all be handled by the same type of optical fiber so long as electrical power is not carried along with the signal. This characteristic is a tremendous logistic benefit, but achievable only if fiber optics are implemented in a standard manner.

Approach

The benefits of using fiber optics were discussed in a paper³ that identified three essential items as being critical to the beneficial implementation

of fiber optics in surface combatants. These are: a system-engineered cable plant that meets the needs of the shipboard systems; components that meet system requirements; and design guidelines to assure that system requirements are met. The concepts for an effective fiber-optic cable plant are discussed below.

Fiber-Optic Cable Plant

The fiber-optic cable plant consists of all of the trunk fiber-optic cables and fiber-optic interconnection equipment within the ship, including connectors, splices, and interconnection boxes. The fiber-optic cable plant is a totally passive system, serving only to transport optical signals physically between equipments (e.g., computers, displays, network nodes). To obtain the greatest ship-wide benefits from fiber-optic upgrades, a ship-wide design approach must be used. Maximum benefits can be obtained by designing the fiber-optic cable plant using a critical compartment interconnection box approach.

Description

The cable plant consists of trunk cable runs and interconnection boxes. Trunk cables that optically interconnect critical compartment interconnection boxes are typically run through

single-fiber connectors or splices within an interconnection box. A family of interconnection boxes is currently in development that will accommodate a maximum of 144 connectors or 432 splices per box. These boxes will also accommodate mixtures of connectors and splices in the same box with variable splices and connector counts. A typical link is shown in Figure 1.

Another term that has been used to describe the cable plant is Intercompartment Cable Service (ICCS). An ICCS was installed on two Ticonderoga-class guided-missile cruisers, USS Valley Forge (CG-50) and USS Mobile Bay (CG-53).

ICCS Concepts

The interconnection box approach to fiber-optic connections takes advantage of the fact that many types of signals can be carried on one type of fiber. This characteristic means that only a few types of components are required for all shipboard applications, which reduces logistic requirements. Since fibers are naturally immune to EMI, electromagnetic pulse (EMP), and other electromagnetic perturbations, multiple fibers may be run in the same cable, any of which will meet the requirements of an equipment interconnection. Thus multiple signals from one system or

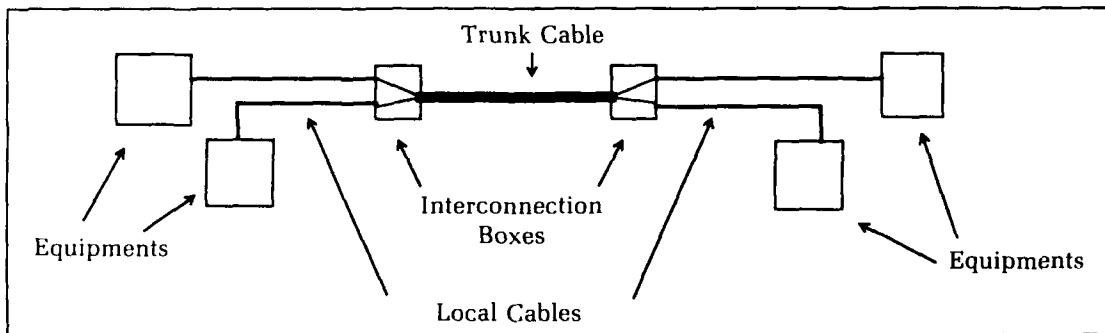


Figure 1. Typical fiber-optic link.

the main cableways. Local equipment cables used to optically connect the equipment to the cable plant at the compartment interconnection box are not typically run through the main cableways. Trunk cables generally have higher fiber counts per cable than do local cables. In previous installations, 8-fiber cables have been used for trunk cables, while 2- or 4-fiber cables have been used for local cables. A 36-fiber cable currently under development may be used in the future for the trunk cable application. The data path is established between equipment by connecting fibers in the local cables to common fibers in a trunk cable via

multiple systems can be run in the same cable, which further reduces cable space and weight requirements. These characteristics, together with the fact that long runs are possible, also reduce routing restrictions.

In addition, because the transfer from electrical cables to fiber-optic cables will save space and weight within the ship's cableways, spare fibers and extra trunk cables can be pulled at the time of ship construction to serve as survivable spares for all critical fiber-optic cables. In the event of fire or battle damage, fiber-optic systems can be manually reconfigured by the ship's crew from the damaged

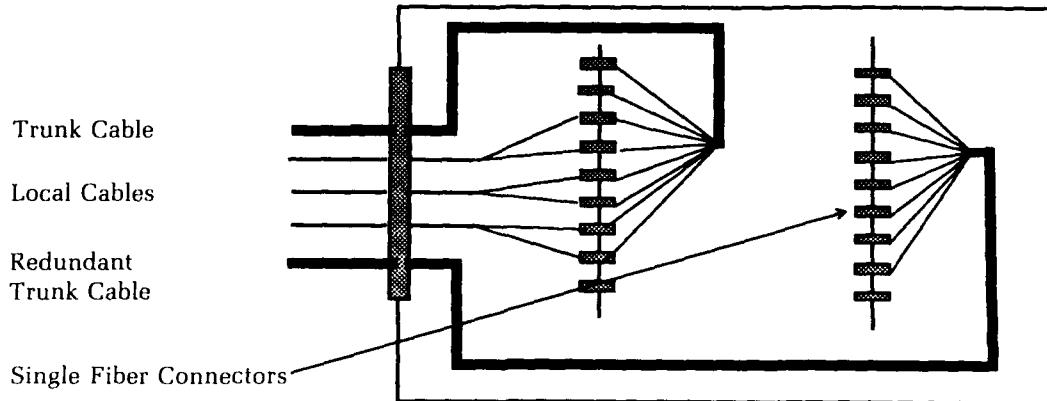


Figure 2. Typical fiber-optic interconnection box configuration.

cables to functional, redundant cables at each interconnection box location (see Figure 2).

Similarly, because the installed fibers will meet the requirements of most future systems, extra cables can also be pulled with the specific purpose of supporting future upgrades. This pre-installation of cable for future upgrades will allow quicker and much less costly additions to the ship's capabilities at later dates. Also, systems can be changed or replaced and still use the same fiber-optic cables for their intended lifetime.

The Naval Sea Systems Command is undertaking a project called Fiber-Optic Topology, the objective of which is to produce documentation and qualify components needed to design, install, maintain, and manage a fiber-optic topology (or cable plant). Much of this information is equally applicable to fiber-optic cable dedicated to one system. The products of this project will be:

- Design Guidelines
- Installation Procedures
- Maintenance Plan
- GENSPEC Update
- Configuration Management Plan
- Navy Training Plan & Curriculum
- Life-Cycle Management Plan
- Qualified Products (Initial set by November 1993)

The importance of standardized components becomes readily apparent as we begin to qualify components. The qualification of components is a long and expensive process, made more complex because of the rigorous surface warfare environment. The full range of specific environments experienced includes operating temperature extremes from -54 to +65°C, humidity levels up to 100 percent relative

humidity (RH), thermal shock, vibration, high impact shock, transportation shock, electromagnetic interference, electromagnetic pulse, electrical transients, lightning, magnetic fields, acoustic noise, inclination, radiation (gamma and neutron), overpressure, gun blast, wind, ice, rain, snow, ultraviolet radiation, dust and sand, salt spray, corrosive gases, explosive gases, fungus, corrosive fluids, water pressure, underwater shock, crush, compression, acceleration, and mechanical stresses such as twisting, bending, and impact. Each component and system installed aboard a Navy surface combatant (when not isolated from specific environments) is expected to be able to withstand these environments without interruption of performance. Each of these environments influences the ultimate design of each component and each item of equipment.

Component Performance

The performance of each component within the surface warfare system environment is limited to changes within some bounds. In the specifications for each component, the base optical performance of that component is specified, as well as the response of the component to each particular environment. Considered separately, the tests that a component must undergo are not extremely hard to pass. However, it is very difficult to design and produce components that will pass the conglomerate of these environmental and materials tests.

Compounding the challenge experienced by manufacturers in developing components that perform as required is the economic issue. The cost of conducting a full set of tests on a particular component can easily run into the hundreds of thousands of dollars. If a manufac-

turer develops more than one type of component, these costs are multiplied. For this reason, it is very hard for small manufacturers to develop products in this market. In some cases it is hard for any manufacturer to justify the costs associated with product development. It has required a close association between the Navy and the manufacturers to obtain the component development that has occurred to date. In some cases the Navy performed some of the tests in order to help the manufacturers keep cost at a minimum.^{4,5} These efforts have been extremely successful, with products available for all of the basic passive components. Most of these components have not been officially tested and qualified, but unqualified products and test data are available from the manufacturers. The component which has required the greatest amount of effort in development is the fiber-optic cable. Optical cable design and manufacturing processes are significantly different from those for electrical cables, and in some cases pose a significantly greater technical challenge for optical cable manufacturers. Specifically, the combination of Navy chemical requirements and shipboard mechanical requirements for the cable jacketing materials presented an extremely difficult problem to manufacturers—one that has only recently been overcome.

Tools and Test Equipment

The requirement for tools and test equipment for installing and repairing fiber-optic cable and connectors makes the standardization of fiber-optic components exceedingly important. Different fiber external diameters require different size connectors. Different types of connectors require different tools and different test equipment adapters. Transmitters and receivers operating at different wavelengths require different test equipment.

Key Cable Plant Component Qualification

The paragraphs below describe the components and point out the first set of components to be qualified.

Optical Fiber

Currently, two standard fibers are recommended for use in Navy systems (per MIL-F-0049291). The first is the 62.5-micron core/125-micron cladding, graded index, multimode fiber developed commercially for intrabuilding and Local Area Network (LAN) applications. This fiber size shows advantages over other multimode fiber sizes in mechanical and environmental sensitivity, and is the least bend-sensitive multimode fiber. For shipboard

interconnection lengths, this fiber is capable of handling data rates up to approximately 1 GHz depending upon the optical source characteristics and the exact link length.

For those applications where the data capacity of the multimode fiber is inadequate, the single-mode fiber is recommended. This fiber was developed commercially for high data rate, long length commercial telephone trunk lines. It has a core of approximately 10 microns and a cladding of 125 microns. The data capacity of this fiber exceeds many GHz.

It should be noted that each of the recommended fibers has the same outside diameter, so the same connector can be used for each application.

Fiber-Optic Cable

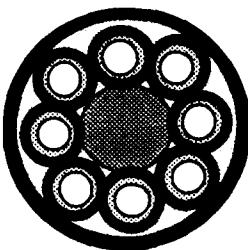
The design of a fiber-optic cable for a surface warfare system is influenced by many factors, including optical performance, environmental and mechanical performance, human factors, and safety. The following sections discuss the benefits and drawbacks of three common fiber-optic cable designs (Figure 3) with regard to those factors. The general requirements that a Navy fiber-optic cable must meet are given in MIL-C-0085045.

Cable Performance

Each of the cable designs has strengths and weaknesses. Any design to be used in a surface warfare system must meet the platform's system requirements. Additionally, the cable must meet minimum levels of performance in safety (e.g., low smoke generation, low toxicity, low halogen content, flame resistance), durability (e.g., able to withstand shock, vibration, mechanical abuse, fluids), ease of installation and repair, and optical performance. The base optical performance of a particular cable design is determined by the optical fiber and is not a primary consideration in choosing a cable design. All the other factors are greatly influenced by the specific design.

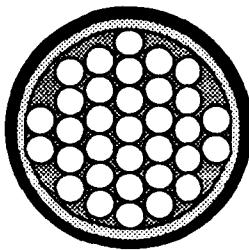
Safety. The safety issue is roughly equivalent for each of the cable designs. If the cable is made of appropriate materials, the cable will be safe. The same materials can be used regardless of the cable design with the exception of the ribbon-cable design. A new material with the appropriate mechanical properties would have to be developed for the inner plastic tube, because the current material will not meet Navy halogen content and toxicity standards.

Each of the designs can be very durable when properly manufactured and, with limited exceptions, could meet all shipboard requirements. The most notable exceptions are



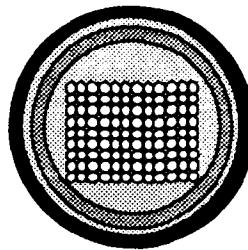
OFCC Cable

A fiber-optic cable consisting of individual, single-fiber cables called Optical Fiber Cable Components (OFCCs) laid with strength members around a central member and over-jacketed for environmental and mechanical protection. OFCC consists of a buffered fiber (900 microns total diameter) surrounded by strength members and a protective jacket with an outer diameter of approximately 2 millimeters.



Stranded Cable

A fiber-optic cable in which the buffered fibers (900 microns total diameter) are stranded down the center of the cable, surrounded by strength members and a protective jacket.



Ribbon Cable

A fiber-optic cable in which multiple-coated fibers (typically 12 each) with an outer diameter of 250 microns are sandwiched in a linear array, called a ribbon, and laid down the center of the cable. A commercial cable may contain up to 17 ribbons. The ribbons are surrounded by an inner plastic tube, strength members, and an outer protective jacket.

Figure 3. Fiber-optic cable designs.

the bending performance of the ribbon-cable design and the waterblocking performance of both the stranded and the ribbon designs.

Waterblocking. Neither the commercially available ribbon nor the stranded cable design is waterblocked with suitable materials (e.g., non-greasy, non-tacky). A method for waterblocking both the ribbon design and the stranded design will need to be developed for use in a surface warfare system. It is expected that the stranded design will show better waterblocking performance than the ribbon design.

Bending. A comparison of bend performance of the different designs indicates that the minimum bend radius of the ribbon design is approximately 2.5 times that of both the stranded and the OFCC designs, and is not small enough to allow unrestricted installation of the ribbon-type cable. If the ribbon cable were waterblocked with appropriate compounds, it is expected that its bend performance would worsen.

Human Factors. The human factors elements for each design are widely varied. The primary area of concern for human factors is in the breakout of individual fibers in the cable plant interconnection (junction) boxes. The OFCC design is easiest to handle because each fiber is contained within its own subcable. The stranded design is almost as easy to use, but the individual fibers are more susceptible to accidental damage because they are not protected as in the OFCC. The ribbon design is the

hardest to handle, because the individual fibers show high susceptibility to damage when separated from the ribbon, necessitating the use of multifiber connection and splicing techniques in which all the fibers within a ribbon are terminated at the same time. The use of multifiber terminations introduces maintenance, reconfiguration, and repair problems.

System Requirements. The system requirements that a fiber-optic cable must meet differ depending upon the platform on which it will be installed. For some surface warfare systems, the primary system requirement is that there shall be no single points of failure within the combat system. This requirement places severe system design limitations on the fiber-optic cable plant design if different combat system elements use common multifiber connectors or splices within it. The multifiber connector or splice becomes a single point of failure for multiple systems. For those particular platforms, all connections within the cable plant must then be single-fiber connectors. Using multifiber connectors or splices for terminating the ribbon-cable design violates this basic principle for those platforms.

Capacity. The ribbon-cable design has the greatest gross fiber capacity, being able to hold 204 fibers in a one-half-inch cable. For the same size cable, the stranded design can accommodate approximately 48 fibers. The OFCC design can accommodate approximately 12 fibers in a one-half-inch cable.

In summary, the OFCC cable design was developed for shipboard applications and meets all shipboard cable requirements. Its only limitation is that the outside diameter increases quickly with increasing fiber count. The OFCC cable design is realistically limited to cables containing fewer than 24 fibers, but in Navy applications will probably never exceed 12 fibers. The stranded cable design is recommended when fiber counts within a cable exceed the realistic limits for OFCC-type cables. A 36-fiber stranded cable based on a commercial cable design is currently being planned. The primary difficulty with this cable design is in meeting the waterblocking requirements. The ribbon-cable design provides the highest density cables with the smallest outside diameter. The ribbon cable has the poorest waterblocking capabilities of the three cable designs, and may violate some system requirements depending upon the platform of application. The ribbon also has very poor bending properties that would force it to be installed using special procedures. Additionally, new materials that meet the Navy toxicity and halogen content requirements for cables would have to be developed for the inner protective tube.

Interconnection Equipment

Implementation of any fiber-optic conversion necessitates the use of fiber-optic interconnection equipment. The type of interconnection equipment needed varies with the specific implementation, but can be divided into three functional groups, namely, single-fiber, light duty connectors or splices; multifiber, heavy duty connectors; and interconnection boxes. These components are described in more detail in MIL-C-83522, MIL-S-0024623, MIL-C-28876, and MIL-I-24728. The design of each of these components was chosen in order to optimize performance for the application in which they are used, as well as to minimize problems due to human factors.

Single-Fiber Light Duty Connectors. The single-fiber connectors (per MIL-C-83522/16) are intended only for use inside interconnection boxes or equipments. They are not intended to be subjected to harsh treatment, although they do meet some minimum requirements in the areas of environmental, mechanical, and chemical performance. The connectors are intended to be durable, yet easy to engage and disengage when reconfiguration is necessary. They have been optimized to provide high-grade optical performance (typical losses of approximately 0.5 dB) while not requiring high skill levels for installation and maintenance. The single-fiber connector design is based on

the commercial ST (straight tip) connector, which has a bayonet coupling for easy engagement and disengagement. To assure high-grade optical performance, tight tolerances are maintained in the connector ferrule manufacturing process, yielding hole sizes with a 1-micron tolerance. The ferrules can be constructed of ceramic, stainless steel, glass, or composite materials, but are typically ceramic, which is preferred because of tighter tolerances and improved environmental performance.

The ST was chosen as the basis for the single-fiber, light duty connector design for a number of reasons. Primary among these were human factors and repeatability. The bayonet design of the ST allows for smaller connector spacings on terminal boards and optimizes maintenance times without sacrificing performance. Furthermore, because of the keyed design, connector repeatability is much greater than that of the other competitive, non-keyed designs.

The environmental performance of the ST was investigated at NAVSWC. The temperature and vibration performance of the connector is comparable to other single-fiber, light duty connectors. The spring-loaded bayonet design makes the connector more susceptible to shock than some other connectors, but the connector does meet the specified requirements for multimode fiber (less than 0.5 dB change in loss during the shock test). For single-mode fiber the connector shows a high shock sensitivity at this time. Work is in process to improve the performance of the connector for both fiber types.

Single-Fiber, Light Duty Splices. The single-fiber splices (per MIL-S-0024623) are also intended for use only inside interconnection boxes or equipments. Their primary purpose is to provide a higher grade optical performance in those cases when the single-fiber connectors are not adequate. In general, they are used when connectors would introduce too much loss into the system, but may also be used when low reflection interfaces are needed, such as for high-speed, single-mode communications systems. The splice is intended to be a permanent connection, but may be de-mated and re-mated to another splice with only slight degradation of performance should emergency reconfiguration become necessary. The single-fiber splice design is based on the commercial rotary splice. It is constructed of a single glass tube, which is broken into two pieces during the installation process. Each half of the tube is terminated in much the same way as a connector, and then the two pieces are reassembled as before they were separated. Since both halves of the splice are from the same glass tube, splice losses from geometry mismatches of the

splice parts are minimized, and an extremely low-loss splice results.

The environmental performance of the single-fiber, light duty splice was also investigated at NAVSWC. The results were excellent; the splice showed no sensitivity to vibration and minimal sensitivity (<0.2 dB change in loss) to temperature and shock.

Multifiber Heavy Duty Connectors. The multifiber heavy duty connectors (per MIL-C-28876) are required to withstand the severe naval environment while showing only limited performance degradation. They are intended to be used to connect equipment in any application where the fiber-optic connector is exposed and not housed within the interconnection box. These connectors are expected to withstand the full rigors of the Navy shipboard environment including shock, vibration, temperature extremes, fluids including salt water, and humidity. The connectors are also expected to withstand the mechanical use and abuse experienced daily by connectors for this application. The design of the connector is based on existing electrical connectors for the same application, but with a redesigned insert for the connector and optical termini instead of electrical contacts. The optical termini are removable to allow limited repair capability and cleaning, if necessary. The termini ferrules are constructed similarly to the light duty connector ferrule and are made of the same materials, typically ceramic. The same types of tolerances are maintained on the ferrule hole size to assure high-grade optical performance. The backshells of the multifiber heavy duty connector are designed to perform all cable strain relief so that no stress is translated to the optical interface of the connector when stress is applied to the cable.

Other multifiber, heavy duty fiber-optic connectors exist in addition to the MIL-C-28876. Some are essentially proprietary, but other MIL-SPEC products are available (e.g., the optical MIL-C-38999). All have shown problems of one type or another to date. Typical problems that have been experienced are high initial insertion loss, low tolerance of mechanical abuse, and greatly increased loss with repeated matings and de-matings. In particular, for the MIL-C-38999, the inability of different manufacturers' products to mate optically is a primary problem, caused by the fact that the connector was designed with tolerances intended for electrical contacts, but insufficient for optical terminations.

Interconnection Boxes. The interconnection boxes (per MIL-I-24728) are also intended to withstand the full rigors of the Navy shipboard environment and, to the maximum extent possible, to protect and isolate the connectors and

60

splices from the environment. The boxes are modular, with current versions containing one, two, or three modules. Each module may contain up to 48 connectors or 144 splices. Connector and splice modules may be mixed in the same box. The connectors are mounted on a flat plate which slides out of the box when it is opened to allow for easy access and to minimize the chance of secondary damage when repairs are performed within the box. The splices are mounted in removable drawers which organize the splices and provide limited mechanical protection. The front cover of the box is hinged such that it may be either completely removed from the box or held out from the box to function as a work table. Both the connector panel and the splice drawers are securely locked into place when the cover is closed and secured. The design of the box structure itself is based on the design of existing lightweight electrical interconnection boxes.

Fiber-Optic System Design Issues

Standardized design guidelines are required because of the way the Navy procures systems. When requests for proposals are based on a system design specification, the impact of the shipboard cable plant on the system design must be specified. Otherwise, respondents may make different assumptions and arrive at designs having vastly different performance and cost. Two such issues are described below.

Environmental Considerations in System Design

In a fiber-optic system for a surface warfare system, optimal design is a function of power budgeting. In a benign environment, only the base component performance in terms of optical loss is taken into account. An additional safety margin is added to the calculated system losses to take into account component aging and any unforeseen environmental or mechanical effects. Typical commercial guidelines recommend that the safety margin be between 3 and 6 dB. For naval applications this margin is inadequate because of the wide range of environments expected and the allowable limits of component degradation in those environments. Furthermore, in the surface warfare system it is not acceptable for a system to go down because of the adverse environment. It is quite possible that the occurrence of the worst environmental scenario will coincide with a combat situation, so it is imperative that the system perform through all realistic environmental scenarios.

In order to account for the environmental perturbations that a shipboard fiber-optic link experiences, a new method of calculating the system power budget was developed.⁶ The base performance of each component in a benign environment is still used to calculate system losses, but additional terms are added to the power budget equation to account for environmental and mechanical effects. In effect, the safety margin was changed from a set value to a variable quantity that changes depending upon the nature and complexity of the system being designed. This new variable margin is defined by the following equation:

$$M = 3 \text{ dB} + \sum_{i=1}^k n_i \Delta_i$$

where n_i is the number of components of type i in the system, Δ_i is an environmental correction factor for components of type i , and a set amount, currently 3 dB, is added to account for transmitter and receiver aging effects. Environmental correction factors for each component have been calculated⁶ by assuming a worst-case set of environments that would act on a system at one particular time and statistically summing the optical responses of that component to each of the environments. The environments considered in this worst-case scenario were temperature, humidity, shock, vibration, bending (cables only), compression (cables only), dust and sand, and mechanical stresses such as impact, twisting (cables only), and tensile stresses (cables only).

As the component responses to an environment are statistical in nature, so also are the environmental correction factors. Therefore, a system designer may choose to what degree (confidence interval) a system will be environmentally resistant and use the corresponding environmental correction factors. Initial recommended values for the mean and standard deviation of the correction factor distribution for the basic shipboard fiber-optic components are listed in Table 1.

The margin calculated by using these values may be interpreted as follows. The mean response of a component represents the mean amount of power that a group of similar com-

ponents will lose in the assumed worst-case environmental scenario. For a group of components in a system, the sum of the individual responses can be interpreted as the amount of additional optical power required by that system, in that scenario, to ensure a baseline performance for some percentage of occurrences. If the mean responses are used, this can be translated to mean that the system will maintain its pre-exposure baseline performance for 50 percent of occurrences of that scenario. For those systems which do not meet their pre-exposure baseline performance, there will not tend to be total loss of the system, but rather some sort of degradation (e.g., increased bit error ratio for digital systems or increased signal-to-noise ratio for analog systems). However, if the performance of enough components in a system lies at the high extremes of the distribution, or if the system shows a very quick degradation of performance with decreased optical power (as in frequency modulated systems), total system loss may occur. The percentage of occurrences for which no system degradation is allowed is a function of the criticality of the system.

Modal Effects on Power Budgets in Multimode Systems

A final subject to be mentioned is the impact of the ship's physical characteristics on the optical design of fiber-optic systems. Shipboard systems tend to be short (10 to 100 meter lengths), highly concatenated applications. Data transmission rates at this time are relatively low (typically less than 10 megabits per second), with expected near-term data rates in the few hundred megabits per second range. Because of these requirements, as well as relative cost and reliability, multimode fiber and light-emitting diodes typically have been the components of choice and probably will remain so for the near future. This combination of components and short length, concatenated systems imposes some unique factors on the system design and analysis process. Procedures and techniques which work in system power budgeting and bandwidth budgeting for long-haul (greater than 2 kilometer) systems do not

Table 1. Mean and Standard Deviation of Environmental Correction Factor Distributions

Component	Distribution Mean	Distribution Standard Deviation
Cable (Single-mode)	1.33/km	0.44/km
Cable (Multimode)	1.35/km	0.45/km
Splice	0.10	0.03
Heavy Duty Connector	0.24	0.08
Light Duty Connector	0.24	0.08

give accurate predictions for short-length systems.^{7,8,9,10} For these types of systems, system performance is a function of the detailed optical characteristics of the optical sources used, which at this time are not predictable. The effect that this has on system design is that designers tend to be overly conservative in system predictions, which in some cases increases system costs. This issue is being addressed by the fiber-optic community, and a generally accepted solution may be near. As single-mode component costs continue to decline and reliability improves, single-mode systems may become the design of choice in the future, eliminating this design issue.

Conclusion

The use of standardized components reduces the complexity of logistic support and the amount of training required for maintenance personnel. System designers should be familiar with standard components and use them to the maximum extent possible. Design guidelines are important in designing fiber-optic hardware for surface warfare systems, especially when a shipboard fiber-optic cable plant is used. The development of these guidelines should be continued, and they should be included in system specifications.

Acknowledgments

The authors wish to acknowledge the support of the Fiber Optics Program Office (SEA 06KR22) and the AEGIS Shipbuilding Program (PMS 400) of the Naval Sea Systems Command. The authors also acknowledge the contributions of the staffs of these offices and the members of the AEGIS Fiber Optic Working Group.

References

1. Owen, D. J. and Paxton, J. G., Jr., "Overview of Fiber Optics Applications to Navy Surface Platforms," *Military Fiber Optics Conference '88-West*, Dec 1988.
2. Morais, Robert M., "Implementation of Fiber Optic Technology in Naval Combatants," *Marine Technology*, Vol. 24, No. 1, Jan 1987, pp. 59-71.
3. Knudsen, D. R., et al., "A Ship-Wide System Engineering Approach for Fiber Optics For Surface Combatants," *Naval Engineers Journal*, May 1990, pp. 127-137.
4. Brown, G. D., et al., "Temperature and Humidity Testing of Fiber Optic Components," *Fiber Optics Reliability: Benign and Adverse Environments III*, SPIE Vol. 1174, Sep 1989.
5. Brown, G. D., et al., "Vibration Testing of Fiber Optic Components," *Fiber Optics Reliability: Benign and Adverse Environments III*, SPIE Vol. 1174, Sep 1989.
6. Brown, G. D. and Anderson, M., "Assigning Loss Budgets to Fiber Optic Components in Adverse Environments," *Fiber Optics Reliability: Benign and Adverse Environments III*, SPIE Vol. 1174, Sep 1989, pp. 177-181.
7. Koyomada, Y. and Yamashita, K., "Launching Condition Dependence of Graded-Index Multimode Fiber Loss and Bandwidth," *Journal of Lightwave Technology*, Vol. 6, No. 12, Dec 1988, pp. 1866-1871.
8. Cibotto, S. and Someda, C. G., "Bandwidth Concatenation in Multimode Fibres: A Review," *Journal of the Institution of Electronics and Telecommunications Engineers*, Vol. 32, No. 4, 1986, pp. 253-258.
9. Reitz, P. R., et al., "A Comparison of Loss Measurement Methods With Observed System Loss in Optical LAN Links," *Fiber Optics: Short-Haul and Long-Haul Measurements and Applications*, SPIE Vol. 559, 1985, pp. 163-169.
10. Brown, G. D., et al., "The Development of Steady State Modal Conditions in Short Length Concatenated Systems," *Fiber Optics Reliability: Benign and Adverse Environments II*, SPIE Vol. 992, 1988, pp. 161-167.

The Authors



DENNIS KNUDSEN received a Ph.D. in physical chemistry from North Dakota State University in 1970. At the Center from 1970 to 1986, he developed shipboard equipment for vapor and radioactive particle detection, methods for testing protective clothing for torpedo workers, and methods for the thermal destruction of hazardous wastes. As head of the Photonics Group in the Combat Systems Department for

the past four years, Dr. Knudsen has directed projects for the AEGIS Shipbuilding Program and the NAVSEA Fiber Optic Program Office. His projects involved fiber-optic component development and system implementation issues. He has been active in the AEGIS Fiber Optic Working Group, serving as the chairman of a group to evaluate fiber optic candidates for future Arleigh Burke class destroyers. Recently he has been involved with focused technology efforts in optical and molecular computing. Dr. Knudsen is a member of the American Chemical Society.



GAIR BROWN graduated summa cum laude in May 1985 from West Virginia Institute of Technology with a B.S. in physics and mathematics. He has worked in the Fiber Optics Group at the Naval Surface Warfare Center since June 1985, concentrating on shipboard cable plant design, specifications and standards, and light launch conditions for shipboard applications. He is currently pursuing an M.S. in

physics from Virginia Polytechnic Institute and State University.



JOSEPH INGOLD received a B.S. degree in physics (1985) and an M.S. degree in electrical engineering with concentration in fiber optics (1990) from Virginia Polytechnic Institute and State University. Since joining the Naval Surface Warfare Center in 1985, he has worked in the Photonics Engineering Group of the Combat System Technologies Branch, performing environmental tests on fiber-optic components and shipboard experiments on fiber-optic sensors and data transfer systems.

An Integrated Artificial Neural System for Target Identification

Thomas Holland, Tom Tarr and Ali Farsaie

A system has been developed which successfully blends the strengths of digital image processing with the advantages of Artificial Neural Systems (ANS) to achieve target identification in non-optimum environments. Infrared (IR) video images of five different military vehicles provide a data base of targets. These images were digitized and preprocessed by means of conventional digital image processing techniques. A novel ANS model was developed to analyze certain features of the images and perform classification of the targets. Tests on actual IR target images demonstrated that the system is capable of correctly identifying any of the five targets to an accuracy of 86 percent, independent of target orientation and location within the scene.

This effort demonstrates a unique technology capable of identifying genuine targets in a degraded-image scene. The system neither requires development of target-specific algorithms, nor is it subject to human/physical limitations. Furthermore, the results demonstrate that this capability is attainable with current state-of-the-art computation and sensor technology.

Introduction

This article describes an effort to develop capability to perform classification of mobile, land-based targets in a man-portable system configuration. The objectives were to: (1) study the performance of an artificial neural system for identification of targets acquired from an infrared imaging device, and (2) investigate integration of an artificial neural system into an advanced surveillance prototype. This article deals with the system-level integration and implementation issues; specific supporting technologies are described in a Naval Surface Warfare Center (NAVSWC) technical report.¹ The article covers seven major sections. The first section presents background information on target recognition problems. The second section covers the nature of the targets used for this study. The third section details the equipment used during this study. The fourth section discusses the Artificial Neural System (ANS) algorithm used for the investigation. The fifth section covers the design methodology of the ANS software development. The sixth section presents the system configuration and describes the military prototype hardware in which the ANS is embedded. The seventh section presents results of tests on field data and offers recommendations and conclusions.

Background

The Special Projects group of NAVSWC's Gun Ammunition Branch has been actively pursuing a United States Marine Corps development program, the goal of which is to produce an advanced observation system consisting of man-portable, multi-sensor, multi-processor based target-acquisition equipment. The intent of this Forward Observer Support Technology program is to demonstrate target observation, provide surveillance for target reconnaissance purposes, and support the direction of indirect fire weapons. The policy of this effort has been to take advantage of the latest technologies in order to provide the functionality of battlefield observa-

tion in a compact, man-portable and automated fashion.

The task of battlefield observation is often carried out under adverse observational conditions and is subject to obscurants between the observer and the area observed. It has been conjectured that the use of an IR sensor (either alone or with a visible light sensor), along with the appropriate processing by an ANS used in conjunction with conventional image processing techniques, would allow the inhibitory effects of most obscurants to be negated. Furthermore, observational capability would be greatly enhanced by reducing the effects of human factors (e.g., fatigue, physical trauma) which further complicate and inhibit an observation mission.

Target Image Data

The data used for this study consisted of images of five United States Marine Corps vehicles acquired at oblique horizontal angles using an infrared line scan imaging device. Images were acquired at midday under bright sunlight with a background of forest and, in the foreground, a grassy field. All vehicles were at normal operating temperature and in operation during imaging.

The complete image set consists of seven images of each of the five vehicles taken at

15-degree increments, with 0 degree being broadside the vehicle and 90 degrees being head-on (Figure 1). The five vehicles imaged were the M60A Main Battle Tank, Light Armored Vehicle (LAV) 8X8, Amphibious Assault Vehicle (AAV7A1), Hummer M998 High-Mobility, Multi-purpose Wheeled Vehicle (HMV), and a 5-ton truck (Figure 2).

These early image acquisitions were selected to offer ideal images in a real environment, the intent being that the system would be able to generalize from these images in order to provide accurate target identification when images are severely degraded. Training and testing of the ANS later in the development identified certain inadequacies in the acquisition of the training set. For example, images were

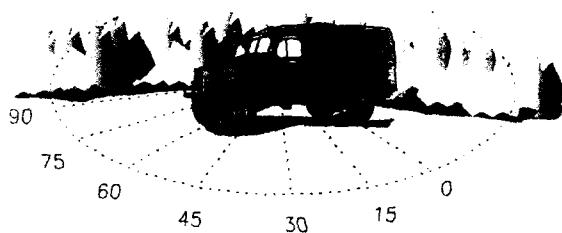


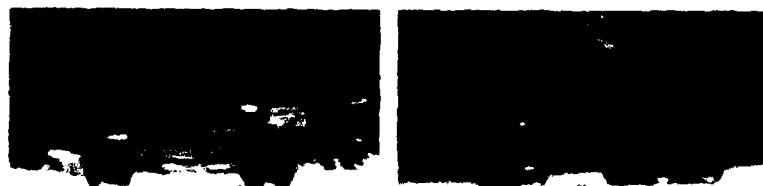
Figure 1. Target aspect angles (degrees).



M60A MBT

LAV 8X8

AAV7A1



M998

Truck

65

Figure 2. Military targets of interest.

sum or vapor created by the initial functioning of the weapon and vapor created by evaporation, with time, of liquid released by the

Figure 5. Degradation to ship's functional departments caused by CW attack.

acquired at a range of approximately 30 meters; however, line scan was varied to fill the scan frame with the target image. This produced scaling artifacts in which relative scales between targets were not maintained.

Equipment Description

Figure 3 illustrates the arrangement of the image acquisition hardware. The IR camera was an Inframetrics Model 522 Infrared Thermal Imaging System consisting of a liquid-nitrogen-cooled, mercury-cadmium-telluride (HgCdTe) image sensor and its associated controlling electronics. The image sensor is line scanned at a 30-Hz frame rate with 200 lines per frame. The infrared spectral range is 3 to 5.6 μm .

Both infrared and visual images were recorded on VHS standard video tape to facilitate image study in the laboratory. Images were converted to digital forms using a Data Translation DT2861 Arithmetic Frame Grabber in an IBM-PC/AT-compatible host, achieving a digital resolution of 512 x 512 pixel elements with 256 gray levels per pixel. Figure 4 presents a functional block diagram of a frame-grabber board. A video signal is converted to a digital format many times each second using

an analog-to-digital converter. Multiplexers (MUX) allow digitized data to come from either video input or the frame-store buffer. Each digital value passes through an input look-up table (LUT) as it comes in, allowing an initial, on-the-fly translation of the data. An arithmetic logic unit (ALU) can then perform a variety of calculations and logic functions on the data, which may be used in conjunction with a Result LUT to provide rapid implementation of many mathematical functions. This process is repeated 512 times for each horizontal video line in each video frame. Each video frame is then stored in a frame-store buffer. The contents of each frame-store buffer may be fed back into the input LUT for later processing by the ALU, or they may be used in conjunction with incoming data. (For instance, the incoming data and a previous frame of the same scene may be passed through an exclusive-OR operation to reveal only that which has changed in the scene.)

The DT2861 frame grabber contains 4 megabytes of frame store memory, allowing storage of 16 images. The onboard ALU and input LUT provide the means to process any of the 16 images stored onboard in real time (approximately 30 times a second), as well as images actively being digitized. This capability

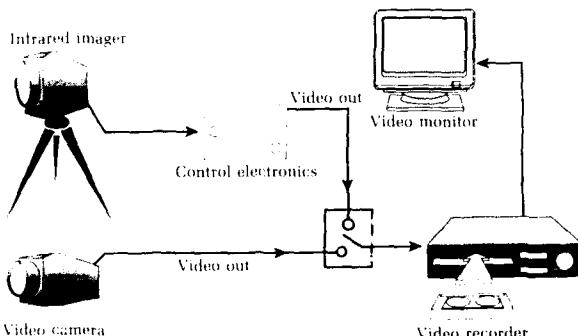


Figure 3. Image acquisition arrangement.

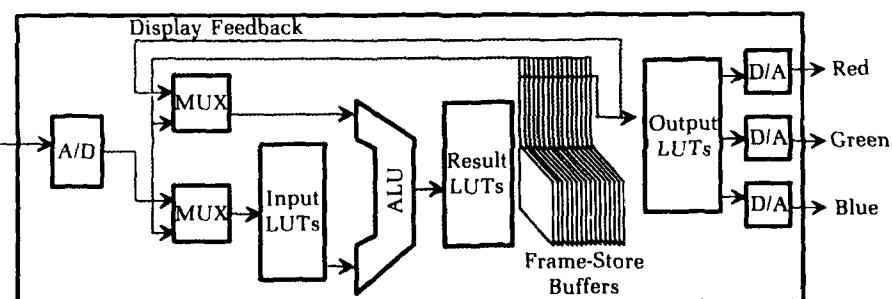


Figure 4. Frame grabber block diagram.

allowed laboratory experimentation with direct applicability to real-time implementation. The frame-grabber board resided within an IBM-PC/AT-compatible computer. ImagePro, an image processing software package from Media Cybernetics, Inc., was used to examine various image parameters in order to determine a suitable preprocessing heuristic. The NTSC (National Television System Committee) standard video from the video cassette recorder was input to the frame grabber and the result displayed on an analog RGB (red-green-blue) monitor.

ANS Algorithm

ANS technology provides a number of tools which could form the basis for a potentially fruitful approach to the recognition problem. It provides powerful collective and computational techniques for designing a robust target recognition system. The technology utilizes a novel pattern recognition process to demonstrate invariant pattern recognition, robustness to noise, obscuration and camouflage.

The elements that combine to determine system performance are: spatial resolution of the sensor; the features extracted from signatures; the structure of the classifier and, in this case, the ANS. With an ANS, important issues have to do with optimizing the network structure in terms of the learning algorithm and the number of connections.

A hybrid network incorporating characteristics of supervised Kohonen,² Hartigan,³ and Adaptive Resonance Theory,⁴ was used. The network is shown schematically in Figure 5. Initially, the network consists of one

neuron (i.e., a single computational element in an artificial neural network) for the first target in the training set. This neuron has a sphere of influence set at some large value. As more examples are presented to the network, the following steps are taken:

1. The radius of the sphere of influence of a neuron may be decreased. This occurs if a training example not of the type represented by the neuron falls within the influence of that neuron.
2. The node center of the neuron is adjusted in the manner described by Kohonen.² This allows the node center of the neurons to seek the statistical center of the data represented by that neuron. A neuron that fires incorrectly has both its radius of influence decreased and its node center shifted.
3. A new neuron may be added. This occurs if a training example does not fall within the influence of any currently active neuron and nearest-neighbor classification of the training example is incorrect. When the new neuron is added, the features of its node center are set to the values of the training example and its radius of influence is set to some large number.

In addition, if the training set is finite, logic can be included to ensure that all neurons identify at least one member of the training set. This can happen if the updating procedure described above pushes the node center too far away from the data early in the training. If the neuron does not fire for any training pattern, it is disabled and used, if needed, at a later time.

Methodology

This effort encompassed definition of hardware, system design and integration, and examination and implementation of an ANS

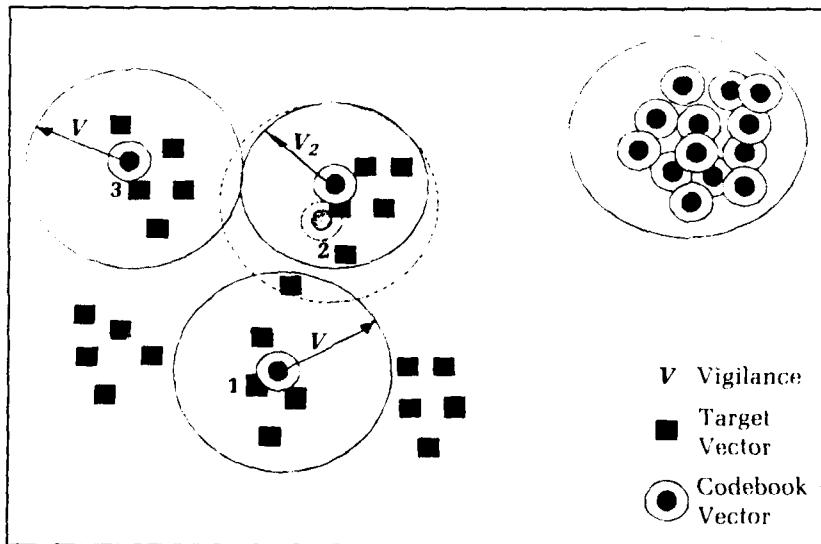


Figure 5. Simplified diagram of the hybrid network.

algorithm to produce a system which could be integrated into an advanced target acquisition prototype. Special attention was given to optimizing input data to the ANS in order to reduce system computational requirements and so increase system throughput and decrease system training time. Figure 6 depicts the methodical approach of system development.

Acquired images were digitized and then digitally processed to eliminate extraneous information. Commercially available image processing hardware and software algorithms allowed the use of established feature extraction methods with near real-time performance; however, human judgment was required to monitor background and foreground elimination in order to prevent excessive loss of target information. A heuristic method was developed to separate the target within the image frame from its surroundings. Histogramming and contrast enhancement were employed in the

heuristic, resulting in a set of binary target images, as shown in Figure 7. The image resolution of 262,144 8-bit pixel elements prohibited a neural network simulation, running on the intended hardware, from processing these images with anything near real-time performance. Earlier experience indicated that ANS training cycles on input matrices of this order proved too lengthy for the time frame of this study. Furthermore, it was the hypothesis of this study that any target of interest would have suitable statistically distinguishable characteristics that would allow conventional image processing methods to reduce the quantity of data needed to determine the target's identity.

This hypothesis led to an examination of various methods of image compression and feature extraction in order to reduce the number of inputs to the ANS. A procedure for selection and extraction of features was

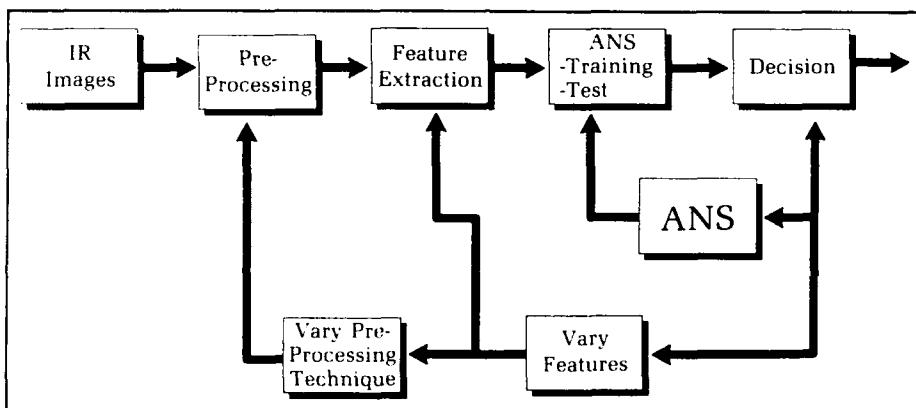


Figure 6. System development.

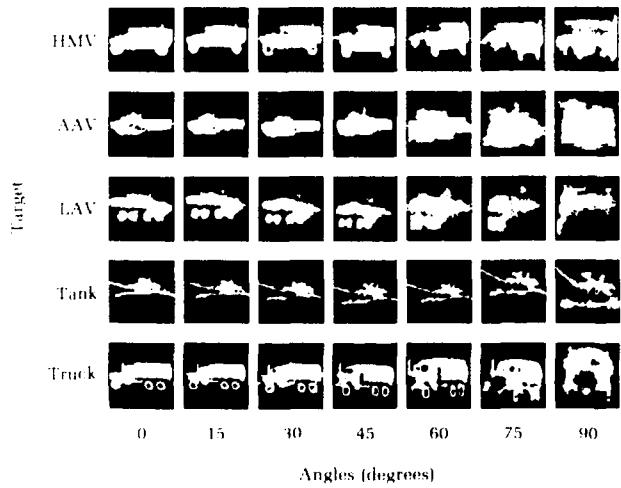


Figure 7. Binary images of targets.

developed. This provided a set of feature vectors called "theta neighbors" for use with the neural network model.

The theta-neighbor vector is made up of multiple elements. The first element is the number of pixels in the target frame that have a value of one. The remainder of the elements are determined by performing logical AND operations. For each pixel in the frame, a logical AND of pixels is performed: the original pixel position ANDed with the pixel rotated by theta degrees about the center of mass of the target. By rotating the target through more theta angles, one can create a more detailed theta-neighbor vector, potentially allowing targets that are similar in appearance to have separable features.

The hybrid network described previously was implemented for target recognition. All targets at certain given angles were used for training. The network exhibited ability to recognize targets during testing viewed from angles not included as part of the training set. As an example, in Table 1 the network was trained on only two angles, zero and 45 degrees, but was tested on five angles, zero through 60 in 15-degree increments. The network correctly identified over 83 percent of the tested targets. This shows that, training on only 40 percent of the data set, an accuracy of over 83 percent

Table 1. Test Results of the Hybrid ANA

Trained Angle (deg)	Tested Angle (deg)	No. of Neurons	% Correct Classification
0, 45	0 - 60	10	83.2
15, 45, 75	0 - 75	19	86.3
0, 30, 45, 60, 90	0 - 90	25	85.7

was achieved using the entire data set. It was also noted that less than half of the 25 allowed neurons were used in this process. When the network was trained using three angles, 15, 45, and 75 degrees, and tested on six angles, zero through 75 in 15-degree increments, over 86 percent correct target identification was achieved. The slight decrease in percent correct when the network was trained on five angles and tested on all seven is believed to be due to the scaling artifact problem which occurred during original data gathering. In general, the hybrid network using theta neighbor extracted features demonstrated great promise for target recognition applications.

System Overview

Figure 8 depicts the system implementation at its highest level. Here an ANS is trained on a representative set of targets maintained in a target data base. Once the ANS is trained, it is transferred to the real-time system. Scenes acquired by the real-time system are preprocessed according to the heuristics developed and then presented to the hybrid network for classification. The output of the network indicates the target identity.

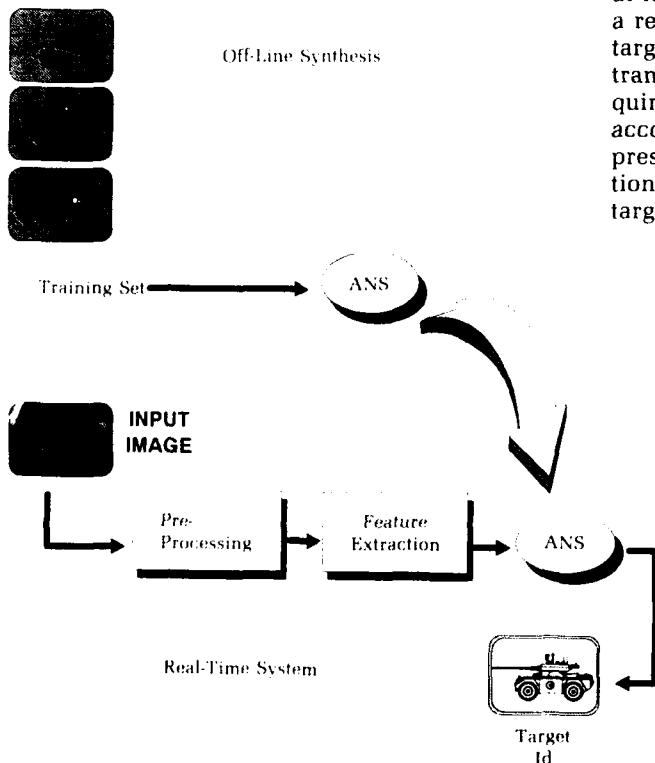


Figure 8. System implementation.

Forward Observer Support Technology, a United States Marine Corps development program being pursued at NAVSWC, is developing man-portable target observation and surveillance equipment for target reconnaissance and the direction of indirect-fire weapons. One of the major subsystems central to this development is the Host Integration Processor (HIP), an IBM-PC/AT-compatible computer system responsible for high-level data handling and user interface to the various components of the system (Figure 9). The HIP was

and neoteric ANS technology to accomplish near real-time target identification. The system demonstrated that this capability is achievable with current off-the-shelf components, and therefore indicates that even better performance may be obtained through the use of emerging processing technologies. The system configuration, heuristic algorithms, and artificial neural system are robust in that they can be implemented with improved performance directly proportional to the improvements in the hardware components. For example, the

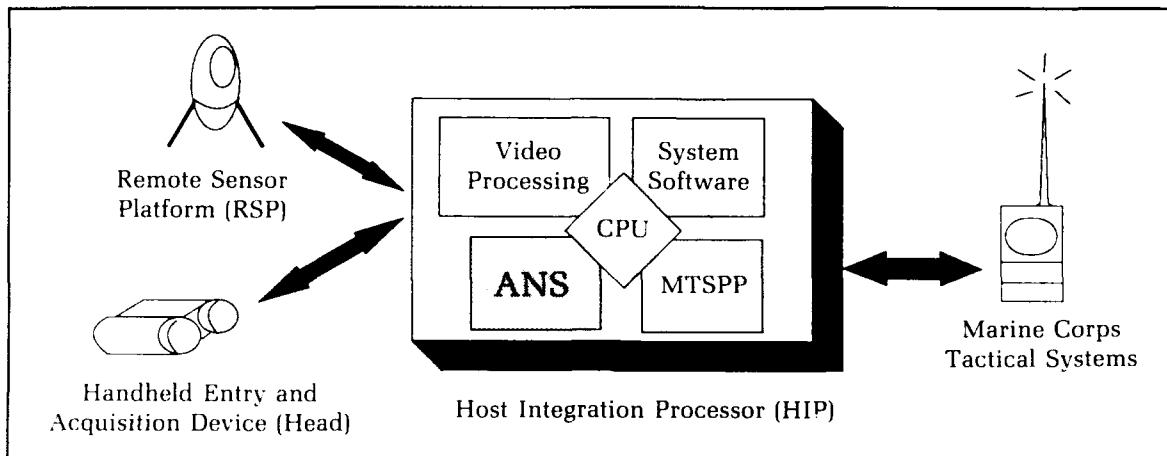


Figure 9. Resulting system configuration.

targeted as a prototype testbed for the ANS-based target identification system for three reasons: the HIP is central to a target observation system that is easily adapted to new observation hardware (e.g., lenses, cameras, image processing hardware); the HIP is at an early stage of development, which facilitates software and hardware integration; and the IBM-PC/AT standard offers a wide selection of hardware and software development tools and allows investigation and development to be pursued with any existing IBM-PC/AT.

Conclusions and Recommendations

The neural network model provides the capability of training the system on a representative subset of a desired target family without requiring that all possible image scenes be presented. Tests on real infrared target images have shown the system to be capable of correctly identifying any of the five targets considered to an accuracy of greater than 86 percent.

The resulting system combined the respective strengths of modern digital image processing

70

current state-of-the-art infrared imaging devices provide image resolution in excess of 300 television lines, three times that used for this study, and so would provide more distinguishable target features for input to the system. Recent advances in digital computer technology have increased the processing speed of typical personal computers by over 500 percent from the "super desktops" of just four years ago. Analog implementations of artificial neural systems are presenting unprecedented associative processing capabilities.

Further work is needed to realize fully the system's capability. Because of the limited scope of this investigation, certain issues could not be considered. The researchers could find no previous studies relating to image-degrading phenomena (meteorological conditions, battlefield smoke, camouflage, etc.) in these terms. Also, nighttime target images need to be included in the target data base. This would eliminate sun heating effects and might prove that daytime images with secondary heating from the sun need not be considered, as each IR image would consist of only target primary heat sources.

Acknowledgments

This article presents the integration and implementation issues resulting from a two-year NAVSWC Independent Exploratory Development program titled, "An Artificial Neural System for Target Identification Using IR Imagery Data."

References

1. Tarr, T., Holland, T., Farsaie, A., *Development of an Artificial Neural System for IR Target Recognition*, NAVSWC TR 91-72, NAVSWC, White Oak, MD, 1 Feb 1991.

The Authors



THOMAS HOLLAND received his undergraduate degree in electrical engineering from Tennessee Technological University and is employed as an electronics design engineer in NAVSWC's U.S. Marine Corps Special Projects Group. His interest in the applications of advanced machine intelligence stems from his graduate studies in electronics, artificial intelligence, and computer architecture at Virginia

Polytechnic Institute and State University. He is currently the program manager for the Marine Corps' Automated Target Detection and Identification Systems program at NAVSWC.

2. Kohonen, T., *Self-Organization and A. sociative Memory, Second Edition*, Springer-Verlag, Co., New York, NY, 1987.
3. Hartigan, J. A., *Clustering Algorithms*, John Wiley & Sons, New York, NY, 1975.
4. Lippman, Richard P., "An Introduction to Computing with Neural Nets," *IEEE ASSP Magazine*, Vol. 4, No. 2, April 1987, pp. 11-13.



TOM TARR received his undergraduate degree in electronics engineering from the University of Maryland. He has taken graduate courses in Johns Hopkins' computer science curricula, specializing in artificial neural systems. He is a senior electronics engineer at NAVSWC, involved in design and development of target recognition systems, and he directs efforts in design, development, and implementation of neural networks. His research interests are computer vision, pattern recognition, and advanced computational techniques in artificial intelligence, neural networks and fuzzy logic.



ALI FARSAIE received his graduate degrees from North Carolina State University, and was an assistant professor at the University of Maryland for six years. This was followed by an Office of Naval Technology post-doctoral fellowship at NAVSWC. He is now a chief engineer in the Center's Concepts and Technology Branch. His research interests include computer vision, target recognition, application of

artificial intelligence to problem solving, and artificial neural networks.

Target State Estimation and Prediction for Tactical Weapons Fire Control

W. D. Blair

The classic problem of tactical weapons fire control is the accurate prediction of the future position and velocity of a given target at the time it is to be intercepted or acquired by the weapon. The prediction of the position and velocity of a target is accomplished by extrapolating an estimate of the current target state a given time into the future with a motion model that is based on the expected behavior of the target. An interacting multiple model (IMM) algorithm with a constant velocity model and two mean-jerk models is proposed for estimating the states of a maneuvering target. The mean-jerk model is the standard constant acceleration model with time-correlated acceleration errors instead of the "white" acceleration errors that are usually assumed. The kinematic constraint for constant speed targets is included in one of the mean-jerk models, while a new technique for modeling the acceleration error relative to the target orientation is included in both mean-jerk models. Also, a new target state prediction algorithm that utilizes the target orientation and maneuverability is proposed for tactical weapons fire control. The prediction algorithm separates the target acceleration estimates into thrust and turning accelerations which are applied throughout the prediction time as parallel and orthogonal to the velocity vector, respectively.

The work described in this article was supported by the Independent Exploratory Development Program, Naval Surface Warfare Center.

Introduction

The classic problem of tactical weapons fire control is the accurate prediction of the future position and velocity of a given target at the time it is to be intercepted or acquired by the weapon. For unguided weapons, the fire control problem is the accurate prediction of the target position at the time of weapon intercept or impact. For this intercept prediction, an accurate prediction of the target velocity is required in order to compute an accurate time-of-flight of the weapon to the target. For guided weapons, the fire control problem is the accurate prediction of the target position and velocity at the time that the weapon is expected to acquire the target. For this acquisition prediction, an accurate prediction of the target velocity provides valuable information that can be used to compute a favorable placement of the weapon relative to the target at acquisition. The prediction problem is depicted in Figure 1, where the inbound missile maneuvers to prevent the outbound missile from intercepting it. The dashed line denotes the predicted trajectory of the target before the maneuver.

The prediction of the future state of a target is accomplished by extrapolating estimates of the current target state a given time into the future with a motion model that is based on the expected behavior of the target. Thus, when the estimate of the current target state is inaccurate, the predicted state for the target will be inaccurate. Therefore, the estimation of the current state of a target is critical to successful fire control.

The problem of estimating the state of maneuvering targets on the basis of position observations has been studied extensively since the mid-1960s.



Figure 1. Illustration of a target (inbound missile) maneuvering to prevent the outbound missile from acquiring it. The broken line denotes the trajectory predicted for fire control.

While the state of a non-maneuvering target can be accurately estimated or tracked with a constant velocity filter, the quality of the position and velocity estimates provided by the constant velocity filter can degrade significantly when the target maneuvers. One of the first works¹ that had a significant and lasting impact on the problem describes a target motion model with an acceleration that is exponentially correlated in time. In that work, a Kalman filter was utilized to estimate the position, velocity, and acceleration of the target. The rationale for the time-correlated acceleration was derived from inertial systems in that a target accelerating at time t is likely to be undergoing a similar acceleration at time $t+\tau$ for sufficiently small τ . However, the assumptions of exponentially correlated, zero-mean accelerations produced a motion with an acceleration that decreases in magnitude during each extrapolation of the state.¹ Thus, when the actual target acceleration is constant or increasing, large errors occur in the state estimates.

In the early 1970s, decision-directed tracking of maneuvering targets² was introduced in response to the demand for algorithms that could provide better tracking of targets performing high g, fast maneuvers. These decision-directed algorithms monitor the tracking errors to detect a maneuver and respond by increasing the modeling error covariance and/or the dimension of the target motion model.^{2,3,4} While these algorithms provide good tracking

performance before and after the maneuver, their performance during and immediately following the maneuver is poor. The problem with these algorithms is that the accelerations during a maneuver are not easily modeled in the sensor reference frame because the accelerations often vary irregularly with time in that frame. The accelerations were modeled in a target-oriented reference frame to reduce the problems associated with modeling the time-varying accelerations.⁵ While that algorithm provides relatively good tracking performance,⁵ the modeling and transformations that must be estimated produce a complicated algorithm. The two-stage Kalman estimator was applied to the tracking of maneuvering targets,⁶ and a significant reduction in the convergence time after a maneuver was achieved. However, the two-stage Kalman estimator⁶ is limited to a constant acceleration model with additive "white" noise errors. The input estimation algorithms^{7,8} were developed to address the problems of tracking time-varying accelerations and convergence time after a maneuver. However, the effective input estimation algorithms require significantly more computations than the previous techniques.

Recently, the idea of introducing a kinematic constraint into the tracking process through a pseudomeasurement was proposed⁹ to remove some of the uncertainty in the estimates of the time-varying accelerations, because the kinematic constraint provides additional

information about the target motion. Use of the kinematic constraint tends to force the acceleration estimates to change in a manner consistent with the dynamics of the target. However, use of the kinematic constraint as originally formulated⁹ results in marginal improvement in the accuracy of the target state estimates. A new formulation of the kinematic constraint for constant speed targets was proposed.¹⁰ The new formulation¹⁰ provides significantly better state estimates than the original one.⁹

While states of a non-maneuvering target can be accurately estimated with a constant velocity filter, the quality of the state estimates degrades significantly when the target maneuvers. Since threats with greater maneuverability can be expected in the future, state estimation algorithms are needed that can provide accurate estimates of the target state during maneuvers. To overcome this problem of estimating the states of maneuvering targets, the interacting multiple model (IMM) algorithm can be used.^{11,12} The IMM algorithm uses multiple models that interact through state mixing to track a target maneuvering through an arbitrary trajectory. The state estimates are mixed according to their model probabilities and the model-switching probabilities that are governed by an underlying Markov chain. The results of previous investigations indicate that the IMM filter is the superior technique for tracking maneuvering targets when the computational requirements of the techniques are considered.^{13,14}

While target trajectory prediction has applications in maneuver detection and aircraft avoidance, it is a vital part of any weapon system because it provides a preview of the target's probable course of travel over a given time interval. However, target trajectory prediction is very difficult because targets often perform evasive maneuvers, and the maneuvers cannot be predicted with great regularity. Trajectory prediction is usually achieved by utilizing the most recent estimate of the target's kinematic state as initial conditions in a state model which is extrapolated forward in time. Two methods commonly used for prediction are the constant velocity and constant acceleration predictors. A constant velocity predictor uses current estimates of target position and velocity while ignoring all acceleration information. The constant acceleration predictor

74 extrapolates the current state estimates assuming the acceleration is constant in each coordinate during the prediction time. This assumption is unrealistic because the acceleration vector is continually changing direction during a maneuver. Trajectory prediction for maneuvering targets can be significantly im-

proved by using the physical limitations of the target or assumptions about the dynamics and/or goal of the target.^{4,15} For example, aircraft motion during a maneuver is generally confined to a plane containing its velocity and acceleration vectors. For constant speed maneuvers, the acceleration vector of the aircraft is orthogonal to the velocity vector,¹⁰ and this restriction on the acceleration vector can be used in the prediction for targets undergoing a constant speed maneuver.¹⁵

In this article, an IMM algorithm is proposed for estimating the state of maneuvering targets for tactical weapons fire control. The proposed IMM algorithm includes a constant velocity motion and two mean-jerk models, which are the standard constant acceleration model with time-correlated acceleration errors instead of the white acceleration errors that are usually assumed. One of the mean-jerk models includes a kinematic constraint, while both mean-jerk models use a new formulation of the acceleration error covariance matrix that reflects the maneuverability of the target relative to its orientation. Also, a new algorithm that utilizes the target orientation and maneuverability is proposed for predicting the future states of maneuvering targets for tactical weapons fire control. The prediction algorithm utilizes the results¹⁵ to separate the target acceleration estimates into thrust and turning accelerations which are applied throughout the prediction time as parallel and orthogonal to the velocity vector, respectively.

This article is organized as follows. First, the problems of target state estimation and prediction are formulated. Then, techniques for modeling the motion of the targets are presented. A description of the IMM algorithm is given along with a discussion of its proposed implementation. The new algorithm for target state prediction is also presented. Last, a summary of the paper and a discussion regarding future work are given.

Problem Formulation

The target state prediction that is optimal in the minimum variance sense is obtained by extrapolating the optimal state estimate of the current state estimate with the expected state model of the target. Thus, the target state estimation and prediction problems will be formulated separately.

Target State Estimation

The dynamics model of a maneuvering target in track is given by

$$\dot{X} = f(X, u, w) \quad (1)$$

$$Z_k = h(X_k, v_k) \quad (2)$$

where X is the state vector, u is the target control vector, w is the process noise vector representing possible deviations in $f(\cdot)$, Z_k is the discrete-time measurement vector at time k , and v_k is the measurement noise vector. The dynamics of the target is a continuous-time process as indicated by Equation (1), where $f(\cdot)$ is a dynamic constraint that defines the motion for the target in the form of a differential equation. The dynamic constraint, which is usually unknown to the tracking system, can differ significantly between targets and change for a common target during the tracking process. As indicated by Equation (2), the measurement process is discrete-time because most sensors used for target tracking record the position and/or radial velocity at a given instant in time.

While $f(\cdot)$ is usually unknown by the tracking system, the major problem with tracking maneuvering targets is that the control vector is not directly observable by the tracking system. When the target applies a control, a bias or lag develops in the estimates of the target state. The control can be included as acceleration in the dynamic constraint $f(\cdot)$, but the acceleration most often varies with time in such a manner that a model cannot be clearly identified during tracking. Thus, the target dynamics are most often modeled as linear in a Cartesian coordinate frame to simplify the filtering and reduce the computations required. Also, for convenience the continuous-time dynamics equation is converted to a discrete-time system. As a result, the dynamics model commonly assumed for a target in track is given by

$$X_{k+1} = F_k X_k + w_k \quad (3)$$

where $w_k \sim N(0, Q_k)$ is the process noise and F_k defines a linear constraint on the dynamics. The target state vector X_k contains the position, velocity, and acceleration of the target at time k as well as other variables used to model the time-varying acceleration. When the target applies a control and the time-varying acceleration is modeled incorrectly, a bias or lag develops in the estimates of the target state. The IMM algorithm^{11,12} can be used to identify a combination of the models in the filter to represent the time-varying acceleration. Dynamics model j of the IMM is denoted by

$$X_{k+1}^j = F_k^j X_k^j + w_k^j \quad (4)$$

where $w_k^j \sim N(0, Q_k^j)$ is the process noise for model j , F_k^j defines a linear constraint on the dynamics imposed by model j , and X_k^j is the state associated with model j .

In this article, the measurement process of Equation (2) will be modeled as linear. While

most sensors used for target tracking measure the target position in a spherical coordinate system, the spherical measurements can be transformed into a Cartesian coordinate frame for processing as a linear function of the target state. The transformation from spherical to Cartesian coordinates is given by

$$x = R \cos E \sin B \quad (5)$$

$$y = R \cos E \cos B \quad (6)$$

$$z = R \sin E \quad (7)$$

where R, B, E are the range, bearing, and elevation, respectively. Thus, the linear measurement process is then given by

$$Z_k = H_k^j X_k^j + v_k \quad (8)$$

where Z_k is the target measurement in the Cartesian coordinate frame and $v_k \sim N(0, R_k)$ is the measurement error. While the measurement errors in spherical coordinates are usually assumed to be Gaussian and uncorrelated in range, bearing, and elevation, the transformation to the rectangular coordinate frame causes the components of v_k to become nongaussian and correlated. Nevertheless, v_k is often assumed to be Gaussian for probability calculations. The R_k is a full matrix, and the equation for computing R_k from the spherical error variances is given by

$$R_k = T \begin{bmatrix} \sigma_R^2 & 0 & 0 \\ 0 & \sigma_B^2 & 0 \\ 0 & 0 & \sigma_E^2 \end{bmatrix} T^T \quad (9)$$

where

$$T = \begin{bmatrix} \cos \bar{E} \sin \bar{B} & \bar{R} \cos \bar{E} \cos \bar{B} & -\bar{R} \sin \bar{E} \sin \bar{B} \\ \cos \bar{E} \cos \bar{B} & -\bar{R} \cos \bar{E} \sin \bar{B} & -\bar{R} \sin \bar{E} \cos \bar{B} \\ \sin \bar{E} & 0 & \bar{R} \cos \bar{E} \end{bmatrix} \quad (10)$$

and $\sigma_R^2, \sigma_B^2, \sigma_E^2$ are the measurement error variances for range, bearing, and elevation, respectively. The $\bar{R}, \bar{B}, \bar{E}$ are the predicted range, bearing, and elevation, respectively.

For linear state and measurement models and Gaussian error processes, the Kalman filter provides target state estimates that are optimal in the maximum likelihood and minimum variance senses. If the error processes are not Gaussian, the Kalman filter provides the state estimates that are optimal in the minimum variance sense. The Kalman filtering equation associated with the state model j in Equation (4) and the measurement model in Equation (8) are given by

Time Update:

$$X_{k|k-1}^j = F_{k-1}^j X_{k-1|k-1}^j \quad (11)$$

$$P_{k|k-1}^j = F_{k-1}^j P_{k-1|k-1}^j (F_{k-1}^j)^T + Q_k^j \quad (12)$$

Measurement Update:

$$S_k^j = H_k^j P_{k|k-1}^j (H_k^j)^T + R_k \quad (13)$$

$$K_k^j = P_{k|k-1}^j (H_k^j)^T (S_k^j)^{-1} \quad (14)$$

$$\tilde{Z}_k^j = Z_k^j - H_k^j X_{k|k-1}^j \quad (15)$$

$$X_{k,k}^j = X_{k|k-1}^j + K_k^j [\tilde{Z}_k^j] \quad (16)$$

$$P_{k|k}^j = [I - K_k^j H_k^j] P_{k|k-1}^j \quad (17)$$

where $X_k^j \sim N(X_{k|k}^j, P_{k|k}^j)$ with $P_{k|k}^j$ denoting the error covariance of the state estimate. The \tilde{Z}_k^j denotes the filter residual, and K_k^j denotes the Kalman gain.

Target State Prediction

Target state prediction is the problem of accurately forecasting the future kinematic state of the target for a given time $t_k + t_p$ when given measurements through the current time t_k . Target trajectory prediction is a very difficult problem because targets often perform evasive maneuvers which cannot be predicted with regularity. It can be shown that the optimal state prediction $X_p(t_k + t_p)$ is found by extrapolating the current optimal state estimate $X_{k|k}$ with the expected state model of the target. This extrapolation process is given mathematically as

$$X_p(t_k + t_p) = g(X_{k|k}, t_p) \quad (18)$$

where $g(\cdot)$ represents the expected motion model of the target. While the expected motion model can be chosen by various methods that utilize assumptions about the dynamics, physical limitations, and/or goal of the target,^{4,5,15,16} the most commonly used motion model for prediction is the motion model used to obtain the estimate $X_{k|k}$. Since accurate and reliable prediction methods are needed for maneuvering targets, all information should be included in the prediction process. For

example, the assumption of constant speed is incorporated in the turning rate predictor developed,¹⁵ while the goal information was incorporated into the minimum control predictor. Thus, the target state prediction problem is the problem of selecting the $g(\cdot)$ that best represents the future motion of the target.

Target Motion Modeling

The IMM algorithm proposed in this article will utilize a nearly constant velocity model and two mean-jerk models. One of the two mean-jerk filters will utilize the kinematic constraint for tracking constant speed targets, while both mean-jerk filters will utilize a target-oriented process noise vector. The first model to be presented is nearly constant velocity, and for a single coordinate the dynamic model is given by

$$F_k^1 = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \quad (19)$$

$$Q_k^1 = \begin{bmatrix} 0.25T^4 & 0.5T^3 \\ 0.5T^3 & T^2 \end{bmatrix} q_k^1 \quad (20)$$

where T is the time interval between measurement k and $k - 1$, q_k^1 is the variance of the acceleration error, and the superscript "1" denotes model 1. The elements of the state vector X_k^1 are position and velocity.

The second model is a mean-jerk model which is the standard constant acceleration model with time-correlated acceleration errors instead of the white noise errors that are usually assumed. The use of the time-correlated acceleration errors represents the fact that the acceleration modeling error at time k is influenced by the acceleration modeling error at time $k - 1$. When a target maneuvers, the acceleration changes in a deterministic manner and the errors in the acceleration estimates at time k will be related to those at time $k - 1$. The continuous-time form of the mean-jerk model in a single coordinate is given by

$$\dot{X}^2 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -\tau \end{bmatrix} X^2 + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} w^2(t) \quad (21)$$

where

$$X^2 = [x \ \dot{x} \ \ddot{x} \ n]^T \quad (22)$$

and n represents the time-correlated acceleration noise. The discrete-time form of the model is given by

$$F_k^2 = \quad (23)$$

$$\begin{bmatrix} 1 & T & 0.5T^2 & \tau^{-3}(0.5(\tau T)^2 - \tau T + 1 - e^{-\tau T}) \\ 0 & 1 & T & \tau^{-2}(e^{-\tau T} - 1 + \tau T) \\ 0 & 0 & 1 & \tau^{-1}(1 - e^{-\tau T}) \\ 0 & 0 & 0 & e^{-\tau T} \end{bmatrix}$$

$$Q_k^2 = B^2(B^2)^T q_k^2 \quad (24)$$

where

$$B^2 = \begin{bmatrix} \tau^{-4}(\frac{1}{6}(\tau T)^3 - 0.5(\tau T)^2 + \tau T - 1 + e^{-\tau T}) \\ \tau^{-3}(0.5(\tau T)^2 - \tau T + 1 - e^{-\tau T}) \\ \tau^{-2}(e^{-\tau T} - 1 + \tau T) \\ \tau^{-1}(1 - e^{-\tau T}) \end{bmatrix} \quad (25)$$

$$X_k^2 = [x_k \dot{x}_k \ddot{x}_k n_k]^T \quad (26)$$

and τ is related to the correlation of the acceleration errors. The time update of the acceleration is given by

$$\ddot{x}_k + 1 = \ddot{x}_k + u_k \quad (27)$$

where

$$u_k = \frac{1}{\tau}(1 - e^{-\tau T}) n_k + \frac{1}{\tau^2}(\tau T - 1 + e^{-\tau T}) w_k^2 \quad (28)$$

Assuming stationary statistics for w_k^2 , (i.e., $\text{VAR}[w_k^2] = q^2$),

$$\text{VAR}[n_k] = \frac{(1 - e^{-\tau T})^2}{\tau^2(1 - e^{-2\tau T})} q^2 \quad (29)$$

Using Equations (28) and (29) and assuming stationary statistics

$$q^2 = \frac{\tau^4(1 - e^{-2\tau T})}{(1 - e^{-\tau T})^4 + (1 - e^{-2\tau T})(\tau T - 1 + e^{-\tau T})^2} q_{acc} \quad (30)$$

where $q_{acc} = \text{VAR}[u_k]$. Thus, Equation (30) can be used to compute q^2 when a certain variance is desired for the acceleration error.

Kinematically Constrained Motion

A kinematic constraint can be utilized as additional information about the target motion to reduce the errors in the estimates of the time-varying accelerations. Using the kinematic constraint tends to force the acceleration estimates to change in a manner that is consistent with the dynamics of the target. A kinematic constraint can be developed for use in tracking constant speed, maneuvering targets. The speed of a target is given by

$$S = (\dot{x}^2 + \dot{y}^2 + \dot{z}^2)^{1/2} \quad (31)$$

For a target moving at a constant speed,

$$\frac{dS}{dt} = 0 \quad (32)$$

or

$$\dot{x}\ddot{x} + \dot{y}\ddot{y} + \dot{z}\ddot{z} = 0 \quad (33)$$

Equation (33) can be written as

$$V \cdot A = 0 \quad (34)$$

where V and A are the target velocity and acceleration, respectively.

The kinematic constraint for constant speed targets is useful information and can be incorporated in the system state in Equation (1) or used as a pseudomeasurement in conjunction with Equation (2). While both approaches are conceptually feasible, the second approach is more attractive because the first changes the state equation in Equation (3) from linear to nonlinear. In the implementation of the extended Kalman filter, including nonlinearities in the measurement equation is computationally less expensive than in the state equation.⁹ Thus, the kinematic constraint in Equation (34) is incorporated into the filter through a pseudomeasurement.¹⁰ The pseudomeasurement equation is given by

$$\frac{V_{k|k}}{S_{k|k}} \cdot A_k + \mu_k = 0 \quad (35)$$

where

$$V_{k|k} = [\dot{x}_{k|k} \dot{y}_{k|k} \dot{z}_{k|k}] \quad (36)$$

$$S = (\dot{x}_{k|k}^2 + \dot{y}_{k|k}^2 + \dot{z}_{k|k}^2)^{1/2} \quad (37)$$

and $\mu_k \sim N(0, R_k^\mu)$. The μ_k is a white Gaussian error process that accounts for the uncertainty in both the $V_{k|k}$ and the constraint. Since the initial estimates of $V_{k|k}$ may be very poor, R_k^μ is initialized with a large value and allowed to decrease as

$$R_k^\mu = r_1(\delta)^k + r_0 \quad (38)$$

where $0 < \delta < 1$, r_1 is a large constant chosen large for initialization, and r_0 is a constant chosen for steady-state conditions.

The filtering equations for this new formulation of the kinematic constraint are given in the following equations with $X_{k|k}^c$ denoting the state estimate after the constraint has been applied, and $P_{k|k}^c$ denoting the associated state error covariance. The filtering equations are

$$X_{k+1|k} = F_k X_{k|k}^c \quad (39)$$

$$P_{k+1|k} = F_k P_{k|k}^c F_k^T + Q_k \quad (40)$$

$$\tilde{Z}_k = Z_k - H_k X_{k|k-1} \quad (41)$$

$$X_{k|k} = X_{k|k-1} + K_k \tilde{Z}_k \quad (42)$$

$$P_{k|k} = [I - K_k H_k] P_{k|k-1} \quad (43)$$

$$X_{k|k}^c = [I - K_k^c C_k] X_{k|k} \quad (44)$$

$$P_{k|k}^c = [I - K_k^c C_k] P_{k|k} \quad (45)$$

where

$$K_k = P_{k|k-1} H_k^T [H_k P_{k|k-1} H_k^T + R_k]^{-1} \quad (46)$$

$$K_k^c = P_{k|k} C_k^T [C_k P_{k|k} C_k^T + R_k^\mu]^{-1} \quad (47)$$

$$C_k = \frac{1}{S_{k,k}} [0 \ 0 \ \dot{x}_{k|k} \ 0 \ 0 \ \dot{y}_{k|k} \ 0 \ 0 \ \dot{z}_{k|k}] \quad (48)$$

This filter using the kinematic constraint as a pseudomeasurement is shown to be uniformly asymptotically stable if the original system without the constraint is uniformly asymptotically stable.¹⁶ Also, if the original system is unbiased $Q_k > 0$, and $R_k^m > 0$, the kinematically constrained filter is unbiased.^{16,17}

Target Oriented Process Noise

Most targets maneuver by controlling their thrust acceleration, which is parallel to the target velocity vector, and/or their lateral acceleration, which is orthogonal to the target velocity vector. Except for missiles in launch or boost phases, the thrust acceleration of most targets is limited by about 1 g, while the lateral acceleration can approach 15 g. Thus, when tracking a maneuvering target, a wider range of uncertainty exists in the expected lateral acceleration than in the expected thrust acceleration. These different uncertainties in this expected acceleration can be included in the target state model of Equation (3) by selecting a process noise covariance Q_k that is dependent on the orientation of the target. A target-oriented process noise covariance is obtained by

$$Q_k = G_k D_k \begin{bmatrix} q^l & 0 & 0 \\ 0 & q^l & 0 \\ 0 & 0 & q^l \end{bmatrix} D_k^T G_k^T \quad (49)$$

where

$$D_k = \begin{bmatrix} \cos\phi \cos\theta & -\sin\theta \cos\theta \sin\phi \\ \sin\theta \cos\phi & \cos\theta \sin\phi \sin\theta \\ -\sin\phi & 0 & \cos\phi \end{bmatrix} \quad (50)$$

$$G_k = \begin{bmatrix} B_k & 0 & 0 \\ 0 & B_k & 0 \\ 0 & 0 & B_k \end{bmatrix} \quad (51)$$

$$\theta = \text{atan}2(\dot{y}_k, \dot{x}_k) \quad (52)$$

$$\phi = \text{atan}2(-\dot{z}_k, S_h) \quad (53)$$

$$S_h = (\dot{x}_k^2 + \dot{y}_k^2)^{1/2} \quad (54)$$

B_k is of the form of Equation (25), q^l is the variance for the expected thrust acceleration, and q^l is the variance for the expected lateral acceleration. For a measurement rate of 4 Hz, typical values for q^l and q^l are $1 \text{ m}^2/\text{s}^4$ and $9 \text{ m}^2/\text{s}^4$, respectively.

Target State Estimation

The IMM algorithm is proposed for estimating the states of a maneuvering target. The proposed IMM algorithm includes a constant velocity model and two mean-jerk models. Both mean-jerk models will utilize a target-oriented process noise covariance matrix, while one of the mean-jerk models will include the kinematic constraint for constant speed targets.

The IMM algorithm consists of a filter for each model, a model probability evaluator, an estimate mixer at the input of the filters, and an estimate combiner at the output of the filters. The multiple models interact through the mixing to track a target maneuvering through an arbitrary trajectory. The flow diagram of an IMM algorithm with two models is given in Figure 2, where $X(k|k)$ is the state estimate based on both models, $X_j(k|k)$ is the state estimate for time k based on model j , $\Lambda(k)$ is the vector of model likelihood at time k , and

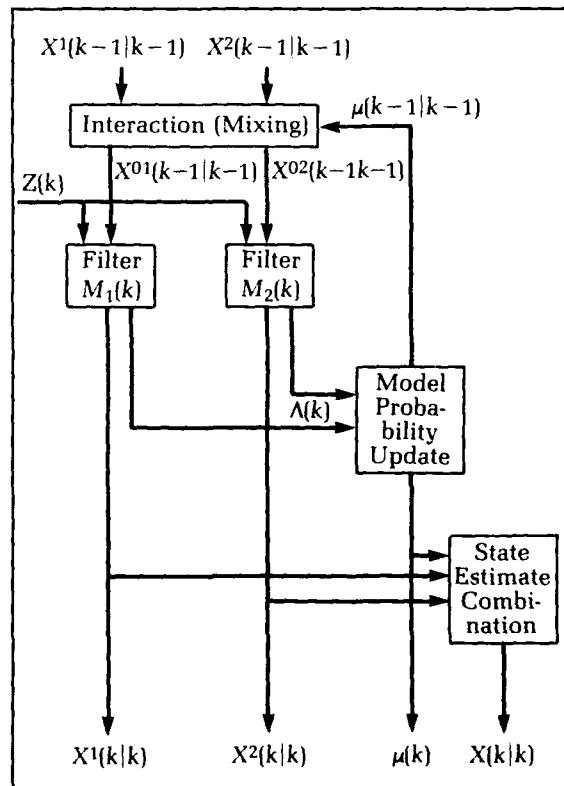


Figure 2. The IMM algorithm.

$\mu(k)$ is the vector of model probabilities at time k when all the likelihoods have been considered. With the assumption that the model switching is governed by an underlying Markov chain, the mixer uses the model probabilities and the model switching probabilities to compute a mixed estimate for each filter. At the beginning of a filtering cycle, each filter uses a mixed estimate and a measurement to compute a new estimate and a likelihood for the model within the filter. The likelihoods, prior model probabilities, and the model switching probabilities are then used to compute new model probabilities. The overall state estimate is then computed with the new state estimates and their model probabilities. For space consideration, subscripts will be used to denote time throughout the paper (i.e., $\mu_k(j)$ is the probability that model j is correct at time k).

The IMM algorithm for tracking with the $N = 3$ models is outlined in the following five steps. A derivation and detailed explanation of the IMM filter are given in the literature.^{12,13}

Step 1: Mixing of State Estimates

The filtering process starts with *a priori* state estimates $X_{k-1|k-1}^j$, state error covariances $P_{k-1|k-1}^j$, and the associated probabilities $\mu_{k-1}(j)$ for each model. The initial state estimate for model j at time k , $M_k(j)$, is computed as

$$X_{k-1|k-1}^{0j} = \sum_{i=1}^N X_{k-1|k-1}^i \mu_{k-1}(i) \quad (55)$$

where

$$\mu_{k-1}(i) = \frac{1}{\bar{c}_j} p_{ij} \mu_{k-1}(j) \quad (56)$$

$$\bar{c}_j = \sum_{i=1}^N p_{ij} \mu_{k-1}(i) \quad (57)$$

and p_{ij} is the assumed transition probability for switching from model i to model j , and \bar{c}_j is a normalization constant. The initial covariance for $M_k(j)$ is computed as

$$P_{k-1|k-1}^{0j} = \sum_{i=1}^N [P_{k-1|k-1}^i + (X_{k-1|k-1}^i - X_{k-1|k-1}^{0j})^T \mu_{k-1}(i)] \\ \cdot (X_{k-1|k-1}^i - X_{k-1|k-1}^{0j}) \quad (58)$$

Step 2: Model Updates

The Kalman filtering equations given in Equations (11) through (17) provide the update algorithm for the models without the kinematic constraint, where $X_{k-1|k-1}^i = X_{k-1|k-1}^{0j}$ and $P_{k-1|k-1}^i = P_{k-1|k-1}^{0j}$. The Kalman filtering

equations given in Equations (39) through (48) provide the update algorithm for the model with the Kinematic constraint, where

$$X_{k-1|k-1}^C = X_{k-1|k-1}^{0j} \text{ and } P_{k-1|k-1}^C = P_{k-1|k-1}^{0j}$$

Step 3: Model Likelihood Computations

The likelihood of $M_k(j)$ is computed with the filter residuals \tilde{Z}_k^j , the covariance of the filter residuals S_k^j , and the assumption of Gaussian statistics. The likelihood of $M_k(j)$ is given by

$$\Lambda_k(j) = \frac{1}{\sqrt{|2\pi S_k^j|}} \exp[-0.5(\tilde{Z}_k^j)^T (S_k^j)^{-1} \tilde{Z}_k^j] \quad (59)$$

Step 4: Model Probabilities Update

The model probabilities are updated as

$$\mu_k(j) = \frac{1}{c} \Lambda_k(j) \bar{c}_j \quad (60)$$

where

$$c = \sum_{i=1}^N \Lambda_k(i) \bar{c}_i \quad (61)$$

Step 5: Combination of State Estimates

The state estimate and covariance for the IMM filter output are obtained from a probabilistic sum of the individual filter outputs as

$$X_{k|k} = \sum_{i=1}^N X_{k|k}^i \mu_k(i) \quad (62)$$

$$P_{k|k} = \sum_{i=1}^N \mu_k(i) [P_{k|k}^i + (X_{k|k}^i - X_{k|k}) \\ \cdot (X_{k|k}^i - X_{k|k})^T] \quad (63)$$

Simulated tracking results are given in Figure 3 to demonstrate the tracking performance of the IMM algorithm against maneuvering targets. The results are compared with the simulated tracking results from a Kalman filter using a constant acceleration model. The IMM filter contains a constant velocity model and two constant acceleration models. Initially, the target is at a range of 50 km and moves with constant velocity from $t=0$ to $t=15$ seconds. The target maneuvers with a constant acceleration of about $3g$ from $t=15$ to $t=30$ seconds. As shown by the root-mean-square (RMS) errors in Figure 3, the IMM algorithm significantly reduces the errors in the position and velocity estimates below that of the Kalman filter. The measurement errors were Gaussian with standard deviations of 8 meters in range and 0.002 radian in bearing and elevation.

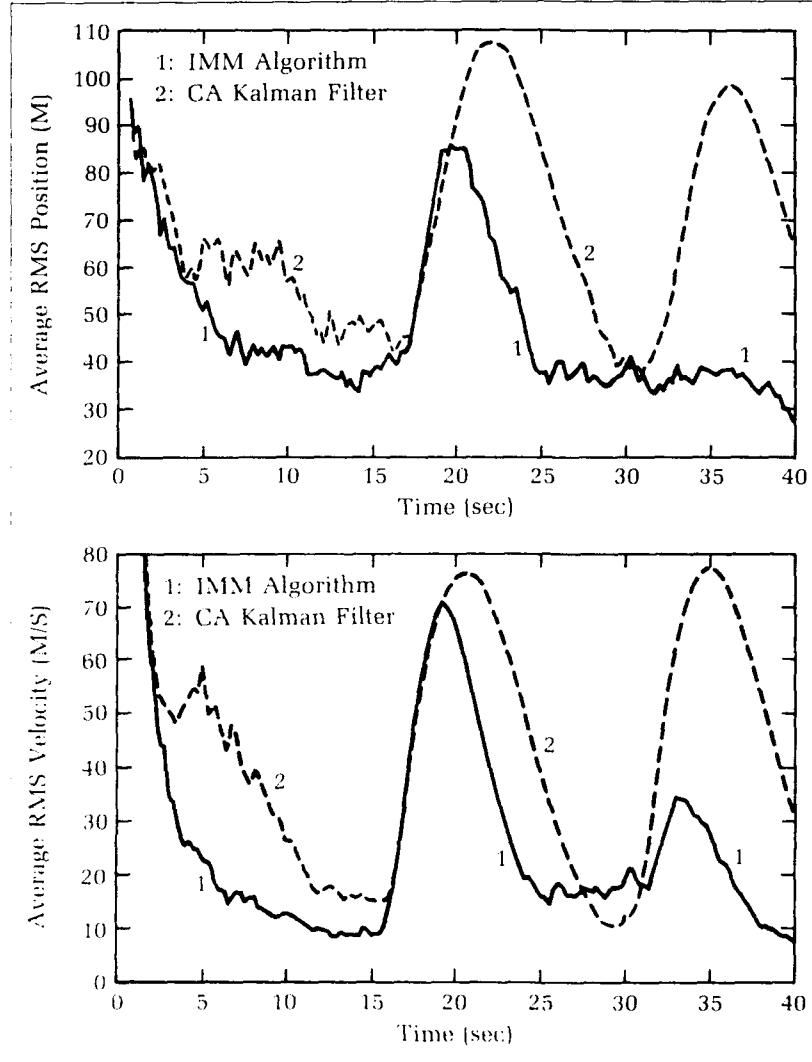


Figure 3. Simulated tracking results comparing the performance of the IMM algorithm and a Kalman filter utilizing a constant acceleration model.

Target State Prediction

The limitations of the constant acceleration predictor led to the development of the turning rate predictor,¹⁵ which uses the same filter output as the constant acceleration predictor to calculate a turning rate for the prediction process. The turning rate predictor assumes the acceleration and velocity vectors of the aircraft are orthogonal to accomplish constant speed prediction for maneuvering targets. In order to predict the future trajectory of targets

- 80 maneuvering with variable speed, the turning rate predictor¹⁵ is extended to accommodate a thrust component of acceleration. The prediction algorithm decouples the acceleration estimates into thrust and lateral components which are applied throughout the prediction process as parallel and orthogonal to the

velocity vector, respectively. The thrust acceleration is given by

$$A_{k,k}^t = \frac{(A_{k,k} \cdot V_{k,k})}{|V_{k,k}|^2} V_{k,k} \quad (64)$$

where

$$A_{k,k} = [\dot{x}_{k,k} \ \dot{y}_{k,k} \ \dot{z}_{k,k}]^T \quad (65)$$

$$V_{k,k} = [\dot{x}_{k,k} \ \dot{y}_{k,k} \ \dot{z}_{k,k}]^T \quad (66)$$

Thus, the lateral component of the acceleration is given by

$$A_{k,k}^l = A_{k,k} - A_{k,k}^t \quad (67)$$

The lateral component of the acceleration can be described with a turning rate given by¹⁵

$$w_{k,k} = \frac{V_{k,k} \times A_{k,k}^l}{|V_{k,k}|^2} \quad (68)$$

where

$$w_{k,k} = [w_{k,k}^x \ w_{k,k}^y \ w_{k,k}^z]^T \quad (69)$$

For constant turning rate prediction, $w_{k,k}$ is assumed to be constant through the prediction time t_p . Since the lateral acceleration changes the direction of the velocity vector, intermediate constant acceleration predictions are performed for small time increments t_S to accomplish the prediction. At each intermediate prediction, the new directions of the thrust and lateral accelerations are computed with the velocity vector. The size of t_S is important because the prediction is an integration process where t_S is the step size. This step interval is related to t_p by the relationship

$$t_p = n t_S \quad (70)$$

where n is the number of increments. After an intermediate prediction is performed for t_S , the turning rate and intermediate predicted velocities are used to calculate new lateral accelerations which are added to the thrust accelerations and inserted into the intermediate state vector for the next intermediate prediction with the constant acceleration predictor. This process of an intermediate prediction followed by an acceleration update is repeated for n iterations. Since the decoupled acceleration predictor is an iterative process, t_S must be chosen to provide acceptable integration of the motion model and minimize the number of repetitions needed for accurate prediction. The predictor is formulated as a sequence of intermediate state predictions at interval i , where $i = 1, \dots, n$, given by

$$X_i(t_k + t_S) = E(t_S)X_i(t_k + i - 1)t_S \quad (71)$$

where $E(t_S)$ is the state transition matrix for constant acceleration motion, $X_i(t_k)$ is the tracking filter output $X_{k,i}$ that initializes the predictor algorithm, and $X_i(t_k + t_S)$ is the output predicted state of the target. Prior to the

intermediate prediction using Equation (71), the intermediate accelerations for $t_k + it_S$ are determined using the turning rate given by Equation (69), the thrust acceleration given in Equation (64), and intermediate velocities $V_p(t_k + it_S)$ by

$$A_p(t_k + it_S) = w_{k,k} \times V_p(t_k + it_S) + \frac{V_p(t_k + it_S)}{|V_p(t_k + it_S)|} |A_{k,k}^l| \quad (72)$$

and inserted into $X_p(t_k + it_S)$. The process is repeated until $X_p(t_k + t_p)$ is computed.

Assuming that the turning rate and/or thrust acceleration are constant is justified for short prediction times, but an extended prediction time requires a time-varying turning rate and thrust acceleration because the magnitude of the accelerations will surely change over the longer prediction times. A time-varying turning rate and/or thrust acceleration are possible modifications to the decoupled acceleration predictor. For example, an exponentially decreasing turning rate and thrust acceleration are given by

$$w_p(i) = w_{k,k} e^{-it_S/\tau_p} \quad (73)$$

$$|A_p^l(i)| = |A_{k,k}^l| e^{-it_S/\tau_p} \quad (74)$$

where $w_{k,k}$ is defined in Equation (68), $A_{k,k}^l$ is defined in Equation (64), and τ_p is the decay rate of the accelerations. Equations (73) and (74) are then used to compute the turning rate and thrust acceleration for the iterative process of Equations (71) and (72).

Summary and Conclusions

A new version of the IMM filter that uses a constant velocity model and two mean-jerk models has been proposed for tactical weapons fire control. The new version is unique because it utilizes a kinematically constrained filter and a target-oriented process noise covariance matrix. While the use of a kinematic constraint in tracking constant speed, maneuvering targets has been studied previously for single model filtering,^{10-15,17} its use in multiple-model filtering is unique to this article. Also, while target state estimation utilizing the target orientation

was considered,⁵ the use of target-oriented process noise as presented in the Target Motion Modeling section herein is unique to this article. A new prediction algorithm that decouples the target acceleration estimates into thrust and lateral accelerations for use in the prediction process was also formulated. Throughout the prediction process, the thrust acceleration is applied parallel to the target velocity vector, while the lateral acceleration is applied orthogonal to the velocity vector. Since the lateral acceleration changes the direction of the velocity vector, the new predictor utilizes an iterative process to change the direction of the thrust and lateral accelerations during the prediction process.

Further research is needed to extend and refine the use of kinematic constraints and target-oriented process noise in state estimation for maneuvering targets. The kinematic constraint for constant speed targets should be extended to targets that maintain a constant speed except when changing altitudes. Also, methods of selecting the parameters that describe R_k^x of Equation (38) are needed for successful use of the kinematic constraint. The target-oriented process noise can be extended to include an error variance for the target acceleration that is orthogonal to the plane of motion of the maneuvering aircraft.

Further research into the decoupled acceleration predictor is needed to extend it to include the goal information that is available to a point defense system. The goal information of the target can be used to shape the predicted trajectory of target. For example, functions $f_1(t)$ and $f_2(t)$ that represent the expected future motion of the target can be used as

$$\begin{aligned} V_p(t_k + it_S) &= f_1(t_k + it_S) w_{k,k} \times V_p(t_k + it_S) \\ &+ f_2(t_k + it_S) \frac{V_p(t_k + it_S)}{V_p(t_k + it_S)} \cdot A_{k,k}^t \end{aligned}$$

where the functions $f_1(t)$ and $f_2(t)$ are based on the current estimate of the target state and the goal information.

References

1. Singer, R. A., "Estimation of Optimal Tracking Filter Performance for Manned Maneuvering Targets," *IEEE Trans. Aero. and Elect. Sys.*, 1970, pp. 473-483.
2. McAulay, R. J. and Delinger, E., "A Decision-Directed Adaptive Tracker," *IEEE Trans. Aero. and Elect. Sys.*, 1973, pp. 229-236.
3. Bar-Shalom, Y. and Birmiwal, K., "Variable Dimension Filter for Maneuvering Target Tracking," *IEEE Trans. Aero. and Elec. Sys.*, Sep 1982, pp. 621-629.
4. Clark, B. L., *Development of an Adaptive Kalman Tracking Filter and Predictor for Fire Control Applications*, NSWC TR-3445, Naval Surface Warfare Center, Dahlgren, VA, 1977.
5. Berg, R. F., "Estimation and Prediction for Maneuvering Target Trajectories," *IEEE Trans. Auto. Cont.*, Mar 1983, pp. 294-304.
6. Alouani, A. T., et al., "A Two-Stage Kalman Estimator for Tracking Maneuvering Targets," Proc. of 30th IEEE Conference on Decision and Control, Brighton, U.K., Dec 1991.
7. Chan, Y. T., Hu, A. G. C., and Plant, J. B., "A Kalman Filter Based Tracking Scheme with Input Estimation," *IEEE Trans. Aero. and Elect. Sys.*, Mar 1979, pp. 237-244.
8. Bogler, P. L., "Tracking a Maneuvering Target Using Input Estimation," *IEEE Trans. Aero. and Elect. Sys.*, May 1987, pp. 298-310.
9. Tank, M. and Speyer, J. L., "Target Tracking Problems Subject to Kinematic Constraints," *IEEE Trans. Auto. Cont.*, Vol. 35, No. 3, Mar 1990.
10. Blair, W. D., Watson, G. A., and Alouani, A. T., "Tracking Constant Speed Target Using a Kinematic Constraint," Proc. 23rd Southeastern Symposium on System Theory, Columbia, SC, 10-12 Mar 1991.
11. Blom, H. A. P., "An Efficient Filter for Abruptly Changing Systems," Proc. 23rd Conf. on Decision and Control, Las Vegas, NV, 1984, pp. 656-658.
12. Blom, H. A. P. and Bar-Shalom, Y., "The Interacting Multiple Model Algorithm for Systems with Markovian Switching Coefficients," *IEEE Trans. Auto. Cont.*, 1988, pp. 780-783.
13. Bar-Shalom, Y., Chang, C. Y., and Blom, H. A. P., "Tracking a Maneuvering Target Using Input Estimation Versus the Interactive Multiple Model Algorithm," *IEEE Trans. Aero. and Elec. Sys.*, 1989, pp. 296-300.
14. Tugnait, J. K., "Detection and Estimation for Abruptly Changing Systems," *Automatica*, 1982, pp. 607-615.
15. Watson, G. A. and Blair, W. D., "Constant Speed Prediction Using A Three Dimensional Turning Rate," Proc. 24th Southeastern Symposium on System Theory, Columbia, SC, 10-12 Mar 1991.

16. Blair, W. D. and Alouani, A. T., "Use of a Kinematic Constraint in Tracking Constant Speed, Maneuvering Targets," Proc. of 30th IEEE Conference on Decision and Control, Brighton, U.K., Dec 1991.
17. Alouani, A. T., Blair, W. D., and Watson, G. A., "Bias and Observability Analysis of Target Tracking Filters Using a Kinematic Constraint," Proc. 23rd Southeastern Symposium on System Theory, Columbia, SC, 10-12 Mar 1991.

The Author



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 NAVSWC in Dahlgren, Virginia, as an electrical engineer. At NAVSWC, his work has involved adaptive estimation for target tracking, multisensor alignment and data fusion, feedback control systems, and neural networks for functional mapping. Mr. Blair is 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.

Innovative Missile Airframe Concepts

Leon Schindel and Samuel R. Hardy

This article presents two examples of how innovative airframe design, integration, and enhanced airframe technologies can improve missile effectiveness. In the first example, a number of concepts are described that can improve the aerodynamic efficiency (lift/drag ratio) and maneuverability of long-range hypersonic missiles. Integration of the various concepts with advanced guidance, propulsion, warhead, and control concepts results in a high-performance, wide-area defense missile concept. In the second example, a conceptual airframe is formulated to meet far-term local-area defense requirements. Lift generated by large dorsal wings enables the missile to achieve fast time responses and high maneuverability. However, due to launcher constraints, critical vibration/autopilot instability limits missile responsiveness. Robust autopilot design and active vibration control technology tasks are under way to alleviate the vibration-induced time-constant limitations.

Introduction

In order to perform its mission, a guided missile must have sufficient range to reach its target, it must be guided to intercept, and it must then destroy the target. In this context, the missile airframe is merely a device for holding the propulsion system, the guidance, and the warhead together. To increase range, you add fuel; to increase velocity, you make the engine thrust greater; to reduce miss distance, you design a more complex guidance system; and to increase kill probability, you use a bigger warhead.

However, the effectiveness of the missile also depends critically on the aerodynamic performance of the airframe. Figure 1, for example, shows payload weight fraction of a typical supersonic missile with airbreathing propulsion as a function of lift/drag ratio. In level flight, the lift on the airframe balances the missile weight, while the engine must supply sufficient thrust to overcome the drag. For a given missile weight, then, increasing the lift/drag ratio (a function of airframe geometry) means that the propulsion system can be smaller and will burn less fuel. At a lift/drag ratio of about 1.7, the payload (warhead and guidance) is zero, so the missile can reach its target but cannot do anything when it gets there. Increasing the lift/drag ratio to 2 results in a payload fraction of about .1, so that a 1000-pound missile would be required to deliver a 100-pound payload. At a lift/drag ratio around 4, however, the payload increases to a respectable 40 percent of the total missile weight. Thus, the airframe shape is of primary significance in the performance of long-range missiles.

Similarly, missile maneuverability and time constant are also functions of airframe configuration. As another example, Figure 2 shows how miss distance depends on airframe maneuverability.¹ The maneuverability is critically dependent on the aerodynamic characteristics of an airframe as well as the missile weight and balance. Thus, a well-designed airframe can be an important contribution to guided missile effectiveness.

In this article, two innovative airframe design concepts will be described. Their application to conceptual long-range and local-area defense missiles will then be indicated. In the first case, the achievement of high lift/drag ratio is of primary importance in transporting a large payload to a

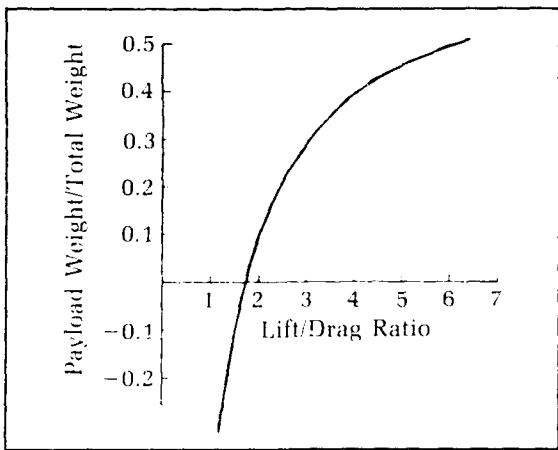


Figure 1. Effect of L/D on cruise missile weight.

distant target. In the latter case, the agility of the missile is of primary importance in the antiair-warfare mission. The dynamic performance of the missile is related to the aerodynamic, structural, and control characteristics of the airframe. In both examples, the value of effective airframe design and integration will be demonstrated.

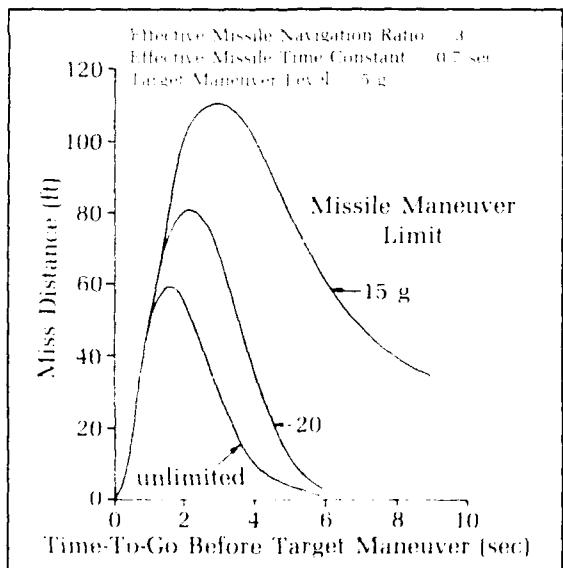


Figure 2. Effect of airframe maneuverability on miss distance.

Long Range Missile Concept

Shown in Figure 3 are a number of airframe design concepts that contribute to high lift/drag ratio or high maneuverability. Each of the con-

cepts will be briefly described. Additional details can be found in an earlier paper.² This missile, designed to cruise approximately 900 nautical miles at Mach number 6 at an altitude of 100,000 feet, incorporates the following features:

- Noncircular body (high lift/drag ratio)
- Propulsion/airframe integration (high lift/drag ratio)
- "Waverider" wings (high lift/drag ratio)
- Sliding canard controls (maneuverability)
- Swiveling plug nozzle (maneuverability)
- Conformal phased-array radar (high lift/drag ratio)

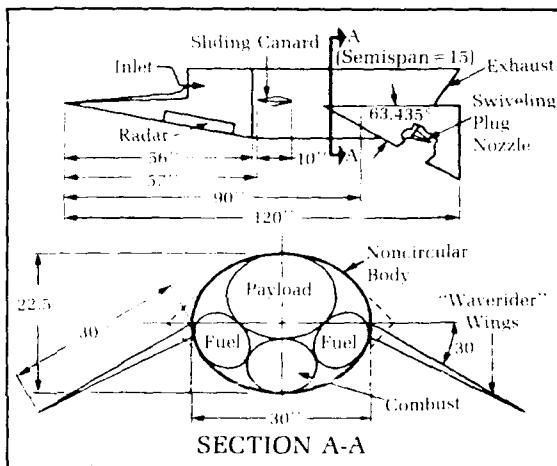


Figure 3. Features of hypersonic airframe.

Noncircular Body

As a general rule, the lift of a body increases with its planform area, while its wave drag increases with cross-sectional area. Thus for high lift/drag ratio, a flattened cross section is desirable. However, constraints imposed by launcher geometry, structural efficiency, and friction drag tend to push the configuration toward a round cross section. The performance analysis assumes a 4:3 ratio of body width to height, but that is probably not optimum. Extending the side edges of the body to a point, as shown by the dotted lines in Figure 3, might increase the lift without having much adverse effect on drag or structural weight.

Propulsion/Airframe Integration

An airbreathing engine (ramjet/scramjet) has two significant performance advantages over a rocket for long-range hypersonic flight. It has a

higher specific impulse, which for long range more than compensates for its heavier engine weight.

The other important consideration is that compression of the air entering the inlet contributes to engine thrust. Only the air which does not enter the inlet contributes to nose pressure drag. Furthermore, since the nozzle is designed to fill the missile base area, base drag is effectively eliminated.

In this concept, then, the airbreathing engine reduces the body drag so that maximum lift/drag ratio occurs at relatively low lift coefficient, permitting flight at reasonable altitudes (below 100,000 ft).

Waverider Technology

The entire configuration could be optimized by waverider techniques for maximum lift/drag ratio.^{3,4,5} However, the main advantage of the waverider at hypersonic speeds is that it traps high-pressure flow on the underside of the configuration. Any triangular wing with supersonic leading edges performs much the same function. This technology is thus incorporated in the design shown in Figure 3.

Sliding Controls

To improve maneuverability, the design incorporates a sliding canard control illustrated in Figure 4. The canard control has the potential for faster missile response and greater missile acceleration than wing or tail controls, arising from the fact that the canard force is in the same direction as the body force, while the force on a tail control opposes the acceleration of the vehicle flight path.

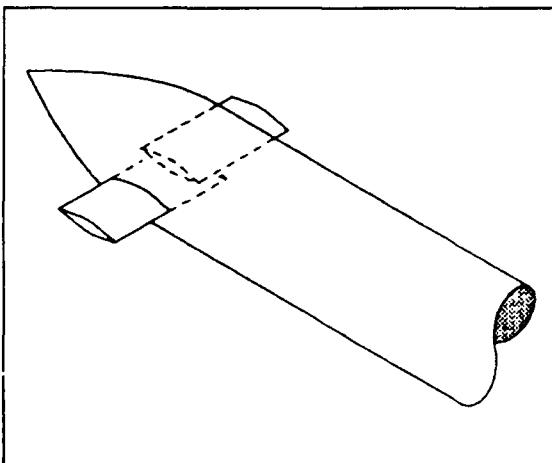


Figure 4. Sliding controls.

Unfortunately, canard deflection must be limited at high missile angle of attack to avoid stall of the control surface. However, in the sliding control concept no rotational deflection motion is allowed. Instead, the sliding control surface is set at some modest deflection angle. The control force is altered by changing the amount of control surface exposed to the flow. The sliding motion of the control surface does not alter the angle of attack of the canard, helping to avoid stall of the control surface. The canard surfaces slide into and out of the body symmetrically for pitch control and asymmetrically for roll. Setting the tail at a small positive angle of attack helps to trim the vehicle, increases the maximum lift/drag ratio, and provides a negative pitching moment when the canards are fully retracted.

The same canard control effectiveness could be achieved by installing conventional deflecting canard surfaces, whose area would be equal to that of the sliding controls in their fully extended position, and deflecting them in the conventional manner. Large control surface area is undesirable, however, because of the probability of encountering high hinge moments requiring correspondingly powerful actuators. Actuators for the sliding controls, on the other hand, need only overcome bearing friction which can be made small; however, this design issue has not been addressed. Thus, the main value of the concept might turn out to be in the development of all-electric missiles, since the low actuator force requirement could make electric motors competitive with hydraulic or pneumatic drivers.

Swiveling Plug Nozzle

A conventional contoured nozzle could be replaced or augmented by a two-dimensional plug nozzle. Furthermore, the plug could be rotated to deflect the thrust vector to provide trim or pitch control. It could be split and the halves deflected differentially to supply rolling moments as shown in Figure 5. Experiments indicate that the thrust can be adequately deflected by such a device.⁶

Guidance/Airframe Integration

To achieve high lift/drag ratio, the airframe must be relatively flat, and hence bank-to-turn guidance is implied. In addition, a conformal phased-array radar permits greater flexibility in the design of a missile nose than would a conventional circular-dish antenna. Interferometric antenna designs might also be used advantageously.

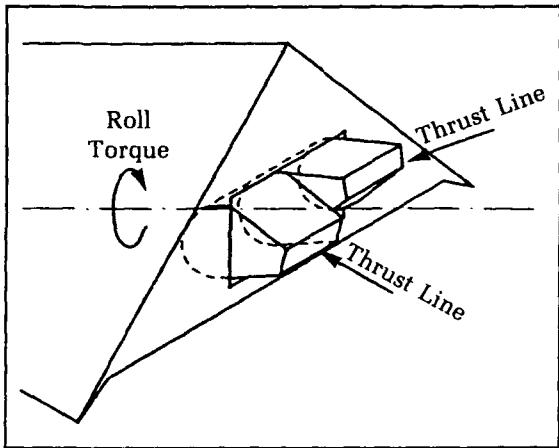


Figure 5. Split two-dimensional plug nozzle.

Performance Estimates

Lift/Drag Ratio

Figure 6 shows plots of estimated lift/drag ratio and corresponding lift coefficient as functions of angle of attack calculated with a simple noncircular body code.⁷ The configuration is shown in Figure 3. Flight is at Mach = 6, and Reynolds number based on body length = 6.6×10^6 . The estimated lift/drag ratio exceeds 4.8 at angles of attack between 3.5 and 6°. The corresponding lift coefficients, based on maximum body cross-sectional area, are .70 to 1.03. The sliding canard is set at 6° deflection with a semispan of 15 inches (1/2 the body width). The wings are set at 2° incidence providing a trimming moment when the canards are retracted.

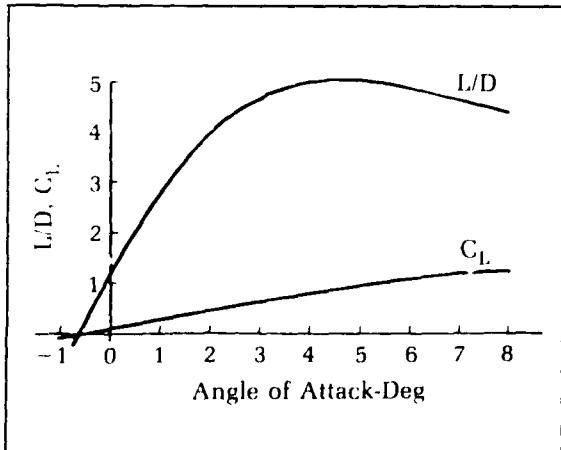


Figure 6. L/D and lift coefficient at $M = 6$.

If the initial weight of the vehicle is assumed to be 1900 pounds, then it can cruise over 900 nautical miles at an altitude of about 100,000 feet with a lift/drag ratio of 4.8 and a lift coefficient of .7 to .9. Table 1 shows a possible weight breakdown, assuming a specific impulse = 800 seconds. Based on a one-dimensional performance estimate,⁸ about 40 percent of the air traversing the nose frontal area is swallowed by the scramjet engine. The nozzle exhaust, fully expanded, fills the missile base area. A specific impulse of about 800 seconds appears appropriate for conventional hydrocarbon fuels.

Table 1. Weight Breakdown

Component	Weight (lb.)
Fuel	470
Warhead	420
Guidance	200
Structure	500
Propulsion	200
Other	110
TOTAL	1900

Range = 900 Nmi
Specific Impulse = 800 Sec.

Ducted-Rocket Booster

The missile is assumed to be boosted vertically from ground to 50,000 feet in order to accelerate it to a ramjet takeover velocity. The booster parameters are shown in Table 2. The

Table 2. Booster Parameters

Fuel Weight	2000 lb.
Structure Weight	200 lb.
Burning Rate	100 lb./sec.
Specific Impulse	300 sec.
Thrust	30,000 lb.
Final Velocity	4150 ft./sec.
Final Altitude	50,000 ft.
Flight Time	20 sec.

assumed specific impulse of 300 seconds is somewhat high for a solid fuel. However, since the rocket exhaust is normally fuel rich, thrust can be augmented by air addition (a "ducted rocket"). This promising technology appears to be currently unexploited.

High Performance Missile Airframe Conceptual Design

The focus of the High Performance Missile project is to provide required technologies to support Local-Area Defense (LAD) missiles that can defeat the threats in the 2010+ time frame. The driving threats are the very low observable subsonic and highly maneuverable supersonic sea-skimming anti-ship missiles. These threats require LAD missiles to have high average speeds and high maneuverability, well beyond current missile capabilities. In addition, significantly shorter terminal homing times are projected even with advanced RF seekers. As a consequence, homing time constants will be approximately one-half of the current capabilities, while maneuver level requirements are nearly doubled. In order to increase ship magazine loadout and ultimately to improve ship firepower, LAD missiles are required to be compatible with the Vertical Launching System (VLS).

The objective of this effort is to determine critical airframe technologies needed to satisfy the requirements and constraints mentioned above.

In this task, conventional aerodynamic controls appeared to be attractive since they take advantage of the high dynamic pressure environment that would exist for most of the missile's flight. Low altitude intercepts and high average velocities are required. It was decided to begin formulation of a concept using aerodynamic tail controls. The airframe was designed to support semi-active RF seeker guidance. It was recognized that there is a fundamental tradeoff between the seeker and the autopilot time constants since both contribute to the terminal homing time constant and miss distance. In previous designs, the autopilot time constant is approximately one-half of the overall missile homing time constant. Based on this, autopilot time constant reductions of approximately 50 percent of current capability are required.

Results from a nonlinear rigid body autopilot simulation (Figure 7) indicate that the body-tail concept sized for the mission could not achieve

88

the desired time constant goal when executing a maximum maneuver step command. Autopilot instability occurs when the tail control actuator rates are not high enough to support the autopilot bandwidth. Current actuators can provide only about one-half of the required rate. It did not seem reasonable to expect actuator technology to double current fin rates without significantly increasing actuator size and cost. Therefore, alternate airframe designs were considered. The sliding canard concept, which addresses this problem, is not applicable to this skid-to-turn missile design.

Although not readily apparent, increasing the aerodynamic gain of the body-tail airframe appeared to be an attractive solution to the fin rate limited response problem. Adding wings is one method of increasing the aerodynamic gain. A missile with wings can meet the maneuver requirements at lower angles of attack. As a result, the pitch rates and corresponding fin deflection rates can be reduced. Autopilot analyses showed that if the normal force coefficient for small angles of attack could be increased by a factor of two, the desired autopilot time constants could be attained. An added benefit of increasing the aerodynamic gain is the reduction of parasitic feedback due to the radome boresight error slope, which can reduce the overall missile homing time constant. Since the VLS constrained the semispan of wings and fins to one-caliber (8 inches) long, dorsal wings were investigated. Based on estimates from AEROPREDICTION,^{9,10} NANC,¹¹ and

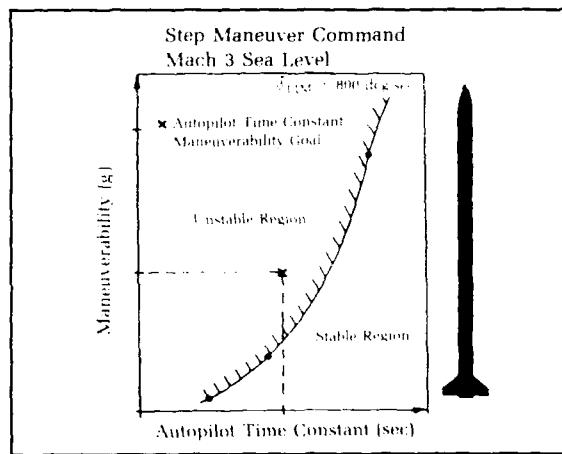


Figure 7. Effect of fin rate limit on body-tail airframe maneuverability.

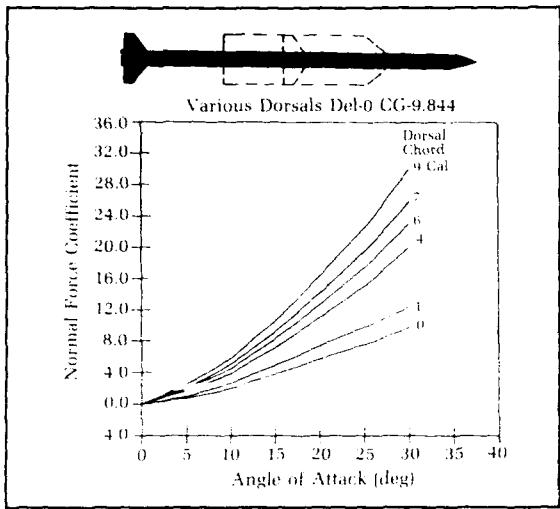


Figure 8. Increase in normal force coefficient for various dorsal lengths.

MISSILE3¹² codes shown in Figure 8, six-caliber chord length dorsals nearly double the normal force (at a given angle of attack) compared to a body-tail missile without a dorsal. It was estimated that the six-caliber chord length dorsal can be positioned on the missile so that the configuration is nearly neutrally stable without blocking warhead fragmentation. A preliminary design of a self-erecting dorsal fin and hinge mechanism was investigated to determine the added weight of the dorsal wings. The dorsals were designed to carry a maximum loading distribution computed using the ZEUS^{13,14} code. The body-dorsal-tail concept was then resized using solid rocket design codes to achieve the range and

average speed requirements. Figure 9 indicates that the body-dorsal-tail concept can achieve the desired time constant goal with state-of-the-art actuators.

The folding dorsal wings create other airframe concerns. Because the spans of the tail and the dorsals are limited by the launcher, the span of the tail cannot be increased to obtain the desired roll control authority. The dorsals increase the induced rolling moments which, in turn, require additional roll control as the angle of attack is increased. However, the simulations indicate that the body-dorsal-tail can meet maneuverability requirements at low angles of attack. Estimates of the induced roll moments and roll control authority were made using the MISSILE3 code and later verified in a wind tunnel test. Adequate roll control authority was confirmed up to moderate angles of attack. Unfolding dorsals during the launch and pitchover flight phases is another concern. The weight and reliability of the hinge mechanism are design issues. Another set of folding fins on the booster is required in order to deploy the dorsals immediately after exiting the VLS. Otherwise, excessive aerodynamic instability will limit the pitchover rates at launch. High pitchover rates are needed to enable the missile to intercept at minimum ranges. It should be pointed out that composite airframe materials were required to meet the performance goals. Figure 10 is an internal drawing of the conceptual airframe.

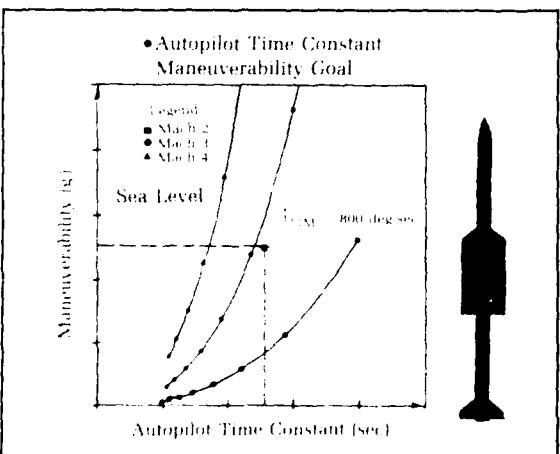


Figure 9. Effect of fin rate limit on body-dorsal-tail airframe maneuverability.

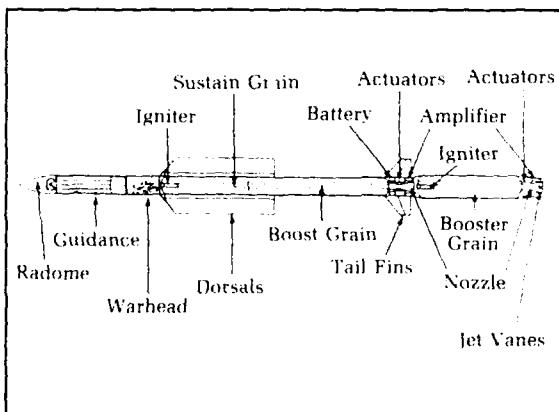


Figure 10. High performance missile, internal layout.

It is also recognized that increasing the autopilot bandwidth in order to achieve the desired time constant will reduce the frequency separation between the first body-bending mode and the autopilot bandwidth. The autopilot rate gyro and accelerometer sense both the rigid

body and natural bending mode vibration of the airframe, thus corrupting the rigid body information that is used by the autopilot to control the airframe. The longer airframes (due to launcher constraints) have lower bending mode frequencies, thus exacerbating the frequency separation problem. Traditionally, when there is sufficient frequency separation between the first bending modes and the autopilot bandwidths, notch filters may be used to cancel the flexible-body feedback that enters the autopilot through the rate gyros and accelerometers. The stability of the autopilot without sufficient separation from the low frequency first bending mode is a critical problem. A linear flexible-body autopilot analysis was used to investigate the stability. The results revealed that the desired autopilot time constant could be obtained at some conditions; however, the gain and phase margins were not adequate for robust performance. Figure 11 shows that even when there is perfect notch filter cancellation of the first bending mode, unwanted dynamics can remain in the control system. Airframe vibration is a critical problem that must be addressed in order to reduce missile time constant by any significant amount. For this reason, three technology approaches were identified.

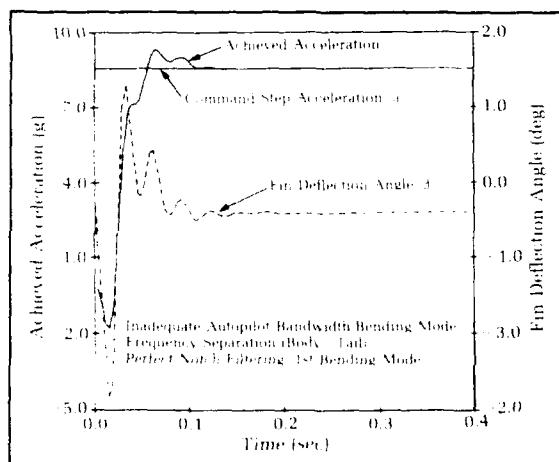


Figure 11. Time response with unacceptable dynamics due to airframe vibration.

The primary problem with airframe vibration is that the natural bending mode damping ratio is usually very small. It is believed that composite fiber-wound airframe components can be fabricated to increase the damping ratio significantly. This solution will require fabrication and experimental modal analysis to verify the actual damping ratios achieved. Since

others, particularly in industry, were already investigating composites, it was decided to concentrate on active vibration control and robust control approaches. These approaches are high risk, but have higher payoffs with many potential applications.

Efforts are being pursued to develop an independent mechanical control system concept that will dampen the natural bending modes. This concept is similar to the approaches being pursued for active vibration control of lightweight space structures. The missile active vibration suppression system is envisioned as an independent system including sensing of the vibration, control logic processing, and actuating components. The missile airframe is a rather heavy structure when compared to the space structures. As a result, a more powerful actuator is required. Magnetostrictive materials are being investigated as actuating components since they have nearly an order of magnitude more striction and energy density than ceramic piezoelectric materials. One possible implementation of the active system is to place the magnetostrictive actuators in missile joints (Figure 12) so that the actuator forces counteract the bending modes. Efforts are ongoing to demonstrate active vibration control feasibility using a single degree-of-freedom system for a simple beam.

Another potential solution to the vibration is to apply less conservative, more robust autopilot design methods. H-Infinity and Mu-Synthesis¹⁵ methods are expected to yield more robust autopilot design. Mu-Synthesis methods will allow stable designs with greater gain and phase margins than the more conservative classical approach.

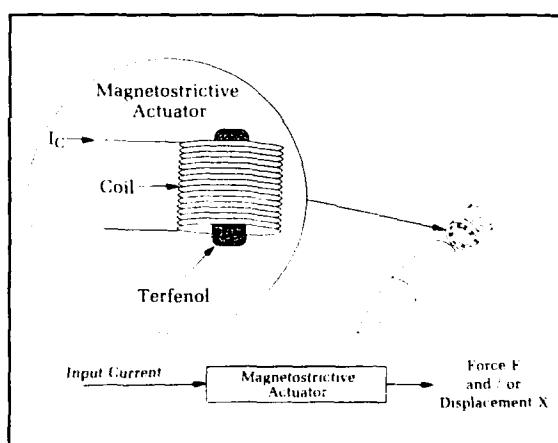


Figure 12. Conceptual airframe vibration suppression system.

Conclusions

Two examples of airframe concepts aimed at improved aerodynamics performance at both long and short range were presented. Enhanced aerodynamics, structures, flight mechanics, and control system design/analysis tools were used in the examples presented. The design tools are a critical part of demonstrating the feasibility of conceptual airframe designs in a cost-effective manner. Effective analysis tools are essential in the testing of innovative airframes by reducing the size of test programs that must be conducted and by estimating performance at conditions that cannot be tested. Innovative airframe design and integration supported by advanced airframe technology can improve missile performance, avoiding much of the cost of increasing propulsion, warhead, or seeker performance.

References

1. Nesline, F. W., "Missile Guidance for Low Altitude Air Defense," AIAA Paper 78-1317, Aug 1978.
2. Schindel, L. H., "Innovative Hypersonic Missile Airframe Design Concepts," AIAA Paper 91-0202, Jan 1991.
3. Schindel, L. H., "Waveriders," Chapter in *Tactical Missile Aerodynamics*, AIAA Progress in Astronautics and Aerodynamics, Vol. 104, Edited by J. N. Nielsen and M. J. Hemsch.
4. Corda, S., and Anderson, J. D., Jr., "Viscous Optimized Waveriders Designed for Axisymmetric Flow Fields," AIAA Paper 88-0369, Jan 1988.
5. Kim, B. S., Rasmussen, M. L., and Jischke, M. C., "Optimization of Waverider Configurations Generated from Axisymmetric Conical Flows," AIAA Paper 82-1299, Aug 82.
6. Krupins, M. V., Ragsdale, W. C. and Schindel, L. H., "Thrust Vectoring with a Two Dimensional Plug Nozzle," Naval Surface Warfare Center, 1984.
7. Schindel, L. H., "A Preliminary Design and Screening Process for Missile Airframe Configurations," AIAA Paper 87-0211, Jan 1987.
8. Schindel, L. H., "Design of High Performance Ramjet or SCRAMJET Powered Vehicles," AIAA Paper 89-0379, Jan 1989.
9. Devan, L., *Aerodynamics of Tactical Weapons to Mach Number Eight and Angle of Attack 180° - Part I - Theory and Application*, NSWC TR 80-346, Oct 1980.
10. Devan, L. and Mason, L., *Aerodynamics of Tactical Weapons to Mach Number Eight and Angle of Attack 180° - Part II - Program and Users Guide*, NSWC TR 81-358, Sep 1981.
11. Devan, L., *Nonaxisymmetric Body Supersonic, Inviscid Dynamic Derivative Prediction*, NSWC TR 89-99, Jun 1989.
12. Lesieurte, D. J., et al., *Prediction of Aerodynamic Characteristics of Cruciform Missiles Including Effects of Roll Angle and Control Deflection*, NEAR TR 360, Aug 1986.
13. Wardlaw, A. B., Jr. and Davis, S. R., *A Second Order Godunou Method for Supersonic Tactical Missiles*, NSWC TR 86-506, Dec 1986.
14. Wardlaw, A. B., Jr. and Priolo, F. J., *Applying the ZEUS Code*, NSWC TR 86-508, Dec 1986.
15. Bibel, J. E. and Stalford, H., "Mu-Synthesis Autopilot Design for a Flexible Missile," AIAA 91-0586, 29th Aerospace Sciences Meeting, Reno, Nevada, Jan 1991.

The Authors



LEON H. SCHINDEL received a B.S., 1945, M.S., 1948, and Sc.D., 1960, in Aerospace Engineering, Massachusetts Institute of Technology and was a member of the research staff, 1948-1967. At NAVSWC since 1967, he served as Head, Aerodynamics Department, 1967-1975; Head, Missiles and Decoys Division, 1975-1976; Associate Head, Advanced Weapons Department, 1976-1978; research associate, Weapon Dynamics

Division, 1978. He has performed research in missile and aircraft aerodynamics, magneto-aerodynamics, vortex-induced oscillations, separation of stores from aircraft, wind tunnel design, and heat and mass transfer. He is a technical consultant on missile aerodynamics, facility design, and related fields. He holds patents on thrust vectoring and radome cooling concepts. Dr. Schindel is a member of AIAA, Sigma Xi, and the Navy Aeroballistics Committee. He currently is developing design codes for air-breathing and rocket-propelled missiles and examining the feasibility of a high-temperature flow facility at NAVSWC.



SAMUEL R. HARDY received a B.S. 1972, M.S. 1980, in Aerospace Engineering Virginia Polytechnic Institute and State University. He has been a member of the Aeromechanics Branch since 1972, where he has developed flight simulations of various weapons, including guided missiles, rockets, and bombs; conducted stability and control wind tunnel tests of rockets, bombs, and missiles, including a hypersonic air-breathing missile. He is a member of AIAA. He currently is principal investigator for Surface Launched Weaponry Aerodynamics and Structures Technology.

The Use of Holographic Interferometry for Flow Diagnostics

W. C. Spring, III, W. J. Yanta and K. U. Gross

The use of interferometry as a viable diagnostics tool for aerodynamics research has been through several cycles of popularity in the last five decades. It has always held great potential for providing an enormous amount of information about a flow field, but the greatest difficulty has been the retrieval of data from the interferograms. With the advent of lasers, high-speed desktop computers with increased memory capacity, and high-resolution image processing techniques, interferometry has become a useful, efficient tool for providing valuable data about complex flow fields around missile and airframe configurations being tested in wind tunnels. This article gives a brief history of interferometry, the development of holographic interferometry, the Naval Surface Warfare Center's (NAVSWC's) contributions to the development of holographic interferometry for aerodynamic testing, its impact on accurate modeling of fluid and gas flows, and some of the areas of future research.

Introduction

Interferometry has been used for many years as a nonintrusive means of measuring densities in a transparent medium,¹ such as the flow around models during wind tunnel testing (Figure 1). The interferometer is sensitive to differences in optical path length between two light beams, for example, where one beam has passed through the wind tunnel and the other has passed around it. These optical-path-length differences are caused by changes in the index of refraction of the transparent gas. These changes of index are directly related to the changes in density of the gas. The path-length differences are made visible by the interference fringes in the image (Figure 2). These fringes are a function of the wavelength of light used to make the interferogram, hence they are a known measurement reference.

Prior to the advent of the laser, classical interferometry had been very difficult, particularly in the relatively hostile environment of wind tunnel test facilities. Interferometry requires a characteristic of light called coherence. Normal light sources, to be sufficiently coherent for interferometry, must be highly filtered, which means that most of the output energy is not available for making the interference patterns.

The laser provides a light source that has high energy levels of highly coherent light. The optical and mechanical constraints of classical interferometry can be greatly eased by the use of lasers. The technique of coherent imaging called holography (also made practical by lasers) can be used to make interferometry even more versatile and convenient. In addition, the availability of high-speed, desktop computers with large memory capacity in the last several years has allowed the use of dedicated data processing systems that make the modern techniques for data reduction from these interferograms practical.

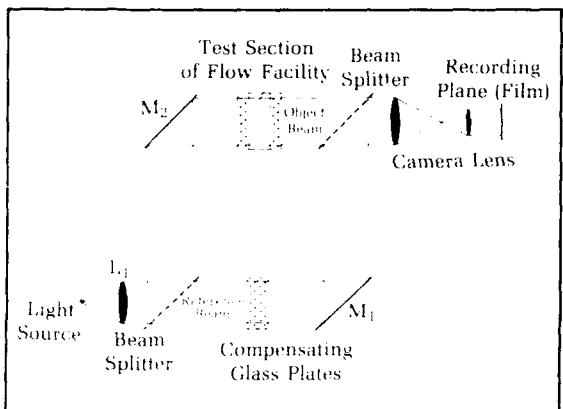


Figure 1. Mach-Zender interferometer schematic.

Holographic Interferometry

Background

Lasers have made a technique known as holography a practical tool for the scientific researcher.^{2,3,4} Holography is the recording of a unique interference pattern on a photo-sensitive plate such that this pattern can be reconstructed to regenerate the original optical wave. In much the same way as with the interferometer, the output from the laser light source is split into two beams. One, the object beam, is reflected off or transmitted through the object; the second, the reference beam, is directed around the object (Figure 3). Both beams are then recombined in the plane of a photo recording medium (usually a film plate). Because of the nature of the interference pattern, both the amplitude and the phase of the light waves are recorded. When the hologram is re-illuminated with a coherent light (or white light in some unique cases) that is a duplicate of the reference beam, the recorded interference pattern will diffract some energy into a beam that is a replica of the object beam. This reconstructed object beam contains all the



Figure 2. Finite fringe interferogram of blunt cone.

amplitude and phase information of the original object beam, and hence looks and acts like the original beam. This is why holography has been called "lensless photography," and it can produce visually dramatic three-dimensional images.

Since the reconstructed wave contains phase as well as amplitude information, images simultaneously reconstructed from two separate holograms can be made to interfere with each other in much the same way as the two beams of an interferometer. This is the basis of the holographic interferometry used at NAVSWC to determine flow-field densities during wind tunnel tests. One hologram is taken through the tunnel test cell before the flow is started, and a

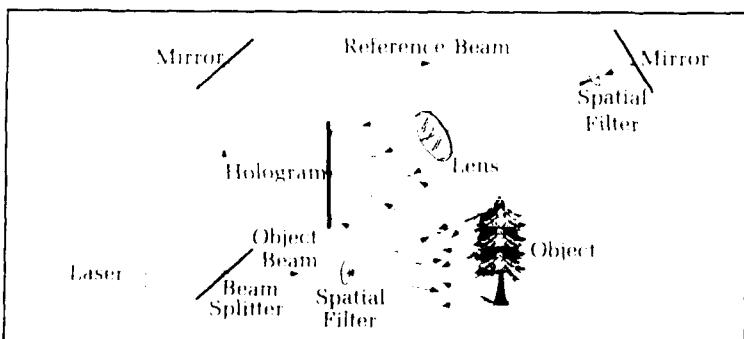


Figure 3. Schematic of hologram construction.

second one is taken after the supersonic flow is established. The two holograms are reconstructed together and the optical path length differences between them, caused by density changes in the airflow, are seen as interference patterns.

Holographic System

To construct holograms of wind tunnel flow fields, a Z-type schlieren system (Figure 4) is used. This optical system is the standard technique for routine flow visualization. The normal light source is replaced with pulsed ruby laser, which provides a high-intensity, short-duration, collimated, coherent beam of light. A partial beamsplitter at the output of the laser splits off a small fraction (typically about 10%) of the energy, which is used to create the holographic reference beam. This beam is directed over the wind tunnel toward the film plane. The remaining laser energy is imaged into the schlieren system, which produces a large diameter collimated beam passing through the tunnel test cell, and then it is focused near the film plane. In the film plane, both beams are recombined to form the hologram.

The special holographic quality ruby laser used as the light source is Q-switched to provide very short duration exposures on the order of 20 nanoseconds (2×10^{-8} seconds). These short exposure times freeze the flow-field density patterns, and also freeze the minute interference patterns that form the hologram. The laser also has mode selectors in the cavity that increase both the temporal and spatial coherence so that high-quality holograms can be made. The ruby laser is constructed such that a small helium-neon laser can be aimed through it for aligning the laser system with the schlieren optics, and so the Q-switched pulse can be monitored with oscilloscopes.

As mentioned earlier, two separate holograms are constructed for each test: the reference

hologram, taken with no flow in the tunnel, and the object (or scene) hologram, taken with the tunnel at the flow conditions to be measured. In an optical sense, these two holograms are the two separate legs of a classical interferometer.

After the plates are processed, they are reconstructed with a helium neon laser (for video digitizing) or an argon laser (for making large negatives). When these plates are reconstructed simultaneously, the differences in optical path length between the two are visible as interference fringes (Figure 2). The interference pattern can be changed by moving one of the holograms with respect to the other, as shown in Figure 5. Also, the scene hologram can be reconstructed by itself to form either a shadowgraph or a schlieren image. Shadowgraph and schlieren are much more common methods of flow visualization, but are less quantitative than interferometry. Thus, holographic interferometry offers the advantage that all three of the classical flow visualization techniques can be employed on the same test.

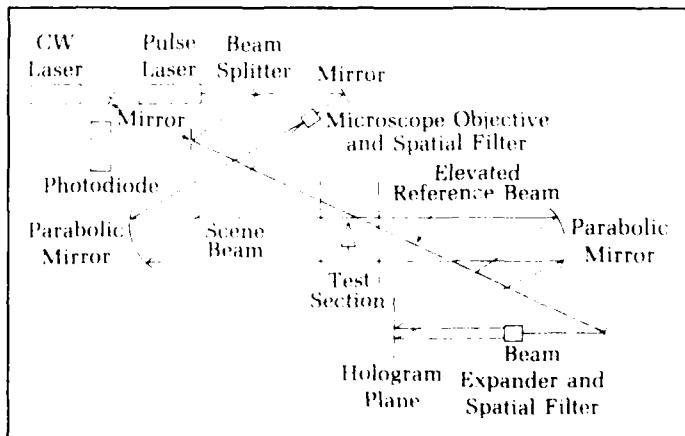
Another advantage of holographic interferometry is that the interference patterns can be generated in the controlled environment of the optics lab at a convenient time after the test, instead of in the relatively hostile, real-time environment of modern hypervelocity wind tunnels. This places fewer demands on the optical setup and data reduction techniques.

Data Reduction

Background

The basic information extracted from the interferograms is the optical path length differences between the object and reference images. The fringe shifts across the field are a direct measure of these path length differences. The classical way to measure this is to count, or track, fringes. The straight-line fringes in the

Figure 4. Schematic of Z-type schlieren system for hologram construction.



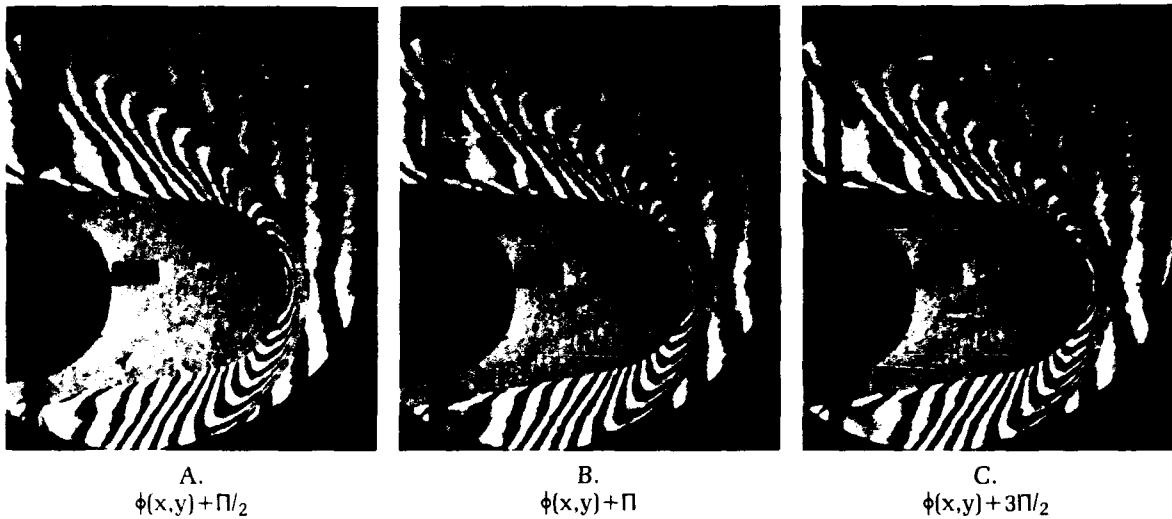


Figure 5. Phase-shifted dual plate interferograms.

free-stream region outside the shock wave, where the gas density is known, are used as reference fringes. The distance the fringes are shifted within the flow field is directly related to the unknown gas density in these areas. In a complex flow field, however, the task of determining these fringe shifts for every point in the field is tedious and time consuming if manually counting is the only method of extracting the data. For years it was this limitation that prevented interferometry from becoming a practical measurement technique.

Data Reduction Technique

The key to the data reduction technique developed at NAVSWC for analyzing flow-field interferograms is the fact that the two holograms are recorded on separate plates. The "dual-plate" method, developed by Havener and Radley,⁵ was applied to the NAVSWC tunnels originally by Hannah and Havener,⁶ and has been used by all other investigators since. This method allows the interference patterns to be continuously varied during the reconstruction. The double-exposure technique, which is the other common method of holographic interferometry, fixes the pattern in place at the time of the exposures. The dual-plate method leads directly to the scheme called "phase-shifting" interferometry. Although not the first to use phase-shifting interferometry, NAVSWC has developed the technique for analysis of flow-field interferograms where sharp density gradients are encountered, such as at supersonic shocks.

The principle of phase-shifting interferometry^{7,8} is that a map of the relative phase of the interference pattern can be calculated from three intensity maps of the

field. The condition is that the background fringes of the three intensity maps must have a constant, known phase shift between them. This is why the ability to shift the fringe pattern during the reconstruction of a pair of holograms of the same test conditions is important. Also required for this technique is a computer-based, high-resolution video digitizing system with the memory and software capability to do image processing and large-array calculations.

The current data reduction capabilities at NAVSWC include a dual plate holder used during reconstruction that was developed by Havener and Radley.⁵ The apparatus provides multiple degrees of freedom to move one plate with respect to the other. The plates are initially aligned to reconstruct an infinite fringe interferogram (see Figure 6); then the object plate is translated horizontally until the desired background fringe spacing is obtained.

The reconstructed interferograms are imaged into a high-resolution video camera, which has gain and contrast adjustments to control the level and range of the output signal. This output is then fed into a PC-based image digitizing system (FrameGrabber from Imaging Technology Corp.). The result is a 512×512 pixel array of digitized intensity values representing the interference pattern.

To capture the images required for the phase-shifting methods, the object plate is driven laterally small fractions of an inch by an accurately controlled piezoelectric micrometer calibrated to produce $\pi/2$ phase shift in the interference pattern. Typically, five images are digitized for each pair of holograms analyzed. They represent phase shifts of $0, \pi/2, \pi, 3\pi/2$, and 2π radians. Once these images have been



Figure 6. Infinite fringe interferogram of blunt cone.

digitized, they can be digitally filtered or processed in a number of ways (by ImageAction software from Imaging Technology), depending on the quality of the input images and the desired output product. Also, the data reduction routines developed at NAVSWC can be run to calculate phase distributions and density values throughout the flow field.

Data Reduction Theory

As shown by Creath,⁸ phase shifting interferometry utilizes three or four (depending on the reduction algorithm selected) digitized images of the intensity fields from the reconstructed interferograms, with $\pi/2$ phase changes between each of the images. The intensity of each of the points (pixels) in the field is:

$$I = I_0[1 + \gamma \cos(\Phi)] \quad (1)$$

where

I = intensity of any given point in the field

I_0 = the dc, or background, intensity

γ = the modulation of the interference fringes

Φ = the phase of the wavefront

The intensity, I , of each point in the field is detected and digitized by the video camera and PC FrameGrabber. The other three values, I_0 , γ ,

and Φ , are unknown, hence the need for at least three digitized images. When the three (or four) intensity fields are recorded with a phase shift of $\pi/2$ between them, three (or four) simultaneous equations for the intensity of any given point in the field (from Equation 1) are as follows:

$$I_1(x,y) = I_0(x,y)[1 + \gamma \cos(\Phi(x,y) + \pi/4)] \quad (2)$$

$$I_2(x,y) = I_0(x,y)[1 + \gamma \cos(\Phi(x,y) + 3\pi/4)] \quad (3)$$

$$I_3(x,y) = I_0(x,y)[1 + \gamma \cos(\Phi(x,y) + 5\pi/4)] \quad (4)$$

$$I_4(x,y) = I_0(x,y)[1 + \gamma \cos(\Phi(x,y) + 7\pi/4)] \quad (5)$$

These can be solved for the phase Φ at that point as follows, where the phase is in radians:

$$\Phi(x,y) = \tan^{-1} \left[\frac{I_3(x,y) - I_2(x,y)}{I_1(x,y) - I_2(x,y)} \right] \quad (6)$$

The general form of Equations 2 through 5 can be solved, but if the phase shifts between the separate intensity measurements are 90° ($\pi/2$), this simplified form results. Also, to solve the general form, the phase shifts between each intensity map must still be known, so it is more expedient to preselect the equally spaced shifts and use the simpler form.

Equation 6 describes a particular algorithm, termed the "three-bucket" method, since three intensity maps are required to calculate the phase map of the field. Several other algorithms have also been examined, namely the "four-bucket," which uses four intensity maps with phase shifts of $\pi/2$, the "three-by-three," which averages two three-buckets, and the Carré, which uses four intensity maps with unknown (but assumed to be linear) phase shifts.

For the four-bucket method, which uses four of the intensity maps, the equation for Φ becomes:

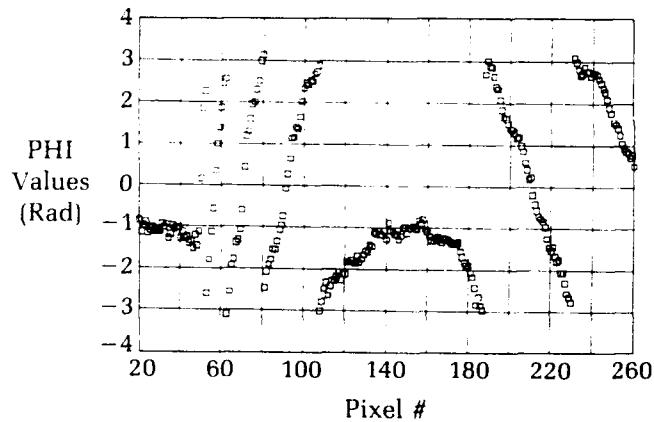
$$\Phi(x,y) = \tan^{-1} \left[\frac{I_4(x,y) - I_2(x,y)}{I_1(x,y) - I_3(x,y)} \right] \quad (7)$$

The three-by-three method averages two three-bucket phase calculations, one using I_1 , I_2 , and I_3 , the second using I_2 , I_3 , and I_4 .

These three methods, by virtue of the addition of arbitrary phase shifts introduced into the mathematics by the intensity map phase shifts, will produce different calculated phase values from the same set of intensity maps. Since the desired input for the density calculations is phase shift relative to the free stream (or other reference), all calculated phase values are normalized to zero in the free stream, or to the other reference value.

The Carré method uses four intensity maps having a phase shift that is not known, but is assumed to be linear. The actual phase shift is a fourth unknown, but can be calculated from the four intensity equations. This method has not yet been successfully integrated into the NAVSWC data reduction programs, since the phase shifts between our intensity maps are well calibrated.

All of these methods produce phase values that are modulo 2π , with a range from $-\pi$ to $+\pi$, which means that a plotted curve of phase values versus distance (radius) will be discontinuous. To produce a continuous phase distribution curve, the calculated phase values are stacked on each other at the discontinuities. This is done with software, by comparing the phase values calculated for adjacent pixels. If the phase difference is greater than π , then 2π is added or subtracted to make the difference less than π . This technique produces smooth phase distribution curves. Typical results for the initially calculated phase and the final phase distributions are shown in Figures 7 and 8. These results are for the flow around a blunt cone which will be described in more detail later in the paper.



Calculation of Density Distributions

Once the phase has been determined, the density distributions may then be calculated. The complexity of the calculations will depend upon the geometry of the phase object (flow field). For a simple 2-D, channel-type flow it can be assumed that the density is constant along the light path through the field of view. The refractive index n and density ρ , at a place where the fringe shift, or phase, is $N(y)$, is given by:

$$N(y)\lambda = (n-n_0)L = K(\rho-\rho_0)L \quad (8)$$

where $n-1 = K\rho$, n_0 and ρ_0 are the refractive index and density, respectively, of a known value of gas, λ is the wavelength of the constructing laser, K is the Gladstone-Dale constant ($0.117 \frac{\text{ft}^3}{\text{slug}}$), and L is the path length through the phenomenon under observation.

$$\text{Finally } \rho = \rho_0 + N(y)\lambda/KL \quad (9)$$

$$\text{or } \frac{\rho}{\rho_0} = 1 + \frac{N(y)\lambda}{\rho_0 KL} \quad (10)$$

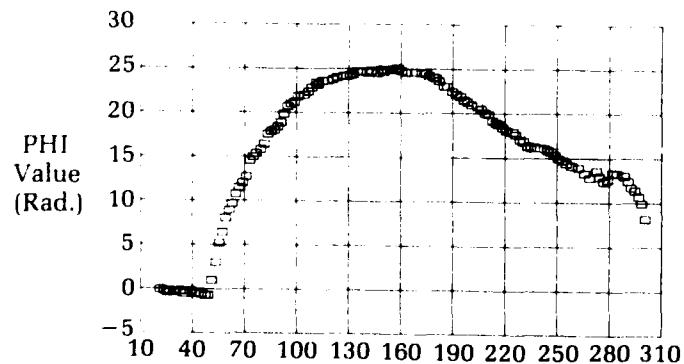


Figure 7. Phase jump technique
(at X = 3.0 in.).

Figure 8. Continuous phase distribution (at X = 3.0 in.).

For axisymmetric flows, the problem is more complex. The phase distribution can be related to changes in the optical path lengths and density for a refractionless, axisymmetric flow field by:

$$N(y)\lambda = 2 \int_y^R \frac{rf(r)}{\sqrt{(f^2-y^2)}} dr \quad (11)$$

where $N(y)$ is the fringe order number (phase shift non-dimensionalized by 2π). This is the functional description of the changes in the optical path length through the flow field. K , again, is the Gladstone-Dale constant, R is the maximum boundary of the flow field, r and y are independent geometric variables, and $f(r)$ is a normalized function for the radial distribution of the index of refraction in the phase volume (normalized to some reference condition n_0), described as follows:

$$f(r) = n(r) - n_0 \quad (12)$$

Since the density function $f(r)$ is the desired quantity, the integral equation must be inverted resulting in:

$$f(r) = \frac{-\lambda}{\pi} \int_r^{\infty} \frac{dN/dy}{\sqrt{(y^2 - r^2)}} dy \quad (13)$$

A variety of methods are used to solve the above equations for the density distribution. The method used at NAVSWC divides the flow region into a series of discrete annular elements (rings) of constant width Δr , as shown in Figure 9 (also see Vest⁴). Since the output from the digitization results in a large number of discrete, equally spaced values of the phase $N(y)$, this method lends itself particularly well to the present setup because the phase value determined at each pixel can represent the value for each ring.

Different assumptions can be made about the density in each of the annular elements, including linear or higher order variation. The simplest is to assume that the density is constant, hence a uniform refractive index exists in each element. This results in a system of equations given by:

$$\sum_{k=i}^{I-1} A_{k,i} f_k = \left(\frac{\lambda}{2 \Delta r} \right) N_i \quad (14)$$

where the coefficients $A_{k,i}$ are given by:

$$A_{k,1} = \sqrt{(k+1)^2 - i^2} - \sqrt{(k^2 - i^2)}$$

and $f_k = n(r_k) - n_0$. Equation 14 represents a set of simultaneous, linear, algebraic equations which must be solved for the unknown values of f_k , the change in index of refraction. The density is then calculated by:

$$\varrho_k = \frac{1}{K} \left[f_k + (n_0 - 1) \right] \quad (15)$$

Applications of Holographic Interferometry

By way of introduction, early holographic experiments that have led up to the dual-plate, phase-shifting technique will be discussed. The first application of holographic interferometry at NAVSWC was a double-exposure, diffusely illuminated setup that was used to plot the three-dimensional shocks from a forward-facing step and a cone model in a Mach 3 nozzle. A continuous-wave Argon laser, diffused by a ground glass, was used for the illumination. Photographs of the reconstructed images along different lines of sight were used to plot the locations of the shocks. Although not rigorous, this experiment proved that the three-

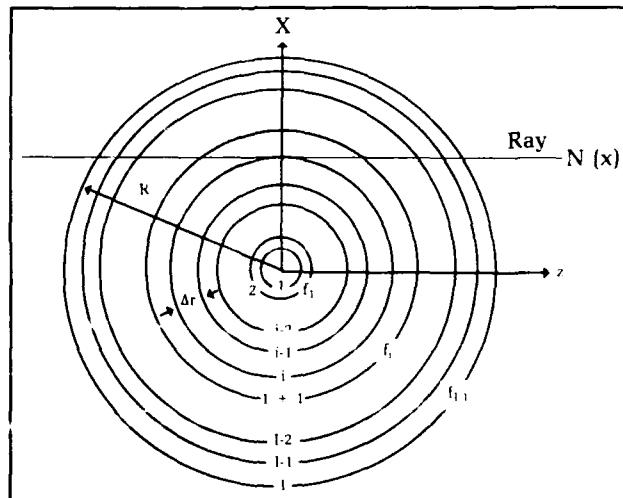


Figure 9. Phase object divided into discrete annular rings.

dimensional characteristics of holography could be used effectively, and the resulting interferograms were the only ones made at NAVSWC that show visually the perspective and parallax that are evident in diffusely illuminated holography. This experiment is described by Ragsdale and Spring.⁹

The next major step was the experiment described by Zien, Ragsdale, and Spring,^{10,11} which reports the conversion of a standard schlieren system for use with interferometry. A pulsed ruby laser was used for these tests, and the double-exposure technique was used to make the interferograms. The experiment was run with a cone mounted asymmetrically in a Mach 2 flow field. The major contribution of this work was in the data reduction.

Reconstructions from a series of holograms, taken as the model was rolled through different aspect angles, were used to calculate the three-dimensional density field. A key to the successful calculations of density was the application by Zien of an area-invariance to extrapolate the fringe shift values in the opaque regions of the interferograms, where the transmitted light is blocked by the solid model.

The dual-plate holographic interferometer was introduced at NAVSWC as a working technique by Hannah and Havener⁶ as part of an effort to automate the data reduction, and was expanded on by Hannah and King.¹² The first attempt at automation was to use a computer-controlled image dissection camera that could scan the interferometric image and track fringes, shock waves, model edges, etc., then perform the required data manipulations to produce phase maps and calculate flow-field densities.

Two-Dimensional Inlet Densities

The first real test of the capabilities of the phase shifting technique was for a two-dimensional flow field in a scramjet inlet investigation in a Mach 4 flow (Figure 10). This test provided a good evaluation of the technique because several other flow diagnostic methods were also used, including Laser Doppler Velocimetry (LDV) and pressure probes, so that the results from the various methods could be compared. This work has been presented in McArthur, et al.¹³ Basically, the dual-plate holograms were reconstructed to form the real-time interferograms, and a series of phase-shifted images was digitized. These images were then converted to phase maps from which densities were calculated. The calculated densities were then compared to densities that had been measured in the inlet with pitot probes, and the two sets of densities compared favorably.



Figure 10. Finite fringe interferogram of scramjet inlet.

Three-Dimensional Axisymmetric Cone Density Measurements

The next challenge for the phase shifting technique was the three-dimensional axisymmetric flow field around a blunt cone (see Figure 11). This problem was selected because

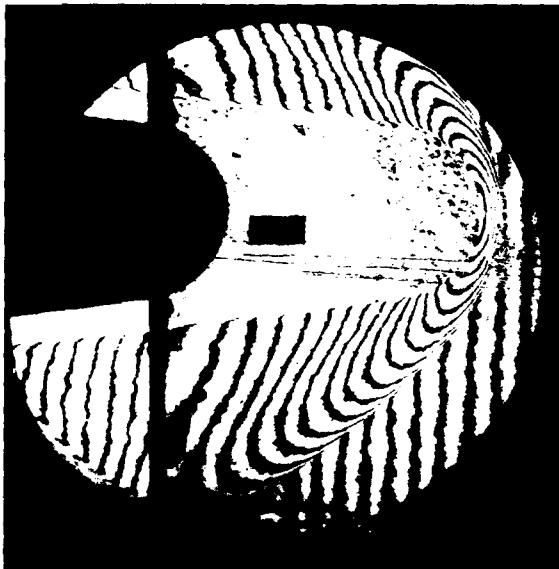


Figure 11. Finite fringe interferogram of blunt cone.

it was a relatively straightforward case and both experimental and computational data (including data from hand-plotted interferograms) were available to compare with the phase-shifted calculations. The Mach number was 2.5 for these tests, and, again, several other techniques, including LDV and pitot probes, were used for comparisons. This program provided the opportunity to compare several different methods of calculating the phase distributions from the intensities, and also to compare several techniques for calculating the densities from the phase distributions. The results, presented in detail,¹⁴ show that the phase-shifting technique does indeed calculate flow-field densities that compared very closely to densities measured experimentally by other methods and to densities predicted by computational fluid dynamics codes. Figure 12, for example, shows the comparison between the phase-shifted densities and densities calculated using a three-dimensional Euler solver by Hsieh and Priolo.¹⁵ These measurements were taken in an axial plane three inches behind the nose tip. This graph (Figure 12) is typical of those taken at several axial locations; all show good agreement except in the region of the shock. Much of the error in this region may stem from the assumption of a refractionless medium. More discussion on this will follow.

The next steps involved trying the phase-shifting methods that had been developed and worked well for the blunt cone for cases that, although still in the relatively simple axisymmetric case, presented more difficult circumstances — namely the sharp cone flow field and the indented nose tip.

Sharp Cone Measurements

The sharp cone in low Mach number (3.0) flow presents a complication because the density changes across the shock wave are

Figure 12. Phase-shift density vs. computational solution.

100

quite small, which means that the fringe shifts will also be small (see Figure 13). Preliminary evidence indicates that the methods developed

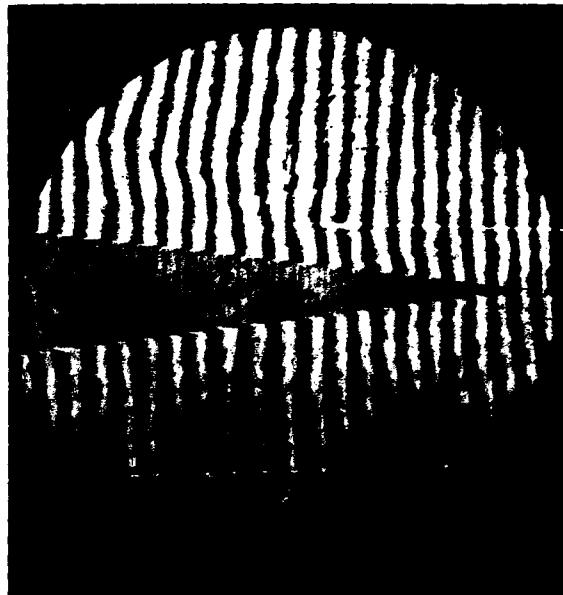
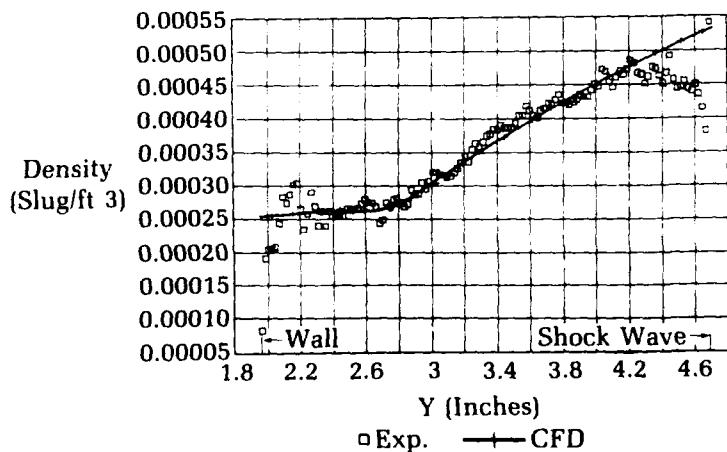


Figure 13. Finite fringe interferogram of sharp cone.

so far will need to be refined for these cases. Possibly a scheme that averages several phase maps will provide greater signal-to-noise discrimination that will increase the confidence in the phase distributions determined under these conditions.

Indented Axisymmetric Cone Measurements

The indented nose shape also presents problems because of the fine detail in the



interference pattern (see Figure 14). These results show that all the fringes must be resolved in the digitized intensity images in order for the phase distributions to be calculated correctly. If all the fringes are not resolved, the phase distributions are lowered by 2π for each lost fringe. A higher resolution digitizing system is one solution to this problem. Another solution is to digitize several enlarged sections of the interferogram so the fine detail can be resolved, and then let the computer put the sections back together to determine the total phase distribution. Equipment will be available at NAVSWC in the near future that will allow further investigation of both of these ideas.



Figure 14. Finite fringe interferogram of indented nosetip.

Future Areas of Investigation

Several areas are currently under investigation and research, with the goal of improving both the setups for construction of the holograms in the wind tunnels and the accuracy of the data reduction.

Experiments have been performed to determine if a technique of localized, or point, reference beam holography, as described by Bachalo and Houser¹⁶ can be used in NAVSWC's hypervelocity tunnel. This technique offers the advantage that both the

reference and object beams are generated from the light energy that passes through the tunnel, thereby simplifying the apparatus and setup for construction of the holograms.

Some other modifications of the current methods include using a collimated rather than a focusing beam for the object beam exiting the schlieren optics, and the use of the "sandwich technique" for recording the two holograms essentially in contact. Both of these changes will make the reconstruction process simpler, more accurate, and less time-consuming.

One of the known causes of errors in data reduction is the assumption that gas densities in the flow field are refractionless. Since the calculations of gas density involve determining the index of refraction of the gas, it is known that changes in this index are going to cause refraction of the light passing through the gas. Also, diffraction of the light waves propagating through the tunnel and optical system contributes to the error in the calculated densities. A theoretical analysis of these problems has been undertaken by Lopez,¹⁷ and work is underway to apply this theory to the real data reduction process. The expectation is that the accuracy of the calculated densities will be significantly improved, particularly in the region of shock and along the model edge.

Some other areas for future consideration have also been mentioned previously. One concerns the resolution on the digitizing system. New equipment will soon be obtained through a NAVSWC SBIR (Small Business Innovative Research) Phase II contract with KMS Fusion of Ann Arbor, Michigan, that will make available a high-resolution camera and digitizing capability. This equipment can be used in conjunction with the computer to digitize in high-resolution areas of the image that can then be processed and recombined at lower resolution to provide phase data for the entire image. This should be helpful in overcoming some of the problems in areas where the present system has limitations, such as in the area of a strong shock wave or the recirculation regions of the indented cone.

Further investigation also is needed for the case of small density gradients, such as the sharp-cone case discussed earlier. An averaging of several sets of intensity maps, or some sort of heterodyning technique that will increase the fringe sensitivity, are possible approaches that will be investigated.

101

Conclusion

This article presents the state of holographic interferometry as used in aerodynamics testing at NAVSWC, past, present, and future. The use

of high-power lasers with long coherence lengths and short exposure times, combined with proper optical setups, has made the production of high-quality interferograms of wind tunnel test conditions a reality. Data reduction from the interferograms has been greatly enhanced by high-resolution video digitizing equipment and with fast, desktop computers with large memory capacity. These have allowed rapid collection and manipulation of large numbers of data points that fully utilize the capabilities of interferometry. The accomplishment of the immediate future goals will make the use of holographic interferometry a straightforward and accurate instrumentation technique for determining flow-field densities in NAVSWC's wind tunnels.

Acknowledgments

The authors would like to acknowledge the NAVSWC Independent Research Program, which has provided most of the funding for the recent work implementing the phase-shifting data reduction technique and for the theoretical analysis of the diffraction corrections. Earlier work on the dual-plate development was sponsored by the Naval Air Systems Command.

References

1. Born, M. and Wolf, E., *Principles of Optics*, Third Edition, Pergamon Press Inc., Long Island City, NY, 1965, p. 313.
2. Gabor, D., "A New Microscope Principle," *Nature*, Vol. 161, 1948, p. 777.
3. Leith, E. N. and Upatnick, J., "Wave Front Reconstruction with Continuous Tone Objects," *Journal of the Optical Society of America*, Vol. 53, No. 12, 1963, p. 1377.
4. Vest, C. M., *Holographic Interferometry*, Wiley & Sons, Inc., New York, NY, 1979.
5. Radley, R. J., Jr. and Havener, A. G., "The Application of Dual Hologram Interferometry to Wind Tunnel Testing," *AIAA Journal*, Vol. 11, Sep 1973, p. 1332.
6. Hannah, B. W. and Havener, A. G., "Applications of Automatic Holographic Interferometry," *International Congress on Instrumentation in Aerospace Simulation Facilities* proceedings, Ottawa, Canada, Sep 1975.
7. Wyant, J. C. and Creath, K., "Recent Advances in Interferometric Optical Testing," *Laser Focus/Electro-Optics*, Vol. 21, No. 12, 1985, p. 118.
8. Creath, K., "Phase-Measurements Interferometry Techniques," *Progress in Optics XXVI*, Edited by E. Wolf, Elsevier Science Publishers B. V., 1988, pp. 350-393.
9. Spring, W. C., III and Ragsdale, W. C., "The Investigation of the Use of Holography in Studying Supersonic Flow in Inlets," *Progress in Astronautics and Aeronautics, Instrumentation for Airbreathing Propulsion*, MIT Press, Cambridge, MA.
10. Zien, T. F., Ragsdale, W. C., and Spring, W. C., III, *Quantitative Applications of Holographic Interferometry to Wind Tunnel Testing*, NOLTR 74-96, Naval Ordnance Laboratory, Silver Spring, MD, Dec 1974.
11. Zien, T. F., Ragsdale, W. C., and Spring, W. C., III, "Quantitative Determination of Three-Dimensional Density Field by Holographic Interferometry," *AIAA Journal*, Vol. 13, No. 7, Jul 1975.
12. Hannah, B. W. and King, W. L., "Extensions of Dual-Plate Holographic Interferometry," *AIAA Journal*, Vol. 15 No. 5, May 1977.
13. McArthur, J. C., et al., "Laser Holographic Interferometric Measurements of the Flow in a Scramjet Inlet at Mach 4," AIAA Paper #89-0043, presented at AIAA 27th Aerospace Sciences Meeting, Jan 1989, Reno, NV.
14. Yanta, W. J., et al., "Phase-Measuring Laser Holographic Interferometry for Use in High Speed Flows," presented at the 13th International Congress on Instrumentation in Aerospace Simulation Facilities, Gottingen, West Germany, 18 Sep 1989.
15. Hsieh, T. and Priolo, F. J., "Generation of the Starting Plane Flow Field Over a Spherically Capped Body," NAVSWC TR 84-484, NAVSWC, Silver Spring, MD, May 1985.
16. Bachalo, W. D. and Houser, M. J., "Optical Interferometry in Fluid Dynamics Research," in *Proceedings of Automated Reduction of Data from Images and Holograms Workshop*, NASA Ames Research Center, Moffett Field, CA, 10 Jan 1985.
17. Lopez, Carlos A., "Diffraction Corrections of Holographic Interferometric Images," AIAA Paper 91-0564, Presented at the 29th Aerospace Sciences Meeting, Reno, NV, Jan 1991.

The Authors



W. C. SPRING, III, graduated from Rochester Institute of Technology in 1963 with a B.S. degree in Photographic Science, and has taken undergraduate and graduate-level courses in math, physics, and optics at the University of Maryland. He has worked at NAVSWC since 1963 in photographic and electro-optical instrumentation. Since 1985 he has specialized in aerodynamic flow visualization techniques,

and is currently responsible for all photo-optical instrumentation in the Center's hypersonic wind tunnels. Systems that he has engineered and operated include black and white and color schlieren, shadowgraph, holographic interferometry, laser vapor screen, fluorescent microtulfts, electron beam, and glow discharge photography, all used primarily for flow diagnostics in wind-tunnel testing. He has also developed multi-camera, high-speed set-ups, including film and video systems, for instrumenting dynamic flight hardware tests. He has been a major contributor to the development of a standardized time-base system to synchronize all the high-speed data recording equipment in the hypervelocity wind tunnel. He is co-author of numerous papers describing tests and experiments that use holographic interferometry for analysis of aerodynamic and engineering tests.



WILLIAM J. YANTA received B.S. (1952) and M.S. (1964) degrees in Aerospace Engineering from the University of Texas. He holds a diploma from the Von Karman Institute for Fluid Mechanics in Brussels, Belgium and a Ph.D. from Catholic University. After working four summers at NAVSWC, he began full-time work in 1965 in the Aerophysics Branch, moving to the Aerodynamics Branch.

He has conducted research in supersonic and hypersonic viscous and inviscid flows, development and application of non-intrusive optical diagnostics, primarily laser doppler velocimetry and holographic interferometry, to be used in high-speed flows. He received the Meritorious Civilian Service Award in 1974. He is presently Chief Technologist in the Aerodynamics Branch, and is a member of the NAVSWC Independent Research Mathematics and Aerodynamics panel, National Aerospace Plane Program Instrumentation Task Force, AIAA Aerodynamics Measurements technical committee, Associate Fellow AIAA, instructor in the Aerospace Engineering Department, University of Maryland, and Adjunct Professor at North Carolina State University.



KIMBERLY U. GROSS came to NAVSWC's Aerodynamics Branch as a Co-op student in 1986. In 1987 she returned after completing her B.S. degree in computer science at the University of Maryland. She currently manages computer systems for the Weapon Dynamics Division and plays a vital role in the smooth operation and maintenance of those systems. A member of the holographic interferometric team, she is solely

responsible for the development of code used in image and data analysis. She also supports the data reduction portion of wind tunnel test programs. While on rotational assignment with the Systems Design Branch, on an independent exploratory development team titled "Eyes-on and Hands-off Display Technologies," she was instrumental in developing ADA code that supported the early stages of the project.

Celsian-Based Ceramics for Advanced Radomes

I. Talmy, D. Haught, E. Wuchina, J. Zaykoski

Research conducted at the Naval Surface Warfare Center has demonstrated that celsian ($\text{BaO} \cdot \text{Al}_2\text{O}_3 \cdot 2\text{SiO}_2$) is a promising candidate for advanced radomes due to extremely good dielectric properties, high melting point, low thermal expansion and good strength. However, the rain erosion and thermal shock resistance of the material need improvement. Toughening of the ceramics using mullite whiskers was investigated and shown to be promising. Two methods, slip casting and cold isostatic pressing, are being evaluated for radome shape processing.

Introduction

Future classes of advanced tactical missiles will require new ceramic materials for radomes to meet stringent requirements arising from greater speeds and longer flight times. The critical need is to develop ceramic materials stable up to at least 1400°C with low and thermally stable dielectric constants (ϵ), low coefficient of thermal expansion (CTE), and high thermal shock and rain erosion resistance as shown in Table 1.

No ceramic material has been developed that can simultaneously meet all of the requirements, especially those for good dielectric properties and rain erosion resistance. High-speed hypersonic radome materials currently used are slip cast fused silica (SCFS) and Pyroceram 9606TM ($2\text{MgO} \cdot 2\text{Al}_2\text{O}_3 \cdot 5\text{SiO}_2$, cordierite). While SCFS has excellent dielectric properties for radomes, the relatively low strength and rain erosion resistance of this material limits its future application. Pyroceram 9606TM is more rain erosion resistant than SCFS, but is not as dielectrically stable with temperature. Development of future radome materials is directed both toward improvement of the mechanical properties of SCFS and identification of new candidate ceramic materials. For example, candidates under investigation are materials in the system $\text{Si}_3\text{N}_4\text{-BN-SiO}_2$ (Nitroxyceram). Nitroxyceram has the potential of being nearly as good as SCFS dielectrically and at least as good if not better than Pyroceram 9606TM in erosion resistance. However, this material requires expensive processing techniques, such as hot isostatic pressing, which are not fully developed.

Celsian ($\text{BaO} \cdot \text{Al}_2\text{O}_3 \cdot 2\text{SiO}_2$) appears to have the potential for meeting the requirements for radome application due to the unique combination of high melting point, low thermal expansion, low dielectric constant, and loss tangent stable over a broad range of temperature and frequency.^{1,2} However, these beneficial properties belong only to the monoclinic modification of celsian which is stable up to 1590°C. Above 1590°C, monoclinic celsian transforms into the hexagonal modification which is stable to the melting temperature (1760°C).³ Even though the hexagonal modification is a high temperature form, it tends to be the first product of solid phase reaction and persists metastably in the whole temperature range. Hexagonal celsian reversibly transforms at 300°C into the low temperature orthorhombic form.^{3,4} The transformation is accompanied by significant volume changes, making hexacelsian unsuitable for high-temperature thermal cycling applications. Transformation of hexagonal celsian into the desirable monoclinic form is promoted by prolonged high-

Table 1. Radome Material Goals

Service Temperature	Up to 1400°C
Dielectric Constant	<9.0 (5.0 Preferred)
Temperature Variation of Dielectric Constant to 1260°C	<7%
Loss Tangent to 1260°C	<0.1
Bending Strength	>35 MPa (5 ksi)
Fracture Toughness	Better Than Slip Cast Fused Silica

temperature heating, hydrothermal treatment, and the presence of impurities or the addition of certain additives (such as B_2O_3 , LiF , Cr_2O_3 , ZrSiO_4).^{5,6}

This article describes the results of research on preparation of high-purity celsian ceramics and their properties.

Preparation and Characterization of Celsian Ceramics

High-purity celsian synthesized from BaCO_3 , Al_2O_3 , and SiO_2 consisted only of the hexagonal modification, even after long-term firing at 1500°C. The hexagonal-to-monoclinic phase transformation of celsian was accomplished using seeding techniques at temperatures below 1300°C.⁷ The successful use of seeding techniques for the preparation of monoclinic celsian is of high importance, since the use of additives promoting the phase transformation can deteriorate the dielectric properties, especially at high temperatures.

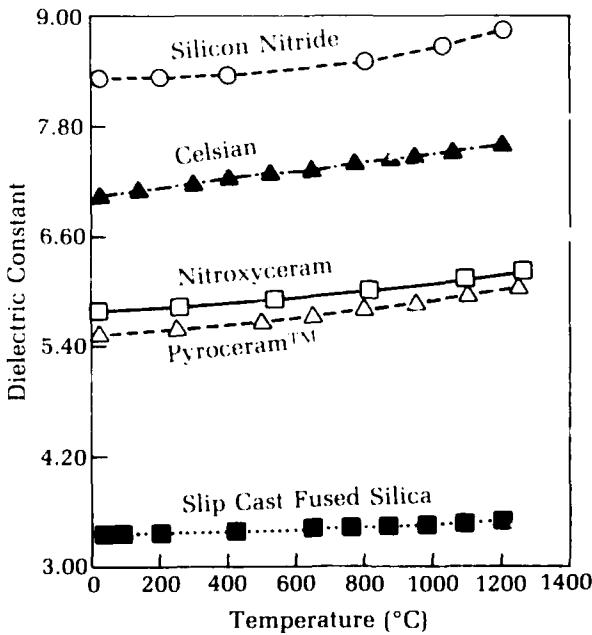
Ceramics prepared from the previously synthesized monoclinic celsian powders were completely densified at 1580°C using pressureless sintering techniques. The ceramics had fine-grained structure with average grain size of about 3 to 5 μm . The properties of celsian ceramics are shown in Table 2. The bending strength measured in 3-point bending was about 100 MPa. The coefficient of thermal expansion was $3.5 \times 10^{-6}/\text{degree C}$ in the 20 to 1200°C range with linear dependence on temperature.

The dielectric properties measured up to 1200°C at 35 GHz⁸ showed that the dielectric constant and its change with temperature are within the required range. The loss tangent was very low and of the same magnitude as slip cast fused silica. At 1200°C, the loss tangent of celsian is 40×10^{-4} , which is negligible for all practical applications. The temperature dependence of dielectric constant of celsian compared to other candidate materials is shown in Figure (1).

Table 2. Properties of Celsian Ceramics

Melting Temperature (°C)	1760
Service Temperature (°C)	Up to 1590
Dielectric Constant (25°C) (1200°C)	6.55 7.00
Temp. Variation of ϵ to 1200°C	6.8%
Loss Tangent to 25°C 1200°C	0.0008 0.0040
Thermal Expansion in 20-1200°C ($\alpha \times 10^6/\text{Degree}$)	3.5
Bending Strength (R.T.) (1000°C)	103 MPa (14.7 ksi) 69 MPa (9.8 ksi)
Fracture Toughness	Needs Improvement

As seen in celsian ceramics meet or exceed the radome material goals for dielectric properties, strength, and thermal expansion. However, single particle impact tests showed that the toughness of the material is insufficient for adequate rain erosion resistance and requires improvement.^{9,10}

**Figure 1.** Temperature dependence of dielectric constant for candidate radome materials.

Toughening of Celsian Ceramics with Mullite Whiskers

Mullite whiskers, developed at the Naval Surface Warfare Center (NAVSWC),^{11,12} were chosen as a candidate reinforcement to improve the fracture toughness of celsian ceramics. Mullite ($3\text{Al}_2\text{O}_3 \cdot 2\text{SiO}_2$) whiskers are promising due to their excellent chemical stability, low thermal expansion, good high-temperature strength and creep resistance, and low dielectric constant. As seen in the celsian-mullite phase diagram (Figure 2),¹³ mullite is compatible with celsian up to 1554°C. At higher temperatures, chemical reactions occur and are accompanied by the formation of a liquid phase and Al_2O_3 . The reinforcement was both introduced into celsian as loose whiskers, and grown in situ during processing.

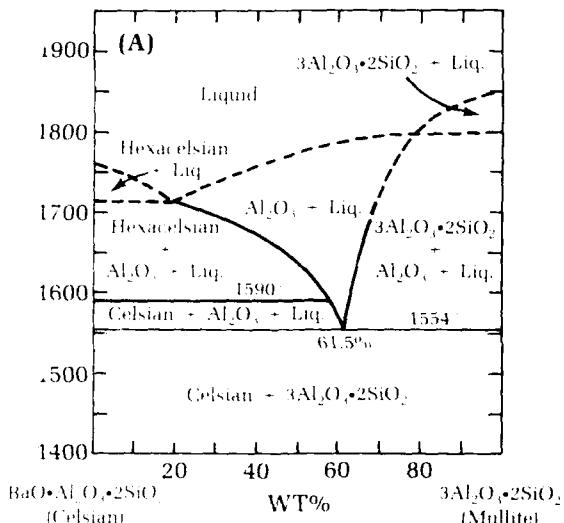


Figure 2. Celsian-mullite phase diagram.

Toughening with Introduced Mullite Whiskers

The whiskers used in this effort were prepared by Dow Chemical Company¹⁴ using the patented NAVSWC process. Unclassified whiskers (as loose aggregates consisting of whiskers with a continuous size distribution) and whiskers classified by cross-sectional dimensions into size ranges of 4 to 12 μm and 12 to 26 μm were mixed with celsian in amounts from 10 to 40 volume percent, pressed into 5 \times 5 \times 60 mm bars, and fired at 1500°C. The temperature was selected to avoid a reaction between celsian and mullite during firing.

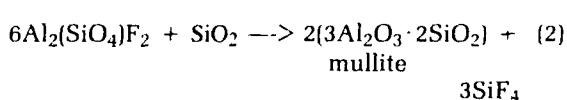
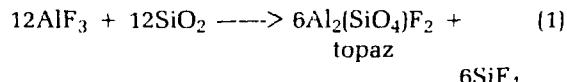
The 3-point bending strength results (Figure 3) for ceramics containing all three whisker classifications show that the strength of

whisker-reinforced celsian ceramics increases with decreasing whisker size. The strength of ceramics with unclassified whiskers was significantly lower than the strength of unreinforced celsian and was slightly higher than the strength of slip cast fused silica (35 MPa). Due to this fact, unclassified whiskers were excluded from further experiments. The introduction of 12 to 26 μm and 4 to 12 μm whiskers resulted in ceramics which were stronger than unreinforced material. The microstructure of celsian containing unclassified and 12 to 26 μm whiskers (Figure 4) shows no evidence of reaction between celsian and mullite. The whisker pull-out observed on the fracture surface of the sample containing 12 to 26 μm whiskers indicates possible toughening.

The thermal shock resistance of the ceramics was characterized by measuring the retained strength after a water quench from 350°C. Figure 5 shows that mullite whiskers increase the thermal shock resistance of celsian ceramics. The samples containing 12 to 26 μm whiskers demonstrated much higher thermal shock resistance than the materials with 4 to 12 μm whiskers. Whisker loading in the 20 to 40 percent range did not have any significant effect on thermal shock resistance for either whisker classification. The measurement of the fracture toughness of mullite whisker reinforced ceramics is in progress.

Toughening with In Situ Grown Mullite Whiskers

In situ growth of whiskers offers several advantages over the introduction of loose whiskers: simplified processing, uniform whisker distribution in the matrix, and elimination of the health hazards associated with loose whiskers. The NAVSWC method for the preparation of mullite whiskers^{11,12} is based on the following reactions:



In reaction (1), AlF_3 and SiO_2 are heated at 700 to 900°C to form topaz, and in reaction (2), topaz is thermally decomposed at 1250 to 1400°C to yield mullite. The whiskers are grown as a result of a vapor-phase chemical reaction. Heating in SiF_4 atmosphere is necessary for the preparation of high-quality mullite whiskers. Up to 75 percent Al_2O_3 can be substituted for AlF_3 in the starting materials without affecting the final product.

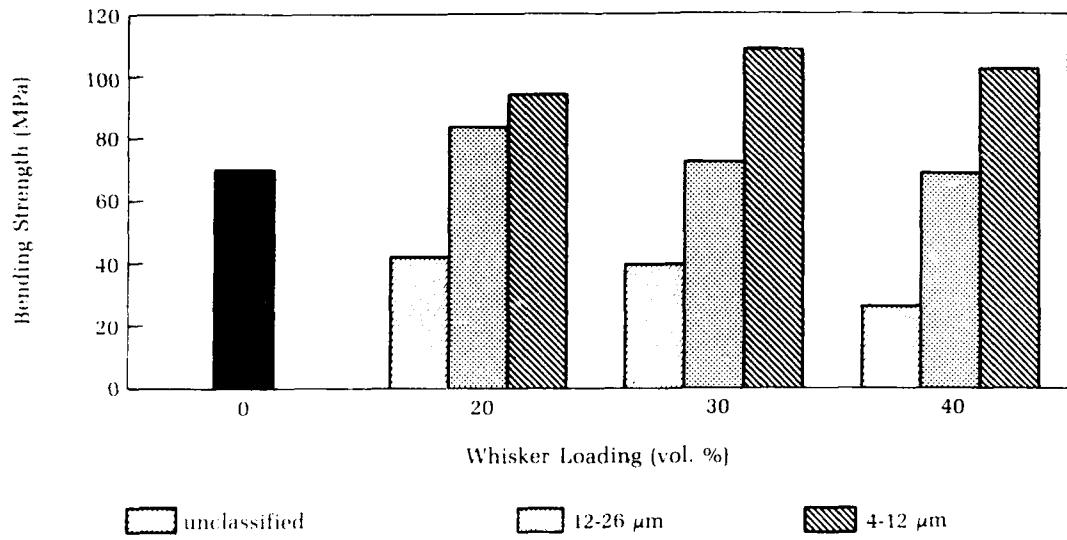


Figure 3. Bending strength of celsian ceramics containing 0-40 vol.% mullite whiskers fired at 1500°C.

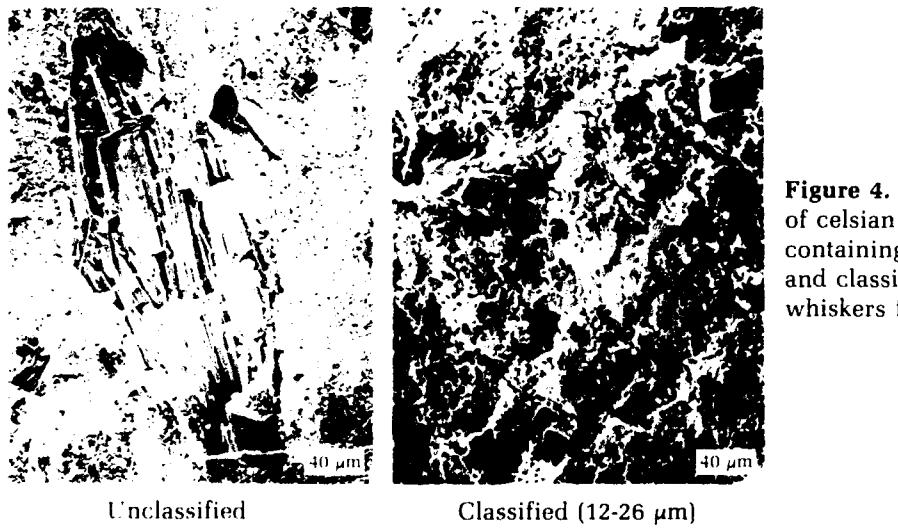


Figure 4. SEM micrographs of celsian ceramics containing unclassified and classified mullite whiskers fired at 1500°C.

In this study, mullite whiskers were grown in celsian ceramics from $\text{AlF}_3 + \text{SiO}_2$ and $\text{AlF}_3 + \text{Al}_2\text{O}_3 + \text{SiO}_2$ (with 50% substitution of Al_2O_3 for AlF_3) mixtures, or from topaz which was previously prepared from AlF_3 and SiO_2 at 800°C according to reaction (1). The use of the $\text{AlF}_3 + \text{Al}_2\text{O}_3 + \text{SiO}_2$ mixture or topaz is advantageous over the $\text{AlF}_3 + \text{SiO}_2$ mixture due to the lower amount of volatiles produced by the reactions (54.2% for $\text{AlF}_3 + \text{SiO}_2$ mixture compared to 29.6% and 28.3% for $\text{AlF}_3 + \text{Al}_2\text{O}_3 + \text{SiO}_2$ mixture and topaz, respectively). The volatiles can interfere with densification of the ceramics during firing.

The whisker precursor powders (to form 15 to 35 volume percent mullite whiskers) and

the celsian powder were mixed, pressed into 5 × 5 × 60 mm bars, and fired at 1500°C in a 20 g/L SiF_4 . The resulting samples were then characterized by density, phase composition, microstructure, strength, and thermal shock resistance.

Mullite whiskers were successfully grown in celsian ceramics from all the precursors used. The typical microstructure of ceramics is shown in Figure 6. The samples with whiskers grown from the $\text{AlF}_3 + \text{SiO}_2$ mixture exhibited the lowest density and strength compared to the other precursors due to the highest content of volatiles. The trend was more pronounced for samples with higher whisker loadings (Figure 7). The highest strength was obtained

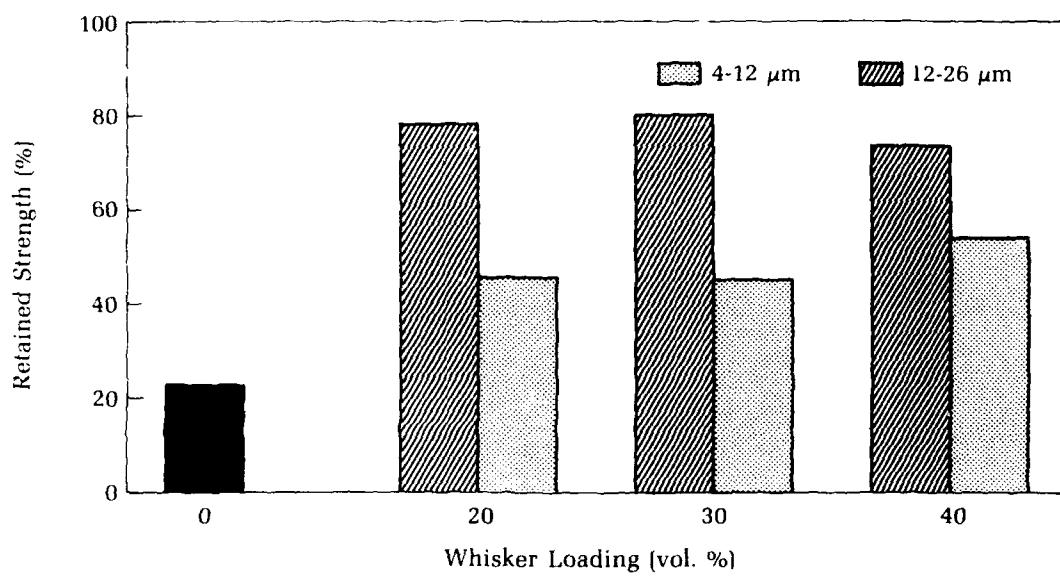


Figure 5. Retained strength (%) of mullite-whiskers reinforced celsian ceramics after thermal shock from 350°C.

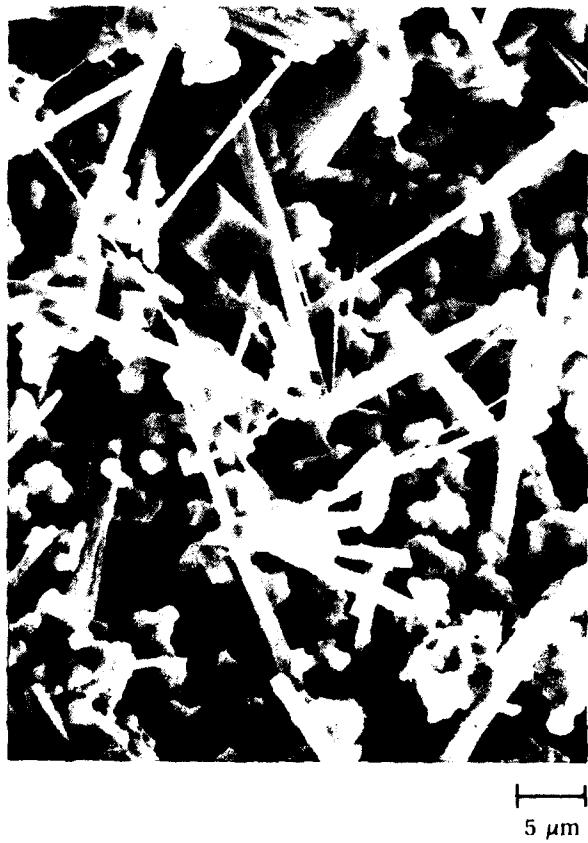


Figure 6. SEM of celsian ceramics with 25 vol.% mullite whiskers grown from $\text{AlF}_3 + \text{SiO}_2$.

for the samples with mullite whiskers grown from topaz. In this case, whiskers were formed as a result of only one chemical reaction (2) compared to two reactions (1) and (2) for the $\text{AlF}_3 + \text{Al}_2\text{O}_3 + \text{SiO}_2$ mixture. This probably resulted in lower stresses in final samples despite the comparable amounts of volatiles for both mixtures. The strength of all samples decreased with increasing whisker loadings. However, the samples with whiskers grown from topaz had strength two times higher than slip cast fused silica even with 35 percent loading.

Thermal shock resistance of *in situ* grown mullite whisker reinforced celsian ceramics was also characterized by the percentage of retained strength after a water quench from 350°C. Figure 8 shows the results for samples containing 15 percent whiskers grown from the $\text{AlF}_3 + \text{Al}_2\text{O}_3 + \text{SiO}_2$ mixture and from topaz. A significant increase in thermal shock resistance was obtained by ceramics with whiskers grown from topaz, illustrating again that topaz is the preferable precursor for *in situ* whisker growth.

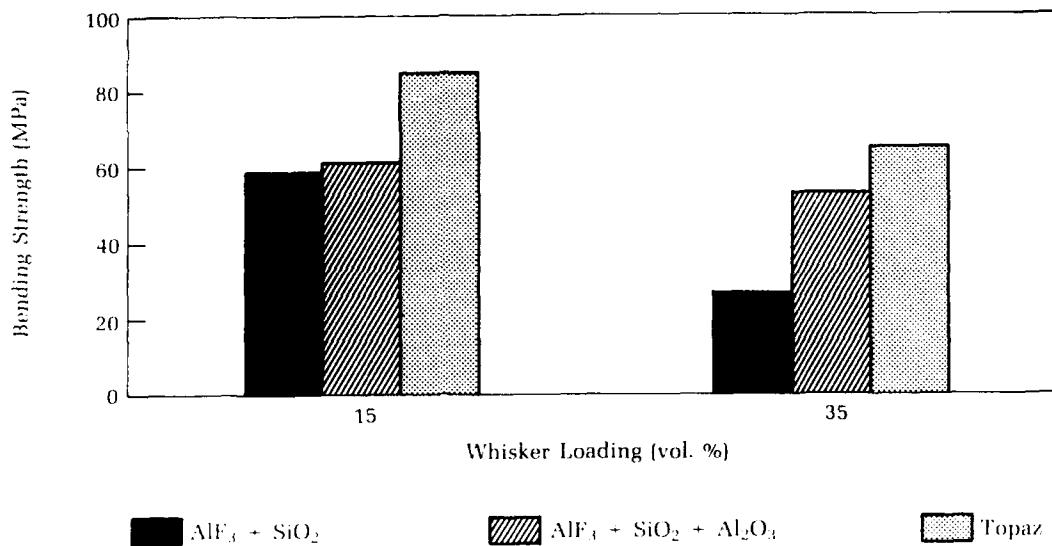


Figure 7. Bending strength of celsian ceramics with mullite whiskers grown from different precursors at 1500°C.

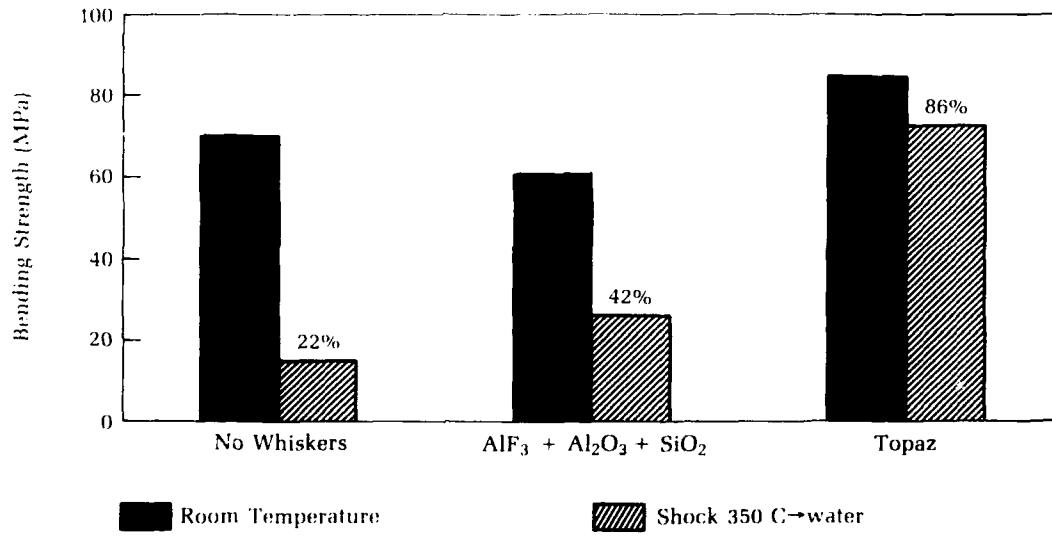


Figure 8. Thermal shock resistance of celsian ceramics with 15 vol.% whiskers grown from different precursors at 1500°C.

Processing of Celsian Radome Shapes

Slip casting and cold isostatic pressing (CIP) of celsian are under development for the processing of radome shapes. Traditional slip casting involves suspending particles in a liquid (usually water) at high solids loading, pouring the suspension into an absorbent mold, and draining the excess slip after the required wall thickness has been cast. The casting is then dried and fired to produce a dense ceramic body.

The preparation of stable ceramic suspensions suitable for casting can be achieved by maximizing the electrostatic surface charges or by adsorbing charged polymeric deflocculants onto particle surfaces. Experiments showed that celsian suspensions can only be stabilized by the second technique, and a commercially available polymer was identified as being optimal for suspension stabilization.

Subscale celsian radomes were cast using the developed slip. However, some problems were encountered during scale-up, such as high casting rate, insufficient green density leading to high shrinkage in firing, and slip thixotropy. It became apparent that optimizing the particle size distribution of the starting celsian powder will eliminate these problems.

Cold isostatic pressing is a promising technique for fabrication of radome shapes, especially from powders containing fibrous (whisker) reinforcement. The method, as applied to making radomes, involves hydrostatic compression of the starting powder, which is held on the surface of a cone-shaped mandrel by a thin, flexible rubber sheath, at pressures ranging from 30,000 to 60,000 psi. The pressed part is then fired to produce a dense ceramic body. Compared to slip casting, the method provides a uniform whisker distribution during fabrication and eliminates the need for costly long-term drying. Tooling for CIPing of celsian was designed, manufactured, and successfully used for fabrication of subscale radomes (1/6 scale) from powder containing 20 percent mullite whiskers. Optimization and further scale-up of both slip casting and CIPing processes are in progress.

Summary

110 Celsian ceramics were prepared and characterized by phase composition, microstructure, physical, mechanical and dielectric properties at room and elevated temperatures. Extremely good dielectric properties, high melting temperature, low thermal ex-

pansion, and high strength make the material a promising candidate for advanced missile radomes. However, the rain erosion and thermal shock resistance of celsian ceramics require improvement because of insufficient toughness. Toughening of ceramics using mullite whiskers is under investigation. The whiskers were both introduced and *in-situ* grown from $\text{AlF}_3 + \text{SiO}_2$ and $\text{AlF}_3 + \text{SiO}_2 + \text{Al}_2\text{O}_3$ mixtures and from topaz. Both methods were shown to be promising for toughening celsian ceramics. Slip casting and cold isostatic pressing are being evaluated for processing of radome shapes. Subscale radomes were successfully fabricated using both techniques.

Acknowledgments

The authors wish to acknowledge Dr. Marriner Norr and Miss Amy Janovsky for their assistance in this project. The work was funded by the Office of Naval Technology Weapons & Spacecraft Materials Block Program and the Naval Surface Warfare Center's Independent Research and Independent Exploratory Development Programs.

References

1. Talmy, I. G. and Haught, D. A., "Phase Equilibria in the System $\text{BaO} \cdot \text{Al}_2\text{O}_3 \cdot 2\text{SiO}_2 \cdot \text{SrO} \cdot \text{Al}_2\text{O}_3 \cdot 2\text{SiO}_2$," Second Navy Independent Research/Independent Exploratory Development Symposium, Vol. I, 1989.
2. Talmy, I. G. and Haught, D. A., "Ceramics in the System $\text{BaO} \cdot \text{Al}_2\text{O}_3 \cdot 2\text{SiO}_2 \cdot \text{SrO} \cdot \text{Al}_2\text{O}_3 \cdot 2\text{SiO}_2$ as Candidate for Radome Applications," Proceedings of the 19th Symposium on Electromagnetic Windows, 1988.
3. Lin, H. C. and Foster, W. R., "Studies in the System $\text{BaO} \cdot \text{Al}_2\text{O}_3 \cdot 2\text{SiO}_2$ I. The Polymorphism of Celsian," The American Mineralogist, Vol. 53, 1-2, 1968, pp. 133-144.
4. Takeuchi, Y., "A Detailed Investigation of the Structure of Hexagonal $\text{BaAl}_2\text{Si}_2\text{O}_8$ with Reference to its α to β Inversion," Mineralogical Journal, Vol. 2, No. 5, 1958, pp. 311-332.
5. Mitev, T., Dimova, T., and Tomova, E., "Effect of B_2O_3 , Cr_2O_3 , MnO_2 and ZrSiO_3 Additions on the Synthesis, Structure, and Properties of Celsian from Bulgarian Raw Materials," Elektroprom-st. Priborost., Vol. 5, No. 8, 1970, pp. 302-303.
6. Moya Corral, J. S. and Garcia, V. A., "Study of the Reaction Between Kaolin and Barium Carbonate," Bol. Esp. Ceram. Vidrio., Vol. 15, No. 6, 1976, pp. 379-381.
7. Talmy, I. G. and Haught, D. A., Ceramics in the System $\text{BaO} \cdot \text{Al}_2\text{O}_3 \cdot 2\text{SiO}_2 \cdot \text{SrO} \cdot \text{Al}_2\text{O}_3 \cdot 2\text{SiO}_2$ For Advanced Radome Application, NSWC TR 89-162, Sep 1989.
8. Ho, W. W., High Temperature Dielectric Tests on Electromagnetic Sensor Window Materials, SC5484.QPR, Rockwell International, Thousand Oaks, CA, May 1988.

9. Adler, W. F. and Flavin, J. W., *Single Particle Impact Damage on Advanced Hypersonic Rain Erosion Resistant Radome Materials - Phase III*, CR-89-1095, General Research Corporation, Santa Barbara, CA, Mar 1989.
10. Adler, W. F. and Flavin, J. W., *Single Particle Impact Damage on Celsian - Phase III*, CR-89-1136, General Research Corporation, Santa Barbara, CA, Sep 1989.
11. Talmy, I. and Haught, D., "Preparation of Mullite Whiskers," U.S. Patent No. 4911902, 27 Mar 1990.
12. Talmy, I. and Haught, D., "Preparation of Mullite Whiskers From AlF_3 , SiO_2 , And Al_2O_3 Powders," U.S. Patent No. 4910172, 20 Mar 1990.
13. Keser, M., Ed., *Phase Diagrams For Ceramists*, The American Ceramic Society, Columbus, OH, 1975, p. 219.
14. Moyer, J. R., Brubaker, B. D., LaBarge, M. S. and Hughes N.N., *Mullite Whisker Development*, NAVSWC TR 90-18, Dow Chemical Company, Midland, MI, Mar 1990.

The Authors



INNA G. TALMY is the senior research ceramist of the High Temperature Materials Group of NAVSWC, which she joined in 1983. She received both M.S. (1957) and Ph.D. (1965) degrees in ceramics science and engineering from the Institute of Chemical Technology in Moscow, USSR. Previously, Dr. Talmy worked both as a research scientist and professor in the ceramic departments of Institutes of

Chemical Technology in Moscow, USSR, and Prague, Czechoslovakia. Her primary research interests are in the field of dielectric ceramics and ceramic-matrix composites. Dr. Talmy directed the development of celsian ceramics as a leading candidate for future radomes. Also under her leadership, mullite whiskers and mullite-whisker felt having high practical importance in ceramic technology were developed. A new generation of candidate radome materials employing chemical (phosphate) bonding for processing is currently under investigation. Her work has resulted in numerous publications and patents.



DEBORAH A. HAUGHT received a B.S. in chemical engineering from Virginia Tech in 1985 and then joined NAVSWC as a member of the High Temperature Materials Group. While working at the Center, Ms. Haught completed work on the M.S. degree in materials science from the University of Virginia in 1989. Her principal involvements are in the development of celsian ceramics for future radomes and the development of mullite whiskers and mullite-whisker felt. These achievements have led to numerous publications and patents. She also has actively participated in the high-temperature superconductivity efforts.



ERIC J. WUCHINA received a B.S. in materials engineering from Virginia Tech and joined the Nonmetallic Materials Branch at NAVSWC in 1988. Mr. Wuchina recently defended his M.S. thesis, "Kinetics of the Hexacelsian to Monoclinic Phase Transformation in Barium Aluminum Silicate," at the Materials Engineering Department, University of Maryland, and plans to pursue a Ph.D. in materials engineering science

from Virginia Tech. Before joining NAVSWC, he was employed at the Oak Ridge National Laboratory as a cooperative education student, where he studied the chemical vapor deposition of nonoxide ceramic coatings. His current research focus involves slip casting of celsian and diamond film technology.



JAMES A. ZAYKOSKI graduated magna cum laude from Wilkes University with a B.S. in Materials Engineering. He joined NAVSWC in 1985. He received an M.S. in Engineering Materials from the University of Maryland in 1988 and is presently pursuing a Ph.D. degree. He has developed processing capabilities for carbon-carbon composites and evaluated new grades of carbon fiber. His current research projects include toughening of ceramics and the development of infrared transparent materials.

An Introduction to Directed Energy Technologies

Lawrence H. Luessen

The threats our naval forces will face in the future are becoming increasingly sophisticated and diverse. Seeker variety, speed, agility, and "stealthiness" are elements of the emerging threat — future missiles will employ assorted seeker options and variants; fly lower and faster; be more maneuverable; and have lower radar cross sections. This, in turn, will greatly reduce detection ranges and engagement times. As the total size of our naval forces decreases, point- and self-defense of individual platforms will become paramount. The bottom line — weapons which offer speed-of-light target interception times, rapid re-targeting capabilities, and large magazine capacities, such as directed energy weapons, may be the only answer to counter the future threat.

Although many claims and counter-claims have been made by the advocates and adversaries of directed energy weapons, the technologies associated with these weapons remain less understood than other more "conventional" weapon technologies. This article, meant to increase this understanding, addresses the following directed energy technologies: charged particle beams, high-energy lasers, and high-power microwaves. A short section is presented on pulsed power technology.

Introduction

The primary threat to U.S. naval ships in the current and future battle scenario is the anti-ship cruise missile (ASCM), a threat that is becoming increasingly sophisticated and diverse. Attack profiles range from sea skimmers to high divers; speeds are exceeding Mach 3; ranges are increasing to beyond 200 nmi; seeker types include radar (radio frequency, or RF), infrared (IR), electro-optic (E-O), anti-radiation (ARM and HARM), and dual-mode (RF/IR); warheads may be nuclear, high explosive, or submunitioned; home-on-jam capability is widely present; and ASCMs can be submarine-, surface-, and air-launched. This increased diversity and sophistication will, in turn, greatly reduce detection ranges and engagement times. As the total size of our naval forces decreases, point- and self-defense of individual platforms will become paramount. Directed energy (DE) weapons, which offer speed-of-light target interception times, rapid re-targeting capabilities, and large magazine capacities, may be the only answer to counter the future ASCM threat.

While many claims and counter-claims have been made by the advocates and adversaries of directed energy weapons, the technologies associated with these weapons remain less understood than other more "conventional" weapon technologies. This article is meant to increase this understanding, and addresses the following directed energy technologies: charged particle beams (CPBs), high-energy lasers (HELs), and high-power microwaves (HPMs). A short section is also presented on pulsed power technology, which may be essential to the successful development and deployment of directed energy weapon concepts.

The material to be presented covers several complex technical areas. We'll first begin with a few "leveling" definitions, followed by individual sections

on each DE technology. A brief discussion of pulsed power technology is then followed by some concluding remarks and thoughts. Table 1 briefly presents the characteristics of each directed energy technology that will be discussed.

Definitions

First, a few "leveling" definitions. Directed energy (DE) is an umbrella term covering technologies that produce particles (charged- or neutral-particle beams, CPBs or NPBs), photons (high- and medium-energy lasers, HELs and MELs), or concentrated electromagnetic (EM) energy (high-power microwaves, HPMs). Directed energy weapons (DEWs) are systems using directed energy, designed primarily to disrupt, damage, or destroy hostile targets. Directed energy warfare includes the operational use, both tactically and strategically, of DE weapons and devices in combat. Electromagnetic waves may be identified within the electromagnetic spectrum (see Figure 1) by frequency or wavelength, i.e., $f = c/\lambda$, where f = frequency, in cycles/sec (or Hz); c = the speed of light ($\approx 3 \times 10^8$ m/sec); and λ = wavelength, in meters, m.

Particle Beams — Charged or Neutral

A particle beam is a directed flow of atomic or subatomic particles. These particles may be either neutral (NPBs) or electrically charged (CPBs); the former are generally used in exo-atmospheric (i.e., "external," or outside the atmosphere) applications, while the latter are the choice for endoatmospheric (i.e., "internal," or within the atmosphere) applications. A particle beam transmits matter, not EM waves,

distinguishing it from other forms of directed energy such as lasers and microwaves. The particles penetrate material and deposit their kinetic energy deeply inside the material. Particle beam systems have the potential for being highly destructive weapons if the beam can be effectively propagated to the target. Upon reaching a target, the particle beam deposits its energy almost instantaneously within the target, resulting in rapid heat build up. In addition, the secondary radiation produced (x-rays and gamma rays) may damage or destroy electronic components in the target. The remaining discussion on particle beams will be confined to CPBs, specifically electron beams, since the Navy's interest in particle beam technology is anti-ship missile defense, an endoatmospheric application.

Components of a CPB Weapon System

Any CPB weapon system has three key subsystems: the particle-beam generator, the beam controller, and the fire-control subsystem.

Particle Beam Generator. This comprises an accelerator and its associated power-generating and power-conditioning subsystems.

Beam Controller. This subsystem points the beam at the target, most probably employing magnetic fields to "steer" the beam as it exits the accelerator.

Fire-Control Subsystem. This subsystem performs the same functions that are needed for a conventional weapon system. It detects and locates the target, passes control to the beam-control subsystem, assesses target kill, and redirects the beam controller to the next target.

While the beam controller and fire-control subsystems are important for a CPB weapon

Table 1. Directed Energy Technologies: Characteristics Summary

Medium	HEL	CPB	HPM	LEL(MEL)
	Photons	Electrons	EM Waves	Photons
Damage Mechanism	Thermal Kill	Kinetic Kill (Plus "Soft" Kill)	"Soft" Kill (Electronics)	Sensor Kill
Mission	Short-Range Anti-Air Warfare/ Anti-Satellite (SRAAW/ASAT)	SRAAW	Anti-Air/ Anti-Surface Warfare (AAW/ASUW) Strike	Anti-Periscope Anti-Missile Seeker Anti-E/O Target

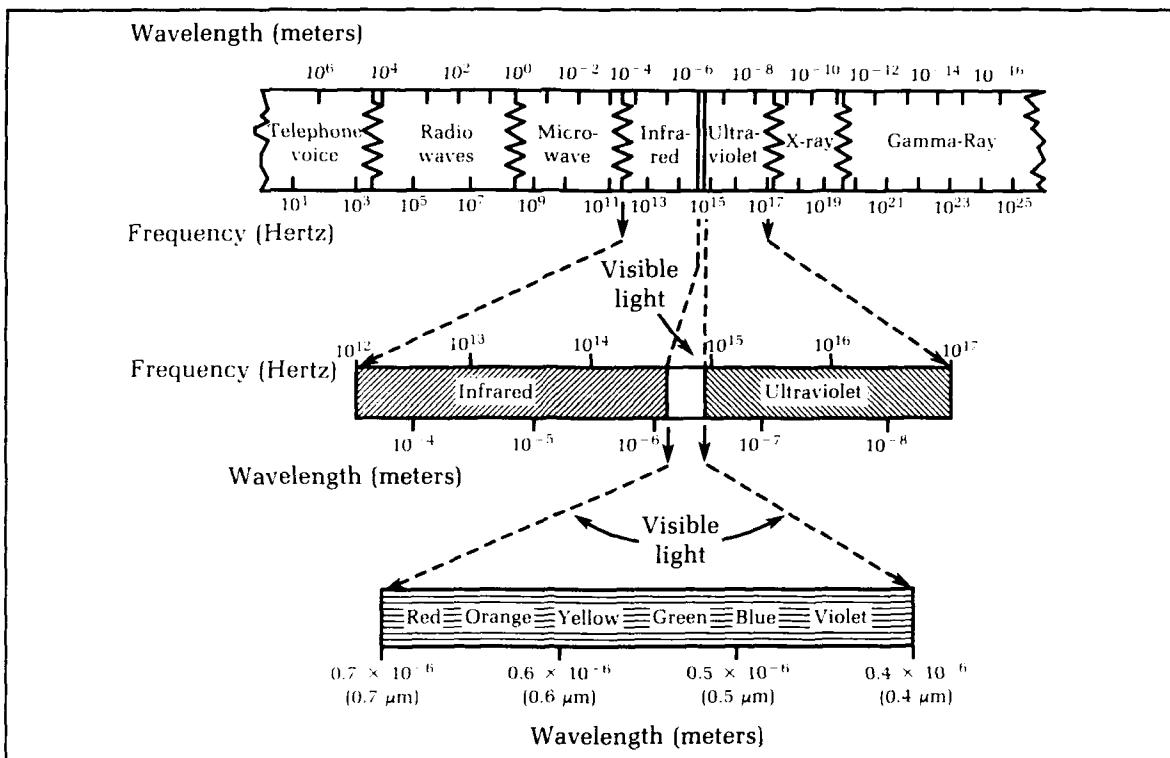


Figure 1. Electromagnetic (EM) spectrum.

system, let's focus on the heart of such a system — the accelerator.

CPB Accelerators

At the heart of any CPB system is an accelerator, a device that increases the speed (or kinetic energy) of charged particles by using electromagnetic fields or waves. Operation of such accelerators is quite simple in theory. Charged particles, e.g., electrons, are produced by applying a strong electric field (voltage) to a material that will emit electrons. By definition, an electron falling unimpeded through a potential difference of 1 volt attains a kinetic energy of:

$$\begin{aligned} E = eV &= 1.60 \times 10^{-19} \text{ coulomb} \times 1 \text{ volt} \\ &= 1.60 \times 10^{-19} \text{ joules} = 1.60 \times 10^{-12} \text{ ergs} \\ &= 1 \text{ electron-volt (eV).} \end{aligned}$$

The electrons are then injected into regions, or accelerating stages, in which they encounter large electric fields, or voltage gradients. As the electrons pass each stage, their velocity increases, and, as they approach the speed of light, relativistic effects cause the mass of the electrons to increase. The mass and velocity determine the energy of the electrons, which determines their penetrating and damaging ability.

114

The lethality of the beam is principally determined by two beam parameters. The first is the energy of the electrons, which is roughly determined by the total voltage the electrons have seen — megaelectronvolts (MeV) to gigaelectronvolts (GeV) for weapon-grade beams. The second is the number of electrons in the beam, which is related to the beam current, and may range from kiloamps (kA) to megamps (MA).

The peak power (peak voltage x peak current) of the beam is extremely high; however, it is produced in short pulses within low repetition-rate bursts to keep the average prime power requirements reasonable.

Types of Accelerators

The majority of particle accelerators are linear, i.e., the accelerating stages are arranged in a straight line, eliminating the need to "bend" the beam. Such accelerators, called linear induction accelerators, or linacs, generally utilize a core of magnetic material (ferrite) that forms a toroidal ring around the accelerating column. The change in flux in the magnetic core induces an axial electric field that accelerates the electron beam down the accelerating column. It therefore can be envisioned as a series of one-to-one pulse transformers threaded by the electron beam.

Unfortunately, such an arrangement tends to be very long. The Advanced Test Accelerator (ATA), developed by the Defense Advanced Research Projects Agency (DARPA) at Lawrence Livermore National Laboratory as a test bed, employs an 85-meter-long linear electron accelerator. The injector produces a 10-kA beam of 2.5-MeV electrons, which are guided by magnetic fields through an accelerator consisting of 200 separate accelerator stages. Each stage increases the electron energy by 0.25 MeV, giving a final beam energy of approximately 50 MeV.

This type of accelerator is obviously not a strong candidate for shipboard use. However, three separate efforts are currently under way to develop alternative compact accelerator concepts. Funded by DARPA and the Navy, all three devices rely on cycling the beam through the accelerating sections to reduce the overall system size and weight. Figure 2 summarizes the differences in these concepts and highlights key issue areas.

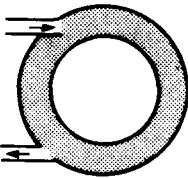
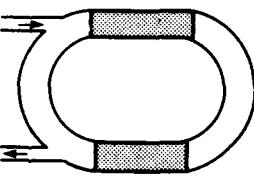
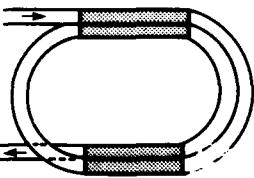
Compact Accelerator Concepts

The Modified Betatron (Figure 2) is a circular device being developed at the Naval Research

Laboratory which reuses the accelerating sections hundreds to thousands of times in accelerating each pulse. Important trade-offs involve the time to accelerate, and, hence, minimum pulse separation, versus accelerator radius, versus voltages. Critical issues include the ability to inject and extract the beam into and out of the ring without introducing instabilities.

The Ion-Focused-Regime Recirculator (Figure 2) is a racetrack concept being pursued by Sandia National Laboratory/Albuquerque which restricts the accelerating sections to straight portions of the device, but passes through each accelerating section six to eight times. The beam is guided around the racetrack by a low-energy electron beam which can be deflected by very low-flux magnets, and which creates a channel for the higher-energy beam. Critical issues include instabilities introduced by multiple beam passes through dielectric cavities, closed-orbit beam transport within the IFR channel, and injection and extraction.

The Spiral-Line Induction Accelerator (Figure 2), being developed by Pulse Sciences, Inc., builds on the ATA technology. In order to avoid injection and extraction issues, the beam passes through the accelerating sections on a

	Modified Betatron	Ion-Focused Regime Recirculator	Spiral Line Induction Accelerator
Layout:			
Accelerating Section	Circular	Straight	Straight
Transport Mechanism	Magnetic	IFR	Magnetic
Core Type	Air Core	Dielectric Cavity	Ferrite Cavity
Type of Accelerating (Slow Pulses)	Continuously Ramped Field	Pulsed to Reset Cavities	Pulsed to Reset Cavities
Injection	No Issue	Minor Issue	No Issue
Extraction	Issue	Issue	No Issue
Pulse Train Formation: Rep Rate (<50 Hz) WIPS* Burst Mode	No Issue Issue	No Issue No Issue	No Issue No Issue
Typical Number of Passes	100 to 1000 (Limit is Acceleration Time vs. Interpulse Spacing)	4 to 8 (Limit is Deterioration of Pulse Shape)	3 to 19 (Limit is Complexity and Number of Beam Lines)

*Wide-interval pulse-separation

Figure 2. Compact accelerator options.

slightly different path for each cycle, thus permitting use of a continuous beam line. The limiting factor is the number of beam lines which can pass through the cavities. Critical issues include designing the cavities and gaps to minimize instabilities and matching fields at transition points.

Other Issues

At least two other issues are important in the development of CPBs for possible anti-ship missile defense (ASMD) applications — satisfactory propagation of the e-beam and kill mechanisms.

Beam Propagation. Once the electron beam pulse is produced, it must propagate satisfactorily through the atmosphere. The high-velocity electrons will collide with air molecules and lose energy. This deceleration of the electrons causes them to radiate, which is responsible for "soft-kill" mechanisms described later. The beam energy decreases as it penetrates the air, until at a certain distance the beam energy is too low for use as a weapon.

Methods to improve the somewhat limited range of CPBs are currently being investigated, including a wide-interval pulse-separation (WIPS) mode of propagation. This uses the fact that air molecules struck by the beam are heated and moved out of the way for a short period of time, creating a rarified "tube" in the atmosphere. A second pulse or beam can propagate through this rarified region before the air density recovers. In this manner, a series of beams can be propagated through the path, "hole boring" so that each beam pulse travels further than the last. It is the nth pulse arriving at the target that must cause the damage — all preceding pulses contribute to the hole boring.

In addition to the rarification of air, the charges left in the "tube" of air will attract the follow-on beams and provide a channel-tracking force to keep them in line and help override the bending effects of the earth's magnetic field and other beam instabilities. However, hole boring is limited by channel overheating, which then causes the channel to become conductive enough to expel later pulses. Therefore, there is a limitation on the number of pulses in the hole-boring pulse train.

Kill Mechanisms. A beam of high-energy electrons is typically a few centimeters in diameter. When this beam strikes a target, energy is deposited deep within the target in several microseconds (much faster than a laser), melting or vaporizing material deep within the target, and creating a thermal-shock mechanism that is difficult to counter or shield against.

In addition to such hard-kill mechanisms, soft-kill effects may also result from the elec-

116

tron beam propagating through air. The moving electrons in the beam create a strong electric and magnetic field, and therefore create electromagnetic-pulse (EMP)-type effects near the beam's axis. This EMP may be large enough to damage electronics within the target even for near-miss conditions.

For beam energies above 100 MeV, the dominant mechanism of energy loss as the electrons are decelerated as they strike air molecules is a process called bremsstrahlung, where electron kinetic energy is converted to forward-directed electromagnetic radiation in the form of gamma rays. The gamma rays, in turn, will continue on to form secondary particles (electron-positron pairs), which will cascade down to lower-energy electrons, neutrons, X-rays, high-energy photons, and RF noise. This radiation forms a narrow cone surrounding the beam and is directed with the beam — again opening up the possibility of soft-kill effects on the target's electronics.

High-Energy Lasers

A laser (light amplification by stimulated emission of radiation) is any device or medium that produces a highly monochromatic, coherent, and directional beam of radiation. The process known as stimulated emission occurs when an atom, in its excited state, is influenced by a photon (of the proper energy and wavelength) and is stimulated to emit a second photon identical in wavelength to the first and direction of travel, and in phase with the first. The radiation frequency varies according to the material used and the method of exciting it.

A laser contains four elements:

- (1) The active medium, a collection of atoms that may be excited to a short-lived higher state of energy known as population inversion (e.g., ruby rods, neodymium-YAG crystals, or dye-doped laser rods).
- (2) The excitation mechanism, a source of energy that moves the atoms from a ground state to an excited state to create population inversion.
- (3) The feedback mechanism, a system that returns a part of the coherent light produced in the active medium to continue the stimulated emission process, i.e., a mirror to redirect the photons.
- (4) The output coupler, which allows a portion of coherent light to leave the laser in the form of the output beam; essentially a "hole" in one end of the lasing cavity and a beam director for pointing and tracking.

Lasers can be classified as solid, liquid, or gas; there are also free-electron lasers. The different types of lasers are shown in Table 2.

Laser Beam Divergence and Focusing

Laser beams spread or diverge after leaving the laser. While light from ordinary sources is emitted in all directions, the light emitted from a laser is confined to a very narrow cone. Nevertheless, as the beam propagates outward, it slowly diverges or fans out. The greater the beam's angle of divergence, the greater will be the diameter of the beam, and the lower the irradiance and radiant exposure (defined shortly), at the target. If R = distance from the output aperture of the laser, θ = full-angle beam divergence in radians, d = initial beam diameter, and d' = the beam's diameter at distance R , then:

$$d' \approx R\theta + d. \quad (1)$$

θ is normally between a milliradian (mrad) and a microradian (μ rads). This divergence corresponds to a 1-meter beam diameter at 1 km for a milliradian and a 1-meter beam diameter at 1000 km for a microradian.

We now need to spend some time on defining the measurement of electromagnetic radiation. The terms that will be introduced apply to both lasers and microwaves.

Radiometry. Radiometry is the science of measuring electromagnetic radiation. Radiant energy is the quantity of energy traveling through space in the form of light waves, usually measured in Joules (J). Radiant power is the radiant energy transferred per unit time, usually measured in watts (W). Irradiance (I), or power density, describes the amount of energy per unit time per unit area, and is usually measured in watts per square meter (W/m^2) or watts per square centimeter (W/cm^2). The size of the area on which the laser beam is concentrated makes a tremendous difference on the

Table 2. Types of Lasers

Class	Type	Example	Excitation
Solid	Insulator	Ruby Nd:YAG Nd:Glass Alexandrite Color Center	Flashlamp
	Semiconductor	Ga:As	Laser Current
Liquid	Organic Solvent	Dye Rare Earth Chelate	Laser
	Inorganic Solvent	Rare Earth Ion	
Gas	Atomic	He:Ne	Electrical Discharge
	Ionic	Argon	
	Molecule	CO N ₂	
		KrF(Eximer) XeCl(Eximer)	
		CO ₂	Sealed Tube Axial Transverse Gas Dynamic
		HF DF	Chemical
Free Electron Laser			

impact of the power delivered — the smaller the cross-sectional area of the target receiving the beam's power, the higher the irradiance, or power density.

Radiant exposure (H), or energy density, is the radiant energy per unit area striking a surface, and is expressed in Joules per square centimeter (J/cm^2). Radiant intensity is a measure of the rate of light, or radiant energy output, leaving a point source per solid angle in space. Radiant intensity is normally measured in watts per steradian (W/sr). The steradian, like the radian, is a unitless quantity, and is the solid angle (Ω) defined in the same manner that a plane angle is defined in radians. It may be thought of as the opening at the tip of a cone which is subtended by a segment of area on a spherical surface. This solid angle is independent of the distance from the source. Just as there are 2π radians in a complete circle, there are 4π steradians in a complete sphere.

Divergence, Beam Diameter, and Spot Size. An important variable in the power of the output beam of a laser occurs across the diameter of the beam. The irradiance, or power density (in watts per unit area), is not the same throughout the cross section of the laser beam. Instead, a Gaussian beam is used to describe irradiance.

The Gaussian beam operates in the transverse EM mode, known as TEM_{00} . Most tactical lasers are designed around TEM. The divergence of the laser beam itself is generally determined by the intrinsic size of the beam within the laser's optical cavity. The maximum irradiance (power density) of a Gaussian beam is at the center of the beam and falls off as $\exp(-(\rho/w)^2)$ with the distance ρ from the center of the beam. The value of ρ for which the beam irradiance decreases to $1/e$ of its value at the center is termed the spot size, w . At one point in the cavity, called the beam waist, the Gaussian beam has its minimum spot size, w_0 . The spot size at the beam waist determines the divergence of the laser beam, since the external beam is really an extension of the internal laser beam within the cavity. The beam diameter is defined as the distance across the center of the beam for which the irradiance exceeds $1/e^2$ of the maximum irradiance.

The beam can be focused, but the minimum focused spot size is limited by diffraction (bending of light) around the largest opening, which is generally the focusing lens. Using diffraction theory, it can be shown that the Gaussian beam from a laser operating in the TEM_{00} mode can be externally focused to a spot size of radius:

$$w_0 = \frac{\lambda}{\pi} \times \frac{f}{w_1}, \quad (2)$$

where f = focal length of the focusing lens used, λ = laser wavelength, and w_1 = the spot size, or radius, of the (collimated) beam incident on the lens, as measured from the beam center to the point where the field amplitude is down by $1/e$ (the irradiance is $1/e^2$ of that at the beam center).

If the lens or aperture is nearly uniformly illuminated, the minimum radius of the focused spot of light is:

$$w_{min} = \frac{1.22 \lambda \times D_T}{A_D}, \quad (3)$$

where D_T = distance to target and A_D = the diameter of the aperture, or focusing lens. In practice, most lasers are worse than this spot-size limit, often by a factor of 10. For long distances, focusing becomes a problem if the spot size must be small. For example, at 10 km, to focus a beam down to a spot size of 1 cm for a CO_2 laser ($\lambda = 10.6$ microns, where 1 micron = 10^{-6} meters = $1 \mu m$) would require a lens 13 meters in diameter. If a Nd:YAG laser ($\lambda = 1.05 \mu m$) is used, the lens diameter would drop to 1.3 meters. This relationship demonstrates the desirability of shorter wavelengths to reduce the optics requirements. However, shorter wavelengths require more precise optical components and alignments.

Beam diameter is essential for calculating irradiance and radiant exposure. If the power remains constant, the smaller the beam diameter, the higher the irradiance (power density); if energy remains constant, the smaller the beam diameter, the higher the radiant exposure (energy density).

Atmospheric Effects on Laser Beam Propagation

Propagation of a laser beam is a complex phenomenon depending on many physical effects that must be considered in the design of a weapon or other laser device.

Water vapor, aerosols (e.g., haze), rain, snow, fog, clouds, and turbulence all have degrading effects upon the laser beam. The degree and type of beam degradation is dependent upon the "atmosphere" through which the laser beam propagates, e.g., a vacuum, linear atmosphere, or nonlinear atmosphere.

Vacuum. In a vacuum, i.e., at very high altitudes for exoatmospheric applications, the beam irradiance is reduced with propagation distance due mainly to diffraction or beam spreading.

Linear Atmosphere. When a laser beam propagates near the earth's surface, there are degradation effects associated with the linear atmosphere, i.e., where the laser beam itself does not alter the normal properties of the atmosphere such as pressure, density, and temperature. Linear effects on the laser beam include beam absorption and scatter caused by molecules and aerosols; beam degradation due to adverse weather conditions such as rain, snow, clouds, or fog; and atmospheric turbulence.

Nonlinear Atmosphere. When the laser beam irradiance is high, additional nonlinear propagation effects occur. They arise because the high-power beam, while passing through the atmosphere, alters the normal properties of the atmosphere, e.g., heats the air. These effects, including thermal blooming, kinetic cooling, and air breakdown, add to the already complex beam degradation taking place due to diffraction and linear atmospheric effects.

Beam pointing and tracking is very difficult for a laser if it is necessary to produce a small spot size on a target and hold it there for several seconds. Accuracy requires that atmospheric effects, such as turbulence, be compensated for by some sort of deformable mirror or adaptive optics. It also requires precise opto-mechanical engineering of the beam director and microradian tracking capability.

Laser Damage Mechanisms

The principal damage from a laser is the rapid build up of heat on the target — the damage mechanism is *thermal*. The energy heats the surface of the target, melts the material, and burns through to the interior, destroying vital components. A focused high-energy laser beam can produce temperatures exceeding those of the sun, causing the target to weaken structurally or melt. However, heating requires time and is not instantaneous as in the case of a CPB. This time period, called dwell time, is dependent on the irradiance (power density) reaching the target, and can vary from tenths of seconds to several seconds.

It should be noted that while HELs are meant for structural kill, low- and medium-energy lasers can also be extremely effective in defeating optical and electro-optical systems. In addition to inflicting temporary to permanent blindness on operators of these systems, these lower-powered lasers may be capable of causing similar results to the E-O systems themselves.

Navy HEL Interests

The Navy is interested in two primary HEI, "weapon" candidate technologies: free-electron lasers (FELs) and chemical lasers. The former may be years away from a practical demonstration, while the Navy's chemical laser efforts have shown some progress over the past several years.

Free-Electron Laser. In an FEL, a stream of electrons is accelerated through a cylinder and subjected to an undulating magnetic field at intervals along the cylinder, i.e., every few centimeters. This undulating field causes the electrons to deflect back and forth — in effect forcing them to vibrate and emit laser radiation as they pass through the cylinder. The wavelength of the laser energy is altered by changing the distance between field reversals. The greatest disadvantage with FELs is that they require a large electron-beam accelerator to generate the electrons, which in turn necessitates a large power supply.

Chemical Lasers. In a chemical laser, two atomic gases are injected into a combustion chamber and ignited. As the gases react, excess energy allows a single electron to transfer back and forth between an atom of one gas and an atom of the other, keeping the new molecule together. As the electron oscillates, it emits radiation in the form of laser light. The Navy's Mid-Infrared Advanced Chemical Laser (MIRACL) (see Figure 3) burns ethylene with nitrogen trifluoride to produce a hot, high-pressure gas in a combustion chamber — much like a rocket engine. The gas expands through nozzles and a small amount of deuterium is added. The resulting reaction produces excited deuterium fluoride molecules from which the lasing occurs. Two mirrors form an optical resonator which extracts the energy from the excited gas in the form of an intense beam of IR radiation. The beam is then transmitted by relay optics to the beam director. In the beam director, an E-O or IR tracker acquires and tracks the target. Servos driven by signals from the tracker drive the optics and point the laser beam at the desired aim point.

Since chemical lasers generate power from the combustion of chemical reactants, they require only modest amounts of electricity.

High-Power Microwaves

Since the development of radar, great emphasis has been placed on the military potential of microwaves. Similar to light, microwaves are electromagnetic waves that share many of the same properties of light. Microwaves travel in a straight line at the speed of light, and can be reflected or absorbed by various materials. Con-

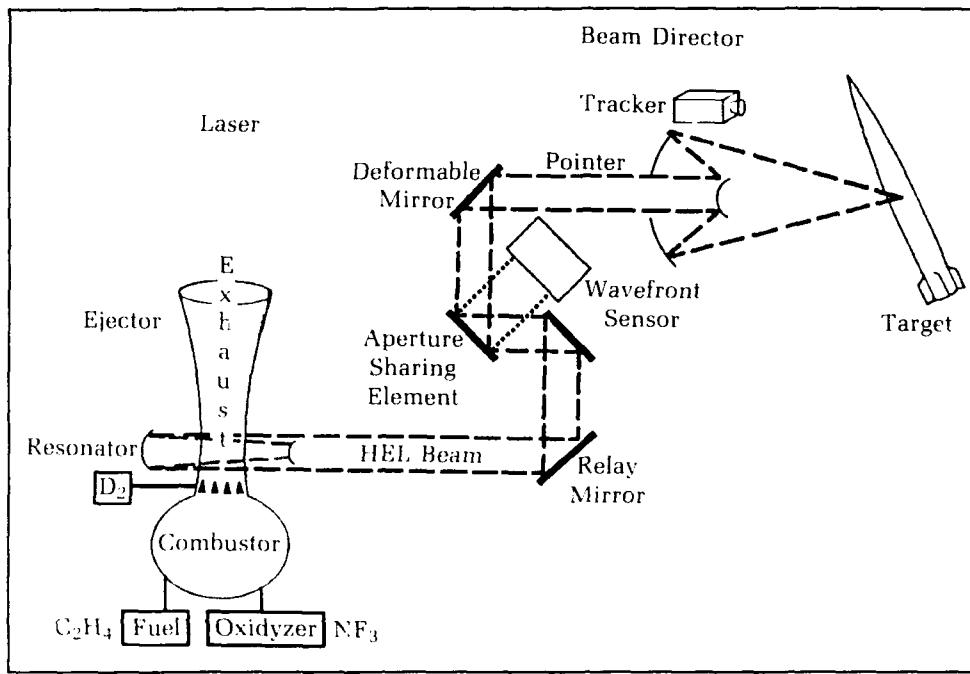


Figure 3. Mid-infrared advanced chemical laser (MIRACL).

ventional microwave systems exist on all naval platforms for many purposes, including communications, navigation, and fire control. In the directed energy area, an HPM might be used to jam or render ineffective a wide range of electronic systems such as radars, communications, and sensors.

An HPM system is distinguished from these existing microwave systems primarily by the magnitude of its peak output power (0.1 to 10 GW). The average power level is 1 MW or more, the energy in a single RF pulse is less than a joule to 10 J or more, and operating frequencies are 100 MHz to above 35 GHz. Like a conventional microwave system, an HPM system consists of a (a) power conditioning system, (b) microwave source, and (c) radiating antenna. If an HPM is to be used as a weapon, there are also target considerations to be examined, i.e., is the potential target susceptible to an HPM source? We'll first look at the HPM source and its antenna, then address target issues.

120 Microwave Sources

Microwave sources for HPM systems deliver high-power, short-duration RF pulses. Devices commonly used include klystrons, magnetrons, backward-wave oscillators, virtual-cathode oscillators (VIRCATORS), and gyrotrons.

Typical sources operate in the 1 to 35 GHz range; capabilities in the 1 gawatt range (peak), with pulse widths of a few nanoseconds, have been demonstrated.

The output of a typical HPM source is normally a pulsed sine wave carrier similar to a pulse radar, but with much higher powers and much lower pulse repetition rates (PRRs) or frequencies (PRFs). A typical HPM pulse train is shown in Figure 4, and includes the critical parameters of carrier frequency (f_c), pulse width (τ), PRF, pulse rise time (t_R), and amplitude.

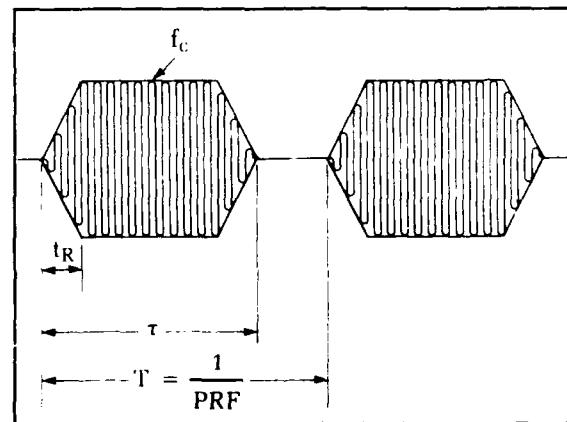


Figure 4. Typical HPM pulse train.

The energy per photon carried by a microwave is directly proportional to the frequency. At the lower frequency bands, wavelengths are much longer, and devices to generate and direct the energy efficiently are much larger. The RF bands defined by the radar and electronic countermeasure (ECM) communities, respectively, are shown in Figure 5. Microwaves are more difficult to focus to a small spot size than lasers since the angular spot size is roughly proportional to the wavelength divided by the diameter of the output aperture or antenna.

Microwave Source Antennas

The rules governing HPM antennas are similar to those for conventional microwave systems. However, because of the high powers, there are some considerations unique to HPM systems. Like conventional antenna design, the physical size of the antenna in wavelengths determines the focusing of the microwave radiation or beam. The larger the antenna, the narrower the beam width, and, thus, the more directional the beam (i.e., more of the available source energy can be focused on the target). For HPMs, the ability of the antenna to focus radiation on a target is limited by two factors: the diffraction limit and air breakdown.

Diffraction Limit. If the peak power output of an HPM source is denoted by P_s and is radiated uniformly in all directions, the power density, P_T (power per unit area), at any distance R from the HPM source can be determined by dividing the transmitted power by the surface area of an imaginary sphere of radius R , i.e.,

$$P_T \text{ (Omnidirectional)} = \frac{P_s}{4\pi R^2}. \quad (4)$$

The power output of a directional antenna is related to P_T of an omnidirectional antenna by the antenna's power gain:

$$P_T \text{ (Directional)} = \frac{P_s G}{4\pi R^2}. \quad (5)$$

Therefore, the power density, P_T (in W/cm²), available at a distance or range, R (in km), from an HPM source with a peak source power, P_s (in W), and antenna directivity or gain, G , is given by the one-way radar equation:

$$P_T = 8 \times 10^{-12} \frac{P_s \times G}{R^2}. \quad (6)$$

The factor G , or directivity, of the microwave source antenna, which is a relative measure of

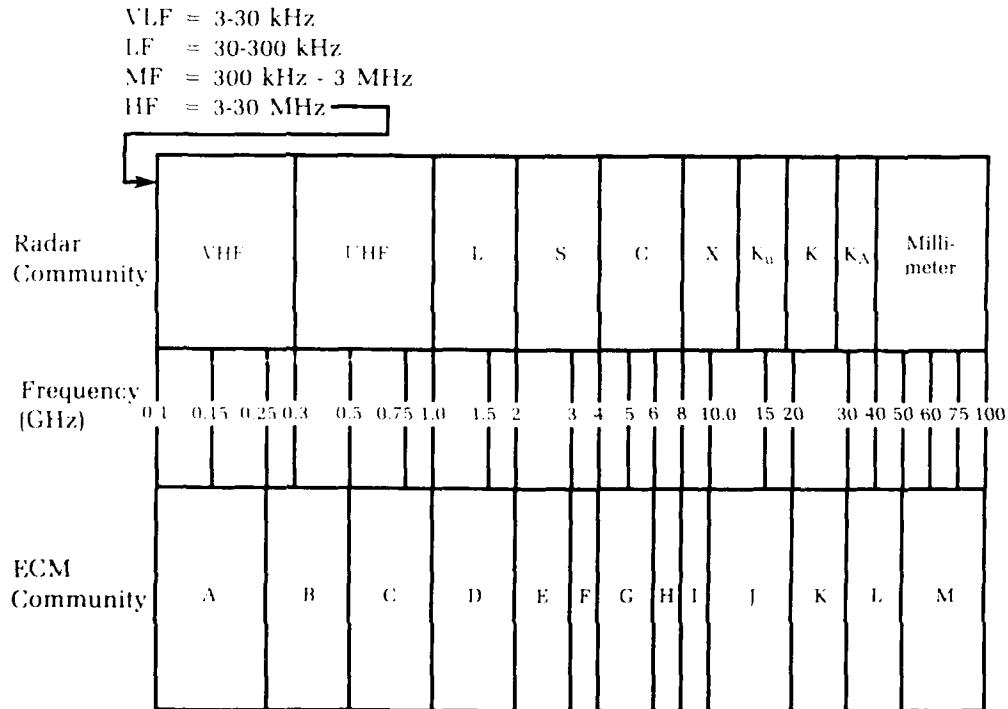


Figure 5. Radio-frequency bands.

the antenna's ability to focus the microwave beam in a particular direction, is commonly called the gain of the antenna. Strictly speaking, gain is differentiated from directivity in that gain takes into account any losses in the antenna. Large values of gain are expressed in decibels (dB). Antenna directivity can be shown to be:

$$G = \frac{4\pi A\eta}{\lambda^2} = \eta \left[\frac{\pi^2 D_A^2}{\lambda^2} \right], \quad (7)$$

where A = antenna area (in cm^2), D_A = the antenna diameter (in cm), η = an efficiency factor to account for losses (usually 0.5 to 0.8), and λ = the radio frequency wavelength (in cm).

Combining equations (6) and (7):

$$P_T = 8 \times 10^{-12} \frac{P_s \eta \pi^2 D_A^2}{R^2 \lambda^2}. \quad (8)$$

Air Breakdown Limitation. One final value related to the microwave source that must be determined is the intensity at the aperture of the microwave source antenna. Assuming once again a source power of P_s (in W) and a source antenna diameter of D_A (in cm), the aperture intensity, P_A (in W/cm^2), is given by:

$$\begin{aligned} P_A &= \frac{P_s}{[\text{Aperture Area}]} = \frac{P_s}{\left[\pi \times \left(\frac{D_A}{2} \right)^2 \right]} \quad (9) \\ &\approx \frac{4 P_s}{\pi D_A^2}. \end{aligned}$$

However, there is a limitation on air breakdown. Clean, dry air breaks down between 1.4 and 2.0 MW/cm² for pulses 100 nsec or longer, which limits values for P_A that can be used by HPM source designers. Therefore, coupling gigawatt microwave sources to an antenna without arcing at the feed aperture poses quite a challenge for HPM systems.

Effective Radiated Power (ERP). ERP, usually expressed in watts, is also employed by the microwave community to describe microwave sources. The ERP can be defined as:

$$ERP = P_s \times G \text{ (dB)}, \quad (10)$$

where G is normally expressed in dB.

122 Target Considerations

The energy coupled into a target is highly dependent on the critical parameters mentioned previously, i.e., frequency, pulse width, pulse shape, intensity of the radiation, and the repetition rate. Each of these parameters can in-

fluence the susceptibility level of a potential target's electronics by either affecting how much energy can reach critical electronics, or the level of energy required to affect the piece-part electronic component. The microwave energy coupled into a target also depends on the coupling path(s) available, namely front-door or back-door coupling.

Front-Door Versus Back-Door Coupling.

Most HPM concepts envisioned and proposed depend on "soft" kill of a target's electronics (i.e., burnout), or upset and disruption of its electronics. HPMs interact with a target in one of two ways: "front-door" or "back-door" coupling. The former refers to any microwave energy entry through intentional paths intended for microwave reception, such as radar front ends, RF or E-O transmit-receive paths, etc. The latter refers to any entry via unintentional paths, e.g., cables, connectors, and seams.

HPM Methodology. Since a system's susceptibility is dependent on the above critical parameters, and the parameter space is so broad, investigating the susceptibility of a system over every possible combination of parameters is almost impossible. Instead, a parametric approach is normally taken where the susceptibility level is determined for representative HPM source parameters.

HPM Susceptibility Analysis. The HPM front-door energy entry path for a typical system is basically through the antenna, down the transmission or signal path, and to the critical sensitive electronic component. The power density, P_T , required to transfer sufficient power through the system to affect the critical component may be determined by a knowledge of the component power damage threshold, P_D , the entry path transmission loss, L , and an effective capture area of the energy port-of-entry or antenna, A_E . The effective antenna area, A_E , sometimes called an effective aperture, can be defined as:

$$A_E = \frac{[\text{Power Reaching Receiver}]}{[\text{Incident Power Density}]} . \quad (11)$$

The effective area is related to the antenna gain by:

$$A_E = \frac{G \times \lambda^2}{4 \pi} . \quad (12)$$

An analytic procedure can then be used to predict/estimate the worst-case susceptibility threshold of a system by the following steps:

(1) Determine by measurement or estimate P_D , the minimum damage power threshold of the component most likely to be damaged, as a function of pulse width, τ , carrier frequency, f_c , and PRF.

(2) Determine by measurement or estimate the transmission loss (i.e., energy attenuation) of the HPM entry path, L, as a function of frequency.

(3) Determine by measurement or estimate the gain of the antenna, G_{max} (for maximum energy coupling), as a function of frequency, aspect angle (azimuth, ϕ , and elevation, θ), and polarization, P.

(4) Convert the antenna gain to an effective capture area using the equation:

$$A_E = \frac{G_{max} \lambda^2}{4 \pi}. \quad (13)$$

(5) Determine the susceptibility threshold of the system (i.e., power density required for damage), P_{TD} , by combining the above parameters into the following equation:

$$P_{TD}(f, \phi, \theta, \tau, P) = \frac{P_D(\tau, f)}{A_E(f, \phi, \theta, P) L(f)}. \quad (14)$$

Other Issues

Several other issues are of concern to the HPM community, including fratricide, power versus energy on target, and narrow-band versus wide-band sources.

Fratricide. Fratricide is of great concern. The antenna must be designed so that side lobes are reduced to protect own ships' personnel and equipment.

Power Density Versus Energy Density. For single-pulse damage effects, the pulse duration (width) and peak power determine the energy deposited into a target's electronic component. However, energy density may be less important for single-pulse HPM weapon concepts than for repetitive (multi-pulse) concepts. In this case, the lethality of an HPM weapon may depend not only on the power density at the target, but also on the pulse shape (particularly pulse rise time) and on fluence, i.e., cumulative energy density deposition. For N pulses on a target (per burst if the weapon operates in a burst mode), with a pulse duration (width) of τ , and power density at the target of P_T in W/cm^2 , the fluence, F, in J/cm^2 , can be expressed as:

$$F = P_T \times \tau \times N. \quad (15)$$

For continuous wave devices, the product $N \times \tau$ indicates the total dwell time, which may be limited by the operational characteristics of the engagement. There is also a trade-off — care must be taken to ensure that pulse separation is less than the time in which a target's electronic component can dissipate or conduct away the heat.

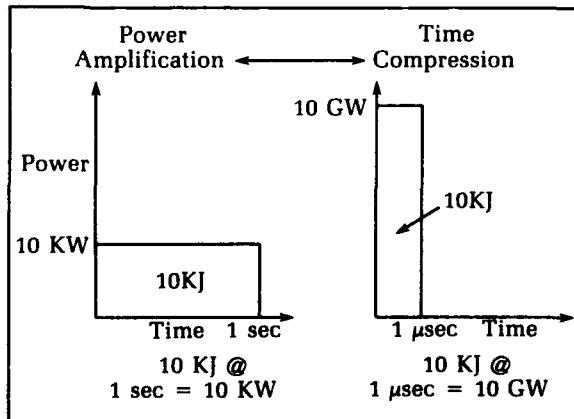
Narrow-Band Versus Wide-Band. Historically, most HPM weapon concepts use narrow-band, single-pulse sources, and emphasize generic burnout of a system's electronic components. These sources are designed to operate in-band to the postulated target system's susceptibility. This is true whether front-door or back-door coupling of the radiated energy is anticipated. However, this does require some prior knowledge of a system's operating characteristics, not unlike the requirements placed on the electronic warfare (EW) community.

More recently, while still investigating burnout effects, the HPM community is beginning to shift its emphasis to generic upset of a system's electronic components, while considering higher carrier frequencies and higher PRFs. At the same time, there has recently been keen interest in wide-band and ultra-wide-band (UWB) technology, both from the detection community, i.e., UWB (impulse) radar, and the ECM community. The generally accepted definition of wide band? The band width, Δf , is commensurate with the center (carrier) frequency, f_0 , i.e., $\Delta f/f_0 \sim 1$.

Like narrow-band, wide-band is dominated by coupling. The energy must be maximized at a potential target's critical frequency and bandwidth. However, even for wide-band effects, generic jamming may mean tunability, i.e., frequency or pulsedwidth agility. There is also a lack of understanding and development of directional wide-band antennas.

Pulsed Power Technology

Finally, a short definition of and introduction to pulsed power is presented. Pulsed power is the compression, in time, of energy. For example, in Figure 6 a capacitor may be charged



123

Figure 6. Pulsed power defined.

from a power supply and store 10 kJ over 1 second, i.e., 10 kW of power. If this same 10 kJ is then discharged (via a "switch") to a load in 1 μ sec, i.e., compressed by 1,000,000 (10^6) in time, 10 GW of power will be delivered to the load. The technologies that are employed to accomplish this compression in time include prime power, i.e., a power "plug"; energy storage, i.e., the storage media; and switching, i.e., what are the volts and amps to be switched, and at what repetition rate?

The development of pulsed power components and their associated technologies is critical to the success of several directed-energy-technology concepts presently being developed. For example, efforts in NAVSWC's Pulsed Power Technology Branch include the development of high-power, high-repetition-rate, high-pressure spark gaps for use in compact e-beam accelerators. Smaller, inexpensive spark-gap-based LC oscillators are also being developed for ECM applications.

In the area of HPMs and EW, a new initiative in photoconductive semiconductor switches has begun. A new technology, the bulk optically controlled semiconductor switch (BOSS), based on copper-compensated semi-insulating GaAs, is being developed with the capability of closing and opening on command with two different laser pulses. When this technology is embedded in a simple RF source called a pulse switch-out generator, synthesis of multiple, high-power RF pulses may be possible with only two switches. This, in turn, will allow true frequency and pulse-width agility in an RF source.

NAVSWC's Pulsed Power R&D Facility, a source of high-voltage prime power (50 KV @ 4 A, or 200 kW), includes 12.5 kJ of energy storage capacity in the filter bank. It is used to test and develop energy storage devices and switching concepts, i.e., pulsed power technologies. However, when a water-based pulse-forming line, an energy storage device, is utilized, it is capable of generating 50-kV, 200-nsec pulses into a 2- Ω load.

Summary and Conclusions

As the missile threat to the surface Navy pushes the limits of conventional ASMDs, directed energy technologies, with their speed-of-light propagation velocities and various hard- and soft-kill mechanisms, may offer tactical alternatives for own-ship defense, either as

an adjunct to a conventional missile defense system or in a stand-alone posture. However, for each advantage articulated for a particular DE technology, at least one, and sometimes several, disadvantages can be presented, namely:

- CPB: Advantages include kinetic energy kill, with energy deposited rapidly (μ secs) and deeply, as well as hard- and soft-kill mechanisms; however, propagation distances are limited, there is a need to develop compact accelerators, and fratricide may be a severe problem.
- HEL: This is a mature technology, with propagation and hard kill demonstrated; an HEL (MIRACL) has been built and operated; however, propagation is severely limited by weather and atmospheric conditions, "long" dwell times (up to a few seconds) are needed for thermal kill, non-linear atmospheric effects exist, and MIRACL is large.
- HPM: Again, this is a mature technology with much available susceptibility data; however, kill mechanisms are not entirely understood nor immediately verifiable, fratricide issues need to be addressed, generic electronics "kill" is not likely, and repetitive systems may be more desirable.

There remain many questions facing the directed-energy community, including pay-off versus investment costs for each individual directed-energy technology, ship compatibility, and near-term applications. For example, in the HPM area, how (or will) ultra-wide-band technology be incorporated into future concepts? Can it truly be used by the detection (radar) community, or is this technology more likely to find applications in the EW community? And the most likely directed energy technology candidate for near-term incorporation into the Navy? Probably some form of an HPM-based system, either as an adjunct in a point-defense suite, or in an ECM (e.g., jamming) role.

As budgets shrink and priorities shift from blue-water tactics to own-ship defense, as faced by the Navy in the Persian Gulf, a decision will undoubtedly be made on which DE technology to pursue and develop to deployment. Although this may mean that some promising technologies go into hiatus for several years, directed energy technologies offer a promising means of countering future threats to U.S. naval forces.

Bibliography

- Berkowitz, B., *Basic Microwaves*, Hayden Book Co., Inc., New York, NY, 1966, pp. 58-59.
- Chesser, Nancy, Ed., *Status Summary — Compact High-Current Electron Accelerators*, Directed Technologies, Inc., Arlington, VA, 16 Nov 1989.
- Directed Energy Warfare Technology and Mathematics Primer*, U.S. Army Combined Arms Command and Fort Leavenworth, Combined Arms Command, Combat Developments, 8 Nov 1990.
- Frieden, D.R., Ed., *Principles of Naval Weapons Systems*, Fourth Printing, 1988, Naval Institute Press, Annapolis, MD, 1985.
- Luessen, L.H., *Directed Energy Weapons System (DEWS) Strategic Business Unit (SBU): FY89-90 Summary*, NAVSWC, Dahlgren, VA, Feb 1990.
- Mazzola, M.S., and Stoudt, D.C., "Executive Summary of the BOSS Development Effort," First Quarter Research (Jul - Sep 1990), NAVSWC, Dahlgren, VA, Oct 1990.
- Merritt, D.L., *High Energy Lasers for Ship Defense*, Space and Naval Warfare Systems Command, Washington, DC.
- Moran, Stuart L., *Pulsed Power Technology Program: Accomplishments in Switching*, NSWC TR 89-259, NAVSWC, Dahlgren, VA, Sep 1989.
- O'Shea, D.C., Callen, W.R., Rhodes, W.T., *Introduction to Lasers and Their Applications*, Addison-Wesley Publishing Co., Inc., Philippines, Second Printing, Dec 1978.
- Stimson, G.W., *Introduction to Airborne Radar*, Hughes Aircraft Company, El Segundo, CA, 1983.
- Tatum, J.T., Berry, M.D., Garver, R.V., "Assessment Methodology for Predicting the Probability of System Failure Due to Radio Frequency Energy," *Fifth National Conference on High-Power Microwave Technology*, United States Military Academy, West Point, NY, 10-15 Jun 1990.

The Author



LAWRENCE H. LUESSEN received his B.A. degree in physics from Gettysburg College and began his career at NAVSWC upon graduating. Selected two years later for the Center's full-time advanced study program, he received an M.S. degree in EE from Duke University.

Mr. Luessen's early assignments included the Navy's first efforts in explosive magnetic flux compression; program manager for the Navy's Special Effects Warhead Program, an early investigation into RF/microwave burnout weapons; and program manager for the Navy's Pulsed Power Technology Program. More recently, Mr. Luessen has been Branch Head of the Directed Energy Branch, and, currently, the Pulsed Power Technology Branch. The Branch's current technology-based efforts include development of high-repetition spark gaps for both directed-energy and electronic-warfare applications; unconventional electronic countermeasures; and research and development of optically controlled photoconductive semiconductor switches. Mr. Luessen has served with numerous technical committees, working groups, and panels, including STRATPLAN 2010, Revolution-at-Sea 2020, the Aegis Technology Round Table Team, and the 1991 Pulsed Power Conference Technical Review Committee. He was also the Chairman of 1990 Power Modulator Symposium, and has directed four NATO Advanced Study Institutes in Europe and the UK, resulting in co-editorship of four technical books. Mr. Luessen is a member of the American Physical Society.

An Electromagnetic Noise Collection and Processing System, "Arctic Research Buoy"

J. F. Scarzello, D. S. Lenko, LCDR G. G. Durante

The Naval Surface Warfare Center (NAVSWC) has developed a cost effective system for continuously acquiring geomagnetic noise data in the magnetosphere's polar and auroral zones in the arctic; the name given to the system is Arctic Research Buoy (ARB). The system was designed to function unattended in remote areas for periods of up to two years. It was also designed to process the recorded data and to transmit selected segments of the data back to the continental United States via satellite telemetry links. This article discusses features of the ARB design and gives some background history on early ARB deployments.

Introduction

Electromagnetic noise associated with the aurora in the arctic has the potential for disturbing many types of electrical systems, including communications systems and several types of sensors that are important components in Navy weapon systems. For investigators to assess the impact of this noise, extensive measurements are needed on frequency of occurrence, duration, correlation over large distances, and variation in intensity. The harsh physical conditions and high cost of operating in the arctic weigh against use of manned measurement stations at remote sites during much of the year. Over the last five years, the Office of Naval Technology has sponsored NAVSWC to develop a unique, unmanned measurement and data collection system that can operate continuously for over a year in the arctic environment. The system is designed to provide a cost effective means for recording electromagnetic noise radiated by the magnetosphere's polar region and auroral zone. The system is also designed to screen the data and transmit selected segments back to the continental United States (CONUS) via satellite link. This article discusses some of the unique features of this system as well as the history behind its development and use to date.

The unique character and amplitude of electromagnetic noise in polar regions are due to the interaction of the solar wind with the earth's magnetic field, as shown in Figure 1.¹ As a consequence of this interaction, there are temporal variations in the ambient magnetic field referred to as geomagnetic noise. The magnetic north and south poles define a "skewed magnetosphere" field distribution. The ambient magnetic field amplitude varies with position on the earth and the earth's orientation to the sun, as well as being a function of solar sunspot activity. The magnetosphere has two distinct regions with respect to arctic geomagnetic noise, the polar cap which tends to be quieter and the auroral zone where disturbances tend to be greatest. It is necessary to collect geomagnetic noise for long periods of time, to characterize it, and to determine the spatial coherence between measurement sites. The system developed by NAVSWC to make these geomagnetic noise measurements is called the ARB.

126

Arctic Research Buoy System

The ARB system is designed to make both triaxial magnetic and triaxial electric field measurements and to be flexible enough to do on-site pro-

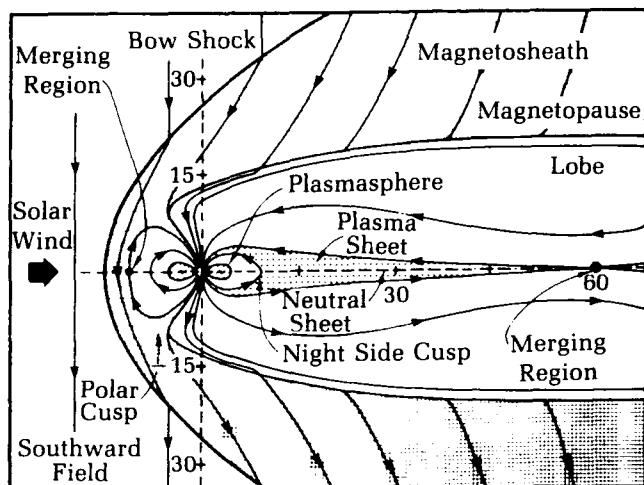


Figure 1. A schematic drawing showing most of the major regions of the magnetosphere. The drawing is roughly to scale and a southward directed interplanetary field is assumed (Reprinted with permission from McPherron, R.L., "Magnetospheric substorms," Rev. Geophys. Space Sci. 17, 1979.)

cessing and screening of the recorded data. This unattended "smart" buoy must operate in the arctic region for at least a year subject to limitations imposed by satellite data rates, available power sources, modest microcomputer processing power, and random access memory data storage size. The ARB system is designed to continuously measure and record the six electromagnetic field components, screen the recorded data for segments of scientific interest based on programmed criteria, and transmit the selected segments to stations in CONUS via satellite link. Once this transmission occurs, the stored data is erased to make room for new data in the system memory. A typical ARB system deployed near an ice

station is shown in Figure 2. The number and type of sensors deployed at any site vary depending on needs and constraints.

The term buoy was adopted because the system must be able to float in order to survive the arctic "summer" in a melt pond, and also because the ARB components look similar to spar buoys in the ocean. Physically, the ARB system is a collection of waterproof packages containing electrical components and other items such as batteries. The square box containing the zinc air batteries appears on the left in Figure 2 along with the solar panels; these are the two primary power systems for the ARB. Individual subsystems are discussed in more detail in the following sections.

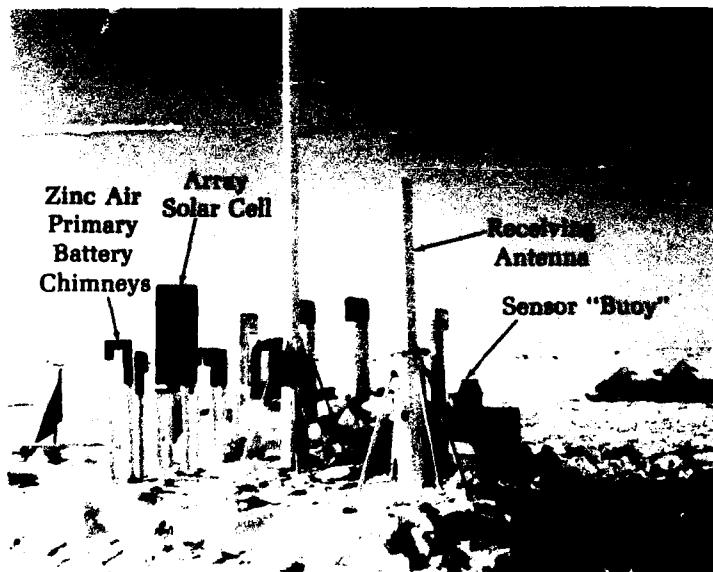


Figure 2. The Arctic Research Buoy: an electromagnetic noise collection and processing buoy system developed by NAVSWC to acquire geomagnetic noise.

Sensor Buoy

The ARB system actually consists of two basic subsystems: the "sensor buoy," which collects and processes the measured data and the "telemetry buoy," which communicates with satellites. As shown in Figure 3, the electric and magnetic field sensors, as well as other environmental sensors, are associated with the sensor buoy package, although the electric field sensors are housed in a separate unit that is cable-connected to the primary sensor buoy.

The analog voltage outputs from the sensors are multiplexed onto a single carrier, run through a 12-bit analog-to-digital (A/D) converter, and sent to a computer in the sensor buoy for storage and processing. The sensor buoy is about 24 feet in length. Figure 4 shows two buoys being uncrated and readied for deployment: the one in the foreground houses the sensors, the one in the background is the primary telemetry buoy, and the two vertical

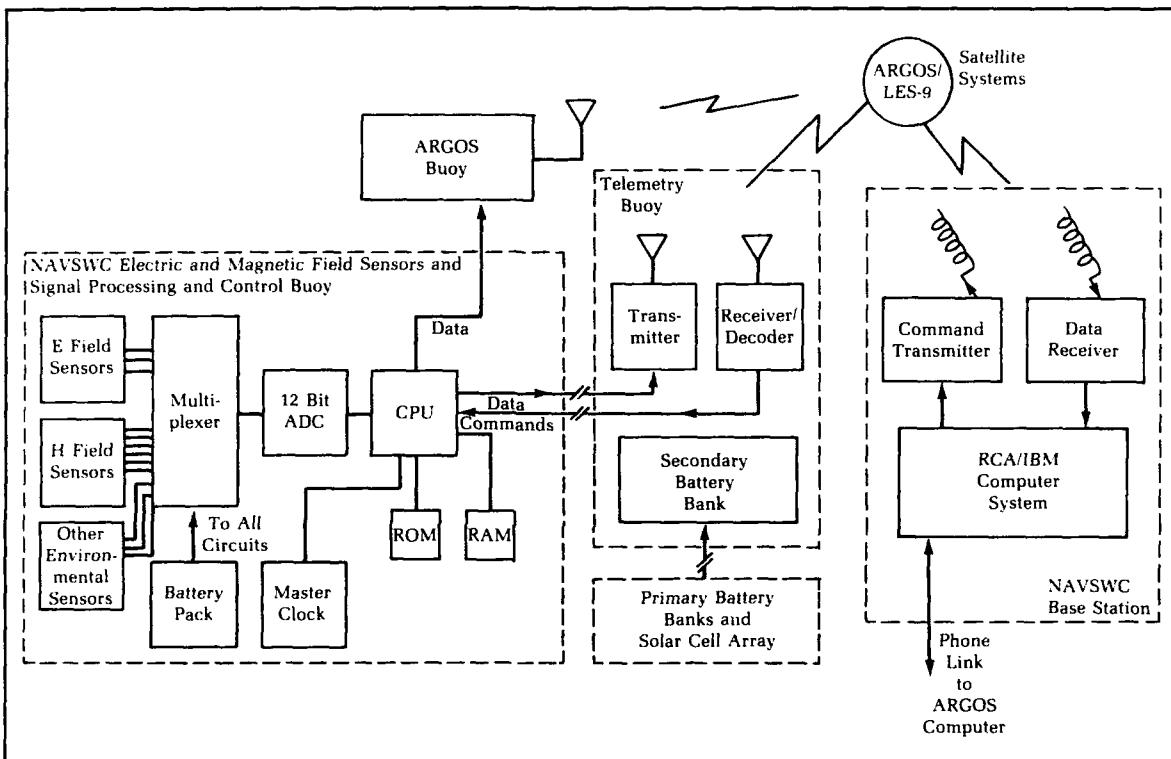


Figure 3. Arctic research buoy system block diagram.

128

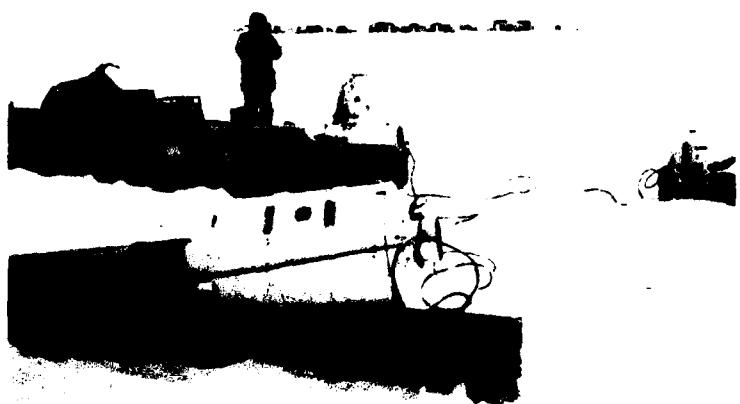


Figure 4. Sensor and telemetry buoys are readied for deployment.

units are the receive (to the left) and transmit antennas. The electric field sensor is placed in an 8-inch hole drilled through the ice so that the electric field sensors can be in the water below the ice canopy. The magnetometers and associated electronics are in the sensor buoy frozen in the ice. The electric and magnetic field sensors are designed with "front-end" calibration systems that can be activated by command. When activated, electric and magnetic fields of known strength are generated; the corresponding voltages are sent through the total system to the computer.

The magnetic sensor used in the system is a triaxial Brown magnetometer which is a very low power (it consumes about 1.5 milliwatts) ring-core fluxgate magnetic sensor with a low noise level; the magnetic sensor package is shown outside its package in Figure 5. In more recent versions of the ARB, the Brown magnetometer is being replaced with a more sensitive Helium-3 total field nuclear magnetic resonance magnetometer.

The electric field sensors are made from silver/silver chloride electrodes mounted in pairs on the ends of 2-meter-long fiberglass tubes. Before deployment, these tubes are all aligned parallel to the main (vertical) axis of the unit buoy so that it can fit through an 8-inch-diameter hole in the ice. After the sensor is properly seated in the hole, two of the arms are allowed to rotate into position to form the horizontal axes. The electric field sensors are isolated from system ground and their outputs are fed into low power amplifiers. The center of the electric field measurement system is in the water about 2 meters below the bottom of the ice. Other sensed information includes tilt, temperature, absolute magnetic field com-

ponents, and system status voltages. For some recent shore-based deployments, long baseline magnetic sensors are located external to an unmanned, computer controlled "sensor" unit.

The computer module, located in the sensor buoy, consists of a low power RCA CMOS 1805 microcomputer, a master absolute time clock, A/D converter, multiplexer, sensor interface board, data memory, and system control software stored in read-only-memory. The USAF Lincoln Experimental Satellite 9 (LES-9) provides a 300-baud command link to the buoys as well a return 2400-baud data link. To take full advantage of the LES-9 link, a number of control options were incorporated including master time clock correction, data selection criteria, timing of data segments to allow correlation with ARBs in other locations, calibration commands, and command-selected data transfer times in an assigned transmission window.

Telemetry Buoy

The telemetry buoy section is the ARB station's link with the outside world. Components of this buoy system are: (1) a receiver and transmitter, (2) an omnidirectional transmitting antenna and an omnidirectional receiving antenna with a preamplifier, (3) secondary gelled electrolyte lead-acid batteries and power regulation circuitry, (4) primary zinc air batteries housed in sealed containers with air-breathing chimneys capped with snow filters that provide about 50 volts with a 2,400-ampere-hour capacity for the satellite telemetry link, and (5) a solar cell array to supplement the primary batteries. Both the sensor buoy and the telemetry buoy systems are designed to



Figure 5. Magnetic sensor package and housing.

have a two-year lifetime. When the system is deployed in the arctic, a plywood ablation shield is placed around all primary ARB components.

Satellite Links

Until recently, the primary telemetry link has been with the USAF's LES-9 satellite. NAVSWC had been assigned a one-hour daily access period to transmit data and commands. The LES-9 was in a geosynchronous circular orbit, inclined 25 degrees to the equator at 106 degrees west longitude. The ARB uses a data link of 2,400 baud, which is carrier-modulated by biphase Differential Phase Shift Keying. The ARB transmits data back after it receives and decodes commands sent at 300 baud with an asynchronous format. The data transmitted typically consists of three "snapshots" of electric and/or magnetic field noise segments which were selected either by time or by amplitude characteristics after screening by the sensor buoy computer. Also transmitted are: (1) relative difference information for each sensor for each 2.5 minute period during the preceding 23 hours, (2), the ambient temperature, and (3) the system status voltages. The LES-9 telemetry link was lost last December when the satellite was moved to support operation Desert Storm.

The ARB sensor buoy is also connected by cable to a nearby ARGOS buoy which provides a communication link with the polar orbiting Naval Oceanographic and Atmospheric Administration (NOAA) ARGOS satellites. The ARGOS buoy data link has very limited capacity, but it does provide information on system location and can relay up to 256 bits of information on vital ARB system status parameters and max/min ARB sensor data. This backup data link proved invaluable when the LES-9 link was no longer available, and important position information transmitted via ARGOS made possible the recovery of two drifting ice stations that otherwise might have been lost.

ARB Base Station

The ARB base station is located at NAVSWC in White Oak, Maryland. It consists of (1) two-directional helical satellite antennas used for receiving and transmitting; (2) a 70-watt UHF transmitter; (3) tunable receiver; and (4) an RCA microcomputer development system that serves as the base station computer with an IBM AT-compatible PC programmed for unattended base station operations. The PC also provides the telephone modem for ARGOS data retrieval and remote emergency satellite access control. Transmitted data segments are

130

"deglitched" of telemetry link dropouts, screened to select the most interesting data, and formatted on 5-1/4-inch floppy disks. This data is provided to researchers to determine spatial coherence and other characteristics of arctic region geomagnetic noise.

Deployments

The ARB system buoy was first deployed using a hard-wired telemetry link during a spring 1984 joint Canadian (Defence Research Establishment Pacific-DREP)/U.S. (NAVSWC) expedition near Resolute, NWT, Canada. The camp was located about 125 miles from the North Magnetic Pole, in the magnetosphere's polar cap region. Several sensor deployment schemes were tried, and the lessons learned resulted in a redesign of the electric field sensor electronics. Much of the data collected in that test period was during a very large geomagnetic storm. The second ARB deployment using the LES-9 satellite data link occurred at the Applied Physics Laboratory, University of Washington's (APL/UW's) Ice Station (APLIS) during ICEX 1-86. Due to equipment failures, telemetry link refurbishment and some ARB system redesigns were again required.

In the spring of 1987, NAVSWC deployed two ARB systems with DREP logistics support, one at an ice station northwest of Alert, and the other on shore about 1000 feet from the Lincoln Sea shoreline. Both systems performed well until unrelated failures occurred about nine months after deployment; the ice station's LES-9 transmitter failed, and the shore station's computer module power regulator transistor shorted. Fortunately, the ice station's backup ARGOS satellite link continued to operate, sending back sensor buoy information and position information; the ARB was recovered by DREP and Canadian Forces helicopter personnel in May 1988. The ARB Alert shore site was repaired and its software upgraded in January 1988 so that modifications could be tested before ICEX 1-88.

Two ARB installations were completed during April 1988 as part of ICEX 1-88. An ARB ice station, similar to the 1987 Alert ARB, was deployed about 2000 feet from the main APLIS camp. At the USAF's Distant Early-Warning line site on Barter Island, Alaska, an ARB shore site was established about 200 feet from the Beaufort Sea shoreline. All four ARB's were operational from late March 1988 to May 1988, when the ARB Alert ice station system was recovered by DREP/CF personnel. DREP, located in Victoria, B.C., Canada, has an LES-9 satellite base station receiver, and acquires

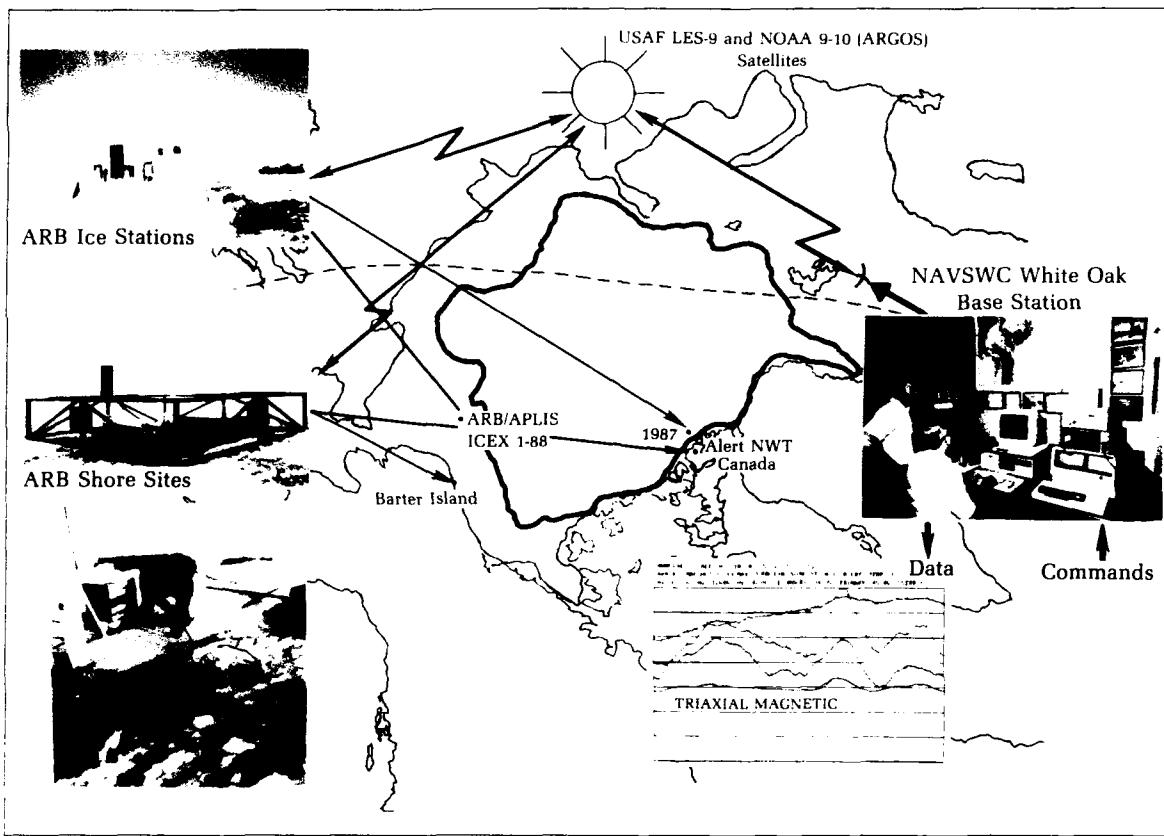


Figure 6. Overview of 1988 ARB network.

the ARB data under the provisions of an international exchange agreement. Figure 6 gives an overview of the ARB network during this time.

The ARB APLIS ice station performed well for four months before abruptly failing. In September 1988, NAVSWC and APL/UW with U.S. Coast Guard (USS Polar Star) support, mounted a repair/recovery attempt which failed due to bad weather (ice, fog) and helicopter navigational errors. The ARB APLIS ice station was recovered on 31 March 1989 as part of the Naval Oceanographic Office White Trident 89 operations. To facilitate the recovery, NAVSWC personnel hurriedly modified 12 AN/SSQ-41B sonobuoys to use as beacons; they were deployed near the ice station by P-3C surveillance aircraft from Oceanographic Development Squadron 8, based at the Naval Air Test Center, Patuxent River, Maryland. The sonobuoys were modified to include an alkaline "D" cell battery pack instead of a sea water battery, and were fitted with a much larger,

bright orange parachute. The acoustic sensor and scuttle timer circuitry were removed, and a programmable timer board was installed so the buoy transmitter would provide periodic RF signals for one week at -40° C. VXN-8, using onboard GPS navigation and NAVSWC's ARGOS-supplied ARB position, located the ARB and dropped a beacon buoy. NAVSWC's recovery team, on an Alaska Air National Guard Black Hawk helicopter, used DF equipment and sighted the beacon buoy about 150 feet from the ARB. The ARB had been attacked by a polar bear and suffered from surface ice ablation. Figure 7 shows that ARB station being dismantled; all high-value equipment was recovered in less than 40 minutes.

Currently, the ARB shore sites at Alert and Barter Island continue to operate, but the telemetry link was lost when the LES-9 satellite was put into a different orbit to support Desert Storm. As part of the remote environmental sensors program in ICEX 91, a third ARB station was recently installed at Pt. Barrow,



Figure 7. The ARB APLIS ice station was recovered as part of NAVOCEANO's White Trident 1989 operations.

Alaska. The stations are configured so that with some minor support from personnel at the sites, the data collection effort can continue. This work is critical because the solar sunspot activity has just passed the maximum in its 11-year cycle and the geomagnetic noise maximum is predicted to peak in the near future.

Data Analysis

In addition to characterizing the geomagnetic noise from measurements at a point, it is also of interest to analyze the spatial coherence between ARB sites. We were fortunate in having four ARB stations functioning at the same time for about a two-month period during 1988; this provided an excellent measure of global, simul-

taneous geomagnetic noise activity. Baseline distances among ARBs varied from 100 to 350 miles between ice stations and corresponding shore sites in both the polar and auroral zones, and 1400 miles between polar and auroral zone stations. Simultaneous noise segment correlations were examined. If a single magnetometer component is selected from an ice station buoy, and a parallel magnetic component from a shore-based triaxial magnetometer, then coherent processing can be used to reduce the noise level. From data processed this way, temporal geomagnetic noise reductions of 20 dB have been demonstrated.

ARB data is being used to help optimize algorithms for various sensor devices currently in use by the fleet, and data is also being stored in a library for future use in designing and evaluating sensor performance.

Future Work

The ARB has been a cost-effective tool for acquiring data from inaccessible regions. Although the current ARB system uses magnetic and electric field sensors, numerous other sensors could be used. Base station and ARB telemetry link upgrades must be made to increase system reliability. The computer module should be upgraded to take advantage of recently developed, powerful, low-power microprocessors and increased memory chip densities. A test to evaluate zinc-air primary battery performance should be refined and battery housings modified to ensure an unobstructed air supply. To fully extract energy from primary batteries, wide-input voltage range DC-to-DC converters must be used with redundant, fail-safe circuitry. Connectors must be rugged, military-type MS connectors protected from the elements by covers or plastic bags. Interconnecting cables with rubber insulating jackets have performed well, but have suffered damage from animals. Applying a coating of quinine or other obnoxious substance might discourage attacks by polar bears and arctic foxes. The hastily modified sonobuoy beacons should be upgraded to include numerous desirable options. Any new ARB ice station should be equipped with an independent, RF-command-actuated VHF or UHF link to facilitate direction finding for repair or recovery.

Acknowledgments

The authors wish to acknowledge the support of the Office of Naval Technology, Dr. A. J. Faulstich and Captain R. Fitch; the early support of the Office of Naval Research, J. Heacock of Code 1125GG; the superb ICEX operations group from the Arctic Submarine Laboratory, Naval Ocean Systems Center, led by Captain M. Dorman; F. Karig, Dr. R. Francois and staff, Applied Physics Laboratory, University of Washington. We appreciate the invaluable cooperation and expertise of Dr. P. Holtham and staff, Canadian

Defense Research Establishment Pacific and Canadian Forces Station Alert; USAF BARMAIN/4700, USAF LES-9/MIT Lincoln Labs; Naval Oceanographic Office, CDR. R. Pentimonti; and VZN-8, LCDR Gore. We acknowledge with great affection the NAVSWC White Oak staff who made this happen, in particular Ronald G. Vermillion, Wayne R. Grine, and Dr. Robert E. Brown.

Reference

1. McPherron, R. L., "Magnetospheric Storms," *Rev. Geophys. Space Sci.*, 17, 1979, pp. 657-681.

The Authors



JOHN F. SCARZELLO received B.S. and M.S. degrees in electrical engineering from New York University. He has worked on resonance magnetometers and geomagnetic noise. At NAVSWC since 1968, he has performed research and development of magnetic materials for NASA satellite magnetometers as well as sensors for Navy tactical intrusion devices (Vietnam), sea mines, vehicle detectors,

special warfare weapons, and arctic region research. He is a member of the Magnetics Society, AGU, and AAAS. Currently, he is conducting R&D on electromagnetic sensors for sea mines and tactical and strategic surveillance systems.



DANIEL S. LENKO received a B.S. degree in electrical engineering, Drexel University, 1975; M.S. in computer science, 1979, and M.S. in electrical engineering, 1982, Johns Hopkins University. At NAVSWC since 1975, he developed a digital compass for installation on the NAVSWC Fort Lauderdale TONGS unit, designed electronics for special operations hardware, and developed time-delay firing devices for the SEAL weapons program. Since 1983, he has been chief arctic research buoy engineer, participating in numerous expeditions for installation, upgrade, and maintenance. He holds seven patents and received a Meritorious Unit Commendation in 1989.



LCDR GARY G. DURANTE, USN, graduated with a B.S. degree in biology from The Citadel in 1980 and received an M.S. in chemistry (chemical engineering) from Auburn University in 1987. From December 1980 to July 1983 he was stationed on USS O'Bannon, and from August 1983 to August 1985 he was assistant course coordinator and instructor at the DD-963 school in Newport, R. I.

LCDR Durante came to NAVSWC, White Oak in Dec 1987 and assumed the duties of Military Assistant for Research and Technology. He also is chairman of the Research Department's Explosives Certification Program and was Deputy Principal Investigator for ICEX-91. In May 91 he becomes the Executive Officer at NAVELEX in Charleston, S.C. He has received the Navy Commendation Medal and the Navy Achievement Medal.

On the Validity of Normal Mode Theory for Modeling Undersea Acoustic Propagation

Juan I. Arvelo, Jr.

This article deals with modeling of acoustic transmission loss in shallow-water environments. The ability to detect targets in shallow coastal waters is receiving increasing emphasis in the light of currently projected low-intensity-conflict scenarios. The most accurate prediction of underwater sound transmission is achieved with the normal-mode theory for a water column modeled as an acoustic waveguide. A Naval Surface Warfare Center (NAVSWC) normal-mode propagation model has been developed to account for sound-speed variation, sediment and subbottom elasticity, surface and bottom roughness, and range dependence of the ocean waveguide. Results of the model are compared with measurements from a laboratory experiment. Good agreement between the predictions and measurements was found to validate the accuracy of the normal-mode model. This model will be used to determine the optimum frequency of operation for an active sonar system in selected environments.

Introduction

Partly as a result of political changes taking place in the world, the emphasis in antisubmarine warfare (ASW) is shifting from deep to shallow water. Shallow-water ASW is difficult both in practice and in theory because of the sound's multiple interactions with the ocean floor and the high variability of the water's sound speed profile. Central to ASW modeling and prediction is the need to have an accurate method for computing propagation loss. The standard ray-trace procedures that work well in deep water are known to break down in shallow water, and surface and bottom interactions become extremely important. The normal-mode approach has been used for many years to model sound propagation in shallow water, but it has become evident that the bottom's elasticity must also be included.

Several other investigators have developed sound propagation models that include bottom elasticity. Schmidt,¹ for example, included shear waves in a Fast Field model that applies only to range-independent environments. Ellis and Chapman² included the elasticity of the bottom in the simple Pekeris waveguide model, resulting in a reasonable agreement with experimental measurements, but they recognized the need for a multilayered model. Collins³ extended the Parabolic Equation model to accommodate bottom elasticity in a range-dependent waveguide, but this model incorporates several assumptions and approximations that must be validated with experimental observations and benchmark calculations. With the development of parallel processing and supercomputers, the propagation approximations in the parabolic equation, finite element, fast field, and ray theory models will become obsolete because exact models, such as the normal-mode approach, can be used for higher frequencies and deeper oceans.

The normal-mode theory gives the exact solution to the problem of sound propagation in a range-independent ocean waveguide. However, the bottom of the ocean is an elastic medium where both compressional and shear waves are created by the compressional waves from the source in the water column.

In the NAVSWC normal-mode model, sound penetration into the bottom is accounted for by including the modes that radiate into the ocean floor. This set of radiating modes has an inherently discrete wave-number spectrum; the complex eigenvalues are determined by a search in the complex k -plane. A semi-infinite elastic basement layer is included in the model, representing the last layer reached by the decaying eigenfunctions. The range dependence of the ocean's acoustic properties and its boundaries are taken into account by dividing this waveguide into range-independent segments. The eigenvalues and eigenfunctions are determined for each segment, and a modified adiabatic method⁴ is used to evaluate the sound transmission loss. Recently, comparisons have been made of the transmission loss with and without the effects of bottom elasticity.⁴ The differences are considerable, and it was concluded that bottom elasticity must be accounted for, particularly when the signal is to be coherently detected, e.g., replica correlation.

In this article, results from the normal-mode sound transmission loss model are compared with experimental measurements in a model tank.⁵ This comparison against measurements taken in a controlled environment serves as convincing evidence that the normal-mode model, including bottom elasticity, is a viable approach in modeling propagation in shallow water.

Theory

Performance evaluation of an active or passive sonar system requires the accurate prediction of the propagation behavior of the signal in the desired ocean environment. In coastal waters, the ocean floor has a slope that complicates transmission loss computation. Since comparisons will be made with laboratory measurements in a flat-bottom environment, only range-independent computations are displayed in this article. Laboratory measurements with a sloping bottom are underway, and comparisons against the range-dependent model will be covered in a future article. Arvelo and Uberall describe the range-dependent normal-mode model.⁴

Transmission loss could be coherently or incoherently computed. The coherent transmission loss as a function of the receiver's horizontal range and depth is given by,

$$TL_c(r,z) = 20 \log_{10} [\pi \rho(z) + \sum_{n=1}^N |u_n(z_0) u_n(z) H_o^{(1)}(k_n r)|] \quad (1)$$

where $\rho(z)$ is the water density at the receiver's depth, $u_n(z_0)$ are the N complex eigenfunctions

at the source's depth, $u_n(z)$ are the N complex eigenfunctions at the receiver's depth, r is the horizontal range between the source and the receiver, $H_o^{(1)}(k_n r)$ is the zero order complex Hankel function of the first kind, and k_n are the N complex eigenvalues representing the water column. The coherent transmission loss accounts for the constructive and destructive interference of the direct-path signal with the multipath arrivals in the water column.

The incoherent transmission loss is given by

$$TL_i(r,z) = 10 \log_{10} [\pi^2 \rho^2(z) \sum_{n=1}^N |u_n(z_0) u_n(z) H_o^{(1)}(k_n r)|^2], \quad (2)$$

where the phase of the signal is ignored to provide a smoother transmission loss curve versus range and depth. The eigenvalues and eigenfunctions satisfy the differential equation,

$$\left[\frac{d^2}{dz^2} + k^2(z) \right] u_n(z) = k_n^2 u_n(z), \quad (3)$$

where $n = 1, 2, \dots, N$ is the mode index, and $k(z) = \omega / c(z)$ is the depth-dependent wave number.

The range-independent water column (Figure 1) is modeled as J-1 water layers with a wave number squared linear in depth in order to obtain Airy functions⁶ as the solutions of Equation (3). If a layer has a constant sound speed, then the wave number is constant and the solutions to Equation (3) are sines and cosines. The elastic layers have constant acoustic properties and the last layer is semi-infinite. The complex eigenvalues are determined by using the Levenberg-Marquardt minimization algorithm⁷ and an up-layer matching technique⁸ within the appropriate boundary conditions at the interface between layers. After the eigenvalues are determined, their respective eigenfunctions are found using the down-layer matching and propagating algorithm.⁹

Results

A three-layer scaled model tank was constructed by Glegg⁵ simulating a shallow-water column with a surficial sediment overlaying a harder substrate rock. An epoxy resin was chosen to simulate the bottom sediment because of its acoustic properties and the ability to reduce air entrapment and inhomogeneities. The substrate was made out of concrete for its acoustic properties, which are similar to those of limestone. Measurements were taken of some acoustic properties of the epoxy resin.

The density of the epoxy is 1.03 gm/cc, the compressional sound speed is 2450 m/s, and

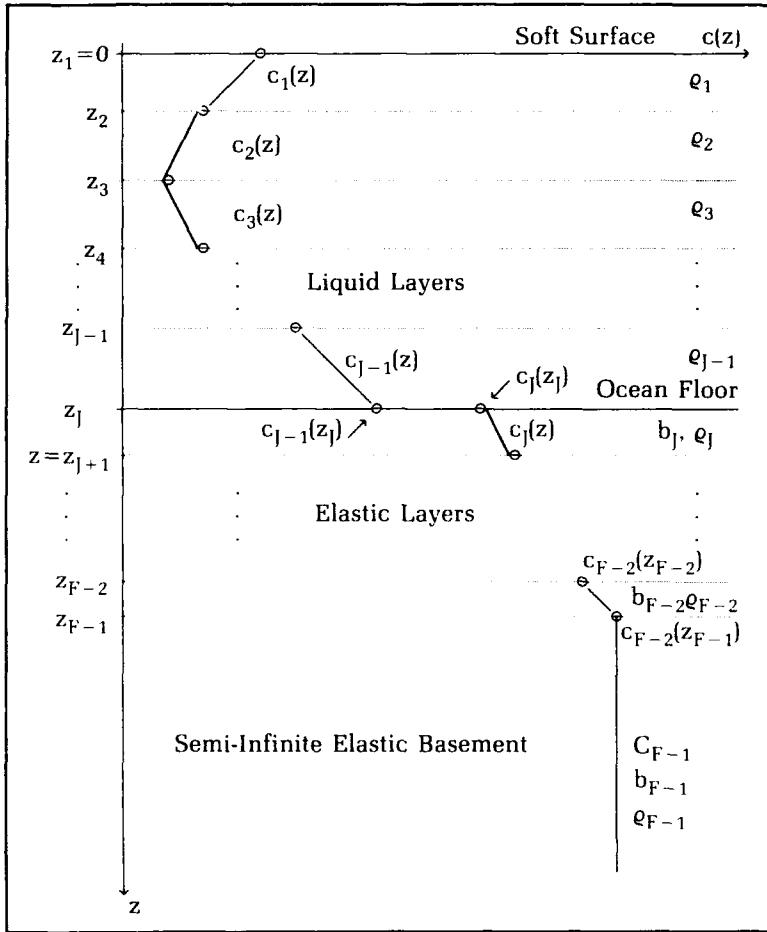


Figure 1. The range-independent ocean wave-guide is divided into horizontal layers representing the water column and the elastic bottom sediments.

the shear sound speed is 1120 m/s. The compressional and shear attenuation coefficients were not measured; therefore, they must be obtained by other means. Hartmann¹⁰ used 2-MHz ultrasonic pulses to measure the sound speeds and attenuation coefficients of cross-linked epoxies at a temperature of 20°C. A significant correlation between the absorption and the sound speeds has been found, hence making it possible to estimate the attenuation coefficients based on their respective measured sound speeds. Figure 2 displays the correlation between the compressional properties of eight mixtures of epoxy. The average error of the linear fit relative to the measured properties is less than one decibel. With a measured compressional sound speed of 2450 m/s, the estimated absorption is about 10.0 ± 1.0 dB/cm at 2 MHz, which is scaled to an attenuation of 0.5 ± 0.05 dB/KHz-m (1.2 ± 0.12 dB/ λ). Figure 3 shows the relationship of the shear properties among the eight mixtures of epoxy. The average error of the linear fit relative to the measured properties is about five decibels;

136

the correlation of the shear properties is not as good as that of the compressional properties because shear properties are more difficult to measure. If the measured shear sound speed is 1120 m/s, then the estimated absorption is about 40.0 ± 5.0 dB/cm at 2 MHz, or 2.0 ± 0.25 dB/KHz-m (2.2 ± 0.28 dB/ λ). Even though the estimated attenuation coefficients of the epoxy involve some uncertainty, a quick sensitivity analysis showed that a 10 percent variation in either attenuation coefficient contributes to less than a one-decibel variation in the estimated transmission loss.

Because the acoustic properties of concrete could not be measured with great accuracy, they were estimated based on the average properties of limestone. Its density is 2.3 gm/cc, its compressional sound speed is 3300 m/s, its compressional absorption is 0.02 dB/KHz-m, its shear sound speed is 2100 m/s, and its shear absorption is 2 dB/KHz-m.

The water column has an average sound speed of 1510 m/s and a density of 1 gm/cc. This water column is 15 cm thick, and

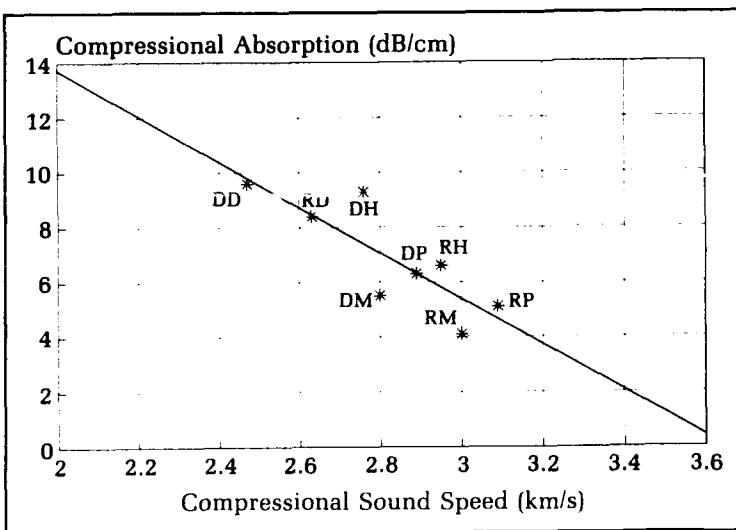


Figure 2. Relationship between the compressional properties of eight cross-linked epoxies at 2 MHz.

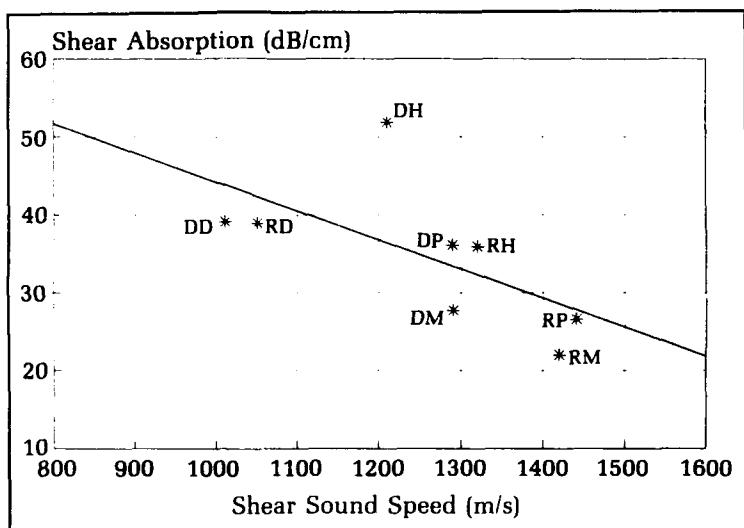


Figure 3. Relationship of shear properties among the eight cross-linked epoxies at 2 MHz.

the epoxy layer is 5.8 cm thick, as shown in Figure 4. Measurements were made with short pulses to ensure that reflections from the walls were negligible. The source is located at 5 cm deep, and measurements were taken at ranges from 1.0 to 2.2 meters in increments of 0.2 meter. More detailed range measurements could not be taken due to physical limitations of the model tank.

Figure 5 displays the calculated coherent transmission loss curve against the measurements at 30 KHz. Note that the measurements follow the same shape of the calculated curve, and the average error of the estimated curve relative to the measurements is 2.0 dB. The 25-KHz transmission loss curve is given in Figure 6. The transmission loss is slightly overestimated by the normal-mode

model. This could be corrected by decreasing the attenuation coefficients of the epoxy resin in the input parameters of the computer model, but the average error of the estimated curve relative to the measurements is 1.7 dB, and errors less than 3 dB are bearable. Figure 7 compares the transmission loss at 20 KHz. Again, the measured values follow the calculated curve very closely, even though some overestimation is apparent. The average error of this estimated curve relative to the measurements is 1.3 dB. It has been observed that the effect of the concrete substrate is negligible in the calculated transmission loss at frequencies of 20, 25, and 30 KHz. Transmission loss calculations with and without the substrate at 15 KHz show that the substrate does affect the sound propagation at this

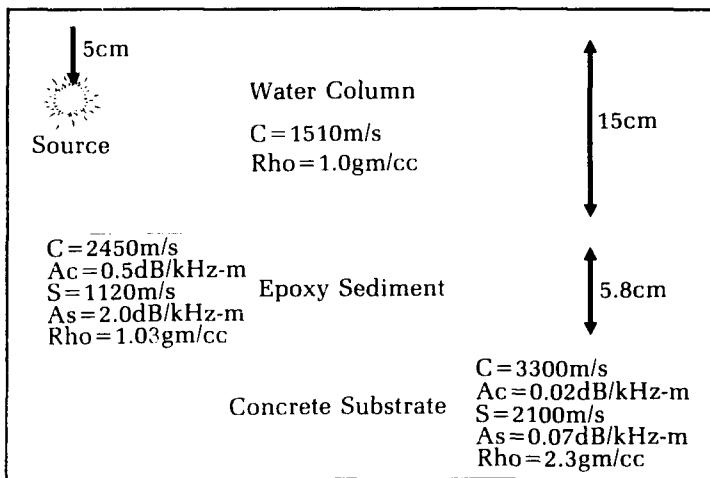


Figure 4. Unscaled model of the tank with the estimated acoustic properties.

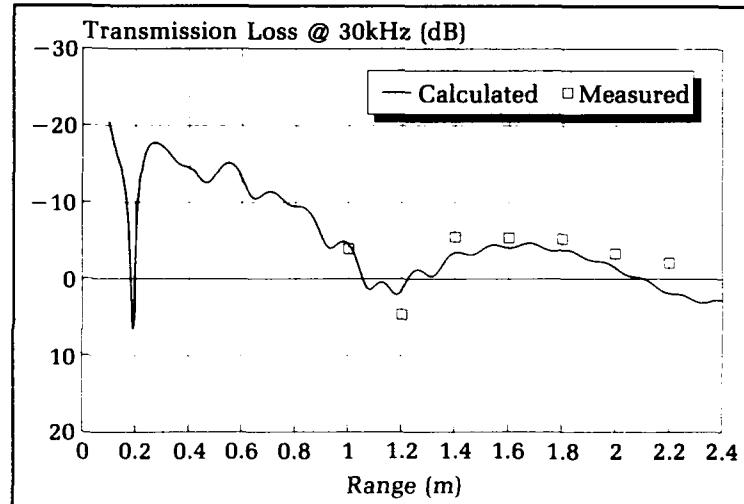


Figure 5. Comparison of transmission loss for a frequency of 30 kHz.

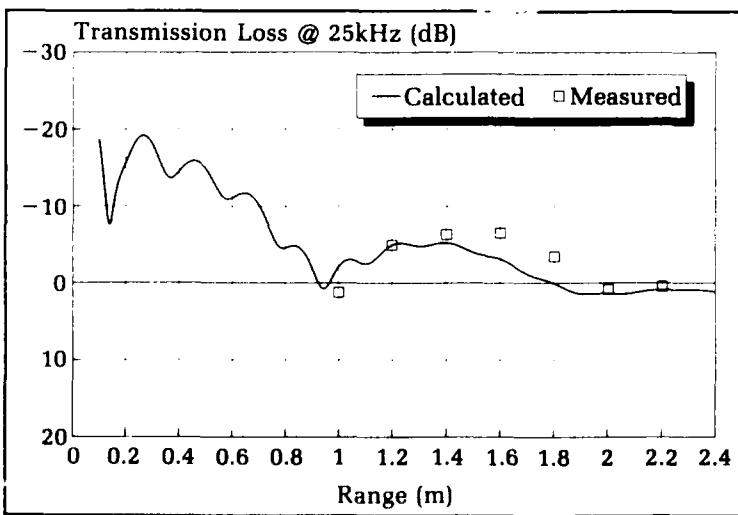


Figure 6. Comparison of transmission loss for a frequency of 25 kHz.

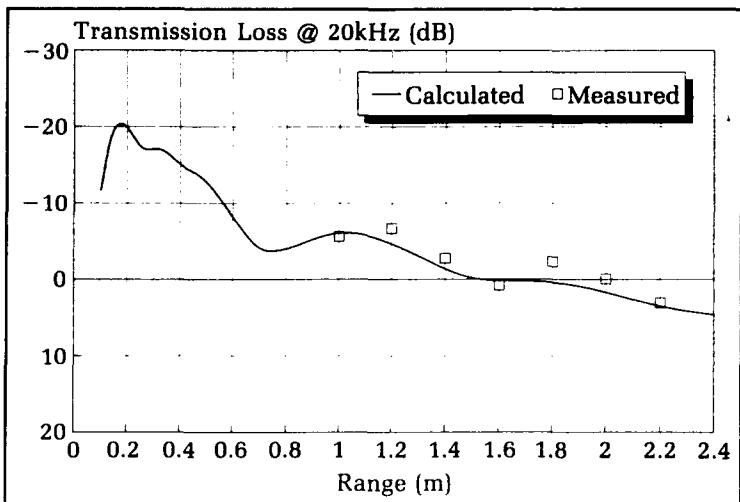


Figure 7. Comparison of transmission loss for a frequency of 20 kHz.

frequency. Figure 8 displays the calculated transmission loss with the effect of the concrete substrate. The normal-mode model still overestimates the transmission loss and its error increases with range. Two causes of this error are possible. The first possibility is that the estimated acoustic properties of the concrete substrate are completely erroneous. What makes this possible is that the concrete does not affect the sound propagation at higher frequencies. The second possibility is that the normal-mode model does not account for the liquid-solid surface wave, and its effect becomes considerable at frequencies near the cut-off frequency.¹¹ An effort to incorporate the surface wave is currently under way.

The total number of normal modes has been found to be given by

$$N \cong \text{INT}(2 f D / C), \quad (4)$$

where f is the frequency in Hertz, D is the depth of the water column, and C is the

minimum sound speed in the water column. This rule-of-thumb equation includes trapped and radiating modes; it neglects the bottom composition, and its accuracy increases with an increasing number of normal modes.

As the number of modes increases with increasing frequency or bottom depth, the required computer time and memory also increase. However, it is not necessary to account for all the modes to arrive at an accurate transmission loss prediction. Figure 9 displays the 238 normalized eigenvalues on the complex plane for a frequency of 1.2 MHz in the same waveguide as shown in Figure 4. The fundamental mode is labeled #1, and the minima at the mode labeled #184 represents the threshold between the trapped and radiating modes. Modes with a larger imaginary component in their respective eigenvalues contribute less to the total transmission loss because their energy is absorbed faster as they propagate in range. Figure 10 shows the convergence of the

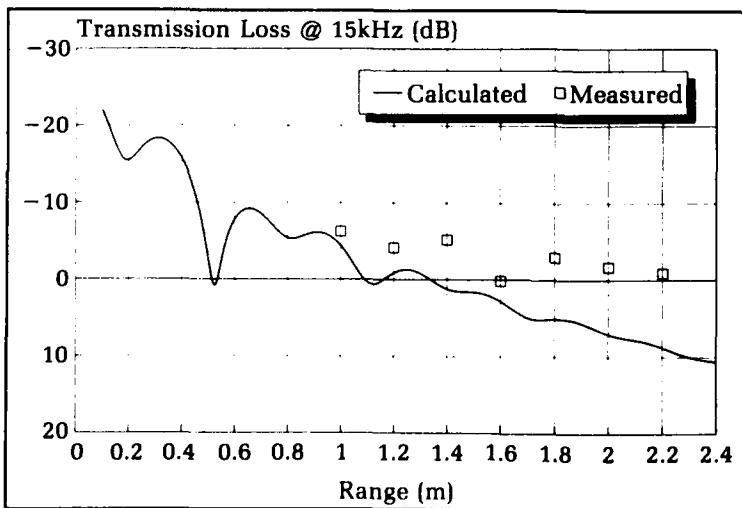


Figure 8. Comparison of transmission loss for a frequency of 15 kHz.

transmission loss as the contribution of each normal mode is added to the calculation. The receiver is 10 cm deep at a horizontal range of one meter from the source. An error smaller than one decibel is anticipated if the first 85 modes are used. This corresponds to neglecting all eigenvalues with an imaginary part greater

are encountered in this one-way transmission loss plot. By the same token, the transmission loss of broadband signals or transients can be predicted by computing the transmission loss for each frequency component of the signal.

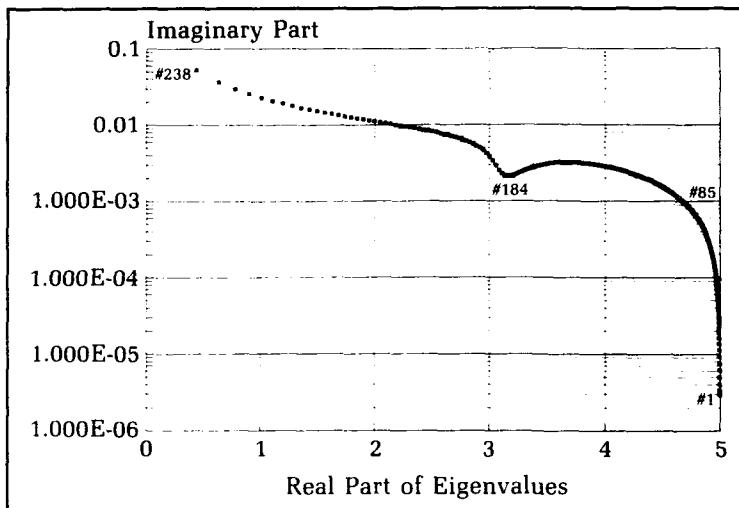


Figure 9. Eigenvalues in the complex K-plane for a frequency of 1.2 MHz and a horizontal range of one meter.

than 0.001 (Figure 11). Using just 85 out of the total 238 modes represents a 64 percent reduction of computer time. Hence, accurate transmission loss at higher frequencies can be obtained with the normal-mode model.

Determining the optimum frequency for the active detection of a target in a known environment is a common problem. Figure 11 is the calculated one-way transmission loss as a function of frequency for the same waveguide in Figure 4. Pulses with a center frequency at about 53 KHz will propagate further than other nearby frequencies. Differences of up to 15 dB

Conclusions

Measurements from a model tank experiment have been used to validate the accuracy of the NAVSWC normal-mode propagation model in a range-independent environment with an elastic bottom. Very good agreement was encountered at frequencies much greater than the cut-off frequency of the waveguide. Some disagreement was found at frequencies near the cut-off frequency; it is suspected that the interface wave plays a significant role in the propagation of VLF sound, and it must be included in the computations.

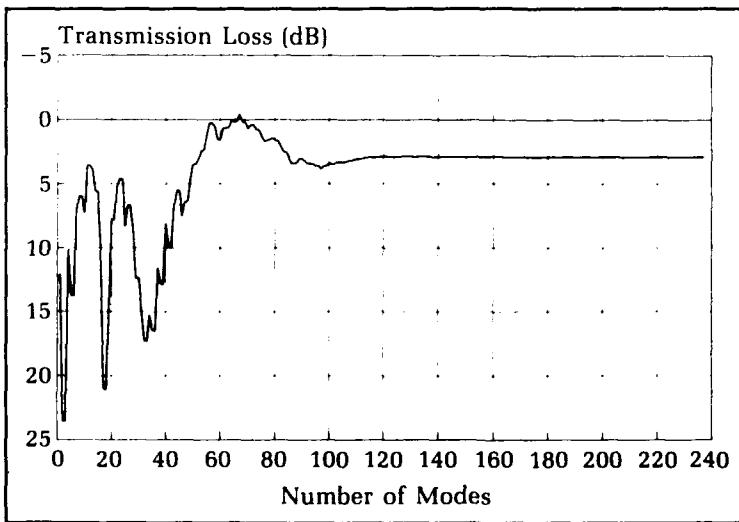


Figure 10. Convergence of the transmission loss for a frequency of 1.2 MHz and a horizontal range of one meter.

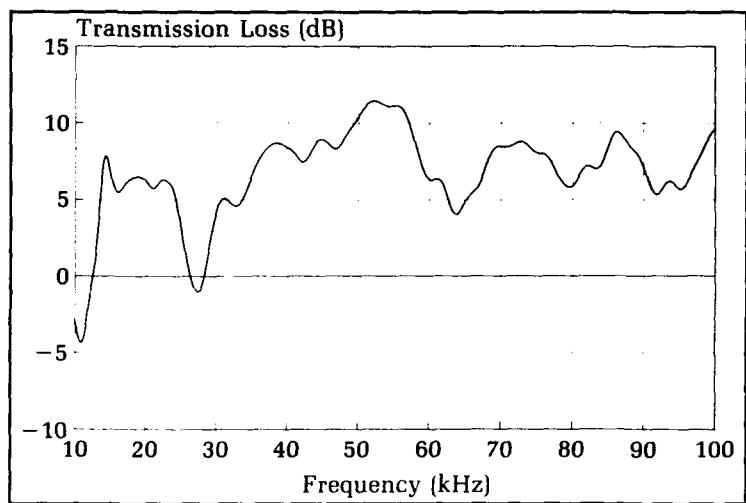


Figure 11. The transmission loss versus frequency for a source depth of 5 cm, a receiver depth of 10 cm, and horizontal range of one meter.

It was also shown that in high frequency or deep water cases, where the number of modes increases linearly, it is not necessary to use all the normal modes due to the convergence nature of the transmission loss equations.

Measurements in a model tank with a sloping elastic bottom are under way. Comparisons against predictions from the range-dependent normal mode model will be published in a future article.

Acknowledgments

The author wishes to thank Dr. Stewart Glegg for providing the transmission loss measurements. This work has been partially funded by the independent research office of NAVSWC and by the Naval Air Systems Command.

References

- Schmidt, H., *SAFARI User's Guide*, SACLANTCEN SR-113, SACLANT Undersea Research Centre, La Spezia, Italy, 1988.
- Ellis, D. and Chapman, D., "A Simple Shallow Water Propagation Model Including Shear Wave Effects," *Acoustical Society of America Journal*, Vol. 78, No. 6, 1985, p. 2087.
- Collins, M., "A Higher-Order Parabolic Equation for Wave Propagation in an Ocean Overlying an Elastic Bottom," *Acoustical Society of America Journal*, Vol. 86, No. 4, 1989, p. 1459.
- Arvelo, J. I. and Uberall, H., "Adiabatic Normal-Mode Theory of Sound Propagation Including Shear Waves in a Range-Dependent Ocean Floor," *Acoustical Society of America Journal*, Vol. 88, No. 5, Nov 1990, p. 2316.
- Hundley, A., et al., "Numerical Modelling and Laboratory Experiments on Underwater Sound Propagation over a Shear Supporting Bottom," *Shear Waves in Marine Sediments*, Kluwer academic publishers, La Spezia, Italy, 1990.
- Schlutzen, Z., Anderson, D., and Gordon, R., "An Algorithm for the Evaluation of the Complex Airy Functions," *Computational Physics Journal*, Vol. 31, No. 60, 1979.
- Morris, A. H., *NSWC Library of Mathematics Subroutines*, NSWC TR 90-21, NAVSWC, Dahlgren, VA, Jan 1990, p. 353.
- Arvelo, J. I., Talmant, M., and Uberall, H., "A Normal-Mode Model of Underwater Propagation including Shear Effects in a Layered Ocean Floor," *Computational Acoustics (Algorithms and Applications)*, Second IMACS Symposium on Computational Acoustics, Princeton, NJ, Mar 1989.
- Arvelo, J. I., *Adiabatic Normal-Mode Theory of Low-Frequency Sound Transmission in a Shallow Range-Dependent Ocean with Seismo-Acoustic Effects from the Elastic Bottom Sediments*, Ph.D. Dissertation, Department of Physics, The Catholic University of America, Washington, DC, 1989.
- Hartmann, B. and Lee, G., "Additive Moduli for Crosslinked Epoxies," *Polymer Science Journal: Polymer Physics Edition*, Vol. 20, 1982, p. 1269.
- Uberall, H., "Surface Waves in Acoustics," *Physical Acoustics*, Vol. 10, Academic Press, Inc., New York, NY, 1973, p. 1.

The Author



JUAN I. ARVELO is a research physicist in the Acoustic Signal Processing Branch of NAVSWC, White Oak. He was born in Brooklyn, NY, in 1959. He received a B.S. degree in physics (magna cum laude) from the University of Puerto Rico, Mayaguez, in 1982, and the Ph.D. degree in applied physics (concentration in underwater acoustics) from the Catholic University of America, Washington, DC, in 1989. He has been engaged in ocean acoustics research, sonar systems design, sound propagation modeling, and the development of acoustic methods for the detection, localization, tracking, and classification of submarines and other underwater targets. Dr. Arvelo is a member of the Acoustical Society of America, the American Institute of Physics, and the Oceanography Society.

A Neural Network for Passive Acoustic Discrimination between Surface and Submarine Targets

Robert H. Baran

This concerns the use of a neural network for classifying acoustic noise emitted by ships in transit past an omnidirectional hydrophone. Relatively noisy surface ships, moving rapidly at medium to long range, radiate noise which superficially resembles that of quieter submarines, moving more slowly and closer to the listening device. The proposed solution features a three-layer neural network in which the inputs represent the evolution of spectral density through time. The outputs indicate which target type (if any) is present. The network maps inputs to outputs by propagating signals forward through a "hidden layer" of neuron-like units. Errors at the output are propagated backward as incremental changes in the strengths of the connections to the hidden units. Because the number of target types and encounter geometries is far greater than could possibly be covered in any representative way by a training set comprising real-world data, the task is to connect the network to a digital simulator and to show that, by training on the output of the simulator, the neural network can learn to pass realistic tests. This article describes a neural network design-and-testing exercise based on a simple model that captures the salient features of the problem. The fact that the network attained greater than 90 percent classification accuracy provides encouragement for further work along these lines.

Introduction

Recent advances in integrated circuits have brought us parallel computers with hundreds of microprocessors in a single package. These advances give rise to computational models that support massive parallelism in a natural way. Neural computing aims at achieving human-like performance in computer systems by developing an analogy to the structure and operation of the central nervous system. Computations are executed by simple, neuron-like processing units which are densely interconnected by "synaptic" links of variable strength (or weight). The resurgence of interest (since about 1982) in these "connectionist models" is due to the inadequacy of algorithm-driven computing and symbolic artificial intelligence (AI) approaches in dealing with real-world problems, such as speech processing and machine vision. Invariant pattern recognition—the ability to recognize an object in spite of various transformations and obscuring influences—comes easily to our central nervous system. Thus, our eyes quickly focus on key objects in a complex image, and speech is understandable to us irrespective of the tonal qualities and accent of the speaker.

The application of neural-network technology to automatic target recognition (ATR) is motivated by the need for this kind of perceptual generalization. Of the many practical problems in defense electronics that have been addressed by neural network solutions in the last few years, close to half have been aimed at improved ATR, which involves the extraction of critical information from complex and uncertain data. Roth¹ has written a survey of neural-network technology for ATR for systems that employ visual or infrared sensors.

This article describes a feed-forward neural network for classifying acoustic signals emitted by ships in transit by an omnidirectional hydrophone. Relatively noisy surface ships, moving rapidly at medium to long range, emit signals which superficially resemble those of quieter submarines, moving more slowly and closer to the listening device. This range-speed ambiguity makes discrimination between surface ships and submarines a hard practical problem that has resisted satisfactory treatment by conventional methods. The solution will rely on the fact that different types of targets can be distinguished by their radiated noise spectra. The complexity of the modeling required to characterize fully the acoustic signatures of surface and submarine targets is exceeded only by the difficulty of translating such models into algorithms for real-time discrimination. Neural networks provide an alternative to the traditional algorithmic approach. Given a comprehensive training set consisting of typical target signatures and the encounter parameters that generated them, a suitable neural network can learn to associate encounter parameters with the inputs in real time.

The neural network approach is motivated by analogy to the sonar classifier (Figure 1) of Gorman and Sejnowski and noted in the history discussion. The present problem can be solved by a similar network architecture, with the outputs indicating which target type (if any) is present. The inputs represent spectral densities computed for each of a number of time lags. One of the strengths of neural network pattern recognizers is their ability to learn by example from training sets comprising real-world data. In the real world, however, the data may be costly to obtain and widely distributed across a number of carefully guarded data bases. It makes sense to test and debug a software model on an unclassified data base and to show that its performance is competitive with the optimal solutions prescribed by classical signal processing theory before using classified information to develop a training set. Moreover, classified data acquisition was sometimes driven by constraints and priorities that translate into a training set that is not truly representative of the operational environment. Dividing the data into subsets, one for training and one for testing, the analyst still has no way to gauge how effectively the trained network will generalize beyond the horizons of the data.

A Brief History of Neural Nets

The neural network technology of today is based largely on the neuroscience of the 1940s. The classical neuron doctrine developed quickly, following Ramon and Cajal's discovery of the neuron using the light microscope (around 1890). The classical neuron integrates the pre-synaptic activity of all the neurons influencing it, sending information in the form of electrical impulses down the one-way path formed by its axon. It took another fifty years to transcribe the concepts into mathematical terms. McCulloch and Pitts,² in 1943, simplified the neuron to an on/off device, either firing impulses at its peak rate or resting quietly. In 1948, Hebb³ theorized that synaptic weights are modified by a reinforcement control procedure, and he argued that synaptic modification constituted the microscopic, physiological basis of adaptation, learning, and behavioral organization.

By 1954, digital computers had been brought to bear on the problems of brain modeling because the mathematical models involved large numbers of variables and their nonlinear interactions. If "intelligent" behavior was going to emerge from networks of McCulloch-Pitts neurons with Hebbian synapses, it would have to be discovered by computer simulation. No one carried this idea so far, so fast as Frank Rosenblatt, whose discoveries were summarized in 1961.^{4,5} Rosenblatt called his neural networks "perceptrons" and tested them for evidence of perceptual generalization. For example, a network trained to recognize the letter "A" on the top part of a grid-like retina spontaneously recognizes "A" on the lower half. The first 300 pages of Rosenblatt's report were devoted to three-layer, series-coupled perceptrons of the kind that will be described here in subsequent sections.

In 1969, M. Minsky and S. Papert authored a book, *Perceptrons*, which is widely regarded as having had a chilling effect on the subject. Minsky, in the third (1988) edition,⁶ recalls that perceptron research and development had already reached a dead end. The revival of perceptron-like models in the 1980s was caused primarily by the widespread

(Continued)

perception that symbolic AI had reached a plateau, and by the availability of cheaper, faster computers with large amounts of random-access memory to store the synaptic weights of large networks. The connectionist models of the 1980s also overcame some of the limitations attributed to the Minsky-Papert perceptron. David Rumelhart⁷ and the PDP Research Group emphasized the importance of having a "hidden layer" of neuron-like units sandwiched between the input and output layers of the network. Their "powerful new result" that drove the progress of artificial neural nets in the late 1980s was an algorithm called back-propagation, which permits three-layer networks to learn internal representations of data sets and sources for which no mathematical model can be written down to specify the correct responses to given inputs. Instead, the neural network learns by example in the course of many passes through a training set.

Gorman and Sejnowski⁸ used a neural network for the classification of active sonar returns from two undersea objects, a metal cylinder and a similarly shaped rock. In their report, "metal cylinder" was transcribed a "mine." The inputs represented the spectrum levels derived from a lofrogram of the echoes, and the two output units registered the classification. (See Figure 1.) Actual field test data was used to create the training set, consisting of input/output pairs from which the network was able to learn to discriminate between the two kinds of echoes. Errors at the output were propagated back to the hidden layer. The Rumelhart/PDP back-propagation technique provided the formulas for modifying the weights in a way that resists and eventually corrects the errors. After the network had learned the elements of the training set, it showed the ability to generalize by classifying, with better than 95 percent accuracy, the remaining echoes in the whole data base from which the training set had been acquired. This and many other demonstrations testify to the power of back-propagation, which has driven the great majority of neural network applications to date.

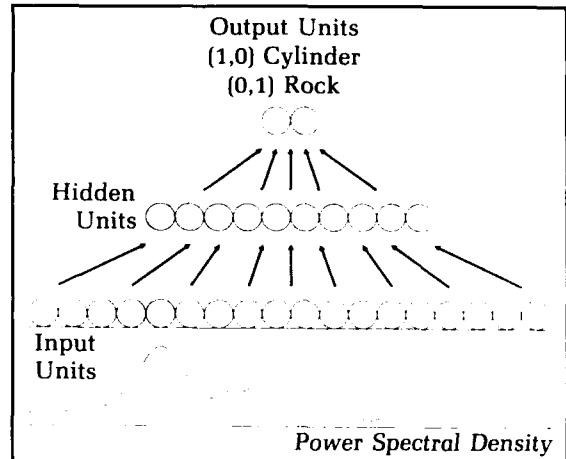


Figure 1. Schematic diagram of the sonar classifier network of Gorman and Sejnowski (1988). The bottom layer has 60 units with their states clamped to the amplitude of the preprocessed signal, shown in analog form below the input layer. Two output units at the top represent the two sonar targets to be identified. The hidden layer allows the network to extract high-order signal features. The connections between the layers of units are represented by arrows.

Therefore, the task is to connect the network to a high fidelity, model-based digital simulator and to show that, by training on the output of the simulator, the neural network can learn to pass realistic tests. This approach was illustrated by Baran^{9,10} in constructing a pattern set for configuring a symmetric (Hopfield) network to classify targets using a different set of sensors and classification criteria. The present case is different by virtue of the technical complexity and procedural difficulties that stand in the way of designing the simulation to achieve the highest possible fidelity. This article describes a neural network design-and-testing exercise based on a simple model that captures the salient features of the problem. Lessons learned in this exercise will bring the practical problem into sharper focus.

Modeling and Simulation

The target is modeled as an acoustic point source that radiates with the same intensity in every direction. Spherical spreading reduces the intensity as the sound travels from the source to the listening hydrophone, and absorption selectively attenuates the higher frequency components of the radiated noise. Ignoring the many effects that would have to be accounted for in a realistic propagation model—like surface reflection, bottom bounce, doppler shift, and an angular dependence of the target's acoustic emissions—the target signature depends on the source level and on the straight-line path followed by the source through the proximity of the listening device. The target signature model follows the evolution of the sound intensity through time t , which is relative to the time of closest approach, when the separation between the source and the listening device is at a minimum. The target signature is a symmetric, bell-shaped curve, the shape of which is determined by the encounter parameters, namely the closest point of approach (c.p.a.) distance, the source intensity, and the target's velocity. (See Figure 2.)

Elements of the Model

Begin with a point source that radiates correlated Gaussian noise with the same intensity in every direction. The intensity at one meter from the source is S_1 . Spherical spreading reduces the intensity to

$$S(R) = S_1 / 4\pi R^2 \quad (1)$$

at range R (meters) from the source. Let H and D be the horizontal and vertical distances, respectively, at time $t=0$. The separation distance at time t is

$$R^2(t) = (Vt)^2 + H^2 + D^2.$$

In particular, the range at $t=0$, at the point of closest approach, is

$$R_0 = (H^2 + D^2)^{1/2}. \quad (2)$$

The target signature,

Relatively noisy surface ships, moving rapidly at medium to long range, emit signals which superficially resemble those of quieter submarines, moving more slowly and closer to the origin. Different kinds of encounters can look the same to a target recognizer that bases its

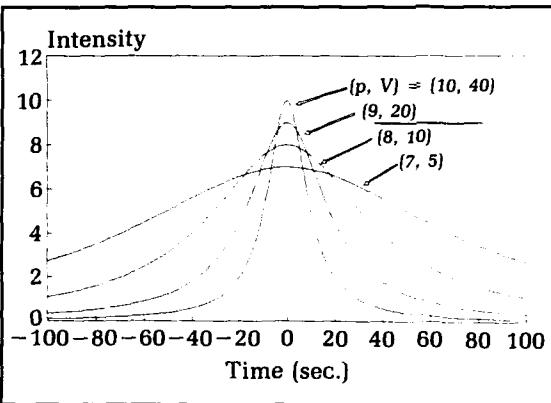


Figure 2. Sound intensity (in arbitrary units) versus time for four point targets with indicated parameters and common closest point of approach distance $R_0 = 400$ meters. Velocities (V) are in m/sec.

$$\begin{aligned} m(t) &= S(R(t)) \\ &= [S_1 / 4\pi R_0^2] / [(V/R_0)^2 t^2 + 1], \end{aligned} \quad (3)$$

is obtained by substituting Equation (2) in Equation (1) and dividing through by $4\pi R_0^2$. The target signatures produced by this equation are symmetrical, bell-shaped curves with peak height

$$\begin{aligned} p &= S(R(0)) \\ &= S_1 / 4\pi R_0^2. \end{aligned}$$

The signature width, which describes the duration of the encounter, is controlled by the parameter

$$q = (V/R_0)^2$$

which appears in the denominator of the expression (3) for $m(t)$. Figure 2 shows four examples of target signatures computed with closest point of approach (c.p.a.) distance $R_0 = 400$ m.

decisions on the temporal variation of sound intensity in a fixed frequency band. This range-speed ambiguity can be resolved by listening to the target through filters tuned to several different frequency bands. The frequency dependence of radiated noise power can be

used to "fingerprint" different ship classes. In the simplest case, a particular type of ship might be identified by certain spectral lines. It can be noted here that surface ships and subs are different by virtue of the latter being much quieter. Regardless of target type, the radiated noise power tends to decline as frequency increases between the nominal limits of 100 Hz (f_l) and 10 KHz (f_h), and the source levels increase with speed across the whole spectrum.

These obvious features were incorporated into a source model in which the radiated noise power, in decibels (dB), is a linear function of the logarithm of the frequency, as shown in Figure 3. The linear model has three

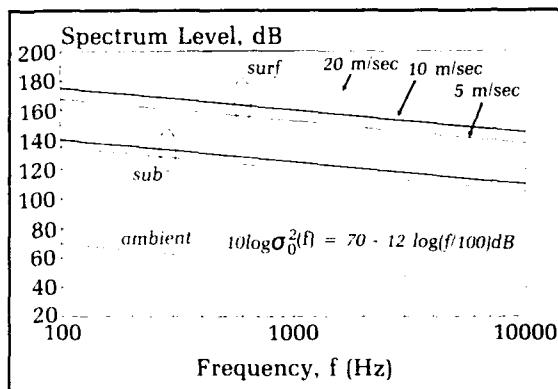


Figure 3(a). Using the linear model (developed below), surface and submarine targets sound the same except for an overall quieting of the latter, which becomes slightly more pronounced as speed increases.

parameters: (a) the peak intensity, which is measured at 100 Hz; (b) a rate of increase in radiated noise power with the ship's velocity; and (c) a roll-off rate that describes the reduction in radiated power with increasing frequency. In Figure 3(a) the roll-off rates of the surface and submarine targets are equal, at 10 dB/decade. The two target types are differentiated mainly by across-the-spectrum disparity in source levels. It is not possible to tell the difference between them without some independent estimate of one of the three natural encounter parameters (S_1 , R_0 , V). Figure 3(b), on the other hand, compares the source levels of two hypothetical surface targets at two different velocities. Increasing the roll-off rate of the surface target to ($c_1 = 15$ dB/decade) makes it possible to discriminate between it and the submarine target. Given a spectral

estimation technique which fits a straight line in the frequency domain to the hydrophone measurements, a target classifier can base its decision on whether the roll-off rate is closer to 10 dB/decade (for the sub) or 15 dB/decade (for the surface ship). Having made this decision, it resolves the range of the target by comparing the evolution of the observed intensity to the known intensity (S_1) at the source.

There may be more powerful algorithms of solving the problem, but their application to real-time signal processing could become more difficult as the source spectra are modeled with increasing sophistication and detail. The next sections will describe the design of a neural

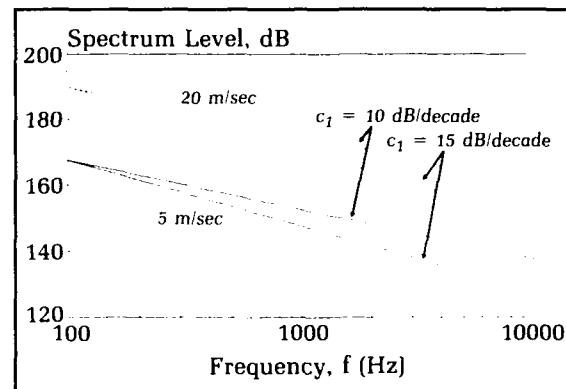


Figure 3(b). Increasing the roll-off rate (c_1) of the surface target's radiated noise power from 10 to 15 dB/decade makes it possible to discriminate it from the nominal submarine target ($c_2 = 10$ dB/decade).

network for implicitly solving the simultaneous nonlinear equations that extract the target type from the target signature. The neural network promises to be robust enough to learn the distinguishing features of realistic spectra. Moreover, the fully trained network is ideally suited to real-time detection and classification.

Range-Speed Ambiguity

Relatively noisy surface ships, moving rapidly at medium to long range, emit signals which superficially resemble those of quieter submarines, moving more slowly and closer to the origin. The range-speed ambiguity is highlighted by rewriting Equation (3) as

$$m(t; p, q) = p/(qt^2 + 1),$$

a scalar function of time which is parameterized by p and q . Imagine a surface ship moving at velocity $V(\text{surf}) = 40\text{m/sec}$ along a straight path that leads to a c.p.a. distance of $R_0(\text{surf}) \approx 2000\text{m}$. The q -parameter for this encounter is $4 \times 10^{-4}/\text{sec}^2$. The same value of q would result from an encounter with a submarine target moving at $V(\text{sub}) = 8\text{m/sec}$ toward a c.p.a. at $R_0(\text{sub}) = 400\text{m}$. The p -parameters will also be the same if

$$S_1(\text{surf})/R_0^2(\text{surf}) = S_1(\text{sub})/R_0^2(\text{sub}),$$

i.e., if the source level corresponding to the submarine is $20\log[R_0(\text{sub})/R_0(\text{surf})] = -14\text{ dB}$ relative to that of the surface target. Different kinds of encounters can look the same to a target recognizer which bases its decisions on the temporal variation of sound intensity in a fixed frequency band. To account for the spectral distribution of the radiated noises from different target types, let $S(f,1;i,V)$ denote the power spectral density at frequency f at 1 meter distance from a point source of type i moving at constant velocity V . The type index determines the source spectrum. If $i = 1$, the point source radiates noise with the characteristics of a surface ship. Case $i = 2$ is a submarine target. Later on it will be convenient to denote the absence of any target by setting $i = 0$. Let the source model be

$$10 \log S(f,1;i,V) = (4)$$

$$a_i + b_i V - c_i \log(f/f_l), f_l \leq f \leq f_h,$$

where the type-specific coefficients a , b , and c are to be defined and f_l and f_h are low- and high-frequency cutoffs. This source model captures two of the most obvious features of ship acoustic spectra. With reference to Figure 3, the radiated noise power increases with velocity, and there is a decreasing trend in the spectrum level as frequency increases. Figure 3(a) was produced with the source level parameters

$$(a_1, b_1, c_1) = (160 \text{ dB}, 1.5 \text{ dB/m/sec}, 10 \text{ dB/decade})$$

and

$$(a_2, b_2, c_2) = (130 \text{ dB}, 1.0 \text{ dB/m/sec}, 10 \text{ dB/decade})$$

with low- and high-frequency cutoffs of 100 Hz and 10 KHz, respectively. The target types

described by these parameters are essentially indistinguishable, since their spectra are the same except for the submarine being quieter over the whole frequency range by a constant (velocity-dependent) factor.

Figure 3(b) shows two surface ship spectra, the first obtained with the same parameters (a_1 , b_1 , c_1) as just noted, and the second obtained with a different roll-off rate, $c_1 = 15 \text{ dB/decade}$. This disparity in the roll-off rates makes it possible to discriminate between the two target types. Let $S(f,R;i,V)$ be the spectral density of the type i source at speed V when measured at a range of R meters, $R > 1$. Then spherical spreading gives

$$S(f,R;i,V) = S(f,1;i,V)/4\pi R^2$$

as in Equation (1). Let $\theta = (i,V,R_0)$ be the vector of encounter parameters. Substituting $R(t)$ for R in the last equation gives an expression for the target signature,

$$\begin{aligned} m(f,t;\theta) &= S(f,R(t);i,V) \\ &= p(f)/(qt^2 + 1), \end{aligned} \quad (5)$$

the same as before except that

$$p(f) = S(f,1;\theta)/4\pi R_0^2$$

is now a function of frequency (and the encounter parameters). Note that

$$p(f)/q = S(f,1;\theta)/4\pi V^2.$$

Defining

$$r = 4\pi p(f)/q,$$

so that

$$S(f,1;\theta) = rV^2,$$

we can equate $10\log(rV^2)$ to the right side of equation (4), obtaining

$$a_i + b_i V - c_i \log(f/f_l) = 10\log(r) + 20\log(V). \quad (6)$$

Given a way of fitting a signature to the time history of the spectral density at f , suitable estimates \hat{p} and \hat{q} can be used to estimate $r/4\pi = \hat{p}/\hat{q}$. Finally, \hat{r} is substituted for r in Equation (6). Taking V as the independent variable, Equation (6) has one solution for each

target type. Figure 4 shows how these solutions are obtained graphically. It assumes a type 1 source at speed $V = 10$ m/sec with a c.p.a. distance of 2000m. The curve is described by the right side of (6). The straight lines plot the left side of the equation for each target type with the same parameters as those used in

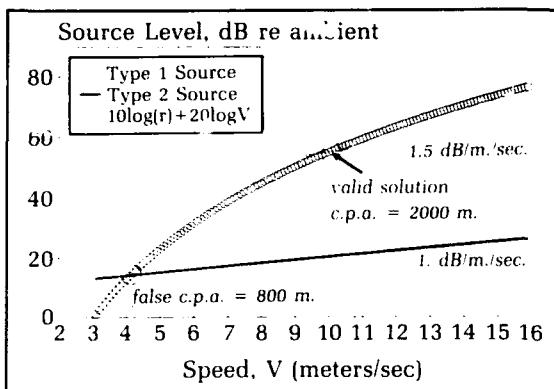


Figure 4. The nonlinear equation (6) has one solution for each target type. This range-speed ambiguity can be resolved by obtaining simultaneous target signatures in each of several frequency bands.

Figure 3(a). The concave curve intersects both straight lines at plausible values of the velocity corresponding to different target types and c.p.a. distances.

The ambiguity can be resolved by repeating the same procedure for each of several frequencies subject to having unequal roll-off rates c_1 and c_2 . Each f gives rise to its own signature, and hence its own estimate \hat{r} . When the correct target type is assumed, the solution to every frequency-specific instance of Equation (6) is the same. Assuming the wrong target type leads to conflicting solutions.

Network Architecture and Training

Virtually every neural network design and training exercise begins with the question of data representation. In the present case, acoustic data needs to be represented at the input of the neural net. The success of the exercise will depend on choosing a data

representation sufficiently detailed to convey the essential characteristics of the target signature, but compact enough to implement. The number of inner product steps (multiply-and-add operations, or "interconnects," in neural network terminology) required to map an input to an output increases as the product of the number of input units times the number of hidden units. More significantly, the time required to train the network using back-propagation may increase rapidly as the size of the network grows.

The problem of invariant pattern recognition arises in the present context when we ask the network to generalize by attributing a target type to an acoustic input which reflects the random variability of the encounter and its acoustic environment. Instead of associating a spectral "icon" with a target-type index, the network must deal with a time-varying spectral density (or lofrogram) which is influenced by the ambient noise as well as the encounter parameters. The availability of cheap, reliable devices for real-time spectral-density computation has made the lofrogram a kind of de facto standard for neural network development in the realm of sonar signal processing. Neural networks have been trained to interpret lofograms in a variety of Navy-sponsored projects. Typically, the analog tapes acquired in a field test are digitized and played through real-time Fourier transformation hardware. The result can be displayed as a moving succession of color-coded scan lines. Each scan line will contain hundreds or thousands of individual components. Therefore, the lofrogram is partitioned into rectangular regions (or bins) and the spectrum levels in each region are averaged. This reduces the pattern to a manageable size. The values associated with the bins are then used as inputs to a trainable neural network which has one input unit for each bin.

The training procedure follows a shortcut from the target signature model to the neural network inputs. Because the spectral densities of the two target types and the ambient noise are presumed given by the equations that produced Figure 3, it is possible to simulate the lofograms by adding random fluctuations to the mean levels specified by the model. The addition of these random fluctuations is accomplished by fairly standard Monte Carlo methods based on an analysis of the mean and variance of the excess power conditioned on the characteristics of the encounter.

Preprocessing and Data Representation

It is necessary to make some concrete assumptions about the preprocessing steps that intervene between the listening hydrophone and the neural network. For each $t = 1, 2, \dots, L-1$ we have a block of serial data, beginning at $t-T/2$ and ending at $t+T/2$, $T < 1$ unit time. Each block is a segment of the time series obtained by sampling and digitizing the output of the hydrophone at a rate of $2B$ samples per unit time. If $2B$ is the Nyquist rate, $B \approx f_h$ is the bandwidth. Each block then contains $2BT$ sampled data points. The blocks are Fourier transformed, each giving rise to a t -dependent power spectral density (PSD). Each of the BT values of the PSD is exponentially distributed with a mean of $\sigma^2(f,t)$, where σ is the r.m.s. noise¹¹ and σ^2 is the noise power. The noise power has an ambient component, $\sigma_0^2(f)$, and a target-generated component which was denoted $m(f,t)$ in equation (5). Thus

$$\sigma^2(f,t) = \sigma_0^2(f) + m(f,t)$$

describes the time- and frequency-dependent mean value of the PSD component. Dividing through by the ambient gives

$$X(f,t) = 1 + m(f,t)/\sigma_0^2(f).$$

the time-dependent excess power spectrum.

It is not generally practical to drive a neural network directly with the logargram

Figure 5 shows how a three-layer, series-coupled perceptron was trained on the output of a simulation based on the model previously described. Following the creation of a network with random initial weights, the training procedure marches through a sequence of cycles. Each cycle follows the same six steps (again with reference to Figure 5):

(1) Generate an encounter parameter vector which specifies target type, c.p.a. distance (R_0), and velocity (V). Let R_0 and V be uniformly distributed between appropriate lower and upper bounds (which may be conditioned on target type). The target type index is zero (i.e., no target) with probability 1/2. If nonzero, the type is 1 (surface) or 2 (sub) equiprobably. The type index determines the source level parameters in Equation (4).

data, since even the simplest problems give rise to X -arrays with hundreds of thousands of elements. Accordingly, the array X is partitioned and the elements of the sub-arrays are summed (or averaged). Each such sum is assigned to one of the input nodes of the neural network. For example, if $B = 10$ kHz, $T = 1$ sec., and $L = 10$ time slices, we have $BTL = 10^5$ elements in the array X with ten rows and 10,000 columns. Let the frequencies be lumped into bins of width B/n , indexed by k , as

$$X_k(t) = \sum_{f=(k-1)B/n}^{kB/n} X(f,t), k = 1, 2, \dots, n. \quad (7)$$

With $B/n = 3$ we have (10 time slices)(3 bins/slice) = 30 nodes in the input layer of our neural network. By the central limit theorems, the distribution of $X_k(t)$ is asymptotically normal with a mean of $\mu(k,t)$ and a variance of $\lambda^2(k,t)$ where, with some approximations,

$$\mu(k,t)/n = 1 + m([(k-1)B/2n,t])/\sigma_0^2([(k-1)B/2n]) \quad (8a)$$

$$\lambda(k,t)/n^{1/2} = \mu(k,t)/n. \quad (8b)$$

Setting $i = 0$ (no target) we have $m = 0$. The network inputs $X_k(t)/n$ are all normally distributed with unit mean and variance in the absence of any target signature.

(2) Compute the target signature from the encounter parameters using Equation (5) for the time- and frequency-dependent mean value of the radiated noise power.

(3) Represent the target signature in a way that can be input to the neural network, making use of Equations (8) to adjust the computer-generated, normal random variables that simulate the excess noise power. This guarantees that the network will never train on the same input more than once, since the probability of generating the same 30 random numbers more than once is negligible.

(4) The network is presented with 30 inputs by subtracting an offset from each of the 30 random variables generated in the last step. Since the mean excess power has unit value when no target is present, subtracting an

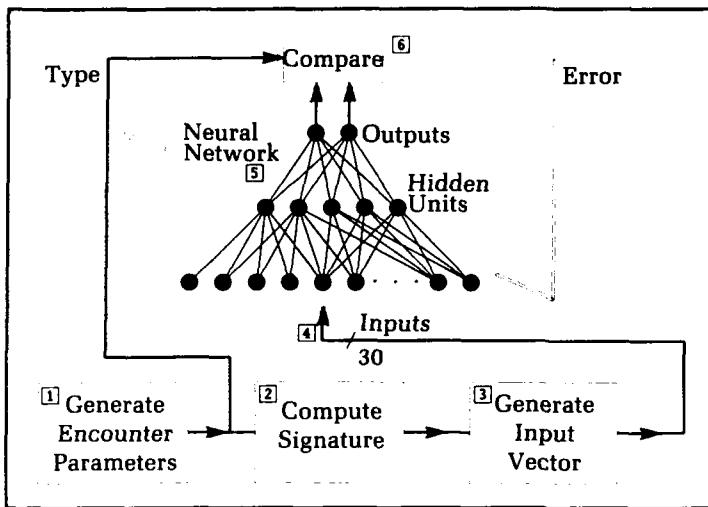


Figure 5. Block diagram of the procedure for training the neural network on the output of a model-based simulation.

offset of 1.0 from each causes the inputs to be equiprobably positive or negative in the absence of a target. When a target is present, these inputs are shifted in the positive direction.

(5) Propagate the inputs forward through the hidden layer to the pair of output units. Figure 6 shows the network architecture in the general case, with N_I inputs, N_H hidden units, and N_O output units. The specifics of the present exercise are $\{N_I, N_H, N_O\} = \{30, h, 2\}$, where the number h of hidden units was variable. The hidden and output layers are composed of binary (0-1) units. The a^{th} hidden unit computes the weighted sum of all 30 inputs (x_1, \dots, x_{30}), the weights being given by a vector of components V_{as} , where $s = 1, 2, \dots, 30$ identifies the input unit. The resulting state (or activation) of the hidden unit is $y_a = 1$ (if the sum is positive) or $y_a = 0$ (if zero or negative). Similarly, the r^{th} output unit assumes the value z_r , which is zero or one depending on the sign of the sum of products $W_{ra}y_a$, $a = 1, 2, \dots, h$. The output convention is $(z_1, z_2) = (0, 0)$ = "00" for no target, "01" for a type 1 target, and "11" for type 2.

(6) Compare the output of the network to the desired states, which correspond to the two digits of the type index. If the output matches the target, return to step (1). If there is an error, modify the weights according to the back-propagation procedure, then return to step (1) above.

In the past several years, back-propagation has become synonymous with the "generalized data rule" introduced by Rumelhart et al.,⁷ who described it as a method of "learning internal representations by error propagation." The trilayer network (in Figures 5 and 6)

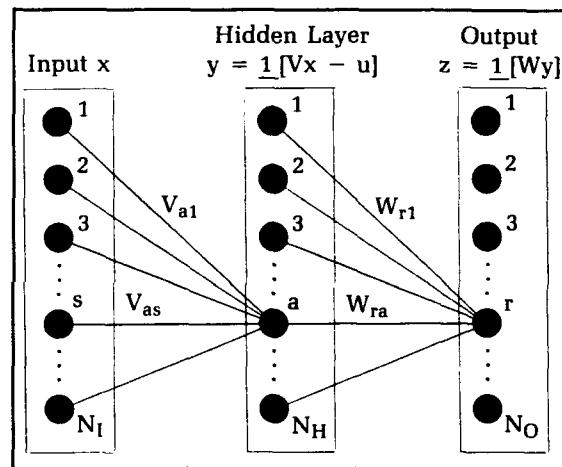


Figure 6. Schematic diagram of a trilayer neural network. The unit step function is $\underline{1}(\phi) = 1$, if $\phi > 0$, or $\underline{1}(\phi) = 0$, if $\phi \leq 0$.

transcribes the input into a pattern of activation in the hidden layer. This pattern of activation (y) is like a code that serves as an intermediate (or internal) representation of the information impressed on the input (x). When the components of the upstream weight matrix (V) are just random numbers, the encoding processes amount to random mixing or scrambling. The "delta rule," also known as the perceptron learning algorithm, was employed by Rosenblatt^{4,5} to decode the internal representation. A given input results in the r^{th} output unit taking on a value (z_r), which initially may differ from the desired or target value (t_r). Error correction is accomplished by incrementally adjusting the components of the

downstream weight matrix (W). The delta rule for adjusting the weight of the connection from the a th hidden unit to the r th output unit is

$$\Delta W_{ra} = gy_a(t_r - z_r),$$

where g , the learning rate parameter, is a small, positive number. Since the hidden unit's activation (y_a) is a binary (0-1) variable, the delta rule does nothing to the connections from hidden units that are turned off ($y_a = 0$) by the input. For those that are turned on ($y_a = 1$), the weight change is proportional to the error. Thus, when a hidden unit excites an output unit that is supposed to be off, the weight of the connection from the hidden unit is reduced. Conversely, when a hidden unit inhibits an output unit that should turn on to the given input, the weight change is in the positive direction. This process of making incremental weight changes, which oppose the errors in classifying the elements of the training set, ultimately corrects all the errors that can be corrected subject to the random coding action of the upstream weight matrix and the hidden units.

What if the "internal representation" is inappropriate to the training set? Since the input-to-hidden layer weights are just random numbers, there can be no guarantee that a solution is attainable by application of the delta rule. For that matter, why include the hidden layer at all if its behavior is not mathematically justifiable? As a psychologist, Rosenblatt was concerned more with nervous system modeling than problem solving. But when Minsky and Papert analyzed the subject⁶ they tacitly did away with the hidden layer in order to achieve mathematical rigor. The result was a model with severe limitations. In the years that followed, this weakened and oversimplified model became the focus of diminishing attention.

When Rumelhart et al.⁷ revived the trilayer, feed-forward architecture, in fact, they provided the theoretical justification for the hidden layer. They disclosed a learning algorithm that could adjust the upstream weights so as to ultimately produce internal representations that constitute a kind of intelligent encoding of the input (instead of a random one). Their "generalized delta rule" applies to networks with hidden units that produce graded responses. Instead of simply turning on or off, the hidden unit's activation is a continuous, differentiable, bounded, increasing function (of the same weighted sum as before). The differentiability of the activation function makes it possible to compute changes to the upstream weights that correct errors through the method of steepest descent. It had been known since

1960 that the delta rule was essentially a steepest descent technique.¹² Yet the assumption of binary hidden units made it impossible to generalize the technique to networks with two layers of modifiable weights:

Although the introduction of hidden units gives a feed-forward network the potential for an arbitrary mapping, before the introduction of [the generalized delta rule], no technique existed for determining the weights of a deterministic, feed-forward network with hidden layers. This was the essence of the credit assignment problem: how to determine the weights of intermediate hidden units so that desired responses occurred as a result of specified inputs.¹

Rosenblatt⁵ addressed this credit assignment problem, which really involves the question of placing blame by attributing errors to hidden units when output units take on the wrong values:

Only one constraint needs to be dropped in order to obtain the most general system of this class: the requirement that the [input-to-hidden unit] connections must have fixed values, only the [hidden-to-output layer] connections being time dependent. In [Chapter 13], variable [input-to-hidden layer] weights will be introduced and the applications of an error-correction procedure will be analyzed. It would seem that a considerable improvement in performance might be obtained if the [upstream weights] could somehow be optimized by a learning process rather than accepting the arbitrary or pre-designed network with which the perceptron starts out. It will be seen that this is indeed the case, provided that certain pitfalls in the design of a reinforcement control procedure are avoided.⁵

With this rationale, Rosenblatt introduced a "back-propagating error correction procedure" for training three-layer perceptrons made with binary units. The main "pitfall" that Rosenblatt avoided was determinism: Given a three-layer, series-coupled perceptron and a training set for which a solution exists, it may be impossible to achieve the solution with any deterministic error correction procedure that obeys the "local information rule" by which only the hidden unit itself and the units to which it is connected can determine the error assigned to it.

Ironically, the generalized delta rule of Rumelhart et al. is a local and strictly deterministic procedure which propagates errors

(corrections) back to the hidden layer. The "pitfall" of getting stuck in local minima of the error function is avoided by injecting "noise" at the input during the training process or by rebooting the process from a different initial weight assignment when convergence gets bogged down.

From the academic viewpoint, the generalized delta rule of Rumelhart et al. is a superior form of back-propagation, since it is based on the gradient descent in squared error. From an engineering viewpoint, there are good reasons to seek a back-propagation algorithm that works with binary units. In most applications the objective is a trained network that accomplishes the desired mapping from input to output. The result of training is a set of weights, which solves the problem when used to connect neuron-like units with specific activation functions. If the same weights are used to connect units with activation functions that deviate from those of the units employed during training, the result will be degraded performance. To achieve extremely fast operation in a weapon system, for example, the neural network may have to be inserted as an analog integrated circuit in which the weights are electronic conductances and the neuron-like units are transconductance amplifiers. It is much harder to fabricate, on a single chip, large numbers of amplifiers with the same ideal, sigmoidal response characteristics than it is to construct hard limiters that merely respond to the sign of the input. For this reason, it is much easier to fabricate neural networks that use binary units. Even if the inserted neurocomputer is a fully digital device, the use of binary units makes it generally possible to represent the weights with fewer bits of accuracy. With these points in mind, the network shown in Figure 5 was trained using Rosenblatt's "back-propagating error correction procedure" (described at the end of the article).

Performance Analysis

The neural network can commit errors of eight different kinds, since three possible target types are classified by two bits at the output of the net. Table 1 assigns a nomenclature to these errors. The output is supposed to discriminate between surface ships and subs only when the right output bit indicates rejection of the null hypothesis, $i = 0$. (With reference to Figure 6, this output bit is z_2 , the state or activation of the second output unit, which is the one on the right when the network is drawn as in Figure 5.) The table shows two kinds of false alarms (FA) and four kinds of false rest (FR). Error types CE correspond to the off-diagonal elements of the "confusion

152

matrix." All kinds of errors are treated equally by the "back-propagation error correction procedure" used in the training process. In the evaluation process, on the other hand, the output "10" in response to an input of type 0 is not an error, because the right (least significant) output bit, which indicates the presence of a target, enables the decision-making process which begins by reading the other output bit. The false alarm probability is $FAP = \Pr\{FA\}$ and the false rest probability (FRP) is similarly defined. The confusion probability, $CEP = P\{CE\}$, is the probability of identifying surface ships as subs or vice versa. Finally, the total probability of error is

$$PE = FAP + FRP + CEP. \quad (9)$$

Table 1. Classification of Errors

Type	Output			
	"00"	"01"	"11"	"10"
0		FA	FA	*
1	FR		CE	FR
2	FR	CE		FR

* This entry is treated the same as "00" (see text).

Detection Experiments

Detection experiments were designed to show that the neural network can learn to distinguish between targets (type 1 or 2) and background ($i = 0$) with error probabilities approaching some point on the detector operating characteristics of the optimum detection receiver prescribed by the classical theory.¹⁴ The detector operating characteristic in general is a concave (downward) curve on the unit square which shows $1 - FRP$ as a function of FAP, usually parameterized by a detectability measure such as signal-to-noise ratio (SNR). Since targets and background are presented equiprobably in the training process, it is not surprising to find that the operating point of the trained net lies in the vicinity of $FRP \approx FAP$. Thus, we use the detection error probability,

$$DEP = FRP + FAP, \quad (10)$$

as the performance figure of merit.

By taking target velocity V sufficiently small, and setting $c_1 = c_2 = 0$ in the source model, one can artificially cause all the inputs to be in-

dependent and identically distributed under both the null hypothesis and the composite alternative ($i \geq 0$). This gives rise to the statistical problem of detecting a shift in the mean of a normal population, subject to a proportionate increase in the variance. Figure 7 shows the theoretical DEP as a function of a nominal SNR which, with reference to Equation (4), is

$$10\log(\text{SNR}_i) = a_i - 20\log R_0 - a_0,$$

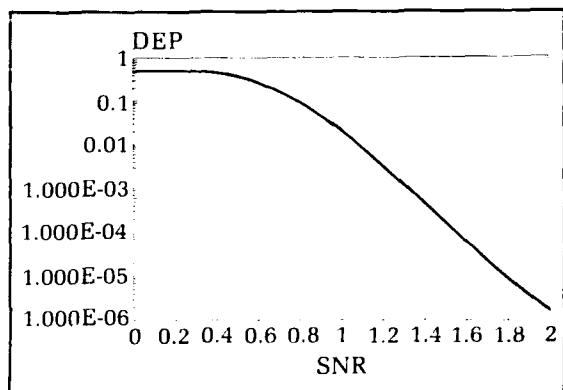


Figure 7. Detection error probability versus signal-to-noise ratio for the optimum detection receiver.

with $a_1 = a_2$. Given the same c.p.a. distance for each encounter, there is no way to distinguish between surface and submarine targets in this degenerate case. Therefore, the output of the target-type indicator unit in the neural network was ignored in both training and evaluation.

The DEP of the neural network approaches the classical result as the training procedure is iterated many times. Figure 8 shows the detection error rate for a net with six hidden units operating on a nominal SNR = 1.42, or $10 \log 1.42 = 1.5$ dB. The error rate is a 500-point moving average computed every 10 cycles of the training procedure. After about 1000 cycles, the error rate falls below 1/500. The time required to reach $\text{DEP} = 1/500$ the first time ranged from about 60 (in the figure) to over 150 cycles in similar trials, which differed only with respect to the seed of the random number generator. (The seed determines both the training input and the initial weights.) Higher values of the SNR gave faster convergence to the 1/500 error level, but make estimation of the limiting DEP very time consuming.

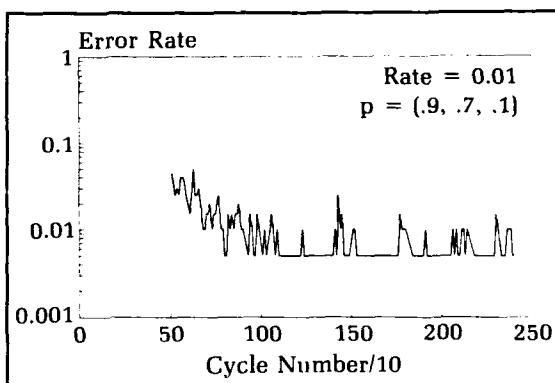


Figure 8. Detection error rate vs. number of cycles for a (30,6,2) BP net with nominal $\text{SNR}_1 = \text{SNR}_2 = 1.5$ dB. The rate is computed every 10 cycles as a 500-point moving average.

Detection-and-Classification Experiments

In this section we discuss some actual tests conducted with the neural network to see if it would, with training, resolve the range-speed ambiguity. The ease of the discrimination task increases with the divergence of the two source spectra, described above in terms of source level parameter vectors (a_i, b_i, c_i) , where $i = 1$ or 2. The source level parameters were the same as given in connection with Figure 3. The roll-off rate parameter c_1 for the surface ship was set at 15 dB/decade in contrast to that of the submarine target ($c_2 = 10$ dB/decade). In order to see whether the neural network could resolve the range-speed ambiguity, the c.p.a. distance and velocity were greater for type 1 targets. Since the surface ship is 30 dB louder than the submarine at the low end of the spectrum, the c.p.a. distances for type 1 targets were made typically 30 times greater than for type 2 targets. In particular, the c.p.a. distance for the surface ship was bounded below by 3000 meters and above by 15 kilometers; but for submarine targets the bounds were 100 meters and 500 meters. Depending on the randomly generated target type index, the parameter R_0 was uniformly distributed on one of the intervals defined by these upper and lower bounds.

The speed V is likewise restricted to a type-specific range in a manner that heightens the ambiguity. One way to "confuse" the network is to let the speed-to-c.p.a. distance ratio (\sqrt{q}) approach unity. In order to keep the speeds of both target types within reasonable bounds, however, the signature generator produced speeds uniformly distributed between 20 and 40 m/sec for surface targets and between 5 and

20 m/sec for subs.

The speeds and distances established by these modeling assumptions give rise to signature widths up to a few hundred seconds. The pre-processing assumed in generating the normal distributions in Equations (8) was made to reflect the signature duration. This was accomplished by using a $T = 1$ second integration time for each time slice and by spacing the slices 20 seconds apart.

After "reconnecting" the target type indicator unit to the neural network, minimization of the total error probability, given by Equation (9), was observed over the course of many training cycles, each involving the presentation of simulated target signatures (or noise) at the input. Figure 9 summarizes the results of detection-and-classification experiments in which the network was given 2500 training cycles to reduce the total error rate. Each plotted point is the average of 10 independent trials differing with respect to the random numbers generated (and the initial weights).

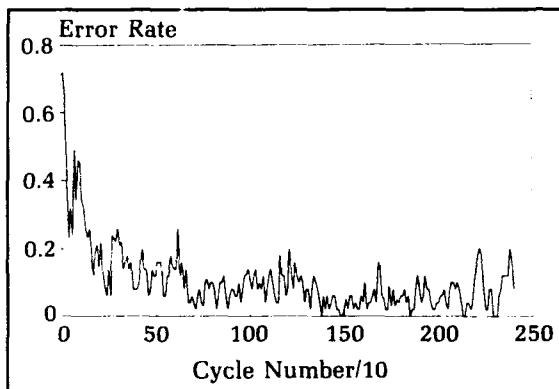


Figure 9. Total error rate vs. number of cycles for a (30,16,2) network trained to discriminate between surface and submarine targets. The error rate is computed every 10 cycles as a 10-point moving average.

Each trial yields a 10-point moving average error rate over 2500 training cycles. The pre-training P_E , from Equation (9), is about $3/4$, under the assumption that each output unit is equiprobably right or wrong. The error rate quickly drops below the 50 percent level. After 500 training cycles the learning curve begins to approach its asymptotic level, $P_E = 0.04$.

These results were obtained with 16 hidden units. Reducing the number of hidden units to 8 or fewer caused the learning curve to flatten in the vicinity of $P_E = 0.15$. Broadening the hidden layer, which caused the execution time to scale up as $(N_I + 2)h$, did not appreciably improve the performance of the net after 2500 cycles.

Similar experiments were performed with larger and smaller values of the minimum SNR, which is the excess power radiated by the submarine target when the c.p.a. distance is maximum and the velocity is set to zero. The upper bounds on the type-specific maximum c.p.a. distances that were noted above correspond to a minimum SNR of 6 dB. Minimum SNR values between 10 dB and 3 dB were established by readjusting the bounds on c.p.a. distances. The limiting total error probability (P_E) was observed to increase gradually as the minimum SNR was reduced, rising from 0.02 (at 10 dB) to 0.35 (at 3 dB).

Conclusions and Recommendations

In the examples considered, the network easily learned to detect the presence of a target of either type, but had minor difficulty in discriminating between surface and submarine targets. This was not surprising, inasmuch as the target signature generation model was adjusted to highlight the range-speed ambiguity. Superb performance could probably have been obtained by presenting the neural net with an easier problem. The fact that the network attained better than 90 percent classification accuracy provides encouragement for further work along these lines. A detailed examination of neural net's performance needs to be preceded by a thorough analysis of classical solutions appropriate to the present case in order to set goals for the total probability of error for parameter combinations of interest. Once again, the neural network approach is regarded as a potential shortcut to near optimal performance, as opposed to a means of exceeding the classically derived optima.

It may be that academic curiosity concerning Rosenblatt's back-propagation technique diverted these efforts from the mainstream of contemporary neural network design practice. Use of hidden units with a differentiable (sigmoidal) activation function and the Rumelhart/PDP gradient descent form of back-propagation might lead readily to better classification accuracy. Indeed, the gradual improvement in total error rate which accompanied the broadening of the hidden layer might suggest that perceptrons with binary threshold units suffer from a reduced capacity relative to today's standard back-propagation nets. On the other hand, if an effective neural network solution can be obtained with binary units, it will be particularly easy to realize high-speed detection and classification in the finished product which, in this case, could take the form of analog integrated circuitry in which the weights are fixed conductances and the neuron-like units are high-gain amplifiers.

Back-Propagation

Rosenblatt (1961, 1962)^{4,5} pioneered the study of learning algorithms for feed-forward neural networks with binary threshold units. His "perceptrons" typically had a "retina" of sensory (S) units feeding forward to an association (A) layer which, in turn, drove the response (R) units. Today it is customary to refer to input units, hidden units, and output units. Mixing the new parlance with the old, most of Rosenblatt's S-A-R perceptrons had input-to-hidden layer (S-A) weights that were fixed while the hidden-to-output (A-R) weights were modified incrementally in the course of many passes through the training set. The asymptote of the individual perceptron learning curve was clearly limited by the suitability of the fixed S-A weights, which were generated by a stochastic rule.

Rosenblatt⁵ (page 292) also used a "back-propagating error correction procedure" for training the S-A weights. The essence of this back-propagation technique was a brief list of rules for assigning errors to hidden (A) units based on their interactions with output (R) units that assumed the wrong state in response to the training input. This back-propagation algorithm is a "supervised" learning algorithm which obtains its feedback from the output units, computing errors by comparing their observed states to preassigned correct values, and propagating corrections back toward the input end of the net if a satisfactory solution cannot be found quickly by making corrections at the output end. The actual modification to the weights is formally the same whether an output unit or a hidden unit is considered. Thus, if the error assigned to a unit is positive, the weights of all connections from active units are increased, eventually turning it on. If the error is negative, the weights of connections from the active units are decreased.

(I) For the r^{th} output unit, let $d_r = t_r - z_r$ be the error, where t_r (equal to 0 or 1) is the desired (target) state and z_r (equal to 0 or 1) is the observed state. Note

$$z_r = \underline{1} \left[\sum_{a=1}^{N_H} W_{ra} y_a \right]$$

in terms of the unit step function

$$\underline{1}[\varphi] = \begin{cases} 1 & \text{if } \varphi > 0 \\ 0 & \text{if } \varphi \leq 0 \end{cases}$$

and the A-R weight matrix W , where

$$y_a = \underline{1} \left[\sum_{s=1}^{N_I} V_{as} x_s - u_a \right]$$

is the state of the a^{th} hidden unit, obtained by feeding the inputs (x) forward through the S-A weight matrix V . Thresholds (u) are ascribed to the hidden units. (Note that d_r assumes one of three values: -1, 0, or +1.)

(II) For the a^{th} hidden unit, error E_a is computed as follows for each training input: Begin with $E_a = 0$.

(IIA) If $y_a = 1$, and a nonzero error d_r differs from the sign of W_{ra} , then decrement E_a by one with probability p_1 .

(IIB) If $y_a = 0$, and a nonzero error d_r agrees in sign with W_{ra} , then increment E_a by one with probability p_2 .

(IIC) If $y_a = 0$, and a nonzero error d_r differs in sign from W_{ra} , then increment E_a by one with probability p_3 .

For all other conditions, E_a is unchanged.

(III) Let $\text{sgn}(\cdot)$ be the sign function. Define $\text{sgn}(0) = 0$. The modification is

$$\Delta W_{ra} = g y_a \text{sgn}(d_r) \text{ for A-R weights or}$$

$$\Delta V_{as} = g x_s \text{sgn}(E_a) \text{ for S-A weights,}$$

using a suitable learning rate parameter $g > 0$.¹⁴ This procedure is illustrated in Figure 10.

While Rosenblatt mainly restricted his attention to binary 0-1 inputs, the formula for the modifications to the upstream weights V can be used on real-valued inputs without any formal change.

The results herein were obtained with $g = 0.01$ and $(p_1, p_2, p_3) = (.9, .7, .1)$, based on study of the few results exhibited by Rosenblatt and on rather extensive numerical experiments in which small perceptrons were trained to classify random patterns of binary digits.¹³

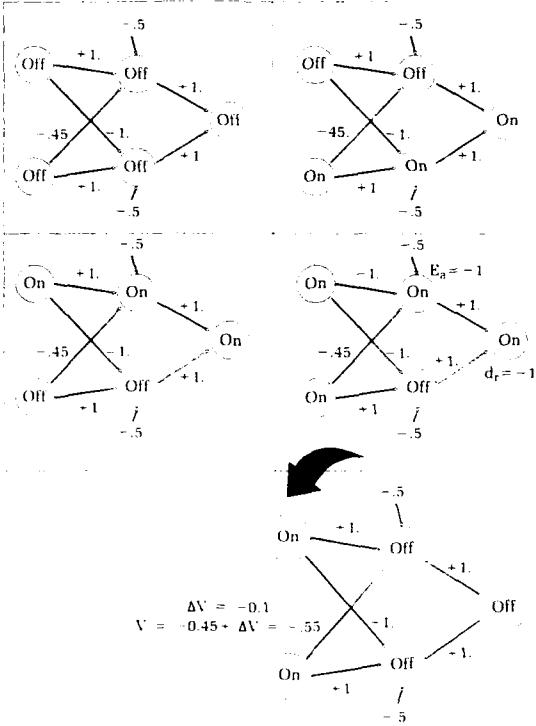


Figure 10. Back-propagation is used to train a 5-unit net to perform the exclusive-or (XOR) logic function. (The XOR, a function of two binary digits, is one when they differ and zero when they are alike.) The net has two input units, two hidden units, and one output. The hidden units have thresholds of 0.5, equivalent to externally-applied biases of -0.5. Weights are noted beside the feed-forward connections. The units are ON and OFF when their states are 1 or 0, respectively. The mapping in the lower right quadrant is incorrect, since $\text{XOR}(1, 1) = 0$. The error at the output, $d_r = -1$, is propagated back to the active hidden unit, giving $E_a = -1$ with probability p_1 (close to one), in accordance with the pertinent rule (IIA) of Rosenblatt's procedure. Then the weight of the connection to the active (ON) hidden unit from the lower input unit is changed by an amount $\Delta V_{as} = g x_s \text{sgn}(E_a) = -g$. Taking $g = 0.1$ for the learning rate, the weight from the lower input to the active hidden unit is reduced by 0.1, from -0.45 to -0.55. This solves the problem, since both hidden units now turn OFF when both inputs are ON.

Acknowledgments

This work was sponsored by the Office of Naval Research through the Naval Surface Warfare Center's Independent Research Program and supported by the Center's Mine Systems Office. The scholarship reflected in this report was made possible by the efforts of the library staff at White Oak. Rosenblatt's *Principles of Neurodynamics*, out of print for thirty years, was obtained from the Defense Technical Information Center (DTIC) as a 1961 Cornell Aeronautical Labs technical report. This report was delivered after a long wait, since the pages were too frail for DTIC to run through ordinary copiers.

References

1. Roth, M. W., "Survey of Neural Network Technology for Automatic Target Recognition," *IEEE Transactions on Neural Networks* 1 (1), 1990, pp. 28-43.
2. McCulloch, W. S. and Pitts, W., "A Logical Calculus of Ideas Immanent in Nervous Activity," *Bulletin of Mathematical Biophysics*, Vol. 5, 1943, pp. 115-124.
3. Hebb, D. O., *The Organization of Behavior*, Wiley, NY, 1948.
4. Rosenblatt, F., *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*, Cornell Aero. Lab., Buffalo, NY, Tech. Report No. VG-1196-G-8, 1961.
5. Rosenblatt, F., *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*, Spartan Books, Washington, DC, 1962.
6. Minsky, M. L. and Papert, S. A., *Perceptrons*, Expanded Edition, MIT Press, 1988.

7. Rumelhart, D. E. and McClelland, J. L., and the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. I. MIT Press, Cambridge, MA, 1986.
8. Gorman, P. and Sejnowski, T., "Analysis of Hidden Units in a Layered Network Trained to Classify Sonar Targets," *Neural Networks* 1 (1), 1988, pp. 75-90.
9. Baran, R., "A Collective Computation Approach to Automatic Target Recognition," *Proceedings IJNCC-89*, Vol. I, Washington, DC, 1989, pp. 39-44.
10. Baran, R., "A Neural Network Approach to Data Fusion in Automatic Target Recognition," *International Journal Neural Nets.: Research and Applications* 1 (2), 1989, pp. 68-77.
11. Korngold, E., *The Periodic Analysis of Sampled Data*, MIT Lincoln Lab. Group 22 Report, 1964-2.
12. Widrow, B., and Lehr, M. A., "Thirty Years of Adaptive Neural Networks: Perceptrons, Madaline, and Backpropagation," *Proceedings of the IEEE* 78 (9), 1990, pp. 1415-1442.
13. Baran, R., "Learning Random Patterns with Rosenblatt's Back-propagating Error Correction Procedure," *Communications on Neural Network* (to appear 1991).
14. Van Trees, H. L., *Detection, Estimation, and Modulation Theory*, Part I, Wiley, NY, 1983.

The Author



ROBERT H. BARAN obtained a B.S. in electrical engineering (with industrial practice) from MIT in 1973, and held several jobs dealing with nuclear hardening of strategic systems. He joined the Naval Surface Warfare Center in 1980, obtaining an M.S. in electrical engineering (communications) from the University of Maryland in 1983. His work as an independent researcher with the Center includes a variety of applied stochastic models with emphasis, since 1987, on neural networks with stochastic dynamics.

Platform Attitude Determination by the Use of Global Positioning System

Alan G. Evans, Bruce R. Hermann and B. Larry Miller

The article describes work performed during the past eight years which established that the Global Positioning System (GPS) is capable of providing users with real-time platform attitude. This capability was demonstrated in an experiment performed during July of 1990 at the Naval Surface Warfare Center (NAVSWC). Real-time azimuth determination of a moving vehicle was achieved using a prototype, dual-antenna GPS receiver developed by Texas Instruments.

The article begins with a brief discussion of early static azimuth and elevation estimates. It continues with results from an astronomical azimuth comparison of conventional stellar measurements with a new technique that employs geodetic azimuths determined from GPS; a dynamic ship's heading test using a prototype three-antenna receiver built by Trimble Navigation, Ltd.; results from the July 1990 real-time dynamic test; and covariance analyses of combined multiple-antenna GPS and inertial navigation systems (INS). The article concludes that two-antenna GPS real-time attitude determination capability is currently available and can significantly enhance the operation of the INS, especially in the azimuth (heading) component.

Introduction

For certain military navigation applications, such as pointing cameras in a reconnaissance aircraft or directing naval guns toward a land target, platform attitude determination becomes an important factor. The Global Positioning System's primary mission is to solve the navigation problem: absolute position and velocity of a single antenna on a vehicle platform in real-time. With the use of multiple antennas and time difference-of-arrival techniques, the capabilities of GPS can be extended to include the determination of platform attitude, or orientation.

This article represents an overview of NAVSWC's efforts to develop a capability to determine a platform's attitude using GPS. Test results obtained in the early 1980s are presented,^{1,2,3} followed by results obtained over longer baselines with the aim of approaching the ultimate GPS orientation accuracy. These tests were conducted jointly with the Defense Mapping Agency (DMA) and the National Geodetic Survey (NGS).⁴ The performance of the first position and attitude GPS receiver is discussed; it was developed through a contract with Trimble Navigation, Ltd., of Sunnyvale, California.⁵ Shipboard testing of this receiver is also reviewed,⁶ along with recent testing conducted at NAVSWC⁷ of a pointing receiver developed by Texas Instruments, Inc. (TI), of Plano, Texas. This receiver was especially designed for precise measurement of the phase difference between two antennas.⁸ After an initialization period, this pointing receiver obtains the azimuth and elevation of the two antennas in real-time. Then, covariance analyses of the attitude accuracy enhancement obtained by combining GPS with an INS are presented. Both a very early (1976) study of a single-antenna case⁹ and a recent multiple-antenna case are summarized. Lastly, the current status of GPS attitude determination development is evaluated.

Early Static Test Results

The first demonstration using GPS to determine attitude was completed in November 1982.¹ This demonstration reprocessed the phase measurement data collected as part of a GPS sensitivity experiment that took place in May 1980.² Here, a single antenna was moved around a right triangular path, stopping at each vertex to collect data for several minutes. The legs of the triangle were approximately 2 meters, and the movement from one vertex to the next was accomplished within 1 minute. The GPS receiver used in this experiment was a single satellite tracker built by Stanford Telecommunications, Inc., and interfaced with data collection hardware and software developed at NAVSWC for DMA. This experiment demonstrated the potential for high accuracy change-in-position estimates when continuous carrier-phase-tracking is achieved. By including several complete circuits of the triangle and enhancing the geometry by tracking two different satellites, orientation estimates for roll, pitch, and yaw, with probable errors of -0.57, -0.28, and 0.03 degree, respectively, were obtained. A similar test in February 1983 was conducted by Magnavox Advanced Products and Systems Company¹⁰ where an antenna was repeatedly moved vertically to obtain a vertical elevation angle.

Although the moving single-antenna procedure requires mechanical motion, it can minimize systematic errors such as different or migrating electrical phase centers inherent to the multiple antenna (coherent phase) procedure. A recommended application for the single moving antenna is to provide the orientation of a spinning satellite.¹¹ An interferometric approach using the differences between coherent phase measurements of at least two antennas was recommended^{11,12,13,14} for various antenna baselines. The interferometric procedures obtain phase measurements at two antennas at the same time instant as shown in Figure 1. The receivers measure the fractional part of the phase measurements (designated by F_A and F_B). These measurements can be related to the difference in range to the satellite (ΔR_c) if the integer number of cycles is obtained. Generally, computer algorithms take advantage of the known antenna baseline length to obtain this quantity.

The advent of coherent and coherently integrated phase-measuring receivers enabled the interferometric attitude determination to become a reality. One early such receiver was the TI4100.¹⁵ Its development was sponsored jointly by DMA and NGS with technical and contractual assistance from the Applied

Research Laboratories of the University of Texas at Austin, NAVSWC, and the Applied Physics Laboratory of Johns Hopkins University.

Static azimuth and elevation were obtained for a 25-meter antenna separation distance³ by connecting two receivers to the same external atomic clock. NAVSWC-developed GEodetic SAtellite Receiver (GESAR) software¹⁶ controlled the phase measurements to take place at the same GPS time (to within the 100-nanosecond synchronization error). Once the integer number of cycles between the phase measurements at each antenna was determined, estimates of azimuth and elevation were obtained at each time mark. A plot of the azimuth error as a function of time is given in Figure 2. The resultant standard deviation of 0.000099 radian, which if translated from a 25-meter baseline to a one-meter baseline represents about a 0.14-degree error, demonstrated the potential accuracy of GPS for attitude determination. The variation of the estimate is thought to be due to signal multipath.¹⁷ Day-to-day comparison of short static baselines has shown¹⁸ that signal multipath is one of the main contributors to GPS attitude estimation errors. Consequently, the accuracy at any given site is very much dependent on the environment of the receiver antennas. More recently, dynamic attitude tests have been performed by combining two GPS receivers as described above.^{19,20,21}

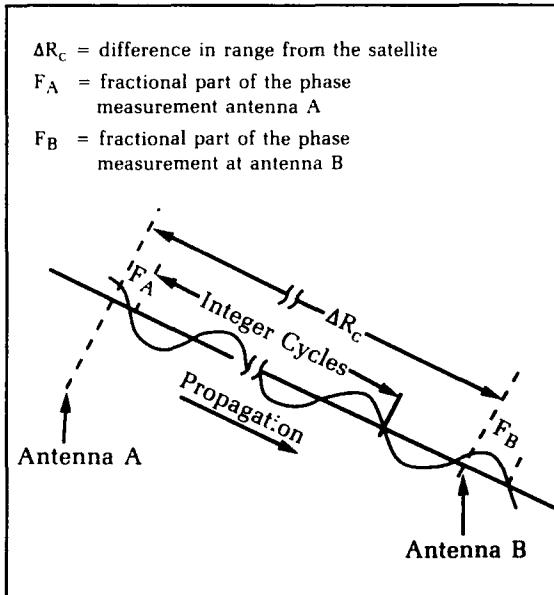


Figure 1. GPS phase measurement at a single-time instant.

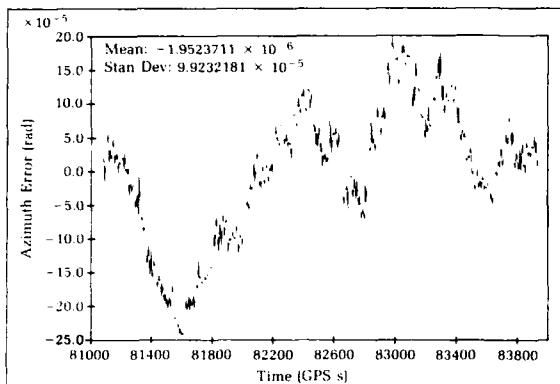


Figure 2. Azimuth estimation error for a static 25-meter baseline.

Longer Baseline Static Tests

To demonstrate further the potential of GPS to determine azimuth, longer (1.2 and 8.3 kilometer) baselines were used. A joint DMA, NGS, and NAVSWC test was conducted at Dahlgren, Virginia, in July 1987.⁴ Using about three hours of static GPS data per day from TI4100 receivers, geodetic azimuths were obtained. These values were converted to astronomic values; the comparisons are given in Table 1. Conventional astro-geodetic values of gravity deflections of the vertical were used to convert the geodetic values. Agreements, -0.27 arcsecond for the 8.3-km baseline (between stations MBRE and RAD9) and -0.97 arcsecond for the 1.2-km baseline (between stations MBRE and HERO), are within the standard deviations of the reference values. The task of determining the reference astronomic azimuths from stellar measurements was delegated to the DMA Geodetic Survey Squadron based in Cheyenne, Wyoming.

The daily estimated values of geodetic azimuth and baseline length are given in Table 2. These results demonstrate the repeatability of

Table 1. Comparison of Azimuths

Stations		GPS Azimuth					Conven-	Diffe-
		Geodetic			$\Delta\alpha$ (Arc)	Astro- nomic (Arc)		
From	To	Deg	Min	Arc		Astro- nomic (Arc)	Difference (Arc)	
MBRE	RAD 9	144	16	21.823	4.103	25.926	26.2	-0.27
MBRE	HERO	293	29	2.322	4.103	6.424	7.4	-0.97

*Derived using GPS geodetic azimuth and conventional astrogeodetic vertical deflections.

the GPS measurements. For the four days, numerical standard deviations for the estimated individual values were 3 millimeters and 1 millimeter for baseline length, and 0.08 arcsecond and 0.31 arcsecond for geodetic azimuth, respectively. The GPS baseline length estimates agreed to about 2.6 parts per million with the electronic distance measuring devices. Here, NGS PHASER software²² and NAVSWC precise post-fitted satellite ephemerides²³ were used for the GPS relative positioning.

Ultimately, the accuracy obtainable when using GPS for attitude determination is dependent on the quality of the satellite ephemerides and clock estimates. Angular errors in satellite position can introduce similar errors in a GPS relative position on a one-to-one basis.²² At GPS altitude, an angular error of 1 arcsecond is equivalent to approximately 124 meters. Testing of the NAVSWC precise post-fitted ephemerides indicates that the satellite orbit errors do not exceed 10 percent of this value. Consequently, any rotation introduced by the post-fitted ephemerides is well below 1 arcsecond. Also, the real-time broadcast ephemerides are limited to 100-meter error, causing less than 1 arcsecond rotation.

Table 2. Repeatability Analysis

Stations		Day of Year	Baseline Length (m)	GPS Azimuth		
				Deg	Min	Arc
MBRE	RAD 9	229	8372.700	144	16	21.889
		230	2.700			21.881
		231	2.705			21.857
		232	2.706			21.865
	MEAN		2.703			21.823
	STD DEV		±0.003			±0.083
MBRE	HERO	231	1153.772	293	29	2.011
		232	3.771			2.622
		233	3.773			2.334
	MEAN		3.772			2.322
	STD DEV		±0.001			±0.306

Trimble Heading and Attitude Receiver

On 28 January 1985, Trimble Navigation, Ltd., started work on a Small Business Innovation Research (SBIR) contract administered by NAVSWC. The objective of the contract was to design and build a prototype receiver that determined both position and attitude.⁵ The

receiver was tested on board USS *Yorktown*, a U.S. Navy guided missile cruiser, in July 1988.⁶

The Heading and Attitude receiver (HAT) used three antennas in a right triangular array, as shown in Figure 3. The primary baseline, used to determine heading and pitch, was 60 centimeters in length. The secondary baseline, used to determine roll, was 40 centimeters. The compact receiver equipment is shown in Figure 4. The equipment used on *USS Yorktown* employed a portable personal computer to record the pseudo-range and phase observations for post-processing. Attitude estimates were performed after the test, but enough information was available in real time to assure that the data collection was successful.

The *USS Yorktown* tests consisted of a static portion performed while the ship was tied up at the dock, and a dynamic portion performed while under way. As part of the dynamic tests, the ship's heading was varied from approximately 40 to 150 degrees in a cyclic pattern within a period of about 15 minutes. Since the GPS receiver and the ship's navigation system kept time independently, this cyclic path allowed an unambiguous comparison to be made between the estimated GPS azimuth values and the results obtained from the ship's INS.

The standard deviation of the difference in heading between the ship's INS and HAT was 1.5 degrees. This result was obtained in the difficult signal multipath environment on *Yorktown* and for a short 60-centimeter primary



Figure 3. Trimble heading and attitude receiver antenna array on *USS Yorktown*. The three antennas define a plane that is fixed on *Yorktown*. The roll, pitch, and yaw movements of the ship are sensed as phase differences by the array. The view is from the deck above the AEGIS radar, facing aft. Note that the guard rails are made of nonreflective fiberglass.

baseline. The pitch and roll estimates were affected by the dilution of precision due to available satellite geometries during the test.

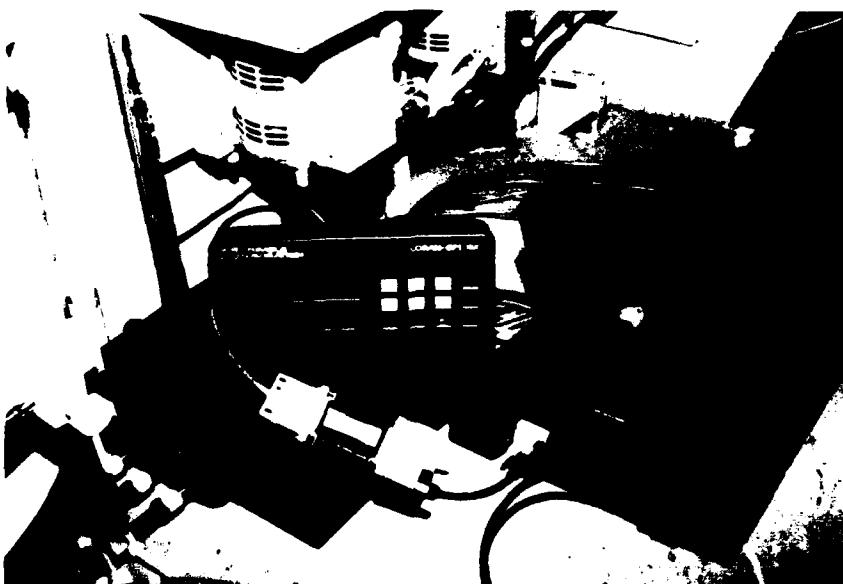


Figure 4. The electronics were located in *USS Yorktown*'s helicopter control room. A laptop portable personal computer (right) communicated with the receiver and recorded all data on floppy disks. The Control Display Unit (CDU) (center) allowed the operator to enter commands into the GPS receiver. The GPS receiver (left, under the CDU) acquired and tracked the signals supplied from the three antennas.

The standard error for pitch, which uses the same measurements as the azimuth estimate, was 4.3 degrees. On the shorter perpendicular baseline, the roll standard deviation was 5.6 degrees.

Due to funding constraints of the SBIR contract, equipment compromises were required. Off-the-shelf hardware was used to measure the phase of the signal from each antenna over a 0.5-second dwell time. The integer number of cycles between the phase measurements needed to be redetermined at every measurement because the equipment did not continuously integrate the phase measurements. The phase reconstruction was successfully handled in the data processing software. Experience with the Trimble HAT receiver has pointed the way toward design improvements that would achieve the greatest payback for the real-time attitude determination problem. The test on USS Yorktown demonstrated the capabilities that GPS offers for determining position and attitude on a low to medium dynamic vehicle such as a ship at sea.

Real-Time Attitude Determination

TI has designed a dual-role capability into the new AN/PSN-9 receiver, which is in production for the U.S. Air Force Joint Program Office. It can operate as a normal navigation unit, or, when modified with a second antenna and additional components, it can become a pointing system.⁸ The modified AN/PSN-9 obtains phase difference measurements directly from up to five satellites. Also, once the initial integer cycle phase differences have been resolved and as long as the receiver continuously tracks each satellite, range differences between all satellites and antennas are available at each observation interval. As part of an internal Independent Exploratory Development project at NAVSWC, the modified AN/PSN-9 was tested in November 1989 and July 1990.

For the static portion of the test, two antennas were placed on top of a beam at a nominal separation distance of 1.0 meter. The beam, shown in Figure 5, was precisely adjusted with a theodolite to coincide with the baseline determined by the astronomic azimuth test described in the Longer Baseline Static Tests section. The alignment error between the beam and the reference azimuth was thought to be on the order of a few arcseconds.

Two antenna types (microstrip patch and drooped turnstile) and three ground plane



Figure 5. Two drooped turnstile antennas (without ground planes) attached to a beam positioned over a precisely known azimuth baseline.

structures (flat plate, choke ring, and no-ground plane) provide six test configurations. The microstrip patch antenna was a Sensor System Model S67-1575-2 airborne L1 GPS antenna. The drooped turnstile antenna was a Dorne and Margolin DM C146 Series L1 GPS antenna packaged for the AN/PSN-9 GPS receiver (see Figures 5 through 8 for more details).⁷

For the static test, the microstrip patch antennas on the choke ring ground planes produce the most consistently accurate results. For the two days of testing, the azimuth errors were -0.02 and -0.04 degree; the elevation errors were 0.07 and -0.05 degree. The corresponding standard deviations for azimuth were 0.06 and 0.08 degree and for elevation were 0.18 and 0.23 degree. The differences in azimuth and elevation accuracies are due to the geometrical differences of the satellite locations. These geometric dilution-of-precision corrections were calculated and explain the test result differences.

For the dynamic tests, Figure 6 illustrates how the beam was attached across the back of a Light Armored Vehicle (LAV) with a 1-meter antenna separation. The surveyed values of the beam azimuth that were taken when the LAV was stopped and the LAV navigation gyro measurements were used to establish a pointing direction reference for the test. Figure 7 shows the theodolite and pointer configuration for obtaining the survey azimuth values; here, the theodolite measures the angle with respect to a second theodolite over a known marker and baseline. Also shown in Figure 7 are the two preamplifier/down-conversion modules that were strapped to the beam to the left of the

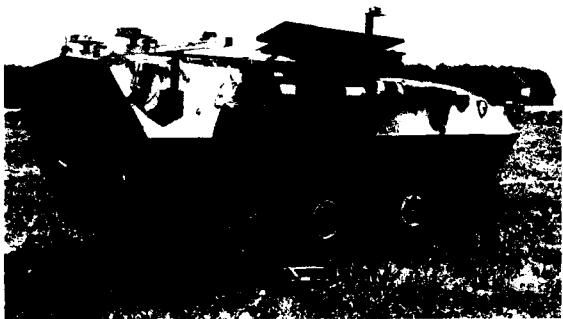


Figure 6. Light armored vehicle with antenna beam attached.



Figure 7. Antenna beam with survey equipment attached. Antennas are microstrip patches at the center of the choke ring ground planes.

antenna/choke ring assembly. Figure 8 shows the ground plane configuration with the microstrip patch antenna; however, the antenna cannot be seen in the photograph.

The LAV was equipped with a Spectra Systems Teldix FOA25 dead-reckoning navigation system²⁴ that includes an odometer, a single-axis gyro, and a map display position indicator. The position and azimuth of the vehicle were transferred in digital format to the map display; for the test, these values also were input to a portable digital computer. Also input to the computer were time-tagged pulses from a TI4100 GPS receiver. The computer software then selected and time-tagged the position and azimuth from the navigation system.

The dynamic test consisted of driving the vehicle over a closed-loop ground track; Figure 9 is a plot of the Teldix dead-reckoning navigation system position data. After an initial surveyed azimuth measurement, the LAV completed two loops of the ground track and stopped for another surveyed azimuth measurement at the HERO site located near Hangar 1 at NAVSWC, Dahlgren, Virginia. Two loops of the ground track took about 10 minutes. Approximately 1 minute after the vehicle stopped, the Teldix FOA25 navigation system detected that the vehicle was no longer in motion and produced constant azimuth values.

The reference data for the dynamic test consisted of azimuth values of the single-axis gyro with a linear correction for bias and drift errors to match the initial and final surveyed azimuth values for each run. This matching of the surveyed angles corrects for the approximately 90-degree rotation of the vehicle heading to the direction of the beam.

A comparison of the microstrip antenna and the flat plate ground plane configuration with

the corrected azimuth of the single axis gyro is given in Figure 10. On the scale used to show the full azimuth change, the LAV reference and GPS pointing results are indistinguishable except for an initial 22-second period that represents the phase ambiguity solution time using the dynamic algorithm.⁸ The beam azimuth plot shown in Figure 10 corresponds to the LAV ground track plot shown in Figure 9. When the LAV is moving east in Figure 10, the beam has a direction of almost zero degree (just after 4:48 for the first loop and 5:03 for the second loop) because the beam is mounted approximately 90 degrees counterclockwise with respect to the LAV navigation system axis. The vehicle is moving west at about 5:00 and 5:05 in the figure. Note the mirrored image of the measured azimuth during the parallel ground track before and after these times.



Figure 8. Antenna beam with microstrip antennas and flat plate ground planes.

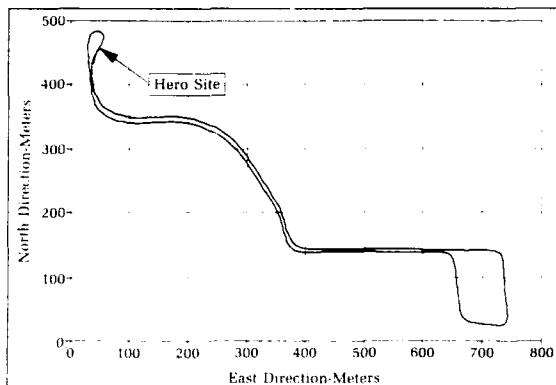


Figure 9. Dynamic test LAV ground track.

Figure 11 shows the GPS azimuth pointing error relative to the corrected LAV navigation truth data; the error has a 0.04-degree mean value and a 0.52-degree standard deviation. Since the large error "spikes" of Figure 11 occurred during maneuvers, they are thought to be experimental errors due to imprecise time alignment of the two measurements. The error standard deviation without these spikes is about 0.2 degree.

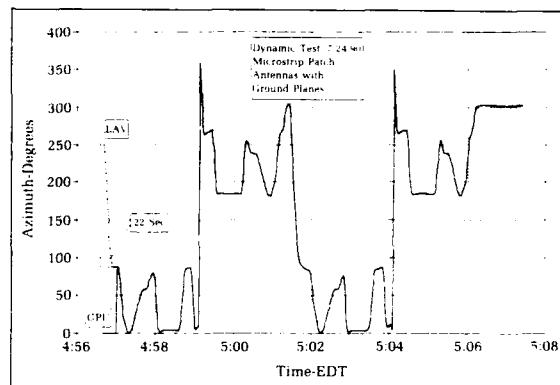


Figure 10. Example of Az pointing solution during dynamic test.

simulated a receiver, with a single antenna and pseudo-range only, on board a carrier-based aircraft. A full 44-state error model was developed, and reduced state suboptimal filters were shown to perform nearly optimally. Since GPS pseudo-range measurements do not directly update platform alignment errors, it is through the filter correlation of attitude errors with position and velocity errors that alignment can be improved. The study also showed that certain flight maneuvers (e.g., flying a horizontal loop) could enhance the correlation through sensed acceleration, and thus aid in reducing attitude errors.

A new covariance study was recently performed by upgrading the original covariance model to include measurements of smoothed pseudo-range or integrated doppler from a single antenna and by modeling phase differenced measurements from multiple GPS receiver antennas.¹⁹ The doppler data may be used to supplement the pseudo-range data, or the pseudo-range data can be sequentially smoothed using continuous doppler data. The revised 23-state error model, which includes states for position, velocity, attitude, receiver clock, and accelerometer and gyro biases and drift, was used to study the effects of the various measurement types for a collection of generic user vehicles. For a dynamic platform, the use of doppler data, even with a single antenna, can significantly reduce attitude errors. This is accomplished by exploiting the system error state correlations between velocity and platform attitude. The use of phase difference measurements from two or three antennas (with nominal 1-meter separation) can significantly reduce attitude errors even for the static case.

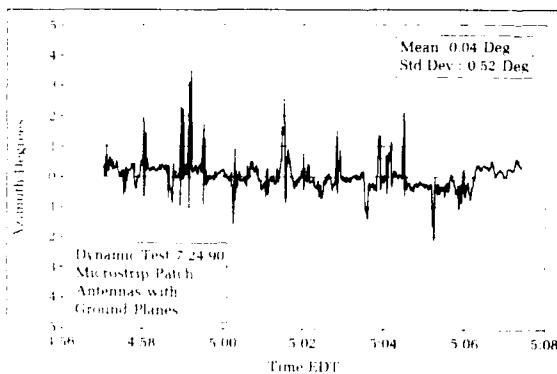


Figure 11. Example of Az pointing error during dynamic test.

164 Covariance Analyses of GPS/INS

Two years before the launch of the first GPS satellite, NAVSWC performed a study of the accuracy enhancements to be gained from combining GPS with an INS.⁹ This early study

Table 3 presents the attitude errors from the covariance analysis for each of four vehicle simulation scenarios of 1-hour duration each. The first scenario is a static case. The second simulation scenario is for a ground vehicle with stop-and-go motion (0 to 60 mph) and occasional right turns; the third case is similar except for a 360-degree turnaround within the first few minutes and later S-turns to demonstrate the correlation enhancement of antenna motion. The fourth scenario is an aircraft taking off, flying a wide out-and-back loop, and landing. All of the cases in Table 3 are for an INS with 1 degree per hour drift. The attitude errors are expressed with respect to the vehicle azimuth and the local vertical. Results are given for phase measurement data at the 1-centimeter and 1-millimeter standard deviation levels; validation tests indicate current capability for phase measurements at this level, if not in a high-signal multipath environment.

The covariance simulations demonstrate the mutual enhancements that can be obtained by combining GPS with inertial navigation. The introduction of GPS observations from a single

antenna significantly reduces the errors in the north and east directions (similar to roll and pitch) compared to those expected from the INS alone, but adds only moderate improvements in azimuth. The addition of the phase difference data class, derived from multiple GPS antennas, significantly reduces the azimuth error.

Summary

Real-time GPS attitude determination has become a practical reality. There are three significant factors that have led to this capability. First, receivers now have increased accuracy for measuring coherent phase or phase difference. In 1984, tests of TI4100 receivers configured to eliminate all sources of noise except those introduced by the receivers themselves, resulted in a measurement accuracy of about 2 millimeters.²⁵ On receivers available today, similar experiments have shown better than an order of magnitude improvement.^{8,26} Second, antennas have improved. Tests at NAVSWC of various antennas and ground plane configurations have demonstrated

Table 3. Attitude Errors from GPS/INS Covariance Analysis with only an Average Quality Gyro (1 deg/hr)*

	Static	Plain Dynamic	Dynamic with Loop	Aircraft
Gyro (no GPS data)	2972/1276	2972/1276	2972/1277	2972/1276
One Antenna				
Pseudo-range data (1 m)	258/1014	258/1001	258/931	259/921
Pseudo-range and Doppler data (1-sec rate-1cm)	162/1010	162/965	162/666	163/801
Multiple Antenna for Single Difference Phase Data (1-m separation)				
All 3 data types (1-cm phase data)	162/447	162/446	162/416	162/437
All 3 data types (1-mm phase data)	162/124	162/125	162/124	162/123

*Results are average RSS level (N and E) and azimuth attitude errors expressed in arcseconds.

significant improvements for the reduction of signal multipath.²⁷ However, signal multipath and electrical phase center variation of the antennas remain the largest error sources for GPS attitude determination. Third, there have been a number of contributions to resolve quickly the integer number of cycles between the phase measurements.^{8,28,29,30,31} Tests of the TI AN/PSN-9, and the TI method of resolving the integer ambiguity described above,⁸ demonstrated that the ability to obtain the proper integer value is very dependent on vehicle dynamics. For example, for a 1-meter antenna separation and a 1-second measurement interval, 5 minutes of data collection were required for the static case. For significant dynamic variation, only 13 seconds were required for integer resolution. A four-antenna array has reduced the integer cycle resolution time to under 5 seconds for the Ashtech 3DF receiver system when five satellites are in view.³¹

In summary, GPS attitude determination has progressed significantly since its initial demonstration. Because the satellite ephemeris errors, satellite clock errors, and propagation effects (such as ionospheric refraction errors) are highly correlated over such short baselines, they do not degrade the GPS attitude determination accuracy. The additional attitude antennas also offer the potential for antenna null steering in a jamming environment.³² Further, multichannel-on-a-chip receiver designs and less costly microstrip antennas, with characteristics that are repeatable unit-to-unit, make the additional hardware required for attitude determination economically feasible. These enhanced capabilities of GPS, especially when combined with an INS, should be considered for all future applications that require attitude determination, including ships at sea, platforms on the ground or in the air, and satellites in space.

Acknowledgments

In a review paper such as this, many topics are touched upon and the authors are indebted to many people. It would be difficult to recognize every contributor in this single acknowledgment. Please refer to the literature cited and the references and acknowledgments therein.

We do, however, want to recognize two people who were instrumental in providing the initiative and guidance for developing a GPS capability in the Space and Surface Systems Division at NAVSWC. They are Richard J. Anderle, currently with General Electric

Company in King of Prussia, Pennsylvania, and Robert W. Hill, Director of the NAVSWC Office of Space Technology.

References

- Evans, Alan G., "Roll, Pitch and Yaw Determination Using a Global Positioning System Receiver and an Antenna Periodically Moving in a Plane," *Marine Geodesy*, Vol. 10, No. 1, 1986.
- Evans, A. G., Hermann, B. R., Fell, P. J., "Global Positioning System Sensitivity Experiment," *Navigation*, Vol. 28, No. 2, Summer 1981.
- Evans, Alan G., "The Global Positioning System: An Alternative to Six Degree-of-Freedom Inertial Navigation," *International Federation of Surveyors Meeting on Inertial, Doppler and GPS Measurements for National and Engineering Surveys*, Munich, Federal Republic of Germany, Jul 1985.
- Evans, A. G., et al., "Demonstration of the Use of GPS to Determine Astronomic Azimuth and Vertical Deflections: Joint Test Description and Results," *Fifth International Geodetic Symposium on Satellite Positioning*, Las Cruces, NM, Mar 1989.
- Kruczynski, Leonard R., et al., "Using GPS to Determine Attitude," *First International Technical Meeting of the Institute of Navigation's Satellite Division: GPS-88*, Colorado Springs, CO, Sep 1988.
- Kruczynski, Leonard R., et al., "Using GPS to Determine Attitude: USS Yorktown Test Results," *Processing of ION GPS-89, The Second International Technical Meeting of the Satellite Division of the Institute of Navigation*, Colorado Springs, CO, Sep 1989.
- Brown, Ronald A. and Evans, Alan G., "GPS Pointing System Performance," *Proceedings of ION GPS-90, The Third International Technical Meeting of the Satellite Division of the Institute of Navigation*, Colorado Springs, CO, Sep 1990.
- Brown, Ronald A. and Ward, Phil, "A GPS Receiver with Built-In Precision Pointing Capability," *IEEE Position Location and Navigation Symposium (PLANS)*, Mar 1990.
- Miller, B. Larry, *Use of the NAVSTAR Global Positioning System to Update the Inertial Navigation System of Carrier-Based Aircraft*, NSWC TR-3437, Mar 1976.
- Joseph, Keith M. and Deem, Paul S., "Precision Orientation: A New GPS Application," *International Telemetering Conference*, San Diego, CA, Oct 1983.
- Maki, Stanley C., "The Global Positioning System: A Boost to Space Users," *Proceedings of ION GPS-89, The Second International Technical Meeting of the Satellite Division of the Institute of Navigation*, Colorado Springs, CO, Sep 1989.
- Spinney, V.W., "Applications of the Global Positioning System as an Attitude Reference for Near Earth Uses," *Institute of Navigation Aerospace Meeting*, Warminster, PA, Apr 1976.
- Ellis, J. F. and Creswell, G. A., "Interferometric Attitude Determination with Global Positioning System," *J. Guidance and Control*, Vol. 2, No. 6, Nov 1979.
- Brown, Allison K. and Bowles, W. M., "Interferometric Attitude Determination Using the Global Positioning System," *Third Geodetic Symposium on Satellite Doppler Positioning*, Las Cruces, NM, Feb 1982.

15. Ward, Phil W., "An Advanced NAVSTAR GPS Geodetic Receiver," *Third Geodetic Symposium on Satellite Doppler Positioning*, Las Cruces, NM, Feb 1982.
16. Darnell, A. Ruth and Meyerhoff, Stan I., "Specification and Formulation for the GESAR Version 1.6 Software," *NSWC Internal Report*, 10 Nov 1987.
17. Evans, Alan G., "Comparison of Pseudorange and Biased Doppler Range Measurements to Demonstrate Signal Multipath Effects," *Proceeding of the Fourth International Geodetic Symposium on Satellite Positioning*, Austin, TX, Apr 1986.
18. Georgiadon, Yola and Kleusberg, Alfred, "Multipath Effects in Static and Kinematic GPS Surveying," *International Association of Geodesy Meeting*, Edinburgh, Scotland, Apr 1989.
19. Evans, A. G., et al., "Incorporation of GPS Information with Inertial Navigation," *Independent Research and Independent Exploratory Development 1990*, Naval Surface Warfare Center, Dahlgren, VA.
20. Nesbo Inge, "Applications of GPS Determined Attitude for Navigation," *First International Technical Meeting of the Institute of Navigation's Satellite Division: GPS-88*, Colorado Springs, CO, Sep 1988.
21. Purcell, George H., Jr., et al., "Measurement of Aircraft Position, Velocity, and Attitude Using Rogue GPS Receivers," *Fifth International Geodetic Symposium on Satellite Positioning*, Las Cruces, NM, Mar 1989.
22. Goad, Clyde C., "Precise Relative Position Determination Using Global Positioning System Carrier Phase Measurements in a Non-Difference Mode," *First International Symposium on Precise Positioning with the Global Positioning System*, National Geodetic Information Center, NOAA, Rockville, MD, Apr 1985.
23. Swift, Everett R., "NSWC's GPS Orbit/Clock Determination System," *First International Symposium on Precise Positioning with the Global Positioning System*, National Geodetic Information Center, NOAA, Rockville, MD, Apr 1985.
24. Teldix FOA25 Vehicle Orientation System Brochure, Spectra Systems, 600 Pine Island Road, Suite 175, Plantation, FL.
25. Evans, A. G., Hermann, B. R., Coco, D. S., and Lynch, J. R., "Collection Tests of an Advanced Geodetic Global Positioning System Receiver," *First International Symposium on Precise Positioning with the Global Positioning System*, National Geodetic Information Center, NOAA, Rockville, MD, Apr 1985.
26. Ashtech, XII GPS Receiver Brochure, Ashtech Inc., 1156 Aster Avenue, Sunnyvale, CA 94086.
27. Evans, Alan G. and Hermann, Bruce R., "A Comparison of Several Techniques to Reduce Signal Multipath from the Global Positioning System," *International Association of Geodesy Meeting*, Edinburgh, Scotland, Apr 1989.
28. Kruczynski, Leonard R., *GPS Heading and Attitude Determining Instrument Investigation and Prototype Development, Final Report*, Trimble Navigation, Ltd., Sunnyvale, CA, Dec 1987.
29. Hatch, Ronald, "Ambiguity Resolution in the Fast Lane," *Proceedings of ION GPS-89, The Second International Technical Meeting of the Satellite Division of the Institute of Navigation*, Colorado Springs, CO, Sep 1989.
30. Brown, R. Grover and Hwang, P. Y. C., "A Kalman Filter Approach to Precision GPS Geodesy," *Navigation*, Winter 1986-87, Vol. 30, No. 4.
31. Ferguson, Kendall, et al., "Three-Dimensional Attitude Determination with the Ashtech 3DF 24-channel GPS Measurement System," *Institute of Navigation National Technical Meeting*, Phoenix, AZ, Jan 1991.
32. Evans, A. G. and Hermann, B. R., "GPS Receiver Processing for Multiple Antenna Gain Control and Attitude Determination," *NAVSWC Independent Exploratory Development 1990 Final Report* (to be published).

The Authors



ALAN G. EVANS has been working in Global Positioning System (GPS) applications at NAVSWC since 1981. He received a B.S.E.E. degree from Widener University in 1964 and M.S. and Ph.D. degrees in electrical engineering from Drexel University in 1967 and 1972, respectively. In the area of GPS, Dr. Evans has contributed to static and dynamic relative positioning and orientation, signal multipath analysis, the determination of geodetic quantities, and receiver development.



BRUCE R. HERMANN received a B.S.E.E. from Bradley University in 1965, an M.S. in 1966 from Colorado State University, and a Ph.D. in electrical engineering in 1972 from the University of Illinois. His Ph.D. dissertation was in the area of radio astronomy. From 1972 to the present, he has been employed at NAVSWC in the Space Sciences Branch. He has been working on various aspects of the GPS since 1978.



B. LARRY MILLER received a B.A. degree from Western Kentucky University in 1967 and an M.A. degree in mathematics from the University of Louisville in 1971. He was a staff meteorologist at Vandenberg AFB until 1972, and has since been employed at NAVSWC in the Space Sciences Branch. Since 1974, he has worked on numerous aspects of the GPS, including satellite constellation design, concept evaluation of ground and space applications, and development of the first spaceborne GPS navigational unit.